Title: Exome sequencing of primary breast cancers with paired metastatic lesions reveals metastasis-enriched mutations in the A-kinase anchoring protein family (AKAPs)

Version: 0 Date: 28 Nov 2016

Reviewer: Loris De Cecco

Reviewer's report:

The manuscript from Kjallquist et al describes a pattern of somatic mutations in different members of A-kinase anchoring protein family enriched in metastases compared to the corresponding paired primary tumors in breast cancer. Their findings are supported by primary/metastasis specimens from an initial case material of 10 patients (cohort1) and confirmed by a series of 20 patients (cohort2). Although, as stated by the authors, cohort2 is from FFPE specimens and the experimental procedures along with the pipelines in the analysis differ from the cohort1, the work holds the potential to advance defining the events leading to the metastasis spread in breast cancer. As a matter of fact, mutations convergent toward the same gene family members or genes sharing the same function are an appealing mechanism as demonstrated by MR Stratton.

Comments:

A limited sample size could drive misleading conclusions; however it is worth mentioning the "control experiment". With the aim to avoid technical biases, good quality genomic DNA from whole blood was subjected to WGA, library constructions, sequencing and compared to the unamplified DNA.

Here, the main concern is that the "control experiment" is an ideal condition using material of too good quality and not reflecting the real experimental setting. The assessment of false discovery rate starting from such an ideal condition could cause overconfidence. In fact, as reported in Table2a/2b most of mutations are G>A. This reviewer is wondering if the authors could explain an high G>A occurrence.

DNA derives from fine needle biopsies (metastases), surgical specimens (primary tumors) and blood (germline control): could this be a systematic source of variability? This reviewer suggest to specify some technical details (DIN from TapeStation, amount of DNA recovered after extraction, starting material for WGA and recovery after WGA) just to be sure no systematic biases are present. The genotype calling metrics in supplementary material and methods confirms an high Ti/Tv. This reviewer strongly suggests to better discuss this point adding the appropriate references.
Pag 7 row 2 and row 17 in "detection of mutations from whole-exome sequencing data". It unclear the filtering >5 variant reads. Are those variants removed? Why did you set the threshold of 5 reads?

The authors claims that the significant enrichment of AKAP mutations are confirmed in both cohorts (pag12, row5-8). Considering that in cohort 2, two (out of 4) variants are tolerated by SIFT prediction, the statement should be rephrased.

Gene expression data from TCGA proved the existence of coexpression clusters among member of AKAP family (AKAP1/3/7/8 and AKAP5/9/10/11/12) related to PAM50 intrinsic subtypes. To this review it not clear how this stratification of AKAP family in two classes supports the mutational findings.

Minor points:

In the exome sequencing of cohort 2, do you know the percentage of duplicates as assessed by Picard?

Pag11 row 15: could you provide the exact p by Fisher's test?

Fig1e at first glance is misleading. The number of mutations are reported for met and pt but TCGA is reported as %. Even if the studies differ largely by sample size, could the authors report the data in fig1e in the same way?

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

No

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

Not relevant to this manuscript

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?
6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license ([http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/)). I understand that any comments
which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal