Reviewer’s report

Title: Genome-wide analysis of three way interplay between gene expression, cancer cell invasion, and anti-cancer compound sensitivity

Version: 2 Date: 10 July 2012

Reviewer: David Rodenhiser

Reviewer’s report:

The purpose of this work is to undertake multi-platform analyses between gene expression, cancer cell invasion and the inherent sensitivities to anti-cancer targets in order to identify a unique gene signature that could predict chemotherapy survival for both lung and breast cancer. This is primarily a study involving in silico analyses of pre-existing genomic data sets along with some invasion assays performed on most of the cell lines from the NCI-60 reference set. There are significant problems with this manuscript both conceptually and in the presentation of the data. At times the manuscript is difficult to follow, with figures that are difficult to decipher and with text that is in places poorly written and data that are inadequately explained. My specific comments are below:

Major Compulsory Revisions

1. In the Background, (para 2) the authors state that they intend to characterize genes associated with invasion heterogeneity and drug sensitivity heterogeneity. Heterogeneity is an important issue in cancer biology these days. Yet I am confused by the authors’ approach to addressing heterogeneity in the cell lines they use. As seen in Figure 2, it appears that there is heterogeneity among cell lines within a tissue type. In fact, there are about 12 of the 60 cell lines that each is significantly higher in their invasiveness than the mean for their tissue type. Without labels, it’s unclear which specific lines these are, but they include 2 RE lines, 1BR, 2OV, 2LC, 1ME, 1PR, 2CO and 1CNS line. It would be interesting, if not essential to analyze expression in these specific lines to see whether the relevance of the 8 gene signature is confirmed. To me this provides a ‘truer’ signature related to the common phenotype of invasiveness in these particular cell lines. Furthermore, regarding Figure 2b, I am unclear as to the relevance of this histogram to the manuscript. What’s your point in including a histogram of this type, and what exactly do you mean by ‘within group residual’; does it refer to tissue of origin? Also regarding this figure and the manuscript in general, since you ultimately focus this paper on lung and breast cancer cohorts, analyses specifically focused on these cell lines would be appropriate.

2. Regarding Figure 3A, the authors need to show on the left axis the location of the dividing line between the probe panels for invasive ability (positive vs
negative) of the NC160 lines. Also, there appear to be common responses in invasive abilities among the cell lines. For instances, the first 14 cell lines appear to share common expression patterns across the genes probes evaluated. Why is this so?; Is it related to the specific cell lines (i.e RE7860, RBT549, etc)? Furthermore, are these specific cell lines the same cell lines that show high invasive characteristics in Figure 2A?

3. Regarding Figure 3b: this Figure is superficial, poorly described, confusing and its purpose is unclear. A better description of the data presented is definitely required. Furthermore, the specific patterns of the eight signature genes should be enlarged and added as a Figure 3C so we can specifically see the ‘complementary’ responses that are described in the text.

4. Figure 4 is a generic, commercial pathway diagram that was copied (and uncredited) from the Metacore webpage. It contains too much irrelevant information. Furthermore, the relevant red labels are difficult to identify and locate. This figure must be redrawn and simplified to support the manuscript. Are the eight signature genes present on this Figure?

5. The probe selection criteria discussed (in Results: Selection of invasion and drug sensitivity associated genes) are poorly described. This section must be re-written.

6. In the Results (Validating gene signature with independent cell lines) you refer to previous work regarding 29 lung cell lines exposed to paclitaxel and docetaxel. Did these include any of the cell lines you used on the NCI-60 panel? A similar question regards the previous study looking at breast cell lines. If so, did you look specifically at your 8-gene signature in relation to these specific cell lines?

7. In the Results (Clinical Outcome prediction; last para), your statement that ‘patients predicted to have higher metastasis potential and to be more drug-resistant had a significantly increased risk for poor survival’ is not surprising and hardly seemed provocative as an adequate interpretation of the results.

8. Overall, the Discussion does not provide a convincing description of the methods applied or the relevance of the data generated. You state in the Discussion (para 2) that a previous study from your group (reference [40]) generated a 4-gene signature for clinical outcome prediction. How do the data generated by the present study correlate with your previous study? In the Discussion (para 3), you infer that combined use of anti-MT drugs with dasatinib / erlotinib may increase efficacy. What literature supports this conjecture? Has this been tested in clinical trials? Furthermore, much of the Discussion reads as a literature review without sufficient integration of the impact of the authors’ new data as it relates to the field.

Minor Essential Revisions
1. In the Results (Gene-Drug heat maps for compound selection) the MYB and
TOB1 genes are mentioned, as are others, yet these cannot be identified or located in Figure 3. Also in this paragraph on page 14, please define or clarify what you mean by saying that the ‘efficacy of the two groups of compounds are complementary’ … perhaps you mean inversely correlated?

**Quality of written English:** Not suitable for publication unless extensively edited

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

**Declaration of competing interests:**

I declare that I have no competing interests.