Reviewer's report

Title: Defining the genomic signature of the parous breast

Version: 1 Date: 21 March 2012

Reviewer: Alain Nepveu

Reviewer's report:

In this study, the authors identify 300 or so probesets that are differentially expressed between parous and nulliparous post-menopausal women. The underlying hypothesis is that some of these genes may provide the molecular basis for the documented lower breast cancer risk associated with early full-term pregnancy. Gene ontology analysis was performed to identify biological processes that are over-represented among differentially regulated mRNAs. The most interesting finding from this analysis was that genes involved in RNA splicing and processing are over-represented among differentially expressed genes. In the discussion, the authors speculate on the mechanisms by which the mRNA processing "reactome" induced by early pregnancy could protect against breast cancer.

This is an interesting study, although the biological relevance of the data cannot be verified. Whether any of the differentially expressed mRNAs has anything to do with the protection against breast cancer is not clear. This is a limitation that is intrinsic to this type of study. Yet, we cannot exclude that future studies might be able to establish a mechanistic link between some of these genes and breast cancer.

Major Compulsory Revisions

I think the manuscript could be improved by dealing with a number of issues. In particular, there is a lack of information regarding the statistical methods that were used. Some of this information may be provided in prior publications from the same group, but it is sometimes not possible from this manuscript to assess the solidity of some statements. For example, for the heatmap presented in Fig. 1, there should be a chi-square analysis to establish that the two groups are indeed different.

Also, additional analysis should be presented (probably in supplemental data) to enable the reader to assess the significance of the data presented. I would have liked to see a heatmap where the two groups are separated. The image thus generated would give an idea of the homogeneity, or lack thereof, of gene expression within each groups.

Also, as a control, one should present a random sampling experiment to assess the average number of differentially expressed genes between any two groups chosen at random. The number of genes interrogated in the microarray is not given, but assuming a total of 30,000 mRNAs, what we observe in Fig. 1 is that
approximately 1% of mRNAs are differentially expressed. Would we observe differential expression of 1% of mRNAs between any 2 groups chosen at random? And would the overrepresented biological functions be the same?

What is the age of the women who contributed samples? There may be a greater difference between women of 55 and 80 years old than between parous and nulliparous women.

What type of Multiple Testing Correction was performed on the data to obtain 305 differentially expressed probesets (208 distinct genes).

What is the Adj. P value?

Table S2 and text on page 6: "Comparison between P and GN revealed that 12 genes (18 probes) were differentially expressed (Table S2)." In Table S2, the p values are low, but the adj. p values are all 0.41. Is this a typo?

Much of the data is presented in Supplemental Data comprised of several Word files. It would make it easier for the readers if all Word files were converted into one single PDF file that can be downloaded at once.

**Level of interest:** An article of limited interest

**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