# Reviewer 1 Comments

**Q1: Please describe the new technology or code (or new application of a known technology or code) reported in this manuscript, and its use.**

The manuscript reports implementation of an R package to process gene expression data to enable translation and testing of a given hypothesis involving perturbation of gene pathways between two systems using in silico experiments.

**Q2: Please highlight the limitations and advantages.**

The manuscript focused on the implementation rather than carefully explaining the method that has been implemented. The manuscript is probably easy to understand by the researchers involved in this project, but it is difficult to be comprehended by an outsider. More details about the used method would be required to judge whether it is statistically valid approach, and what its limitations might be.

**Q3: Are there objective errors or fundamental flaws? If yes, please detail your concerns.**

The main idea to use SEM for hypothesis testing and adopt this approach to experiments involving perturbations of gene expression is sensible. However, more details are required to describe the method including all details about processing steps into the main manuscript, and I may suggest to move the implementation details to supplementary.

**Q5: Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List)**

The paper is difficult to understand as many important details are not given.

*Abstract*: please explain what inputs are necessary to perform the analysis, define or explain perturbation of gene expression pathways, explain what is meant by gene activities, explain what statistical significance refers to. Some statements used in 'Contributions to the field' may be added to Abstract, but there are again unclear statements: a basic SEM model (how does it look like?), relationships among end-points (what are these points?), how to briefly explain how the functional hypothesis can be generated?

*Introduction*: how are SEM models fitted? - a reference may be enough, why t-score can be used as activity metric? Implementing bootstrap random sampling is probably not that difficult. More importantly, there should be better literature survey outlined in Introduction, and also summary of contributions and advantages of the proposed method compared to other similar methods.

*Methods*: please add more details what has been implemented in SEMIP package or software, e.g. explain a 3-node fitting problem, it may help to add a paragraph describing what biochemical processes are considered, what type of data are assumed in the analysis, explain what is meant by system response was exemplified, the role of t-score in transferring knowledge between two stochastic systems is unclear (and this point seems to be critical for understanding the paper), how can bootstrap simulation eliminate unrelated gene signatures? Why running the bootstrap over 1000 samples is sufficient? Why not 100 or 10000? What is multicore hardware needed?

*Results*: line 152: sometimes ... help ... proposed new hypothesis - when does it help and when it does not? Why are the results provided in zipped file, are they so large? line 165: A Use Case of ...., line 190: ... out hypothesis ... it is unclear what is being referred to

*Discussion*: line 204: how different is your package from MplusAutomation? Are there any other similar R packages? What advantage your packages bring compared to these other software?

*Overall*:

1. Please add details focusing on the science behind rather than the implementation.
2. It helps enormously to explain what is being modeled, what type of data are assumed, what type of hypothesis can be assumed etc.
3. Add some numerical results demonstrating the statistical validity of the developed software.
4. Please proofread the paper for some occasional English writing errors.

# Reviewer 3 Comments

**Q1: Please describe the new technology or code (or new application of a known technology or code) reported in this manuscript, and its use.**

The authors developed an R Shiny application to aid in the transfer of perturbations in gene expression pathways from one system to another for determining casual inference of molecular interactions in silico. The authors used a 3-node PGR-GATA2-SOX17 gene network as a use case to evaluate the potential of using putative downstream genes of GATA2 as surrogate reporters of GATA2 activity.

**Q2: Please highlight the limitations and advantages.**

*Strength*: This manuscript presents a useful tool that can facilitate hypothesis generation and testing and allow bench scientists to perform analyses through a user-friendly interface.

*Limitations*: The content of the manuscript sometimes focus too much on procedural details and not enough on the purpose of the analysis, consideration of assumptions and interpretation of results. Schema illustrations also need improvement.

**Q3: Are there objective errors or fundamental flaws? If yes, please detail your concerns.**

no comment

**Q5: Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List)**

***Major comments:***

* Line 73-75: More details about the t-score should be added here, i.e. the assumptions and observed data for the t-test.

The main motivation using a “t-score” was to achieve the cross-species projection from a model animal (mice or rats) experiment to another species or human when a perturbation was not directly applicable (Wu, S.P. et al. 2015).  With a model animal (mice or rats) experiment, normally the animals are randomly assigned into two groups, where one group will receive “placebo” and/or no treatment and another group will receive the perturbation treatment. Experimental measurement will be properly collected from both groups (i.e. gene expression profile from a Microarray experiment) according to some thresholds followed by a statistical analysis (t-test). Significantly changed genes/probes (signatures) will be obtained from this analysis with directionality (up/down regulation). Such a group of genes/probes are deemed collectively as the “signature profiling” of biological responses to a particular perturbation. This finding and information will be projected into another animal system (i.e. human) of interest bearing the assumption that the experimental animal of interest would respond similarly if the perturbation were applied.

In a separate experiment of interest that is done with species of interest (i.e. human), the homologous genes of those previously identified genes/probes from the experimental species will be selected, and the experimental measurement from this study (GEO accession: GSE58144, (Koot et al. 2016))  will be used. The directionality information will be used to group these genes into two separate groups. A normal t-statistics will be calculated from these two groups of measurement to represent the responses in a new species and new experiment set up of interest. Since this calculation was originated from a standard t-test statistics and the term “T-score” was coined firstly by Wu, S.P. et al (Wu, S.P. et al. 2015) and have been widely used in other research projects.

* Line 96: Regarding the two bootstrapping methods, do they have different assumptions, computational costs and/or test power? Any recommendations for when either method is preferred?

***Minor comments:***

* Line 70: projects -> projected
* Line 73: Such an information -> Such information
* Line 262: Figure 1 is a little confusing. The green shape is not a rectangle but was referred to as one. Varying both shape and color without appropriate annotation is confusing/distracting. The red boxes, dashed vs solid, do they have different meaning?
* Supplementary Figure 1
  + - Why is SOX17 in brackets?
  + - The top two thicker blue arrows seem to indicate the same processing step, but the text annotations are different, which is confusing and distracting.
  + - Figure legend: “The resulting shrunken GATA2 gene list or reduced GATA2 [gene list] then restored by the same number of irrelevant genes are tested in the SEM model.”
* Source code: Coding style in the source code could use some standardization.