Reviewer's report

**Title:** Relationship between circulating tumor cells and epithelial to mesenchymal transition in early breast cancer.

**Version:** 3 **Date:** 25 March 2015

**Reviewer:** Nadia Dandachi

**Reviewer's report:**

In this manuscript, Mego and colleagues correlated the presence of CTCs with the expression of EMT transcription factors in tumor tissues of breast cancer patients. Overall, CTCs were detected in 24% of patients with 12% expressing only epithelial markers, 15% expressing only EMT markers and 2% expressing both epithelial and EMT markers. They found no correlation between EMT transcription factors TWIST and SLUG in breast tumors and presence of CTCs suggesting that EMT markers do not play a major role in CTC release.

The authors have addressed an interesting topic, however technical details on CTC enrichment are not clear and need to be addressed together with some minor issues:

1) In the methods section, the authors state that they used CD45 depletion for CTC enrichment as previously described (Line 133). In the paragraph “RNA extraction” (Line 135) they state that they used RosetteSep kit to perform CD45 depletion. However, the paper cited used first EPCam-depleted PB followed by Ficoll-hypaque. Then they performed CD45 depletion using magnetic beads coated with CD45. So it is not clear, which technique was used for the CD45 depletion in the present manuscript. For clarification please describe in more detail which technique was used in this paper.

2) In our own experience RosetteSep has a low sensitivity compared to other CTC enrichment techniques. Do the authors have any data on sensitivity of this method? The major problem with this technique is that CD45 depleted cells do not necessarily contain only CTCs. So, concluding that EMT expression defines the presence of CTC with EMT phenotype is critical.

In this same context, CTC definition (LINE 162) is not clear. Again the definition of CTCs with EMT phenotype is questionable. The authors need to address this issue and discuss it.

3) Line 175

The authors write that two pathologist performed review of slides, but three initials are listed.

4) Highest expression levels of CK19 and EMT markers are not informative. Please add median and range.
5) In the result section, the paragraph “CTC detection” is not clear, at times even confusing and therefore needs revision. An additional table with analyzed genes and results might be helpful to better understand results.

More specifically:
Line 242-244. The number of genes listed is four, but results are presented for 5 genes?
Line 244-247. Numbers presented in this section do not correspond with numbers in the abstract.
Line 247: What do the authors mean with the term overlap in expression of EMT markers?

6) In the discussion section, the authors describe the limitations of their study. However, they also should include the fact that mRNA and protein levels do not always correlated (e.g. post translational modifications)

7) In the tables, it would be helpful for the reader to include which method was used for presented data (protein versus mRNA data)

8) For IHC staining, please add information on which controls were used (negative and positive).

9) The authors mention in the discussion section that intratumoral heterogeneity could have influenced mRNA expression results. In this context, did the authors notice any staining differences of EMT markers when using IHC that could support this assumption?

10) There are some spelling errors that need to be checked. For example, Line 288, 310, 318, etc.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**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