Reviewer’s report

Title: Regulatory network analysis reveals the oncogenesis roles of feed-forward loops and therapeutic target in T-cell acute lymphoblastic leukemia

Version: 0 Date: 03 Oct 2018

Reviewer: Federica Eduati

Reviewer's report:

Xia and colleagues investigate the importance of miRNA-mediated feed-forward loops (FLL) in T-cell acute lymphoblastic leukemia (T-ALL), focusing on differentially expressed mRNA and miRNA when comparing T-ALL and normal T cell samples. They identify core FLL involved in regulation of cell cycle and development of T-ALL and use this information to suggest potential therapeutic targets. The authors address an important topic, such as the investigation of the regulatory mechanisms underlying the development and treatment of T-ALL, however there are some major issues that should be addressed.

1a) I am not convinced of the robustness of the results, given that the paper is based on differential analysis of miRNA derived from the comparison of 48 T-ALL patient samples vs 2 healthy donors (for both CD34+ and CD4+CD8+ normal T-cell subsets), therefore with a very unbalanced sample size and with only 2 healthy samples. Additional computational analysis is required at least to show that differentially expressed miRNAs (DEMs) have comparable expression in the 2 healthy samples (to make sure that the differential expression is not driven by only one sample). Approaches like the one described by Zambelli et al. NAR, 2018 https://www.ncbi.nlm.nih.gov/pubmed/29390085 could also prove useful to understand if the results of the differential expression analysis are confirmed or if they are rather due to individual samples.

1b) Additionally, given availability of miRNA expression data for T-ALL cell lines (from the same publication of the 48 T-ALL patients and 2 healthy donors (18)), I am wondering why the author did not follow the same approach used to define differential gene expression (DEG), i.e. comparing T-ALL cell lines vs normal, T-ALL patient vs normal and looking at the intersection. In Results authors actually state that they used "T-ALL samples (patients and cell lines)“, where the cell lines data used, and if yes how?

2a) The authors claim that their network analysis is essential to identify feed-forward loops and therapeutic targets, however no validation is made of the role of the regulatory mechanisms of the feedforward loops, but they rather discuss the relevance of individual elements of the module (or of gene sets), which were identified simply by differential expression analysis (without need
for network interpretation, apart for the definition of the hub nodes). For example, the authors mention that downregulation of CDC25A induce cell cycle arrest and miR-21-5p knock down induces apoptosis, but what is the their role as part of the FOXM1-miR-21-5p-CDC25A FFL? This is a coherent FFL and this kind of module has been known to function as a surveillance mechanism, is there any evidence this might be the case for this particular FFL?

2b) Similarly for the interpretation of the association with drug response, the authors observe that CDC25A, core target gene of FFL FOXM1-miR-21-5p-CDC25A is negatively correlated with drugs inhibiting T-ALL (GSK-J4 and Teniposide), but what is the role of the FFL in this? For how it is currently presented in the paper (see also point 3a), the network information seems to have been used only to filter for highly connected nodes in the definition of the subnetwork used to study development of T-ALL. The network-based interpretation should be better explained/exploited.

3) The paper is missing important details in the Methods section, especially in the description of "Generation of network and analysis of hub components":

a) How is the network reconstructed? Reference to in-house scripts and previous publication is not sufficient, essential details to understand the nature of the inferred network should be provided here. For example, was the network derived purely from databases and prediction algorithms (as described in (13)) or did they also use expression correlation to filter the false-positive regulatory interactions (as described in (20))? Additionally, providing open source script to the community would be highly appreciated and would guarantee reproducibility of the results.

b) How was the subnetwork (Fig 4) derived and how was the pathway/GO term enrichment done? The description in methods "We selected top 5 TF and miRNA and combined with GO terms results to find core TFs and miRNAs" is not sufficient to understand and reproduce the author's findings.

c) Usage of GSCALite tool is also poorly described (e.g. was the Spearman correlation between gene expression and drug sensitivity computed pan-cancer or specifically for ALL?)

Other minor comments:

- p7 53-57: I would not highlight as a result that the "T-ALL and normal samples were segregated well by the expression level of TFs" in Fig. 2B, since this is an obvious consequence of the fact that only differentially expressed (between T-ALL and normal) TFs are shown here.
- **p8 7-11:** "Our network contained 486 edges, which consisted of 132 DEG and 12 miRNA", please rephrase with the current structure "which" seems to refer to "edges".

- **p8 11-19** not clear (probably because the corresponding methods section is not complete - see comment 3a) and need grammar check, e.g. "FOXM1 was only one TF" should be "FOXM1 was the only TF"; "SOX4 regulated about 66% genes" should be "SOX4 regulated about 66% of the genes". Similar grammar issues are also elsewhere in the manuscript.

- **p9 27** refers to Figure 6B, which has no panel marked with B (I understand that the authors refer to the bottom part of the figure).

- Results section "Potential drug indicators of genes in the regulatory network" could be expanded going deeper in the interpretation of the results. For example, how should positive vs a negative correlation be interpreted? Correlation seems to have been computed pan-cancer and based on the GDSC and CTRP gene expression data (not clear, see comment 3c). If this is the case, are positively and negatively regulated genes up- or down-regulated in the T-ALL dataset and can this tell us more about expected sensitivity/resistance to the drug? And can this be linked to positive/negative regulators identified in the network (to give an actual network interpretation of the results)?

In the same section (p10 15-19) the authors state "we found MCM2 was upregulated in the T-ALL samples and showed negative correlation with 39 drugs, which implied MCM2 may serve as a potential therapeutic target". My interpretation would be that if MCM2 is negatively correlated with drug A, then up regulated samples (i.e. T-ALL) are more sensitive to the drug (lower IC50), but it is not clear to me why MCM2 could be seen as a potential therapeutic target based on the above observation.

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

No

**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.

No
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.

I recommend additional statistical review

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

Needs some language corrections before being published

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/). 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