Reviewer's report

Title: Gene expression profiles of breast biopsies from healthy women identify a group with claudin-low features

Version: 1 Date: 12 August 2011

Reviewer: Lance David Miller

Reviewer's report:

This work explores the transcriptional programming of normal breast tissue and seeks to characterize the natural molecular variation inherent to the human breast. The work is driven largely by unsupervised hierarchical cluster analysis of normal breast gene profiles and biological inferences drawn from gene set/ontology enrichment analyses and statistical associations between cluster organization and gene signature-based biological classification.

The work is of potential importance to the field of breast cancer, as a better understanding of the transcriptional wiring of normal breast tissue, and particularly the inherent variation therein, could lead to new insights into: mechanisms of breast carcinogenesis, the tumor microenvironment, and predictive models of breast cancer susceptibility.

Of particular interest and potential clinical relevance are the observations that some breast tissue samples show enrichment of certain stem cell- or progenitor cell-like properties, exhibit strong similarity to the poor outcome-associated, claudin-low breast cancer subtype, and display molecular properties that seem to be significantly associated with parity.

Major Compulsory Revisions:

1) The Cluster 1 samples were found to be highly associated with the claudin-low breast cancer subtype based on the centroid classification method of Prat, et al, 2010. Similarly, the authors investigated the relationship between the normal breast specimens and the other intrinsic breast cancer subtypes yet the results have not been presented. How the other intrinsic subtypes relate to the cluster variation within Cluster 2 could provide important insights into the underlying variation. These data are relevant and should be presented (at least in supplement).

2) The presentation of gene ontology enrichment results (eg, Figure 2 and Figure S1) should include p-values (after correction for false discovery) and the number of genes comprising the enriched fraction.

3) The malignancy risk predictor of Chen, et al (2009) associates with Cluster 1 with high statistical significance. This implies (and is elaborated in the discussion section) that the Cluster 1 patients might have a reduced risk of developing breast cancer. However, the cluster 1 samples also resemble the claudin-low, poor-outcome associated form of breast cancer. These concepts are at odds and should be reconciled in the Discussion.
4) The Discussion and Conclusions sections do not clearly articulate the authors’ interpretations and conclusions. For example, the messages in the last four paragraphs of the Discussion are obscure; their purpose seems only to describe the limitations of the study rather than offer concrete interpretations. Both sections are in need of better organization and clarity of thought. The Conclusions section should focus less on restating the results and more on what the findings may mean.

Minor Essential Revisions:
1) In the Results section titled, “Unsupervised hierarchical clustering”, what other methods and gene filtering parameters were investigated? These should be enumerated – perhaps in the Methods section.

Discretionary Revisions:
1) In the Abstract, the methods section is vague, and there is a run-on problem with the second sentence of the Abstract’s results section.
2) In the Discussion section, the sentence beginning, “The breasts of nulliparous breasts...” should be re-worded.
3) In the Discussion section, the sentence beginning, “This study does not give enough power to conclude, and the association between...” should be re-worded.
4) In the Discussion section, the sentence containing, “...different biopsies from whole one tumor share similar...” should be re-worded.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests