**The calling of differential genetic interactions between conditions needs to be improved. At the moment, if I understand correctly, they use a Z-score cut-off in each condition to define interactions in that condition and then they simply count whether an interaction was found or not in each condition. This is likely to call false positive differential interactions because of fluctuations around the Z-score threshold. I would like to see quantitative comparisons of the interaction strengths across the conditions and direct statistical tests for the differences in interaction between / across the conditions and of course to control the global false discovery rate.**

We thank the reviewer for their feedback on the statistical calling of genetic interactions. The Z-score cutoff used meant that the most extreme deviations around the expected double mutant fitness in each condition were called as genetic interactions. In practice, because of the expected high frequency of genetic interactions in this dataset from previous publications, this simple cutoff is likely to work well for prioritizing potential interactions for individual study, as we have demonstrated. However we do agree that beyond prioritization, more statistical rigor in the calls would greatly help others interpret and analyze the (differential) interactions in the BFG-GI data itself. To allow for better future analysis, we have re-vamped our calling of GIs and have updated the Methods (lines XX to XX) to reflect this. Our updated pipeline is based on three principles:

* The genetic interaction scores should be internally consistent. That is, we should expect few or no interactions amongst neutral-neutral and neutral-DNA damage pairs (“neutral pairs”), but some under DNA damage pairs. This point may seem trivial, but in practice requires a careful definition of fitness from the sequencing data. We had noticed that in some cases, our previous GI scoring would exhibit a positive bias for neutral pairs, and we have redefined our fitness metric to avoid such cases. Moreover, we now use the distribution of GIs amongst neutral pairs as an empirical null model, allowing us to estimate a false discovery rate for each interaction as suggested.
* The genetic interaction scores should incorporate uncertainty in the measurements. In some cases, it is possible to obtain a large GIS because of uncertainty in fitness estimates associated with low read counts. We have incorporated an error model which normalizes the genetic interaction score by several modeled sources of uncertainty, and work with this score instead of the raw GI directly.
* The genetic interaction scores should be externally consistent. Given the above two considerations, we have now chosen a 5% ‘internal’ FDR cutoff based on the uncertainty-normalized genetic interaction scores. As a sanity check, we see that the called interactions validate well in the St. Onge et al data, in both the MMS (red) and no MMS (black) condition. We estimate a ~70-78% external validation rate for positive interactions, and ~85-87% external validation rate for negative interactions. We have included this score for each interaction, so that other cutoffs may be chosen by the reader.

