Reviewer's report

Title: EMT is the Dominant Program in Human Colon Cancer

Version: 2 Date: 10 September 2010

Reviewer: Qiang Yu

Reviewer's report:

- Major Compulsory Revisions

This study provides an interesting finding showing that EMT gene signature represents a significant gene program in colon cancer development. It also suggests that EMT signature can be used to classify colon cancer and has prognostic values. Given cancer EMT is increasingly associated with disease progression and resistance to chemotherapy, identification of an EMT signature in colon cancer may provide insights in colon cancer progression and useful guidance for chemotherapy. However, my major concern is about the data documentation and presentation. Since this is a genomic study, the large data sets used in this study are not accessible and so is difficult to assess the significance of the data analysis. The supplementary tables simply provide the gene list, without showing the raw data or expression values.

- Major Compulsory Revisions

Fig. 1, The PC1 gene set was identified through non-supervised approach in two colon cancer datasets. The authors further correlated the PC genes to in vitro established “EMT” signature. This is an important analysis step towards the key conclusion in this study, and the data for this analysis should be shown in the main figures. In addition, the supplementary figures can’t be opened up so it is difficult to assess the significance of this result.

Fig2, the authors further assess the correlation of known EMT related genes to EMT phenotype. However, many EMT driver genes, such as those that regulate CDH1, including Snail, Slug, ZEB1/2 were not seem to be 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.

Fig 3, miR200a/b are correlated with epithelial but anti-correlated with mesenchymal phenotypes. The authors propose that miR200 may achieve this via its negative regulation of ZEB1/2 and CDH1 expression. So does ZEB1/2 expression behaves 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.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.