**Summary**

We are grateful for the thoughtful comments on our manuscript, and we have worked diligently to clarify the remaining reviewer points. We have also made efforts to include the additional analyses suggested by the editor. In total, there have been many additional efforts made to address reviewer comments and we have provided specific responses (in black) to reviewer comments (in blue) below.

**Editor**

**Major Concerns:**

In particular, as I read over their comments and the manuscript, I feel you have presented an interesting new approach, especially in regard to the potential of inferring a cell's metabolic uptake, secretion and internal metabolism, but the validation in those regards remain incomplete, with a lack of gene essentiality evaluation (as can be found using multiple datasets on paired and defined growth conditions for E. coli) and limited validation substrate uptake/secretion based on just a few conditions.

We have added comparisons of substrate uptake predictions in defined media across platforms, demonstrating that RIPTiDe is able to infer these conditions based on contextualized model activity (Table S7).

We also performed a gene essentiality analysis and have shown that RIPTiDe outperforms other approaches (Table S3). We believe performing an analysis of true gene essentiality on a contextualized model where any amount of functionality has been pruned or altered is ill-advised. Only when all possible metabolic pathways and potential compensatory mechanism are present, and only the available exchange reactions have been manipulated to reflect media conditions, are gene essentiality screens truly reliable. In accordance with this concern, we have renamed the concept in the text as conditional importance, and performed a comparison between other platforms.

The resultant model from RIPTiDe specifically only includes what is the most likely momentarily active metabolic pathways given the data provided. This is not to say that an organism doesn’t have another pathway to compensate for the loss of some functionality predicted to be active using RIPTiDe. We utilized essentiality in this instance to demonstrate the most likely forms of metabolism that are needed to perform optimally given the growth conditions encountered by the bacterium. This consideration becomes especially important in environments where understanding the active metabolic strategy may be important to other phenotypes as in certain types of bacterial pathogenesis.

**Reviewer #1:**

**Overview:**

I thank the authors for clarifying the questions I had in the previous round. It has now become more clear to me that the main purpose of the method is to use gene expression data to infer the metabolic context in situations where environmental conditions are not known. As far as I understand, the authors have not proved (or claimed) that this method would outperform a normal pFBA simulation when the medium composition is well defined. However, I am afraid that, at least from my side, another (hopefully short) round of review is still necessary.

**Major Concerns:**

In figures 1A and 5A the reaction coefficient is higher for higher transcript levels. Shouldn't this be the other way around ? I think that in general the paper is missing a clear mathematical description of the whole algorithm.

The reviewer is correct and we have corrected the figures to better reflect the actual coefficient assignment. The method assigns a higher weight (corresponding to a lower coefficient) for a higher transcript level during the pruning phase, and we over-simplified these ideas in the figure. We appreciate the reviewer’s attention to this point.

On page 17 it says this is a bilevel optimization method. I do not think this is the case (at least in the common use of the word in this field), since it only uses a single objective function and is solved as a simple LP problem.

We have added additional clarification to this section to indicate exactly how the multiple phases of optimization are implemented. We now refer to our approach as sequential optimization.

In figure 3C, I do not understand how it is possible that some genes (five, according to the main text) are essential using the max parsimony simulation, and no longer essential in RIPTiDe. Since the max parsimony simulation is less constrained than RIPTiDe, there is no relaxation of the problem that would allow an essential gene to become non-essential.

Maximum parsimony is far more constrained than RIPTiDe. Transcriptomic evidence simply allows for a relaxation of these strict constraints on a reaction-by-reaction basis through RIPTiDe. The language surrounding this distinction as been subsequently modified to make this point more clear.

In page 25 the authors say that RIPTiDe correctly predicts the lower growth rates observed experimentally compared to those predicted with the pFBA method. This claim is not exactly correct. The lower growth rates are simply a consequence of the fact that in RIPTiDe the authors relax growth rate to 80% of the maximum value. This is the result of a manually selected threshold, and not a prediction of the method.

The pFBA method requires the identical 80% minimum threshold to be applied to the objective function. This constraint is therefore evenly applied to all cases presented in the growth rate predictions. We have added this to the text in this section to clarify this point.

In the abstract the authors make a claim on the accuracy of the method. I think this claim should be removed, since at no point the authors actually validate the predictive accuracy of the model by comparing the flux predictions with experimental flux or metabolite data. All the case-studies provided only show that the context-based predictions of the method seem to be biologically meaningful and can help provide mechanistic insight into metabolic regulation mechanisms.

We understand the reviewers concern and have adjusted the language in the abstract accordingly. We have also tried to better reflect this point when discussing gene essentiality.

**Reviewer #3:**

**Major Concerns:**

Even if I agree that a CRISPR-based validation strategy has some issues, I would nevertheless rather trust gene essentiality screens over the predictions of any metabolic model. Metabolic models are only useful if they are able to recapitulate to some extend observations made for a given well experimental conditions. This allows having more confidence in predictions made for other conditions.

In this revision, we have added experimental validation to the computational analyses with data from substrate utilization and gene essentiality screens.

I still think the claims from this paper have to be nuanced as the claims only are supported by a benchmarking study from 2014 that was performed in yeast. Please change to:

“But some of these approaches were recently shown in one study to generate less accurate metabolic predictions in yeast than a transcript-agnostic method of flux minimization (pFBA), which identifies the most efficient/economic patterns of metabolism given certain growth constraints.”

There is a specific statement that this paper validated the approach in yeast (see 2nd to last paragraph in the introduction).

If RIPTiDE responds to a problem that is not addressed by the other algorithms then should more emphasize this fact and be more neutral towards shortcomings of other and especially if you do not prove that your algorithm is exempt from the latter or if you do not show it your algorithm better results in real-life experiments. Of course, the use of arbitrary thresholds has a large impact on the output models and that’s the reason why a lot of authors of algorithms have published data integration workflows even before the papers you cite. I can indicate citations that show that these discretization methods perform better than continuous data integration. But that’s not the point. The point is that using an integration like the expression ratio function is not that different from an arbitrary threshold. Another function would produce very different models. Further, I agree that none algorithms outperform the other in all the settings this idea should, by the way, be perceived when reading your text. But again it is not a question about discretized versus continuous based algorithms in general. The question is: Is your algorithm performing better than existing algorithms for matching tasks. Using continuous data is no guarantee that it does. Further I am not sure how suitable CORDA is for flux prediction I would have categorized it as model building algorithms such as MBA, mCADRE or FASTCORE. That’s the reason I suggested the use of RegrEX. Taking into consideration that Matlab is not free I suggested tINIT which was in Python. Although tINIT is not a true flux prediction algorithm, it is more recent than GIMME and iMAT. There is no publication refereeing the python version of tINIT because I guess they just translated the algorithm from matlab to python and that does not deserve a publication. Fine if you want to restrict yourselves to algorithms that were coded in python but I maintain then you should really nuance your claims. Because you are avoiding testing algorithms that are more recent and that were designed to perfom similar tasks and might outperform your algorithm. Taken together I suggest you modify the following statements:

“While these approaches have been shown to create models that contain increased amounts of known tissue-specific metabolic pathways, they do not inherently generate fully functional models of metabolism nor were they reflective of momentary metabolic trends that would be active under specific conditions. “ Remove or nuance or back these claims with evidence. To my knowledge, none of the existing algorithms produce fully functional models and they are of course reflective of momentary metabolic trends as long as the data supports these specific conditions. These algorithms were shown, among others, to capture metabolic variations between cancer and healthy tissues only based on transcriptomic data from the TCGA. What you can say is that these algorithms, unlike the previous ones, aim to extract from a GENRE a consistent subnetwork that is more specific to a given context by maximizing reactions that are supported by omics in this condition of interest rather than predicting fluxes. Please see Machado et al, for an explanation of the differences between the two approaches.

In this revision we have included additional benchmarking assessments as well as revised the text to reflect nuances referenced by the reviewer.

Furthermore, FASTCORE , mCADRE and MBA have very similar strategies and were, therefore, aggregate in one family: The MBA family by Estevez et al. So I do not understand why MBA is cited with the flux prediction algorithms. I am not quite longer sure for Colijn’s or the other algorithms but Lee’s algorithms take as input absolute values from transcriptomic data. Please be more precise in your statements. Please also discuss RegrEX, for disclosure, I am not an author of this algorithm but it does not use thresholds and also does not maximize the consistency between the data and the fluxes. Again you have to be more nuanced in your statements.

“We have also integrated functionality to instead utilize one or more metabolic tasks (single reactions) as constraints which must achieve positive flux in order to accommodate larger-scale models where a single defined cellular objective may be a biologically invalid assumption (i.e. human tissue). For a more detailed description of the algorithm, refer to the Methods section.” Remove that from the Introduction and put that in the methods. I miss the computational demands of RIPTiDE, how long does it take for a reconstruction? Can it work on a non-curated model? You talk about human tissue but can RIPTiDE deal with bigger models such as Recon 3D. If it can’t then I would remove every reference to human tissues. The first example only provides little information, the conclusion one can draw is that transcriptomics data help in the given example with simulated data. Of course this setting is designed to favor RIPTIDE and algorithm using transcriptomic information over pFBA so I am not sure what does this example proves. I guess that any algorithm using transcriptomic data would pass the test. I would be more interested in knowing if the tool could do the same with real data and a non-curated reconstruction.

We have now added additional analysis benchmarking against the RIPTiDe method presented here.

RIPTiDe takes an average of 33 seconds to complete on the model of *E.coli* used here for benchmarking. RIPTiDe also completes using Recon3D (the largest human model we had available) in 2 minutesnand 30 seconds. Both considerably faster than many other approaches. We have now added this information to the text and generated a new supplemental figure.

Other algorithms do not necessarily pass the test of the simplified model example. Both CORDA and RegrEx fail using the prescribed thresholds, which is now reflected in the text and table S6.

I fail to see how this example shows that RIPTiDE outperforms GIMME, IMAT, and CORDA. What is the metric to test it? CORDA was indeed able to produce a model that was capable of growth, however 1001 out of 2583 reactions were included in the resulting model and the growth rate was not significantly altered. I do not understand the point you want to make The fact that a model is more easily studied is not a strong argument. The authors of CORDA do not believe that a model has to be the most compact one. It could make sense that an organism retains a certain level of plasticity to allow for adaptation. If the data supports the activation of other pathways than the ones required by the biomass function then maybe your test is not good. I guess even Ecoli is more complex than the solution of an FBA. Did you look how many reactions for CORDA are not supported by the data and how many for your algorithm? And again as stated before CORDA is not the best choice here. Further, the expression does not correlate with activity, therefore, I do not see why the fact that GIMME and IMAT reintroduce hundreds of reactions, is a problem. The low expression does not mean that the reaction is inactive and hundred of reactions is really a low number. Are the genes linked to every reaction of the RIPTiDE model highly expressed or does it also includes genes with low expression? Further, I do not know for the Python implementation but in Matlab, the functions have an option to produce functional models, therefore, I would not insist on that especially as I am not sure that the python version was coded by the authors of GIMME and iMAT.

As referenced above, we have revised this submission to include more comparisons to validating experimental data and to add appropriate nuance to comparisons with other algorithms.

The authors state that they chose GIMME and iMAT [11,12], due to their combined prevalence across the literature. Older algorithms, of course, accumulate more citations than recent ones. It does not prove that they are not obsolete. Remove this statement.

We have revised the text to be responsive to this concern from the reviewer.

For figure 3, I guess other flux prediction algorithms that use transcriptomic data would also capture some differences in this experimental setting. Can you at least test GIMME, IMAT and CORDA for these experiment setting.

We have added this analysis (See table S7); we also provided comparisons of MBA and RegrEx.

Different topologies of reaction inclusion were found among each of the contextualized models; however none were completely unique and a core set of metabolic reactions that were necessary for optimal growth across all conditions (347 reactions; Fig. 3B). Again would be interested to see with publicly available data if you have enrichment for essential genes among the genes controlling these reactions. I understand that predicted essential genes are not necessarily required for survival I would expect that they would have been selected. It the predicted essential genes have really an impact on the growth it should be possible to validate it by in experimental data. A list of 159 core genes without validation have only a limited interest.

As mentioned above, we have included a description of several validation data sets including substrate utilization and essential genes as referenced by the reviewer here.