Author's response to reviews

Title: A gene expression profile for detection of sufficient tumour cells in breast tumour tissue: microarray diagnosis eligibility

Authors:

Paul Roepman (paul.roepman@agendia.com)
Arenda Schuurman (arenda.schuurman@agendia.com)
Leonie JMJ Delahaye (leonie.delahaye@agendia.com)
Anke T Witteveen (anke.witteveen@agendia.com)
Arno N Floore (arno.floore@agendia.com)
Annuska M Glas (annuska.glas@agendia.com)

Version: 2 Date: 29 May 2009

Author's response to reviews: see over
Dear Dr. Alexandersson,

Thank you for reviewing our paper “A gene expression profile for accurate tumour cell percentage scoring of breast tumour tissue: microarray diagnostic eligibility”.

We have read and addressed all issues raised in the review and hereby submit a revised manuscript.

In response to the Editors requirements we have addressed all the comments of the reviewers. See attached a point-by-point response to the concerns and suggestions.

We hope the answers and alterations made to the paper are satisfactory and will lead to the acceptance for publication of our manuscript in BMC Medical Genomics

Sincerely,

Anuska Glas

MS: 6385402902504104 - Referee comments

Editorial comments

We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. Please ensure that your revised manuscript conforms to the journal style. It is important that your files are correctly formatted.

Data availability and deposition - Please deposit your data in Gene Expression Omnibus (GEO) and state the relevant accession numbers as pointed out by Referee 4. We have uploaded our data to GEO and will be made publicly available upon publication of the manuscript under the accession number GSE16201. The GEO ID is now provided in the methods section on page 4: "Full genome gene expression measurement of all samples analyzed in this study is publicly available at the Gene Expression Omnibus (GEO) with accession number GSE16201."

Referee #1

In this work, the authors describe the discovery and validation of a gene-expression based classifier capable of distinguishing low and high percentage of tumor cell tissue in frozen breast tumor samples. Given the subjective nature of pathological review for tumor cell percentage, and the requirement of tumor-section staining which is not easily adaptable for core and fine-needle
biopsies, an accurate gene-expression based predictor of tumor cell percentage could hold merit as a rapid and quantitative alternative to conventional methods. Overall, this version of the manuscript is a bit short on details and interpretation of results, but the experimental design is mostly sound, and the findings are intriguing and worthy of further investigation.

Major Compulsory Revisions
1) The Methods section is scant on details. Where did the tumor samples come from? By what criteria were they selected (other than tumor cell percentage)? Was the use of these specimens approved by an institutional review board? How were the microarray data processed and normalized? What was the rationale for using a nearest-mean classifier?

The rationale for using a nearest mean classifier was based on studies we have previously performed and which resulted in valid gene classification on independent samples. (e.g. van ‘t Veer et al, Nature 2002, Roepman et al, Nat Genet 2005, Roepman et al, 2009 Clin. Cancer Res.). Moreover, we have shown that a nearest-mean classifier is very robust for technical changes, is suitable for use in a diagnostic setting (Glas et al, BMC Genomics) and has been approved by regulatory bodies such as the FDA (MammaPrint test which is based on a nearest-mean classification algorithm). Therefore, our goal was to develop a TCP profile based on a similar procedure that can be adopted relatively easily for diagnostic use.

Regarding data processing and normalization, we have added the following text to outline more details about the procedure: page 4, 1st paragraph: "Microarray slides were scanned using an Agilent G2565AA scanner and were quantified using Feature Extraction software (version 9.5, Agilent). Gene expression log-ratio data was obtained from non-background subtracted sample/reference signal and was normalised using a lowess global normalisation procedure."

We have added the following text on page 4: "Four hundred and three frozen tumor samples or tumor samples preserved in RNALater from breast cancer patients were used. All samples were de-identified and were part of research implementation studies for MammaPrint. No patient data was collected, only the tumor cell content of the sample was determined."

2) The training and testing logic for constructing the classifier appears sound up until the final selection of 13 genes. The authors explain that of their initial set of 35 genes capable of optimal classification in the first training set, those with large variation in an additional set of 70 tumors with high (>50%) TCP were removed, resulting in 13 genes for setting the classification thresholds/weights (in the training set) and validating in the test set. How these 13 genes were selected is a bit ambiguous. What variation threshold was applied and why?

We agree with the reviewer that the methods section is somewhat limited on details regarding the final gene selection procedure. To clarify the selection of the final 13 genes from the first set of 35 genes we have added the following lines on page 4: "Twenty-two genes that showed a variation in gene expression greater than 0.4 (stdev across 70 samples) were excluded from further analysis. This stringent threshold was used to ensure a stable profile in tumours with a relative high tumour cell percentage.". Please also see other reviewer comments regarding additional issues that have now been included in the methods sections.

3) The authors do not provide a clear interpretation of how the TCP classifier should be used with respect to its predictive performance. In the title it is said to be for “accurate tumor cell percentage scoring…”, and it is suggested in the abstract that it essentially replaces histopathological analysis for initial tumor percentage scoring. However, the predictive
performance of the classifier does not demonstrate accuracy in scoring tumor cell percentage, per se, but rather indicates an ability to distinguish low TCP from medium/high TCP better than chance.

As correctly pointed out by the reviewer, the development of the TCP profile was aimed towards assessment of sufficient tumour cells in a biological sample to be used for microarray diagnostic. Although the TCP profile is capable in correctly identifying samples with a too low tumor cell percentage, the results in this study do not automatically indicate that the profile outperforms pathological tumor cell percentage scoring. We have therefore changed all notification regarding the accuracy of the profile for detection of tumor cell percentages into its capability in detecting samples with a sufficient high tumor cell content.

Regarding this issue we have changes the title of our manuscript into "A gene expression profile for detection of sufficient tumour cells in breast tumour tissue: microarray diagnosis eligibility". Furthermore, we have made the following changes: page 2: "The developed 13-gene profile will provide an additional tool for accurate and more objective assessment for tumour cell content within breast cancer tissue." into "The developed 13-gene profile provide an objective tool for assessment whether a breast cancer sample contains sufficient tumour cells for microarray diagnosis.", and on page 7: "Thus we believe that the... score based on formalin-fixed H/E stained slides." into "Since the TCP profile is based on transcriptional levels of high TCP related gene expression within fresh or frozen tumour tissue, we believe that the developed gene profile likely gives a better indication whether a sample is suitable for microarray diagnostics compared to a pathological tumour cell percentage scoring on formalin-fixed H/E stained slides.", and on page 8: "...for objective assessment of tumour cell content..." into "...for objective assessment of sufficient tumour cell content...".

While it is clear by the overall accuracy and kappa score that conventional scoring and the TCP predictor are associated with one another, a more relevant question is what does it mean that 22% of the low TCP samples were incorrectly classified as medium TCP (30-49% TCP) or high TCP (>50% TCP)? And the fraction of medium and high TCP incorrectly classified as low TCP is not disclosed or discussed. These questions are of practical concern (ie, if the utility of the classifier lies in its ability to distinguish low TCP from medium/high TCP tumors), and should be considered in the context of conventional histopathological review variation. Which is more reliable, an approximate 10% average variation in TCP scoring attributable to inter-pathologist disagreement, or the performance of the classifier given its classification error in distinguishing low from medium/high TCP tumors? Some analysis addressing this perspective would provide much needed context for interpreting the utility of the classifier.

As suggested by the reviewer, we have added the false-positive and false-negative rates of the TCP gene profile compared to pathological classification to the results section on page 6: "The profile showed a false-positive classification rate of 5% (high-TCP samples classified as low by the profile) and a false-negative classification of 21% (low-TCP samples classified as high by the profile)". We have further outlined the performance of the TCP profile by analysis of the index on a continuous scale compared to continuous pathological scoring. This has been added to the results section on page 6: "Analysis of continuous TCP profile index and continuous pathological tumor cell percentage scoring resulted in a significant association between both methods ($R^2=0.48$, Wilcoxon $P<0.001$, AUC 0.90) and indicated that the profile might also be useful for indication of accurate tumor cell percentage instead of a low-, medium- or high-TCP classification."
We have further acknowledged the utility of the profile in the discussion of our manuscript and the issue regarding which method is better/more reliable: a pathological TCP assessment able to identify tumor cell percentages up to 10% increments that will delay microarray processing or a quick TCP-profile analysis able to distinguish low from medium/high TCP samples. We have added to following paragraph in the discussion on page 7: "The utility of the gene profile lies in its capability to identify tumours with a high percentage of tumour cells compared to tumours with insufficient tumour content for subsequent microarray diagnostics. Conventional histopathological review results in tumour cell percentage scorings up to 10% increments but is laborious and requires an experienced (in-house) pathologist. The profile, on the other hand, is able to distinguish samples with low-, medium- or high-tumour cell content. Although the 13-gene profile provides a more qualitative measurement compared to the quantitative pathological assessment, it is a value tool for an objective TCP scoring based on transcriptional levels that can quickly identify samples suitable for diagnostics."

Also have we changes the following sentence on page 7, 2nd paragraph: "...will remain necessary for detection of ductal carcinoma in situ (DCIS) and necrosis to define the suitability of the specimen," into "...will remain necessary for detection of ductal carcinoma in situ, necrosis and a detailed assessment about the percentage of tumour cells to define the suitability of the specimen,“.

Minor Essential Revisions
1) In the last paragraph of the results section, it is stated that, “The difference between medium and high TCP samples was not significant...as the majority of the medium TCP samples are correctly classified as high TCP by the molecular profile.” Cross-classification (“medium” classified as “high”) is really “incorrect” classification, but I think the point is more at the fact that medium and high TCP samples, while viewed by the classifier in a similar way, still remain measurably different than the low TCP samples on the whole. Perhaps the authors can rewrite this sentence for greater clarity.

We have re-written the last paragraph of the results section to clarify the observed differences low, medium and high classification and have removed the claim of ‘correct classification of medium-TCP samples as high-TCP samples by the profile’, as indicated by the reviewer. We have changes the following text: "Statistical analysis of the profile....are correctly classified as high TCP by the molecular profile." into "Although the difference in index between medium and high TCP samples was borderline significant (P=0.02), the majority of medium TCP samples (78%) were classified as high TCP by the molecular profile. This result indicates that both medium and high TCP harbour a different tumour cell related gene profile compared to low TCP samples."

Discretionary Revisions
1) As prognostic gene expression predictors of outcome in breast cancer have been developed using the same (or similar?) microarray platform as was used by the authors, it might be interesting to attempt to evaluate how the TCP classifier can add value to the performance of such a prognostic signature. Presumably, tumors with suboptimal TCP will confound performance of a prognostic signature, as the signature would not be reliably assessed in those tumors. If the TCP classifier can identify tumors predicted to have low TCP, then might the exclusion of those tumors enhance the classification performance of the prognostic signature in question?
Very interesting suggestion. However, the great majority of samples analyzed using our prognostic signature (MammaPrint) has a tumor cell percentage over 50%. We have previously investigated whether pathological tumor cell percentage showed any association with the prognostic power. These results (unpublished) indicated no such effect. It would be very interesting to see how the TCP profile would perform but unfortunately, for the samples analyzed in this cohort we do not have survival data available.

2) Does tumor histology (ductal, lobular, mucinous, tubular, medullary) make a difference in classifier performance?
This is a very good point as it is known that histopathological tumor cell percentage scoring is in general harder for lobular carcinomas than for ductal carcinomas (reference or not??). Unfortunately these data are not available for most patients. The majority of the patients are ductal carcinomas.

3) It would be nice to demonstrate that the predictor’s performance is robust to issues related to choice of microarray normalization procedure and final gene selection method. For example, does 10, 15, 20 genes have essentially the same test-set performance as the 13 selected genes? We have investigated the issue regarding gene selection stability and gene set performance within the identified set of 35 genes. Interestingly, selection of just random 2 genes within this set already results in a relative high performance (mean accuracy >80 percent). This suggests that for the final classifier generally all genes are of equal importance. These results are included in the results section on page 5.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Referee #2

This paper addresses an interesting question of using microarray data to estimate the tumor cell percentage (TCP) using microarrays. Although this question is not urgent right now, but this issue will be important if the use of microarrays for breast cancers become extensive, which is part of the mission of Agenda where these authors work. Even though this issue is not of concern in current medicine, I could see the value of these studies. To achieve this goal, the authors first use the estimated TCP from pathologists as “golden standards” and phenotypes for supervised analysis on the training datasets. From these supervised analysis, a small subset of 13 genes (all highly expressed in tumors with TCP) were put forward due to their ability first in the cross-validation sets and then in an independent cohort of patients to test the ability of these genes in predicting the TCP in more breast cancer patients. Although I can see the value of this study, I have some some major concerns which need to be addressed.

1. Theoretically, the TCP can be represented by the degree of gene expression of either epithelial
cells (for breast carcinoma) or proliferation cluster (for mitotic activities) as surrogate. It may be of value for the authors to compare the performance of 13 genelists with these two possible gene signatures to represent the TCP.

We agree with the reviewer that the biological origin of the gene expression measured by the developed TCP profile can be a valuable starting point of research towards the overlap with known factors linked to breast cancer. The study described in this manuscript was focused towards identification of the optimal set of genes that is able to discriminate samples with high TCP compared to low TCP, with its main focus on a methodological issue. While we do acknowledge the scientific interest in the overlaps between previously associated tumor-cell specific gene expression, we believe that this lies outside the scope of the current study. We do however have addressed other issue regarding the biological background of the identified 35- and 13-gene set as indicated by the reviewer comments below.

2. The 13 gene listed in Table I is of limited obvious biological information other than ANAPC7 (associated with proliferation). Previous text in the result mentioned several cancer-related processes associated with the 35 gene TCP. What are the genes in the 35 genes which account for these GO term enrichment? I am pretty sure that no such GO enrichment would be found for the 13 gene list in Table I. It is of concern of using 13 genes without no obvious biological linkage to represent a obvious histopathological phenotypes (TCP) under microscopy. For the reasons mentioned by the reviewer, we have performed the GO analysis on the larger set of 35 genes. Indeed the set of 13 genes looks less related to known factors involved with tumor development. However we believe that this is less relevant to the performance of the profile as it has been validated on a large independent sample cohort. We therefore decided to continue with the set of thirteen. Nevertheless, to give more insights in the biological annotation of the profile associated gene expression we have inserted in the results section the names of all genes present in the GO category significantly enriched in the profile and related to cancer progression. The first paragraph on page 6 has been changed into: "Biological function analysis indicated that the 35 TCP associated gene set was enriched for genes associated with cancer related processes such as cellular morphology (MGAT3, AQP1, TPM3, MYLK, SFRP1, NTRK2: p=1.7e-3), cell-to-cell signalling (JAM2, LOC338328, OGT, MYLK, NTRK2, SNAP29: P=1.7e-3), cellular movement and invasion (MGAT3, TPM3, MYLK, SFRP1, NTRK2, CCL15: P=1.7e-3) and cellular survival and apoptosis (TPM3, PRR13, OGT, MYLK, SFRP1, P=3.6e-5)."

3. Although the authors claims that this 13 gene score system can separate the tumors with low, medium and high TCP reasonable well. But the actual data between these groups had significant overlap. More relevant analysis may be to use the TCP gene expression score to separate the tumors into 3 groups – high, medium and low TCP score and see whether this classification will correctly predict the rating of pathologists.

We agree with the reviewer that the medium-TCP group and the classification of these samples remains questionable. As suggested we have repeated classification by the gene profile into three groups to investigate whether medium-TCP sample might form a separately identified class. Unfortunately, this analysis resulted in an overall poorer performance compared to classification into two group (data not shown). This results could already have been guessed by looking at figures 2D and 3A which indicate that the medium-TCP group has a wide variation in TCP profile indexes. An additional issue regarding the medium-TCP group is that this grouping is based on a pathological scoring which, as indicated in the paper, suffers from a variance up to
plus or minus 10-15%. We therefore decided to include the "intermediate" medium-TCP group in order to keep the low-TCP and high-TCP analysis as clean as possible.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Referee #3

General comment
The authors developed a gene-expression signature correlated with tumor cell percentage (TCP) in breast cancer tissues to assess sample quality for subsequent application of diagnostic gene-expression assay like MammaPrint. There are some concerns in the signature development and its practical usefulness as described below.

Major comments
(1) A methodological weakness is that the signature was trained based on the histologically defined TCP, which was described as inconsistent and subjective. In fact, according to Fig.1B, TCP varies particularly at its lower range, where more reliable measurement would be required (the variation looks greater than +/-10% within this range of TCP). To address this issue of uncertainty of TCP in training the signature, one potential way would be to create pools of RNA samples extracted from either of pure tumor or pure non-tumor tissues, mix them in a series of different TCPs, and profile them to define the signature according to the known and definite TCP. The signature then could be tested in an independent set of actual patient samples.
We agree with the fact that the development based on a pathological scoring that itself suffers from inconsistency and subjectivity might result in a suboptimal gene profile, as one can therefore argue that the developed TCP-profile also suffers from these factors. However, our main goal was not to develop a better tool of TCP assessment but a method that can be performed independent of pathological expertise and gives an objective scoring based on transcriptional gene levels instead of the number of tumor cells. The use of a robust cross validation procedure that included a mimicked pathological variation on the model should, in principle, be able to select a set of genes that do no/suffer less from this variation. We have outlined this issue in the discussion on page 7 and added the following paragraph: "Since the gene profile was developed in such a way to mimic a pathological scoring which has been described as inconsistent and subjective, one might argue that the TCP profile also suffers from these factors. The main goal of this study, however, was not to develop a more accurate tool for TCP assessment but an objective method that can be performed independent of pathological expertise and which is based on transcriptional gene levels instead of the number of tumor cells. The use of a robust cross validation procedure that included a mimicked pathological variation was therefore included in the selection model that should, in principle, select a set of genes that are robust to this variation. Nevertheless, the 10 percent misclassification between the gene profile and pathological scoring might partly be attributed to this phenomenon."
The classifier development procedure does include a direct comparison of a set of high-TCP samples with a set of low-TCP samples. Although the sample have not physically been pooled before hybridization, the analysis shown in figure 2B by comparing high- to low-TCP samples have proven to be useful for development of a strong profile. We appreciate the suggestion of the reviewer.

(2) Even if the signature accurately captured TCP, the high success MammaPrint prediction rate around 80% even in the “low TCP” group may indicate that TCP is not an optimal predictive marker to decide the subsequent application of MammaPrint. One question would be whether a signature can be trained based on the success of MammaPrint prediction instead of histologically defined TCP.

Very good point. Although MammaPrint currently requires a minimal percentage of tumor cells, as the profile has been developed using samples with relative high TCP, this does not mean that MammaPrint will not work on low TCP samples. We agree with the reviewer that a different type of signature or other selection criteria linked to MammaPrint success rate might be a better indicator for samples eligibility. We are currently setting up such a study and are anxious to see those results. For this study, the development of a breast cancer TCP profile will also be useful for other diagnostic gene profiles on breast cancer (in development). This way, all breast tumor samples can be scored for a sufficient amount of tumor cells. Nevertheless, we agree that a more specific signature can likely be developed specifically for MammaPrint eligibility.

Minor comments
(1) Please provide GEO accession number for the microarray platform and datasets described in the manuscript.

The GEO accession number (GSE16201) has now been provided.

(2) It is not clear how the prediction analysis was performed (e.g., was the signature trained and tested based on 2~3-class comparison or continuous value of TCP?, how was the “pathologist variation” incorporated in the model?). It would be better to overview it in Methods section.

We have inserted more details into the material and methods section concerning the TCP associated gene selection procedure which was based on a continuous value analysis, and how we included pathological variation into the supervised learning model. Page 4, 2nd paragraph, added sentences: "Within each CV loop, two-thirds of the samples were randomly selected as training samples and their TCP was adjusted randomly by either -10, 0 or +10 percent to represent pathological variation, as indicated above. For all genes the association between transcriptional levels and continuous pathological TCP was determined (Pearson correlation) and the top 200 genes (100 with positive association and 100 with negative association) were selected for inclusion in the classifier model.”

(3) Page 5, last paragraph: It is not clear what is the point (this part may not be necessary).

We have included these results to show the stability of the classifier performance if only a subset of genes is used for classification. We believe that this is a good method to prove that the complete set of genes is not overtrained on the training samples, but is based on a real association of all gene members within the profile to the TCP scoring. As this issue was requested by reviewer #1, we decided to keep this paragraph into the manuscript.
(4) Figure 2B: Please provide x-axis. We have updated figure 2.

(5) Page 7, last line: It may not necessarily be RT-PCR. Agree, but RT-PCR will likely be most practical for single-gene readout in a diagnostic setting. We therefore want to limit the suggestion to RT-qPCR only.

(6) Page 8, line 3: It would not be clear if the proposed approach really reduce the cost depending on the actual implementation of the TCP prediction assay. Although direct costs might not decrease due to the implementation of the profile for TCP assessment, we are confident that it will reduce the processing time, thereby improving the efficiency and capacity of our microarray diagnostic pipeline and reducing the (indirect) costs.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Referee #4

The paper by Roepman, P. et al. addresses an outstanding issue in the microarray field: what is the percent tumor in any given sample, and are there gene expression patterns reflective of that percentage. This group takes a straight-forward approach, and correlates pathologist’s scoring of tumor percentage with gene expression patterns. Further, this group uses their refined gene set in a large independent dataset to show there is significant correlation between pathology-based and gene expression predicted tumor content.

Discretionary Revision: This work is well presented and described. The use of the validation cohort confirms their findings. It would be ideal to use samples in which there is a “spike-in” of tumor versus non-tumor RNA in order to have tight control over tumor percentage (and correlate that to their signature) as an additional validation method, but this study meets the level of confirmation necessary for publishing without that data.

We thank the reviewer for this good suggestion on how to better control and confirm the tumor cell percentage measurement. Similarly as suggested by reviewer 3, we also believe that an artificial composition of TCP will be the ideal proof-of-concept. However, as also indicated by both reviewers, the current study contains sufficient data to confirm that the TCP profile is correlation with pathological TCP. Future research to further optimize the tumor cell percentage issue in relation to microarray diagnostic will definitely consider these suggestions.

It would also be useful in Figure 1 to provide further examples.
While we do have additional data regarding the variation between pathological observed directly from a HE-section compared to a scoring based on a scanned image of a HE section, we unfortunately have no additional examples which sufficient different time-points and different pathologists. Although we feel that our data together with data described in literature is convincing enough to support the fact that currently TCP scoring is a rather subjective procedure, we agree that additional examples would be helpful to make our message even stronger.

Major Compulsory Revision:
Before acceptance for publication the data in its entirety needs to be placed on GEO. I could not find the data under the lead or last authors name. Please provide GEO ID so I can reference this work.
We have uploaded our data to GEO and will be made publicly available upon publication of manuscript under the accession number GSE16201. The GEO ID is now been provided in the materials and methods section on page 4: "TCP associated gene expression data....Agendia website" has been replaced by: "Full genome gene expression measurement of all samples analyzed in this study is publicly available at the Gene Expression Omnibus (GEO) with accession number GSE16201.

Overall, this is a well-performed study that could be useful to investigators for identifying tumor samples with low percentage of actual tumors. Its utility in “normalizing” samples based on tumor content will provide an additional measure of sample quality.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.