Reviewer's report

Title: Differential Expression of Cancer Associated Proteins in Breast Milk Based on Age at First Full Term Pregnancy and Length of Nursing

Version: 1 Date: 21 November 2011

Reviewer: Dana Bovbjerg

Reviewer's report:

1. Is the question posed by the authors well defined? The authors are to be commended for investigating an important and understudied area - the mechanisms underlying the positive association between the age of women at the time of their first full term pregnancy (FFTP) and their risk of subsequently developing breast cancer. Their hypothesis is very reasonable (quoting the abstract): “…that proteins linked to breast cancer would be differentially expressed in human milk collected at three time points during lactation based on: 1) age at FFTP, and 2) length of time that a woman nurses.” However, the second predictor is confusing. The authors should simply indicate the specifics of the three time points in the hypothesis, since it is the sample collected at weaning that may be predicted by the variable duration of lactation preceding it.

2. Are the methods appropriate and well described? The methods are generally appropriate and well described, with two important exceptions that require major compulsory revisions. First, citations are not provided for each of the different protein biomarker assays. The absence of such citations is particularly problematic for the assessment of KLKs, which were not based on commercially available kits. These are straightforward assays, but if this is the first description of these methods in the literature, more detail is required so that the study could be replicated. For example: What were the sources of each of the monoclonal antibodies? What were the sources of the proteins used to generate the polyclonal antibodies? Which KLK assays used which methods? Second, the statistics are inappropriate. The initial analyses (Tables 2 &3) were conducted separately within each of the FFTP age groups, but addressing the central question of the paper requires a comparison between the two age groups, which requires having FFTP as a variable in analyses that include all participants in a single model with FFTP age group as a predictor. The decision to dichotomize the FFTP age variable is also not well justified here, and the choice of age ranges made by the authors actually excludes several older participants from the initial analyses. These individuals are included in later analyses (see figures) making it impossible to compare across the different results since the samples are not the same. The problem of different samples in the analyses also arises for the T-test comparisons of the pairs of time points. Differences between the proteins that show significant effects of FFTP age for comparisons between BL and two months, while not showing such effects of BL and W could simply be due to the different participant mix in the two analyses (note the large differences in the sample sizes for those two analyses in Table 3). Finally, the analyses also
need to examine the parity status of participants as a predictor. That is, the authors need to address the possibility that women who have not previously given birth may show a different pattern of responses in their milk proteins, or this variable may interact with FFTP age (i.e., the effect of FFTP age on protein levels may differ based on parity status).

3. Are the data sound? The assays used to assess protein levels appear to be sound, but it would strengthen the paper to compare the levels found here to previous reports to the extent possible.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition? Yes, except as noted above.

5. Are the discussion and conclusions well balanced and adequately supported by the data? No. See the concerns above regarding data analyses, which make interpretation of findings problematic.

6. Are limitations of the work clearly stated? The authors do note several limitations, but should also emphasize the small sample sizes for several of the follow-up analyses (e.g., methylation data). Sample sizes should be reported for all tables and figures.

7. Do the authors clearly acknowledge any work upon which they are building? With the exceptions noted above, yes.

8. Do the title and abstract accurately convey what has been found? Yes, with the exceptions noted above.

9. Is the writing acceptable? Overall the manuscript is well written.

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

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests.