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)?

Those two labels represent two different processes. We have now included an explanation for the difference in the Two-class Bootstrap simulation section (Line 266-272). Here we would like to test the impact of an upstream regulator on its downstream targets (N) using two simulation strategies. In the “elimination without replacement” process, we attempted to eliminate the same number (N) of irrelevant genes from the test gene list (i.e. GATA2 significant gene list in Figure 3), and then continued with the following SEM modeling steps. On the other hand, in the “elimination with replacement” process, we first eliminated the regulator’s target genes from the test gene list, and then randomly selected the same number of “irrelevant genes” from the gene pool (indicated by the blue cylinder in Figure 3, All genes except tested genes). The randomly selected irrelevant genes were next put back into the shrunken list to restore to the same number of genes as the initial test gene list (i.e. GATA2 significant gene list in Figure 3) and subsequently went through the following SEM modeling steps.

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 apologize for not fully addressing this question in the initial response. As explained in responding to your further comment #1these two methods do have different objectives.

No, they do not have different computational cost. The number of bootstraps would be the same. The two bootstrap methods do not test power. They simply provide a non-parametric way to assess the probability of detecting a gene regulatory network via the SEM by chance. The bootstrap with replacement is preferable as it models the null hypothesis more readily since the downstream targets removed are replace with randomly selected. Thus, network sizes remain constant.

Additional minor comments:

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

~~Revised accordingly.~~ It is corrected

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.~~ It is corrected

3. Line 227, typo "uwe"

~~Revised accordingly.~~ It is corrected

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

It is corrected