RESPONSE LETTER

The authors would like to thank for the comments made to our manuscript *'Metabolic and hepatic effects of liraglutide, obeticholic acid and elafibranor in diet-induced obese mouse models of biopsy-confirmed nonalcoholic steatohepatitis'.* We have carefully addressed the criticisms point-by-point, our response is indicated below in italics. Changes in the manuscript are highlighted in yellow.

REVIEWER 1

Tølbøl KS and colleagues provided evidences about the therapeutic effects of three different drugs (liraglutide, OCA and elafibranor) in the context of biopsy-proven experimental non-alcoholic steatohepatitis (NASH). They evaluated the efficacy of the above-mentioned compounds in two different strains of high-fat fed mice namely wild type (C57BL/6J) and genetically obese (ob/ob) mice. They analyzed the metabolic effects of the drugs (body weight, lipid profile) as well as the liver damage both from a biochemical (ALT, AST) and histological point of view (steatosis, inflammation and fibrosis). Moreover, they corroborated their findings by analyzing the hepatic transcriptome. As expected, they found that all three drugs ameliorated somehow the different pathological aspects (metabolic and/or histological) involved in NASH evolution. The present manuscript is potentially interesting taking in account the urgent need to define an efficient therapeutic regimen for a successful treatment of human NASH. Presently, one of the most important goal within the liver research field is to find a therapy aiming to avoid the progression of NASH toward advanced phases such as cirrhosis and hepatocellular carcinoma (HCC).

1. Major Comments: In my opinion, to allow an easier reading of the paper, the authors should insert in the introduction section some hints concerning the molecular mechanisms/pathways targeted by the different drugs classes. Otherwise, it is quite difficult for the reader to clearly understand the experimental results provided in the following sections.

Response: We thank the reviewer for bringing this to our attention. To distinguish between targeted mechanisms, we have now slightly extended the Introduction to describe the principal mechanisms of action of the three compounds tested (p.4, line 17- p. 5, line 6).

The inflammation plays a crucial role in the pathogenesis of NASH and on the same vein a growing body of evidences suggested as immune responses may drive the worsening of the hepatic injury during the evolution of NAFLD. On the base of these premises, I believe that for an overall understanding of the data presented in the manuscript (for example transcriptome profile), the authors should give just a quick overview about the critical role of the immunity (see for instance World J Hepatol. 2015 Jul 8;7(13):1725-9 "Is there a role for adaptive immunity in nonalcoholic steatohepatitis?" and Immune Netw. 2016 Jun;16(3):147-58 "The Immune Landscape in Nonalcoholic Steatohepatitis."). Even if, it is not the target of the present manuscript, however this aspect is quite important considering the strict connection between FXR signaling and immune system function. In fact, as the authors mentioned, it is well established the anti-inflammatory action mediated by FXR agonist (see for instance Curr Pharmacol Rep. 2017 Apr;3(2):92-100). I

guess that the impact of the pharmacological treatment on the systemic inflammation should be evaluated. For instance, they could measure by ELISA the circulating levels of pro-inflammatory cyto/chemokines such as TNF-α and MCP-1. I guess it could be also relevant for clinical purposes.

Response: We agree that the Introduction will be improved by indicating current knowledge on the complex pathogenesis of NASH. We have now included a paragraph in the beginning of the introduction that briefly mentions most current hypotheses (with relevant references), including alterations in the innate and adaptive immune system (p. 3, line 3-7). The transcriptomics applied in the present study clearly indicates upregulation of hepatic pro-inflammatory cytokines and recruitment of various immune cells, making these disease mechanisms and signalling pathways relevant in the context of discussing the various mechanism of action of the three compounds tested. These effects are therefore considered in the Discussion and compared to histological evidence of inflammation in the two mouse models of NASH employed. However, it is beyond the scope of the present study to specifically address anti-inflammatory mechanism of action of the individual compounds, and we do therefore not perceive it relevant to also measure the corresponding protein levels of selected inflammatory markers.

2. Minor Comments: I would use the same nomenclature for the genes in the transcriptome analysis images: Figure 1I, the authors reported MCP-1 instead in Figure 5E they used CCl2 (please choose only one). Moreover, I guess in figure 1I Ccr should be replaced by Ccr2. Why they did not use the same order and the same genes to display? It would be much easier to compare the figures. In figure 1I there is showed Mac-2 while in figure 5E appears Lgals3. Please add both.

Response: Figure 1 and 5 have been modified to provide consistency in the gene nomenclature, thank you.

REVIEWER 2

1. Tølbøl et al. demonstrated the efficacy of liraglutide, obeticholic acid and elafibranor using C56 and ob/ob mice fed AMLN diet, showing histology of NASH. Basically, experimentations are well-done and data interpretation is adequate. However, results and discussion sections are too long and need improvement.

Response: Considering the number of parameters measured, the Results sections is rather small (3½ A4 page). We have carefully weighted each and every element in the Discussion with reference to the impact of the individual findings. It must be acknowledged that the present report represents a comparative comprehensive pharmacological study as it describes effects of three different compounds in two individual mouse models of NASH, which warrants a thorough discussion of the results. Furthermore, as it was also the aim of the study to provide information on the translatability of the results obtained with reference to corresponding clinical findings, this must be taken into account when considering the length of the Discussion. It should be noted that aspects of clinical translatability are very important and often not included in preclinical pharmacological studies in the context of NASH. In total, the Discussion should be viewed as fully balanced with respect to the impact and implications of the findings.

2. Discussion sections: The authors need to state the discussion more clearly. For example, I recommend to state the efficacy of drugs according to histological components of NASH, such as steatosis, inflammation, and fibrosis.

Response: These individual histopathological hallmarks of NASH are indeed considered and carefully discussed. As the study comprises head-to-head comparison of three important drugs in advanced clinical development for NASH, we discuss and compare the effects of each individual compound in the two models of mouse NASH. As also experienced in the clinic, the compounds have very different pharmacological effects, and thus also show different impact on NASH pathology. This aspect is also reflected in the Discussion.

3. The authors should add the data of diet intake amount and mention whether it was basically similar or matched between the groups.

Response: This information is now given in the Results section where relevant (p.11, line 13-15 & p.12, line 23-p. 13, line 1).

4. There are some discrepancies between biochemical/histological data and mRNA expression. For example, elafibranor attenuated fibrosis score but the expression of collagen 1a1 and galectin 3 was unchanged. The authors should comment this point.

Response: It is important to distinguish between the two very different methods for determining liver histopathology in DIO-NASH and ob/ob-NASH mice. Hepatopathological scores are based on qualitative clinical assessments of liver pathology. In contrast, Col1a1/galectin-3 levels are determined by quantitative immunohistology. This makes these two sets of parameters not well-suited for direct comparisons. Therefore, the liver pathology scores will not always correlate to quantitative changes in histopathology and vice versa. The disparity between qualitative and quantitative histological data is discussed where relevant, e.g. 'In contrast, total Col1a1 and Galectin-3 immunoreactivity was unaffected in elafibranor-treated DIO-NASH mice due to the liver hypertrophic effect' (p. 18, line 23-24). Essentially, this indicates that increased liver mass afforded by elafibranor treatment will result in increased total amounts of Col1a1 and Galectin-3 (as compared to NASH control mice).

5. Also, serum ALT was unchanged or increased but NAS improved (Figure 1,2 elafibranor; Figure 7 OCA). The authors should mention the reason of this discrepancy or the validity of ALT measurement.

Response: Plasma ALT and AST levels were measured using high-quality commercial assay kits with proper assay controls, and ALT/AST data are therefore fully valid. NAFLD and NASH is typically suspected based on raised ALT levels together with other clinical and biochemical features. The usual observed biochemical pattern in hepatic steatosis due to NAFLD is increased levels of transaminases, with ALT levels exceeding those of AST. However, the progression of NASH is often accompanied by an increase in AST levels with a resultant rise in the AST:ALT ratio, and a reversal of the AST/ALT ratio to >1 has been reported to better predict the presence of severe fibrosis, but

not NASH per se (Rinella, JAMA 313:2263-2273, 2015). As can be inferred from the AST and ALT data in Figure 2 and 6, both transaminases are equally upregulated in DIO-NASH and ob/ob-NASH mice. Hence, changes in plasma AST levels do not dominate in these two models, which could be due to their mild to moderate disease phenotype. This also signifies that drug-induced reductions in ALT and AST levels cannot predict treatment efficacy on liver histopathology in these two models, and should thus be considered as one marker of several markers affected by individual treatment. This interpretation is now briefly indicated in the Discussion (p. 14, line 14-17).

6. Figure 4C and 4E, Figure 7F and 7G: How did you measure collagen 1a1 and galectin contents (g)?

Response: Because all treatment paradigms affected total liver weight, quantitative data on liver histology (liver fat, Col1a1 and galectin3 content) was expressed as whole-liver amounts by multiplying individual terminal liver weight with the corresponding liver fractional area of the individual histology marker. This approach is indicated in the Materials & Methods section (p. 8, line 4-8).

REVIEWER 3

Authors analyzed the effects of liraglutide, OCA, and elafibranor in 2 different NASH models, namely DIO-NASH and ob/ob-NASH mice. it seems that the manuscript data is taken from a thesis. Therefore, there are too many unnecessary data. if authors want to emphasize the differential effect of drugs on 2 different model result and discussion must be organized accordingly. if authors want to compare the 3 drugs in different animal models, then they must give relevant data and make conclusions about the drugs. The results in conclusions should be clearly stated by omitting unrelevant data. i.e data reveals that elafibranor improves weight, biochemical marker, inflammation and fibrosis in both model, while other 2 drugs has partial effects on these parameter.

Response: We strongly dissociate ourselves from the reviewer's comments, as the present manuscript represents high-quality original research by experts within preclinical NASH science. The reviewer should rather acknowledge the thorough description of the individual disease phenotype in the two mouse models of NASH disease, and appreciate that this makes these two models particularly applicable for preclinically testing compounds for potential anti-NASH efficacy. Moreover, the comparative pharmacological study applied to these two models of NASH provides unprecedented insight into the different pharmacology of three important compounds in current advanced clinical development for NASH. There is increasing demand for clinically translatable animal models of NASH that (1) more closely recapitulates the clinical disease phenotype, and (2) exhibit treatment effects of relevant anti-NASH compounds similar to that observed in the clinic. Our data suggest that the two mouse models of NASH, in combination with the application of baseline biopsy-confirmed liver pathology, fully meets these criteria.

REVIEWER 4

Authors indicated time course of changes in metabolic parameters and histopathological findings in liver in detail, and that effect of three drugs on those phenotypes in two types NASH model mice. The findings will contribute to understanding of pathophysiological alterations in NASH model mice, and the experimental design will be helpful to improve the screening of candidate drugs for NASH in preclinical trials. Thus, it will be suitable for the publication in this journal. However, author should answer the following two points.

1. Can the NASH score adapted to human disease be used to NASH model mice? Author used Kleiner's NASH score, which is produced for diagnosis of human NASH, in this study. However, histopathological findings in NASH model mice is inconsistent with that in human, especially low-grade fibrosis. If you modified the scoring method, it should be described in method section.

Response: There is an increasing consensus that the human NAFLD scoring system (NAS) is instrumental in the evaluation of qualitative changes in liver histopathology in rodent models of NASH (e.g. Liang et al., PLoS One 9: e115922, 2015; Hansen et al., Drug Discovery Today 22: 1707-1718, 2017). In the present study, NAS and fibrosis staging was applied according to the clinical criteria outlined by Kleiner et al. (indicated in the Materials and Methods section, p.7, line 3-4) - without modifications.

2. Steatosis in vehicle group of ob/ob-NASH mice was more severe than that in other groups. Did individual mice randomly divide to four groups?

Response: As indicated in the Materials and Methods section (page 7, line 10-12), DIO-NASH and ob/ob-NASH animals, respectively, were randomized and stratified to treatment based on baseline mean fibrosis and steatosis score. Although individual scores will always vary within each treatment groups, the stratifications procedure assures that all treatment groups had similar baseline mean fibrosis and steatosis score prior to treatment start. As for ob/ob-NASH (and DIO-NASH) mice, it can be observed that baseline mean steatosis was similar (grade 3) across all experimental groups (Suppl. Figure 2 & 3). This procedure is only possible due to the unique application of biopsy-confirmed liver histopathology in the present study.

3. Fig.6 was not looked. The display in figure legends get corrupted.

Response: Figure 6 is indeed included and embedded in the manuscript (page 40). Legends to Figures (including Figure 6) are indicated on page 22 in the manuscript.

REVIEWER 5

The manuscript of 'Metabolic and hepatic effects of liraglutide, obeticholic acid and elafibranor in diet-induced obese mouse models of biopsy-confirmed nonalcoholic steatohepatitis' presents some informative and interesting findings about the therapeutic efficacy of liraglutide, obeticholic acid and elafibranor in rodent nonalcoholic steatohepatitis, such as the improvements in hepatic

steatosis, inflammation, and liver firbosis. Because of the limited approach to clinical interference, these achievements could be valuable to both pharmaceutical research and clinical application.

1. Major comments. According to the results obtained from present study, liraglutide, obeticholic acid and elafibranor demonstrate therapeutic effects in different pathological characteristics of nonalcoholic steatohepatitis. For example, elafibranor and OCA reduce hepatic steatosis and inflammation in both DIO-NASH and ob/ob-NASH. Elafibranor attenuates liver fibrosis. Liraglutide improves scores of steatosis, fibrosis and inflammation. However, there is weak evidence for the drug comparison. Revision in this aspect is then suggested.

Response: It is unclear what the reviewer refers to when stating that there is weak evidence for the drug comparison. As the study comprises meticulous head-to-head comparisons of three important drugs in advanced clinical development for NASH, we thoroughly discuss and compare the effects of each individual compound in the two models of mouse NASH. As also experienced in the clinic, the compounds have very different pharmacological effects, and thus also show different impact on NASH pathology. This aspect is also reflected in the Discussion.

2. The effects of liraglutide are reported to be 'Liraglutide improves steatosis scores in DIO-NASH mice only, but reduces total steatosis, fibrosis and inflammation levels in both models'. Is there any conflict?

Response: There is no conflict in these data sets. It is important to distinguish between the qualitative (scoring) and quantitative assessments (histochemical and immunohistochemical stainings) of histopathology in DIO-NASH (and ob/ob-NASH) mice. This makes these two sets of parameters not well-suited for direct comparisons. Therefore, the liver pathology scores will not always correlate to quantitative changes in histopathology and vice versa. The disparity between qualitative and quantitative histological data is discussed where relevant.

3. It's unclear whether liraglutide, obeticholic acid and elafibranor exert their effect in a dose-dependent manner. Therefore, the rationality of dose selection should be clarified, at least be discussed carefully.

Response: The compound doses used in the present study are within the dose range reported efficacious in other mouse models of diet-induced obesity, diabetes and NASH. This is now indicated with relevant references in the Materials & Methods section (p. 7, line 13-14). It should be noted that we discuss potential dose-response effects (e.g. for OCA, p. 18, line 1-7).

4. The Abstract draw a conclusion of 'Diet-induced mouse models of biopsy-confirmed NASH show distinct treatment effects of liraglutide, OCA, and elafibranor, being in general agreement with corresponding findings in clinical trials for NASH'. Then, what's the novelty for these findings?

Response: The present paper addresses the translatability of the two obese mouse models of NASH in the context of treatment effects of three compounds in current advanced clinical development for NASH. Drug pharmacodynamics between humans and animal models of the relevant disease are particularly important to compare when selecting the right animal model for preclinical

evaluation of novel drugs. The novelty in the present findings argues for comparable metabolic and histological effects in the two mouse models of NASH and findings in corresponding clinical trials, which strongly supports the utility of these models in preclinical drug development for NASH. This is already mentioned in the Conclusion (p. 19), but is now also added to the abstract (p.2, line 21-23).

5. Minor comments. Limited information has been exhibited in the result of Abstract. It will be appreciated for authors to revise it, if possible, with more numerical details.

Response: We appreciate this view by the reviewer. Unfortunately, the WJG format requirements on information to be indicated in the abstract, as well as the extensive data package provided in the report, limits the word counts in the Results paragraph of the abstract, and we can therefore only provide the information already indicated.