Reviewer's report

Title: EMT is the Dominant Program in Human Colon Cancer

Version: 2 Date: 29 September 2010

Reviewer: Chia-Huey Ooi

Reviewer's report:

General comment: This is a very interesting work which attempts to address the issue of molecular subtyping of colon cancer into biologically and clinically relevant subgroups.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

The authors mentioned the possibility of targeted therapy near the end of the paper. 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 exhibiting low PC1 scores.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Figure 1: Need labels to clearly specify that samples are actually rows and genes are columns – currently the clustergrams are a bit confusing.

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

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.

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 17.”
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.

Origin of EMT signature: Lung (signature) <-> colon (samples on which signature was mapped): tissue specificity issues. 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 from Jechlinger et al., Oncogene (2003) 22, 7155–7169. This signature was obtained in the context of mammary epithelial cells (EpH4). In contrast the signature used in this paper was obtained by stratifying lung cancer cell lines into two groups based on static baseline expression of only two EMT markers (CDH1, VIM). If the correlation between PC1 and an independent EMT signature is comparable to the correlation between PC1 and the lung cancer cell 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).

Figure 4B-D: Would an EMT signature predict survival/recurrence just as well as PC1? An advantage of a static EMT signature is that there is no need to compute a PC1 for each dataset, and the markers of 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.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.