Accepted Manuscript

HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant

William S. Mason, Upkar S. Gill, Samuel Litwin, Yan Zhou, Suraj Peri, Oltin Pop, Michelle L.W. Hong, Sandhia Naik, Alberto Quaglia, Antonio Bertoletti, Patrick T.F. Kennedy

PII: S0016-5085(16)34808-9
DOI: 10.1053/j.gastro.2016.07.012
Reference: YGAST 60585

To appear in: *Gastroenterology*
Accepted Date: 7 July 2016

Please cite this article as: Mason WS, Gill US, Litwin S, Zhou Y, Peri S, Pop O, Hong MLW, Naik S, Quaglia A, Bertoletti A, Kennedy PTF, HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant, *Gastroenterology* (2016), doi: 10.1053/j.gastro.2016.07.012.

TITLE: HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant

SHORT TITLE: Immunopathology in immune tolerant CHB

AUTHORS: William S. Mason[1], Upkar S. Gill[2], Samuel Litwin[1], Yan Zhou[1], Suraj Peri[1], Oltin Pop[3], Michelle L.W. Hong[4], Sandhia Naik[5], Alberto Quaglia[3], Antonio Bertoletti[4] & Patrick T.F. Kennedy[2]

AUTHOR AFFILIATIONS: [1]Fox Chase Cancer Center, Philadelphia, PA, United States, [2]Hepatology, Centre for Immunobiology, Blizard Institute, Barts and The London School of Medicine & Dentistry, QMUL, London, UK, [3]Histopathology, Institute of Liver Studies, Kings College Hospital, London, UK, [4]Emerging Infectious Diseases Program, Duke-NUS Graduate Medical School, Singapore, [5]Department of Paediatric Gastroenterology & Hepatology, The Royal London Hospital, Barts Health NHS Trust, London, UK.

CORRESPONDING AUTHOR: William S. Mason, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497

Tel: 215-728-2462; e-mail: William.Mason@fccc.edu

GRANT SUPPORT: Institutional funding from Fox Chase Cancer Center to WSM; Wellcome Trust Clinical Research Training Fellowship (107389/Z/15/Z) to USG; Fox Chase Cancer Center Core Grant (CA006927) to SL, YZ and SP; Singapore Translational Research (STaR) Investigator Award (NMRC/STaR/013/2012) to AB; Barts and The London Charity Large Project grant (723/1795) to PTFK.

ABBREVIATIONS: CHB, chronic hepatitis B; CPA, collagen proportionate area; FS, Ishak fibrosis state; IA, immune active; IT, immune tolerant; HAI, histological activity index; HBeAg, hepatitis B envelope antigen; HBcAg, hepatitis B core antigen; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma; NI, necro-inflammatory stage; PBMC, peripheral blood mononuclear cells.

DISCLOSURES: The authors have no relevant disclosures or conflicts of interest

AUTHOR CONTRIBUTIONS: Study concept & design: **WSM, AB, PTFK**; Generation, acquisition, collection of data/performed experiments: **WSM, USG, SL, OP, MH**; Analysis & Interpretation of data: **WSM, USG, SL, YZ, SP, OP, MH, AQ, AB, PTFK**; Provision of patient samples/technical/material support: **USG, SN, PTFK**; Statistical analysis: **WSM, USG, SL, OP**; Drafting, revision & writing of manuscript: **WSM, USG, AB, PTFK**; All authors provided critical input and approved the manuscript.

ABSTRACT

Background & Aims: Chronic infection with hepatitis B virus (HBV) progresses through different phases. The first, called the immune-tolerant phase, has been associated with lack of disease activity. We examined HBV DNA integration, clonal hepatocyte expansion, HBV antigen expression, and HBV-specific immune responses in patients in the immune-tolerant phase to assess whether this designation is appropriate or if there is evidence of disease activity.

Methods: We studied HBV DNA integration, clonal hepatocyte expansion, and expression of hepatitis B surface antigen and core antigen in liver tissues from 26 patients with chronic HBV infection (14–39 years old); 9 patients were positive for hepatitis B e antigen (HBeAg) in the immune-tolerant phase and matched for age with 10 HBeAg-positive patients with active disease and 7 HBeAg-negative patients with active disease. Peripheral blood samples were collected and HBV-specific T cells were quantified for each group.

Results: Detection of HBV antigens differed among groups. However, unexpectedly high numbers of HBV DNA integrations, randomly distributed among chromosomes, were detected in all groups. Clonal hepatocyte expansion in patients considered immune-tolerant was also greater than expected, potentially in response to hepatocyte turnover mediated by HBV-specific T cells, which were detected in peripheral blood cells from patients in all phases of infection.

Conclusions: We measured HBV specific T cells, HBV DNA integration, and clonal hepatocyte expansion in different disease phases of young patients with chronic hepatitis B, with emphasis on the so-called immune tolerant phase. A high level of HBV DNA integration and clonal hepatocyte expansion in patients considered immune tolerant indicated that hepatocarcinogenesis could be underway— even in patients with early-stage chronic HBV infection. Our findings do not support the concepts that this phase is devoid of markers of disease progression or that an immune response has not been initiated. We propose that this early phase be called a high replication, low inflammation stage. The timing of therapeutic interventions to minimize further genetic damage to the hepatocyte population should be reconsidered.

KEY WORDS: HBsAg, anti-viral immunity, HBV replication, hepatocyte proliferation

INTRODUCTION

Chronic hepatitis B (CHB) virus infection acquired at birth or in early childhood typically progresses through an early disease phase characterized by normal serum alanine aminotransferase (ALT) and high titer viremia (EASL & AASLD guidelines).[1] Patients can remain in this phase of CHB for several decades. Historically perceived as disease-free, these patients are considered 'immune tolerant' (IT) and thus excluded from therapy based on international treatment recommendations (EASL & AASLD). Classically, the IT phase is followed by a period of immune active (IA) liver disease, characterized by hepatic flares of increased inflammatory activity with elevated ALT levels, where patients are deemed to meet treatment criteria.

A question in the management of chronic HBV infection is whether antiviral treatment should be withheld until the development of persistently elevated serum ALT. Arguments against treatment in the IT phase have centered on drug cost, potential selection for drug resistant virus, and toxicity associated with long-term therapy.[2] Historically, a stronger argument against treatment has been the perceived lack of disease activity and suppression of antiviral immunity, but the validity of these arguments, which in a clinical setting normally rely on serological assays without liver histology, is unclear. For instance, the mechanism of hepatocyte destruction (e.g., apoptosis versus necroptosis) might change during the course of CHB, influencing ALT levels in a manner not reflecting the amount of cell destruction.[3,4]

The notion that events potentially leading to cumulative liver damage, including HCC initiation and promotion, are absent in IT patients has been contested by recent immunological data, which do not support clear differences between phases of CHB.[1,4-6] We have previously shown that HBV exposure in utero does not induce a generic state of immunological tolerance,[7] and also, that HBV-specific T cell responses in young patients labeled IT are not inferior to those seen in their peers with IA disease differentiated only by ALT elevation.[5] Recent data from CHB adults confirms that HBV-specific immunological parameters are no different between these two disease phases.[8] Further evidence against an inert immunological response in IT patients has come from a study demonstrating an increased innate immune gene signature in IT patients[6] and from virological data showing sequence evolution of HBV with increasing age in a cohort of IT patients.[9]

The presence of immunological activity and high levels of HBV replication in what is considered the IT phase may promote cumulative liver damage, since hepatocytes appear to constitute a closed, self-renewing cell population, as reported in animal studies investigating both syngeneic hepatocytes and transplanted human hepatocytes.[10,11] First, normal hepatocytes as well as hepatocytes with markers of senescence were able to proliferate to maintain liver mass during injury.[12] Second, recent evidence suggests that so-called liver progenitor/stem cells (e.g., oval cells) either do not have a significant role in liver regeneration[13-15] or conversely, if they do have a role in regeneration, are first formed via de-differentiation of mature hepatocytes.[10] Though some of these issues are still contested,[16-18] the overall conclusion that hepatocytes are primarily self-renewing seems valid. Consequently, epigenetic and genetic dysregulation, including damage via HBV DNA integration, might increase over time.

In the present study, we performed a comprehensive analysis of clinical and virological parameters in patients considered IT, and in age-matched IA non-cirrhotic HBeAg positive(+) and negative(-) CHB patients. We also assessed the frequency of HBV DNA integration and clonal hepatocyte expansion across all patient groups. Integration of HBV DNA into chromosomal DNA during chronic infection is one of the factors believed to contribute to or reflect mutagenesis leading to hepatocarcinogenesis. Importantly, using duck hepatitis B virus, integration was found to occur at double strand breaks, probably due to non-homologous end joining, and the frequency of mutagenesis during repair of double stranded breaks was 10 times as frequent as HBV DNA integration at the site.[19] Thus, HBV DNA integration frequency may significantly underestimate the mutation frequency in hepatocytes. Errors during repair of double stranded DNA breaks are considered important in human oncogenesis.[20]

A recent study showed that virus integration and hepatocyte expansion may be present in the IT phase, but this phenomenon was not studied in detail and age-matched controls were not available.[21] In the present study, we compared HBV integration frequency and clonal hepatocyte expansion in young patients considered IT, and aged-matched IA HBeAg(+) and HBeAg(-) controls. Since HBV DNA integration occurs at random sites in host DNA, virus/host DNA junctions serve as markers of hepatocyte lineages, and the multiplicity of virus/cell DNA

junctions from liver tissue can be used to calculate clonal hepatocyte expansion. Finally, differential HBV antigen expression in hepatocytes, as well as HBV-specific immune responses were determined across the disease phases to test the validity of what is labeled IT CHB.

MATERIALS AND METHODS

Patient samples & Study design

Twenty-six patients were recruited and categorized into CHB phases using established clinical characteristics: measurements of serum transaminases (ALT), serological parameters, including HBsAg, HBeAg, anti-HBeAg and virus titers (EASL & AASLD): Immune tolerant (IT) (n=9); HBeAg(+) immune active (IA) (n=10); HBeAg(-) immune active (IA) (n=7) *(Table 1).* The patients were further assessed by liver biopsy. HBV DNA levels (virus titers) in serum samples were quantified by real-time PCR (Roche COBAS AmpliPrep/COBAS Taqman HBV test v2.0-dynamic range 20 to $1.7x10^8$ IU/ml-Roche molecular diagnostics, Pleasanton, CA) and HBsAg by Abbott Architect (Abbott Diagnostics, Abbot Park, IL). Serum was tested for HBeAg and anti-HBe with a chemiluminescent microparticle immunoassay (Abbott Architect). HBV genotype was also recorded. Ishak fibrosis stage (FS) and necro-inflammatory (NI) scores from liver biopsies were also determined. Whole blood was taken at the time of liver biopsy. PBMC were isolated by Ficoll-Hypaque density gradient centrifugation and cryopreserved for immunological analysis. Liver biopsy specimens, surplus to diagnostic requirements, were stored at -80°C for subsequent DNA extraction. Tissue samples taken for diagnostic histological examination were formalin-fixed, paraffin-embedded and used for immune-histochemical staining. Written informed consent was obtained from all patients. The study was approved by the local ethics committee (Barts and The London NHS Trust Ethics Review Board) and the Institutional Review Board of the Fox Chase Cancer Center.

In vitro expansion of HBV-specific T cells

Frozen PBMCs isolated from fresh heparinized blood by Ficoll-Hypaque density gradient centrifugation were thawed and resuspended in AIM-V medium with 2% pooled human AB serum (serum AIM-V). For HBV-specific T cell expansion, panels of synthetic peptides (15-mers, with 10 amino acids overlap, 313 in total) were pooled in 4 mixtures covering the whole HBV proteome. After 10 days of *in vitro* expansion, the presence of T cells responding to HBV peptide stimulation were determined by measuring the frequency of T cells producing IFN-γ with intracellular cytokine staining (ICS) or ELISPOT assays as previously described[22] (*Supplementary Materials & Methods*).

Immunohistochemistry & Image Analysis

Adequate specimens of Formalin-fixed and paraffin-embedded tissue from 19/26 patients (*Table 1*) were available for immunohistochemistry (IHC) (*Supplementary Materials & Methods*).

Slides were imaged using a Leica DM6000 B microscope (Leica Biosystems, Newcastle, UK) equipped with a Leica DFC300 FX camera (Leica Biosystems, Newcastle, UK). A variable number of serial micrographs were taken from each Sirius red stained slide to cover the entire tissue. Tissue and collagen areas were measured on each micrograph using the ImageJ software (Bethesda, Maryland, USA) (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsbweb.nih.gov/ij/, 1997-2015) and a protocol described on the ImageJ web page (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA,http://rsbweb.nih.gov/ij/docs/examples/stained-sections/index.html, 2015) following previous calibration. Total tissue and collagen areas were then calculated for each biopsy (*Supplementary Materials & Methods*).

Results were assessed and plotted using GraphPad Prism 6 Trial Version (GraphPad Software, SanDiego, USA). The following tests were performed: Shapiro-Wilk normality test, Mann-Whitney, Kolmogorov-Smirnov, and Spearman correlation.

Extraction and inverse PCR analysis of liver DNA

Two to three ~1 mm pieces of each liver biopsy were cut, and nucleic acids extracted. Inverse PCR was designed to detect the right hand junction of integrations occurring between host DNA and HBV double stranded linear DNA (HBV dslDNA) *(Figure 1A),* the primary substrate for viral DNA integration.[23, 24] To design PCR primers, and determine endonuclease cleavage sites for detection of the right hand virus/cell junction fragments*,* the predominant HBV sequence in the liver of each patient was determined by PCR amplification and sequencing of fragments covering the region from nts ~1193 to ~1860 on the HBV genome.[25] HBV sequences were numbered according to Galibert et al.[26] (accession number V01460).

Prior to inversion, high MW DNA (≥10-20 kbp) was purified by low-melt agarose gel electrophoresis, to reduce cccDNA contamination. The DNA was then digested by addition of NcoI-HF (NEB) and incubation for 30 min at 37℃. Nc oI-HF was heat inactivated for 20 min at 80℃, and the DNA recovered using the QIAquick PCR purification kit. The DNA fragments were then circularized by incubation with T4 DNA ligase (Figure 1A).[27] Prior to use for PCR, the circularized DNA was suspended in 40µl NEB buffer 4 supplemented with BSA (NEB) and linearized by digestion at 65℃ with BsiHKAI (NEB). Molecules potentially derived from intra-molecular ligation of residual cccDNA (e.g., between the authentic NcoI site and a distal NcoI "star" site in cccDNA) or from cccDNA deletion mutants (PCR conditions were not adequate to amplify full-length cccDNA) were cleaved with SphI (NEB) to reduce their amplification during inverse PCR. (For several samples, it was necessary to use different restriction enzymes, because of differences in HBV DNA sequence (*Supplementary Table 1* and *Figure 1A*). See *Supplementary Materials & Methods* for additional details.

Following inversion, endpoint dilution, and nested PCR, the products were subjected to electrophoresis in 1.3% agarose gels containing E-buffer and 0.5µg/ml of ethidium bromide (*Figure 1B*). Bands were excised from the gel and sequenced with the F2 or R2 primer, as previously described.[28] The junction of viral with cellular DNA was located using the GCG program FASTA. Junctions repeated in different wells were identified by comparing cell sequences immediately adjacent to virus/cell junctions, using Sequencher version 5.0.1 (Gene Codes Corporation) *(Supplementary Materials & Methods).*

Quantifying host DNA in liver biopsy extracts

Host DNA was quantified by real time qPCR of epsilon globin DNA (accession number M81361), as previously described.[25] A PCR amplified epsilon globin DNA was used as a control. The cell equivalents of DNA extracted from each biopsy are summarized in *Supplementary Table 2*.

Statistical analyses: Quantifying virus/cell junctions by end-point dilution

As illustrated in *Figure 1,* inverted DNA samples were serially diluted into 96 well PCR plates.

Typically, 5-10µl of inverted DNA, representing a small fraction of the original DNA sample (~5-10%), was added to 170-175µl of PCR reaction mix in well A1. After mixing, 60µl was serially diluted into 120µl of reaction mix in wells B1 through G1. Well H1 contained 120µl of reaction mix, but no DNA sample, and served as a negative control. 10µl aliquots of the reactants in column 1 were then distributed to columns 2 to 12 and subjected to nested PCR. 95% confidence intervals for clone sizes determined using end-point dilution were calculated using the fortran program Sim19 *(Supplementary Materials & Methods).*

Modeling the Clonal Expansion of hepatocytes

The program Csize8 was devised to predict the size of hepatocyte clones created after birth, as a consequence of liver growth and random hepatocyte turnover. Liver growth was assumed to be linear during the growth phase. Hepatocyte turnover during growth and in the full size liver were assumed to occur as a result of random death of hepatocytes with a rate constant, k. In the adult liver, death and regeneration were assumed to occur at the same rate, to maintain liver size. In the simulations presented here, k was assumed to be the same for the growing and adult liver *(Supplementary Materials & Methods).*

RESULTS

Evidence of HBV-specific T cell responses in patients in the immune tolerant phase of CHB

HBV-specific T cells were detected in all 3 patient groups, IT, HBeAg(+), and HBeAg(-) IA disease (*Table 1*). Using HBV-specific peptides spanning the entire HBV proteome, T cells were expanded *in vitro* and assayed for both intracellular cytokine staining and ELISPOT *(Figure 2A).* The quantity of HBV-specific T cells in terms of magnitude (number of cells recognizing a single HBV peptide mixture, *Figure 2B*) or the ability to recognize different mixtures of HBV peptides (*Figure 2C*) were comparable among the three patient cohorts. Consistent with our previous data,[5] patients classified as IT did not show any significant difference in circulating HBV-specific T cells in comparison with CHB patients classified as IA in relation to their virological and clinical features *(Table 1; Groups 2 & 3).* Serum ALT levels were significantly lower in the IT group compared to the other groups. Despite this, differences in immune response of patients across the disease phases were not detected. ALT is often considered a surrogate of immune activity; however, as noted earlier, we and others have previously demonstrated that ALT does not 'benchmark' the HBV immune response.[5, 8] The comparable levels of peripheral HBV-specific T cell responses in IT patients with those in the other two groups suggested that infected hepatocytes might be targeted for T cell mediated destruction in all patients including those diagnosed as IT. For this reason, we analyzed whether clinical phases could be distinguished by differences in the intrahepatic compartment. Immunohistochemistry analyses, measurements of HBV DNA integration and hepatocyte turnover were performed to determine if IT patients were different from the other patients studied.

A larger fraction of nuclear HBcAg positive hepatocytes are found in immune tolerant CHB

Liver tissue from 19/26 patients [Group 1, IT n=8; Group 2, HBeAg(+) IA n=5; HBeAg(-) IA n=6] was double stained for detection of HBcAg and HBsAg. Significant differences were found in the level of nuclear HBcAg positive hepatocytes in IT patients (Group 1; mean 30.1%) compared to the other groups (mean 0.92% and 0%; Groups 2 and 3 respectively) (IT vs. IA, $P<.005$) *(Figure 3A, B).* Interestingly 7/8 IT patients had >18% nuclear-HBcAg positive

hepatocytes *(Figure 3A)*; conversely, no patient exceeded ~3% positivity in the other groups irrespective of virus titer. HBsAg staining alone, the classical ground glass appearance reported on HBV tissue, was significantly higher in HBeAg(-) IA disease (Group 3) compared with IT (Group 1) (*P*=.004), but was not significantly different between Groups 1 and 2 *(Figure 3A, B).* These findings are consistent with previous work, which reported that nuclear HBcAg positive hepatocytes predominated in the IT phase in children.[29] The reason for this finding remains unclear.

Despite the significant difference in nuclear HBcAg positive hepatocytes between patient groups, there was no overall difference in Ishak fibrosis stage (*Table 1*), collagen proportionate area (CPA) *(Figure 3C, D)* or histological activity index (HAI) *(Table 1; Figure 3E, F)* underscoring the limitations of standard histological assessment and clinical parameters used alone or even in combination to define phases of CHB.

Integrated HBV DNA was identified in chromosomes of all patients

Five hundred and ninety two different virus/cell junctions were detected overall, using inverse PCR. 500 could be mapped to unique sites on human chromosomal DNA (208 for group 1, 195 for group 2, 97 for group 3) (*Supplementary Table 3*; *Supplementary Results*). Of these 500 integration sites, 246 were located within potentially transcribed regions, including 217 mRNA encoding regions and 29 non-coding RNAs. 231 of the integration sites mapped to introns, 13 to exons, one at an intron/exon boundary, and one mapped within a gene (uncertainty in the exact junction site precluded the exon/intron distinction). Of the protein coding genes with integrated HBV DNA, ~70% appeared to be transcribed in the liver. Protein expression in the liver has been reported for ~45% of these (www.genecards.org).[30] 4/29 regions with integrated HBV DNA that specified non-coding RNAs also appeared to be expressed in liver. For most, it was unclear if expression occurred in hepatocytes or other liver cells. The remaining 92/592 integrations were located in repeated DNA sequences and/or could not be mapped. Our results are likely to underestimate the true number of unique HBV integration sites in the DNA samples; that is, single or low copy clones might be obscured by competing amplification of high copy number clones *(Figure 1B).* Notably, multiple integrations were found on every chromosome except Y *(Table 2; Supplementary*

Table 3). Integration sites are illustrated in *Figure 4A*. Using the Chi-Squared Test, we were unable to reject the null hypothesis that the integration frequency on chromosomes *(Figure 4A; Supplementary Table 3),* including Y (13 of 26 patients were male), was proportional to their length (*P*=0.195). No significant differences were seen between patient groups *(Figure 4B; IT patients).*

The average frequency of total integrations in groups 1 through 3, respectively, including those in hepatocyte clones, ranged from $1.5x10^9$ to $5x10^9$ per liver of $5x10^{11}$ hepatocytes *(see Supplementary Table 2 for individual patients).* Importantly, integration is prevalent in patients considered IT. Because just a small fraction of each DNA sample was assayed, we could only make a minimum estimate of the unique integration sites among the total. The data suggested at least ~$5x10^6$ distinct integration sites are present in a liver of $5x10^{11}$ hepatocytes in each patient group. This high number of possible sites means that a liver of $5x10^{11}$ hepatocytes would contain at least one hepatocyte in which a particular gene would be mutated in each patient group including those characterized as IT, not just age matched controls with more advanced liver disease. (We could not demonstrate any correlation with HBsAg or HBV DNA levels and total integrations in the whole study cohort; thus, the extent to which integration might contribute to HBsAg production in the 3 patient groups remains unclear).

Clonal hepatocyte expansion

Because the hepatocyte population appears self-renewing, death and regeneration will lead to loss of some cell lineages and clonal expansion of others to maintain liver mass. To determine if IT patients have elevated hepatocyte turnover, possibly due to anti-HBV immune killing, we investigated if these patients had evidence of hepatocyte clones that were similar in size to those found in late phases of CHB with HCC.[28] Simultaneously, we asked if similar levels of clonal hepatocyte expansion were present in our three age-matched patient groups. Insertional mutagenesis and expression of HBV genes from integrated DNA are potential initiation events in hepatocarcinogenesis, as is repair of double stranded DNA breaks by non-homologous end joining in the absence of HBV integration.[19] Enhanced hepatocyte turnover could be promotional,[31] by facilitating clonal expansion of subsets of hepatocytes, including

but not limited those with preneoplastic mutations. Large hepatocyte clones were seen in all three patient groups (*Figure 5*). The difference in maximum clone sizes between groups 1 and 3 was statistically significant (*P*=.0015), as was the difference between groups 2 and 3 (*P*=.014) (*Table 4*; *Figure 5*); the difference between Groups 1 and 2 did not reach statistical significance (*P*=0.36) (Wilcoxon 2-sided Rank Sum Test).

Hepatocyte clone sizes were larger in immune tolerant patients than predicted by a model of random hepatocyte turnover

As discussed, hepatocyte turnover in the liver should lead to increasing clonality, with loss of some hepatocyte lineages and expansion of others. To determine if the large clones (*Figure 5A-C*) could be explained by random death and compensatory division of hepatocytes, to maintain liver mass, a computer simulation, Csize8 *(Supplementary Materials & Methods)*, was used. We assumed that hepatocytes proliferate (and die) in the adult liver with a rate constant k=0.0015/day (0.15%/day), 3 times the fraction of hepatocytes in the S phase (0.0005) in healthy adult liver at any given time.[32] We also assumed that infection occurred at birth and that the liver size increases 10-fold during maturation. The maximum expected clone sizes in the patients studied (age range 14-39 years) increased, with age, from ~400 to ~600 hepatocytes *(Figure 5E).* This range would increase from ~800 to ~1200 if the rate constant for hepatocyte death increased to k=0.004/day (0.4%/day), and ~1600 to 2800 with a rate constant of k=0.01/day (~1% of hepatocytes killed/day) *(Figure 5E).*

Maximum observed clone sizes exceeded clone sizes predicted for random liver turnover (0.015%/day)[32] for 6/9 IT patients (Group 1), 6/10 in HBeAg(+) IA patients (Group 2), and 7/7 in HBeAg(-) IA patients (Group 3) *(Figure 5E; Supplementary Table 4)*. For a turnover of 0.04% per day, excess turnover was observed in 2/9 patients in Group 1, 5/10 in Group 2, and 7/7 in Group 3. 3/10 patient samples in Group 2 and 5/7 in Group 3 exceeded predictions even for a daily turnover of 1%. While differences in maximum predicted clone sizes and observed sizes may appear small, it is important to note that the amount of hepatocyte destruction and replacement in the model that is necessary, for example, to give a maximum clone size of 600 (k=0.0015) vs. 2800 (k=0.01) hepatocytes, after 39 years, is 21 vs. 142 livers worth of hepatocyte death and replacement. In summary, a model of random death and

regeneration of hepatocytes at a level estimated for healthy liver did not provide a consistent explanation for maximum clone sizes observed in 6/9 IT patients, which was also true of 6/10 patients in Group 2 and all patients in Group 3. The differences might be more extreme, because the modeling assumes all clones are detected, not just those with integrations. These analyses suggest a selective process for hepatocyte turnover can occur in all groups. This might result, for example, from emergence of hepatocyte clones that are resistant to T cell killing, or because some hepatocyte lineages are more responsive to growth signals to divide, to maintain liver mass.

DISCUSSION

We have demonstrated that HBV DNA integration and clonal hepatocyte expansion were similar in patients considered IT to those that have HBeAg(+) IA CHB. These results raise questions about the perception that the IT phase is 'disease-free', as well as the premise upon which treatment decisions are made. In line with our previous work and recent publications in the field, we feel that the term 'high replicative low-inflammatory' (HRLI) CHB more accurately reflects this early disease phase, and thus should now be adopted into clinical practice.[4-6,8,33]

CHB is the leading cause of primary liver cancer worldwide and despite the lack of robust data to support this notion,[1] the current consensus is that HCC risk does not increase in the majority of patients until there is perturbation in serum ALT, interpreted as a sign of immune activity. There are, however, studies supporting the development of HCC in the absence of advanced liver disease. The REVEAL study demonstrated an association between high viral load and HCC development, independent of cirrhosis, thus pertinent to the study population here.[34] The data presented here suggest that an approach to management which excludes HRLI patients (formerly considered IT) from treatment may be flawed, as HBV specific T cells as well as extensive clonal hepatocyte expansion are already present in this early phase of CHB. Evidence from both animal and human studies demonstrate that clonal hepatocyte expansion is a major risk factor for HCC,[35] moreover, HBV DNA integration, a potential initiating event for HCC, was found to be prevalent not just in later stages of CHB, but also in the HRLI phase. The presence of both HBV DNA integration and clonal hepatocyte expansion in this early phase of CHB are thus at odds with the concept of a 'disease-free' state.

These data are consistent with recent studies and our previous findings,[5,6,8] which dispute the idea that so-called IT patients are immunologically inert and, therefore, fundamentally distinct from HBeAg(+) IA disease. In the current study, we confirmed the presence of HBV-specific responses in the HRLI and later phases of CHB by dual modality (ELISPOT and Intracellular Cytokine Staining) (*Figure 2)*. These data reinforce the fact that there is no quantifiable difference in antiviral immunity between the HRLI phase and HBeAg(+) IA patients. Furthermore, these findings were verified by detailed analysis of the liver compartment of the patients studied. In keeping with the HBV specific response in the periphery, we

demonstrated few if any differences in liver histology (*Figure 3)*. Based on serological assessment, patients labeled IT had similar levels of fibrosis, CPA and HAI as those considered to have IA disease.

In addition, we could demonstrate differences in the level of nuclear core expression; being significantly higher in those considered IT compared with HBeAg(+) and HBeAg(-) IA patients; in contrast, HBsAg positive hepatocytes were preferentially found in HBeAg(-) IA patients (*Figure 3*). This mosaic distribution of HBV antigens in different hepatocytes and phases of HBV infection might reflect different virological or immunological features that need further characterization. A recent study suggested that hepatocytes expressing high HBcAg may have higher level of HBV replication and higher cccDNA content than HBsAg expressing hepatocytes.[36] However, the biological significance of the diverse HBV antigen patterns detectable in the different categories of CHB remains unclear.

An important issue is the number of different HBV DNA integration sites in the livers of the three patient groups, which will determine the numbers of host genes potentially mutated by HBV integration, and also may be an indirect indicator of the number of double strand DNA breaks repaired by non-homologous end joining, which is also potentially mutagenic.[19,20] Our primary goal was to estimate hepatocyte clone size using end point dilution assays; thus, we can only make a minimum estimate for the number of unique HBV integration events. The real number may in fact be much larger, but interestingly, the number estimated for all three patient groups (at least ~$5x10^6$ per liver in all three groups) would be sufficient, if uniformly distributed across the human genome in a liver of $5x10^{11}$ hepatocytes, to place integrated HBV DNA within any 1000 nt region in the genome of at least one hepatocyte. (Note that mutation by incorrect repair of double stranded DNA breaks may be 10-times more frequent.[19]) Thus, the potential to mutate and alter expression of any host gene in at least one hepatocyte appears very high across the disease phases. Some of these integrations may be procarcinogenic.

To explore the concept that patients considered IT may require earlier treatment we also investigated clonal hepatocyte expansion in these patients and compared it to IA CHB (*Figure 5*). The rationale was that clonal hepatocyte expansion in mutated hepatocytes would

contribute to tumor promotion.[31,35] To the extent that the hepatocyte population is self-renewing, and undergoing random death and regeneration, it is possible to relate cumulative hepatocyte turnover to maximum hepatocyte clone sizes. Compared to our predictions, actual clone sizes in HBeAg(-) IA disease (group 3) appeared excessive, similar to those in HCC patients, even assuming a relatively high hepatocyte death rate of 1.0% per day (*Figure 5*).[32] In contrast, HBeAg(+) IA patients (Group 2) appeared to have much lower hepatocyte turnover and were not significantly different than those considered IT *(Figure 5).* Nonetheless, average hepatocyte clone sizes in both groups 1 and 2, exceeded predictions for normal liver turnover (k=0.0015). Indeed, in some of these patients, very large clone sizes were detected, which can only be explained by assuming a selective growth or survival advantage for hepatocytes (*Supplementary Table 4*). This was also noted in a study of non-tumorous liver samples from non-cirrhotic HCC patients.[21,25] In brief, our data suggest that clonal hepatocyte expansion, an HCC risk factor,[31,35] is active across all the phases of CHB studied here.

This study confirms the presence of HBV-specific T cell responses and the significant extent of HBV DNA integration/cell mutagenesis along with clonal hepatocyte expansion in the HRLI phase and across the disease phases. These findings further challenge the notion of an IT phase devoid of disease progression, raising questions about the timing of therapeutic intervention to minimize genetic damage to the hepatocyte population and reduce the promotional role in carcinogenesis of elevated hepatocyte turnover. As the risk of HCC may already be present in the HRLI phase, these data make a compelling case to consider antiviral therapy in these patients. Future studies are required to explore the merits of earlier treatment to prevent disease progression and the development of HCC in CHB.

ACKNOWLEDGEMENTS

We are grateful for helpful discussions and advice from Drs. Richard Katz, Christoph Seeger and Glenn Rall (Fox Chase Cancer Center) and Jesse Summers (University of New Mexico), and to Anita Cywinski (Fox Chase Cancer Center DNA sequencing facility) for the DNA sequencing required to carry out this study.

WSM acknowledges the kind support of Christoph Seeger for providing laboratory space and to Christoph Seeger, John Gricoski, and Robert Beck for encouragement to carry out his portion of the study.

PTFK acknowledges the patients, their families and the staff at The Royal London Hospital who have supported and contributed to this work.

REFERENCES

1. Zoulim F, Mason WS. Reasons to consider earlier treatment of chronic HBV infections. Gut 2012;61:333-336.
2. Gill US, Zissimopoulos A, Al-Shamma S, et al. Assessment of bone mineral density in tenofovir-treated patients with chronic hepatitis B: can the fracture risk assessment tool identify those at greatest risk? J Infect Dis 2015;211:374-382.
3. Maini MK, Boni C, Lee CK, et al. The role of virus-specific CD8(+) cells in liver damage and viral control during persistent hepatitis B virus infection. J Exp Med 2000;191:1269-1280.
4. Bertoletti A, Kennedy PT. The immune tolerant phase of chronic HBV infection: new perspectives on an old concept. Cell Mol Immunol 2015;12:258-263.
5. Kennedy PT, Sandalova E, Jo J, et al. Preserved T-cell function in children and young adults with immune-tolerant chronic hepatitis B. Gastroenterology 2012;143:637-645.
6. Vanwolleghem T, Hou J, van Oord G, et al. Re-evaluation of hepatitis B virus clinical phases by systems biology identifies unappreciated roles for the innate immune response and B cells. Hepatology 2015;62:87-100.
7. Hong M, Sandalova E, Low D, et al. Trained immunity in newborn infants of HBV-infected mothers. Nat Commun 2015;6:6588.
8. Park JJ, Wong DK, Wahed AS, et al. Hepatitis B Virus-Specific and Global T-Cell Dysfunction in Chronic Hepatitis B. Gastroenterology 2016;150:684-695.
9. Wang HY, Chien MH, Huang HP, et al. Distinct hepatitis B virus dynamics in the immunotolerant and early immunoclearance phases. J Virol 2010;84:3454-3463.
10. Tarlow BD, Pelz C, Naugler WE, et al. Bipotential adult liver progenitors are derived from chronically injured mature hepatocytes. Cell Stem Cell 2014;15:605-618.
11. Grompe M. Liver stem cells, where art thou? Cell Stem Cell 2014;15:257-258.
12. Wang MJ, Chen F, Li JX, et al. Reversal of hepatocyte senescence after continuous in vivo cell proliferation. Hepatology 2014;60:349-361.
13. Schaub JR, Malato Y, Gormond C, et al. Evidence against a stem cell origin of new hepatocytes in a common mouse model of chronic liver injury. Cell Reports 2014;8:933-939.
14. **Yanger K, Knigin D**, Zong Y, et al. Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. Cell Stem Cell 2014;15:340-349.

15. Marongiu F, Serra MP, Sini M, et al. Cell turnover in the repopulated rat liver: distinct lineages for hepatocytes and the biliary epithelium. Cell Tissue Res 2014;356:333-340.
16. Font-Burgada J, Shalapour S, Ramaswamy S, et al. Hybrid periportal hepatocytes regenerate the injured liver without giving rise to cancer. Cell 2015;162:766-779.
17. Wang B, Zhao LL, Fish M, et al. Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver. Nature 2015;524:180-185.
18. Miyajima A, Tankaka M, Itoh T. Stem/progenitor cells in liver development, homeostasis, regeneration, and reprograming. Cell Stem Cell 2014;14:561-574.
19. Bill C, Summers J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. Proc Natl Acad Sci USA 2004;101:11135-11140.
20. Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci USA 2003;100:12871-12876.
21. Tu T, Mason WS, Clouston AD, et al. Clonal expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. J Viral Hep 2015;22:737-753.
22. Tan AT, Loggi E, Boni C, et al. Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. J Virol 2008;82:10986-10997.
23. Yang W, Summers J. Integration of hepadnavirus DNA in infected liver: evidence from a linear precursor. J Virol 1999;73:9710-9717.
24. Gong SS, Jensen AD, Chang CJ, et al. Double-stranded linear duck hepatitis B virus (DHBV) stably integrates at a higher frequency than wild-type DHBV in LMH chicken hepatoma cells. J Virol 1999;73:1492-1502.
25. Mason WS, Liu C, Aldrich CE, et al. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. J Virol 2010;84:8308-8315.
26. Galibert F, Mandart E, Fitoussi F, et al. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E.coli. Nature 1979;281:646-650.
27. Mason WS, Low HC, Xu C, et al. Detection of clonally expanded hepatocytes in chimpanzees with chronic hepatitis B virus infection. J Virol 2009;83:8396-8408.
28. Summers J, Jilbert AR, Yang W, et al. Hepatocyte turnover during resolution of a transient hepadnaviral infection. Proc. Natl. Acad. Sci. USA 2003;100:11652-11659.

29. Hsu H-Y, Lin Y-H, Chang M-H, et al. Pathology of chronic hepatitis B infection in children: With special reference to the intrahepatic expression of hepatitis B virus antigens. Hepatology 1988;8:378-382.
30. Stelzer G, Dalah I, Stein TI, et al. In-silico human genomics with GeneCards. Hum Genomics 2011;5:709-717.
31. Farber E, Sarma DS. Hepatocarcinogenesis: a dynamic cellular perspective. Lab Invest 1987;56:4-22.
32. Mancini R, Marucci L, Benedetti A, et al. Immunohistochemical analysis of S-phase cells in normal human and rat liver by PC10 monoclonal antibody. Liver 1994;14:57-64.
33. Gish RG, Given BD, Lai CL, et al. Chronic hepatitis B: Virology, natural history, current management and a glimpse at future opportunities. Antiviral Res 2015;121:47-58.
34. Chen CJ, Yang HI, Su J, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 2006;295:65-73.
35. Marongiu F, Doratiotto S, Montisci S, et al. Liver repopulation and carcinogenesis: Two sides of the same coin? Am J Pathol 2008;172:857-864.
36. **Zhang X, Lu** W, Zheng Y, et al. In situ analysis of intrahepatic virological events in chronic hepatitis B virus infection. J Clin Invest 2016;126:1079-1092.

Author names in bold designate shared co-first authorship.

FIGURE LEGENDS

Figure 1: Inverse PCR detection of integrated HBV DNA. A) Strategy for detection of integrated HBV DNA and clonal hepatocyte expansion. Inverse PCR, as used by Summers *et al.,*[25,28] was designed to detect the right hand junction of integrations of HBV dslDNA, the predominant precursor for integration, into host DNA.[23,24] Following cleavage and ligation (*Figure 1A*), the DNA samples were serially diluted and subjected to nested PCR using the indicated forward and reverse primers (*Figure 1B*). Primers are indicated in *Supplementary Table 1* and Materials and Methods. (*Figure 1A* modified from reference by Mason *et al.*[25]). B) Gel electrophoresis of inverse PCR products. Samples from nested PCR, carried out in a 96 well tray, were subjected to gel electrophoresis in a 1.3% agarose gel. PhiX phage DNA digested with HaeIII was used as a size marker (M). The fraction of the initial DNA sample distributed across each row of 12 wells is indicated. Bands were picked from the last 5 rows, not including the negative control, and subjected to DNA sequencing to identify virus/cell DNA junctions. For instance, the circled bands arise from a single hepatocyte clone; other clones were also identified by DNA sequencing (not highlighted).

Figure 2: Profile of HBV-specific T cell responses in all patient groups.
Patient PBMC were analyzed by ELISPOT and intracellular cytokine staining (ICS) for IFN-γ. (A) Evidence of HBV-specific T cell responses by ELISPOT and ICS against the Core, Envelope and Polymerase proteins, for each patient in the groups studied; Shaded black – positive HBV-specific T cell response; unshaded squares – negative HBV-specific T cell response, shaded grey – sample not done. (B) Comparison of spot forming units (SFU) by ELISPOT, in each patient, in the different groups; immune tolerant (IT) (shaded black), HBeAg(+) IA (shaded grey) and HBeAg(-) IA (unshaded). Bars represent the number of SFU cells in response to HBV core, envelope, and polymerase peptide pools. (C) Number of HBV peptide pools recognized by HBV-specific T cells obtained in the indicated patients.

Figure 3: Differential nuclear core antigen staining but similar fibrosis and inflammatory indices between CHB phases.
Formalin-fixed and paraffin-embedded tissue was analyzed with immunohistochemistry for HBcAg and HBsAg positive hepatocytes, along with quantification of fibrosis and histological activity indices for each patient. (A) Percentage of HBcAg positive hepatocytes (left panel)

and HBsAg positive hepatocytes (right panel) in each group; IT (open circles), HBeAg(+) IA (open squares) and HBeAg(-) IA (open triangles). Each point represents 1 patient, data shown as mean with SEM, as error bars. (B) Immunostaining identifying HBcAg positive hepatocytes (brown) and HBsAg positive hepatocytes (pink) from representative patients from each patient group (*Table 1)* (100x); IT (left panel), HBeAg(+) IA (middle panel) and HBeAg(-) IA (right panel). Inset shows magnified image (400x). (C) Ishak Fibrosis stage (left panel) and collagen proportionate area (right panel) of patients studied in each phase of CHB, data shown as mean with SEM, as error bars. (D) Sirius red staining of liver tissue from representative patients in each phase; IT (left panel), HBeAg(+) IA (middle panel) and HBeAg(-) IA (right panel). (E) Histological activity index scores; (from left to right – Interface hepatitis, Confluent necrosis score, Focal lytic necrosis, apoptosis & focal inflammation score and Portal inflammation score) of patients studied in each phase of CHB, data shown as mean with SEM, as error bars, (F) Identification of the inflammatory infiltrate as shown in (E) from representative patients in each phase of CHB; IT (left panel), HBeAg(+) IA (middle panel) and HBeAg(-) IA (right panel). Significant changes marked with asterisks, **P*<.05; ***P*<.01; ****P*<.001; ns=not significant

Figure 4: Sites of HBV DNA integration on human chromosomes.

A) Integration sites are summarized from all three patient groups *(Table 1)* by vertical lines. Results include the 208 from IT disease patients (Group 1), 195 from HBeAg(+) IA disease (Group 2), and 97 from HBeAg(-) IA disease (Group 3). Groups 1 (IT) and 2 [HBeAg(+)]; integrations were found on all chromosomes except Y. The single Y chromosome integration was from a patient from group 3. No group 3 patient integration sites were mapped to chromosomes 15 and 16. B) Integration sites in Group 1 patients - IT phase. Integration site details are shown in *Supplementary Table 3*. Clone sizes: *>5,000 and #>20,000.

Figure 5: Hepatocyte clones detected in all patient groups

Hepatocyte clones in (A) IT disease (Group 1), (B) HBeAg(+) IA disease (Group 2) and (C) HBeAg(-) IA disease (Group 3). Clone sizes were estimated as described. *(Figure 1, Materials and Methods and Supplementary Materials & Methods)*. The point estimates for clone size were calculated using the program Sim19 *(Supplementary Materials & Methods).* Clones are grouped by increasing size for each patient, and patients within a group are

arranged by increasing age from left to right. D) Mean of the maximum clone size for each patient within a group. Geometric means were calculated using the point estimates in *Supplementary Table 4*. HCC data are from a published analysis of clone sizes in non-tumorous liver from a group of 5 non-cirrhotic HCC patients.[25] (E) Predicted maximum clone sizes vs. age. These were calculated using the Csize8 program *(Materials and Methods and Supplementary Materials & Methods)*, for 3 different daily rate constants for hepatocyte turnover; k=0.0015/day (0.15%) - (black dashed line); k=0.004/day (0.40%) – (grey dashed line) and k=0.01/day (1.00%) - (solid black line). The adjacent corresponding bars indicate the geometric mean hepatocyte clone size, for each patient group in (D), for comparison against the predicted maximum clone size.

Table 1: Patient Characteristics

Group 1: Immune Tolerant	Sex	Age	ALT IU/L	HBV Geno-type	HBeAg/ anti-HBe	HBV DNA log IU/ ml	HBsAg titer log IU/ml	Fibrosis Stage (/6)	HAI (/18)	Peripheral T cell analysis *	IHC & Image analysis *
Pt. 1	F	15	36	E	+/-	8.69	5.22	2	3	Yes	Yes
Pt. 2	M	17	29	C	+/-	9.17	4.59	1	2	Yes	Yes
Pt. 3	F	18	18	B	+/-	8.42	4.68	0	2	Yes	Yes
Pt. 4	M	18	38	D	+/-	9.71	5.16	2	2	Yes	Yes
Pt. 5	M	22	40	E	+/-	8.66	4.36	3	3	Yes	Yes
Pt. 6	F	24	38	C	+/-	8.58	4.52	1	2	No	No
Pt. 7	F	28	30	E	+/-	7.60	4.57	1	4	No	Yes
Pt. 8	F	30	32	C	+/-	8.51	4.89	1	3	Yes	Yes
Pt. 9	F	39	31	B	+/-	8.52	4.55	1	2	No	Yes
Group 2: HBeAg(+) IA											
Pt. 10	M	14	70	D	+/-	8.80	4.17	2	3	Yes	Yes
Pt. 11	M	14	99	A	+/-	8.19	4.11	3	4	No	No
Pt. 12	F	16	63	D	+/-	7.06	2.67	1	3	No	No
Pt. 13	F	17	127	D	+/-	7.98	3.02	3	5	Yes	Yes
Pt. 14	M	19	89	C	+/-	8.49	4.82	3	3	Yes	Yes
Pt. 15	F	23	172	A	+/-	8.32	4.19	2	7	No	No
Pt. 16	M	25	77	B	+/-	8.36	4.76	1	2	Yes	Yes
Pt. 17	M	25	59	D	+/-	8.19	5.13	1	3	No	No
Pt. 18	F	28	161	C	+/-	7.09	2.29	1	6	Yes	Yes
Pt. 19	F	29	68	B	+/-	8.59	5.09	1	4	No	No
Group 3: HBeAg(-) IA											
Pt. 20	M	23	113	D	-/+	3.64	4.09	2	2	Yes	Yes
Pt. 21	F	25	29	E	-/+	4.19	3.73	1	2	Yes	Yes
Pt. 22	M	26	55	D	-/+	2.62	4.15	1	2	No	No
Pt. 23	M	26	118	D	-/+	3.94	5.09	0	1	Yes	Yes
Pt. 24	M	26	110	C	-/+	6.31	3.82	4	5	No	Yes
Pt. 25	F	27	23	D	-/+	6.70	4.03	1	2	Yes	Yes
Pt. 26	M	29	81	C	-/+	8.22	4.49	2	5	No	Yes

Group 1: HBeAg positive/HBeAb negative; ALT ≤40 (median 32 IU/L); HBV DNA ≥ 7.50 (median 8.58 log IU/ml)
Group 2: HBeAg positive/HBeAb negative; ALT >40 (median 83 IU/L); HBV DNA >7.00 (median 8.26 log IU/ml)
Group 3: HBeAg negative/HBeAb positive; ALT ≥ 40 with HBV DNA at any level, or if ALT ≤40 with HBV DNA >3.3 (median ALT 81 IU/L; median HBV DNA 4.19 log IU/ml)

* Columns indicating whether or not peripheral T cell analyses and IHC were carried out. T cell results are presented in Figure 2 and IHC in Figure 3. All samples were analyzed for HBV DNA integration and clonal hepatocyte expansion.

HAI, histological activity index

Table 2: Observed and Expected Integration Sites per Chromosome

Chromosome	Chromosome length	Integration Sites	
		Observed	Expected
1	$2.49x10^8$	44	40.9
2	$2.43x10^8$	47	39.9
3	$1.98x10^8$	34	32.6
4	$1.90x10^8$	39	31.2
5	$1.82x10^8$	26	29.8
6	$1.71x10^8$	28	28.1
7	$1.59x10^8$	25	26.2
8	$1.45x10^8$	19	23.8
9	$1.38x10^8$	21	22.7
10	$1.34x10^8$	21	22.0
11	$1.35x10^8$	19	22.2
12	$1.33x10^8$	18	21.9
13	$1.14x10^8$	20	18.8
14	$1.07x10^8$	16	17.6
15	$1.02x10^8$	13	16.8
16	$9.03x10^7$	9	14.8
17	$8.33x10^7$	12	13.7
18	$8.04x10^7$	10	13.2
19	$5.86x10^7$	19	9.6
20	$6.44x10^7$	13	10.6
21	$4.67x10^7$	7	7.7
22	$5.08x10^7$	8	8.3
X	$1.56x10^8$	25	19.2
Y	$5.72x10^7$	1	2.3

Expected integration sites per chromosome were calculated assuming that the incidence of integration was proportional to chromosome length. Integration incidence for the X and Y chromosome were adjusted to account for the equal numbers of males and females in the patient population.

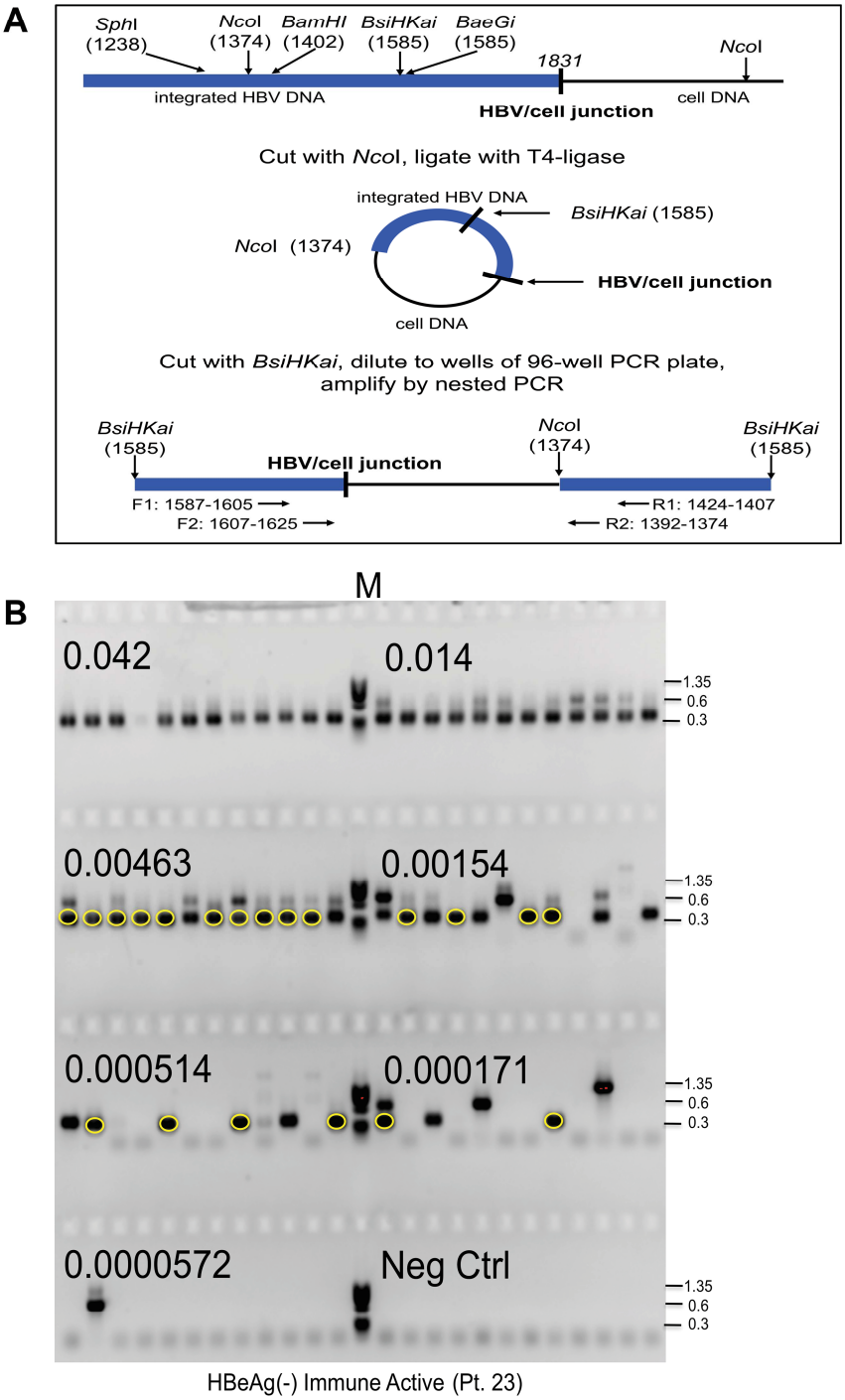
A
SphI (1238)
NcoI (1374)
BamHI (1402)
BsiHKai (1585)
BaeGi (1585)
1831
NcoI
integrated HBV DNA
HBV/cell junction
cell DNA
Cut with NcoI, ligate with T4-ligase
integrated HBV DNA
BsiHKai (1585)
NcoI (1374)
HBV/cell junction
cell DNA
Cut with BsiHKai, dilute to wells of 96-well PCR plate, amplify by nested PCR
BsiHKai (1585)
HBV/cell junction
NcoI (1374)
BsiHKai (1585)
F1: 1587-1605
F2: 1607-1625
R1: 1424-1407
R2: 1392-1374
B
M
0.042
0.014
0.00463
0.00154
0.000514
0.000171
0.0000572
Neg Ctrl
1.35
0.6
0.3
HBeAg(-) Immune Active (Pt. 23)

Figure 1

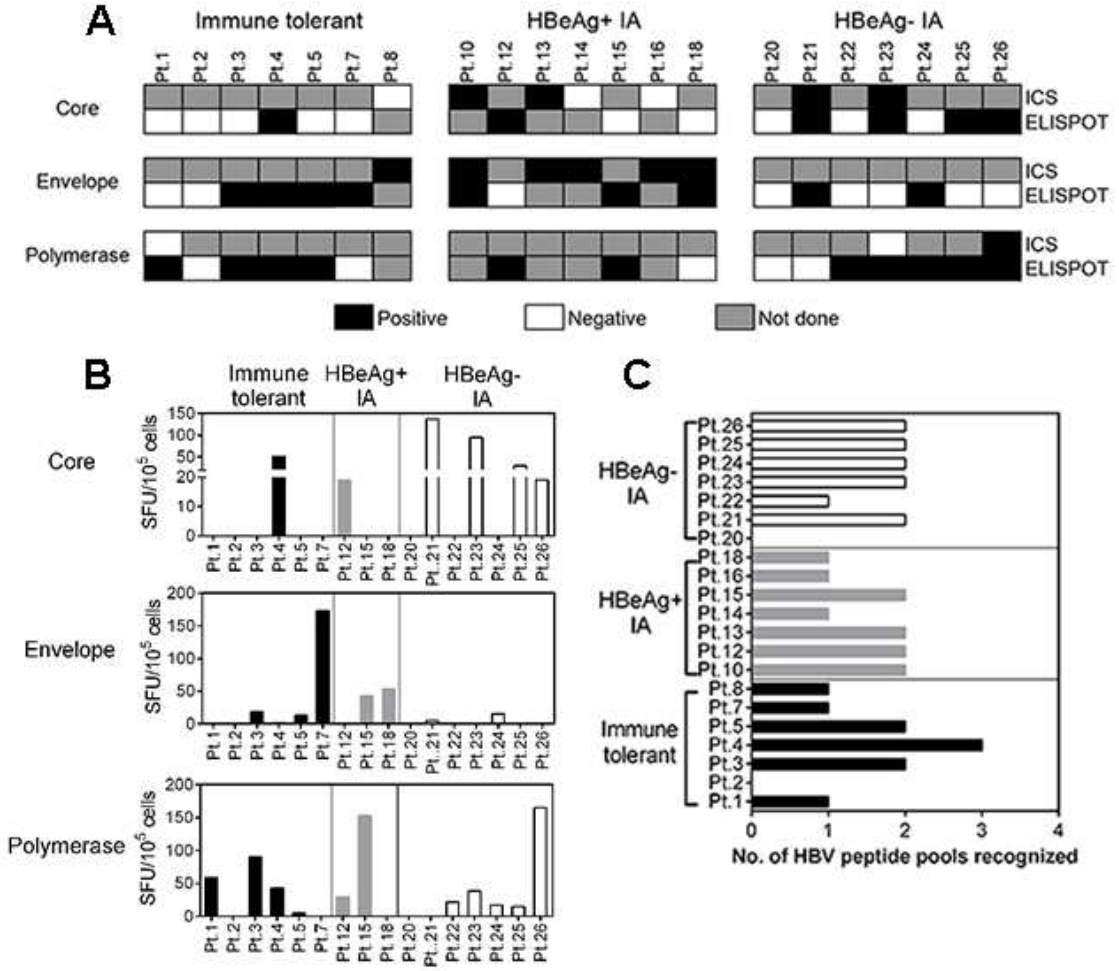
A
Immune tolerant
HBeAg+ IA
HBeAg- IA
Pt.1 Pt.2 Pt.3 Pt.4 Pt.5 Pt.7 Pt.8
Pt.10 Pt.12 Pt.13 Pt.14 Pt.15 Pt.16 Pt.18
Pt.20 Pt.21 Pt.22 Pt.23 Pt.24 Pt.25 Pt.26
Core
Envelope
Polymerase
ICS
ELISPOT
Positive
Negative
Not done
B
Immune tolerant
HBeAg+ IA
HBeAg- IA
SFU/10^5 cells
C
HBeAg- IA
HBeAg+ IA
Immune tolerant
No. of HBV peptide pools recognized

Figure 2

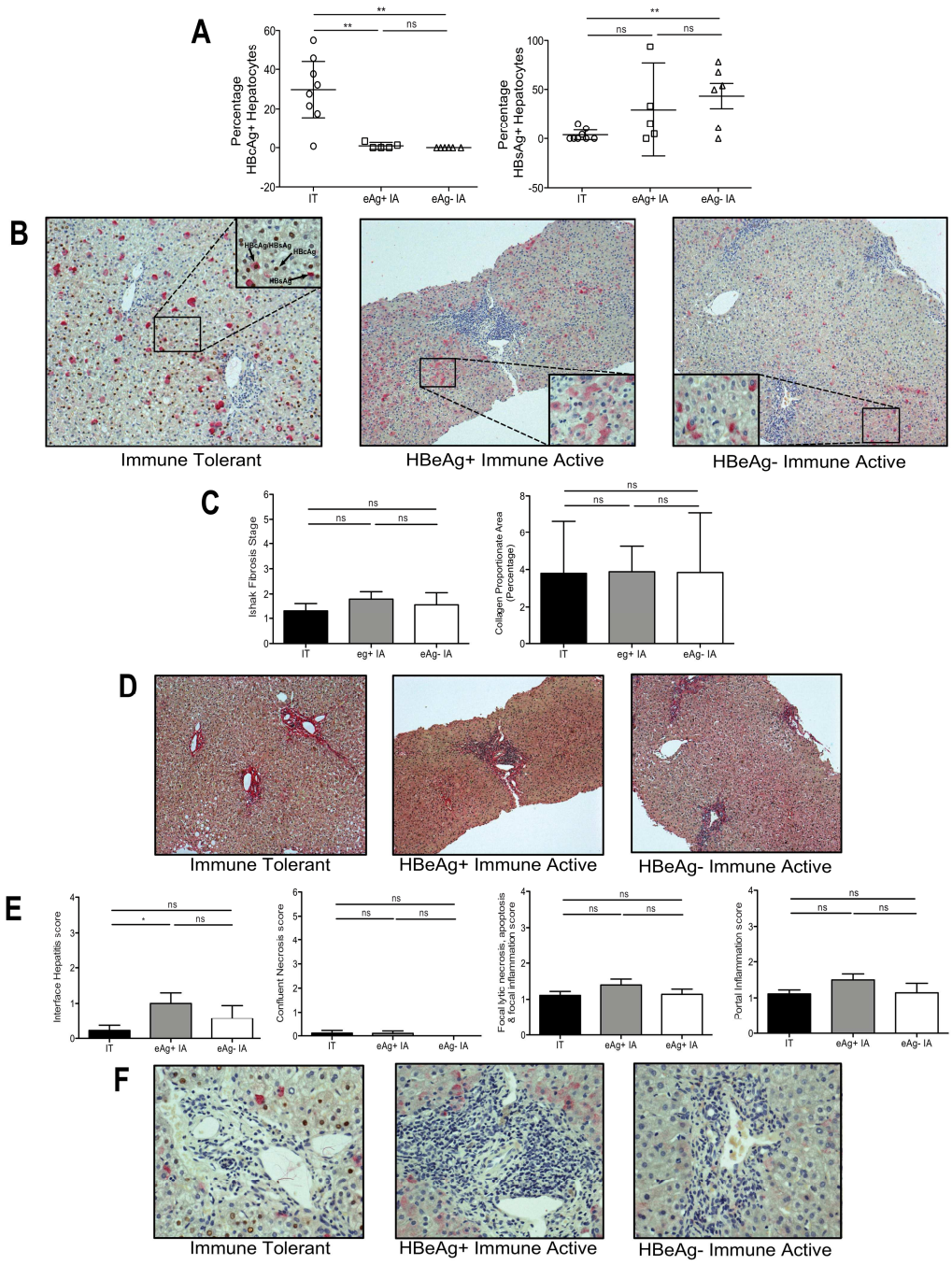
A
Percentage HBcAg+ Hepatocytes
Percentage HBsAg+ Hepatocytes
IT
eAg+ IA
eAg- IA
B
Immune Tolerant
HBeAg+ Immune Active
HBeAg- Immune Active
C
Ishak Fibrosis Stage
Collagen Proportionate Area (Percentage)
D
Immune Tolerant
HBeAg+ Immune Active
HBeAg- Immune Active
E
Interface Hepatitis score
Confluent Necrosis score
Focal lytic necrosis, apoptosis & focal inflammation score
Portal Inflammation score
F
Immune Tolerant
HBeAg+ Immune Active
HBeAg- Immune Active

Figure 3

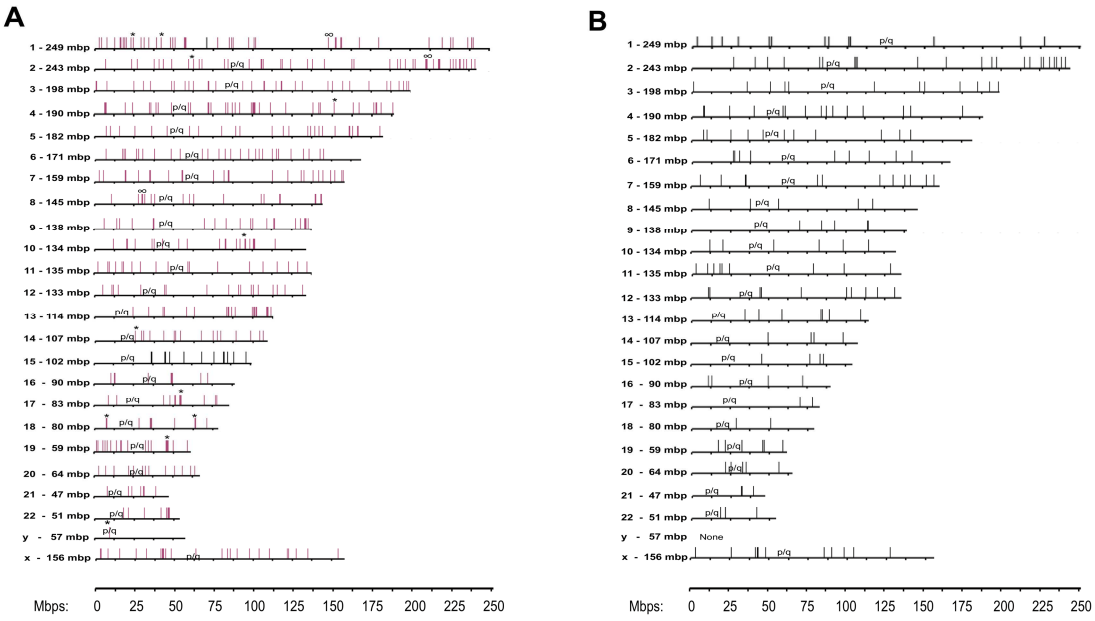

Figure 4

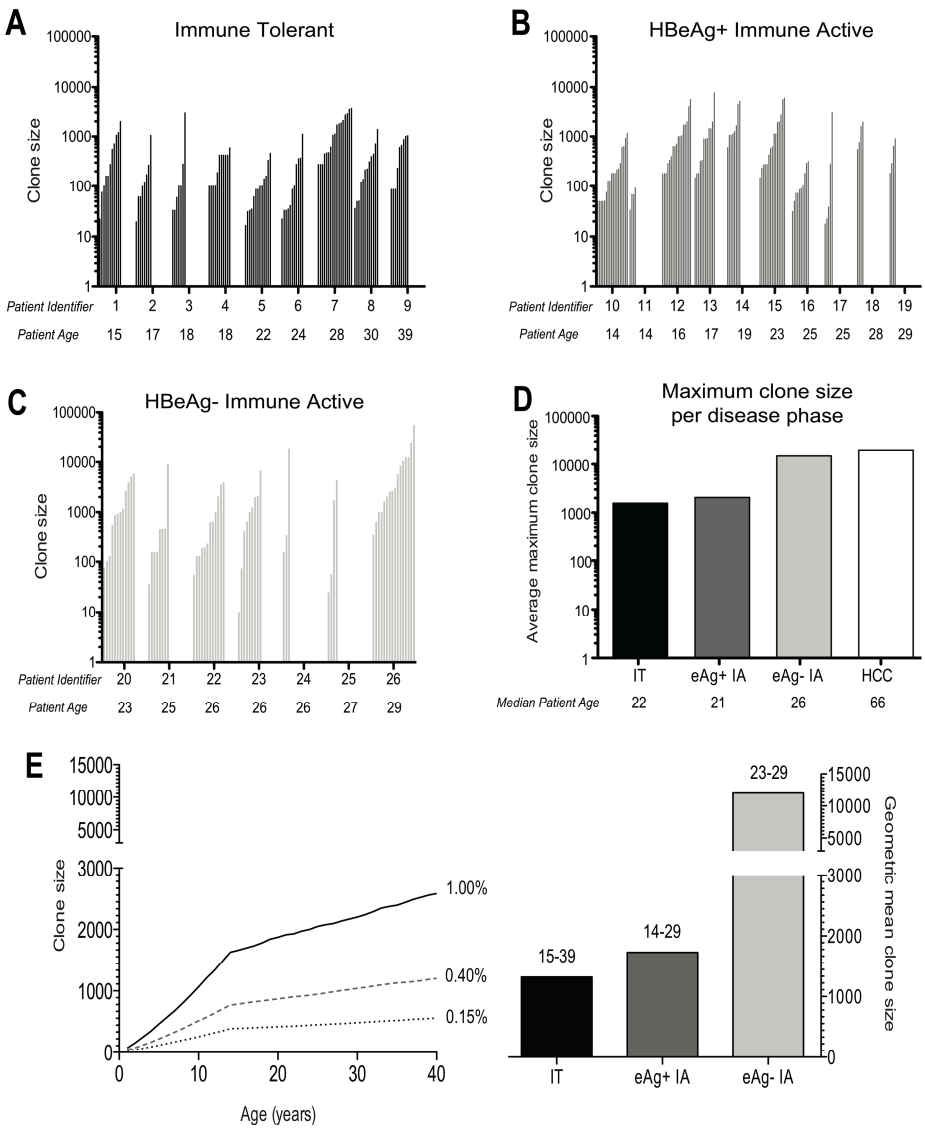
A
Immune Tolerant
Clone size
100000
10000
1000
100
10
1
Patient Identifier 1 2 3 4 5 6 7 8 9
Patient Age 15 17 18 18 22 24 28 30 39
B
HBeAg+ Immune Active
Clone size
Patient Identifier 10 11 12 13 14 15 16 17 18 19
Patient Age 14 14 16 17 19 23 25 25 28 29
C
HBeAg- Immune Active
Clone size
Patient Identifier 20 21 22 23 24 25 26
Patient Age 23 25 26 26 26 27 29
D
Maximum clone size
per disease phase
Average maximum clone size
IT eAg+ IA eAg- IA HCC
Median Patient Age 22 21 26 66
E
Clone size
15000
10000
5000
3000
2000
1000
0
1.00%
0.40%
0.15%
0 10 20 30 40
Age (years)
23-29
14-29
15-39
IT eAg+ IA eAg- IA
Geometric mean clone size

Figure 5

SUPPLEMENTARY MATERIALS & METHODS

In vitro expansion of HBV-specific T cells

HBV proteome: core (35) envelope (72), X (29) and polymerase (177) peptides of differing HBV genotypes (HBV A/D and HBV B/C) were available. A detailed list of the peptides used to stimulate PBMCs of the patients was published in Tan *et al.*[1] and matched according to the infecting genotype. Twenty percent of PBMCs were first stimulated with 10 µg/ml of the different overlapping peptide mixtures from the respective HBV genotypes for 1 hour at 37℃, then washed and resuspended at 3.0 x 10^6 cells/ml before co-culturing with the remaining PBMCs in serum AIM-V supplemented with interleukin-2 (IL-2, R&D systems, Abingdon, UK) (20 IU/ml), seeded at 1 ml/well in 24-well plates.

Intracellular cytokine and IFN-γ ELISPOT assays

IFN-γ ELISPOT assays were performed as previously described,[1] using a panel of 313 overlapping peptides covering the full HBV proteome sequence pooled in the described mixtures. HBV-specific T cell responses were analyzed in IFN-γ ELISPOT assays after short-term peptide-specific polyclonal T cell expansion (10-days). Briefly, 96-well plates (Multiscreen-HTS Millipore, Billerica, MA) were coated overnight at 4℃ with 5 µg/ml capture mouse anti-human IFN-γ monoclonal antibody (1DIK, Mabtech, Sweden). Plates were then blocked with AIM-V supplemented with 10% heat inactivated fetal calf serum (FCS) for 30 minutes at room temperature. $5x10^4$ cells from short-term polyclonal T cell lines were seeded per well, in duplicates for each individual peptide mixture. Plates were incubated for 18 hours at 37℃ in the presence or absence of peptides (at a final concentration of 5µg/ml). After this incubation, plates were developed using the alkaline phosphatase substrate (5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium chloride; BCIP/NBT, KPL, MD) according to the recommended protocol from Mabtech. The colorimetric reaction was stopped after 10-15 minutes by washing with distilled water. Plates were air-dried and spots were counted using an automated ELISPOT reader (Immunospot, CTL, OH). The number of peptide specific IFN-γ secreting cells

was calculated by subtracting the non-stimulated control value from the stimulated sample. Positive controls consisted of PBMC stimulated with the Phorbol-Myristate-Acetate (10ng/ml) and Ionomycin (100 ng/ml). Wells were considered positive when the SFU is above 5 and at least 2 times the mean of unstimulated control wells (3 wells/patient).

For intracellular cytokine staining, *in vitro* expanded PBMC were incubated in medium alone (control) or with viral peptides (5µg/ml) for 5 hours in the presence of brefeldin A (10 mg/ml). After washing, the cells were stained with anti-CD8 Pe-Cy7 and anti-CD3 PerCp-Cy5.5 mAb (BD Biosciences) for 30 min at 4℃, fixed, and permeabilized using Cytofix/Cytoperm™ Fixation/Permeabilization solution (BD Biosciences, San Jose, CA), according to the manufacturer's instructions. Cells were stained with anti-IFN-γ-PE (BD Biosciences, San Jose, CA) for 30 min on ice, washed, and analyzed by flow cytometry. Positive responses were considered as those with a frequency of IFN-γ-producing T cells at least twice the frequency found in unstimulated cells and where values exceeded 0.1% of total T cells.

Immunohistochemistry & Image Analysis

Two 4µm-thick serial sections were cut from each tissue block using a Leica RM2235 rotary microtome (Leica Biosystems, UK) and picked up on poly-l-lysine coated slides. The first section was stained with Sirius red according to standard staining protocols (Liver Histopathology Department, Institute of Liver Studies, King's College Hospital), followed by HBsAg immunohistochemistry on a fully automated IHC and ISH Leica BOND-MAX immunostainer (Leica Biosystems, Newcastle, UK) using a Leica Bond Polymer Refine Detection kit (code DS9800, Novocastra, Newcastle, UK). On the second slide double epitope immunohistochemistry for HBsAg and HBcAg was performed using the same immunostainer, the Polymer Refine Detection kit for HBcAg, and the Leica Bond Polymer Refine Red Detection kit (code DS9390, Novocastra, Newcastle, UK). Rabbit polyclonal anti-HBcAg antibody (Dako); (concentration 1:10,000) and mouse monoclonal anti-HBsAg (Dako), (concentration 1:600) were used. Slides

were then dehydrated with alcohol, cleared with xylene and cover slipped with DPX (Leica Biosystems, UK).

Collagen Proportionate area

Collagen proportionate area (CPA) was calculated using the following formula: Total Collagen Area/Total Tissue Area × 100. Total parenchymal area was determined as the difference between total tissue area and collagen area. Hepatocytes were counted on 10 random high power fields (HPF) within the parenchymal area, and the total number of hepatocytes per biopsy was estimated as the number of hepatocytes/μm^2 times the total parenchymal area. Numbers and percentages of HBcAg-, HBsAg-, and double-positive hepatocytes were assessed. For cluster analysis, a cluster was defined as at least two adjacent HBcAg+ and/or HBsAg+ hepatocytes (data not shown). The number, size, and composition of clusters in terms of HB positivity, as well as the number of isolated HBcAg+, HBsAg+, double-positive, and negative hepatocytes, were assessed on 5 random HPF within the parenchyma of each biopsy. The Modified Hepatic Activity Index (HAI) and Ishak Fibrosis Stage were assessed by a liver histopathologist (AQ) who was blinded to the clinical data.

Extraction of liver DNA and inverse PCR analysis

In brief, each biopsy sample was placed in 400µl of 0.05M TRIS-HCl, pH 7.8, 0.01M EDTA, 0.1M NaCl, 5% (w/v) SDS, 1mg/ml proteinase-K, and incubated for 2 hours at -55°C with occasional vortexing for a few seconds. After cooling, the sample was extracted with an equal volume of phenol:CCl_3 (1:1), and nucleic acids in the aqueous phase were precipitated by addition of 20µg of dextran carrier and 2 volumes of 100% ethanol. Samples were stored for at least 16 hours at -20°C, and nucleic acids were collected by centrifugation, washed once with 1ml of 100% ethanol, vacuum dried, dissolved in EB buffer (Qiagen), and stored at -80°C.

Preparative electrophoresis in 1% low-melt agarose gels containing E buffer (0.04M TRIS-HCl, 0.02M Na Acetate, 1 mM EDTA, pH 7.2, 0.5µg/ml ethidium bromide) was carried out as described previously.[2] The region of the gel containing high molecular weight DNA (≥10-20kbp), as visualized by ethidium bromide staining, was excised and equilibrated overnight with the appropriate New England Biolabs (NEB) restriction endonuclease buffer (Supplementary Table 1). Restriction digestion followed by intra-molecular ligation with T4 DNA ligase was carried out as illustrated in Figure 1A. The DNA was then suspended to 450µl in T4 DNA ligation buffer (NEB) and incubated for 2 hours at room temperature with 500 units T4 DNA ligase to facilitate intra-molecular ligation, to join the NcoI site in viral DNA the nearby NcoI site in cell DNA (Figure 1A). The T4 DNA ligase was heat inactivated for 20 min at 72℃, and the products recovered by ethanol precipitation. The circularized DNAs were then cleaved by restriction endonuclease digestion (Supplementary Table 1, Figure 1A) to produce linear molecules in which the virus/cell junction is flanked by viral DNA sequences. Following serial dilution into microtiter trays (12 wells per dilution), nested PCR was carried out using the HBV specific forward primers F1 and F2 and reverse primers R1 and R2 (Supplementary Table 1, Figure 1A). The F1 primer sequence (nts 1587 to 1605) was 5'-TTCGCTTCACCTCTGCACG-3' (#1380). The F2 sequences (nts 1607-1625) were 5'-CGCATGGAGACCACCGTGA-3' (#1385); 5'-CGCATGGAAACCACCGTGA-3' (#1529); 5'-CGCATGGCGACCACCGTGA-3' (#1532); 5'-TGCATGGAAACCACCGTGA-3' (#1533); 5'-CGCATGGAGGCCACCGTGA (#1535). The R1 primer sequences (nts 1424-1407) were 5'-AAAGGACGTCCCGCGCAG-3' (#1383); 5'-AAAGGACGTCCCGCGAAG-3' (#1395); 5'-AAAGGACGTCCCTCGCAG-3' (#1530); 5'-AAAGGACGTCCCCCGCAG-3' (#1534). The R2 primers sequences (nts 1392-1374) were 5'-CACAGCCTAGCAGCCATGG-3' (#1382); 5'-TACAACCTAGCAGCCATGG-3' (#1393); 5'-CACACCCTAGCAGCCATGG-3' (#1437). The primer sets used for each patient are summarized in Supplementary Table 1. Reaction 1 was carried out using the AmpliTaq Gold reagents (Applied Biosystems) with 0.25units polymerase per 10µl reaction. ~0.5µl was transferred using a pin replicator to reaction 2. The 2nd PCR reaction was carried out using 10µl of GoTaq reagents (Promega) with 0.25units polymerase per 10µl reaction.

The right hand virus/cell junction, on the viral genome, was detected for 500 of 592 integrations; for the remainder, uncertainty in sequence reads near the virus/cell junction or insertion in repeated DNA prevented precise mapping. ~92% of the viral junctions with host DNA were between nt 1688 and nt 1829, the 5' end of HBV minus strand DNA (data not shown), similar to previous reports[3, 4]. Integration sites on the human genome were located using BLAST of human genome build GRCh38. Location of integration junctions to introns or exons was determined using the Integrated Genome Browser[5] genome version H_Sapiens_Dec_2013.

Statistical analyses: Serial dilution of DNA samples for inverse PCR.

In this situation, the starting material contains a finite number of different virus/cell junctions. The problem is to estimate the copy number, *n*, for a repeated junction using the results from serial dilutions. From the point of view of quantifying *n* for a particular virus/cell junction, there are two technical problems. First, at lower dilutions, its presence in a given PCR well may be obscured by more abundant virus/cell junctions. Thus, not all dilutions may be informative with respect to copy number, and we were often required to estimate clone sizes from the higher sample dilutions, where there were only a few positive wells. This competition problem also means that we may totally fail to detect some repeated fragments because they are of low abundance compared to others. Though we have seen instances of clone sizes as low as ~20, these will be obscured when very large clones are present in a sample.

A Monte Carlo approach was therefore used to devise a statistical model to give a clone size range with 95% confidence, as well as a best-fit point estimate. For example, if there was no competition, then we might see that the seven rows with increasing, 3-fold sample dilutions contained 12,12,12,9,7,3 and 1 positive wells, respectively, for a given virus/cell junction fragment. That is, a data vector, $data_0$, with the values (12,12,12,9,7,3,1). To simulate this, we considered selection of a fraction *f* of the virus/cell junctions for distribution to 12 wells at each dilution of *n* virus/cell junctions,

including the initial dilution into the first 12 wells. The probability would be 2/3 for the next six 3-fold dilutions. If informative data is obtained from all 7 wells, then in the simulation, the vector (k_1,k_2,…,k_7) would be obtained for an assumed value of *n* and compared to the data using a least squares fit, S, as follows:

$$S_{01} = \sum_{i=1}^{7} w_i (k_i - data_{0i})^2 ,$$

where *w* = (√729, √243, √81, √27, √9; √3, √1) is used to weight the data at each dilution, on the assumption that the lower dilutions have higher value. S_{01} measures the fit for the first of a series of random trials for a given value of *n*. S_{02},…$S0_{0,1000}$ for the same value of *n* are similarly produced in subsequent trials. All of these values are averaged to give $\underline{S}_{0,}$ which measures the average fit of random process output to the given data for this value of *n*.

To determine the fit of *n* to the data, a single random trial of the process is executed with *n* and the output used as a simulated data vector, $data_1$. The process described above is now run using $data_1$ and used to produce an average least squares fit, $\underline{S}_1$. The process is repeated again, starting with a new simulated data vector, $data_2$, to produce $\underline{S}_2$, and so on using data sets $data_3$ through $data_{99}$. These 100 trials give rise to average least squares fits $\underline{S}_1$ through $\underline{S}_{99}$ to simulated data sets for the value *n*. Each value of $\underline{S}$ for the simulated data sets is compared to $\underline{S}_0$. If the tally $S_j \geq S_0$ for j = 1 through 99 is less than 5, then the hypothesis that $data_0$ came from a process determined by *n* can be rejected at the 5% level.

To obtain a confidence region and best fit for *n*, a binary search can now be used to vary *n* and decide which values are in a, say, 95% confidence region. For example, with *f* = 0.1, and $data_0$ = (12,12,12,9,7,3,1) we found a best fit estimate of *n* = 6779 and that *n* is in the range 3405-13747 with 95% confidence. In practice, a vector may have the values (-,-,-,-,7,3,1), where the negative sign indicates that no information was available at the these dilution. The program, Sim19, is available upon request (Samuel.Litwin@fccc.edu).

Modelling of Clonal Expansion of hepatocytes.

The simulation Csize8 was designed to study a liver sample of up to 10,000,000 cells, with the object of tracking the number and sizes of distinct clones of cells. Hepatocytes at time zero were each given a unique identifier. This is based on the assumption that infection is neonatal. These hepatocytes then divide to expand liver size (in the simulations presented here, the liver was assumed to increase 10-fold in size, from 800,000 to 8,000,000 million hepatocytes, during the first fourteen years of life). During growth, and later, random hepatocytes die at a fixed rate and are replaced by replication of other random hepatocytes. Hepatocyte death leads to loss of some unique identifiers; *i.e.*, reduction of the total number of distinct clones. However, selection of random cells to replicate will increase the sizes of other clones. The number of clones of each size from 1 to the maximum clone size is finally tabulated at the end of the run. Fortran code or the compiled program is available upon request to Samuel.Litwin@fccc.edu or William.Mason@fccc.edu.

Bibliography

1. Tan AT, Loggi E, Boni C, et al. Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. J Virol 2008;82:10986-10997.
2. Mason WS, Low HC, Xu C, et al. Detection of clonally expanded hepatocytes in chimpanzees with chronic hepatitis B virus infection. J Virol 2009;83:8396-8408.
3. Tu T, Mason WS, Clouston AD, et al. Clonal expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. J Viral Hepatitis 2015;22:737-753.
4. Mason WS, Liu C, Aldrich CE, et al. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. J Virol 2010;84:8308-8315.
5. Nicol JW, Helt GA, Blanchard SGJ, et al. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 2009;25:2730-2731.

SUPPLEMENTARY RESULTS

Integrated HBV DNA was identified in chromosomes of all patients.

An inverse PCR protocol was designed to detect virus/host DNA junctions that occurred near the 3' end of HBV dslDNA (Figure 1), and to estimate the sizes of hepatocyte clones emerging subsequent to viral DNA integration.[1-3] After DNA inversion, the samples were serially diluted into microtiter trays. Virus/host junctions were then amplified by nested PCR, using HBV specific primers. The PCR products were detected by agarose gel electrophoresis, as illustrated in Figure 1B (Group 3, patient 23, Table 1). In this example, all DNA bands from the 5 highest sample dilutions were excised and sequenced to map the virus/cell junctions. Sequences were then compared using Sequencher software (Gene Codes Corp., Ann Arbor, MI), to identify virus/cell junction fragments resulting from the same integration event. For example, the circled bands (Figure 1B) all revealed the same virus/cell DNA junction. Assuming this virus/cell junction occurs once per hepatocyte, it is present in a clone of ~12,000 hepatocytes (Supplementary Table 3, patient 23, clone 82). In fact, this clone was larger than ~12,000, because the junction was also detected in an adjacent fragment of the liver biopsy from this patient. The remaining bands in Figure 1B represented either smaller clones, or virus/host junctions that occurred only once.

Overall, two distinct integrations into the same gene occurred six times, including "cell adhesion molecule L1-like" (gene symbol CHL1) and "alcohol dehydrogenase 1B (class I), beta polypeptide" (ADH1B) on chromosome 3, "vav 2 guanine nucleotide exchange factor" (VAV2) on chromosome 9, "myosin XVI" (MYO16) and "sodium leak channel, non selective" (NALCN) on chromosome 13, and "mbt domain containing 1" (MBTD1) on chromosome 17. Only the two chromosome 17 integrations, in MBTD1, were observed in the same patient, 16. In addition to integrations within genes, another 46 integrations mapped within 10,000 nts of a gene, 101 between 10,000 and 100,000 nts, and 103 between 100,000 and 1,000,000 nts (Supplementary Table 3).

While integrants appeared to be distributed at random across most chromosomes, three possible exceptions were noted. Using the Komogorov-Smirnov test, a higher than expected accumulation of the integrations on chromosomes 2 and 13 appeared near their right hand ends, from nt 187112119 to the right hand end of chromosome 2,

and from nt 83040027 to right hand end of chromosome 13 (Figure 4A). The lack of integrations adjacent to the centromere on chromosome 1, between nts 101626622 and 147668340, also appeared statistically significant, occurring only 62 out of 10,000 times in simulations of random distributions of integration sites across this chromosome.

Bibliography

1. Mason WS, Liu C, Aldrich CE, et al. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. J Virol 2010;84:8308-8315.
2. Yang W, Summers J. Integration of hepadnavirus DNA in infected liver: evidence from a linear precursor. J Virol 1999;73:9710-9717.
3. Gong SS, Jensen AD, Chang CJ, et al. Double-stranded linear duck hepatitis B virus (DHBV) stably integrates at a higher frequency than wild-type DHBV in LMH chicken hepatoma cells. J Virol 1999;73:1492-1502.

Supplementary Table 1: Restriction endonuclease digestions and primers for Inverse PCR

	1st Cut	2nd Cut	3rd Cut	Primers for Inverse PCR			
Group 1: Immune Tolerant				**F1**	**F2**	**R1**	**R2**
Pt. 1	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1530	1382
Pt. 2	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1437
Pt. 3	NcoI-HF	BaeGI	SphI	1380	1385	1383	1382
Pt. 4	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 5	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1534	1382
Pt. 6	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1437
Pt. 7	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1530	1382
Pt. 8	BamHI-HF	BaeGI	NcoI	1380	1385	1383	1382
Pt. 9	NcoI-HF	BaeGI	SphI	1380	1385	1383	1382
Group 2: HBeAg(+) IA							
Pt. 10	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 11	NcoI-HF	BsiHKAI	SphI-HF	1380	1533	1395	1393
Pt. 12	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 13	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 14	NcoI-HF	BsiHKAI	SphI-HF	1380	1529	1383	1382
Pt. 15	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1395	1393
Pt. 16	NcoI-HF	BaeGI	SphI	1380	1385	1383	1382
Pt. 17	NcoI-HF	BsiHKAI	SphI-HF	1380	1535	1530	1382
Pt. 18	BamHI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 19	NcoI-HF	BaeGI	SphI	1380	1385	1383	1382
Group 3: HBeAg(-) IA							
Pt. 20	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 21	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1530	1382
Pt. 22	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 23	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382
Pt. 24	BamHI-HF	BsiHKAI	NcoI-HF	1380	1532	1383	1382
Pt. 25	NcoI-HF	BaeGI	SphI	1380	1385	1383	1382
Pt. 26	NcoI-HF	BsiHKAI	SphI-HF	1380	1385	1383	1382

Digestion 1 was carried out after equilibration of gel slices with NEB4. Digestion 2 with BsiHKAI and SphI (or NcoI) was carried out in NEB4 supplemented with BSA. BaeGI and SphI (or NcoI) digestion 2 was carried out in NEB1. The relevant cleavage sites on the HBV genome for BamHI, BaeGI, BsiHKAI, NcoI and SphI are nt 1402, 1585, 1585, 1374, and 1238, respectively (relative to HBV accession number V01460). PCR primer sequences are presented in Materials and Methods.

Supplementary Table 2: HBV DNA Integration in Liver Biopsies

Group 1: Immune Tolerant	Total cells[1]	Distinct integrants / total integrants	Total integrants per liver	Distinct integrants sites per liver	Fraction of hepatocytes with integrated HBV DNA
Pt. 1	$4.7x10^{5}$	17/3134	$3.3x10^{9}$	$1.8x10^{7}$	$9.5x10^{-3}$
Pt. 2	$3.1x10^{6}$	34/2067	$3.4x10^{8}$	$5.5x10^{6}$	$9.7x10^{-4}$
Pt. 3	$2.2x10^{6}$	23/2254	$5.2x10^{8}$	$5.2x10^{6}$	$1.47x10^{-3}$
Pt. 4	$1.9x10^{6}$	41/3100	$8.2x10^{8}$	$1.1x10^{7}$	$2.3x10^{-3}$
Pt. 5	1.09x10	28/1833	$1.3x10^{7}$	$1.3x10^{7}$	$2.4x10^{-3}$
Pt. 6	1.47x10	28/2134	$7.3x10^{8}$	$8.8x10^{6}$	$2.1x10^{-3}$
Pt. 7	$5.6x10^{6}$	46/22300	$1.98x10^{9}$	$4.1x10^{6}$	$5.7x10^{-3}$
Pt. 8	$2.4x10^{6}$	23/3168	$6.5x10^{8}$	$4.7x10^{6}$	$1.87x10^{-3}$
Pt. 9	$7.2x10^{6}$	23/4634	$8.7x10^{8}$	$1.6x10^{6}$	$8.7x10^{-4}$
Group 2: HBeAg+ IA					
Pt. 10	1.28x10	36/3088	$1.21x10^{9}$	$1.4x10^{7}$	$3.5x10^{-3}$
Pt. 11	$2.4x10^{6}$	15/300	$6.2x10^{7}$	$3.1x10^{6}$	$1.79x10^{-4}$
Pt. 12	1.85x10	27/9408	$2.5x10^{9}$	$7.3x10^{6}$	$7.3x10^{-3}$
Pt. 13	$4.6x10^{6}$	35/11280	$1.21x10^{9}$	$3.7x10^{6}$	$3.5x10^{-3}$
Pt. 14	$5.5x10^{6}$	19/18900	$1.7x10^{9}$	$1.7x10^{6}$	$4.9x10^{-3}$
Pt. 15	$3.6x10^{6}$	33/16900	$2.4x10^{9}$	$4.6x10^{6}$	$6.7x10^{-3}$
Pt. 16	$2.4x10^{6}$	31/1840	$3.9x10^{8}$	$6.6x10^{5}$	$1.12x10^{-3}$
Pt. 17	$3.3x10^{5}$	9/811	$1.24x10^{9}$	$1.4x10^{7}$	$3.6x10^{-3}$
Pt. 18	$4.7x10^{6}$	7/11616	$1.25x10^{9}$	$7.5x10^{5}$	$3.6x10^{-3}$
Pt. 19	$9.2x10^{5}$	9/2016	$1.09x10^{9}$	$4.9x10^{6}$	$3.1x10^{-3}$
Group 3: HBeAg- IA					
Pt. 20	$3.3x10^{6}$	22/12528	$1.89x10^{9}$	$3.3x10^{6}$	$5.4x10^{-3}$
Pt. 21	1.91x10	21/7080	$1.86x10^{9}$	$5.5x10^{6}$	$5.3x10^{-3}$
Pt. 22	$9.5x10^{5}$	20/12960	$6.8x10^{9}$	$1.1x10^{7}$	$1.96x10^{-2}$
Pt. 23	$2.4x10^{6}$	19/17568	$3.7x10^{9}$	$4.0x10^{6}$	$1.05x10^{-2}$
Pt. 24	$8.3x10^{5}$	6/9008	$5.4x10^{9}$	$3.6x10^{6}$	$1.55x10^{-2}$
Pt. 25	$3.9x10^{6}$	5/3952	$5.0x10^{8}$	$6.4x10^{5}$	$1.44x10^{-2}$
Pt. 26	$4.4x10^{6}$	28/74878	$8.54x10^{9}$	$3.2x10^{6}$	$2.44x10^{-2}$

Total integrants in a biopsy were corrected for the fraction of the biopsy that was analyzed. Distinct integrations were not corrected, since we did not know if analyzing a larger fraction of a biopsy specimen would reveal new integration sites. Thus, distinct integration sites may represent a minimum estimate of those actually present. Total and distinct integration sites per liver were estimated assuming that human liver contains $5x10^{11}$ hepatocytes. Total cells in a biopsy, as determined by qPCR (Materials and Methods), include non-hepatocytes. Subsequent calculations were corrected with the assumption that 70% of total cells are hepatocytes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	**Table S3: HBV DNA Integration Sites***																	
2	**patient number**	**Group**	**Age**	**sequence no**	**clone no.**	**Clone size**	**Chrm**	**Chrm insert sites**	**gene symbol**	**insert lccation**	**Gene polarity**	**gene to left**	**distance to left hand gene**	**gene to left polarity**	**gene to right**	**distance to right hand gene**	**gene to right polarity**	**cell junction sequence (lower to higher sense)**
3	1	1	15	4042	253	2035	X	89717084				CPXCR1	-962300	+	TGIF2LX	204856	+	AGCCATATTTCTGTTCTT--HBV
4	1	1	15	4050			3	150124112				TMEM183B	-140746	-	LINC01213	114406	+	GCCAGCAGGAGCTTTGCT--HBV
5	1	1	15	3209	185	23	12	102998812				ASCL1	-38296	+	C12orf42	238778	-	AAAATGTAACAAGAGTC--HBV
6	1	1	15	3186	182	282	22	20814502	PI4KA	intron	-							HBV--AGTTTTCGAAGGCTGAGG
7	1	1	15	3183	184	78	1	155314931	FDPS	intron	+							TGGTGGTGGCTGAAACAAGATTGTC-HBV
8	1	1	15	3203	181	165	2	196661734	CCDC150	intron	+							HBV--ATTGCCCTTAATAGTACTT
9	1	1	15	3202	183	165	4	6942259	TBC1D14	intron	+							ACTGAATTATTTCATGCCTTC--HBV
10	1	1	15	3188			7	156990872				NOM1	-17690	+	MNX1	13980	-	HBV--nnGACTGTGTCTATCATTGGG
11	1	1	15	3191			6	142418190	ADGRG6	intron	+							TAATCCCAGCTACTACTCn--HBV
12	1	1	15	3199			13	43244958	ENOX1	intron	-							within 130 nts to right if 43244958
13	1	1	15	3208			6	114662146				HS3ST5	-599269	-	>1000000			HBV--GGTGGTTTAAGCAAAATAAAT
14	1	1	15	3334	190	103	19	45113949	PPP1R37	intron	+							HBV--TGGCTCTGCTTGCTCTTT
15	1	1	15	3337	188	732	2	187112119				ZSWIM2	-262949	-	CALCRL	229843	-	GCAGTTATGTATATGTTTCT--HBV
16	1	1	15	3368	189	1078		0										
17	1	1	15	3339	187	571	1	12184317	TNFRSF1B	intron	+							HBV--TCATTGTTTAGCCATGAC
18	1	1	15	3338	186	1223	15	44884789				TRIM69	-117017	+	C15orf43	71903	+	HBV--AAAAAAGGAAAAAAAAACTCATTTGGAGT
19	2	1	17	3296	206	20	1	2970771				TTC34	-181180	-	LINC00982	88845	-	GCTGATAACAGCCGGGC--HBV
20	2	1	17	3291	205	102	4	84240667	LOC101928978	intron								HBV--nCCAAACATATTAATTGGGAT
21	2	1	17	3274	204	275	1	84772117				SSX2IP	-81560	-	LPAR3	41285	-	CCCGTGTATCATCTAAGAGG--HBV
22	2	1	17	3260				0										
23	2	1	17	3264			X	97542265	DIAPH2-AS1	intron	-							AGTCAGAAAGTCTTCnnnnnnnnnnnnnnnnnnnnnnnn--HBV
24	2	1	17	3272				0										
25	2	1	17	3276			12	99976747	ANKS1B	intron	-							TTAATAAGAAACATGAT--HBV
26	2	1	17	3278			4	176197346				SPATA4	-1675	-	ASB5	16328	-	TAAAGGGAAAAGTCAAAGTGnnnnnn--HBV
27	2	1	17	3283			3	147148414				PLSCR5	-542198	-	ZIC4	237633	-	GGCACATAGACCAATGGAA--HBV
28	2	1	17	3290				0										
29	2	1	17	3303	200	176	5	45712223				HCN1	-16105	-	>1000000			HBV--TGGAATAATATGCTTGC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
30	2	1	17	3324	201	118	2	193675671				PCGEM1	-898772	+	LOC101927406	668597	+	HBV--AAATCAGAGATAACAGCATCC
31	2	1	17	3310	202	63	13	83919439				SLITRK1	-37046	-	LINC00333	221162	+	HBV--TTTTAAAAAATATAACT
32	2	1	17	3332	203	63	4	111109946				PITX2	-467823	-	>1000000			ATATATATATATTAAAATTnn--HBV
33	2	1	17	3298				0										
34	2	1	17	3299			2	164157155				FIGN	-421152	-	GRB14	335250	-	HBV--AACTATGATAGTGGTGA
35	2	1	17	3304				0										
36	2	1	17	3306			3	197990949	LMLN	intron	+							HBV--nnAATGACNGATNATTTTCTTTGGTGTCTG
37	2	1	17	3313			18	28103851	CDH2	intron	-							TACTTTTATATGGCTGCA-HBV
38	2	1	17	3316			1	100312508				RTCA	-19739	+	CDC14A	39958	+	AGTCATTTCTGTGTGATGGCTnnnnnnn--HBV
39	2	1	17	3317			X	40821270				MED14OS	-82569	+	LOC100132831	9946	-	HBV--AGTAAGTACCAATGGC
40	2	1	17	3318			4	87006259	AFF1	intron	+							within 40 nts to left of nt 80192447
41	2	1	17	3325				0										
42	2	1	17	3329				0										
43	2	1	17	3331				0										
44	2	1	17	3533	216	1087	5	9591420				SNHG18	-41123	+	TAS2R1	37576	-	ACATTAAAATCAAGAACTT--HBV
45	2	1	17	3535			14	79214645	NRXN3	intron	+							HBV--AAATCTGGATGTGGATCC
46	2	1	17	3540				0										
47	2	1	17	3541				0										
48	2	1	17	3542			2	83784072				LOC1720	-926303	+	FUNDC2P2	506609	+	within 110 nts to right of nt 83784072
49	2	1	17	3544			10	20668718				MIR4675	-116672	+	NEBL	111255	-	HBV--CTGTGTCCCCACCCGA
50	2	1	17	3546			19	46739731	STRN4	intron	-							GGCCCAACTTAGACACTTA--HBV
51	2	1	17	3555			2	231222422	ARMC9	intron	+							HBV--CAGCCTAATAAGGTTCCT
52	2	1	17	3563			4	58975164				>1000000			>1000000			HBV--CACTATTGATAATGTCAA
53	3	1	18	3803	222	34		0										
54	3	1	18	3812	220	103	3	117606720				LINC00901	-674482		>1000000			HBV--AATGACACATCTAGTCT
55	3	1	18	3819	221	103	11	128419044				>1000000			ETS1	39716	-	HBV-AGTGTATCAGGTAT
56	3	1	18	3806			3	198096501				ANKRD18DP	-15830	-	FAM157A	55864	+	TGTATGAATG TATGAAAAC--HBV
57	3	1	18	3807			6	132078040				CTGF	-126868	-	MIR548AJ1	37151	-	HBV--CCTCAACCCAAGACCC
58	3	1	18	3809			6	26014093				TRIM38	-26764	+	HIST1H1A	2938	-	TTACCTTCCTCCCTTGGA--HBV
59	3	1	18	3816			10	114424498				AFAP1L2	-19720	-	ABLIM1	6611	-	TGAGCGTAAAGCTTAGn-HBV
60	3	1	18	3817			7	141417781	TMEM178B	intron	+							HBV--nnnnnnnnnnnnAGCAGTTACTTTGTGCC
61	3	1	18	3820				0										
62	3	1	18	3824			19	31760568				THEG5	-167018	+	LINC01533	265293	+	HBV--AGAAGTGGGCATTGGCAC
63	3	1	18	3827				0										

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
64	3	1	18	3828			11	77870013	AAMDC	intron	+							TACCTAACCC TGAGACAGnnnnnnnnnnnn--HBV
65	3	1	18	3826			X	47283158				USP11	-34830	+	ZNF157	87441	+	HBV--TTTCTGCACATTTGTGTCA
66	3	1	18	3780	223	3024		0										
67	3	1	18	3914	225	34	20	56183987				CBLN4	-178515	-	MC3R	64744	+	HBV-CAATTACCCAGTCTTA
68	3	1	18	3933	224	61	14	76815170				ANGEL1	-2230	-	LRRC74A	11211	+	HBV--GTATAAACTCTCAATG
69	3	1	18	3940	226	286		0										
70	3	1	18	3913			12	43966271	TMEM117	intron	+							AAGAAGGGGATAAGAGn--HBV
71	3	1	18	3938			3	192177304	FGF12	intron	-							HBV--AGGCCTATAAAATTAT
72	3	1	18	3944			13	58018244				PCDH17	-289313	+	LOC101926897	147582		HBV--TGGTGACACATACCTGTAGT
73	4	1	18	3572	218	433	22	17962139	MICAL3	intron	-							CTTGCTGACTGCACATATCACATGnn--HBV
74	4	1	18	3592	219	193	5	59451480	PDE4D	intron	-							HBV--GATAACCGACATATAC
75	4	1	18	3567	217	611	2	233776390	MROH2A	intron	+							HBV--CTGTTGCCCAGGCTGGA
76	4	1	18	3566			11	17429330	ABCC8	intron	-							HBV--ATCCCACCCTCAAGGCAA
77	4	1	18	3583			1	18438305				IGSF21	-59822	+	KLHDC7A	42624	+	HBV--GGGCTCAAATCCCTGTTCACAGC
78	4	1	18	3584			16	12584020				SNX29	-9731	+	CPPED1	75778	-	HBV--nCTAAGTGTCCCTCCTGTGGT
79	4	1	18	3586			1	86733852	SH3GLB1	intron	+							CTGTGGTGGGAGTATTTATA--HBV
80	4	1	18	3588			2	82108916				LOC100507201	-641970	-	LOC1720	747886	+	HBV--AATTTCTATTGGGATTCACC
81	4	1	18	3394	194	103		0										
82	4	1	18	3385	191	102	10	53190646				MBL2	-418946	-	PCDH15	612124	-	TTTCAAAAAATTAATTTTTA--HBV
83	4	1	18	3397	192	433	9	92115984				SPTLC1	-510	-	LOC100128076	16849	+	within 40 nts to left of 92115984
84	4	1	18	3383				0										
85	4	1	18	3388			21	30829731	KRTAP7-1	at beginning of exon	-							HBV--TGTGAAGGGTAAGTTACCCA
86	4	1	18	3389			2	225661861				NYAP2	-7843	+	LOC646736	480932	+	HBV--TTACAGACATGCACCAC
87	4	1	18	3391				0										
88	4	1	18	3401				0										
89	4	1	18	3403			7	34592630	NPSR1-AS1	intron	-							within 40 nts to right of nt 34592630
90	4	1	18	3406			5	80192447	SERINC5	intron	-							within 200 nts to right of nt 80192447
91	4	1	18	3408			8	37724134	LOC101929622	intron	+							HBV--ACTCCAGCCCTCCTCTGAGC
92	4	1	18	3409				0										
93	4	1	18	3410			X	42702365				CASK	-779331	-	PPP1R2P9	75000	-	GCTTTCTATATGTTGATGACTn--HBV
94	4	1	18	3381			15	76549357	SCAPER	intron	-							HBV--CAAATGTCCAACAAT
95	4	1	18	3414	196	103	6	101585725	GRIK2	intron	+							GGAAAAAGGTGACACAT--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
96	4	1	18	3426	199	103	12	119899421				CIT	-22130	-	CCDC64	90422	+	TTATAGCCCTACTAGAGTTGn--HBV
97	4	1	18	3424	199	433	14	49382149				>1000000			RPS29	194522	-	CCCTATTGACTCCAGT--HBV
98	4	1	18	3427	195	433	5	25236337				LOC340107	-395754	-	>1000000			ACCCTGTCAACCACTAT--HBV
99	4	1	18	3432	197	433		0										
100	4	1	18	3421			3	311592	CHL1	intron	+							TCACCTGGAGGAGAGATAGCAnnn--HBV
101	4	1	18	3428			12	11576834				LINC01252	-12433	+	ETV6	73019	+	GTAAGATGGG AGGTTTTACA--HBV
102	4	1	18	3433				0										
103	4	1	18	3434			19	22193758	ZNF676	intron	+							TAGGACCTTGTGCTTTTCCTA--HBV
104	4	1	18	3441			5	134702101	SEC24A	intron	+							HBV--TTATCACAATATCTTGAAAAC
105	4	1	18	3447			1	50290811				ELAVL4	-87026	+	DMRTA2	127853	-	AACATGCCCCATACCTGCn--HBV
106	4	1	18	3449				0										
107	4	1	18	3451				0										
108	4	1	18	3452			20	21547651				NKX2-2	-33625	-	LOC101929625	22372	+	HBV--GTGCATTCATTCAGAGGT
109	4	1	18	3453			21	31409049	TIAM1	intron	-							HBV--nnnnnnCATTGCTCCCACACTCAT
110	4	1	18	3454			13	101380239	NALCN	intron	-							AGAATTCAAGACTAACAA--HBV
111	4	1	18	3443			10	11454352				CELF2	-117679	+	USP6NL	6157	-	HBV--CCAGGCCTCCCTCACTG
112	5	1	22	3837	228	34	18	50627176	MAPK4	intron	+							HBV--TTGGACCAGCCTTGGGGAATGG
113	5	1	22	3859	231	17	7	18845205	HDAC9	intron	+							HBV--CTCACTCACCTGTTAATTT
114	5	1	22	3868	230	36	2	227134742	COL4A4	intron	-							HBV--nGAAAGAATGACTCACTCACG
115	5	1	22	3857	229	103	7	130562490	COPG2	intron	-							ATGTATTGCCTATGACAAG--HBV
116	5	1	22	3831	227	342	2	106990648				ST6GAL2	-103541	-	MIR548AU	159016	-	AAACACTGACACATGTTGG--HBV
117	5	1	22	3840			20	24835069				SYNDIG1	-168452	+	CST7	114160	+	HBV--TCACTCAGTTTGTGTTTTGGT
118	5	1	22	3846				0										
119	5	1	22	3852				0										
120	5	1	22	3856				0										
121	5	1	22	3858			6	91787314				CASC6	-96886	-	>1000000			GTAGAAAACAATAGGATTT--HBV
122	5	1	22	3861				0										
123	5	1	22	3866				0										
124	5	1	22	3867			4	91384710	CCSER1	intron	+							GCAACAAAATTACAGATTGT--HBV
125	5	1	22	3870			13	83040027				>1000000			RNU6-67P	258043	-	TGAGTACTCA AAAATGATnnn-HBV
126	5	1	22	3871				0										
127	5	1	22	3889	233	89	2	40345643	SLC8A1	intron	-							HBV--TTGAAAACCCTCTATGCAGCC
128	5	1	22	3872	232	472	15	85124435	PDE8A	intron	+							HBV-- CCACCCTCCTCCCCAGCTC
129	5	1	22	3896			12	10664319	STYK1	intron	-							CACATCCCTAACTTCCTATT--HBV
130	5	1	22	3948	235	145	5	36000096	UGT3A1	intron	-							TGGAGAGGCCCAATGGAAT--HBV
131	5	1	22	3962	238	63		0										

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
132	5	1	22	3965	238	89	3	172866200				ECT2	-44726	+	SPATA16	23156	-	GCTTTTATAGTTGAGGCC--HBV
133	5	1	22	3949	234	102	3	59285332				C3orf67	-543323	-	FHIT	463977	-	TACCTGGGTTTGAACACTGTTnn--HBV
134	5	1	22	3952	237	32	7	34326405				BMPER	-170533	+	NPSR1-AS1	20106	-	HBV--ATCCACAAAACACCTCC
135	5	1	22	3960	236	165	16	13311889				SHISA9	-76479	+	ERCC4	-610142	+	HBV--AAAATTAAATGAACTAAAAT
136	5	1	22	3963			2	238285839	PER2	intron	-							HBV--AGCCCTGCAGGATTTTACAACC
137	5	1	22	3967				0										
138	5	1	22	3972			5	122419019	SNCAIP	intron	+							within 30 nts to right of nt 122419019
139	5	1	22	3977			21	38923345	LOC400867	intron	-							TGTTTCTGAATTACATGTC--HBV
140	6	1	24	4330	240	89	19	7905028	MAP2K7	intron	+							GTACCTCTCGCTTCACACACAC--HBV
141	6	1	24	4316	241	103	2	6468883				LINC01247	-93461	-	LINC00487	260284	-	AAACAACACAGCATTCTCTAAA--HBV
142	6	1	24	4332	240	368	9	36836065	PAX5	intron	-							HBV--CTGCCAAGGCTGGGCAGTCGTT
143	6	1	24	4312			9	127214489	RALGPS1	intron	+							CGATTTCCTTTGTATTTTATnnnnnnnnnnnnnnn-HBV
144	6	1	24	4321			1	96939372				PTBP2	-124323	+	DPYD	138371	-	within 700 nts to left of nt 96939372
145	6	1	24	4322			1	70708693	LOC101927244	intron	+							GTCAGACAAGTAATGACTGnnnnnnnn--HBV
146	6	1	24	4325			4	178108873				LINC01098	-118123	+	>1000000			CTGATTTAATAATGTATATnnnnnnnnnn--HBV
147	6	1	24	4327			2	117314398				>1000000			DDX18	500280	+	HBV--TAATTATAGCTTGTCCTC
148	6	1	24	4331				0										
149	6	1	24	4265	244	34	4	177597656				AGA	-155153	-	LINC01098	131100	+	ACTTAGAAGCTTTACCTC--HBV
150	6	1	24	4267	244	34	4	6436575	PPP2R2C	intron	-							TGTTACTAGAGTTCCCTnnnnnnnnn--HBV
151	6	1	24	4282	244	23	18	35811667				GALNT1	-99833	+	MIR187	93150	-	TTTATGTCCAACAGAC--HBV
152	6	1	24	4299	243	281	1	38590990				LINC01343	-367223	-	RRAGC	260624	-	AAAAGAAAAAAATTAACTTAG--HBV
153	6	1	24	4286	242	1133	10	109779347				>1000000			XPNPEP1	85418	-	TTTGGGCCATTACAGTATCAGC--HBV
154	6	1	24	4288				0										
155	6	1	24	4291			3	197076755	DLG1	intron	-							GTACAAAGGCCCAGCCTAGCA--HBV
156	6	1	24	4293			10	37470521				LINC00993	-123493	+	MTRNR2L7	130916	-	HBV--AGGCAGGAGAATCGCTTGA
157	6	1	24	4248	246	36	15	96814536				SPATA8	-28921	+	LINC00923	928079	-	TTTTTTGTGTTTTTAAAAAG C--HBV
158	6	1	24	4241	247	42	16	48633860				N4BP1	-23651	-	CBLN1	645543	-	HBV--TGGGTCCAGACCACTTAAG
159	6	1	24	4239	245	381	1	225490526	ENAH	intron	-							HBV--ATCGCACCACTGCACTCC
160	6	1	24	4237			2	137801750				THSD7B	-124033	+	LOC101928273	77003	-	TTGAGATGTTTTTCAGACTT--HBV
161	6	1	24	4238			10	20246294	PLXDC2	intron	+							TGTGTGTCTGTGTGACGAnnnn--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
162	6	1	24	4240			X	42769983				CASK	-846949	-	PPP1R2P9	7382	-	CACCTCCATAACATACTT--HBV
163	6	1	24	4244			13	109667013				MYO16	-460177	+	LINC00676	61260	+	CCCCCGCGTACAGGCACCCC--HBV
164	6	1	24	4246			3	7363687	GRM7	intron	+							GGTCTCATCCAAT--HBV
165	6	1	24	4257				0										
166	6	1	24	4259			5	160337249	CCNJL	intron	-							GGTCTTGAACTCCTGGGCTG--HBV
167	6	1	24	4262			X	7273398	STS	intron	+							HBV--CAAGTCATAGGACTGAAAA
168	7	1	28	3092	173	1078	16	10557512	EMP2	intron	-							HBV--nnAAAGTTTTTCCTTGG
169	7	1	28	3088	169	1149	1	49161242	AGBL4	intron	-							HBV--TCTGGAGCCGAGTACATGGTTCCT
170	7	1	28	3093	168	2136	X	2621766	CD99P1	intron	-							HBV--TTATTGATTACTCATGTAC
171	7	1	28	3104	172	3730	7	84597817				LOC101927378	-13499	+	SEMA3D	397738	-	HBV--ACTATTCACAATAGCAAAAG
172	7	1	28	3106	170	2919	1	28945637	EPB41	intron	+							AGTGTTTAATATTCTGTT--HBV
173	7	1	28	3108	171	3583	4	60356734				>1000000			MIR548AG1	565884	+	within 40 nts to left of nt 101626622
174	7	1	28	3085			1	101626622				LINC01307	-249309	+	OLFM3	175944	-	ACACTGCTTCCTGCATCCCAGC--HBV
175	7	1	28	3087			9	133048728	GTF3C5	intron	+							within 150 nt to left of nt 133048728
176	7	1	28	3089			4	141763033				IL15	-29046	+	INPP4B	265748	-	CTCACACCTGTAATCCTAGCn--HBV
177	7	1	28	3090				0										
178	7	1	28	3094			X	42098007				CASK	-174973	-	PPP1R2P9	679358	-	HBV--nnnnnnnnnnnnnnTAAACTCGGCCCAGTCCTAGTT
179	7	1	28	3096			22	41681252	NHP2L1	intron	-							TCAAAATTTATTTTTGTTCTTA--HBV
180	7	1	28	3097			6	37762238				MDGA1	-64248	-	ZFAND3	57292	+	ATTTCTGAGATAGGACTTGnn--HBV
181	7	1	28	3098			2	58948077	LINC01122	intron	+							AAATTTGCAT CCACAACCATCATA--HBV
182	7	1	28	3103			9	83598638				FRMD3	-60205	-	IDNK	24410	+	GAGCCAGGGTCAGAGATGA--HBV
183	7	1	28	3109				0										
184	7	1	28	3115			4	137562796				PCDH18	-30298	-	LINC00616	464626	-	HBV--AATTCTCAATCATTTTCCTTT
185	7	1	28	3490	212	458	12	45245364	ANO6	intron	-							TTAATAATTGTTAATATA--HBV
186	7	1	28	3468	211	488	3	50252682	GNAI2	intron	+							AGGACCAGCCTGGCCAACTTG--HBV
187	7	1	28	3456	208	1741	11	2162764				INS	-1555	-	TH	1164	-	CTGAGCCATGCCACAGCnn--HBV
188	7	1	28	3465	209	1910	20	34711080	TP53INP2	exon	+							GCACCAAGGGAGTGTGCA--HBV
189	7	1	28	3460	210	1826		0										
190	7	1	28	3459	207	2763	4	24192173				PPARGC1A	-302096	-	MIR573	328018	-	TGGGAAAAATGTTG--HBV
191	7	1	28	3469			2	240851310				KIF1A	-31002	-	AGXT	17434	+	GGTCCTCCTGCCCCACAT--HBV
192	7	1	28	3473			13	33533086	STARD13	intron	-							within 224 nts to right of nt 33533086

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
193	7	1	28	3474			7	81645681	LOC100128317	intron								within 246 nts to left of nt 81645681
194	7	1	28	3476			12	131196616	LINC01257	intron	+							ACAACGTTCACTTTTGTT--HBV
195	7	1	28	3478			4	7378006	SORCS2	intron	+							HBV--TACCCCGATTTTCATGAAG
196	7	1	28	3488			13	88267974				LINC00397	-457454	-	LINC00433	272854	+	HBV--AAAGAGCTTCAGCACAGCA
197	7	1	28	3491			17	77440729	SEPT9	intron	+							GTCTGCACAGGTGCCATC--HBV
198	7	1	28	3493			7	121815754				FAM3C	-419386	-	PTPRZ1	57350	+	HBV--ATTCTGAACATCACTAAT
199	7	1	28	3118	176	282	7	138078614	AKR1D1	intron	+							HBV--CTCAAAGCTATATTCTC
200	7	1	28	3124	177	282	4	73404654	ALB	intron	+							CTAAGGAAAGTGCAAAG--HBV
201	7	1	28	3147	178	283	16	49649494	ZNF423	intron	-							TCAGCACATATTTAGGAT--HBV
202	7	1	28	3127	174	484	9	113373775	HDHD3	exon	-							HBV--AATTATCCCCAACATGG
203	7	1	28	3151	175	635	11	9570069				ZNF143	-41545	+	WEE1	3611	+	HBV--AAACAAAACATTGTTAAG
204	7	1	28	3125			8	116876744	RAD21-AS1	exon	+							HBV--nCAGCCATAAAAAAGAAAAA
205	7	1	28	3134			3	34918259	LOC101928135	intron	-							HBV--ATCCCACAATAGGAAA
206	7	1	28	3135			3	184416743				CHRD	-26908	+	EPHB3	145055	+	CAGGTGCGCCATAACT--HBV
207	7	1	28	3137				0										
208	7	1	28	3139			4	99319381	ADH1B	intron	-							within 65 nts to right of nt 99319381
209	7	1	28	3143			X	25732927				ARX	-716979	-	MAGEB18	405415	+	AAACCATAATATACTTCCT--HBV
210	7	1	28	3145			7	151906852				PRKAG2-AS1	-27629	+	GALNTL5	49526	+	HBV--GGCCATTTTCACAATATA
211	7	1	28	3146			4	38669922	KLF3	intron	+							HBV--AAGATGTCCTCTTAA
212	7	1	28	3148				0										
213	7	1	28	3150			19	58353102	A1BG	exon	-							HBV--GAGTCTCCAGGTGGGC
214	8	1	30	2190	89	124	2	26707285	KCNK3	intron	+							CCCTGGAATGGAGTGCG--HBV
215	8	1	30	2198	87	229	3	61826795	PTPRG	intron	+							HBV--TGATAGCCAAATATAAAACGTTC
216	8	1	30	2160	88	220	2	105319156	TGFBRAP1	intron	-							AAAAATAAAAAAATCAACAn--HBV
217	8	1	30	2164				0										
218	8	1	30	2166			11	97872304				>1000000			>1000000			GCTGGCTGAGCCCAGCA--HBV
219	8	1	30	2168			16	71569561	TAT-AS1	intron	-							HBV--TAACTCCTGCGCTCAAGTG
220	8	1	30	2176			7	5495114	FBXL18	intron	-							HBV--AGGAAAAGGCTCCCACTGC
221	8	1	30	2185			5	141370942	PCDHGB3	exon	+							HBV--AGGAGAACCTGGATGGCAG
222	8	1	30	2281	94	52	2	145043331	TEX41	intron	+							HBV--GAAGTATTCCAAGATAACTTCT
223	8	1	30	2301	93	37	6	26933909	GUSBP2	intron	-							TTTCTACAGG TCTGAAGGTGnnnnnn-HBV
224	8	1	30	2300	91	320		0										
225	8	1	30	2278	90	1416	2	48690599	LHCGR	intron	-							CGTAGTTCTCTGAATAAGTTCT--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
226	8	1	30	2283			X	85244654	ZNF711	intron	+							AACCTTTGGAATGCGCGAGGG--HBV
227	8	1	30	2310			11	18736493	PTPN5	intron	-							HBV--TTTATGTCTAAATCCTCA
228	8	1	30	2376	99	51	6	29796259	LOC554223	intron	+							CCTTGGCTAGAAAGAGGTCAT--HBV
229	8	1	30	2372	97	141		0										
230	8	1	30	2385	96	456	2	235320752				SH3BP4	-265038	+	AGAP1	173336	+	HBV--AGAAGAGTAACCTCACGG
231	8	1	30	2371	98	407		0										
232	8	1	30	2365	95	727		0										
233	8	1	30	2394			12	112612606				PTPN11	-102693	+	MIR1302-1	82427	-	HBV--CCAGGCTGGAGTGCAATGA
234	8	1	30	2401			9	69450228	APBA1	intron	-							within 50 nts to right of 72065144
235	8	1	30	2403			2	105251384				GPR45	-7917	+	TGFBRAP1	13005	-	HBV--TTGTCTGAGATGTGGGGA
236	8	1	30	2404				0										
237	9	1	39	3498	213	680	X	127219938				PRR32	-398153	+	ACTRT1	831023	-	HBV--CTACATGTTTCCTTTTT
238	9	1	39	3497	215	1030	14	97905846				LOC100129345	-219188	-	LINC01550	19763	-	TTAGTGTCTCTCTGCCC--HBV
239	9	1	39	3503	214	621	8	55581800				XKR4	-55649	+	TMEM68	156943	-	within 40 nts to right of nt 55581800
240	9	1	39	3504				0										
241	9	1	39	3507			20	32473272	NOL4L	intron	-							within 40 nts to right of nt 32473272
242	9	1	39	3514			12	70661649	PTPRR	intron	-							HBV--CCAAACCCAGCACGGGTGAGGG
243	9	1	39	3520			8	107067262				ABRA	-297018	-	ANGPT1	182219	-	CAGGCAACAGGCAATA--HBV
244	9	1	39	3522			17	69641717	LINC01483	intron	+							TTTGTTTAAGGGGAA--HBV
245	9	1	39	3526			10	82316697	NRG3	intron	+							HBV--TCTACTTCTAAAAAGGA
246	9	1	39	3033	161	89	11	13471813				BTBD10	-8516	-	PTH	20240	-	HBV-nACATTCTCTATGTGGAAAGG
247	9	1	39	3035	162	89	5	65472122	ADAMTS6	intron	-							within 40 nts to left of nt 65472122
248	9	1	39	3037	163	89	2	214293678	SPAG16	intron	+							HBV--GTGGAGCTCATGGTTTGCT
249	9	1	39	3014			X	103653329				TCEAL1	-22381	+	MORF4L2	22168	-	within 368 nts to right of nt 103653329
250	9	1	39	3019			11	23679528				CCDC179	-819102	-	LUZP2	817441	+	HBV--AAGAGGAGCTTTATTGAG
251	9	1	39	3020			5	7013040				MIR4278	-185119	-	MIR4454	256262	-	HBV--CTTGATGATTAGTCAT
252	9	1	39	3034			19	16397387	EPS15L1	intron	-							HBV--GCTGTTTGTCTATTTTAC
253	9	1	39	3040			15	83141577	HDGFRP3	intron	-							CTCCATGACTGACAGTTTTnnn--HBV
254	9	1	39	3045			1	211474161				LINC00467	-41626	+	RD3	2360	-	HBV--GATAGAGTAGTTGGGGACT
255	9	1	39	3047				0										
256	9	1	39	3065	166	238	8	10720956				C8orf74	-20363	+	SOX7	2811	-	HBV--CAACCCAGCACAACAAATGGAT
257	9	1	39	3055	164	1063	1	227011883	CDC42BPA	intron	-							within 40 nts to right of nt 227011883
258	9	1	39	3082	165	881	10	99810494	ABCC2	intron	+							HBV-- AGTGTGTGGGATTGGGAA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
259	9	1	39	3068			2	217797054				DIRC3	-40461	-	TNS1	2734	-	HBV--CTGTCTCGAAAGAAAAAAAGAAAA
260	10	2	14	2541	109	52	9	108587789				>1000000			ACTL7B	266799	-	HBV--TTTTTTCCTGAAATTCCAGGTGAC
261	10	2	14	2548	108	51	5	118860870	DTWD2	intron	-							HBV--TTAGAAATGATTTTGC
262	10	2	14	2551	107	131	5	15736111	FBXL7	intron	+							ATGTAAAGAAAAGATAATTT A--HBV
263	10	2	14	2535			2	202829835	ICA1L	intron	-							ACTTTCTAAGTTTTGCTT--HBV
264	10	2	14	2538			8	59680145				TOX	-560937	-	CA8	508718	-	CTCCAGCATTTTGAGTA--HBV
265	10	2	14	2539			10	79033393	ZMIZ1-AS1	intron	-							HBV--nnACCGGTTTACATCATTCA
266	10	2	14	2543				0										
267	10	2	14	2544			X	80115837				TBX22	-84068	+	CHMP1B2P	112651	-	TGATTCTCATTCTCTCTCT--HBV
268	10	2	14	2546			20	30421656				MLLT10P1	-18194	-	DEFB115	836007	+	TGCATATGGAATGTCTG--HBV
269	10	2	14	2641	114	51	15	56886666	LOC145783	exon	-							CATAATATAGCACTTTCT--HBV
270	10	2	14	2619	111	187	7	27605977	HIBADH	intron	-							HBV--CAAACAAGTTATTCATCCCA
271	10	2	14	2648	112	187	22	46582920				CELSR1	-45750	-	GRAMD4	43840	+	HBV--CTTTCTCTCTCTTTTGTCT
272	10	2	14	2639	113	131	4	35618971				>100000			ARAP2	447026	-	HBV--nnnnnnnnnnnGCCTAAAGCCCTTTGTTC
273	10	2	14	2655	110	229	9	113532006	RGS3	intron	+							GAGGAGCTGGGTGTC--HBV
274	10	2	14	2612			15	82533803				GOLGA6L17P	-7319	+	RPS17	2946	-	HBV--TCAAGCGATCCTCCTGCCT
275	10	2	14	2618				0										
276	10	2	14	2633			1	239589547				>1000000			CHRM3	39525	+	ATATAGTTTTATGTAATTTCC--HBV
277	10	2	14	2634			13	108744898	MYO16	intron	+							ACTACTTAAACCTTACC--HBV
278	10	2	14	2637			4	163110338				MIR4454	-16710	-	NAF1	16369	-	within 227 nts to right of nt 163110338
279	10	2	14	2654			20	64198457	MYT1	intron	+							CCTGATGCCT CAGCACCTGCn--HBV
280	10	2	14	2714	122	77	1	156213591	PMF1-BGLAP	intron	+							TCCTGCCGCAGCTTCCCAA--HBV
281	10	2	14	2710	123	51	1	220059053	BPNT1	intron	-							TGGATTGTGGTTTTTCTAG--HBV
282	10	2	14	2739	125	187	11	29116520				MIR8068	-638972	-	KCNA4	893220	-	TGCTCTCCTACACAGTAAGCAA--HBV
283	10	2	14	2728	120	220	6	78898764	IRAK1BP1	exon	+							HBV--TTGTAAGACAAAATTT
284	10	2	14	2731	121	297	9	135085520	OLFM1	intron	+							HBV--TAAGGATTGGGTGCCAGGCA
285	10	2	14	2747	124	642		0										
286	10	2	14	2750	118	622	14	34499424				SPTSSA	-37162	-	EAPP	16504	-	within 40 nts to right of nt 34499424
287	10	2	14	2727	119	938	16	13012870	SHISA9	intron	+							HBV--AACTTTCAAAAGGGT
288	10	2	14	2751	126	1183	7	55119726	EGFR	intron	+							HBV--AACCCACCCTGCCCTGGTT
289	10	2	14	2707			6	116104675	NT5DC1	intron	+							within 73 nts to right of nt 116104675

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
290	10	2	14	2715			19	1594102				MBD3	-1341	-	UQCR11	3052	-	HBV--CCGGGTTCAAGTAATTCTCCTG
291	10	2	14	2732			22	30891961	OSBP2	intron	+							TGGTCAGGGAGCAGCTTTGC--HBV
292	10	2	14	2736			12	90641768				LINC00936	-929816	+	LINC00615	276254	+	CCTCTTTTGCCCAAGTGGnnn--HBV
293	10	2	14	2744			13	62942327				LINC00448	-134968		LINC00376	240773	-	HBV--TTAATAATTGAAATTAATAATT
294	10	2	14	2749			9	24112457				ELAVL2	-286392	-	IZUMO3	430757	-	HBV--AGTGATTCTCCTGCCTCA
295	10	2	14	2753				0										
296	11	2	14	3716	276	69	8	139110753				COL22A1	-196747	-	KCNK9	490084	-	HBV--TGGCAACATCCTCAGGCACA
297	11	2	14	3718	277	94	12	84639238				>1000000			SLC6A15	220249	-	HBV--AGTTACTTCATAAATGTTGA
298	11	2	14	3712			3	93470513				>1000000			PROS1	402523	-	TTTGATGTAATTTTATATTTT--HBV
299	11	2	14	3724			17	48056784	NFE2L1	uncertain	+							within ~300 nts to right of nt 48056784
300	11	2	14	3730			1	56829818				C1orf168	-10122	-	C8A	24951	+	within ~650 nts to right of nt 56829818
301	11	2	14	3735				0										
302	11	2	14	3680	268	34	19	6696480	C3	intron	-							ACTCCTTCTGCAGGGTGAGTGA--HBV
303	11	2	14	3678				0										
304	11	2	14	3681				0										
305	11	2	14	3683			21	23640805				D21S2088E	-255970	-	LOC101927869	663744	-	TTTCTTATAATGGTGTATTTAT--HBV
306	11	2	14	3702	269	69	10	25705185	LINC00836	intron	+							HBV--nAGTGGTGTAATCTCGGCTCAC
307	11	2	14	3695				0										
308	11	2	14	3696			14	89211096	FOXN3	intron	-							HBV--AGTTTCATTACCTTTCAAC
309	11	2	14	3700			13	109021744	MYO16	intron	+							within 40 nts to left of nt 109021744
310	12	2	16	2581	130	183	3	24420709	THRB	intron	-							HBV--GGCAAGGTTGCAGAGTAAAAGGA
311	12	2	16	2597	132	642	2	23042543				LOC102723362	-504256		KLHL29	342883	+	AATAACTCTTTATACAGTTAn--HBV
312	12	2	16	2666	131	1008	15	35925508				DPH6-AS1	-66507	+	MIR4510	1347	+	GTGCAGCGGAGAGTGACCCAn--HBV
313	12	2	16	2589	129	2000	6	67321866				>1000000			LOC102723883	733484		HBV--nAGGTTAGACTTTTTGGTAC
314	12	2	16	2555	127	1733		0										
315	12	2	16	2567	128	1058	5	163114132				GABRG2	-958593	+	CCNG1	323438	+	ACAGGCAGAATTTCGAAT--HBV
316	12	2	16	2562			11	58955636	GLYATL1	exon	+							HBV--CAGCTGGATGTCTCTTATTC
317	12	2	16	2570				0										
318	12	2	16	2576				0										
319	12	2	16	2582			3	126580557	TXNRD3NB	intron	-							TGTGTGGATTCCAATCCCCA--HBV
320	12	2	16	2583			5	111924181	NREP	intron	-							GCTGGGCCTTGGATTTTGAAA--HBV
321	12	2	16	2690	137	187	6	70563526				FAM135A	-2352	+	SDHAF4	3395	+	CTAGGTTGCTGCAGTTTCATn--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
322	12	2	16	2667	134	342	8	143235361				GPIHBP1	-18191	+	ZFP41	11459	+	HBV--nGATTACAGGATCACACACA
323	12	2	16	2682	133	723	2	201601445	ALS2CR11	intron	-							HBV--AGAGAAGAGCAAGATAGAG
324	12	2	16	2700	135	407	10	105121512	SORCS3	intron	+							ATTGATTTCTCATTTCGTTTTT--HBV
325	12	2	16	2677			3	130938818	ATP2C1	intron	+							within 242 nts to right of nt 130938818
326	12	2	16	2762	144	187	3	98942382				DCBLD2	-41056	-	COL8A1	696213	+	within 40 nts to right of nt 98942382
327	12	2	16	2797	140	297	1	167529560				CD247	-10950	-	CREG1	11453	-	HBV--nnAGTTGTATTAGCTTTCTATTGC
328	12	2	16	2764	141	643	19	45585702				OPA3	-838	-	GPR4	4062	-	TCAGAAATAAATTTAAGGAC--HBV
329	12	2	16	2803	142	1008	6	135385661	AHI1	intron	-							CTATGAAACCTGTGAATTTnnnnnnnnn--HBV
330	12	2	16	2759	138	4023	9	98882830				GALNT12	-32749	+	COL15A1	60882	+	TTTCTTTCCCCTGTCCTTTC--HBV
331	12	2	16	2802	139	1733	18	70984299				GTSCR1	-333442	-	LINC01541	535664	-	ATGCCACTTACTAGGTGGACAAA--HBV
332	12	2	16	2809	143	5596	1	41476045				SCMH1	-233902	-	EDN2	2279	-	HBV--TCTTGCTCACGAGAGTCCTCCC
333	12	2	16	2765			2	97732650	ZAP70	intron	+							CATCTGTTGGCTCTTGGAGAAnn--HBV
334	12	2	16	2768				0										
335	12	2	16	2771				0										
336	12	2	16	2792				0										
337	12	2	16	2810			2	37547391				QPCT	-174564	+	CDC42EP3	94490	-	HBV--TCGTGGGGTGACAAAGGGG
338	13	2	17	1959	36	340	X	15162293				MOSPD2	-240956	+	ASB9	81693	-	CCAATAAGCTAGATTCATTnn--HBV
339	13	2	17	1921	34	907	14	106076891				ADAM6	-104622	-	LINC00226	210782	+	HBV--ATAAAATGTATATGCA
340	13	2	17	1955	36	1456	22	30596440	PES1	intron	-							HBV--CTCTGAAAGTGCTGGGATTA
341	13	2	17	1953	35	2000	3	194843049				LOC100507391	-60881	+	XXYLT1	225234	-	CTGTAATCCCAGCACTTT--HBV
342	13	2	17	1918			2	218668259	RNF25	exon	-							approx. HBV--nCCAGAGTGAAGCAGACATACTGT
343	13	2	17	1922			6	151518831	CCDC170	intron	+							GCAGGTTTTTATTAAGGACT--HBV
344	13	2	17	1925			19	11034381	SMARCA4	intron	+							within 272 nts to right of nt 11034381
345	13	2	17	1926			18	7331602				LRRC30	-99558	+	PTPRM	236216	+	HBV--nTCTGCCTCCCGGGTTCA
346	13	2	17	1927			X	63317858				>1000000			SPIN4	29369	-	within 270 nts to left of nt 63317858
347	13	2	17	1928			X	154096037	MECP2	intron	-							~ATGTTACTACTGAGGCCGTA--HBV
348	13	2	17	1930			8	104886603				LRP12	-297579	-	ZFPM2	432315	+	AAATCCTCTCTTTAGC--HBV
349	13	2	17	1940			10	42917286				BMS1	-82349	+	LINC01264	61730	-	HBV--AAAAAATTGAAAACAGC
350	13	2	17	1943			1	56570009	PPAP2B	intron	-							HBV--CTTATATAGTGTTTGTCA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
351	13	2	17	1952			22	45236843	KIAA0930	intron	-							CCCCTTCTCTGCACCA--HBV
352	13	2	17	1966			11	122647359				MIR100HG	-444297	-	UBASH3B	8330	+	AGACCCTTCAGGGACTGCCT--21n--HBV
353	13	2	17	1887	39	938	1	4160502				LINC01346	-207919	+	LOC284661	251548	+	~GTCATGTCTACCCTCTGCAGGGT--HBV
354	13	2	17	1880	40	1456	16	33963253				ENPP7P13	-178978		LINC00273	195331	-	CCCTTGCCTCTTGGCGC--HBV
355	13	2	17	1895	38	907	12	14259231				GRIN2B	-279143	-	ATF7IP	106400	+	AGGATAACAGTAATAACCGCnn--HBV
356	13	2	17	1874	37	7967	19	44757015	BCL3	at intron-exon junction	+							HBV--AGACACCGCTCCACCTGG
357	13	2	17	1866			7	144526314	TPK1	intron	-							HBV--TGAAAAAGATTTTGAGATAG
358	13	2	17	1875			5	89178690				MEF2C-AS1	-146314	+	LINC01339	979648		HBV--GGGCAGAGAAATGTGATATG
359	13	2	17	1876			15	36350632				MIR4510	-423709	+	C15orf41	228970	+	within 95 nts to left of nt 36350632
360	13	2	17	1907			18	64044086				SERPINB8	-54712	+	LINC00305	35922	-	HBV--GGTTTGTTCTTTCACTGT
361	13	2	17	1914				0										
362	13	2	17	1790	28	187	20	2333235	TGM3	intron	+							TTCTGTCTCCAGAGAAACTGnnnnnnnnnnnnnn--HBV
363	13	2	17	1777	25	152		0										
364	13	2	17	1786	27	182	21	8218570	LOC100507412	intron	+							GGATTATGACTGAACGCCTC--HBV
365	13	2	17	1792	26	327	16	49022236				N4BP1	-412027	-	CBLN1	255681	-	GCGGGATTTGGGGTCCTTC--HBV
366	13	2	17	1780				0										
367	13	2	17	1787			9	13895619				LINC01235	-464290	-	LINC00583	32351	+	HBV--nnnAGCAAGTCACAAAAATTC
368	13	2	17	1795			9	5896764	MLANA	intron	+							CTTTGTGGTAACTGTAAT--HBV
369	13	2	17	1798			20	60862076				MIR548AG2	-297451	+	LOC100506470	216987	+	HBV--nnGATGCTGATTCAGTT
370	13	2	17	1800			22	46675898	GRAMD4	intron	+							GCCCCTGCCTCGGATGC--HBV
371	13	2	17	1806				0										
372	13	2	17	1814			3	56852545	ARHGEF3	intron	-							HBV--TGTTTTTTCAAATTCTCTA
373	14	2	19	1120	11	616	5	166605125				>1000000			CTB-7E3.1	300096	-	HBV--ATAGCCTTATCCTGTGCTT
374	14	2	19	1122	10	1149		0										
375	14	2	19	1090	8	5163	18	6873174	ARHGAP28	intron	+							within 144 nts to right of 6873174
376	14	2	19	1114	9	4450	20	45127011				WFDC12	-2546	-	PI3	47887	+	AATATCTCTTTGGGATCCTG--HBV
377	14	2	19	1106				0										
378	14	2	19	1107			10	94808237	CYP2C19	intron	+							HBV--GGCCAGTATAATGT
379	14	2	19	1126				0										
380	14	2	19	1408	15	1674	6	18042875				KIF13A	-55252	-	NHLRC1	77611	-	HBV--GAAGGACCCCCTCCGCTGC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
381	14	2	19	1407			8	30152539				MBOAT4	-7855	-	DCTN6	3757	+	HBV--nnnGCACGTGCCACCATGCCTCGC
382	14	2	19	1416			2	210601408	CPS1	intron	+							within 102 nts to right of nt 210601408
383	14	2	19	1421			5	165964705				>1000000			CTB-7E3.1	940516	-	within 40 nts to right of nt 165964705
384	14	2	19	1426			12	98467233				SLC9A7P1	-10088	-	LOC643770	18310	-	TAAAATTTCTCTTGCAG--HBV
385	14	2	19	1149	13	1078	2	191820569				NABP1	-132047	+	SDPR	13736	-	TAACATTTTACAGAGCCC--HBV
386	14	2	19	1152	13	1078	3	117825136				LINC00901	-892898		>1000000			HBV--TGTTAGGTCATCATTTTT
387	14	2	19	1133	12	1272		0										
388	14	2	19	1129			4	82746892	SCD5	intron	-							within 150 nts to right of nt 82746892
389	14	2	19	1131			11	115683192				CADM1	-178669	-	LINC00900	72140	-	within 175 nts to left of nt 115683192
390	14	2	19	1141				0										
391	14	2	19	1156			2	162841092				KCNH7	-2345	-	FIGN	768459	-	within 150 nts to right of nt 162841092
392	15	2	23	4037	257	151	9	133881250	VAV2	intron	-							CACAGAGAGATGCGGCCn--HBV
393	15	2	23	4006	255	238	15	48030022				SEMA6D	-255799	+	SLC24A5	90949	+	TTTCAACTCTTCCTGCCTGCnnnnnnn--HBV
394	15	2	23	4026	257	282		0										
395	15	2	23	4005	256	1149	7	55452043				LANCL2	-18301	+	VOPP1	18564	-	HBV--TACCAATAGTAGGAC
396	15	2	23	3986	254	587	4	48345308	SLAIN2	intron	+							HBV--CAGAGTAAATAATCCCAGTT
397	15	2	23	3997			11	46715331				ZNF408	-9370	+	F2	3861	+	HBV--CCCAGCTATTTGGGAG
398	15	2	23	4007			3	156586391				SSR3	-31191	-	TIPARP-AS1	86779	-	HBV--GTTTAGGCTTGGCCTCATCACA
399	15	2	23	4021			1	56978586				C8B	-12446	-	DAB1	19319	-	within 40 nts to left of nt 56978586
400	15	2	23	4029			1	47331883				STIL	-17736	-	CMPK1	1913	+	within 450 nts to left of nt 47331883
401	15	2	23	4030			14	103885822				LINC00637	-27773	+	C14orf2	26465	-	HBV--CCTAGCCCCTACCCTTCTC
402	15	2	23	4032			4	180692756				>1000000			LINC00290	371333	-	within 1200 nts to left of nt 180692756
403	15	2	23	4035			10	36055897				PCAT5	-254977	+	>1000000			HBV--GTCCATTCTTCTCATC
404	15	2	23	4038			1	238589902				LOC100130331	-661583	+	>1000000			within ~750 nts to right of nt 238589902
405	15	2	23	4065	259	282	3	106417444				CBLB	-548401	-	LINC00882	692345	-	HBV--TGTAAACAAACATAAAGACGCA
406	15	2	23	4094	262	277	7	46572890				IGFBP3	-651618	-	TNS3	702263	-	HBV--ACCCTGTGAATTATGGTGC
407	15	2	23	4069	264	433	18	35238602				MAPRE2	-95134	+	ZNF397	2427	+	within 162 nts to left of nt 35238602
408	15	2	23	4070	258	1949	5	143874163				HMHB1	-53444	+	YIPF5	288054	+	HBV--ATAAATCTTGCTGTAAGGAAAC
409	15	2	23	4102	260	635		0										
410	15	2	23	4087	261	1149	12	115153653				TBX3	-469489	-	MED13L	804922	-	within 150 nts to left of nt 115153653
411	15	2	23	4096	263	5500	3	41060244				ZNF621	-520449	+	CTNNB1	139208	-	TTGAAGATCAGCTGACTGTAG--HBV
412	15	2	23	4080			14	50308450	L2HGDH	intron	-							HBV--ATACACAGATGGCAAAGAAGCAC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
413	15	2	23	4082			7	74481611	GTF2IRD1	intron	+							within 40 nts to left of nt 74481611
414	15	2	23	4086				0										
415	15	2	23	4098			2	217439396	DIRC3	intron	-							ACAACAACGCAGTGAGATAAA--HBV
416	15	2	23	4101			19	13956220	DCAF15	exon	+							within 40 nts to left of nt 13956220
417	15	2	23	4158	266	2000	4	71601151				SLC4A4	-29064	+	GC	140542	-	HBV-TTTAAATTAAAACAGTTTTAGGCA
418	15	2	23	4183	267	2743	3	187060777	ST6GAL1	intron	+							TTGGCAGCTAGAAGTTGAAATT--HBV
419	15	2	23	4152	265	5999	10	95731408	ENTPD1	intron	+							TGTAGCCGACCCTCTG--HBV
420	15	2	23	4162			2	118067070				CCDC93	-52907	-	INSIG2	21403	+	within 75 nts to left of nt 118067070
421	15	2	23	4182			X	122468185				>1000000			GRIA3	716057	+	TGTTTTTCCGATAGATnnnn--HBV
422	15	2	23	4184			4	99313535	ADH1B	intron	-							HBV--ATTTTTTAACTAAAAATTAATAA
423	15	2	23	4185			8	62536589	NKAIN3	intron	+							CTGTGGGGACACATCTAGn--HBV
424	15	2	23	4186			1	235916277				LYST	-32569	-	NID1	59554	-	TCAACCTCTGGAATTTTGA--HBV
425	15	2	23	4189			1	77621722	ZZZ3	intron	-							HBV--CAGGAGGATTTCCTGTGCC
426	16	2	25	1327	22	74	8	81218770				PAG1	-106702	-	FABP5	61712	+	ACACCTGGCTAACTTTTTAT--HBV
427	16	2	25	1294	18	92	14	67211985	FAM71D	intron	+							HBV--GCATTCTTCTGCCAGGAGA
428	16	2	25	1293	19	104	10	91904675				TNKS2	-39200	+	FGFBP3	1912	-	HBV--CAGTGCAAAATATTTTCTG
429	16	2	25	1307	21	187	16	67667074	C16orf86	intron	+							CGCTCAGACCTCCGAGGTn--HBV
430	16	2	25	1336	20	297	2	123807565				>1000000			CNTNAP5	217721	+	GACACACAGAGCTATGTGG--HBV
431	16	2	25	1295	17	322	8	27291710	TRIM35	intron	-							HBV-nnnnnATAAACAATAGTTTATTT
432	16	2	25	1297			22	2083373				>1000000			>1000000			CAAGAGCAAGACTCTACCTCAAnnnnnnnnn--HBV
433	16	2	25	1299			19	33855613				KCTD15	-39852	+	LSM14A	316833	+	within 68 nts to left of nt 33855613
434	16	2	25	1319			17	51215233	MBTD1	intron	-							TAATCATGTG GTAAGGTTAGn--HBV
435	16	2	25	1322			13	24214004	SPATA13	intron	+							HBV--ATATGTTTGGTTTCAGA
436	16	2	25	1342			4	120612857				MAD2L1	-546123	-	PRDM5	79055	-	within 40 nts to right of nt 120612857
437	16	2	25	1447	16	87	4	19740344				>1000000			SLIT2	513219	+	AACAGTATCTCCATTTT--HBV
438	16	2	25	1428			6	105969056				PREP	-565932	-	PRDM1	117263	+	TTGTTAATTGCACCACTTC--HBV
439	16	2	25	1430			15	88988718				MFGE8	-75286	-	ABHD2	99431	+	AGGGGAATTGCAATAGAGAAA--HBV
440	16	2	25	1431			7	84719888				LOC101927378	-145045	+	SEMA3D	275667	-	within 19nts to left of nt 84719888
441	16	2	25	1435			13	99366339	UBAC2	intron	+							HBV--GTTTTGAATTCTGAGTTTTTCCCCA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
442	16	2	25	1441			11	59932181				TCN1	-65613	-	OOSP1	10697	+	within 40 nts to right of nt 59932181
443	16	2	25	1449			3	114575127	ZBTB20	intron	-							CAGAACTAGGACTAGAAACT--HBV
444	16	2	25	1452			15	68279727	FEM1B	intron	+							HBV--ATTCTTTTTTAGGTTTAAACA
445	16	2	25	1684	23	51	X	83326924				>1000000			POU3F4	181336	+	GCATGGGAGAACTGAGTTAGC--HBV
446	16	2	25	1670	24	74	11	8636670	TRIM66	intron	-							TCTGATATCTCTTGAAATCTTTn--HBV
447	16	2	25	1679	24	32	13	111000244				ANKRD10	-85137	-	ARHGEF7	115282	+	AGCTGNGAAGTCATGCAGCnnnnnnnnnnnnnnnnnn--HBV
448	16	2	25	1662			4	166433955				TLL1	-329498	+	SPOCK3	299429	-	AATTAAACTACATGCATTT--HBV
449	16	2	25	1664				0										
450	16	2	25	1671			17	76492339	RHBDF2	intron	-							TGGAGACAGCATAACCCAG--HBV
451	16	2	25	1672			17	51221932	MBTD1	intron	-							TTACAAATCTATGAGAAATCAnnnnn--HBV
452	16	2	25	1675				0										
453	16	2	25	1689			9	75945996	PCSK5	intron	+							TGTTGATTTGGTGACTGTT--HBV
454	16	2	25	1691			1	33546240	CSMD2	intron	-							HBV--AATCCCATTATTTTTCAGGGA
455	16	2	25	1696			6	85461810	NT5E	intron	+							HBV--AGGCCTGGCACCCCTCTCTCT
456	16	2	25	1700				0										
457	17	2	25	4133	251	23	3	30662750	TGFBR2	intron	+							CTAGAAAATTATCATGGGC--HBV
458	17	2	25	4106	249	39	X	121427083				MIR3672	-56030		>1000000			GTATTATTTAGAACCATTATn--HBV
459	17	2	25	4119	250	18	1	15673431				RSC1A1	-11709	+	PLEKHM2	10900	+	TGTTGGCTGTTGTTTCTGC--HBV
460	17	2	25	4123	248	3105	10	83077447				NRG3	-90268	+	>1000000			TTAGGAATTCTAGCAGAACAnnnnn-HBV
461	17	2	25	4134			17	53822748				C17orf112	-835096	+	KIF2B	129	+	TTTCTCGCCATGATCCGGA--HBV
462	17	2	25	4136			3	107279636	LINC00883	intron	+							CAAAACAATTTTGAAACAGC--HBV
463	17	2	25	4195	252	286	8	29637026				DUSP4	-286748	-	LINC00589	84233	-	ATGGCTAACATGGTGAAACnn--HBV
464	17	2	25	4194			1	22316535				MIR4418	-50235	+	ZBTB40	135315	+	CCCTTTAGGTCCTCACAn--HBV
465	17	2	25	4200			3	48634649	SLC26A6	intron	-							HBV--nnAGCTTTCACCAGTCAGGAA
466	17	2	25	4149			13	43863539	CCDC122	intron	-							HBV--AAGAAAGTATAAACTGTTT
467	17	2	25	4150			8	34897319				LINC01288	-32521	+	UNC5D	338137	+	CTAGAATTTGGAAACTATCT--HBV
468	17	2	25	4151			2	65831213				SPRED2	-398691	-	MIR4778	527035	-	within 40 nts to right of nt 65831213
469	18	2	28	2154			4	99208221	LOC100507053	intron	-							HBV--nAGATTTTTAAGTAACTTCC
470	18	2	28	2157			3	228861	CHL1	intron	+							TAATGGTGGA TTTGACCnnnnnnnnn--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
471	18	2	28	2237	102	561	6	144643887	UTRN	intron	+							HBV--nnnnnnnnTTTTTTAAAATTGCTTTA
472	18	2	28	2232	101	1628	1	152152524				TCHH	-37070	-	RPTN	1070	-	HBV--TCAATAGTCCCNGGG
473	18	2	28	2235	103	1976	7	132456666	PLXNA4	intron	-							HBV--CCCAGGATGGGCCAGG
474	18	2	28	2233				0										
475	18	2	28	2322	100	780	14	30602427	G2E3	intron	+							TAATTATTTTTTTGTGTACT--HBV
476	19	2	29	2901	152	187	12	5179216				KCNA5	-133227	+	LOC101929584	54779	-	CCATCCTAGACTGTCCTGn--HBV
477	19	2	29	2916	151	295	5	180023238	RNF130	intron	-							HBV--nTATGTAGATATGCAGGAGTTA
478	19	2	29	2902	149	652	17	13949850				HS3ST3A1	-348721	-	CDRT15P1	74647	+	HBV--nnnCATTTGAACTCCCTTGTG
479	19	2	29	2909	150	923	4	141492489				LOC100507639	-159872		IL15	144106	+	HBV--AGTCATTCCTGTTCCA
480	19	2	29	2904			1	6882297	CAMTA1	intron	+							AACAGGGAGCATATGGCCTGT--HBV
481	19	2	29	2906			15	45445807				C15orf48	-12358	+	SLC30A4	36673	-	within 40 nts to left of nt 45445807
482	19	2	29	2910			13	100363997	PCCA	intron	+							AGCATATTCATCCTGGCTGGCnn--HBV
483	19	2	29	2911			11	134107675	JAM3	intron	+							within 168 nts to right of nt 134107675
484	19	2	29	2919			3	71405086	FOXP1	intron	-							CTTCCAAATTTCCACACA--HBV
485	20	3	23	1242	33	77	19	5695130	LONP1	intron	-							HBV--nnnnnATGGTATCTTAGATGCCGCT
486	20	3	23	1259	14	5895	14	25412634				STXBP6	-362337	-	>1000000			HBV--AGTACGGCCGACTCCAGTAGGGA
487	20	3	23	1252			2	209661342	MAP2	intron	+							HBV--nnTCCACACCAACCCCCCAAG
488	20	3	23	1261			4	188076224				ZFP42	-71175	+	TRIML2	15047	-	HBV--CAACTCTCCTGCCTCAGCCT
489	20	3	23	1274			7	2634434	TTYH3	intron	+							CTGCCTCCCCTGGTGAGGAGTTGC--HBV
490	20	3	23	1282			20	7007038				BMP2	-226775	+	LINC01428	139428	-	within 130 nts to right of nt to right of 7007038
491	20	3	23	1652	47	561	4	104040223				TACR3	-320407	-	CXXC4	428082	-	CTCAGCCCCAAATCTCCTTA--HBV
492	20	3	23	1635	49	1008	8	142522537	ADGRB1	intron	+							TCTGTTGGGGGCTTCAG--HBV
493	20	3	23	1657	48	1171	3	71755312				GPR27	-135	+	PROK2	16342	-	HBV--nGATTGTACTGACTCCTTTGGG
494	20	3	23	1643	43	878	14	74508971	LTBP2	intron	-							HBV--GAGGAAGACAGCCGATGGC
495	20	3	23	1637	45	2682	3	160892148	PPM1L	intron	+							CGGAACTTAAAATAAAATAA A--HBV
496	20	3	23	1610	42	5162	1	24169532	IFNLR1	exon	-							HBV--nnnnCATAGAACATAGCAGCTCCTT
497	20	3	23	1626				0										
498	20	3	23	1639			2	43087375				LOC102723854	-47832	-	ZFP36L2	136943	-	HBV--nCCCAGAAGATTCCCTGGATCC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
499	20	3	23	4346	275	103	13	101318155	NALCN	intron	-							HBV--TTTCTCACCCGCTGGGCTC
500	20	3	23	4380	274	131	11	38776300				LINC01493	-121051		>1000000			TCAAACTATCCCTGTTTGCAG--HBV
501	20	3	23	4372	273	938	4	100518926				EMCN	-633	-	LINC01216	-141352	-	CATGTCTCATTTTAGCTGATC--HBV
502	20	3	23	4352	272	3960	19	16582352	MED26	intron	-							CCAGGCCCTTTGCAGGGATT--HBV
503	20	3	23	4344				0										
504	20	3	23	4357			6	53339039	ELOVL5	intron	-							CACATTTCTTTTTTAAAA--HBV
505	20	3	23	4370				0										
506	21	3	25	1019	2	448	1	87599199				LINC00152	-177688	+	MIR4435-2	30560	+	TAGAAAAAAAACAGAGAA--HBV
507	21	3	25	1003	1	9340	4	150947664	LRBA	intron	-							CTACTTTTATTCAAATTAGTn--HBV
508	21	3	25	1023			19	35718988	KMT2B	intron	+							within 40 nts to left of nt 35718988
509	21	3	25	1027			5	160700946	ATP10B	intron	-							TGCCCAAGACTTTGATTnnnnn--HBV
510	21	3	25	1033			4	40213910	RHOH	intron	+							HBV--GTGGCACACACCTGTAATCC
511	21	3	25	1076	7	36	5	138825876	CTNNA1	intron	+							TGGTGCATGTTAAAATTTACTAGC-HBV
512	21	3	25	1048	5	157	14	28830705				LINC01551	-35911	+	PRKD1	746532	-	AAGGCAAGAG CAACTTAAn--HBV
513	21	3	25	1047	6	157	14	53355603				LOC101927620	-198075	+	MIR5580	592823		HBV--TTCCTCTTCACTCCCTCTGG
514	21	3	25	1069	7	157		0										
515	21	3	25	1059	4	468	6	124090679	NKAIN2	intron	+							HBV--TTCTTCACCTCAGTGAAAC
516	21	3	25	1049	3	468	19	1362668	MUM1	intron	+							GAGTGAGGGATGGCCCAGTGnnnnn--HBV
517	21	3	25	1056			2	231520846				LINC00471	-6507	-	NMUR1	2313	-	HBV--AAGGGTAGGCATATGC
518	21	3	25	1073			1	16174389				EPHA2	-18302	-	ARHGEF19	23714	-	HBV--TCCGTCTCCCCACTTCCTGGA
519	21	3	25	1077			6	112053921				FYN	-180469	-	WISP3	153	+	HBV--CTAGAGAGGACCCGGAG
520	21	3	25	1082			14	42981436				>1000000			>1000000			ATAAAAAGTAATTCAGTn--HBV
521	22	3	26	1351	51	129	7	19465891				FERD3L	-320470	-	TWISTNB	229570	-	CTAGGGTTAATTATAAGCAG--HBV
522	22	3	26	1460	53	131	22	46083145				PRR34-AS1	-24623	+	MIRLET7BHG	2851	+	CTTTTAAGACCCCACATG--HBV
523	22	3	26	1368	55	187	4	71522210	SLC4A4	intron	+							within 160 nts to right of nt 71522210
524	22	3	26	1366	54	55		0										
525	22	3	26	1458	52	229	2	228922189				SPHKAP	-740591	-	PID1	103442	-	AATCGTCATGCTGTCTTCCCAGA--HBV
526	22	3	26	1463	56	642	4	34525030				LOC101928622	-485137		>1000000			HBV--AGAAACCTGTTCAAATGTA
527	22	3	26	1353	50	1018	10	106279691				LOC101927549	-459358		SORCS1	293971	-	ATTCTTTAGAAGACATTT--HBV
528	22	3	26	1352			5	91362806				LUCAT1	-48404	-	ARRDC3	5917	-	GTCAGGTTCTGCTTCTAACACTA--HBV
529	22	3	26	1376				0										

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
530	22	3	26	1550	60	194	X	112390159	ZCCHC16	intron	+							HBV--ATTTAGGATCGTGGCCAGGTG
531	22	3	26	1571	61	642		0										
532	22	3	26	1566	58	2062	1	152568994				LCE3E	-2241	-	LCE3D	10389	-	HBV--TTTTTTTTCTTTAGCACTTTG
533	22	3	26	1536	57	4011	13	86522191				SLITRK6	-722843	-	MIR4500HG	921795	-	HBV--CCAGTCAGAACAGCTGGATATTA
534	22	3	26	1581	59	3580	13	84077248				SLITRK1	-195741	-	LINC00333	63353	+	HBV--CTTCCCAGAGGATAAAAGG
535	22	3	26	1541				0										
536	22	3	26	1572			9	133971570	VAV2	intron	-							HBV--GTGACACTGGCCTCCTGCCCAG
537	22	3	26	1590			4	88503047	HERC5	intron	+							HBV--GTTTTTCTGTTTTCTATTTTAT
538	22	3	26	1591			9	130611406	FUBP3	intron	+							TGCCTGATTTGGGGAGA--HBV
539	22	3	26	1596			6	47571362	CD2AP	intron	+							CAAAAAGTAGACTGCACAGG--HBV
540	22	3	26	1604			1	179363185				SOAT1	-4506	+	AXDND1	2534	+	HBV--nnnnnnnnAGGAAAATAAAGAAAAAAAG
541	23	3	26	1996	80	74	21	21342520	NCAM2	intron	+							AGCCTCTCCAGTATGGCTnnnnnnnnnnnn--HBV
542	23	3	26	1970	79	2120	9	16239116	C9orf92	intron	-							HBV--AACCTGGTGGGAAGTAATTGA
543	23	3	26	1983			10	58643783	BICC1	intron	+							within 10 nts to right of nt 58643783
544	23	3	26	1990			5	136506566				TRPC7	-141090	-	SPOCK1	468731	-	HBV--nnnGGCTAGGGACCTGGAGTCTG
545	23	3	26	2056	83	399	18	38531347				MIR4318	-874132	+	LINC00669	675576	-	CATACCCCTCTCACCTCCCA--HBV
546	23	3	26	2016	83	652	2	134257821	MGAT5	intron	+							HBV--nnnnnATTAGTTTGCAGTCTCTTTTT
547	23	3	26	2044	84	1008	20	50917312	ADNP	intron	-							HBV--CAAGTCAGTCAGATGATT
548	23	3	26	2023	81	1241	7	112447138	IFRD1	intron	+							HBV--TGAATCAAAGTCATCAAT
549	23	3	26	2012	82	36781	1	147668340	ACP6	intron	-							TCAACCATAAAGGCAGAAG--HBV
550	23	3	26	2032			9	100084919	ERP44	intron	-							HBV--AAAAAGCTAAGGCAACA
551	23	3	26	2033			7	148773668	CUL1	intron	+							CGCTGCGTGTGTCTTCTC--HBV
552	23	3	26	2037				0										
553	23	3	26	2041			21	29369825				BACH1	-7929	+	BACH1-IT2	2675	+	TTCTCAGGCCTTGGGTCC--HBV
554	23	3	26	2055			1	17703791				ARHGEF10L	-5916	+	ACTL8	51521	+	HBV--GCAGTGGTGCGAACCCGGC
555	23	3	26	2064	86	2000	17	54406378				KIF2B	-581166	+	TOM1L1	494312	+	AAAGATTAATAGTGGTGATT--HBV
556	23	3	26	2088			6	6255684	F13A1	intron	-							AGTGGCAAAGCATGATn--HBV
557	23	3	26	2093			17	9077838	NTN1	intron	+							HBV--GACCGGTGTTCAATGAAT
558	23	3	26	2096			12	91829905				DCN	-646876	-	C12orf79	155070	-	TGCCCACTTTGGCCTCCC--HBV
559	24	3	26	2413	104	18469		0										
560	24	3	26	2445			10	89427337				IFIT5	-6335	+	SLC16A12	2956	-	~TCCCTTGTGT GAACAGTTCn--HBV
561	24	3	26	2470	105	157	8	138771194	COL22A1	intron	-							HBV--nnnnCTCCCTGCACCTCACCAGG
562	24	3	26	2483	106	337	17	43882099	MPP2	intron	-							within 40 nts to right of nt 43882099

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
563	24	3	26	2485			6	97144652	MMS22L	intron	-							ATTTAGGGAAAGTGGCTATnnn--HBV
564	24	3	26	2490				0										
565	25	3	27	2530	145	25	3	76457907				ZNF717	-672324	-	ROBO2	582235	+	GGTTGGGGCGACTGGGACACAG--HBV
566	25	3	27	2836	148	57	9	37277108	ZCCHC7	intron	+							CCAAGTATAGACTCATT--HBV
567	25	3	27	2849	147	1733	X	134523203				HPRT1	-22535	+	MIR450B	16981	-	HBV-nnnnnnnnGTAGCTGGGATTACAGGCATGT
568	25	3	27	2847	146	4320	2	67602716				ETAA1	-192315	+	LOC101927701	193337	-	GATAGAGATATAAAATTCTG--HBV
569	25	3	27	2818			X	3084718	ARSF	intron	+							HBV--nAAAAGGACCTAGATA
570	26	3	29	1530	69	340	X	31805791	DMD	intron	-							HBV--AAATATTTTTTCCCCTGGA
571	26	3	29	1497	64	643	20	12456031				BTBD3	-529436	+	LOC101929486	409172	-	TAAAGTGTCTAATTCA--HBV
572	26	3	29	1503	66	1008	13	102146298	FGF14	intron	-							HBV--TGGTGGCTGACAAATTATTTT
573	26	3	29	1469	62	2670	5	151860329	GLRA1	intron	-							GAACACTTGTTTTCATCAT--HBV
574	26	3	29	1495	65	2062	1	31238043	NKAIN1	intron	-							HBV--CCTCTGCAATCCTCAGGGATA
575	26	3	29	1486			19	48962442				BAX	-644	+	FTL	2866	+	ACAGAGAGGAAGGGACCTnn--HBV
576	26	3	29	1516			6	81890783				LINC01526	-76591	-	IBTK	279455	-	HBV--ACCACACACGGTCAAACAA
577	26	3	29	1523				0										
578	26	3	29	1525			6	103795610				>1000000			HACE1	932482	-	HBV--AAAGAAAAAAAAACCCTATGG
579	26	3	29	1533			12	29009007				CCDC91	-458841	+	FAR2	139995	+	GAACCTGTGATTCCTATTT--HBV
580	26	3	29	1828	78	1008	6	18822120				MIR548A1	-250240	+	LOC101928519	246422		ATCACTAATAGGTAATCTATT--HBV
581	26	3	29	1823	76	1624	6	17025890				ATXN1	-264400	-	STMND1	76367	+	TGCCATGCTGGGTTTCA--HBV
582	26	3	29	1835	71	3032	X	44242614	EFHC2	intron	-							HBV--AAAATTCAAACTTGAAACT
583	26	3	29	1855	72	2536	1	19390206	CAPZB	intron	-							HBV--AAGAAATCACATTAAGTGT
584	26	3	29	1821	70	8389	17	55013573	STXBP4	intron	+							GTGCCTTTGGATAATTTTAGTCC--HBV
585	26	3	29	1833	154	10501	7	156783566	LMBR1	intron	-							HBV--nnGACTGTGTCTATCATTGGGTTGGT
586	26	3	29	1853	74	24258	2	209999103	UNC80	exon	+							TAGCCATTTTATAGGTTAAA--HBV
587	26	3	29	1861	73	12438	18	63467741				VPS4B	-45222	-	SERPINB5	9169	+	CTTTCTACTCGATGGGTGCCAA--HBV
588	26	3	29	1830			8	117071845	SLC30A8	intron	+							AGATTTACCATTTTAATAAACATG--HBV
589	26	3	29	1854			11	106701739	GUCY1A2	intron	-							HBV--ATTTATTGTGTATTTTCAGA
590	26	3	29	2964	156	5713	2	62405869				B3GNT2	-181138	+	TMEM17	94351	-	HBV--TATATGCAAATGATGATAATAAT
591	26	3	29	2968	155	12556	Y	8834727				TTTY11	-17345	-	RBMY1A3P	482333	-	GTCAGTTCCTCACCCCT--HBV
592	26	3	29	2939	153	56293	8	31564713				WRN	-391611	+	NRG1	75038	+	HBV--AGCTACCATTAGACCCAGC
593	26	3	29	2945				0										
594	26	3	29	2973			4	100034797	LOC256880	intron								TGTACAATAGTCAGGGATGT--HBV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
595	*Integration sites are listed by patient age within each patient group (IT, gp1; HBeAg(+) IA, gp2; HBeAg(-), gp3). A "0" in the insert site location means that we were unable to locate a unique integration site based on the cell sequence at the virus/cell junction. A string of n's at the virus/cell junction means that the indicated number of bases differed between the consensus human DNA sequence (GRCh38) and the observed sequence. For instance, HBV--nnGACTGTGTCTATCATTGGGTTGGT means that two bases at the virus/cell junction aligned neither with published HBV sequences nor the human sequence GRCh38. For integrations that did not map to either introns or exons, we determined the distance to the nearest genes to the left or right of the integration (within 1,000,000 base pairs).																	

Supplementary Table 4: Observed and Predicted Maximum Sizes of Hepatocyte Clones

Group 1: Immune Tolerant	Age	Largest Clone (95% confidence interval)	Predicted Size (k=0.0015) (95% confidence interval)	Predicted Size (k=0.004) (95% confidence interval)	Predicted Size (k=0.01) (95% confidence interval)
Pt. 1	15	2035 (899-3974)	392 (384-401)	786 (771-801)	1663 (1625-1702)
Pt. 2	17	1087 (423-1987)	403 (394-412)	824 (807-842)	1737 (1699-1775)
Pt. 3	18	3024 (1404-6369	408 (399-417)	839 (821-856)	1783 (1744-1822)
Pt. 4	18	611 (212-1458)	408 (399-417)	839 (821-856)	1783 (1744-1822)
Pt. 5	22	472 (170-1234)	430 (423-437)	907 (889-926)	1932 (1890-1973)
Pt. 6	24	1133 (465-2205)	441 (436-452)	931 (912-950)	2002 (1963-2041)
Pt. 7	28	3730 (1021-10306)	469 (461-477)	1007 (986-1027)	2136 (2095-2178)
Pt. 8	30	1416 (689-3182)	484 (476-494)	1043 (1022-1064)	2202 (2156-2247)
Pt. 9	39	1063 (479-2677)	552 (540-565)	1187 (1162-1212)	2555 (2495-2614)
Group 2: HBeAg(+) IA					
Pt. 10	14	1183 (240-3904)	388 (379-396)	770 (755-785)	1632 (1595-1670)
Pt. 11	14	94 (36-197)	388 (379-396)	770 (755-785)	1632 (1595-1670)
Pt. 12	16	5596 (1811-24563)	397 (388-406)	807 (791-823)	1699 (1661-1738)
Pt. 13	17	7967 (2915-18429)	403 (394-412)	824 (807-842)	1737 (1699-1775)
Pt. 14	19	5163 (2318-13361)	414 (406-423)	855 (838-871)	1842 (1803-1880)
Pt. 15	23	5999 (2711-11001)	438 (430-446)	916 (897-936)	1973 (1934-2013)
Pt. 16	25	322 (120-1391)	451 (444-459)	949 (930-968)	2050 (2009-2091)
Pt. 17	25	3105 (1384-6212)	451 (444-459)	949 (930-968)	2050 (2009-2091)
Pt. 18	28	1976 (901-6234)	469 (461-477)	1007 (986-1027)	2136 (2095-2178)
Pt. 19	29	923 (386-2031)	476 (467-484)	1021 (1000-1042)	2171 (2125-2217)
Group 3: HBeAg(-) IA					
Pt. 20	23	5895 (2448-10727)	438 (430-446)	916 (897-936)	1973 (1934-2013)
Pt. 21	25	9340 (4824-16022)	451 444-459)	949 (930-968)	2050 (2009-2091)
Pt. 22	26	4011 (1702-7735)	457 (449-465)	966 (947-985)	2076 (2040-2113)
Pt. 23	26	36781 (16898-54565)	457 (449-465)	966 (947-985)	2076 (2040-2113)
Pt. 24	26	18469 (8782-35639)	457 (449-465)	966 (947-985)	2076 (2040-2113)
Pt. 25	27	4320 (1953-9098)	464 (456-472)	989 (969-1008)	2099 (2056-2143)
Pt. 26	29	56293 (21517-112562)	476 (467-484)	1021 (1000-1042)	2171 (2125-2217)

Hepatocyte clones sizes (point estimates and 95% confidence interval) were determined from end point dilution data (e.g., Figure 1B) using the program Sim19. Predicted sizes were determined for different daily rates of hepatocyte destruction (k) using the program Csize8. The programs are described in Materials and Methods/Supplementary Methods & Materials. Shaded areas indicate that the maximum observed clone size in a patient could not be explained either by a low level of hepatocyte turnover thought to characterize healthy adults, 0.15% per day (k=0.0015) (3x the number of S phase hepatocytes)[31] or, where indicated, by higher turnover rates of 0.4% or 1% per day (k=0.004; k=0.01).