Reviewer's report

**Title:** Association of High Obesity with PAM50 Breast Cancer Intrinsic Subtypes and Gene Expression

**Version:** 3  
**Date:** 23 November 2014

**Reviewer:** Maria Elena Martinez

**Reviewer's report:**

In this cross-sectional study, investigators assessed the role of obesity and breast tumor gene expression and intrinsic subtypes. Analyses were conducted combining two populations: LACE and Pathways studies. Body mass index derived using self-reported weight and height “around the time of diagnosis”, which was defined as 21 months prior to and two months after diagnosis. BMI was then categorized into 5 categories ranging from underweight (<18.5) to highly obese (35+). The PAM50 assay was used to obtain gene expression levels and the intrinsic breast tumor subtypes.

Strengths of the study include the well characterized populations and unique study question, among others. However, there are several weaknesses, some already acknowledged by the investigators, which limit the interpretation of the results. These are outlined below:

**Major Compulsatory Revisions**

A substantial concern relates to the limited sample size for some of the main and several of the stratified analyses presented in the manuscript. It is difficult to appreciate the degree of precision in the tables since no data on cell sizes are presented. For some tables, the reader has to do a bit of calculation to derive at these; in others, it is not possible to derive these, at least to this reader (i.e., no. of pre- and post-menopausal cases for each intrinsic type, related to Table 4). This reviewer is left with difficulty interpreting results for the following groups:

1. Using table 1 to derive sample sizes, it appears that findings for highly obese women appear to be based on a total of 183 cases. This number is then split to take into account the gene expression and intrinsic subtypes. These are then further split to stratify by menopausal status.

2. Pre- and post-menopausal differences for just about all comparisons, although specifically problematic for those in table 4. For example, it appears that the data in this table for basal like associations in pre-menopausal women based on about 40 cases total, distributed among 5 BMI categories. A similar problem occurs for HER2 positive cases.

3. It is nearly impossible to say anything about the underweight women. For example, in table 4, it appears that the number of basal-like underweight women category is only 2 cases. How can analyses be conducted stratifying by menopausal status? Perhaps my figures are incorrect, which is why more data are needed in the tables.
4. Given these sample size limitations, perhaps the authors should consider streamlining their analytic approach to results where precision is not a major issue.

Minor Essential Revisions

1. The pooling of two studies that have key differences is potentially problematic. One recruited early-stage cases 15 or so years ago from a mixed Utah population (some from the cancer registry and some from KPNC) of patients enrolled within 2 years of diagnosis. The second recruited women more recently with presumably has no stage eligibility criteria (although this is not specifically noted) and enrolled within 2 months of diagnosis; importantly, this study comprised of a diverse racial/ethnic population. Given these differences and the aims of the study, investigators need to provide justification for combining the studies. At the very least, a variable for study site should be included in the multivariate models and sensitivity analyses should be conducted to see whether major findings apply to both study populations.

2. Please provide more data in the tables so that the reader can easily get a sense of the cell sizes.

3. A factor not related to sample size per se is the classification of Asian women into the BMI categories using the same cut-points as those for non-Asian women. Can the authors comment on the appropriateness of this approach?

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I have no conflicts of interest.