New comments from Reviewer1:

The paper somewhat improved, but it is still difficult to read by someone not directly involved in this project. My recommendation is to focus on Abstract and Introduction to clearly describe the key ideas. Then the rest of the paper can be more easily understood. Specifically, when I read abstract, the meaning or purpose of the following terms is unclear to me:

We thank reviewer’s comments and have revised the abstract and introduction based on the reviewer’s suggestions. We hope that the revised abstract and introduction will be clearer to describe our key ideas.

- line 30: a model system (a biological system?), a human tissue (omit this term completely?), a T-score calculation (too generic, T-score of what?)

We have addressed them accordingly in the newly revised abstract.

- line 31: complex human system (complex human biology?),

Yes, it was meant for the complext human biological system.

- line 33: specimen (in expression assay?)

Yes, specimen can be human tissues as study object in the experiment.

 - line 36: T-score (of what?)

T-score is what was calculated from the gene expression matrix between two groups, essentially two-tailed inverse of the *t*-distribution (details in method) .

- line 39: If the SEM is limited to 3 nodes, it would be useful to make a comment here whether it is too limiting, or what other applications can be considered, and whether 2 input and 1 output variable are always assumed

Yes, a 3-node SEM model is the most popular choice. There are many other SEM models with more nodes that fit different hypotheses. In our practice, we focus on a 3-node SEM model to test our hypotheses and explore potential relationship among other “factors”, which are represented as the variables/nodes in the SEM model.

- line 43: not sure if removing 'in silico' would actually improve the clarity of this sentence

We chose the word “*in silico*” because it helps us explore hypothesis generation. Not only does it save resources, it helps us narrow the candidates and ultimately saves us significant time for experiments.

Can we just remove “in silico” in this sentence?

To illustrate further, I am looking at the supplementary file.

-  page 1: 'in the SEMIPs project' more clear would be 'in the proposed SEMIPs method' and 'we tested the relationship among 3 variables assuming a three-node SEM model' and it would be better to clarify what variables are assumed; next sentence says 'each of these factors' but it is unclear what factors are meant, since no factor was mentioned previously; and again, T-score is mentioned, but it is unclear what quanitity the T-score is calculated for

We have revised the text according to the suggestions.

Furthermore, in Figure 1, where is the SEM considered? It may be useful to emphasize it.

We made revisions to Figure 1 by enclosing the SEM model with a green rectangle. We also revised the legend accordingly by “through a 3-node SEM model indicated by the green rectangle.”

In Figure 3, what are the white circles overlaping labels?

We have included an explanation to the figure 3 legend, “The while circles represent targeted subsets of genes to be eliminated in both bootstrap simulations.”

Figure 4 makes it clear what 3-variable SEM looks like, perhaps the general version of this model should be shown and explained earlier in the manuscript.

We have revised the figure 4 legend for clarity: “Model fit statistics for the joint regulation of SOX17 gene expression by GATA2 and PGR in the GEO GSE58144 dataset was illustrated in the 3-node SEM. Two exogenous variables were “Gene Signature of GATA2 Direct Downstream Targets” and “PGR Gene Signature” respectively, and one endogenous variable was “SOX17 Expression Levels”.”