Follow-up to previous comments from Reviewer3:

1. Figure 3 (supplementary Figure 1 of initial submission), could you please clarify if the top thicker blue arrows, one labeled "Random exclude N genes", the other labeld "Exclude downstream target genes",indicate the same processing step? If yes, could you please label them using the same text. If no, could you please detail the difference in the Two-class Bootstrap Simulation section (line 256)?

Thanks for the follow up question. The reason we labeled them differently was that those were two different processes, but they do end up with “N” number of genes eliminated, and we see whether the confusion comes from. The starting point is that we try to test the impact of an upstream regular on its downstream target. In the “elimination without replacement” process, we attempt to eliminate same number (N) of irrelevant genes, the continue with the following SEM modeling steps etc. On the other hand, in the “elimination with replacement” process, we firstly eliminate “actual downstream target genes (N)”, and then randomly select same number of “irrelevant genes” from the pool – indicated by the blue cylinder and put them back into the shrunken list to restore back to the same number of genes as the “GATA significant gene list” followed by the following SEM modeling steps.

The two simulation strategies have different objectives.

We have updated the manuscript for the Two-class Bootstrap section line xxx

2. My earlier question about the two bootstrapping methods are not adequately addressed. In my question, I meant to ask if the two methods have different computational costs, and if they have different test power. In what situtations would you prefer one method over the other?

We are sorry for the oversight. As explained in responding to your further comment #1these two methods do have different objectives.

The “elimination without replacement” process was design to test whether randomly eliminating the same number of the irrelevant “downstream genes” render the same SEM testing result as those from the “real downstream target genes”. From such 1000 bootstrap simulation, we shall have an empirical distribution of “statistics” we are interested, a significance result will be obtained if our actual measure falls in any of the significant tails.

The “elimination with replacement” process, since we store the “original gene list” i.e. “GATA significant gene list” by supplying the same number of genes that we eliminated from the initial set. This will further provide us information on whether what we observe is from “eliminating N number of genes” or from “eliminating the target genes”. This procedure does take a little bit more computational resource, but it helps us to rule out the unnecessary question and it is necessary.

Additional minor comments:

1. Line 36, move "two-sided t-statistic" to line 30, where "T-score" is first mentioned.

Revised accordingly.

2. Line 57, suggest to remove "unbias[ed]ly", unless the authors would clarify what kind of bias is of concern and how the gene signature constructed this way are un biased.

Revised accordingly.

3. Line 227, typo "uwe"

Revised accordingly.

4. Figure 3, "Random[ly] exclude N genes", "Random[ly] draw (N) genes"

Do we need to change this??