Author's response to reviews

Title: Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status

Authors:

Xi Zhao Ph.D. (xi.cameron@me.com)
Einar Rødland Ph.D. (einarro@ifi.uio.no)
Therese Sørlie Ph.D. (Therese.Sorlie@rr-research.no)
Hans Kristian M. Vollan M.D. (hans.kristian.vollan@me.com)
Hege G. Russnes M.D., Ph.D. (hegrus@rr-research.no)
Vessela N. Kristensen Ph.D. (v.n.kristensen@medisin.uio.no)
Ole Christian Lingjærde Ph.D. (ole@ifi.uio.no)
Anne-Lise Børresen-Dale Ph.D. (a.l.borresen-dale@medisin.uio.no)

Version: 4 Date: 23 June 2013

Author's response to reviews: see over
Palo Alto, June 20, 2013

Dear editors of the BMC Cancer:

We hereby submit the revised manuscript “Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status”. In this letter, please find our point-by-point response to referees' comments and explanations of how the manuscript has been changed.

We strongly believe that our study fits well with the scope of BMC Cancer and is well suited to the interest of your audiences.

Thanks for your consideration.

Kindest regards,

Xi Zhao (Corresponding author)
Response to comments on “Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status” (BMC Cancer: MS 8553457659452288)

Below, we list in details our responses to each of the reviewers’ comments (cited by the quotes) and the changes we have made to the manuscript.

Respond to the comments/suggested revisions from reviewer #1:

“General comment

Zhao et al. present a meta-analysis of prognostic signatures for breast cancer. The study is based on two previously published series of gene expression data, each of them including ~1000 patients. The authors report on a strong influence of hormone receptor status and of follow-up time on the performance of the gene signatures.

While the analysis of follow-up times is new, performance dependence of the signatures on the molecular subtype of breast cancer is well-known for a long time (e.g. Desmendt et al., CCR 2008). There is a technical difficulty related to a low number of patients and events in the group of long follow-up times.

Major concerns

The major problem with this paper lies in the fact that the important new prognostic signatures for breast cancer, i.e. Endopredict (Filipits et al. Cancer Res. 2011, Dubsky et al. Ann Oncol. 2012), has not been included in the study. At the San Antonio meeting 2012 it was reported that the EP in particular keeps its predictive power over a period of ten years. This is of great relevance for the quality of an assays. The authors should include the EP to reflect the current situation. This would improve the paper dramatically.”

EndoPredict (EP) signature has been added to our analysis, presented in the main text (Fig.1-3, Fig.S2, Fig.S3, Fig.S5, Fig.S6, Table 1-3 and Table S1, Table S3-S6) as well as additional analysis in the Supplement. We found that the EP signature's predictive power also drops over time, and that it is quite similar to PAM50 and GGI in most respects. The authors for the EP signature [1, 2] analyzed the 10-year relapse/survival, but did not actually address the predictive power of the signature during particular time periods, such as after 5 years of follow-up.

“Comparison of Intrinsic and PAM50 in terms of concordance (in %) would be more transparent. Further, these classifications should be compared with the results of immunohistochemistry (and in-situ hybridization) for ER, PgR and HER2.”

Figure 1SA gives the concordance between Intrinsic and PAM50; the overlap percentages are added in the text and the figure captions. The indications are consistent with those of previous publications in various datasets. Consistently, the most striking trend is that these two signatures agree to each other on “basal-like” subtype; while around half of the LumAs identified by intrinsic signature is classified
as “LumB” in PAM50. However, our analysis aims to take the comparison further, where we use centroid correlation to quantify the consistence on the classification of a particular subtype (Figure 1S B) and use a novel statistic “D score” to quantify the distance for a misclassified sample by PAM50 with respect to intrinsic signature. The analysis presented in Figure 1S provided a more thorough comparison of the subtype signatures and more importantly in a quantitative manner.

We also added the comparisons between subtype assignments with receptor marker IHC status in the supplement (Supplement Figure 2 and 3) and discussed in the main text. For Intrinsic signature, there were 62% IHC ER-positive samples were assigned to Luminal subtypes; 79% IHC triple-negative samples were assigned to Basal-like subtype and 55% IHC HER2-positive samples were assigned to HER2-enriched subtype (Supplement Figure 2). For PAM50, there were 64% IHC ER-positive samples were assigned to Luminal subtypes; 79% IHC triple-negative samples were assigned to Basal-like subtype and 50% IHC HER2-positive samples were assigned to HER2-enriched subtype (Supplement Figure 3).

“The observation that the performance of gene signatures depends on hormone receptor status is not new. The related literature (e.g. Desmedt et al., CCR 2008) needs to be evaluated, cited and discussed.”

Thanks for pointing this out. We now have included discussion in context of Desmedt et al., CCR 2008 (See the main text under Discussion):
The indications from our study that prognostic power of gene signatures depend on ER-status, has previously been reported by Desmedt et al. [3]. They used a gene module score to estimate HER2 and ER activity, and used this to split the samples by HER2 status, and the HER2-negative were further split by ER status, resulting in three groups. We used ER and HER2 status based on IHC where available, or imputed from gene expression if not. Since we did not see a substantial effect of HER2 status on DMFS or time dependency (Figure S5), we did not focus on stratification based on the HER2 status.

“As this is another important molecular marker for breast cancer, the dependence of performance on HER2 status should be investigated in parallel to the dependence on ER status.”

We carried out our analysis by stratifying HER2 status. However, we did not observe time dependency on the prognostic power of gene signature. Furthermore, the differences in term of the gene signature prognosis between HER2+ and HER2-subgroup were not profound. We briefly described the analysis on HER2 and the observations in the main text.

“Interpretation of follow-up times appears difficult: These can be interpreted biologically in presence of events (late relapse vs. early relapse). However, in absence of an events, follow-up times should be considered as external, observational factors.”

We are not sure we understand the question correctly, and apologize if we are misaddressing it.
It is not the follow-up times in themselves that are analyzed and interpreted, but the relapse frequencies at different time intervals after treatment. The follow-up times determine which patients are included at any time-point of the Cox analyses. As such, when we analyze different follow-up time intervals, we are concerned with the frequencies of relapse events amongst patients still being under study (ie not previously relapsed and still followed up).

The main difficulty of the approach is that in the latter follow-up time periods, ie 5-10 years and >10 years, patients with previous relapse have been removed from the study material, so the latter time periods consist of a selected set of patients: ie the >10 years analyses consider risks of relapse amongst those who survived 10 years without relapse.

If the risk signatures predict early versus late relapse, rather than just predict relapse versus no relapse, this would appear as higher risk in the high-risk group during the early time periods as the early relapses occur, while in the latter time periods there could be higher risk in the low-risk group than in the high-risk group as the late relapses occur. We are suggesting that this may in part explain our results.

“The number of patients at risk should be added to Table 3. The impact of sample sizes and the number of events on the significance of results should be discussed in detail.”

We now added the number of patients at risk in Table 3. We divided the follow-up time into three intervals: first 5 years, 5-10 years, and beyond 10 years. Patients experiencing an event before the start of the interval were excluded, while those that remained at risk at the end of the time interval were censored. Therefore, the numbers of patients at risk at 0 year, 5 years and 10 years are the number of patients with no missing information on followup time and event status at each of the time points, respectively.

A rule of thumb is that Cox models should be used with a minimum of 10 outcome events per predictor variable [4, 5], although another study suggested that this rule can be relaxed [6]. We considered the sample size and the number of events are sufficient in our analysis on the first two follow-up time intervals, as well as the last time interval in ER-positive group. The number of events in the last time interval (beyond 10 years) for the ER-negative group (n = 5 in both Affy cohort & METABRIC set) is a concern for prognostic power of the gene signature being reliably studied. However, we consider our analysis to be reliable since the general indications we obtained from the two included datasets are similar: decreasing prognostic power over time, and the direction of the odds ratio associated with the gene signatures. (Also briefly discussed in the main text under Discussion.)

“The good prognosis of the patients with poor prognosis signatures after >10 years follow-up in the ER-negative subtype is an interesting observation. The patients characteristics and the treatment characteristics of this subgroup should be analyzed and discussed.”

Although it is interesting to observe that higher risk scores were associated with lower
risks for distant metastasis in this group of patients, this is based on a small number of events (n=5) and the protective effects (HR<1) were only borderline significant. Thus, we do not consider this protective effect a reliable finding, and are reluctant to make a point out of it. Rather, we consider this emphasize once again the need to build ER-strata specific gene signatures for prognostic / predictive purpose. (Also briefly discussed in the main text under Discussion.)

We compared patient characteristics between good and poor prognostic groups within the ER-/beyond 10 year followup samples. There is more G1 in the good prognostic group, and more HER2+ patients in the bad prognostic group.

“The analysis of the METABRIC data should be presented in more detail (e.g. as in Table 3).”

We now have reported the detailed time dependency analysis on the METABRIC data in Table S5 (format similar to Table 3).
Respond to the comments/suggested revisions from reviewer #2:

“Authors examine various gene signatures and their relation to prognosis over time and ER status.

Conclusions are not really novel (differences between ER status has been noted many times before, signatures are highly correlated).

Inclusion of the meta-bric data is interesting, but more interesting would have been a comparison with the gene classifier included in the Curtis publication.”

We did not include the classification for molecular subtype proposed in Curtis study in comparison of other signatures for the following reasons (which also presented in the main text). Curtis classification is based on integrative clustering both gene expression and copy number data through a joint latent model. The majority of samples in our main analysis did not have copy number data available, while evaluating the Curtis classification on the METABRIC set together with other signatures would bias the results since the Curtis classification was trained on the same set.

“I am not sure the conclusion regarding histologic grade can be correctly drawn - the strong correlations between the signature and grade would make a cox model unstable to evaluate this. A change in Liklihood ratio should be used instead”

For signatures with strong predictive power, e.g. PAM50, EP and GGI, the explanatory power (as measured by the chi-square) did not diminish greatly by the inclusion of histological grade in the model (from Model 1 to Model 2 in Table S4), and remained large. Correlations between histological grade and the signatures should therefore not be a problem. If correlations had been strong enough to make the Cox parameter estimates unstable, this would have been visible as a drop in marginal explained variation (the marginal chi-square value) of the signatures towards zero, which is not the case. Note that the chi-square values remain stable even in cases where, due to correlations between covariates, parameter estimates become unstable.

The chi-square values reported are derived from likelihood ratios. If the changes in likelihood ratios were meant to refer to how the marginal predictive power of the signatures are affected by the addition of histological grade, then this is exactly the drop in marginal chi-square for the signatures from Model 1 to Model 2. In all cases of significance, the histological grade captures only a small portion of the relapse rate explained by the risk signatures.

Since the explanatory power of several of the signatures remains strong even after adjusting for histological grade, it is surprising that the time trend is so heavily influenced. This is why we state that the time trend seems somehow related to histological grade.
Reference:


