Reviewer's report

Title: Prognostic Values of TGF-Beta Isoforms and Receptors mRNA levels in Breast Cancer

Version: 1 Date: 6 February 2014

Reviewer: Rosemary Akhurst

Reviewer's report:

1. Is the question posed by the authors well defined? The authors need to formulate a question in the background. As it stands there is not hypothesis or question,

2. Are the methods appropriate and well described? Concerns about multiple testing and slight cherry picking in Fig 3. The need to cherry pick in the first place is not clear, as the outcome seems fairly robust regardless of thresholds for high and low expression. Maybe the authors could comment on this,

3. Are the data sound? RT-PCR validation data is not shown.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition? N/A

5. Are the discussion and conclusions well balanced and adequately supported by the data? No

6. Are limitations of the work clearly stated? They could be more clearly stated, and direct comparison to general concepts and other manuscripts comparing TGF beta levels in breast cancer should be discussed.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

8. Do the title and abstract accurately convey what has been found? Not certain until reviewers address questions about prognosticative value of TGFB1 TGFB2 and TGFB3.

9. Is the writing acceptable? yes

This manuscript would be of interest to individuals interested in targeting TGF betta for breast cancer, since it is a survey of RNA expression levels of the three ligands and two receptors in published data sets from TCGH, and correlation with nodal status and with overall survival. There is some validation of data by qRT-PCR using additional BCa tumors from an Australian cohort (n=71).

The overall conclusion is that is high TGFBR1 and low TGFB2 mRNA levels in tumours were associated with a worse prognosis for patients, which appears like a sound and useful observation. The authors also state that “the prognostic value of lymph node metastasis status can be significantly improved by assessing the TGFB1, TGFB2 and TGFB3 mRNA levels of the primary tumours”. However they do not explain the magnitude of this improvement, nor what they are using as
their base level for prognostication e.g. PAM50? This needs more clarification and a quantitative approach to the added value of including the assessment of TGFB1, TGFB2 and TGFB3 mRNA levels of the primary tumours.

The authors do not adequately address the limitations on the study of RNA only. They do not sufficiently survey the literature of the prognostic value of markers of TGF beta signaling in cancer e.g. that by Reiss group Cancer Res 2002. Alterations of Smad signaling in human breast carcinoma are associated with poor outcome: a tissue microarray study. The authors need to undertake a thorough literature search, and place their data in the light of others, including de Kruijf et al 2013.

The RT-PCR validation data should be shown graphically. Rather than presenting Table 3, it would be more informative have a histogram of distribution of expression levels (med low high ) for each gene within the primary tumor of each BCa class. This could simple be:

y axis = number of tumors and x axis = tumor class and gene. Each bar could include number of tumors categorized as low, med and or high for that gene. This gives a visual of how many tumors examined in each class and the relative distribution of expression levels within and between tumor classes. Alternatively a box and whisker plot of log transformed RNA levels for each tumor class. Indicate number of tumors within each class.

Figure 1 B. This should be omitted for clarity, and include only those correlations r2 > 0.5. Moreover, the p values need to be corrected for multiple comparisons.

Figure 2. What do the P values calculated by Kruskal-Wallis test refer to: i.e. what is the null hypothesis?

Figure 3A is very difficult to comprehend, as multiple manipulations of the data are shown in the same graph, which also raises the issue of multiple comparisons. What is there to gain by using any value other than the 40th percentile for high versus low expression (i.e. the right hand data point in each graph)? The most significant values are found predominantly in this larger group, since smaller sample sizes (on the left) lower significance.

Why have the authors selected to show Kaplan Meier plots in B-F using different percentile cut-offs for inclusion of low and high expression? This seems like cherry-picking?

It would be valuable to visualize a Kaplan Meier for TGFB2 by nodal status.

Regarding the conclusion that higher levels if TGFB1 TGFB3 are associated with poorer outcome, the authors should discuss their findings in the light of the consensus view that expression of TGF beta ligands, particularly TGFB1, are generally thought to increase with tumor progression and be associated with higher risk for progression. Since the data in the manuscript appear to go against this consensus view, the authors should discuss this in more depth. Or are the studies on ligands, apart from TGFB2, inconclusive due to multiple testing issues? The authors might also like to speculate on the therapeutic significance of high TGFB R1 levels correlating with worse prognosis.
Validation by some protein data (IHC?) would be helpful, although not essential because, as the authors point out, there is considerably post-transcriptional control of TGF beta ligands.

Discussion: the authors state that “Our results suggest that the prognostic value of lymph node metastasis status can be significantly improved by assessing the TGFB1, TGFB2 and TGFB3 mRNA expression levels of the primary tumours, i.e. high TGFB1 and TGFB3 mRNA levels provide better prognosis for patients diagnosed with regional lymph node metastases, and high TGFB2 mRNA levels provide better prognosis for patients with lymph node negative diseases”. Significantly improved compared to what? And how more accurate would this predictor be compared to standard predictors, and Pam50?

All requests are mandatory. The suggestion of undertaking some IHC is optional.

**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.

**Declaration of competing interests:**

I am on the SAB or Isarna Therapeutics. However, I do not feel that this affects my judgment of the data