Author’s response to reviews

Title: A 25-gene classifier predicts overall survival in resectable pancreatic cancer

Authors:

David Birnbaum (david.birnbaum10@gmail.com)
Pascal Finetti (finettip@ipc.unicancer.fr)
Alexia Lopresti (alexia.lopresti@inserm.fr)
Marine Gilabert (gilabertm@ipc.unicancer.fr)
Flora Poizat (poizatf@ipc.unicancer.fr)
Jean-Luc Raoul (raouljl@ipc.unicancer.fr)
Jean-Robert Delpero (delperojr@ipc.unicancer.fr)
Vincent Moutardier (moutardierv@ipc.unicancer.fr)
Daniel Birnbaum (daniel.birnbaum@inserm.fr)
Emilie Mamessier (emilie.mamessier@inserm.fr)
François Bertucci (bertuccif@ipc.unicancer.fr)

Version: 1 Date: 28 Jul 2017

Author’s response to reviews:

July 27, 2017

Dear Editor-in-Chief,

Please find a revised version of our research paper by Birnbaum and colleagues, entitled “A 25-gene classifier predicts overall survival in resectable pancreatic cancer” that we had submitted to BMC Medicine for publication.

We thank the two reviewers and the Editor for their positive and helpful comments, which have been taken into account as follows.
Reviewer #1’s (Francisco Real) comments and authors’ responses

In this ms. Birnbaum et al. perform an analysis of a large dataset of published mRNA expression studies to identify a poor prognosis signature. They perform an extreme case-case comparison, select 25 best performers among 1400 differentially expressed genes and apply this signature to the whole dataset, the individual datasets and analyze associations with clinical/pathological factors and with survival. They compare the performance of the signature with that of other previously published classifiers and the data support the notion that this signature outperforms the published ones in the multivariable analyses. The work will be of interest to the community and should foster independent prospective validation.

1. Non-ductal tumors represent 10% of the whole series. They should be analyzed separately and a precise description of the tumor types should be provided.

We thank the reviewer for his comment that led us to further analyze the original papers and associated supplementary data. We found that all Collisson’s and Chen’s samples were PDAC (ductal tumors), and not « missing data ». We were also able to better precise the pathological types of many « other » pathological types in the TCGA and ICGC sets. In this revised version, the pathological types are now available for 548 samples and include 537 ductal adenocarcinoma types (98%) and 11 other types (2% : 8 neuroendocrine tumors, 2 acinar cell carcinomas, and 1 intraductal tubulopapillary neoplasm).

We have thus modified Table 1, Table S4, Table 3 (new revised Table 4), and Table 4 (new revised Table 5). We have also modified the correlations of classifications with the pathological type, notably in the uni- and multivariate analyses, without any change regarding the independent prognostic value of our classifier.

There were only 6 non-ductal samples in the validation set of 562 samples, representing 1% of cases. Interestingly, all were classified in the “LTS-like” class, in agreement with the better prognosis of neuroendocrine tumors. Interestingly, the survival analysis limited to the 504 informative ductal samples showed the same results. In the Discussion, on line 4, page 14 (PP), we have added the following sentences : « The pathological type (other vs ductal) was not significant (HR=0.36 with p=0.151) because of the small percentage (1%) of “other” types. Interestingly, all six “other” type samples were classified in the “LTS-like” class, in agreement with the better prognosis of neuroendocrine tumors. Importantly, the OS analysis was not modified when limited to the 504 ductal samples with 27% 2-year OS in the “STS-like” and 48% in the “LTS-like. »

2. Figures 2a and 2b are very difficult to read and they should be drawn in a different manner.
We agree.

The Figure 2 and its legend have been modified.

3. The training set is remarkably small compared with the test set; this may reduce the ability of the used strategy to capture the best genes for the classifier.

We have added a comment in the Discussion.

On line 11, page 13 (PP), the sentence "The learning set included only 39 samples, but was carefully designed for containing two groups with distinct aggressiveness, a group of long-term survivors after surgery (LTS) and a group of short-term survivors (STS).” has been replaced by the following sentence: "The learning set, which included only 39 samples, was remarkably small compared with the validation set; this might have reduced our ability to capture the best genes for the classifier. However, it was carefully designed for containing two groups with distinct aggressiveness, a group of long-term survivors after surgery (LTS) and a group of short-term survivors (STS), and for containing samples profiled using the same technology (RNA-Seq). Such design likely explains the large number of genes (1,400) differentially expressed between the two patients’ groups despite the correction for the multiple testing hypothesis, and the robustness of our final signature in the validation set.”

4. Although it is said that a comparison of the performance of the 25 gene-classifier in comparison with random signatures was analyzed, I cannot find that detailed information in the manuscript. Only a short statement is provided in the Results section.

We do not agree with this comment because this analysis was already detailed in the initial version in the “Material and Methods” section (8 lines from line 8 to line 15, page 8 PP) and of course in the “Results” section (5 lines from line 25, page 10 to line 4, page 11 PP). There was also one line in the Abstract and 4 lines in the “Discussion” section (from line 7 to line 10, page 14 PP).

5. Why use a two-tiered classifier rather than developing a 25-gene based score for continuous classification?

We agree with the reviewer. We have also tested our 25-gene signature as a continuous score. In univariate and multivariate prognostic analyses, this score was significant, suggesting that our signature will warrant further testing not only as a two-tiered classifier, but also as a continuous prognostic score.
In the Conclusion, on line 6, page 17 (PP), the sentence « The testing of our signature in the current prospective trials of adjuvant and neoadjuvant chemotherapy trials is warranted. » has been replaced by the following sentences « the testing of our signature in the current prospective trials of adjuvant and neoadjuvant chemotherapy trials is warranted, and should be tested not only as a two-tiered classifier, but also as a continuous score. Indeed, a continuous score based on the expression of 25 genes showed significant prognostic value (data not shown) in univariate analysis (HR for death of 2.84 (95%CI 2.06-3.91), p=1.96E-10) and in multivariate analysis (HR for death of 3.25 (95%CI 2.11-4.99), p=7.42E-08). »

6. A pathway analysis of the 1400 genes should be provided in the main text as a Table.

We agree.

A new Table 2 has thus been added, and the following tables have been renumbered.

On line 5, page 10 (PP), the sentence ”The associated GO biological processes are shown in Additional file 3 (Table S3).” has been replaced by the following sentence : ”All associated GO biological processes are shown in Additional file 5 (Table S3), and the top 40 processes are shown in Table 2”.

7. Table 3: do the results presented correspond to associations or to correlations, as stated in the title? I believe that these are associations.

We agree.

We have replaced « correlations » and « correlated » with « associations » and « associated » in the Text and the corresponding Tables and titles.

8. The authors may not have access to the full Australian ICGC dataset but indeed this would be the largest single dataset in which to test the classifier. It is not correct to say that the largest study corresponds to 102 cases.

Indeed, we could not have access to the full Australian ICGC expression dataset. The corresponding GEO dataset contains only 91 samples out of 232 and no clinical data is available.

Regarding the issue of the « largest study », we agree with the reviewer. In fact, we meant the largest study using supervised analysis. But it is true that the largest prognostic study, which used unsupervised analysis for identifying prognostic molecular subtypes, is the Australian ICGC study by Bailey and colleagues with 328 samples. On line 5, page 13 (PP), the sentence « The
paucity of tumor specimens available for analysis explains the relatively small number of samples profiled in previous prognostic studies, 102 in the largest one [20]. » has been replaced by the following sentence: « The paucity of tumor specimens available for analysis explains the relatively small number of samples profiled in previous prognostic studies, 102 in the largest one [20] that used supervised analysis, and 328 in the Australian ICGC study [25] that identified prognostic molecular subtypes by unsupervised analysis.”

9. A statement should be included in the Discussion indicating that, despite the very high p value, the HR for the signature in both the univariable and multivariable analyses is relatively low, around 2, and therefore of uncertain clinical value.

We agree.

On line 6, page 17 (PP), the sentence “The testing of our signature in the current prospective trials of adjuvant and neoadjuvant chemotherapy trials is warranted.” has been replaced by the following sentence: “Despite the very high p-values, the HR for death was relatively low, around 2, in both univariate and multivariate analyses, and therefore of uncertain clinical value. However, we think that the testing of our signature in the current prospective trials of adjuvant and neoadjuvant chemotherapy trials is warranted.”

10. The discussion of specific genes in the signature is not useful, in my view: most likely it does not reflect causal effects but associations. The pathway discussion may be more relevant.

We agree that genes of our signature do not automatically reflect causal effects. However, it was interesting to find for some of them some functional or clinical data that may be related to their association with survival, as recognized and appreciated by the second reviewer (see his fourth comment “well description of functional analysis of the classifier 25 genes in discussion »). Thus, we have conserved this part in the discussion. However, on line 19, page 14 (PP), we have added the following sentences: “Whether the 25 classifier genes are causative of the phenotype in a biological sense or reflect another associated phenomenon remain to be explored. However, it was interesting to find some genes already reported as associated with cancer biology and/or clinical outcome of cancer patients.”

11. I am not convinced about the statement in the conclusion regarding the possible usefulness of the signature in determining whether patients should undergo immediate surgery vs.
neoadjuvant chemotherapy since PDAC is most likely a systemic disease in the vast majority of cases.

In the Conclusion, we just suggest (line 6, page 17 PP) that the next step will be to test the prognostic performance of our signature in the ongoing prospective trials of adjuvant and neoadjuvant chemotherapy. We remind that, even if PDAC is likely a systemic disease at the time of diagnosis, neoadjuvant chemotherapy is not a standard treatment in resectable pancreatic carcinoma and several patients who display early post-operative relapse could have been spared from toxicity of unnecessary surgery. In the future, prognostic features will be necessary to help the difficult choice between immediate surgery or neoadjuvant chemotherapy. We just suggest (line 7, page 17 PP) that our 25-gene classifier might help, but of course will require prospective validation of its clinical utility before any application in clinical routine. On line 11, page 17 (PP), we have added the following sentence “Of course, the clinical utility of this approach will have to be prospectively demonstrated before any use in clinical routine.”

Reviewer #2’s (Dung-Tsa Chen) comments and authors’ responses

Summary: The manuscript is well organized and written. Several issues are needed to be addressed.

Strengths:
1. Excellent introduction of current clinical issue of pancreatic cancer and potential utility of genomic signature.
2. Significant effort to collect and synchronize the 9 valuable gene expression data of pancreatic cancer.
3. Comprehensive analysis of univariate and multivariate analysis with comparison to clinical variables and other molecular subtypes.

We thank the reviewer for his positive comments.

Weaknesses:
1. Training set is based on TCGA to test OS>36 months vs. 2-6 months OS. How are the results if other datasets are used as the training set? Or are these 1,400 genes using TCGA
also significant in other datasets when comparing OS >36 months vs. 2-6 months OS? What are the distributions (sample size) of other datasets regarding OS: >36 months and 2-6 months? Any justification to use TCGA instead of other dataset as training set?

The training set for supervised analysis was not based on TCGA only, but on TCGA and ICGC. It was carefully designed according two combined criteria: containing two patients’ groups with distinct outcome, a group of long-term survivors after surgery (LTS, with OS>36 months) and a group of short-term survivors (STS, with 2-6 months OS), and containing samples profiled using the same technology (RNA-Seq). Given the distribution of LTS and STS samples in the other data sets and the technology used, the TCGA set (22 samples; RNA-Seq) and Bailey’s set (17 samples; RNA-Seq) were retained for the learning set. As suggested by the reviewer, we have added in Table S1 (Additional file 1) the distributions of all sets regarding the LTS and STS samples.

As suggested, we have also tested the ability of the 1,400 genes (differentially expressed between LTS and STS in the learning set) for classifying the 67 LTS and STS samples from other data sets. Out of the 67 samples classified, 49 (76%) were accurately classified, suggesting the robustness of these genes in independent sets (p=7.68E-05, Fisher’s exact test). On line 6, page 10 (PP), we have added the following sentences: “The robustness of those genes was tested by testing their ability for classifying the LTS and STS samples from the other independent data sets. Out of the 67 samples classified, 49 (76%) were accurately classified, suggesting strong robustness (p=7.68E-05, Fisher’s exact test).”

2. It is unclear how different datasets are pooled. It seems dataset is normalized individually. Since each dataset may use different platform of microarray/RNA-seq, special treatment is performed on probeset level to convert into gene level data for later data analysis. However, it is unclear if any cohort adjustment (or batch correction) is performed prior to data pooling. In other words, how the validation cohorts are pooled. Explanation is needed.

We agree and have improved this point.

Each data set was normalized individually. Regarding the learning set, the TCGA and the ICGC sets, based on the same RNA-Seq technology, were pooled before supervised analysis by using COMBAT (empirical Bayes) as batch effects removal method, included in the inSilicoMerging R/Bioconductor package. The final merged set included 15,291 genes in log2-transformed data. The accuracy of normalization was controlled by principal component analysis (PCA) that we display now in the new Additional file 2 (Figure S1). This information has been added on line 14, page 7 (PP) as follows: “Samples of the learning set were pooled before supervised analysis by using COMBAT (empirical Bayes) as batch effects removal method, included in the inSilicoMerging R/Bioconductor package. The final merged set included 15,291 genes in log2-
transformed data. The accuracy of normalization was controlled by principal component analysis (PCA) (Additional file 2: Figure S1).

By contrast, regarding the validation set, each of the normalized sets was treated separately, notably for classifying the samples according to the molecular subtype classifiers and our prognostic signature. Once classified, all samples were gathered before the statistical analyses.

On line 6, page 8 (PP), the sentence “Its robustness was assessed in the independent validation set (N=562) by classifying each sample as “STS-like” or “LTS-like”.” has been replaced by the following sentence: “Its robustness was assessed in the independent validation set (N=562) by classifying each sample in each data set separately as “STS-like” or “LTS-like”.” And on line 12, page 7 (PP), the sentence “The 562 other samples with available survival data from the other data sets were pooled and used as an independent validation set.” has been replaced by the following sentence: “The 562 other samples with available survival data from the other data sets were gathered and used as an independent validation set.”

3. It has become a standard to provide function codes with output of data analysis to ensure reproducibility and transparency. Since R software is used in the study, authors should consider Rmarkdown using knitr package to generate the report, such as from data processing from probeset to gene level, data pooling, gene signature development of 1400 and 25 genes in training set and the validation in test set, as well as the comparison to clinical variables and three molecular subtype classifiers in univariate and multivariate analyses.

A Sweave report describing the analysis of gene expression data and the associated statistical analysis has been created and is available as Supplementary File. This information has been added on line 9, page 9 (PP) with the following sentence: “A Sweave report describing the analysis of gene expression data and the associated statistical analysis is available as Additional file 3 (Supplementary Text).”

4. Evaluation of the random noise gene seems biased. The 25-gene signature is not directly derived from the whole 15,261 genes. It is the result after tuning up the 1,400 genes, meaning it is selected from the 1,400 genes, not from the 15,261 genes. Thus, the evaluation should be based on the 1,400 genes.

We have redone the analysis as suggested by the reviewer. Rather than generating the 100,000 random 25-gene signatures from the 15,261 genes, we have generated them from the 1,400-gene list identified by supervised analysis. We found the same result: none of the random signatures was more significant than the data-derived 25-gene signature, suggesting that this latter represented an optimal prognostic combination.
On line 10, page 8 (PP), the sentence “A resampling scheme was used to generate 100,000 random 25-gene signatures within the 15,261 genes included in the normalized and pooled learning set.” has been replaced by the following sentence: “A resampling scheme was used to generate 100,000 random 25-gene signatures within the 1,400 genes differentially expressed identified by supervised analysis in the learning set”.

On line 25, page 10 (PP), the sentence “To assess the likelihood of our 25-gene signature as a non-random signature, we generated by a resampling scheme 100,000 random 25-gene signatures extracted from the learning set and tested their prognostic value in the validation set.” has been replaced by the following sentence: “To assess the likelihood of our 25-gene signature as a non-random signature, we generated by a resampling scheme 100,000 random 25-gene signatures from the list of 1,400 genes differentially expressed and tested their prognostic value in the validation set.”

5. The algorithm from 1400 genes to 25 genes is unclear. It seems LASSO is used to finalize the 25 genes using the glmnet package. However, how is the penalty regularization parameter λ value finalized? Detailed explanation is needed.

The penalty regularization parameter λ value was finalized by using the lambda.min that is the value of Lambda giving minimum mean cross-validated error (lambda.min was 0.0153).

On line 4, page 8 (PP), we have added the following sentence: « The λ value was finalized by using the lambda.min that is the value of lambda giving minimum mean cross-validated error (lambda.min was 0.0153). »

6. The TCGA data as training set is for PDAC. However, some of the other datasets may not be PDAC. This may create heterogeneous issue. Please justify.

As explained in our response to the first comment from the first reviewer, in this revised version, the pathological types are now available for 548 samples and include 537 ductal adenocarcinoma types (98%) and 11 other non-ductal types (2% : 8 neuroendocrine tumors, 2 acinar cell carcinomas, and 1 intraductal tubulopapillary neoplasm). There were only 6 other non-ductal types (1%) in the validation set, and all were classified in the “LTS-like” class, in agreement with the better prognosis of neuroendocrine tumors. Furthermore, the survival analysis limited to the 504 ductal samples in the validation set showed results similar to those observed with all samples. Such data have been integrated in the revised version, as detailed in our response to the first comment from the first reviewer.
7. How many of the 25-gene signature overlap with other signatures and the three molecular subtypes.

We have done this analysis and added the results.

On line 13, page 10 (PP), we have added the following sentences: « We assessed the gene overlap between our 25-gene signature and the three molecular subtype classifiers [16,25,26] and five other signatures recently published that displayed robust and independent prognostic value [15,17,20,22,30]. As shown in Additional file 6 (Figure S2), there was no overlap with the five signatures, and the overlap with the molecular subtype classifiers was very low (0 gene with Collisson, 1 with Moffitt stroma, 2 with Moffitt tumor, and 3 with Bailey). »

A new Additional file (Figure S2) has been added.

Editorial Policies and authors’ responses

In accordance with BioMed Central editorial policies and formatting guidelines, all manuscript submissions to BMC Medicine must contain a Declarations section which includes the mandatory sub-sections listed below. Please refer to the journal's Submission Guidelines web page for information regarding the criteria for each sub-section (https://bmcmedicine.biomedcentral.com/).

The formatting of our initial and revised versions, notably regarding the Declarations section, agrees with the journal’s guidelines.

As you can see, we have answered all questions raised by the Reviewers and modified the manuscript as suggested. We hope that this improved version will meet with your approval for publication in BMC Medicine.

Sincerely yours,

Pr François BERTUCCI