

Very strong presentation, and the committee really appreciated your attention to detail on the strengths and weaknesses of the various approaches under consideration. In general this seems like a strong project in terms of possible impact and being an appropriate amount of work. Committee members had several comments that we offer in hopes of strengthening any future implementation of the project, which we include below, but we'd like to emphasize that in general this was an extremely strong preliminary exam performance, so we're really just at the stage of enjoying the science and wanting to give our best recommendations.

First, we'd like to recommend that early in a project lifecycle, it is important to learn all of the details of the project in terms of likely false positive/negative rates, needed performance to be useful, etc.

We are also concerned about the feasibility of the proposed work simply due to the number of different disciplines represented. In practice the proposed research only seems feasible as a PhD project if the metabolomics could be handed off to a collaborator for much of the performance and compound identification.

Additional comments:

\*It may be difficult to directly test the aim 1 hypothesis as currently written; model performance could be due to functional differences that are unrelated to functional redundancy. It may be useful to have separate predictions about functional redundancy and predictive power of functional profiling.

\*Why use OTUs vs. SNVs?

\*Why pathways vs. protein functions? Pathway annotation adds an extra layer of uncertainty, may result in sparse data (that's my experience with metacyc).

\*Could transcriptomics be a good alternative (or addition?) to metabolomics? It'd be more sensitive and would be easier link to metagenomics (though as pointed out during our discussion it has its own limitations).

\*We'd like to encourage you to think about disconnect between the coverage of metabolome by pathway databases and by experimental LC-MS methods and the alternative ways to annotate metabolites, including additional databases and including unknown metabolites into the analysis.