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 time.

**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 now included the detailed explanations of T-score, bootstrap approach, and SEM method in the revised manuscript. The revised texts could be found in:

1. Lines XXX-YYY: Detailed description of the t-score.
2. Lines XXX-YYY: Detailed explanation about the bootstrap method.
3. Lines XXX-YYY: The biological application of the method
4. Lines XXX-YYY: Additional discussion about the potential limitations of this App and comparison between this App and MplusAutomation.

**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 have moved some implementation details to supplementary material and methods sections. 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 XXX-YYY:

*“Gene expression is controlled by multiple regulators and by the interaction among these factors’ activities. Data from genome-wide gene expression assays enable a mathematical estimation of molecular activities via a projection from the gene signature of a non-human system to gene expression profiles of human system using a T-score calculation. This transformation will help us understand the complex human system and have potential clinical applications. With the quantification of a given gene activity in each individual specimen, structural equation modeling (SEM) was able to determine the concurrent regulatory effects of two or more upstream regulators on levels or activities of a downstream reporter gene. Here we developed an R Shiny application, termed “Structural Equation Modeling of In silico Perturbations (SEMIPs)” to compute a two-sided t-statistic, or T-score as a surrogate gene activity in a given human specimens, which can be used in either correlation studies between outcome variables of interest or subsequent model fitting on multiple variables. This application implements a 3-node SEM model that consists of two upstream regulators as input variables and one downstream reporter as an outcome variable to examine the significance of interactions among them. SEMIPs enables scientists of non-bioinformatic background to examine the genetic interactions among the three variables in silico. As a case example for SEMIPs, we showed that putative direct downstream genes of the GATA2 transcription factor are sufficient to infer GATA2’s activities in silico for the conserved PGR-GATA2-SOX17 genetic network in the human uterine endometrium.”*

*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 XXX-YYY:

*“To determine the relations among multiple variables, structural equation modeling (SEM) is one of the statistical techniques to indicate the strength of influence among variables by getting an overall fit of model with existing data. The fit of the model can be assessed using various criteria, including the root mean square error of approximation (RMSEA), along with a 90% confidence interval, the Comparative Fit Index (CFI), the Tucker-Lewis Fit Index (TLI), and the standard root mean square residual (SRMR). For the RMSEA, the general rule of thumb is that values <.05 indicate close fit, values between .05 and .10 indicate marginal fit, and values >.10 indicate poor fit (MacCallum RC et al. Psychol Methods. 1996). For both the CFI and the TLI, a value of 1 indicates perfect fit, and the general rule of thumb is that values >.90 indicate adequate fit (Hu L et al. Psychol Methods. 1998, Hu L et al. Structural Equation Modeling. 1999). Also, SRMR values <.08 indicate a very good fit between the model and the data. Therefore, SEM offers a statistical framework to make casual inferences about the causality of multiple variables in a system.*”

Further explanation of using the t-score as an activity metric is now included in Lines XXX-YYY:

“*While gene expression data in public repositories provide a valuable resource for investigators to infer regulatory processes {Edgar, 2002 #9}, the causal relationships among variables of interest are not always directly measurable in a system. Moreover, it is challenging to test the knowledge obtained from experimental model systems in human due to undetermined clinical outcomes and ethical considerations. Genome-wide gene expression assays on human specimens allow observations of correlations among the gene expression levels as well as between RNA abundances and phenotypic outputs. Meanwhile, these assays can also determine the downstream targets of a factor of interest in model systems that are relevant to the particular type of human specimen via genetic or pharmacological perturbations. The resulting gene signature, manifested by the behavior of these downstream target genes in response to a perturbation, could unbiasly serve as a surrogate of the activity of the factor of interest in a given context. Assuming that gene functions are preserved between human tissues and relevant model systems, the degree of similarity between the gene signature of the factor of interest and the specimen’s gene expression profile could be quantitatively estimated by a T-score calculation to represent activities of the factor of interest in the targeted specimen [PMID: 25295534, 19666588, 19490893, 18757322]. This scoring system have been employed to establish correlations between the prognosis outcome and manifestation of activities of the factor of interest in corresponding tumors [PMID: 25295534, 23201680, 19666588, 19490893, 18757322]. The T-score calculation has also been utilized to determine the association among activities of factors of interest or between the activities of an upstream regulator and levels of its downstream targets within a set of human specimens [PMID: 26356605, 27783953]. Results of these studies demonstrated applications of such a surrogate score of molecular activities in investigation of gene functions and inference of regulatory processes.*”

The advantage of this SEMIPs application and its contribution to scientists who have limited bioinformatic background is stated in Lines XXX-YYY:

“*We were motivated to develop a Structural Equation Modeling of In silico Perturbations (SEMIPs) Shiny application to facilitate casual inference of gene regulatory processes, especially on multifactoral impacts on outcome variables concurrently. SEMIPs enables quantification of a projected activity metric (T-score) and allows users to fit desired SEM models using variables of interest. For hypothesis generation purpose, SEMIPs provides two different bootstrap random sampling procedures (elimination with or without replacement) to test the significance of a model after removing a subtest of downstream targets that are pertinent to pathways of interest in the gene signature (PMID: 18757322). Previously, the T-score system and SEM were applied to gene expression data to evaluate gene interactions that regulate the progesterone signaling pathway in the mouse uterus and inference of the gene regulation processes in human uterine specimens {Rubel, 2016 #1}. SEMIPs streamlines this process and allows scientists of limited bioinformatic background to perform computations and analyses through a user-friendly interface.*”

*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 XXX-YYY):

“*This app also provides a 3-node model fitting function using structural equation modeling to test the joint regulation of a target gene by two upstream regulators in silico.*”

Additional descriptions on the 3-node model can be found in Lines XXX-YYY:

“*SEMIPs is designed to test concurrently contributions of regulatory effects of two upstream regulators “Fac1” and “Fac2” on the expression of a downstream reporter gene “Endpoint”. Meanwhile, two-directional interactions between the two upstream regulators are also examined. Under this structure, users could test the relationships among the gene expression levels of all three variables. If a hypothesis is involved testing of molecular activities of two upstream regulators, gene signatures of the upstream regulators are first projected to a gene expression matrix of human specimens of interest (e.g., an expression dataset that are derived from human biopsies) through the T-score calculation function. The resulting T-scores will serve as the surrogate molecular activities to test for the manifestation of the proposed genetic network in human specimens via model fitting.*”

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. T-scores’ biological meaning is now described in Lines XXX-YYY:

“*The T-score was employed to project molecular activities of a gene of interest from a model system experiment to human specimens where a perturbation was not directly applicable [PMID: 25295534, 19666588, 19490893, 18757322]. In a model system, the biological replicates 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 measurements will be properly collected from both groups (i.e., gene expression profile from a genome wide gene expression experiment). Significantly changed genes/probes (signatures) will be obtained from this analysis according to pre-determined thresholds followed by a statistical analysis with directionality (up/down regulation). Such a list of genes/probes are deemed collectively as the “gene signature” of biological responses to a particular perturbation in a given context such as cell or tissue types of interest. And these downstream target genes of the perturbed molecule are referred as “signature genes” of the molecule of interest. This gene signature information will be projected into the human specimen of interest bearing the assumption that the biological behavior of the gene of interest is conserved between the chosen model system and the human specimens.*

*In the gene expression dataset (i.e., human) that molecular activities of the factor of interest on individual samples are to be estimated, the orthologs of the signature genes were first identified and grouped based on the directionality of the signature genes. The T-scores of individual samples in the dataset were calculated by a normal t-statistics from these two groups of measurements to derive a single number as a quantitative surrogate of molecular activities of interest. Specimens with T-scores larger than 0, which share a similar signature gene expression profile from the model system, were classified as having gene signature activities and vice versa.*”

A description of T-score calculation is included in Lines XXX-YYY:

“*The T-score calculation requires the input of two components, a normalized gene expression matrix of the human specimens and a gene signature of the factor of interest. To generate the normalized gene expression matrix of human tissues, such as microarray or RNAseq data, the expression values of each gene were centered to the median across all samples. If the gene had multiple probes or transcripts, the probe/transcript with the highest variation (standard deviation) was chosen to represent that gene. The gene signature was first determined by identifying downstream target genes whose RNA abundance are associated with the levels of the upstream regulator in a given set of statistical criteria. These associated downstream targets were further subgrouped based on the positive (up-regulated signature) or negative (down-regulated signature) correlations on the RNA abundance between the upstream regulator and the downstream targets. The T-score was then calculated based on the following formula:*

*Tscore = d\*TINV(p, df);*

*Where as,*

*d =1, if the average expressions of homologous genes of up-regulated signature genes is larger than the average expressions of homologous genes of down-regulated signature genes). Otherwise, d = -1.*

*TINV: the function of inverting t statistic.*

*p: p value of 2 tailed t-test of the expressions of homologous genes of up-regulated signature genes and the expressions of homologous genes of down-regulated signature genes with equal variance.*

*df: degree of freedom; total number of the homologous genes of signature genes minus 2.*”

The revised text reads as follows on additional explanation of bootstrap approach is included in Lines XXX-YYY:

*Biological signaling is often transduced by a cascade of downstream effectors in a hierarchical manner. The gene signature of an upstream regulator is usually a summary presentation of activities of multiple downstream effectors whose mRNA abundance may or may not be altered upon stimulations. In silico dissection of the contribution of effectors to the upstream regulators’ effect has been utilized previously by removing genes that reflect the effector’s activities from the upstream regulator’s gene signature (PMID: 18757322).*

**Regarding questions on “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?”**

Both bootstrap methods are non-parametric with no assumption of the population distribution, therefore sufficient large amount of simulation will provide us empirical distribution where can be consulted for statistics testing. It largely depends how much the “downstream genes target” eliminated will be impacted by the upstream regulator revealed from the SEM fitting. 100 could work but in our exercise 1000 rounds ensures a stable empirical distribution curve. In our implementation, we rely on the parallel process to conduct the bootstrap simulation, therefore a multicore hardware is recommended. As the example shown in the manuscript on KEGG pathway analysis with 28 categories, it can take up to a couple of hours to finish this step. If no multicore is detected, the application to execute a serialized solution which can take much longer time to complete. **(done)**

*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 XXX-YYY:

“*For the hypothesis generation purpose, a subset of genes that are associated with pathways of interest or downstream effectors could be removed from the upstream regulator’s gene signature as a in silico perturbation to infer the potential impact of losing the downstream signaling on the activities of the upstream regulator (PMID: 18757322). Based on the SEM model, a presumed relationship can be tested in humans by determining the significance of the inference via a non-parametric bootstrap resampling framework. The resulting perturbed pathways would help to prioritize experimentations in model systems. These workflow steps are shown within the dotted rectangle on the right side of Figure 1 with three major features implemented in the SEMPIPs App as the function tabs when the Shiny App is launched (Figure 2).”*

The ”hypothesis” in question was originally described in lines XXX-YYY”: “*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 XXX-YYY):

“*This finding suggests that the expression levels of GATA2 direct downstream targets, a subset of the full GATA2 regulated genes, can mathematically serve as surrogate reporters of the GATA2 activities in the human endometrium tissues, which supports that observing gene expression patterns of GATA2 direct downstream target genes is able to reflect GATA2’s activities in this context.*”

*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 agree that the MplusAutomation’s authors have done a good job in “mirroring” the commercially available software “Mplus” and implement this modeling in open-source R. It is designed to automate three major aspects of latent variable modelling, (1) create a group of models (2) run them in batches (3) allow extracting the model fitting statistics. Our SEMIPs has a similarity to MplusAutomation, where we implement SEM model in R instead of Mplus for the computational flexibility and backend automation consideration. We use the lavaan package, a highly credited/cited package exists in the research community since 2012 to implement the SEM model and extract all the statistics from the modeling output. But, our main goal in this research anchors on the biology and provide our web-lab scientists a convenient tool to explore their novel hypothesis and test the validity of their thought, and most importantly helps with hypothesis generation process. The main advantage is that our application is designed to use Rshiny to render a user’s friendly web front end. It allows web-lab scientists with limited bioinformatics skills to use the platform for this biological hypothesis testing and generation as we exemplified in the manuscript.

*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 also 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 time.

**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 included the detailed explanations of T-score, bootstrap approach, and SEM method in the in the revised manuscript. The revised texts could be found in:

1. Lines XXX-YYY: Detail explanation about the t-score.
2. Lines XXX-YYY: Detail explanation about the bootstrap method.
3. Lines XXX-YYY: The biological application of the method
4. Lines XXX-YYY: Additional discussion about the potential limitations of this App and comparison between this App and MplusAutomation.

The figure 1 and figure legend has 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.

An elaboration on the t-score has been included in lines XXX-YYY:

“*The T-score was employed to project molecular activities of a gene of interest from a model system experiment to human specimens where a perturbation was not directly applicable [PMID: 25295534, 19666588, 19490893, 18757322]. In a model system, the biological replicates 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 measurements will be properly collected from both groups (i.e., gene expression profile from a genome wide gene expression experiment). Significantly changed genes/probes (signatures) will be obtained from this analysis according to pre-determined thresholds followed by a statistical analysis with directionality (up/down regulation). Such a list of genes/probes are deemed collectively as the “gene signature” of biological responses to a particular perturbation in a given context such as cell or tissue types of interest. And these downstream target genes of the perturbed molecule are referred as “signature genes” of the molecule of interest. This gene signature information will be projected into the human specimen of interest bearing the assumption that the biological behavior of the gene of interest is conserved between the chosen model system and the human specimens.*

*In the gene expression dataset (i.e., human) that molecular activities of the factor of interest on individual samples are to be estimated, the orthologs of the signature genes were first identified and grouped based on the directionality of the signature genes. The T-scores of individual samples in the dataset were calculated by a normal t-statistics from these two groups of measurements to derive a single number as a quantitative surrogate of molecular activities of interest. Specimens with T-scores larger than 0, which share a similar signature gene expression profile from the model system, were classified as having gene signature activities and vice versa.*”

* 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?

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 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 did use expression of SOX17, therefore, it is put in the 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.

Thanks for pointing this out, we have explained in the “revised” figure and “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 made in the supplemental figure.

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

We think this is an excellent suggestion. The code was written in our internal format. We have included more comments in the code so it would be easier for readers to follow. If there is a source code standard which you would like us to follow, could it be possible you could provide us more specific requirements or point us to an example?