**Responses to Reviewer #1 comments**

We first thank the reviewer for the helpful comments. In the revised manuscript, we have incorporated the reviewer’s comments.

Basically, the authors have addressed most of the comments and criticisms I have raised in my review of the original manuscript. However, there are some issues, which this reviewer likes to mention:

1. Regarding question 6, the addition of Supplementary Table S12a is appreciated. However, sample UCSD.CRC.020 is designated as “benign rectal mucosa”, hence it was not a colorectal carcinoma. UCSD.CRC.029 is a liposarcoma, which does also not correspond to a colorectal carcinoma. UCSD.LC.002 is a metastatic melanoma, thus it is not a lung cancer but a metastasis from a skin tumor. UCSD.LC.015 is “benign fibrous tissue with scar and mild chronic inflammation”. Therefore, some samples appear to be benign, whereas other samples are in fact metastases from other sites. In spite of this, the liposarcoma has clustered together with the colon and the melanoma with the lung (Figure 4); the same applies to the two aforementioned normal samples, which have clustered together with the lung or colon samples. Furthermore, it appears that, according to the values in Table S6, these tissue samples were indeed not identified as “not colorectal cancer” or “not lung cancer”, respectively, by the plasma analyses. Regarding Figure 4d, the authors mention “a small number (N<5) of outliers; do these outliers include the aforementioned samples? These issues should be clarified.

Response: We thank the reviewer for the deep consideration from the reviewer. We noticed the result is consistent that these benign tissues (UCSD.CRC.020: benign rectal mucosa and UCSD.LC.015: chronic inflammation) always have lower MHL compared with average MHL of CRC or LC. In addition, we didn’t do the sample cluster in Figure 4, we only did the feature (MHB) cluster to select MHBs with tissue-specific MHL signal. Finally, the outliner we mentioned in the manuscript are referred to normal plasmas samples, since we notice some normal plasmas have higher average MHL compared with the average value of CRC and LC plasmas. However, what don’t have enough information for these normal plasmas, we cannot make further discussion to the outliers.

2. Does Figure 1a really show a “schematic overview of data generation and analysis”? Regarding the prediction model training, where is the part using 29 WGBS and 14 RRBS data sets? Supplementary Figure 11 is more informative than Figure 1a.

Response: The tissue-of-origin analysis used the prediction model generated by the 43 WGBS and RRBS datasets. In lines 308-310, as well as Figure 5b, we described these datasets. Figure 1a focused on the main parts of this manuscript, while Supplementary Figure 11 was supposed to describe the steps of analyses. We collected these 29 WGBS (8 our own WGBS, 18 Roamap and 3 WBC) and 14 RRBS (10 tissue samples of ENCODE including different biological replicates) in Supplementary Table S12i

3. Furthermore, the upper part of Figure 1b was probably copied from the UCSC genome browser, and the text “Txn Factor ChIP” was misplaced within the layered H3K27Ac curves. Instead of “Txn Factor Chip” “Transcription Factor ChIP-seq” might be more informative.

Response: We thank the reviewer for noticing and pointing out this mistake in the Figure1b. We have removed the misplaced text and changed the wording for Figure 1b.

4. The color scale of Figure 1c is still not explained.

Response: Consider the comments of question 3 of Reviewer #2, Figure 1c related section (Figure 1c, Supplementary Figure 2 and Supplementary Table 1b) was removed from the study to avoid false positive result.

The color scale is relative scaled densities (z value in the following fudgeit function) while the maximum of the value of the scale is the fourth root maximum local density (default of smoothScatter function in R) within the whole Figure and then transfer the density scales to color scales. In order to reduce the complexity, we remove the bar scales and just indicate the density with the color in which red indicate higher density and blue indicates low density.

fudgeit <- function(){

xm <- get('xm', envir = parent.frame(1))

ym <- get('ym', envir = parent.frame(1))

z <- get('dens', envir = parent.frame(1))

colramp <- get('colramp', parent.frame(1))

fields::image.plot(xm,ym,z, col = colramp(256), legend.only = T, add =F)

}

5. In Fig. 1c the “Primary tumor tissue” panel has a black line above the “1.0” at the y-axis and the dotted line does not seem to be at 0.9. It appears that the dotted and the black lines got out of place. In the legend of Figure 1C what does the following mean: “500,000 adjacent CpG loci in MHB regions were randomly sampling and the attenuation of the the r2 with the distance of the CpG loci in different scenario shown different characteristics.” There is no “yellow dot line” (it is rather a red hue)?

Respones: We thank the reviewer for the correction. The black line was removed and the dotted lines was adjusted for the position and we replace the yellow with red in the legend. Meanwhile, we replace our previous statement with “the negative correlation between the r2 and the distance of the CpG loci was observed in different scenario” so that it will be more explicit.

6. In Supp. Fig. 2 the color scale is missing. The red region in the right lower corner indicates that a large number of CpGs are in linkage equilibrium (despite the small distance of 150 bp) and this appears to be very different from the plots shown in Fig. 1C. The authors should comment on this. Figure 3b is not referenced in the text.

Response: To be consistent with Question 5’s change, we are not about to add the color scale since smoothscatter plot usually don’t need color scale to indicate the 2D distribution.

As for the Supp. Fig. 2, reviewer 2 have mentioned some kind of batch effect between different groups. We have added the remind about the risk of the batch effect in the manuscript as “…with the exclusion of the batch effect as much as we can and the result was validated by another independent WGBS data from kidney cancer” in line 92 and 93.

Figure 3b has been referenced in line 191, page 5.

7. On page 6 lines 255 and 259: the wording that normal plasma contains “residual” or “low tumor contribution” is awkward. As the blood has been obtained from individuals without known cancer, no plasma DNA fragments should be derived from a tumor. The authors should rather call it something like “low/residual plasma fragments with a tumor MHB signature”.

Response: We thank the reviewer for the great suggestion. the wording has been replaced in lines 256 and 261

8. Table S6C has probably the wrong heading, as it summarizes lung cancer samples and not colorectal samples. Abbreviations should be spelled out when they are used for the first time. For example, VMR is first used on page 3, line 98, but only spelled out on page 3, line 132 (this also applies to other abbreviations). LAD and LOCK regions should also be spelled out.  
  
Response: We thank the reviewer for the great suggestion. Abbreviations has been checked again and was spelled out before the usage.

**Responses to Reviewer #2 comments**

We first thank the reviewer for the helpful comments. In the revised manuscript, we have incorporated the reviewer’s comments.

The authors have adequately addressed my concerns through their revisions, and I am satisfied with the manuscript proceeding. A few remaining points:   
1. Line 39: "A number of studies...". You should provide some citations at the end of this sentence, not comprehensive but to the best such small studies.

Response: Yes, we thank the reviewer’s suggestion. corresponding references have been added.

2. Line 45: "genome-wide" better than "full-genome"

Response: Yes, we have made the change to “genome-wide”

3. Line 88-92: I continue to be quite concerned that batch effects could cause these differences, despite the additional analyses. Have you taken care that the read length distribution differences between these experiments performed at different sites/studies might not contribute to this? The p-value is kind of meaningless because it's just driven by the large numbers and doesn't provide me any reassurance that this is not batch. My inclination is to suggest just deleting these analyses because I don't think that they add much to the story and are in my view not super believable if the samples were generated/processed/sequenced at different sites, esp. given the subtlety of the differences, e.g. 94.8% vs 91.2%. If you are not going to delete them, at least explicitly remind the reader that there is a risk of batch effects, however careful you were.

Response: We agree with the reviewer. We have added the remind about the risk of the batch effect in the manuscript as “…with the exclusion of the batch effect as much as we can and the result was validated by another independent WGBS data from kidney cancer” in line 92 and 93.

4. Line 94 - The sentence starting "Gene Ontology" is not very clear, e.g. what is a 'cancer loss linkage region'?

Response: We replace our previous statement with ‘Gene Ontology enrichment analysis to MHB regions whose r2 is decayed compared with normal shown significantly associated with number of cancer related pathway and functions’ so that it will be more explicit.

5. I suggest a careful reading and editing to be more circumspect particularly around self-promotion of the claims. This always hurts more than helps from the perspective of the reader and you'd do be more cautious. One example is "very accurate prediction" at line 363 (what is justifiably considered very accurate in this context depends on a lot of factors, better just to say 'accurate'). It may seem like a minor point (and is) but I think it would help if you go through carefully and revise other such instances.

Response: Thank you for your great suggestion and we have gone through the manuscript to identify and correct these minor points.