Author's response to reviews

Title: Qualitative and Quantitative Proteomic Analysis of Nipple Aspirate Fluid from Women with Early-Stage Breast Cancer using Isotope-coded Affinity Tags and Tandem Mass Spectrometry

Authors:

Timothy M Pawlik (tpawlik1@jhmi.edu)
David H Hawke (dhawke@mdanderson.org)
Yanna Liu (yliu@mdanderson.org)
Savitri Krishnamurthy (skrishna@mdanderson.org)
Herbert Fritsche (hfritsche@mdanderson.org)
Kelly K Hunt (khunt@mdanderson.org)
Henry M Kuerer (hkuerer@mdanderson.org)

Version: 2 Date: 4 January 2006

Author's response to reviews: see over
January 4, 2006

Peter Newmark, MD
Editor-in-Chief
BMC-Series Title

RE: REVISED MANUSCRIPT MS: 6144244858247860
Qualitative and Quantitative Proteomic Analysis of Nipple Aspirate Fluid from Women with Early-Stage Breast Cancer using Isotope-coded Affinity Tags and Tandem Mass Spectrometry

Dear Dr. Newmark:

We would like to thank you and the reviewers for your careful review of our manuscript. We believe that the comments provided valuable suggestions for revising and strengthening the manuscript. We have specifically addressed each comment below and have revised the manuscript accordingly.

Reviewer 1:
General Comments
This paper is of significant interest because breast cancer still imposes significant healthcare burden and there is still no sensitive non-invasive method for early breast cancer detection. Identification of biomarkers with sufficient sensitivity for early detection of breast cancer remains a priority for many biomedical researchers.

Specific Comments

1. Under the Methods section in the abstract it will be useful to include the number of donors used in the discovery/exploratory phase and the number used for validation by western analysis. As requested, this information has been added to the abstract Methods section.

2. In the Methods section additional information is needed. For example, on page 7 paragraph 1, the volume and type of buffer into which the supernatant was suspended should be indicated. Also, from the last line of page 7 to the first line of page 8 “...two kinds of wash buffer...” are mentioned. Readers will want to know which buffer(s) were used to be able to repeat the experiment.

More detailed information on the types of buffers used in the experiments has been added to the Methods section on pages 7 and 8.
3. In the Results section, on page 10, the authors report that “In three cases, the signal from either the tumor bearing NAF or disease-free NAF could not be detected and a ratio could not be computed”. The meaning and the relevance of this sentence is not clear. This sentence has been accordingly revised to read: “In three cases, a peptide signal could only be convincingly detected in one, not both states (diseased versus not), so no ratio could be calculated.”

4. In the Results section, on page 11 (western blot analysis section), the sentence starting from line 4, “On western analysis, vitamin D – Binding protein expression ...” is not consistent with the data presented as Figure 5. The data suggests that Vit D-binding protein expression is higher in cancer than in volunteers, whereas the sentence suggests that it is 2X higher in the volunteers compared to the cancer.

The Reviewer is correct. An error was made in the writing of this sentence. The sentence has been revised to reflect the data in Figure 4, which shows that Vitamin-D binding protein (VDBP) expression was overexpressed in patients with early-stage breast cancer compared with healthy volunteers.

5. In Table 1, what is the significance of the column heading “Score”. The significance of this ICAT program probability score should be explained as a footnote to the Table for the benefit of readers that are not familiar with the terminology.

Scores are as reported by the ProICAT program. The score represents a “quality of match” assessment much like a Mascot score, or Xcorr from Sequest. In general, higher scores are associated with a more “quality” or “better” match. In addition to the scoring provided by the ProICAT program, we manually curated all spectra to confirm the matches. A footnote has been added to Table 1 with this information.

6. It is unclear to me why the 3 proteins with a ratio of 999 are included.

As noted above, in three cases, a peptide signal could only be convincingly detected in one, not both states (diseased versus not), so no ratio could be calculated. The “999” was an artifact of the software. We have revised the table so that “N/A” appears instead.

7. The relevance of figure 2 is unclear. Is there any significance in finding out that there was about a 50:50 distribution in the number of over-expressed and under-expressed expressed proteins in NAF?

The purpose of Figure 2 was to emphasize that very few proteins were deregulated, and the detected proteins have a fairly normal distribution (as expected).

8. Based on the Text and Figure legends, the figure titled as Figure 4 should be Figure 3B and that titled as Figure 5 should be Figure 4.

The Reviewer is correct. Although the Figures were “uploaded” as Figures 3B and Figure 4, the electronic submission process has erroneously re-named the figures as Figure 4 and Figure 5. The electronic submission process does not appear to allow sub-sets of figures (i.e. Fig 3a and Fig 3b) to be labeled separately.

Reviewer 2:
General Comments
In this manuscript the authors address the identification of biomarkers for breast cancer by the application of the ICAT methodology in the protein analysis of nipple aspirate fluid (NAF). Overall, this is an interesting study demonstrating the application of quantitative proteomic approaches in the investigation of the NAF proteome.
Specific Comments
1. In the list of proteins presented in Table 1 several were identified based on 1 peptide. The criteria and specifically, number of mis cleavage/s allowed, ppm accuracy or MS and MS/MS tolerance, and allowed modifications should be provided. In addition the method of calibration should be described. It is recommended that the molecular weight as well as the % coverage of the identified proteins be provided.

The ICAT experiment by design produced fewer peptides, but the requirement of containing Cys increases the confidence of the one peptide hits when manually curated (as these were). The human subset of the NCBI database was searched using ProICAT allowing 1 missed cleavage (trypsin) and 0.3 and 0.5d windows for MS/MS and MS resp. Modifications allowed are fairly typical: ICAT on Cys, Met oxidation. We respectfully disagree with the Reviewer regarding the importance of providing the molecular weights and % coverage. The molecular weights can be easily obtained from the NCBI database and would add little to the discussion. The % coverage is also not particularly informative because of the restriction on peptides to contain cysteine (especially compared with peptide-mass-fingerprinting). For the one-peptide hit of interest, VDBP, a second peptide, not containing Cys was identified using Mascot (data not shown). This information has been added to the Results section on page 11. Calibration: QSTAR was externally calibrated, using the doubly and triply charged peaks of the peptide neurotensin. This information was added to the Methods section on page 8.

2. Some of the ratios described in Table 1 (for example 1.01, 1.05, 0.92) appear to be not indicative of differential expression. Usually in these methodologies, a cut-off of 15-20% change is utilized for differential expression. The authors should elaborate on the variability of the methodology and provide the criteria they used to assemble the list of differential expressed proteins.

Table 1 is a list of proteins identified with reasonable certainty in order of relative strength of assignment, not a list of differentially expressed proteins. We agree with the Reviewer that about a 20% cut-off has been reported with standards using stable isotope methods. In the current study, VDBP is up about 80%. In the current study, a peptide with a ratio of 1.56 was about 1 standard deviation above the median ratio.

3. In the background the sentence “This is mainly because MS is not an inherently quantitative technique and does not allow for the specific identification of individual peptides” should be rewritten. The first part of the sentence is true (i.e. MS is not quantitative) but the second part is not-MS allows for protein/peptide identification.

This sentence specifically refers to MS profiling experiments such as SELDI in which MS-1 only is performed, not MS/MS experiments wherein fragmentation data is acquired. For clarification purposes, we have revised this sentence on page 4.

Reviewer 3:
General Comments
None

Specific Comments
1. The authors should perform a reciprocal labeling to validate the list of differential proteins. This could confirm if the same list of proteins showed up even in reciprocal labeling.

We agree with the Reviewer that a general approach to validate labeling strategies would be reciprocal labeling. However, the samples are no longer available and to obtain fresh samples and process these samples is not feasible at the current time. It is important to note that for the peptide hit of interest (VDBP), we identified another (not Cys-containing) peptide by Mascot, and also confirmed the presence of VDBP by Western blot analysis.
2. No statistical analysis on the ratio of heavy:light peaks was shown, therefore, one might question if a difference with less than two-fold in expression was statistically significant. As noted by Reviewer 2, a standard cut-off of 15-20% change is utilized for differential expression. In the current experiments, a peptide with a ratio of 1.56 was about 1 standard deviation (SD) above the median ratio. This information has been included in the Statistics section in the Methods on page 9. We rarely see expression ratios much higher than 2 SD in serum experiments. Of note, the peptide hit of interest (VDBP) was up about 80%.

3. The authors should show results from a validation immunoassay that correlated with ICAT finding in couple markers.
We agree that this may be interest as a follow-up experiment. However, we believe that we conclusively identified a peptide hit of interest (VDBP) using three different experimental approaches (Cys-containing ICAT, not Cys-containing peptide by Mascot, and also confirmed the presence of VDBP by WB). As the samples are no longer available, performing an immunoassay would require obtaining new NAF samples, which we are not prepared to do at this time.

4. Number of singleton peaks was not listed. They might be the better markers to use with bigger difference in expression than the marginal two-fold in the listed markers.
Some of the proteins at the bottom of the list in Table 1 were effectively “singletons”. In our hands, however, “singletons” were quite rare. As indicated, we did not observe proteins that changed by > 2x.

5. Proteins at the bottom of table 1 had low ProCAT scores and may represent false positive identification.
We agree with the Reviewer that the proteins at the bottom of the list in Table 1 are clearly the least secure matches, which is why they are on the bottom of that table. The point of the Table is not to emphasize the protein matches at the bottom of the list, even though these identifications were checked with manual curation. Rather, given that the ProCAT scores are presented alongside the protein identifications, we wanted to present the data in aggregate for the readers’ own interpretation. A cautionary sentence to this effect has been included in the discussion on page 12.

6. The authors need to give an account on the omission of other known cysteine-containing breast cancer proteins from nipple aspirate such as i) PSA and ii) gross cystic disease fluid protein 15.
These proteins were not detected in this analysis. Many other proteins (EGFR, PI3K etc.) also were not. That is why we did not discuss them.

7. Was it optimized by scanning only the two most intense ions in MS/MS analysis?
As noted on page 8 of the Methods section, data was acquired selecting the two most intense ions from each survey spectrum, subject to the previous target list and minimum intensity constraints.

8. Page 11, the statement that “menopausal status did not affect the expression of vit-D binding protein…” was not valid.
We have removed the sentence referring to menopausal status and expression of vit-D binding protein.

9. The legend of Figure 5 was missing.
There is no figure 5. Although the Figures were “uploaded” as Figures 3B and Figure 4, the electronic submission process has erroneously re-named the figures as Figure 4 and Figure 5. The electronic submission process does not appear to allow sub-sets of figures (i.e. Fig 3a and Fig 3b) to be labeled separately.

9. There were three proteins in Table 1 in which the ratio was 999. Did it mean a singleton peak or were they typos? 
In three cases, a peptide signal could only be convincingly detected in one, not both states (diseased versus not), so no ratio could be calculated. The “999” was an artifact of the software. We have revised the table so that “N/A” now appears instead.

Thank you in advance for your time in considering our revised manuscript.

Sincerely,

Timothy M. Pawlik, MD, MPH
Assistant Professor
Department of Surgery
Johns Hopkins Hospital

Henry M. Kuerer, MD, PhD
Director, Breast Surgical Oncology Training Program
Associate Professor of Surgical Oncology
The University of Texas M. D. Anderson Cancer Center