Dear Dr. Ren,

We thank you for the invitation to revise our manuscript and thank the Reviewers for their time and appreciate all their thoughtful critiques. We have made substantial changes in our manuscript based on the two Reviewers’ comments and suggestions. Our point-by-point responses to each of the Reviewers’ comments (highlighted in blue) are included in the interactive review forum as well as the rebuttal letter.

We hope that our revised manuscript is now suitable for publication in Frontiers in Genetics. Thank you again for all your help in the review of our manuscript and your consideration.

# Reviewer 1 Comments

Reply to reviewer 1:

We greatly appreciate the reviewer for the constructive comments and insightful suggestions. We have made substantial changes in the manuscript based on the comments and suggestions you provided and have now included our point-by-point responses in this response letter.

**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.

Thank you for your nice summary.

**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.

Thank you for pointing this out. We have made substantial changes and included the detailed explanations of T-Score, bootstrap approach, and SEM methods and results in the revised manuscript. The revised texts could be found in:

1. T-Score: Detailed description of the T-Score method in Lines 99 – 126, and explanation of T-Score results in Lines 204 - 236
2. SEM model: Detailed description of the SEM model and background in Lines 129 – 149, and results explanation in Lines 239 - 254
3. Bootstrap method: Detailed explanation about the bootstrap method in Lanes 152 – 162, and results explanation in Lines 257 - 272
4. The biological application of the SEMIPs method: Enhancement of overviews in Lanes 89 – 96 and Lines 181 – 201. A use case application in Lines 275 – 306
5. Additional discussion about the potential limitations of this App and the comparison between this App and MplusAutomation could be found in lines 314 - 328

**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.

As suggested by the reviewer, we rearranged the manuscript by adding more biology and result explanations in the main text and shifted the weight of implementation details to supplementary materials and methods sections. Since our focus is to present a user’s friendly interface for the non-bioinformatic oriented bench scientists to test their hypotheses, we have also included the more detailed description of the T-Score, bootstrapping and the SEM method, as well as additional discussion of the results in the main text.

**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?

As suggested, we have revised the abstract to include more specific information for better delivery of the concept to readers. These changes can be found in Lines 28 - 45.

*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.

The SEM model fitting is now included in Lines 131-141:

Further explanation of using the T-Score as an activity metric is now included in Lines 99 - 126:

The advantages of this SEMIPs application and its contribution to scientists who have limited bioinformatic background are stated in Lines 74 – 84.

*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?

The Methods section has been extensively revised to describe the tools that are implemented in the SEMIPs application. For example, the 3-node fitting system was designed for testing the relationship among 3 members of a gene regulation network (Lines 91 - 93). Additional descriptions on the 3-node model can be found in Lines 183 – 191.

T-Scores served as a quantitative surrogate of molecular activities of a gene in a given biological context where the actual molecular activities could not be measured directly. A description of T-Score calculation is included in Lines 99 – 117 and the T-Scores’ biological meaning is now described in Lines 204 – 225.

The revised text reads on additional explanation of bootstrap approach is included in Lines 257 – 262.

***Regarding the questions on “how can bootstrap simulation eliminate unrelated gene signatures and why running the bootstrap over 1000 samples is sufficient.”:*** Bootstrap simulation is primarily designed for hypothesis generation study on a known/defined downstream targets regulated by an upstream regulator or perturbation. The downstream targets often consist of a group of genes. To assess the significance of the upstream regulator, we attempt to remove those downstream targets from the original pool, which leads to altered SEM results. In each bootstrap round, it will randomly select “same number of gene signatures”, eliminate them from the original pool, and then proceed with the above-mentioned modeling steps till the end. Both bootstrap methods are non-parametric with no assumption of the population distribution; therefore, sufficiently large amount of simulation will provide us empirical distribution where can be consulted for statistics testing. It largely depends on the magnitude of the impact from the upstream regulator revealed from the SEM fitting when the “downstream genes targets” are eliminated. When the impact is minimal, fewer rounds of the bootstrap will be sufficient. Sometimes, a smoother fit can be reached when we reach to 1000 rounds as shown in this sample QQ plot (Figure X1).

Chart, line chart

Description automatically generated

Figure X1, A sample QQ plot of three bootstrap simulation results at 100, 500, and 1000 rounds.

We started with 100 rounds, assessed the fitting curve, and often found insufficient. It largely depends on how many downstream genes target will be eliminated, as we increase the bootstrap rounds, we get smooth fitting as shown in the following graph. In our exercise 1000 rounds ensures a stable empirical distribution curve.

It is mainly required for the bootstrap simulation steps to speed up the process. It uses two R packages, parallel and doParallel; it detects available computing cores in real time and requests half of the available cores to conduct the simulation job. Take the example we showed in the manuscript on KEGG pathway analysis with 28 categories, it can take up to a couple of hours to finish this step. Therefore we suggest a multicore hardware equipment. If no multicore is available, it will execute serialized process.

*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

A revised statement on the hypothesis generation application of SEMIPs is now included in lines 193 – 201.

The ”hypothesis” in question was originally described as “*we hypothesize that expression levels of GATA2’s direct downstream targets reflect its activities in silico.*” We have rephrased the sentence in question to improve readability (lines 298 -302):

*Response to “Why are the results provided in zipped file, are they so large?”*

The SEM fitting results include both model fitting statistics (SEMfitting.txt) and three-node image (SEMplot.png), therefore, they are put into a “zipped file”. These two are different file type, we choose to use a zipped file, which can be downloaded by the users then unzipped. It is not large, and only for the convenient purpose. We have modified the codes accordingly to address reviewer’s comment as well.

*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?

We have included the information in the discussion section in lines 319 - 328:

*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.

We appreciated the reviewer’s constructive comments. We have provided specific information based on the suggestions and extensively revised the abstract, introduction, methods, results, and discussion. In addition, we had our colleagues proofread the revised manuscript.

# Reviewer 3 Comments

Reply to reviewer 3:

We’d like to thank the reviewer for the constructive comments and insightful suggestions. We have made substantial changes in the manuscript based on the comments and suggestions you provided and have now included our point-by-point responses in this response letter.

**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.

Thank you for your the nice summary.

**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.

Thank you for pointing this out. We have made substantial changes and included the detailed explanations of T-Score, bootstrap approach, and SEM methods and results in the revised manuscript. The revised texts could be found in:

1. T-Score: Detailed description of the T-Score method in Lines 99 – 126, and explanation of T-Score results in Lines 204 - 236
2. SEM model: Detailed description of the SEM model and background in Lines 129 – 149, and results explanation in Lines 239 - 254
3. Bootstrap method: Detailed explanation about the bootstrap method in Lanes 152 – 162, and results explanation in Lines 257 - 272
4. The biological application of the SEMIPs method: Enhancement of overviews in Lanes 89 – 96 and Lines 181 – 201. A use case application in Lines 275 – 306
5. Additional discussion about the potential limitations of this App and comparison between this App and MplusAutomation could be found in lines 314 - 328

The Figure 1 and figure legend have been updated following the suggestion.

**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.

A description of T-Score calculation is included in Lines 99 – 117 and the T-Scores’ biological meaning is now described in Lines 204 – 225.

* 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?

Thank you for this suggestion. We have included a new session in the supplemental materials primarily address the detail implementation of two-class bootstrap simulations.

Response to “Do the two bootstrap methods have different assumptions?”

Both bootstrap methods are non-parametric so there are no assumptions regarding the distribution of the data. However, there is an assumption that the population is infinite, or sufficiently large such that that the effect of taking a sample is essentially negligible.

Response to “Do the two bootstrap methods have computational cost?”

Yes, there is a computational cost for each method. The more bootstraps are performed, the longer the analysis takes but the closer the estimated parameter is to the true value.

Response to “Do the two bootstrap methods test power?”

No. The bootstrap methods ascertain the significance of the test.

***Minor comments:***

* Line 70: projects -> projected

It is corrected.

* Line 73: Such an information -> Such information

It is corrected.

* 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?

As suggested, we have revised the Figure 1 and its legend to avoid the confusion.

* Supplementary Figure 1
  + - Why is SOX17 in brackets?

When the “brackets” are used, they refer to “gene expression value” instead of “projected activity”. In this model, we used expression of SOX17, therefore, it is put in the brackets (now Figure 3).

* + - The top two thicker blue arrows seem to indicate the same processing step, but the text annotations are different, which is confusing and distracting.

Thanks for pointing this out. We have explained the steps in the revised figure 3 and the revised figure legend

* + - 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.”

Correction was made in the Figure 3.

* Source code: Coding style in the source code could use some standardization.

This is an excellent suggestion. Here are some details on how we follow the coding standard and make the corresponding changes to address your recommendation.

Our development focuses on the user’s friendly “web-application”; therefore, we have followed the best practice recommendation by shinyapp development <http://shiny.rstudio.com/>. To address reviewer’s suggestion, we further consult the recommendations in Mastering Shiny by Hadley Wickham (2020 O’Reilly Media) and follow their best practice suggestions to achieve a better standard.

We use a separate helpers.R file to list all the dependency libraries for SEMIPs. The file will provide user a look up table and it will be invoked once the app is launched.

We use the github as our version control protocol. Currently we have a private repository at <https://github.com/NIEHS/SEMIPs> and plan to make it publicly available after the manuscript is published. All the modifications and future modification will be documented through github repository.

We have added additional comments in the code so that it would be easier for readers to follow. We add header in each file to provide project information, github repository location, author and other important information. After cleaning the code of the program, we roll out a new version on github.

To achieve the clean code standard, we keep all four main tabs separate for convenient code management, and use the same postfix to “ui” and “server” files. We used the “global variables” to keep track of the information that is needed for each separate UI.

For the web css style, we follow css standard and keep page theme and template in a separate folder.