Author’s response to reviews

Title: Differential Proteomic Comparison of Breast Cancer Secretome Using a Quantitative Paired Analysis Workflow

Authors:

Giselle Brunoro (gisellebrunoro@gmail.com)
Paulo Carvalho (paulo@pcarvalho.com)
Valmir Barbosa (valmir@cos.ufrj.br)
Dante Pagnoncelli (dantepagnoncelli@gmail.com)
Claudia De Moura-Gallo (cgallo8@yahoo.com.br)
Jonas Perales (jperales@ioc.fiocruz.br)
René Zahedi (rene@proteincetner.com)
Richard Valente (rhv4u@ioc.fiocruz.br)
Ana Gisele Neves-Ferreira (anag@ioc.fiocruz.br)

Version: 1 Date: 21 Oct 2018

Author’s response to reviews:

Dear Dr. Rice,

We are glad to know that our manuscript is potentially publishable in BMC Cancer. We thank you and both reviewers for this opportunity and for carefully revising our work. Please find reflected below a point by point response to the suggestions and criticisms that were raised. Furthermore, the PatternLab for Proteomics computer environment released a new build (v. 4.1) and we found it best to review our results with this latest version of the software. All updates to the manuscript are marked with track changes throughout the text. Additionally, all supplemental material and figures were updated as well.
Reviewer 1

This manuscript describes a proteomic study of paired nipple aspirate fluid samples from patients with unilateral breast cancer. The study is focused on the breast cancer secretome that contains proteins enriched in the tumor microenvironment. A novel bioinformatics software, PatternLab, is used to assess peptide level differences with disease state. 88 proteins are identified as being differentially abundant and the biological origins of these proteins is described. Overall the paper is interesting and the strategy to use nipple aspirate fluid and perform pairwise studies to identify breast cancer biomarkers is attractive and highly complementary to other biomarker studies in the literature. I would recommend that the following points be addressed:

1. The paper focus of the paper is slightly convoluted because a new bioinformatics tool is described and the tool is applied to the nipple aspirate fluid analysis. Is this the first time that PatternLab has been described in the literature? If so, it would be good to further qualify the tool. For example, the information obtained by traditional, more standard/established, proteomics analysis tools such as Proteome Discoverer should be compared to the information obtained from Pattern Lab.

PatternLab for Proteomics is a widely adopted general purpose proteomics platform that has been around for about 10 years. Its latest protocol was published in 2016 on Nature Protocols [1] and there is a reference in our manuscript. PatternLab has been extensively compared to other tools and, according to a study from the Chinese Academy of Science, was shown to be the one reporting the least amount of errors [2]. In this manuscript, we used PatternLab to perform identification and quantitation; we then described a new module that is able to perform the paired analysis. We compared this new approach to a standard t-test approach and showed improvements in sensitivity.

2. There are many points in the sample collection that need further clarification. Although the NAF collection has been previously described, a brief description of how the samples were collected should be given here. It says that the eligibility criteria included that all subjects were "to be on menopause." Does that mean the subjects were currently menopausal? Given the ages, this is probably not the case. Perhaps this means that the subjects were post-menopausal? In Table 1, "Birth control use" needs to be clarified. Does this mean they every used birth control? What kind of birth control? The same comments apply for hormone replacement. Also in Table 1, how is breastfeeding defined? Is there a minimum duration of breast feeding? Lastly in Table
1. why were samples collected from the three individuals without breast cancer? Did they have some other benign lesion?

The reviewer made important queries that deserve further clarification in the manuscript. To better describe the sample collection, we added to section “2.1. Sample collection”, the following sentence: “Briefly, the breast was gently massaged from the chest wall toward the nipple for 5 min followed by warm compress for equal time. The nipple fluid was then aspirated using breast pumps and the fluid droplets were collected using a 10 µL micropipetter (Gilson, Inc., Middleton, WI, USA). NAF samples were immediately diluted (10-fold with phosphate buffered saline pH 7.4) and centrifuged at 250 xg for 10 min at 6 ºC, and the supernatant was collected and stored at -80 ºC. NAF protein concentration was determined using the bicinchoninic acid protein assay kit (Sigma-Aldrich, St. Louis, MO, USA)”. Indeed, the patients were all post-menopausal and the motivation for choosing these samples was to avoid variations due to the presence of ovarian hormones. The text regarding eligibility criteria was amended. The data presented in “Birth control use/hormonal replacement” column from Table 1, depicts the answer that each individual gave regarding whether she used or not oral contraceptives/hormone replacement pills. The questionnaire did not contemplate the barrier contraceptives (e.g. female or male condoms, cervical caps, diaphragms, intrauterin device) possibly used over reproductive years of life, nor the patch, topical or vaginal hormone replacement therapy after menopause. Additionally, the type of therapy was not informed given that patients usually had a hard time remembering the name of the drug. A description was included as a footnote in Table 1. The information regarding breastfeeding was informed by the patient and reports if the patient breastfed, or not, the baby for at least 30 consecutive days (Table 1 was modified accordingly). Finally, the samples from both breasts of 3 three individuals without breast cancer, and no clinical or image indication of benign or malignant breast lesion, were also collected to perform a differential analysis between samples from right and left breasts. The motivation for this additional analysis was to evaluate the chance of false positive identifications since, in principle, NAF samples originating from healthy right and left breasts are subjected to the same physiological body control, and there should be no reason for having differentially abundant proteins between these paired samples.

3. In the sample preparation, how was protein extracted from the nipple aspirate fluid?

The fluid droplets were collected, immediately diluted 10 times in phosphate buffered saline, and centrifuged to collect only the supernatant, devoid of exfoliated cellular content (separated from the cell pellet). Only the supernatant proteins were used for analysis (no further extraction was needed).
4. The differentially abundant proteins between cancer and non-disease breasts included 87/88 proteins that were not in the Plasma Proteome Database. These proteins look like a proteomic profile of plasma. Are the nipple aspirate samples bloody? If so, this would be a huge confounding factor in the analysis. Also, the results that that several immunoglobulins are not included in the Plasma Proteome Database. That seems very strange given that immunoglobulins are a major component of plasma proteins.

Qualitative similarities between NAF and plasma are well known; hence, we can understand why the reviewer perceived “a proteomic profile of plasma” in our work. Some examples are the presence of albumin, immunoglobulin classes, zinc-α2-glycoprotein, transferrin, and prolactin-induced protein [3, 4]. However, the dynamic ranges for NAF and plasma samples are distinct. Immunoglobulins correspond to 30% of NAF total protein content, posing as one of its most abundant protein [5]. Additionally, according to SDS-PAGE analysis, Sánchez and colleagues proposed the categorization of NAF by the prevalence of the presence of zinc-α2-glycoprotein, apolipoprotein D, prolactin-induced protein, lactoferrin, α-lactalbumin, and lysozyme C, which are predominant proteins in NAF, showing a particular relative composition [6]. NAF can be translucid or colored (white, yellow, brown, and black). Darker-colored NAFs were previously related to tobacco use and high concentration of cholesterol and lipids whereas lighter-colored ones were associated with lactose [7, 8]. As for the NAF samples analyzed in this study, the majority were translucid or white-colored, and only one presented bloody exudate characteristics. In our sample comparison and statistical setup, for a protein to be found as differentially abundant, the quantification data had to be consistently measured across samples of distinct patients. Since the identification of proteins typically identified in plasma is expected, the literature has recurrently matched NAF data to the Plasma Proteome Database [4, 9]. Finally, it was indeed unexpected that some immunoglobulins identified in this work were not present in the Plasma Proteome Database, even though they have associated protein entries in the Uniprot database. However, since these entries are classified as “automatically annotated and not reviewed” in Uniprot, they might still not be eligible for inclusion in the Plasma Proteome Database.

5. The validation, of even a single protein, by an orthogonal method such as Western Blot in cancer/normal nipple aspirate fluids would really boost the confidence in the findings in this paper.

We have followed the reviewer's recommendation and performed an SRM experiment, which is now described in the new section entitled “2.6 Selected Reaction Monitoring (SRM)”. We assayed for increased abundancy of 4 glycolytic enzymes (pyruvate kinase, glyceraldehyde-3-
phosphate dehydrogenase, triosephosphate isomerase, and fructose-bisphosphate aldolase A), 4 proteins from the complement cascade (complement C5, complement C3, complement factor B, and complement factor H), and 4 proteins involved with platelet activation and signaling (alpha-2-macroglobulin, apolipoprotein A-I, fibronectin, and annexin A5). Nine of the selected targets had their previously observed fold-changes corroborated by this orthogonal method, even though only three of them (pyruvate kinase, alpha-2-macroglobulin, and complement factor B) also displayed a p-value < 0.05, following statistical analysis by Paired Analyzer. Finally, the data for the last three targets selected (fibronectin, glyceraldehyde-3-phosphate dehydrogenase, and triosephosphate isomerase) were discarded due to technical issues of unclear origin [signal below limit of quantification (LOQ), peptide co-elution etc.].

Reviewer 2

The work presented is another computational pipeline approach to examine differences in NAF between matched tumor-bearing and non-tumor-bearing breasts in patients. The samples collected offer a unique way to look at how NAF changes in breast cancer patients. The use of NAF to study breast health and breast cancer has been studied since the 90's and has yet to be developed into a routine clinical diagnostic. Doing so would have a significant impact in the field. Whether this approach will lead to a robust diagnostic for detection of early presence of biomarkers is unclear. Given that previous studies have robustly shown signatures such as uPA + PAI-1 + TF have a near 100% accuracy in separating pre-cancerous and cancerous tissues in the challenging setting of patient heterogeneity, the impact of this work may be limited. Furthermore, whilst new potential markers were identified computationally in this manuscript, no follow-up orthogonal validation has been undertaken to confirm the reliability of the computational approach.

Indeed, NAF samples have been used to study breast diseases for almost four decades. However, one of the aims of our work was to pinpoint alterations that may contribute to a better understanding of the disease; additionally, due to the complexity of this proteome, it is our understanding that we can push beyond a single minimal set of discriminants (as the one established based on tissue sample analysis). Another important contribution of this study is the description of an original paired proteomic pipeline that could be applied to any other paired biological samples, therefore increasing the power of the statistical analysis. Here, we have clearly demonstrated that the paired analysis of breast secretome with a tailored statistical analysis could overcome the individual heterogeneity challenge. As we have already pointed out
to Reviewer 1, we have applied SRM to 12 of the potential markers described in this work and have orthogonally validated 3 of them. Finally, we are aware that further analysis with a different/bigger sample set will be necessary before proposing a new panel of diagnostic proteins, especially for the early detection of the disease.

REQUESTED REVISIONS:

Modification searches in the proteomic pipeline only include carbamidomethylation of cysteine and oxidation of methionine. Given the majority of the factors in NAF are secreted, they should include other common extracellular modifications such as glycosylation, and hydroxylation of prolines, etc. to increase coverage.

We have chosen to limit the number of variable modifications in our searches since they could amount not only to an unreasonable increase in computational search space but also in higher false-positive identifications. Furthermore, specifically for glycosylation, the number of mass difference values that would account for all permutations of glycan residues/structures would require a dedicated glycoproteomic approach, which was not within the scope of this work. On the other hand, we acknowledge that substantial NAF protein glycosylation has been reported [10, 11] and we do agree with the reviewer that the study of extracellular amino acid modifications of NAF proteins should be pursued. However, this is part of an ongoing work in our group, where a dedicated sample preparation for NAF glycoproteomic studies is being developed.

Whilst the authors contextualise some of the more novel components they find in the text, additional orthogonal validation of their presence is required to validate the accuracy of the approach.

We have obliged to both reviewers recommendation for an orthogonal validation. To that effect, SRM experiments for 12 protein targets were done: 3 were confidently validated; 6 were “border-line” validations, and 3 could not be detected/validated. For further details, we invite the reviewer to read both the reply to the 5th question, raised by the 1st reviewer, and the updated version of this manuscript.


