Author's response to reviews

Title: Integration of Transcript Expression, Copy Number and LOH Analysis of Infiltrating Ductal Carcinoma of the Breast.

Authors:

Lesleyann Hawthorn (lhawthorn@mcg.edu)
Jesse Luce (jesse.luce@roswellpark.org)
Leighton Stein (leighton.stein@roswellpark.org)
Jenniffer Rothschild (jenniffer.rothschild@roswellpark.org)

Version: 4 Date: 9 July 2010

Author's response to reviews:

Dear Editors, I have submitted the data to GEO and it has received an accession number cited on page 11. I have corrected the sentence suggested by reviewer 2 highlighted on page 12. I am resubmitting figure 4a and 4b in a zipped format. If it will not upload please provide an email where I can send it. Thank you very much for all your efforts

Reviewer 1

Major compulsory revisions

1. It appears that the number of samples used is confusing. We have altered the abstract to state that 14 samples were used in the integrated analysis of copy number and gene expression. This is highlighted in the abstract, page 2. Table 1 also shows which samples were analyzed on each of the different platforms.

2. The control samples were obtained from 4 normal adjacent tissues of the patients in the study. Again, in table 1 the patient numbers reflect which control tissues match which tumor samples. In order to clarify this further I have added a sentence highlighted on page 5 detailing that the control samples are matched normal tissues adjacent to the tumors.

3. Reviewer 1 seems to have several concerns regarding Figure 3 and refers to the “spikiness”. We can only assume that has failed to understand the accompanying figure legend. The length of the bars indicates the number of samples which show the actual loss or gain and NOT the actual copy number. To remedy this we have replaced figure 3 to show actual mean copy numbers.

4. We are aware that the number of samples in our study is relatively low but many copy number analysis papers have been published using fewer samples including several by Reviewer 1 which analyzed 4 cell lines (PMID: 19863778). Another by this reviewer assay a group of oral cancers using an n=5 (PMID: 19633365) and a recent publication describing HPV-associated chromosomal alterations in squamous cell carcinomas of cervix and head and neck with n=10
5. Reviewer 1 states that a pool of normal samples should have been used that were generated by the same lab and from the same sample types and not a Partek reference for copy number analysis. We are uncertain what this comment refers to. It is clearly stated in the Materials and Methods section under Copy Number Data Analysis that “the baseline used was generated from the 250K Mapping 270 HapMap panel” This is a set of 270 samples is a common baseline for use in CGH studies as it involves several different races and represents a large normal population (PMID: 19322034, ). We agree that an optimal situation would involve the analysis of tumor/normal pairs but unfortunately, we did not have matching normal tissue for the majority of our samples. There is also some concern in the literature about using normal adjacent tumors due to field effect cancerization which has been documented in breast tissues.

6. The CNAs that we report here are not CNVs. We have carefully checked our data against those contained the CNV database (http://projects.tcag.ca/variation/ng42m_cnv.php). In many cases CNVRs map into our larger regions of copy number alterations. In one instance, on chromosome 4p16.1, an amplified region overlaps with a reported CNVR and we have added a paragraph to describe this instance highlighted on page 9.

7. The reviewer suggests that we have used rudimentary calling algorithms to detect copy number alterations and suggests we use more sophisticated and published software. As described in the Materials and Methods section, we have used the Genomic Segmentation Algorithm available in PARTEK. This algorithm is similar to HMM segmentation but instead of searching for regions from a specified list of states, the Genomic Segmentation Algorithm finds breakpoints in the data. This is a commonly used algorithm for the detection of CNAs (PMID: 20084286, PMID: 20048075, PMID: 20142589, PMID: 1930695, PMID: 19701073, PMID: 19755429, PMID: 19436329). The cutoffs were generated after processing the data using the Genomic Segmentation algorithm. This algorithm has several advantages over HMM. Segmentation does not require binning. Cancer derived samples are likely to contain different populations of cells, which may not display the same copy number variations. For this reason, copy number often will not fall into biological predicted bins and occasionally becomes a continuous variable. Segmentation looks for changes in genomic abundance, not regions of a specific copy number state, enabling segmentation to be highly effective in cases where tissue heterogeneity can lead to non-integer copy number intensities. The Genomic Segmentation Algorithm does not bin the regions into predefined states; instead regions will be called with a mean at any copy number state with no redefined normal bin filtered out. The different segments are then regions of local stability of copy number and each region is compared to the expected normal value and assigned a likelihood of being a CAN.

8. After segmenting the genome of a sample into different regions of copy number, the algorithm compares the region to a normal copy number with two
one-sided t-tests and reports the significance of the CNA. These p-values are used to filter out regions of change that are rare or due to noise. Noise is significant in copy number data so that the algorithm does not consider normal at a diploid number of 2 but instead is considered a range around 2. We used a test for significance as 2.3 for amplifications and 1.7 as a cutoff for deletions.

9. The reviewer suggests that we make a distinction between copy number gain and amplification. Since we have included the mean copy number of each event in supplemental table 1 we have changed any CNA event with a mean copy number of >4.5 an amplification and anything below that number is defined as a gain. The results of our analysis are highlighted on page 8. We have also incorporated this information into the overlay of LOH and copy number this information is also included in table 2

Minor Essential Revisions

1. We have cut the description of integrative genomics in the introduction section and added appropriate references. This adjustment is highlighted on page 4.

2. The discussion has been rewritten and only 2-3 genes are discussed.

3. We have removed this remark—the edited paragraph is highlighted on page 3

Reviewer 2

Minor Essential Revisions

1. We apologize for the quality of these images but the BMC website restricted upload sizes for images to 20MB.

2. We will resubmit these using an alternative method. We will resubmit the uncropped versions of the figures, again we were attempting to decrease the size.

3. Page numbers have been added to the manuscript.

4. The description of RMA has been corrected See highlighted text on page 5, we also added GC correction to the pre-processing of the data.

5. We used 10 probesets as a minimum number of SNPs showing the CNA. With the 250K array the mean inter-marker distance is ~10 Kb so we are identifying fragments of 100Kb or larger. This means that we will not detect any microdeletions or amplifications. We have added a statement to this effect on page 6 (highlighted).

6. We have reset our cutoff values at 5 samples for copy number and 3 for LOH. With this adjustment, our data becomes very similar to other published data.
7. The het rate has been adjusted from >0.07 to <0.07
8. We have rewritten the paragraph and it is highlighted on page 7
9. We have corrected the sentence, highlighted on page 7
10. The sentence was removed.
11. Corrections are highlighted on page 7.
12. Corrections are highlighted on page 7
13. Corrections are highlighted on page 7
14. This section has been re-written
15. Corrections made
16. The LOH discussion has been re-written
17. Since this is difficult to accomplish in PARTEK, we have included the raw data in the form of normalized gene expression values in Supplemental table 5. The tumors with the CN were then compared to these expression values to make sure that they did show corresponding increases or decreases with the CN gain or loss. This is highlighted on page 12
18. We are in the process of submitting the data to GEO
19. We have changed our wording in the abstract (page three) and conclusion (page 17)

Major Compulsory Revisions

1. We had originally thought that because our sample numbers were low, the most significant contribution would be to present those findings that were very frequent. A few of the reviewers were not comfortable with this approach, therefore we reanalyzed the data and set our cutoffs much lower (5/22 samples for copy number alterations and 3/22 for LOH events)

2. We have changed the cutoff rates for copy numbers as follows: homozygous deletions (<0.5), loss (<1.7) gain (>2.3) and amplifications (>4.5).

3. We ran our data through PennCNV as suggested see page 7 and 11 highlighted. Comparing the data with that generated by Partek did not show significant differences except that the PennCNV calls were fewer. To make sure that our LOH data was reliable we visually examined the allele ratio plots for each event and made calls based on the appearance of the plots in neighboring regions. We have included the PennCNV data as a Supplemental table (3).

Reviewer 3
Major Compulsory Revisions

1. We have clarified the number of samples analyzed using the different platforms. See highlighted region in abstract and materials and methods. Table 1 shows which samples were analyzed using the different platforms.

2. We have presented information in the materials and methods and in table one regarding the sources of control tissues (see highlighted region, page 4).

3. We agree with this reviewer that an optimal approach would have been to have matched normal samples for these analyses, however this was a legacy set of tumors and we did not have matched normal samples. Many studies have been published that use the HAPMAP samples as a normal baseline for LOH studies. PMID: 17564968, PMID: 18664255, PMID: 17564968

4. We detected a lower number of regions of copy number loss because we reported only those regions occurring in 50% or more of the samples. This approach led this reviewer to believe that there is systematic bias in the data. This is not the case and the bias was introduced by our requirement events in a large percentage of the samples. To rectify this, we have reduced our cutoff values to 5/22 tumors instead of 11. We have replaced figure 3 and Supplemental table one to show that deletions are at least as common as amplifications but do not occur in as many tumors. This finding is also echoed by Andre et al, 2009 where the authors reported the region showing the most frequent amplifications ranged from 20-56% while the most frequent regions of loss ranged from 20-31%.

5. We have rewritten this section, please see highlighted region on the reanalysis of the data we do have some regions showing copy number loss and LOH, however, copy neutral loss is a much more common event in cancer PMID: 17564968.

Minor Essential Revisions

1. We have resubmitted all the figures at higher resolution.

2. The text has been proof read.