Dear Ze-Mao Gong,

In this letter we address the reviewer's comments about our manuscript, entitled "Analysis of HBV preS1 variability and prevalence of the rs2296651 polymorphism in a Spanish population" (Id. 36089), point-by-point. In keeping with these comments, we changed the title of the current version of the manuscript and revised the aims and the discussion according to an accurate comment by reviewer 02608938. We have attempted to further illustrate the rationale for studying sequence variation/conservation in the preS1 domains in the introduction. Lastly, we thoroughly reviewed the entire manuscript to resolve some language issues raised by the reviewers with rewriting of sections in some cases.

Regarding this last point, we noted that some of the reviewer's concerns resulted from confusion related to the writing. Hence, the entire paper has been read and revised to enhance the precision of the writing. Part of the former aim has been incorporated in the body of the introduction and materials and methods so that the rationale of the study is more clearly evident. The aim itself has been shortened to its main components for better comprehension. The short form of the term "the preS1 domain of the hepatitis B virus large envelope protein that interacts with NTCP" (ie, *NTCP-interacting domain*), which we coined for this study to enhance readability, has been explained in appropriate places of the paper and abstract. A paragraph has been added to the conclusions to address the reviewers' concerns about the relatively small number of patients and novelty of the results. Finally, we rewrote the core tip to make the main points of the paper more easily understood by an overall readership, not only those involved in HBV research. All changes have been tracked in Word. We can send a document with all the changes visible if you need it.

In accordance with the journal's academic rules and norms, we have added the biostatistics statement in materials and methods section (please see page 13 lines 16 to 19 in the revised manuscript), and described the statistical tests used (please see the footnote below Table 1, page 36 in the revised manuscript). We have also responded to concerns about the relatively small number of patients, sensitivity of the PCR amplifications, and the novelty of the results.

We hope that in its present form you will find our manuscript suitable for publishing in *World Journal of Gastroenterology* and look forward to your response.

Sincerely,

Francisco Rodríguez Frías, PhD

Reviewer 00052339

Study of factors that influence HBV-NTCP interactions in a Spanish population: HBV preS1 variability and prevalence of the rs2296651 polymorphism Casillas R et al. This manuscript showed the results of polymorphism of PreS1 region of HBV and human NTCP gene because of the possibility of interaction between PreS1 and NTCP gene product. The results shown here are not novel and definite. The following points should be addressed:

#1 The PCR method should be reconstructed, for example, other primer settings or nested PCR. The rate of successful PCR were very low such as 18 out of 41. Hence, the results of these 18 cases may have much bias.

We regret that the description in the methods led to confusion about the performance of the PCR technique used and the patients studied. The sequencing study was not done in all 41 CHB patients (Group A). It was only done in the 18 patients with a viral load ≥4 logIU/mL, the sensitivity of the technique. We have changed the writing in Methods to make this clearer: "Two main analyses were performed. First, the variability and conservation of the preS1 region of the LHBs was analyzed in the 18 patient samples from Group A with a viral load greater than >4 logIU/mL. This is the sensitivity limit of the PCR to amplify that region. Then we determined the sequence of preS1 by next-generation sequencing (NGS) based on ultra-deep pyrosequencing (UDPS) on the GS-Junior platform (454 Life sciences-Roche, Branford, USA)." (please see page 9, lines 15 to 20 in the revised manuscript) However, the rs2296651 SNP

(S267F) was determined in all 41 CHB patients (as well as the total of patients in the other 2 groups), as it did not depend on viral load and, being an analysis of prevalence, we wished to have a larger representation of patients for more robust results.

#2 What role is Group B? Did the author try to amplify PreS1 region using the serum or liver tissue of Group B? If Group B just gave the white blood cells or lymphocytes to investigate rs2296651 SNP (S267F), it must be combined to Group C as a control. I think Group B may be divided into two subgroups depending on the titer of anti-HBc Ab. If the anti-HBcAb is high, the patient may be HBV carrier with loss of HBsAg.

Group B patients tested negative for HBsAg but positive for anti-HBc antibodies, an indication that the infection had cleared. We included this group to specifically study the prevalence of the rs2296651 single nucleotide polymorphism (SNP) in a group with resolved HBV infection to determine whether HBV resolution could be related to the presence of that polymorphism.

The preS1 region was only analyzed in the 18 serum samples from CHB patients, as mentioned above, because we were not able to amplify this region in group B patients. We only had a serum sample available from those patients and since they resolved HBV infection, they did not show detectable viremia, thus amplification of the HBV genome could not be performed from those samples.

Finally, the SNP was analyzed in serum in all participants included, but none of them showed the SNP. Hence, it would not be necessary to divide group B patients into subgroups according to anti-HBc Ab titer for further comparison.

#3 The results of SNP of rs2296651 region were not novel.

In the study by Ho et al (Ho RH et al. J Biol Chem 2004;279:7213-22) the SNP rs2296651 was not detected in Americans of European, African or Hispanic origin. However, there is little data on its prevalence in our geographic area (typically classified as having an intermediate HBsAg prevalence, with a mainly Caucasian

population of European-Mediterranean origin). Since synthetic myristoylated lipopeptides share the same amino acid sequence with the preS1 domain that interacts with NTCP (NTCP-interacting domain), it is reasonable to think that the S267F NTCP variant could affect their attachment to this receptor, which might influence the effectiveness of the related therapies in some patients. Keeping this in mind, we believed it was worth studying the prevalence of rs2296651 in a representative sample of patients from our area, who are potentially eligible to receive NTCP blocking therapy. We have added some text to the conclusions about the local relevance of the findings (please see page 21, lines 18 to 27 in the revised manuscript).

Reviewer 02607378

Manuscript has now been significantly strengthened with improved clarity.

Reviewer 02608938

Casillas in the largely revised manuscript describe their studies of variation/conservation of the receptor interacting domain in the large envelope protein of HBV as well as the prevalence of rs2296651 in NTCP, the HBV receptor, in a Spanish population. Although authors well addressed most previous concerns, there are still some critical issues that should be taken care of before this manuscript can be published.

Specific comments:

#1 The critical argument for the aim of the current study, which is described in the last paragraph of introduction, is that "The degree of sequences variability in a domain is an indication of the extent to which sequence conservation is important in the domain."

- **First, this statement needs support of references with evidence.**

We based this statement on the generally accepted fact that when a domain is essential for maintaining the protein function, the sequence and structure of this domain should be conserved in order to ensure that this function will continue being properly performed. We have cited the book chapter by Roberts K et al. to support the statement (Roberts K, et al. Reporting of clinical genomics test results. In: Kulkarni S, Pfeifer J. Clinical Genomics. A guide to clinical next generation sequencing. USA: Elsevier Academic Press; 2015. page 219–31) which states that *Assessing the evolutionary conservation of amino acid at which a substitution occurs can also assist with variant classification. […] Variants that occur at non-conserved residues are generally considered less likely to be pathogenic, while sequence conservation may indicate that the residue is important for proper protein structure or function* (please see reference 19 on page 7 line 31 of the revised manuscript).

- **Second, the so-called domain in the entire manuscript means one on the protein while only genomic sequences were examined experimentally in the current study. Authors should discuss this and bear in mind that only sense mutation may result in amino acid sequence change.**

What we analyzed in the present study were amino acid sequences. After obtaining the nucleotide haplotypes, we classified them according to their HBV genotype by phylogenetic analysis in order to differentiate sequence variations from genotype-related polymorphisms, and then separated them into different fasta files on this basis. The nucleotide haplotypes in each of these files were then translated into amino acid sequences, and those with the same sequence were recollapsed into amino acid haplotypes and their frequencies were updated. We used these haplotypes to study both the NTCP-interacting and virion morphogenesis domains of the preS1 region.

Silent mutations were not taken into account in our analysis because they were not counted as mutations after translation of the nucleotide haplotypes into amino acid sequences, as they did not cause any amino acid change. The method for analyzing the preS1 domains is explained in more detail in Patients and Methods (Amino acid

variability/conservation in the preS1 region section, page 11 line 14 to page 13 line 19).

- **Third, change(s) of amino acid sequence even in the function/active protein domain may or may not impact the functional interaction. The current study did not perform relevant studies and thus authors should be causations when interpreting the results.**

It is true that since we did not perform *in vitro* functional tests, we cannot obtain strong conclusions from the amino acid conservation/variation events we observed in our study, and we have attempted to bear this in mind in the Discussion.

Our results showed a high degree of conservation between preS1 aa positions 2 to 21 (genotype D numeration), especially between positions 9 and 21. Assuming the premise that the degree of sequence variability in a domain is an indication of the extent to which sequence conservation is important in the domain (justified above), these observations were consistent with results of an *in vitro* study by Glebe et al. (Glebe D, et al. Gastroenterology 2005;129:234–45) (please see the Discussion page 17 line 24 to page 18 line 1 in the revised manuscript). Nonetheless, we are unable to know the specific manner in which all segments of the sequences in the preS1 domain that interacts with the NTCP spatially arrange to interact with this receptor; therefore, in the manuscript, we suggested that preS1 structural simulations could help to clarify the role of each segment in the NTCP interactions and the sequence conservation requirements for their respective functions (please see the Discussion page 18 lines 5 to 7 in the revised manuscript).

We found that the virion morphogenesis domain had a smaller percentage of genotype-associated viral polymorphisms than the NTCP-interacting domain in the genotype A to H consensus sequences obtained from the 86 HBV genome sequences downloaded from GenBank, whereas genotype-unrelated changes above 1% of the quasispecies were lower in the NTCP-interacting domain, with both domains being most highly conserved in genotype C and most highly variable in genotype E. In the

manuscript we only describe these relevant observations, but again, we suggest that structural simulations would likely be helpful to understand the reasons for the sequence conservation and variability in these two domains (please see Discussion, page 19, lines 1 to 3 in the revised manuscript). We have attempted to be cautious about the interpretation of these observations in the Discussion.

Similarly, for proline residues, which generally showed a high degree of conservation, and serine residues, which were particularly conserved in the C-terminal virion morphogenesis domain, or changed to threonine or tyrosine (all potentially phosphorylatable amino acids), we proposed a hypothesis to explain the high degree of conservation based on the physical-chemical properties of those amino acids. However, as what we obtained was just an observation, we acknowledged that site-directed mutagenesis experiments with modification of the proline and serine residues could clarify the function of these two conserved amino acids in both the preS1 essential domains analyzed (please see Discussion, page 19 lines 19 to 21 in the revised manuscript).

#2 Current study actually examined the sequence conservation on viral genome and SNP in subject for viral-receptor interaction, but no any "factor" was examined and no study was performed to test the impact of sequence change on the interaction. Therefore, the title is not appropriate with "factor" that influences … . Instead, it should be something such as "Analyses of HBV preS1 variability and prevalence of the rs2296651 polymorphism in a Spanish population".

We thank the reviewer for pointing out this imprecision. We have changed the title in the current version of the manuscript according to this suggestion (please see the title page). We also reviewed the text and eliminated the term "factors" that influence… to refer to the impact of the variability of the preS1 NTCP-interacting domain and prevalence of the rs2296651 SNP on the interaction between the HBV preS1 region or preS1-derived synthetic lipopeptides and NTCP, since we actually did not phenotypically test the impact of the sequence change on this interaction. To arrange

this we have rewritten the aims of the study (please see the introduction, page 8 lines 23 to 30), the discussion (page 17, lines 12 to 17), and Research frontiers in the Comments section (page 22 lines 19 to 27) of the revised manuscript.

#3 The rationale to study sequence variation/conservation in the HBV genome should be further illustrated. It should be clarified that whether one individual can be infected by multiple genotypes of HBV, whether infected viral genomes may have various mutations etc. and what is the impact of HBV subgenotype on this variation/conservation examined here.

We have included some background in the introduction to clarify the rationale of the study. HBV infection depends on the N-terminal preS1 region (here referred to as the NTCP-interacting domain) which would compete with the molecules of synthetic myristoylated lipopetide analogues to attach to NTCP in NTCP-blocking therapies. Therefore, the sequence variation/conservation in this domain could have an impact on the response to these therapies, e.g. by modifying the affinity of the preS1 NTCP-interacting domain to attach to this receptor, thereby changing the dynamics of competition with synthetic myristoylated lipopeptide-analogues. We have added some explanation about this in the introduction of the study (please see page 7 line 29 to page 8 line 6 in the revised manuscript).

Although Sanger sequencing of the preS1 region only showed a single genotype for each patient (please see Table 2 in page 37 of the revised manuscript), two patients showed more than one genotype by the much more sensitive next-generation sequencing, as has been added to the legend of Supplementary Table 2 (please see page 6, lines 13 and 14 in the supplementary material). Mutations and variations of each amino acid position of haplotypes in relation to the wild-type amino acid of the same HBV genotype in the haplotypes are shown in Supplementary Table 2.

Unfortunately, we cannot determine the subgenotype of the haplotypes. Although the results obtained from partial sequencing of the HBV genome are useful for determining HBV genotype, they are not appropriate for reliably determining HBV

subgenotype, for which phylogenetic analysis of the nucleotide sequences of the whole HBV genome is required (Pourkarim MR, et al. World J Gastroenterol 2014;20: 7152-68).

#4 Please clarify how PCR fragment sizes were determined based on the genomic position on page 9 since the sizes following the nt position are very confusion. In most cases, for example, 2844-56 means 2844 to 2856. Here, it appears that authors want to describe a region starting at position 2844, passing the end of the genome and reaching position 56. However, the genome size or the last nt should be provided considering varied sizes of different HBV genotypes.

As the reviewer inferred, we intended to describe a region starting at nucleotide position 2844 in genotype A, 2838 in genotypes B, C, D, E and H and 2837 in genotype F, passing the end of the genome and reaching nucleotide position 56 (the same for all genotypes). The differences in the length of this region between different genotypes are due to genotype-specific insertions and deletions, especially in genotype D. In order to clarify the numbering and the length of the PCR fragments obtained in each genotype we changed the dash between the numbers of positions by to (e.g. 2844-56 has now been written as 2844 to 56), and as requested by the reviewer we have added the size of the HBV genome in the different genotypes analyzed (please see page 10 lines 15 to 23 of the revised manuscript).

#5 English writing has been largely improved, but still some faults remain. Right terms should be used consistently through the manuscript. Authors should go through the manuscript carefully again and clearly tell readers what they want to describe using right English. Some, but not all, examples are listed below.

1) It sounds better to change the AIM of Abstract to "Determine variability/conservation of the receptor-interacting domain in the large envelope proteins (LHBs) of hepatitis B virus (HBV) and prevalence of rs2296651 (S267F)

polymorphism in sodium-taurocholate cotransporter polypeptide (NTCP), the HBV receptor, in a Spanish population."

We agree that the aim in the previous version of the manuscript was a bit curt. This was because the *World Journal of Gastroenterology* guidelines specified that the aim had to be no longer than 20 words (actually the aim in the previous version was already a bit longer than that). The problem is that we are using long terms that have to be defined from the beginning and this task alone already requires almost 20 words! In the current version, we have expanded the aim to fit the reviewer's suggestion and other requirements to define abbreviations. We hope that the aim in the revised version is now clear and the length acceptable in the opinion of the journal editors. Please see the re-written aim on page 4, lines 2 to 10 of the revised manuscript.

2) In Methods of Abstract, NTCP should not be explained again in which, "included and" can be deleted. Or HBV receptor-interacting domain should be used to maintain the consistency of wording. For your information, searching "NTCP-interacting domain" in Pubmed database or google did not yield any result showing this grouped term. In addition, viron vs. host should be clarified, such as "Variability …. Of HBV receptor-interacting domain (aa2-48 in viral genotype D)...". "…, both in the HBV preS1 region," should be deleted. "by next-generation sequencing" should be removed to after the "compared" and add "of viral genome in 18 CHB…".

We greatly appreciate the reviewer's comments to help with the terms and have tried to improve the writing in the individual points raised. In relation to the grouped term *NTCP-interacting domain*, it is true that it is not a standard term to indicate "the pre-S1 domain of the hepatitis B virus large envelope protein that interacts with NTCP", but we had to repeat this term so often that we sought a shorter designation for it. We have now attempted to clearly define the term the first time it appears in the abstract (please see page 4, line 8), the main text (page 7, lines 27 and 28) and the comments (page 22, lines 24 and 25) in the revised manuscript. We have also attempted to use this term consistently throughout the manuscript (marked in red).

We have also clarified when we refer to the virus and when to the host, especially with the term genotype, by adding the terms viral or HBV before genotype where we judged that it would facilitate comprehension (marked in red throughout the manuscript). In the abstract, we also deleted "…, both in the HBV preS1 region," as suggested by the reviewer (please see page 4 lines 16 to 25 in the revised manuscript).

Finally, we could not adhere to the last point raised by the reviewer, which we believe was raised because of our imprecise writing leading to confusion. First, in our view it is important to mention in methods that we used next-generation sequencing to analyze variability and conservation in this region, as this provides this study with added value: this analysis could have been performed by much less sensitive techniques, such as molecular cloning, but while next-generation sequencing enabled us to reliably detect variants present in proportions until 0.25% of the circulating HBV quasispecies (please see Materials and Methods in the main text, page 11 lines 9 to 12 in the revised manuscript), with molecular cloning we would have sequence thousands of clones to be able to achieve the same level of sensitivity, which is actually not cost-effective and extremely time-consuming. For these reasons, we believe it is essential to mention the next-generation sequencing technique. Second, we did not add "of viral genome" before "in 18 CHB patients". Although we did obtain nucleotide sequences from the HBV genome, we translated the sequences into amino acids. Hence, we were not analyzing variability and conservation of nucleotide sequences from the HBV genome but instead, the variability and conservation of amino acid sequences from the preS1 region of the HBV large envelope proteins. We regret the confusion and have mentioned this important point in the methods section of the abstract to facilitate proper comprehension of what we did in the study (please see page 4 lines 16 to 25 in the revised manuscript).

3) In Results of Abstract, NTCP should be replaced by receptor to be consistent with text above. "HBV preS1 NTCP interacting domain" is an incorrect term and no result was produced from searching this grouped term from Pubmed. The conservation is among different viral genomes, but not between aa 9-21,

meaning that "the conservation of aa 9-21 region/sequence is stable among examined viral genomes". It is not "referring to", but "in".

We believe this has been answered in the first paragraph of our response to point 2. The abbreviated term *NTCP-interacting domain*, was coined to stand for "the pre-S1 domain of the hepatitis B virus large envelope protein that interacts with NTCP", a cumbersome phrase that had to be repeated. We now realize that it was not properly introduced in the original version and have attempted to clearly introduce it in the abstract (please see page 4, line 8), the main text (page 7, lines 27 and 28) and comments (page 22, lines 24 and 25) in the revised version, and we have tried to use it consistently throughout the manuscript (marked in red).

As the reviewer pointed out, we acknowledge that conservation of amino acids 9 to 21, and in fact the entire NTCP-interacting domain, is actually among the viral genomes analyzed. In order to clarify this aspect, we rewrote the beginning of results in the abstract as "The HBV preS1 NTCP-interacting domain showed a high degree of conservation among the examined viral genomes" (please see the page 4 lines 29 and 30 in the revised manuscript). We would prefer not to add that "the conservation of aa 9-21 region/sequence is stable" because we cannot actually ensure whether the conservation is stable, because we did not do a longitudinal study of the same group of patients over a certain time period. Here we just wish to state that the degree of conservation of the entire NTCP-interacting domain among the viral genomes obtained from the samples analyzed was high. We also clarified this in the results of the main text (please see Results in the main text, page 14 line 24 in the revised manuscript).

4) In the introduction, the 2nd paragraph on page 6, "… , encoded by" should be " … , encoded by the SLC10A1 gene and located on chromosome 14, as …".

We have made the change as suggested by the reviewer, please see page 7 line 15 in the revised manuscript.

Reviewer 03664122

Interactions between hepatitis B virus (HBV) and sodium-taurocholate cotransporting polypeptide (NTCP) in hepatocytes could be influenced by the HBV preS1 region and/or the NTCP variability in HBV infected patients. Obviously, the analysis of such variations is important for the appropriate use of anti-HBV drugs which blocks HBV entry by coupling with NTCP or HBV preS1 proteins.

#1 The absence of rs2296651 polymorphism in the studied Spanish population is of local importance, as the particular mutation was reported predominantly at Asian population. Subsequently, the number of CHB patient samples with analyzed preS1 sequence is insufficient to drive rigorous conclusions about variability/conservation of that region.

We agree that the number of CHB patient samples in which we analyzed the variability of the amino acids in the preS1 domain that interacts with NTCP and the preS1 virion morphogenesis domain is a limitation of the study. The 18 samples included yielded haplotypes of most HBV genotypes (A to F and H). However, even with the diversity of the viral strains analyzed, they remain insufficient to make strong conclusions about the variability and conservation in the NTCP-interacting and virion morphogenesis domains. Furthermore, the genotypes represented in larger numbers of our patients (the most prevalent in our area) are overrepresented in relation to the less frequent ones. Our study should be considered an exploratory one whose findings would need to be confirmed in a larger patient population. We have mentioned this in the Discussion (please see page 20, lines 21 to 27 in the revised manuscript) and have attempted to be cautious about the interpretation and conclusions. We have added a paragraph to the conclusions (final paragraph) to further state that the findings of the study are limited and relevant locally (please see page 21, lines 18 to 27 in the revised manuscript).

#2 In general I do not see this research as basic study, but consider as excellent methodological approach.

We thank the reviewer for this kind comment on the methodologies used in this study. We considered this research as a basic study mainly because we believe that the conservation findings in the NTCP-interacting and virion morphogenesis domains, e.g. the high conservation of residues 9 to 21 in the NTCP-interacting region and proline and serine/potentially phosphorylatable residues, deepen our knowledge of the molecular biology of these two essential domains of the preS1 LHBs region. In addition, the findings provide a starting point for structural and phenotypic studies aimed at studying the mechanisms of HBV infectivity, and investigating markers of the response to NTCP blocking therapies. For example, variability in the NTCP-interacting domain might change the affinity of this domain to bind NTCP, modifying the dynamics of competition with synthetic myristoylated lipopeptide analogues to attach to this receptor, although this must be demonstrated in phenotypic *in vitro* studies.

#3 I highly recommend to improve the method of HBV DNA isolation from serum samples of patients with lower viral load by use of larger sample volumes and organic /inorganic extraction methods, instead of Qiagen kit with columns in this particular case.

We agree that a better approach would be to use larger sample volumes and nucleic acid extraction methods that are more efficient than the Qiagen kit with columns. Unfortunately, the most efficient viral DNA isolation method available in our laboratory at the time this study was performed was the Qiagen kit. Because of the limited amount of serum available (no more than 1 mL) from samples obtained for routine clinical analysis (mentioned in Materials and methods, page 9 lines 26 to 28 of the revised manuscript), we do not have sufficient sample volume to try repeating HBV-DNA isolation with volumes larger enough to obtain significant increases in the HBV-DNA isolated, and other protocols.

The authors performed analysis of both viral (preS1) and host (rs2296651 polymorphism -the NTCP variant, S267F associated with reduced ability to interact with preS1) factors in conjunction with disease status- chronic hepatitis B versus resolved hepatitis B, in Spanish population.

The authors developed very modern methodology and comprehensive data analysis allowing to perform such analysis with high accuracy. The manuscript provides adequate details of methodical approach and good description of data analysis. The developed an in-house FRET-based real-time PCR method is potentially applicable to different platforms for determining the prevalence of the rs2296651 SNP in any patient.

#4 The obvious weak point is insufficient sample number of preS1 DNA (only 18 out of 46 CHB patients) to drive any reliable conclusions about conservation/variability of preS1. Especially, within genotypes, where only 2-5 samples from each genotype were analysed. For example, in case of more rare C genotype, out of 3 patients studied, no any variation was found, as it was possible to understand from the provided data. Therefore it is questionable if all C genotype patients did not have a common source of HBV infection, which is the reason of minimal variation found.

As indicated in comment #1, we agree that the number of samples in which we analyzed the variability of the LHB preS1 NTCP-interaction and virion morphogenesis domains is low, and this is a limitation of the study which could be partially mitigated by the fact that the 18 samples included yielded sequences from most HBV genotypes (A to F and H), and approximate the results on the variability and conservation obtained in both domains to the real overall variability of the HBV in these domains. For this reason, we acknowledged in the discussion that the study should be considered a preliminary approach whose results should be confirmed in further studies with larger populations (please see page 20, lines 21 to 27 in the revised manuscript) and have added a new paragraph to the conclusions in this regard (please see page 21, lines 18 to 27 in the revised manuscript).

The samples included were selected from the population attending the outpatient clinic of our center, Vall Hebron University Hospital (Barcelona, Spain). There was no

relationship between the HBV genotype C patients included in the study, so we ruled out a common source of HBV infection. All of them are of similar age and it seems likely that they would have acquired HBV infection through vertical transmission (from infected mother to child).

#5 I strongly recommend for authors to improve the protocol for HBV DNA extraction from serum, using larger volume of sample (0,5-1 ml serum instead of 0,2 ml) and organic or inorganic extraction methods instead of Qiagen kit with columns, as it is described in Changotra H. and Sehajpal P.K. 2013. (Changotra H. and Sehajpal P.K An improved method for the isolation of hepatitis B virus DNA from human serum. 2013, Indian J. Virol, 24(2): 174-179).

We are extremely grateful to the reviewer for this comment and the reference. We are currently trying to improve our HBV-DNA extraction method in order to amplify samples with viral loads <4 logIU/mL, and the study suggested describes several protocols that could be useful for us. Unfortunately, as was indicated in comment #3, due to the limited amount of available serum sample (no more than 1 mL of serum from samples obtained for routine clinical analysis, as explained in Materials and methods, page 9 lines 26 to 28 of the revised manuscript), we do not have sufficient sample volume to try to obtain significant increases in the HBV-DNA by repeating HBV-DNA isolation.

Nevertheless, as was mentioned above, this is an exploratory study, which we hope to expand with future experiments with larger and more specific groups of chronic hepatitis B patients, for which the suggestions and bibliography recommended by the reviewer will be valuable. In addition, in these future studies we also aim to perform functional *in vitro* studies to confirm the variability/conservation data obtained by next-generation sequencing.

The presentation of results and discussion are relevant and correct. The results obtained are logic and in general agreement with already known data about preS1 conservation (Glebe et al, 2005; reference 24 in the manuscript) and NTCP

polymorphism (references in the manuscript 18, 19, 20). The mentioned weaknesses of the study as well as future research directions were pointed out and discussed by authors.

#6 Basically no essential new data provided. Additionally, the study of preS1 polymorphism in CHB patients only, without comparison with the viral sequences from hepatitis B resolved patients due to difficulties to obtain viral DNA from such patients, is also of limited value for assessment of impact of preS1 variability to disease manifestation and treatment.

We agree with the reviewer that due to the limited number of chronic hepatitis B (CHB) patients the results from our study are just a preliminary description of the variability and conservation of preS1 NTCP-interacting and virion morphogenesis domains in CHB patients. The comparison of those patients with hepatitis B resolved patients would be certainly interesting, however in that patient group serum HBV-DNA levels were undetectable (below 20 IU/mL); therefore, the only way to compare both chronic hepatitis B and hepatitis B resolved patients would be obtaining HBV-DNA from retrospective serum samples with sufficient HBV-DNA levels (which were not available), or to compare HBV-DNA isolated from biopsies obtained in both chronic hepatitis B and hepatitis B resolved patients, which are much more difficult to obtain than serum samples, technically difficult, and may raise ethical concerns.

Despite the limitations of the present study, we believe it lays the foundation for future studies investigating the essential HBV preS1 region of LHBs, by the development of a reliable high-throughput protocol based on next-generation sequencing. This protocol provides a detailed picture of the variability/conservation of the preS1 NTCP-interacting and virion morphogenesis domains, which enabled us to observe, for example, the high degree of conservation of proline and serine residues, particularly in the C-terminal virion morphogenesis domain. Based on physical-chemical properties of those amino acids, we hypothesized that proline residues could stabilize the structure of both preS1 domains, and the tendency to keep serine, threonine or tyrosine in specific positions of the virion morphogenesis domain suggests that they could be phosphorylated, a phenomenon already

demonstrated in the preS1 region in LHBs of duck hepatitis B virus (DHBV) model both *in vitro* and *in vivo* (Grgacic et al. Virology 1994; 68: 7344-7350, Grgacic et al. J Gen Virol 1998; 79: 2743–2751, Borel et al. Virology 1998; 242: 90–98, Rothmann et al. J Virology 1998; 72: 10138–10147), but not in human HBV (please see Discussion, page 19 lines 4 to 21 in the revised manuscript). Site-directed mutagenesis experiments with modification of the proline and serine residues could clarify the function of these two conserved amino acids in both essential preS1 domains analyzed, a goal we hope to investigate in the future.

Finally, we demonstrated a low to null prevalence of the rs2296651 SNP in our area (typically classified as having an intermediate HBsAg prevalence, with a mainly Caucasian population of European-Mediterranean origin, as mentioned in the previous question). We have only found one very recent study (Ezzikouri et al. BMC Infectious Diseases 2017;17:99), aimed at determining whether there is an association of the S267F NTCP variant with HBV infection status in Moroccan patients, who have an ethnical origin closer to ours than those included in previous studies, and in whom the SNP rs2296651 was also absent. It is reasonable to consider that the S267F NTCP variant could interfere in attachment of synthetic myristoylated lipopeptides (molecules that have the same amino acid sequence as the preS1 NTCP-interacting domain) to this receptor. Therefore these results are potentially useful locally as they indicate that treatments based on the inhibition of HBV entry in hepatocytes by NTCP block therapies would be useful in the CHB population from our geographical area.

Nevertheless, the methodological approach is unique and simultaneous analysis of interplay of both viral and host factors important for HBV entry is original and potentially interesting research which have to be conducted with larger sample numbers for HBV preS1 and more possible NTCP polymorphisms analysis.