Reviewer's report

Title: L1CAM is preferentially expressed in triple-negative breast cancers and is inversely correlated with Androgen receptor

Version: 3
Date: 10 October 2014
Reviewer: Hugo Aria-Pulido

Reviewer’s report:

The manuscript “L1CAM is preferentially expressed in triple-negative breast cancers and is inversely correlated with Androgen receptor” by Doberstein et al. reports the detection of L1 and AR by IHC in two cohorts of breast cancer samples with supplementary work performed using published data sets to examine the L1 and AR expression levels in breast tumors, and additional functional work on 6 cancer cell lines to report that L1CAM is under AR control, and concluding that L1 may be a diagnostic tool and that L1CAM expression ‘could be causally related to the bad prognosis of TNBCs’.

While the data are certainly of clinical interest, there are a few issues that need to be carefully addressed.

Major Compulsory Revisions

1. The data focused on TNBC, but there is not exact information about the characterization or classification of samples as TNBC. For example, the supplemental table shows ER and PR but nothing about HER2 and, most important, TNBC samples.

2. The Conclusion about L1 diagnosis and prognosis needs strong support by running a multivariate analysis to determine whether L1 stands alone as an independent prognostic factor among the classical known prognostic risk factors or not.

Minor Essential Revisions:

In general, the description in Materials and Methods is a bit vague without complete/exact information about number of samples and cell lines analyzed by each technique (IHC, WB, qRTPCR, and FACS). If the Authors want to save space, please add that info as Supplementary Data to avoid losing critical data. Similar problems are in the Results section (see below). I think that making too many subdivision (basal, TNBC, and non-basal TNBC) is confusing and diverts the attention of the ‘take home’ message, and open more question, for example, would addition of L1 IHC expression to the 5 biomarkers (ref. 14) improve better the classification and prognosis of basal-like tumors and/or TNBC?

Page 5, line 13: ‘these features’. What features you refer to?

Page 6, line 9: Is L1 clinically established? As far as I know, it is not. This sentence needs clarification.
Page 7, line 10: Only the Innsbruck cases were possibly TNBC (because there is not data on HER2 for both data sets. Please provide that info here as well as in the Suppl. Table).

Page 7, line 9-down: No information on ER, PR, HER2, L1 IHC (antibody concentrations, time and temperature, etc.) for IHC were provided as well as for scoring. What is the idea of the E-cadherin work for this work?

Page 9, line 8: Please provide detailed information of primers here or as supplementary data with indication of the primer location per genebank reference sequence.

Line 15: Stats need better and more detailed description.

Page 10, line 2-down: This paragraph needs extensive edit because is confusing as is. The text moves freely from mRNA to protein which is confusing. For example, 5% of 1002 is about 51 cases by mRNA, but what was the % of L1 at the protein level? Figs read protein not mRNA expression, what was represented here or can you add the mRNA expression in a similar way? What is the correlation with TN (avoid basal subdivision)?

Line 8: Suggestion: Change 'as expected' to 'in agreement' (to avoid biasing the data).

Line 12: Define how you got the cut-off for this data set. Can you show the numbers (%) for each group? 1D gives the impression that there are more than 50 L1 positive cases and that's what you reported initially, 5% of 1002 cases or ~50 cases. Please clarify.

Line 13: Graph reading is different? KM graphs usually provide Survival probability. You have % DFS/OS and cumulative DFS/OS. Why such differences? Show # of pts at risk below the x-axis. Do the same for other KM curves elsewhere.

Line 15: Graph E representation is misleading because it looks like half (50%, not 5%) were dissimilar.

Line 18: TNBC vs. non-TNBC? Can you keep the data presentation as in Fig. 1A? Changing from ER,PR/HER2 types to LumA/B, Her2, TNBC with subdivisions of basal and non-basals TNBC is confusing and distracts/dilutes the data presentation.

Discretionary Revisions: Keep both the Introduction and Discussion short and to the point.

Level of interest: An article of importance in its field

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

Statistical review: Yes, and I have assessed the statistics in my report.

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