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)
Anniika Antonsson (Anniika.Antonsson@qimrberghofer.edu.au)
Peter T Simpson (p.simpson@uq.edu.au)
Luis Vitetta (l.vitetta@uq.edu.au)

Version: 3
Date: 5 May 2015

Author's response to reviews: see over
Response to the reviewers’ comments

We sincerely thank the reviewers for the thorough review and valuable comments, which allowed us to significantly improve the quality of this manuscript.

The spelling and grammatical errors pointed out by Dr Rosemary Akhurst have been corrected. Our point-by-point response to Dr Guy Brock’s 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 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
Response to Dr Guy Brock’s Comments

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

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 http://www.researchgate.net/post/How_can_you_combine_different_published_expression_datasets_and_analyze_them_in_R 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.
RESPONSE 1: When combining the GEO data sets, the differences in the clinical characteristics of the cohorts were noted. In the Methods, we have specified the details as how the datasets were combined, which should be appropriate to prepare the datasets for further analyses.

“The Hatzis cohort (GSE25066) was mainly composed of HER2-negative tumour specimens [16]. The Ivshina cohort (GSE4922) was of unselected population [17]. The Schmidt (GSE11121) and Wang (GSE2034) cohorts were both composed of the tumour specimens from lymph node-negative patients [18, 19]. The Symmans cohort (GSE17705) was composed of ER-positive tumour specimens [20].”

“The gene expression data of the 5 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 were log2-transformed and standardized to mean = 0 and SD = 1. The “GEO cohort, All” was compiled by merging the log2-transformed and standardized gene expression data and the clinical data from the Hatzis and Ivshina cohorts. The Schmidt and Wang cohorts were added to the lymph node-negative subset of the “GEO cohort, All” and formed “GEO cohort, N neg.”; and the Symmans cohort was added to the ER-positive subset of the “GEO cohort, All” and formed “GEO cohort, ER pos.”. The $T \leq 2$ cm, $T > 2$ cm, lymph node-positive and ER-negative subsets of the “GEO cohort, All” formed “GEO cohort, $T \leq 2$ cm” “GEO cohort, $T > 2$ cm”, “GEO cohort, N pos.” and “GEO cohort, ER neg.”, respectively.”

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 should a change actually takes place.

**RESPONSE 2:** We have now evaluated the PH assumption using `cox.zph` function for each Cox model in Figure 4. In the Figure, “The shapes were coloured red if the corresponding proportional hazards assumption was not met (P < 0.1).” Inserts were used in the Figure 4 to better display the overlays. As suggested, we also added Figure 5, which used time-dependent Cox regression analysis using a Heaviside function approach to evaluate the associations between the gene expression variables and patients’ survival.

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.

**RESPONSE 3:** The Figure (now Figure 6) has been modified as suggested to show the results that were based on comparing the “high x%” to the “low 1-x%”, i.e. using only one cut point instead of two. The description for the KM curves has been added in the Statistical Analysis section. The representative KM curves are plotted for the cut points that produce the lowest Wald test p values for HR.
Minor Comments

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 TGFβ: those that respond to TGFβ and are suppressed (X1), those that do not respond to TGFβ (X2) and those that respond to TGFβ but progress (X3).” Can the authors shed some light on predicting which patients would fall into each of the three categories?

RESPONSE 4: We have discussed this in Discussion. The proposed classification is based on the knowledge that: “Transforming growth factor beta (TGFβ) signalling is involved in the maintenance of tissue homeostasis and suppression of premalignant tumour cells, however, when the regulations are circumvented, TGFβ signalling can be advantageously exploited by the tumour cells to promote tumour progression and metastasis [1].”

“Assuming that we can group breast cancer patients according to the tumour cells’ responses to TGFβ, and evaluate the associations between the expression levels of TGFB1, TGFB2, TGFB3, TGFBR1 and TGFBR2 and patients’ clinical outcomes such as overall and relapse-free survival in each group using Cox proportional hazards regression model, then we should expect negative hazard ratios for X1 tumours, 0 HRs for X2 tumours and positive HRs for X3 tumours, respectively. Obviously, the HRs for different genes should be different. Unfortunately, we are still not able to separate the three groups of tumours. However, for a cohort composed of patients with mixed groups of tumours, a significant negative HR implicates that there are more patients with X1 tumours, a significant positive HR implicates that there are more patients with X3 tumours, and a non-significant HR implicates that there are more patients with X2 tumours, the patients are heterogeneous or the population is too small. Thus the association between the expression levels and patients’ clinical outcomes may serve as an indicator of the proportion of the three types of tumours.”
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?

**RESPONSE 5:** The sentence has been removed.

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

**RESPONSE 6:** We will take the suggestion for future works, however, we are not able to produce time-dependent ROC curves or C-indexes at this stage.

Pg. 12: The sentence “Higher TGFB1 mRNA expression levels were observed in the tumours compared with the adjacent normal tissues and in the tumours from lymph node-positive patients compared with 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.

**RESPONSE 7:** The sentence has been corrected to “Higher TGFB1 mRNA expression levels were observed in tumours compared with adjacent normal tissues and in tumours from lymph node-positive patients compared with tumours from lymph node-negative patients.”