Reviewer’s report

Title: TGF-beta isoforms and receptors mRNA expression in breast tumours: prognostic values and clinical implications

Version: 2
Date: 22 December 2014

Reviewer: Guy Brock

Reviewer’s report:

The authors perform a meta-analysis of several data sets (TCGA, 5 GEO data sets, and their own data) to investigate the prognostic value of TGF# expression in breast cancer tumors. I believe the authors’ results shed some important light on the relationship between TGF# isoforms and breast cancer outcomes. However, there are important methodological issues that need to be addressed by the authors.

Major Comments

1. The statistical analysis of the newly included GEO data sets is questionable. First, the pre-processing method used of log2-transforming and then standardizing to mean 0 and SD 1 for each data set is inappropriate since the study populations differ (e.g., one data set is predominantly HER2 (-) patients and two others consisted of only lymph node (-) patients) and normalizing in this fashion would eliminate any differences in gene expression associated with these different study populations. The authors need to use more appropriate methods for meta-analysis of gene expression studies. Some references and software (R packages) are detailed at the following url:

   http://www.pitt.edu/~tsengweb/MetaOmicsMethods.htm, and some recent examples of application papers include De Cecco et al. “Comprehensive gene expression meta-analysis of head and neck squamous cell carcinoma microarray data defines a robust survival predictor” (Ann Oncol (2014) 25 (8): 1628-1635) and Buehler et al. “Meta-analysis of microarray data identifies GAS6 expression as an independent predictor of poor survival in ovarian cancer” (Biomed Res Int. 2013;2013:238284). It seems several strategies are possible for meta-analysis – i.e. the authors could combine the data (see http://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html and the discussion thread here

   https://www.researchgate.net/post/How_can_you_combine_different_published_expression_data_sets_for_some_practical_and_easily_implemented_R_code) and then analyze the composite data, or the authors could fit separate Cox models to each of the data sets and use appropriate meta-analysis methods to combine the results. Regardless, an appropriate and justifiable approach needs to be used.

2. The Cox regression analysis of looking at different periods of follow-up time from 1 to 10 years is a little unorthodox and the results are difficult to interpret in some cases. The authors are suggested to instead check whether each gene
expression variable is time-dependent (e.g., by using the cox.zph function in the survival package in R). If the gene expression variable violates the proportional hazards assumption then the authors can create a time-dependent version of the variable (e.g., by creating a cut-point at a given time) and then re-fit the model (which does require re-formatting the data to the counting-process style, see http://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf). The process can be repeated by re-evaluating the PH assumption and checking whether another cut-point is needed (i.e., more time intervals). This is superior to the current approach in that 1) the authors can formally test whether the HR does change over time and 2) the authors can check where a change actually takes place.

3. The authors analysis of looking at multiple cut-points and multiple years of follow-up (Figure 5) is overly-complex (and the multiple cut-points for the KM curves is not specified in the ‘Statistical Analysis’ section). Further, I'm not sure about the clinical utility of a result based on comparing the ‘highest 35%’ to the ‘lowest 35%’, since therapeutic decision making would need to be applied to all patients (e.g., what about the middle 30% in the above case?). Therefore it is preferable to use cut-points applied to all patients (>35% vs. # 35%). Formal testing procedures for determining a cut-point can be found in e.g. Section 8.6 ‘Discretizing a Continuous Covariate’ of Klein and Moeschberger’s ‘Survival Analysis’ test (2nd edition). This basically amounts to selecting the cut-point with the largest log-rank statistic and using an appropriate method to determine the p-value. Or, the authors can simply use the median expression value (or the median + quartiles if more divisions are needed). The number of follow-up years evaluated can be determined from the time-dependent Cox regression analysis.

Minor Comments

1. In the Discussion it is stated that “Thus we may assume that there are generally three types of tumour cells in terms of the responsiveness to TGF#: those that respond to TGF# and are suppressed (X1), those that do not respond to TGF# (X2) and those that respond to TGF# but progress (X3).” Can the authors shed some light on predicting which patients would fall into each of the three categories?

2. Pg. 10 (top): The sentence “The consistent results from the independent datasets mutually validated themselves” needs more elaboration – how exactly did the results from the independent datasets (do the authors mean the GEO data sets here?) validate the findings from the other data sets?

3. The authors should include some time-dependent ROC curves or C-indexes to evaluate the prognostic ability of the TGF-# isoforms.

4. Pg. 12: The sentence “Higher TGFB1 mRNA expresssion levels were observed in the tumours compared with the dajecant normal tissues and in the tumours from lymph node-positive patients compared with the the tumours from lymph node positive patients.” Do the authors “lymph node negative patients” at the end of the sentence? Also note several misspellings in the sentence.
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests