Author's response to reviews

Title: Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study

Authors:

Annemieke W.J. van Winden (A.W.J.vanWinden@umcutrecht.nl)
Marie-Christine W. Gast (Marie-Christine.Gast@slz.nl)
Jos H. Beijnen (Jos.Beijnen@slz.nl)
Emiel J.Th. Rutgers (E.Rutgers@nki.nl)
Diederick E. Grobbe (D.E.Grobbee@umcutrecht.nl)
Petra H.M. Peeters (P.H.M.Peeters@umcutrecht.nl)
Carla H. van Gils (C.vanGils@umcutrecht.nl)

Version: 2 Date: 6 November 2008

Author's response to reviews: see over
November 6th, 2008

Dear Mrs. Norton,

We would like to thank you for reviewing our manuscript “Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study” for publication in BMC Medical Genomics. We have revised the manuscript, taking into account the comments of the Reviewers. You will find our point-by-point response to the comments of the Reviewers attached.

Yours faithfully,

Annemieke van Winden, MSc
Carla H. van Gils, PhD
In answer to the specific concerns of the Reviewers:

**Reviewer:** Paul Townsend

**Minor Essential Revisions:**

1. Difference in population should not be a confounding factor for biomarkers if they are true ones. Why do the authors believe this?

*Differences in populations should indeed not be a confounding factor. However, it is possible that since breast cancer is such a heterogeneous disease [1] different biomarkers are needed for the different breast cancer subtypes. When patient populations are different in composition of those (molecular) subtypes, they may be yielding different true biomarkers. Since molecular tumor characteristics have seldomly been described in proteomics studies, it is yet unknown whether different breast cancer subtypes are related to different proteins. Better description of the molecular tumor characteristics of patient populations in the literature and investigation of the relation with peak intensities in future studies may clarify this (see discussion on page 20). We changed study population for patient population*

2. Also, would this difference in the expression of these three markers relate to the disease stage?

*Breast lesions are believed to progress in a linear fashion through the sequential stages of the disease. The progression to malignant breast disease is associated with accumulation of an increasing number of genetic mutations, as well as changes in the expression of cell cycle-related and apoptosis-related proteins [2-5]. It is therefore likely that other proteins as well, are only expressed in a certain stage of the disease. Consequently, those proteins will only be good biomarkers for breast cancer in specific stages of the disease. Differences in stage of disease between patient populations then may lead to a difference in discriminative power of a protein between studies.*

3. Protein profiling: Detector attenuation of 100Da is too low, and will lead to matrix related peaks detection which might affect other peaks intensities. Did the authors try different intensities?

*Obviously detector attenuation of 100 Da is much to low. This had to be 1000 Da, it was a typographic error. We corrected this in the manuscript.*

4. What is the rationale if you excluded the peaks below 3000Da in your analysis?

*We noticed that in the spectra the noise from the matrix passed through up to 3,000 Da despite the detector attenuation. Since we did not want to include matrix related peaks in our analysis we excluded this region from the analysis. Furthermore, we were especially*
interested in the discriminative peaks reported by Li et al. [6] which were not in this region. Therefore it was possible to exclude this region. See methods, section protein profiling on page 8.

5. Also external calibration should have been undertaken with the Biorad ‘All in one’ Protein standard as this would cover the masses of interest. What standard peaks were used for calibration undertaken by the authors? This is important information needed for such a technical paper.

For the external calibration in our study we used the Bio-Rad ‘All in one’ standard peptide mixture, containing the peptides; vasopressin (1084.3 Da), somatostatin (1637.9 Da), dynorphin (2147.5 Da), ACTH 1-24 (2933.5 Da), bovine insulin beta-chain (3495.9 Da), human insulin (5807.7 Da) and recombinant hirudin (6963.5 Da). We added this information to the manuscript; see methods, section protein profiling, page 7. We used the ‘All in one’ standard peptide mixture since we were especially interested in the proteins reported by Li et al. [6], with masses of 4.3 kDa, 8.1 kDa and 8.9 kDa. The peptides in the peptide standard cover the low mass range we were interested in better than the proteins in the protein standard. Proteins in the ‘All-in-one’ protein standard except one (7.0 kDa) have a mass higher than 12 kDa, these would not cover sufficiently the mass region we were interested in.

6. To estimate the reproducibility of these duplicates the median coefficient of variance (CV) was calculated per m/z. What were your CV ranges?

The inter-quartile ranges of the CV's were 15-62% [median CV: 31%], 7-47% [median CV: 17%], 5-22% [median CV: 11%] and 5-23% [median CV: 13%] for m/z 4276, 4292, 8129 and 8941, respectively. This information was added to the manuscript; see results, section peak detection on page 13.

7. Did you use reference sample between the runs?

We did not use a reference sample between the runs. It is known that analyzing samples in different runs on different days introduces day-to-day variation [7-9]. To exclude the effect of any differences caused by day of sample preparation we adjusted for this using a logistic regression analysis. We categorized the subjects by tertiles of peak intensity separately for samples prepared on day 1 and day 2. Afterwards, the subjects in the same category in the two groups were combined. See methods, section data analysis, on page 9.

8. Data analysis: Did the authors investigate if there were any obvious sub grouping in the unsupervised clustering methods?

Since our focus was from the beginning on validating the proteins discovered by Li et al. [6], we did not perform a unsupervised clustering method.

9. Were the samples prepared manually?
Yes, the samples were prepared manually.

10. Why are there 2 missing values in the menopausal status? Similarly, there are 6 missing values in p53 status.

Information on menopausal status was obtained through examination of the medical records. However, for two women (aged 48 and 58 years) the medical record did not provide this information. P53 expression was determined by pathological examination. For 6 subjects the p53 status was not reported in the pathological report in their medical record. These subjects were diagnosed in The Netherlands Cancer Institute (NKI), however their surgery was performed in another hospital. In those hospitals, the p53 status of the tumor was not determined. See results, section study population on page 11.

11. Results: This reviewer believes it is a relevant part of the study to know the menopausal status of the controls.

We agree with the reviewer that information on the menopausal status of the controls is relevant. Unfortunately, this information was not collected at inclusion and could not be retrieved afterwards. We made a rough approximation of the menopausal status by regarding controls under the age of 50 years as premenopausal and those aged 50 years or older as postmenopausal. None of the four peaks was related with this variable in the controls. Although menopausal status of the controls was unknown we frequency matched the cases and controls for age and adjusted the relation between breast cancer and the four peaks for age as well.

12. 31% is quite a high CV for the peak 4276Da. Any reasons for it to be so high?

The CV for the 4276 Da peak is indeed higher than the CV for the other peaks. This may be caused by differences in median intensity of these peaks. Peaks with low intensities usually have higher CVs than peaks with high intensities since reproducibly measuring low intensities is more difficult. Furthermore, beside the 4276 Da peak we also detected its oxidized form. It may be possible that this oxidation process (occurring in the tube of the SELDI-TOF) was not equal between duplicates. However, the CV for the 4292 Da peak (the oxidized protein) was not as high as for the 4276 Da peak (17%). We do not have any other plausible explanation for this higher CV.

13. In terms of peak intensity using SELDI, generally it is the signal to noise (S/N) that should be considered when differentiating cases from controls. It would be more relevant to refer to S/N between the different peaks on different days –Table 2-, as well as other data provided.

We assume that the Reviewer meant that this would be a method to correct for day of sample preparation. However, by categorizing the subjects according to tertiles of peak intensity per day of sample preparation, in the logistic regression analysis we were
already able to correct for this. The results of the logistic regression analysis were similar to the results of the Mann-Whitney u test, regarding the discriminative power of the individual peaks. This shows that the discriminative power of the peaks was not explained by the variation caused by day of sample preparation.

14. A multivariate analysis to build a classification module between cases and controls would add strength to the results.

We performed a backward logistic regression analysis in which we included simultaneously the four peaks representing the peaks reported by Li et al. [6] (continuous). Peaks were removed from the model if they did not statistically significantly contribute to the discrimination of cases and controls. M/z 8129 was removed from the model. M/z 4276, m/z 4292 and m/z 8941 all contributed statistically significantly to the discrimination of cases and controls. We added this information to the manuscript (see results, section relationships between peak intensities and the presence of breast cancer on page 14).

15. Protein identity: Identification of the 4.3 and 8.1 peaks is crucial to confirm their identification and not just by mass correlation. Have the authors attempted this?

We did not only base the identity of the proteins on similarity in mass but also on the condition under which these proteins were detected. For our study we analyzed the same matrix (serum) on the same array type with the same protocol as was used for the detection of the proteins identified in the validation study by Li et al.[10]. By doing this we did not only select those proteins with the same mass, but also those binding under the same conditions to the chip, indicating similar pI. We added this information to the manuscript (see methods, section data analysis on page 8). We have not pursued direct identification on m/z 8129, since we did not find this protein to be discriminative between case and controls. For m/z 4276 and m/z 4292 we had also other indications to assume that they represent the 4.3 kDa ITIH4 fragment and its oxidized form. We found three other peaks in our study which had masses similar to the masses of three other ITIH4 fragments previously described by Song et al. [11] and Villanueva et al. [12]. The intensities of these peaks were highly correlated with the intensities of m/z 4276 and m/z 4292. This high correlation in intensity is only expected when these proteins originated from the same protein, in this case ITIH4. See results, section protein identity on page 15 and 16.

16. Discussion: Were any of the other differentially expressed proteins found in the other studies?

In the first study by Li et al. [6] these three proteins were the best discriminative proteins. They do not describe any other proteins to be discriminative between breast cancer and healthy controls. The aim of the study of Mathelin et al. [13] was to validate the proteins found in the study by Li et al. [6]. They do not show any information on other proteins detected in their study either. In the validation study by Li et al. [10] they also only focussed on the three proteins previously reported in the first study by Li et al. [6]. In our
study we choose to report the masses of the other discriminative proteins as well. This 
was done to be able to compare results with future studies.

17. The text needs editing towards the end and possibly explaining the future 
work in this field.

We made the discussion more clear by adding the part about storage temperature and 
storage duration to the part about collecting samples of cases and controls in the same 
timeframe. By doing this we also removed some overlapping parts. We also added some 
recommendations one should pay attention to in future studies. See discussion on page 
18, 19 and 20.

Discretionary Revisions:
1. The authors concluded that the M/z 4276 and 4292 represent the same 
protein, were they present in the same spectra? The authors speculate on 
oxidised forms of ITIH4 – did you try and prove this. How do you explain the 
mass difference?

We conclude that m/z 4276 is the same fragment of ITIH4 as detected in the first study 
by Li et al. [6] and that m/z 4292 is this same fragment of ITIH4, but now its oxidized 
form. We assume this because the mass difference between these two peaks, which is 
16 Da, is the exact mass of an oxygen-atom. Moreover, the intensities of these proteins 
are highly correlated (Pearson R² = 0.834; p<0.001) (they are present in about the same 
ratio in every spectrum).

Beside, we found three other peaks in this study which masses (m/z 3156, m/z 3270 and 
m/z 3965) highly corresponded with the theoretical masses of three other fragments of 
ITIH4 previously described by Villanueva et al. [11] and Song et al. [12] (3158 Da, 3274 
Da and 3972 Da). The intensities of two of these supposed ITIH4 fragments found in our 
study (m/z 3270 and m/z 3965) were highly correlated with the intensities of m/z 4276 
and m/z 4292 (m/z 3270 and m/z 4276: R² = 0.812, m/z 3270 and m/z 4292: R² = 0.722, 
m/z 3965 and m/z 4276: R² = 0.821 and m/z 3965 and m/z 4292: R² = 0.756 [p<0.001 for 
all]). This high correlation in intensity is only expected when these proteins originated 
from the same protein, in this case ITIH4. We added this information about correlation to 
the manuscript. See results, section protein identity on page 15 and 16.

2. In the methods, was -30 the same storage temperature in Li’s original study?

No, the storage temperature of the samples analysed in the first study by Li et al. was 
-80 °C [6]. The storage temperature of the samples analysed by Mathelin et al. was -20 
°C [13] en -30 °C for the samples analysed in the validation study by Li et al. [10]. See 
discussion on page 19.

Reviewer: Dalibor Valik

Minor comments:
As the authors point out significant preanalytical differences have been observed in SELDI TOF studies. Furthermore, peaks they four significant are located within a suspicious mass range 3500 – 5000 amu. Since they very probably reflect some kind of concomitant epiphenomenon the authors may also have looked at different blood-derived material such as various kinds of plasma to assess the relevance of their results. Since this is not possible in the retrospective study, authors may choose to discuss this point and may consider planning future prospective work using not only serum as probably – in my opinion – the least suitable material. We have previously shown that clot-activator collecting devices may significantly contribute to the low-molecular peak spectrum using this technology (Valik et al., Clin Chem. 2006). In line with this authors shall describe and discuss specimen sampling in more particular – specifically, what kind of collection devices were used and subsequent possible preanalytical influence thereof.

We read the article by Valik et al. in Clinical Chemistry with much interest. The samples analysed in our study were collected with BD Vacutainer® SST plastic serum tubes with clot activator and gel (Becton-Dickinson, Franklin Lakes, NJ, USA). We added this information to the manuscript. See methods, section study population on page 6. Our samples were allowed to clot for 30 minutes. Although it might be possible that the low mass peaks found in our study are the result of epiphenomena, we would like to point out that samples of both cases and controls analyzed in our study were collected and pre-analytically treated in exact the same way. Since we can assume that changes in protein profiles will occur independently from disease status (i.e. case or control) any changes in the protein profiles caused by the sample handling are unlikely to have caused the differences in protein expression we observed between cases and controls. Furthermore, the aim of this study was to validate the difference in expression of the three proteins found by Li et al. in serum. Since the matrix which is used in a study mainly determines which part of the proteome can be investigated, we had to use the same matrix to be able to detect the same proteins.

Furthermore, I would suggest to replace the words “up-regulated and down-regulated” using clear words “increased and decreased”, respectively, as being more relevant for serum, plasma or blood-utilizing studies and do not confound medical readers with terms originating from more experimentally structured, yet not always clear, molecular biology and cell biology studies. Logically – here the relation between breast cancer and i.e., ITIH4 or C3adesArg#8 is not “up/downregulation” implying causation but merely increase/decrease.

We agree with the reviewer on this and replaced the words “up-regulated and down-regulated” for “increased and decreased” throughout the manuscript.

References


