Reviewer's report

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

Version: 3
Date: 24 June 2015

Reviewer: Guy Brock

Reviewer's report:

The authors have presented a nice revision of the manuscript. The pros include the completeness of presentation, the nice graphical summaries, and the connection of their results to previous findings. The cons include a questionable manner for pre-processing the gene expression values and the potential for over-interpretation due to the myriad of cut-points and time-intervals evaluated (details below). I believe this manuscript could be suitable for publication; however the authors have to address the issues outlined below.

Major Comments

1. I am still not convinced that the pre-processing of the GEO gene expression values is the correct way to go. Specifically, a z-score calculated for one patient sample reflects the deviation from the mean of that sample, and if the patient populations of the different data sets differ then the z-scores will in fact be measuring different things. If all the patient populations are similar then this method will work for expression measurements obtained using different arrays (e.g. Affymetrix vs. two-color, etc.). Did the authors verify whether TGF-beta expression was related to any of the clinical / tumor pathology / molecular characteristics that differed between the patient populations (e.g., the Hatzis cohort consisted primarily of HER2-neg tumors)? If so, then the authors need to take this into consideration when combining data from different sources. Also it seems the z-scores were calculated separately for each individual data set, but not e.g. by ER pos / neg and T # 2 cm / > 2 cm strata. If z-scores were done separately according to these patient sub-groups it might be easier to interpret because then they are calculated with respect to a homogenous (with respect to the clinical / pathology factor at hand) population. However, I still think the authors need to take into consideration the suggestions I made in my previous comments. At the very least, they need to provide strong justification for why the method they chose is appropriate, preferably backed by prior literature.

2. The authors present Kaplan-Meier curves associated with the TGF-beta isoforms in Figure 6 corresponding to the cut-point with the greatest separation (statistical significance). However the p-values here are misleading because they do not account for the number of cut-points evaluated. This can be done using e.g. the method outlined in Section 8.6 ‘Discretizing a Continuous Covariate’ of Klein and Moeschberger’s ‘Survival Analysis’ test (2nd edition), or simply using a Bonferroni or FDR correction (though FDR correction should ideally account for
Minor Comments

1. The authors should explain what the heaviside function is since the typical reader will not be familiar with it. Also why did the authors select 3 years as the cut-point for the heaviside function? Was this the point at which the proportional hazards assumption was violated? Also the only place where the authors discuss the PH assumption is in the Figure 5 caption where they state ‘The proportional hazards assumption was met for each model.’ Does this mean the PH assumption was met over the entire time period or it was met for each of the separate time intervals <3 years and #3 years?

2. In the discussion of the three types of tumors (X1, X2, and X3, in terms of responsiveness to TGF-beta) the authors state ‘positive’ and ‘negative’ HRs, which should be positive and negative log HRs. Also the ‘0 HRs for X2 tumours’ would be better stated as ‘log HRs of zero’. Last, I find the ‘X1’, ‘X2’, etc. notation a little confusing as this is usually reserved for mathematical / statistical models. Suggest changing it to ‘type 1’, ‘type 2’, etc. unless this will cause ambiguity elsewhere.

Level of interest: An article of importance in its field

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