Author's response to reviews

Title: Extracellular Matrix Signatures of Human Primary Metastatic Colon Cancers and their Metastases to Liver.

Authors:

Alexandra Naba (anaba@mit.edu)
Karl R Clauser (clauser@broadinstitute.org)
Charles A Whittaker (charliew@MIT.EDU)
Steven A Carr (scarr@broadinstitute.org)
Kenneth K Tanabe (KTANABE@partners.org)
Richard O Hynes (rohynes@mit.edu)

Version: 4
Date: 3 June 2014

Author's response to reviews: see over
June 3rd, 2014

Dear editor,

Thank you for giving us the opportunity to resubmit our manuscript with minor revisions. We have uploaded a revised version of the manuscript, modified to clarify certain points of the manuscripts and to address the reviewers’ comments. We believe we have addressed all their concerns, as described in our point-by-point response (see below).

Of note, the raw mass spectrometry data deposited at MassIVE: ftp://MSV000078555@massive.ucsd.edu will only become available upon acceptance of the paper, in the meantime you can use the following credentials (username: MSV000078555 / dataset password: matrisome) to access it. Feel free to share these credentials with the reviewers.

Thank you again for considering our manuscript. We hope you will find our revisions satisfactory and our article suitable for publication in BMC Cancer.

With best regards,

Alexandra Naba and Richard Hynes
REVIEWER 1:

Major concerns:
As the author mentioned, there are variation among different clinical dataset. There are also variation and only a small overlapping between the proteins author identified and the clinical database. Therefore, it is still questionable whether the proteins (HPX, SPP1 and COMP, etc) authors identified actually play roles in the liver metastasis in vivo. I recommend authors to make some supplemental experiment confirm their conclusion. Authors could knock down the candidate gene by shRNA or ectopic expression gene in the colon cancer cell line, then inject these cell lines into nude mice. In the final, account the liver metastasis node number and make Statistics comparison. Then based on this result, authors could make the related conclusion.

The experiments suggested by the reviewer are, in principle, interesting. However, we do not have a colon cancer transplantation assay up and running in the lab and we know from prior experience with mammary cancer experiments that such experiments would likely take a year or more to complete. Thus, this set of experiments seems beyond the scope of the present paper. The editor also acknowledged that these experiments would be "too demanding and not realistic" for inclusion in the revision of the manuscript.

Minor concerns:
1. Please describe clearly what are the steps 1, 2, 3, 4 in figure1A? In figure legend or method part, there is no description about the procedure.
   The detailed protocol has been described previously and is referenced (#13). We have now added a sentence in the Methods section to identify the extraction steps used to enrich ECM proteins from the samples (see page 4).

2. Align the number 1,2,3,4 with each blot band. Furthermore, in the last blot of Figure 1A, the Actin is not aligned with GAPDH. Please provide new figure with clear indication.
   Thank you. We have now aligned the number with the corresponding lanes of the immunoblots (see Figure 1).

3. It seems that author selects actin and GAPDH as the marker for the extraction from cytoplasm, Pan-histones as the marker for the extraction from nucleus.
   Please explain why there is no any marker in step 4 and ECM-rich Fraction of the normal colon extract. And the level of GAPDH is inconsistent with the level of actin in metastasis extract. There is no actin in the ECM-rich fraction of the normal liver extract, and GAPDH is
missing in step 4 and ECM-rich Fraction. The level of Pan-Histones is inconsistent in each step, which may be caused by inappropriate protein concentration. Therefore, please make the experiment again and provide convincing blot data.

The enrichment of ECM proteins and concomitant depletion of intracellular components is based on the relative solubility of proteins in the different buffers. This differs among different tissues – in some cases the histones and actin are more readily extracted than in others. Normal colon and liver both show very good extraction of actin and histones leaving none in the final ECM-enriched extract – the tumor samples in contrast show more retention of these markers. We have added a sentence to the Methods section to clarify this point (see page 4).

4. In the first paragraph of the result part, the author describe like “….Figure 1A shows the efficiency of the sequential extraction protocol leading to significant enrichment of collagen I …”. Since we know that the molecular weight of Collagen I is 138KD, why the band of Collagen I is shifting so much. Please make sure that you used the correct antibody or your samples are not degraded before running the gel. Other issue is that the word “significant” should be cautious in this sentence, because there is no statistical comparison in the related figure.

We confirm that we used an antibody directed against collagen I for the immunoblots. Collagen I typically runs at around 120K on gels because of its unusual amino acid composition. We would also like to point out that since we are probing native collagen I from in vivo tissues, the collagen undergoes post-translational modifications and cross-linking, which is why one observes smeared and variable patterns on immunoblots.

5. In the first paragraph, authors describe like this “….Concomitant depletion of intracellular proteins (actin, GAPDH, histones) in the final ECM-enriched samples….”. Why there are actin and GAPDH in the sample of colon tumor ECM-rich Fraction.

See discussion above under point 3 and note that enrichment is not meant to imply purity.

In the second paragraph of the result part, authors described as follow “reproducibility for normal colon and colon tumor samples from patient 2 (Additional File 2).”. But in the related figure, “Additional File 3” was labeled. Please correct this error.

Thank you. We corrected the reference to the additional file in the manuscript (see page 6).
6. In the result part, authors described their data as follow“….identified subsets of tumor-specific proteins: 37 proteins were characteristic of the colon tumor matrisome, 7 proteins were characteristic of the metastasis matrisome and 23 proteins were characteristic of both primary tumors and metastases….”. From their mass spectrum comparison, they identified 37 proteins are related with colon tumor formation, 7 proteins are related with metastasis, and 23 protein are related with both of them. After analysis in the clinical dataset, only three proteins (HPX, SPP1 and COMP) are closely associated with metastasis, only four genes (MMP1, MMP2, MMP11, and LEFTY1) are associated with colon cancer formation. Therefore, only a little of protein ID author identified are associated with clinical survival. Most of proteins authors identified have nothing to do with the clinical survival. How do author explain this based on their MS data.

The first statement in the results applies to our proteomics data. After comparisons with gene expression data from clinical samples, not all the proteins we correlate with tumor or metastasis formation show clear associations. That is exactly as expected – mRNA levels and protein levels are well known not always to correlate (as we discuss in the manuscript). Some also drop out simply because they were not included in the array data. Therefore, we report those listed as being correlated with clinical outcome. The other proteins that we report may also correlate but corroborative data are not provided by the mRNA expression data – validation of whether or not they correlate with clinical outcome will have to await more extensive analysis at the protein level – e.g., by IHC (as mentioned in the manuscript).

7. From the Mass spectrum result author identified, a lot of known proteins related with liver metastasis in other studies are not appeared in the authors’results. For example, L1CAM (1), Cadherin-17 (2), periostin (3), TMPRSS4 (4)

References:
Our study focused on components of the extracellular matrix. In this regard, we did detect perioistin (POSTN) but the expression of perioistin was not specific to colon tumors and was found in all four sample types analyzed (see Table 1); we have added a sentence to the text to discuss this and references (see page 9). The three other proteins cited: L1CAM, Cadherin-17 and TMPRSS4 are all transmembrane proteins - it was not our goal to study such components and they were not included in our analyses.

REVIEWER 2:

1. The premise of this manuscript discussed in the introduction involved the translation of proteome signatures to clinic, however this entire manuscript was merely descriptive of the signatures elucidated from three patients. Is there a way to connect the findings with patient prognosis or survival. Key signature patterns identified could be probed using IHC. 

2. Maybe a discussion of these findings could predict metastasis might be useful. The referee is correct that further correlation with prognosis, survival and metastasis will require extensive IHC analyses. We are working on that but, as the referee is probably aware, many of the necessary antibodies are either unavailable or do not work well for IHC. We are currently screening and validating antibodies as well as working on generating new antibodies against the proteins from our signatures. This is a lengthy process and beyond the scope of this manuscript, which is intended to demonstrate that one can discover candidate markers for such an approach using a limited number of patient samples.

3. Related to point one. If you on find the same genes patterns using gene expression arrays, what is the benefit of doing proteomics for diagnosis? A discussion might be useful. After comparisons with gene expression data from clinical samples, not all the proteins we correlate with tumor or metastasis formation show clear associations. That is exactly as expected – mRNA levels and protein levels are well known not always to correlate (as we discuss in the Ms). As indicated above and in the text, some also drop out because they were not included in the array data. Therefore our proteomic data provide a different and arguably more valuable set of markers since proteins are the operative molecules in the tumor microenvironment and changes occur at that level that are not reflected at the mRNA level because of post-transcriptional processes (translation, stability etc.). We have added a sentence to the Conclusions section to address this point (see page 11).
4. Further clarification of the significance of “leading edge” in statistical analysis of Figure 4 C-D might be useful. The abstract needs to be improved with some mention of the signature highlights.

We have now defined the significance of “leading edge” in the GSEA Methods section (see page 5) and we have mentioned the clinical correlations in the abstract.