Reviewer comments:  
  
Reviewer #1  
Basically, the authors have addressed most of the comments and criticisms I have raised in my review of the original manuscript.   
We appreciate the careful reading of our manuscripts by this reviewer and are glad that we were able to address most of his/her previous comments and criticisms.

However, there are some issues, which this reviewer likes to mention:   
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.   
Could this be due to our analyses not being able to distinguish specific types of cancer well?

The figures and their legends still can be improved:   
Does Figure 1a really show a “schematic overview of data generation and analysis”? Regarding the prediction model training, where is the part using 28 WGBS and 14 RRBS data sets? Supplementary Figure 11 is more informative than Figure 1a.   
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. The numbers appear to be incorrect, since I was only able to count 8-9 RRBS ENCODE datasets which are among the 10 tissues, and 28+14 is only 42.

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

The color scale of Figure 1c is still not explained. 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.   
Yes, the reviewer is correct and we have corrected the Figure 1c.

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)?   
We have revised the Figure legends & made sure that all other Figure legends do not have the same mistake.

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.   
The reviewer have probably gotten confused here, because the color scale does not indicate the level of linkage equilibrium but actually the density of points.

Figure 3b is not referenced in the text.   
Yes, this figure should be referenced around line 192.

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”.   
Table S6C has probably the wrong heading, as it summarizes lung cancer samples and not colorectal samples.

Yes, we have made the necessary corrections.  
  
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.  
Yes, we have made the necessary corrections for abbreviations.  
  
Reviewer #2  
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.   
Yes, the references have been added.

2. Line 45: "genome-wide" better than "full-genome"   
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.   
We have checked and the average mapped read lengths were all above 80 and up to 109. So the regions up to 80 bp in distance would still be valid for consideration without the effects of read length. We also believed that while the samples were generated/processed/sequenced at different sites, they should be randomly selected sites. As such, we had to assume that the 500,000 random sampling of sites would be close to the true distribution. This large number should also means that a small difference could be significant & important.

4. Line 94 - The sentence starting "Gene Ontology.." is not very clear, e.g. what is a 'cancer loss linkage region'?   
We have reworded this sentence to be clearer.

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.

Thank you and we have gone through the manuscript to identify and correct these minor points.