EMBO reports

Transforming primary human hepatocytes into hepatocellular carcinoma with genetically defined factors

Zhiwu Jiang, Lin Cheng, Zhiping Wu, Linfu Zhou, Haitao Wang, Qilan Hong, Qiting Wu, Youguo Long, Yunlin Huang, Gaoqi Xu, Yao Yao, Zhaoyang Tang, Zhenfeng Zhang, Lili Yang, Wei Luo, Jie Yang, Likun Gong, Pentao Liu, Xinwen Chen, ShuZhong Cui, Qi Zhang, Yinxiong Li, and Peng Li

DOI: 10.15252/embr.202154275

Corresponding author(s): Peng Li (li_peng@gibh.ac.cn) , Yinxiong Li (li_yinxiong@gibh.ac.cn)

Review Timeline:

Submission Date:	4th Nov 21
Editorial Decision:	15th Nov 21
Revision Received:	27th Jan 22
Editorial Decision:	16th Mar 22
Revision Received:	2nd Apr 22
Accepted:	7th Apr 22

Editor: Achim Breiling

Transaction Report: This manuscript was transferred to EMBO reports following peer review at The EMBO Journal.

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Dear Dr. Li,

Thank you for transferring your manuscript to EMBO reports. I now went through your manuscript, the referee reports from The EMBO Journal (attached again below) and your revision plan (point-by-point response). Referees #2 and #3 have raised remaining concerns and suggestions to improve the manuscript, or to strengthen the data and the conclusions drawn.

EMBO reports emphasizes novel functional over detailed mechanistic insight, but asks for clear physiological relevance of the findings, and strong experimental support of the major conclusions. It will thus be necessary that in a revised manuscript you address all the points questioning the main conclusions of the study, and all technical concerns, or points regarding the experimental design, model systems used, or data presentation.

Given the constructive referee comments, we would like to invite you to further revise your manuscript with the understanding that all referee concerns must be addressed in the revised manuscript and in a detailed point-by-point response (as indicated in your revision plan). Acceptance of your manuscript will depend on a positive outcome of a final round of review at EMBO reports and will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

When submitting your revised manuscript, please also carefully review the instructions that follow below.

PLEASE NOTE THAT upon resubmission revised manuscripts are subjected to an initial quality control prior to exposition to re-review. Upon failure in the initial quality control, the manuscripts are sent back to the authors, which may lead to delays. Frequent reasons for such a failure are the lack of the data availability section (please see below) and the presence of statistics based on n=2 (the authors are then asked to present scatter plots or provide more data points).

When submitting your revised manuscript, we will require:

1) a .docx formatted version of the final manuscript text (including legends for main figures, EV figures and tables), but without the figures included. Please make sure that changes are highlighted to be clearly visible. Figure legends should be compiled at the end of the manuscript text.

2) individual production quality figure files as .eps, .tif, .jpg (one file per figure), of main figures and EV figures. Please upload these as separate, individual files upon re-submission.

The Expanded View format, which will be displayed in the main HTML of the paper in a collapsible format, has replaced the Supplementary information. You can submit up to 5 images as Expanded View. Please follow the nomenclature Figure EV1, Figure EV2 etc. The figure legend for these should be included in the main manuscript document file in a section called Expanded View Figure Legends after the main Figure Legends section. Additional Supplementary material should be supplied as a single pdf file labeled Appendix. The Appendix should have page numbers and needs to include a table of content on the first page (with page numbers) and legends for all content. Please follow the nomenclature Appendix Figure Sx, Appendix Table Sx etc. throughout the text, and also label the figures and tables according to this nomenclature.

For more details please refer to our guide to authors:
http://www.embopress.org/page/journal/14693178/authorguide#manuscriptpreparation

See our guide for figure preparation:
http://wol-prod-cdn.literatumonline.com/pb-assets/embo-site/EMBOPress_Figure_Guidelines_061115-1561436025777.pdf

See also the guidelines for figure legend preparation:
https://www.embopress.org/page/journal/14693178/authorguide#figureformat

3) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

4) a complete author checklist, which you can download from our author guidelines (https://www.embopress.org/page/journal/14693178/authorguide). Please insert page numbers in the checklist to indicate where the requested information can be found in the manuscript. The completed author checklist will also be part of the RPF.

Please also follow our guidelines for the use of living organisms, and the respective reporting guidelines:
http://www.embopress.org/page/journal/14693178/authorguide#livingorganisms

5) that primary datasets produced in this study (e.g. RNA-seq, ChIP-seq and array data) are deposited in an appropriate public

database. This is now mandatory (like the COI statement). If no primary datasets have been deposited in any database, please state this in this section (e.g. 'No primary datasets have been generated and deposited').

See also: http://embor.embopress.org/authorguide#datadeposition

Please remember to provide a reviewer password if the datasets are not yet public.

The accession numbers and database should be listed in a formal "Data Availability " section (placed after Materials & Methods) that follows the model below. Please note that the Data Availability Section is restricted to new primary data that are part of this study.

Data availability

The datasets produced in this study are available in the following databases:

- RNA-Seq data: Gene Expression Omnibus GSE46843 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46843)
- [data type]: [name of the resource] [accession number/identifier/doi] ([URL or identifiers.org/DATABASE:ACCESSION])

*** Note - All links should resolve to a page where the data can be accessed. ***

Moreover, I have these editorial requests:

6) We strongly encourage the publication of original source data with the aim of making primary data more accessible and transparent to the reader. The source data will be published in a separate source data file online along with the accepted manuscript and will be linked to the relevant figure. If you would like to use this opportunity, please submit the source data (for example scans of entire gels or blots, data points of graphs in an excel sheet, additional images, etc.) of your key experiments together with the revised manuscript. If you want to provide source data, please include size markers for scans of entire gels, label the scans with figure and panel number, and send one PDF file per figure.

7) Our journal encourages inclusion of *data citations in the reference list* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at: http://www.embopress.org/page/journal/14693178/authorguide#referencesformat

8) Regarding data quantification and statistics, please specify, where applicable, the number "n" for how many independent experiments (biological or technical replicates) were performed, the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values in the respective figure legends. Please provide statistical testing where applicable, and also add a paragraph detailing this to the methods section. See:
http://www.embopress.org/page/journal/14693178/authorguide#statisticalanalysis

9) Please note our reference format:
http://www.embopress.org/page/journal/14693178/authorguide#referencesformat

10) Please provide the abstract in present tense and make sure it has not more than 175 words.

11) For microscopic images, please add scale bars of similar style and thickness to all the microscopic images, using clearly visible black or white bars (depending on the background). Please place these in the lower right corner of the images. Please do not write on or near the bars in the image but define the size in the respective figure legend.

Please also note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised manuscript. We will thus need the ORCID of co-corresponding author Yinxiong Li. Please find instructions on how to link the ORCID ID to the account in our manuscript tracking system in our Author guidelines:
http://www.embopress.org/page/journal/14693178/authorguide#authorshipguidelines

I look forward to seeing a revised form of your manuscript when it is ready. Please use this link to submit your revision:
https://embor.msubmit.net/cgi-bin/main.plex

Best,

Achim Breiling

Achim Breiling
Editor
EMBO Reports

Referee #1:

My concerns have been largely met, the authors have made a substantial effort to improve their work and I don´t see any reason why this should not be published.

Referee #2:

The authors failed to delineate new mechanisms using their humanized mouse model. In their response, the correlation of the secreted human ALB in serum and in vivo repopulation of PHHs in Table S1 was not consistent with the previous study (PMID: 17664939), indicating that the transplantation system is not well established.

In addition, the gene mutation combination was not representative of human HCCs. I don't think this study modeling the initiation and progression of human HCCs.

Referee #3:

I realize that the two other reviewers were more enthusiastic than me. Still, I do not feel that the model presented is a good model for human HCC for the same main reasons I outlined in the first place: 1. the combination of mutations is not representative of human HCC; 2. the retained wt p53 is against the dogma. There are also some technical caveats that are not resolved as outlined below.
The following comments address the authors' responses to my comments, according to the numbering in their PBP response:
1. I accept their interpretation that a two-day expansion period would likely not make a huge difference and that it is likely that most of the selection process occurs in situ.
2. I remain skeptical. WNT pathway mutations are not the only way to induce myc overexpression and it is unlikely that the degree of myc overexpression in beta-catenin mutated HCC is similar to that observed with their artificial myc overexpression system. Similarly, the authors wish to regard all RTK and PI3K mutant tumors as similar to RAS mutant tumors. This is wrong, at least in my mind. A better choice of the driver combinations in the first place would have circumvented the need for such argumentation.
3. The clustering analysis is compelling and I agree that it supports the authors' claim that their model recapitulates aggressive human HCC.
4. I'm afraid that it is wrong to overexpress a mutP53 gene while maintaining 2 intact WT p53 alleles. In their answer the authors go at length to explain why there is no senescence. While their explanation is interesting, it ignores the elephant in the room - that their tumors have two intact wt p53 alleles and it's not clear how their model works. The correct way to do this is to inactivate p53 either genetically or with siRNA.
5. The in vivo CART cell experiments are indeed impressive.
6. OK
7. The staining is better. Notably, the Ki-67 index is way over 50% which is uncommon in human HCC, which rarely displays such high proliferation indices.
8. It is not OK to choose the cutoff that gives the lowest p value, without correcting for multiple comparisons (basically with 364 patients, the KM plotter will make 363 comparisons). I expect that with that after correcting the FDR will be higher than 0.05.
9. The staining is still not convincing. FAP is a type II transmembrane protein. As such, for true staining one would expect at least a hint of membrane enhancement of the stain. The protein atlas pictures are also not perfect. It should not be used as a reference - it is a site that mostly tests the antibodies from one company. For protein atlas to be a useful guide, one should look at tissues were a protein is known to be expressed - in FAP's case, that would be stromal cells of the endometrium, and alpha cells in the pancreas. Looking at their pancreas stains, they show nice alpha cell staining, but significant background staining, higher than that observed in the HCCs - so I wouldn't rely on this.

Dear Dr. Breiling,

We thank the reviewers for their time and constructive critiques. By responding to all of their suggestions, we have carried out an additional experimental work and extensive revisions to address their concerns. Accordingly, we have addressed each of the concerns raised by the reviewers, and respond to their queries in a point-by-point response below. We feel that the new manuscript contains more compelling data to support the main conclusions.

Referee #1:

my concerns have been largely met, the authors have made a substantial effort to improve their work and I don´t see any reason why this should not be published.

Our response: We appreciate the feedback.

Referee #2:

The authors failed to delineate new mechanisms using their humanized mouse model.

Our response: We agree that we did not delineate new mechanisms using the humanized mouse model. However, a main point of this manuscript is to show that for the first time we can transform primary human hepatocytes into aggressive HCC *in situ*. We have admitted that new mechanisms underlying how $TP53^{R249S}$, MYC, and $KRAS^{G12D}$ cooperatively transform primary human hepatocytes (PHHs) into HCC need to be addressed in a following study in the discussion of the revised manuscript at page 15, lines 1-2.

In their response, the correlation of the secreted human ALB in serum and in vivo repopulation of PHHs in Table S1 was not consistent with the previous study (PMID: 17664939), indicating that the transplantation system is not well established.

Our response: Though the transplantation system may not be as good as the previous study, the sophistication of transplantation system is not a key point in this study or has little effects on the strength of our data, as tumor growth is a positive selection process. Instead, we want to show readers that overexpression of $TP53^{R249S}$, MYC, and $KRAS^{G12D}$ transforms PHHs into HCC in vivo. Nevertheless, the reconstitution protocol needs to be further optimized. We have admitted this point as a limit of this study in the discussion at page 15, lines 13-14.

In addition, the gene mutation combination was not representative of human HCCs. I don't think this study modeling the initiation and progression of human HCCs.

Our response: We agree that the gene mutation combination was not representative of human HCCs. However, as the reviewer#3 agrees, our clustering analysis (Figure 5F-G) shows that iHCC samples recapitulate aggressive human HCC. Thus, we have changed the title of the revised manuscript as "Transforming primary human hepatocytes to hepatocellular carcinoma with genetically defined factors".

Referee #3:

I realize that the two other reviewers were more enthusiastic than me. Still, I do not feel that the model presented is a good model for human HCC for the same main reasons I outlined in the first place: 1. the combination of mutations is not representative of human HCC; 2. the retained wt p53 is against the dogma. There are also some technical caveats that are not resolved as outlined below.

The following comments address the authors' responses to my comments, according to the numbering in their PBP response:

1. I accept their interpretation that a two-day expansion period would likely not make a huge difference and that it is likely that most of the selection process occurs in situ.

Our response: We appreciate the feedback.

2. I remain skeptical. WNT pathway mutations are not the only way to induce myc overexpression and it is unlikely that the degree of myc overexpression in beta-catenin mutated HCC is similar to that observed with their artificial myc overexpression system. Similarly, the authors wish to regard all RTK and PI3K mutant tumors as similar to RAS mutant tumors. This is wrong, at least in my mind. A better choice of the driver combinations in the first place would have circumvented the need for such argumentation.

Our response: We agree with the critique that the combination of $TP53^{R249S}$, MYC, and $KRAS^{G12D}$ was not representative of human HCCs. We have admitted this point as a limit of study in the discussion of revised manuscript at page 14, lines 23-25.

3. The clustering analysis is compelling and I agree that it supports the authors' claim that their model recapitulates aggressive human HCC.

Our response: We appreciate the comment.

4. I'm afraid that it is wrong to overexpress a mutP53 gene while maintaining 2 intact WT

p53 alleles. In their answer the authors go at length to explain why there is no senescence. While their explanation is interesting, it ignores the elephant in the room - that their tumors have two intact wt p53 alleles and it's not clear how their model works. The correct way to do this is to inactivate p53 either genetically or with siRNA.

Our response: We agree with the critique. According to the referee's advice, we inactivated p53 by CRISPR/Cas9 in PHHs, where $TP53^{R249S}$, MYC, and $KRAS^{G12D}$ were overexpressed and orthotopically transplanted these genetic modified PHHs into NSIF mice for tumor watch (Figure EV3B, C). Tumors (3 out of 4 mice) were observed in the recipient mice transplanted with PHHs overexpressing MTK with or without deletion of TP53 (Figure EV3D). These results demonstrate that overexpression of $TP53^{R249S}$, MYC, and $KRAS^{G12D}$ induced PHHs with WT p53 inactivation to transform into iHCC in vivo. We have provided these results in the revised manuscript at page 8, lines 7-11.

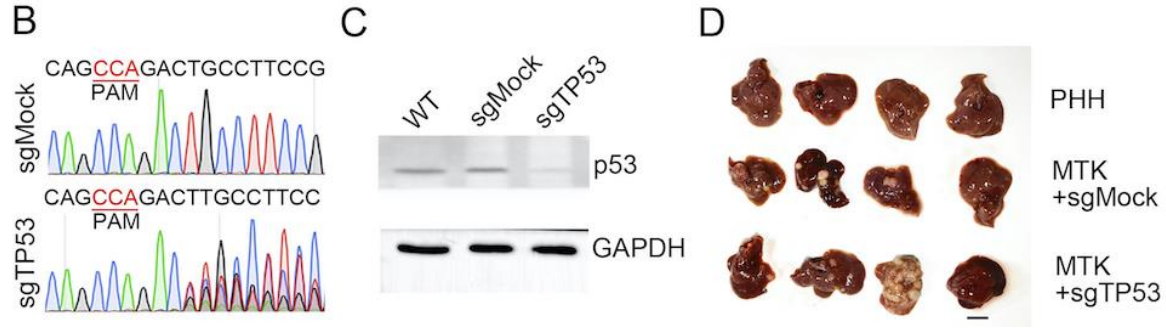

Figure EV3B-D: (B) DNA sequencing confirmed mutations of TP53 in the genomic DNA of sgTP53 transduced PHH cells. (C) Expression of p53 was detected in the PHH cells transduced with sgMock or sgTP53 by Western blotting. (D) Representative images of in situ liver carcinomas derived from PHHs transduced with a combination of MTK with or without deletion of TP53 in NSIF mice. 3 out of 4 mice were observed tumor in both MTK+ sgMock and MTK+ sgTP53 groups after 11 weeks. n = 4 for each group.

5. The in vivo CART cell experiments are indeed impressive.

Our response: We appreciate the feedback.

6. OK

Our response: We appreciate the feedback.

7. The staining is better. Notably, the Ki-67 index is way over 50% which is uncommon in human HCC, which rarely displays such high proliferation indices.

Our response: The high Ki-67 index as high proliferation indices is correlated with the aggressiveness of iHCC, as suggested by the referee#3. We have admitted that Ki-67 index in iHCC is not common in primary human HCC as a limit of the study in discussion of the revised manuscript at page13, lines 23- 25.

8. It is not OK to choose the cutoff that gives the lowest p value, without correcting for multiple comparisons (basically with 364 patients, the KM plotter will make 363 comparisons). I expect that with that after correcting the FDR will be higher than 0.05.

Our response: Thanks for this suggestion. We have re-analyzed TCGA datasets using the GEPIA tool rather than the KM plotter tool, according to the reviewer's previous suggestions. We found that four out of the top 20 genes (SLC34A2, FBN2, FOLR1, and SLC39A10) were still associated with poor prognosis in HCC patients. We have added Kaplan-Meier analysis of these four genes in HCC patients (Figure EV5A) in the revised manuscript at page 11, lines 24-25.

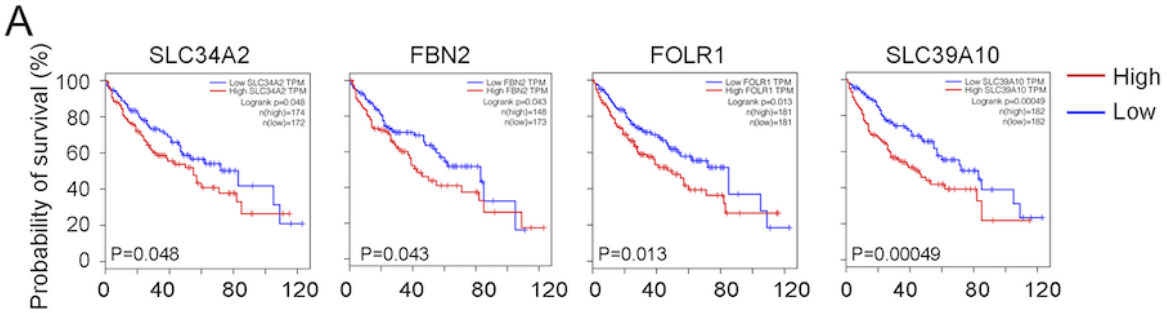

Figure EV5A: Kaplan-Meier analysis of the TCGA-HCC (TCGA-LIHC) cohorts based on the expression levels of SLC34A2, FBN2, FOLR1, and SLC39A10, in the cohort samples base on GEPIA site.

9. The staining is still not convincing. FAP is a type II transmembrane protein. As such, for true staining one would expect at least a hint of membrane enhancement of the stain. The protein atlas pictures are also not perfect. It should not be used as a reference - it is a site that mostly tests the antibodies from one company. For protein atlas to be a useful guide, one should look at tissues were a protein is known to be expressed - in FAP's case, that would be stromal cells of the endometrium, and alpha cells in the pancreas. Looking at their pancreas stains, they show nice alpha cell staining, but significant background staining, higher than that observed in the HCCs - so I wouldn't rely on this.

Our response: Thanks for this critique. We have purchased two more FAP antibodies from Abcam (anti-FAP antibody: ab207178) and CST (FAP (E1V9V): #66562). We found that the FAP antibody from Abcam is more specific than CST's FAP antibody. We have added new IHC images of iHCC and PHH with the new FAP antibody (ab207178) in the revised manuscript (Figure EV5B).

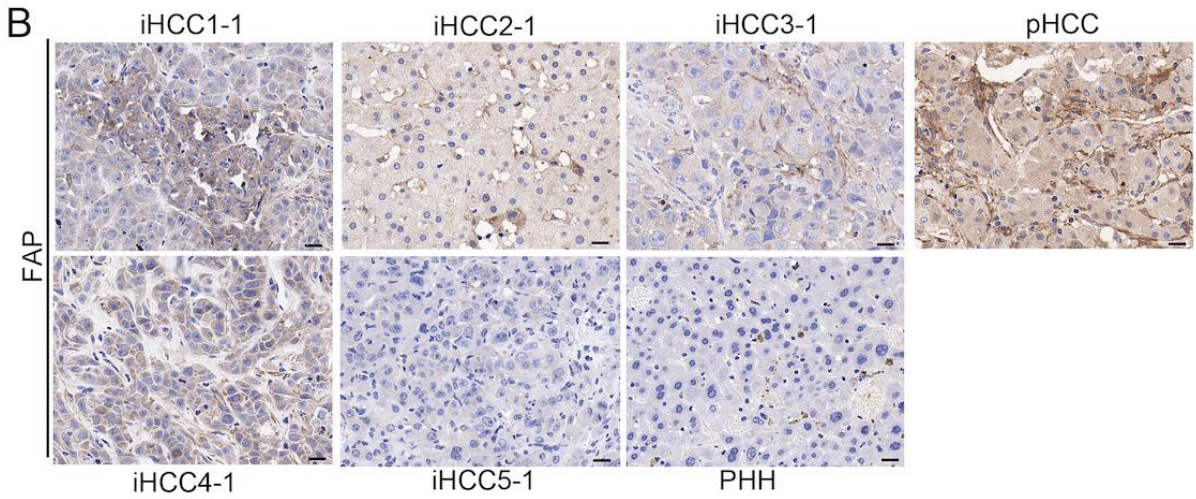

Figure EV5B: Representative IHC staining of FAP in a normal liver (PHH), primary HCC (pHCC) and five MTK-transduced iHCC tissues (iHCC1-1, iHCC2-1, iHCC3-1, iHCC4-1, and iHCC5-1) that were derived from five different donors (PHH1~5). Scale bars, 20 μm.

Dear Dr. Li,

Thank you for the submission of your revised manuscript to our editorial office. I have now received the report from the referee that was asked to re-evaluate the study, you will find below. As you will see, the referee now supports the publication of your study in EMBO reports. Nevertheless, s/he has some suggestions to improve the manuscript, I ask you to address in a final revised manuscript. Please also provide a p-b-p-response addressing these remaining points.

Moreover, I have these editorial requests I also ask you to address:

- Please have your final manuscript carefully proofread by a native speaker. There are still some typos and grammatical errors present.

- Please make sure that the number "n" for how many independent experiments were performed, their nature (biological versus technical replicates), the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values is indicated in the respective figure legends (for main and EV figures), and that statistical testing has been done where applicable. Please avoid phrases like 'independent experiment' or 'independent sample', but clearly state if these were biological or technical replicates. Please add complete statistical testing to all diagrams. Please also indicate (e.g. with n.s.) if testing was performed, but the differences are not significant.

- We would need some adjustments to the Appendix file. Please remove the authors and affiliations from the first page. It is sufficient to state here 'Appendix for: "Transforming primary human hepatocytes to hepatocellular carcinoma with genetically defined factors", followed by the Table of contents (TOC). Please put here a complete TOC ('figure of content' does not make sense), first the Figures, then the Tables with page numbers. Moreover, please move the legends below each figure. This is much easier to follow for the reader.

- As the few Western blots shown are significantly cropped, could you please provide the source data for the blots (main and EV figures). The source data will be published in a separate source data file online along with the accepted manuscript and will be linked to the relevant figure. Please submit the source data (scans of entire blots) together with the final revised manuscript. Please include size markers for the scans of entire blots, label the scans with figure and panel number, and send one PDF file per figure.

In addition, I would need from you:
- a short, two-sentence summary of the manuscript (not more than 35 words).
- two to four short bullet points highlighting the key findings of your study
- a schematic summary figure (in jpeg or tiff format with the exact width of 550 pixels and a height of not more than 400 pixels) that can be used as a visual synopsis on our website.

I look forward to seeing the final revised version of your manuscript when it is ready. Please let me know if you have questions regarding the revision.

Best,

Achim Breiling
Senior Editor
EMBO Reports

Referee #1:

The authors have adequately answered all my concerns and questions. I only have a few minor comments which do not require any additional experiments.

- p7 line 20: add statistical analysis to show that this is statistically significant (it clearly is but better state this).

- Fig. 4C: the CK19 epxression is not typical of HCC, it is more a bad prognostic marker. It may also suggest a mixed lineage tumor. However, the morphology (as in 3E) is more typical of classic HCC. Maybe show another representative figure, or else change the text in line with a marker for prognosis rather than for HCC in general.

- Fig. 5E: the dsDNA staining looks too much to me (I would expect to see very few foci in a positive cell, and this only in a few cells, but this is not within my specialty). On the other hand, I am not sure that this finding is needed for the manuscript's claim. I leave it to the authors' discretion to decide whether or not they want to leave this in, depending on their level of confidence in their staining.

- The sentence: "Though KRAS mutant have not been reported in HCC tumors that harbor p53 mutations or MYC amplification in TCGA database, we have identified that 126 tumors with TP53 mutations, 140 tumors with mutations in the RTK/RAS/PI3K signaling pathways (KRAS, MET, PIK3CA, PTEN, NF1, and NRAS), and 197 tumors with the MYC/WNT pathways (CTNNB1, AXIN1, APC, and MYC) from 348 HCC tumors obtained from the TCGA database." needs to be edited as this important sentence does not make a conclusion coming after "though... we have identified that...". i.e. state that while KRAS mutations are not common, you believe that they are replaced by other sources of RTK activation.

Dear Dr. Breiling,

We thank the reviewers for their time and constructive critiques. By responding to all of their suggestions, we have addressed each of the concerns raised by the editor and reviewers, and respond to their queries in a point-by-point response below.

Moreover, I have these editorial requests I also ask you to address:

- Please have your final manuscript carefully proofread by a native speaker. There are still some typos and grammatical errors present.

Our response: We are sorry for the typos and grammatical errors. The revised manuscript has been proofread and edited by Professor Jean Paul Thiery from Guangzhou Laboratory.

- Please make sure that the number "n" for how many independent experiments were performed, their nature (biological versus technical replicates), the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values is indicated in the respective figure legends (for main and EV figures), and that statistical testing has been done where applicable. Please avoid phrases like 'independent experiment' or 'independent sample', but clearly state if these were biological or technical replicates. Please add complete statistical testing to all diagrams. Please also indicate (e.g. with n.s.) if testing was performed, but the differences are not significant.

Our response: We appreciate the advice. We have corrected the inappropriate statistical description according to your instructions in revised manuscript.

- We would need some adjustments to the Appendix file. Please remove the authors and affiliations from the first page. It is sufficient to state here 'Appendix for: "Transforming primary human hepatocytes to hepatocellular carcinoma with genetically defined factors", followed by the Table of contents (TOC). Please put here a complete TOC ('figure of content' does not make sense), first the Figures, then the Tables with page numbers. Moreover, please move the legends below each figure. This is much easier to follow for the reader.

Our response: We appreciate the suggestion. We have removed the authors and affiliations from the first page, added the manuscript title followed by the Table of Contents, and moved the legends below each figure in the Appendix file.

- As the few Western blots shown are significantly cropped, could you please provide the source data for the blots (main and EV figures). The source data will be published in a separate source data file online along with the accepted manuscript and will be linked to the relevant figure. Please submit the source data (scans of entire blots) together with the final revised manuscript. Please include size markers for the scans of entire blots, label the scans with figure and panel number, and send one PDF file per figure.

Our response: We have uploaded original source data for the blots (Figure 3I, and Figure EV3C), according to your instruction, together with the final revised manuscript. However, we are sorry that the original blot of Figure 1C lacks protein ladders. We thus repeated the western blot experiment and replaced the original Figure 1C with the new figure and uploaded the source data for the blot of Figure 1C in the revised manuscript.

In addition, I would need from you:

- a short, two-sentence summary of the manuscript (not more than 35 words).
- two to four short bullet points highlighting the key findings of your study
- a schematic summary figure (in jpeg or tiff format with the exact width of 550 pixels and a height of not more than 400 pixels) that can be used as a visual synopsis on our website.

Our response: We have submitted these files together in the synopsis file together with the revised manuscript.

Referee #1:

The authors have adequately answered all my concerns and questions. I only have a few minor comments which do not require any additional experiments.

- p7 line 20: add statistical analysis to show that this is statistically significant (it clearly is but better state this).

Our response: We appreciate the suggestion. According to the referee's advice, we add that tumorigenic rate of MTK-transduced PHHs is significantly higher than that of 16.6% (2 out of 12 mice) MT-transduced PHHs (Figure 4A and Table 1). We have provided statistical analysis in the revised manuscript at page 30, lines 11-12.

- Fig. 4C: the CK19 expression is not typical of HCC, it is more a bad prognostic marker. It may also suggest a mixed lineage tumor. However, the morphology (as in 3E) is more typical of classic HCC. Maybe show another representative figure, or else change the text in line with a marker for prognosis rather than for HCC in general.

Our response: Thanks for the advice. We have change the description as “Notably, iHCC samples (iHCC1-1 and iHCC3-1) expressed cytokeratin-19 (CK19) and EpCAM, which are expressed in hepatic progenitors (Rao et al, 2008) and are considered as poor prognostic markers (Chan et al, 2014; Govaere et al, 2014; Llovet et al, 2015) in HCC patients (Figures 4C, EV3A).” in the revised manuscript at page 8, lines 3-6.

- Fig. 5E: the dsDNA staining looks too much to me (I would expect to see very few foci in a positive cell, and this only in a few cells, but this is not within my specialty). On the other hand, I am not sure that this finding is needed for the manuscript's claim. I leave it to the authors' discretion to decide whether or not they want to leave this in, depending on their level of confidence in their staining.

Our response: Thanks for the suggestion. To avoid confusion, we have deleted the original Fig. 5E from the revised manuscript.

- The sentence: "Though KRAS mutant have not been reported in HCC tumors that harbor p53 mutations or MYC amplification in TCGA database, we have identified that 126 tumors with TP53 mutations, 140 tumors with mutations in the RTK/RAS/PI3K signaling pathways (KRAS, MET, PIK3CA, PTEN, NF1, and NRAS), and 197 tumors with the MYC/WNT pathways (CTNNB1, AXIN1, APC, and MYC) from 348 HCC tumors obtained from the TCGA database." needs to be edited as this important sentence does not make a conclusion coming after "though... we have identified that...". i.e. state that while KRAS mutations are not common, you believe that they are replaced by other sources of RTK activation.

Our response: We appreciate the advice. We have modified this sentence as “Though KRAS mutations have not been reported in HCC tumors that harbor p53 mutations or MYC amplification in TCGA database, we have identified that 126 tumors with TP53 mutations, 140 tumors with mutations in the RTK/RAS/PI3K signaling pathways (KRAS, MET, PIK3CA, PTEN, NF1, and NRAS), and 197 tumors with the MYC/WNT pathways (CTNNB1, AXIN1, APC, and MYC) from 348 HCC tumors obtained from the TCGA database (Figure EV4B and Dataset EV2). Therefore, while KRAS mutations are not

common, we believe that they are replaced by other sources of RTK activation.” in the revised manuscript at page 11, lines 1-7.

Dr. Peng Li
Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences
Center for Cell Regeneration and Biological Therapies
190 kaiyuan avenue
Guangzhou, Guangdong 510530
China

Dear Dr. Li,

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

At the end of this email I include important information about how to proceed. Please ensure that you take the time to read the information and complete and return the necessary forms to allow us to publish your manuscript as quickly as possible.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. As you are aware, this File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

If you do NOT want this File to be published, please inform the editorial office within 2 days, if you have not done so already, otherwise the File will be published by default [contact: emboreports@embo.org]. If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case." Please note that the author checklist will still be published even if you opt out of the transparent process.

Thank you again for your contribution to EMBO reports and congratulations on a successful publication. Please consider us again in the future for your most exciting work.

Yours sincerely,

Achim Breiling
Editor
EMBO Reports

****

THINGS TO DO NOW:

Please note that you will be contacted by Wiley Author Services to complete licensing and payment information. The required 'Page Charges Authorization Form' is available here: https://www.embopress.org/pb-assets/embo-site/er_apc.pdf

You will receive proofs by e-mail approximately 2-3 weeks after all relevant files have been sent to our Production Office; you should return your corrections within 2 days of receiving the proofs.

Please inform us if there is likely to be any difficulty in reaching you at the above address at that time. Failure to meet our deadlines may result in a delay of publication, or publication without your corrections.

All further communications concerning your paper should quote reference number EMBOR-2021-54275V3 and be addressed to emboreports@wiley.com.

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Peng Li
Journal Submitted to: EMBO reports
Manuscript Number: EMBOR-2021-54275V2

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Pilot studies were used for estimation of the sample size to ensure adequate power. In most of the experiments, 3 to 10 /samples was sufficient to identify differences between groups with a 5% significance level.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Pilot studies were used for estimation of the sample size to ensure adequate power. In most of the experiments, 3 to 10 mice/samples was sufficient to identify differences between groups with 5% significance level.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	No data were excluded.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Allocation into experimental groups was done randomly
For animal studies, include a statement about randomization even if no randomization was used.	Allocation into experimental groups was done randomly
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	Allocation into experimental groups was done randomly
4.b. For animal studies, include a statement about blinding even if no blinding was done	Allocation into experimental groups was done randomly
5. For every figure, are statistical tests justified as appropriate?	For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).
Is there an estimate of variation within each group of data?	NO

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumo

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
https://osp.od.nih.gov/biosafety-biosecurity-and-emerging-biotechnology/
http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	YES

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Antibodies information were given at material and methold section.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	293T cells were purchased from ATCC which are mycoplasma free.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	ICR mice were purchased from Bejing Vital River Laboratory Animal Technology Ltd. NSI mice were generated by direct targeting NOD/SCID mice via TALEN to IL2rg. NSIF mice were generated by direct targeting NSI mice via TALEN to fah gene. All mice were used at 8-12 weeks of age. Mice were housed in the SPF grade animal facility of the Guangzhou Institution of Biomedicine and Health, Chinese Academy of Science (GIBH, CAS, China). All animal experiments were approved by the Institutional Animal Care and Use Committee of Guangzhou Institutes of Biomedicine and Health (IACUC-GIBH).
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	All animal experiments were approved by the Institutional Animal Care and Use Committee of Guangzhou Institutes of Biomedicine and Health (IACUC-GIBH).
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	YES

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	Primary human hepatocytes were purchased from Bioreclamation IVT (Baltimore, MD, USA). All experiments were approved by the Institutional Animal Care and Use Committee of Guangzhou Institutes of Biomedicine and Health (IACUC-GIBH).
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	This study did not involve human research participants
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	This study did not involve human research participants
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	This study did not involve human research participants
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	This study did not involve human research participants
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	This study did not involve human research participants
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	This study did not involve human research participants

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	RNA-Seq data: Gene Expression Omnibus GSE143165 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE143365).
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	no data
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	no data
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	no novel code for computational models.

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NO