Author's response to reviews

Title: EMT is the Dominant Program in Human Colon Cancer

Authors:

Andrey Loboda (andrey_loboda@merck.com)
Michael Nebozhyn (Michael_nebozhyn@merck.com)
James Watters (james_watters@merck.com)
Carolyn Buser (carolyn_buserdoepner@merck.com)
Peter Shaw (Peter_Shaw3@merck.com)
Pearl Huang (pearl_huang@merck.com)
Laura Van't Veer (vantveerl@cc.ucsf.edu)
Rob A.E.M. Tollenaar (r.a.e.m.tollenaar@lumc.nl)
David Jackson (jackson@lifebiosystems.com)
Deepak Agrawal (deepak.agrawal@moffitt.org)
Hongyui Dai (hongyue_dai@merck.com)
Timothy Yeatman (timothy.yeatman@moffitt.org)

Version: 4 Date: 9 November 2010

Author's response to reviews: see over
Response to Review of Ms. “EMT is the Dominant Program in Human Colon Cancer”

Version: 2 Date 10 September 2010

Dear Sir:

We are in receipt of the reviewer’s comments for the manuscript entitled “EMT is the Dominant Program in Human Colon Cancer”. We are appreciative of the reviewer’s comments and have made every effort to address each and every comment as per below:

Reviewer # 1:

1. “My major concern is about the data documentation and presentation...the large data sets used in this study are not accessible...” To address this comment, we have uploaded into GEO all of the molecular data derived from the cell lines used to create the EMT signature. In addition, we have submitted the raw data representing the EMT signature genes for all of the samples used in the study. We have also submitted the NKI colon dataset used for validation of our signature. The German dataset is already in the public domain.

2. “Fig 1...the authors further correlated the PC genes to in vitro established “EMT” signature. This is an important step towards the key conclusion in this study, and the data analysis should be shown in the main figures.” To address this comment, we have moved supplementary figures 1 & 2 into Figure 1 in order for the reader to see the full evolution of the EMT signature. We have amended the figure legend to reflect this change.

3. “Fig 2, the authors further assess the correlation of known EMT related genes to the EMT phenotype. However, many EMT driver genes, such as those that regulate CDH1, including Snail, Slug, ZEB 1/2 were not included. This is an important gene set need to be evaluated and discussed. In addition I am not sure how RAS expression can fit in this context, though Ras mutation has been previously shown to be anti-associated with EMT. Without knowing the mutation status of Ras in these tumors, it is difficult to define the Ras dependency in these tumors and its correlation to EMT.” Many EMT driver genes are actually included in Figure 2 such as SNAI2, TWIST1/2, SMAD1/3 but many more are also in the PC1 and EMT scores. These genes are seen in Supplemental Figures 4a,4b, 4c. We agree with the comments about RAS and have removed this analysis from the figures and the text of the manuscript.
4. “Fig 3. MIR 200a/b are anti-correlated with the Mesenchymal phenotype. The authors propose that MiR200 may achieve this via its negative regulation of ZEB 1/2 and CDH1 expression. So does ZEB1/2 expression behave the opposite in the database in relationship to EMT and CDH1 is the same as miR200a/b? These data should be shown in parallel with miR200 to support the claim. In addition, expression profiles of 415 miRNAs in 49 stage 1-IV colon cancers should be provided as the supplementary information to allow readers to assess the data analysis and significance.” We agree with these comments and have now included the EMT driver genes in the table with the MiRs to show the inverse correlations as expected (Supplementary Table 2). In addition, we have included the MiR expressions in supplementary Tables 2,3.

Reviewer #2:

1. “The paper would be enhanced if the authors can show that, for example, TGF-beta inhibitors work on cancer cells showing high PC1 scores, and, on the other hand, Myc inhibitors target cancer cells inhibiting low PC1 scores.” We agree that this would be an attractive experiment, but unfortunately, it is beyond the scope of this manuscript.

2. “Fig 1. Needs labels to clearly specify that samples are actually rows and genes are columns—currently the clustergrams are a bit confusing.” This issue has been addressed by adding labels to the figure explicitly indicating that samples are rows and genes are columns.

3. “Page 3. Definition of “intrinsic” PC1 signature is unclear: We set out to identify the most differentially expressed genes, and…” Need to clarify how the “most differentially expressed genes” were defined-standard deviation/variance across samples, or genes that are significant component of PC1? If the latter, what are the cutoffs to determine “significant components”? Moreover, since PCA is unsupervised, the membership of PC1 signature will hardly be 100% identical from dataset to dataset. It’d be good if the authors would document the difference (which I suspect would be slight but not totally non-existent) in the signature at least between the two datasets in Figure 1.” The second sentence in the Results section has been rewritten to avoid ambiguity. Current version explicitly states that ~5000 genes were selected to be the most correlated (either positively or negatively) to PC1, defined by the p-value<1e-15 and associated with absolute value of Pearson correlation coefficient of at least 0.4.

4. “Page 4-5. “The significant finding was that the unsupervised PC1 signature, which represented an “intrinsic” subtype classifier of colon cancer, appeared to be driven by a core EMT program of up-and down-regulated genes (Supplementary Table 2). In fact 92% of
probes mapped to EMT UP gene set (genes that were up-regulated in mesenchymal vs. epithelial lung cell lines) were positively correlated with PC1 and 82% of probes from EMT DOWN gene set (genes that were respectively down-regulated), corresponding to Fisher exact test p-value of 2 x 10^{-16}.” Again, since the PC1 signature would differ between datasets, one would need Fisher exact test p-value for each dataset. It is also not clear which dataset this p-value is referring to. A minor note: The superscripted 17 seems to be a typo”. We have now explicitly indicated the dataset for which analysis was done and appended a p value to each analysis in the text of the manuscript. We also have removed the typographical error.

5. “Have the authors looked at how PC1 correlates to other EMT signatures? An example is a signature containing genes translationally regulated during the EMT process itself for Jechlinger et al., Ocogene (2003) 22, 7155-7169... If the correlation between PC1 and an independent EMT signature is comparable to the correlation between PC1 and the lung cancer lines signature, then the authors would be able to conclude that the association between PC1 and EMT is general and not tissue specific or worse, signature specific (e.g., will work with lung-origin signature mapped to colon samples but not with breast-origin signature mapped to colon samples)”. We have looked at the noted dataset and find a correlation. This analysis has now been included in the supplementary data (see Supplementary Figures 8,9).

6. “Figure 4B-D. Would an EMT signature predict survival/recurrence just as well as PC1? An advantage of static EMT signature is that there is no need to compute a PC1 for each dataset, and the markers are fixed, whereas a principal component is dependent on the gene expression dataset. The same can be said of proliferation or RAS signatures. It’d good to have a section in Discussion on why PC1 should be used instead of EMT signature alone or in combination with other signatures as predictive marker of outcome in colon cancer cases”. We now show that the EMT signature can predict recurrence for Stages 2,3 as well as all stages. These data are shown in Supplementary Figure 10,11.