nature portfolio

Peer Review File

Multi-omic analysis of hepatocellular carcinoma reveals aberrant cis-regulatory changes and dysregulated retrotransposons with prognostic potential

Corresponding Author: Professor Danny Leung

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
While the overall results of the integration strategy are interesting, some of the original ideas such as the epigenetic landscape between normal liver and HCC, had already started to be explored in previous publications. Despite potentially interesting results, it is an extremely long manuscript regarding all main figures (8) and supporting material. Understandably, the authors want to show the many results collected. However, the integration of the results, which are complex per se, is rather long and very descriptive, and the authors focus on too many details and examples of altered methylation/transcription throughout the results, often switching between a main figure/supplementary figures/supplementary tables, making the manuscript hard to read and follow properly. With so much information input, the reader often loses focus on the interesting results.
NOTE: It would help the reader to have the Supplementary Tables numbered/identified according to main text.

1. More focused results and size reduction of the Results sections to the main highlights/results are highly recommended. The manuscript should be shortened to aid the communication of the most important findings.

2. In Extended Data Fig. 4. Characterization of dysregulated typical enhancers and SEs in HCC. a Volcano plot…" it is not totally clear what the volcano plot data refers to, i.e., the authors mention "Activated… and inactivated typical enhancers…" but based on what data? H3K27ac enrichment? H3K9me3? Integration of both/other data? On panel b) of Extended Data Fig. 4 then the box plots refer to H3K27me3 and H3K9me3 data, but in panel a) is not clear; thus, please clarify.

3. The authors have based the results on data coming from solely 7 patients. Several of the findings are shown as potentially translatable (e.g., HERVE-int as "a potential prognostic biomarker and predictor of immunotherapy response") but are not validated in any patients outside the small cohort. Given the known heterogeneity present in HCC, the authors have not discussed anywhere in the manuscript the limitations of such short cohort.

4. Have the authors searched in public databases for example, for RNAseq datasets with prognostic value as metadata to confirm findings such as expression levels of retroposons like HERVE-int? This would have been very interesting to see. Otherwise, the many suggestions of applicability in prognosis and drug selection in the Discussion sections are merely highly speculative. Please comment.

5. The authors have compared the HCC with its normal counterpart (TAN) within each patient. While it is understandable that the authors wanted to control for the variability between individuals, it is known that normal cells in the neighbourhood of a tumour are not entirely normal and often reflect changes that have occurred during the progression of the "neighbour" tumour. The phenomenon is often referred to as "field cancerization" or "field effect". Assuming that this is occurring in the TANs of the analyzed patients, the findings may not be present in the livers of healthy individuals and/or other important findings may have been missed due to this lack of comparison. The authors have not addressed this limitation anywhere in the manuscript. What about comparing the HCC data with Normal liver profiles from Healthy individuals? Does the presented data still stand?

Reviewer #2

(Remarks to the Author)
Review of “Multi-omic analysis of hepatocellular carcinoma reveals aberrant cis-regulatory changes and dysregulated retrotransposons with prognostic potentials” by Cheng et al. In this study, the authors performed multi-omic profiling of hepatocellular carcinoma (HCC) patient samples from the tumor itself and non-tumor adjacent regions with the ultimate goal of understanding how these epigenetic marks collaborate to regulate genes and retrotransposons. The identify both the gene GPC3 and the retrotransposon HERVE-int as being epigenetically deregulated in HCC. The authors elegantly manipulate expression of these two factors through a CRISPRi-based approach targeting associated enhancers demonstrating that these enhancers are, at least in part, responsible for the altered expression level in HCC cells. They then further focus upon links between HERVE-int and clinical outcomes. Overall, this study is technically well done and convincing. However, this is tempered by a variety of already published literature on GPC3 and HERVE-int in HCC with specific studies examining epigenetic regulation of these two factors (e.g. GPC3: PMID39725192 HERVE-int PMC11889587) A handful of minor comments are below.
Comments
On line 683, the authors mention that a log2 fold change of -0.01 was used as significant. Is this correct, or a typo?
Would it be beneficial to include either “intergenic” or “whole genome” as an annotation to Figure 2b (similar to Figure 2e)?
On a related note, do boxplots suppress the bimodal distribution that we typically see with DNA methylation? Perhaps histograms/density plots of each feature would be more illustrative?
Seeing as Figure 2d selected for hyper DMRs, the distribution between TAN and HCC is preselected for, making this graph redundant. Perhaps a heatmap or binned scatterplot could highlight the baseline level of methylation in TAN relative to HCC tissue in these specific regions.
In Figure 2f, does the methylation level of the protein-coding genes in PMDs impact the expression (and therefore explain the difference between TAN/HCC)?
In Figure 2f/g, are the PMDs linked to the specific samples in which expression is shown? I ask this because of the heterogeneity that the authors observe at the histone level (Figure 1c/d) and wonder if there is also heterogeneity of PMDs in each sample, which may influence the expression pattern.
Was any validation of the imputation method performed? E.g. take a sample that worked well, artificially remove a percentage/all of the data, impute, and compare? I say this because of how 865T H3K4me3 is very clearly different from all tumors in Figure 3d. It sounds like maybe this was performed from lines 711-720, but unclear if this performed well (i.e. is there some imputation quality score/metric like R2). More additional detail could be used for imputation as well. Were all samples used, or just those from the same cohort (e.g. impute TAN only from other TAN)?
Is there any empirical evidence to suggest that the inactivated super-enhancer in ADAMTS2 (Figure 4f) loops to the promoter (e.g. GeneHancer)? The authors use HepG2 HiC in Figure 5c, but presumably that interaction is lost due to HepG2 being transformed and the SE lost.
Is a non-tumor/tumor difference in GPC3 expression observed in TCGA? Figure 5b shows that LIHC is relatively highly expressed relative to other tissues, but unclear if that significance between non-tumor/tumor is there. On a related note, line 969 expresses that p-values are shown for panel b in Figure 5, but I do not observe any.
HERVE-int has been shown to be deregulated in HCC through SETDB1/H3K9me3 (PMC11889587). A discussion of this should be done. Similarly, published studies on the epigenetic regulation of GPC3 could be expanded upon.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
In this revised version, Cheng et al have significantly overhauled the manuscript to address all concerns of reviewer 2.

Response to Reviewers' comments:

Reviewer #1:

> While the overall results of the integration strategy are interesting, some of the original ideas such as the epigenetic landscape between normal liver and HCC, had already started to be explored in previous publications. Despite potentially interesting results, it is an extremely long manuscript regarding all main figures (8) and supporting material. Understandably, the authors want to show the many results collected. However, the integration of the results, which are complex per se, is rather long and very descriptive, and the authors focus on too many details and examples of altered methylation/transcription throughout the results, often switching between a main figure/supplementary figures/supplementary tables, making the manuscript hard to read and follow properly. With so much information input, the reader often loses focus on the interesting results.
>
> NOTE: It would help the reader to have the Supplementary Tables numbered/identified according to main text.

We thank the reviewer for taking the time to evaluate our submission. We appreciate that our manuscript was lengthy. We would emphasize that the collection of datasets generated will be a valuable resource to other researchers. Therefore, while the results of the initial integrative analysis may be expected, it is important for the context of understanding the subsequent results and confirming the quality of the data. Regardless, we have attempted to shorten and make the manuscript more concise by removing non-essential pieces of analysis and moved more overview and descriptive results to the supplemental information.

1. More focused results and size reduction of the Results sections to the main highlights/results are highly recommended. The manuscript should be shortened to aid the communication of the most important findings.

We have tried to streamline our presentation and shorten the manuscript. We feel it is still important to include the general DNA methylation and transcriptomic analyses but agree that it was taking up too much of the main results. We have condensed some of those results and moved them to **New Supplementary Figures 2 and 3**.

2. In Extended Data Fig. 4. Characterization of dysregulated typical enhancers and SEs in HCC. a Volcano plot…" it is not totally clear what the volcano plot data refers to, i.e., the authors mention "Activated… and inactivated typical enhancers…" but based on what data? H3K27ac enrichment? H3K9me3? Integration of both/other data? On panel b) of Extended Data Fig. 4 then the box plots refer to H3K27me3 and H3K9me3 data, but in panel a) is not clear; thus, please clarify.

We thank the reviewer for pointing this out. We have now added more details to the figure legends to improve the clarity. For example, we have modified the figure legend of **Supplementary Fig. 4a** to "Volcano plot showing dysregulated typical enhancers based on differential H3K27ac enrichment in HCC". Similar edits were made to other figure legends.

3. The authors have based the results on data coming from solely 7 patients. Several of the findings are shown as potentially translatable (e.g., HERVE-int as "a potential prognostic biomarker and predictor of immunotherapy response") but are not validated in any patients outside the small cohort. Given the known heterogeneity present in HCC, the authors have not discussed anywhere in the manuscript the limitations of such short cohort.

We appreciate the reviewer's point about the heterogeneity present in HCC patients. Indeed, this was our observation of differential reactivation of typical enhancers and super-enhancers across patients. Our multi-omic analyses aimed to explore mechanisms rather than survey large cohorts. To this end, we carried out functional studies in cell lines to understand how epigenetic changes impact the activities of the non-coding elements that we are focusing on. In our current submission, we included CRISPRi and additional shRNA knockdown experiments to show that the *GPC3* super-enhancer is normally repressed by DNA methylation, which when derepressed is subsequently activated by transcription factors such as CEBPα (**New Figure 3e**).

We should also emphasize that our conclusions were not only drawn from the 7 pairs of samples. The results about *GPC3* regulation, HERV-E-int lncRNA expression and the potential prognostic values of retrotransposons all involved analyses of The Cancer Genome Atlas-Liver Hepatocellular Carcinoma (TCGA-LIHC) cohort.

4. Have the authors searched in public databases for example, for RNAseq datasets with prognostic value as metadata to confirm findings such as expression levels of retroposons like HERVE-int? This would have been very interesting to see. Otherwise, the many suggestions of applicability in prognosis and drug selection in the Discussion sections are merely highly speculative. Please comment.

We thank the reviewer for his/her insight. We agree that further meta-analysis with the addition of treatment response would be a crucial next step. Unfortunately, to the best of our knowledge, there are currently no available datasets that are appropriate for such analyses. While there are some studies that analyzed responders versus non-responders to combined atezolizumab and bevacizumab therapy in HCC, their method of transcriptomic profiling is not informative for our analyses. For example, Zhu *et al.* Nat. Cancer 2022 (PMID: 35739268) generated capture RNA-seq datasets for a large cohort. However, the exon enrichment excludes retrotransposons. Therefore, we have not been able to find additional publicly available datasets to include.

Our current discussion on the prognostic and predictive values of retrotransposons is based on the expression of established markers for such traits. Our analysis of the TCGA cohort revealed significant correlation between HERV-E-int and the expression of these genes. In the current submission, we further compared the expressions of genes comprising of the ateozolizumab+bevacizumab response signature (ABRS) (PMID: 35739268) between HERVE-int-High and -Low patients. We found significantly higher expression of the majority of these genes in HERVE-int-High patients. (**New Figure 6i**). This provides an additional line of support for the predictive potential of retrotransposon activities.

5. The authors have compared the HCC with its normal counterpart (TAN) within each patient. While it is understandable that the authors wanted to control for the variability between individuals, it is known that normal cells in the neighbourhood of a tumour are not entirely normal and often reflect changes that have occurred during the progression of the "neighbour" tumour. The phenomenon is often referred to as "field cancerization" or "field effect". Assuming that this is occurring in the TANs of the analyzed patients, the findings may not be present in the livers of healthy individuals and/or other important findings may have been missed due to this lack of comparison. The authors have not addressed this limitation anywhere in the manuscript. What about comparing the HCC data with Normal liver profiles from Healthy individuals? Does the presented data still stand?

We agree with the reviewer's point that the TAN samples are not "normal" tissue. Rather they serve as a control for epigenomic analyses, which is standard practice. We did not have access to healthy liver samples as controls in this study. However, we did include normal liver data from the Roadmap Epigenomics Project in our analysis. This was shown in the majority of the relevant figures (For examples, see **New Figures 2f**, **3a**, **5b**, **Supplementary Figs. 4d**, **7a**, and **9a,b**). We should note that all the epigenomic and transcriptomic changes we report here were observed in HCC and not in TAN or normal liver datasets.

Reviewer #2 (Remarks to the Author):

> Review of “Multi-omic analysis of hepatocellular carcinoma reveals aberrant cis-regulatory changes and dysregulated retrotransposons with prognostic potentials” by Cheng et al. In this study, the authors performed multi-omic profiling of hepatocellular carcinoma (HCC) patient samples from the tumor itself and non-tumor adjacent regions with the ultimate goal of understanding how these epigenetic marks collaborate to regulate genes and retrotransposons. The identify both the gene GPC3 and the retrotransposon HERVE-int as being epigenetically deregulated in HCC. The authors elegantly manipulate expression of these two factors through a CRISPRi-based approach targeting associated enhancers demonstrating that these enhancers are, at least in part, responsible for the altered expression level in HCC cells. They then further focus upon links between HERVE-int and clinical outcomes. Overall, this study is technically well done and convincing. However, this is tempered by a variety of already published literature on GPC3 and HERVE-int in HCC with specific studies examining epigenetic regulation of these two factors (e.g. GPC3: PMID39725192 HERVE-int PMC11889587) A handful of minor comments are below.

We thank the reviewer for evaluating our submission. While we recognize that some of the initial findings are similar to those already published, we would like to clarify that our main findings are different to those mentioned by the reviewer. Firstly, we fully acknowledge and have cited the prior studies that report GPC3 as an HCC molecular marker and immunotherapy target. Our study did not attempt to validate this claim, rather, we aimed to leverage the epigenomic data to elucidate the molecular mechanism by which this HCC-specific biomarker gene becomes dysregulated. We demonstrate that the increased transcription of *GPC3* results from concomitant loss of DNA methylation at its *cis*-regulatory elements and the activation of a fetal liver super-enhancer. In our current submission, we added additional experiments and analyses to determine that the super-enhancer is regulated by the transcription factor CEBPα. These results have implications into how GPC3 immunotherapy could be potentially improved by combining epigenetic inhibitors, which will be the focus of future studies. In regard to the HERVE-int element, to the best of our knowledge, this has not been previously reported. The study that the reviewer mentioned looks at how SETDB1 silences specific HERV elements in HCC cell lines. Those are not the same elements that are being studied here (see response below). It is a common misconception that retrotransposons are regulated ubiquitously as a group. However, depending on the subfamily and individual element’s sequence and location, their regulation is distinct. We have amended the text to distinguish these points. We show here specific instances of the expression levels HERVE-int and THE1D subfamilies, which are repressed predominantly by DNA methylation, can significantly stratify HCC patient survival and are correlated with molecular features indicative of better response to

immunotherapy. In our current submission, we have added additional analyses (**New Figure 6i**) to further describe this feature.

Comments

> On line 683, the authors mention that a log2 fold change of -0.01 was used as significant. Is this correct, or a typo?

We thank the reviewer for spotting this error. We have now corrected it to show log2 fold change < -1.

> Would it be beneficial to include either “intergenic” or “whole genome” as an annotation to Figure 2b (similar to Figure 2e)? On a related note, do boxplots suppress the bimodal distribution that we typically see with DNA methylation? Perhaps histograms/density plots of each feature would be more illustrative?

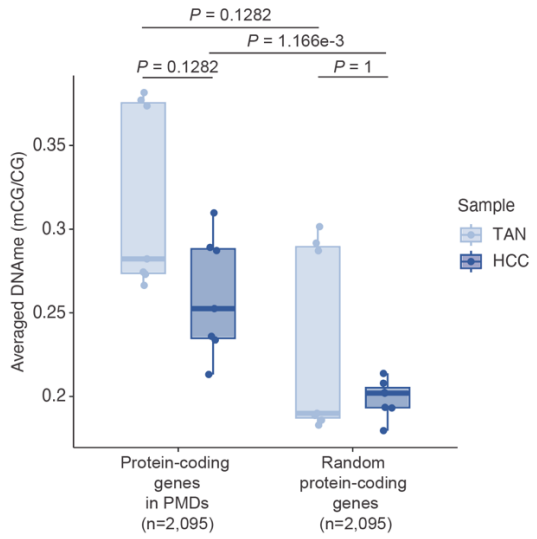
We appreciate the reviewer’s suggestion and have now included “intergenic” annotations in the **new Supplementary Fig. 2a**.

> Seeing as Figure 2d selected for hyper DMRs, the distribution between TAN and HCC is preselected for, making this graph redundant. Perhaps a heatmap or binned scatterplot could highlight the baseline level of methylation in TAN relative to HCC tissue in these specific regions.

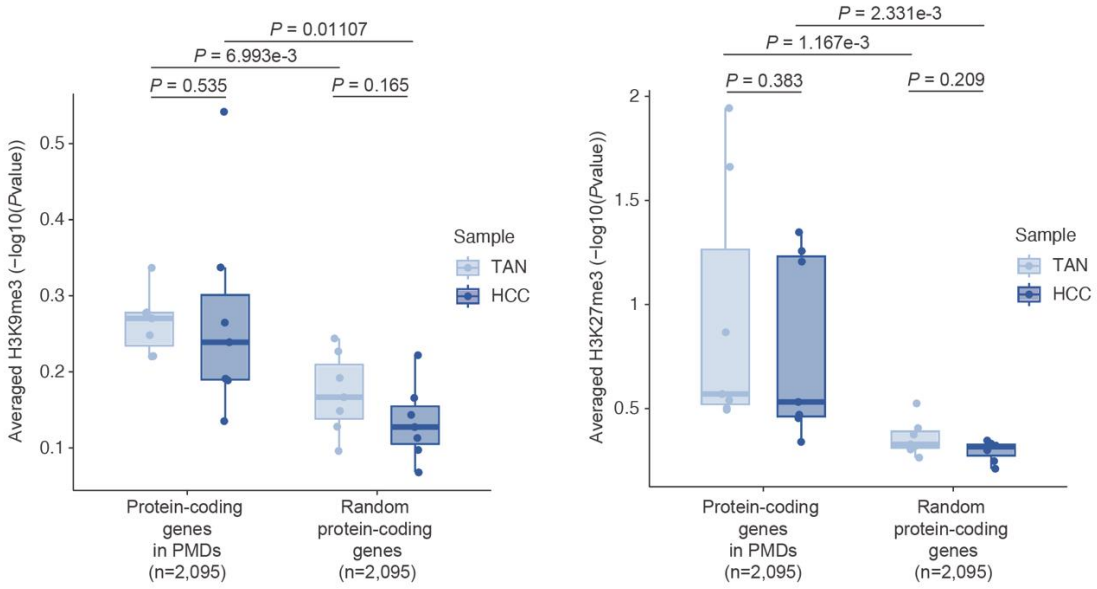
We agree with the reviewer’s comment and have removed this plot.

> In Figure 2f, does the methylation level of the protein-coding genes in PMDs impact the expression (and therefore explain the difference between TAN/HCC)?

We thank the reviewer for this question. We observed a significantly higher mCG/CG ratios at promoter regions of genes within PMDs as compared number-matched, randomly selected genes. However, we detected no significant difference between HCC and TAN (see figure below).



As the reviewer pointed out, the genes within PMDs showed significantly lower expression in HCC samples. Therefore, we also analyzed the levels of H3K9me3 and H3K27me3 at these loci. Similarly, we found no significant difference between HCC and TAN (see figures below). Hence, we conclude that the transcriptional differences are not a result of increased deposition of repressive epigenetic modifications in PMDs. In accordance with Reviewer 1's comments, we have elected not to include these results as they function to confirm previously published features of PMDs.



In Figure 2f/g, are the PMDs linked to the specific samples in which expression is shown? I ask this because of the heterogeneity that the authors observe at the histone level (Figure 1c/d) and wonder if there is also heterogeneity of PMDs in each sample, which may influence the expression pattern.

To address the reviewer's question, we analyzed the variability of PMDs. Indeed, we observed heterogeneity of PMDs between HCC samples (**New Supplementary Fig. 2e**). We found that while there are many PMDs that are shared among all patients, a substantial number are restricted to individual patients. Given that the CREs that show variable activity are not differentially enriched with repressive modifications (**New Supplementary Fig. 4b)**, the heterogeneity of PMDs is likely not related to this finding. While the mechanism that gives rise to the heterogeneity of PMDs is interesting, we feel it falls beyond the scope of this study.

Was any validation of the imputation method performed? E.g. take a sample that worked well, artificially remove a percentage/all of the data, impute, and compare? I say this because of how 865T H3K4me3 is very clearly different from all tumors in Figure 3d. It sounds like maybe this was performed from lines 711-720, but unclear if this performed well (i.e. is there some imputation quality score/metric like R2). More additional detail could be used for imputation as well. Were all samples used, or just those from the same cohort (e.g. impute TAN only from other TAN)?

We appreciate the reviewer's question. We have validated the performance of imputation with the suggested approach. In brief, we processed ChIP-seq datasets and fitted the data to the ENCODE2018-Core model. We enumerated three complementary metrices to compare the mean-square error (MSE) between the observed and imputed signals (baseline MSE, baseline averaged MSE, and global MSE). Overall, for H3K4me3 and H3K27me3, the MSE for the left-out samples were comparable to that of the non-left-out samples, suggesting that the imputed signals are reliable (**New Supplementary Fig. 1b**). For H3K9me3, the MSE for the left-out samples were slightly worse than the non-left-out samples (**New Supplementary Fig. 1b**). This was potentially due to the algorithm not performing as well for broad marks. Taken together, we were confident with the Avocado-imputed ChIP-seq datasets for downstream analyses.

> Is there any empirical evidence to suggest that the inactivated super-enhancer in ADAMTS2 (Figure 4f) loops to the promoter (e.g. GeneHancer)? The authors use HepG2 HiC in Figure 5c, but presumably that interaction is lost due to HepG2 being transformed and the SE lost.

To address the reviewer's question, we looked at the significant interactions defined by Hi-C. We do observe significant interactions between the super-enhancer and the *ADAMTS2* promoter. We have also added the GeneHancer track to the updated screenshot (**New Figure 2f**). It should be noted that while the *ADAMTS2* SE is lost in HepG2 cells, this is not a general phenomenon. In the case mentioned by the reviewer (**Former Figure 5c**), the *GPC3* SE is also defined in HepG2 cells. Therefore, the transformation does not result in loss of all SEs and the Hi-C contacts are concordant with the epigenomic signatures. We have also included the H3K27ac ChIP-seq track in the new screenshot to illustrate this point (**New Figure 3b**).

> Is a non-tumor/tumor difference in GPC3 expression observed in TCGA? Figure 5b shows that LIHC is relatively highly expressed relative to other tissues, but unclear if that significance between non-tumor/tumor is there.

We thank the reviewer for this suggestion. We did detect a significant increase of *GPC3* expression in HCC as compared to TAN in the TCGA-LIHC cohort. We have now included this analysis along with the expression within our own cohort in the **New Supplementary Fig. 5a**.

> On a related note, line 969 expresses that p-values are shown for panel b in Figure 5, but I do not observe any.

We thank the reviewer for pointing out this mistake. We previously did not include a statistical test for this figure. In our current submission, we conducted pairwise comparisons between LIHC and all other cancer types individually. The difference was significant for all comparisons. We have added this information to the figure legend.

> HERVE-int has been shown to be deregulated in HCC through SETDB1/H3K9me3 (PMC11889587). A discussion of this should be done. Similarly, published studies on the epigenetic regulation of GPC3 could be expanded upon.

We thank the reviewer for this comment. As mentioned above, the HERVE-int elements described by Igarashi *et al* (PMC11889587), although belonging to the same subfamily, are not the same element. The genomic location for the HERVE-int described in our study (hg38: chr5:86,570,295-86,570,683) and the two previously described in the aforementioned study (hg38: chr5:86,565,889-86,566,230 and chr5:86,566,604-86,567,311) are different. Nevertheless, we have included discussion of this study and the mentioned GPC3 report in the updated manuscript.