CHARACTERISATION OF HEPATITIS B VIRUS DNA INTEGRANTS IN LIVER OF SOUTHERN AFRICAN BLACKS WITH HEPATOCELLULAR CARCINOMA.

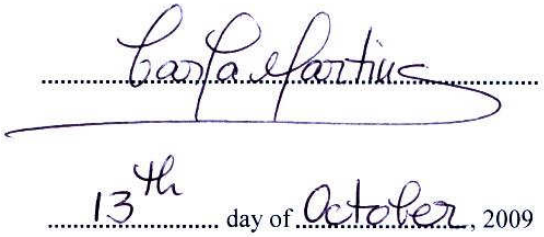
Carla Suzana Pinto Martins-Furness

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfillment of the requirements for the degree of Doctor of Philosophy

Johannesburg, 2009

DECLARATION

I, Carla Suzana Pinto Martins-Furness declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.



13th day of October, 2009

DEDICATION

In memory of my earthly father:

Manuel Antonio Martins
15.10.1936
12.08.2007

and to the glory of my heavenly Father,
in whose Son I can do all things.

PRESENTATIONS

A powerpoint presentation:

CHARACTERISATION OF HEPATITIS B VIRUS DNA INTEGRANTS IN LIVER OF SOUTHERN AFRICAN BLACKS WITH HEPATOCELLULAR CARCINOMA.

Was presented at:

Symposium: "*New insights in HBV diversity, pathogenesis, diagnosis and treatment*"
12-14 December 2007. Ghent University, Het Pand, Onderbergen 1, Gent, Belgium. Hosted by: Professor A. Kramvis, Department of Medicine, Hepatitis Virus Diversity Research Program, University of the Witwatersrand, South Africa and Professor G. Leroux-Roels, Center for Vaccinology, Department of Clinical Biology, Microbiology and Immunology, Ghent University Hospital, Ghent, Belgium.

International Symposium: "*Hepatitis B virus genotypes ~ from an academic question to the clinic*"
30 July – 1 August 2008. Origins Centre, University of the Witwatersrand, Johannesburg, South Africa. Hosted by: Professor A. Kramvis, Department of Medicine, Hepatitis Virus Diversity Research Program, University of the Witwatersrand, South Africa and Professors M. Mizokami and Y. Tanaka, Department of Clinical Molecular Informative Medicine, University of Nagoya, Nagoya, Japan.

ABSTRACT

Hepatitis B virus (HBV) is hyperendemic in sub-Saharan Africa with a correspondingly high incidence of hepatocellular carcinoma (HCC). In 80 – 90 % of the HCC cases, HBV is found integrated, contributing to HCC initiation and progression through direct mutation of host DNA, and expression of truncated viral, and hybrid viral-host proteins; which may cause increased cell division, apoptosis, aberrations in growth cycle and proliferation, and transformation. This study aimed to determine the proportion of tumours with integration of HBV and whether this was single or multiple; to ascertain the sites of HBV integration relative to functional cellular genes and to establish whether any integrants were expressed.

The tumour (T) and non-tumorous (NT) tissues studied were from 18 (+3) black southern African males with HCC, mainly miners from South Africa and surrounding countries. DNA was analysed by Southern hybridisation (SH) with single and double digests of restriction endonucleases with different numbers of target sites on the HBV genome (*Eco*RI, *Bam*HI, and *Hind*III), and at different DNA concentrations, using a full genome HBV probe. The full length (FL)-PCR, inverse (I)-PCR and *Alu*-PCR techniques were attempted because PCR enables the characterisation of large numbers of HBV integrants to be processed less laboriously, rapidly, cheaply and with the use of less tissue than library construction. In order to study the expression of integrants in an African HCC patient, we designed and used a novel adaptation of the differential gene expression technique. RNA was extracted from the HCC tissue, reverse transcribed, amplified by PCR with ^{35}SdATP, and oligo d$(T)_{15}$/HBV-specific primers. Amplicons were resolved by polyacrylamide gel electrophoresis (PAGE), purified individually, re-amplified and sequenced.

SH revealed multiple integrants in 9/11 individuals (82 %) with 10 being the largest number of distinct clonally expanded integrants in any one HCC. In the only matched T/NT pair studied, both tissues contained integrants, with the T having more distinct clonal expansions. In 3 HCCs, DNA from different areas of the tumour did not yield the same pattern of integration upon subsequent analysis. The PLC/PRF/5 cell line was found to contain 6 or possibly 9 integrants, in accordance with previous publications. No HBV integrants were amplified by either I-PCR, *Alu*-PCR or FL-PCR. With the use of the novel adaptation of the differential gene expression technique we report, for the first time, the possible expression of an integrant from an African HCC. A 396 bp rearranged, hybrid mRNA, constituting 322 bp of HBV sequence comprising the end portion of the *X* gene, DR1 (a hot-spot for integration), and the beginning of the core gene (nucleotides 1820 to 2142), was flanked by 74 bp of DNA sequence which appeared rearranged, comprising 22bp of sequence that could not be reliably mapped, 33 bp of possible *Homo sapiens VH1* region (AJ007320) and 19 bp of

adjacent sequence. When the sequence was submitted to Genbank at a later date it was revealed that the region of the AJ007320 *VH1* sequence to which it originally mapped appeared to derive from the plasmid pPCR-Script Amp SK(+). A number of considerations make it unlikely that the flanking DNA of this possible integrant was from a HBV DNA cloned into pPCR-Script Amp SK(+), but the possibility cannot be entirely eliminated. The possibility of the integrant being legitimate can also not be excluded as portions or all of the flanking sequence may be a part of the human genome that may not have been banked with GenBank yet. Furthermore, the T45.3 HBV sequence was found to cluster with the HBV genotype E isolates, found primarily in the populations of sub-Saharan Africa, and the T45.3 subject was a Mozambican male who had migrated to South Africa. There was no genotype E clone in the laboratory at the time. The possibility exists that more of the HBV genome could be present upstream to the genomic DNA flank of this integrant. Specific cases of rearrangement of *X* gene integrants, with 2 sections (or more) of HBV DNA separated by genomic DNA have been reported in the literature. The *X* gene is the most frequently integrated region of the viral genome.

The sequence obtained for this integrant was in a well conserved region of the HBV genome and upon phylogenetic analysis of the integrant was found to cluster with the HBV genotype E isolates relative to a representative selection of sequences of HBV isolates from the international databases. The integrant was derived from a sample that originated from a Mozambican male who had migrated to South Africa. Unpublished data indicate that 15.7% of HBV isolates from Mozambique belong to genotype E. Of all the HBV genotypes, E is the most prevalent worldwide, found primarily in the populations of sub-Saharan Africa – an estimated 20% of HBV chronic carriers or approximately 60 million people, and may be the most important genotype globally. This is the first description of a possible genotype E integrant and its expression.

In conclusion, SH results corresponded with published literature in that 80 % of samples had integration and all had multiple integrants. SH results also indicated that 3 HCCs contained more than one hepatocyte clone. We showed that the number of integrants detected in the PLC/PRF/5 cell line varies between studies. I-PCR FL and *Alu*-PCR methods were shown to have shortcomings, and not to be easily reproducible. These methods have been employed to analyse mainly non-African HCCs, primarily containing single integrants. The presence of multiple integrants per sample makes characterisation of such integrants more complicated. It is clear from the present study that methods developed to study single integrants in HCC tissue are not suited to study multiple integrants, which are found in HCC tissue from Africans.

ACKNOWLEDGEMENTS

I thank and acknowledge the following people:

For their input to my life and thesis:

My parents, who always believed in me, supported my dreams, and always encouraged me to be the best I can be. I could not have asked for better. Your love has shaped my life.

My husband, who has suffered in silence these many years as I diligently worked on this project while ignoring him. You support, patience, love and encouragement have meant the world to me.

For their input to my thesis:

Professors MC Kew and A Kramvis for their assistance and support, financial, theoretical and technical. A special vote of thanks goes to Professor Kramvis for her invaluable advice.

All my fellow students, and the staff of the Molecular Hepatology Research Unit, especially Caroline who has been an excellent sounding board and friend.

Professor H. Will for his kind donation of a head-to-tail dimer of the complete genome of HBV genotype D (subtype *ayw*, GenBank accession number V01460).

Professors J. Alexander; H. Nakabayashi for their kind donations of the PLC/PRF/5 and Huh7 cell lines, respectively.

I acknowledge the following funding bodies:

The Poliomyelitis Research Foundation
National Research Foundation (NRF)
Wits University Research Committee
HE Griffin Charitable Trust
Cancer Trust
The Molecular Hepatology Research Unit and the University of the Witwatersrand, where the research was performed.

The Bilaterale (internationale) Wetenschappelijke en Technologische Samenwerking (BTWS)-South African NRF bilateral agreement, for funding my trip to Belgium to present at the joint symposium on "*New insights in HBV diversity, pathogenesis, diagnosis and treatment*".

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A2BP1 Ataxin 2-binding protein 1 gene

ANKH Ankylosis gene

Anti-HBc Antibody to the HBV core antigen

Anti-HBs Antibody to the HBV S antigen

APC Adenomatosis polyposis coli

APCL Adenomatous polyposis-coli like gene

AXIN1 Axis inhibitor 1 gene

BBX Bobby sox gene

BCK Blood and cell DNA extraction kit

BCL2L2 B-cell-lymphoma-like 2 gene

BCP HBV basic core promoter region

BIRC3 Baculoviral IAP repeat-containing 3 gene

CA California

CASPR3 Cell recognition molecule contacin-associated protein gene

cccDNA Covalently closed circular DNA

CCNA2 Cyclin A2 gene

CCT Chaperonin-containing TCP1, subunit 8 (theta) gene

CDC42EP5 CDC42 effector protein (Rho GTPase-binding) 5 gene

CDK4 Cyclin dependant kinase 4

cDNA Copy DNA

c-*fms* Macrophage colony-stimulating factor gene

c-*fos* Cellular homologue of the v-fos oncogene isolated from FBJ-MSV and FBR-

ChHBV Chimpanzee hepatitis virus

CHML Choroideraemia-like (Rab escort protein 2) gene

Coh Cohesive end region

CTLs Cytotoxic T-lymphocytes

CTNND2 Catenin delta-2 gene

DD-PCR Differential display PCR

DEPC Diethyl pyrocarbonate

DHBV Duck HBV

DIAPH3 Diaphonous homolog 3 gene

DMEM Dulbecco's Modified Eagle's Medium

DNAJB6 DnaJ (Hsp40) homolog, subfamily B, member 6 gene

DR1 HBV direct repeat one

DR2 HBV direct repeat two

DR(s) Direct repeat(s)

EDTA Ethylenediamine tetra acetic acid

EMEM Eagles Minimum Essential Medium

EMX2 gene homologue of the *Drosophila empty spiracles* (*ems*) head gene

ER Endoplasmic reticulum

EtBr Ethidium bromide

EtOH Ethanol

EVER2 Epidermodysplasia verruciformis 2 gene

EYA3 Eyes absent 3 gene

FALZ Bromodomain PHD finger transcription factor gene

FBS Foetal Bovine Serum

FDFT1 Farnesyl-diphosphate farnesyltransferase 1 gene

FL Full length (in reference to the HBV genome)

FL-PCR Full length polymerase chain reaction (in reference to the HBV genome)

FLT3 FMS-related tyrosine kinase 3 gene

FN1 Fibronectin 1 gene

FOXF2 Forkhead box F2

GA Liver mitochondrial glutaminase/ breast cell glutaminase gene

GAPDH Glyceraldehyde-3-phosphate-dehydrogenase

GCHFR GTP cyclohydrolase 1 feedback regulatory protein gene

GiHBV Gibbon hepatitis virus

GRE Glucocorticoid-receptor element

GRID2 Glutamate-receptor, ionotrophic, delta 2 gene

GSHV Ground squirrel hepatitis virus

HBcAg HBV core antigen

HBeAg HBV e antigen

HBsAg HBV surface antigen

HBV Hepatitis B virus

HBx HBV X protein

HCC Hepatocellular carcinoma

hTERT Human telomerase reverse transcriptase gene

IGF2R Insulin-like growth factor 2 receptor

IGF2 Insulin-like growth factor 2 gene

IP3R2 Inositol 1,4,5-triphosphate receptor type 2 gene

I-PCR Inverse-PCR

IRAK2 IL-IR-associated kinase 2

IRF2 Interferon regulatory factor 2 gene

ITP3R1 Inositol 1,4,5-triphosphate receptor type 1 gene

ITPKB Inositol 1,4,5-triphosphate-3 kinase B gene

JNK Jun-activated kinase/stress-activated protein kinase pathway

KCNB2 Potassium voltage-gated channel, Shab-related subfamily, member 2 gene

KIF3C Kinesin family member 3C gene

KLF13 Kruppel-like factor 13 gene

LAIR2 Leukocyte-associated immunoglobulin-like receptor 2 gene

LHBs Large HBV surface protein

LHFPL3 Lipoma HMGIC fusion partner-like 3 gene

LOH Loss of heterozygosity

MACF1 Microtubule-actin crosslinking factor 1 gene

MCM8 Minichromosome maintenance protein-related gene gene

MEM Minimal Essential Medium

MGMT O-6-methylguanine-DNA methytransferase gene

MHBs Middle HBV surface protein

MHBs^{t} Truncated middle HBV antigen polypeptide

MIG6 Mitogen-inducible gene 6

MK Mevalonate kinase gene

***MLL2* (*MLL4*)** Myeloid/lymphoid or mixed lineage leukaemia 2 gene

MLL5 Myeloid/lymphoid or mixed lineage leukaemia 5 gene

MOPS 4-Morpholinepropanesulfonic acid

MUC16 Mucin 16 gene

MWM(s) Molecular weight marker(s)

NA Not applicable

NCF1 Neutrophil cytosolic factor 1 gene

NEDD4L Ubiquitin-protein ligase NEDD4L-like; neural precursor cell expressed, developmentally downregulated gene 4-like gene

NER Nucleotide excision DNA repair

NL 'Non-HBV' liver

NMP Nuclear matrix protein p84 gene

NT non-tumorous tissue

NTRK2 Neurotropic tyrosin receptor kinase 2 gene

OCIA OCIA domain-containing 1 gene

ODZ2 Homolog of odd Oz 2 *Drosophila* gene

ORF(s) Open reading frame(s)

ORM1 Orosomucoid 1 gene

OuHV Orangutan hepatitis virus

P/C Phenol chloroform DNA extraction method

p42MAPK1 p42 Mitogen-activated protein kinase 1 gene

PAGE Polyacrylamide gel electrophoresis

PARK7 Parkinson disease gene

PBMCs Peripheral blood mononuclear cells

PCR Polymerase chain reaction

PDE11A Phosphodiesterase 11A gene

PDE6D Phosphodiesterase 6D, cGMP-specific, rad, delta gene

PDGFB Platelet derived growth factor beta polypeptide gene

PDGFRB Platelet derived growth factor receptor beta precursor gene

PDGR Platelet derived growth factor gene

pgRNA Pregenomic RNA

PI-3-K Phosphatidylinositol-3-kinase

PITPNC1 Posphatidylinositol transfer protein, cytoplasmic 1 gene

PRE Postranscriptional regulatory element

RAB30 Ras oncogene family member 30

RAI17 Retinoic acid-induced gene 17

RAR-β Retinoic acid receptor beta

RB1 Retinoblastoma 1 gene

RBMY RNA-binding motif Y chromosome gene

RE/s Restriction enzyme/s

RSA Republic of South Africa

RT Reverse transcription

RT-PCR Reverse transcription PCR

S/O Salting out DNA extraction method

SCARA3 Scavenger receptor class A, member 3 gene

SDS Sodium dodecyl sulphate

SERCA1 Sarco-endoplasmic reticulum calcium ATPase gene

SEST3 Sestrin 3 gene

SFXN5 Sideroflexin 5 gene

SH Southern hybridization

SHBs Small HBV surface protein

SITA Alpha 2,3 sialyltransferase gene

SMAD2 Human homologue 2 of the Drosophila Mad gene

SMAD4 Human homologue 4 of the Drosophila Mad gene

SMOC1 SPARC-related modular calcium-binding 1 gene

SOCS4 Suppressor of cytokine-signaling 4 gene

SPG4 Spastic paraplegia 4 gene

SRP46 Splicing factor, arginine/serine rich 2B gene

SRPK2 Serine protein kinase 2 gene

SSC Sodium chloride sodium citric acid

STAC SH3 & cystein rich domain gene

STX3A Syntaxin 3A gene

T Tumour tissue

TACC1 Transforming acedic coiled-coil-containing protein 1 gene

TCEA3 Transcription factor SII-related protein 4 gene

TE Tris-EDTA

TERF2 Telomeric repeat-binding factor 2 gene

TGF-β Transforming growth factor-beta

TIAM1 T-cell lymphoma invasion and metastasis 1 gene

Tm Annealing temperature

TNF Tumour necrosis factor-induced protein gene

TRAP Thyroid hormone receptor-associated protein-150 alpha gene

TRUP Thyroid hormone-uncoupling protein gene

TS Tumour suppressor

TTC30A Tetratricopeptide repeat domain 30A gene

UK United Kingdom

URF4 unidentified reading frame 4

USA United States of America

UV Ultra violet

VH1 *Homo sapiens* heavy chain variable region gene

VNTRs Variable nucleotide tandem repeats

WBSCR1 Williams-Beuren syndrome (WS) chromosome region 1 gene

WHO World Health Organisation

WHV Woodchuck hepatitis virus

WMHBV Woolly monkey hepatitis virus

WT1 Wilm's tumour gene

XPO4 Exportin 4 gene

YY1 Yin and Yan transcription initiation factor

ZNF521 Zinc finger protein 521 gene

ε HBV encapsidation signal or epsilon

Units

bp Base pairs

kb Kilobases

cm^2 Centimeters square

cpm Counts per minute

g Grams

µg Micrograms

mg Milligrams

µl Microlitres

ml Millilitres

µM Micromolar

mM Millimolar

pmol Picomole

x g Multiples of the acceleration due to gravity (g)

1.0 INTRODUCTION

1.1 Hepatitis and Hepatitis B Virus – a brief history

Hepatitis or liver inflammation is recognized in ancient times, where Mesopotamian healers record patients with yellowing of the skin and eyes. Transmission of the disease through contaminated blood is however, only recorded in 1885 (Lürman, 1885: cited in MacCallum, 1971 - original article in German) and in 1947, MacCullum distinguishes two forms of hepatitis, that he terms A, for infectious hepatitis, and B for homologous serum hepatitis (MacCallum and Bauer, 1947). In 1963 the Australia antigen, which is later shown to be associated with hepatitis, is discovered in the serum of an Australian aborigine (Blumberg *et al.*, 1964, 1965). This is the hepatitis B virus (HBV) surface antigen (HBsAg), which is subsequently shown to occur only in individuals infected with HBV (Okochi and Murakami, 1968; Prince 1968). However, it is not until 1970 that the complete HBV particle is isolated (Dane *et al.*, 1970). The production of the successful HBsAg vaccine against HBV is licensed in 1981 (Blumberg and Millman, 1972; ACIP, 1985).

1.2 HBV Infection and Hepatocellular Carcinoma

HBV infection which most often results in spontaneous recovery but may run a fulminant course, causes acute hepatitis, chronic hepatitis, and asymptomatic liver disease (Redeker, 1975; Lee 1997). Chronic infection with the virus is shown to correlate geographically with hepatocellular carcinoma (HCC) (Figure 1.1), and it is currently estimated that approximately 350 million individuals are infected worldwide (WHO estimates 2002). HCC is one of the most common cancers, comprising 85 % of malignant tumours of the liver (Hussain *et al*., 2001; Kao and Chen 2002; Zhu, 2003), and ranks fifth in men, eighth in women, and third in terms of annual mortality rate (Parkin *et al*., 2008). HBV DNA is found integrated in 80-100 % of all HCCs (Shafritz and Kew, 1981; Bréchot *et al*., 2000; Murakami *et al*., 2005). The reason for the variable outcome of HBV infection (Sommer *et al*., 1997) and the development of HCC (Szmuness, 1978) has not yet been satisfactorily explained, although it has been postulated that the host immune response, and HBV sequence variability, may play a role (Günther *et al*., 1998).

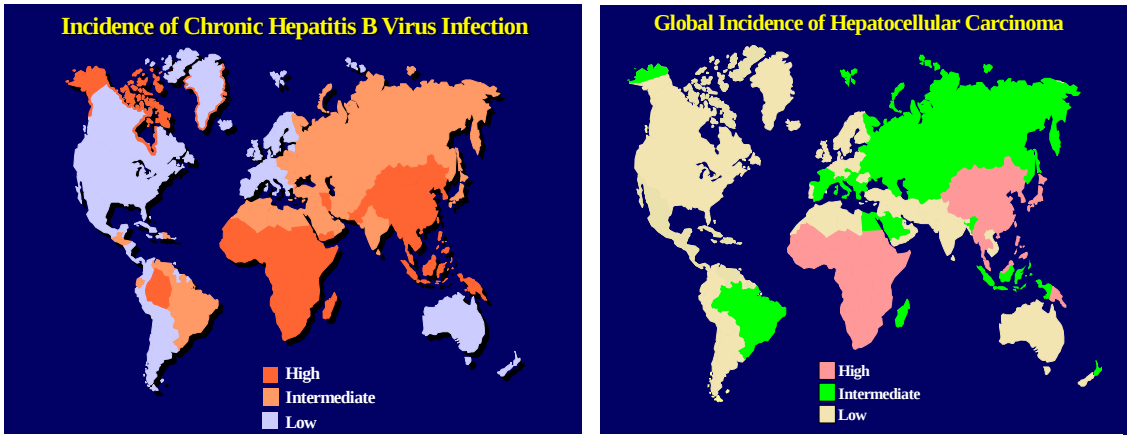


Figure 1.1: Geographical distribution of HCC and chronic HBV infection.

Slides reproduced from a presentation by kind permission of Professor M.C. Kew.

Transmission of HBV occurs primarily by percutaneous exposure to infected blood, or blood products, or semen (Szmuness *et al*., 1975; Alter *et al*., 1975, 1977; Bancroft *et al*., 1977). In areas of high endemicity (defined as such if 8 % or more of the population is chronically infected with the virus), the primary route of transmission is parenteral. The risk is high (70-90 %) and lower (10-40 %) when mothers are HBeAg positive (Beasley *et al*., 1977, 1981, 1982) and negative, respectively (Stevens *et al*., 1979; Xu *et al*., 1985). Infants infected perinatally have a 90 % chance of remaining chronically infected (Beasley and Hwang, 1983; Xu *et al*., 1985; Franks *et al*., 1989; Hurie *et al*., 1992). In sub-Saharan Africa the major route of infection is horizontal spread in childhood. Of these children 40-90 % will become chronically infected (Kew *et al*., 1987). The exact mode of such transmission remains unclear. A case/control study performed in South Africa revealed that HBsAg positive carriers have a 9-33 times increased risk of developing HCC (Kew, 2002a) than uninfected persons. Becoming a carrier early in life allows for the 20-30 year latency period, which exists between initial infection with HBV, and development of HCC (Matsubara and Tokino 1990). The cancer often occurs in several members of the same family, all of which are chronically infected with the virus (Tong *et al*., 1979).

Factors other than infection with HBV may also contribute to HCC. Alcohol-induced cirrhosis and the dietary intake of high levels of iron, ingested chemical carcinogens such as

aflatoxin, all increase the risk of developing the cancer (Edmondson and Steiner, 1954; Lustbader *et al*., 1983; Bassett *et al*., 1986; Bressac *et al*., 1991; Hsu *et al*., 1991; Kato *et al*., 1996; Smela *et al*., 2002; Kew and Asare, 2007). Hepatitis C virus also causes HCC and co-infection with the two viruses increases the risk of HCC development (Benvegnu *et al*., 1994; Ye, *et al*., 1994; Kirk *et al*., 2004).

Treatment of HCC is long-term, expensive, and mostly unsuccessful. The cancer may be delayed with the use of HBV-inhibiting drugs, but once these are stopped, HBV replication resumes in the majority of cases (Gong *et al*., 1999). Surgical resection of the tumour is not always possible, and tumour recurrence occurs in 46-83% of cases in 5 years and in 73-92% of cases by the tenth year (Someya *et al*., 2006). Recurrence in large HCCs occurs after transplantation in 0-54% of patients after 5 years (Ng *et al*., 2009). Currently the best defence against HBV infection is vaccination, which decreases the prevalence of chronic infection by 50-93 % in children (Chang *et al*., 1997; Ni *et al*., 2001; Tsebe *et al*., 2001; Goldstein *et al*., 2002; Ni *et al*., 2007). Unfortunately, because of logistical factors, high cost, and lack of education in most developing countries (which are also often those where the virus is endemic) vaccination is currently only implemented for roughly one third of the world's infants, and infection and death rates remain high (Van Herck and Van Damme, 2008). To exacerbate the problem, vaccination is unsuccessful in 3-4 % of individuals. This may be as a result of vaccine adjuvants and HBV viruses with genetic polymorphisms and mutations (Carman *et al*., 1990).

1.3 HBV

1.3.1 Classification

The family *Hepadnaviridae* comprises two genera: the *Avihepadnaviruses*, which infect birds, and the *Orthohepadnaviruse*s, which infect mammals. The human HBV is the prototype member of the family (Summers, 1981) and the other primate HBVs are: chimpanzee (ChHBV) (Vaudin *et al*., 1988) woolly monkey (WMHBV) (Lanford *et al*., 1998) orangutan (OuHV) (Warren *et al*., 1999) gibbon (GiHBV) (Norder *et al*., 1996; Lanford *et al*., 2000) and gorilla (Grëthe *et al*., 2000). Other orthohepadnaviruses include the woodchuck hepatitis virus (WHV) (Summers *et al*., 1978) and ground squirrel hepatitis virus (GSHV) (Marion *et al*., 1980a). The avihepadnaviruses infect domestic ducks (DHBV) (Mason *et al*., 1980) and the

heron (Sprengel *et al*., 1988).

1.3.2 Animal Models

The role of HBV in HCC initiation and progression is substantiated by studies of non-human *Hepadnaviridae* in their respective hosts. HCC develops in all woodchucks, neonatally infected with WHV (Bréchot, 1987; Korba *et al*., 1989; Gerin 1990) and in 30 % of ground squirrels chronically infected with GSHV for five to six years (Marion *et al*., 1983, 1986). Many such tumours test positive for viral integration. Data gathered from investigations into animal hepadnaviruses in their respective hosts has proven invaluable as studies on the molecular biology of HBV have historically been hampered by the virus's relatively stringent tissue and species specificity.

1.3.3 HBV Genotypes and Their Distribution

There are eight HBV genotypes. These are defined by a divergence in their complete nucleotide sequence of more than 8.0 % (Okamoto *et al*., 1988; Norder *et al*., 1992a; Kramvis *et al.*, 2008) and more than 4 % at the S gene level (Norder *et al*., 1992b). They display a distinctive pattern of geographical distribution, and are designated A to H (Okamoto *et al*., 1988; Norder *et al*., 1992b; Magnius and Norder 1995; Stuyver *et al*., 2000). There is some relationship between the eight genotypes and the serological subtypes (serotypes) (Table I). In areas endemic for HBV a higher proportion of variable genotypes occurs (Bowyer *et al*., 1997; Carman *et al*., 1997).

Table I: Relationship between HBV genotypes and serotypes.*

Genotype	Distribution	Relationship to serotype
A	Pandemic but mainly in North America, north-west Europe and central Africa, Venezuela	*adw2, ayw1, ayw2, adw4*
B	Eastern Asia, Venezuela	*adw2, ayw1, adr*
C	Predominates in East Asia, Korea, Australia, China and Japan	*adrq+, adrq, ayr, adw, adw2, ayw2, ayw3, adwr*
D	Pandemic but more common in the Mediterranean, the Middle East and India	*ayw2, ayw3, ayw4, adw3*
E	West Africa, Nigeria and Western and Central Africa, The Gambia	*ayw4*
F	Native South Americans and Polynesia, France and Alaska	*adw4, adw4q-, adw2, ayw4*
G	France and the United States of America	*adw2*
H	Native Indians of Central America and in California and Mexico	*adw4*

* Adapted from Kramvis *et al*., 2005a

1.3.4 Composition of HBV

Three types of HBV particles are found in human serum, the complete virion or Dane particle (42 nm) (Figure 1.2), and two non-infectious particles, which are spherical (22 nm) and filamentous (22nm diameter, length variable) in shape and are composed entirely of HBV surface proteins and lipids from the hepatocyte membrane (Dane *et al*., 1970; Robinson and Lutwick, 1976), and may act as decoys for antibodies to the envelope proteins.

The Dane particle comprises an outer envelope of viral S proteins and host lipids, and an inner core consisting of a 27 nm spherical or icosahedral shaped nucleocapsid (Robinson and Lutwick 1976), composed of 21 kD phosphoprotein (HBcAg), with the viral genome at its centre (Figure 1.2) (Almeida *et al*., 1971; Robinson, 1974; Robinson and Greenman, 1974; Hruska *et al*., 1977; Landers *et al*., 1977).

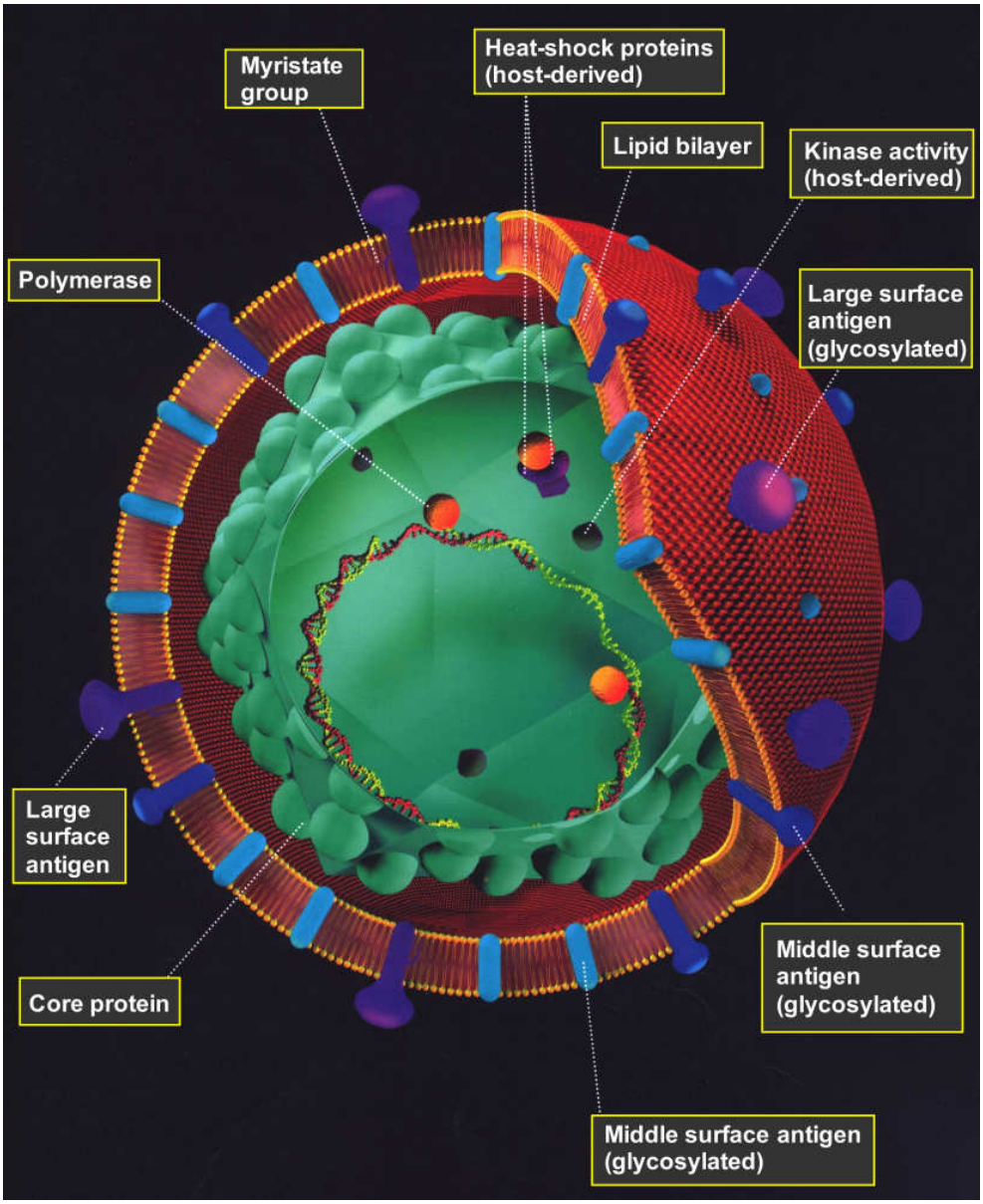


Figure 1.2: Diagrammatic representation depicting the HBV Dane particle.

Reproduced from Lai *et al*., 2002.

1.3.5 The Genome and Expressed Proteins of HBV

The HBV genome is a partially double stranded 3.2 kb DNA molecule (molecular weight $2.1x10^6$) (Figure 1.3) (Summers *et al*., 1975; Sninksy *et al*., 1979; Delius *et al*., 1983).

The minus/full-length (FL) strand undergoes transcription to yield the viral mRNAs and is covalently linked to a polymerase protein (Gerlich and Robinson, 1980) at its 5' terminus. The complementary plus/incomplete strand is secured to the minus strand at the 5' end by a 224 bp cohesive sequence resulting in a circular viral genome **(**Figure 1.3**)** (Sattler and Robinson, 1979). The cohesive terminus has 11 bp direct repeats (5'T-T-C-A-C-C-T-C-T-G-C3') on either side, termed DR1 and DR2, respectively, which are conserved *cis*-acting elements (Galibert *et al*., 1979; Ono *et al.*, 1983). The plus strand is attached to an oligoribonucleotide at its 5' end (Will *et al*., 1987) and does not always terminate at the same 3' nucleotide position each time the viral genome is synthesised. It is therefore of variable length, creating a gap region in the genome.

The extremely compact genome contains four partially overlapping open reading frames (ORFs) (Galibert *et al.*, 1979): the *S* gene encompassing the *preS* (divided into p*reS1*, and *preS2*) and *S* regions; the *preC/C* gene; the *P* gene and the *X* gene **(**Figure 1.3**)**.

Transcription and translation of the virus is firmly regulated by four promoters (*preS1*, *preS2*, *C* and *X*) and the orientation independent enhancer elements I and II **(**Figure 1.3**)** (Shaul *et al*., 1985; Yee 1989; Matsubara and Tokino1990). There is also a GRE (glucocorticoid-receptor element) (Tur-Kaspa *et al.*, 1986). mRNAs range in size from 0.7 to 3.5 kb (Guo *et al*., 1991; Yaginuma *et al*., 1993; Chisari, 2000) and all terminate at the only poly-adenylation signal on the genome. This conserved TATAAA sequence is thought to be responsible for RNA cleavage and poly-adenylation (Montell *et al*., 1983; Nevins 1983; Zarkower *et al*., 1986; Chen *et al*., 1995). When transcripts are generated from the upstream core/pregenomic promoter (Section 1.3.7), which is in close proximity to the adenylation site (Russnak and Ganem 1990), the poly-adenylation signal is ignored on the first pass of the host RNA polymerase II and is only honoured the second time around. This occurs for two reasons, firstly the poly-adenylation signal is inefficient because it is noncanonical and secondly the proximity of the 5' end of the transcript weakens application of the signal.

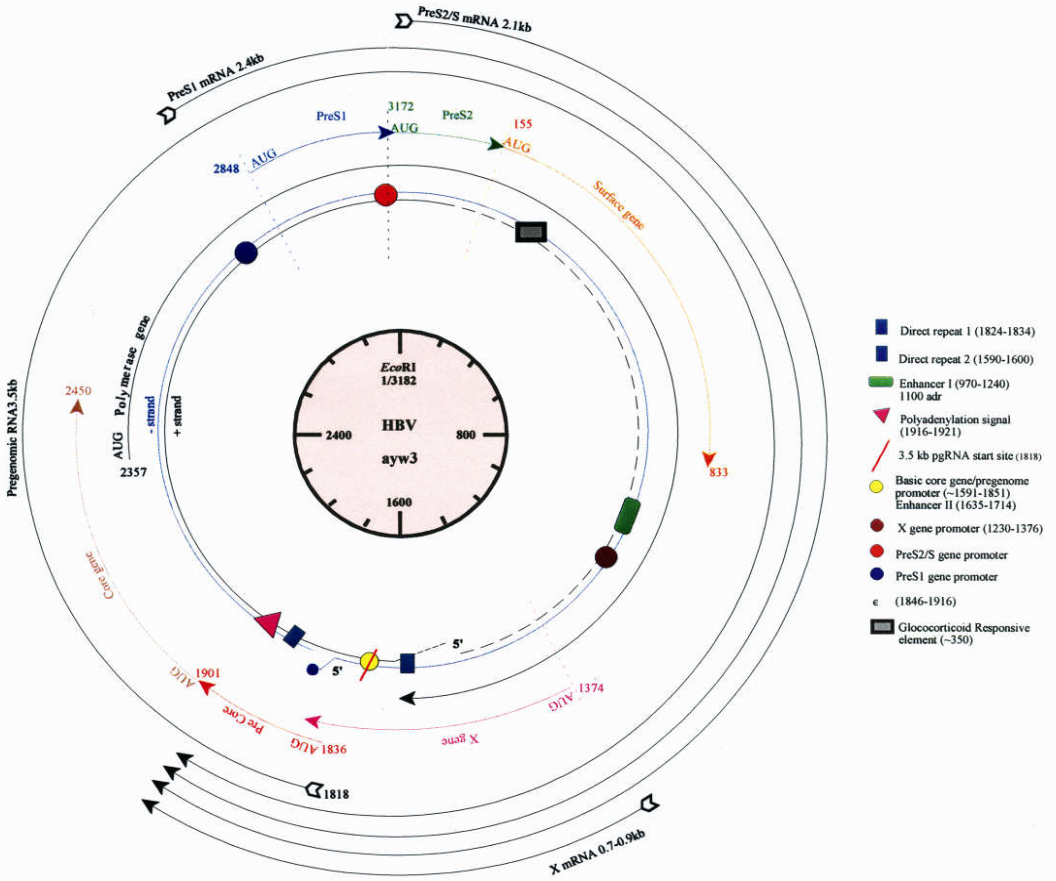


Figure 1.3: The HBV genome structure and genetic organisation.

Key: Numbering of nucleotides begins at the *Eco*RI site and is as per HBV subtype *ayw*.

Adapted from: Matsubara and Tokino, 1990, with modifications from Tiollais *et al.*, 1985*;* Tur-Kaspa *et al*., 1986; Kramvis and Kew MC, 1998.

At the second pass the efficiency is enhanced by the increased distance to the 5' end and the presence of RNA elements which strengthen signal usage. These RNA sequences are found only on the read-through transcript, because the DNA sequences that encode them reside upstream of the core promoter (Buckwold *et al*., 1996; Vyas and Yen**,** 1999). This feature is essential for the generating of pregenomic RNA (pgRNA).

1.3.5.1 *preS/S* ORF

Structure and Proteins

The *preS/S* ORF contains three in-frame ATG start sites (Heermann *et al*., 1984; Eble *et al*., 1986) and the *preS1* and *preS2* promoters (Figure 1.3). The *preS1* promoter regulates transcription of the *preS1*, *preS2* and *S* regions (Cattaneo *et al*., 1983, 1984; Standring *et al*., 1984). The *preS2* promoter drives transcription of a family of mRNAs 2.1 kb long (Cattaneo *et al*., 1983, 1984; Standring *et al*., 1984; Persing *et al*., 1985) (Figure 1.3).

S gene transcripts contain a region known as the postranscriptional regulatory element (PRE), which initiates their transport into the cytoplasm but prevents them from being spliced without affecting their transcription initiation or cytoplasmic RNA stability (Huang and Liang, 1993; Huang and Yen, 1994). S protein is the major component of the viral envelope and spontaneously assembles into 20 nm particles, which are secreted. Each viral envelope is constructed from some 100 S monomers, cross-linked by inter-chain disulphide bonds (Vyas *et al*., 1972). There are four major serological subtypes of S protein **(**section 1.3.6), which are antigenically heterogenous (Stibbe and Gerlich, 1983; Heermann *et al*., 1984; Michel *et al*., 1984; Milich *et al*.,1985; Neurath *et al*., 1985). Hence HBV vaccines are directed at *S* and *pre-S2* epitopes.

Dane particles are comprised mainly of LHBs proteins, while spherical and filamentous particles contain little or no LHBs, making them non-infectious (Heermann *et al*., 1984). Middle (MHBs) and small (HBsAg) proteins are produced in excess. The middle and large (LHBs) HBV surface antigens may be involved in attachment and entry into the hepatocyte. MHBs supports virus-host interaction and has trans-activating capacity (Kann, 2002).

1.3.5.2 *preC/C* ORF

Structure and Proteins

The precore/core (*preC/C*) ORF codes for the core protein (HBcAg) and for the HBV 'e' antigen (HBeAg). The core promoter, and both enhancers, assisted by a number of cellular transacting factors and nuclear receptors, regulate the transcription of precore mRNA and pgRNA (Will *et al*., 1987; Yaginuma *et al*., 1987a; Lopez-Cabrera *et al*., 1990; Guo *et al*., 1993; Raney *et al*., 1997; Yu and Mertz 1997). The pgRNA (3.5 kb) is longer than the complete genome because of a terminal redundancy resulting from transcription initiation at DR1, upstream of the poly-adenylation site (section 1.3.7) (Summers and Mason, 1982). This results in duplicate copies of the DR1 and epsilon (ε) (section 1.3.7) sequences (one copy at the 3' end and the other at the 5' end) in the pgRNA (section 1.3.7). The polymerase is expressed as a result of ribosomal scanning of the pgRNA (Jean-Jean *et al*., 1989).

HBc dimer associations result in the production of the HBV nucleocapsid (Robinson and Lutwick, 1976), within which the HBV genome and polymerase is contained. The capsid transports the HBV genome to and into the nucleus. Plus-strand DNA synthesis occurs within the capsid (Kann, 2002). The basic core promoter (BCP) region (approximately 1744-1851) is involved in transcription factor-binding and exact initiation of transcription of core mRNAs (Yuh *et al*., 1992).

HBeAg is not a component of the virus but is a soluble antigenic protein secreted into the serum and expressed on the cell surface. The core gene is preceded by a short in-phase ORF that encodes the precore protein, which is virtually identical in amino acid sequence to the core polypeptide except for an extra 29 amino acid residues at the amino terminus (Matsubara and Tokino, 1990). This protein has a hydrophobic transmembrane domain, which allows it to travel to the Golgi apparatus and the endoplasmic reticulum (ER) lumen (Ou *et al*., 1986; Bruss and Gerlich 1988). During entry into the ER, 19 residues of the protein are cleaved, and the remaining 10 residues prevent the protein from self-assembling into nucleocapsids. In the Golgi complex, the carboxyl end is cleaved at variable regions. The purpose of this complex series of modifications is unknown. It has been proposed that HBe antigen may function as an immune system suppressor (Chen *et al*., 2004).

1.3.5.3 *P* ORF

Structure and Protein

The *P* ORF covers 80 % of the HBV genome and overlaps the other three ORFs. It encodes the viral polymerase protein, which possesses polymerase (P), reverse transcriptase and RNaseH functions (Toh *et al*., 1983; Bartenschlager and Schaller, 1988; Schödel *et al.*, 1988; Khudyakow and Makhov, 1989; Radziwill *et al*., 1990; Bartenschlager and Schaller, 1992). The polymerase is responsible for minus and plus-strand synthesis and encapsidation of the pgRNA. HBV nucleocapsids and polymerase may also be involved in the transport of the HBV genome through the nuclear pore (Jilbert *et al*., 2002; Kann 2002).

The P protein has four domains. The N terminus (primase domain) functions primarily in priming synthesis of the viral minus strand, after which it remains attached to the strand at its 5' end (section 1.3.7). The second domain is a spacer region, while the RNA/DNA dependent polymerase activity of the protein resides in its third domain (Lanford *et al*., 1999). The protein exhibits polymerase activity only in the presence of the HBV encapsidation signal (epsilon, ε) present in the pgRNA (section 1.3.7) and then only when metal ions are available to act as co-factors (Bartenschlager and Schaller, 1992; Tavis *et al*., 1998; Urban *et al*., 1998). The RNase activity of the protein is in its fourth and last domain (Chang *et al*., 1990; Radziwill *et al*., 1990).

1.3.5.4 *X* ORF

Structure and Protein

The *X* gene is conserved among the mammalian hepadnaviruses (Lo *et al*., 1988). It has its own promoter and encodes the X protein (HBx), which is important in establishing infection *in vivo* but not *in vitro* (Yaginuma *et al*., 1987a; Blum *et al*., 1992; Chen *et al*., 1993; Zoulim *et al*., 1994) and is a transcriptional transactivator (Balsano *et al*., 1994). Its transcription is regulated by HBV enhancer I (Antonucci and Rutter 1989; Raney *et al*., 1990; Ohno *et al*., 1997). Although it is generally accepted that the *X* gene transcript is 0.7 kb in size, mRNAs ranging from 0.7-0.9 kb may occur (Guo *et al*., 1991). The HBx 17 kDa X protein is between 146 and 154 amino acids in size (depending on the HBV strain) (Kidd-Ljunggren *et al*., 1995). Extensive research has been done on the *X* gene and its protein, but its exact function remains elusive and the precise molecular mechanisms by which it acts are still uncertain. One fact is

however undisputed, HBx has an inordinate number of indiscriminate interactions with a variety of miscellaneous proteins, leading to several often apparently conflicting outcomes. HBx has been shown to interfere in regulatory pathways such as *Src*, *NH_2-terminal kinase* (*JNK*), *phosphatidylinositol-3-kinase* (*PI-3-K*) (Doria *et al*., 1995; Klein *et al*., 1997; Shih *et al*., 2000; Diao *et al*, 2001a) and *c-Jun* (Benn *et al*., 1996). It promotes unregulated cell growth by interfering in the *ras-raf-map* kinase pathway and preventing apoptosis (Benn and Schneider 1994; Wang *et al*., 1997) and has also been reported to activate the protein kinase C signalling pathway (Kekulé *et al*., 1993) and to have ribo/deoxy ATPase activity (de-Medina *et al*., 1994). HBx acts as a transcriptional co-activator and effector of various *cis*-acting DNA elements of cellular and viral genes (Spandau and Lee 1988; Levrero *et al*., 1990; Lee and Rho, 2000; Nijhara *et al*., 2001; Kim and Rho 2002; Shamay *et al*., 2002; Su *et al*., 2007) and may affect certain cellular promoters, thereby *trans*-activating cellular genes. Which promoters these are remains largely unknown. HBx further hinders basal transcription mechanisms and transcription factors (Lin *et al*., 1997, 1998) by inhibiting nucleotide excision and the repair of damaged cellular DNA (Groisman *et al*., 1999; Jia *et al*., 1999), possibly through its physical interaction with the proteins involved in mediating the process (Lee *et al*., 1995; Capovilla *et al*., 1997).

Of particular interest is the relationship between HBx and the tumour suppressor (TS), *p53*. The X protein has been documented to repress *p53* promoter activity (Kwon and Rho, 2003) by binding to the C terminus of the p53 protein, and forming a protein-protein complex, thereby inhibiting its DNA sequence-specific binding and transcriptional transactivation of cellular DNA sequences.

The net result is that HBx can modify cell proliferation, cause genomic transformation, and both inhibit and induce apoptosis (Shih *et al*., 2000; Diao *et al*., 2001a and b; Kim *et al*., 2001b). Transgenic mice experiments have shown over-expression of X and in one case concurrent initiation and development of HCC (Kim *et al*., 1991; *Koike et al*., 1994; Slagle *et al*., 1996; Terradillos *et al*., 1997; Yu *et al*., 1999).

1.3.6 Serology

The course of HBV infection can be monitored serologically with the use of laboratory tests for the different viral markers (Figure 1.4).

Once an individual is infected with the HBV an incubation period of 4-12 weeks occurs before the first antigen (HBsAg) is detected. In acute infection (Figure 1.4A), this is followed by the early acute stage with the emergence of HBeAg and increased HBsAg levels. In the acute phase, as symptoms begin to appear, total anti-HBc and anti-HBc IgM emerge and their levels rise, while HBeAg decreases and disappears as the patient progresses into the sero-conversion period of the disease. HBsAg levels begin to decrease during the acute phase and are non-detectable once sero-conversion is complete. The patient enters the convalescent period 3-6 months later, where anti-HBc total levels stabilise while anti-HBc IgM decrease. Anti-HBe initially increases before declining. In early recovery almost constant anti-HBc total levels may be observed, together with very gradually delining levels of anti-HBe and rising levels of anti-HBs. Recovery takes years and differs from early recovery in that levels of anti-HBs peak and slowly decrease over time. In acute HBV infection viral DNA is detectable only in the early acute and acute phases of infection (Hodinka, 1999).

Chronic infection arises (Figure 1.4B) when the acute stage of the disease is prolonged from 1 week to 6 months and HBsAg, HBeAg and anti-HBc peak and remain constant thereafter. If this continues for more than 6 months the patient has a chronic HBV infection. HBeAg levels may decrease during the acute phase and disappear well into the chronic stage which may be marked by the appearance of anti-HBe with levels increasing with time. In chronic infection HBV DNA remains present indefinitely (Hodinka, 1999).

Individuals with HCC are positive for HBsAg in as many as 85 % of cases in those areas where HCC is common and HBV is endemic, versus control subjects where the likelihood is 15 % or less. HBsAg is less frequently found in patients with cirrhosis or chronic hepatitis than in those with HCC (Arbuthnot and Kew, 2001).

Nine different serological subtypes of HBV are defined, based on S antigen reactivity with subtype-specific antibodies. The determinants comprise the '*a*' determinant, common to all

subtypes, and the mutually exclusive subdeterminants '*d*' or '*y*'; '*w*' or '*r*'. The nine serological subtypes are designated *ayw* (1, 2, 3 and 4), *ayr*, *adw* (2 and 4) and *adr* (*adrq+* and *adrq-*) (Swenson *et al*., 1991; Blitz *et al*., 1998). This is not the only way of differentiating HBV strains or genetic variablity. Advances in molecular biology techniques, computers and data handling programs, have permitted HBV to be classified according to genotype (section 1.3.3).

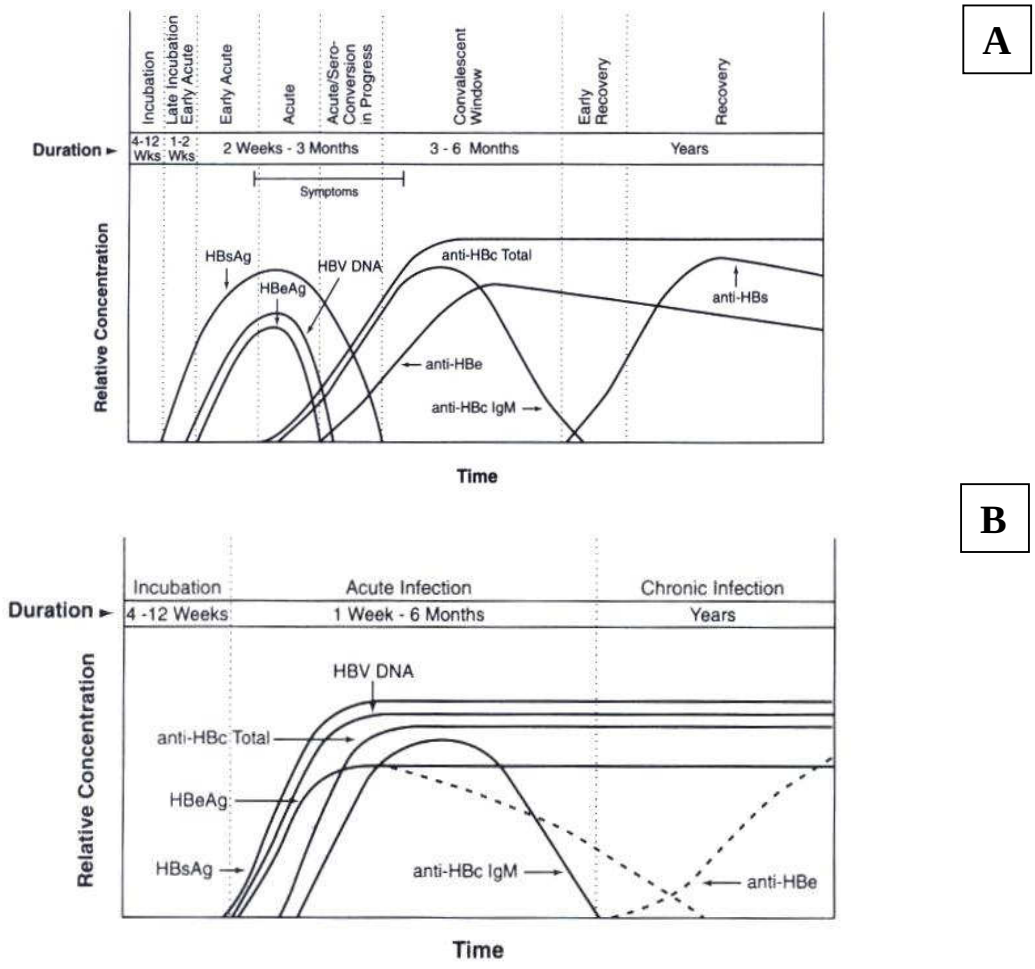


Figure 1.4: Serologic patterns observed during acute and chronic HBV infection.

Graphs taken from: Hodinka RL. In *Viral Hepatitis, Diagnosis, Therapy and Prevention*, Ed. Steven Specter, 1999 pages 200 & 202.

1.3.7 HBV Replication and Life Cycle (Figure 1.5)

Upon infection, the virus enters the cell membrane of the hepatocyte by way of as yet unknown cell surface receptor/s (Leistner *et al*., 2008). It sheds its envelope and the core particle travels to the nucleus, where the partially double stranded viral genome is converted into a covalently closed circular DNA (cccDNA). Unlike the retroviruses, hepadnaviruses do not integrate into the host genome as part of their viral life-cycle. They do not have their own viral integrase, but under certain circumstances they appear to integrate randomly making use of the host's own enzymes (Arbuthnot and Kew, 2001). Viral mRNAs required for expression of X and surface proteins, and larger than genome transcripts, are transcribed from the cccDNA and exit the nucleus into the cytoplasm. Transcription is co-ordinated by cellular transcriptional activators found in hepatocytes. The pgRNA is translated into the core (C) protein (HBcAg) and the viral polymerase/reverse transcriptase (P), wherease HBeAg is translated from the precore mRNA because pgRNA lacks the ATG start codon for HBeAg (Figure 1.5 [1]) (Enders *et al*., 1987; Nassal *et al*., 1990). The encapsidation of pgRNA is triggered by the binding of the viral polymerase to its 5' end (Figure 1.5 [2]), which is folded into the stem-loop ε structure (Bartenschlager *et al*., 1990; Junker-Niepmann *et al*., 1990; Bartenschlager and Schaller 1992). The core proteins accumulate in the cytoplasm, and at a critical concentration (~0.8) µM self-assemble to form a capsid (Cohen and Richmond, 1982;

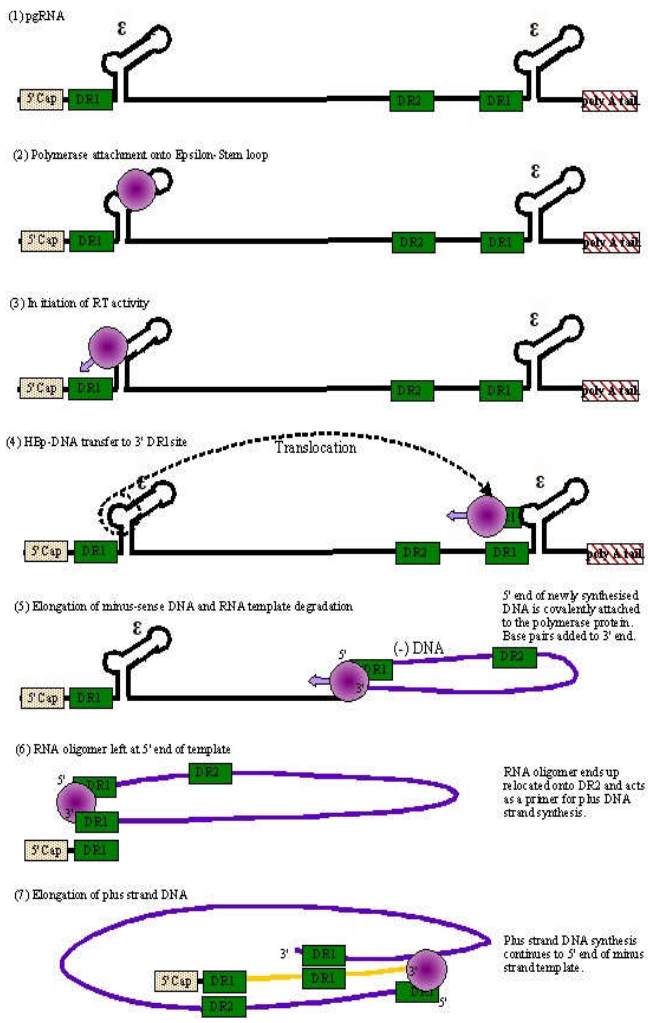


Figure 1.5: Diagrammatic representation of the replication of the HBV genome.
Reproduced with modifications from: Lai *et al*., 2002.

Miyanohara *et al*., 1986; Junker-Niepmann *et al*., 1990; Seifer and Standring, 1995). The ε/polymerase interaction is followed by the packaging of one pgRNA/polymerase per capsid (Bartenschlager *et al*., 1990) and the subsequent activation (Figure 1.5 [3]) of the reverse transcription activity of the polymerase (Wang *et al*., 1994a). Ribosome-mediated suppression is thought to prevent pgRNAs with the epsilon start codon from being packaged (Nassal *et al*., 1990). Positive and negative strand synthesis occurs in the core particle (Figure 1.5). The ε stem-loop functions as the template for the synthesis of the first three nucleotides of the minus strand DNA (Figure 1.5 [1]-[3]), after which nascent minus strand and attached polymerase are transferred to the 3' DR1 on the pgRNA template (Figure 1.5 [4]), where DNA synthesis reinitiates (Hirsch *et al*., 1990). As the minus strand is being reverse-transcribed the pgRNA template is degraded by the RNAse H activity of the viral polymerase (Figure 1.5 [5]), leaving the last 18 ribonucleotides to serve as a primer for plus strand synthesis, and the primer translocates to DR2 (Figure 1.5 [6]) where synthesis of the DNA plus strand commences. The process continues until the free nucleotides within the capsid become depleted or steric hindrance occurs, resulting in the partially double-stranded genome (Nassal and Schaller, 1996) (Figure 1.5 [7]). The nucleocapsids, now containing mature viral DNA, may find their way to the nucleus where they release their DNA. Those capsids not headed for the nucleus interact via their HBc proteins with S protein at certain areas of the Golgi complex and become enveloped. From here the virions exit the cell by exocytosis (Kramvis and Kew, 1998; Wands, 2004).

1.4 Mechanisms of HBV-Related HCC and the Primary Biochemical Pathways Modified in Human HCC

The classical observed features of cancer are: loss of cellular differentiation, lack of spatial inhibition and growth of the affected organ beyond its boundary, immortalisation of the cancer cell, metastasis of cells within and beyond the organ of origin. Mutant cells with growth advantages are selected for and develop into clones while those with mutations detrimental to their survival undergo apoptosis. Massive cellular death and increased cell turnover results in mechanical damage to the liver – an indirect mechanism of HCC development (section 1.5.1.6). It is now generally accepted that cancer initiation and progression occur as a result of cumulative mutations. This is particularly true of HBV-related carcinogenesis where no single mechanism contributes to its aetiology and numerous and divergent processes may be involved.

The accumulation of mutations necessitates DNA replication and cell division. Mature, healthy hepatocytes divide very infrequently. Their lifetime in adults is approximately 400 days (Matsubara and Tokino, 1990). On the other hand, HBV infected hepatocytes are routinely targeted and eliminated by the host immune system. Hepatocyte regeneration occurs, allowing for increased DNA replication and cell turnover. The selection of mutant cells with growth advantages and their development into clones follows (Matsubara and Tokino, 1990). Thus genes involved in the pathways of DNA damage repair, apotosis, cellular growth and regeneration and control of the cell cycle would be likely candidates for mutation in HCC.

Ozturk, (1999), suggests exactly that: altered genes may be present in large numbers in HCC, but will have low individual frequencies of mutation. However, the affected genes tend to be involved in one or more common growth regulatory pathway/s that prove to be vital to the cell. Several of these genes were grouped in four main pathways (Figure 1.6): the *p53* pathway (DNA damage response), the *RB1* pathway (cell cycle control), the *TGF-β* pathway (growth inhibition and apoptosis), and the *β-catenin*/adenomatosis polyposis coli (*APC*) pathway (morphogenesis and signal transduction). These pathways are likely related, and could represent individually a distinct step of hepatocellular carcinogenesis. Regrettably, it is not yet known exactly what events will lead to HCC initiation, and whether these might accumulate in a specific order in HCC progression (Ozturk, 1999).

The controlled cell cycle guarantees faithful duplication of all cellular components in their appropriate order and correct distribution into daughter cells. In interphase the G1 and G2 phases of the cell cycle are characterised by intense metabolic activity, cell growth and differentiation. There are checkpoints at the G1 and G2 phases at which cell cycle arrest may occur in case of errors. The crucial point in terms of faithful DNA duplication is at the S phase. In mammalian cells extracellular growth factors, mitogen antagonists, differentiation inducers and spatial cues are taken into account before a cell enters S phase (Hartwell and Weinert 1989). It it therefore imperative that cells retain their sensitivity to these stimuli and respond appropriately. If problems are detected the cell cycle can arrest. Once S phase is initiated, the cycle runs to completion.

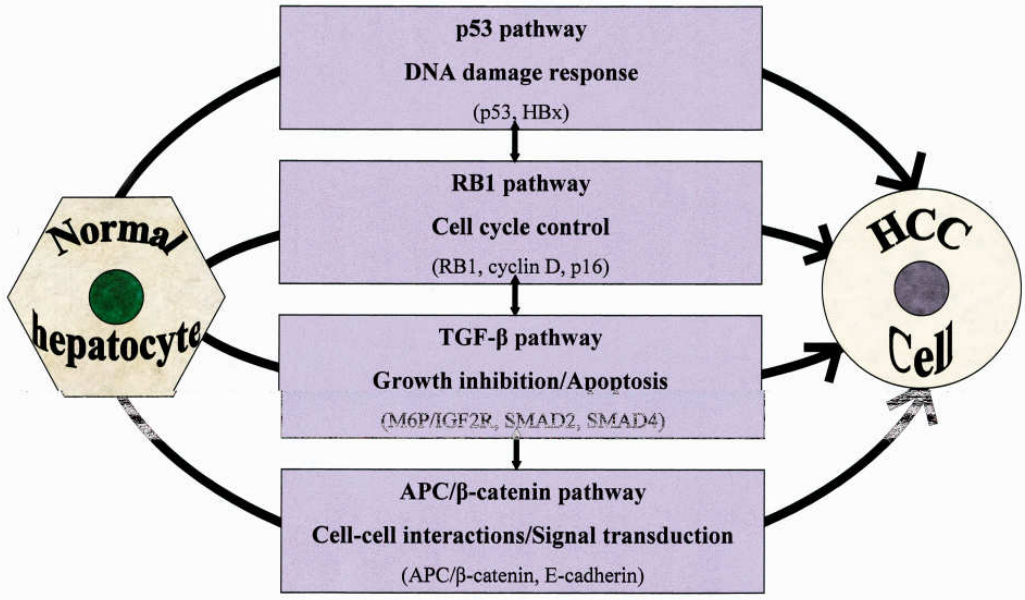


Figure 1.6: Main regulatory pathways altered in human hepatocellular carcinomas.
Reproduced from: Ozturk, 1999.

The prevention of propagation of damaged DNA to daughter cells is one of the main functions of the p53-activated signal pathways, which appear to delay G1 to S phase progression, allowing time for DNA repair. The p53 protein also forms part of the transcription-repair complex TFIIH involved in nuclear excision repair (Ko and Prives, 1996) and may be involved in inhibiting critical helicase activities required for DNA replication and recombination, thus prohibiting single-strand recombination intermediates that could result in gene duplication, amplification, and activation of oncogenes (Levine *et al*., 1994; Wang *et al*., 1995). The p53 protein may also induce apoptosis if DNA damage is excessive (Lowe *et al*., 1993) or if the cellular environment appears abnormal.

Cyclins, cyclin dependent kinases (CDKs) and certain other genes (eg. cell division control genes) are the key regulators of G1 and G2. Cyclins can be modified at cell cycle checkpoints in a manner which induces them to halt the progression of the cell cycle. Checkpoint pathways are influenced by availability of nutrients and growth factors, and can in turn be stimulated to regulate levels or ability of cyclins to activate CDKs (Levine *et al*., 1994). The clue to one

possible involvement of the *p16INK4A, Cyclin D, RB1* gene products in human HCC is their link to the CDK4 polypeptide, with which they interact, in one way or another, at the G1/S transition phase of the cell cycle. At G1/S cyclin D and CDK4 protein form a complex, whereas the retinoblastoma protein is the CDK4 substrate. The p16 protein in turn is known to inhibit CDK4 activity (Serrano *et al*., 1981). Individually, mutations in each of these genes occurs in 10 to 20 % of HCCs; combined they may cause deregulation of the cell cycle in 30 % or more of HCCs (Nakamura *et al.*, 1991; Fujimoto *et al.*, 1994; Nishida *et al.*, 1994; Ashida *et al.*, 1997). Similarly, the *M6P/IGF2R* (insulin-like growth factor II receptor) and human homologues 2 and 4 of the Drosophila Mad (*SMAD2*, *SMAD4*) gene products are involved in the regulation of transforming growth factor beta (*TGF-β*), which acts in hepatocytes as a growth inhibitor and apoptosis inducer (De Souza *et al.*, 1995; Derynck *et al.*, 1998). *SMAD2* and *SMAD4* mediate the activity of *TGF-β*, and mutations in these genes occur with 0 to 2 % and 0 to 6 % frequency, respectively in HCCs (Kawate *et al.*, 1999a and b; Yakicier *et al.*, 1999). *M6P/IGF2R* has been implicated in *TGF-β* activation, and 18-33 % of HCCs contain mutations of this gene (De Souza *et al.*, 1995; Piao *et al.*, 1997a).

Jointly these genes result in modification of the *TGF-β* pathway in approximately 25 % of HCCs.

This pathway is also involved in suppression of pRB phosphorylation resulting in cell cycle arrest at G1 and inducing cells to exit the cell cycle and differentiate further (Weinberg, 1995). pRB as a suggested component of the cellular ‘generational clock’, which acts to record the number of divisions before a cell differentiates, dies or irreversibly exists the cell cycle, may result in cellular immortalisation if inactivated. Complete absence of the RB1 protein (pRB) would therefore result in unrestricted cell growth.

E-cadherin protein is the primary adhesion molecule in epithelium (Shimoyama *et al*., 1989). Cell-cell adhesion and cell invasion of surrounding tissue are two of the main features that characterise the development of malignant tumour. Loss of e-cadherin correlates with decreased cell-cell adhesion, cellular phenotypic changes, development of invasive properties and metastatic potential (Takeichi, 1991). In HCC intrahepatic metastasis is common as is intrahepatic recurrence after liver transplantation (Nagao *et al*., 1990; Iwatsuki *et al*., 1991).

The *β-Catenin* protein in turn interacts with both the APC and E-cadherin proteins, all of which appear to be involved in intercellular interactions (Peifer, 1997; Hirohashi, 1998). The β-catenin protein may have an additional role as a transcription regulator (Peifer, 1997). Somatic mutations of *β-catenin* occur in 19-26 % of HCCs (de La Coste *et al.*, 1998; Miyoshi *et al.*, 1998). *APC* mutations are relatively rare in HCC (Kita *et al.*, 1996), whereas *E-cadherin* has been shown to be mutated in HCC (Slagle *et al.*, 1993; Kanai *et al.*, 1997). These observations indicate that the *β-catenin*/*APC* gene pathway appears to be modified in at least 30 % of HCCs.

1.5 Integration and HCC

Integrants comprising a complete and uninterrupted HBV genome are rarely isolated/cloned/identified in HCC specimens. One entire HBV genome, with an additional 25 % of HBV DNA was documented by Dejean *et al.*, (1983). Generally integrants can be of two types:

- simple, consisting of single HBV genomes, without rearrangement but with deletions (Nakamura *et al.*, 1988)
- complex, comprising multiple HBV genome copies, with at least one virus-virus junction, often with complicated rearrangements resulting from post-integration recombination (Ogston *et al.*, 1982; Dejean *et al.*, 1984; Mizusawa *et al.*, 1985; Yaginuma *et al.*, 1985; Ziemer *et al.*, 1985; Hino *et al.*, 1986; Berger and Shaul 1987; Nagaya *et al.*, 1987; Shih *et al.*, 1987; Yaginuma *et al.*, 1987b; Zhou *et al.*, 1988; Hino *et al.*, 1989).

Such integrants have also been found in non-tumorous tissue adjoining HCCs, and in patients with chronic hepatitis alone (Koshy *et al.*, 1981; Shafritz *et al.*, 1981; Hino *et al.*, 1984). Generally it is not the type of integrant that varies between these conditions and HCC, but the subsequent additional mutations caused by the presence of the integrant and not directly evident from the initial event. Such integrant clones may be subjected to selection by the host's immune response. As previously stated, the exact point at which a non-tumorous liver becomes a HCC is as yet undefined.

Viral replication cannot occur from these disrupted integrated sequences as the viral genome sequence is not preserved.

In the viral genome certain sites are favoured for integration. Over half of all documented integrants (section 1.5.4) have one of their junctions at the viral 5' cohesive end sequence, within the DRs (Dejean *et al*., 1984; Koch *et al*., 1984a; Mizusawa *et al*., 1985; Yaginuma *et al*., 1985; Ziemer *et al*., 1985; Hino *et al*., 1986; Nagaya *et al*., 1987; Yaginuma *et al*., 1987b; Shih *et al*., 1988; Zhou *et al*., 1988) and the cohesive end region is considered a mutational 'hot spot' (Quade *et al*., 1992). Integration is proposed to occur during active viral replication (Nakamura *et al*., 1988) (section 1.5.4). The second viral-host junction is non-specific with respect to the virus, and is thought to comprise merely a recombination of HBV and cellular DNA (Robinson, 1994). The preferential integration of *preC/C* mutant viruses in HCC tissues has been documented. Of those integrants analysed by Zhong *et al*., (2000) 65 % contained *preC/C* mutations, comprising the 1896 stop codon, base substitutions, missense mutations and even deletion of the core sequence.

Although there is no common host genome site for HBV integration because it generally appears to occur randomly (Nagaya *et al*., 1987), it has been proposed that certain genes (eg. *hTERT*) and chromosomes such as chromosomes 3, 11p and 17 (Bowcock *et al*., 1985; Rogler *et al*., 1985; Hino *et al*., 1986; Minami *et al*., 2005; Murakami *et al*., 2005) (Table II), may be targeted more often than others, and therefore integration may not be entirely random.

A number of lesions, which may have occurred immediately or at some time post-integration, have been documented at integration sites (Tokino *et al*., 1987; Zhou *et al*., 1988; Takada *et al*., 1990; Tokino *et al*., 1991; Pineau *et al*., 1996).

- *microdeletions* (10bp) and larger deletions of cellular DNA (Nagaya *et al*., 1987; Tokino *et al*., 1991; Tokino and Matsubara 1991) resulting in loss of heterozygocity (LOH)
- *translocations* where the viral DNA joins regions of two different chromosomes (Hino *et al*., 1986; Möröy *et al*., 1986)
- *rearrangements* such as reversed and duplicated sequences containing both viral and genomic DNA (Koshy *et al*., 1981; Shafritz *et al*., 1981; Hino *et al*., 1984; Tsuei *et al*., 1994; Simon and Carr, 1995). Rearrangements in integrated viral DNA may also have occurred before integration (Takada *et al*., 1990).
- *amplification* of cellular DNA at the integration site (Hatada *et al*., 1988)
- *allelic imbalance* (AI); *polyploidy* and *point mutations* far from the site of integration

As in other cancers, the genes suffering these changes are likely to be oncogenes, TS genes, or genes involved in DNA repair (Figure 1.6) (Ozturk, 1999; Murakami *et al*., 2005).

1.5.1 Making Sense of Integration Events

From previous studies, the frequency of integrated HBV DNA in HBV-related HCCs is at least 80 % (Bréchot *et al*., 2000; Murakami *et al*., 2005). However, the true percentage of HCCs with integrated HBV DNA may well approach 100 % (Murakami *et al*., 2005).

In order to facilitate the understanding of the events leading to HBV-induced hepatocarcinogenesis, Robinson, (1994) catalogues four main possibly oncogenic events, which occur with frequency of at least 25 % and are not necessarily mutually exclusive:

- *insertional mutagenesis*, often coupled with DNA deletions, with the possible *cis*-activation of cellular genes or the generation of hybrid proteins with altered function
- *transactivation* of cellular genes by HBV proteins
- *amplification* of cellular Oncogenes
- *alteration of TS genes* or interference of HBx with TS proteins and the subsequent effects on apoptosis and DNA repair.

Events not easily contained in these four groups have been listed under ‘other’ (section 1.5.1.5).

1.5.1.1 Insertional Mutagenesis

Insertional mutagenesis occurs in the woodchuck, where, in at least one-half of infected woodchucks, the WHV enhancer region has integrated in, or close to, the *c-myc* and *N-myc* proto-oncogenes (Möröy *et al*., 1986; Hsu *et al*., 1988; Etiemble *et al*., 1989; Möröy *et al*., 1989; Fourel *et al*., 1990, 1992). This results in increased levels of *myc* and *myc*-WHV hybrid transcripts.

HBV-induced insertional mutagenesis has been documented for a number of genes with different repercussions (Table II). These studies reinforce the notion that despite HBV integration being random, it appears sometimes to favours certain chromosomes, areas of

repetitive sequence, and possibly certain genes. Whether this is in any way linked to its mechanism of integration remains uncertain.

In a 2005 study, Murakami *et al.*, identified 32 genes where HBV had integrated close to or in the gene sequence (Table II). These, together with 10 genes they had previously published, comprised a total of 15 cancer-related genes and 25 cellular genes. Although the effect of the integration event was not established in the majority of cases, it became evident that the genes appeared to belong to distinct cellular pathways, namely calcium-signalling-related genes (*SERCA1, ITPR1, ITPR2, SMOC1*) (Table II) 60s ribosomal protein encoding genes (*L7a, L14, L17*), and platelet derived growth factor and mixed lineage leukaemia encoding genes (*PDGR receptor β; PDGF β*; *MLL2; MLL4*) (Table II). Five genes were involved in apoptosis control (*hTERT* repeatedly; *TRAP150α; SCARA3; MAPK1; BCL2L2*) (Table II), and two were TS genes (*APCL; MGMT*) (Table II). Thus in the majority of cases (89.7 %), a cell growth advantage could have resulted from the integration event. They also discovered HBV integrants in or close to 19 sequences potentially coding for unknown proteins.

Table II: HBV integration into, or in the region of, known human genes and the effect of the integration event (if known).

Gene name/ symbol*	Gene product function/protein family	Sample type	Integrant position relative to host gene	Region of HBV integrated	Effect of integration event	References
RAR-β Retinoic acid receptor beta RARB HGNC 9865	Regulates genes involved in cell growth and differentiation	Human HCC and erythro-cytes	*S* gene fused in-frame with most of the host gene sequence	*S*	Altered regulation of normally poorly expressed *RAR-β* gene resulting in truncated chimeric protein and possible contribution to cellular transformation	Dejean *et al*., 1986, 1990; Garcia *et al*., 1993
Cyclin A2 (*CCNA2*) Cyclin A2 HGNC 1578	Cell division	Human HCC and rat kidney cells	In-frame in 2nd intron	*Pre-S2*/to middle *S*	Strong expression of stable hybrid protein-containing signals for cyclin degradation leading to possible increase in the rate of cell division	Wang *et al*., 1990, 1992; Berasain *et al*., 1998
Mitochondrial gene ***URF4*** Unidentified reading frame 4	Encode components of respiratory-chain NADH dehydrogenase[1]	PLC/PRF/5	In promoter	3125 bp integrant	Possible expression of hybrid protein, effect unknown	Koch *et al*., 1989
3' to the ***β-globin*** gene cluster	*β-globin* genes are expressed during tissue development in adults[2]	Human HCC	In semi-repetitive sequences TBG41, of the *Kpn*I[3] family, 3kb 3' to the β-globin gene cluster	Junction: 1886 (*adr*) and pre-core 1970 at cohesive end region	Possible production of HBsAg and truncated HBx, possible core/polymerase-host hybrid protein, effects unknown	Quade *et al*., 1992
ITPKB Inositol 1,4,5-triphosphate-3 kinase B ITPKB HGNC 6179	Calcium regulation	One each in PBMCs of 3 HBV-related chronic hepatitis patients	73 kb downstream	*X* and *preC/C*	unknown	Murakami *et al*., 2004
RAB30 Member of *RAS* oncogene family	Growth-related protein	PBMCs	4 kb downstream	DR1 and *preC/C*		

TNF Tumour necrosis factor-induced protein TNF HGNC 11892	Involved in organisation and function of immune system	Liver tissue of patient with acute hepatitis	In intron between 2nd and 3rd exons	*X*	unknown	Murakami *et al.*, 2004
MK Mevalonate kinase MK TAIR AT5G27450	Mevalonate phosphorylation Regulation of cell growth	PLC/PRF/5	Between gene sequences and promoter and regulatory elements	*preS2/S*	Over-expression of hybrid transcripts driven by HBV promoter and over-production of functionally active MK protein possibly resulting in aberrant growth	Graëf *et al.*, 1994
NTRK2 Neurotropic tyrosin receptor kinase 2 NTRK2 HGNC 8032	Cell survival, growth, differentiation	Human HCC	13 kb downstream	*X*	Possible modification of expression of cellular gene, possible modified function of HBx	Paterlini-Bréchot *et al.*, 2003; Murakami *et al.*, 2005
IRAK2 IL-IR-associated kinase 2 IRAK2 HGNC 6113	Ras-dependent cell-signalling	Human HCC	6 kb upstream*[A] intron downstream from exon 3*[A]	*X*	Possible modification of expression of cellular gene, possible modified function of HBx	Paterlini-Bréchot *et al.*, 2003; Murakami *et al.*,2005
p42MAPK1 p42 mitogen-activated protein kinase 1 MAPK1 MGI 1346858	Cell-signalling (Ras/MAPK pathway)	Human HCC	20 kb upstream*[B] 22 kb downstream*[B]	*X*	Possible modification of expression of cellular gene, possible modified function of HBx	Paterlini-Bréchot *et al.*, 2003
IP3R2 (ITPR2) Inositol 1,4,5-triphosphate receptor type 2 ITPR2 HGNC 6181	Calcium channels in ER membrane; cell transformation and progression	Human HCCs	45 kb upstream*[C] 44 kb downstream	*X*	Possible modification of expression of cellular gene, possible modified function of HBx	Paterlini-Bréchot *et al.*, 2003; Murakami *et al.*, 2005
ITP3R1 (ITPR1) Inositol 1,4,5-triphosphate receptor type 1 ITPR1 HGNC 6180	Calcium channels in ER membrane; cell transformation and progression	Human HCC	Within 39th exon	*X*	Integrant contained a 3' truncated X gene whose protein had cell proliferation, viability and transformation control capabilities	Tu *et al.*, 2001; Paterlini-Bréchot *et al.*, 2003

SITA *(ST3GAL6)* Alpha 2,3 sialyltransferase ST3GAL6 HGNC 18080	Protein glycosylation and tumour invasion, cell differentiation	Human HCC	3 kb upstream*[D] 4 kb downstream* [D]	*X*	Integrant contained a 3' truncated X gene whose protein had cell proliferation, viability and transformation control capabilities	Paterlini-Bréchot *et al.*, 2003; Murakami *et al.*, 2005
TRUP (RPL7A) Thyroid hormone-uncoupling protein TRUP HGNC 10364	Interferes with thyroid hormone and retinoic acid receptors inhibiting their interaction with DNA	Human HCC	8 kb downstream	*X*	Integrant contained a 3' truncated X gene whose protein had cell proliferation, viability and transformation control capabilities	Tu *et al.*, 2001; Paterlini-Bréchot *et al.*, 2003
hTERT Human telomerase reverse transcriptase TERT HGNC 11730	Catalyses telomere synthesis reactions	Human HCC	5.2 kb upstream*[E] 10.8 kb upstream*[E]	*X*	Possible modification of expression of cellular gene, possible modified function of HBx	Gozuacik *et al.*, 2001; Paterlini-Bréchot *et al.*, 2003; Murakami *et al.*, 2005
		2 Human HCCs	0.3 downstream	*X*	Possible influence on apoptosis	Murakami *et al.*, 2005
			16 kb downstream	*Core* gene		
		HuH4	In promoter	*Pre-S1* and HBV enhancer sequence	Cis activation of *hTERT* gene expression	Kekulé *et al.*, 1990; Horikawa and Barrett, 2001
		Human HCC	577bp of upstream promoter region, in-frame	HBV Genotype C (2318-2396; 2898-3215; 1-545)	Resulting in upregulation of *hTERT*	Ferber *et al.* in 2003
		SNU449 cell line	Intron 2 (35782-36418)	2 HBV fragments in opposite directions (1451-1805; 1216-436 Genotype C)		

SERCA1 (ATP2A1) Sarco-endoplasmic reticulum calcium ATPase ATP2A1 HGNC 811	Regulation of calcium levels in cell; neural signal transmission, muscle contraction, cell proliferation and apoptosis	Human HCC	3rd exon	*X*	HBV X/SERCA1 hybrid protein with C-terminally truncated SERCA1. Transcription driven by HBV X gene promoter via cis-activation. Hybrid protein localised to ER, causing calcium depletion & apoptosis	Chami *et al.*, 2000, 2001; Gozuacik *et al.*, 2001; Murakami *et al.*, 2005
TRAP150α Thyroid hormone receptor-associated protein-150 alpha	Coactivator of nuclear receptor involved in transcription	Human HCC	In intron upstream to 6th exon	*X*	Possibly involved in apoptosis	Gozuacik *et al.*, 2001; 2003 Murakami *et al.*, 2005
hMCM8 Minichromosome maintenance protein-related gene MCM8 HGNC 16147	Control of DNA replication	Human HCC	Intronic sequence downstream of 5th exon	*X*	c-terminally truncated hMCM8 protein produced, may affect control of cell proliferation	
FR7	New gene on 12p, expressed in foetal and liver tumour tissues	Human HCC	0.4 kb upstream	*X*	Chimeric transcript; effect unknown	
NMP Nuclear matrix protein p84	Control of the cell cycle	Human HCC	Intronic sequence Upstream from exon 11	*X*	Proposed possible disruption of cell cycle control	
RBMY (RBMY1A1) RNA-binding motif Y chromosome RBMY HGNC 9912	Expressed in male germ cells only; possible signal transduction and RNA activation	Childhood Human HCC	Intron 6, separating N-terminal RNA-binding motif from C-terminal auxiliary domain	Junction at 1817 just prior to DR1 sequence	Normally silent *RBMY*-activated and expressed in tumour but not in NT	Tsuei *et al.*, 2002
TACC1 Transforming acedic coiled-coil-containing protein 1 TACC1 HGNC 11522	Belongs to a family of proto-oncogenes[4]	Human HCC	15 kb downstream	*X*	unknown	Murakami *et al.*, 2005

IRF2 Interferon regulatory factor 2 IRF2 HGNC 6117	Interferon regulation	Human HCC	Intronic sequence downstream from 2nd exon	*Core* gene	unknown	Murakami *et al*., 2005
MUC16 Mucin 16 MUC16 HGNC 15582	Transmembrane region and tyrosine phosphorylation site	Human HCC	Intronic sequence upstream from 2nd exon	*S*	unknown	
PDGFRB Platelet derived growth factor receptor beta precursor; beta platelet derived growth factor receptor PDGFRB HGNC 8804	Growth factor	Human HCC	Intronic sequence upstream from 14th exon	*X*	unknown	
PDGFB Platelet derived growth factor beta polypeptide (simian sarcoma viral [v-sis] oncogene homologue) PDGFB HGNC 8800	Growth factor	Human HCC	530 kb upstream	*X*	unknown	
APCL (APC2) Adenomatous polyposis-coli-like *APC2 HGNC 24036*	APC homology	Human HCC	15 kb upstream	*X*	unknown	
GA (GLS2) Liver mitochondrial glutaminase/ breast cell glutaminase GLS2 HGNC29570	Glutaminase proteins/proto-oncogene	Human HCC	Intronic sequence downstream from 17th exon	*X*	unknown	
MGMT O-6-methylguanine-DNA methytransferase MGMT HGNC 7059	DNA reparation	Human HCC	6 kb downstream	*X*	unknown	

EVER2 (TMC8) Epidermodysplasia verruciformis 2 TMC8 HGNC 20474	Integral membrane protein located in the ER	Human HCC	Within 2nd exon	*X*	unknown	Murakami *et al.*, 2005
FN1 Fibronectin 1 FN1 HGNC 3778	Cell adhesion, morphology and surface architecture	Human HCC	14 kb downstream	*X*	unknown	
NEDD4L Ubiquitin-protein ligase NEDD4L-like; neural precursor cell expressed, developmentally downregulated gene 4-like NEDD4L HGNC 7728	Ubiquitin ligase	Human HCC	0.4 kb upstream	*X*	unknown	
NCF1 Neutrophil cytosolic factor 1 NCF1 HGNC 7660	NADPH oxidase	Human HCC	Intronic sequence downstream 5th exon	*X*	unknown	
CCT Chaperonin-containing TCP1, subunit 8 (theta) CCT HGNC 1614	Cell cycle regulation	Human HCC	36 kb downstream	*X*	unknown	
TCEA3 Transcription factor SII-related protein 4 TCEA3 HGNC 11615	Binds to RNA polymerase II and stimulates nuclease activity	Human HCC	7.4 kb upstream	*X*	unknown	
SCARA3 Scavenger receptor class A, member 3 SCARA3 HGNC 19000	Apoptotic regulation	Human HCC	Intronic sequence downstream from 6th exon	*X*	unknown	

GCHFR GTP cyclohydrolase 1 feedback regulatory protein GCHFR HGNC 4194	Key enzyme of tetrahydrobiopterin biosynthesis	Human HCC	2 kb downstream	*X*	unknown	Murakami *et al*., 2005
BIRC3 Baculoviral IAP repeat-containing 3 BIRC3 HGNC 591	Apoptosis inhibitor	Human HCC	22 kb downstream	*X*	unknown	
CHML Choroideraemia-like (Rab escort protein 2) CHML HGNC 1941	Geranylanylation of most Rab proteins	Human HCC	5 kb upstream	*X*	unknown	
ORM1 Orosomucoid 1 ORM1 HGNC 8498	Ubiquitin protein	Human HCC	Within 1st exon	*X*	unknown	
BCL2L2 BCL-like 2 BCL2L2 HGNC 995	Apoptotic regulation	Human HCC	Within 4th exon	*X*	unknown	
RAI17 (ZMIZ1) Retinoic acid-induced gene 17 ZMIZ1 HGNC 16493	Co-regulator of androgen receptor	Human HCC	6 kb upstream	*X*	unknown	
CASPR3 (CNTNAP3) Cell recognition molecule CASPR3 CNTNAP3 HGNC 13834	Cell differentiation	Human HCC	Intronic sequence upstream of 11th exon	*X*	unknown	
SFXN5 Sideroflexin 5 SFXN5 HGNC 16073	Rab11-interacting protein	Human HCC	70 kb downstream	*X*	unknown	
SMOC1 SPARC-related modular calcium-binding 1 SMOC1 HGNC 20318	Calcium regulation	Human HCC	2.5 kb downstream	*X*	unknown	

KIF3C Kinesin family member 3C KIF3C HGNC 6321	Kinesin family	Human HCC	Intronic sequence upstream of 4th exon	*X*	unknown	Murakami *et al.*, 2005
GRID2 Glutamate receptor, ionotrophic, delta 2 GRID2 HGNC 4576	Inontrophic glutamate receptor	Human HCC	0.9 kb upstream	*X*	unknown	
WBSCR1 (EIF4H) Williams-Beuren syndrome (WS) chromosome region 1 EIF4H HGNC 12741	Microdeletion at this locus results in WS (neurological disorder; abnormalities: behavioural, physical, cognitive)[5,6]	Serum of acute hepatitis patient	Within *WBSCR1* gene (7q11.23)	Portion of *X* and *S*	unknown	
MLL2 (MLL4)[*G] Myeloid/lymphoid or mixed lineage leukaemia 2 MLL2 HGNC 7133	Transcriptional regulator	Human HCC	12 kb downstream	*X*	unknown	
MLL4 (MLL2)[*G] Myeloid/lymphoid or mixed lineage leukaemia 4 MLL4 HPRD 06017; MIM 606834	Transcriptional regulator. Member of *MLL* gene family located on chromosome 19q13.1	Human HCC	26 kb upstream	*X*	unknown	Murakami *et al.*, 2005
		4 Human HCCs	Intron 3 Resulting in translocation of *MLL4* gene (t17;19)(p11;q13.1)	1 possible FL HBV; X gene around DR1 incl. promoter	In-frame HBx/MLL4 (introns 4 and 5) fusion proteins sup-pressed 11 genes (*AVIL*;*KERA*;*UBXD1*;*PIAS3*; *MYBPC2*;*PITPNM*;*EHD2*;*GJB1*; *WASL*;*TNRC6C*;*BC1D10B*)*[F] involved in actin-binding and polymerisation, signal transduction, cytokinesis, endocytosis and unknown functions	Saigo *et al.*, 2008

MLL5 Myeloid/lymphoid or mixed lineage leukaemia 5 MLL5 HGNC 18541	Likely indirect transcriptional regulator Member of *MLL* gene family located on chromosome 7q22[7]	Human HCC	98 kb upstream	*X*	Direct effect unknown, but no evidence that HBV-X integration was responsible for HCC in these patients	Toyoda *et al.*, 2008
FOXF2 Forkhead box F2 **FOXF2 HGNC 3810**	Transcriptionally activates several lung-specific genes[8]		39 kb upstream			
ANKH Ankylosis ANKH HGNC 15492	Likely involved in phosphate level control[9]		177 kb upstream			
TIAM1 T-cell lymphoma invasion and metastasis 1 TIAM1 HGNC 11805	Involved in T-cell lymphoma		21.5 kb upstream			
***SPG4* (*SPAST*)** Spastic paraplegia 4 SAPST HGNC 11233	Member of AAA family (ATPases associated with variety cellular activities) Autosomal dominant, spastin[10]		Intron			
DIAPH3 Diaphonous homolog 3 DIAPH3 HGNC 15480	Homolog to *Drosophila* gene		Intron			
FDFT1 Farnesyl-diphosphate farnesyltransferase 1 FDFT1 HGNC 3629	Cholesterol synthesis[11] (also called squalene synthase)		Intron			

ZNF521 Zinc finger protein 521 ZNF521 HGNC 24605	Regulator of diverse immature cells (likely involved in nuclear remodeling, histone deacetylation, trascriptional co-repression)[12]	Human HCC	23 kb upstream	*X*	Direct effect unknown, but no evidence that HBV-X integration was responsible for HCC in these patients	Toyoda *et al.*, 2008
KIAA1181 (ERGIC1) Endoplasmic reticulum-golgi intermediate compartment 32 kDa protein ERGIC1 HGNC 29205	New cycling protein[13]		38 kb downstream			
XPO4 Exportin 4 XPO4 HGNC 17796	Nuclear export[14]		12.6 kb upstream			
DNAJB6 DNAJ (Hsp40) homolog, subfamily B, member 6 DNAJB6 HGNC 14888	Molecular chaperone[15]		4 kb downstream			
A2BP1 Ataxin 2-binding protein 1 A2BP1 HPRD 16091; MIM 605104	Binds ataxin-2[16]		31.9 kb upstream			
STX3A (STX3) Syntaxin 3A STX3 HGNC 11438	Intracellular membrane fusion events/part of SNARE (Soluble N-ethylmaleimide-sensitive factor Attachment protein Receptor) machinery[17]		29 kb downstream			

FLT3 FMS-related tyrosine kinase 3 FLT3 HGNC 3765	Tyrosine kinase	Human HCC	20 kb downstream	*X*	Direct effect unknown, but no evidence that HBV-X integration was responsible for HCC in these patients	Toyoda *et al.*, 2008
MACF1 Microtubule-actin crosslinking factor 1 MACF1 HGNC 13664	Microtubule-actin crosslinking		Intron			
PDE6D Phosphodiesterase 6D, cGMP-specific, rad, delta PDE6D HGNC 8788	Phosphodiesterase	Human HCC	6 kb upstream	*X*	unknown	Murakami *et al.*, 2005
PDE11A Phosphodiesterase 11A PDE11A HGNC 8773	Phosphodiesterase	Human HCC	Intron 17	*X*	Exact effect unknown, possible deregulation of gene expression causing cell death or survival and uncontrolled growth.	Minami *et al.*, 2005
AXIN1 Axis inhibitor 1 AXIN1 HGNC 903	Regulatory protein of Wnt signaling pathway – vertebral dorsal ventral axis formation (possible TS gene)		Intron 2			
SRPK2 Serine protein kinase 2 SRPK2 HGNC 11306	Serine/threonine kinase		Intron 17			
KLF13 Kruppel-like factor 13 KLF13 HGNC 1367	Transcription factor involved in haematopoietic development[18]		Intron 1			
KCNB2 Potassium voltage-gated channel, Shab-related subfamily, member 2 KCNB2 HGNC 6232	Delayed rectifier potassium channel[19]		Intron 5			

BBX *Bobby sox* BBX HGNC 14422	Hypothetical protein gene, homolog of *Drosophila* gene	Human HCC	Intron 3	*X*	Exact effect unknown, possible deregulation of gene expression causing cell death or survival and uncontrolled growth.	Minami *et al*., 2005
AL713702 (FAM19A1) AL713702 HGNC 21587	TAFA family member – likely brain-specific chemokine or neurokine		Intron			
CTNND2 Catenin delta-2 CTNND2 HGNC 2516	Component of adherens junction complex		Intron 20			
PITPNC1 Phosphatidylino-sitol transfer protein, cytoplasmic 1 PITPNC1 HGNC 21045	Phosphatidylinositol transfer from one membrane compartment to another (NM_012417)		Intron 14			
EYA3 Eyes absent 3 EYA HGNC 3521	Homolog of *Drosophila* gene, phosphatase activity in man		Intron 14			
BC026191 (ENDOD1) Endonuclease domain-containing 1 BC026191 HGNC 29129	Protein coding gene[20] (AB385387)		Intron 1			
ODZ2 Homolog of odd Oz 2 *Drosophila* gene ODZ2 HGNC 29943	Possible transcriptional regulator in man		Intron 7			
LHFPL3 Lipoma HMGIC fusion partner-like 3 LHFPL3 HGNC 6589	Tetraspan transmembrane protein[21]		8.6 kb downstream			

TTC30A Tetratricopeptide repeat domain 30A TTC30A HGNC 25853	Protein coding gene[23] (NM_152275)	Human HCC	150 kb upstream	*X*	Exact effect unknown, possible deregulation of gene expression causing cell death or survival and uncontrolled growth.	Minami *et al*., 2005
STAC SH3 and cystein rich domain STAC HGNC 11353	Possible protein coding gene (AK313493)		114 kb upstream			
TERF2 Telomeric repeat-binding factor 2 TERF2 HGNC 11729	Component of vertebrate telomeres; protective role in cellular senescence		168 kb upstream			
OCIA (OCIAD1) OCIA domain-containing 1 OCIAD1 HGNC 16074	Novel protein gene (NM_001079842)		6 kb upstream			
PARK7 Parkinson disease PARK7 HGNC 16369	Positive regulator of androgen receptor-dependent transcription		22 kb downstream			
MIG6 Mitogen-inducible gene 6 or (ERRFI1) ERBB receptor feedback inhibitor 1 ERBB HGNC 18185	Cytoplasmic protein upregulated with cell growth[24]		4 kb downstream			
CDC42EP5 CDC42 effector protein (Rho GTPase-binding) 5 CDC42 HGNC 17408	Regulates formation of F-actin-containing structures through interaction with downstream effector proteins[25]		1.5 kb upstream			

LAIR2 Leukocyte-associated immunoglobulin-like receptor 2 LAIR2 HGNC 6478	Unknown function	Human HCC	28 kb upstream	*X*	Exact effect unknown, possible deregulation of gene expression causing cell death or survival and uncontrolled growth.	Minami *et al*., 2005
FALZ (*BPTF*) Bromodomain PHD finger transcription factor FALZ HGNC 3581	Possible transcription regulator		148 kb upstream			
SRP46 (*SFRS2B*) Splicing factor, arginine/serine rich 2B SFRS2B HGNC 16988	Essential splicing factor[26]		28 kb downstream			
SEST3 (*SESN3*) Sestrin 3 SESN3 HGNC 23060	Antioxidant modulator of peroxiredoxins[27]		75 kb downstream			
SOCS4 Suppressor of cytokine signaling 4 SOCS4 HGNC 19392	Likely cytokine-inducible negative regulator of cytokine-signalling[28]		61 kb downstream			

References:

1 Chomyn *et al*., 1985
2 Roberts *et al*., 1972; Maniatis *et al*., 1981
3 Cunningham *et al*., 1996; Hattori *et al*., 1985
4 Cully *et al*., 2005
5 Williams *et al*., 1961
6 Beuren *et al*., 1962
7 Deng *et al*., 2004
8 Mahlapuu *et al*., 1998
9 Zaka *et al*., 2006
10 Erzberger and Berger, 2006
11 Fünfschilling *et al*., 2007
12 Bond *et al*., 2008
13 Breuza *et al*., 2004
14 Lipowsky *et al*., 2000
15 Qiu *et al*., 2006
16 Kiehl *et al*., 2000
17 Sharma *et al*., 2006
18 Outram *et al*., 2008
19 Vasan *et al*., 2007
20 Quintana *et al*., 2008
21 Kalay *et al*., 2006
22 Zody *et al*., 2006
23 Ota *et al*., 2004
24 Wick *et al*., 1995
25 Hirsch *et al*., 2001
26 Soret *et al*., 1998
27 Kopnin *et al*., 2007
28 Bullock *et al*., 2007
29 Murakami *et al*., 2005
30 Paterlini-Bréchot *et al*.,2003
31 Gozuacik *et al*., 2001

PBMCs – peripheral blood mononuclear cells

NOTES:

*Where possible the NCBI nucleotide database primary source number has been given below the name of each gene. If the official symbol for a particular gene differs on the NCBI site to that given in the original publication (listed in the last column) the one given on the NCBI site is listed alongside in brackets.
http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccoreanditool=toolbar Accessed 15 July 2008.

*[A-E]Murakami *et al*., 2005 (Ref. 29) documents HBV integration into or close to 42 cellular genes, 10 of which they had previously reported (Paterlini-Bréchot *et al*., 2003 [Ref. 30]; Gozuacik *et al*., 2001 [Ref. 31]). The genes are listed in ref. 29 in Table II, p1165, and on a supplementary table on the Web (address below) and when reporting more than one integration event (eg. *hTERT* gene) the gene is listed in their tables more than once for every event. There appear to be several discrepancies, between Ref. 29, and the previously published data in Ref. 30 and Ref. 31. The authors do not make clear whether these discrepancies derive from a difference in the interpretation of their results or if they are in fact referring to separate integration events. *[A]The *IRAK2* gene is reported to have an integrant 6 kb upstream of the start codon (Ref 30) and an integrant downstream of exon 3 (Ref 29). *[B]The *p42MAPK1* gene is reported to have an integrant 20 kb upstream of the start codon (Ref 30) and then 22 kb downstream of the gene (Ref 29). *[C]The *IP3R2* gene has an integrant 45 kb upstream of the start codon (Ref. 30) and then 44 kb downstream of the gene (Ref. 29). *[D]The *ST3GALVI* has an integrant 3 kb upstream of the start codon (Ref. 30) and then 4 kb downstream of the gene (Ref 31). *[E]Ref. 29 reports 3 separate integration events in the *hTERT* gene (10.8 kb upstream of the *hTERT* promoter; 0.3 kb downstream of the *hTERT* gene; 16 kb downstream of the gene). However, there may be a fourth such event, reported in Ref. 30, of an integration 5.2 kb upstream of the gene. Data in the table above is reported as given in the references listed in the far right-hand column, and in Murakami *et al*., (2005), supplementary table at: http://www.gutjnl.com/supplemental

*[F]Advillin; Keratocan; UBX domain-containing 1; Protein inhibitor of activated STAT3; Fast-type myosin-binding protein C; Phosphatidylinositol-transfer protein membrane-associated; EH-domain-containing 2; Connexin 32; Wiskott-Aldrich syndrome-like; Trinucleotide repeat-containing 6c; TBC1 domain family, member 10B

*[G] Murakami *et al.*, 2005 report *MLL2* and *MLL4* as two different genes. The NCBI nucleotide database records them as the same gene (15 July 2008).

This led them to propose a mechanism of HCC development:

(1) HBV integrates into a cellular gene during the acute phase of viral infection, resulting in gene changes;
(2) viral/cellular gene oncogenic activity possibly provides the cell with selective growth advantage (over viral propagation) during chronic phase of infection;
(3) increasing accumulation of genetic changes during liver cell proliferation may result in HCC (Murakami *et al*., 2005).

Other known sites of integration include chromosome 1p36 (Simon and Carr, 1995) chromosomes 3 (Dejean *et al*., 1986) 6 (Hino *et al*., 1986), 15q22-q23 (Bowcock *et al*., 1985) and 18 (Bowcock *et al*., 1985; Hino *et al*., 1986).

HBV integrants in human HCC are often found in repetitive DNA sequences (Ogston *et al*., 1982; Nagaya *et al*., 1987; Shih *et al*., 1987, 1988; Ogata *et al*., 1990; Quade *et al*., 1992; Tsuei *et al*., 1994) (eg. minisatellite sequences), which are found associated with cellular genes or gene clusters (Dover, 1989) and are known to be involved in activation and repression of transcription (Takeda *et al*., 1989; Trepicchio and Krontiris, 1993) and the binding of regulatory proteins (Trepicchio and Krontiris, 1993). Although certain disorders and cancers are known to be caused by mutation of minisatellite repeat sequences (Novelletto *et al*., 1994; Lim *et al*., 2006; Baca-Garcia *et al*., 2007; Vairaktaris *et al*., 2007; Buisine *et al*., 2008; Maruta *et al*., 2008), it has yet to be demonstrated that a mutation in a repetitive sequence, generated by the integration of HBV DNA, has resulted in the incorrect regulation of a cellular gene (or genes) with a subsequent oncogenic outcome (Robinson, 1994). However, expansion/contraction of microsatellite sequences at two or more loci (microsatellite instability) (Kazachkov *et al*., 1998) is considered indicative of genomic instability and replication errors (Gao *et al*., 1994), and may be important in a subset of HCCs (Sheu *et al*., 1999).

Integration may occur more often in repeating sequences such as *Alu* repeats, satellite III sequences, α-satellite DNA and minisatellites or variable nucleotide tandem repeats VNTRs because their GC content is very similar to that of HBV DNA (Ogston *et al*., 1982; Shaul *et al*., 1984; Yaginuma *et al*., 1985; Nagaya *et al*., 1987; Shih *et al*., 1987, 1988; Ogata *et al*., 1990; Chen *et al*., 1994). Almost all integrants in the PLC/PRF/5 cell line are located in GC

rich isochors (Zerial *et al.* 1986). Such sequences have a 52 % GC level which is very close to the 49 % GC content of HBV DNA (Ono *et al*., 1983).

1.5.1.2 Transactivation

The two HBV genes most commonly integrated in HCCs, and tumour-derived cell lines are *X* and *S*, which have transactivating activity. The *X* region is more commonly integrated than *S* (Table II) and neither gene integrates in its entirety. In *X* gene integrants, the viral junction is often found near or within the *X* open reading frame, close to one cohesive end of the viral DNA, resulting in the interrupted or most often 3' truncated form of the gene (Wollersheim *et al*., 1988; Chen *et al*, 2000), while retaining its transactivating function (Balsano *et al*., 1991). Moreover, the integration and truncation of a region of the *preS2/S* sequence bestows transcriptional activation function upon the resulting truncated middle HBV antigen ($MHBs^t$) polypeptide (Caselmann *et al*., 1990; Kekulé *et al*., 1990).

Some of these truncated HBx and preS2/S proteins, as well as the viral/host hybrid proteins, are expressed in HCCs (Table II) and may induce expression of oncogenes or dormant genes, or silencing of TS genes. In one study the *preS* region of HBV was found integrated in-frame with the retinoic acid receptor beta (*RAR-β*) gene, but no expression of a fusion/hybrid protein was confirmed (Dejean *et al*., 1986) (Table II). In a HCC from a Chinese nine-year boy a partially deleted and rearranged integrant was found containing both *X* and *S* regions oppositely orientated. The *X* gene was c-terminally truncated, but retained transactivating activity (Tsuei *et al*., 1994). In Kimbi *et al*., (2005), a partially deleted and rearranged integrant is reported in the *WBSCR1* gene (Table II). Expressed truncated HBx has also been shown in the G1, G2, G3 and G5 phases of the cell cycle, suggesting a further possible involvement in the pathogenesis of HCC (Chen *et al*., 2000). An interesting observation made by Nagaya *et al*., (1987), is that when the upstream sequence of the *X* gene suffers deletion, the viral enhancer is brought closer to the virus-cell junction, increasing the possibility of its effect upon a cellular gene.

Other regions of the HBV genome may be integrated as often as the *X* and *S* genes but are not as readily detected, because they are more prone to deletion or gross mutation/rearrangement, or because cells containing regions such as the core and polymerase gene (whose proteins

show intrinsic toxicity) have little chance of survival and proliferation (Ziemer *et al*., 1985; Chen *et al*., 1988, 2000). The majority of primary tumours investigated by Diamantis *et al.*, (1992), and Paterlini *et al.*, (1995), were found to contain integrated *X* gene regions that expressed mRNA. However almost none of them showed transcription of other integrated HBV genes. Furthermore in the PLC/PRF/5 cell line, unlike in HBV virions or HBV infected tissues, there is stable methylation of some integrants and definite transcription from those integrated HBV sequences where little or no methylation is present, such as the *S* gene integrants. However, the integrated core gene sequences are methylated almost in their entirety, and no production of HBcAg occurs (Miller and Robinson, 1983).

1.5.1.3 Amplification of Cellular Oncogenes

Amplification and increased transcription of the c-*myc* gene is documented in ground squirrels without mutation of, or viral integration within, the gene (Transy *et al*., 1992). In the woodchuck this appears to be entirely absent, but it is documented in man, albeit rarely (Trowbridge *et al*., 1988; Transy *et al*., 1992).

1.5.1.4 Alteration of TS genes

LOH by deletion has been frequently detected on chromosomes 1p, 4q, 5q, 8p, 10q; 11p, 13q, 16q, 17p, 22q in HCCs (Rogler *et al*., 1985; Wang and Rogler, 1988; Buetow *et al*., 1989; Fujimori *et al*., 1991; Simon *et al*., 1991; Emi *et al*., 1992; Nose *et al*., 1993; Takahashi *et al*., 1993; Yeh *et al*., 1994; Kuroki *et al*., 1995; Yumoto *et al*., 1995; Nagai *et al*., 1997). HBV DNA integrated at chromosome 11 p13 results in a 12 kb deletion in a HCC (Rogler *et al*., 1985). The involvement of tumour-suppressor genes at sites of LOH seems likely. In particular, the retinoblastoma (*RB1*) gene (Harbour *et al*., 1988; Lee *et al*., 1988; Bookstein *et al*., 1990; Ashida *et al*., 1997) the Wilm's tumour (*WT1*) gene (Gessler *et al*., 1992; Piao *et al*., 1997b), the adenomatosis polyposis coli (*APC*) gene (Piao *et al*., 1997b); the *E-cadherin* gene (Tsuda *et al*., 1990; Vleminckx *et al*., 1991; Sheu *et al*., 1999) and the breast cancer-related gene 2 (*BRCA2*) (Katagiri *et al*., 1996) contain mutations or deletions in human carcinomas. In poorly differentiated HCCs cumulative LOH has been frequently reported (Piao *et al*., 1997a and b), possibly reflecting the multi-step nature of cancer progression (Tamura *et al*., 1997), although it remains to be shown whether some or all of these deletions contribute to the

initiation or development of the actual tumour.

p53 gene mutations are of the most common found in human HCC and occur in 30-60 % of tumours and in some HCC cell lines (Robinson, 1994; Puisieux and Ozturk, 1997). These changes contain deletions, LOH rearrangements, point mutations, and altered gene expression without loss of an allele. In some cases the deleted DNA is as a result of a HBV integration event (Hino *et al*., 1986; Slagle *et al*., 1991) although, loss or deletion of the *p53* gene is not more common in HCCs with HBV integrants than in those without (Robinson, 1994).

1.5.1.5 Other

Integration events exist that cannot be easily slotted into one of the four categories as designated by Robinson. HBV integrants are unstable, with the initial integrated HBV DNA sometimes suffering post-integration mutation/rearrangement, often involving the cellular flanking sequences (Nagaya *et al*., 1987; Simon and Carr, 1995). Such rearrangements are documented in HCC progression in chronically infected patients and integrants have also been found in non-tumorous tissue adjoining HCCs (Koshy *et al*., 1981; Shafritz *et al*., 1981; Hino *et al*., 1984). Integrants may even be lost altogether, again possibly involving the cellular flanking sequences (Hino *et al*., 1986; Slagle *et al*., 1991). This could account for those HCCs where no detectable integrants exist (~10-15 % of HBV-related HCCs), although this is not proven. As mutagenic events bring the liver ever closer to HCC, chromosomal instability increases (Moradpour and Blum, 2005).

Speculation that integrants behave as "hit and run mutagens", where the original HBV integrant may translocate to another region of the genome (Hino *et al*., 1986; Möröy *et al*., 1986) taking a portion of the cellular DNA along, is substantiated by the findings of integrants flanked by different chromosomes; such as t(17;18) (Hino *et al*., 1986), t(X;17) (Tokino *et al*., 1987), t(7;17) (Meyer *et al*., 1992), t(3;8) (Pineau *et al*., 1996) and t(17;19) (Saigo *et al*., 2008). One of the integrant's flanks reported in Pineau *et al*., (1996), is located at 8p23, in the vicinity of the *carboxypeptidase N* gene, whose the product is present in large amounts in normal liver, but known to be regularly deleted in HCCs. The second flank, at 3q27-29 has been documented in different types of cancers involving a number of types of translocations and is known to be involved in the control of cellular growth. In this case it was possible that a

proto-oncogene may have been activated or a hybrid product produced. In either case, a growth advantage may have arisen in the clonal expansion, resulting from the fusion of the heavily liver-transcribed carboxypeptidase N gene with the cell growth control region (Pineau *et al*., 1996).

In a human HCC investigated by Hino *et al*., (1986), a translocation between chromosomes 17 and 18 is documented. It is postulated that this resulted from mediation by two HBV DNA sequences integrated within each of the two chromosomes. A section of approximately 1.3.kb of cellular DNA appears to have been lost from chromosome 18, but any direct oncogenic outcomes of the mutation are not evident.

In the PLC/PRF/5 cell line there are reports of amplification and transposition of integrated HBV along with its flanking cellular sequences (Koch *et al.*, 1984b; Zeimer *et al.*, 1985). Evidence indicates the duplication, transposition, fragmentation, rearrangement and subsequent divergence of integrants even after integration (Ziemer *et al*., 1985). It is not established however, whether the transposition occurred in the primary tumour or while the cell line was being established.

1.5.1.6 Indirect Influence of HBV on HCC

There exists a line of thought that a significant number of HBV-related HCCs arise as a result of persistent injury to the liver and the consequences of chronic viral infection (eg. inflammation), rather than from a precise viral mechanism (Chisari and Ferrari, 1995 a and b). This can be inferred from the fact that in about 15 % of HCCs, the HBV DNA is at too low a level to be detected by SH and many HBV infected individuals show no hepatic dysfunction. Even in those with acute hepatitis B, symptoms are observed a number of weeks after high-level viremia. Furthermore, transgenic mice containing HBV S region sequences develop HCC only if substantial or moderate liver injury is concurrently observed (Cerutti, 1988; Chisari, 1988). Furthermore, the virus is not directly cytopathic and hepatocellular necrosis, inflammation and regeneration occur as a result of the hosts own immune response. It was originally thought that cytotoxic T-lymphocytes (CTLs) are responsible for viral clearance through the targeting and killing of infected hepatocytes, thereby leading to necrosis (Chisari, 1997). This is however not the complete picture. It has been shown in the transgenic mouse

model that all infected hepatocytes need not be eliminated to ensure viral clearance (Guidotti *et al*., 1994), but rather a cytokine-mediated mechanism, working at the posttranscriptional level, can prevent HBV replication and gene expression. The subsequent substantial decrease in viral antigens would lead to a diminished host immune response, allowing some HBV particles to persist. These would gradually resume active infection thereby arousing a heightened immune response once again. In HBV chronically infected humans, this could explain the cycles of active disease broken by periods of quiescence (Hodinka, 1999). HBV may also be indirectly involved through the generation of oxidants during hepatic inflammation caused by chronic hepatitis (Moraes *et al*., 1989, 1990; Hodinka, 1999).

1.5.2 HCC and Clonality

In cirrhotic liver, HCCs develop in adenomatous foci produced within regenerative nodules of liver cells (Robinson, 1994). On scrutinising these nodules investigators found one or more integrant clones, each with its own singular pattern of integration (Aoki and Robinson, 1989) indicating selective amplification of a single hepatocyte with integrated virus (Rogler and Summers, 1984).

Clones may differ between non-tumorous and tumorous liver tissues of the same individual. Several clones may develop concurrently in a HCC, or a number may occur individually over time (Sheu *et al*., 1993; Matsumoto *et al*., 2001). This is logical, as it has been documented (section 1.4) that no specific integration event or series of events in particular cause HCC. Integration is therefore primarily random (sections 1.5 and 1.5.1.1) and as such the integration events in a single cell will not mirror those in a different cell. Each cell may develop into a clone. In each clone, the resulting genomic changes, caused or influenced by the initial integration event and subsequent integration events, will result in different repercussions at the cellular level. Those cells that respond well to the selection pressures already present will thrive, whereas those with lethal mutation/s will die – and be 'removed' from the 'clone population'. One can therefore surmise that a clone in non-tumour tissue is simply non-tumorous because the integrant event/s it contains have neither resulted in lethal repercussions, nor have they crossed over the nebulous boundary of cumulative mutation that will push them into HCC. However, this may simply be a matter of time. This is bourne out by the generally accepted view that early stage HCCs are polyclonal in origin, (Sheu *et al*., 1993)

whereas late stage tumours are primarily monoclonal, although there have been reports of the occasional polyclonal advanced stage HCC (Esumi *et al*., 1989). However, intrahepatic spread, and metastasis in late stage tumours may erroneously indicate monoclonality (Blum *et al*., 1987; Sheu *et al*., 1993). In this case the tumorous mass has progressed to the metastatic phase. Various tumour cells will have broken away from the original clone and settled at various sites within the same liver, generating more than one clone with the same integration pattern. Any other clones originally present may have been lost over time or may not have responded as well to selection pressure, in which case they may have been engulfed and obscured by a faster growing clone-mass nearby. Identifying them may not be possible as the two clones would have to be differentiated by microscopy and shown by Southern Hybridisation to have different integration profiles. Monoclonal HCCs rarely contain one integrant (Esumi *et al*., 1989). Usually three to four integrants are present, although as many as ten or more have been reported, with each event being at a different site in the host genome and having occurred independently. Nonetheless, HCCs containing large numbers of integrants are rare. This is indicative of a possible mechanism of counter-selection, operating in hepatocytes with too many integration events (Matsubara and Tokino, 1990).

The old debate as to whether multiple integration events occur during hepatocyte progression to HCC or whether this is restricted to the period of active viral replication only, still endures (Matsubara and Tokino, 1990). Whether or not HBV replication is supported in HCCs also remains a matter of some controversy. Thus far integrants studied have been unable to support viral replication because of mutation but HCCs may contain replicating episomal HBV. A HCC in a study by Chen *et al*., (1982), contains both integrated and extrachromosomal DNA. Simon and Carr, (1995), document viral replication in a non-tumorous tissue, but not in its associated primary HCC. Nazarewicz *et al*., (1977), propose that replication is absent in poorly differentiated HCC (ie. late stage tumours) but can be present in early carcinomas.

At times multicentric occurrence of HCC occurs in patients with primary liver disease caused by HBV. Multicentric HBV-related HCC can be diagnosed by making use of the integration ‘pattern’ (clonality) of HBV within a tumour (Matsumoto *et al*., 2001). The generation of many clones with different integration patterns, within and between individuals, bears testimony to the fact that HBV integration may contribute to HCC development via a number of different mechanisms interfering with growth and metabolic pathways within the cell.

Clinically, different clones may show varied metastatic abilities and responses to chemotherapy regimes (Esumi *et al*., 1989).

1.5.3 Integration in the PLC/PRF/5 cell line

The PLC/PRF5 cell line is possibly the most studied hepatoma cell line. Other than the present PhD study and the one by Kimbi *et al*., (2005), studies on the PLC/PRF/5 cell line comprise the only other work where integrants from a southern African black individual are characterised. The PLC/PRF/5 cell line currently contains nine integrants per cell (Zerial *et al*., 1986), none comprising the FL genome of the virus, but rather rearranged fragments (Alexander *et al*., 1976; Edman *et al*., 1980; Koshy *et al*., 1983; Shaul *et al*., 1984; Ziemer *et al*., 1985). They express the HBV S protein (Alexander *et al*., 1976) from two integrated HBV S gene sequences which contain TATA like box and S gene promoter regions, but lack the HBV poly-adenylation signal. HBsAg must therefore derive from a hybrid mRNA resulting from integrants positioned in-frame with a cellular poly-adenylation signal (Koch *et al*., 1984a; von Loringhoven *et al*., 1985; Ziemer *et al*.,1985). It is not known whether the integrated HBV DNA was originally expressed in the host liver. Other virus-host chimeric transcripts have been documented in PLC/PRF/5 (Chakraborty *et al*., 1980; Ou and Rutter, 1985) but their function remains unknown and no certain proof exists that these were categorically involved in transformation. The PLC/PRF/5 integrants appear to derive from four different *adw* (Ziemer *et al*., 1985) strains, a subtype prevalent in Mozambique (Courouce-Pauty *et al*., 1983).

In 1980, Edman *et al*., showed PLC/PRF/5 cell line integrants with junctions in 1400-2600 bp of the viral genome. Koch *et al*., (1984a), in turn analysed four integrants. All were integrated at the single stranded gap region of HBV but integration into the cellular genome was random. These findings provided support for a model of integration (section 1.5.4).

1.5.4 Types of HBV Integrants and Models of Integration

As initial studies of HBV integrants progressed and became more numerous, the information they yielded spawned various models of integration (Dejean *et al*., 1983; Koike *et al*., 1983; Dejean *et al*., 1984; Koch *et al*., 1984b; Shaul *et al*., 1984; Miller *et al*., 1985; Mizusawa *et*

al., 1985; Yaginuma *et al*., 1985; Ziemer *et al*., 1985; Hino *et al*., 1986; Nagaya *et al*., 1987; Shih *et al*., 1987) some of which are conflicting and most of which have since been discounted. The prevailing models are discussed here. There is no one proposed mechanism of integration that can explain every type of integrant and models are not necessarily mutually exclusive.

Integrants can be thought of as falling into one of two groups: those integrants that have one or both of their end junctions, lying within the cohesive end region, and those that have neither end within this region. The majority of integrants in HBV-related human HCCs (about 67 %), have one or both of the virus cell junctions at DR1 or DR2 or within the viral cohesive end region (Dejean *et al*., 1984; Koch *et al*., 1984a; Mizusawa *et al*., 1985; Yaginuma *et al*., 1985; Ziemer *et al*., 1985; Hino *et al*., 1986; Nagaya *et al*., 1987; Shih *et al*., 1987). These cohesive end region integrants (termed Coh) have been grouped into four main types (I, II, III and IV) by Shih *et al*., (1987), based on their strand polarity and end specificity (Figure 1.7).

- *Type I*, the second most common type, integrate into the host's DNA in the core-surface-X gene orientation, with DR1 sequences at the fixed end (Figure 1.7.I). Shih *et al*., (1987), proposed a model by which such integrants might arise (section 1.5.4.2).

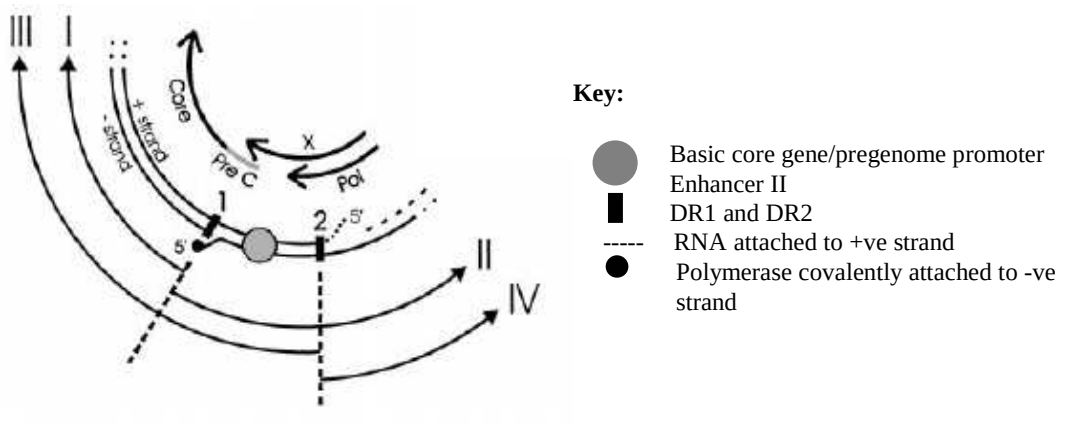


Figure 1.7: Diagrammatic representation of the four types of Coh Integrants (I, II, III, IV) often found in HBV related HCC.

Reproduced from: Arbuthnot and Kew, 2001.

- *Type II*, the most common type (46.6 %), integrates in the *X-surface-core* gene orientation, with the cohesive end region or DR1 repeat sequence present before the *X* gene (Figure 1.7.II) (Nagaya *et al*., 1987). Such integrants are explained by the models of Yaginuma *et al*., (1985), and the relaxed circle where cellular DNA invades into the single-stranded gap region (section 1.5.4.2). Thereafter, displacement synthesis of the cohesive overlap region occurs.

- *Type III* integrates into the host's DNA in the *core-surface-X* gene orientation, but has the cohesive end region or DR2 repeat sequence present before the core gene (Figure 1.7.III).

- *Type IV* is the same as type II, but the DR2 sequence is present at the fixed end (Figure 1.7.IV). Such integrants are rare and can be explained by the relaxed circle model (section 1.5.4.2) (Shih *et al*., 1987). Of the four types, I and II are three times more common than III and IV (Arbuthnot and Kew, 2001).

1.5.4.1 Intermediates of Replication as Substrates of Integration

Currently the models of strand invasion and recombination between viral intermediates of replication and cellular DNA prevail in explaining Coh integrants. There are two reasons for supposing that the most likely substrates of integration are viral replication intermediates (Nagaya *et al*., 1987; Shih *et al*., 1987). Firstly, the fact that a large proportion of integrants are of the Coh type and that synthesis of both strands of HBV initiates at the DR sequences, which are also required for the switching of the template strand during pgRNA reverse transcription. Secondly, that in Coh type integrants, one of the ends of the integrated HBV DNA molecule lies within the cohesive end region, but the other end lies at a variable position within the HBV genome. This variable end can be explained if the pre-integration substrates are replication intermediates of different sizes. The variable end may also result from a post-integration rearrangement, but this requires a secondary event to the integration itself and integrants exist where little or no rearrangement has occurred, and the second end of the integrated DNA is variable (Yaginuma *et al*., 1985; Dejean *et al*., 1986).

Very little is known about the mechanism of illegitimate recombination but in numerous integrants the cellular and viral sequence at or near the integrant junctions show no evidence of overall homologies. This is indicative of illegitimate recombination, usually associated with free DNA ends, which enhance the process 10 to 100-fold (Folger *et al*., 1984) and can be involved in strand invasion (Cunningham *et al*., 1979). Furthermore, illegitimate recombination in eukaryotes appears to frequently result in the non-uniform deletion of cellular DNA (Henderson, 1987), as is often seen in integrants and suggests an integration mechanism that does not involve a specific integrase. Single stranded linear replicative intermediate HBV DNA is thought to initiate the recombination with host DNA. The minus strand is most commonly the one that invades, and if invading from the 5' end, type II integrants occur (Nagaya *et al*., 1987). If invasion occurs from the 5' end of plus strand, a type III integrant results. Theoretically, invasion by the 3' end of both positive and negative HBV strands would result in Coh integrants of types I (Nagaya *et al*., 1987) and IV, respectively. A model of single-stranded positive strand HBV DNA integration remains equivocal, as such DNA has never been observed in the hepatocyte nucleus (Shih *et al*., 1987).

1.5.4.2 The Relaxed Circle

In 1983, Koshy *et al*., proposed a model of integration based upon certain sequence homologies between the viral DNA and flanking genomic DNA. In actively replicating hepatocytes the HBV genome may align with the cellular genome at such homologous sites and once the replication fork approaches, the DNA polymerase switches from the genomic template to the single stranded gap of the viral one, joining the two. Recombination follows, resulting in an integrant. This mechanism explains type IV integrants.

The triple-stranded structure and circular roll-in model of Shih *et al*., (1987), is similar to the relaxed circle model and explains type I integrants. It proposes that the terminal redundancy resulting in the triple stranded region of the relaxed circular form of HBV DNA, found in the nucleus of infected hepatocytes (Miller and Robinson, 1984) may play a crucial part in the process of integration.

However, the model makes provision for the fact that the relaxed circular form may not be essential, or enough for successful integration. If viral replication in the hepatocyte is

somehow hindered, it is likely that integration rather than replication will occur (Shih *et al*., 1987).

1.5.4.3 Other Possible Mechanisms and Substrates of Integration

Hino *et al*., (1989), suggested that integrants not generated from replication intermediates may result from topoisomerase I cleavage of cccDNA around DR1, with the resulting products undergoing integration. This profuse cellular enzyme may target the cohesive terminus of HBV because its DNA sequence is one of its preferred cleavage substrates. The idea is supported by *in vitro* studies of WHV integration mediated by this enzyme (Wang and Rogler, 1991). Normally involved in DNA transcription and replication, topoisomerase I, cleaves a single strand of the DNA duplex thereby relieving the superhelical tension that arises during these processes (Wang, 1985). Cleavage of the partially double-stranded HBV genome in the cohesive overlap linearises the viral DNA. The exposed ends of the linear molecule become available targets for host nucleases, and are 'nibbled at'. The HBV DNA is now linear and capable of ligating to the host genomic DNA. Integrants are documented with microdeletions at their virus-cell genome junctions (Dejean *et al*., 1984), lending further support to this hypothesis (Wang and Rogler 1991).

The complex structure of many integrants initially led researchers to conclude that these had suffered extensive mutagenic events post integration. However, in 1982, Rogler and Summers found free HBV rearranged genomes with structures similar to those of complex integrants, which they named ‘novel form molecules’ (Rogler and Summers, 1982).

1.6 Rationale/Justification for This Study

HBV-related HCC is prevalent in areas endemic for the virus, including sub-Saharan Africa where, numerous individuals die of this cancer annually, and the manifold causes involved in the initiation and progression of HCC are still not fully understood (Kew *et al*., 1987; Günther *et al*., 1998; Kew *et al*., 2002a; Parkin *et al*., 2008). There is no known cure or cost-effective treatment regimen (Gong *et al*., 1999) (section 1.2). Vaccination is still logistically problematic and futile for those individuals already infected with the virus (Van Herck and Van Damme, 2008). There exists a very real need for further study of HBV, the molecular

mechanisms behind its mode of replication; relationship between viral variants and course of the disease; the oncogenic potential of integrated HBV and the development of new methods with which to effectively combat chronic HBV infection (Nassal & Schaller, 1996). Furthermore, in studying integration sites, as yet unidentified cellular genes involved in cellular differentiation and proliferation pathways may be discovered. Studies on the PLC/PRF/5 cell line comprise one line of work where a few integrants from a southern African black individual are characterised, however, it is highly likely that integration events in a cell line, under controlled laboratory conditions, will not mimic exactly the repercussions of integration events in individuals. Most studies on HCCs (Table II) have been primarily concerned with tissue from Europe, China and Japan. Whether the integration events in these populations are the same, similar or distinct to those operating in the southern African populations is unknown. Other than the present PhD study only one integrant from a southern African black individual has been partially characterised. Further research is therefore important in order to gain improved insights into the process of HBV-related HCC. Results could allow for a comparison to be made against previously published data and thereby contribute to an understanding of the molecular aetiology of HCC, at least in the context of the southern African populations.

All the carcinomas in this study were late stage. Studying HCC in southern Africa has several logistical problems, the main one being that the majority of HCC tumours are seen in individuals from rural black populations. The work done here was performed on tissue from one of the few available HCC archives in Africa.

The objective of this study was therefore: to characterise the pattern of HBV integration in tumour and non-tumorous liver obtained at autopsy or laparotomy from southern African blacks with HCC; and the aims were to:

1. determine the proportion of tumours with integration and single or multiple integrants
2. ascertain the sites of HBV integration relative to functional cellular genes
3. establish whether any integrants were expressed in the tumours

In short, research is required to establish the exact manner in which HBV integration contributes to the aetiology of HCC in the southern African populations, especially as it becomes ever more evident that no single mechanism may be the sole culprit in most cases, but that the processes are likely to be numerous and divergent.

1.6.1 Thesis Structure

The chapters of this thesis have been structured so as to divide the experimental work into parts, allowing for greater clarity. There is no standard method for the detection of integrants. Previous to the initiation of this study most of the work done on HBV integrants had been performed by screening of genomic libraries constructed from the DNA of individual patients. This technique, although excellent in terms of the data it can yield, is both laborious and time-consuming. Large amounts of tissue are usually required and few integrants are characterised at any one time. In HBV-related HCCs it can also prove frustrating because of the presence of coexisting non-integrated free virus, which can confound the identification of genuine integrated viral sequences.

The development of laboratory techniques, which promised the speedy processing of large numbers of tumours from small amounts of material, seemed an auspicious approach to the problem. However, these techniques had never been applied to African HCCs, which, contrary to most HCCs found elsewhere, contain multiple integration events.

Chapter 2 of this thesis, details the optimisation of various techniques applied in the attempt to determine the proportion of tumours with integration and single or multiple integrants, and to ascertain the sites of HBV integration relative to functional cellular genes (aims 1 and 2 section 1.6); while chapter 3 reports the results obtained for aims 1 and 2. Chapter 4 details results obtained for aim number 3: determining whether any integrants are expressed in the tumours.

1.6.2 Outline of Materials and Methods Employed

The structure of this thesis is unorthodox but has been employed to facilitate the understanding of work performed in characterising integrants from southern African HCCs. As various

methods were used, some of which were abandoned, and the work is divided into several sections, a figure is supplied below (Figure 1.8) to facilitate the understanding of the process.

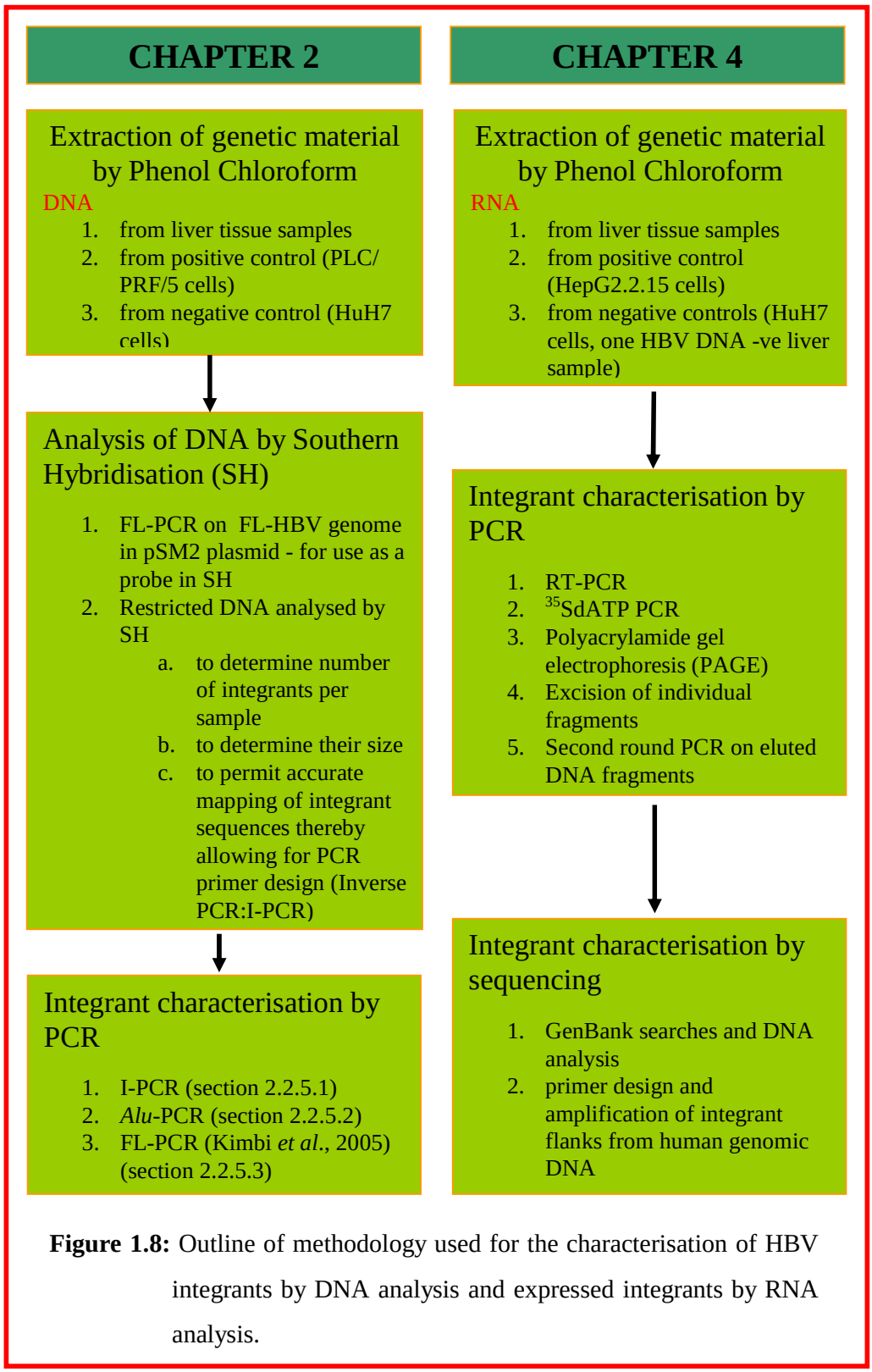


Figure 1.8: Outline of methodology used for the characterisation of HBV integrants by DNA analysis and expressed integrants by RNA analysis.

2.0 OPTIMISATION OF LABORATORY TECHNIQUES FOR THE DETECTION AND CHARACTERISATION OF INTEGRANTS (DNA ANALYSIS)

2.1 Introduction

When we initiated this study, the most common method for determining the number, clonality and total percentage integration in HCCs was Southern blotting and hybridisation (SH), which had been used since the early 1980s. Furthermore, the small number of integrants characterised had been primarily investigated by cloning, and with the use of DNA libraries. Investigations in man, cell lines and animals models, have shown that integration could include deletions of cellular DNA, LOH (often involving TS genes), translocations, rearrangements, transposition, fragmentation, amplification of cellular oncogenes, and point mutations and gene mutations resulting in oncogene activation, TS gene inhibition, or post-integration mutation/rearrangement and aberrant expression of cellular genes and HBV/host hybrid proteins (section 1.5 & Table II). Integration was known to occur often in repeating sequences, with the *S* and *X* genes being the most commonly integrated and the DRs a hot-spot for integration events (sections 1.5.1.1 & 1.5.1.2 & 1.5).

In order to fully characterise integrants in the southern African population (aims 1 and 2 section 1.6) certain facts had to be established beforehand. These included:

(1) the percentage of HCCs containing integrants, so that an adequate number of samples could be investigated to allow for the greatest number of integrants to be characterised
(2) the samples which contained integrants, and which did not (these would be excluded), and of those that did, how many integrants were present in each tissue
(3) the size of individual integrants since the integrant size would affect the feasibility of successful PCR amplification.
(4) a map of each integrant in order to obtain information for primer design enabling further characterisation by PCR (section 2.2.5.2; Appendix A, A15-A17).

These requirements were met by restriction enzyme (RE) digestion of genomic DNA and analysis by SH (sections 2.2.4.1 & 2.2.4.2-2.2.4.3). Further characterisation of integration events (ascertaining the sites of HBV integration relative to functional cellular genes - aim number 2 section 1.6), was also attempted by Full Length (FL) PCR (Günther *et al.*, 1998). This PCR was not originally designed to amplify HBV integrants, but rather as a method of amplifying the complete HBV genome in one PCR reaction from a HBV template DNA. However, this PCR had previously been used in the characterisation of an integrant/genomic DNA junction from the serum of an acute hepatitis patient (Kimbi *et al.*, 2005) (Table II). It was applied again in this study to DNA extracted from tumour tissue to see whether it might be a useful method of routinely detecting HBV integrants in human genomic DNA.

2.2 Materials and Methods

2.2.1 Selection of Sample Tissues and Controls

2.2.1.1 Subjects

These were black southern African males with HCC, mainly miners from South Africa and surrounding countries. The greater proportion was Mozambican, as was evident from the names. Subjects ranged in age from 20-40 years, with the majority being between 30 and 40. As the subjects were migrant workers, there was a paucity of information concerning individual's personal and medical history. Ethics clearance was obtained from the University of the Witwatersrand Human Research Ethics Committee (Medical) (Appendix F).

2.2.1.2 Liver Tissues

Liver tissue was obtained at autopsy or laparotomy, frozen in liquid nitrogen (African Oxygen Ltd, Germinston, RSA) and stored at -70 °C. The tumours were classified as late stage based on the size of the primary tumour, the presence of metastatic spread beyond the liver, the patients' general condition and death shortly after diagnosis.

Tissue from 18 of these HCCs and adjoining non-tumorous liver tissue, when available, was cut on dry ice and at ambient temperature, and used for DNA, extraction (section 2.3.1, Table III). A further 3 samples were used for RNA studies (Chapter 4). These were T40, T45 and

T110. Those 3 samples did not undergo the same experimental procedures as the 18 listed in Table III because the amount of tissue available only allowed for the extraction of sufficient genetic material for reverse transcription PCR (RT-PCR) and was only sufficient for a few amplifications. These samples were used in a concurrent study and the tissue for T45 became depleted. Sample tissue was removed from the -70^{0}C freezer for short periods of time and placed on dry ice so that it could be cut at ambient temperature with the use of a sterilised hacksaw, and sterile blades (Swann-Morton, Sheffield, UK). Every attempt was made to obtain pieces of sample (about 5 g) as quickly as possible, so as to prevent extensive defrosting and compromising of the DNA in the remaining portion of the sample.

The corresponding serum for each tissue sample had previously been screened for HBV markers: HBsAg, anti-HBs, anti-HBc, HBeAg, and anti-HBe using commercially available radioimmunoassays (Abbott Laboratories, Abbott Park, Illinois, USA) (AUSRIA II for HBsAg; AUSAB for anti-HBs; CORAB for anti-HBc; HBeAg, and anti-HBe).

2.2.1.3 Cell Lines

Frozen cell stocks of the liver cell lines PLC/PRF/5 (Alexander *et al.*, 1976) and HuH7 (Nakabayashi *et al*., 1982) were the kind donations of J. Alexander (Department of Molecular and Cell Biology, Microbiology Unit, University of the Witwatersrand, RSA), and H. Nakabayashi (Division of Pathology, Cancer Institute [H.N.K.M., J.S.] and Department of Orthopedics [T.Y.] Okayama University School of Medicine, Japan), respectively. The cell lines were cultured and used for total DNA and RNA extraction.

The HuH7 cell line contains no HBV integrants and was used as a negative control for integrant detection. This cell line was established circa 1982 from the surgically removed, well-differentiated HCC of a 57 year old Japanese male. Doubling time is approximately 35-44 hours, chromosome number ranges from 50-59, and the karyotype is stable (Nakabayashi *et al.*, 1982).

The PLC/PRF5 cell line was established circa 1973 from autopsy tissue of the HCC of a 24 year old Mozambican male with chronic HBV infection (Alexander *et al*., 1976). It was originally known as the Alexander cell line. The PLC/PRF/5 cell line currently contains nine

integrants per cell, none comprising the FL genome of the virus, but rather rearranged fragments (section 1.5.3). The cells resemble hepatocytes in culture and do not produce HBV particles but do express the HBV S protein.

RNA derived from HepG2.2.15 (Sells *et al*., 1987) was used as a positive PCR control in RT-PCR, and HBV-specific PCR off cDNA template. This cell line was established circa 1987 by transfection of the HCC HepG2 cell line (from the HCC of a 15 year old Caucasian male) (Aden *et al*., 1979) with four 5' to 3' tandem copies of FL HBV genome. The cells contain chromosomally integrated, relaxed circular, covalently closed, and incomplete copies of the viral genome and support the production of HBV replicative intermediates, and the assembly and secretion of complete infectious (Dane) viral particles.

2.2.2 Tissue Culture

Cell cultures (Appendix A section A1.1) were prepared in a laminar flow hood (LABOTEC, Bio-flow Model 660, Midrand, RSA), in a sterile laboratory (UV), exclusively reserved for tissue culture. All regulations for the correct procedure for sterile tissue culture were observed. The cell lines were maintained at 37 ºC in a Forma Scientific (Socotia, NY, USA) water jacketed incubator (model 3164) with 5 % CO_2. Media and reagents were purchased from BioWhittaker (Walkersville, Maryland USA), GIBCO (Life Technologies, Carlsbad, CA, USA) or Cambrex BioScience (Wakerville, MD, USA). These monolayer cell lines were cultured in 80 cm^2 vented flasks (Nunclon). At 80 % confluency, cells were subcultured (Appendix A section A1.2) to prevent nutrient and space depletion and cell death, or they were harvested (Appendix A section A1.3) for DNA or RNA extraction. Fresh stocks of each cell line were made routinely to ensure backups in case of contamination, or accidental loss/death of cells (Appendix A section A1.4). Stocks were stored at -70 °C. For further details see section A1, Appendix A.

2.2.3 DNA Extraction for the Characterisation of Integrants by DNA Analysis

2.2.3.1 Phenol Chloroform (P/C) Method of DNA extraction

This was according to Sambrook *et al.,* (1989), with modifications. For a detailed protocol see Appendix A section A2.

Avoiding and Detecting Possible Contamination

To ensure that no cross-contamination occurred between sample tubes, DNA from the PLC/PRF/5 cell line, which is known to carry 9 HBV DNA integrants, was always extracted on a separate day to any other sample, including DNA from the HBV DNA negative cell line HuH7. Although reagents for P/C prepared in bulk, they were aliquoted into amounts sufficient for the extraction of 9 samples concurrently.

2.2.4 Detection of Integrants by SH (E.M. Southern, 1975)

2.2.4.1 RE Digestion of Genomic DNA and Integrated HBV DNA

The number and size of integrants present in a sample can be established by digesting the DNA with *Hind*III, which does not cut HBV isolates from black southern Africans, and analysing products by SH. Restriction enzyme (RE) digests of these integrants, with various other endonucleases, further enables the construction of hypothetical maps of each integrant, or at least one integrant junction. Maps may be surmised by knowing the exact location of RE cleavage sites on the DNA of interest (in this case the HBV genome – Figure A1, Appendix A). Digestion with a particular RE allows for the sizing of those integrated HBVfragments generated by the particular RE. The fragments can then be reconstructed into a map by locating their positions on the HBV genome via the position of the RE cleavage site, thereby allowing for the design of primers at the cleavage sites. Digestion with single REs as well as combinations of the same REs allows for more precise mapping (section 2.3.3.4).

The REs used were: *Eco*RI, *Bam*HI and *Hind*III. The HBV sequence V10460 (Genbank Accession Number July 2007 - sequence originally characterised by Galibert *et al.*, 1979) was used as a probe in SH and as the restriction digest control.

The selection criteria for the REs were: (1) that the cleavage sequence on human genomic DNA be present very frequently, such that the digested fragments yield the classical smear pattern observed upon agarose gel electrophoresis, (2) that the cleavage sequence be infrequently present on the HBV genome – ie. ideally at a single site but two or three sites is also acceptable, (3) that one RE have no cutting sites on the HBV genome, and (4) possibly that they have been previously documented in the literature as having been successfully used in HBV digestion for SH or the design of PCR primers (Bréchot *et al*., 1980; Chakraborty *et al*., 1980; Edman *et al*., 1980). In particular, REs were selected from Bruni *et al.*, (1995), as it was hoped that the I-PCR technique could be applied to the characterisation of integrants in this study.

For further details, **see section A3.1,** Appendix A.

Digestion Reactions

Digests were carried out as per manufacturer's instructions, with modifications from Sambrook *et al*., (1989), p9.32. To ensure that DNA was digested to completion a digestion control was included as well as other controls (section 2.3.3.1).

Both single and double digests were performed. Single digests involved the use of a single RE whereas double digests were performed with the enzyme combinations. For detailed protocols see Appendix A section A3.1.

2.2.4.2 Fragmentation of DNA for Transfer and SH

All control reactions were placed to one side of the gel, away from the test samples, to reduce chances of contamination, and prevent obscuring of the signal from the test samples. Before the gels were prepared for SH, a 10 µl aliquot of DNA digest reaction was separated by electrophoresis on a short (20 cm x 10 cm) trial gel (0.8 % agarose gel D1-LE agarose (Hispanagar, Burgos, Spain), in 1xTris-boric acid EDTA buffer (Appendices A and D), and 3 % EtBr (Sigma-Aldrich) at 0.3mA/cm^2, for 2 hours at ambient temperature in order to confirm

complete digestion. For detailed protocols see section A3.2, Appendix A.

The DNA was transferred to neutral membrane, Hybond N (AEC Amersham), by capillary action, or with the use of a 'Semi-dry electrophoretic blotting system' (Sigma-Aldrich). The exact protocol for pre-treatment of the gel prior to blotting depended on which method was employed to transfer the DNA (Appendix A, sections A3.2.1 & A3.2.2.). Once transfer was complete the membrane was hybridized to the radioactively labeled HBV probe.

2.2.4.3 Probe used in SH

The probe used in SH was kindly donated by Professor H. Will (Appendix B) and comprised a head-to-tail dimer of the complete genome of HBV genotype D (subtype *ayw*, GenBank accession number V01460) (Galibert *et al.*, 1979) cloned via an *Eco*RI cleavage site into the plasmid pSM2 (Sommer *et al.*, 1997) in JM109 *Eschericia coli* strain. The sequence characterised in Galibert *et al*., (1979), clusters with the South African D genotypes which comprise 25 % of the isolates existing in southern Africa (Kimbi *et al*., 2004). This clone was available in the laboratory and provided a convenient probe for SH (Appendix B). The probe was amplified by FL-PCR (with primers P1 and P2 Appendix C) from the cloned HBV DNA template in plasmid pSM2. PCR product generated one FL HBV genome (Appendix B). The 3.2 kb band excised from the agarose gel following electrophoresis in order to eliminate primer DNA and residual pSM2 plasmid, before labeling (see Appendix A, section A3.5).

2.2.5 Optimisation of PCR for the Characterisation of Integrants

Primer sequences were obtained from the literature (Appendix C), except for the 1374+ primer used for cDNA amplification, which was designed with the aid of the Website tool: http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-primer.html. Primer design incorporated the standard conditions required for the design of an optimal set of primers.

All primers were synthesized by Invitrogen, Life Technologies, Carlsbad, California, USA; or Inqaba Biotechnical Industries (Pty) Ltd, Hatfield, RSA. DNA polymerases with proofreading functions were selected for PCR, and appropriate ones were selected for long template (~40 kb) amplification. PCRs were optimized with annealing temperatures within a 10 ºC range of

Tm, and various $MgCl_2$ and DNA/RNA concentrations were optimized as necessary. All PCRs were carried out under strict conditions, as outlined by Kwok and Higuchi, (1989), so as to prevent cross-contamination and false-positive results. The different stages of DNA preparation, namely extraction, amplification and electrophoresis were carried out in different rooms.

2.2.5.1 I-PCR (Bruni *et al*., 1995)

DNA is restricted with the correct RE (Figures A3 & A4, Appendix A) and ligated, yielding a circular template. Primers are orientated such that the DNA synthesis is primed towards the opposite ends of the template, hence the term "Inverse PCR" (Figures A3 & A4, Appendix A). After amplification by I-PCR, sequencing of the product permits characterization of a section of the integrated HBV DNA. From this data cellular primer/s directing DNA polymerisation towards integrated viral DNA, and specific to the integration under investigation, are designed (fixed flanking primers). The fixed flanking primers are used in fixed-flanking-primer PCR (FFP-PCR) with genomic template. Re-amplification of the FFP-PCR product can be performed if necessary. A restriction digest of the product is performed with the original respective RE, and analysed on a 2% agarose gel to confirm product specificity. The integrant is characterised by automated sequencing**.** For a more details see Appendix A section A4.

2.2.5.2 PCRs for the direct amplification of HBV integrants off genomic DNA template

This technique may involve up to 5 rounds of amplification, comprising: *Alu*-PCR and digestion with uracil DNA glycosylase (UDG); Touchdown PCR (product may be visualized by SH); and Nested/heminested PCRs I, II and III (Figure A5, Appendix A).

First round: "*Alu* PCR"

The first round PCR is derived from a technique developed by Nelson *et al*., (1989, 1991) (Figure A5), which permitted the amplification of human DNA against a background of phage, yeast or rodent genomic DNA. The protocol was adapted by

Minami *et al*., (1995), for the amplification of HBV DNA against the background of a human genome.

In principle: the PCR requires the use of a HBV specific primer (Appendix C), and a primer for the most common SINE (short interspersed nuclear element) repeat sequence (Appendix C) in the human genome (~5% of total genome mass): the *Alu* repeat (Szmulewicz *et al.,* 1998). *Alu*-repeat sequences are found in the human genome on average every 4 kb, and are called *Alu* because they can be restricted with the RE *Alu*I. The primers used to amplify the *Alu*-repeat sequences were originally designed to: hybridise either the 5' or 3' end of the *Alu*-repeat, be highly conserved (Britten, 1994) and to have a Tag sequence at the 5' end.

In theory, this PCR will amplify HBV-genomic DNA junction/s that lie sufficiently close (ie. ≤40 kb), either up or downstream from an *Alu* repeat sequence, while avoiding amplifications between *Alu* sequences (Figures A5 & A6, Appendix A). For a detailed protocol see section A4.2, Appendix A.

Second round (Touchdown) PCR (Don *et al*., 1994)

"Touchdown" PCR was initially devised to eliminate mis-priming by one or both primers during PCR reactions involving the amplification of a specific gene sequence from a complex genome template. Mis-priming results in spurious bands, which when visualized by agarose gel electrophoresis, are often found to prevail over the desired product. This occurs with greater prevalence when the sequence of interest is present in small amounts, as with HBV integrants.

In "Touchdown" PCR the annealing temperature (or a temperature above it) is decreased by a unit of temperature (eg. 1ºC) every second cycle for 10ºC, at which temperature 10 cycles are performed. "Any difference in Tm between the correct and incorrect 'annealings' will give an advantage of 2-fold per cycle, or 4-fold per ºC, to the correct product, all else being equal" (Don *et al*., 1994). For a detailed protocol see section A4.2.2, Appendix A.

Further amplification

"Touchdown" PCR may yield a smear of fragments. To generate discrete products a third round of amplification may be required. If so, 1 µl of "Touchdown" PCR product is used as template in nested or hemi-nested PCR with internal primers (MD60 and Tag5) (Appendix C). Two further rounds of amplification may still be required before final discrete, sufficiently concentrated PCR products, are available for further analysis (P. Paterlini-Bréchot - personal communication). For "fourth round" PCR MD26C sense primer (Poussin *et al*., 1999) (Appendix C) and anti-sense Tag primer are required. Amplification conditions are identical to 2nd round PCR. For fifth round PCR MX2 sense primer (Appendix C) and Tag anti-sense primer can be used.

PCR optimization and controls

Optimisation and positive control:

In the study reporting the original adaptation of Nelson's technique to the amplification of HBV integrants the PCR conditions were "optimized using clone 20BB (Wang *et al*., 1992), [consisting of] … an 18 kb human genomic DNA fragment that contains 3 kb of HBV sequence integrated in human cyclin A gene and an *Alu* repeat about 1600 bp downstream of the viral-host junction." DNA from this clone was "mixed with human genomic DNA to obtain about a 2 kb band using a modification of long PCR protocols" (Minami *et a*l., 1995).

Negative controls:

Three negative controls were included in these amplification reactions to ensure that HBV DNA was being amplified: (1) water blanks were inserted as the first (tube 1) and the last tube/reactions in the PCR (control for contamination); (2) HuH7 DNA was included as tube 2 (control for HBV DNA primer specificity); (3) DNA extracted from the liver tissue of a normal (HBV free) subject (organ donor) was included as tube 3 (control as per HuH7 DNA).

2.2.5.3 FL-PCR

The FL HBV genome PCR protocol, developed by Günther *et al*., (1998), allows for the efficient amplification of the entire HBV genome in one PCR. For details of this technique see Appendix A, section A4.3.

Templates for FL-PCR

1. total DNA isolated from subjects (genomic template)
2. the head-to-tail dimer HBV genotype D **(**Appendix B) for use as a probe in SH (sections 2.2.4.3 & 2.2.4.1)

Controls

DNA extracted from the HuH7 cell line, as well as best quality water, was used as a template in FL-PCR for blank or negative controls. Usually a blank/water control was positioned at the beginning (tube 1) and end of a PCR (last tube), and blanks were placed between every sample. The pSM2/HBV clone template (Appendix B) was used in genomic DNA FL-PCR as the positive control (fragment size ~3.2 kb).

2.3 Results

2.3.1 Selection of Samples for Analysis

The first step in analysing integrants from the southern African black population was to select suitable samples from the available tissues. This selection was based on four primary criteria.

1. *Presence of HCC* The presence of HCC in the samples of liver tissue from patients known to have died as a result of the tumour were examined by Professor M. Kew, Dora Dart Professor of Medicine, Faculty of Health Sciences of the University of the Witwatersrand and the supervisor of the research project, and Professor A. Paterson from the Department of Anatomical Pathology of the Faculty of Health Sciences of the University of the Witwatersrand, with the naked eye and under the microscope to confirm that they contained HCC tissue. If HCC tissue was confirmed, those specimens were used in the study. If the specimens also contained non-tumorous liver tissue they were of particular value (section 2.2.1.1 & 2.2.1.2).

2. *Presence of HBV* Archived samples from individuals who were positive for HBsAg were chosen (section 2.2.1.1). In these samples, HBV DNA and possibly HBV DNA integrants would likely be present. One patient, (No. 14) was HBsAg negative, but positive for anti-HBsAg and anti-HBcAg. As the number of available samples was limited, it was decided to keep this particular one in the study, considering that HBV integrants have been reported in patients who have resolved HBV infection serologically (Kew 2002b; Kimbi *et al*., 2005).

3. *Amount of available tissue* It was supposed that several DNA extractions might have to be performed on any one sample in order to obtain sufficient genetic material. Samples with at least 50 g of tissue or more were chosen.

4. *Condition of available tissues* Liver tissue from 18 individuals was selected (we had originally proposed to use 10-15). Of these, samples were accepted or rejected based on the availability of tumour (T) and non-tumorous (NT) tissue, and on DNA yield. Four samples yielded very little, poor quality, or no DNA, even after three attempts at extraction. In these cases a second set of extractions were performed on tissue from another area of the original sample. If this did not yield good quality DNA/RNA, that sample was excluded. In total 11/18 were selected for analysis (Table III). A further 3 samples were used for RNA analysis (Chapter 4 – see also section 2.2.1.2).

Table III: Selection of samples based on availability of tumour (T) or non-tumorous tissue (NT).

	Sample number																	
T	3		9	10		14	15	16	17		30	31	32	37	39	69	70	87
NT		6		10	11	14		16		24						69	70	87
A/R	A	R	A	R	R	A	A	R	R	R	A	A	A	A	A	A	A	R

Key: T – tumour tissue NT – non-tumorous tissue; A/R – accepted/rejected

2.3.2 DNA Extraction

Good quality DNA in sufficient quantity was extracted by the PC method from liver tissues (Figure 2.1) and cell lines.

2.3.3 Detection of Integrants by SH

2.3.3.1 SH Technique Optimisation

Samples were digested with endonucleases *Eco*RI, *Bam*HI, and *Hind*III, and subjected to SH. Initially RE digestion was performed on 10 µg of genomic DNA at the appropriate temperature for the particular RE, for 16 hours. However, the DNA concentration was increased to 20 µg as per some published protocols (Chen *et al*., 1986) because HBV integrants occur in lower copy number than host genes, and are therefore more difficult to detect, requiring a probe of high sensitivity. The REs chosen would theoretically enable: (1) the number, and approximate size of HBV integrants present in each paired tumour/non-tumour sample to be established; (2) the construction of a hypothetical RE map for each integrant, from which primers for PCR could be designed.

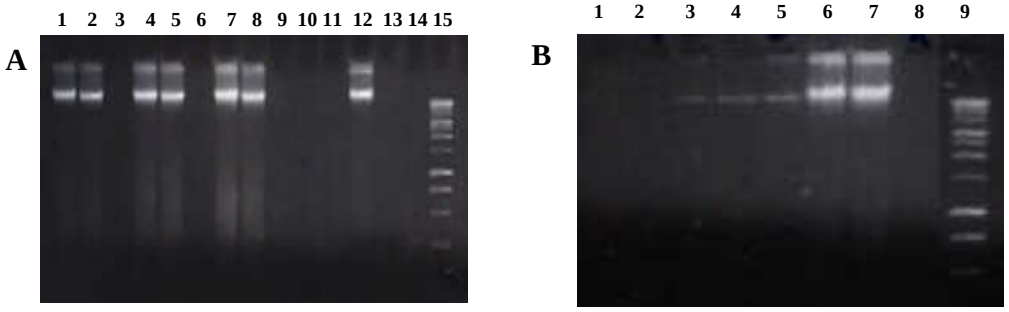


Figure 2.1: Agarose gel resolution of T14 DNA extracted by the P/C technique.

Figure 2.1A: Estimation of DNA concentration against commercially available standards. Lanes **1,2,4,5,7,8**: T14 DNA**; 3,6,9**: negative water controls for the extraction procedure; **10,11,14**: empty; **12,13**: 100 ng and 5 ng lambda DNA respectively; **15**: 1 kb molecular weight marker (Promega). Aliquots of 100 ng, and 5 ng of commercially available lambda DNA (Promega) were included to allow approximate estimation of DNA concentrations.

Figure 2.1B: Serial dilutions of T14 DNA allow for approximate estimation of DNA concentrations. *(Composite photograph)* Lanes **1-7**: T14 DNA dilutions 1:500, 1:300, 1:200, 1:100, 1:50, 1:10, 1:1; **8**: negative water control for the extraction procedure; **9**: 1 kb MWM (Promega). DNA concentration was calculated from the faintest visible band (10 ng), ie. lane 2, in this case 0.3 µg/µl.

It was confirmed that endonucleases had no unspecific endonuclease and exonuclease activity in digests lasting 16 hours, as per the manufacturer's catalogue (Nucleotide positions numbered as per RE cutting site nucleotide position 1 - Galibert *et al*., 1979). RE cutting sites were confirmed using GenDoc and a compilation of 175 sequences representative of worldwide HBV isolates courtesy of Dr Anna Kramvis (MHRU, Wits University Medical School, 7 York Rd Parktown, RSA).

To ensure complete digestion, conditions were optimised using 20 μg of the negative control HuH7 DNA combined with 10^{-5} μg of the HBV FL genomic DNA used as probe in SH (Sambrook *et al*., 1989, p 9.32). For SH digested DNA from samples were resolved on agarose gel alongside digested and undigested probe DNA and DNA from the positive control PLC/PRF/5 cell line (containing 9 integrants when digested with *Hind*III). Digests were deemed complete when even DNA smears with signature bands for the particular RE were present on agarose gels for digested genomic DNA, and undigested genomic DNA was in the high molecular weight region of the gel, with no appearance of smearing. Furthermore, all the predicted fragments for digested probe and PLC/PRF/5 DNA had to be present in the digested samples, and a 3.2 kb fragment and high molecular weight fragment had to be present on autoradiographs for the undigested probe and PLC/PRF/5 DNA respectively (Figure 2.2).

Figure 2.2: Agarose gel resolution of *Eco*RI digested DNA for the SH technique. *Lanes **1:** HuH7; **2:** Hep3, **3:** PLC/PRF/5; **4:** (3 mm wide) 1 kb molecular weight marker (Promega).* The DNA was evenly digested (A: lanes 3-6) and the same signature for the *Eco*RI RE is clearly visible for both the samples and the controls.

Initially the DNA in some samples appeared to be incompletely digested, as was apparent from slightly shortened smears and incomplete RE signatures as well as undigested HBV probe in the control reactions. This could have resulted from variations in the local concentration of DNA within the tube, where DNA forms secondary structures and remains inaccessible to REs. Similarly the half-life of the REs used is 2 hours. To improve DNA digestion, the DNA was resuspended on a gently rotating shaker for several hours before digestion as recommended in Sambrook *et al.*, (1989). The amount of enzyme used was increased from the manufacturer's recommended dose of 1 unit/1 ug DNA to up to 4 units/1 ug. The RE was added in three doses at two hourly intervals - 'spiking of the enzyme' (section 2.2.3.1) - thereby extending the period of maximum enzyme activity. These measures resolved the problem.

As the volume of the digestion reactions often exceeded the well capacity of the agarose gels, DNA was precipitated with ethanol and re-suspended before resolution by electrophoresis. Care was taken to ensure that as little of the DNA as possible was lost during this procedure, however as the SH results were to be qualitative and not quantitative, a small amount of DNA

loss did not constitute a problem. To ensure maximum DNA transfer to the membrane, capillary transfer was performed as per Sambrook *et al*., (1989), but over a period of 3 days. During this time, buffer was 'topped up' and the gel and buffer tank kept well sealed with the use of plastic film to prevent dehydration of the gel. Later the purchase of a 'Semi-dry electrophoretic blotting system' allowed us to accomplish the transfer successfully in 5 hours.

2.3.3.2 Probe Labelling and Hybridisation

Labelling of the probe (section 2.2.4.3; Appendix A section A3.5) initially yielded a low specific activity (approximately $4x10^6$ cpm/μg) and a probe of high specific activity was required for integrant detection. The literature suggested 2-8 x 10^8 cpm/μg (Chen *et al*., 1986; Zerial *et al*., 1986). Initially doubling the time of the labelling reaction increased the efficiency of labelling by a factor of 2. Increasing the amount of DNA used in the labelling reaction to a maximum of 200 ng improved labelling by a factor of 5. However, increasing both DNA concentration and duration of labelling reaction, only increased labelling efficiency by a factor of 3. The labelling reaction was then attempted with TE buffer instead of the deionised distilled water suggested in the kit protocol. This improved labelling by a factor of 10 and probe of 5-8 x 10^8 cpm/μg was obtained for hybridisation. At this point it was found that when labelling with TE buffer, the difference between 100 ng and 200 ng of DNA was a factor of 2, rather than a factor of 5. It was decided that as the labelling efficiency was already 10x higher, the further increase by a factor of 2 was not strictly necessary, and using less DNA was more economical.

2.3.3.3 Rapid Hybridisation Solution (QuikHyb) versus Church and Gilbert Hybridisation Buffer

The QuikHyb® hybridization solution (Stratagene) was used for the hybridisation step as it required a total hybridization time of 1-2 hours. However, upon exposure of the autoradiographs for up to 3 days, the fragments were barely visible. In order to extend the time of hybridization to 16 – 20 hours, QuikHyb® hybridization solution was replaced with Church and Gilbert Hybridisation Buffer recommended in the Hybond N Manual (AEC Amersham International). Optimal signals were obtained on a 20x20 membrane with a hybridisation step in 50 ml of buffer, 8 x 10^8 cpm/μg of labeled probe, at 65 ^{0}C overnight.

2.3.3.4 Interpretation of SH Results

It was assumed that if a target site/s for a particular RE was/were present in an integrant, two fragments from each site would be generated, each capable of binding the HBV probe. SH results would be interpreted as follows: (1) unrestricted sample DNA – band in high molecular weight DNA region of blot indicated the presence of integrated HBV DNA; band in region 3.1-3.3kb indicated the presence of episomal HBV (2) sample DNA restricted with *Hind*III – bands indicate individual integrants, the number of which corresponded to the number of bands as Hind*III* does not cut HBV DNA; (3) *Eco*RI, *Bam*HI – two or more bands showed the presence of the site in the particular integrant (section 2.2.4.1). Each sample was digested with each RE, and with combination of the 3 endonucleases.

Initially molecular weight markers were labeled with radioactive isotope before being resolved alongside the samples on the agarose gels. After being photographed, the lanes containing the markers were not cut off the gels before transfer. This enabled the markers to be visualised on the autoradiographs, permitting better sizing of any fragments present. Thereafter, two trial hybridisation reactions were performed with samples and unlabeled molecular weight marker DNA on the membrane and radioactively labeled HBV as probe. These were stripped and exposed to confirm the absence of any visible bands. Thereafter they were re-hybridised with both radioactively labeled HBV probe and the same molecular weight markers as probes. The results were comparable between the blots, whichever method was employed. There were no extraneous fragments when the molecular weight markers were used as probes, indicating that HBV probe was binding specifically and that the molecular weight marker DNA was not hybridising to potentially homologous regions on the human genomic DNA. Furthermore the lanes containing DNA from the human HBV DNA negative cell line HuH7 remained devoid of any bands in all autoradiographs proving that neither the molecular weight marker (MWM) nor the HBV probe DNAs were binding to human genomic DNA. All remaining SH analyses were then performed with the use of both radioactively labeled HBV and molecular weight marker DNA as probes.

2.3.4 PCRs for the Characterisation of Integrants

2.3.4.1 I-PCR

I-PCR was not performed (Discussion - section 2.4.4.1).

2.3.4.2 PCRs for the direct amplification of HBV integrants off genomic DNA template

The conditions used in first round PCR were from the 1995 publication by Minami *et al*., with primers HB1 and A5/A3, and later again by Chami *et al*., (2000), (Figures A5 & A6, Appendix A). HB3 A5/A3 PCRs could not be attempted as the published sequence for primer HB3 was incorrect and did not map anywhere on the HBV genome. Several attempts to obtain the correct sequence were unsuccessful so an alternative primer to HB3 (FX3 Kawai *et al*., 2001 – Appendix C) was used in combination with each of the reverse primers (A5/A3) respectively. The location of HB3 was unknown so the criteria for the replacement primer were that it (1) be a primer in the reverse orientation to HB1, (2) be relatively conserved, (3) be in a favourable location of the HBV genome (4) be compatible to the forward primer. FX4 (Kawai *et al.*, 2001) was thus chosen, located in an ideal position (1846-1865) close to an integration hot-spot and priming towards the X gene. The correct sequence for the HB3 primer was eventually published in 2004, by Murakami *et al*., and was located at 1257-1288.

Before performing PCR on sample DNA, an attempt was made to establish whether any of the 9 PLC/PRF/5 cell line integrants might be located sufficiently close to an *Alu*-repeat to amplify using this technique. If so, this would serve as a useful positive PCR control, and PCR conditions could be optimised with minimal use of sample DNA. A number of DNA concentrations were used. The products obtained with PLC/PRF/5 DNA were resolved on 0.8% agarose, and slight smears were observed in several lanes (Figure 2.3). Minami *et al.*, (1995), found that in certain subjects, while their 20BB PCR control generated a distinct 2 kb band, there were no other visible products other than possibly a smear. These could only be visualised by SH (with a HBV probe) of "Touchdown" PCR products, sometimes as a smear, at other times as discrete fragments. Further rounds of amplification should be performed in order to attempt to generate discrete fragments, but only if there is a product visible on SH after Touchdown PCR, therefore SH was performed on the products. The SH technique itself was successful, as the molecular weight markers (also used as transfer and hybridisation controls) were present, but no distinct HBV specific product was detected. The smears observed on the agarose gel were most likely non-specific amplification, possibly between *Alu*-*Alu* repeat sequences, or non-specific binding of HBV primers to genomic DNA. The PCR reaction thus failed to amplify any possible integrants from the PLC/PRF/5 cell line.

Without a positive control is was not possible to optimise conditions for the PCR.

Alu-PCR was performed with primers HB1 and FX1 in combination with A5 and A3 and respective nested primers on DNA from HCC samples T87, T37, T30, T69. The PCR reactions generated only smears (Figure 2.3) and SH showed an absence of HBV DNA amplification. Decreasing the amount of enzyme and using a high salt buffer resulted in increased specificity and reduced smearing, but smears are still evident in lanes 2-7 and especially in lanes 3-6. This PCR reaction should amplify discrete fragments specific to HBV sequences, however often a smear is evident resulting from some amplification of fragments located between *Alu* repeat sequences. Upon confirmation of the presence of HBV DNA in the smear by SH, discrete fragments can be generated with further amplification by PCR.

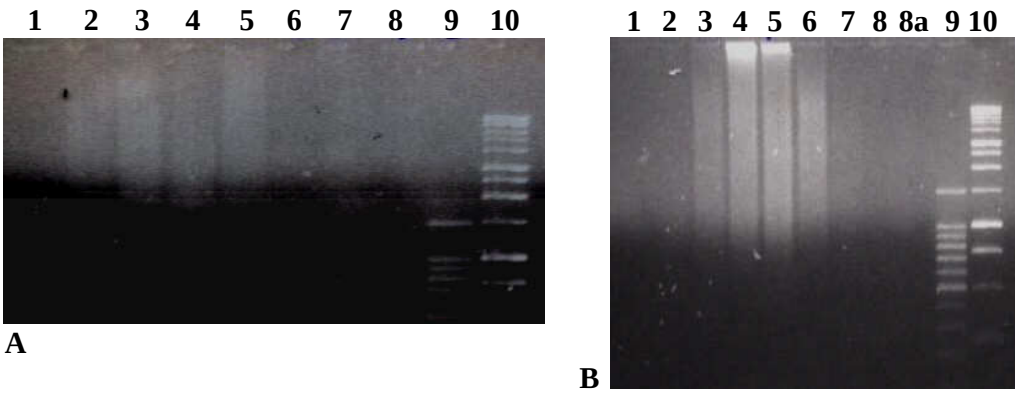


Figure 2.3: Agarose gel resolution of nested *Alu*-HBV PCR product, performed with primers FX1/A5 (PCRI - A) FX2/Tag5 (PCRII - B).

*Lanes **1&8:** negative water controls; **2:** PLC/PRF/5; **3:** T87; **4:** T37; **5:** T30; **6:** T69; **7:** HuH7; **8a:** empty; **9&10:** 100 bp and 1 kb MWMs (Promega) respectively.*

2.3.5 Amplification of integrants by FL-PCR

Of the several PCR products generated only some would be derived from HBV DNA, making it necessary to distinguish fragments unique to specific samples in order to localise possible integrants. HuH7 DNA, and DNA extracted from "normal" liver tissue or the blood of a healthy HBV negative individual, was amplified alongside HCC DNA samples. Upon

resolution on agarose gel, unique fragments, absent in the normal controls and other samples, were excised from the gel and sequenced using the original PCR primers (Figure 2.4). The fragments amplified from all samples were found to contain only genomic DNA and this approach was abandoned (Discussion - section 2.4.4.3). This was clear on the autoradiograph as only the DNA marker and PLC/PRF/5 samples generated the expected bands. The HuH7 negative control and the sample lanes were blank, showing a complete lack of any hybridisation of sample DNA to the radioactively labelled HBV probe (data not shown).

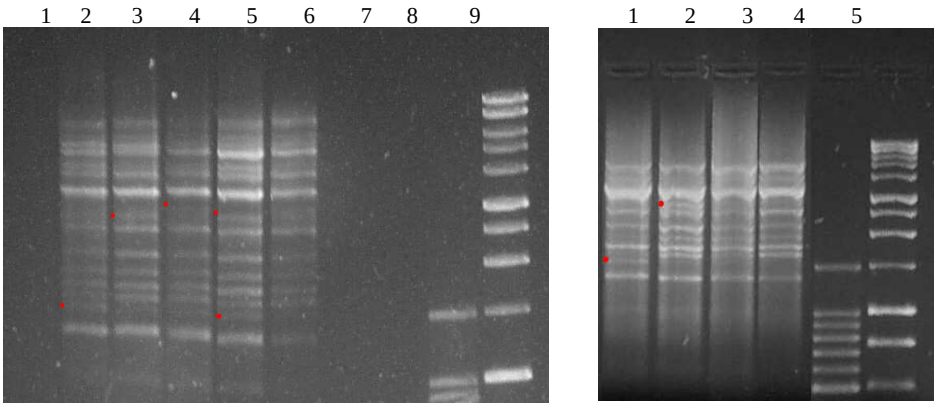


Figure 2.4: Agarose gel resolution of DNA fragments generated by the FL-PCR technique (composite photos).

2.4A: *Lane **1:** negative water control; **2:** T87; **3:** T37; **4:** T30; **5:** PLC/PRF/5; **6:** HuH7; **7:** negative water control; **8:** empty; **9-10:** 100bp and 1kb MWMs (Promega) respectively.*

2.4B: *Lane **1:** T9; **2:** T69; **3:** PLC/PRF/5; **4:** HuH7; **5-6:** 100bp and 1kb MWMs (Promega) respectively.* The red dots indicate bands that appear to be unique to a sample or that may be found in one or two samples but are absent in the HuH7 control (or normal control when used). These bands could indicate possible integrants and were candidates for further analysis.

2.4 Discussion

2.4.1 Selection of samples for analysis

All tumours were at an advanced stage and because of the paucity of available records (most samples had been obtained at autopsy) no attempt was made to correlate mutations with clinical or other features of the individuals. A substantial portion of the southern African population is rural, patients are only seen by a medical practitioner at very advanced stages of the disease and follow-up on any of them is virtually impossible. Therefore, attempting to draw any conclusions requiring information dependent on stage of HCC development is not possible because of the advanced stage of the tumours.

In this study, considering the need for SH, various PCR techniques and possible cloning and expression studies, it was decided at the start that samples used should comprise at least 50 g of tissue in case further DNA/RNA extractions be required in the later stages of the work.

Eighteen HCC tumours or tumour/non-tumour pairs were initially used. At a later date some samples were included (section 2.2.1.2) for which very little tissue was available. One such sample yielded a partial integrant and could not be further investigated because of the depletion of the tissue.

2.4.2 DNA and RNA Extraction from Tissue Samples

SH of RE digested DNA was performed in order to establish which samples contained integrants and what percentage of the whole these comprised. The results should also permit the design of RE maps from which PCR primers could be designed (sections 2.2.5.1 & 2.3.4.1). For SH, DNA should not be sheared so as to ensure correct digestion with REs, which is vital in establishing whether an integrant is intact and in generating accurate maps. As 100-300 μg of DNA was required, several extractions were simultaneously performed, each on 0.5 g of tissue, by the P/C (Sambrook *et al*., 1989) protocols. Different extracts from the same sample, extracted by the same method were pooled. This introduced both a major advantage and several disadvantages. These 0.5 g samples were not always obtained from the same area/nodule on the tumour (section 1.5.2). The advantage was, that should the original tissue comprise various separate clonal expansions of integrants (Robinson, 1994), these would all be

represented in the DNA/RNA extract. If individual extractions were not pooled, the different extracts might contain different sets of clonally expanded integrants and would not be representative of the total population.

Disadvantages of this approach were that by pooling different extracts, which could possibly have derived from different clonal expansions, or areas containing no integrants, one would be "diluting" down the integrants in the total mix, making them more difficult to detect. This was however, unavoidable, and it seemed preferable to have sufficient material to work with.

2.4.3 SH

SH remains the best method for the initial detection of integrated HBV DNA in host DNA. In this technique the detection of a target sequence is affected by the amount of complementary sequence available for hybridisation, in terms of the number of copies of each integrant present, the amount of DNA used, and the transfer efficiency. Technically sensitivity is less than 0.1 pg of DNA complementary to a probe radiolabeled to a high specific activity ($>10^9$ cpm/ug) with α-^{32}P dCTP, after several days of autoradiographic exposure (Sambrook *et al*., 1989). Theoretically therefore, some integrants may be missed: namely those integrants present in very few cells, or very small integrants (less than ~100 bp) (specificity of probe-binding should increase with the length of complementary sequence). Nevertheless, because of the drawbacks of the PCR methods of integrant detection (sections 2..4.4.2 & 2.4.4.3), SH was successfully optimized and found to be the most reliable method of integrant detection in the present study.

2.4.4 PCRs for the Characterisation of Integrants

2.4.4.1 I-PCR

The I-PCR was not used as fragments generated by RE digestion could not reliably be assigned to any one of the multiple integrants present in each sample. The original publication had used European HCCs in which single integrants are usually the norm (Giacchino *et al*., 1987).

2.4.4.2 PCRs for the direct amplification of HBV integrants off genomic DNA template

Further characterisation of integration events (ascertaining the sites of HBV integration relative to functional cellular genes - aim number 2 section 1.6), was also attempted by *Alu* PCR with second round (Touchdown) PCR (Don *et al*., 1994). Drawbacks to this technique include missing those integrants positioned too far from *Alu* repeat sequences to be successfully amplified by PCR, and those integrants in which the primer sequence has been deleted. However, this series of PCRs was successfully used to amplify HBV integrants for characterisation, at least in patients with single integrants (Paterlini-Bréchot personal communication) (Table II section 1.5.1.1) (Chami *et al*., 2000, 2001; Gozuacik *et al*., 2001; Paterlini-Bréchot *et al*., 2003; Murakami *et al*., 2004, 2005). These PCRs had to be optimised for southern African samples.

The original protocol developed by Nelson *et al*., (1989, 1991), was adapted by Minami *et al*., (1995), for the amplification of HBV DNA against the background of a human genome. They optimized PCR conditions with the use of a clone (20BB) (Wang *et al*., 1992), mixed with human genomic DNA and with the use of a modification of long PCR protocols (Minami *et al*., 1995). Minami *et al*., (1995), found that in certain subjects, the 20BB control generated only one product - a distinct 2 kb band. Upon further amplification with "Touchdown" PCR however, other products became visible. These were analysed by SH with a HBV probe. The products were sometimes a smear, at other times discrete bands. If a smear was evident, Minami *et al*., (1995), performed further rounds of amplification in order to attempt to generate discrete bands.

After attempting the *Alu*-PCR using DNA from PLC/PRF5 cells and several HCC samples, no HBV-specific products were visualised on SH in this study, but rather a smear was generated, suggesting that non-specific amplification was the cause, either from incorrect orientation of the primers (the entire PCR protocol requires that 4 independent reactions be performed per sample because of the possible orientation of the primers in relation to the HBV integrant and *Alu* sequence), or non-specific primer-binding. The 'Touchdown' PCR was then performed to generate discrete bands, and products analysed by SH. No discrete products were visualised, indicating the need for further optimisation of the amplification conditions.

To successfully optimise the PCRs and amplify integrant sequences without the help of a positive control is difficult and laborious, especially when two rounds of PCR are required to generate discrete products and a further SH required to visualise them. Furthermore, this is extremely difficult in samples where multiple integrants are present and partial PCR products of various integrants may be obtained. Fragments amplified by *Alu*-PCR and made visible by SH might still be too dilute for successful sequencing and require a further series of downstream amplification reactions. With each new round of PCR the combination of primers for subsequent reactions increases exponentially (Dr Paterlini-Bréchot – personal communication). In Minami *et al.*, (1995), the authors state that: "Efficiency of the method should be enhanced by fine adjustment and [the] application of long PCR protocols; however, long PCR, at present, seems more sensitive to any change in amplification conditions, including primer design and the sequence to be amplified". Not an ideal recommendation for a PCR where the first round conditions required adjustment and 'fine adjustment' and no positive control was available. Generating such a control would prove extremely time consuming. It was not possible to obtain a control elsewhere. After several attempts at, and various changes to, the initial PCR conditions, *Alu* and Touchdown PCR (followed by SH) continued to show unsuccessful amplification. Without a positive control it was impossible to re-design primers and optimize conditions for the PCR, and it was decided to proceed with another approach.

Currently, other than in the lab in which HBV-*Alu* PCR was developed, and two other collaborating laboratories with affiliated authors (Murakami Y; Tu H.) one other study (Kawai *et al.*, 2001) in the literature reports the use of HBV-*Alu* PCR to successfully amplify and characterise integrants. Kawai *et al.*, (2001), reported 10-14/17 HBsAg positive HCCs to have integrated HBV DNA by the *Alu*-PCR method. However, they detected the integrants by SH and did not report sequencing of the products. There also appeared to be more than one product per sample. This, jointly with the low concentration of the PCR product may have made the products difficult to sequence. The technique may not therefore be as easily reproducible as the initial reports suggest.

2.4.4.3 FL-PCR Technique

The PCR approach to characterising HBV integrants in the host ideally enables large numbers of integrants to be processed less laboriously, rapidly, cheaply and with the use of much less tissue than library construction (Chen *et al*., 1994). With libraries there is also the confusion generated from the incorporation of the free (non-integrated) HBV DNA, which competes with integrant DNA during ligation. SH is useful for the initial analysis of possible integrants but does not allow for precise characterisation. Furthermore, it requires the use of relatively large amounts of DNA versus the minute amounts required for PCR. PCR however, does have the main disadvantage that certain integrants can be missed because exact primers cannot be chosen for an unknown target sequence. PCR must therefore be attempted with non-specific primers, or with primers targeting a known sequence ligated to the unknown target prior to amplification. No standard protocol exists for either method.

The FL-PCR technique was initially developed to amplify the complete HBV genome in a single reaction (Günther *et al.*, 1998). In the laboratory of Professor Kew, a study on HBV variants, involving this technique, had shown that the primers P1 and P2 (Appendix C) used to amplify HBV DNA, had amplified what appeared to be a portion of a possible HBV integrant. It was also discovered that these HBV specific primers could also randomly hybridise to chromosomal DNA. The integrant was further analysed with the use of two subgenomic PCRs (Kimbi *et al*., 2005) and was found to have integrated at 7q11.23 in the *WBSCR1* gene. The HBV DNA comprised a portion of the *S* and *X* genes in opposite orientation with the right junction mapping to 1774 (at the cohesive overlap terminus) at a site for Topoisomerase I cleavage. It was not possible to map the left junction. This happenstance indicated that the FL-PCR technique could possibly be used to amplify HBV-genomic DNA junctions and it was therefore applied to the HCC samples in this study. In this PCR fragments larger than entire genome length could be considered integrants. Those smaller than entire genome length, could be splice variants; replicative intermediates or integrants containing deletions.

The FL-PCR produced various fragments several of which (especially those which appeared to differ between samples) were isolated and sequenced. In all cases they proved to be purely genomic DNA, although most of the sequences were not found to have homologous counterparts in EMBL/GenBank, a scenario that is sometimes reported in the literature

(Murakami *et al*., 2004). From these results it became apparent that although an integrant can be isolated in this manner (Kimbi *et al*., 2005) the likelihood remains low, as the integrant sequence will be indistinguishable from other fragments apparently unique to samples, which are purely genomic DNA. The genomic DNA would be of human/host origin and derives from the fact that the PCR primers are binding too indiscriminately. From this it becomes apparent that the FL-PCR is a less effective method for cloning HBV-host integration sites than those PCR methods dependent on primer anchoring by mapping of RE sites on the integrant sequence or HBV genome (eg. I-PCR, RS-PCR, ligation PCR). The recognition of HBV DNA by the HBV specific PCR primers used in Kimbi *et al*., (2005), appears to occur too infrequently to allow for cloning of HBV-genomic junctions from most HBV-positive HCCs. A large (unknown) number of fragments would have to be screened before a single integrant was found making the technique clumsy and impractical for this purpose. A better solution would be to analyse the PCR products by SH to establish which bands included HBV DNA before proceeding to sequence analysis. SH of the samples in this study however, indicated a lack of HBV DNA in the fragments generated by FL-PCR.

3.0 DETECTION AND CHARACTERISATION OF INTEGRANTS (DNA ANALYSIS)

3.1 Introduction

SH was successfully optimized in order to study HBV integration in liver derived from southern African HCC patients. This technique described in the previous chapter (section 2.2.4) was used to establish the:

- presence or absence of integrants
- percentage of HCCs containing integrants
- number and size of integrants per sample and the size of individual integrants.

3.2 Results

3.2.1 Integration in the HCCs

Multiple integrants were detected in 9 out of 11 individuals (82 %) (Figures 3.1, 3.2 and Table IV), with T3 having the largest number (10) (Figure 3.1 – lane 2) of clonally expanded integrants. In lane 3 of Figure 3.1 the high-molecular weight band is evident and a smear is present towards the middle and bottom of the lane, possibly indicating hydrolysis, however the smear is not tapering and it was not present on the agarose gel. Faint smudges hint at possible integrants, but these are not present in the *Hind*III lane. The more likely explanation of this result is that the smear is caused by salt, possibly trapping residual RNA.

In the only matched pair (T/NT14) studied, both the tumour and non-tumorous tissues contained integrants, with the tumour tissue having more distinct clonal expansions (Table IV). No integrants were found in T31 and T32.

Table IV: List of tissue samples (with their antigen status) and PLC/PRF/5 DNA, analysed by SH and the number and size of integrants found.

Patient No.	Sample	Antigen status	No. of clonally expanded integrants	Size of integrants in kb		
1	T3	HBsAg +ve HBeAg -ve	10	>10.0 7.0 6.0	5.0 3.5 3.0	2.7 2.6 2.3 1.0
2	T9	HBsAg +ve	2	>10.0 6.8		
3	T14	HBsAg –ve anti-HBsAg +ve anti-HBcAg +ve	4	8.0 3.0 2.5 2.1		
	NT14	HBsAg –ve anti-HBsAg +ve anti-HBcAg +ve	2	2.4 2.0		
4	T15	HBsAg +ve	5	unable to determine size (see Fig. 9)		
5	T30	HBsAg +ve	5	6.4 5.4 3.0	2.2 1.2	
6	T31	HBsAg +ve HBeAg –ve	0	NA		
7	T32	HBsAg +ve	0	NA		
8	T37	HBsAg +ve	7	9.9 7.8 6.9 5.9	4.9 3.6 3.2	
9	T39	HBsAg +ve HBeAg –ve	Diffuse pattern of integration	NA		
10	T69	HBsAg +ve	5	>10.0 7.5 3.7	3.5 3.3 band intensities differ	
11	T70	HBsAg +ve	6	3.9 3.5 3.2	2.8 2.0 1.75	
PLC/PRF/5		HBsAg +ve	6 with 3 possible others	>12.0 12.0 7.7	5.0 4.5 1.9?	1.7? 1.4?

Key:

T - tumour tissue
NT - non-tumorous tissue
kb - kilobases
NA - not applicable

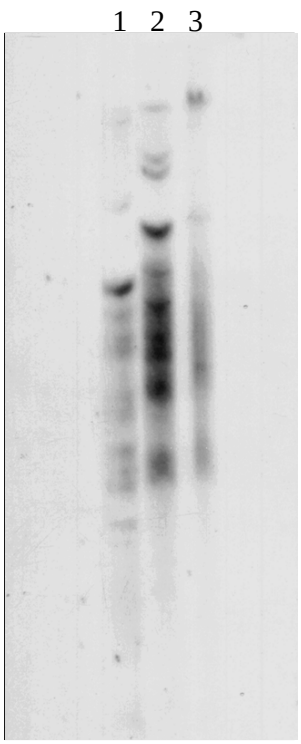


Figure 3.1: SH result for T3.

*Lane **1:** T3 Eco*RI *digested; **2:** T3 Hind*III *digested; **3:** T3 undigested.*

T39 had a diffuse pattern of probe binding. There was hybridisation of the HBV probe, but not so as to generate distinct bands. Rather an even smear in the background of the *Hind*III lane was visible, while other lanes and the autoradiograph as a whole remained free of any background hybridisation. The smear was not tapered towards the bottom, and was absent in the agarose gel, showing a lack of DNA hydrolysis. This could be indicative of integration sites differing with different hepatocytes. It is highly unlikely that this was because of replicative intermediates as no free-form HBV was evident in any of the samples.

In the case of T9, T14 and T70 (Table IV) DNA extracted from different areas of the tissue sample, did not yield the same pattern of integration upon secondary or tertiary analysis. It was found that in the case of these three samples, integrants clearly visible on one blot were not

present on another blot performed from a different DNA extraction, despite the DNA from both extractions having been of good quality, and the positive and negative controls on both blots having given the expected results. T9 had two integrants or none, and T14 had either none or 4, depending on the blot. T70 in particular had previously been found to contain one integrant when this sample was analysed overseas (personal communication, Dr C. Bréchot; INSERM, CHU Necker, Paris, France). In the present study it was shown to contain no integrants in one experiment and 6 in another.

Undigested DNA analysed by SH showed no fragments in the 3.1 to 3.2 kb region (established with radioactively labeled size markers) in any of the samples, indicating a lack of FL HBV DNA, and hence viral replication, in these tissues. The same applies to bands that could possibly indicate cccDNA and nicked cDNA.

It was not possible to construct RE maps of integrants because the fragments generated by RE digestion could not reliably be assigned to any one of the multiple integrants present in each sample (Figure 3.1). Double digests, in which the respective REs were combined, did not help to elucidate a RE map of the integrated HBV DNAs (Figure 3.2). The PLC/PRF/5 DNA has 6-9 fragments visible in the *Hind*III digested DNA – lane 3. T37 (lane 6) was shown to contain 7 integrants. At first glance, the undigested T37 appears slightly hydrolysed, however it does not taper and was not hydrolysed on the agarose gel. The slight smear is likely the result of DNA overload. T37 *Bam*HI digest (lane 7) yielded 8 fragments. T37 *Eco*RI digest yielded 12 fragments (lane 8). The band patterns are too complex for a map to be drawn as it is impossible to determine accurately which of the 8 and 12 fragments respectively was generated from each of the 7 fragments produced by *Hind*III digestion (See Appendix A, Figure A4). The double digests (lanes11-13) did not help to elucidate the RE maps for any of the integrants. Both the *Hind*III/*Bam*HI digest for example, yielded fewer fragments than there were integrants.

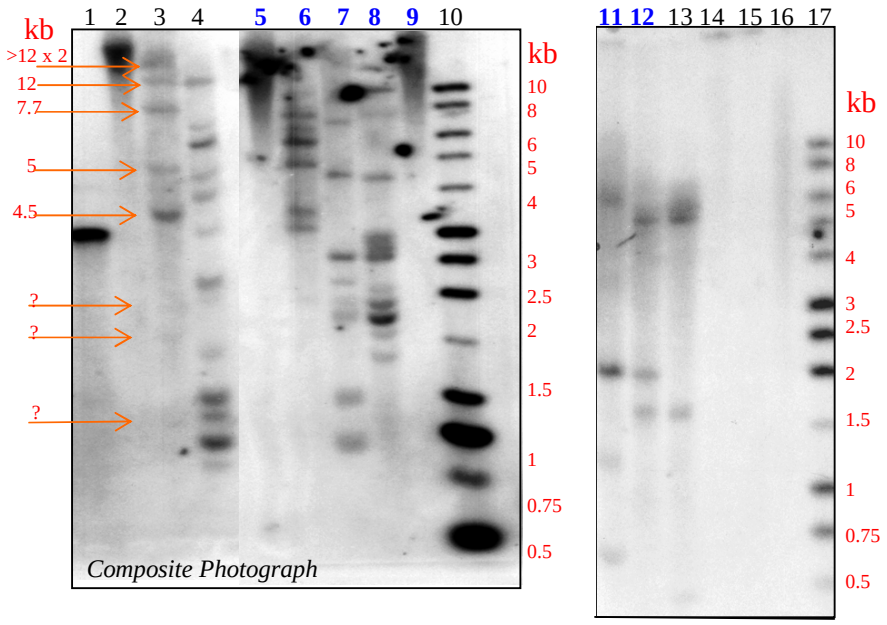


Figure 3.2: Autoradiographic exposure of SH results for PLC/PRF/5 cell line and T37 DNA, single and double RE digests.

HBV probe *Lane* ***1****: Hind*III*;* ***PLC/PRF/5 DNA*** *Lanes 2-4:* ***2:*** *undigested,* ***3:*** *Hind*III*,* ***4:*** *Bam*HI*;* **Sample** ***T37*** *Lanes 5-13:* ***5:*** *undigested,* ***6:*** *Hind*III*,* ***7:*** *Bam*HI*,* ***8:*** *Eco*RI*,* ***9:*** *Sty*I*,* ***11:*** *Hind*III*/Bam*HI*,* ***12:*** *Hind*III*/Eco*RI*,* ***13:*** *Bam*HI*/Eco*RI *; Lanes* ***14-16:*** *empty; Lanes* ***10*** *&* ***17:*** *Lambda DNA MWM (Promega).*

3.2.2 Integration in the PLC/PRF/5 and HuH7 cell lines

The HuH7 cell line, in agreement with the published literature (Nakabayashi *et al*., 1982) did not contain any HBV integrated DNA (Table IV).

The PLC/PRF/5 cell line was found to contain 6 or possibly 9 integrants (3 fragments were very faint) (Figure 3.2 and Table V), in agreement with previous reports (Zerial *et al*., 1986) (Table V), although the sizes of some of the integrants differed from those published.

Table V: Sizes of integrants in the PLC/PRF/5 cell line as established by SH.

Fragment Size (bp)	>12000	>12000	12000	7700	~5000	4500	1900	1700	1400
D/M/F	D	D	D	D	M	D	F	F	F

Key: band intensity on the autoradiograph dark (D), medium (M) or faint (F)

3.3 Discussion and Conclusion

3.3.1 PLC/PRF/5 and HuH7

The HuH7 cell line, in accordance with the published literature (Nakabayashi *et al*., 1982) does not contain any HBV integrated DNA and therefore served as a reliable negative control.

In agreement with previous reports (Zerial *et al*., 1986) the PLC/PRF/5 cell line, which served as the positive control, was found to contain 6 or possibly 9 integrants (Figure 3.2). Of the total fragments detected, 6 correlated in size with others documented in the literature. The integrants from the PLC/PRF/5 cell line are well documented in the literature (Table VI), but discrepancies exist between studies on the integrant status of the PLC/PRF/5 cell line DNA in terms of integrant number and size. The restriction patterns in the present study are in good agreement with those published by Chen *et al*., (1982), and several fragments agree in size with those of other publications (highlighted in Table VI).

Various studies have reported the number and size of integrants in the PLC/PRF/5 cell line and as is evident in Table VI, although the pattern is found to be similar between research groups, the number and size of the bands differ. Bréchot *et al*., (1980), (Table VI Ref 2) claimed that

their results were in agreement with those of Marion *et al*., (1980b), (Table VI Ref 5), although the number and size of the restriction DNA fragments were different. The former documented 5 fragments when PLC/PRF/5 DNA was restricted with *Hind*III, and 8 fragments when digested with *Eco*RI (Table VI). Fragments were sized by comparing to a *Bam*HI digested cloned HBV genome and the *Hind*III digested DNA of plasmid 8plac5cI857S7. Marion *et al.*, (1980b), on the other hand, found only three fragments in the PLC/PRF/5 DNA when this was restricted with *Hind*III. Fragments were sized by comparison to *Hind*III digested phage PM2 DNA and the HBV genome. In the study performed by Chakraborty *et al*., (1980), (Table VI Ref 1), four fragments were detected when digested with *Hind*III and four differently sized fragments when digested with *Eco*RI. The length of the fragments was established by comparison to the RE generated fragments of the pBR322 plasmid. Edman *et al*., (1980), (Table VI Ref 3) documented 6 fragments (4 200 to > 30 000 bp) upon digestion with *Hind*III and 10 with *Eco*RI. An intense band was observed at the 3 000 bp mark in the *Eco*RI digested DNA (the band sizes from this study listed in the table are estimations as per the autoradiographs shown in the original publication and are not conclusively stated as being of this size by the authors of the publication). Fragments were sized by comparison to the 3 200 and 1 400 bp fragments generated from a cloned HBV genome. Common to all these studies was the complete absence of discrete bands in undigested PLC/PRF/5 DNA.

SH was used in all the studies and protocols were virtually identical except for the probe activity and this may explain the discrepancies in the number of fragments reported. A higher number of integrants was detected when higher probe activity was used. As done in the present study, Zerial *et al*., (1986), used higher activity probe than that used in the majority of the other studies (8×10^8 cpm/µg DNA versus 1-2 x 10^8 cpm/µg DNA) and up to 9 integrants were detected.

Other differences may involve DNA transfer, where transfer may have been slightly more successful in some studies than in others, or blots exposed for slightly longer, and the fainter integrant bands may not have been clearly visibly. Small discrepancies in band sizes can be attributed to human error, as these fragments were approximately sized against radioactive markers of known size, rather than mathematically calculated. A good example is the dark 3 kb fragment obtained upon *Eco*RI digestion (Table VI row 16).

However, another and possible reason for the discrepancies (both in number and size of integrants) is that integrants may have undergone transposition and/or mutation. Bréchot *et al*., (1980), examined the RE pattern for the PLC/PRF/5 at both the 100^{th} and 140^{th} passages and found it to be identical. However, mutation and transposition events may require a significant time period in which to occur (ie. more than 40 passages) and the generation of significantly different restriction digest patterns would probably have resulted from differences in selection pressure in different laboratories, despite the obvious desire to have maintained culturing conditions as close as possible to those previously reported. These factors would be acting synergistically as, were it solely a matter of time, one would expect the later work done on the cell line (Koshy *et al*., 1983; Chen *et al*., 1986; Zerial *et al*., 1986) to have revealed more integrants than the earlier studies (Bréchot *et al*., 1980; Chakraborty *et al*., 1980; Edman *et al*., 1980; Marion *et al*., 1980b), which is not the case. All nine integrants detected in this cell line may not have resulted from independent integration events but rather from amplifications, deletions and translocations of the four sequences proposed to have integrated originally (Ziemer *et al*., 1985; Zerial *et al*., 1986; Kuo *et al*., 1998).

Table VI: A comparison of band sizes obtained with SH in the PLC/PRF/5 cell line when digested with *Hind* III and *Eco*RI between this study and other published works (where fragment sizes were published).

***Hind*III digested PLC/PRF/5 DNA fragment sizes-bp**									***Eco*RI digested PLC/PRF/5 DNA fragments-bp**				
1						>40000					≥12000		
2			>30000 x2							10000			10000
3						30000						~9500	
4		24000				24000					~8000	~8000	8000
5							23100			7800			
6		17000				17800	~18000	>12000x2		7000			7000
7		12000						~12000				~6900	
8				10700		10700						~6500	
9				10500		10500	~10500				~5700		
10			≥10000						4650				
11			9000							4000			~4100
12								~7 700		3800		~3800	
13	6650	6500			6500 F		~6500			3500	~3600	~3600	
14				6000		6000			3250				
15	5550 F		5500								~3200		3300
16					5000 D			~5000		3000 D	~3000 D	~3000	~3100 D
17	4500							~4500	2800			~2800	~2700
18						4300	~4300			2400	~2500		~2600
19		4200	4200										~2300
20	3600				3500 D				2100			~2100	
21						1900		~1900			~2000		
22								~1700			~1500		
23								<1400					
Ref	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**this study**	**1**	**2**	**3**	**7**	**this study**

KEY:
Fragments (bp) in common with other studies are highlighted in yellow
Band intensity on the autoradiograph is indicated as dark (D), medium (M) or faint (F)

REFERENCES

1. Chakraborty *et al*., (1980) *Nature* **286:**531.
2. Bréchot *et al*., (1980) *Nature* **286:**533.
3. Edman *et al*., (1980) *Nature* **286:**535.
4. Koshy *et al*., (1983) *Cell* **34:**215
5. Marion *et al*., (1980b) *J. Virol*. **33:**795
6. Zerial *et al*., (1986) *Nuc. Acids Res*. **14(21):**8373.
7. Chen *et al*., (1986) *Br. J. Exp. Path*. **67:**279.

3.3.2 South African HCC samples

Originally 18 samples were chosen for this study and 11 were eventually analysed by SH, only one was a T/NT pair. Overall, the SH results for the samples in this study correlated with those previously reported in the literature. Of the 11 samples analysed, 9/11 contained integrated HBV DNA, a total of ~81.8 %. These results were consistent with previously published reports where HBV DNA is found integrated in 80-100 % of HCCs worldwide (Shafritz *et al*., 1981; Bréchot *et al*., 2000; Murakami *et al*., 2005), and in black Africans where 80 % integration was reported in HBV-related HCCs (Bréchot *et al*., 1981; Shafritz and Kew, 1981; Buendia *et al*., 1992).

As expected, most of the samples (9/11) had multiple (between 2 and 10) integrants per sample. Usually three to four integrants are present, although as many as ten or more have been previously reported (Matsubara and Tokino 1990). Most African and Chinese HCCs contain multiple integrants (Matsubara and Tokino 1990) whereas in childhood and European HCCs, single integrants are usually the norm (Tanaka *et al*., 1986; Giacchino *et al*., 1987). Clonal expansion was observed in 7/11 samples and has been reported to occur in both early (Sheu *et al*., 1993) and late stage HCCs (Blum *et al*., 1987).

The T/NT14 pair from a HBsAg-ve individual had integrants in both the tumour and the non-tumorous tissue. Integrants have been documented in NT tissue, usually in HBsAg +ve patients, but also in those with markers of past infection (Kew, 2002b). Shafritz *et al*., (1981), showed integrated HBV DNA in anti-HBs+, HBsAg-ve black Africans. In HBsAg -ve patients, virions may still be present but may be missed because of a lack of sensitivity in the laboratory detection techniques employed, the formation of HBsAg-antiHBs immune complexes may render the HBsAg undetectable, or because the virions present are S gene mutants.

There were no fragments detected in the 3.1 to 3.2 kb region in the undigested samples in this study. However, should any characterised amplicons derive from free virus rather than integrated HBV DNA, this would later be resolved upon sequencing, as integrants should contain both HBV DNA and flanking genomic host sequences. Generally if no bands are visible in the undigested lane of a SH autoradiograph, those bands visible in the *Hind*III digestion lane are distinguished as true integrants and not free replicative HBV DNA (Ogston

et al., 1982). Furthermore, some authors have postulated that such HBV DNA may be present but not visible in the undigested lane because of having been trapped in the high molecular weight (HMW) DNA (Ogston *et al*., 1982). All the tissues studied were late stage tumours and replication of HBV is reported to be absent in such late stage HCCs by some groups (Nazarewicz *et al*., 1977; Matsubara and Tokino, 1990; Puisieux *et al*., 1993; Simon and Carr, 1995), but not others that maintain that the HCCs contain free replicating HBV (Chen *et al*., 1982). It has been proposed that the absence of replication may be the result of the poor differentiation of the HCCs (Nazarewicz *et al*., 1977).

In the case of two of the late stage tumours (T9, T14) integrants clearly visible on one blot were not present on another blot performed from a different DNA extraction, despite the DNA from both extractions having been of good quality and the controls on both blots having given the expected results and blank controls having confirmed an absence of contamination. In one case (T70) the sample clearly had 6 integrants (two of which were represented by dark bands of approximately the same intensity sized at 3.5 and 3.2 kb) but only one integrant had previously been documented during another study. Experimental error may be the cause. However, if the previous study documented one integrant it would be expected that the second band would be resolved, considering the two darkest bands were virtually same intensity and only 0.3 kb apart. Furthermore, on comparing the protocols used in both studies they were found to be very similar, and identical in terms of the amount of DNA resolved and transferred, as well as in the intensity of labeling of the probe (Paterlini-Bréchot; personal communication). Moreover, the other two samples (T9, T14) contained some integrants in one blot but no integrants in a second blot using DNA from a different extraction, even though both blots had been subjected to the same SH protocol and the controls worked, therefore experimental error is not the explanation.

This result is supported by the literature. Esumi *et al*., (1989), reported a case of a polyclonal advanced stage HCC where four sections of the tumour were examined. Three sections gave the same SH profile, while the SH profile of the fourth section contained an extra integrant. It was also determined that, on occasion, presumed monoclonal tumours are actually biclonal or multiclonal, and may contain one or more subpopulations of tumour cells without detectable HBV DNA (Esumi *et al*., 1989). In cirrhotic liver, HCCs develop within nodules, each of which may contain more than one hepatocyte clone, which in turn has a distinctive pattern of

integration (Aoki and Robinson, 1989; Robinson, 1994; Yasui *et al*., 1997). It is thought that early stage hepatocellular carcinomas are primarily of polyclonal origin (Sheu *et al*., 1993) while late stage HCCs are primarily monoclonal.

4.0 DETECTION OF EXPRESSED INTEGRANTS (RNA ANALYSIS)

4.1 Introduction

The third aim of the present study was to demonstrate expression, if any, of integrants in tumours derived from southern African HCC patients, by analysing total and messenger RNA.

The first demonstration of the expression of an integrated HBV genome in a HCC was in 1984 (Yaginuma *et al*., 1984). Since then a number of studies have documented integrants whose expression has resulted in:

- truncated proteins
- hybrid proteins primarily incorporating those regions of the HBV that are most commonly integrated (the HBV *S* and *X* gene sequences) (Section 1.5.1.2)
- over-expressed hybrid proteins driven by HBV promoters and
- over-production of host proteins from the effect of viral promoters (Table II section 1.5.1.1).

In the majority of cases, the ultimate outcome of these events remains unknown. In a few instances, however, the consequences have been deduced or proven. These include the following possibilities: contribution to cellular transformation, increase in the rate of cell division, aberrant cell growth, interference with cell proliferation, viability and transformation control capabilities, disruption of cell cycle control and apoptosis (Table II section 1.5.1.1). Expression of integrated HBV DNA has also been shown in the PLC/PRF/5 cell line (Section 1.5.3). It has therefore become increasingly evident that expression of HBV integrants may play an important role in the development and progression of HCC (Wang *et al*., 1990; 1992; Graëf *et al*., 1994; Berasain *et al*., 1998; Chami *et al*., 2000, 2001; Gozuacik *et al*., 2001; Tu *et al*., 2001; Paterlini-Bréchot *et al*., 2003).

The expression of integrants has thus far been studied mostly indirectly. Integrants have been detected by DNA analysis. Integrants and the adjoining host DNA are excised, cloned and expressed in an appropriate system, in an effort to establish the exact effect of the integration

event on cellular pathways, such as the cell cycle or apoptosis, and possibly on initiation and progression of HCC. Despite the increasing number of integrants analysed in recent years, only a few of these have been successfully functionally characterised integrants (Table II section 1.5.1.1). None of these integrants have been from African HCCs.

In this study, we attempted to detect expressed integrants directly, by analyzing RNA and mRNA from HCCs for HBV/host hybrid transcripts. In this manner, the limitations of the previously attempted PCR techniques could be overcome. To our knowledge, only three other studies exist where a technique has been modified or developed to investigate RNA from HCCs for expressed integrants (Paterlini *et al*., 1995; Poussin *et al*., 1999; Chen *et al*., 2000). These studies however, did not prove conclusively that the expressed mRNA definitely derived from an integrated HBV DNA, although in Poussin *et al*., (1999), it was highly likely (section 4.4.4.5).

In order to study the expression of integrants in a southern African HCC, a novel adaptation of the differential gene expression technique was designed and applied.

4.2 Materials and Methods

An outline of the techniques used in the analysis of RNA extracted from southern African HCCs is depicted in Figure 4.1.

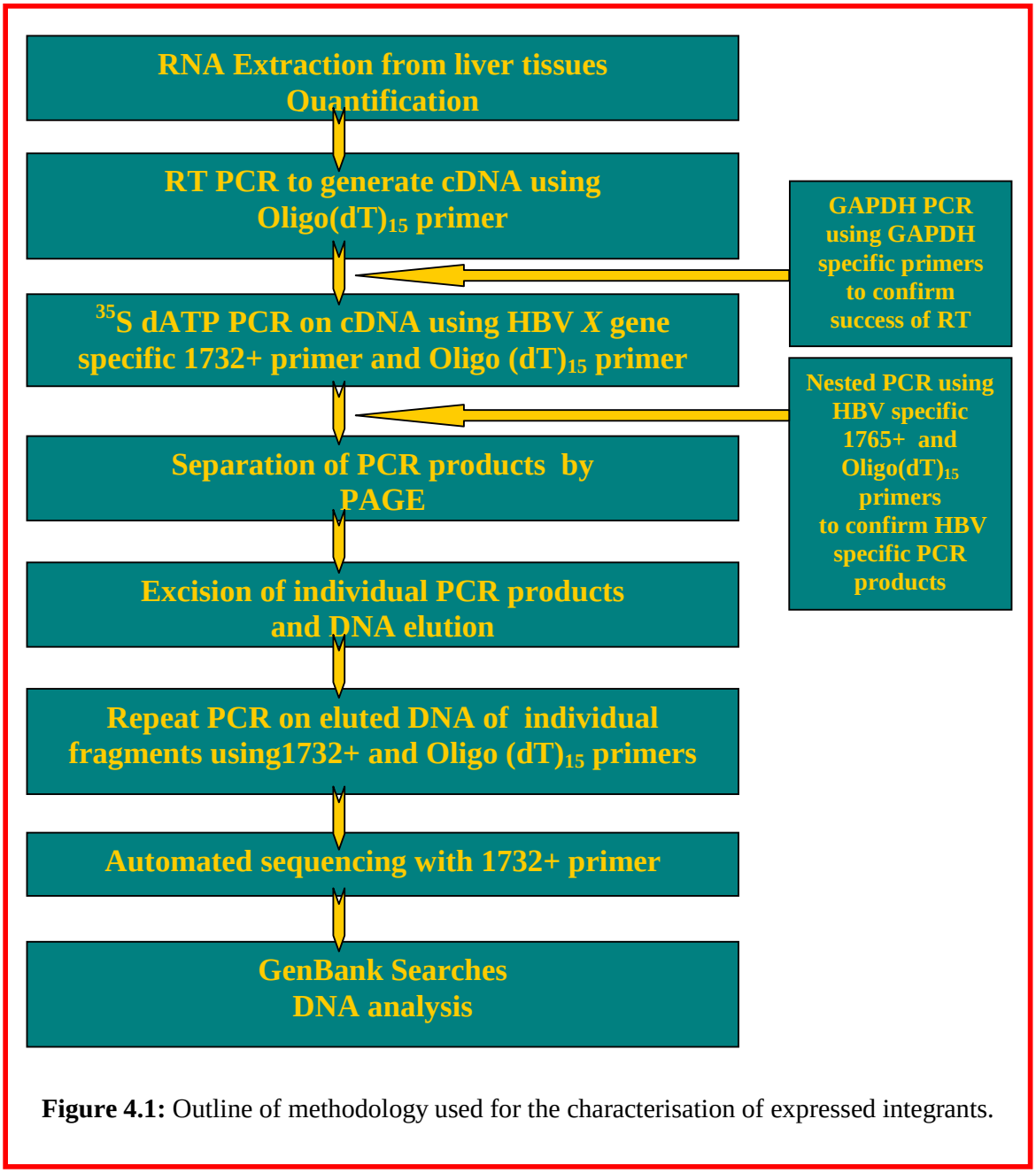


Figure 4.1: Outline of methodology used for the characterisation of expressed integrants.

4.2.1 RNA Extractions

RNA was extracted from frozen liver tissues (sections 2.2.1.1, 2.2.1.2), the HuH7 cell line, and the HBV replicating cell line HepG2.2.15 (Sells *et al.,* 1987) (section 2.2.1.3) using the classic Guanidium-Acid-Phenol Method (Chomczynski and Sacchi, 1987) or the RNeasy® Mini Kit (QIAgen) as per the kit protocol. RNA from a HBV negative liver (organ donor), and from the cell line HuH7, was used as a negative PCR control, while that of HepG2.2.15 was used as a positive control. PLC/PRF/5 RNA was not used as a positive control because the RNA required is HBV RNA, and the PLC/PRF/5 cell line, unlike HepG2.2.15, does not express all HBV RNAs but only that of the *S* gene.

4.2.1.1 Maintaining RNA Integrity

Both the tissue and cell samples were prevented from thawing during weighing and homogenizing by ensuring that all steps were performed on ice, and in liquid nitrogen (African Oxygen). Extraction of RNA involved batches of no more than 7 samples at a time, to ensure as short a handling time as possible. This prevented RNA degradation, change in gene expression, and subsequent false negatives by PCR.

The standard conditions and precautions for RNA extraction were followed, namely: the creation of ribonuclease free environment and reserving reaction components specifically for RNA work only (Appendix A section A5).

4.2.1.2 Guanidium-Acid-Phenol Method for the Extraction of RNA from Tissues or Cells (with Minor Modifications)

A 100 mg piece of liver tissue was cut quickly, on ice, from an interior section of the still frozen sample. The tissue was ground in a sterile, ribonuclease free mortar and pestle, with liquid nitrogen. While homogenizing, the liquid nitrogen was not allowed to evaporate leaving the sample exposed to ambient temperature. Thereafter liquid nitrogen was allowed to evaporate and 1 ml of denaturing solution (4 M guanidium thiocyanate; 25 mM sodium citrate, pH 7.0; 0.5 % sarcosyl; 100 mM 2-mercaptoethanol) (Appendix D) was added at ambient temperature. For harvested tissue culture HepG2.2.15 cells, 100 µl of denaturing solution was used per 10^6 cells. The homogenate was transferred to an ribonuclease free 4 ml disposable

polypropylene tube and the following reagents were added in order, with thorough mixing by inversion after each addition: 0.1 ml 2M sodium acetate, pH 4.0; 1 ml water saturated phenol (nucleic acid grade) (BDH, West Chester, PA, USA); 0.2 ml chloroform:isoamyl alcohol (49:1) (BDH). Once all reagents had been added the tube was shaken vigorously for 10 seconds. The sample was cooled on ice for 15 minutes; and centrifuged at 10 000 x *g* at 4 ºC, until the phases had separated completely. The aqueous phase was transferred to a fresh tube, an equal volume of isopropanol was added, and the sample was placed at -20 ºC for approximately 16 hours; followed by centrifugation at 10 000 x *g* for 20 minutes at 4 ºC. The supernatant was discarded, the RNA pellet dissolved in 300 µl of denaturing solution, and transferred to a fresh 2 ml tube. The RNA was precipitated by addition of an equal volume of ice-cold isopropanol at -20 ºC for 2 hours. The sample was centrifuged at 14 000 x *g* for 10 minutes at 4 ºC and the supernantant discarded. The pellet was washed twice with 500 µl of 75 % EtOH, re-centrifuged, and left to air dry for 10 minutes. The RNA was dissolved in 50 µl DEPC–treated water, with incubation in a dry block (Hägar) at 65 ºC for 10 minutes and was stored in 15µl aliquots, to avoid repeated freezing and thawing, at -70 ºC.

DNA digestion: RNA was rid of DNA contamination by digestion with 1 unit of deoxyribonuclease I (DNaseI) (SIGMA Bio Sciences) enzyme for every 1 µg of RNA, DNaseI, 1 x buffer (200 mM Tris-HCl pH8.3, 20 mM $MaCl_2$) (SIGMA Bio Sciences), RNase free water to a total volume of 50 µl; and incubated at 37 ºC for 30 minutes. Inactivation of the digest and removal of digested DNA was performed with the addition of an equal volume of RNA standard water-saturated P/C(49:1)isoamyalcohol in a 3 to 1 ratio (BDH). The reaction was vortexed vigorously for 10 seconds, incubated on ice for 10 minutes and centrifuged at 10.000 x *g* at 4 ºC for 15 minutes. The aqueous phase was transferred to an RNase free tube, and the RNA precipitated with 0.3 M NaAc, and 2 volumes 100 % EtOH at -70 ºC for 1 hour. After centrifugation at 10.000 x *g* and 4^0C for 15 minutes, the pellet was briefly rinsed in cold 70 % EtOH. RNA was air dried for a few minutes and re-suspended in RNase free water to a final concentration of 0.5-2.5 µg/µl.

4.2.1.3 RNA Quantification

RNA concentrations were obtained with the use of spectrophotometry and quartz cuvettes cleaned with 3 % H_2O_2 for 15 minutes and calculated as follows: RNA concentration = 40

µg/ml x A260 x dilution factor. The dilution factor used was 200 in DEPC treated water. Purity A260/A280 was 1.8 (ideal range 1.8-2.1). (Protocol Online: http://www.protocol-online.org/prot/u7GeneralLaboratoryTechniques/Quantitation/RNA Quantitation/
and RNA Isolation and Quantification Protocol: http://www.ucc.ie/ucc/depts/physio/MolPhysiol/ MolPhys/ Protocols/pdfs/001.pdf).

4.2.2 RT-PCR

DNA and RNA polymerases with proofreading functions were selected for PCR, and appropriate ones were selected for long template (~40 kb) amplification. RNA was reverse transcribed using AMV Reverse Transcriptase (Promega) as per manufacturers instructions, or the ImpromII™ Reverse Transcription System (Promega). The ImpromII™ Reverse Transcription System kit protocol has been optimized to produce FL cDNA from rare or long messages. The PCR components were defrosted and assembled on ice, and no more than 8 samples (including controls) were processed in order to reduce handling time to less than one hour.

4.2.2.1 RT-PCR Controls

The controls included:

- *negative control 1:* one extraction without cells or tissue, these were replaced by nuclease-free water, confirming absence of any contaminating material in the extract (extraction control);
- *negative control 2*: for each sample two RT-PCRs were performed, one containing template and reverse transcriptase enzyme; the other containing template and no reverse transcriptase – this permitted confirmation that amplification was off a RNA template and not off a contaminating DNA template; The expressed T45.3 sequence was generated from RNA. DNA was digested with DNaseI, and there was a lack of amplification in reactions lacking the reverse transcriptase enzyme. These controls were performed with primers for the beta-catenin gene in the concurrent study from which the aliquot of RNA was 'borrowed' and were not repeated with HBV specific primer because of the limited supply of tissue (Elmileik *et al*., 2005).

- *negative control 3*: template replaced by nuclease-free water, confirming absence of any contaminating material inadvertently introduced during the RT-PCR set-up (RT-PCR control)
- *negative control 4*: HuH7 RNA template, which contains no HBV RNA, confirming specificity of PCR;
- *negative control 5:* RNA template extracted from HBV negative liver tissue, which contains no HBV RNA, confirming specificity of PCR;
- *positive control:* template was RNA extracted from HepG2.2.12 cells, liver cell line known to express HBV, and HBV mRNA.

RT-PCR

(1) Experimental: 1 µg of total or mRNA; 0.5 µg Oligo $(dT)_{15}$ (Appendix C); nuclease free water to a total volume of 5 µl; (2) Positive control: 1ng total or mRNA from HepG2.2.15 cells; remainder as per experimental; (3) Negative control: no RNA – volume replaced with RNase-free water; remainder as per experimental. The positive control was placed in tube 8, after the RT-PCR negative/blank control, to ensure it was not carried over to sample amplification reactions. Once RNA, primer and water had been combined, the reactions were placed at 70 ºC for 5 minutes on a dry block, and chilled in ice water for 5 minutes. Tubes were centrifuged for 10 seconds to precipitate condensation, and placed in ice water until the reaction mix was added.

Reaction mix: This was combined in a 1.5 ml nuclease-free polypropylene tube, on ice. A 15 µl aliquot of mix per sample was prepared, which would result in a final reaction volume of 20 µl. Mix: 4.0 µl of 5 x reaction buffer; 1.5 mM $MgCl_2$; 0.5 mM each dNTP; 20 units of Recombinant RNasin® Ribonuclease inhibitor; ImpromII™ reverse transcriptase (Promega) 1 µl (enzyme concentration undisclosed); nuclease free water to final volume. A separate reaction mix was combined for the reverse transcription negative control where the ImpromII™ reverse transcriptase was replaced by 1 µl of nuclease free water. The reaction mix was dispensed in 15 µl aliquots into the 0.5 ml nuclease-free polypropylene tubes containing the 5 µl RNA/primer/water on ice, and overlaid with a drop of mineral oil. Amplification was performed in a pre-heated Robocycler® Temperature Cycler (Stratagene), and equilibrated at 25 ºC for 5 minutes. Extension was at 42-55 ºC for 1 hour, followed by inactivation of the reverse transcriptase at 70 ºC for 15 minutes.

4.2.3 Amplification of the Housekeeping Gene *Glyceraldehyde-3-phosphate Dehydrogenase* (GAPDH) from cDNA Template (Weinberg *et al.*, 2000)

The housekeeping gene, *GAPDH* is constitutively expressed (Fort *et al*, 1985), and served as a control for successful RT-PCR.

4.2.3.1 PCR

10 µl RT-PCR product/cDNA; 1 x reaction buffer (TAKARA SHUZO CO.); 25 µM each primer (Appendix C); 0.2 mM dNTP; and RNase free water to a total volume of 49 µl; 1 unit of *TaKaRa Ex Taq*^{TM} DNA polymerase (TAKARA SHUZO CO.).

Amplification was performed in a TECHNE Thermal Cycler (Staffordshire, UK). An initial denaturation step at 96 ºC for 10 minutes was followed by 56 ºC for 1 minute. The reaction was paused and the *TaKaRa Ex Taq*^{TM} DNA polymerase added to each reaction tube. Amplification: 30 cycles of: 94 ºC for 1 minute; 52 ºC for 30 seconds; 72 ºC for 1 minute and 30 seconds; with a final extension step at 72 ºC for 10 minutes. The product was resolved on a 1 x TBE / 1 % / 3 % EtBr agarose gel.

Controls: all the control reactions from the previous (RT-PCR) amplification reaction were included, along with fresh water blanks to ensure no contamination was inadvertently introduced in this PCR.

4.2.4 Amplification of Integrants Using *X* Gene Primers 1732+ and 1765+

4.2.4.1 Amplification of Integrants Using *X* gene Primer 1732+, From cDNA Template or DNA PCR Product Purified From Polyacrylamide Gels

Given that almost 25 % of integrants include DR repeat regions of the HBV genome (Shih *et al*., 1987; Robinson *et al*., 1992), it was logical to use an *X* gene primer. The PCR reaction consisted of 5 µl cDNA; 2.5 units of *TaKaRa Ex Taq*^{TM} DNA polymerase; 1 x reaction buffer; 1 µM HBV X gene sense primer 1732 (Appendix C); residual Oligo $(dT)_{15}$ primer from RT-PCR; 0.2 and 0.8 mM dNTP; and RNase free water to a total volume of 25 µl. Amplification

was performed in a TECHNE Thermal Cycler. An initial denaturation step was performed at 94 ºC for 3 minutes, followed by 35 cycles of 94 ºC for 30 seconds; 62 ºC for 50 seconds; 70 ºC for 1 minute and 30 seconds; and a final extension at 72 ºC for 10 minutes. Conditions for re-amplification of PCR products extracted from polyacrylamide gels were identical with the following exceptions: final dNTP concentrations are 20 µM instead of 2-4 µM, and no isotope is added (Zhu and Liang, 1997).

Amplification of Integrants Using Nested *X* Gene Primer 1765+

This reaction amplified products from the previous round of PCR (1732) for further confirmation.

Five µl PCR product (1732 PCR); 2.5 units of *TaKaRa Ex Taq*™ DNA polymerase; 1 x reaction buffer; 1.25 µM HBV *X* gene sense primer 1765 (Appendix C); 1.25 µM Oligo $(dT)_{15}$ primer; 0.8 mM dNTP; and RNase free water to a total volume of 25 µl. Cycling conditions were as before, with the exception of the annealing temperature which was 58 ºC. Conditions for re-amplification of PCR products extracted from polyacrylamide gels were identical with the following exceptions: final dNTP concentrations are 20 µM instead of 2-4 µM, and no isotope was added (Zhu and Liang, 1997).

Controls: as for *GAPDH* PCR (section 4.2.3). For DNA extracted from the polyacrylamide gels a further control was included. A DNA extraction was performed on an area of gel which should have been DNA free. If 5 µl of this extract failed to amplify, it was concluded that no contamination was introduced during the electrophoresis or gel DNA extraction procedures.

4.2.4.2 Amplification of Integrants Using 1732+ α^{35}SdATP Radioactive Isotope

This reaction enabled the visualization of the PCR product on polyacrylamide gels/autoradiographs, in order that the different sized fragments might be excised, purified, amplified and sequenced separately.

The PCR comprised: 5 µl of cDNA (for 1732 PCR); 2.5 units of *TaKaRa Ex Taq*™ DNA polymerase; 1 x reaction buffer; 1.25 µM HBV *X* gene sense primer 1732 (sequence Appendix C); 1.25 µM Oligo $(dT)_{15}$ primer; 0.8 mM each dCTP, dGTP, dTTP; 1 µl of 1:5 dilution of

α^{32}S dATP; and RNase free water to a total volume of 25 µl. Cycling conditions were as before (section 4.2.4.1).

4.2.5 Polyacrylamide Gel Electrophoresis (PAGE) and Extraction of DNA from the Polyacrylamide Gel

The radioactive 1732 PCR product was fragmented on 6 % polyacrylamide denaturing gels (Appendix D) in order that the different products might be separated and extracted from the gel for re-amplification. This technique was a slight modification of that used in Gene Differential Display (DD), adapted from Zhu and Liang, 1997.

A total of 20 µl of the PCR product was mixed with 15 µl of loading dye (Appendix D). A gap of 1 cm was left between samples to ensure that no cross-contamination occurred. One hundred bp and 1 kb DNA molecular weight markers (Promega, Madison, USA) were radioactively labeled (Appendix D) and 5 µl of each was mixed with 5 µl loading dye and loaded at the end of the gel. Gels were electrophoresed at 60 W for 7 hours, transferred to Whatman 3MM paper (Kent, UK), and covered with clean plastic film. Gels were dried on top of a second sheet of Whatman 3MM paper for 2 hours at 80 ºC, on a Bio-Rad Gel Dryer 583. Thereafter gels were placed in an Okamoto cassette and exposed to autoradiography film for 48 hours. Film was developed in an automatic developer for 30 seconds.

In order to excise the bands of interest, the first autoradiograph was wiped clean with 3 % H_2O_2, to avoid carry over of contaminating DNA to the gel. The bands of interest, numbered on the records autoradiograph, were marked with ink on the film taped to the gel, and their positions transferred onto the gel by piercing through the film and gel with a new sterile needle for each band. The bands were individually excised using a fresh sterile blade each time, and placed in 2 ml polypropylene tube with using 3 % H_2O_2 sterilised tweezers. The tubes were labeled according to numbering on the record film. Each band yielded a section of gel/filter paper approximately 3 cm long by 0.4 mm wide. A 200 µl aliquot of ddH_2O was added to each tube, and the gel/paper squashed into the tube bottom. After a 5 minute soak, another 200 µl water was added, followed by a second soak for 30 minutes. The tube cap was tightly closed, sealed with parafilm (Pechinery Plastic Packaging, Menasha, USA), and the sample/band boiled at 99 ºC on a dry heating block for 15 minutes. A centrifugation step at 2000 x *g* was

performed to collect condensation, pellet paper and gel debris. The supernatant was transferred to a fresh tube, with 40 µl of 3 M NaOAc, 10 µl of20 mg/ml glycogen (Sigma-Aldrich), 1.8 ml 100 % molecular grade EtOH; and placed at –20 ºC overnight. The sample was centrifuged at 2000 x *g* for 10 minutes at 4 ºC to pellet DNA, and the supernatant discarded. The pellet was rinsed in 85 % ice cold EtOH, briefly centrifuged, and residual EtOH aspirated. The DNA pellet was dissolved in 10 µl of sterile ddH_2O, at room temperature on a rotating wheel. A 5 µl aliquot of this DNA was used for re-amplification by PCR, while the remainder was stored at –20 ºC for future use. For a more detailed explanation and a diagrammatic representation of this please refer to the Appendix A section A6 and Figure A7.

4.2.6 Gel Electrophoresis of RNA

Five µg of RNA was visualized by AGE on 1 % D1-LE agarose (Burgos, Spain) in DEPC-treated ddH_2O, 0.66 M formaldehyde, 1 x MOPS (Appendix D) gels. To ensure that the RNA migrated only with respect to its molecular weight, 5 µg (up to 4.7 µl) of RNA was denatured in 2 µl of 5 x MOPS buffer, 3.3 µl of formaldehyde, and 10 µl of formamide and heated in a dry block at 55 ºC for 15 minutes. Each sample was combined with 2 µl of loading buffer (50 % glycerol; 1 mM EDTA pH 8.0; 0.4 % bromophenol blue; 0.4 % xylene cyanol) and resolved by electrophoresis at 5 V per cm of gel length, until the bromophenol blue dye (Appendix D) had migrated to approximately 80 % of the gel length. Gels were stained in 0.5 µg/ml EtBr/DEPC treated water for 1 hour, de-stained with gentle shaking in DEPC treated water for 16 hours and visualized and photographed on an UV transilluminator. RNA concentration was confirmed against RNA samples of known concentration, and sized against 3µl of RNA markers (range 281-6.583 bp, 9 bands) (Promega). RNA was considered to be un-degraded if the 28S and 18S subunits were present, and the 28S rRNA band appeared twice as intense as the 18S rRNA subunit band. mRNA could often be seen as a very faint smear in the lanes.

4.2.7 Sequencing

Sequencing reactions were analysed by automated sequencing at Inqaba Biotechnical Industries (Pty) Ltd, Hatfield, RSA with a Spectrumedix SCE2410 genetic analysis system with 24 capillaries (SpectruMedix LLC in Pennsylvania, USA).

4.2.7.1 Sequence data analysis

Sequences were read using CHROMAS 145.exe (Free of charge from: http://www.technelysium.com.au/chromas14x.html accessed 2004-2006) and aligned using GenDoc (current version 1.0-beta5, Free of charge from: http://gendiapo.sourceforge.net). Matched to GenBank (http://www.clcbio.com/index.php?id=442).

4.3 Results: Detection of Expressed Integrants (RNA Analysis)

4.3.1 Study for RNA Analysis

This study was performed to establish whether expression of integrated HBV DNA could be detected in any of the southern African samples. RNA that had been extracted for another study was reverse transcribed for T14, T24, T31, T40, T45, T69, T110. The reverse transcription reaction was assessed by successful amplification of the housekeeping gene *GAPDH* (section 4.2.3). cDNA was amplified with the use of the HBV-specific primer (1732+) located in the *X* gene, chosen because of its proximity to the preferred viral integration site DR1 and the cohesive overlap "hot-spot" region for integration (Dejean *et al*., 1984; Quade *et al*., 1992). mRNAs expressed off an integrated HBV DNA would theoretically generate a PCR product of integrated HBV DNA containing the 1732+ sequence, coupled to genomic sequence with a termination codon, polyadenylation signal, and poly(A) tail. Transcripts from free HBV genomes would also amplify with this technique. In late stage HCCs this is unlikely to occur considering that the replication of HBV in these is reported to be absent (Nazarewicz *et al*., 1977; Matsubara and Tokino, 1990). However, even if such transcripts are amplified they are distinguishable from integrant transcripts as they will be the standard size HBV transcripts, and will not be found adjacent to host genomic DNA. Legitimate integrant transcripts can be traced back to the host DNA.

In this study, the PCR reaction amplified four of the seven RNA extracts (T14, T31, T110, T45), two of which had multiple products (T31, T45) (Figure 4.2).

A hemi-nested PCR was performed using 1765+/Oligo $(dT)_{15}$ primers. The products that had amplified successfully in the initial PCR again amplified well, while those that initially amplified poorly or failed to amplify again amplified poorly or failed to amplify (T24, T31 and

T69). This indicated that amplicons for the T14, T31, T40 and T45 samples were most likely HBV-specific (Figure 4.3).

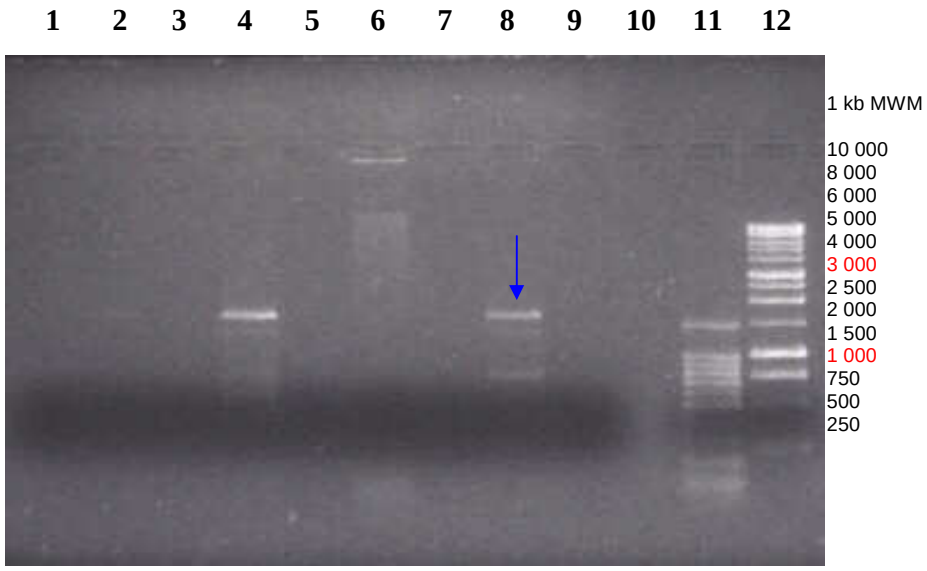


Figure 4.2: Agarose gel resolution of PCR product using 1732/Oligo d(T)$_{15}$ primers.

*Lanes **1,3,5,7,10:** Empty; **2:** T14; **4:** T31; **6:** T110; **8:** T45; **9:** PCR negative (water) control; **11and12:** 100 bp and 1 kb MWMs respectively.*

PCR amplicons were generated for T14 (1600 bp), T31 (1650 bp, 850 bp, 600 bp), T110 (40 kb), and T45 (1550 bp, 1400 bp, 800 bp). Sample T110 showed slight smearing, possibly indicative of non-specific amplification.

Figure 4.3: Agarose gel resolution of hemi-nested PCR product with primers 1765+/Oligo $d(T)_{15}$from cDNA.

Lanes **1:** PCRI water control; **2:** T14 (1600 bp); **3:** T24; **4:** T31 (1650 bp); **5:** T69; **6:** T110; **7:** T40; **8:** T45 (1550 bp); **9:** PCRII water control; **10:** 1 kb molecular weight marker (Promega).

Nested PCR successfully amplified fragments in T14; T31; T40; T45.

The PCR amplification of the samples T14, T31, T110 and T45 was repeated using 1732+ and Oligo $(dT)_{15}$ primers, but this time the dATP nucleotide was substituted with radioactive $\alpha^{35}S$ dATP (Figure 4.4). Three of the four samples amplified as before but T110 failed to amplify and was discarded. These three samples were resolved by polyacrylamide gel electrophoresis (PAGE).

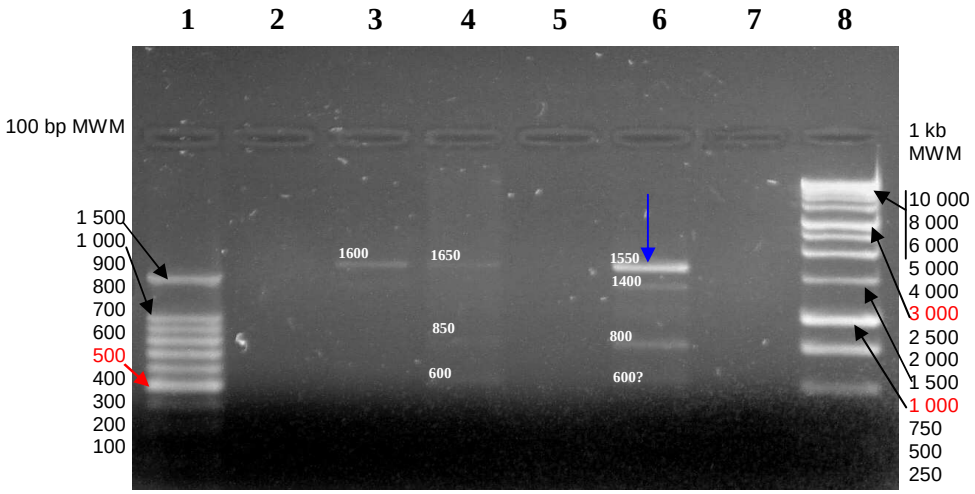


Figure 4.4: Agarose gel resolution of radio-active PCR fragments generated with primers 1732+/Oligo d(T)$_{15}$ and α^{35}S dATP.

*Lanes **1&8:** 100 bp and 1 kb MWMs (Promega) respectively; **2&7:** PCR water controls; **3:** T14; **4:** T31; **5:** T110; **6:** T45.* T14, T31 and T45 yielded the same fragments as before (1600 bp; 1650 bp, 850 bp, 600 bp; and 1550 bp, 1400 bp, 800 bp respectively and possibly 600 bp?).

It could be argued that there was a chance that some replicating virus could persist in these samples, despite no amplification occurring in those reactions lacking reverse transcriptase. It was therefore reassuring to find that of the products generated by 1732+/Oligo (dT)$_{15}$, only two (T31 and T45: 850 and 800 bp products) corresponded to the expected size of the HBV *X* gene mRNAs generated by replicating virus (0.7-0.9 bp) (Guo *et al*., 1991; Yaginuma *et al*., 1993; Chisari, 2000). Whether this RNA product was generated off a viral genome or a HBV integrant, would be resolved upon sequencing. The fragments on the autoradiograph were compared to those from the agarose gel in terms of size and comparative intensity (Figures 4.5, 4.6 and Table VII). Fragments of the correct size, intensity, and sufficient concentration, were good candidates for sequencing (Table VII). Fragment T45.3 (1550 bp) was particularly promising as it was an intense band.

Upon PAGE, the number of fragments for each sample increased as the resolving power of polyacrylamide is greater than that of agarose. The 1732+ PCR blank controls were also

subjected to autoradiography, to confirm the absence of contamination and residual DNA. The PAGE (Figure 4.5) distorted during electrophoresis and the fragments were not in straight lines. The 1600 bp fragment present on the agarose gels for T14 was evident (red dots have been placed on the photograph to help indicate the location of the fragments). There was also a somewhat smaller fragment present which was very faint and therefore not used. Such fragments are expected as the resolving power of polyacrylamide is greater than that of agarose. All three of the fragments (1650 bp; 850 bp; 600 bp) apparent on the agarose gels for T31 were present, as were the 3 fragments for T45 (1550 bp; 1400 bp; 800 bp). There were an additional 4 fragments (>1550 bp) present for T45. All fragments were of approximately the same intensity as when viewed on the agarose gel. Fragment T45 (1550 bp) was an intense band.

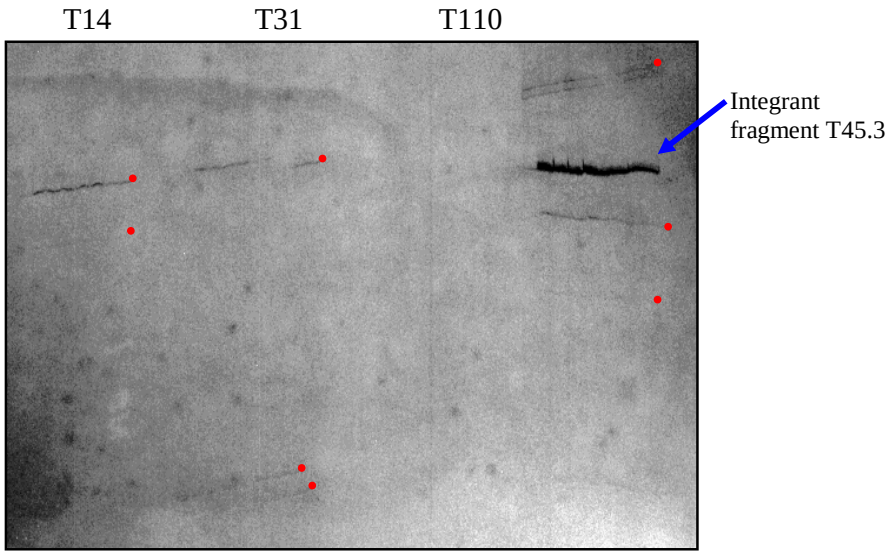


Figure 4.5: Autoradiograph of polyacrylamide gel resolution of radio-active PCR products generated with primers 1732+/Oligo $d(T)_{15}$ and $\alpha^{35}S$.

Lanes T14; T31; ***3:*** *T110;* ***4:*** *T45.* (Red dots have been placed on the photograph to help indicate the location of the fragments). For this gel several wells were combined to form one large well (~ 5 cm long) for each sample.

The DNA from each fragment on the polyacrylamide gel was eluted. The amplicons resolved on the polyacrylamide gel were numbered from the largest to the smallest as T14.1 (1600 bp); T31.1 (1650 bp), T31.2 (850 bp) etc. to facilitate keeping track of the fragments during elution

and further amplification. Each eluted DNA was re-amplified with the 1732+/Oligo d(T)$_{15}$ PCR and resolved on an agarose gel (Figure 4.6). The PCR products were as follows: T14.1 (lane2 - 1600 bp), T31.3 (lane 5 - 600 bp), T45.2 (1550 bp -lane 8) and T45.3 (1400 bp - lane 9) - the fragments are present and the sizes correct. In lane 3 the fragment of 1600 bp (T31.1) is absent. This reaction was discarded. The fragments in lanes 4 (550 bp), 6 (750 bp), 7 (750 bp) and 10 (500 bp) were not the expected sizes of T31.2 (850 bp), T45.1 and T45.2 (both >1550 bp) and T45.4 (800 bp). These samples were discarded because they were the wrong size, possibly indicating non-specific priming.

Figure 4.6: Agarose gel resolution of 1732+/Oligo d(T)$_{15}$ PCR products, eluted from polyacrylamide gel, and re-amplified with the same primers.

Lanes: ***1&12:*** *PCR negative controls,* ***2:*** *T14.1;* ***3:*** *T31.1;* ***4:*** *T31.2;* ***5:*** *T31.3;* ***6:*** *T45.1;* ***7:*** *T45.2;* ***8:*** *T45.3;* ***9:*** *T45.4;* ***10:*** *T45.5;* ***11****: gel negative control;* ***13:*** *empty;* ***14&15:*** *100 bp and 1 kb molecular weight markers respectively (Promega).*

The T45.3 fragment T45.3 generated a good quality sequence upon automated sequencing with the primer 1732+. Of the initial (1550 bp) amplified (Figure 4.7), 397 bp was successfully sequenced, with 322 bp comprising HBV genome encoding the end portion of the *X* gene and the beginning of the core gene (nucleotides 1820 to 2142) (Appendix E). The remaining 74 bp was not HBV sequence. Upon phylogenetic analysis, the partial sequence of the T45.3 integrant was found to cluster with the HBV genotype E isolates (Figure 4.8).

Table VII: Product sizes of samples amplified by 1732+/Oligo $(dT)_{15}$ radioactive PCR, resolved by agarose gel electrophoresis and visualised by autoradiography.

Sample designation	PCR product size (bp)	Fragment number	Decision	Result
T14	1 600	T14.1	sequence	N1 sequence
T31	1 650	T31.1	sequence	genomic
	850	T31.2	sequence	N1 sequence
	650	T31.3	sequence	genomic
T45	>1 550	T45.1 nvoa	Discard-insufficiently concentrated	NA
	>1 550	T45.2 nvoa	Discard-insufficiently concentrated	NA
	1 550	T45.3	sequence	HBV integrant
	1 400	T45.4	sequence	N1 sequence
	800	T45.5	sequence	N1 sequence

KEY: nvoa = not visible on agarose gel; NA = not applicable

N1 (multiple) sequences are unreadable because of their superimposition. Inqaba Biotechnology, proposed that these are generated when more than one round of PCR amplification is performed and the second round of amplification involves 35 cycles or more. Such sequences were most likely produced as a consequence of mis-priming of the DNA, upon re-amplification with the Oligo $(dT)_{15}$ primer.

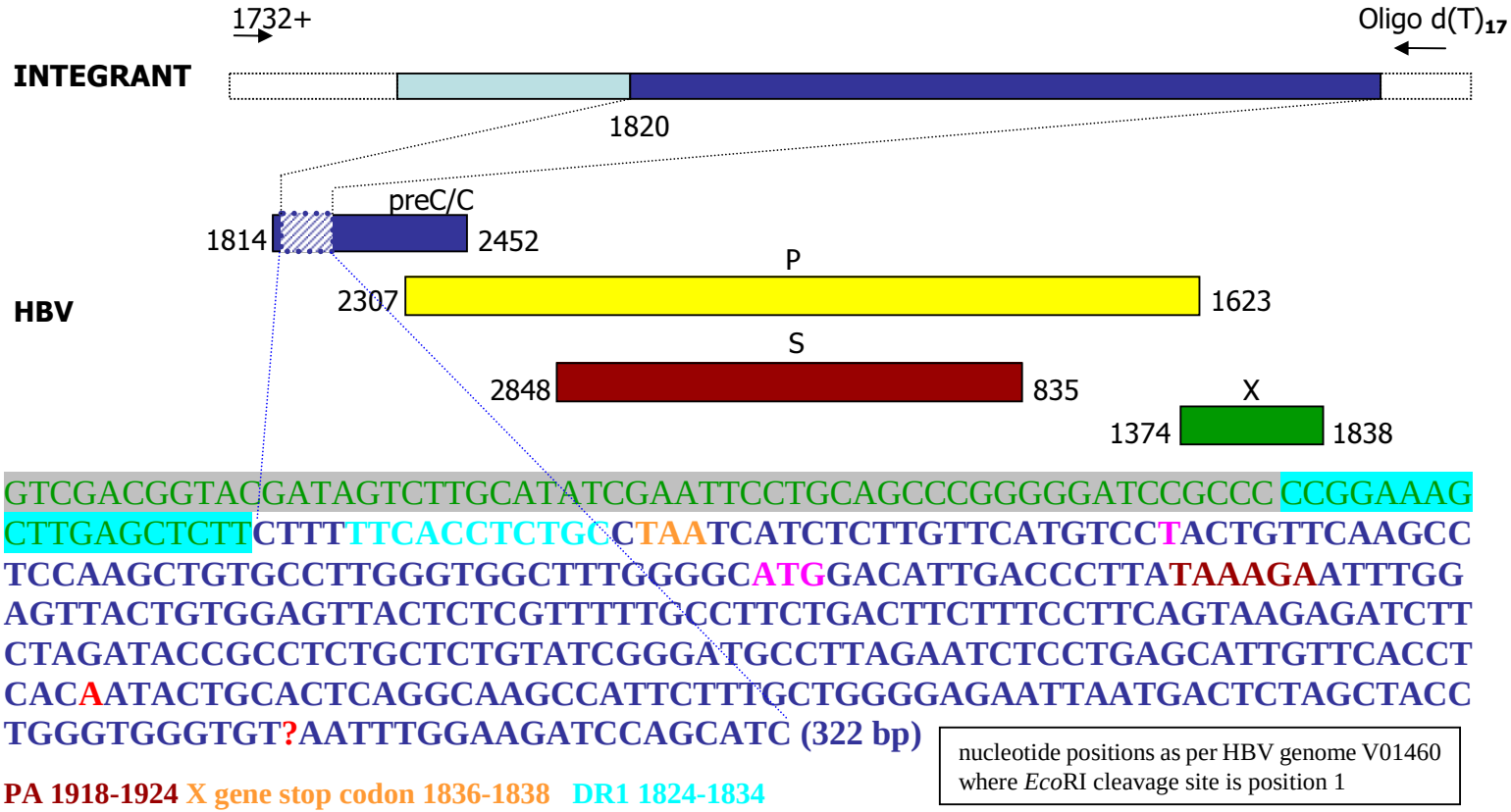


Figure 4.7: Diagrammatic representation of HBV sequence from T45 fragment number 3.

Gendoc alignment and Genbank Blast searches revealed 322 bp of HBV sequence (bp 1820-2143), phylogenetically clustering with the genotype E isolates. The remainder of the sequence appeared rearranged, with 55 bp mapping to the *Homo sapiens* heavy chain variable region (4e-13) of the *VH1* gene (Identities = 53/57) (92%) or possibly to cloning vector pPCR-Script Amp SK(+) (4e-13) (Identities = 53/57) (92%) and 19 bp mapping to chromosome 16 clone CTD-322SF13 (2.9) or possibly HBV DNA , (Identities 17/17) (100%). There appears to be an A at position 2054 instead of a C and a T missing at position 2122.

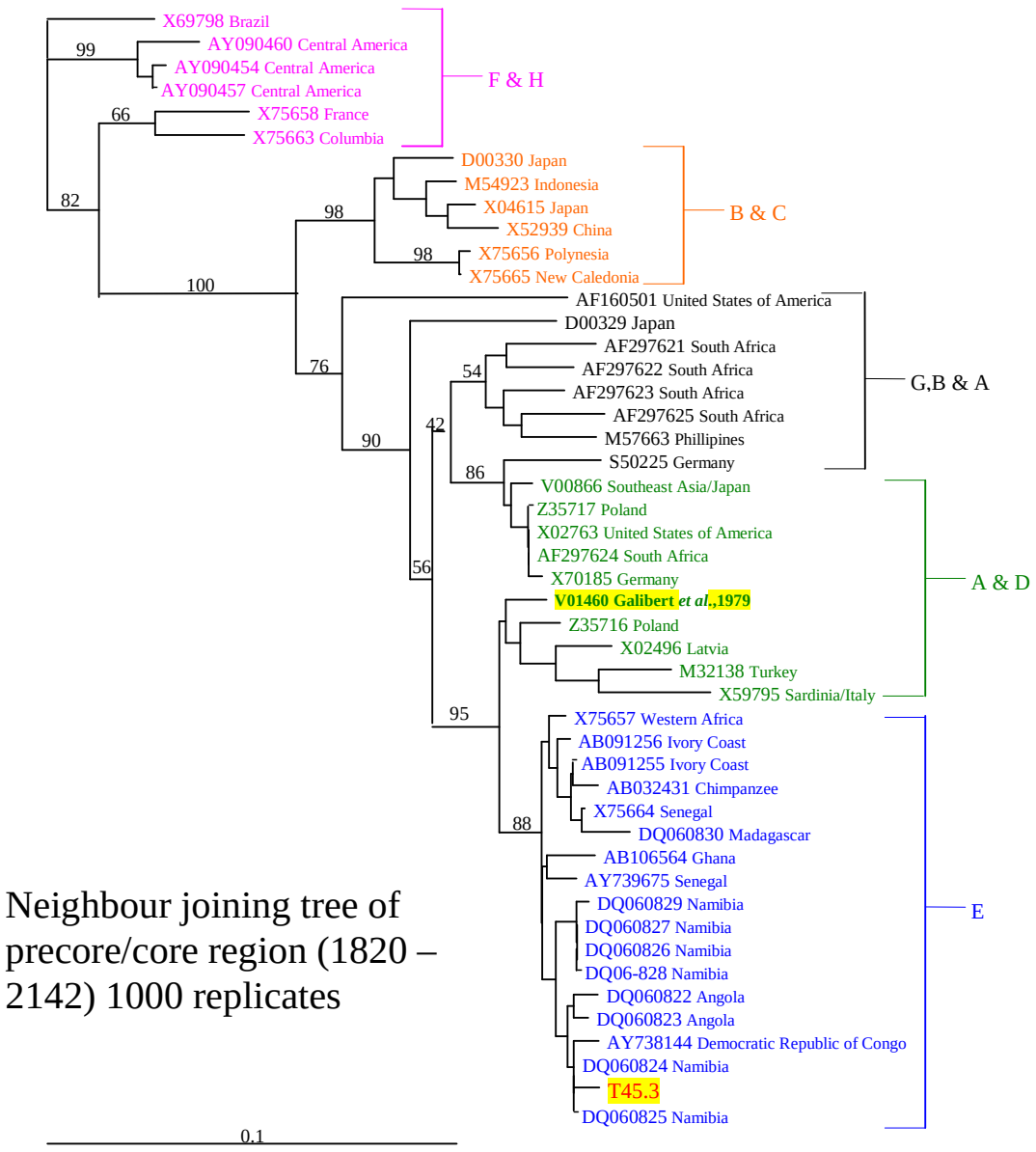


Figure 4.8: Phylogenetic tree showing the relative positioning of the partial sequence of the T45.3 to a representative selection of international HBV isolates. The partial sequence of the T45.3 clustered with the HBV E genotype isolates. The D genotype sequence originally published by Galibert *et al*., (1979) clusters with the South African D genotype.

Further characterisation of the DNA was attempted. It was possible to use a sense primer further along the sequence already obtained, in order to extend the sequence downstream. Amplification with the 1898(+)/Oligo $(dT)_{15}$ primers yielded a definite product, but unfortunately it was insufficiently concentrated for sequencing (Figure 4.9). A PCR was also attempted with the 1732/1966 primers, but with only half the amount of cDNA template as the remainder had been depleted. Unfortunately there was very little cDNA remaining, and the initial RNA extract yielded only enough RNA for these amplifications. As previously stated this sample was also used in another study and the tissue became depleted (Elmileik *et al*., 2005). The 1732/1966 PCR failed to amplify, despite the original sequencing data having indicated the presence of the sequence used for the 1966 primer. Further analysis of T45.3 became impossible.

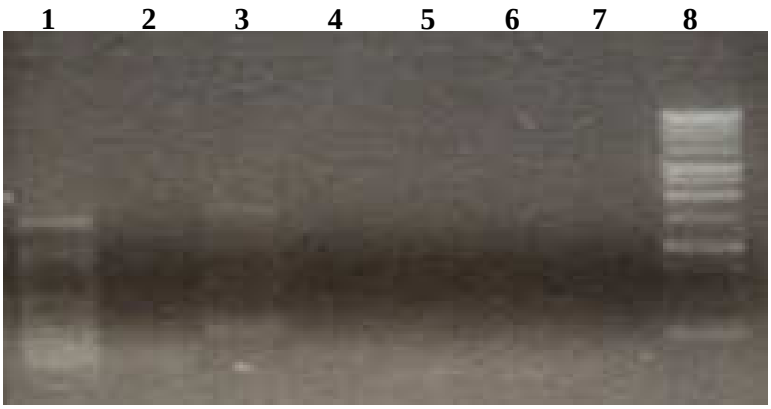


Figure 4.9: Agarose gel resolution of PCR product for T45. *Lanes:* ***1&8:*** *100 bp & 1 kb MWMs (Promega);* ***2-4:*** *1898+/Oligo d(T)$_{15}$ PCR;* ***5-7:*** *1732+/1966- PCR;* ***2,4,5,7:*** *negative (water) controls,* ***3:*** *T45.3 (1650 bp);* ***6:*** *T45.3.*Amplification with the 1898(+)/Oligo $(dT)_{15}$ primers produced a 1650 bp product (lane 3).

4.4 Discussion and Conclusion

4.4.1 Differential Display (DD)

Kim *et al*., (2001a), used DD-PCR cloning (Huang and Hsu, 1995; Scuric *et al*., 1998) to study the expression profile of cellular genes in HCC and Kuo *et al*., (1998), used the same method in cell lines. To our knowledge, this is the first time that DD has been used to attempt to characterise the expression of HBV integrants in HCC tissue. A method was devised to study cloned HBV genes and their expression in transfected cell lines (Dickens, 2002). The rationale is similar to that applied for DD. Fragments of interest are amplified from reverse transcribed RNA with the use of an Oligo d(T) primer and HBV gene-specific primer. The fragments would be compared between samples and between samples and 'normal' controls (comprising either normal liver or HBV negative liver cell lines). In standard DD only the fragments that differ between samples are subjected to further analysis, however in this case it was decided to amplify very concentrated fragments as well, as very faint amplicons could be products of non-specific amplification, while darker bands (more concentrated amplicons) should represent true integrants, the sizes of which could not be predicted as there would not be a standard length for the genomic sequence adjacent to the integrated HBV DNA. Initially two rounds of PCR were attempted. One PCR prior to fragment isolation and one PCR post-fragment isolation. A round of hemi-nested PCR (with a second HBV-specific primer - 1765) was also performed to help confirm the specificity of the first round PCR product.

4.4.2 Amplification with hemi-nested PCR

The hemi-nested PCR using 1765+/Oligo $(dT)_{15}$ primers amplified four of the seven RNA extracts (T14, T31, T110, T45), two of which had multiple products (T31, T45). Poorly amplified products from this PCR were either the result of non-specific amplification or else they lacked the 1765+ sequence. There were certain products that amplified well, which were later shown by sequencing to have been non-specific. Non-specific priming could be therefore a shortcoming of this technique.

4.4.3 Reverse trancriptase negative PCR control

The expressed T45.3 sequence was generated from RNA. DNA was digested with DNaseI, and

there was a lack of amplification in reactions lacking the reverse transcriptase enzyme. These controls were performed with primers for the beta-catenin gene in the concurrent study from which the aliquot of RNA was 'borrowed' and were not repeated with HBV specific primer because of the limited supply of tissue (Elmileik *et al*., 2005). That DNA was effectively eliminated in the sample could be intimated by the fact that these samples, did not amplify the housekeeping beta-catenin gene in these samples. Furthermore, it was proven in all subsequent PCR reactions with other samples prepared in the same way that no product was generated when amplified with HBV specific primer for those reactions initially lacking reverse transcriptase.

4.4.4 The sequence of T45.3

This DNA was derived from the HCC of a Mozambican adult male. This specimen had not been one of the tissues initially analysed by SH because of the small amount of tissue available for this sample. T45.3 comprised 74 bp of flanking DNA and 322 bp of positive strand HBV DNA (Figure 4.7, Appendices E and G), which integrated in the 5'-3' orientation of the HBV genome. It could not be established whether the integrant had integrated in-frame. The flanking DNA appeared rearranged, consisting of 55 bp of *Homo sapiens* immunoglobulin heavy chain variable region (*VH1*) (92%, 4e-13) (AJ007320.1 last accessed 28 September 2009, Appendix G page G6) in the reverse complementary orientation, or possibly the cloning vector pPCR-Script Amp SK(+) (92%, 4e-13) (U46017.1 last accessed 28 September 2009, Appendix G page G6) (Appendix G pages G2 to G22); with a further 19 bp mapping to *Homo sapiens* chromosome 16 clones CTD-3225F13 and 66H6 (100%, probability 2.9) (AC126538.1 & AC007014.1 - last accessed 28 September 2009, Appendix G page G26 & G30) and two HBV clones (LT6 :DQ464177.1; DQ464176.1; and LT5 DQ336691.1 - last accessed 28 September 2009, Appendix G page G26 & G29) (100%, probability 0.19) (Appendix G pages G23 to G32).

Over time major structural rearrangements and deletions, possibly mediated by DR 1 and 2, may occur in integrated HBV DNA (Shih *et al*., 1987; Schlüter *et al*., 1994). Chen *et al*., (2000), report expressed integrated *X* gene with various deletions and multiple nucleotide mutations. Two nucleotide differences were evident in the HBV sequence of T45.3: an A at position 2054 instead of a C and a T missing at position 2122. These mutations were only

detected in the forward sequence and could not be confirmed by sequencing in the reverse orientation because of the lack of PCR product.

4.4.4.1 The T45.3 'heavy chain variable region' sequence

The initial analysis of the T45.3 sequence mapped 33 bp of the flanking sequence to the *VH1* gene (AJ007320.1) (100%, 5e-09 – last accessed 28 September 2009, Appendix G page G11 & G13) (Identities = 33/33) (100%), which had previously been characterised from the mRNA for the *VH1* region clone KC4, by Souto-Carneiro and Krenn (unpublished). Their sequence was complementary and in the reverse orientation to the sequence of T45.3. Subsequent to this, and much later, when the T45.3 sequence was submitted to Genbank it was revealed that a region of the AJ007320 *VH1* sequence appeared to derive from the plasmid pPCR-Script Amp SK(+) (U46017.1) (100%, 5e-09 – last accessed 28 September 2009, Appendix G page G11 & G14) (Identities = 33/33) (100%) used by Souto-Carneiro and Krenn to clone the particular *VH1* sequence AJ00320.1 (last accessed 28 September 2009, Appendix G page G16 & G17). Subsequent analysis of the flanking DNA showed that this was a second possibility. Part of the T45.3 flanking sequence could derive from the plasmid pPCR-Script Amp SK(+), as this plasmid was in use in the laboratory where T45.3 was isolated. Therefore the possibility that this portion of the T45.3 sequence could be pPCR-Script Amp SK(+) sequence cannot be excluded although it cannot be concusively proven (section 4.4.3.2 and Appendix G pages G2 to G22).

4.4.4.2 T45.3 as a legitimate integrant

The T45.3 sequence is arguably an integrant rather than a PCR artifact for five reasons.

1. Part of the sequence integrated was that of the *X* gene, the most frequently integrated region of the viral genome (Wang *et al*., 1991; Diamantis *et al*., 1992; Wang *et al*., 1994b).
2. The integrated region of the HBV genome was at nucleotide 1824 (nucleotide position of HBV genome V01460 where *Eco*RI cleavage site is position 1) close to the DR1 11-bp sequence (nucleotides 1824-1834) a preferred site for HBV integration and the junction lay within the cohesive-end overlap, a hot-spot for integration (Quade *et al*., 1992).

3. The integrated viral DNA was adjacent to non-HBV DNA at HBV nucleotide 1820, possibly indicating that it was not generated from a HBV mRNA, but rather from a chimera mRNA produced from genomic DNA and adjacent HBV integrated DNA. Rearrangement and mutation of integrated HBV genomes, as well as their flanking sequences, is well documented in HCC (Zhou *et al*., 1988; Meyer *et al*., 1992; Hino and Kajino, 1994; Pineau *et al*., 1996; Poussin *et al*., 1999; Wang *et al*., 2004) and the evidence suggests that such mutations are post-integration (secondary) events (Rogler and Summers, 1982) although in some cases there has been the indication that mutant/rearranged genomes may be preferentially integrated (Rogler and Summers, 1982; Kew *et al*., 1993). Specific cases of rearrangement of *X* gene integrants, with 2 sections (or more) of HBV DNA separated by genomic DNA were documented in 1994 by Schlüter *et al*., and in 2004 and 2005 by Murakami *et al*., who also reported eight and seven integrants incorporating portions of the *X* and part of the pre-core/core ORF. A further 9 integrants comprising *X* gene DNA and genomic DNA in or near a cellular gene were documented in 2003 by Paterlini-Bréchot *et al*., by Murakami *et al*., (2005), and 10 integrants were documented in 2002, by Tsuei *et al.*, most of which contained the DR1 sequence.
4. The T45.3 mRNA was unlikely to have been generated from free viral DNA, as the amplicon was of the wrong size for standard HBV mRNAs or splice variants, given the position of the 1732 primer. Moreover, HBV replication is reported to be absent in late stage HCCs (Nazarewicz *et al*., 1977; Simon and Carr, 1995) and by extension, transcription and translation should also be lacking.
5. The 1732+ and 1765+ PCR reactions could not have been non-specific and would have amplified off HBV sequence further upstream (section 4.4.3.2). This is characteristic of highly rearranged integrants (section 4.4.3.4).
6. The possibility of the integrant being legitimate cannot be excluded, as portions or all of the flanking sequence may be a part of the human genome that may not have been banked with GenBank yet. Furthermore, the T45.3 HBV sequence was found to cluster with the HBV genotype E isolates, found primarily in the populations of sub-Saharan Africa, and the T45.3 subject was a Mozambican male who had migrated to South Africa. There was no genotype E clone in the laboratory at the time.

Additional amplifications of T45 RNA were performed (Figure 4.9) with primers 1898/Oligo d(T) and 1732/1966, to attempt further characterisation of the integrant. A faint amplicon was generated with the 1898/Oligo d(T) primers but it was insufficiently concentrated for sequencing. Fresh cDNA gave the best amplification results and at this point this cDNA had been used extensively over several weeks. The 1732/1966 PCR failed to amplify, despite the original sequencing data having indicated the presence of the sequence used for the 1966 primer, but this was not unexpected as the reaction was performed with only half the amount of cDNA template (the remainder had been depleted). Unfortunately the initial RNA extract yielded only enough cDNA for these amplifications. This sample was also used in another study and the tissue became depleted.

4.4.4.3 T45.3 Flanking sequence as pPCR-Script Amp SK(+)

As stated in section 4.4.3.1 the possibility exists that part of the sequence (33 bp) adjacent to the HBV DNA was generated from the plasmid pPCR-Script Amp SK(+) (U46017.1) (100%, 5e-09 – last accessed 28 September 2009, Appendix G page G11 & G14) (Identities = 33/33) (100%). This vector was originally used by Souto-Carneiro and Krenn (unpublished) when cloning mRNAs of the *VH1* gene and as it was in use in the laboratory where this PhD study was performed. The first 55 bp (Figure 4.7) of the flanking sequence of T45.3 map to base pairs 674 to 728 of the pPCR-Script Amp SK(+) plasmid (U46017.1) with Identities 53/57 (92%) with a probablility of 4e-13 (Appendix G page G6), exactly the same statistics as for the mapping of the sequence to the *VH1* gene base pairs 283 to 229. Furthermore, base pairs 229 to 108 of the *VH1* gene sequence banked by Souto-Caneiro and Krenn mapped to the plasmid sequence base pairs 728 to 699 (Identities 79/80) (98%) with a probablility of 4e-31. On closer inspection it was noted that base pairs 23-55 of T45.3 map to both the *VH1* and pPCR-Script Amp SK(+) with 100% homology and probability 5e-09. Base pairs 1-22 do not map to pPCR-Script Amp SK(+) with 100% homology but do map to a number of sequences with extremely high e values (ie. probability of 17 and above). They cannot therefore be reliably mapped (Appendix G pages G2 to G22).

In supposing that the sequence concerned is pPCR-Script Amp SK(+), several issues must be considered:

- the 1732 primer would have had to have primed the plasmid DNA non-specifically in two separate PCRs, or the HBV sequence had to have recombined with the plasmid

DNA so that the complementary primer sequences (1732 and 1765) were located upstream to the PCR product; Non-specific priming of 1732 and 1765 to the plasmid did not occur. The 1732+ primer maps to this plasmid only with a 7 base pair mismatch for an 18 bp primer (position 386-404 on the plasmid). The case would be the same for the 1765+ primer which would only function as an internal primer if it primed off the plasmid sequence at 1305-1324 or 1752-1772, but again with 8 bp mismatches for a 21 bp primer, both despite annealing temperatures of 67 and 61 °C respectively;

- the Oligo d(T) primer would have had to prime three separate PCR reactions non-specifically off a stretch of adenosine nucleotides rather than a poly-A tail;
- all the blank water controls for all three PCRs performed showed a lack of contamination in the amplifications;
- reverse transcriptase negative PCR reactions had shown a lack of contaminating DNA in the sample, albeit with the beta-catenin primers;
- the plasmid sequence is not contiguous with the HBV sequence but is separated from it by 19 bp of DNA that is neither of plasmid nor HBV origin;
- despite the presence of pPCR-Script Amp SK(+) in the laboratory in which the PhD was performed, there was no cloned HBV genotype E in the laboratory at that time – that work began a year later. This makes it unlikely that the flanking DNA was from a HBV DNA cloned into pPCR-Script Amp SK(+), but the possibility cannot be entirely eliminated.

4.4.4.4 T45.3 chromosome 16/HBV sequence

The 17/19 bp (bp 56-74 Figure 4.7) were found to map to HBV clones LT6 (DQ464177.1; DQ464176.1) and LT7 (DQ336691.1) with probability 0.19, (Identities 17/17) (100%) (Appendix G page G26), as well as *Homo sapiens* chromosome 16 clones CTD-3225F13 (AC126538.1) and 66H6 (AC007014.1), with probability 2.9, (Identities 17/17) (100%) (Appendix G page G26). The LT5 and LT 6 clones comprised PCR amplified full length genomes (2966 and 3149 bp) of the same HBV sequence respectively, beginning at bp 1821. The last 217 bp of clone LT6 are extremely unusual (Appendix G page G31 & G32). Base pairs 3091 to 3108 (17 bp) map exactly to the 17 bp designated as bp 56-74 (highlighted in light blue – Figure 4.7) in this thesis. This 17 bp sequence does not however, map to any other HBV sequence on the GenBank Database, or anywhere else on the HBV genome. Base pairs

3110 to 3132 (22 bp) map to bp 1-22 of the HBV sequence of the same clone, and bp 3133 to 3149 (16 bp) map to bp 2971 to 2987 and 3056 to 3071 of the same clone with 1 bp mismatch. Clone LT5 has 2966 bp identical to LT6 but is missing bp 2967-3090 present in LT6. It does however, contain the same 58 bp of sequence on the end as does LT6 – namely the 17 bp designated as bp 56-74 (highlighted in light blue – Figure 4.7) in this thesis and the same 22 bp, followed by 16 bp, although these sequences are absent elsewhere on this clone. It is not evident from this work (Pollicino *et al*., 2007) whether these two clones derived from the same patient or two different patients. The probability is that this represents a PCR artifact. This highlights the difficulties and potential pitfalls of repeated PCR amplification in the molecular biology laboratory (AC126538.1 and AC007014.1 – last accessed 28 September 2009, Appendix G pages G23 to G32).

4.4.4.5 Expression of T45.3

The T45.3 sequence is thought to be expressed because the PCR product was generated from DNA free RNA, digested with DNaseI (section 4.2.1.). However, it presupposes that the established T45.3 sequence originally contained the upstream portion of HBV sequence containing the 1732+ primer sequence. Subsequent rearrangement of the integrant resulted in a segment of the *VH1* gene being inserted in the complementary reverse orientation inbetween the HBV sequence containing the 1732+ primer sequence and the established T45.3 sequence obtained in the sequencing reaction. The presence of these primers upstream is highly probable as two independent 1732+ PCRs yielded strong bands of the same size of the T45.3 sequence. Furthermore the 1765+ nested primer also produced a correctly sized product. This would not have occurred by mis-priming. Furthermore, specific cases of rearrangement of *X* gene integrants, with 2 sections (or more) of HBV DNA separated by genomic DNA were documented in 1994 by Schlüter *et al*., and in 2004 and 2005 by Murakami *et al*., who also reported eight and seven integrants incorporating portions of the *X* and part of the pre-core/core ORF. A further 9 integrants comprising *X* gene DNA and genomic DNA in or near a cellular gene were documented in 2003 by Paterlini-Bréchot *et al*., by Murakami *et al*., (2005), and 10 integrants were documented in 2002, by Tsuei *et al.,* most of which contained the DR1 sequence.

Transcribed chimeric genomic/HBV integrants are well documented in the literature (Wollersheim *et al*., 1988; Dejean and de Thé, 1990; Kobayashi *et al*., 1997; Chami *et al*., 2000; Horikawa and Barrett, 2001; Wang *et al*., 2004; Saigo *et al*., 2008). Most such integrants contain truncated or mutated *X* genes that retain their function, some of which may be implicated in HCC progression (Chami *et al*., 2000, 2001) or show direct carcinogenic activity.

Expressed integrants may contribute to the process of carcinogenesis in a number of ways, such as inducing cell proliferation, pro-transforming effects, chimeric proteins that induce apoptosis, and expressed *X* gene integrants have transactivating potential (Wollersheim *et al*., 1988; Tsuei *et al*., 1994; Chami *et al*., 2001; Paterlini-Bréchot *et al*., 2003).

Three other studies document expressed HBV integrants by RT-PCR (Paterlini *et al*., 1995; Poussin *et al*., 1999; Chen *et al*., 2000). The initial work (Paterlini *et al*., 1995) was performed on HBsAg –ve patients with HCC. HBV *S*, *X* and *C* transcripts were analysed by RT-PCR. There were a significant number of *X* products detected but no *S* or *C* sequences. The second study by Poussin *et al*., (1999), duplicated the methods of Paterlini *et al*., (1995), but also analysed *X* gene sequences by SH and HBx expression. However, in both studies the use of 2 HBV-specific primers for each PCR resulted in PCR products containing only HBV sequence in 12 patients. The assumption that these mRNAs derived from integrated HBV rested on the fact that the patients were HBsAg –ve, and that 5/9 *X* gene sequences were interrupted or highly mutated. Theoretically, in HBsAg –ve patients, replicating virus should be absent, however, virions may still be present but may be missed because of a lack of sensitivity in the laboratory detection techniques employed, the formation of HBsAg-antiHBs complexes (rendering the HBsAg undetectable), or because the virions present are quiescent or *S* gene mutants (Shafritz *et al*., 1981). A disrupted *X* gene sequence has been documented in free HBV, from which 6 *X*-*C*-specific transcripts of differing sizes were detected (Rho *et al*., 1989; Kim *et al*., 1992). Therefore, disrupted *X* sequences are not restricted to integrants alone, although disrupted integrated *X* sequences have commonly been documented, and probably occur more frequently than disrupted *X* sequences in free virus. The likelihood that the expressed *X* sequences amplified derived from integrated HBV was high but there was no conclusive proof. Furthermore, 4/9 *X* gene sequences showed no evidence of mutation and could have derived from free virus genomes. The third study (Chen *et al*., 2000) attempted to amplify expressed *X* gene sequences from serum and liver tissue in 5HBsAg +ve patients with

the use of various *X*-specific primers. The rationale here was that sequences found to be expressed in the liver but not the serum could be presumed integrated, especially if disrupted. As argued before, a disrupted sequence cannot automatically be assumed to be integrated. Furthermore it is known that HBsAg +ve tumours contain free replicative HBV genomes (Michalak *et al*., 1994; Yotsuyanagi *et al*., 1998; Marusawa *et al*., 2000). The assumption that sequences found to be expressed in the liver but not the serum could be presumed integrated is worrisome, as several investigators have reported that PCR sensitivity differs according to sample type, ie. detection of HBV *S* region is more sensitive in serum samples, whereas *X* region detection is more sensitive in liver tissue samples (Jilg *et al*., 1995; Villa *et al*., 1995; Takeuchi *et al*.,1997; Koike *et al*., 1998). The method performed in the present study would be advantageous over the above methods, in that potentially, expressed integrants will amplify along with at least one of their flanks, and once sequenced, can be back-tracked to the host genome and mapped.

4.4.4.6 The partial HBV sequence in T45.3

Using phylogenetic analyses, the 322 bp integrated T45.3 HBV sequence was found to cluster with the HBV genotype E isolates (Figure 4.8). When mapped to Genbank the sequence mapped to HBV E genotype isolates 0262-09 and 0121-20 (DQ060825.1 and DQ060824.1) (Kramvis *et al*., 2005a) (4e-162; Identities 320/322; 99%) and HBV E genotype isolates M and K precore/core protein gene (DQ297468.1 and DQ297467.1) (Mathet *et al*., 2005, Unpublished) (4e-162; Identities 320/323; 99%) (Appendix G pages G33 to G43). The region mapped encompasses 1820-2142 (end of *X* gene beginning of *C* gene), which is a conserved region of the HBV genome among the mammalian hepadnaviruses (Lauder *et al*., 1993; Mizokami *et al*., 1997). This is the possible first characterisation of a genotype E HBV integrant and its expression in HCC.

Of all the HBV genotypes, E is the most prevalent worldwide, found primarily in the populations of sub-Saharan Africa – an estimated 20% of HBV chronic carriers or approximately 60 million people (Kew, 1992; Mulders *et al*., 2004; Kramvis and Kew, 2007), and may be the most important genotype internationally. Despite its wide geographic distribution and high prevalence, its genetic diversity is low, indicative of a short evolutionary history in humans (Odemuyiwa *et al*., 2001; Kramvis and Kew, 2007), and its origin remains

obscure. Its sequence has been extensively studied in recent years in order to attempt to understand the spread of its strains, which along with its evolution, is poorly understood (Hübschen *et al*., 2008). Those strains of genotype E detected in regions outside of Africa, such as Europe and America originate from carriers of African origin (Kramvis *et al*., 2005a), except for one, isolated from a Spanish native, which intimates that the genotype can spread in the European population (Echevarria *et al*., 2005).

The T45.3 sample was from a Mozambican male who had migrated to South Africa. Although published data indicated that genotype E was relatively rare in South Africa, comprising less than 3 % of the total population of HBV isolates (Kew *et al*., 2005), recent work indicates that the E genotype may be significantly more prevalent in South Africa than previously thought (Dr A. Kramvis PC, unpublished data). In The Gambia, 8/10 patients with HCC carried genotype E (Mendy *et al*., 2008). This is consistent with the fact that the most prevalent genotype in chronic carriers in this country appears to be genotype E *ayw*4 (93.9%) (Mendy *et al*., 2008). It is accepted that chronic carrier status correlates with HCC prevalence and a 20-30 year latency period occurs before the appearance of the cancer (Matsubara and Tokino, 1990).

The distribution of genotype E includes a crescent extending from Senegal to Namibia (Maïga *et al*., 2005; Kramvis and Kew 2007) and to the Central African Republic in the East (Mulders *et al*., 2004), with those sequences from the southern end of the crescent (South-west Africa and the DRC) clustering together (Kramvis and Kew, 2007) and separately from those from the Northern part of the crescent. The higher genetic diversity of the isolates from the north suggests that they have been present in the population for longer and subsequently spread south (Hübschen *et al*., 2008).

Genotype E has unique characteristics differentiating it from the other genotypes found in Africa, namely genotype A (subgenotype A1) in southern and eastern Africa and genotype D in northern African countries around the Mediterranean Sea (Kramvis *et al*., 2005a).

Genotype E strains have a unique genome length of 3212 nucleotides and despite their low genetic variability (S gene 0.73%, entire genome 1.71%) (Hübschen *et al*., 2008) there are a number of new mutations of the preS/S gene. There is a deletion of one amino acid at the amino end of the pre-S1 region. This differentiates genotype E and genotype G (118 amino

acids) from genotypes A,B,C,F,H (with 119 amino acids) and genotype D (108 amino acids) (Kramvis *et al*., 2005b). Genotype E strains have the signature motif $Leu^{3}SerTrpThrValProLeuGluTrp^{11}$ and signature amino acids Thr^{18}, Arg^{38}, His^{44}, Thr^{52}, Met^{83}, Lys^{85} and Thr^{108}. Most fully sequenced genotype E strains to date, contain Met^{83}, the consequence of which is a new translational start codon downstream from the pre-S1 start codon (Kramvis *et al*., 2005a), resulting from an A-G mutation at position 3095 within the D region of the S2 promoter (Moolla *et al*., 2002). This may result in the production of an elongated MHBs protein (317 amino acids versus 281) whose translation would be regulated by the S1 promoter, and might interfere with the translation of the downstream MHBs. All but two of the genotype E isolates, which have Ile^{15}, have His^{15} at that position. The pre-S2 region appears to have no amino acids that are specific to genotype E (Kramvis *et al*., 2005a).

The 'a' determinant of the HBV is the most divergent region of the E genotype in relation to genotype A (from which monoclonal antibodies are prepared), with a difference of 5.7% at the nucleotide level and 8.5% at the amino acid level (equivalent to 8 amino acids) (Olinger *et al*., 2007). All genotype E isolates to date, have serological subtype, *ayw*4 characterised by Arg^{122}, Lys^{160} and Leu^{127} residues of the S gene. Genotype E strains also have Asn^{146} instead of Thr in the core region; and 20 nucleotides upstream from the core gene stop codon resides the sequence CAGCTTCC, which translates into $Pro^{178}Ala^{179}$ instead of $Arg^{178}Glu^{179}$, a motif shared with genotype F, G and H and non-human primate hepadnaviruses (Takahashi *et al*., 2000; Kramvis *et al*., 2005a). There is a Ser to Gly substitution at nucleotide 22 in the X region of genotypes E and G. The amino acids Ile^{11} and Ile^{118} of the terminal protein of the polymerase are genotype E-specific and some isolates have also been found to carry the specific amino acid Asn^{108}(Arauz-Ruiz *et al*., 2002). There appear to be 8 amino acids unique to genotype E in the spacer region, namely Met^{14}, Glu^{16}, His^{21}, Arg^{52}, Asp^{55}, Lys^{88}, Asn^{110} and His^{111}, while the reverse transcriptase contains only one genotype-specific amino acid substitution, Met^{164} (Kramvis *et al*., 2005a).

There is a paucity of data on genotype E and its role in the clinical manifestation of HBV infection. The unique characteristics of genotype E and its geographical distribution, which is restricted to a large proportion of the African continent, where the virus is hyperendemic, makes it vital that this genotype is studied comprehensively, especially its hepatocarcinogenic potential.

5.0 CONCLUSION

At the time when this study began there was limited data concerning the details of HBV integration in HCC in the southern African populations. The literature documented integration of the virus in at least 80 % of tumours, world-wide (Shafritz *et al*., 1981; Bréchot *et al*., 2000; Murakami *et al*., 2005). HBV integration was thought to contribute to HCC initiation and progression in a number of ways, from direct mutation of the DNA of the host, resulting in genomic instability, to de-regulation of cellular pathways and cellular viability and proliferation. These studies however, had been primarily concerned with HCCs from Europe, China and Japan. Whether the integration events in these populations were the same, similar or distinct to those operating in the southern African populations was unknown. The objective of this study was thus to characterise the pattern of HBV integration in T/NT liver from southern African blacks with hepatocellular carcinoma. We aimed to (1) determine the proportion of tumours with single/multiple integrants; (2) map the sites of integration relative to functional cellular genes. These results could allow for a comparison to be made against previously published data and thereby contribute to an understanding of the molecular aetiology of HCC, at least in the context of the southern African populations. It was therefore necessary to investigate a reasonable number of tumours (15) with corresponding non-tumorous tissue where possible, and a reasonable number of integrants (~15). To this end library construction, a laborious and time-consuming technique requiring large amounts of tissue, and proven to be most effective in the study of HBV integration events, was impractical. At the time, a number of methods involving PCR had recently been published, which promised the speedy processing of several tumours from small amounts of material.

The tissue samples used in this study were from one of the few available HCC tissue collections in Africa. They were late stage tumours, which did not pose a problem as we did not hope to attempt to draw any conclusions requiring information dependent on the stage of HCC development. The paucity of records (samples were obtained at autopsy) also impeded any effort to correlate mutations with clinical or other features of the individuals. The amount of tissue available was limited for certain samples and sometimes of poor quality because of recurrent freezing and thawing. Many had been depleted of either the tumour or non-tumorous component, when ideally both should be used. Despite these difficulties, it was a worthwhile

and necessary study and a concerted effort produced sufficient amounts of good quality DNA and RNA for analysis.

This SH technique permitted the determination of the proportion of tumours with single/multiple integrants. Overall, the results for samples correlated with those previously reported in the literature. As expected for African HCCs, 80 % of the southern African samples analysed here had integration and all of them had multiple integrants. One HBsAg-ve sample had integrants, in both the tumour and the non-tumorous tissue. This had been previously documented in black Africans (Shafritz *et al*., 1981; Kew 2002b). In at least one sample, SH patterns indicated that the HCC contained more than one hepatocyte clone, each with its own pattern of integration.

The PLC/PRF/5 cell line was found to contain 6 and possibly 9 integrants, in accordance with previous reports (Zerial *et al.*, 1986). It was interesting to note that the sizes of integrants differed, not only between this study and published work, but between a number of published studies as well. It was concluded that this was as a result of integrants having suffered amplifications, deletions and translocations. This is corroborated in the literature (Ziemer *et al*., 1985; Zerial *et al*., 1986; Kuo *et al*., 1998).

A new method based on the DD technique was developed in this study for the characterisation of expressed HBV integrants from mRNA. The method is an improvement in relation to previously proposed methods in that potentially, expressed integrants will amplify along with at least one of their chromosomal flanks, and once sequenced, can be back-tracked to the host genome and mapped. Some degree of non-specific priming may be observed and could be a shortcoming of this technique.

A possible integrated, transcribed chimeric genomic/HBV sequence of genotype E was found.

The flanking genomic DNA appeared rearranged, comprising 22 bp of sequence that could not be reliably mapped, 33 bp of *Homo sapiens VH1* region (AJ007320) and 19 bp of adjacent sequence . Subsequent to this, and much later, when the sequence was submitted to Genbank it was revealed that the region of the AJ007320 *VH1* sequence to which it originally mapped appeared to derive from the plasmid pPCR-Script Amp SK(+). A number of considerations

make it unlikely that the flanking DNA was from a HBV DNA cloned into pPCR-Script Amp SK(+), but the possibility cannot be entirely excluded. The possibility of the integrant being legitimate can also not be excluded as portions or all of the flanking sequence may be a part of the human genome that may not have been banked with GenBank yet. The ambivalent results obtained highlight the potential pitfalls of database searches, especially with short stretches (<100 bp) of sequence, which one should be cognisant of. The T45.3 HBV sequence was found to cluster with the HBV genotype E isolates, found primarily in the populations of sub-Saharan Africa, and the T45.3 subject was a Mozambican male who had migrated to South Africa. There was no genotype E clone in the laboratory at the time.

The fact that the HBV sequence is genotype E is not surprising as this genotype occurs in western and central Africa (Bekondi *et al*., 2007; Kramvis and Kew 2007) and is probably the most abundant genotype worldwide. Although published data indicated that genotype E was relatively rare in South Africa, comprising less than 3 % of the total population of HBV isolates (Kew *et al*., 2005), recent work indicates that the E genotype may be significantly more prevalent in South Africa than previously thought (Dr A. Kramvis PC, unpublished data). Furthermore, in a recent study conducted in The Gambia, 8/10 patients with HCC carried genotype E (Mendy *et al*., 2008). This is consistent with the fact that the most prevalent genotype in chronic carriers in this country appears to be genotype E *ayw*4 (93.9%) (Mendy *et al*., 2008). It is accepted that chronic carrier status correlates with HCC prevalence and a 20-30 year latency period occurs before the appearance of the cancer (Matsubara and Tokino, 1990). All specimens of the E genotype analysed to date cluster together and are very well conserved (1.2 % intragroup divergence) (Mulders *et al*., 2004; Kramvis *et al*., 2008).

Structural rearrangements and mutations of integrants and their flanking sequences are well documented in the literature (Shih *et al*., 1987; Zhou *et al*., 1988; Meyer *et al*., 1992; Hino and Kajino, 1994; Pineau *et al*., 1996; Poussin *et al*., 1999; Wang *et al*., 2004; Moradpour and Blum, 2005). The integrated sequence comprised base pairs 1820-2142 of the HBV *X* and *pre-cC/C* genes and included the DR1 repeat. This region (within the cohesive-end overlap) is a hot-spot for HBV integration (Quade *et al*., 1992). The *X* gene is also the most frequently integrated region of the viral genome (Wang *et al*., 1991; Diamantis *et al*., 1992; Wang *et al*., 1994b). The possibility exists that more of the HBV genome could be present upstream to the genomic DNA flank of this integrant. Specific cases of rearrangement of *X* gene integrants,

with 2 sections (or more) of HBV DNA separated by genomic DNA have been reported (Schlüter *et al.*, 1994; Tsuei *et al.*, 2002; Paterlini-Bréchot *et al*., 2003; Murakami *et al*., 2004). Transcribed chimeric genomic/HBV integrants are well documented in the literature (Wollersheim *et al*., 1988; Dejean and de Thé, 1990; Kobayashi *et al*., 1997; Chami *et al*., 2000; Horikawa and Barrett, 2001; Wang *et al*., 2004). Expressed integrants may contribute to the process of carcinogenesis in a number of ways, such as inducing cell proliferation, pro-transforming effects, chimeric proteins that induce apoptosis, and expressed *X* gene integrants with transactivating potential (Wollersheim *et al*., 1988; Tsuei *et al*., 1994; Chami *et al*., 2001; Paterlini-Bréchot *et al*., 2003; Anzola, 2004).

No standard method for the characterisation of HBV integrants (expressed or otherwise) existed or exists. Previously published PCR methods were employed in this study, and were shown to have shortcomings, and not to be easily reproducible. These methods have been employed to analyse non-African HCCs, mainly containing single integrants. The presence of multiple integrants per sample makes characterisation of such integrants more complicated.

In short: SH results corresponded with published literature. Results in the PLC/PRF/5 cell line vary between studies in relation to experimental error and human interpretation.

In this study a new technique for the characterisation of expressed integrants directly from mRNA, which is an improvement in relation to the only other previously proposed method. A possible integrated transcribed chimeric genomic/HBV sequence of genotype E was found. This genotype has never previously been reported as integrated. This highly rearranged sequence comprises a portion of the HBV *X*, DR1 and *preC/C* genes, the hot-spot for HBV integration.

Further research, into the exact manner, in which HBV integration contributes to the aetiology of HCC remains essential. Currently, other than the two partially characterised integrants, one from the present PhD and the other published in Kimbi *et al*., (2005), characterisation of the integrants in the PLC/PRF/5 cell line comprises the only other information available on the sequences and consequences of integration events in southern African blacks with HCC.

APPENDIX A

PROTOCOLS APPENDIX

A1 Tissue Culture (section 2.2.2)

A1.1 Culturing of frozen stocks

Frozen stocks were initially cultured in 25 cm^2 tissue culture flasks (Nunclon). Appropriate medium (at 37 °C) was added to a 1 ml cryogenic vial containing 1 000 000 cells, thawed at 37 °C, one drop at a time, until a total volume of 2 ml was obtained. Cells were resuspended by gentle pipetting and transferred to a 10 ml polypropylene tube (Nunclon), where the medium was slowly topped up, 1 ml at a time until a total of 10 ml was obtained. The cells were centrifuged at 2 000 x g for 5 minutes, and the supernatant removed. Cells were resuspended by gentle pipetting, in 10 ml of fresh medium, placed in the 50 cm^2 culture flask with vented lid, and allowed to incubate at 37 °C in a 5 % CO_2 incubator (Forma Scientific) with moisture (ddH_2O tray). The cells attached to the flask surface after 24 hours. The depleted medium was removed, the cells were rinsed with 1 ml 10 % PBS at 37 °C to remove cell debris, and any remaining depleted medium; and 5 ml fresh medium was added, every second day. Once 80 % confluency had been achieved the cells were transferred to an 80 cm^2 flask, and sometimes thereafter to a 175 cm^2 flask. Depleted medium was replaced every second day with 25 ml (80 cm^2 flask), and 75 ml (175 cm^2 flask) fresh medium, after rinsing with 5 ml and 10 ml 10 % PBS at 37 °C, for the 80 cm^2, and 175 cm^2 flasks respectively. All reagents were warmed to 37 °C in a waterbath, before addition to the cells.

PLC/PRF/5 cell line was cultured in Eagles Minimum Essential Medium (EMEM) (GIBCO) with Earles buffered salt solution (EMEM pH 7.5), supplemented with: (1) 10 % foetal bovine serum (FBS) inactivated at 56 °C for 30 minutes, to remove complement; (2) 50 units/ml penicillin/streptomycin; (3) 1 % non-essential amino acids; (4) 100 mM sodium pyruvate.

HuH7 cell line was cultured in Dulbecco's Modified Eagle's Medium (DMEM) (GIBCO) with four and a half M glucose, and 4mM L-glutamine, supplemented as for the EMEM medium, with the exclusion of the sodium pyruvate.

HepG2.2.15 cell line was cultured in Minimal Essential Medium (MEM) (Cambrex Bio Science Walkerville, MD, USA) supplemented with 10 % FBS and 380 µg/ml of G418.

A1.2 Subculturing

Depleted medium was removed from the flask and the cells rinsed twice with 10 % PBS at 37 °C to remove cell debris, and any remaining depleted medium, which might inhibit the action of the cell detachment agent, trypsin (Cambrex). Cells were detached from the flask surface with 1 ml (50 cm^2 flask), 3 ml (80 cm^2 flask), and 5 ml (175 cm^2 flask) of 1x trypsin at 37 °C, the optimal temperature for activation of trypsin, for 5 minutes.

Cells were transferred to a 15 ml polypropylene tube (Nunclon), centrifuged at 2 000 x g for 5 minutes, and the supernatant removed. The pellet was resuspended in 5 ml warm 10 % PBS with gentle pipetting to remove all traces of trypsin and debris, centrifuged at 2 000 x g for 5 minutes, and resuspended in 5 ml warm medium with gentle pipetting. Resuspended cells were transferred to the appropriate sized flask, topped up with media, and returned to the incubator. The amount of cells seeded per 80 cm^2 flask was $1x10^5$, and 80 % confluency was achieved within 2-4 days.

A manual cell count of 1 x 10^6 for the 80 cm^2 flask, was obtained with the use of a haemocytometer (SIGMA Bio Sciences, Saint Louis, Missouri, USA), trypan blue stain (Cambrex) and a microscope (Nikon Optiphot, Japan) at 100 x magnification.

A1.3 Harvesting

Once cells attained 80 % confluency they were harvested. In order to detach the cells from the flask 0.1 % trypsin was added and incubated for 5 minutes at 37 ºC. Thereafter, 5 ml of medium containing serum, was added to inactivate the trypsin, and the cells were pelleted by centrifugation at 30 000 x g for 5 minutes, the supernatant aspirated, and after two washes in 1 x PBS (Appendix D) and centrifugation, the PBS was aspirated. The PBS washes ensured complete removal of the culture medium which could inhibit the lysing buffer, and diluted the lysate, resulting in decreased DNA/RNA yield. The pellet was immediately used for DNA/RNA extraction or snap frozen in liquid nitrogen (African Oxygen, Germinston, RSA) and stored at -70 °C for up to 6 weeks.

A1.4 Making stocks from mammalian cell lines

Stocks of each cell line were made routinely to ensure backups in case of contamination or accidental loss/death of cells. The cells were counted, and harvested with the use of trypsin as previously described. After centrifugation at 2 000 x g for 5 minutes, the supernatant was removed, and the pellet re-suspended in 10 ml of PBS by gentle pipetting. The suspension was once again centrifuged at 2 000 x g for 5 minutes, supernatant removed, and the PBS wash repeated, followed by a final centrifugation at 2 000 x g for 5 minutes, and removal of supernatant. The pelleted cells were re-suspended in 10 ml of freezing medium (Appendix D) by adding it slowly 1 ml at a time with gentle pipetting to disperse the pellet. The cells were aliquoted into sterile 1.5 ml cryotubes (Nunclon) at 1 ml freezing media/cell mix (1 000 000 cells per cryotube). The cell/freezing medium mix was frozen gradually so as not to lyse the cells, by first placing the full tubes wrapped in cotton wool at 4 °C for 2 hours, then at -20 °C for 4 hours, and lastly at -70 °C for 2 hours before removing the cotton wool. Stocks were stored at -70 °C for up to two years.

A2 Phenol Chloroform (P/C) Method of DNA extraction (section 2.2.3.1)

A 500 mg tissue sample, or 1 x 10^6 cells, was used per extraction. The tissue was finely cut with a sterile blade before homogenisation. Both tissue and cells were homogenised in a sterile mechanical homogeniser (Ultra Turrax®, Polytron® pt 3000, Kinematica AG, Littau, Switzerland) with 1 ml lysing buffer (Appendix D). Thereafter they were transferred to a 50 ml sterile polyethylene tube (Nunclon, Kamstrupvej, Roskilde, Denmark), where the lysing buffer was gradually topped up to 10 ml while the tissue/cells were further disrupted using a sterile glass rod. A further 2 % SDS (Appendix D) was added and the tube placed on a tilting platform for gentle agitation at 37 °C for 20 minutes. Successive washes were performed by addition of 10 ml tris-saturated phenol (pH 7) with vigorous shaking, and centrifugation at 6 000 x *g* for 15 minutes at 4 °C. The aqueous phase was aspirated, while avoiding the interphase as much as possible, and transferred to a sterile 50 ml polyethylene tube. The P/C wash was repeated, and the aqueous phase combined 10 ml of chloroform-isoamylalcohol (24:1) and shaken vigorously. Centrifugation was performed at 6000 x *g* for 10 minutes at 4 °C, the aqueous

phase recovered and DNA precipitated by addition of 2.5 volumes ice-cold ethanol with gentle swirling in order to aggregate the DNA strands. If no DNA was visible, the sample was placed at -20 °C for 16 hours and centrifuged at 6000 x *g* for 30 minutes. DNA was resuspended in 10 ml Tris-EDTA buffer (pH 8) overnight on a gently rotating wheel at 4 °C. The DNA was treated twice with 200 µg/ml Proteinase K, at 60^0C for 2 hours with gentle shaking, in order to remove residual protein; the P/C washes were repeated to remove impurities and digested proteins and the DNA was resuspended to a concentration of 0.5-2 µg/µl in TE buffer (pH 8) overnight at 4 °C. If DNA proved difficult to resuspend, it was placed overnight in a heating block at 60 °C. If it precipated out of solution when spun down on a tabletop microcentrifuge for a few seconds, it was further diluted with Tris-EDTA buffer (50 µl at a time) at 60 °C for 1 hour, and then overnight on a gently rotating wheel at 4 °C. This was repeated until all the DNA was in solution.

A2.1 Quantification of DNA

The extent of hydrolysis and estimated DNA concentration were visually assessed by agarose gel electrophoresis. The quality, purity and concentration of the DNA was analysed with the use of a Philips PU 8700 series ultra violet (UV)/visible spectrophotometer.

Readings were taken using a serially diluted commercially available DNA standard (Qiagen) of known concentration to confirm that the spectrophotometer was reading accurately. DNA concentration was determined by measuring its absorbance at 260 nm and 280 nm and applying the formula:

DNA concentration in µg/ml = 50 µg/ml x A_{260} x dilution factor;

and purity by the ratio of the readings taken at 260 nm and 280 nm, where:

280/260 = X.

If the ratio is greater then 0.75 the DNA is free of significant amounts of protein and is of an acceptable purity (Sambrook *et al.*, 1989, p 9.19).

Quantification was confirmed by fractionating a DNA dilution series (1:500; 1:300; 1:100; 1:50; 1:10; 1:1) alongside four DNA standards of known concentration (5ng, 50ng, 100 ng, 500 ng) on a 1 % agarose gel (5µg EtBr/100ml of gel) (Sigma-Aldrich). The band that was

just visible contained 10 ng of DNA (Sharp *et al.*, 1972 as cited in Sambrook *et al.*, 1989 p6.2).

A3 SH (section 2.2.4)

A3.1 Restriction Enzyme Digestion of genomic DNA for SH (section 2.2.4.1)

A3.1.1 Rationale

A map of each integrant is required, in order to obtain information for primer design, thereby enabling further characterisation of integrants by PCR. Maps may be surmised by knowing the exact location of restriction enzyme (RE) cleavage sites on the DNA of interest (in this case the HBV genome – Figure A1). Digestion of genomic DNA containing integrants, with a particular RE, and its analysis by SH with the use of radioactively labeled HBV probe, allows for the sizing of those integrated HBVfragments generated by the particular RE. The fragments can then be reconstructed into a map by locating their supposed positions on the HBV genome via the position of the RE cleavage site. Digestion with single REs as well as combinations of the same REs allows for more precise mapping. The process of mapping an integrant becomes impossible however, when more than one integrant is present.

A3.1.2 Cleavage sites for REs *Eco*RI, *Bam*HI and *Hind*III:

(Nucleotide positions numbered as per RE cutting site nucleotide position 1 - Galibert *et al.*, 1979).

The *Eco*RI (Boëhringer Mannheim, Mannheim, Germany) restriction endonuclease has one target site (at nucleotide 1) (Galibert *et al.*, 1979; Marion *et al.*, 1980) except in the following isolates, AY233279AC (RSA) at nucleotides 2775-2776, AY090460 at nucleotides 2769-2770 F, and SAX75663 at nucleotides 2770-2771 and the H subtype isolates which do not occur in RSA. The V01460 sequence is cut by *Eco*RI only once at nucleotide 1.

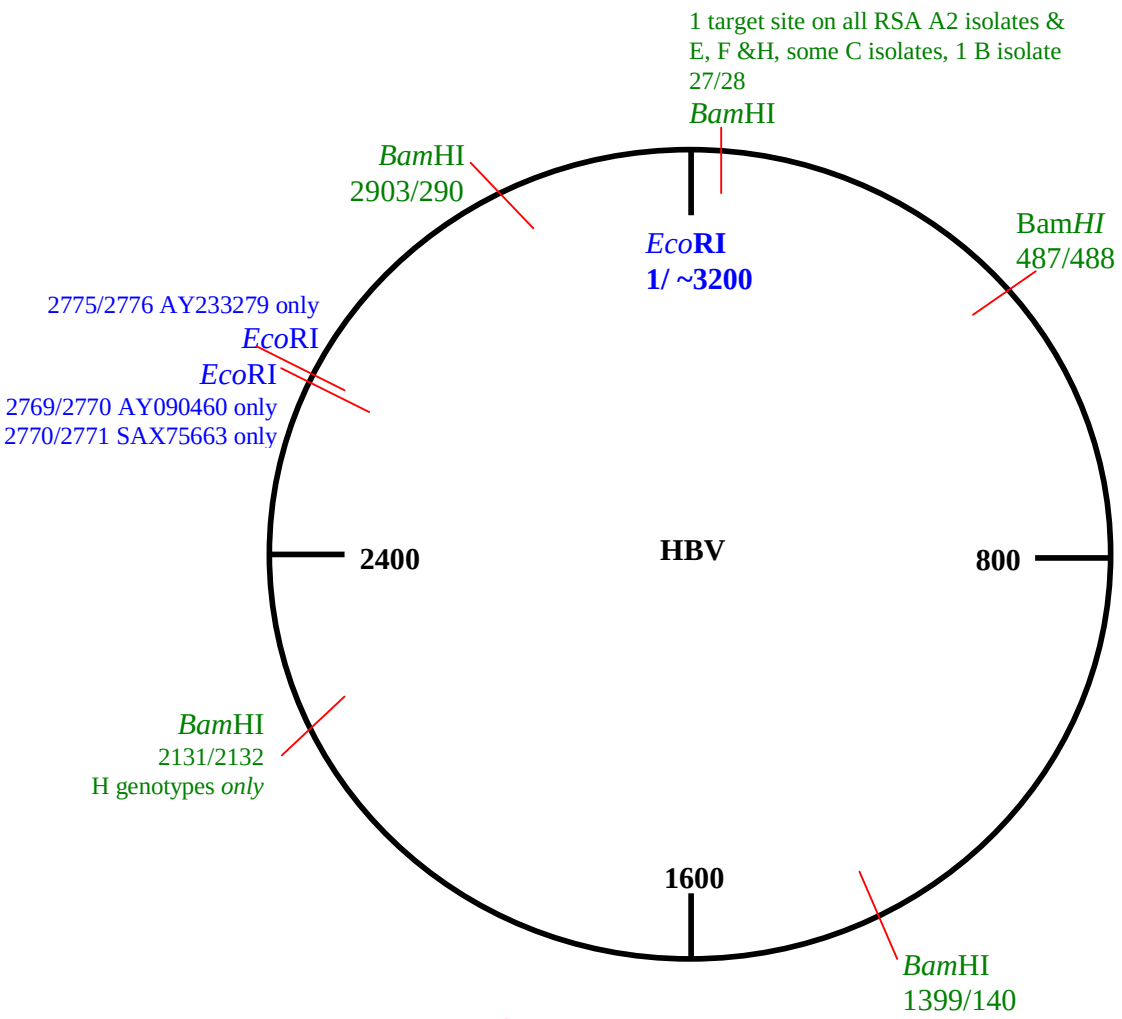


Figure A1: Diagrammatic representation of the HBV genome showing the positions of the *Eco*RI, *Bam*HI and *Hind*III restriction enzyme cleavage sites. In those cases where further cleavage site/s are known for an isolate they are shown at the correct positions and the isolates in which they occur are annotated alongside.

The *Bam*HI (AEC Amersham, Little Chalfont, Buckinhamshire, UK) restriction endonuclease one target site on all the RSA A2 isolates, the E, F, H, some C and one B genotype at nucleotide position 27-28. It produces a further cut in the H genotypes at nucleotide position 2131-2132. The V10460 sequence is cut 3 times by *Bam*HI at nucleotide positions 487-488, 1399-1400, 2903-2904.

The *Hind*III (Boëhringer Mannheim, Mannheim Germany) restriction endonuclease has no target sites in sequence V10460 or in any of the RSA isolates, but should be used with care when investigating: C genotypes from Australia and China (AB048704; AB048705; X75656) isolates from the UK (X80925) and Greece (XX97849); F isolates from Argentina (AF223963) Costa Rica (090459) Al-Salvador (AY090461) and H isolates from central America (AY090454; AY090457) a B isolate from Japan (AB073855) and a D genotype from Italy (X59795) and Greece (X97849) - where it cuts once.

Enzymes were checked for absence of unspecific endonuclease and exonuclease activity in digests lasting 16 hours. RE cutting sites were confirmed using Gendoc and a complilation of 175 sequences representative of worldwide HBV isolates courtesy of Dr Anna Kramvis (MHRU, Wits Univerity Medical School, 7 York Rd Parktown, RSA).

A3.1.3 Materials and Methods for Single and Double Digest Reactions

A3.1.3.1 Single Digests

Initially 10 µg of genomic DNA was restricted as recommended in the literature. However, as the SH results obtained were not optimal, 15 µg, and 20 µg of DNA were also restricted for use in SH. Best results were obtained at a genomic DNA concentration of 20 µg. The DNA aliquots of genomic DNA from liver, Huh7 and PLC/PRF/5 cell lines, 1ng probe (FL HBV genotype D), and a genomic DNA (20 µg)/probe DNA (1 ng) control were each individually combined in a polypropylene tube with nuclease-free water (Promega, Madison, WI, USA) to a total final reaction volume of 50 µl, and 1 x RE buffer compatible to the particular RE and incubated at 4 ^{0}C for 4 hours, on a rotating wheel, to ensure complete and even suspension of the DNA. Variations in the local concentration of DNA lead to uneven digestion as DNA secondary structures are inaccessible to REs. Twenty units of the respective RE was added to the digest, gently mixed by careful pipetting and incubated at 4ºC for 2-3 minutes. The digest

reaction was transferred to an incubator (Labcon) with gentle shaking, set at the optimal temperature for activity of the particular endonuclease (37 ºC) for 2-4 hours, when the activity of the RE began to decrease (half-life 2 hours) at which time a second aliquot of 10 units of the endonuclease was added, and the reaction allowed to proceed for another 12 hours. This "spiking" of the enzyme ensured consistent activity over a longer period. Enzymes were inactivated at 65 ºC on a dry block (Hägar, Pinelands, USA) for 15 minutes.

A3.1.3.2 Double digests

Double digests were performed with the enzyme combinations *Eco*RI/*Bam*HI, *Eco*RI/*Hind*III; *Bam*HI/*Hind*III. Twenty-five µg of DNA was combined in a 1.5 ml polypropylene tube with 1 x SuRE/Cut buffer B (10 mM Tris-HCl; 100 mM NaCl; 5 mM $MgCl_2$; 1 mM 2-mercaptoethanol; pH8 at 37EC) (Boëhringer Mannheim) (in which all four enzymes have 80-100 % activity) and nuclease-free water (to a total final reaction volume of 100 µl). The protocol was as for single digests with the following modifications: digests were spiked 3 times at 2 hourly intervals with 30 + 20 + 40 units of each enzyme instead of 20 + 10 units; after the last addition of enzyme the reaction was allowed to proceed for 16 hours.

A3.1.3.3 Controls:

Undigested DNA

An undigested aliquot of every DNA sample was prepared alongside digestion reactions. All components of the digest reaction were present in these tubes, with the exception of the RE, which was replaced with nuclease-free water. This reaction was subjected to the same conditions as the experimental samples, and the undigested DNA was resolved by electrophoresis alongside its respective digested DNA pair as a control for DNA shearing or degradation.

HuH7 DNA

The DNA from this cell line contains no integrants and was used as a negative control (section 2.2.3) in SH.

PLC/PRF/5 DNA

The DNA from this cell line contains up to 9 integrants (section 2.2.3) and was used as a positive control for DNA transfer and successful hybridisation. The 20 µg *Hind*III

digested PLC/PRF/5 DNA served also as a control for the detection of integrated HBV sequences at the level of 1 copy/cell (Chen *et al*.,1986).

Digestion Controls

Aliquots of the HBV FL genome probe PCR (section 2.4.1.2) were also digested with each of the four REs, to ensure complete DNA digestion. These controls comprised a series of digests containing 20 µg genomic DNA, and a very small amount of probe DNA (eg. $1x10^{-5}$, $1x10^{-6}$, $1x10^{-7}$ µg of plasmid) (Sambrook *et al.*, 1989 p9.32) and were resolved on the agarose gels for SH, with an undigested sample of the same. These positive controls also helped to indicate efficient DNA transfer, and successful hybridization.

A3.2 Fragmentation of DNA for Transfer and SH (section 2.2.4.2)

After digestion DNA was concentrated with 0.2 M NaCl and 2.5 x volumes ice cold 100 % EtOH at -20ºC overnight, recovered by centrifugation at 12 000 x *g* for 30 minutes and re-suspended in 20-40 µl nuclease-free water. Thereafter it was heated to 56ºC for 2-3 minutes to disrupt base-pairing between protruding cohesive termini and combined with 20 µl of 0.4 % gel loading buffer (Appendix D). The glycerol content of the loading buffer was increased to >50 % glycerol with 10 % glycerol (Sigma-Aldrich Life Science, St. Louis, Missouri, USA) to ensure no DNA was lost to the running buffer. The 1 kb and 100 bp DNA ladders (Promega) were de-phosphorylated and labeled with radioisotope at the 5' ends with the use of T4 Polynucleotide Kinase (Promega) (Appendix D) and run on the gels to allow accurate band size determination on the final autoradiograph. They also served as controls for DNA transfer.

DNA was resolved by electrophoresis at <1 V/cm of gel, for a total of 16 hours on 0.6 % agarose gels (D1-LE, Hispanagar, Burgos, Spain), with 3 % EtBr. Samples were photographed on a Bio-Rad Gel Documentation System Universal Hood II (Bio-Rad, Hercules, CA, USA) and blotted onto Hybond N membrane (Amersham International, Little Chalfont, Buckinghamshire, England).

The DNA/gel was pre-treated by depurination (0.25 M HCl for 10-20 minutes), denaturation (1 M or 1.5 M NaCl/0.5 M NaOH for 20-45 minutes), neutralization (3 M NaCl/1 M Trizma base

(pH 7.5) or 0.5 M Tris-base (pH8.0), 0.5 % SDS for 30 minutes), and soaking (20 x SSC or 0.3 x TBE in running buffer for 20-60 minutes). The DNA was transferred to neutral membrane, Hybond N (AEC Amersham), by capillary action, or with the use of a 'Semi-dry electrophoretic blotting system' (Sigma-Aldrich). The exact protocol for pre-treatment of the gel prior to blotting depended on which method was employed to transfer the DNA (sections A3.2.1, A3.2.2). Once transfer was complete, the membrane was rinsed in 2 x SSC (capillary transfer) or 0.3 x TBE (Electro-Blotter transfer), and baked in an oven at 80ºC for 2 hours, to allow cross-linking of the DNA phosphate backbone with the membrane fibres. Successful DNA transfer was confirmed by staining of the agarose gel in 0.02 % EtBr/deionised distilled water for 20 minutes at ambient temperature, and viewing on an ultraviolet transilluminator (UVP). The absence of DNA on the agarose gel indicated efficient transfer.

A3.2.1 DNA transfer by capillary action

The gel was soaked in three stages (with a nuclease-free water rinse between each stage). Stage 1: depurination of the gel with 0.25 M HCl solution (Appendix D) for 10-20 minutes, or until the loading dye turned from blue to a yellow-green colour. Stage 2: denaturation of the DNA in the gel with 1.5 M NaCl/0.5 M NaOH solution (Appendix D) for 45 minutes, until the loading dye reverted from the yellow-green colour, back to blue. Stage 3: neutralization of gel pH with 3 M NaCl/1 M Trizma base (pH 7.5) solution (Appendix D) for 30 minutes. All stages were performed at ambient temperature. Before DNA transfer to the membrane, the gel was saturated in 20 x SSC for 20 minutes. The DNA was transferred to the Hybond N membrane by capillary action for 24-48 hours (Figure A2.2).

A3.2.2 DNA transfer by Semi-dry electrophoretic blotting system

As per manufacturer's instructions: Gels were depurinated by soaking in 0.25 M HCl for 20 minutes, denatured in 0.5 M NaOH, 1.0 M NaCl for 20 minutes, and neutralised twice in 0.5 M Tris-base (pH8.0), 0.5 % SDS for 30 minutes. Before assembly, the gels were soaked in running buffer 0.3 x TBE (Appendix D) for 1 hour. The membrane was cut to size and soaked in deionised distilled water for 10 minutes, then in 0.3 x TBE for 60 minutes. The blotter was assembled, and DNA was transferred to the membrane at room temperature for 4 hours at 50 mA (Figure A2.1).

A3.3 Hybridisation with QuikHyb® Rapid Hybridization Solution (Stratagene, La Jolla, California, USA)

The QuikHyb® hybridization solution (Stratagene) was used for SH using radioactively labeled probe (Appendix A section A3.5). This solution enabled total hybridization time to be reduced to 1-2 hours.

After baking the membrane was briefly rinsed in deionised distilled water, rolled lengthwise in nylon mesh and placed in a conical hybridization tube (Hybaid, Hampshire, UK). Fifteen ml of QuikHyb® solution was used per 20 x 20 cm^2 of blot and pre-hybridisation performed in a hybridisation micro-oven (Hybaid) at 65 ºC for 10-20 minutes. Eight x 10^7 cpm/µg of [α-^{32}P]dCTP labeled probe was combined with 1 mg of sonicated salmon sperm DNA (SIGMA Bio Sciences). The double stranded probe was boiled at 99ºC in a dry block (Hägar) for 2 minutes and the membrane hybridized at 65 ^{0}C for 2 hours. Thereafter membrane washes were performed as follows: 2 x SSC/0.1 % SDS for 15 minutes at ambient temperature twice; 0.1 x SSC / 0.1 % SDS for 30 minutes at 60 ºC x 1 for a high stringency wash. The membrane was sealed in plastic and exposed to autoradiography film (Amersham Biosciences, Buckinghamshire, UK) in an Okamoto cassette with intensifying screen at -70 ºC for 3 days or 2-3 weeks as necessary.

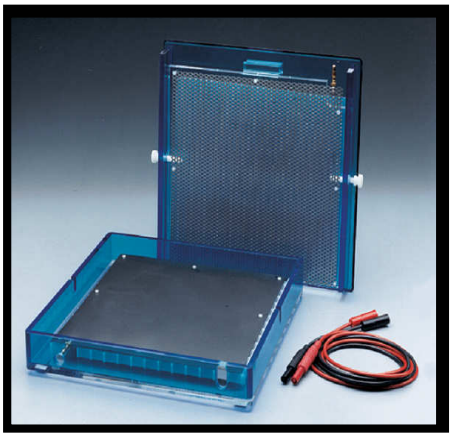


Figure A2.1: The semi-dry Electrophoretic Blotting System. The Sigma model B2529 is a plastic box (blotting chamber) inside which are contained two steel plates which act as electrodes, a stainless steel cathode and platinum coated anode. The blotting chamber is closed on top with a safety cover into which the power leads can be inserted. (SIGMA, St. Louis, MO). Photograph accessed on 1.8.2005, from the Web site: http://www.Sigmaaldrich.cim/catalog/search/ProductDetail/SIGMA/B2529

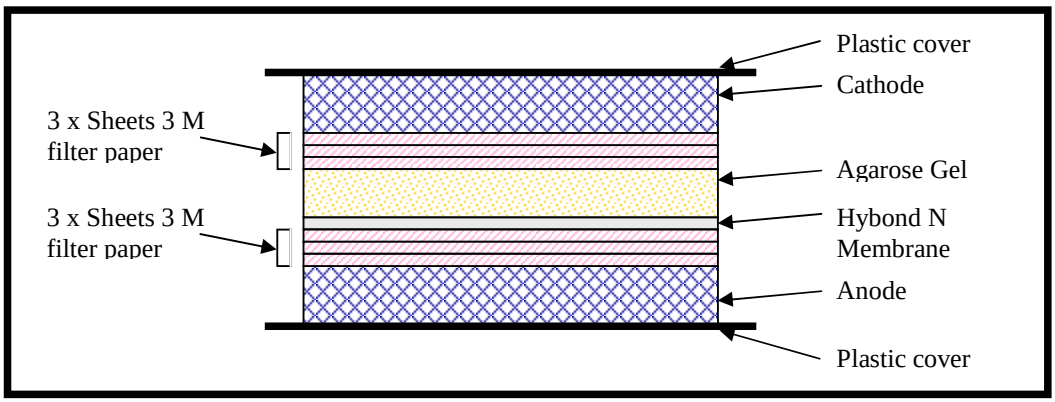


Figure A2.2: Schematic representation of the semi-dry Electroblotting Unit DNA transfer procedure. The anode sits on the bottom of the blotting chamber. Onto this is placed the mylar mask which shields the uncovered portion of the anode. Three sheets of 3M filter paper, soaked in running buffer are placed on to the 'window' cut into the mask, directly in contact with the anode. The membrane is soaked in nuclease-free water for 10 minutes, saturated in 0.3 x TBE for 20 minutes, and placed atop this. The gel (6 mm thick) is soaked in 0.3 x TBE FOR 60 minutes, and placed DNA side down, on the membrane. A pipette is rolled over it to remove trapped air bubbles. A further three sheets of 3 M filter paper, soaked as before, are added, and the cathode screwed into place, until the entire ensemble is secure. The leads are attached to the connectors protruding from the cathode and anode, and a low current (15V @ 50 mA) applied for 4 hours at room temperature. The DNA travels downwards, driven by the current from the cathode to the anode, and becomes trapped on the membrane surface. After transfer the membrane is rinsed in 0.3 x TBE and baked at 80 ºC for 2 hours, in an oven. (Taken from: Sigma-Aldrich Semi-dry Electroblotting Unit Manual, SIGMA).

A3.4 Hybridisation with Church and Gilbert Buffer as per manufacturer's instructions for Hybond-N membrane (AEC Amersham) (section 2.2.4)

The Hybond N membrane was prepared as described previously and pre-hybridised with 50 ml of Church and Gilbert Hybridisation Buffer (Hybond N Manual, Amersham International) in a Hybaid hybridisation micro-oven at 65 ºC, with rotation, for 30 minutes.

HBV DNA FL genome probe was prepared as described in section 2.3.3.2 and Appendix A section A3.5. As this area of the work involved the use of radioactive isotopes, the international rules for the handling of radioactive material were observed.

A 50 ng aliquot of FL HBV genome probe (GenBank number V01460) was radioactively labeled with 17 pmol [α-^{32}P]-dCTP, specific activity 3000 Ci/mMol (Amersham Pharmacia Biotech UL Limited, Little Chalfont, Buckinghamshire, England) using the Megaprime™ DNA labeling systems (Amersham International) as per manufacturer's instructions. For 50 ml of Church and Gilbert hybridisation buffer, HBV probe with total activity of 6-8 x 10^7 cpm was used (Hybond N Manual, Amersham International, Little Chalfont, Buckinghamshire, England). The appropriate amount of probe required to yield 6-8 x 10^7 cpm was calculated and combined with 100 µl of 100 mg/ml salmon sperm DNA (Sigma Bio Sciences), denatured at 99 ºC in a dry block (Hägar) for 5 minutes, and snap cooled on ice for 5 minutes. The probe was mixed with 1 ml of warm Church and Gilbert hybridisation buffer, and added to the remaining 49 ml in the hybridization tube. The membrane was returned to the hybridisation oven at 65 ºC for overnight hybridisation.

Non-specifically hybridized probe was removed by successive membrane washes: 2 washes 8-15 minutes each in 2 x SSC / 0.1 % SDS at ambient temperature; one wash for 10-30 minutes at 65 ºC in 1 x SSC / 0.1 % SDS; one rinse at ambient temperature in 0.1 x SSC.

Membranes were sealed in plastic and exposed to autoradiography film (Amersham Biosciences, Buckinghamshire, UK) as described previously.

Once scored, blots were stripped of all probe by washing the membrane twice in boiling 0.1 % SDS for 40 minutes, and rinsing it three times in boiling 2 x SSC. Thereafter, membranes were sealed in plastic and placed at -20 ºC to prevent dehydration, and preserve the DNA.

A3.5 Probe labeling with Megaprime™ DNA Labeling System for SH (section 2.2.4.3)

The probe was placed in a 2 ml polypropylene tube, together with 5 µl of random nonamer primers, and TE buffer to a total volume of 33 µl. This mixture was denatured at 99 ºC in a dry block for 5 minutes. The reagents were mixed by pipetting, placed on ice, and combined with 10 µl labeling buffer (dATP, dGTP, dTTP, in Tris-HCl pH7.5, 2-mercaptoethanol, $MgCl_2$ – concentrations patented), 2 units DNA polymerase I Klenow fragment (in 50 mM potassium phosphate pH6.5, 10 mM 2-mercaptoethanol, 50 % glycerol), and 17 pmol [α-^{32}]P-dCTP, specific activity 3000 Ci/mMol. The mixture was incubated in a dry block at 37 ºC for 5 minutes, 1 µl of 0.1 M spermidine (Sigma-Aldrich) was added, mixed by pipetting, and the tube returned to 37 ºC for a further 15 minutes. The reaction was stopped by heating to 99 ºC for 5 minutes, condensation precipitated by brief centrifugation, and the mixture placed on ice. Fifty µl of TE buffer was added to the labeled probe, to adjust the total volume to 100 µl.

The Sephadex columns were prepared by loading 1 ml syringes (plugged with 1 mm of filter wool [Aquarium Design, Hayward, CA, USA]) with 5 % (50 x grade) 1 x TE saturated Sephadex (Sigma-Aldrich). The Sephadex was compacted to the 0.9 ml mark by successive centrifugation steps at 6 000 x g for 5 minutes. Before the probe was loaded onto the columns, these were calibrated with 2 x 100 µl of 1 x TE (Appendix D) buffer. A 1.5 ml polypropylene tube was placed beneath the column to collect the flow-through/labeled probe.

The probe was loaded onto the Sephadex column, and centrifuged at 3 500 x g for 5 minutes. A 100 µl aliquot of TE buffer was added to the column and incubated at room temperature for 5 minutes. The column was centrifuged as before. The probe, devoid of most unincorporated isotope, was collected in the tube below. The volume was adjusted to 1 ml with 1 x TE buffer.

A 10 µl aliquot of probe was combined with 5 ml of Insta-Gel Plus scintillation gel (Perkin Elmer) in a scintillation tube, vigorously shaken, and counts per minute (cpm) measured on a liquid scintillation counter (Perkin Elmer) for 1 minute.

A4 PCRs for the Characterisation of Integrants (section 2.2.5)

A4.1 Inverse Polymerase Chain Reaction (I-PCR) (Bruni *et al.*, 1995) (section 2.2.5.1)

This PCR technique relies on the specific amplification (and later cloning) of the integrated DNA via an initial pair of specific primers, which can only be designed with some fore-knowledge of the sequence of the HBV integrant. The sequence and positioning of the primers is vital so as to exclude non-integrated episomal and replicative HBV DNA.

The results of the Southern Hybridisation technique supplied both an idea of the number of integrants present in a sample, as well as their possible approximate size. The results of the single and double RE digests could further enable the construction of a hypothetical map of each integrant or at least one integrant junction, thereby allowing for the design of primers close to the conserved specific HBV restriction sites (Figure A3 & A34). An as yet undigested aliquot of the sample DNA would be restricted with the correct RE, and ligated, yielding a circular template (Figure A3B, C & D). The primers would be orientated such that that the DNA synthesis would be primed towards the opposite ends of the template, hence the term "Inverse PCR" (Figure A3B & A4D). After amplification by I-PCR, sequencing of the product would permit characterization of a section of the integrated HBV DNA. From this data cellular primer/s directing DNA polymerisation towards integrated viral DNA, and specific to the integration under investigation, could be designed (fixed flanking primers). The fixed flanking primers would be used in fixed-flanking-primer (FFP)-PCR with genomic template (Figure A4D). Re-amplification of the FFP-PCR product could be performed if necessary (Figure A4D). A restriction digest of the product would be performed with the original respective RE, and analysed on a 2% agarose gel confirm product specificity. Automated sequencing of the product would permit full characterization of the integrant.

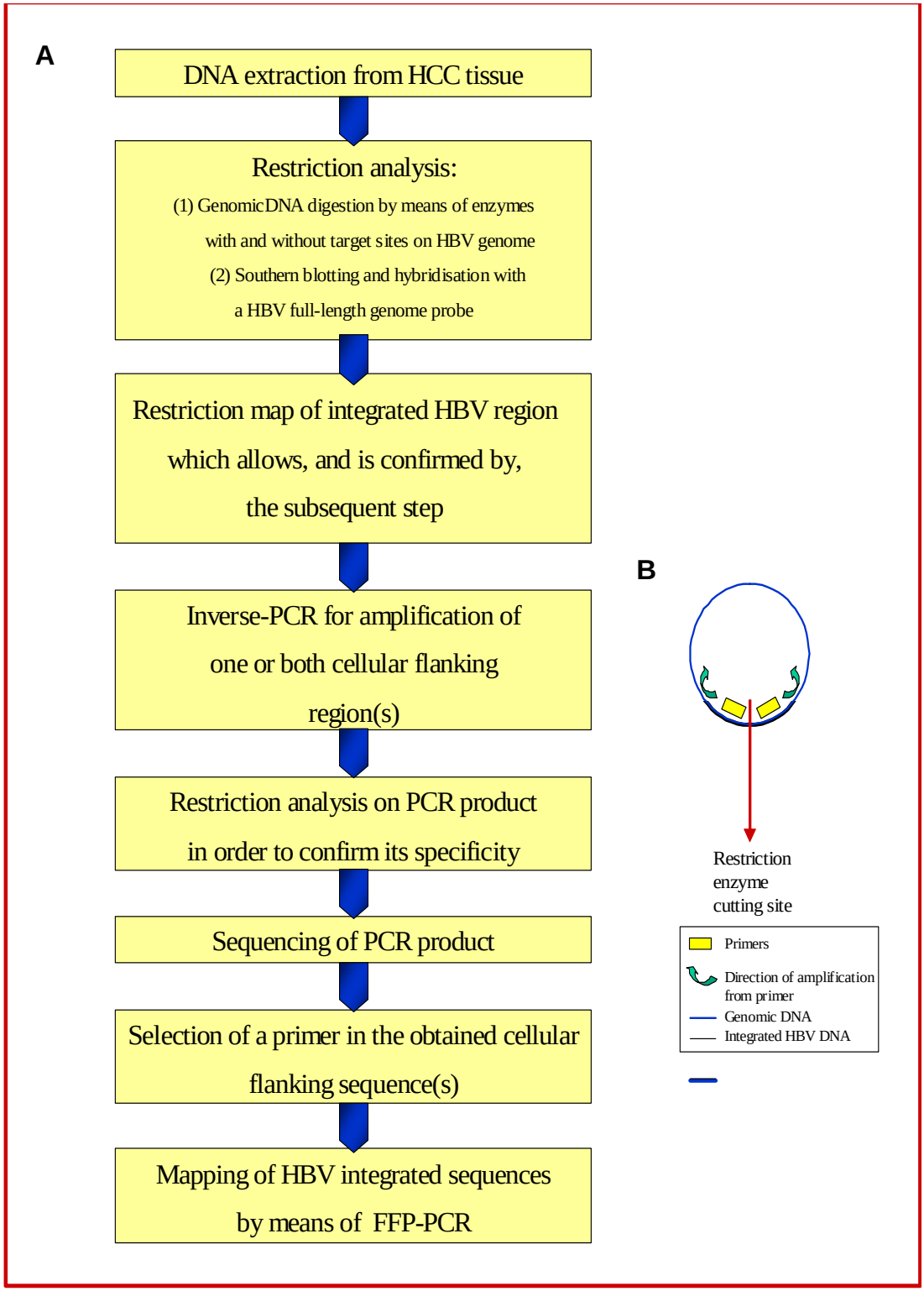


Figure A3: Outline of the procedure for the characterisation of HBV integrants by I-PCR. Reproduced from: Bruni *et al*., 1995. See text for details.

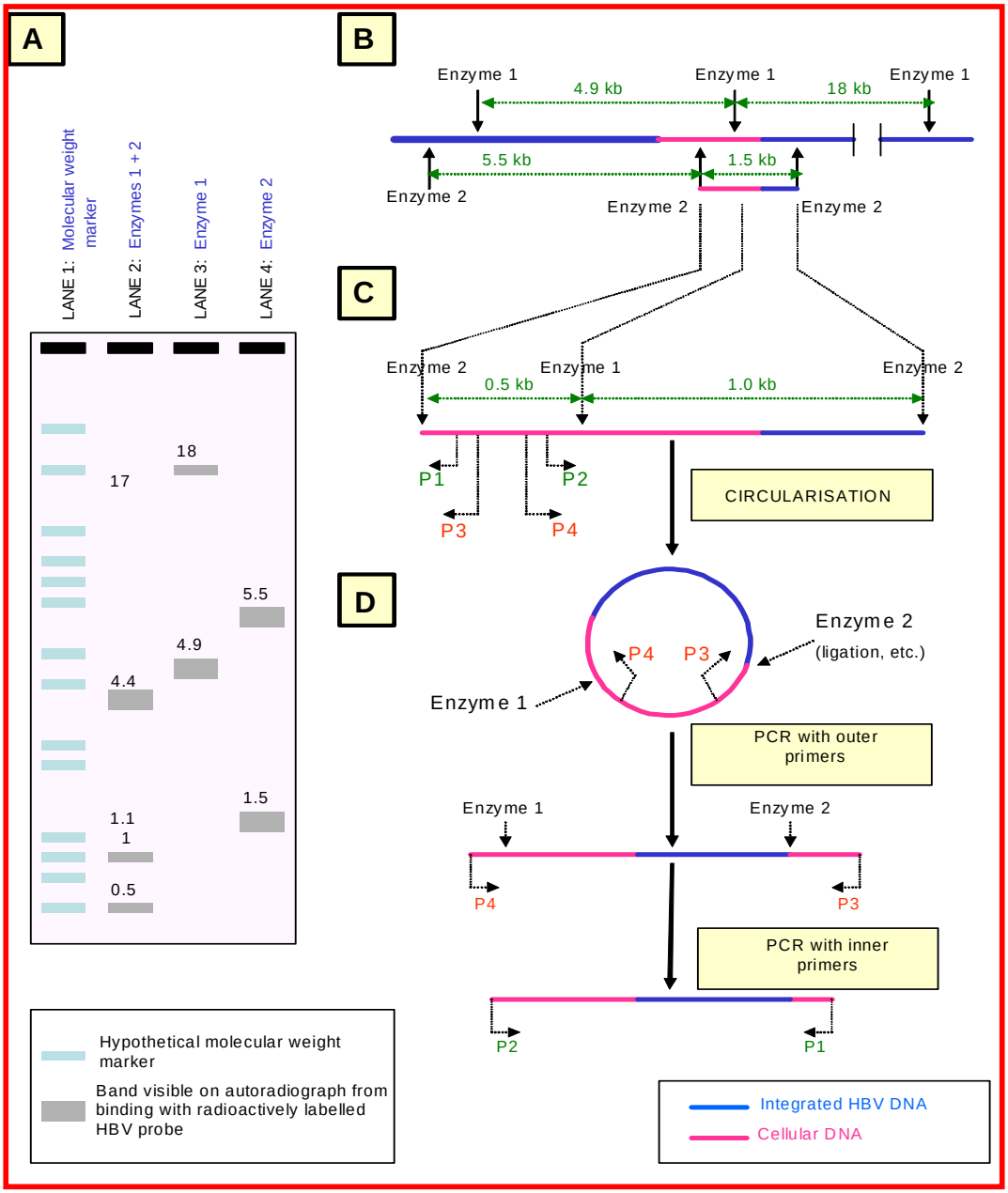


Figure A4: Outline of I-PCR concept for characterisation of HBV DNA integrants. SH autoradiograph of total sample DNA extracted from HCC tissue containing HBV integrant/s (**A**): DNA is digested with two REs and detected by ^{32}P-labeled HBV probe. Sizes of restriction fragments hybridised to probe are estimated with the use of commercially available molecular weight markers in adjacent lane. Hypothetical RE map **(B):** results of the Southern hybridisation autoradiograph are used to construct a hypothetical restriction enzyme map of the HBV integrant. Outline of IPCR technique **(C):** schematic representation of method of I-PCR used to amplifiy viral integrant junction/s (example shown for one junction). Primers are arbitrarily designated P1 to P4 with arrows representing 5'-3' orientation.

Reproduced from: Bruni *et al.*, 1995. See text for further details.

A4.2 PCRs for the direct amplification of HBV integrants off genomic DNA template (section 2.2.5.2)

The strategy of amplification for the characterization of HBV integrants from genomicDNA by *Alu*-PCR involves several rounds of PCR, only the first of which makes use of the *Alu* sequence based primers (Figure A5).

A4.2.1 First round: "*Alu* PCR" (Figure A6)

Ten cycles of amplification are performed in a final reaction volume of 50 µl comprising 100 ng of DNA; 25 mM Tris (pH 8.9); 40 mM potassium acetate (Promega); 2.5 mM magnesium chloride (Promega); 4% glycerol (Promega, Madison, USA); 200 µM each dNTP (Promega), 10 pmol *Alu*-repeat specific primer (Invitrogen, Life Technologies); 100 pmol of HBV primer (Invitrogen, Life Technologies); 1 unit Taq DNA polymerase (Invitrogen, Life Technologies); 0.04 units high-fidelity $Vent_R\beta$ DNA polymerase (New England Biolabs, Ipswich, MA, USA). Cycling was performed in a Perkin-Elmer thermal cycler (Emeryville, CA, USA) as follows: initially a hot start at 80ºC, addition of the magnesium chloride; denaturation at 94ºC for 30 seconds; annealing at 59ºC for 30 seconds; extension at 72ºC for 3 minutes; with the initial denaturation step at 94ºC for 60 seconds.

Uracil DNA glycosylase treatment of PCR primers: primers from the first round of PCR were destroyed by the direct addition of 1 unit (one unit catalyzes the release of 1 nmol of free uracil in one hour at 37EC from ^{3}H-poly-dU) of the enzyme UDG (GIBCO BRLβ, Life Technologies, Carlsbad, CA, USA) to each PCR reaction, and incubation at 37 ºC for 30 minutes; followed by heating to 94 ºC for 10 minutes to break DNA strands at apurinic dUTP sites (Figure A5).

A4.2.2 Second round (Touchdown) PCR (Don *et al*., 1994)

Ten pmol of each internal primer (eg. HB2/Tag5 or Tag3, Appendix C) was added to each PCR reaction from first round PCR, and a further I unit of ThermoPol *Taq* DNA Polymerase (New England Biolabs) with 0.04 units of $Vent_R\beta$ DNA Polymerase. Final volume was 51 µl, expected PCR product size unknown. Amplification was performed in a Perkin-Elmer

thermal cycler: denaturation at 94ºC for 30 seconds; annealing at AºC (Appendix C) for 30 seconds; extension at 70ºC for 3 minutes. The temperature was decreased by 1 degree Celsius every second cycle until A-10ºC (Appendix C) was attained, after which 20 cycles were performed at 45ºC. A final extension step at 72ºC for 8 minutes completed the protocol.

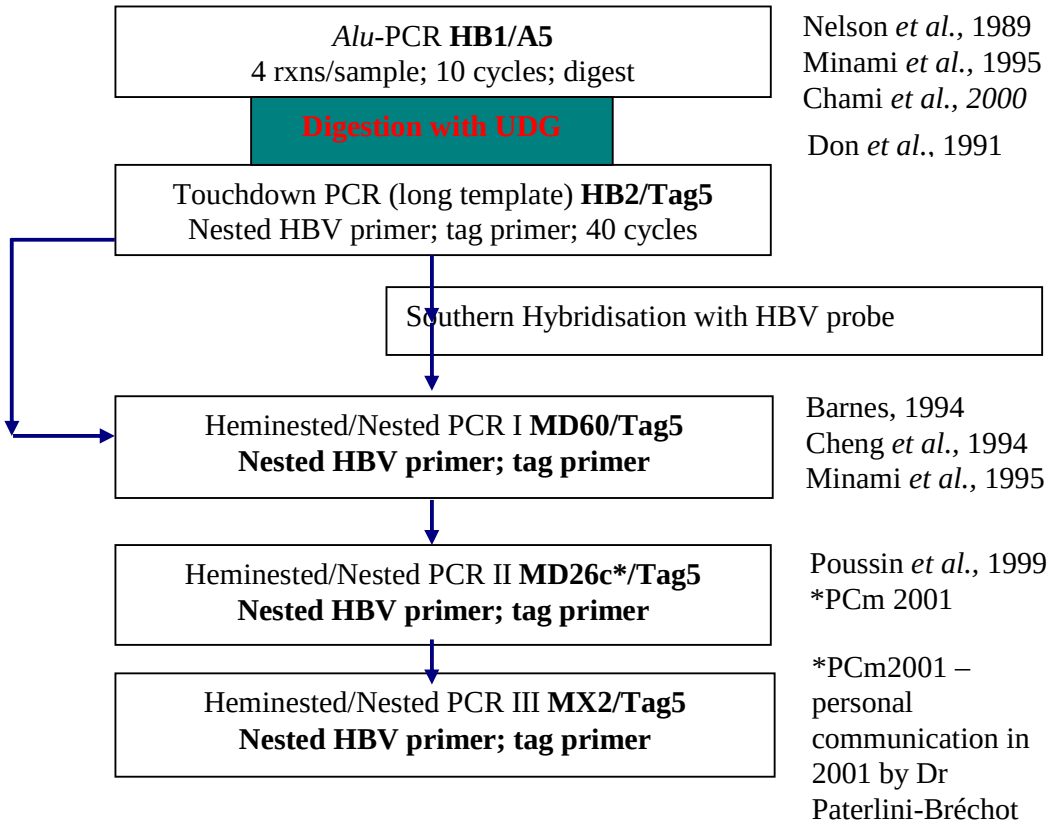


Figure A5: Outline of amplification strategy for the characterization of HBV integrants from genomicDNA. (Primer names are indicated in bold, for sequences refer to Appendix C)

The procedure begins with *Alu*-PCR, in which HBV-genomic DNA junction/s that lie sufficiently close (ie. ≤40 kb), either up or downsteam from an *Alu* repeat are amplified, while amplifications between *Alu* sequences are avoided. The primers for this PCR are constructed with dUTP instead of dTTP, and are destroyed after 10 cycles by a digestion step with UDG. Touchdown PCR follows (product may be visualized by Southern hybridisation); and a further three rounds of amplification, Nested/heminested PCRs I, II and III, may be performed if necessary.

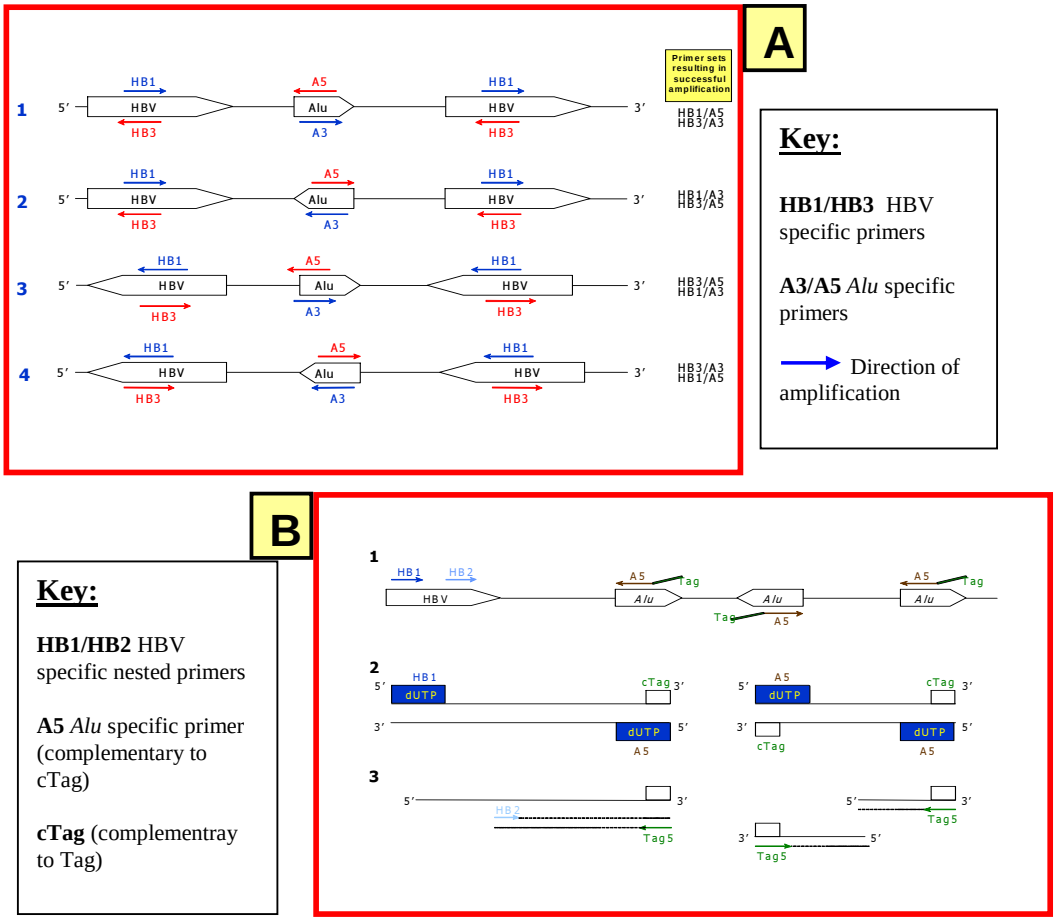


Figure A6: Outline of primer design strategy for *Alu* PCR (Minami *et al*., 1995).

(A) In order to achieve amplification, the orientation of the HBV integrant and *Alu* repeat sequences must be considered in relation to each other, and primers must be designed accordingly. The annealing sites of the HBV specific primers HB1, HB3 and nested primers HB2 (for HB1), HB4 (for HB3), and the *Alu*-repeat sequence specific primers A3/A5 and nested primers Tag3/Tag5 are indicated **(A1-4 & B 1-3)**. The PCR begins with 10 cycles of amplification, performed in four separate PCR reactions, using the 4 primers: 2 HBV-specific (one sense and one anti-sense, HB1 & HB3) and 2 *Alu*-repeat sequence specific primers (one sense and one anti-sense, A5 & A3). These primers are combined in the four possible permutations: HB1/A3; HB1/A5; HB3/A3; and HB3/A5, thus ensuring that every integration scenario possible has the potential to be amplified **(A 1-4, B1)**. The HB1/HB3 and A5/A3 primers are synthesized with dUTP instead of dTTP. **(B2)** After the initial 10 cycles, a (UDG) digestion step is performed, which results in the cleavage of HB1 and A5. **(B3)** Thereafter another 40 cycles of amplification are performed with the use of the HBV internal primers HB2, and Tag5 which is an internal primer complementary to the complementary Tag (cTag) sequence of A5. The rationale for the use of primers HB3, A3 and their internal primers HB4, Tag3 is identical. Those PCR products resulting from amplification between two *Alu*-repeat sequences will have a Tag sequence at one end only. Further amplification of these products after UDG treatment will thus be inefficient, while the product containing HBV and Tag sequences will amplify logarithmically. Theoretically, Tag primer extension products could anneal to each other and amplify, but this is rare (Minami *et al*. 1995). Refer to Appendix C for primer sequences. Reproduced from: Minami *et al*., 1995.

The total number of cycles was 40. Products were separated by agarose gel electrophoresis, followed by Southern blotting and hybridization with a full-length HBV genome probe.

A4.3 FL-PCR (section 2.2.5.3)

A4.3.1 Amplification off Total DNA (genomic template)

Per sample: 5µg of genomic DNA was combined, on ice, with 1 x reaction buffer containing 15 mM $MgCl_2$; 200 µM dNTP mixture (Promega), 1 µM each of sense (P1) and anti-sense (P2) primers; and nuclease free water, to a total volume of 20 µl. This was overlaid with 40 µl mineral oil.

A mixture comprising 3.75 units Expand High Fidelity deoxynucleoside-triphosphate deoxynucleotidyltransferase (Taq Polymerase) (Roche, Mannheim, Germany) in storage buffer (20 mM Tris-HCL, pH 7.5, 100 mM KCl, 1 mM Dithiothreitol, 0.1 mM EDTA, 0.5 % Nonidet P40, 0.5 % Tween 20, 50 % glycerol); 0.5 µl reaction buffer; and nuclease free water to a total volume of 5 µl; was prepared on ice.

Samples were heated in a pre-heated RoboCycler® temperature Cycler (Stratagene, La Jolla, CA, USA), to 94ºC for 2 minutes, and 58ºC for 30 seconds. Five µl of the enzyme mixture was then added to each sample, while the temperature was maintained at 58ºC. Thereafter samples were passed through 40 cycles of amplification comprising 10 cycles of 94ºC for 40 seconds, 58ºC for 90 seconds, 72ºC for 3 minutes; and 30 cycles of identical conditions, with the exception of the 72ºC step, where the time was increased by 2 minutes with every 10 cycles.

A4.3.2 Amplification off pSM2/HBV Clone Template for use as a probe in SH

All conditions as for genomic template, with the exception that 5 ng of the template DNA, were used. Both restricted human genomic DNA and FL-PCR products were subjected to SH (sections 2.6, 2.4) to screen for integrants with this FL-HBV genome probe.

Products were analysed by using 0.8 % agarose gel electrophoresis, and the 3.2 kb (FL HBV genome product) band excised on a UV transilluminator (UVP), and purified with the

MinElute QIAgen gel extraction kit (protocol as per manufacturer's instructions) for use as a probe in SH.

A5 Creating a ribonuclease-free environment (RNA Analysis Handbook, Promega, 2004) (section 4.2.1.1)

RNA work was carried out in an area, and with labware (pipettes, ice buckets, apparatus, plasticware etc.), dedicated exclusively for that purpose. Latex gloves were used at all times to prevent contamination of bacteria or mould from the skin.

Working surfaces were treated with Recombinant RNasin® Inhibitor (patented, Promega), which inhibits most ribonucleases, and which is guaranteed not to interfere with downstream applications. Once treated surfaces were wiped with clean paper towelling, 3 % H_2O_2, and copious amounts of Diethyl pyrocarbonate (DEPC)-treated water (Appendix D).

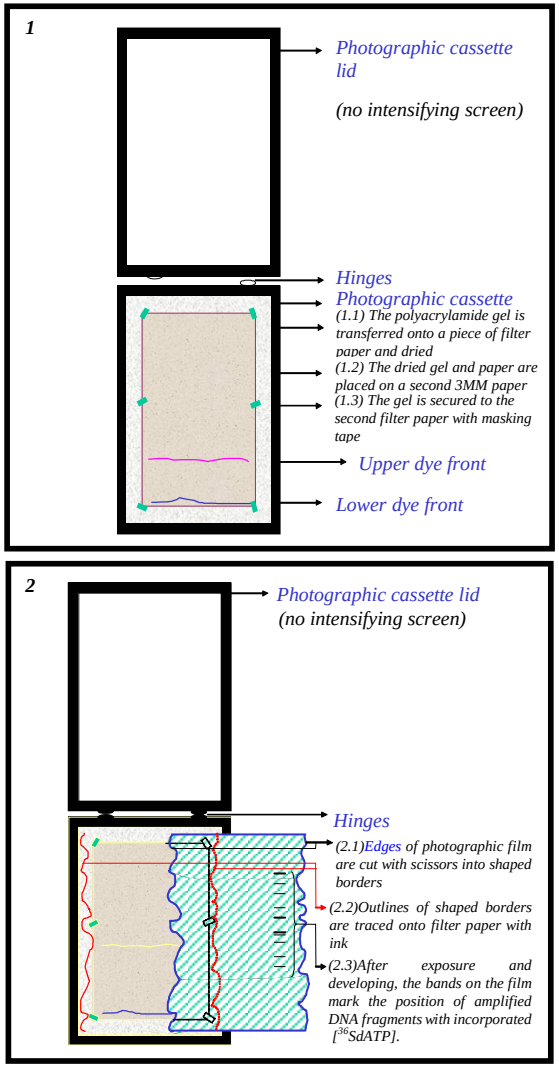
All reagents used in RNA work were exclusively reserved for this purpose, and handled with the use of sterile technique, baked spatulas and sterile paper weighing boats. Solutions were treated with DEPC (by overnight incubation and stirring at room temperature and autoclaving at 121 ºC for 30 minutes to inactivate DEPC). Tris-based buffers, whose free amino group would react with the DEPC, thus inhibiting their buffering action, were prepared with ribonuclease free tested Tris, DEPC-treated water, and correctly treated glassware etc.

Sterile, disposable ribonuclease and deoxyribunuclease free plastic-ware was used. Non-disposable glassware was washed with Vircon (Roche) soap and tap water, rinsed in tap water, rinsed in 70 % EtOH (Roche) to remove residual soap, and rinsed several times in deionised distilled water; followed by 3 rinses in DEPC-treated water (Appendix D). Each item was wrapped in foil, autoclaved, and baked for 16 hours at 250 ºC.

Items such as electrophoresis troughs were soaked for 15 minutes in 3 % H_2O_2 and rinsed 3 times in DEPC-treated water before use.

A6 Polyacrylamide gel electrophoresis (PAGE) and extraction of DNA from the polyacrylamide gel (sections 4.2.4.1, 4.2.5)

The PCR generated fragments were resolved by PAGE (Figure A7). The polyacrylamide gel was transferred onto filter paper, dried (Figure A7:1.1-1.3) placed in a photographic cassette with an intensifying screen (Figure A7:1) and exposed to autoradiographic film (Figure A7:2). The edges of the film were shaped, the film secured to the filter paper and the pattern at the edges traced onto the filter paper for future aligning purposes. After developing, the bands on the film were pierced through to the polyacrylamide gel (Figure A7:3) to mark the positions of the fragments (Figure A7:2.1-2.3; A7:3.1-3.3). Fragments were excised from the polyacrylamide gel with a sterile blade and the DNA boiled off in distilled water (Figure A7:3.4) and re-amplified by PCR. By extracting and amplifying the DNA from each fragment separately, the various products of one PCR can be sequenced individually.



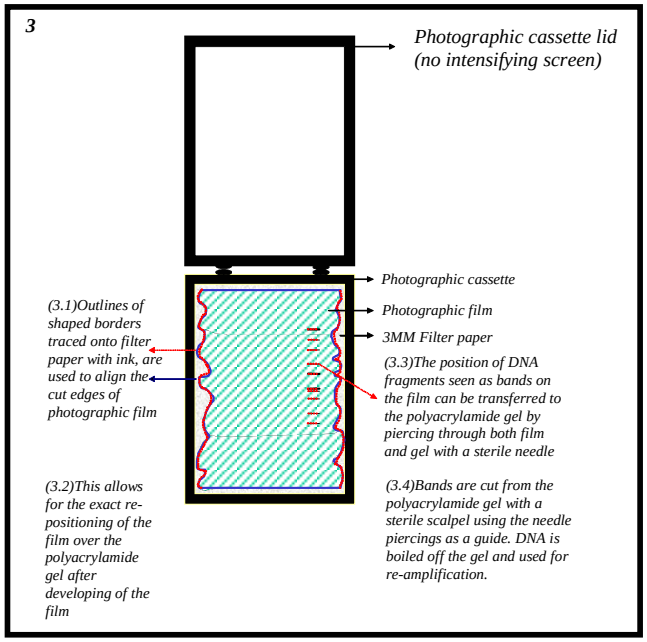


Figure A7: Diagrammatic representation of procedure for the marking and excision of bands from polyacrylamide gels. Once excised, the gel/paper is soaked and boiled in water to enable recovery of the DNA for further amplification. By extracting and amplifying DNA from each band separately, the various products of one PCR can be sequenced individually.

APPENDIX B

Diagrammatic representation of plasmid pSM2

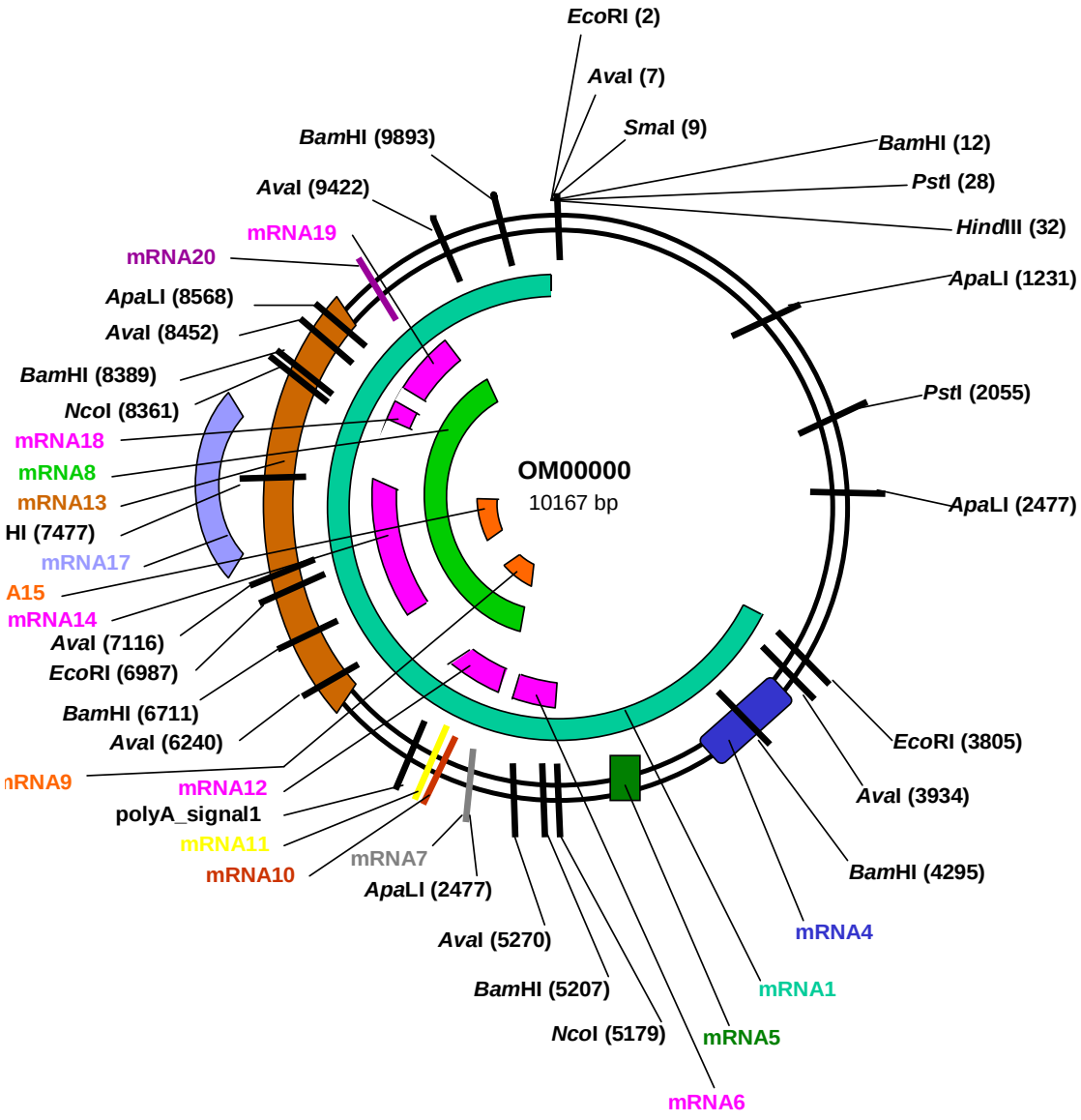


Figure B1: Plasmid pSM2 contains an *Eco*RI head-to-tail dimer of full-length HBV genotype D, subtype *ayw* (Galibert *et al.*, 1979) which was cloned via an *Eco*RI site (Sommer *et al.*, 1997). Reproduced from a map kindly supplied by Professor H. Will (Heinrich-Pette-Institut fur experimentalle Virologie und Immunologie an der Universitat Hamburg, Germany).

APPENDIX C

PRIMER SEQUENCES

Table VIII: DNA oligonucleotide sequences for PCR.

Name and position* if applicable	Sequence	Tm (˚C); 50mM (Na^+)
HBV-specific primers		
[1]HB1+ (1133-1154)	5' - ACAUGAACCUUUACCCCGUUGC - 3'	62
[1]HB2+ (1165-1193)	5' - GCGCTGCAGTGCCAAGTGTTTGCTGACGC - 3'	76
[1]HB3- (2489-2510)	5’ – GAGUUCUUCUUCUAGGGGACCU - 3’[A]	62
HB4-	Sequence never published	
[2]FX1 (1243-1272)	5' – ACCTTTGTGGCTCCTCTGCCGATCCATACT – 3’	73
[2]FX2 (1276-1295)	5’ – GAACTCCTAGCAGCTTGTTT – 3’	56
[2]FX3 (1847-1866)	5’ – TGAACAGTAGGACATGAACA – 3’	54
[2]FX4 (1871-1900)	5’ – GCCCCAAAGCCACCCAAGGCACAGCTTGGA – 3’	77
[1&3]A3-	5' - AGUGCCAAGUGUUUGCUGACGACUGCACUC CAGCCUGGGCGAC - 3'	84
[1]A5-	5' - CAGUGCCAAGUGUUUGCUGACGCCAAAGUG CUGGGAUUACAG - 3'	81
[4&5]MD60 (1258-1279)	5' - CTGCCGATCCATACTGCGGAAC - 3'[B]	66
[4]MD26C (1607-1625)	5' - CATGGAGACCACCGTGAAC - 3'	60
[4]MX2 (1639-1662)	5' - TGCCCAAGGTCTTACATAAGAGGA - 3'	64
[6]1374+ (1374-1393)	5’- ATGGCTGCTAGGCTGTACTG -3’	60
[7]1732+ (1730-1747)	5’ - CTGGGAGGAGTTGGGGGA - 3’	61
[7]1765+ (1763-1781)	5’ - GGTCTTTGTACTAGGAGGCTG - 3’	61
[8]1898+ (1896-1916)	5’- GGCATGGACATTGACCCGTA – 3’	54
[7]1966- (1946-1966)	5’ - GTCAGAAGGCAAAAACGAGAG – 3’	60
[9]P1 (1820-1841)	5’- CCGGAAAGCTTGAGCTCTTCTTTTTCACCTC TGCCTAATCA -3’ (HBV sequence underlined)	55

[9]P2 (1813-1853)	5'- CCGGAAAGCTTGAGCTCTTCAAAAAGTTGCATG GTGCTGG -3' (HBV sequence underlined)	56
Tag primers		
[1]Tag3	5' - CCAAGTGTTTGCTGACGACTGCA - 3'	65
[1]cTag3	5' - TGCAGTCGTCAGCAAACACTTGG - 3'	65
[1]Tag5	5' – CAAGTGTTTGCTGACGCCAAAG – 3'	62
[1]cTag5	5'- CTTTGGCGTCAGCAAACACTTG - 3'	62
Primers for reverse-amplification of mRNA		
Oligo $d(T)_{15}$	5' - TTTTTTTTTTTTTTT - 3'	24
[11]Anchored Oligo $d(T)_{17}$	Oligo d $(T)_{17}$ VN, where V is A, G or C, and N is A, G, C or T	36-40
Beta-Galactosidase Gene Primers		
[10]GAPDH +	5' - CCCTTCATTGACCTCAACTACATG - 3'	64
[10]GAPDH -	5' - CATGCCAGTGAGCTTCCCGTTCAG - 3'	69

(+) sense (-) anti-sense

* Numbers in brackets denote nucleotide position of primer in relation to HBV genome V01460 where *Eco*RI cleavage site is position 1.

[A]This primer was a sense primer and should have been anti-sense. Attempts to confirm that the incorrect sequence had been published were unsuccessful.

[B]Later publications (Murakami *et al*., 2004; Saigo *et al*., 2008) equated the HB3 primer with MD60, and published the MD60 sequence as that of HB3. In this case MD60 would not be used as a nested primer for HB3.

Note:

Primers:

1. were synthesized commercially by:
   Integrated DNA Technologies (Coralville, Iowa, USA), GIBCOBRL Life Technologies (Paisley, Scotland, UK), and Sigma (Saint Louis, Missouri, USA)
2. were designed with the use of:
   http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-primer.html
3. Tm values were calculated by Net primer Tm calculator for 50mM (Na^+)
   http://www.basic.northwestern.edu/biotools/oligocalc.html

References:

1 Minami *et al*., 1995
2 Kawai *et al*., 2001
3 Chumakov *et al*., 1992
4 P. Paterlini-Bréchot – personal communication
5 Murakami *et al*., 2004
6 Unpublished - primer designed in house
7 Kramvis *et al*., 1996
8 Owiredu *et al*., 2001
9 Günther *et al*., 1998
10 Weinberg *et al*., 2000
11 Zhu & Liang 1997

APPENDIX D

RECIPES

Acrylamide solution 40%

38 g Acrylamide
2 g BIS
5 g Amberlite mixed bed resin
50 ml ddH_2O
Cover container with foil

Stir at room temperature for 30 minutes
Filter through Whatmann filter paper
Make up to 100 ml volume with ddH_2O

Agarose Gels for RNA Resolution

Prepare a 1% agarose/formaldehyde gel (containing 0.5 µg ethidium bromide)
20.0 ml 5x MOPS buffer
72.0 ml DEPC-treated ddH_2O
1.0 g agarose, molecular biology grade

Heat to boiling (in a microwave oven) and cool to about 55°C
In a fume hood, add 17.6 ml 37% formaldehyde and 5 µl of 10 mg/ml ethidium bromide to the mixture
Gently mix
Pour the gel and allow to cool
Prepare the RNA samples by mixing 1 part RNA sample with two parts RNA sample buffer to a total volume of 10-30 µl, depending on how much volume can be loaded on the gel
Heat the sample to 65°C for 5 minutes and then cool on ice for 2 minutes
Add 2 µl of RNA loading buffer
Pre-run the gel for 10 minutes in 1 x MOPS buffer prior to loading the samples
Load the RNA samples and run the gel at 4-5 V/cm
Continue electrophoresis until the bromophenol blue dye has migrated at least 2/3 to ¾ of the length of the gel

Bromophenol Blue Dye

20 mg of 0.1% bromophenol blue dye
4 ml of 0.1M EDTA
10 ml of 50% glycerol

Deionised Formamide

Pour formamide into a 2 L beaker
Add 5% amberlite MB-1
Stir for 30 minutes at room temperature
Remove fluid formamide by filtering through a Buchner funnel and Whatmann filter paper
Repeat the filtering step
Aliquot into 100 ml botlles
Store at -20ºC

DEPC treated Water

1 ml DEPC
1 L ddH_2O

Add the DEPC to the ddH_2O while stirring gently
Cap the bottle loosely and leave to stir gently overnight
Autoclave
Store at room temperature

Freezing medium
A total volume of 50 ml of medium requires:

Complete medium
50 ml Commercially available MEM, DMEM, or EMEM) appropriate for cell line.
Heat to 37ºC
Require approximately 10 ml for 4 000 000 cells

FBS (complement inactivated)
5 ml FBS for a final concentration of 20% (complete medium already contains 10% FBS)
Pre-heat to 37ºC

Glycerol or DMSO
5 ml required for a final concentration of 10%

Combine complete medium, FBS, glycerol or DMSO, and stir well
Make up fresh every time

Glycerol Tolerant Buffer

108 g Tris base
36 g Taurine
2 G $Na_2EDTA.2H_2O$
ddH_2O to 500 ml
Filter sterilise

Modified Church and Gilbert Buffer for SH

0.5M phosphate buffer consisting of:
720ml of 0.5M Na_2HPO_4
280ml of 0.5M NaH_2PO_4
pH 7.2
7% w/v SDS
10 mM EDTA

Gel dye for differential display

95% formamide
10mM EDTA (pH 8.0)
0,01% Xylene Cyanol
0,01% Bromophenol Blue
Store at room temperature

Luria Broth

10 g bacto-tryptone
5 g bacto-yeast extract
5 g NaCl
ddH_2O to 900 ml
pH to 7.5 with NaOH
Aliquot into 100 ml bottles
Autoclave immediately

Luria Broth Agar Plates

1 g Tryptone
0.5 g Yeast extract
1 g NaCl
ddH_2O to 95 ml
1.5 g Agar
Autoclave and cool to 48ºC
Add 100 µl (50 mg/ml) selective antibiotic (usually ampicillin or tetracycline), in hood, using sterile technique
Pour into sterile culture plates (to thickness 1.5 cm) in hood using sterile technique
Allow to set
Dry lid down in 37ºC incubator overnight (also to check for contamination)
Store at 4ºC for up to 1 month

Lysing Buffer for Phenol Chloroform DNA Extraction

420.42 g of 7 M urea
17.53 g of 0.3 M NaCl
3.72 g of 10 mM EDTA
10 ml of 10 mM Tris (pH7.5)
ddH_2O to 1 L

MOPS Buffer 5x

0.2 M MOPS (pH 7.0)
0.05 M Sodium Acetate
0.005 M EDTA (pH 8.)

83.72 g MOPS (free acid)
8.23 g Sodium acetate
1.6 L DEPC H_2O
20 ml DEPC-treated 0.5 M EDTA
adjust pH to 7.0 with 10 N NaOH
Make up to 2 L volume with DEPC H_2O
Dispense into 200 ml aliquots and autoclave

Phosphate Buffered Saline (PBS)
8 g NaCl
0.2 g KCl
1.44 g Na_2HPO_4
0.24 g KH_2PO_4
ddH_2O to 800 ml
pH to 7.4 with HCl
ddH_2O to 1 L
Autoclave

Polyacrylamide Denaturing Gel (0.6%)

31.5 g Urea
3.75 ml Glycerol tolerant buffer
15 ml 40% Polyacrylamide
26.25 ml ddH_2O
Stir until urea dissolves (~30 Minutes)
450 µl Ammonium persulphate
18.75 µl TEMED

RNA Loading Buffer

50% high grade Glycerol
1 mM EDTA
4% bromophenol blue

Prepare in DEPC-treated H_2O
Dispense into single use aliquots
Store at -20°C

RNA Sample Buffer

10 .0 ml deionized formamide
3.5 ml 37% formaldehyde
2.0 ml 5x MOPS buffer

Dispense into single aliquots
Store in tightly sealed, screw cap tubes, at -20°C for up to 6 months

SDS Buffer 10%

Autoclave 1 L ddH_2O
Dissolve 100 g SDS in 900 ml of this water with gentle heat
Adjust pH to 7.2 with HCl
Make up to volume with remaining autoclaved ddH_2O

SSC Buffer 20x

350.7 g NaCl
44.13 g Na citrate
Make up to 2 L with ddH_2O
Autoclave

TBE Buffer 10x

108 g Tris base
55 g Boric acid
7.44 G Na_2EDTA
Add ~800 ml ddH_2O
pH to 8.0
Make up to 1 L volume with ddH_2O

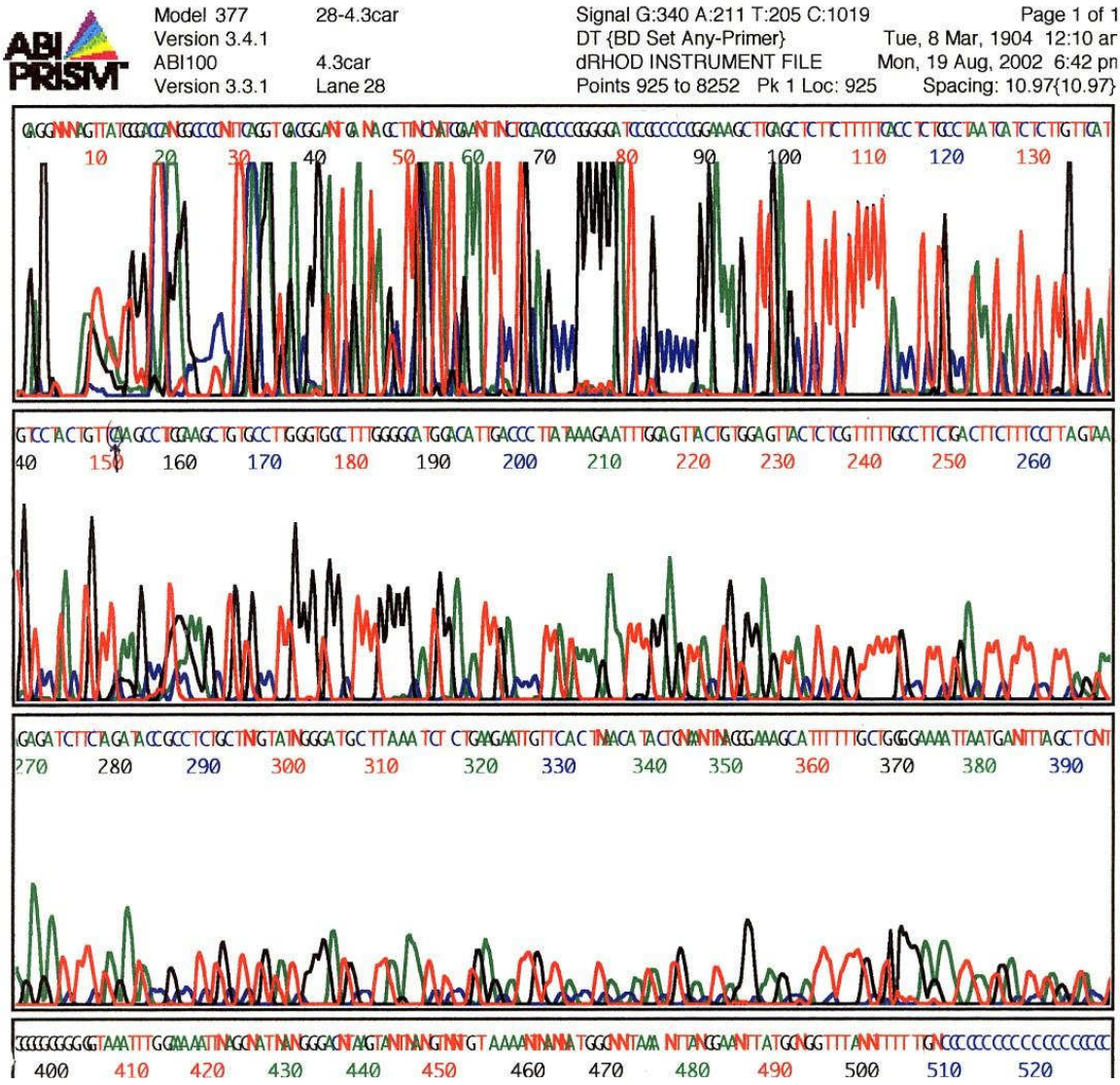
TE Buffer (pH 8) 1x

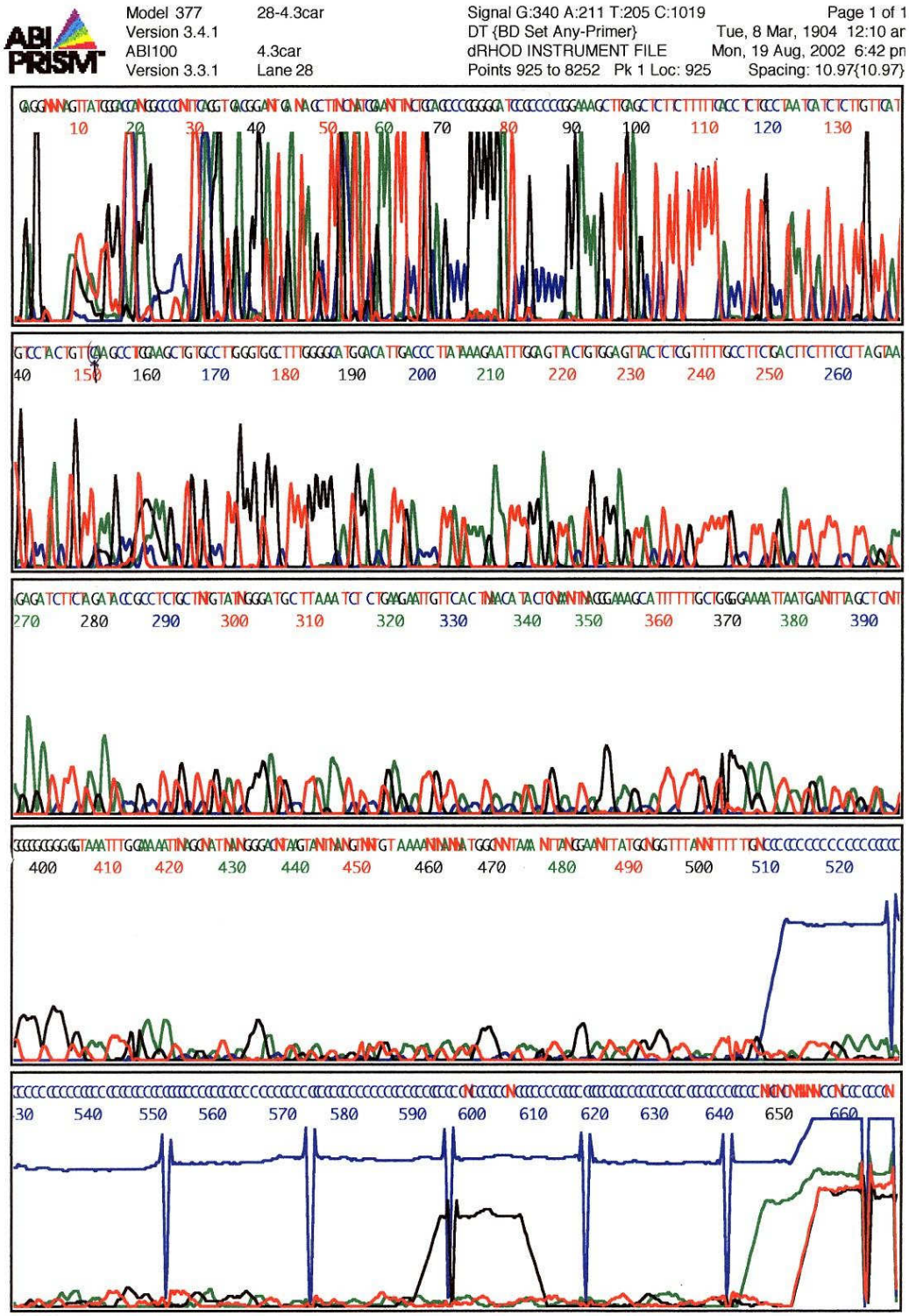
1 ml of 1 M EDTA
10 ml of 1 M Tris (pH 8.0)
Add ~800 ml ddH_2O
pH to 8.0
Make up to 1 L volume with ddH_2O
Autoclave

APPENDIX E

CHROMAS RESULTS FOR

SEQUENCE OF T45.3




Model 377
Version 3.4.1
ABI100
Version 3.3.1
28-4.3car
4.3car
Lane 28
Signal G:340 A:211 T:205 C:1019
DT {BD Set Any-Primer}
dRHOD INSTRUMENT FILE
Points 925 to 8252 Pk 1 Loc: 925
Page 1 of 1
Tue, 8 Mar, 1904 12:10 ar
Mon, 19 Aug, 2002 6:42 pn
Spacing: 10.97{10.97}

APPENDIX F

ETHICS CLEARANCE CERTIFICATES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Martins-Furness

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER M071048**
PROJECT	Characterisation of Hepatitis B Virus DNA Integrants in Liver of Southern African Blacks with Heptocellular Carcinoma
INVESTIGATORS	Ms C Martins-Furness
DEPARTMENT	Department of Medicine
DATE CONSIDERED	07.10.26
DECISION OF THE COMMITTEE*	APPROVED UNCONDITIONALLY

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 07.10.30 **CHAIRPERSON** .....
(Professors PE Cleaton-Jones, A Dhai, M Vorster, C Feldman, A Woodiwiss)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Prof M Kew

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10005, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Kew

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER 40351**
PROJECT	The aetiology and pathogenesis of hepatocellular carcinoma in Black Africans
INVESTIGATORS	Prof C Kew
DEPARTMENT	Medicine
DATE CONSIDERED	04.03.19
DECISION OF THE COMMITTEE*	Approved unconditionally (old number 939135 - renewal of grant)

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Prof MC Kew

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10005, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX G

GENBANK DATA SEARCHES FOR SEQUENCE T45.3

T45.3 bp 1-55

BLAST Basic Local Alignment Search Tool

Edit and Resubmit Save Search Strategies Formatting options Download

Nucleotide Sequence (55 letters)

Results for: lcl|7243 None(55bp)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|7243
Description
None
Molecule type
nucleic acid
Query Length
55

Database Name
nr
Description
All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program
BLASTN 2.2.21+ Citation

Reference
Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Search Parameters

Program	blastn
Word size	28
Expect value	10
Hitlist size	100
Match/Mismatch scores	1,-2
Gapcosts	0,0
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Database

Posted date	Sep 27, 2009 5:41 PM
Number of letters	28,993,070,215
Number of sequences	9,897,685
Entrez query	none

Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	1.33271	1.28
K	0.620991	0.46
H	1.12409	0.85

Results Statistics

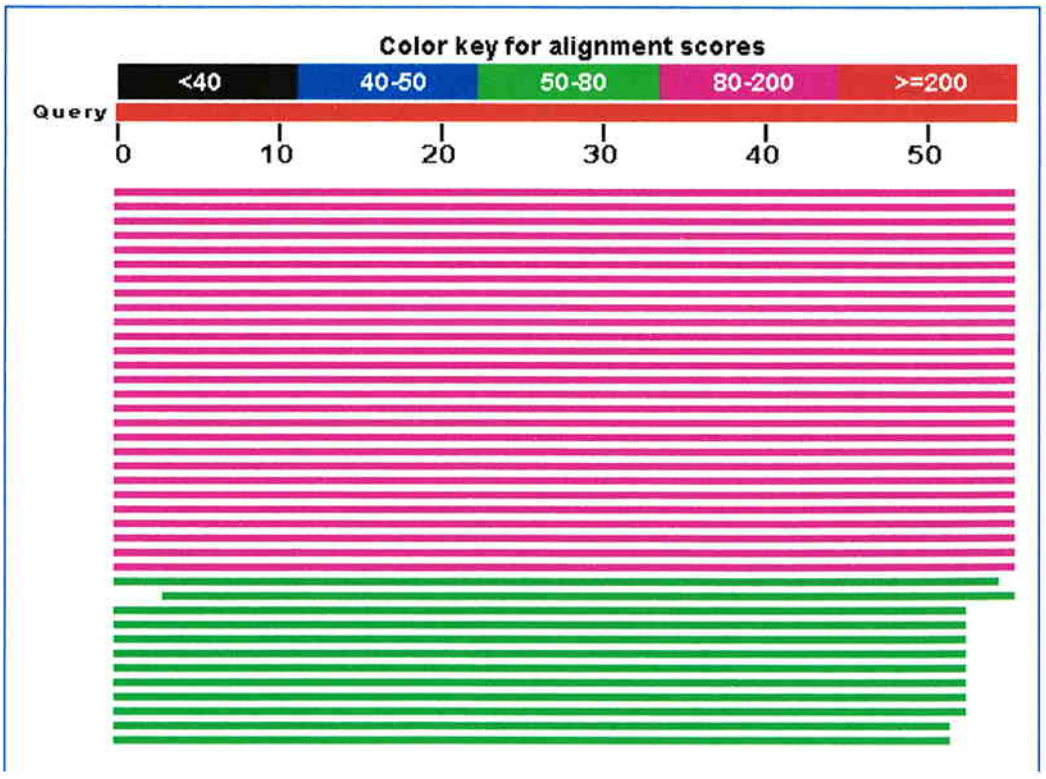
Length adjustment	29
Effective length of query	26
Effective length of database	28706037350
Effective search space	746356971100
Effective search space used	746356971100

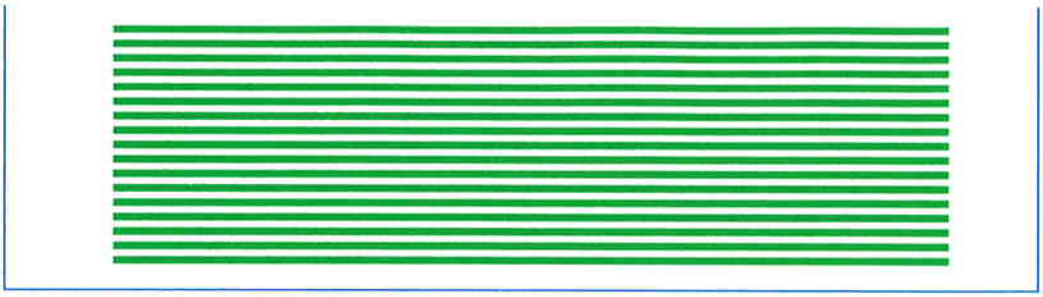
Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.





Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

AM087268.1	Uncultured ectomycorrhiza (Thelephoraceae) 18S rRNA gene (partial), ITS1, 5.8S rRNA gene, ITS2 and 25S rRNA gene (partial), clone 11-3C	80.5	80.5	100%	4e-13	92%
AB241267.1	Verasper moseri DNA, microsatellite marker Vemos64	80.5	80.5	100%	4e-13	92%
AB241258.1	Verasper moseri DNA, microsatellite marker Vemos43	80.5	80.5	100%	4e-13	92%
AB114435.1	Pseudorasbora parva DNA, microsatelite marker PA19	80.5	80.5	100%	4e-13	92%
AJ620495.1	Mus musculus partial mRNA for non-selective cation channel TRPV1 (trpv1 gene)	80.5	80.5	100%	4e-13	92% U G
AF327894.1	Cloning vector pExchange 1, complete sequence	80.5	80.5	100%	4e-13	92%
AJ567368.2	Vitis vinifera samdc gene for S-adenosylmethionine decarboxylase proenzyme, mRNA	80.5	80.5	100%	4e-13	92% U G
AF076778.1	Cloning vector pCMVTAG5c, complete sequence	80.5	80.5	100%	4e-13	92%
AF076777.1	Cloning vector pCMVTAG5b, complete sequence	80.5	80.5	100%	4e-13	92%
AF076312.1	Cloning vector pCMVTAG5a, complete sequence	80.5	80.5	100%	4e-13	92%
AF076311.1	Cloning vector pCMVTAG4c, complete sequence	80.5	80.5	100%	4e-13	92%
AF076310.1	Cloning vector pCMVTAG4b, complete sequence	80.5	80.5	100%	4e-13	92%
AF073000.1	Cloning vector pCMVTAG4a, complete sequence	80.5	80.5	100%	4e-13	92%
AF072999.1	Cloning vector pCMVTAG3c, complete sequence	80.5	80.5	100%	4e-13	92%
AF072998.1	Cloning vector pCMVTAG3b, complete sequence	80.5	80.5	100%	4e-13	92%
AF072997.1	Cloning vector pCMVTAG3a, complete sequence	80.5	80.5	100%	4e-13	92%
AF072540.1	Cloning vector pCMVTAG2c, complete sequence	80.5	80.5	100%	4e-13	92%
AF072539.1	Cloning vector pCMVTAG2b, complete sequence	80.5	80.5	100%	4e-13	92%
AF072538.1	Cloning vector pCMVTAG2a, complete sequence	80.5	80.5	100%	4e-13	92%
AJ007320.1	Homo sapiens mRNA for immunoglobulin heavy chain variable region, clone KC4	80.5	80.5	100%	4e-13	92% U
AF067646.1	Cloning vector pCMV-scriptEX, complete sequence	80.5	80.5	100%	4e-13	92%
AF028239.1	Mammalian expression vector pCMV-Script, complete sequence	80.5	80.5	100%	4e-13	92%
Y13161.1	Sus scrofa microsatellite DNA isolate pGC.60	80.5	80.5	100%	4e-13	92%
U46018.1	Cloning vector pCRSCRIPT Cam, complete sequence	80.5	80.5	100%	4e-13	92%
U46017.1	Cloning vector pPCR-Script Amp SK(+), complete sequence	80.5	80.5	100%	4e-13	92%
X91477.1	Uncultured bacterium partial 16S rRNA gene (clone group A54n)	80.5	80.5	100%	4e-13	92%
X91293.1	Bacterial sp. partial 16S rRNA gene (clone group G41)	80.5	80.5	100%	4e-13	92%
AJ132233.1	Enterobacter gergoviae glgC (partial) and glgA (partial) genes, strain 57-7	78.7	78.7	98%	2e-12	92%
AB241259.1	Verasper moseri DNA, microsatellite marker Vemos44	75.0	75.0	94%	2e-11	92%
AY303671.1	Broad host range expression vector pMHE7Tc, complete sequence	75.0	75.0	94%	2e-11	92%
AY303669.1	Broad host range expression vector pMHE7, complete sequence	75.0	75.0	94%	2e-11	92%
AY299693.1	Broad host range expression vector pMHE2, complete sequence	75.0	75.0	94%	2e-11	92%
AJ277653.1	Escherichia coli plasmid pGBG1	75.0	75.0	94%	2e-11	92%
AF179626.1		75.0	75.0	94%	2e-11	92%

	Expression vector pGP100, complete sequence				
AF179625.1	Shuttle vector pCS513, complete sequence	75.0	75.0	94%	2e-11 92%
AF178582.1	Expression vector pCS316 complete sequence	75.0	75.0	94%	2e-11 92%
AB005666.1	Homo sapiens mRNA for GTPase-activating protein, complete cds	75.0	75.0	94%	2e-11 92%
FJ834451.1	Expression vector pTL80, complete sequence	73.1	73.1	92%	7e-11 92%
FJ834450.1	Expression vector pTL85, complete sequence	73.1	73.1	92%	7e-11 92%
FJ602701.1	Expression vector pFJ1.1-yk1350a08, complete sequence	73.1	73.1	92%	7e-11 92%
FJ410917.1	Binary vector pNML54, complete sequence	73.1	73.1	92%	7e-11 92%
FJ380063.1	Suicide vector pMQ200, complete sequence	73.1	73.1	92%	7e-11 92%
FJ380062.1	Shuttle vector pMQ175, complete sequence	73.1	73.1	92%	7e-11 92%
FJ386489.1	Cloning vector pEX302, complete sequence	73.1	73.1	92%	7e-11 92%
AB464951.1	Secale cereale DNA, rye-species specific DNA element, clone: lambda11-2,1.5subclone	73.1	73.1	92%	7e-11 92%
AB304876.1	Shuttle vector pRS446TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304875.1	Shuttle vector pRS446PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304874.1	Shuttle vector pRS446GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304873.1	Shuttle vector pRS446ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304872.1	Shuttle vector pRS445TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304871.1	Shuttle vector pRS445PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304870.1	Shuttle vector pRS445GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304869.1	Shuttle vector pRS445ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304868.1	Shuttle vector pRS444TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304867.1	Shuttle vector pRS444PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304866.1	Shuttle vector pRS444GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304865.1	Shuttle vector pRS444ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304864.1	Shuttle vector pRS436TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304863.1	Shuttle vector pRS436PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304862.1	Shuttle vector pRS436GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304861.1	Shuttle vector pRS436ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304860.1	Shuttle vector pRS435TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304859.1	Shuttle vector pRS435PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304858.1	Shuttle vector pRS435GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304857.1	Shuttle vector pRS435ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304856.1	Shuttle vector pRS434TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304855.1	Shuttle vector pRS434PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304854.1	Shuttle vector pRS434GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304853.1	Shuttle vector pRS434ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304852.1	Shuttle vector pRS433TEF DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304851.1	Shuttle vector pRS433PGK DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304850.1	Shuttle vector pRS433GAP DNA, complete sequence	73.1	73.1	92%	7e-11 92%
AB304849.1	Shuttle vector pRS433ADH DNA, complete sequence	73.1	73.1	92%	7e-11 92%
EU546822.1	Cloning vector pMQ125, complete sequence	73.1	73.1	92%	7e-11 92%
EU546821.1	Cloning vector pMQ124, complete sequence	73.1	73.1	92%	7e-11 92%
AM940001.1	Cloning vector pIMC	73.1	73.1	92%	7e-11 92%
AM940000.1	Cloning vector pIMK	73.1	73.1	92%	7e-11 92%
EU257522.1	Cloning vector pNC-GFP, complete sequence	73.1	73.1	92%	7e-11 92%
EU257520.1	Cloning vector pTOR, complete sequence	73.1	73.1	92%	7e-11 92%
EU257519.1	Cloning vector pSAM, complete sequence	73.1	73.1	92%	7e-11 92%

AM884385.1	Cloning vector pGHGWY	73.1	73.1	92%	7e-11 92%
AM884384.1	Cloning vector pGHGWG	73.1	73.1	92%	7e-11 92%
AM884383.1	Cloning vector pGHGWC	73.1	73.1	92%	7e-11 92%
EU186089.1	Binary vector pCLEAN-G144, complete sequence	73.1	73.1	92%	7e-11 92%
EU186088.1	Binary vector pCLEAN-G146, complete sequence	73.1	73.1	92%	7e-11 92%
EU186087.1	Binary vector pCLEAN-G147, complete sequence	73.1	73.1	92%	7e-11 92%
EU147786.1	Binary vector pGPro2, complete sequence	73.1	73.1	92%	7e-11 92%
AM884388.1	Cloning vector pGBGWY	73.1	73.1	92%	7e-11 92%
AM884387.1	Cloning vector pGBGWG	73.1	73.1	92%	7e-11 92%
AM884386.1	Cloning vector pGBGWC	73.1	73.1	92%	7e-11 92%
AM884382.1	Cloning vector pGKGWC	73.1	73.1	92%	7e-11 92%
AM884381.1	Cloning vector pGKGWG	73.1	73.1	92%	7e-11 92%
AM884380.1	Cloning vector pGKGWY	73.1	73.1	92%	7e-11 92%
EU048869.1	Cloning vector pGreenII 0229 62-SK, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048868.1	Cloning vector pGreenII 62-SK, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048867.1	Cloning vector pGreenII 0229, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048866.1	Cloning vector pGreenII 0179, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048865.1	Cloning vector pGreenII 0029 62-SK, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048864.1	Cloning vector pGreenII 0029, T-DNA region	73.1	73.1	92%	7e-11 92%
EU048863.1	Cloning vector pGreenII 0800, T-DNA region	73.1	73.1	92%	7e-11 92%

BLAST Basic Local Alignment Search Tool

- Your search parameters were adjusted to search for a short input sequence.

Edit and Resubmit Save Search Strategies Formatting options Download

Nucleotide Sequence (33 letters)

Results for: lcl|6357 None(33bp)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|6357
Description
None
Molecule type
nucleic acid
Query Length
33

Database Name
nr
Description
All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program
BLASTN 2.2.21+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Search Parameters

Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

Database

Posted date	Sep 27, 2009 5:41 PM
Number of letters	28,993,070,215
Number of sequences	9,897,685
Entrez query	none

Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	1.37406	1.37406
K	0.710603	0.710603
H	1.30725	1.30725

Results Statistics

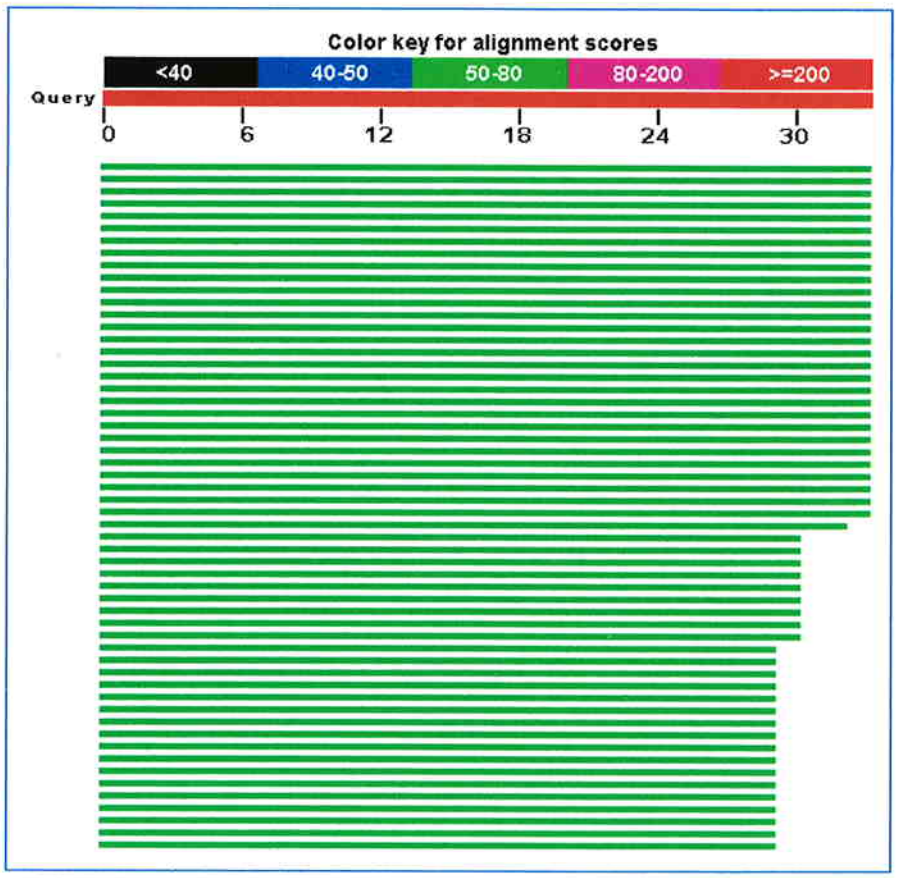
Length adjustment	20
Effective length of query	13
Effective length of database	28795116515
Effective search space	374336514695
Effective search space used	374336514695

Graphic Summary

Distribution of 104 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

AM087268.1	Uncultured ectomycorrhiza (Thelephoraceae) 18S rRNA gene (partial), ITS1, 5.8S rRNA gene, ITS2 and 25S rRNA gene (partial), clone 11-3C	65.9	65.9	100%	5e-09	100%	
AB241267.1	Verasper moseri DNA, microsatellite marker Vemos64	65.9	65.9	100%	5e-09	100%	
AB241259.1	Verasper moseri DNA, microsatellite marker Vemos44	65.9	65.9	100%	5e-09	100%	
AB241258.1	Verasper moseri DNA, microsatellite marker Vemos43	65.9	65.9	100%	5e-09	100%	
AB114435.1	Pseudorasbora parva DNA, microsatelite marker PA19	65.9	65.9	100%	5e-09	100%	
AJ620495.1	Mus musculus partial mRNA for non-selective cation channel TRPV1 (trpv1 gene)	65.9	65.9	100%	5e-09	100%	U G
Y14981.1	Brassica oleracea fae1 gene	65.9	65.9	100%	5e-09	100%	
AF327894.1	Cloning vector pExchange 1, complete sequence	65.9	65.9	100%	5e-09	100%	
AJ567368.2	Vitis vinifera samdc gene for S-adenosylmethionine decarboxylase proenzyme, mRNA	65.9	65.9	100%	5e-09	100%	U G
AF076778.1	Cloning vector pCMVTAG5c, complete sequence	65.9	65.9	100%	5e-09	100%	
AF076777.1	Cloning vector pCMVTAG5b, complete sequence	65.9	65.9	100%	5e-09	100%	
AF076312.1	Cloning vector pCMVTAG5a, complete sequence	65.9	65.9	100%	5e-09	100%	
AF076311.1	Cloning vector pCMVTAG4c, complete sequence	65.9	65.9	100%	5e-09	100%	
AF076310.1	Cloning vector pCMVTAG4b, complete sequence	65.9	65.9	100%	5e-09	100%	
AF073000.1	Cloning vector pCMVTAG4a, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072999.1	Cloning vector pCMVTAG3c, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072998.1	Cloning vector pCMVTAG3b, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072997.1	Cloning vector pCMVTAG3a, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072540.1	Cloning vector pCMVTAG2c, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072539.1	Cloning vector pCMVTAG2b, complete sequence	65.9	65.9	100%	5e-09	100%	
AF072538.1	Cloning vector pCMVTAG2a, complete sequence	65.9	65.9	100%	5e-09	100%	
AJ007320.1	Homo sapiens mRNA for immunoglobulin heavy chain variable region, clone KC4	65.9	65.9	100%	5e-09	100%	U
AF067646.1	Cloning vector pCMV-scriptEX, complete sequence	65.9	65.9	100%	5e-09	100%	
AF028239.1	Mammalian expression vector pCMV-Script, complete sequence	65.9	65.9	100%	5e-09	100%	
Y13161.1	Sus scrofa microsatellite DNA isolate pGC.60	65.9	65.9	100%	5e-09	100%	
U46018.1	Cloning vector pCRSCRIPT Cam, complete sequence	65.9	65.9	100%	5e-09	100%	
U46017.1	Cloning vector pPCR-Script Amp SK(+), complete sequence	65.9	65.9	100%	5e-09	100%	
X91477.1	Uncultured bacterium partial 16S rRNA gene (clone group A54n)	65.9	65.9	100%	5e-09	100%	
X91293.1	Bacterial sp. partial 16S rRNA gene (clone group G41)	65.9	65.9	100%	5e-09	100%	
AJ132233.1	Enterobacter gergoviae glgC (partial) and glgA (partial) genes, strain 57-7	63.9	63.9	96%	2e-08	100%	
AY303671.1	Broad host range expression vector pMHE7Tc, complete sequence	60.0	60.0	90%	3e-07	100%	
AY303669.1	Broad host range expression vector pMHE7, complete sequence	60.0	60.0	90%	3e-07	100%	
AY299693.1	Broad host range expression vector pMHE2, complete sequence	60.0	60.0	90%	3e-07	100%	
AJ277653.1	Escherichia coli plasmid pGBG1	60.0	60.0	90%	3e-07	100%	
AB045011.1	Homo sapiens hmGluR2 gene for metabotropic glutamate receptor type 2, complete cds	60.0	117	90%	3e-07	100%	E
AF179626.1	Expression vector pGP100, complete sequence	60.0	60.0	90%	3e-07	100%	
AF179625.1	Shuttle vector pCS513, complete sequence	60.0	60.0	90%	3e-07	100%	
AF178582.1	Expression vector pCS316 complete sequence	60.0	60.0	90%	3e-07	100%	
AB005666.1	Homo sapiens mRNA for GTPase-activating protein, complete cds	60.0	60.0	90%	3e-07	100%	U E G
AB447355.1	Expression vector pTY2b DNA, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ834451.1	Expression vector pTL80, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ834450.1	Expression vector pTL85, complete sequence	58.0	58.0	87%	1e-06	100%	
AB512772.1	Gene silencing vector pSilent-Dual1 DNA, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ602701.1	Expression vector pFJ1.1-yk1350a08, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ410917.1	Binary vector pNML54, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ380063.1	Suicide vector pMQ200, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ380062.1	Shuttle vector pMQ175, complete sequence	58.0	58.0	87%	1e-06	100%	
FJ386489.1	Cloning vector pEX302, complete sequence	58.0	58.0	87%	1e-06	100%	

AB464951.1	Secale cereale DNA, rye-species specific DNA element, clone: lambda11-2,1.5subclone	58.0	58.0	87%	1e-06	100%
FM200852.1	Recombinase expression vector pCre-ble, complete sequence	58.0	58.0	87%	1e-06	100%
EU606202.1	Neomycin resistance cassette pNeo4 NeoTet (neoTet) gene, complete cds	58.0	92.2	87%	1e-06	100%
AB304876.1	Shuttle vector pRS446TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304875.1	Shuttle vector pRS446PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304874.1	Shuttle vector pRS446GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304873.1	Shuttle vector pRS446ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304872.1	Shuttle vector pRS445TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304871.1	Shuttle vector pRS445PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304870.1	Shuttle vector pRS445GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304869.1	Shuttle vector pRS445ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304868.1	Shuttle vector pRS444TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304867.1	Shuttle vector pRS444PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304866.1	Shuttle vector pRS444GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304865.1	Shuttle vector pRS444ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304864.1	Shuttle vector pRS436TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304863.1	Shuttle vector pRS436PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304862.1	Shuttle vector pRS436GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304861.1	Shuttle vector pRS436ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304860.1	Shuttle vector pRS435TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304859.1	Shuttle vector pRS435PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304858.1	Shuttle vector pRS435GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304857.1	Shuttle vector pRS435ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304856.1	Shuttle vector pRS434TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304855.1	Shuttle vector pRS434PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304854.1	Shuttle vector pRS434GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304853.1	Shuttle vector pRS434ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304852.1	Shuttle vector pRS433TEF DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304851.1	Shuttle vector pRS433PGK DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304850.1	Shuttle vector pRS433GAP DNA, complete sequence	58.0	58.0	87%	1e-06	100%
AB304849.1	Shuttle vector pRS433ADH DNA, complete sequence	58.0	58.0	87%	1e-06	100%
EU546822.1	Cloning vector pMQ125, complete sequence	58.0	58.0	87%	1e-06	100%
EU546821.1	Cloning vector pMQ124, complete sequence	58.0	58.0	87%	1e-06	100%
EU546815.1	Cloning vector pMQ2, complete sequence	58.0	58.0	87%	1e-06	100%
AM940001.1	Cloning vector pIMC	58.0	58.0	87%	1e-06	100%
AM940000.1	Cloning vector pIMK	58.0	58.0	87%	1e-06	100%
EU327977.1	Binary vector pCCRE, complete sequence	58.0	58.0	87%	1e-06	100%
EU327976.1	Binary vector pBCRE, complete sequence	58.0	58.0	87%	1e-06	100%
AM889285.1	Gluconacetobacter diazotrophicus PAl 5 complete genome	58.0	146	87%	1e-06	100%
EU257522.1	Cloning vector pNC-GFP, complete sequence	58.0	58.0	87%	1e-06	100%
EU257520.1	Cloning vector pTOR, complete sequence	58.0	58.0	87%	1e-06	100%
EU257519.1	Cloning vector pSAM, complete sequence	58.0	58.0	87%	1e-06	100%
AM884385.1	Cloning vector pGHGWY	58.0	58.0	87%	1e-06	100%
AM884384.1	Cloning vector pGHGWG	58.0	58.0	87%	1e-06	100%
AM884383.1	Cloning vector pGHGWC	58.0	58.0	87%	1e-06	100%
EU186089.1	Binary vector pCLEAN-G144, complete sequence	58.0	58.0	87%	1e-06	100%
EU186088.1	Binary vector pCLEAN-G146, complete sequence	58.0	58.0	87%	1e-06	100%
EU186087.1	Binary vector pCLEAN-G147, complete sequence	58.0	58.0	87%	1e-06	100%
EU147786.1	Binary vector pGPro2, complete sequence	58.0	58.0	87%	1e-06	100%
AM884388.1	Cloning vector pGBGWY	58.0	58.0	87%	1e-06	100%
AM884387.1	Cloning vector pGBGWG	58.0	58.0	87%	1e-06	100%
AM884386.1	Cloning vector pGBGWC	58.0	58.0	87%	1e-06	100%

```
            |||||||||||||||||||||||||||||||||
Sbjct  715  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>gb|AF073000.1|AF073000 Cloning vector pCMVTAG4a, complete sequence
Length=4319

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  715  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>gb|AF072999.1|AF072999 Cloning vector pCMVTAG3c, complete sequence
Length=4328

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  747  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  715
```

>gb|AF072998.1|AF072998 Cloning vector pCMVTAG3b, complete sequence
Length=4327

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  746  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  714
```

>gb|AF072997.1|AF072997 Cloning vector pCMVTAG3a, complete sequence
Length=4326

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  745  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  713
```

>gb|AF072540.1|AF072540 Cloning vector pCMVTAG2c, complete sequence
Length=4325

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  744  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  712
```

>gb|AF072539.1|AF072539 Cloning vector pCMVTAG2b, complete sequence
Length=4324

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  743  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  711
```

>gb|AF072538.1|AF072538 Cloning vector pCMVTAG2a, complete sequence
Length=4323

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  742  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  710
```

>emb|AJ007320.1| Homo sapiens mRNA for immunoglobulin heavy chain variable region, clone KC4
Length=308

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
```

```
            |||||||||||||||||||||||||||||||||
Sbjct  261  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  229
```

>gb|AF067646.1|AF067646 Cloning vector pCMV-scriptEX, complete sequence
Length=4307

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  715  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>gb|AF028239.1|AF028239 Mammalian expression vector pCMV-Script, complete sequence
Length=4278

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Minus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  715  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>emb|Y13161.1| Sus scrofa microsatellite DNA isolate pGC.60
Length=1350

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Plus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  46   ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  78
```

>gb|U46018.1|CVU46018 Cloning vector pCRSCRIPT Cam, complete sequence
Length=3399

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Plus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  696  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  728
```

>gb|U46017.1|CVU46017 Cloning vector pPCR-Script Amp SK(+), complete sequence
Length=2961

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Plus

Query  1    ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
            |||||||||||||||||||||||||||||||||
Sbjct  696  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  728
```

>emb|X91477.1| Uncultured bacterium partial 16S rRNA gene (clone group A54n)
Length=338

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Plus

Query  1   ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
           |||||||||||||||||||||||||||||||||
Sbjct  65  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  97
```

>emb|X91293.1| Bacterial sp. partial 16S rRNA gene (clone group G41)
Length=322

```
 Score = 65.9 bits (33),  Expect = 5e-09
 Identities = 33/33 (100%), Gaps = 0/33 (0%)
 Strand=Plus/Plus

Query  1   ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  33
           |||||||||||||||||||||||||||||||||
Sbjct  34  ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  66
```

>emb|AJ132233.1| Enterobacter gergoviae glgC (partial) and glgA (partial) genes, strain 57-7
Length=1536

```
 Score = 63.9 bits (32),  Expect = 2e-08
 Identities = 32/32 (100%), Gaps = 0/32 (0%)
 Strand=Plus/Plus

Query  1   ATATCGAATTCCTGCAGCCCGGGGGATCCGCC  32
           ||||||||||||||||||||||||||||||||
```

```
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Minus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  764  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  710
```

>emb|AJ007320.1| Homo sapiens mRNA for immunoglobulin heavy chain variable region, clone KC4
Length=308

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Minus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  283  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  229
```

>gb|AF067646.1|AF067646 Cloning vector pCMV-scriptEX, complete sequence
Length=4307

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Minus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  737  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>gb|AF028239.1|AF028239 Mammalian expression vector pCMV-Script, complete sequence
Length=4278

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Minus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  737  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  683
```

>emb|Y13161.1| Sus scrofa microsatellite DNA isolate pGC.60
Length=1350

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Plus

Query  1   GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
           |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  24  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  78
```

>gb|U46018.1|CVU46018 Cloning vector pCRSCRIPT Cam, complete sequence
Length=3399

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Plus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  674  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  728
```

>gb|U46017.1|CVU46017 Cloning vector pPCR-Script Amp SK(+), complete sequence
Length=2961

```
 Score = 80.5 bits (43),  Expect = 4e-13
 Identities = 53/57 (92%), Gaps = 4/57 (7%)
 Strand=Plus/Plus

Query  1    GTCGACGGTA-CGAT-AGTCTTGCATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  55
            |||||||||| |||| || |||| ||||||||||||||||||||||||||||||||
Sbjct  674  GTCGACGGTATCGATAAG-CTTG-ATATCGAATTCCTGCAGCCCGGGGGATCCGCCC  728
```

>emb|X91477.1| Uncultured bacterium partial 16S rRNA gene (clone group A54n)
Length=338

```
LOCUS       AJ007320                308 bp    mRNA    linear
PRI 22-JUN-1998
DEFINITION  Homo sapiens mRNA for immunoglobulin heavy chain
variable region,
            clone KC4.
ACCESSION   AJ007320
VERSION     AJ007320.1  GI:3250843
KEYWORDS    heavy chain; immunoglobulin; variable region; VH1
gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates;
Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Souto-Carneiro,M.M. and Krenn,V.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 308)
  AUTHORS   Souto-Carneiro,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) Souto-Carneiro M.M.,
Pathologisches
            Institut, Wuerzburg University, Josef-Schneider-Str.
2, D- 97080
            Wuerzburg, GERMANY
FEATURES             Location/Qualifiers
     source          1..308
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /isolate="donor 194"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="KC4"
                     /sex="female"
                     /cell_type="B-lymphocyte"
                     /tissue_type="lowerleg 'sheat"
                     /rearranged
                     /note="rheumatoid arthritis patient"
     gene            <1..>308
                     /gene="VH1"
     CDS             <1..>237
                     /gene="VH1"
                     /codon_start=1
                     /product="immunoglobulin heavy chain
variable region"
                     /protein_id="CAA07460.1"
                     /db_xref="GI:3250844"
```

/translation="VSCKASGYTFTGYFIHWVRQAPGQGLEWMGWISANNGNTHYAHK
LQGRVTMTTDTSTRTAYMDLRSLRSDDTAIYYGRI"
V region 1..237
/gene="VH1"
/product="immunoglobulin heavy chain
variable region"
ORIGIN

1 gtctcctgca aggcttctgg atacaccttc accggctact ttatacactg
ggtgcgacag

61 gcccctggac aagggcttga gtggatggga tggatcagtg ctaacaatgg
aaatacacac

121 tatgcacata aactccaggg cagagtcacc atgaccacag acacatccac
gaggacagcc

181 tacatggacc tgaggagcct gagatctgac gacacggcca tatattatgg
gcggatcccc

241 cgggctgcag gaattcgata tcaagcttat cgataccgtc gacctcgagg
gggggcccgg

301 tacccaat

Notes:

Corresponds to pPCR Script Amp SK(+) with 100% homology

Corresponds to T45.3 sequence bp 23-55 with 100% homology

BLAST Basic Local Alignment Search Tool

- Your search parameters were adjusted to search for a short input sequence.

Edit and Resubmit Save Search Strategies Formatting options Download

Nucleotide Sequence (22 letters)

Results for: lcl|40741 None(22bp)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|40741
Description
None
Molecule type
nucleic acid
Query Length
22

Database Name
nr
Description
All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program
BLASTN 2.2.21+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Search Parameters

Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

Database

Posted date	Sep 27, 2009 5:41 PM
Number of letters	28,993,070,215
Number of sequences	9,897,685
Entrez query	none

Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	1.37406	1.37406
K	0.710603	0.710603
H	1.30725	1.30725

Results Statistics

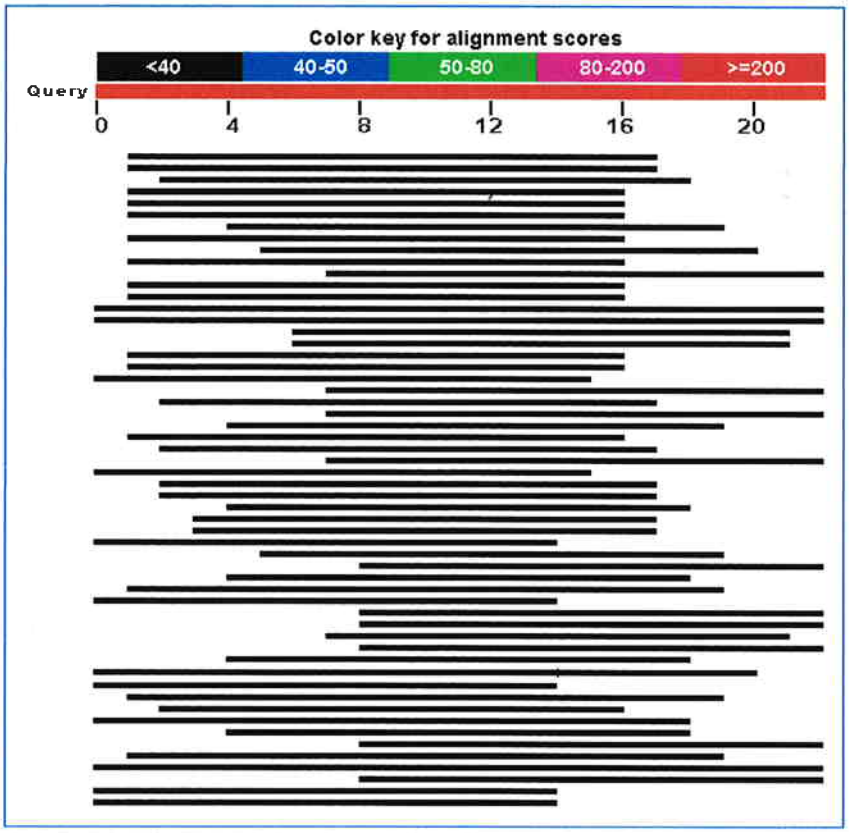
Length adjustment	19
Effective length of query	3
Effective length of database	28805014200
Effective search space	86415042600
Effective search space used	86415042600

Graphic Summary

Distribution of 103 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

AB231255.1	Melanotus koikei gene for 28S ribosomal RNA, partial sequence	32.2	32.2	72%	17	100%	
AB231253.1	Melanotus cete cete gene for 28S ribosomal RNA, partial sequence	32.2	32.2	72%	17	100%	
CP000148.1	Geobacter metallireducens GS-15, complete genome	32.2	32.2	72%	17	100%	
CP001715.1	Candidatus Accumulibacter phosphatis clade IIA str. UW-1, complete genome	30.2	30.2	68%	69	100%	
XM_002426526.1	Pediculus humanus corporis eukaryotic translation initiation factor 4B, putative, mRNA	30.2	30.2	68%	69	100%	G
CP001638.1	Geobacillus sp. WCH70, complete genome	30.2	30.2	68%	69	100%	
NG_005634.3	Homo sapiens E2F transcription factor 3 pseudogene 1 (E2F3P1) on chromosome 17	30.2	30.2	68%	69	100%	G
AP008955.1	Brevibacillus brevis NBRC 100599 DNA, complete genome	30.2	30.2	68%	69	100%	
AE009439.1	Methanopyrus kandleri AV19, complete genome	30.2	30.2	68%	69	100%	
CP001364.1	Chloroflexus sp. Y-400-fl, complete genome	30.2	30.2	68%	69	100%	
FJ211062.1	Neospora caninum 24B (24B) gene, complete cds	30.2	30.2	68%	69	100%	
CP000909.1	Chloroflexus aurantiacus J-10-fl, complete genome	30.2	30.2	68%	69	100%	
CP000680.1	Pseudomonas mendocina ymp, complete genome	30.2	30.2	68%	69	100%	
XM_001416344.1	Ostreococcus lucimarinus CCE9901 gamma tubulin (TubG) mRNA, complete cds	30.2	30.2	100%	69	95%	G
CP000582.1	Ostreococcus lucimarinus CCE9901 chromosome 2, complete sequence	30.2	30.2	100%	69	95%	
XM_001363270.1	PREDICTED: Monodelphis domestica similar to Cas-Br-M (murine) ecotropic retroviral transforming sequence b, transcript variant 2 (LOC100011932), mRNA	30.2	30.2	68%	69	100%	G
XM_001363186.1	PREDICTED: Monodelphis domestica similar to Cas-Br-M (murine) ecotropic retroviral transforming sequence b, transcript variant 1 (LOC100011932), mRNA	30.2	30.2	68%	69	100%	G
AB231257.1	Mucromorphus miwai miwai gene for 28S ribosomal RNA, partial sequence	30.2	30.2	68%	69	100%	
CP000142.2	Pelobacter carbinolicus DSM 2380, complete genome	30.2	30.2	68%	69	100%	
AB223388.1	Aspergillus oryzae cDNA, contig sequence: AoEST0220	30.2	30.2	68%	69	100%	
XM_745154.1	Aspergillus fumigatus Af293 RNA polymerase III transcription factor TFIIIC subunit (Tfc4) (AFUA_1G04980), partial mRNA	30.2	30.2	68%	69	100%	G
AY864330.1	Chrysodeixis chalcites nucleopolyhedrovirus, complete genome	30.2	30.2	68%	69	100%	
AY651069.1	Yarrowia lipolytica fructose-1,6-bisphosphatase (FBP1) gene, promoter region and 5' UTR	30.2	30.2	68%	69	100%	
AC015911.8	Homo sapiens chromosome 17, clone RP11-1094M14, complete sequence	30.2	30.2	68%	69	100%	
AP006627.1	Bacillus clausii KSM-K16 DNA, complete genome	30.2	30.2	68%	69	100%	
AY666015.1	Grouper iridovirus, complete genome	30.2	30.2	68%	69	100%	
CR382127.1	Yarrowia lipolytica strain CLIB122 chromosome A complete sequence	30.2	30.2	68%	69	100%	
AL117559.1	Homo sapiens mRNA; cDNA DKFZp566O1624 (from clone DKFZp566O1624)	30.2	30.2	68%	69	100%	U E
M29149.1	N.crassa DNA (clone pEC25) before exposure to (RIP) repeat-induced point mutation process	30.2	30.2	68%	69	100%	
M29151.1	N.crassa DNA (clone pBJ20) after exposure to (RIP) repeat-induced point mutation process	30.2	30.2	68%	69	100%	
BN001301.1	TPA: TPA_reasm: Aspergillus nidulans FGSC A4 chromosome I	28.2	28.2	63%	271	100%	
XM_002582277.1	Uncinocarpus reesii 1704 conserved hypothetical protein, mRNA	28.2	28.2	63%	271	100%	G
CP001712.1	Robiginitalea biformata HTCC2501, complete genome	28.2	28.2	63%	271	100%	
CP001688.1	Halomicrobium mukohataei DSM 12286, complete genome	28.2	28.2	63%	271	100%	
GQ446419.1	Uncultured prokaryote clone DeadSeal_IOBCFE001_18-A05-Fwd genomic sequence	28.2	28.2	63%	271	100%	
NW_003038009.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000234	28.2	28.2	63%	271	100%	
NW_003020078.1	Penicillium chrysogenum Wisconsin 54-1255 complete genome, contig Pc00c18	28.2	28.2	63%	271	100%	
XM_002501109.1	Micromonas sp. RCC299 chromosome 3 hypothetical protein (MICPUN_57383) mRNA, complete cds	28.2	28.2	81%	271	94%	G
CP001626.1	Rhizobium leguminosarum bv. trifolii WSM1325 plasmid pRl32504, complete sequence	28.2	28.2	63%	271	100%	
FN357716.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000425	28.2	28.2	63%	271	100%	
FN357525.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000234	28.2	28.2	63%	271	100%	
FN357361.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000070	28.2	28.2	63%	271	100%	
FN357336.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000045	28.2	28.2	63%	271	100%	
FN357293.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000002	28.2	28.2	63%	271	100%	

CP001227.1	Rickettsia peacockii str. Rustic, complete genome	28.2	56.5	90%	271	100%	
CP001612.1	Rickettsia africae ESF-5, complete genome	28.2	28.2	63%	271	100%	
CP001324.1	Micromonas sp. RCC299 chromosome 3, complete sequence	28.2	28.2	81%	271	94%	
CP001323.1	Micromonas sp. RCC299 chromosome 2, complete sequence	28.2	28.2	63%	271	100%	
AP009153.1	Gemmatimonas aurantiaca T-27 DNA, complete genome	28.2	56.5	81%	271	100%	
CU928129.15	Zebrafish DNA sequence from clone CH73-39O22 in linkage group 5, complete sequence	28.2	28.2	63%	271	100%	
AC235194.1	Glycine max strain Williams 82 clone GM_WBb0013G10, complete sequence	28.2	28.2	63%	271	100%	
AP009380.1	Porphyromonas gingivalis ATCC 33277 DNA, complete genome	28.2	56.5	81%	271	100%	
CP001053.1	Burkholderia phytofirmans PsJN chromosome 2, complete sequence	28.2	28.2	100%	271	90%	
AY536043.3	Lyngbya majuscula CCAP 1446/4 putative pyruvate flavodoxin dehydrogenase gene, partial cds; bidirectional hydrogenase structural gene cluster, complete sequence; hypothetical protein, XisH (xisH), XisI (xisI), hypothetical proteins, HoxW (hoxW), and hypothetical proteins genes, complete cds; and hypothetical protein gene, partial cds	28.2	28.2	63%	271	100%	
CP001029.1	Methylobacterium populi BJ001, complete genome	28.2	28.2	63%	271	100%	
CP001032.1	Opitutus terrae PB90-1, complete genome	28.2	28.2	63%	271	100%	
XM_001804828.1	Phaeosphaeria nodorum SN15 hypothetical protein partial mRNA	28.2	28.2	63%	271	100%	G
XM_001801684.1	Phaeosphaeria nodorum SN15 hypothetical protein partial mRNA	28.2	28.2	63%	271	100%	G
XM_001793893.1	Phaeosphaeria nodorum SN15 hypothetical protein partial mRNA	28.2	28.2	63%	271	100%	G
CU678735.1	Synthetic construct Homo sapiens gateway clone IMAGE:100020404 3' read CLINT1 mRNA	28.2	28.2	63%	271	100%	
XM_001758137.1	Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_161231) mRNA, complete cds	28.2	28.2	63%	271	100%	U G
CP000937.1	Francisella philomiragia subsp. philomiragia ATCC 25017, complete genome	28.2	28.2	63%	271	100%	
CP000766.1	Rickettsia rickettsii str. Iowa, complete genome	28.2	28.2	63%	271	100%	
AM889285.1	Gluconacetobacter diazotrophicus PAl 5 complete genome	28.2	28.2	63%	271	100%	
EU145676.1	Stethorus sp. JAR-2007 28S ribosomal RNA gene, partial sequence	28.2	28.2	63%	271	100%	
AM778940.1	Microcystis aeruginosa PCC 7806 genome sequencing data, contig C310	28.2	28.2	63%	271	100%	
CP000683.1	Rickettsia massiliae MTU5, complete genome	28.2	28.2	63%	271	100%	
CP000848.1	Rickettsia rickettsii str. 'Sheila Smith', complete genome	28.2	28.2	63%	271	100%	
CP000847.1	Rickettsia akari str. Hartford, complete genome	28.2	28.2	63%	271	100%	
CR927050.34	Zebrafish DNA sequence from clone DKEY-217G21, complete sequence	28.2	28.2	63%	271	100%	
XM_001649113.1	Aedes aegypti inositol 1,4,5-trisphosphate receptor partial mRNA	28.2	28.2	63%	271	100%	U G
CP000789.1	Vibrio harveyi ATCC BAA-1116 chromosome I, complete sequence	28.2	28.2	63%	271	100%	
XM_001640807.1	Nematostella vectensis predicted protein (NEMVEDRAFT_v1g159083) mRNA, complete cds	28.2	28.2	63%	271	100%	U G
XM_001608633.1	Babesia bovis hypothetical protein (BBOV_I000210) mRNA, complete cds	28.2	28.2	63%	271	100%	G
XM_001549280.1	Botryotinia fuckeliana B05.10 predicted protein (BC1G_11879) partial mRNA	28.2	28.2	63%	271	100%	G
CP000764.1	Bacillus cereus subsp. cytotoxis NVH 391-98, complete genome	28.2	28.2	63%	271	100%	
CP000740.1	Sinorhizobium medicae WSM419 plasmid pSMED02, complete genome	28.2	28.2	63%	271	100%	
CP000738.1	Sinorhizobium medicae WSM419, complete genome	28.2	28.2	63%	271	100%	
CP000698.1	Geobacter uraniireducens Rf4, complete genome	28.2	28.2	63%	271	100%	
XM_001402323.1	Aspergillus niger CBS 513.88 hypothetical protein (An04g10040) partial mRNA	28.2	28.2	63%	271	100%	G
XM_363059.2	Magnaporthe grisea 70-15 hypothetical protein (MGG_08643) partial mRNA	28.2	28.2	63%	271	100%	G
CP000656.1	Mycobacterium gilvum PYR-GCK, complete genome	28.2	28.2	63%	271	100%	
CP000655.1	Polynucleobacter necessarius subsp. asymbioticus QLW-P1DMWA-1, complete genome	28.2	28.2	63%	271	100%	
CP000586.1	Ostreococcus lucimarinus CCE9901 chromosome 6, complete sequence	28.2	28.2	63%	271	100%	
XM_001276096.1	Aspergillus clavatus NRRL 1 aminopeptidase, putative (ACLA_000820), partial mRNA	28.2	28.2	63%	271	100%	G
XM_001269661.1	Aspergillus clavatus NRRL 1 RNA polymerase III transcription factor TFIIIC subunit (Tfc4), putative (ACLA_029660), partial mRNA	28.2	28.2	63%	271	100%	G
CU207366.1	Gramella forsetii KT0803 complete circular genome	28.2	28.2	63%	271	100%	
CP000423.1	Lactobacillus casei ATCC 334, complete genome	28.2	28.2	63%	271	100%	
XM_001195560.1	PREDICTED: Strongylocentrotus purpuratus similar to type-1 corticotropin-releasing hormone receptor alpha isoform (LOC585532), mRNA	28.2	28.2	63%	271	100%	U G
XM_001176861.1	PREDICTED: Strongylocentrotus purpuratus similar to p34cdc2,	28.2	28.2	63%	271	100%	U G

	transcript variant 1 (LOC575965), mRNA						
XM_001177005.1	PREDICTED: Strongylocentrotus purpuratus similar to p34cdc2, transcript variant 2 (LOC575965), mRNA	28.2	28.2	63%	271	100%	UG
XM_785357.2	PREDICTED: Strongylocentrotus purpuratus similar to type-1 corticotropin-releasing hormone receptor alpha isoform (LOC585532), mRNA	28.2	28.2	63%	271	100%	UG
XM_776322.2	PREDICTED: Strongylocentrotus purpuratus similar to p34cdc2 (LOC575965), mRNA	28.2	28.2	63%	271	100%	UG
NM_067458.2	Caenorhabditis elegans SCP-Like extracellular protein family member (scl-23) (scl-23) mRNA, complete cds	28.2	28.2	63%	271	100%	UEG
NM_001054513.1	Oryza sativa (japonica cultivar-group) Os02g0724500 (Os02g0724500) mRNA, complete cds	28.2	28.2	63%	271	100%	UG
XM_001209553.1	Aspergillus terreus NIH2624 conserved hypothetical protein (ATEG_06867) partial mRNA	28.2	28.2	63%	271	100%	G
DQ889502.1	Human herpesvirus 1 strain HF clone 10, partial sequence	28.2	28.2	63%	271	100%	
BX284684.8	Zebrafish DNA sequence from clone DKEY-14I10 in linkage group 8 Contains part of the plxna3 gene for plexin A3 and two CpG islands, complete sequence	28.2	28.2	81%	271	94%	
CP000352.1	Ralstonia metallidurans CH34, complete genome	28.2	28.2	63%	271	100%	
AP007171.1	Aspergillus oryzae RIB40 DNA, SC011	28.2	28.2	63%	271	100%	

T45.3 bp 56-74

BLAST Basic Local Alignment Search Tool

- Your search parameters were adjusted to search for a short input sequence.

Edit and Resubmit Save Search Strategies Formatting options Download

Nucleotide Sequence (19 letters)

Results for: lcl|20163 None(19bp)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|20163
Description
None
Molecule type
nucleic acid
Query Length
19

Database Name
nr
Description
All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program
BLASTN 2.2.21+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Search Parameters

Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

Database

Posted date	Sep 27, 2009 5:41 PM
Number of letters	28,993,070,215
Number of sequences	9,897,685
Entrez query	none

Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	1.37406	1.37406
K	0.710603	0.710603
H	1.30725	1.30725

Results Statistics

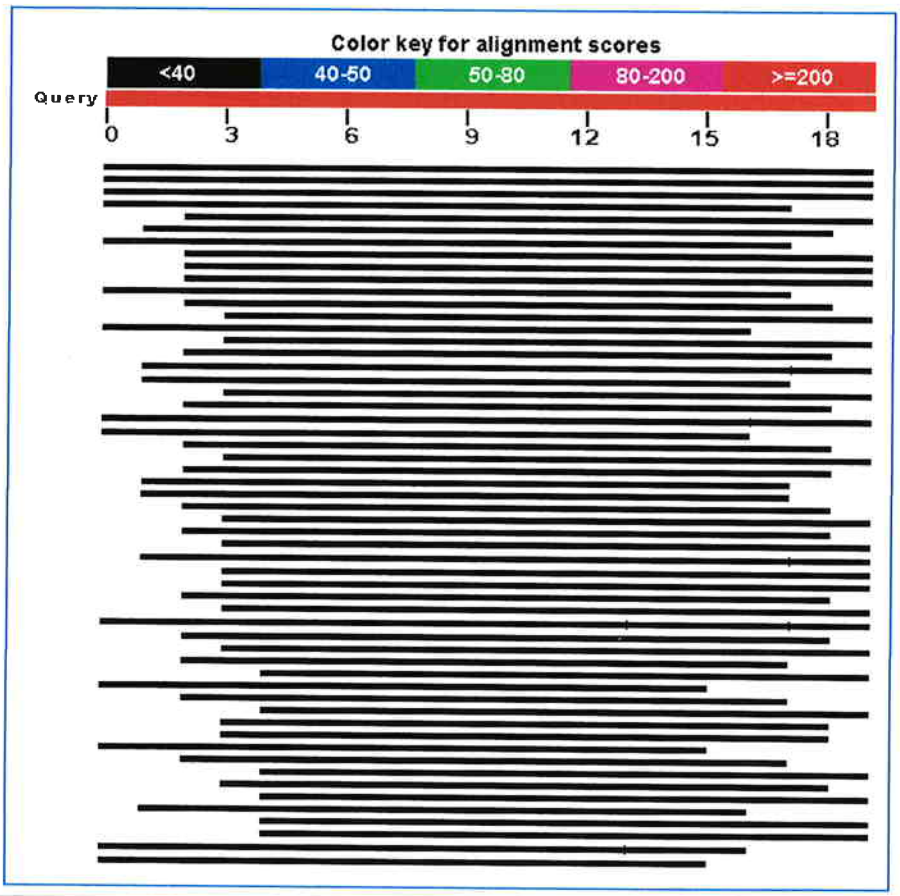
Length adjustment	17
Effective length of query	2
Effective length of database	28824809570
Effective search space	57649619140
Effective search space used	57649619140

Graphic Summary

Distribution of 142 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

DQ464177.1	Hepatitis B virus isolate LT6 clone 4 HBcAg protein and HBsAg protein (S) genes, complete cds	38.2	66.4	100%	0.19	100%	
DQ464176.1	Hepatitis B virus isolate LT6 clone 2 HBcAg protein and HBsAg protein (S) genes, complete cds	38.2	38.2	100%	0.19	100%	
DQ336691.1	Hepatitis B virus isolate LT5 clone 7 HBcAg and HBsAg (S) genes, complete cds	38.2	38.2	100%	0.19	100%	
CU633901.1	Podospora anserina genomic DNA chromosome 1, supercontig 2	34.2	34.2	89%	2.9	100%	
XM_001767234.1	Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_230074) mRNA, complete cds	34.2	34.2	89%	2.9	100%	G
XM_001752228.1	Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_174045) mRNA, complete cds	34.2	34.2	89%	2.9	100%	U G
XM_001704768.1	Giardia lamblia ATCC 50803 Elongation factor 2 (GL50803_17570) mRNA, complete cds	34.2	34.2	89%	2.9	100%	G
CT027701.11	Zebrafish DNA sequence from clone CH211-209D1 in linkage group 16, complete sequence	34.2	34.2	89%	2.9	100%	
AC126538.1	Homo sapiens chromosome 16 clone CTD-3225F13, complete sequence	34.2	34.2	89%	2.9	100%	
AC007014.1	Homo sapiens chromosome 16 clone 66H6, complete sequence	34.2	34.2	89%	2.9	100%	
D29835.1	Giardia intestinalis mRNA for elongation factor 2, partial cds	34.2	34.2	89%	2.9	100%	
BN001305.1	TPA: TPA_reasm: Aspergillus nidulans FGSC A4 chromosome V	32.2	60.5	84%	12	100%	
XM_002502501.1	Micromonas sp. RCC299 chromosome 5 predicted protein (MICPUN_50300) mRNA, complete cds	32.2	32.2	84%	12	100%	G
CP001649.1	Desulfovibrio salexigens DSM 2638, complete genome	32.2	86.7	84%	12	100%	
CP001326.1	Micromonas sp. RCC299 chromosome 5, complete sequence	32.2	32.2	84%	12	100%	
CU856378.10	Zebrafish DNA sequence from clone CH73-44P7 in linkage group 7, complete sequence	32.2	32.2	84%	12	100%	
CP001336.1	Desulfitobacterium hafniense DCB-2, complete genome	32.2	137	94%	12	100%	
XM_002178919.1	Phaeodactylum tricornutum CCAP 1055/1 predicted protein, mRNA	32.2	32.2	84%	12	100%	U G
AC189384.2	Brassica rapa subsp. pekinensis clone KBrB052E10, complete sequence	32.2	32.2	84%	12	100%	
AC199142.9	Canis familiaris, clone XX-240A15, complete sequence	32.2	32.2	84%	12	100%	
AL954747.1	Nitrosomonas europaea ATCC 19718, complete genome	32.2	58.5	100%	12	100%	
XM_001272348.1	Aspergillus clavatus NRRL 1 conserved hypothetical protein (ACLA_065570), partial mRNA	32.2	32.2	84%	12	100%	G
BT026825.1	Gasterosteus aculeatus clone CFW60-H01 mRNA sequence	32.2	32.2	84%	12	100%	U
CR937049.4	CH230-323M7, complete sequence	32.2	32.2	84%	12	100%	
NM_126411.1	Arabidopsis thaliana F-box family protein (AT2G03610) mRNA, complete cds	32.2	32.2	84%	12	100%	U E G
AP008230.1	Desulfitobacterium hafniense Y51 DNA, complete genome	32.2	84.7	84%	12	100%	
XM_384526.1	Gibberella zeae PH-1 hypothetical protein partial mRNA	32.2	32.2	84%	12	100%	G
AC006836.7	Arabidopsis thaliana chromosome 2 clone F19B11 map RNSI, complete sequence	32.2	62.4	84%	12	100%	E
AL591038.9	Human DNA sequence from clone RP11-120E1 on chromosome 9 Contains the 5' end of gene MGC23427 and two CpG islands, complete sequence	32.2	32.2	84%	12	100%	
AL117205.2	Caenorhabditis elegans YAC Y116A8A, complete sequence	32.2	32.2	84%	12	100%	
CR513789.22	Zebrafish DNA sequence from clone CH211-212J20 in linkage group 22, complete sequence	32.2	32.2	84%	12	100%	
BA000001.2	Pyrococcus horikoshii OT3 DNA, complete genome	32.2	60.5	94%	12	100%	
AL954148.15	Zebrafish DNA sequence from clone CH211-212L17 in linkage group 22 Contains the nr5a2 gene for nuclear receptor subfamily 5 group A member 2 and four CpG islands, complete sequence	32.2	32.2	84%	12	100%	
AP006708.1	Lotus japonicus genomic DNA, chromosome 3, clone: LjT09I23, TM0590c, complete sequence	32.2	32.2	84%	12	100%	
AP003800.3	Oryza sativa Japonica Group genomic DNA, chromosome 7, BAC clone:OJ1014_E09	32.2	32.2	84%	12	100%	
AF250225.1	Helicobacter pylori IceA2 (iceA2) and CATG-specific methyltransferase (hpyIM) genes, complete cds	32.2	32.2	84%	12	100%	
AE000782.1		32.2	111	100%	12	100%	

	Archaeoglobus fulgidus DSM 4304, complete genome						
AB018435.1	Amphibacillus xylanus gene for alkyl hydroperoxide reductase C, NADH oxidase, complete cds	32.2	32.2	84%	12	100%	
AF020707.1	Helicobacter pylori methylase HpyI (hpyIM) gene, complete cds	32.2	32.2	84%	12	100%	
BN001307.1	TPA: TPA_reasm: Aspergillus nidulans FGSC A4 chromosome VII	30.2	30.2	78%	46	100%	
FN393075.1	Saccharomyces cerevisiae EC1118 chromosome X, EC1118_1J11 genomic scaffold, whole genome shotgun sequence	30.2	30.2	78%	46	100%	
FP236842.1	Erwinia pyrifoliae strain Ep1/96 complete chromosome	30.2	30.2	78%	46	100%	
CP001629.1	Desulfomicrobium baculatum DSM 4028, complete genome	30.2	30.2	78%	46	100%	
BT097479.1	Soybean clone JCVI-FLGm-22G13 unknown mRNA	30.2	30.2	78%	46	100%	U
XM_002524312.1	Ricinus communis type I inositol polyphosphate 5-phosphatase, putative, mRNA	30.2	30.2	78%	46	100%	G
XM_002518470.1	Ricinus communis beta-glucosidase, putative, mRNA	30.2	30.2	78%	46	100%	G
XM_002499864.1	Micromonas sp. RCC299 chromosome 2 SNF2 super family (MICPUN_56440) mRNA, complete cds	30.2	30.2	78%	46	100%	G
CP001431.1	Neorickettsia risticii str. Illinois, complete genome	30.2	30.2	78%	46	100%	
XM_002422280.1	Candida dubliniensis CD36 patatin-like serine hydrolase, putative (CD36_33890) mRNA, complete cds	30.2	30.2	78%	46	100%	G
NM_126406.5	Arabidopsis thaliana ubiquitin-protein ligase (AT2G03560) mRNA, complete cds	30.2	30.2	78%	46	100%	G
CP001614.2	Teredinibacter turnerae T7901, complete genome	30.2	56.5	78%	46	100%	
GQ220325.1	Vitis vinifera strain PN40024 mitochondrion, partial genome	30.2	30.2	78%	46	100%	
AC213131.2	Oryza glaberrima clone OG_BBa0031E23, complete sequence	30.2	30.2	78%	46	100%	
AC117096.5	Rattus norvegicus NOVECTOR CH230-268L9 (Children's Hospital Oakland Research Institute Rat (BN/SsNHsd/MCW) BAC library) complete sequence	30.2	30.2	78%	46	100%	
CP001389.1	Rhizobium sp. NGR234, complete genome	30.2	82.8	84%	46	100%	
CP001323.1	Micromonas sp. RCC299 chromosome 2, complete sequence	30.2	30.2	78%	46	100%	
CP001114.1	Deinococcus deserti VCD115, complete genome	30.2	30.2	78%	46	100%	
BX548175.1	Prochlorococcus marinus MIT9313 complete genome	30.2	56.5	89%	46	100%	
FM204883.1	Streptococcus equi subsp. equi 4047, complete genome	30.2	56.5	84%	46	100%	
FJ708746.1	Arabidopsis thaliana leucine-rich repeat receptor-like protein kinase (LRR-RLK) mRNA, partial cds	30.2	30.2	78%	46	100%	U G
CP000857.1	Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594, complete genome	30.2	58.5	84%	46	100%	
AE008691.1	Thermoanaerobacter tengcongensis MB4, complete genome	30.2	30.2	78%	46	100%	
AE008923.1	Xanthomonas axonopodis pv. citri str. 306, complete genome	30.2	82.8	100%	46	100%	
XM_002331779.1	Populus trichocarpa predicted protein, mRNA	30.2	30.2	78%	46	100%	G
XM_002317422.1	Populus trichocarpa predicted protein, mRNA	30.2	30.2	78%	46	100%	G
CP001087.1	Desulfobacterium autotrophicum HRM2, complete genome	30.2	82.8	89%	46	100%	
FM992695.1	Candida dubliniensis CD36 chromosome R, complete sequence	30.2	30.2	78%	46	100%	
BC151172.1	Mus musculus myb-like, SWIRM and MPN domains 1, mRNA (cDNA clone MGC:184072 IMAGE:9088061), complete cds	30.2	30.2	78%	46	100%	U G
BC150946.1	Mus musculus myb-like, SWIRM and MPN domains 1, mRNA (cDNA clone MGC:183857 IMAGE:9087857), complete cds	30.2	30.2	78%	46	100%	U G
CP001390.1	Geobacter sp. FRC-32, complete genome	30.2	112	78%	46	100%	
CP001358.1	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774, complete genome	30.2	86.7	84%	46	100%	
XM_002174310.1	Schizosaccharomyces japonicus yFS275 nucleoporin Nup186, mRNA	30.2	30.2	78%	46	100%	G
CP000855.1	Thermococcus onnurineus NA1, complete genome	30.2	112	89%	46	100%	
AE009949.1	Streptococcus pyogenes MGAS8232, complete genome	30.2	30.2	78%	46	100%	
FM179380.1	Vitis vinifera complete mitochondrial genome, cultivar Pinot noir clone ENTAV115	30.2	30.2	78%	46	100%	
CP000829.1	Streptococcus pyogenes NZ131, complete genome	30.2	30.2	78%	46	100%	
CU855892.8	Zebrafish DNA sequence from clone CH73-131O13 in linkage group 14, complete sequence	30.2	30.2	78%	46	100%	
FJ147546.1	Uncultured bacterium clone RI41 membrane-bound nitrate reductase (narG) gene, partial cds	30.2	30.2	78%	46	100%	
FJ147542.1	Uncultured bacterium clone RI37 membrane-bound nitrate reductase (narG) gene, partial cds	30.2	30.2	78%	46	100%	

AC232527.1	Brassica rapa subsp. pekinensis clone KBrB084N21, complete sequence	30.2	30.2	78%	46	100%	
XM_002093176.1	Drosophila yakuba GE20905 (Dyak\GE20905), mRNA	30.2	30.2	78%	46	100%	G
XM_002086404.1	Drosophila yakuba GE22862 (Dyak\GE22862), mRNA	30.2	30.2	78%	46	100%	G
CP001100.1	Chloroherpeton thalassium ATCC 35110, complete genome	30.2	56.5	78%	46	100%	
AC229605.1	Brassica rapa subsp. pekinensis clone KBrH059N21, complete sequence	30.2	60.5	78%	46	100%	
AP010486.1	Lotus japonicus genomic DNA, chromosome 3, clone: LjT05N10, TM2057, complete sequence	30.2	30.2	78%	46	100%	
AP010485.1	Lotus japonicus genomic DNA, chromosome 3, clone: LjB25M18, BM2091, complete sequence	30.2	30.2	78%	46	100%	
CP000967.1	Xanthomonas oryzae pv. oryzae PXO99A, complete genome	30.2	82.8	100%	46	100%	
BX950229.1	Methanococcus maripaludis strain S2, complete sequence	30.2	30.2	78%	46	100%	
NM_103869.3	Arabidopsis thaliana ATMTK (ARABIDOPSIS THALIANA S-METHYL-5-THIORIBOSE KINASE); S-methyl-5-thioribose kinase (ATMTK) mRNA, complete cds	30.2	30.2	78%	46	100%	U E G
NG_007481.1	Homo sapiens complement component (3b/4b) receptor 1 (Knops blood group) (CR1) on chromosome 1	30.2	30.2	78%	46	100%	
XM_001861253.1	Culex quinquefasciatus conserved hypothetical protein, mRNA	30.2	30.2	78%	46	100%	G
CP000961.1	Shewanella woodyi ATCC 51908, complete genome	30.2	82.8	89%	46	100%	
AM920689.1	Xanthomonas campestris pv. campestris complete genome, strain B100	30.2	82.8	89%	46	100%	
XM_001749905.1	Monosiga brevicollis MX1 predicted protein MONBRDRAFT_38958 mRNA, complete cds	30.2	30.2	78%	46	100%	G
CU367257.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	
CU358650.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	
CU364510.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	
CU353601.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	
CU362623.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	
CU362562.1	Aphanomyces euteiches cDNA	30.2	30.2	78%	46	100%	

Alignments Select All Get selected sequences Distance tree of results Multiple alignment NEW

>gb|DQ464177.1| Hepatitis B virus isolate LT6 clone 4 HBcAg protein and HBsAg protein (S) genes, complete cds
Length=3149

Sort alignments for this subject sequence by:
E value **Score** **Percent identity**
Query start position **Subject start position**

Score = 38.2 bits (19), Expect = 0.19
Identities = 19/19 (100%), Gaps = 0/19 (0%)
Strand=Plus/Plus

```
Query  1     CCGGAAAGCTTGAGCTCTT  19
             ||||||||||||||||||| 
Sbjct  3091  CCGGAAAGCTTGAGCTCTT  3109
```

Score = 28.2 bits (14), Expect = 181
Identities = 14/14 (100%), Gaps = 0/14 (0%)
Strand=Plus/Minus

```
Query  6     AAGCTTGAGCTCTT  19
             ||||||||||||||
Sbjct  3007  AAGCTTGAGCTCTT  2994
```

>gb|DQ464176.1| Hepatitis B virus isolate LT6 clone 2 HBcAg protein and HBsAg protein (S) genes, complete cds
Length=3149

Score = 38.2 bits (19), Expect = 0.19
Identities = 19/19 (100%), Gaps = 0/19 (0%)
Strand=Plus/Plus

```
Query  1     CCGGAAAGCTTGAGCTCTT  19
             ||||||||||||||||||| 
Sbjct  3091  CCGGAAAGCTTGAGCTCTT  3109
```

>gb|DQ336691.1| Hepatitis B virus isolate LT5 clone 7 HBcAg and HBsAg (S) genes, complete cds
Length=2966

Score = 38.2 bits (19), Expect = 0.19
Identities = 19/19 (100%), Gaps = 0/19 (0%)
Strand=Plus/Plus

```
Query  1     CCGGAAAGCTTGAGCTCTT  19
             ||||||||||||||||||| 
Sbjct  2932  CCGGAAAGCTTGAGCTCTT  2950
```

>emb|CU633901.1| Podospora anserina genomic DNA chromosome 1, supercontig 2
Length=1337546

Features flanking this part of subject sequence:
922 bp at 5' side: unnamed protein product
1505 bp at 3' side: unnamed protein product

Score = 34.2 bits (17), Expect = 2.9
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Plus

```
Query  1        CCGGAAAGCTTGAGCTC  17
                ||||||||||||||||| 
Sbjct  1190979  CCGGAAAGCTTGAGCTC  1190995
```

>ref|XM_001767234.1| Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_230074) mRNA, complete cds
Length=11643

GENE ID: 5930420 PHYPADRAFT_230074 | hypothetical protein [Physcomitrella patens subsp. patens] (10 or fewer PubMed links)

Score = 34.2 bits (17), Expect = 2.9

```
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Plus

Query  3     GGAAAGCTTGAGCTCTT  19
             |||||||||||||||||
Sbjct  3328  GGAAAGCTTGAGCTCTT  3344


>ref|XM_001752228.1| UG Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_174045)
mRNA, complete cds
Length=399

 GENE ID: 5915452 PHYPADRAFT_174045 | hypothetical protein
[Physcomitrella patens subsp. patens] (10 or fewer PubMed links)

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Minus

Query  2    CGGAAAGCTTGAGCTCT  18
            |||||||||||||||||
Sbjct  123  CGGAAAGCTTGAGCTCT  107


>ref|XM_001704768.1| G Giardia lamblia ATCC 50803 Elongation factor 2 (GL50803_17570)
mRNA, complete cds
Length=2697

 GENE ID: 5697690 GL50803_17570 | Elongation factor 2
[Giardia lamblia ATCC 50803] (10 or fewer PubMed links)

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Minus

Query  1    CCGGAAAGCTTGAGCTC  17
            |||||||||||||||||
Sbjct  617  CCGGAAAGCTTGAGCTC  601


>emb|CT027701.11| D Zebrafish DNA sequence from clone CH211-209D1 in linkage group
16, complete sequence
Length=115132

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Plus

Query  3      GGAAAGCTTGAGCTCTT  19
               |||||||||||||||||
Sbjct  94763  GGAAAGCTTGAGCTCTT  94779


>gb|AC126538.1| D Homo sapiens chromosome 16 clone CTD-3225F13, complete sequence
Length=188015

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Minus

Query  3     GGAAAGCTTGAGCTCTT  19
             |||||||||||||||||
Sbjct  3414  GGAAAGCTTGAGCTCTT  3398


>gb|AC007014.1|AC007014 D Homo sapiens chromosome 16 clone 66H6, complete sequence
Length=167080

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Minus

Query  3      GGAAAGCTTGAGCTCTT  19
               |||||||||||||||||
Sbjct  60821  GGAAAGCTTGAGCTCTT  60805


>dbj|D29835.1|GIAEF2  Giardia intestinalis mRNA for elongation factor 2, partial cds
Length=2457

 Score = 34.2 bits (17),  Expect = 2.9
 Identities = 17/17 (100%), Gaps = 0/17 (0%)
 Strand=Plus/Minus

Query  1    CCGGAAAGCTTGAGCTC  17
            |||||||||||||||||
Sbjct  530  CCGGAAAGCTTGAGCTC  514


>tpe|BN001305.1| D TPA_reasm: Aspergillus nidulans FGSC A4 chromosome V
Length=3145033
```

Sort alignments for this subject sequence by:
E value **Score** **Percent identity**

>gi|102621374|gb|DQ464177.1| Hepatitis B virus isolate LT6 clone 4 HBcAg protein and HBsAg protein (S) genes, complete cds

TTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTG
GCTTTAGGACATGGACATTGACCCTTATAAAGAATTTGGAGCTTCCGTGGAGTTACTCTCGTTTTTGCCT
TCTGACTTCTATCCTTCAGTACGAGATCTTCTAGATACCGCCTCAGCTCTGTTTCGGGAAGCCTTAGAGT
CTCCTGAGCATTGTTCACCTCACCATACTGCACTCAGGCAAGCAATTCTTTGCTGGGGGAATCTAATGAC
TCTAGCTACCTGGGTGGGTGCTAATTTGGAAGATCCAGCATCTAGAGACCTAGTAGTCAGTTATGTCAAC
ACTAATATAGCCCTAAAGTTCAGGCAACTCTTGTGGTTTCACATTTCTTGTCTCACTTTTGGAAGAGAAG
TAGTTGTAGAGTATTTGGTGTCTTTCGGAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAATGC
CCCTATCCTATCAACACTTCCGGAGACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACT
CCCTCACCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTAAATCTCGGGAATCTCAATGTT
AGTATTCCTTGGACTCATAAGGTGGGGAGCTTTACTGGGCTTTATTCTTCTACTGTACCTGTCTTTAATC
CTCATTGGAAAACACCATCTTTTCCTAACATACATTTACACCAAGACATTATCAAAAAATGTGAACAGTT
TGTAGGCCCACTCACAGTTAATGAGAAAAGAAGATTGAAATTGATTATGCCTGCTAGGTTTTATCCAAAT
GCTACCAAATATTTACCATTGGATAAGGGTATTAAACCTTATTATCCAGAACATCTAGTTAATCATTACT
TCCAAACTAGACACTATTTACACACTCTATGGAAGGCGGGTATATTATATAAGAGAGAAACAACACATAG
CGCCTCATTTTGTGGGTCACCATATTCTTGGGAACAAGATCTACAGCATGGGGCAGAATCTTTCCACCAG
CAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAGAGCAAACACCGCAAATCCAGAT
TGGGACTTCAATCCCAACAAGGACACCTGGCCAGACGCCAACAAGGTAGGAGCTGGAGCATTCGGGCTGG
GTTTCACCCCACCGCACGGAGGTCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATGCTACAAACTTTGCC
AGCAAATCCGCCTCCTGCCTCCACCAATCGTCAGTCAGGAAGGCAGCCTACCCCGCTGTCTCCACCTCTG
AGAAACACTCATCCTCAGGCCATGCAGTGGAACTCCACAACCTTCCACCAAACTCTGCAAGATCCCAGAG
TGAGAGGCCTGTATTTCCCTGCTGGTGGCTCCAGTTCAGAAACAGCAAACCCTGTTCCGACTACTGCCTC
TCCTTTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCACTGAACATGGAGAACATCACATCAGGACTC
CTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATCCTCACAATACCGCAGAGTC
TAGACTCGTGGTGGACTTCTCTCAGTTTTCTAGGGGGAAACACCGTGTGTCTTGGCCAAAATTCGCAGTC
CCCAACCTCCAATCACTCACCAACCTCCTGTCCTCCAACTTGTCCTGGATATCGCTGGATGTGTCTGCGG
CGTTTTATCATCTTCCTCTTCATCCTGCTTCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAAG
GTATGTTGCCCGTTTGTCCTCTAATTCCAGGATCCTCAACCACCAGCACGGGACCATGCCGAACCTGCAC
GACTCTTGCTCAAGGAACCTCTATGTATCCCTCATGTTGCTGTACCAAACCTTCGGACGGAAATTGCACC
TGTATTCCCATCCCATCATCCTGGGCTTTCGGAAAATTCCTATGGGCGTGGGCCTCAGCCCGTTTCTCCT
GGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTAT
ATTGATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCCCTTTTTACCGTTGTTACCAATT
TTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAAAGATGGGGTTATTCTTTAAATTTTAT
GGGGTATGTCATTGGAAGTTATGGGTCCTTGCCACAAGAACACATCGTACAAAAAATCAAAGAGTGTTTT
AGAAAACTTCCTGTTAACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTAGGTTTTG
CTGCCCCTTTCACACAATGTGGTTATCCTGCTTTAATGCCCTTGTATGCATGTATTCACTCTAAGCAGGC
TTTCACTTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGG
CAACGGCCAGGTCTATGCCAAGTGTTTGCTGATGCAACCCCCACTGGTTGGGGCTTGGCCATGGGCCATC
AGCGCATGCGTGGAACCTTTCTGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGC
TCGCAGCAGGTCTGGAGCAAACATTATCGGGACTGATAACTCTGTTGTCCTCTCCCGCAAATATACATCG
TTTCCATGGCTGCTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGG
CGCTGAATCCCGCGGACGACCCTTCTCGGGGTCGCTTGGGACTCTCTCGTCCCTTTCTCCGTCTGCCGTT
CCGACCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCAGTGCCTTCTCATCTGCCGGACCGT
GTGCACTTCGCTTCACCTCTGCCTAATCATGCCCAGCATCATGCAACTTTTTGAAGAGCTCAAGCTTGCT
GGAAAGCTGGAGGTCTTCTTTTTCACCTCTGCCTAATCAGCAGACCCCAGCACCATGCAACTTTTTGAAG
AGCTCAATTTCCGGAAAGCTTGAGCTCTTCGTTTTCACCTCTGCCTAATCATGCCCAGCACCATGCAAC

Notes:

CGGAAAGCTTGAGCTCTT maps to bp 56-74 of T45.3

TTTTCACCTCTGCCTAATCAT maps to bp 2-22 at the beginning of the cloned sequence LT6

GCCCAGCACCATGCAAC maps to bp 2971 to 2987 also in the cloned sequence LT6

T45.3 bp 75-323

BLAST Basic Local Alignment Search Tool

Edit and Resubmit Save Search Strategies Formatting options Download

Nucleotide Sequence (322 letters)

Results for: lcl|49827 None(322bp)

Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|49827
Description
None
Molecule type
nucleic acid
Query Length
322

Database Name
nr
Description
All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program
BLASTN 2.2.21+ Citation

Reference
Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Search Parameters

Program	blastn
Word size	28
Expect value	10
Hitlist size	100
Match/Mismatch scores	1,-2
Gapcosts	0,0
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Database

Posted date	Sep 27, 2009 5:41 PM
Number of letters	28,993,070,215
Number of sequences	9,897,685
Entrez query	none

Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	1.33271	1.28
K	0.620991	0.46
H	1.12409	0.85

Results Statistics

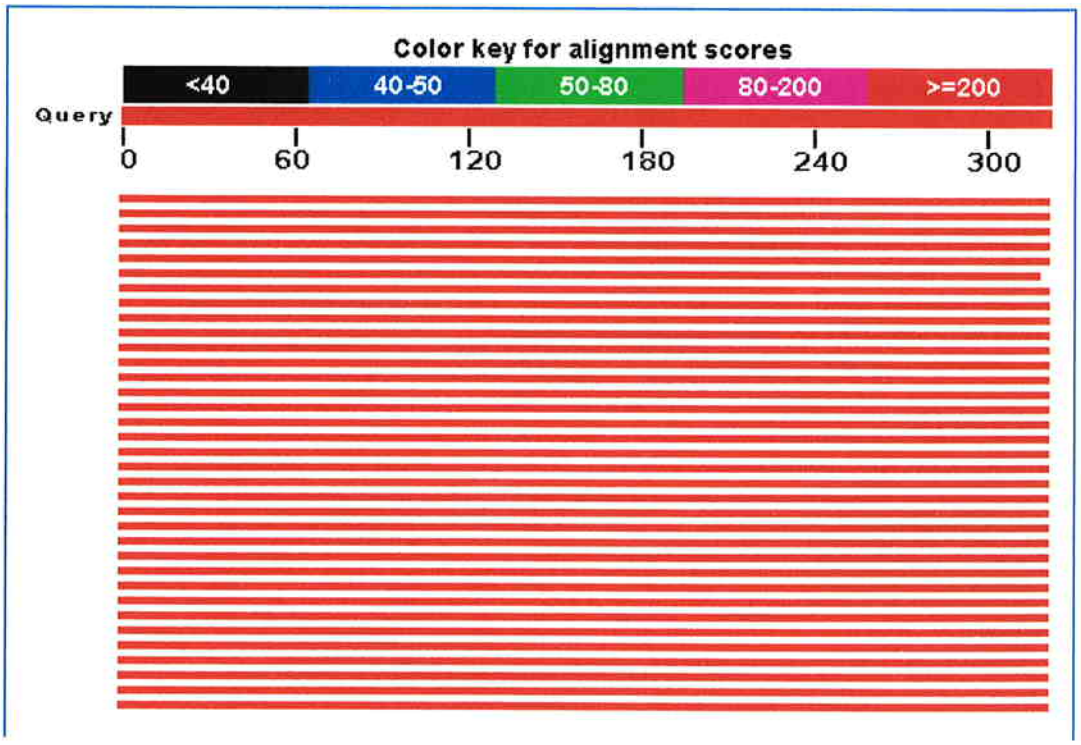
Length adjustment	31
Effective length of query	291
Effective length of database	28686241980
Effective search space	8347696416180
Effective search space used	8347696416180

Graphic Summary

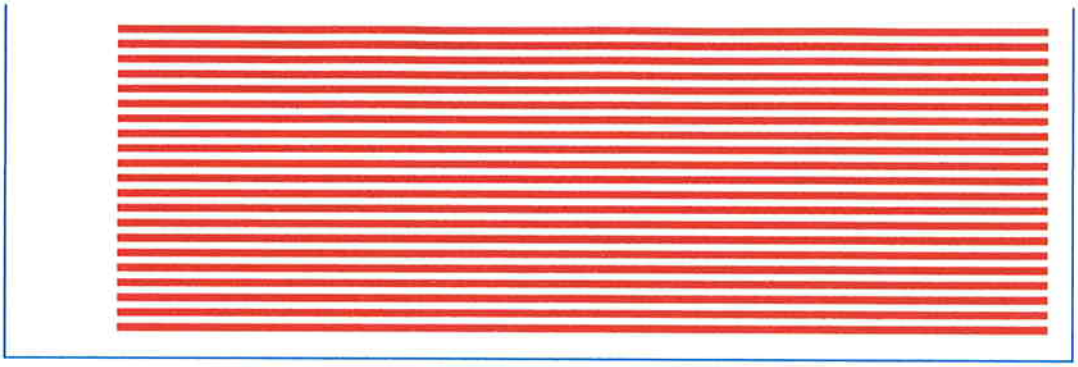
Distribution of 100 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.



G36



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

DQ060825.1	Hepatitis B virus isolate 0262-09, complete genome	579	579	100%	4e-162	99%
DQ060824.1	Hepatitis B virus isolate 0121-20, complete genome	579	579	100%	4e-162	99%
DQ297468.1	Hepatitis B virus isolate M precore/core protein gene, partial cds	579	579	100%	4e-162	99%
DQ297467.1	Hepatitis B virus isolate K precore/core protein gene, partial cds	579	579	100%	4e-162	99%
AP007262.1	HBV genotype E DNA, complete genome, isolate: HB-JI411F	573	573	100%	2e-160	98%
EU620071.1	Hepatitis B virus isolate Irish-16 X protein (X) gene, partial cds; and precore/core protein (C) and core protein (C) genes, complete cds	568	568	99%	8e-159	98%
AY738147.1	Hepatitis B virus clone B11, complete genome	568	568	100%	8e-159	98%
AY738145.1	Hepatitis B virus clone B5, complete genome	568	568	100%	8e-159	98%
AY738144.1	Hepatitis B virus clone B3, complete genome	568	568	100%	8e-159	98%
DQ060823.1	Hepatitis B virus isolate 8504-91, complete genome	568	568	100%	8e-159	98%
AM110883.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12255	562	562	100%	4e-157	98%
AY738146.1	Hepatitis B virus clone B9, complete genome	562	562	100%	4e-157	98%
AY739674.1	Hepatitis B virus clone A1-33, complete genome	562	562	100%	4e-157	98%
AY935700.1	Hepatitis B virus isolate TH-HBV-012, complete genome	562	562	100%	4e-157	98%
DQ060828.1	Hepatitis B virus isolate 0443-03, complete genome	562	562	100%	4e-157	98%
DQ060827.1	Hepatitis B virus isolate 0317-04, complete genome	562	562	100%	4e-157	98%
DQ060826.1	Hepatitis B virus isolate 0433-20, complete genome	562	562	100%	4e-157	98%
DQ060822.1	Hepatitis B virus isolate 8510-91, complete genome	562	562	100%	4e-157	98%
AB194947.1	Hepatitis B virus DNA, complete genome, isolate: CMR954	562	562	100%	4e-157	98%
AF350200.1	Hepatitis B virus isolate Gam1713 precore/core protein gene, complete cds	562	562	100%	4e-157	98%
AF350150.1	Hepatitis B virus isolate Gam1182F41 precore/core protein gene, complete cds	562	562	100%	4e-157	98%
GQ161816.1	Hepatitis B virus isolate GU1214, complete genome	556	556	100%	2e-155	97%
AM494717.1	Hepatitis B virus complete genome, isolate CAR187	556	556	100%	2e-155	97%
AM494715.1	Hepatitis B virus complete genome, isolate CAR171	556	556	100%	2e-155	97%
AM494712.1	Hepatitis B virus complete genome, isolate CAR150	556	556	100%	2e-155	97%
AM494708.1	Hepatitis B virus complete genome, isolate CAR138	556	556	100%	2e-155	97%
AM494707.1	Hepatitis B virus complete genome, isolate	556	556	100%	2e-	97%

	CAR137				155	
AM494706.1	Hepatitis B virus complete genome, isolate CAR131	556	556	100%	2e-155	97%
AM494704.1	Hepatitis B virus complete genome, isolate CAR112	556	556	100%	2e-155	97%
AM494702.1	Hepatitis B virus complete genome, isolate CAR106	556	556	100%	2e-155	97%
AM494699.1	Hepatitis B virus complete genome, isolate CAR082	556	556	100%	2e-155	97%
AM494693.1	Hepatitis B virus complete genome, isolate CAR039	556	556	100%	2e-155	97%
AM494690.1	Hepatitis B virus complete genome, isolate CAR008	556	556	100%	2e-155	97%
AM494689.1	Hepatitis B virus complete genome, isolate CAR006	556	556	100%	2e-155	97%
EU239226.1	Hepatitis B virus isolate PW6, complete genome	556	556	100%	2e-155	97%
AB274974.1	Hepatitis B virus s gene for S protein, complete cds, strain:AB60	556	556	100%	2e-155	97%
AM110903.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAE306	556	556	100%	2e-155	97%
AM110885.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12106	556	556	100%	2e-155	97%
AM110882.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12243	556	556	100%	2e-155	97%
AM110875.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate Lag561	556	556	100%	2e-155	97%
AM110862.1	Hepatitis B virus preC/C gene for preC/core protein, isolate NIE24175	556	556	100%	2e-155	97%
AM110861.1	Hepatitis B virus preC/C gene for preC/core protein, isolate NIE24028	556	556	100%	2e-155	97%
AM110859.1	Hepatitis B virus preC/C gene for preC/core protein, isolate NIE24106	556	556	100%	2e-155	97%
DQ060829.1	Hepatitis B virus isolate 0350-03, complete genome	556	556	100%	2e-155	97%
AB201289.1	Hepatitis B virus genes for X protein, precore/core ORF, complete cds, isolate:Benin269	556	556	100%	2e-155	97%
AF350151.1	Hepatitis B virus isolate Gam1184F41 precore/core protein gene, complete cds	556	556	100%	2e-155	97%
X75657.1	Human hepatitis virus (genotype E, Bas) preS1, preS2, S, C, X, antigens, core antigen, X protein and polymerase	556	556	100%	2e-155	97%
GQ161828.1	Hepatitis B virus isolate GU1399, complete genome	551	551	100%	8e-154	97%
GQ161817.1	Hepatitis B virus isolate GU630, complete genome	551	551	100%	8e-154	97%
GQ161804.1	Hepatitis B virus isolate GU1428, complete genome	551	551	100%	8e-154	97%
GQ161798.1	Hepatitis B virus isolate GU1222, complete genome	551	551	100%	8e-154	97%
GQ161789.1	Hepatitis B virus isolate GU1169, complete genome	551	551	100%	8e-154	97%
GQ161784.1	Hepatitis B virus isolate GU766, complete genome	551	551	100%	8e-154	97%
GQ161781.1	Hepatitis B virus isolate GU200, complete genome	551	551	100%	8e-154	97%
GQ161776.1	Hepatitis B virus isolate GU1178, complete genome	551	551	100%	8e-154	97%

GQ161771.1	Hepatitis B virus isolate GU896, complete genome	551	551	100%	8e-154	97%
EF103285.1	Hepatitis B virus isolate Ran128, complete genome	551	551	100%	8e-154	97%
AM494713.1	Hepatitis B virus complete genome, isolate CAR154	551	551	100%	8e-154	97%
AM494710.1	Hepatitis B virus complete genome, isolate CAR144	551	551	100%	8e-154	97%
AM494709.1	Hepatitis B virus complete genome, isolate CAR143	551	551	100%	8e-154	97%
AM494705.1	Hepatitis B virus complete genome, isolate CAR115	551	551	100%	8e-154	97%
AM494701.1	Hepatitis B virus complete genome, isolate CAR097	551	551	100%	8e-154	97%
AM494697.1	Hepatitis B virus complete genome, isolate CAR063	551	551	100%	8e-154	97%
AB274972.1	Hepatitis B virus s gene for S protein, complete cds, strain:AB250	551	551	100%	8e-154	97%
AB274971.1	Hepatitis B virus s gene for S protein, complete cds, strain:AB295	551	551	100%	8e-154	97%
AB274970.1	Hepatitis B virus s gene for S protein, complete cds, strain:AB291	551	551	100%	8e-154	97%
AM110909.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch36	551	551	100%	8e-154	97%
AM110904.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate CAE375	551	551	100%	8e-154	97%
AM110896.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate CAE193	551	551	100%	8e-154	97%
AM110894.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate CAE40	551	551	100%	8e-154	97%
AM110887.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12179	551	551	100%	8e-154	97%
AM110886.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12195	551	551	100%	8e-154	97%
AM110880.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE12260	551	551	100%	8e-154	97%
AM110873.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24258	551	551	100%	8e-154	97%
AM110870.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24261	551	551	100%	8e-154	97%
AM110869.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24259	551	551	100%	8e-154	97%
AM110868.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24255	551	551	100%	8e-154	97%
AM110867.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24233	551	551	100%	8e-154	97%
AM110866.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24195	551	551	100%	8e-154	97%
AM110865.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24186	551	551	100%	8e-154	97%
AM110858.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24142	551	551	100%	8e-154	97%
AM110857.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24146	551	551	100%	8e-154	97%
AM110853.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24095	551	551	100%	8e-154	97%
AM110852.1	Hepatitis B virus partial preC/C gene for	551	551	100%	8e-	97%

	preC/core protein, isolate NIE24093				154	
AM110848.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24008	551	551	100%	8e-154	97%
AM110808.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate CAEch60	551	551	100%	8e-154	97%
AM110806.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch57	551	551	100%	8e-154	97%
AM110805.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch55	551	551	100%	8e-154	97%
AY739675.1	Hepatitis B virus clone A1-35, complete genome	551	551	100%	8e-154	97%
AB201290.1	Hepatitis B virus genes for X protein, precore/core ORF, complete cds, isolate:Benin293	551	551	100%	8e-154	97%
AB091256.1	Hepatitis B virus DNA, complete genome, isplate:ABI-212	551	551	100%	8e-154	97%
AB091255.1	Hepatitis B virus DNA, complete genome, isplate:ABI-129	551	551	100%	8e-154	97%
AM110908.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch13	549	549	100%	3e-153	97%
GQ161780.1	Hepatitis B virus isolate GU50, complete genome	547	547	100%	1e-152	97%
AM110915.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch52	547	547	100%	1e-152	97%
AM110906.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch4	547	547	100%	1e-152	97%
AM110905.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch3	547	547	100%	1e-152	97%
AM110850.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate NIE24072	547	547	100%	1e-152	97%
AM110804.1	Hepatitis B virus preC/C gene for preC/core protein, isolate CAEch53	547	547	100%	1e-152	97%
AM110914.1	Hepatitis B virus partial preC/C gene for preC/core protein, isolate CAEch48	545	545	100%	4e-152	96%

Alignments Select All Get selected sequences Distance tree of results Multiple alignment NEW

>gb|DQ060825.1| Hepatitis B virus isolate 0262-09, complete genome
Length=3212

Score = 579 bits (313), Expect = 4e-162
Identities = 320/323 (99%), Gaps = 1/323 (0%)
Strand=Plus/Plus

```
Query  1     CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1820  CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  1879

Query  61    TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1880  TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  1939

Query  121   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1940  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  1999

Query  181   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2000  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  2059

Query  241   GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2060  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  2119

Query  301   GTAA-TTTGGAAGATCCAGCATC  322
             |||| ||||||||||||||||||
Sbjct  2120  GTAAATTTGGAAGATCCAGCATC  2142
```

>gb|DQ060824.1| Hepatitis B virus isolate 0121-20, complete genome
Length=3212

Score = 579 bits (313), Expect = 4e-162
Identities = 320/323 (99%), Gaps = 1/323 (0%)
Strand=Plus/Plus

```
Query  1     CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1820  CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  1879

Query  61    TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1880  TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  1939

Query  121   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1940  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  1999

Query  181   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2000  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  2059

Query  241   GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2060  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  2119

Query  301   GTAA-TTTGGAAGATCCAGCATC  322
             |||| ||||||||||||||||||
```

```
Sbjct  2120  GTAAATTTGGAAGATCCAGCATC  2142
```

>gb|DQ297468.1| Hepatitis B virus isolate M precore/core protein gene, partial cds
Length=489

```
 Score =  579 bits (313),  Expect = 4e-162
 Identities = 320/323 (99%), Gaps = 1/323 (0%)
 Strand=Plus/Plus

Query  1    CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  44   CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  103

Query  61   TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  104  TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  163

Query  121  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  164  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  223

Query  181  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  224  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  283

Query  241  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  284  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  343

Query  301  GTAA-TTTGGAAGATCCAGCATC  322
            |||| ||||||||||||||||||
Sbjct  344  GTAAATTTGGAAGATCCAGCATC  366
```

>gb|DQ297467.1| Hepatitis B virus isolate K precore/core protein gene, partial cds
Length=489

```
 Score =  579 bits (313),  Expect = 4e-162
 Identities = 320/323 (99%), Gaps = 1/323 (0%)
 Strand=Plus/Plus

Query  1    CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  44   CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  103

Query  61   TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  104  TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  163

Query  121  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  164  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  223

Query  181  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  224  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  283

Query  241  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  284  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  343

Query  301  GTAA-TTTGGAAGATCCAGCATC  322
            |||| ||||||||||||||||||
Sbjct  344  GTAAATTTGGAAGATCCAGCATC  366
```

>dbj|AP007262.1| HBV genotype E DNA, complete genome, isolate: HB-JI411F
Length=3212

```
 Score =  573 bits (310),  Expect = 2e-160
 Identities = 319/323 (98%), Gaps = 1/323 (0%)
 Strand=Plus/Plus

Query  1     CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1820  CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  1879

Query  61    TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
```

```
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||||||
Sbjct  1880  TGCCTTGGGTGGCTTTGGGACATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  1939

Query  121   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1940  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  1999

Query  181   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2000  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  2059

Query  241   GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2060  GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  2119

Query  301   GTAA-TTTGGAAGATCCAGCATC  322
              |||| ||||||||||||||||||
Sbjct  2120  GTAAATTTGGAAGATCCAGCATC  2142
```

>**gb|EU620071.1|** Hepatitis B virus isolate Irish-16 X protein (X) gene, partial cds; and precore/core protein (C) and core protein (C) genes, complete cds
Length=761

```
 Score =  568 bits (307),  Expect = 8e-159
 Identities = 316/320 (98%), Gaps = 1/320 (0%)
 Strand=Plus/Plus

Query  1     CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  129   CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  188

Query  61    TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
              |||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Sbjct  189   TGCCTTGGGTGGCTTTGGGACATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  248

Query  121   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  249   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  308

Query  181   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  309   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  368

Query  241   GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  369   GCACTCAGGCAAGCCATTCTTTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGGT  428

Query  301   GTAA-TTTGGAAGATCCAGC  319
              |||| ||||||||||||||||
Sbjct  429   GTAAATTTGGAAGATCCAGC  448
```

>**gb|AY738147.1|** Hepatitis B virus clone B11, complete genome
Length=3212

```
 Score =  568 bits (307),  Expect = 8e-159
 Identities = 319/324 (98%), Gaps = 3/324 (0%)
 Strand=Plus/Plus

Query  1     CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1820  CTTTTTCACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTG  1879

Query  61    TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGTTACTGTG  120
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1880  TGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTG  1939

Query  121   GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1940  GAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTAAGAGATCTTCTAGATACC  1999

Query  181   GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACAATACT  240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2000  GCCTCTGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTCACCTCACCATACT  2059

Query  241   GCACTCAGGCAAGCCATTCTT-TGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGG  299
              |||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Sbjct  2060  GCACTCAGGCAAGCCATT-TTGTGCTGGGGAGAATTAATGACTCTAGCTACCTGGGTGGG  2118
```

REFERENCES

1 ACIP - Advisory Committee on Immunization Practices. (1985) Recommendations for protection against viral hepatitis. *MMWR. Morbidity and Mortality Weekly Report* **34:**313-324, 329-335.

2 Aden DP, Fogel A., Plotkin S, Damjanov I, Knowles BB. (1979) Controlled synthesis of HBsAg in a differentiated liver carcinoma-derived cell line. *Nature* **282:**615-616.

3 Alexander JJ, Bey EM, Geddes EW, Lecatsas G. (1976) Establishment of a continuously growing cell line from primary carcinoma of the liver. *South African Medical Journal* **50:**2124-2128.

4 Almeida JD, Rubenstein D, Stott EJ. (1971) New antigen-antibody system in Australia-antigen-positive hepatitis. *The Lancet* **2:**1225-1227.

5 Alter HJ, Holland PV, Purcell RH. (1975) The emerging pattern of post-transfusion hepatitis. *American Journal of Medical Science* **270:**329-334.

6 Alter HJ, Purcell RH, Gerin JL, London WT, Kaplan PM, McAuliffe VJ, Wagner J, Holland PV. (1977) Transmission of hepatitis B to chimpanzees by hepatitis B surface antigen-positive saliva and semen. *Infection and Immunity* **16:**928-933.

7 Antonucci TK, Rutter WJ. (1989) Hepatitis B virus (HBV) promoters are regulated by the HBV enhancer in a tissue-specific manner. *Journal of Virology* **63:**579-583.

8 Anzola M. (2004) Hepatocellular carcinoma: role of hepatitis B and hepatitis C viruses proteins in hepatocarcinogenesis. *Journal of Viral Hepatitis* **11:**383-393.

9 Aoki N, Robinson WS. (1989) State of hepatitis B viral genomes in cirrhotic and hepatocellular carcinoma nodules. *Molecular Biology and Medicine* **6:**395-408.

10 Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. (2002) Genotype H: A new Amerindian genotype of hepatitis B virus revealed in Central America. *Journal of General Virology* **83:**2059-2073.

11 Arbuthnot P, Kew M. (2001) Hepatitis B virus and hepatocellular carcinoma. *International Journal of Experimental Pathology* **82:**77-100.

12 Ashida K, Kishimoto Y, Nakamoto K, Wada K, Shiota G, Hirooka Y, Kamisaki Y, Itoh T, Kawasaki H. (1997) Loss of heterozygosity of the retinoblastoma gene in liver cirrhosis accompanying hepatocellular carcinoma. *Journal of Cancer Research and Clinical Oncology* **123:**489-495.

13 Baca-Garcia E, Vaquero-Lorenzo C, Diaz-Hernandez M, Rodriguez-Salgado B, Dolengevich-Segal H, Arrojo-Romero M, Botillo-Martin C, Ceverino A, Piqueras JF, Perez-Rodriguez MM, Saiz-Ruiz J. (2007) Association between obsessive-compulsive disorder and a variable number of tandem repeats polymorphism in intron 2 of the serotonin transporter gene. *Progress in Neuro-psychopharmacology & Biological Psychiatry* **31:**416-420.

14 Balsano C, Avantaggiati ML, Natoli G, De Marzio E, Will H, Perricaudet M, Levrero M. (1991) Full-length and truncated versions of the hepatitis B virus (HBV) X protein (pX) transactivate the *c-myc* protooncogene at the transcriptional level. *Biochemical Biophysical Research Communication*s **176:**985-992.

15 Balsano C, Billet O, Bennoun M, Cavard C, Zider A, Grimber G, Natoli G, Briand P, Levrero M. (1994) Hepatitis B virus X gene product acts as a transactivator in vivo. *Journal of Hepatology* **21:**103-109.

16 Bancroft WH, Snitbhan R, Scott RM, Tingpalapong M, Watson WT, Tanticharoenyos P, Karwacki JJ, Srimarut S. (1977) Transmission of hepatitis B virus to gibbons by exposure to human saliva containing hepatitis B surface antigen. *Journal of Infectious Diseases* **135:**79-85.

17 Bartenschlager R, Schaller H. (1988) The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome linked protein) believed to prime reverse transcription. *EMBO Journal* **7:**4185-4192.

18 Bartenschlager R, Junker-Niepmann M, Schaller H. (1990) The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *Journal of Virology* **64:**5324-5332.

19 Bartenschlager R, Schaller H. (1992) Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *The EMBO Journal* **11:**3413-3420.

20 Bassett MLO, Halliday JW, Powell LW. (1986) Value of hepatic iron measurements in early hemochromatosis and determination of the critical iron level associated with fibrosis. *Hepatology* **6:**24-29.

21 Beasley RP, Trepo C, Stevens CE, Szmuness W. (1977) The e antigen and vertical transmission of hepatitis B surface antigen. *American Journal of Epidemiology* **105:**94-98.

22 Beasley RP, Hwang L-Y, Lin C-C, Stevens CE, 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 state. Initial report of a randomised double-blind placebo-controlled trial. *The Lancet* **2:**388-393.

23 Beasley RP, Hwang L-Y, Lin C-C, Leu M-L, Stevens CE, Szmuness W, Chen K-P. (1982) Incidence of hepatitis B virus infections in preschool children in Taiwan. *The Journal of Infectious Diseases* **146:**198-204.

24 Beasley RP, Hwang L-Y. (1983) Postnatal infectivity of hepatitis B surface antigen-carrier mothers. *The Journal of Infectious Diseases* **147:**185-190.

25 Bekondi C, Olinger CM, Boua N, Talarmin A, Muller CP, Faou AL, Venard V. (2007) Central African Republic is part of the West-African hepatitis B virus genotype E crescent. *Journal of Clinical Virology* **40:**31-37.

26 Benn J, Schneider RJ. (1994) Hepatitis B virus HBx activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signaling cascade. *Proceedings of the National Academy of Sciences of the United States of America* **91:**10350-10354.

27 Benn J, Su F, Doria M, Schneider RJ. (1996) Hepatitis B virus HBx protein induces transcription factor AP-1 by activation of extracellular signal-regulated and c-Jun N-terminal mitogen-activated protein kinases. *Journal of Virology* **70:**4978-4985.

28 Benvegnu I, Fattovich G, Noventa F, Tremolada F, Chemello L, Cecchetto A, Alberti A. (1994) Concurrent hepatitis B and C virus infection and risk of hepatocellular carcinoma in cirrhosis. A porspective study. *Cancer* **74:**2442-2448.

29 Berasain C, Patil D, Perara E, Huang SM, Mouly H, Bréchot C. (1998) Oncogenic activation of a human cyclin A2 targeted to the endoplasmic reticulum upon hepatitis B virus genome insertion. *Oncogene* **16:**1277-1288.

30 Berger I, Shaul Y. (1987) Integration of hepatitis B Virus: Analysis of unoccupied sites. *Journal of virology* **61:**1180-1186.

31 Beuren AJ, Apitz J, Harmjanz D. (1962) Supravalvular aortic stenosis in association with mental retardation and a certain facial appearance. *Circulation* **26:**1235-1240.

32 Blitz L, Pujol FH, Swenson PD, Porto L, Atencio R, Araujo M, Costa L, Monsalve DC, Torres JR, Fields HA, Lambert S, Van Geyt C, Norder H, Magnius LO, Echevarria JM, Stuyver L. (1998) Antigenic diversity of hepatitis B virus strains of genotype F in Ameridians and other population groups from Venezuela. *Journal of Clinical Microbiology* **36:**648-651.

33 Blum HE, Offensperger W-B, Walter E, Offensperger S, Wahl A, Zeschnigk C, Gerok W. (1987) Hepatocellular carcinoma and hepatitis B virus infection: molecular evidence for monoclonal origin and expansion of malignantly transformed hepatocytes. *Journal of Cancer Research Clinical Oncology* **113:**466-472.

34 Blum HE, Zhang ZS, Galun E von Weizsacker F, Garner B, Lliang TJ, Wands JR. (1992) Hepatitis B virus X protein is not central to the viral life cycle in vitro. *Journal of Virology* **66:**1223-1227.

35 Blumberg BS. (1964) Polymorphisms of the serum proteins and the development of iso-precipitins in transfused patients. *Bulletin of the New York Academy of Medicine* **40:**377-386.

36 Blumberg BS, Alter HJ, Visnich S. (1965) A "New" Antigen in Leukemia Sera. *The Journal of the American Medical Association*; **191:**541-546.

37 Blumberg BS & Millman I. (1972) Vaccine Against Viral Hepatitis and Process, US Patent Office No. 3,636.191 (1972).

38 Bond HM, Mesuraca M, Amodio N, Mega T, Agosti V, Fanello D, Pelaggi D, Bullinger L, Grieco M, Moore MA, Venuta S, Morrone G. (2008) Early hematopoietic zinc finger protein-zinc finger protein 521: a candidate regulator of diverse immature cells. *The International Journal of Biochemistry and Cell Biology* **40:**848-854.

39 Bookstein R, Shew J-Y, Chen P-L, Scully P, Lee W-H. (1990) Suppression of tumorigenicity of human prostate carcinoma cells by replacing a mutated RB gene. *Science* **247:**712-715.

40 Bowcock AM, Pinto MR, Bey E, Kuyl JM, Dusheiko GM, Bernstein R. (1985) The PLC/PRF/5 human hepatoma cell line. II. Chromosomal assigment of hepatitis B virus integration sites. *Cancer Genetics Cytogenetics* **18:**19-26.

41 Bowyer SM, Van Staden L, Kew MC, Sim JG. (1997) A unique segment of the hepatitis B virus group A genotype S identified in isolates from South Africa. *Journal of General Virology* **78:**1719-1729.

42 Breuza L, Halbeisen R, Jenö P, Otte S, Barlowe C, Hong W, Hauri HP. (2004) Proteomics of endoplasmic reticulum-Golgi intermediate compartment (ERGIC) membranes from brefeldin A-treated HepG2 cells identifies ERGIC-32, a new cycling protein that interacts with human Erv46. *The Journal of Biological Chemistry* **279:**47242-47253.

43 Britten RJ. (1994) Evolutionary selection against change in many *Alu* repeat sequences interspersed through primate genomes. *Proceedings of the National Academy of Science of the United States of America* **91:**5992-5996.

44 Bréchot C, Pourcel C, Louise A, Rain B, Tiollais P. (1980) Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. *Nature* **286:**533-535.

45 Bréchot C, Hadchouel M, Scotto J, Fonk M, Potet F, Vyas GN, Tiollais P. (1981) State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and -negative liver diseases. *Proceedings of the National Academy of Sciences of the United States of America* **78:**3906-3910.

46 Bréchot C. (1987) Hepatitis B virus (HBV) and hepatocellular carcinoma. HBV DNA status and its implications. *Journal of Hepatology* **4:**269-279.

47 Bréchot C, Gozuacik D, Murakami Y, Paterlini-Bréchot P. (2000) Molecular basis for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). *Seminars in Cancer Biology* **10:**211-231.

48 Bressac B, Kew M, Wands J, Ozturk M. (1991) Selective G to T mutations of *p53* gene in hepatocelluar carcinoma from southern Africa. *Nature* **350:**429-431.

49 Bruni R, Argentini C, D'Ugo E, Giuseppetti R, Rapicetta M. (1995) A PCR-based strategy for rapid mapping of hepadnavirus integrated sequences in hepatocellular carcinomas. *Journal of Virological Methods* **52:**347-360.

50 Bruss V, Gerlich WH. (1988) Formation of transmembraneous hepatitis B e-antigen by contranslational in vitro processing of the viral precore protein. *Virology* **163:**268-275.

51 Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH. (1996) Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *Journal of Virology* **70:**5845-5851.

52 Buendia MA. (1992) Hepatitis B viruses and hepatocellular carcinoma. *Advances in Cancer Research* **59:**167-226.

53 Buetow KH, Murray JC, Israel JL, London WT, Smith M, Kew M, BlanquetV, Bréchot C, Redeker A, Govindarajah S. (1989) Loss of heterozygosity suggests tumor suppresssor gene responsible for primary hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **86:**8852-8856.

54 Buisine MP, Wacrenier A, Mariette C, Leteurtre E, Escande F, Aissi S, Ketele A, Leclercq, Porchet N, Lesuffleur T. (2008) Frequent mutations of the CA simple sequence repeat in intron 1 of EGFR in mismatch repair-deficient colorectal cancers. *World Journal of Gastroenterology* **14:**1053-1059.

55 Bullock AN, Rodriguez MC, Debreczeni JE, Songyang Z, Knapp S. (2007) Structure of the SOCS4-ElonginB/C complex reveals a distinct SOCS box interface and the molecular basis for SOCS-dependent EGFR degradation. *Structure* **15:**1493-1504.

56 Capovilla A., Carmona S, Arbuthnot P. (1997) Hepatitis B virus X protein binds damaged DNA and sensitizes liver cells to ultraviolet irradiation. *Biochemical Biophysical Reseach Communications* **232:**255-260.

57 Carman WF, Zanetti AR, Karayiannis P, Waters J, Manzillo G, Tanzi E, Zuckerman AJ, Thomas HC. (1990) Vaccine-induced escape mutant of hepatitis B virus. *The Lancet* **336:**325-329.

58 Carman WF, Van Deursen FJ, Mimms LT, Hardie D, Coppola R, Decker R, Sanders R. (1997) The prevalence of surface antigen variants of hepatitis B virus in Papua New Guinea, South Africa, and Sardinia. *Hepatology* **26:**1658-1666.

59 Caselmann WH, Meyer M, Kekulé AS, Laluer U, Hofschneider PH, Koshy R. (1990) A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. *Proceedingss of the National Academy of Sciences of the United States of America* **87:**2970-2974.

60 Cattaneo R, Will H, Hernandez N, Schaller H. (1983) Signals regulating hepatitis B surface antigen transcription. *Nature* **305:**336-338.

61 Cattaneo R, Will H, Schaller H. (1984) Hepatitis B virus transcription in the infected liver. *The EMBO Journal* **3:**2191-2196.

62 Cerutti PA. (1988) Response modification creates promotability in multistage carcinogenesis. *Carcinogenesis* **9:**519-526.

63 Chakraborty PR, Ruiz-Opazo N, Shouval D, Shafritz DA. (1980) Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. *Nature* **286:**531-533.

64 Chami M, Gozuacik D, Saigo K, Capiod T, Falson P, Lecoeur H, Urashima T, Beckmann J, Gougeon ML, Claret M, le Maire M, Bréchot C, Paterlini-Bréchot P. (2000) Hepatitis B virus related insertional mutagenesis implicates *SERCA*1 gene in the control of apoptosis. *Oncogene* **19:**2877-2886.

65 Chami M, Gozuacik D, Lagorce D, Brini M, Falson P, Peaucellier G, Pinton P, Lecoeur H, Gougeon M-L, le Maire M, Rizzuto R, Bréchot C, Paterlini P. (2001) *SERCA*1 truncated proteins unable to pump calcium reduce the endoplasmic reticulum concentration and induce apoptosis. *The Journal of Cell Biology* **153:**1301-1314.

66 Chang LJ, Hirsch RC, Ganem D, Varmus HE. (1990) Effects of insertional and point mutations on the functions of the duck hepatitis B virus polymerase. *The Journal of Virology* **64:**5553-5558.

67 Chang MH, Chen CJ, Lai MS, Hsu H-M, Wu T-C, Kong M-S, Lliang D-C, Shau W-Y, Chen D-S. (1997) Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. Taiwan Childhood Hepatoma Study Group. *The New England Journal of Medicine* **336:**1855-1859.

68 Chen DS, Hoyer BH, Nelson J, Purcell RH, Gerin JL. (1982) Detection and properties of hepatitis B viral DNA in liver tissues from patients with hepatocellular carcinoma. *Hepatology* **2:**42S-46S.

69 Chen HS, Kaneko S, Girones RW, Anderson RW, Hornbuckle WE, Tennant BC, Cote PJ, Gerin JL, Purcell RH, Miller RH. (1993) The woodchuck hepatitis virus X gene is important for establishment of virus infection in woodchucks. *The Journal of Virology* **67:**1218-1226.

70 Chen I-H, Huang CJ, Ting L-P. (1995) Overlapping initiator and TATA box functions in the basal core promoter of hepatitis B virus. *Journal of Virology* **69:**3647-3657.

71 Chen J-Y, Harrison TJ, Lee C-S, Chen D-S, Zuckerman AJ. (1986) Detection of hepatitis B virus DNA in hepatocellular carcinoma. *British Journal of Experimental Pathology* **67:**279-288.

72 Chen J-Y, Harrison TJ, Lee C-S, Chen D-S, Zuckerman AJ. (1988) Detection of hepatitis B virus DNA in hepatocellular carcinoma: analysis by hybridisation with subgenomic DNA fragments. *Hepatology* **8:**518-523.

73 Chen J-Y, Harrison TJ, Tsuei D-J, Hsu T-Y, Zuckerman AJ, Chan TS, Yang C-S. (1994) Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in the hepatocellular carcinoma cell line HCC36. *Intervirology* **37:**41-46.

74 Chen MT, Billaud JN, Sällberg M, Guidotti LG, Chisari FV, Jones J, Hughes J, Milich DR. (2004) A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proceedings of the National Academy of Science of the United States of America* **101:**14913-14918.

75 Chen WN, Oon CJ, Leong AL, Koh S, Teng SW. (2000) Expression of integrated hepatitis B virus X variants in human hepatocellular carcinomas and its significance. *Biochemical and Biophysical Research Communications* **276:**885-892.

76 Chisari FV. (1988) New model systems for hepatitis B virus research. *Laboratory Investigation; a journal of technical methods and pathology* **59:**155-157.

77 Chisari FV, Ferrari C. (1995a) Hepatitis B virus immunopathology. *Springer Seminars in Immunopathology* **17:**261-281.

78 Chisari FV, Ferrari C. (1995b) Hepatitis B virus immunopathogenesis. *Annual Review of Immunology* **13:**29-60.

79 Chisari FV. (1997) Cytotoxi T cells and viral hepatitis. *Journal of Clinical Investigation* **99:**1472-1477.

80 Chisari FV. (2000) Rous-Whipple Award Lecture. Viruses, immunity, and cancer: lessons from hepatitis B. *American Journal of Pathology* **156:**1117-1132.

81 Chomyn A, Mariottini P, Cleeter MWL, Ragan CI, Matsuno-Yagi A, Hatefi Y, Doolittle RF, Attardi G. (1985) Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase . *Nature* **314:**592-597.

82 Chomczynski P, Sacchi N. (1987) The single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols* **1:**581-585.

83 Chumakov IM, Le Gall I, Billaut A, Ougen P, Soularue P, Guillou S, Rignault P, Bui H, De Tand MF, Barillot E, Abderrahim H, Cherif D, Berger R, Le Paslier D, Cohen D. (1992) Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genetics* **1:**222-225.

84 Cohen BJ, Richmond JE. (1982) Electron microscopy of hepatitis B core antigen synthesized in *E. coli*. *Nature* **296:**677-679.

85 Courouce-Pauty AM, Plancon A, Soulier JP. (1983) Distribution of HBsAg subtypes in the world. *Vox Sang* **44:**197-211.

86 Cully M, Shiu J, Piekorz RP, Muller WJ, Done SJ, Mak TW. (2005) Transforming acidic coiled coil 1 promotes transformation and mammary tumorigenesis. *Cancer Research* **65:**10363-10370.

87 Cunningham JM, Jane SM. (1996) Hemoglobin switching and fetal hemoglobin reactivation. *Seminars in Hematology* **33:**9-23.

88 Cunningham RP, Shibata T, DasGupta C, Radding CM. (1979) Single strands induce recA protein to unwind duplex DNA for homologous pairing. *Nature* **281:**191-195.

89 Dane DS, Cameron CH, Briggs M. (1970) Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *The Lancet* **1:**695-698.

90 Dejean A, Bréchot C, Tiollais P, Wain-Hobson S. (1983) Characterisation of integrated hepatitis B viral DNA cloned from a human hepatoma and the hepatoma-derived cell line PLC/PRF/5. *Proceedings of the National Academy of Sciences of the United States of America* **80:**2505-2509.

91 Dejean A, Sonigo P, Wain-Hobson S, Tiollais P. (1984) Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat. *Proceedings of the National Academy of Sciences of the United States of America* **81:**5350-5354.

92 Dejean A, Bougueleret L, Grzeschik KM, Tiollais P. (1986) Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. *Nature* **322:**70-72.

93 Dejean A, de Thé H. (1990) Hepatitis B as an insertional mutagene in a human hepatocellular carcinoma. *Molecular Biology and Medicine* **7:**213-222.

94 de La Coste A, Romagnolo B, Billuart P, Renard CA, Buendia MA, Soubrane O, Fabre M, Chelly J, Beldjord C, Kahn A, Perret C. (1998) Somatic mutations of the *beta-catenin* gene are frequent in mouse and human hepatocellular carcinomas. *Proceedings of the National Academy of Sciences of the United States of America* **95:**8847-8851.

95 Delius H, Gough NM, Cameron CH, Murray K. (1983) Structure of the hepatitis B virus genome. *Journal of Virology* **47:**337-343.

96 De-Medina T, Haviv I, Noiman S, Shaul Y. (1994) The X protein of hepatitis B virus has a ribo/deoxy ATPase activity. *Virology* **202:**401-407.

97 De Souza AT, Hankins GR, Washington MK, Orton TC, Jirtle RL. (1995) *M6P*/*IGF*2R gene is mutated in human hepatocellular carcinomas with loss of heterozygosity. *Nature Genetics* **11:**447-449.

98 Deng LW, Chiu I, Strominger JL. (2004) MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. *Proceedings of the National Academy of Science of the United States of America* **101:**757-762.

99 Derynck R, Zhang Y, Feng XH. (1998) Smads: Transcriptional activation of TGF-beta responses. *Cell* **95:**737-740.

100 Diamantis ID, McGandy CE, Chen T-J, Liaw Y-F, Gudat F, Bianchi L. (1992) Hepatitis B *X*-gene expression in hepatocellular carcinoma. *Journal of Hepatology* **15:**400-403.

101 Diao J, Khine AA, Sarangi F, Hsu E, Iorio C, Tibbles LA, Woodgett JR, Penninger J, Richardson CD. (2001a) X protein of hepatitis B virus inhibits Fas-mediated apoptosis and is associated with up-regulation of the SAPK/JNK pathway. *The Journal of Biological Chemistry* **276:**8328-8340.

102 Diao J, Garces R, Richardson CD. (2001b) X protein of hepatitis B virus modulates cytokine and growth factor related signal transduction pathways during the course of viral infection and hepatocarcinogenesis. *Cytokine and Growth Factors Reviews* **12:**189-205.

103 Dickens, C. (2002) Investigating changes in liver cell gene expression by differential display. Thesis (M.Sc.)--University of the Witwatersrand, Science Faculty (Molecular and Cell Biology), 125 leaves.

104 Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1994) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19:**4008.

105 Doria M, Klein N, Lucito R, Schneider RJ. (1995) The hepatitis B virus HBx protein is a dual specificity cytoplasmic activator of Ras and nuclear activator of transcription factors. *The EMBO Journal* **14:**4747-4757.

106 Dover GA. (1989) DNA fingerprints: victims or perpetrators of DNA turnover? *Nature* **342:**347-348.

107 Eble BE, Lingappa VR, Ganem D. (1986) Hepatitis B surface antigen: an unusual secreted protein initially synthesized as a transmembrane polypeptide. *Molecular and Cellular Biology* **6:**1454-1463.

108 Echevarria JM, Avellon A, Magnius LO. (2005) Molecular epidemiology of hepatitis B virus in Spain: identification of viral genotypes and prediction of antigenic subtypes by limited sequencing. *Journal of General Virology* **76:**176-184.

109 Edman JC, Gray P, Valenzuela P, Rall LB, Rutter WJ. (1980) Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. *Nature* **286:**535-538.

110 Edmondson HA, Steiner PE. (1954) Primary Carcinoma of the liver: a study of 100 cases among 48,900 necropsies. *Cancer* **7:**462-503.

111 Elmileik H, Paterson AC, Kew MC. (2005) *Beta-catenin* mutations and expression, 249 serine p53 tumor suppressor gene mutation, and hepatitis B virus infection in southern African Blacks with hepatocellular carcinoma. *Journal of Surgical Oncology* **91:**258-263.

112 Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maaeda Y, Tsuruta K, Miyaki M, Nakamura Y. (1992) Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung. *Cancer Research* **52:**5368-5372.

113 Enders GH, Ganem D, Varmus HE. (1987) 5'-terminal sequences influence the segregation of ground squirrel hepatitis virus RNAs into polyribosomes and viral core particles. *Journal of Virology* **61:**35-41.

114 Erzberger JP, Berger JM. (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual Review of Biophysics and Biomolecular Structure* **35:**93-114.

115 Esumi M, Tanaka Y, Tozuka S, Shikata T. (1989) Clonal state of human hepatocellular carcinoma and non-tumorous hepatocytes. *Cancer Chemotherapy and Pharmacology* **23(Suppl):**S1-S3.

116 Etiemble J, Möröy T, Jacquemin E, Tiollais P, Buendia MA. (1989) Fused transcripts of *c-myc* and a new cellular locus, *hcr* in a primary liver tumor. *Oncogene* **4:**51-7.

117 Ferber MJ, Montoya DP, Yu C, Aderca I, McGee A, Thorland EC, Nagorney DM, Gostout BS, Burgart LJ, Boix L, Bruix J, McMahon BJ, Cheung TH, Chung TK, Wong YF, Smith DI, Roberts LR. (2003) Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (*hTERT*) gene in liver and cervical cancers. *Oncogene* **22:**3813-3820.

118 Folger K, Thomas K, Capecchi M. (1984) Analysis of homologous recombination in cultured mammalian cells. *Cold Spring Harbour Symposium on Quantitative Biology* **49:**123-138.

119 Fort P, Marty L, Piechaczyk M, el Sabrouty S, Dani C, Jeanteur P, Blanchard JM. (1985) Various rat tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multigenic family. *Nucleic Acids Research* **13:**1431-1432.

120 Fourel G, Trepo C, Bougueleret L, Henglein B, Ponzetto A, Tiollais P, Buendia MA. (1990) Frequent activation of *N-myc* genes by hepadnavirus insertion in woodchuck liver tumours. *Nature* **347:**294-298.

121 Fourel G, Transy C, Tennant BC, Buendia MA. (1992) Expression of the woodchuck N-myc2 retroposon in brain and in liver tumors is driven by a cryptic *N-myc* promoter. *Molecular and Cellular Biology* **12:**5336-5344.

122 Franks AL, Berg CJ, Kane MA, Browne BB, Sikes RK, Elsea WR, Burton AH. (1989) Hepatitis B infection among children born in the United States to Southeast Asian refugees. *The New England Journal of Medicine* **321:**1301-1305.

123 Fujimori M, Tokino T, Hino O, Kitegewa T, Imamura T, Okamoto E, Mitsunobu M, Ishikawa T, Nakagama H, Harada, Yagura M, Matsubara K, Nakamura Y. (1991) Allelotype study of primary hepatocellular carcinoma. *Cancer Research* **51:**89-93.

124 Fujimoto Y, Hampton LL, Wirth PJ, Wang N-J, Xie J-P, Thorgeirsson SS.(1994) Alterations of tumor suppressor genes and allelic losses in human hepatocellular carcinomas in China. *Cancer Research* **54:**281-285.

125 Fujiwara Y, Emi M, Ohata H, Kato Y, Nakajima T, Mori T, Nakamura Y. (1993) Evidence for the presence of two tumor suppressor genes on chromosome 8p for colorectal carcinoma. *Cancer Research* **53:**1172-1174.

126 Fünfschilling U, Saher G, Xiao L, Möbius W, Nave K-A. (2007) Survival of adult neurons lacking cholesterol sythesis in vivo. *BMC Neuroscience* **8:**1-9.

127 Galibert F, Mandart E, Fitoussi F, Tiollais P, Charnay P. (1979) Nucleotide Sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. *Nature* **281:**646-650.

128 Gao X, Wu N, Grignon D, Zacharek A, Liu H, Salkowshi A, Li G, Sakr W, Sarkar F, Porter AT, Chen YQ, Honn KV. (1994) High frequency of mutator phenotype in human prostatic adenocarcinoma. *Oncogene* **9:**2999 -3003.

129 Garcia M, de Thė H, Tiollais P, Samarut J, Dejean A. (1993) A hepatitis B virus pre-S-retinoic acid receptor beta chimera transforms erythrocytic progenitor cells in vitro. *Proceedings of the National Academy of Science of the United States of America* **90:**89-93.

130 Gerin JL. (1990) Experimental WHV infection of woodchucks: an animal model of hepadnavirus-induced liver cancer. *Gastroenterologia Japonica* **25:**38-42.

131 Gerlich WH, Robinson WS. (1980) Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. *Cell* **21:**801-809.

132 Gessler M, König A, Bruns GA. (1992) The genomic organisation and expression of the *WT*1 gene. *Genomics* **12:**807-813.

133 Giacchino R, Pontisso P, Navone C, Alberti A, Dini G, Facco F, Ciravegna B. (1987) Hepatitis B virus (HBV) DNA-positive hepatocellular carcinoma following hepatitis B virus infection in a child. *Journal of Medical Virology* **23:**151-155.

134 Goldstein ST, Alter MJ, Williams IT, Moyer LA, Judson FN, Mottram K, Fleenor M, Ryder PL, Margolis HS. (2002) Incidence and risk factors for acute hepatitis B in the United States, 1982-1988: implications for vaccination programs. *The Journal of Infectious Diseases* **185:**713-719.

135 Gong ZJ, De Meyer S, Clarysse C, Verslype C, Neyts J, De Clercq E, Yap SH. (1999) Mycophenolic acid, immunosuppressive agent, inhibits HBV replication in vitro. *Journal of Viral Hepatitis* **6:**229-236.

136 Gozuacik D, Murakami Y, Saigo K, Chami M, Mugnier C, Lagorce D, Okanoue T, Urashima T, Bréchot C, Paterlini-Bréchot P. (2001) Identification of human cancer related genes by naturally occurring hepatitis B virus DNA tagging. *Oncogene* **20:**6233-6240.

137 Gozuacik D, Chami M, Lagorce D, Faivre J, Murakami Y, Poch O, Biermann E, Knippers R, Bréchot C, Paterlini-Bréchot P. (2003) Identification and functional characterisation of a new member of the human Mcm protein family: hMcm8. *Nucleic Acids Research* **31:**570-9.

138 Graëf E, Caselmann WH, Wells J, Koshy R. (1994) Insertional activation of mevalonate kinase by hepatitis B virus DNA in a human hepatoma cell line. *Oncogene* **9:**81-87.

139 Grëthe S, Heckel J-O, Rietchel W, Hufert FT. (2000) Molecular epidemiology of hepatitis B virus variants in non-human primates. *The Journal of Virology* **74:**5377-5381.

140 Groisman IJ, Koshy R, Henkler F, Groopman JD, Alaoui-Jamali MA. (1999) Downregulation of DNA excision repair by the hepatitis B virus-x protein occurs in p53-proficient and p53-deficient cells. *Carcinogenesis* **20:**479-483.

141 Guidotti LG, Martinez V, Loh YT, Rogler CE, Chisari FV. (1994) Hepatitis B virus nucleocapsid particles do not cross the hepatocellular nuclear membrane in transgenic mice. *Journal of Virology* **68:**5469-5475.

142 Günther S, Sommer G, Von Breunig F, Iwanska A, Kalinina T, Sterneck M, Will H. (1998) Amplification of full-length hepatitis B virus genomes from samples from patients with low levels of viremia: frequency and functional consequences of PCR-introduced mutations. *Journal of Clinical Microbiology* **36:**531-538.

143 Guo W, Wang J, Tam G, Yen TS, Ou J. (1991) Leaky transcription termination produces larger and smaller than genome size hepatitis B virus X gene transcripts. *Virology* **181:**630-636.

144 Guo W, Chen M, Yen TS, Ou JH. (1993) Hepatocyte-specific expression of the hepatitis B virus core promoter depends on both positive and negative regulation. *Molecular and Cellular Biology* **13:**443-448.

145 Harbour JW, Lai S-I, Whang-Peng J, Gazdar AF, Minna JD, Kaye FJ. (1988) Abnormalities in structure and expression of the human *retinoblastoma* gene in SCLC. *Science* **241:**353-357.

146 Hartwell LH, Weinert TA. (1989) Checkpoints: controls that ensure the order of cell cycle events. *Science* **246:**629-634.

147 Hatada I, Tokino T, Ochiya T, Matsubara K. (1988) Co-amplification of integrated hepatitis B virus DNA and transforming gene *hst*-1 in a hepatocellular carcinoma. *Oncogene* **3:**537-540.

148 Hattori M, Hidaka S, Sakaki Y. (1985) Sequence analysis of a *Kpn*I family member near the 3' end of human *beta-globin* gene. *Nucleic Acids Research* **13:**7813-27.

149 Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH. (1984) Large surface proteins of hepatitis B virus containing the *pre-s* sequence. *Journal of Virology* **52:**396-402.

150 Henderson A. (1987) Chromosome accomodation to integration of foreign DNA. *Chromosomes Today* **9:**12-21.

151 Hino O, Kitagawa T, Koike K, Kobayashi M, Hara M, Mori W, Nakashima T, Hattori N, Sugano H. (1984) Detection of hepatitis B virus DNA in hepatocellular carcinomas in Japan. *Hepatology* **4:**90-95

152 Hino O, Shows TB, Rogler CE. (1986) Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17;18 translocation. *Proceedings of the National Academy of Sciences of the United States of America* **83:**8338-8342.

153 Hino O, Ohtake K, Rogler CE. (1989) Features of two hepatitis B virus (HBV) DNA integrations suggest mechanisms of HBV integration. *Journal of Virology* **63:**2638-2643.

154 Hino O, Kajino K. (1994) Hepatitis virus-related hepatocarcinogenesis. *Intervirology* **37:**133-135.

155 Hirohashi S. (1998) Inactivation of the E-cadherin-mediated cell adhesion system in human cancers. *American Journal of Pathology* **153:**333-339.

156 Hirsch DS, Pirone DM, Burbelo PD. (2001) A new family of Cdc42 effector proteins, CEPs, function in fibroblast and epithelial cell shape changes. *The Journal of Biological Chemistry* **276:**875-883.

157 Hirsch RC, Lavine JE, Chang L-J, Varmus HE, Ganem D. (1990) Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* **344:**552-555.

158 Hodinka RL. Laboratory Diagnosis of Viral Hepatitis. Chapter 8. *In: Viral Hepatitis, Diagnosis, Therapy, and Prevention*, Ed. Steven Specter. Humana Press Inc., New Jersey USA. 1999 p 200, 202.

159 Horikawa I, Barrett J. (2001) cis-Activation of the human telomerase gene (*hTERT*) by the hepatitis B virus genome. *Journal of the National Cancer Institute* **93:**1171-1173.

160 Hruska JF, Clayton DA, Rubenstein JL, Robinson WS. (1977) Structure of hepatitis B Dane particle DNA before and after the Dane particle DNA polymerase reaction. *Journal of Virology* **21:**666-672.

161 Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang N-J, Harris CC. (1991) Mutational hotspot in the *p53* gene in human hepatocellular carcinomas. *Nature* **350:**427-428.

162 Hsu T, Möröy T, Etiemble J, Louise A, Trepo C, Tiollais P, Buendia MA. (1988) Activation of *c-myc* by woodchuck hepatitis virus insertion in hepatocellular carcinoma. *Cell* **55:**627-635.

163 Huang J, Liang TJ. (1993) A novel hepatitis B virus (HBV) genetic element with Rev response element-like properties that is essential for expression of HBV gene products. *Molecular and Cellular Biology* **13:**7476-7486.

164 Huang LR, Hsu HC. (1995) Cloning and Expression of *CD24* Gene in Human Hepatocellular Carcinoma: A Potential Early Tumour Maker Gene Correlates with *p53* Mutation and Tumor Differentiation. *Cancer Research* 55:4717-4721.

165 Huang Z-M, Yen TS. (1994) Hepatitis B virus RNA element that facilitates accumulation of surface gene transcripts in the cytoplasm. *Journal of Virology* **68:**3193-3199.

166 Hübschen JM, Andernach IE, Muller CP. (in press) Hepatitis B virus genotype E variability in Africa. *Journal of Clinical Virology* (2008), doi:10.1016/j.jcv.2008.08.018

167 Hurie MB, Mast EE, Davis JP. (1992) Horizontal transmission of hepatitis B virus infection to United States-born children of Hmong refugees. *Pediatrics* **89:**269-273.

168 Hussain SA, Ferry DR, El-Gazzaz G, Mirza DF, James ND, McMAster P, Kerr DJ. (2001) Hepatocellular Carcinoma. *Annals of Oncology* **12:**161-172.

169 Iwatsuki S, Starzl TE, Sheahan DG, Yokoyama I, Demetris AJ, Todo S, Tzakis AG, Van Thiel DH, Carr B, Selby R, Madariaga J. (1991) Hepatic resection versus transplantation for hepatocellular carcinoma. *Annals of Surgery* **214:**221-228.

170 Jean-Jean O, Weimer T, de Recondo AM, Will H, Rossignol JM. (1989) Internal entry of ribosomes and ribosomal scanning involved in hepatitis B virus *P* gene expression. *Journal of Virology* **63:**5451-5454.

171 Jia L, Wang XW, Harris CC. (1999) Hepatitis B virus X protein inhibits nucleotide excision repair. *International Journal of Cancer* **80:**875-879.

172 Jilbert AR, Burrell CJ, Triatni M, Kamm M. (2002) Hepatitis B Virus Replication. In: *Hepatitis B Virus*. Lai and Locarnini Eds. International Medical Press 2002, London, UK, pp47.

173 Jilg W, Sieger E, Zachoval R, Schatzl H. (1995) Individuals with antibodies against hepatitis B core antigen as the only serological marker for hepatitis B infection: High percentage of carriers of hepatitis B and C virus. *Journal of Hepatology* **23:**14-20.

174 Junker-Niepmann M, Bartenschlager R, Schaller H. (1990) A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *The EMBO Journal* **9:**3389-3396.

175 Kalay E, Li Y, Uzumcu A, Uyguner O, Collin RW, Caylan R, Ulubil-Emiroglu M, Kersten FF, Hafiz G, van Wijk E, Kayserili H, Rohmann E, Wagenstaller J, Hoefsloot LH, Strom TM, Nürnberg G, Baserer N, den Hollander AI, Cremers FP, Cremers CW, Becker C, Brunner HG, Nürnberg P, Karaguzel A, Basaran S, Kubisch C, Kremer H, Wollnik B. (2006) Mutations in the lipoma HMGIC fusion partner-like 5 (*LHFPL5*) gene cause autosomal recessive nonsyndromic hearing loss. *Human Mutation* **27:**633-639.

176 Kanai Y, Ushijima S, Hui AM, Ochiai A, Tsuda H, Sakamoto M, Hirohashi S. (1997) The *E-cadherin* gene is silenced by CpG methylation in human hepatocellular carcinomas. *International Journal of Cancer* **71:**355-359.

177 Kann M. (2002) Structural and Molecular Virology. In: *Hepatitis B Virus*. Lai and Locarnini Eds. International Medical Press 2002, London, UK, pp14-20.

178 Kao JH, Chen DS. (2002) Recent research progress in hepatocellular carcinoma. *Journal of the Formosan Medical Association* **101:**239-248.

179 Katagiri T, Nakamura Y, Miki Y. (1996) Mutations in the *BRCA*2 gene in hepatocellular carcinomas. *Cancer Research* **56:**4575-4577.

180 Kato J, Kobune M, Kohgo Y, Sugawara N, Hisai H, Nakamura T, Sakamaki S, Sawada N, Niitsu Y. (1996) Hepatic iron deprivation prevents spontaneous development of fulminant hepatitis and liver cancer in Long-Evans Cinnamon rats. *Journal of Clinical Investigation* **4:**923-929.

181 Kawai S, Yokosuka O, Imazeki F, Maru Y, Saisho H. (2001) State of HBV DNA in HBsAg-negative, anti-HCV-positive hepatocellular carcinoma: existence of HBV DNA possibly as nonintegrated from with analysis by *Alu*-HBV DNA PCR and conventional HBV PCR. *Journal of Medical Virology* **64:**410-418.

182 Kawate S, Fukusato T, Ohwada S, Watanuki A, Morishita Y. (1999a) Amplification of *c-myc* in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and *p53* overexpression. *Oncology* **57:**157-163.

183 Kawate S, Takenoshita S, Ohwada S, Mogi A, Fukusato T, Makita F, Kuwano H, Morishita Y. (1999b) Mutation analysis of transforming growth factor beta type II receptor, Smad2, amd Smad4 in hepatocellular carcinoma. International Journal of *Oncology* **14:**127-131.

184 Kazachkov Y, Yoffe B, Khaoustov VI, Solomon H, Klintmalm GB, Tabor E. (1998) Microsatellite instability in human hepatocellular carcinoma: relationship to *p53* abnormalities. *Liver* **18:**156-161.

185 Kekulé AS, Lauer U, Meyer M, Caselmann WH, Hofschneider PH, Koshy R. (1990) The *preS2/S* region of integrated hepatitis B virus DNA encodes a transcriptional transactivator. *Nature* **343:**457-461.

186 Kekulé AS, Lauer U, Weiss L, Luber B, Hofschneider PH. (1993) Hepatitis B virus transactivator HBx uses a tumour promoter signalling pathway. *Nature* **361:**742-745.

187 Kew MLC, Kasdsianides C, Berger EL, Song E, Dusheikko GM. (1987) Prevalence of chronic hepatitis B virus infection in pregnant black women living in Soweto. *Journal Medical Virology* **22:**263-268.

188 Kew MC. (1992) Chronic hepatitis B virus infection and hepatocellular carcinoma in Africa. *SA Journal of Science* **88:**524-528.

189 Kew MC, Miller RH, Chen HS, Tennant BC, Purcell RH. (1993) Mutant woodchuck hepatitis virus genomes from virions resemble rearranged hepadnaviral integrants in hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **90:**10211-10215.

190 Kew MC. (2002a) Hepatitis B virus as a cause of hepatocellular carcinoma, with particular reference to sub-Saharan Africa. *South African Journal of Epidemiology and Infection* **17:**34-39.

191 Kew MC. (2002b) Hepatitis B virus in the etiology of hepatocellular carcinoma. In: *Viruses and Liver Cancer*. E. Tabor (editor). Elsevier Science B.V. Amsterdam, The Netherlands. p17-29.

192 Kew MC, Kramvis A, Yu MC, Arakawa K, Hodkinson J. (2005) Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu speaking sub-Saharan Africans. *Journal of Medical Virology* **75:**513-521.

193 Kew MC, Asare GA. (2007) Dietary iron overload in the African and hepatocellular carcinoma. *Liver International* **6:**735-741.

194 Khudyakow YE, Makhov AM. (1989) Prediction of terminal protein and ribonuclease H domains in the *P* gene product of hepadnaviruses. *FEBS Letters* **243:**115-118.

195 Kidd-Ljundggren K, Oberg M, Kidd AH. (1995) The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. *Journal of General Virology* **76:**2119-2130.

196 Kiehl TR, Shibata H, Pulst SM. (2000) The ortholog of human *ataxin-2* is essential for early embryonic patterning in C. elegans. *Journal of Molecular Neuroscience* **15:**231-241.

197 Kim CM, Koike K, Saito I, Miyamura T, Jay G. (1991) HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature* **351:**317-320.

198 Kim JH, Rho HM. (2002) Activation of the human transforming growth factor alpha (*TGF-alpha*) gene by the hepatitis B viral X protein (HBx) through AP-2 sites. *Molecular and Cellular Biochemistry* **231:**155-61.

199 Kim M-Y, Park E, Park J-H, Park D-H, Moon W-S, Cho B-H, Shin H-S, Kim D-G. (2001a) Expression profile of nine novel genes differentially expressed in hepatitis B virus-associated hepatocellular carcinomas. *Oncogene* **20:**4568-4575.

200 Kim SH, Hong SP, Kim SK, Lee WS, Rho HM. (1992) Replication of a mutant hepatitis B virus with a fused *X-C* reading frame in hepatoma cells. *Journa of General Virology* **73:**2421-2424.

201 Kim YC, Song KS, Yoon G, Nam MJ, Ryu WS. (2001b) Activated *ras* oncogene collaborates with HBx gene of hepatitis B virus to transform cells by suppressing HBx-mediated apoptosis. *Oncogene* **20:**16-23.

202 Kimbi GC, Kramvis A, Kew MC. (2004) Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *Journal of General Virology* **85:**1211-1220.

203 Kimbi GC, Kramvis A, Kew MC. (2005) Integration of hepatitis B virus DNA into chromosomal DNA during acute hepatitis B. *World Journal of Gastroenterology 7:*6415-6421.

204 Kirk GD, Lesi OA, Mendy M, Akano AO, Sam O, Goedert JJ, Hainaut P, Hall AJ, Whittle H, Montesano R. (2004) The Gambia liver Cancer study: Infection with hepatitis B and C and the risk of hepatocellular carcinoma in West Africa. *Hepatology* **39:**211-219.

205 Kita R, Nishida N, Fukuda Y, Azechi H, Matsuoka Y, Komeda T, Sando T, Nakao K, Ishizaki K. (1996) Infrequent alterations of the *p161INK*4A gene in liver cancer. *International Journal of Cancer* **67:**176-180.

206 Klein NP, Schneider RJ. (1997) Activation of Src family kinases by hepatitis B virus HBx protein and coupled signalling to ras. *Molecular and Cellular Biology* **17:**6427-6436.

207 Ko LJ, Prives C. (1996) *p53*: puzzle and paradigm. *Genes and Development* **10:**1054-1072.

208 Kobayashi S, Saigoh K, Urashima T, Asano T, Isono K. (1997) Detection of hepatitis B virus x transcripts in human hepatocellular carcinoma tissues. *Journal of Surgical Research* **73:**97-100.

209 Koch I, Hofschneider PH, Koshy R. (1989) Expression of a hepatitis B virus transcript containing fused mitochondrial-like domains in human hepatoma cells. *Virology* **170:**591-594.

210 Koch S, Freytag von Loringhoven A, Kahmann R, Hofschneider PH, Koshy R. (1984a) The genetic organisation of integrated hepatitis B virus DNA in the human hepatoma cell line PLC/PRF/5. *Nucleic Acids Research* **12:**6871-6886.

211 Koch S, Freytag von Loringhoven A, Hofschneider PH, Koshy R. (1984b) Amplification and rearrangement in hepatoma cell DNA associated with integrated hepatitis B virus DNA. *The EMBO Journal* **3:**2185-2189.

212 Koike K, Kobayashi M, Mizusawa H, Yoshida E, Yaginuma K, Taira M. (1983) Rearrangement of the surface antigen gene of hepatitis B virus integrated in the human hepatoma cell lines. *Nucleic Acids Research* **11:**5391-5402.

213 Koike K, Moriya K, Iino S, Yotsuyanagi H, Endo Y, Miyammura T, Hurakawa K. (1994) High level expression of hepatitis B virus HBx gene and hepatocarcinogenesis in transgenic mice. *Hepatology* **19:**810-819.

214 Koike K, Kobayashi M, Gondo M, Hayashi I, Osuga T, Takada S. (1998) Hepatitis B virus DNA is frequently found in liver biopsy samples from hepatitis C virus-infected chronic hepatitis patients. *Journal of Medical Virology* **54:**249-255.

215 Kopnin PB, Agapova LS, Kopnin BP, Chumakov PM. (2007) Repression of sestrin family genes contributes to oncogenic *Ras*-induced reactive oxygen species up-regulation and genetic instability. *Cancer Research* **67:**4671-4678.

216 Korba BE, Wells FV, Baldwin B, Cote PJ, Tennant BC, Popper H, Gerin JL. (1989) Hepatocellular carcinoma in woodchuck hepatitis virus-infected woodchucks: presence of viral DNA in tumor tissue from chronic carriers and animals serologically recovered from acute infections. *Hepatology* **9:**461-470.

217 Koshy R, Maupas P, Muller R, Hofschneider PH. (1981) Detection of hepatitis B virus-specific DNA in genomes of human hepatocellular carcinoma and liver cirrhosis tissues. *Journal of General Virology* **57:**95-102.

218 Koshy R, Koch S, Freytag von Loringhoven A, Kahmann R, Murray K, Hofschneider PH. (1983) Integration of hepatitis B virus DNA: evidence for integration in the single-stranded gap. *Cell* **34:**215-223.

219 Kramvis A, Bukofzer S, Kew MC. (1996) Comparison of Hepatitis B Virus DNA extractions from serum by the QIAamp blood kit, GeneReleaser, and the phenol-chloroform method. *Journal of Clinical Microbiology* **34:**2731-2733.

220 Kramvis A and Kew MC. (1998) Structure and function of the encapsidation signal of hepadnaviridae. *Journal of Viral Hepatitis* **5:**357-367.

221 Kramvis A, Kew MC, Francois G. (2005a) Hepatitis B virus genotypes. *Vaccine* **23:**2409-2423.

222 Kramvis A, Restorp K, Norder H, Botha JF, Magnius LO, Kew MC. (2005b) Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and Madagascar reveals low genetic variability. *Journal of Medical Virology* **77:**47-52.

223 Kramvis A, Kew MC. (2007) Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatology Research* **37:**S9-S19.

224 Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. (2008) Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *Journal of Medical Virology* **80:**27-46.

225 Kuo K-W, Yang P-Y, Huang Y-S, Shieh D-Z. (1998) Variations in gene expression and genomic stability of human hepatoma cells integrated with hepatitis B virus DNA. *Biochemistry and Molecular Biology International* **44:**1133-1140.

226 Kuroki T, Fujiwara Y, Nakamori S, Imaoka S, Kanematsu T, Naiamura Y. (1995) Evidence for the presence of two tumour-suppressor genes for hepatocellular carcinoma on chromosome 13q. *British Journal of Cancer* **72:**383-385.

227 Kwok S, Higuchi R. (1989) Avoiding false positives with PCR. *Nature* **339:**237-238.

228 Kwon JA, Rho HM. (2003) Transcriptional repression of the human *p53* gene by hepatitis B viral core protein (HBc) in human liver cells. *The Journal of Biological Chemistry* **384:**203-212.

229 Lai C-L, Locarnini S, Kann M, Shih C, Jilbert A, Burrell CJ, Triatni M, Delaney WE IV, Isom HC, Zoulim F, Le Guerhier F, Seignères B, Mason W, Litwin S, Farrell J, Dienstag J, Goodman Z, Bowden S, Kao J-H, Chen D-S, Lau GKK, Gish RG, Liang R, Kew M, Yuen M-F, Thompson S. (2002) Inner front cover from: *Hepatitis B Virus*. Lai and Locarnini Eds. International Medical Press 2002, London, UK.

230 Landers TA, Greenberg HB, Robinson WS. (1977) Structure of hepatitis B Dane particle DNA and nature of the endogenous DNA polymerase reaction. *Journal of Virology **23:**368-76.*

231 Lanford RE, Chavez D, Brasky KM, Burns RB, Rico-Hesse R. (1998) Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proceedings of the National Academy of Science USA* **95:**5757-5761.

232 Lanford RE, Kim YH, Lee H, Notvall L, Beames B. (1999) Mapping of the hepatitis B virus reverse transcriptase TP and RT domains by transcomplementation for nucleotide priming and by protein-protein interaqction. *Journal of Virology* **73:**1885-1893.

233 Lanford RE, Chavez D, Rico-Hesse R, Mootnick A. (2000) Hepadnavirus infection in captive gibbons. *Journal of virology* **74:**2955-2959.

234 Lauder IJ, Lin H-J, Lau JYN, Siu T-S, Lai C-L. (1993) The variability of the hepatitis B virus genome: statistical analysis and biological implications. *Molecular Biology and Evolution* **10:**457-470.

235 Lee DS, Huh K, Lee EH, Lee DH, Hong KS, Sung YC. (1997) HCV and HBV coexist in HBsAg-negative patients with HCV viraemia: possibility of coinfection in these patients must be considered in HBV-high endemic area. *Journal of Gastroenterology and Hepatology* **12:**855-861.

236 Lee EY-H, Lee P, To H, Shew J-Y, Bookstein R, Scully P, Lee W-P. (1988) Inactivation of the *retinoblastoma* susceptibility gene in human breast cancers. *Science* **241:**218-221.

237 Lee TH, Elledge SJ, Butel JS. (1995) Heptitis B virus X protein interacts with a probable cellular DNA repair protein. *Journal of Virology* **69:**1107-1114.

238 Lee SG, Rho HM. (2000) Transcriptional repression of the human *p53* gene by hepatitis B viral X protein. *Oncogene* **19:**468-471.

239 Leistner CM, Gruen-Bernhard S, Glebe D. (2008) Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cell Microbiology* **10:**122-133.

240 Levine AJ, Perry ME, Chang A, Silver A, Dittmer D, Wu M, Welsh D. (1994) The 1993 Walter Hubert Lecture: The role of the *p53* tumor-suppressor gene in tumorigenesis. *British Journal of Cancer* **69:**409-416.

241 Levrero M, Balsano C, Natoli G, Avantaggiati ML, Elfassi E. (1990) Hepatitis B virus X protein transactivates the long terminal repeats of the human immunodeficiency virus types 1 and 2. *Journal of Virology* **64:**3082-3086.

242 Lipowsky G, Bischoff FR, Schwarzmaier P, Kraft R, Kostka S, Hartmann E, Kutay U, Görlich D. (2004) Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. *EMBO Journal* **19:**4362-4371.

243 Lim MH, Kim HW, Paik KC, Cho SC, Yoon DY, Lee HJ. (2006) Association of the DAT1 polymorphism with attention deficit hyperactivity disorder (ADHD): a family-based approach. *American Journal of Medical Genetics: Part B, Neuropsychiatric Genetics* **141:**309-311.

244 Lin Y, Nomura T, Cheong J, Dorjsuren D, Iida K, Murakami S. (1997) Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. *Journal of Biological Chemistry 272:*7132-7139.

245 Lin Y, Tang H, Nomura T, Dorjsuren D, Hayashi N, Wei W, Ohta T, Roeder R, Murakami S. (1998) The hepatitis B virus X protein is a co-activator of activated transcription that modulates the transcription machinery and distal binding activators. *Journal of Biological Chemistry* **273:**27097-27103.

246 Lo SJ, Chien ML, Lee YH. (1988) Conserved HBV *X* gene. Characteristic of the *X* gene of hepatitis B virus. *Virology* **167:**289-292.

247 Lopez-Cabrera M, Letovsky J, Hu KQ, Siddiqui A. (1990) Multiple liver-specific factors bind to the hepatitis B virus core pregenomic promoter: trans-activation and repression by CCAAT/enhancer binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **87:**5069-5073.

248 Lowe SW, Ruley HE, Jacks T, Housman DE. (1993) *p53*-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74:**957-967.

249 Lurman A. [Eine icterusepdemie.]. (1885) *Berlin Klinische Wochenschrift* **22:**20-23. **Original Article in German. Cited in:** MacCullum FO. (1971) 1971 International Symposium on Viral Hepatitis. Historical Perspectives. *Cancer Medical Association Journal* Special Issue 106:423-426.

250 Lustbader ED, Hann HW, Blumberg BS. (1983) Serum ferritin as a predictor of host response to hepatitis B virus infection. *Science* **220:**423-425.

251 MacCallum F, Bauer D. (1947) Homologous serum hepatitis. *Lancet* **ii:**691-692.

252 MacCullum FO. (1971) 1971 International Symposium on Viral Hepatitis. Historical Perspectives. *Cancer Medical Association Journal* Special Issue 106:423-426.

253 Magnius LO, Norder H. (1995) Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the *S*-gene. *Intervirology* **38:**24-34.

254 Mahlapuu M, Pelto-Huikko M, Aitola M, Enerbäck S, Carlsson P. (1998) FREAC-1 contains a cell-type-specific transcriptional activation domain and is expressed in epithelial-mesenchymal interfaces. *Developmental Biology* 15:183-195. *Erratum* in: *Developmental Biology* (1999), **207:**476.

255 Maïga I, Le Faou A, Muller CP, Venard V. (2005) Unexpected high prevalence of hepatitis B and HIV infections in Malian medical students. *European Journal of Clinuical Microbiology and Infectious Disease* **24:**501-502.

256 Mandahl N, Höglund M, Mertens F, Rydholm A, Willén H, Brosjö O, Mitelman F. (1994) Cytogenetic aberrations in 88 benign and borderline adipose tissue tumors. *Genes Chromosomes Cancer* **9:**207-215.

257 Maniatis T, Fritsch EF, Lauer J, Lawn RM. (1981) Molecular genetics of human haemoglobins. *Annual Reviews in Genetics* **14:**145-178.

258 Marion PL, Oshiro L, Regnery DC, Scullard GH, Robinson WS. (1980a) A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *Proceedings of the National Academy of Sciences of the United States of America* **77:**2941-2945.

259 Marion PL, Salazar FH, Alexander JJ, Robinson WS. (1980b) State of hepatitis B viral DNA in a Human Hepatoma Cell Line. *Journal of Virology* **33:**795-806.

260 Marion PL, Knight SS, Salazar FH, Popper H, Robinson WS. (1983) Ground squirrel hepatitis virus infection. *Hepatology* **3:**519-27.

261 Marion PL, Van Davelaar MJ, Knight SS, Salazar FH, Garcia G, Popper H, Robinson WS. (1986) Hepatocellular carcinoma in ground squirrels persistently infected with ground squirrel hepatitis virus. *Proceedings of the National Academy of Sciences of the United States of America* **83:**4543-4546.

262 Marusawa H, Uemoto S, Hijikata M, Ueda Y, Tanaka K, Shimotohno K, Chiba T. (2000) Latent hepatitis B virus infection in healthy individuals with antibodies to hepatitis B core antigen. *Hepatology* **31:**488-495.

263 Maruta Y, Okayama N, Hiura M, Suehiro Y, Hirai H, Hinoda Y. (2008) Determination of ancestral allele for possible human cancer-associated polymorphisms. *Cancer Genetics Cytogenetics* **180:**24-29.

264 Mason WS, Seal G, Summers J. (1980) Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *Journal of Virology* **36:**829-36.

265 Mathet VL, Cuestas ML, Ruiz V, Minassian ML, Rivero CW, Trinks J, Daleoso G, León LM, Sala A, Corach D, Oubiña JR. (2005) Characterization of two hepatitis B virus isolates ascribed to the E genotype in Argentina: an unusual detection in a family of African descendants even in presence of high titers of anti-HBs antibodies. (Unpublished)

266 Matsubara K, Tokino T. (1990) Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. *Molecular Biology and Medicine* **7:**243-260.

267 Matsumoto Y, Fujii H, Matsuda M, Kono H. (2001) Multicentric occurrence of hepatocellular carcinoma: diagnosis and clinical significance. *Journal of Hepatobiliary and Pancreatic Surgery* **8:**435-440.

268 Mendy M, D'Mello F, Kanellos T, Oliver S, Whittle H, Howard CR. (2008) Envelope protein variability among HBV-infected asymptomatic carriers and immunized children with breakthrough infections. *Journal Of Medical Virology* **80:**1537-1546.

269 Meyer M, Wiedorn KH, Hofschneider PH, Koshy R, Caselmn WH. (1992) A chromosome (17:7) translocation is associated with a hepatitis B virus integration in human hepatocellular carcinoma DNA. *Hepatology* **15:**665-671.

270 Michalak TI, Pasquinelli C, Guilhot S, Chisari FV. (1994) Hepatitis B virus persistence after recovery from acute viral hepatitis. *Journal of Clinical Investigations* **93:**230-239.

271 Michel M-L, Pontisso P, Sobczak E, Malpièce Y, Streeck RE, Tiollais P. (1984) Synthesis in animals cells of hepatitis B surface antigen particles carrying a receptor for polymerized human serum albumin. *Proceedings of the National Academy of Science USA* **81:**7708-7712.

272 Milich DR, McNamara MK, McLachlan A, Thornton GB, Chisari FV. (1985) Distinct *H-2*-linked regulation of T-cell responses to the *pre-S* and *S* regions of the same hepatitis B surface antigen polypeptide allows circumvention of nonresponsiveness to the S region. *Proceedings of the National Academy of Science USA* **82:**8168-8172.

273 Miller RH, Robinson WS. (1983) Integrated hepatitis B virus DNA sequences specifying the major viral core polypeptide are methylated in PLC/PRF/5 cells. *Proceedings of the National Academy of Sciences of the United States of America* **80:**2534-2538.

274 Miller RH, Robinson WS. (1984) Hepatitis B virus DNA forms in nuclear and cytoplasmiic fractions of infected human liver. *Virology* **137:**390-399.

275 Miller RH, Lee SC, Liaw YF, Robinson WS. (1985) Hepatitis B virus DNA in infected human liver and in hepatocellular carcinoma. *Journal of Infectious Diseases* **151:**10814-1092.

276 Minami M, Poussin K, Bréchot C, Paterlini P. (1995) A novel PCR technique using *Alu*-specific primers to identify unknown flanking sequences from the human genome. *Genomics* **29:**403-408.

277 Minami M, Daimon Y, Mori K, Takashima H, Nakajima T, Itoh Y, Okanoue T. (2005) Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis. *Oncogene* **24:**4340-4348.

278 Miyanohara A, Imamura T, Araki M, Sugawara K, Ohtomo N, Matsubara K. (1986) Expression of hepatitis B virus core antigen gene in *Saccharomyces cerevisiae*: synthesis of two polypeptides translated from different initiation codons. *Journal of Virology* **59:**176-180.

279 Miyoshi Y, Iwao K, Nagasawa Y, Aihara T, Sasaki Y, Imaoka S, Murata M, Shimano T, Nakamura Y. (1998) Activation of the *beta-catenin* gene in primary hepatocellular carcinomas by somatic alterations involving exon 3. *Cancer Research* **58:**2524-2547.

280 Mizokami M, Orito E, Ohba K, Ikeo K, Lau JYN, Gojobori T. (1997) Constrained evolution with respect to gene overlap of hepatitis B virus. *Journal of Molecular Evolution* **44:**S83-S90.

281 Mizusawa H, Taira M, Yaginuma K, Kobayashi M, Yoshida E, Koike K. (1985) Inversely repeated integrated hepatitis B virus DNA and cellular flanking sequences in the human hepatoma-derived cell line huSP. *Proceedings of the National Academy of Sciences of the United States of America* **82:**208-212.

282 Montell C, Fisher EF, Caruthers MH, Berk AJ. (1983) Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. *Nature* **305:**600-605.

283 Moolla N, Kew M, Arbuthnot P. (2002) Regulatory elements of hepatitis B virus transcription. *Journal of Viral Hepatology* **9:**323-331.

284 Moradpour D, Blum HE. (2005) Pathogenesis of hepatocellular carcinoma. European *Journal of Gastroenterology and Hepatology* **17:**477-83.

285 Moraes EC, Keyse SM, Pidoux M, Tyrrell RM. (1989) The spectrum of mutations generated by passage of a hydrogen peroxide damaged shuttle vector plasmid through a mammalian host. *Nucleic Acids Research* **17:**8301-8312.

286 Moraes EC, Keyse SM, Tyrrell RM. (1990) Mutagenesis by hydrogen peroxide treatment of mammalian cells: a molecular analysis. *Carcinogenesis* **11:**283-293.

287 Möröy T, Marchio A, Etiemble J, Trépo C, Tiollais P, Buendia MA. (1986) Rerrangement and enhanced expression of *c-myc* in hepatocellular carcinomma of hepatitis virus infected woodchucks. *Nature* **324:**276-279.

288 Möröy T, Etiemble J, Bougueleret L, Hadchouel M, Tiollais P, Buendia MA. (1989) Structure and expression of *hcr*, a locus rearranged with *c-myc* in a woodchuck hepatocellular carcinoma. *Oncogene* **4:**59-65.

289 Mulders MN, Venard V, Njayou M, Edorh AP, Bola Oyefolu AO, Kehinde MO, Muyembe Tamfum JJ, Nebie YK, Maïga I, Ammerlaan W, Fack F, Omilabu SA, Le Faou A, Muller CP. (2004) Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. *Journal of Infectious Diseases* **190:**400-408.

290 Murakami Y, Minami M, Daimon Y, Okanoue T. (2004) Hepatitis B virus DNA in liver, serum and peripheral blood mononuclear cells after the clearance of serum hepatitis B virus surface antigen. *Journal of Medical Virology* **72:**203-214.

291 Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Bréchot C, Paterlini-Bréchot P. (2005) Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* **54:**1162-1168.

292 Nagai H, Pineau P, Tiollais P, Buendia AM, Dejean A. (1997) Comprehensive allelotyping of human hepatocellular carcinoma. *Oncogene* **14:**2927-2933.

293 Nagao T, Inoue S, Yoshimi F, Sodeyama M, Omori Y, Mizuta T, Kawano N, Morioka Y. (1990) Postoperative recurrence of hepatocellular carcinoma. *Annals of Surgery* **211:**28-33.

294 Nagaya T, Nakamura T, Tokino T, Tsurimoto T, Imai M, Mayumi T, Kamino K, Yamamura K, Matsubara K. (1987) The mode of hepatitis B virus DNA integration in chromosome of human hepatocellular carcinoma. *Genes and Development* **1:**773-782.

295 Nakabayashi H, Taketa K, Miyano K, Yamane T, Sato J. (1982) Growth of human hepatoma cell lines with differentiated functions in chemically defined medium. *Cancer Research* **42:**3858-3863.

296 Nakamura T, Tokino T, Nagaya T, Matsubara K. (1988) Microdeletion associated with the integration process of hepatitis B virus DNA. *Nucleic Acids Research* **16:**4865-4873.

297 Nakamura T, Iwamura Y, Kaneko M, Nagawa K, Kawai K, Mitamura K, Futagawa T, Hayashi H. (1991) Deletions and rearrangements of the *retinoblastoma* gene in hepatocellular carcinoma, insulinoma and some neurogenic tumors are found in a study of 121 tumors. *Japanese Journal of Clinical Oncology* **21:**325-329.

298 Nassal M, Junker-Niepmann M, Schaller H. (1990) Translational inactivation of RNA function; discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell* **63:**1357-1363.

299 Nassal M, Schaller H. (1996) Hepatitis B virus replication - an update. *Journal of Viral Hepatitis* **3:**217-226.

300 Nazarewicz T, Krawczynski K, Slusarczyk J, Nowoslawski A. (1977) Cellular localization of hepatitis B virus antigens in patients with hepatocellular carcinoma coexisting with liver cirrhosis. *Journal of Infectious Diseases* **135:**298-302.

301 Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT. (1989) *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proceedings of the National Academy of Scienc eof the United States of America* **86:**6686-6690.

302 Nelson DL, Ballabio A, Victoria MF, Pieretti M, Bies RD, Gibbs RA, Maley JA, Chinault AC, Webster TD, Caskey CT. (1991) *Alu*-primer polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: Identification of clones associated with a disease locus. *Proceedings of the National Academy of Science of the United States of America* **88:**6157-6161.

303 Neurath AR, Kent SB, Strick N, Taylor P, Stevens CE. (1985) Hepatitis B virus contains the *pre-S* gene-encoded domains. *Nature* **315:**154-156.

304 Nevins, JR. (1983) The pathway of eukaryotic mRNA formation. *Annual Reviews in Biochemistry* **52:**441-466.

305 Ni Y-H, Chang M-H, Huang L-M, Chen H-L, Hsu Y, Chiu T-Y, Tsai K-S, Chen D-S. (2001) Hepatitis B virus infection in children and adolescents in a hyperendemic area: 15 years after mass hepatitis B vaccination. *Annals of Internal Medicine* **135:**796-800.

306 Ni YH, Huang LM, Chang MH, Yen CJ, Lu CY, You SL, Kao JH, Lin YC, Chen HL, Hsu HY, Chen DS. (2007) Two decades of universal hepatitis B vaccination in Taiwan: impact and implication for future strategies. *Gastroenterology* **132:**1287-93.

307 Nijhara R, Jana SS, Goswami SK, Rana A, Majumdar SS, Kumar V, Sarkar DP. (2001) Sustained activation of mitogen-activated protein kinases and activator protein 1 by the hepatitis B virus X protein in mouse hepatocytes *in vivo*. *Journal of Virology* **75:**10348-10358.

308 Nishida N, Fukuda Y, Komeda T, Kita R, Sando T, Furukawa M, Amenomori M, Shibagaki I, Nakao K, Ikenaga M, Ishizaki K. (1994) Amplification and overexpression of the *cyclin D*1 gene in aggressive human hepatocellular carcinoma. *Cancer Research* **54:**3107-3110.

309 Ng KM, Yan TD, Black D, Chu FC, Morris DL. (2009) Prognostic determinants for survival after resection/ablation of a large hepatocellular carcinoma. *HPB (Oxford)* **11:**311-320.

310 Norder H, Hammas B, Löfdahl S, Couroucé A-M, Magnius LO. (1992a) Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *Journal of General Virology* **73:**1201-1208.

311 Norder H, Couroucé A-M, Magnius LO. (1992b) Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *Journal of General Virology* **73:**3141-3145.

312 Norder H, Ebert JW, Fields HA, Mushahwar IK, Magnius LO. (1996) Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* **218:**214-223.

313 Nose H, Imazeki F, Ohto M, Omata M. (1993) *p53* gene mutations and 17p allelic deletions in hepatocellular carcinoma from Japan. *Cancer* **72:**355-360.

314 Novelletto A, Persichetti F, Sabbadini G, Mandich P, Bellone E, Ajmar F, Squitieri F, Campanella G, Bozza A, MacDonald ME, Gusella JF, Frontali M. (1994) Polymorphism analysis of the huntingtin gene in Italian families affected with Huntington disease. *Humna Molecular Genetics* **3:**1129-1132.

315 Odemuyiwa SO, Mulders MN, Oyedele OI, Ola SO, Odaibo GN, Olaleye DO, Muller CP. (2001) Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa. *Journal of Medical Virology* **65:**463-469.

316 Ogata N, Tokino T, Kamimura T, Asakura H. (1990) A comparison of the molecular structure of integrated hepatitis B virus genomes in hepatocellular carcinoma cells and hepatocytes derived from the same patient. *Hepatology* **11:**1017-1023.

317 Ogston CW, Jonak GJ, Rogler CE, Astrin SM, Summers J. (1982) Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from hepatocellular carcinomas of woodchucks. *Cell* **29:**385-394.

318 Ohno H, Kaneko S, Kobayashi k, Murakami S. (1997) Human hepatitis B virus enhancer 1 is responsive to human interleukin-6. *Journal of Medical Virology* **52:**413-418.

319 Okamoto H, Tsuda F, Sakugawa H, Sastrosoewinjo RI, Imai M, Miyakawa Y, Mayumi M. (1988) Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *Journal of General Virology* **69:**2575-2583.

320 Okochi K, Murakami S. (1968) Observations on Australia antigen in Japanese. *Vox Sanguinis* **15:**374-385.

321 Olinger CM, Weber B, Otegbayo JA, Ammerlaan W, van der Taelem-Brulé N, Muller CP. (2007) Hepatitis B virus genotype E surface antigen detection with different immunoassays and diagnostic impact of mutations in the *preS/S* gene. *Medical Microbiology and Immunology* **196:**247-252.

322 Ono Y, Onda H, Sasada R, Igarashi K, Sugino Y, Nishioka K. (1983) The complete nucleotide sequences of the cloned hepatitis B virus DNA: Subtype adr and adw. *Nucleic Acids Research* **11:**1747-1757.

323 Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe K, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, Hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase TO, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakebe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Yamazaki M, Watanabe K, Kumagai A, Itakura S, Fukuzumi Y, Fujimori Y, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senba T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Watanabe M, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Nakamura Y, Ohara O, Isogai T, Sugano S. (2004) Complete sequencing and characterisation of 21,243 full-length human cDNAs. *Nature* Genetics **36:**40-45

324 Ou J-H, Rutter WJ. (1985) Hybrid hepatitis B virus-host transcripts in a human hepatoma cell. *Proceedings of the National Academy of Sciences of the United States of America* **82:**83-87.

325 Ou J, Laub O, Rutter W. (1986) Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. *Proceedings of the National Academy of Sciences of the United States of America* **83:**1578-1582.

326 Outram SV, Gordon AR, Hager-Theodorides AL, Metcalfe J, Crompton T, Kemp P. (2008) KLF13 influences multiple stages of both B and T cell development. Epub. *Cell Cycle* **7:**2047-2055.

327 Owiredu WK, Kramvis A, Kew MC. (2001) Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B surface antibody alone: possible association with recombination between genotypes A and D. *Journal of Medical Virology* **64:**441-454.

328 Ozturk M. (1999) Genetic aspects of hepatocellular carcinogenesis. *Seminars in Liver Disease* **19:**235-242.

329 Parkin DM, Sitas F, Chirenje M, Stein L, Abratt R, Wabinga H. (2008) Part I: Cancer in Indigenous Africans--burden, distribution, and trends. *The Lancet Oncology* **9:**683.692.

330 Paterlini P, Poussin K, Kew M, Franco D, Bréchot C. (1995) Selective accumulation of the *X* transcript of hepatitis B virus in patients negative for hepatitis B surface antigen with hepatocellular carcinoma. *Hepatology* **21:**313-321.

331 Paterlini-Bréchot P, Saigo K, Murakami Y, Chami M, Gozuacik D, Mugnier C, Lagorce D, Bréchot C. (2003) Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human *telomerase* gene. *Oncogene* **22:**3911-3916.

332 Peifer M. (1997) *Beta-catenin* as oncogene: the smoking gun. *Science* **275:**1752-1753.

333 Persing DH, Varmus HE, Ganem D. (1985) A frameshift mutation in the *pre-S* region of the human hepatitis B virus genome allows production of surface antigen particles but eliminates binding to polymerized albumin. *Proceedings of the National Academy of Sciences of the United States of America* **82:**3440-3444.

334 Piao Z, Choi Y, Park C, Lee WJ, Park JH, Kim H. (1997a) Deletion of the *M6P/IGF2r* gene in primary hepatocellular carcinoma. *Cancer Letters* **120:**39-43.

335 Piao Z, Kim H, Jeon B, Lee WJ, Park C. (1997b) Relationship between loss of heterozygosity of tumour suppressor genes and histologic differentiation in hepatocellular carcinoma. *Cancer* **80:**865-872.

336 Pineau P, Marchio A, Terris B, Mattei M-G, Tu Z-X, Tiollais P, Dejean A. (1996) A t(3;8) chromosomal translocation associated with hepatitis B virus integration involves the carboxypeptidase N locus. *Journal of Virology* **70:**7280-7284.

337 Pollicino T, Raffa G, Costantino L, Lisa A, Campello C, Squadrito G, Levrero M, Raimondo G. (2007) Molecular and functional analysis of occult hepatitis B virus isolates from patients with hepatocellular carcinoma. *Hepatology* **45:**277-285.

338 Poussin K, Dienes H, Sirma H, Urban S, Beaugrand M, Franco D, Schirmacher P, Bréchot C, Bechot PP. (1999) Expression of mutated hepatitis B virus *X* genes in human hepatocellular carcinomas. *International Journal of Cancer* **80:**497-505.

339 Prince AM. (1968) An antigen detected in the blood during the incubation period of serum hepatitis. *Proceedings of the National Academy of Sciences of the United States of America* **60:**814-821.

340 Puisieux A, Ponchel F, Ozturk M. (1993) *p53* as a growth suppressor gene in HBV-related hepatocellular carcinoma cells. *Oncogene* **8:**487-490.

341 Puisieux A, Ozturk M. (1997) TP53 and hepatocellular carcinoma. *Pathologie-Biologie (Paris)* **45:**864-870.

342 Qiu XB, Shao YM, Miao S, Wang L. (2006) The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences* **63:**2560-2570.

343 Quade K, Saldanha J, Thomas H, Monjardino J. (1992) Integration of hepatitis B virus DNA through a mutational hot spot within the cohesive region in a case of hepatocellular carcinoma. *Journal of General Virology* **73:**179-182.

344 Quintana A, Molinero A, Borup R, Nielsen FC, Campbell IL, Penkowa M, Hidalgo J. (2008) Effect of astrocyte-targeted production of IL-6 on traumatic brain injury and its impact on the cortical transcriptome. *Developmental Neuroscience* **68:**195-208.

345 Radziwill G, Tucker W, Schaller H. (1990) Mutational analysis of the hepatitis B virus *P* gene product: domain structure and RNase H activity. *Journal of Virology* **64:**613-620.

346 Raney AK, Milich DR, Easton AJ, McLachlan A. (1990) Differentiation-specific transcriptional regulation of the hepatitis B virus large surface antigen gene in human hepatoma cell lines. *Journal of Virology* **64:**2360-2368.

347 Raney AK, Johnson JL, Palmer CN, McLachlan A. (1997) Members of the nuclear receptor superfamily regulate transcription from the hepatitis B virus nucleocapsid promoter. *Journal of Virology* **71:**1058-1071.

348 Redeker AG. (1975) Viral hepatitis: clinical aspects. *American Journal of Medical Science* **270:**9-16.

349 Rho HM, Kim K, Hyun SW, Kim YS. (1989) The nucleotide sequence and reading frames of a mutant hepatitis B virus subtype *adr*. *Nucleic Acids Research* **17:**2124.

350 Roberts AV, Weatherall DJ, Clegg JB. (1972) The synthesis of human haemoglobin A_2 during erythroid maturation. *Biochemical and Biophysical Research Communications* **47:**81-87.

351 Robinson WS, Greenman RL. (1974) DNA polymerase in the core of the human hepatitis B virus candidate. *Journal of Virology* **13:**1231-1236.

352 Robinson WS, Clayton DA, Greenman RL. (1974) DNA of a human hepatitis B virus candidate. *Journal of Virology* **14:**384-391.

353 Robinson WS, Lutwick LI. (1976) The virus of hepatitis, type B (First of two parts). *The New England Journal of Medicine* **295:**1168-1175.

354 Robinson WS, Lutwick LI. (1976) The virus of hepatitis, type B. (Second of two parts). *The New England Journal of Medicine* **295:**1232-1236.

355 Robinson WS. (1992) The role of hepatitis B virus in the development of primary hepatocellular carcinoma: Part I. *Journal of Gastroenterology and Hepatology* **7:**622-638.

356 Robinson WS. (1994) Molecular events in the pathogenesis of hepadnavirus-associated hepatocellular carcinoma. *Annual Review of Medicine* **45:**297-323.

357 Rogler CE, Summers J. (1982) Novel forms of woodchuck hepatitis virus DNA isolated from chronically infected woodchuck liver nuclei. *Journal of Virology* **44:**852-863.

358 Rogler CE, Summers J. (1984). Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from a chronically infected liver. *Journal of Virology* **50:**832-837.

359 Rogler CE, Sherman M, Su CY, Shafritz DA, Summers J, Shows TB, Henderson A, Kew M. (1985) Deletion in chromosome 11p associated with a hepatitis B integration site in hepatocellular carcinoma. *Science* **230:**319-322.

360 Russnak R, Ganem D. (1990) Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes and Development* **4:**764-776.

361 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. *Human Mutation* **29:**703-708.

362 Sambrook J, Fritsch EF, Maniatis T. (1989*) Molecular Cloning: A Laboratory Manual.* Second Edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New york, USA. p6.2; 9.14-9.19; 9.32.

363 Sattler F, Robinson WS. (1979) Hepatitis B viral DNA molecules have cohesive ends. *Journal of Virology* **32:**226-233.

364 Schlüter V, Meyer M, Hofschneider PH, Koshy R, Caselmann WH. (1994) Integrated hepatitis B virus *X* and 3' truncated *preS/S* sequences derived from human hepatomas encode functionally active transactivators. *Oncogene* **9:**3335-3344.

365 Schödel FT, Weimer T, Will H, Sprengel R. (1988) Amino acid sequence similarity between retroviral and *E. coli* RNase H and hepadnaviral gene products. *AIDS Research and Human Retroviruses* **4:**9-11.

366 Schröder CH, Zentgraf H. (1990) Hepatitis B virus related hepatocellular carcinoma: chronicity of infection - the opening to different pathways of malignant transformation? *Biochimica et biophysica acta* **1032:**137-156.

367 Scuric Z, Stain SC, Anderson WF, Hwang JJ. (1998) New member of aldose reductase family proteins overexpressed in human hepatocellular carcinoma. *Hepatology* **27:**943-950.

368 Seifer M, Standring DN. (1995) Assembly and Antigenicity of Hepatitis B Virus Core Particles. *Intervirology* **38:**47-62.

369 Sells MA, Chen M-L, Acs G. (1987) Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. *Proceedings of the National Academy of Science of the United States of America* **84:**1005-1009.

370 Serrano M, Hannon GJ, Beach D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of *cyclin D/CDK4*. *Nature* **366:**704-707.

371 Shafritz DA, Shouval D, Sherman HI, Hadziyannis SJ, Kew MC. (1981) Integration of hepatitis B virus DNA into the genome of liver cancer cells in chronic liver disease and hepatocellular carcinoma. *The New England Journal of Medicine* **305:**1067-1073.

372 Shafritz DA, Kew HC. (1981) Identification of integrated hepatitis B virus DNA sequences in human hepatocellular carcinomas. *Hepatology* **1:**1-8.

373 Shamay M, Barak O, Doitsh G, Ben-Dor I, Shaul I. (2002) Hepatitis B virus pX interacts with HBXAP, a PHD finger protein to coactivate transcription. *Journal of Biological Chemistry* **277:**9982-9988.

374 Sharma N, Low SH, Misra S, Pallavi B, Weimbs T. (2006) Apical targeting of syntaxin 3 is essential for epithelial cell polarity. *The Journal of Cell Biology* **173:**837-848.

375 Shaul Y, Ziemer M, Garcia P, Crawford R, Hsu H, Valenzuela P, Rutter WJ. (1984) Cloning and analysis of integrated hepatitis virus sequences from a human hepatoma cell line. *Journal of Virology* **51:**776-787.

376 Shaul Y, Rutter WJ, Laub O. (1985) A human hepatitis B viral enhancer element. *The EMBO Journal* **4:**427-430.

377 Sheu JC, Huang G-T, Chou H-C, Lee P-H, Wang J-T, Lee H-S, Chen D-S. (1993) Multiple hepatocellular carcinomas at the early stage have different clonality. *Gastroenterology* **105:**1471-1476.

378 Sheu JC, Lin Y-W, Chou H-C, Huang G-T, Lee H-S, Lin Y-H, Huang S-Y, Chen C-H, Wang J-T, Lee P-H, Lin J-T, Lu FJ, Chen DS. (1999) Loss of heterozygosity and microsatellite instability in hepatocellular carcinoma in Taiwan. *British Journal of Cancer* **80:**468-476.

379 Shih C, Burke K, Chou MJ, Zeldis JB, Yang CS, Lee CS, Isselbacher KJ, Wands JR, Goodman HM. (1987) Tight clustering of human hepatitis B virus integration sites in hepatomas near a triple-stranded region. *Journal of Virology* **61:**3491-3498.

380 Shih CC, Stoye JP, Coffin JM. (1988) Highly preferred targets for retrovirus integration. *Cell* **53:**531-537.

381 Shih WL, Kuo ML, Chuang SE, Cheng AL, Doong SL. (2000) Hepatitis B virus X protein inhibits transforming growth factor beta-induced apoptosis through the activation of phosphatidylinositol 3-kinase pathway. *The Journal of Biological Chemistry* **275:**25858-25864.

382 Simon D, Knowles BB, Weith A. (1991) Abnormalities of chromosome 1 and loss of heterozygosity on 1p in primary hepatomas. *Oncogene* **6:**765-770.

383 Simon D, Carr BI. (1995) Integration of hepatitis B virus and alteration of the 1p36 region found in cancerous tissue of primary hepatocellular carcinoma with viral replication evidenced only in noncancerous, cirrhotic tissue. *Hepatology* **22:**1393-1398.

384 Simoyama Y, Hirohashi S, Hirano S, Noguchi M, Shimosato Y, Takeichi M, Abe O. (1989) Cadherin cell-adhesion molecules in human epithelial tissues and carcinomas. *Cancer Research* **49:**2128-2133.

385 Slagle BL, Zhou Y-Z, Butel JS. (1991) Hepatitis B virus integration event in human chromosome 17p near the *p53* gene identifies the region of the chromosome commonly deleted in virus-positive hepatocellular carcinomas. *Cancer Res*earch **51:**49-54.

386 Slagle BL, Zhou Birchmeier W, Scorsone KA. (1993) Deletion of the *E-cadherin* gene in hepatitis B virus-positive Chinese hepatocellular carcinomas. *Hepatology* **18:**757-762.

387 Slagle BL, Lee T, Medina D, Finegold M, Butle J. (1996) Increased sensitivity to the hepatocarcinogen diethylnitrosamine in transgenic mice carrying the hepatitis B virus *X* gene. *Molecular Carcinogenesis* **15:**261-269.

388 Smela ME, Hamm ML, Henderson PT, Harris CM, Harris TM, Essigmann JM. (2002) The aflatoxin B_1 formamidopyrimidine adduct plays a major role in cauring the types of mutations observed in human hepatocellular carcinoma. *Proccedings of the National Academy of Science USA* **99:**6655-6660.

389 Sninsky JJ, Siddiqui A, Robinson WS, Cohen SN. (1979) Cloning and endonuclease mapping of the hepatitis B viral genome. *Nature* **279:**346-348.

390 Someya T, Ikeda K, Saitoh S, Kobayashi M, Hosaka T, Sekazi H, Akuta N, Suzuki F, Suzuki Y, Arase Y, Kumada H. (2006) Interferon lowers tumor recurrence rate after surgical resection or ablation of hepatocellular carcinoma: a pilot study of patients with hepatitis B virus-related cirrhosis. *Journal of Gastroenterology* **41:**1206-1213.

391 Sommer G, Günther S, Sterneck M, Otto S, Will H. (1997) A new class of defective hepatitis B virus genomes with an internal poly(dA) sequence. *Virology* **239:**402-412.

392 Soret J, Gattoni R, Guyon C, Sureau A, Popielarz M, Le Rouzic E, Dumon S, Apiou F, Dutrillaux B, Voss H, Ansorge W, Stévenin J, Perbal B. (1998) Characterisation of SRp46, a novel human SR splicing factor encoded by a PR264/SC35 retropseudogene. *Molecular and Cellular Biology* **18:**4924-4934.

393 Southern EM. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98:**503-517.

394 Souto-Carneiro, M.M. and Krenn,V. (unpublished) Homo sapiens mRNA for immunoglobulin heavy chain variable region, clone KC4. Genbank Accession No. AJ007320, Version AJ007320.1 GI:3250843. Direct Submission. Submitted (16-JUN-1998) Souto-Carneiro M.M., Pathologisches Institut, Wuerzburg University, Josef-Schneider-Str. 2, D- 97080 Wuerzburg, GERMANY.

395 Spandau DF, Lee CH. (1988) Trans-activation of viral enhancers by the hepatitis B virus X protein. *Journal of Virology* **62:**427-434.

396 Sprengel R, Kaleta EF, Will H. (1988) Isolation and characterisation of a hepatitis B virus endemic in herons. *Journal of Virology* **62:**3832-3839.

397 Standring DN, Rutter WJ, Varmus HE, Ganem D. (1984) Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. *Journal of Virology* **50:**563-571.

398 Stevens CE, Neurath RA, Beasley RP, Szmuness W. (1979) HBeAg and anti-HBs detection by radioimmunoassay: correlation with vertical transmission of hepatitis B virus in Taiwan. *Journal of Medical Virology* **3:**237-241.

399 Stibbe W, Gerlich WH. (1983) Structural relationships between minor and major proteins of hepatitis B surface antigen. *Journal of Virology* **46:**626-628.

400 Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schnazi RF, Rossau R. (2000) A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *Journal of General Virology* **81:**67-74.

401 Su JM, Lai XM, Lan KH, Li CP, Chao Y, Yen SH, Chang FY, Lee SD, Lee WP. (2007) X protein of hepatitis B virus functions as a transcriptional corepressor on the human telomerase promoter. *Hepatology* **46:**402-413.

402 Summers J, O'Connell A, Millman I. (1975) Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Proceedings of the National Academy of Sciences of the United States of America* **72:**4597-4601.

403 Summers J, Smolec JM, Snyder R. (1978) A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proceedings of the National Academy of Sciences of the United States of America* **75:**4533-4537.

404 Summers J. (1981) Three recently described animal virus models for human hepatitis B virus. *Hepatology* **1:**179-183.

405 Summers J, Mason WS. (1982) Replication of the genome of a hepatitis B—like virus by reverse transcription of an RNA intermediate. *Cell* **29:**403-415.

406 Swenson PD, Riess JT, Krueger LE. (1991) Determination of HBsAg subtypes in different high risk populations using monoclonal antibodies. *Journal of Virological Methods* **33:**27-28.

407 Szmulewicz MN, Novick GE, Herrera RJ. (1998) Effects of *Alu* insertions on gene function. *Electrophoresis* **19:**1260-1264.

408 Szmuness W, Much MI, Prince AM, Hoofnagle JH, Cherubin CE, Harley EJ, Block GH. (1975) On the role of sexual behaviour in the spread of heptitis B infection. *Annals Internal Medicine* **83:**489-495.

409 Szmuness W. (1978) Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. *Progress in Medical Virology* **24:**40-69.

410 Takada S, Gotoh Y, Hayashi S, Yoshida M, Koike K. (1990) Structural rearrangement of integrated hepatitis B virus DNA as well as cellular flanking DNA is present in chronically infected hepatic tissues. *Journal of Virology* **64:**822-828.

411 Takahashi K, Kudo K, Ishibashi H, Hirata Y, Niho Y. (1993) Frequent loss of heterozygosity on chromosome 22 in hepatocellularr carcinoma. *Hepatology* **17:**794-799.

412 Takahashi K, Brotman B, Usuda S, Mishiro S, Prince AM. (2000) Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: Implications for an origin of HBV. *Virology* **267:**58-64.

413 Takeda J, Ishii S, Seino Y, Imamoto F, Imura H. (1989) Negative regulation of human insulin gene expression by the 5'-flanking region in non-pancreatic cells. *FEBS Letters* **247:**41-5.

414 Takeichi M. (1991) Cadherin cell adhesion receptors as a morphogenetic regulator. Science 251:1451-1455.

415 Takeuchi M, Fujimoto J, Niwamoto H, Yamamoto Y, Okamoto E. (1997) Frequent deletion of hepatitis B virus *X*-gene DNA in hepatocellular carcinoma and adjacent liver tissue in hepatitis B surface antigen-negative patients. *Digestive Diseases and Sciences* **42:**2264-2269.

416 Tamura S, Nakamori S, Kuroki T, Yo S, Furukawa H, Ishikawa O, Imaoka S, Nakamura Y. (1997) Association of cumulative allelic losses with tumour aggressiveness in hepatocellular carcinoma. *Journal of Hepatology* **27:**669-676.

417 Tanaka T, Miyamoto H, Hino O, Kitagawa T, Koike M, Iizuka T, Sakamoto H, Ooshima A. (1986) Primary hepatocellular carcinoma with hepatitis B virus-DNA integration in a 4-year-old boy. *Human Pathology* **17:**202-204.

418 Tavis JE, Massey B, Gong Y. (1998) The duck hepatitis B virus polymerase is activated by its RNA packaging signal, epsilon. *Journal of Virology* **72:**5789-5796.

419 Terradillos O, Billet O, Renard C, Levy R, Molina T, Briand P, Buendia M. (1997) The hepatitis B virus *X* gene potentiates *c-myc* induced liver oncogenesis in transgenic mice. *Oncogene* **14:**395-404.

420 Tiollais P, Pourcel C, Dejean A. (1985) The Hepatitis B virus. *Nature (London)* **317:**489-495.

421 Toh H, Hayashida H, Miyata T. (1983) Sequence homology between retroviral reverse transcriptase and putative polymerase of hepatitis B virus and cauliflower mosaic virus. *Nature* (London) **305:**827-829.

422 Tokino T, Fukushige S, Nakamura T, Nagaya T, Murotsu T, Shiga K, Aoki N, Matsubara K. (1987) Chromosomal translocation and inverted duplication associated with integrated hepatitis B virus in hepatocellular carcinomas. *Journal of Virology* **61:**3848-3854.

423 Tokino T, Tamura H, Hori N, Matsubara K. (1991) Chromosome deletions associated with hepatitis B virus integration. *Virology* **185:**879-882.

424 Tokino T, Matsubara K. (1991) Chromosomal sites for hepatitis B virus integration in human hepatocellular carcinoma. *Journal of Virology* **65:**6762-6764.

425 Tong MJ, Weiner JM, Ashcavai MW, Vyas GN. (1979) Evidence for clustering of hepatitis B virus infection in families of patients with primary hepatocellular carcinoma. *Cancer* **44:**2338-2340.

426 Toyoda H, Kumada T, Kaneoka Y, Murakami Y. (2008) Impact of hepatitis B virus (HBV) *X* gene integration in liver tissue on hepatocellular carcinoma development in serologically HBV-negative chronic hepatitis C patients. *Journal of Hepatology* **48:**43-50.

427 Transy C, Fourel G, Robinson WS, Tiollais P, Marion PL, Buendia MA. (1992) Frequent amplification of *c-myc* in ground squirrel liver tumors associated with past or ongoing infection with a hepadnavirus. *Proceedings of the National Academy of Sciences of the United States of America* **89:**3874-8.

428 Trepicchio WL, Krontiris TG. (1993) IGH minisatellite suppression of USF-binding-site and Eu-mediated transcriptional activation of the adenovirus major late promoter. *Nucleic Acids Research* **21:**977-985.

429 Trowbridge R, Fagan EA, Davison F, Eddleston ALWF, Williams R, Linskens MHK, Farzaneh F. (1988) Amplification of the *c-myc* Gene Locus in a Human Hepatic Tumour Containing Integrated Hepatitis B Virus DNA. In: *Viral Hepatitis and Liver Disease*, ed. AJ Zuckerman, pp764-768. New York: Liss.

430 Tsebe KV, Burnett RJ, Hlungwani NP, Sibara MM, Venter PA, Mphahlele MJ. (2001) The first five years of universal hepatitis B vaccination in South Africa: evidence for elimination of HBsAg carriage in under 5-year-olds. *Vaccine* **19:**3919-26.

431 Tsuda H, Zhang W, Shimosato Y, Yokota J, Terada M, Sugimura T, Miyamura T, Hirohashi S. (1990) Allelic loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **87:**6791-6794.

432 Tsuei D-J, Hsu T-Y, Chen J-Y, Chang M-H, Hsu H-C, Yang C-S. (1994) Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in a childhood hepatocellular carcinoma. *Journal of Medical Virology* **42:**287-293.

433 Tsuei DJ, Chang MH, Chen PJ, Hsu TY, Ni YH. (2002) Characterisation of integration patterns and flanking cellular sequences of hepatitis B virus in childhood hepatocellular carcinomas. *Journal of Medical Virology* **68:**513-521.

434 Tu H, Bonura C, Giannini C, Mouly H, Soussan P, Kew M, Paterlini-Bréchot P, Bréchot C, Kremsdorf D. (2001) Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Research* **61:**7803-7810.

435 Tur-Kaspa R, Burk R, Shaul Y, Shafritz D. (1986) Hepatitis B virus DNA contains a glucocorticoid-responsive element. *Proceedings of the National Academy of Sciences of the United States of America* **83:**1627-1631.

436 Urban M, McMillan DJ, Canning G, Newell A, Brown E, Mills JS, Jupp R. (1998) In vitro activity of hepatitis B virus polymerase: requirement for distinct metal ions and the viral epsilon stem-loop. *Journal of General Virology* **79:**1121-1131.

437 Vairaktaris E, Serefoglou ZC, Yapijakis C, Vassiliou S, Nkenke E, Avgoustidis D, Vylliotis A, Stathopoulos P, Neukam FW, Patsouris E. (2007) The platelet glycoprotein Ibalpha VNTR polymorphism is associated with risk for oral cancer. *Anticancer Research* **27:**4121-4125.

438 Van Herck K, Van Damme P. (2008) Benefits of early hepatitis B immunization programs for newborns and infants. *Pediatric Infection and Disease Journal* **27:**861-869.

439 Vasan RS, Larson MG, Aragam J, Wang TJ, Mitchell GF, Kathiresan S, Newton-Cheh C, Vita JA, Keyes MJ, O'Donnell CJ, Levy D, Benjamin EJ. (2007) Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. *BMC Medical Genetics* 8(Suppl 1):S2.

440 Vaudin M, Wolstenholme AJ, Tsiquaye KN, Zuckerman AJ, Harrison TJ. (1988) The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee. *Journal of General Virology* **69:**1383-1389.

441 Villa E, Grottola A, Buttqafoco P, Trande P, Merighi A, Fratti N, Seium Y, Cioni G, Manenti F. (1995) Evidence for hepatitis B virus infection in patients with chronic hepatitis C with and without serological markers of hepatitis B. *Digestive Diseases and Sciences* **40:**8-13.

442 Vleminckx K, Vakaet L Jr., Mareel M, Fiers W, Van Roy F. (1991) Genetic manipulation of *E-cadherin* expression by epithelial tumour cells reveals an invasion suppressor role. *Cell* **66:**107-119.

443 von Loringhoven AF, Koch S, Hofschneider PH, Koshy R. (1985) Co-transcribed 3' host sequences augment expression of integrated hepatitis B virus DNA. *The EMBO Journal* **4:**249-255.

444 Vyas GN, Rao KR, Ibrahim AB. (1972) Australia antigen (hepatitis B antigen): a conformational antigen dependent on disulfide bonds. *Science* **178:**1300-1301.

445 Vyas GN, Yen TSB. Hepatitis B Virus: Biology, Pathogenesis, Epidemiology, Clinical Description, and Diagnosis. Chapter 3. *In: Viral Hepatitis, Diagnosis, Therapy, and Prevention*, Ed. Steven Specter. Humana Press Inc., New Jersey USA. 1999 p 39, 53.

446 Wands JR. (2004) Prevention of Hepatocellular Carcinoma. *The New England Journal of Medicine* **351:**1567-1570.

447 Wang G-H, Zoulim F, Leber EH, Kitson J Seeger C. (1994a) Role of RNA in enzymatic activity of the reverse transcriptase of hepatitis B viruses. *Journal of Virology* **68:**8437-8442.

448 Wang HD, Trivedi A, Johnson DL. (1997) HepatitisB virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signalling pathway. *Molecular and Cellular Biology* **17:**6838-6846.

449 Wang HP, Rogler CE. (1988) Deletion in human chromosome arms 11p and 13q in primary hepatocellular carcinomas. *Cytogenetics Cell Genetics* **48:**72-78.

450 Wang HP, Rogler CE. (1991) Topoisomerase I-mediated integration of hepadnavirus DNA in vitro. *Journal of Virology* **65:**2381-2392.

451 Wang J. (1985) DNA topoisomerases. *Annual Reviews in Biochemistry* **54:**665-697.

452 Wang J, Chenivesse X, Henglein B, Bréchot C. (1990) Hepatitis B virus integration in a *cyclin A* gene in a hepatocellular carcinoma. *Nature* **343:**555-557.

453 Wang J, Zindy F, Chenivesse X, Lamas E, Henglein B, Bréchot C. (1992) Modification of *cyclin A* expression by hepatitis B virus DNA integration in a hepatocellular carcinoma. *Oncogene* **7:**1653-1656.

454 Wang W, London WT, Feitelson MA. (1991) Hepatitis B x antigen in hepatitis B virus carrier patients with liver cancer. *Cancer Research* **51:**4971-4977.

455 Wang XW, Forrester K, Yeh H, Feitelson MA, Gu J, Harris CC. (1994 b) Hepatitis B virus X protein inhibits p53 sequence-specific DNA binding, transcriptional activity, and association with transcription factor ERCC3. *Proceedings of the National Academy of Sciences of the United States of America* **91:**2230-2234.

456 Wang XW, Yeh H, Schaeffer L, Roy R, Moncollin V, Egly JM, Wang Z, Freidberg EC, Evans MK, Taffe BG, Bohr VA, Weeda G, Hoeijmakers JH, Forrester K, Harris CC. (1995) p53 modulation of TFIIH-associated nucleotide excision repair activity. *Nature Genetics* **10:**188-195.

457 Wang Y, Lau SH, Sham JS-T, Wu M-C, Wang T, Guan X-Y. (2004) Characterisation of HBV integrants in 4 hepatocellular carcinomas: association of truncated *X* gene and hepatocellular carcinogenesis. *Oncogene* **23:**142-148.

458 Warren KS, Heeney JL, Swan RA, Heeriyanto, Verschoor EJ. (1999) A new group of hepadnaviruses naturally infecting orangutans (Pongo pygmaeus). *Journal of Virology* **73:**7860-7865.

459 Weinberg M, Passman M, Kew M, Arbuthnot P. (2000) Hammerhead ribozyme-mediated inhibition of hepatitis B virus *X* gene expression in cultured cells. *Journal of Hepatology* **33:**142-151.

460 Weinberg RA. (1995) The retinoblastoma protein and cell cycle control. *Cell* **81:**323-330.

461 Wick M, Bürger C, Funk M, Müller R. (1995) Identification of a novel mitogen-inducible gene (*mig-6*): regulation during G1 progression and differentiation. *Experimental Cell Research* **219:**527-35.

462 Will H, Reiser W, Weimer T, Pfaff E, Buscher M, Sprengel R, Cattaneo R, Schaller H. (1987) Replication strategy of human hepatitis B virus. *Journal of Virology* **61:**904-11.

463 Williams JCP, Barratt-Boyes BG, Lowe JB. (1961) Supravalvular aortic stenosis. *Circulation* **24:**1311-1318.

464 Wollersheim M, Debelka U, Hofschneider PH. (1988) A transactivating function encoded in the hepatitis B virus *X* gene is conserved in the integrated state. *Oncogene* **3:**545-552.

465 Xu ZY, Francis DP, Liu CB, Purcell RH, Duan SC, Chen Rj, Wen YM. (1985) Prevention of hepatitis B virus carrriage of infants using HBV vaccine in Shanghai. Preliminary report of a randomized double-blind placebo-controlled trial. *Chinese Medical Journal (Engl)* **98:**623-626.

466 Yaginuma K, Kobayashi H, Yoshida E, Kobayashi M, Koike K. (1984) Direct evidence for the expression of integrated hepatitis B virus DNA in human hepatoma tissue. *Gann*. (Japanese Journal of Cancer Research) **75:**743-6.

467 Yaginuma K, Kobayashi M, Yoshida E, Koike K. (1985) Hepatitis B virus integration in hepatocellular carcinoma DNA: Duplication of cellular flanking sequences at the integration site. *Proceedings of the National Academy of Sciences of the United States of America* **82:**4458-4462.

468 Yaginuma K, Shirakata Y, Kobayashi M, Koike K. (1987a) Hepatitis B virus (HBV) particles are produced in a cell culture system by transient expression of transfected HBV DNA. *Proceedings of the National Academy of Sciences of the United States of America* **84:**2678-2682.

469 Yaginuma K, Kobayashi H, Kobayashi M, Morishima T, Matsuyama K, Koike K. (1987b) Multiple integration site of hepatitis B virus DNA in hepatocellular carcinoma and chronic active hepatitis tissues from children. *Journal of Virology* **61:**1808-1813.

470 Yaginuma K, Nakamura I, Takada S, Koike K. (1993) A transcription initiation site for the hepatitis B virus *X* gene is directed by the promoter-binding protein. *Journal of Virology* **67:**2559-2565.

471 Yakicier MC, Irmak MB, Romano A, Kew M, Ozturk M. (1999) *Smad*2 and *Smad*4 gene mutations in hepatocellular carcinoma. *Oncogene* **18:**4879-4883.

472 Yasui M, Harada A, Nonami T, Takeuchi Y, Taniguchi K, Nakao A, Takagi H. (1997) Potentially multicentric hepatocellular carcinoma: clinicopathologic characteristics and postoperative prognosis. *World Journal of Surgery* **21:**860-864.

473 Ye B, Shen J, Xu Y. (1994) Etiologic study on the relationship between HBV, HCV, and HCC. *Zhonghua Liu Xing Bing Xue Za Zhi* **15:**131-134. (Article in Chinese)

474 Yee J. (1989) A liver-specific enhancer in the core promoter region of human hepatitis B virus. *Science* **246:**658-670.

475 Yeh S-H, Chen P-J, Chen H-L, Lai M-Y, Wang C-C, Chen DS. (1994) Frequent genetic alterations at the distal region of chromosome 1p in human hepatocellular carcinomas. *Cancer Research* **54:**4188-4192.

476 Yotsuyanagi H, Yasuda K, Iino S, Moriya K, Shintani Y, Fujie H, Tsutsumi T, Kimura S, Koike K. (1998) Persistent viremia after recovery from self-limited acute hepatitis B. *Hepatology* **27:**1377-1382.

477 Yu DY, Moon HB, Son JK, Jeong S, Yu SL, Yoon H, Han YM, Lee CS, Park JS, Lee CH, Hyun BH, Murakami S, Lee KK. (1999) Incidence of hepatocellular carcinoma in transgenic mice expressing the hepatitis B virus X-protein. *Journal of Hepatology* **31:**123-132.

478 Yu X, Mertz JE. (1997) Differential regulation of the *pre-C* and pregenomic promoters of human hepatitis B virus by members of the nuclear receptor superfamily. *Journal of Virology* **71:**9366-9374.

479 Yuh CH, Chang YL, Ting LP. (1992) Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus. *Journal of Virology* **66:**4073-4084.

480 Yumoto Y, Hanafusa T, Hada H, Morita T, Ooguchi S, Shinji N, Mitani T, Hamaya K, Koide N, Tsuji T. (1995) Loss of heterozygosity and analysis of mutation of *p53* in hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology* **10:**179-185.

481 Zaka R, Stokes D, Dion AS, Kusnierz A, Han F, Williams CJ. (2006) P5L mutation in Ank results in an increase in extracellular inorganic pyrophosphate during proliferation and nonmineralizing hypertrophy in stably transduced ATDC5 cells. *Arthritis Research & Therapy* **8:**R164.

482 Zarkower D, Stephenson P, Sheets M, Wickens M. (1986) The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA in vitro. *Molecular and Cellular Biology* **6:**2317-2323.

483 Zerial M, Salinas J, Filipski J, Bernardi G. (1986) Genomic localization of hepatitis B virus in a human hepatoma cell line. *Nucleic Acids Research* **14:**8373-8386.

484 Zhong S, Chan JY, Teo W, Tam JS, Johnson PJ. (2000) Frequent integration of precore/core mutants of hepatitis B virus in human hepatocellular carcinoma tissues. *Journal of Viral Hepatitis* **7:**115.

485 Zhou YZ, Slagle BL, Donehower LA, van Tuinen P, Ledbettter DH, Butel JS. (1988) Structural analysis of a hepatitis B virus genome integrated into chromosome 17p of a human hepatocellular carcinoma. *Journal of Virology* **62:**4224-4231.

486 Zhu AX. (2003) Hepatocellular carcinoma: are we making progress? *Cancer Investigations* **21:**418-428.

487 Zhu W, Liang P. (1997) Detection and isolation of differentially expressed genes by differential display. In: Boultwood J. (ed) Methods in Molecular Biology Volume 86: Gene isolation and mapping protocols. Humana Press Inc., Totowa, NJ. Chapter 15:211-220.

488 Ziemer M, Garcia P, Shaul Y, Rutter WJ. (1985) Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. *Journal of Virology* **53:**885-892.

489 Zody MC, Garber M, Sharpe T, Young SK, Rowen L, O'Neill K, Whittaker CA, Kamal M, Chang JL, Cuomo CA, Dewar K, FitzGerald MG, Kodira CD, Madan A, Qin S, Yang X, Abbasi N, Abouelleil A, Arachchi HM, Baradarani L, Birditt B, Bloom S, Bloom T, Borowsky ML, Burke J, Butler J, Cook A, DeArellano K, DeCaprio D, Dorris L 3rd, Dors M, Eichler EE, Engels R, Fahey J, Fleetwood P, Friedman C, Gearin G, Hall JL, Hensley G, Johnson E, Jones C, Kamat A, Kaur A, Locke DP, Madan A, Munson G, Jaffe DB, Lui A, Macdonald P, Mauceli E, Naylor JW, Nesbitt R, Nicol R, O'Leary SB, Ratcliffe A, Rounsley S, She X, Sneddon KM, Stewart S, Sougnez C, Stone SM, Topham K, Vincent D, Wang S, Zimmer AR, Birren BW, Hood L, Lander ES, Nusbaum C. (2006) Analysis of the DNA sequence and duplication history of human chromosome 15. *Nature* **440:**671-675.

490 Zoulim F, Saputelli J, Seeger C. (1994) Woodchuck hepatitis virus X protein is required for viral infection in vivo. *Journal of Virology* **68:**2026-2030.

WEBSITES ACCESSED

WHO. http://www.who.int. Hepatitis B.World Health Organization Department of Communicable diseases Surveillance Reponse.WHO/CDS/CSR/LYO/2002.2: Hepatitis B. 2002. Accessed 11 June 2008.

NCBI

References:

http://www.ncbi.nlm.nih.gov/sites/entrez Accessed 2001- 2009.

Gene searches:

http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore&itool=toolbar Accessed 2001- 2009.

Primer design Websites:

http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-primer.html Accessed 2002-2004.

Sequencing data analysis software:

Sequences:

CHROMAS 145.exe, Free of charge from:

http://www.technelysium.com.au/chromas14x.html Accessed 2004-2006.

Alignments: Gendoc (current version 1.0-beta5). Free of charge from:

http://gendiapo.sourceforge.net Accessed 2004-2009.

Genbank:

http://www.clcbio.com/index.php?id=442 Accessed 2004-2009.

RNA Quantitation:

Protocol Online:

http://www.protocol-online.org/prot/u7General_Laboratory_Techniques/Quantitation/RNA Quantitation/ Accessed 5 July 2004.

RNA Isolation and Quantification Protocol:

http://www.ucc.ie/ucc/depts/physio/MolPhysiol/MolPhys/ Protocols/pdfs/001.pdf Accessed 5 July 2004.