Author's response to reviews

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

Authors:

Chenfeng Chen (chenfeng.chen@uqconnect.edu.au)
Kong-Nan Zhao (k.zhao@uq.edu.au)
Paul Masci (p.masci@uq.edu.au)
Sunil R Lakhani (s.lakhani@uq.edu.au)
Annika Antonsson (Annika.Antonsson@qimrberghofer.edu.au)
Peter T Simpson (p.simpson@uq.edu.au)
Luis Vitetta (l.vitetta@uq.edu.au)

Version: 4
Date: 2 August 2015

Author's response to reviews: see over
Response to Dr Guy Brock’s comments

We sincerely thank Dr Guy Brock for his thoughtful comments and valuable suggestions. Our point-by-point response to the comments has been provided below. Each response starts with a numbered RESPONSE and is underlined. In each response, the sentence copied from the manuscript is always placed within the bold quotation marks.

Sincerely,

Chenfeng Chen (on behalf of all authors)

The University of Queensland, School of Medicine, Brisbane, Australia

Email: chenfeng.chen@uqconnect.edu.au
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.

RESPONSE 1: In the last version of the manuscript, the Hatzis and Ivshina cohorts were merged and named as “GEO cohort, All”, which was used to represent unselected population. The Ivshina cohort was of unselected population; the Hatzis cohort was mainly composed of HER2-negative tumour specimens (95.5%). To avid biased result, The Hatzis cohort has now been removed from the analysis. Only the Ivshina cohort was used to represent unselected population. The z-scores were calculated separately for different subsets of the Ivshina cohort (stratified by tumour size, presence or absence of regional lymph nodes and ER status). Please see below for details.
The Ivshina cohort (GSE4922) was of unselected population [16]. The Schmidt (GSE11121) and Wang (GSE2034) cohorts were both composed of the tumour specimens from lymph node-negative patients [17, 18]. The Symmans cohort (GSE17705) was composed of ER-positive tumour specimens [19].

The gene expression data of the four GEO datasets were all based on Affymetrix U133A and U133A&B array sets. The data of the probe sets 203084_at (TGFB1), 220406_at (TGFB2), 209747_at (TGFB3), 206943_at (TGFBR1) and 208944_at (TGFBR2) from each GEO dataset and each subset of the Ivshina dataset (stratified by tumour size, presence or absence of regional lymph nodes and ER status) were log2-transformed and standardized to mean = 0 and standard deviation (SD) = 1. The Ivshina cohort was renamed as “GEO cohort, All”, which was used to represent unselected population. The log2-transformed and standardized gene expression data of the Schmidt cohort, the Wang cohort and the lymph node-negative subset of the Ivshina cohort were merged to form “GEO cohort, N neg.”; and the log2-transformed and standardized gene expression data of the Symmans cohort and the ER-positive subset of the Ivshina cohort were merged to form “GEO cohort, ER pos.”. The T ≤ 2 cm, T > 2 cm, lymph node-positive and ER-negative subsets of the Ivshina cohort were named as “GEO cohort, T ≤ 2 cm”, “GEO cohort, T > 2 cm”, “GEO cohort, N pos.” and “GEO cohort, ER neg.”, respectively.

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 the dependent hypotheses).

**RESPONSE 2:** “Bonferroni correction for multiple testing was applied where appropriate and corrected p-values were given the symbol P’. For comparison, Storey’s false discovery rate (FDR) procedure, a less stringent multiple testing correction method, was also applied where appropriate [23].” In Figure 6, E and F survived Bonferroni correction, while C and D did not. However, C and D survived FDR correction. In the revised manuscript, Bonferroni corrected p-values have been provided in the Figures or in the context where appropriate.

**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?

**RESPONSE 3:** A reference [22] that demonstrates extension of the Cox proportional hazards model for time-dependent variables using a Heaviside function approach has been provided in the Methods. The decision of using 3 years as the cut point for the Heaviside function is based on graphical assessment of Figure 4. In Figure 4, “The border of shape was coloured red if the corresponding proportional hazards assumption was not met (P < 0.1).” In Figure 4, the red bordered shapes were seemed to appear either in the period of < 3 years or in the period of ≥ 3 years, though this involves subjective judgement. After using Heaviside function approach, “The proportional hazards assumption was met (P ≥ 0.1) 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.

**RESPONSE 4:** In the discussion, “positive HRs”, “negative HRs” and “0 HRs” have been corrected to “positive log HRs”, “negative log HRs” and “log HRs of zero”, respectively. The notations “X1”, “X2” and “X3” have been changed to “type 1”, “type 2” and “type 3”, respectively.