Reviewer Response: "A Landscape of Metabolic Changes Across Tumor Types"

Reviewers' comments:

Reviewer #1: In the manuscript entitled "The Landscape of Metabolic Changes Across Tumor Types" Reznik et al. attempt to harmonize metabolomics data from different cancer studies to identify metabolic traits common to different tumor types.
The manuscript is well written and the methodology used is innovative.

We thank the reviewer for the positive comments.

The study suffers from two main limitations: the scarcity of large-scale metabolomic data available, and the uncertainty in the identification and absolute quantification of metabolites intrinsic to the mass spectrometry-based metabolomics approach. Because of these limitations, the biological conclusions of the study are observational.
Before publication, the authors should support some of the novel metabolic observations by coupling covariation analysis using transcriptomic datasets. For example they should assess if the intratumoral levels of kynurenine and glutamate associate with the expression of IDO1 and GLS respectively.

We interpreted the comment as having two basic components: **A) Does transcriptional data mirror the metabolomic findings?** and **B) Is there a way to independently corroborate our metabolomic findings?**

Addressing question A), We (Hakimi,Reznik, *et al* 2016) and others (*e.g.* Gerosa et al, Cell Systems 2016, Chubukov *et al,* Molecular Systems Biology 2013) have shown that while metabolites and enzyme levels jointly regulate metabolic flux (in a complex, non-linear fashion), the abundances of each are not necessarily correlated to each. For example, in ccRCC, we showed that both at an individual metabolite/enzyme level and at a pathway level, there is no correlation between changes (tumor vs. normal) when comparing enzymes/metabolites (Hakimi et al, Cancer Cell 2016). This intuitively makes sense, because both quantities may be subject to different kinds of regulation, and may adjust independently to achieve a desired flux.

Despite this previous work and the null expectation, we investigated whether our findings associating key metabolites with clinical features might also be evident using transcriptomic data. To do so, obtained TCGA transcriptomic data for pathological stage (grade was not available) for kidney and breast cancers (the two cancer types where had the highest numbers of samples with clinical and metabolomics data). We then

identified (with pathway informatics) the adjacency of metabolites and enzymes (*i.e.* which metabolites are used/produced by each enzyme). We then asked whether metabolites associated with clinical stage were also preferentially surrounded by genes associated with clinical stage. For three metabolomic datasets (Kidney, Breast Tang, Breast Terunuma), we found no such statistically significant enrichment. More specifically, while metabolites associated with stage were indeed surrounded by some genes which were differentially expressed in high stage vs. low stage patients, metabolites **not associated** with clinical stage were also surrounded by such genes. Indeed, the proportion of stage-associated genes surrounding the two sets of metabolites was nearly identical. Despite not finding statistically significant results, the tools for **this analysis are available for others to build on/enhance in the github repository in the analysis/Run_ClinicalMetabolomicsvsExpression.R script.**

B) On the other hand, there are opportunities within our dataset to independently validate some of our metabolomic findings. In particular, because we have 2 or more independent studies of 3 cancer types, this enables us to evaluate whether our findings can be independently validated. In particular we:

- Show in a **newly revised Figure S2A** that for breast, pancreatic, and prostate cancers, we find highly similar metabolomic changes between tumor and normal samples across replicate studies of the same cancer type.
- We also show in **Figure S2B** that for an independent study of a fourth cancer type, clear-cell renal cell carcinomas where we were only able to obtain information on tumor/normal changes (*i.e.* only fold changes and p-values, from Li *et al*, Nature 2014, fructose-1,6-bisphosphatase opposes renal carcinoma progression), we again find a high degree of statistically significant agreement.
- We also introduced a novel approach for evaluating the contribution of technical noise to our data. We have now used pathway information to identify metabolites in our dataset which are metabolically adjacent to each other (*i.e.* there exists an enzyme which produces/consumes each of the metabolites in a "metabolically adjacent pair"). We then calculated pairwise Spearman correlations between each metabolite in our dataset, and compared the distributions of correlation coefficients between pairs of adjacent metabolites/pairs of non-adjacent metabolites. Across all 9 studies with sufficient sample size except one (bladder cancer), we observe an enrichment for highly positive correlations between adjacent metabolites compared to non-adjacent metabolites. This data is now reported in **SI Table 1: Adjacent Metabolite Analysis and Supplementary Figure S2C and S2D** for 2 example studies (breast and kidney)**.** This finding bolsters the case that technical noise is not overpowering signal in our data.

When discussing the association between the abundance of metabolite in normal tissues and tumor grading, the Authors should comment on the contribution of infiltrating cancer cells into adjacent tissue. Vice versa, the presence of stromal and immune cells can dilute or enrich for specific metabolites, independently of the nutrient shuttling model hypothesized for some tumor types.

We agree with the reviewer, and have now included a discussion of infiltrating cancer cells/tumor impurities and the microenvironment on **Page 17** and **Figure S6.** In general, data is not available for both metabolite levels and tumor infilitration. However, we were able to obtain estimates of stromal/immune infiltrate for 23 breast cancer samples in our dataset (also profiled by the TCGA). While no metabolites were correlated to either stromal/immune infiltration scores after multiple hypothesis correction, several metabolites were approaching statistical significance. We comment on these metabolites (*e.g.* NADH and succinate), and remark in the manuscript that increased sample size may indeed identify metabolites that may be markers for specific types of tumor impurity.

Reviewer #2: The study by Reznik et al. investigates large scale changes in the metabolite levels associated with malignant transformation across a range of metabolomics studies and multiple cancer types. By first developing a pipeline that enables a systematic analysis across independent studies, the authors conclude that there exists significant heterogeneity in metabolomics changes between tumor and normal tissues. The authors also find some metabolites to be recurrently abundant in the majority of tumors types, and correlate their levels with clinical severity.

The concept of a pan-cancer analysis of metabolite levels is of general importance to the field. Furthermore, the study addresses a major difficulty of directly comparing metabolite levels across different studies. While the biological insights from the paper are relatively limited, the pipeline and the dataset can serve as a useful resource for future studies. However, some important questions need to be addressed: 1) To what extent are the authors analyzing technical noise inherent to individual metabolomics studies and independent of cancer/tissue type? 2) In general, how does one interpret the altered metabolite profiles? different metabolic phenotypes are often capable of producing similar metabolomics patterns 3) Are there general organizational principles that unify the presented results (i.e. do metabolite changes cluster primarily by tissue type, genotype etc.)?

We thank the reviewer for the kind commentary. We address each of the three points above below, as they are all related to specific comments/suggestions.

More specific comments/suggestions are given below.

1. In Figure S3, the correlation between metabolite fold-changes within studies of similar tumor types is confusing. The authors need to provide the raw correlations coefficients, as the Fisher's Exact test seems only to account for similar directionality of changes. If these correlations are weak, do they reflect true biological differences between similar cancer types, or primarily technical noise associated with individual studies? This is important to understand in order to draw meaningful biological conclusions from the data. One way in which the authors can investigate the contribution of noise, while comparing multiple studies, is to look at within-study correlations between metabolite levels.

We agree with the reviewer. We initially implemented the Fisher test in **Figure S2** (was previously S3) because we were interested in the direction of change of statistically significant differentially abundant metabolites. We have now revised the Figure to include both the Fisher test for significantly changed metabolites (now labeled with red color), as well as a Spearman correlation for all fold changes. Notably, the Spearman correlations strengthen our finding of consistency across studies of the same cancer type.

We also agree with the reviewer regarding the importance of investigating the contribution of noise to our findings. We have now used pathway information to identify metabolites in our dataset which are metabolically adjacent to each other (*i.e.* there exists an enzyme which produces/consumes each of the metabolites in a “metabolically adjacent pair”). We then calculated pairwise Spearman correlations between each metabolite in our dataset, and compared the distributions of correlation coefficients between pairs of adjacent metabolites/pairs of non-adjacent metabolites. Across all 9 studies with sufficient sample size except one (bladder cancer), we observe an enrichment for highly positive correlations between adjacent metabolites compared to non-adjacent metabolites. This data is now reported in **SI Table Adjacent Metabolite Analysis and Supplementary Figure S2 (**was previously S3) for 2 example studies (breast and kidney)**.** This finding bolsters the case that technical noise is not overpowering signal in our data.

2. In calculating the median absolute deviation (MAD), should the MAD values be normalized for total abundance within tumors and normal samples? One would expect that tissue types with higher overall metabolite levels in tumors would have higher overall variability as well.

The intuition of the reviewer is correct, our measure of "variability" (defined as the ratio of median absolute deviation (MAD) in the tumor vs MAD in the normal) was biased. Specifically, if tumors are preferentially enriched for metabolites with increased abundance, then unless proper normalization is implemented tumors will show a corresponding increase in variability. This is for obvious reasons undesirable, and we have amended our measure of variability below, and in the **Methods on Pages 21**.

The question which we are ultimately interested in is whether *tumors as a group* are more variable than *normals as a group*. Thus, we redefined our measure of variability by first separating tissue samples into a tumor group and a normal group. For each metabolite, we implement the following procedure. Within each group (e.g. tumor), we normalize the abundance of each metabolite to the median abundance *within the group* (*e.g.* in tumor samples for a given metabolite *x* we find the median abundance of *x* only in tumor samples, and then divide all measurements of *x* in a tumor sample by *m* and take the log2 ratio*;* this procedure is then separately repeated for normal samples). Note that the introduction of a log2 ratio ensures that proportional decreases are weighted equally to proportional increases (*i.e.* a 5-fold decrease relative to median is equal to a 5-fold increase).

After the above normalization, MAD is calculated for normalized tumor and normalized metabolite abundances as before, and the ratio of the MAD scores is calculated. This procedure is repeated for all metabolites in a study, and the distribution of the MAD ratios is tested for whether it is enriched for higher MAD values in tumor or normal tissues.

The above normalization now addresses whether tumors are inherently more/less variable than normal tissues. Importantly, after implementing this normalization, our results remain unchanged: kidney and breast tumors show increased variability relative to normal tissue, while pancreatic tumors show decreased variability. The kidney cancer finding is notable because, although kidney cancer has roughly equal proportions of differentially more/less abundant metabolites relative to normal tissue, we found that the tumors were extremely variable compared to normal tissue (MAD ratio shifted far to the right). This suggests that while kidney tumors may not generically have higher levels of key metabolites, the variation in the levels of those metabolites is higher.

3. In general, the authors should provide some basic comments and interpretations into the pathway-level metabolic alterations. For example, why would metabolites involved in sugar metabolism be high, whereas metabolites in fatty acids biosynthesis be lower? How does one interpret the heterogeneity in the central carbon metabolism - is there relative accumulation (depletion) upstream (downstream) of important branch points of glycolysis, such as the oxidative PPP or serine biosynthesis? Do elevated lactate levels reflect increased glycolytic flux or preferential shunting of pyruvate to lactate under PDH inactivation? Some baseline interpretations need to be provided in addition to the table of pathway level results.

We agree with the reviewer, interpretation of our findings is critical. We have now added 4 paragraphs of text discussing changes in central carbon metabolism on **Pages 10-12.** These paragraphs (1) discuss the central challenges of inferring flux from metabolite levels, (2) provide evidence supporting an increase in glycolytic flux as the primary driver of increases in lactate levels in tumors, (3) dissect cancer-type-specific patterns of changes in TCA cycle metabolites, specifically in prostate and kidney cancers, and (4) propose that cancer cells may increasingly rely on beta-oxidation as an anapleurotic resource for the TCA cycle. We have also included a discussion of 2HG as it relates to tumor grade on **Page 14**, see comment below.

4. The results correlating metabolite levels to pathological grade are interesting. One interesting finding is that the directionality of association changes in different tumor types. Again, the authors should provide some additional commentary. For example, the pathophysiology of 2-HG and its downstream effects are particularly well-studied. Is there any data to explain why in particular contexts this metabolite would lead to increased or decreased aggressiveness (breast and kidney cancer versus glioma and prostate cancer).

The reviewer raises an intriguing question. In the text, we now discuss the opposing associations for 2HG on **Page 14**, emphasizing 2 points:

First, while elevation of 2HG is known to modulate the epigenetic landscape (*e.g.* hypermethylation of the genome), the link between epigenetic changes and tumor aggressiveness is complex and likely tumor-type-dependent.

Second, we propose that the actual chemical identity of 2HG may be distinct across these tumor types. In particular, 2HG may be produced via distinct mechanisms in

gliomas vs. RCCs/breast cancers (here we ignore prostate cancer, which has been described to harbor activating *IDH* mutations, but very rarely). Specifically, neomorphic mutations to *IDH1/2* induce production of the D-enantiomer of 2HG in gliomas. In contrast, RCC/breast tumors do not (canonically) harbor *IDH* mutations; rather, it is possible that the 2HG in these tumors is actually of the (more potent) L-enantiomer and induced by elevated hypoxia. This is in fact supported by data showing that hypoxia (RCC tumors are driven by induction of HIF) induces production of L-2HG (Intlekofer 2017, Nature Cell Bio), and direct evidence of L-2HG in RCC tumors by Shim et al, 2014 Cancer Discovery. We emphasize in the text that to actually evaluate this hypothesis, further enantiomer-specific metabolomic profiling in these tumors will need to be undertaken.

5. The correlation of metabolite levels in adjacent normal tissue to tumor clinical grade is an interesting idea. However, the results regarding putrescine in prostate cancer are quite puzzling. Can the authors reconcile the opposite correlations of putrescine in tumor versus normal tissue and connect these results to physiological/pathophysiological roles of putrescine in cancer? The possibility of cross-feeding of metabolites from normal adjacent tissue into tumors is also interesting. Are there metabolites that are positively correlated to clinical grade in both tumor and normal tissue? Have the authors specifically investigated known cross-feeding interactions (i.e. alanine in the case of PDAC)?

Again, we found this comment very intriguing. We decomposed it into four separate but related questions:

A) *Do many metabolites show positive correlations between tumor and normal tissues?* We calculated the (spearman) correlation for each metabolite in matched tumor/normal samples. Doing so, we indeed found a number of metabolites whose abundances were correlated in matched tumor/normal tissues (**Supplementary Table Tumor/Normal Correlation Scores).** Notably, many of these metabolites appear to be metabolites which circulate physiologically through the body (for example, 1,5-anhydroglucitol, which may be used as a measure of glycemic variability, and the ketone body 3-hydroxybutyrate, which may produced in the liver and distributed systemically). Therefore, we believe that some, but not all, of the correlations in metabolite abundance that we observe are driven by whole-body physiological variation.

B) *Are there metabolites showing the same direction of clinical correlation in both tumor and normal tissue?* We queried our clinical results to identify such metabolites, and identified many (see **SI Figure 6A and Page 15**). These include

(but are not limited to) a positive association between arginine levels in both breast tumor and normal tissue, and a negative correlation between choline levels and kidney tumor and normal tissue (reported in the **Supplementary Table 2,** and **viewable interactively on our website, sanderlab.org/pancanmet).**

B) *Do we find opposite correlations in tumor/normal tissues, such as those seen with putrescine, puzzling?* We do not. Indeed, we suspect that if nutrient shuttling exists, it may manifest as either negative or positive correlation between metabolites levels in tumor/normal samples (or no correlation at all). The precise mechanism underlying the shuttling phenomenon could lead to different associations between metabolite levels. For example, a positive correlation may arise if excess production in the normal tissue is shuttled to the tumor passively (*e.g.* carried by blood vessels). Similarly, a negative correlation may arise if a fixed amount of the metabolite is produced by the normal tissue, and an external factor (*e.g.* the expression of a nutrient transporter in the normal tissue) increases/decreases shuttling to the tumor tissue. We have included a discussion of this point on **Page 18.**

C) *Have we investigated known cross-feeding interactions such as alanine in the case of PDAC?* We did indeed look specifically at the case of alanine in PDAC. We had access to 3 independent studies of PDAC, and we observed a strong correlation in one (Pancreatic Kamphoorst), a weak correlation in second (Pancreatic Zhang1) and no correlation in the third (Pancreatic Zhang2). We briefly mention this point on **Page 15-16.**

Reviewer #3: This is a timely and interesting study, gathering, organizing and presenting metabolomics profiling across cancers. The manuscripts describes and discusses the challenges inherent to such techniques, and to the best of my knowledge this is the first pan-cancer study of this kind.
One major value of the study it to examine the variability of metabolomics data, and talk about the requirements for standardization in the analysis of such data. The study also publishes the finding in a browsable interface, which adds to the potential impact.
Whilst the results are intriguing, with just some specific points requiring further clarification (see below), one significant shortcoming of this study is that the implications of the discoveries are not tested nor fully discussed. This is mentioned again in the specific points below, but in general this results in a missed opportunity to contextualize the work to the bigger picture of metabolic reprogramming in cancer.

Specific comments:

* Mistakes in references and missing key references

- The references start from 9; to check the order
We have corrected this, thank you.

- Line 42: "Instead, much of the work in cancer metabolism has focused on examination of changes in the expression of metabolic enzymes, and the identification of patterns of metabolic gene expression common to many cancer types" References to high impact studies in the field seem to be missing e.g. the landmark paper by Possemato et al (Nature 476, 2011). The manuscript would greatly benefit from citing major studies in the field.
We have amended the text to include reference to this and other landmark papers in the field on **Pages 2-3.**

- Line 54: "Furthermore, many functional and common genetic alterations in cancer involve metabolic enzymes (e.g. IDH, SDH, FH) or regulators of metabolism (e.g. MTOR, VHL)." Missing references e.g. for IDH1 mutations a recent key study by Jalbert et al (Scientific reports 7, 2017), and also they are completely omitting certain types of recurrent genetic alterations e.g. recent large pancancer analysis by Haider et al (Genome Biology 17, 2016).
We have now added these and other references on **Pages 2-3**, thank you for the suggestion.

* Lack of clarity in data accessibility

- The web service is intriguing yet not all information is accessible, so the potential utility is unclear. Could the whole analysis be reproduced via the web site? It would be helpful to clearly state in the manuscript which data will be accessible in the website and which omitted, a table might help the reader to understand this.

We provide a web application for **interactive exploration** of results (those described in the manuscript, as well as those not described) rather than a web service from which readers might analyze data directly. Importantly, we **separately provide input data, informatic pipeline, analysis, and visualization code at our code repository (https://github.com/dfci/pancanmet_analysis)**, so that readers can fully replicate and build on our findings. This is now clarified in the text on **Pages 3-4**.

More specifically, the web application has the ability to explore 1) metabolite data stratified by metadata data (e.g. HMDB classifications) and by study, 2) univariate and bivariate fold changes by metabolite and by study, and 3) clinical associations by metabolite and by study.

- Linked to the previous point, it would help to state also where other information/data, not available via the website, will be saved/accessible by the community.

We added additional text to the introduction **on Pages 3-4** to highlighting the interactive website and its capabilities, as well as the code repository containing the code for the whole analysis and data not available from the exploratory website.

- If some data is not accessible, the reasons should be clearly discussed and the impact accessed e.g. analyses can be browsed but not reproduced, etc.

This is not the case. We strive to make the work as transparent as possible and data/code for the whole analysis is available from the analysis code repository (https://github.com/dfci/pancanmet_analysis). This is now mentioned at several points in the text, including on **Pages 3-4 and 18.**

- The computational framework/workflow for the analysis is very interesting, and discussed very clearly, but it does not seem accessible for reuse and reproducibility checks; a clear link to this should be provided.

We make all code available and highlight the code repository (https://github.com/dfci/pancanmet_analysis) in several sections of the manuscript, including the introduction and discussion, *e.g.* **Pages 3-4 and 18.**

* Not fully developed discussion

- The major weakness of this manuscript is the lack of context and critical discussion provided. For a paper that aims to produce a major pancancer resource this aspect is crucial. The report of previous findings is limited and often biased, and key papers in the field are missing (see above). This is a missed opportunity to critically examine the results obtained here with respect to the state of the art, and to highlight common findings but also differences where they exist.

We agree with the reviewer, and have spent a great deal of effort adding context and reflection to the manuscript. Some specific points are highlighted below:

A) We have added the relevant references to critical prior work, *e.g.* on **Pages 2-3**.
B) We have now added 4 paragraphs of text discussing changes in central carbon metabolism across cancer types on **Page 10-12.** These paragraphs (1) discuss the central challenges of inferring flux from metabolite levels, (2) provide data supporting an increase in glycolytic flux as the primary driver of increases in lactate levels in tumors, (3) dissect cancer-type-specific patterns of changes in TCA cycle metabolites, specifically in prostate and kidney cancers, and (4) propose that cancer cells may increasingly rely on beta-oxidation as an anapleurotic resource for the TCA cycle.
C) We have added text elaborating on our clinical findings on **Page 14**. In particular, we discuss the biochemistry underlying 2-HG production, and how this may lead to different directions of association between 2-HG and tumor grade across several cancer types.
D) We have elaborated further on our findings relative to a prior meta-analysis completed by Goveia *et al* on **Page 9**. In particular, we highlight some of the common findings between the two studies, as well as some of the key differences.

* Results/Methods need clarification

- The meta-analysis framework is well justified; however the authors did not test the applicability of such approach. An analysis assessing whether the dataset are sufficiently uniform for this approach would benefit the manuscript.

We appreciate the question from the reviewer. We are critically aware of data normalization issues throughout our analysis, and for this reason we elected to use strictly rank based tests for the meta-analysis (and for all other analyses in the manuscript). For these tests, there are no assumptions on the distribution or scale of the metabolite data, except that the measurements obey rank-ordering (*i.e.* a higher value corresponds to a higher abundance). Because our data obeys rank-ordering (indeed if it did not, all of the analyses in this manuscript would be invalid), our approaches are appropriate. We have clarified this in the manuscript on **Page 13.**

- The number of samples considered changes throughout the manuscript. Whilst this is clearly due to missing cases and not all cases having combined tumour-normal tissues,

the number of samples used should be clear for every analysis presented. The data ought to be in each figure legend or on each plot.

We agree with the reviewer and have now **added text to the legend of each figure (including SI Figures)** indicating which data was used and the size of it (except of course Figure 1 which actually depicts the sample sizes and Figure 2 which is just a methodological figure).

- Linked to the previous point, it is not always clear which analyses are paired and which are not. To avoid misunderstanding, my suggestion would be to introduce a statement in each Figure.

We again agree with the reviewer and have also **added text to the legend of each figure (including SI Figures)** indicating whether the analysis was done with/without regard to matching/pairing from the same patient.

- For some cancer types less cases are available. Does this introduce a bias in the analysis? Whilst this is done for one of the analyses, it would be good to quantify this for all major analyses, to reassure this is not the case.

The reviewer is correct. The other major analysis in the manuscript (aside from the tumor/normal analysis) is the association with clinical data. We have now compared the number of findings in the clinical analysis with sample size, with results available in **SI Figure S3.** Unlike the differential abundance analysis, we find that the number of findings in the clinical analysis is indeed correlated to sample size. This likely indicates that our clinical analysis is biased towards cancer types with more samples and generically underpowered, and that increasing the sample size (which will happen as more studies are published in the future) will lead to an increased sensitivity and a higher number of findings.

- In some analysis Mann-Whitney is used, in other Spearman, and the rational is not always clear. For example, in Figure S3 Spearman test might be more informative.

We agree with the reviewer. We initially implemented the Fisher test in Figure S3 because we were interested in the direction of change of statistically significant differentially abundant metabolites. We have now **revised Figure S2** (previously S3) and its **caption** to include both the Fisher test for significantly changed metabolites (now labeled with red color), as well as a Spearman correlation for all fold changes.

Notably, the correlations strengthen our finding of consistency across studies of the same cancer type.

- The analysis of Figure 3 and Figure S3 are discussed to support, respectively, the agreement between studies and the difference between cancer types. However, the tests and presentation of data are different. The analysis presented in Figure S3 is less clear than the one presented in Figure 3. Irrespective of which presentation is chosen, the same one should be provided to justify the conclusions of agreement in one case and differences in the other.

Here, we feel that we may have confused the reviewer with the purpose of the two figures, and thus the reason why the analyses are different. In Figure 3, we lay out a "landscape" of alterations across many cancer types, where the purpose is to see whether metabolites are commonly differentially abundant across *different cancer types* (*e.g.* kidney and prostate cancers). Indeed, there we find that such metabolites exist. In contrast, Figure S2 (previously S3) asks whether independent, replicative studies of *the same cancer* (*e.g.* two prostate cancer studies) produce identical results. Each figure is important in it's own right, so we have left them in place, but we have added a clarifying sentence on **Page 8.**

- It is not clear how the initially small library size and the fact that the sampling of metabolites is inherently biased would affect a median normalization approach. In other words, it is not clear if the median abundance is expected to be stable across samples (e.g. tumour samples versus normal). Since the wrong assumptions on normalization could introduce artefacts, this point would benefit from a clarification. A plot of the un-normalized data and comparison with a second normalization approach would also help to clarify.

This is a critical point. The purpose of our computational pipeline is to jointly analyze *pre-normalize*d published metabolomics data, not raw mass spectra. None of the studies we used provided (to our knowledge) raw spectra. In all cases, we apply a data-cleaning approach to the provided normalized data, essentially doing three things: 1) imputing the data as needed, 2) standardizing it to the median measurements as needed for visualization purposes only, 3) standardizing nomenclature. We then merge the data together into a joint dataset, to which we apply robust, non-parametric methods to handle heterogeneous data that may have been normalized in different (and perhaps incorrect) ways.

We have emphasized the above point in both the introduction (**Pages 4-5) and Methods (Page 19).**

We further describe that because our analytical approach makes no assumptions on the underlying distribution of the data, it is robust to many normalization procedures. However, we also emphasize on **Pages 4-5** that *incorrect* normalization by the original producers of the data (which is not detectable by us) will obviously lead to erroneous results from our code.

- Up to 60% of metabolites are found differentially abundant in some cancer types: how is a median normalization approach compatible with this?

The median normalization approach we implement is “within-metabolite,within-study” (see **Figure 2**). Because of these restrictions, the normalization actually has no effect on any differential abundance tests (these tests are non-parametric and simply rely on the rank-ordering of the abundance of the metabolite in each sample, so dividing each measurement by any constant will have no effect on the test). This standardization is implemented to make the data more readable and visualizable (and again, does not change the results of any analysis). We clarify and emphasize this point on **Pages 4-5** and in the **Methods on Page 19.**

The reviewer does raise an important point though: some studies have an exceptionally high proportion of differentially abundant metabolites. We mention this in the text on **Pages 7**, and speculate that (for the breast cancer data) it could be due to a large amount of imputation of the normal tissue data. Because we did not generate the original data (nor do we have or analyze the spectra), we cannot definitively say what is causing the phenomenon. However, our analytical methods (non-parametric differential abundance tests) are robust and not affected by large numbers of differentially abundant genes.

- Line 156 "Cancers share common genetic alterations despite..": this is introduced but not developed. The discussion and link of these results with results from previous studies showing genetic and genomic alterations is stated twice (here and in the introduction) but never developed.

We have added text to expand on the point and its importance to the treatment of cancer on **Page 8.**

- The overlap with previous findings from other metabolomics studies is stated but not clear. This is important given the convincing discussion the authors provide on variability of these assays. In figure S6 some metabolites are provided which were significant in one or two previous studies, and the results between the previous studies are compared, but it is not clear which of these metabolites were also significant in the present study as nomenclature does not match and is not always provided.

We have now elaborated on and clarified the comparison with prior findings in Figure S6. Briefly, Goveia *et al* used a meta-analysis approach to identify metabolites which were recurrently elevated/depleted in tumors relative to normal tissues. They assign a score to each metabolites, with positive (negative) scores for metabolites recurrently elevated (depleted). We now elaborate on **Page 9** on the similarities of our findings (*e.g.* recurrent elevation of lactate, taurine, depletion of myo-inositol, and a general elevation of metabolites (as opposed to depletion) in tumors). Importantly, **the analysis can be exactly replicated** by running the appropriate script in our code, which is provided to all readers on GitHub. (it is named RunCheckGoveia.R)

- Use of nomenclature is confusing. For example, kynurenine, one of the top hits in this study, is not an official ID or name. It does though correspond to at least 5 metabolites and it is not clear which one is referred to in this study.

We have chosen a single unique name for each metabolite, which is used through the manuscript. We have now added the Metabolite ID Dictionary to **Supplementary Table 1: Metabolite ID Dictionary** shows the IDs across a number of databases (when possible) to help guide users on the metabolite described in the text (e.g. Kynurenine, PubChem:161166, KEGG:C00328, HMDB00684, CHEBI:57959). We have also added text pointing to this resource at the end of the introduction on **Page 3.**

- The authors describe a big annotation effort which is excellent and needed. However, the IDs of the metabolites are not provided in key results table and figures, and the main results are provided using non official IDs. It would be extremely useful to provide a table where the aliases are mapped to the official IDs.

We now **provide a mapping in Supplementary Table 1: Metabolite ID Dictionary**; see previous comment. We have also added text pointing to this resource at the end of the introduction on **Page 3.**

- A pathway analysis is provided and this is the stronger part of the manuscript, but the use of the methods presented here, with respect to more traditional enrichment tests is

not discussed nor justified. The methods to estimate significance (p-values) are not clearly stated.

We do not test the statistical significance of our pathway analysis results. There are several reasons for this, but the most important is that the KEGG metabolite pathways were relatively small (~10 metabolites/pathway), and for many of these pathways, less than half of the constituent metabolites were even measured. This seriously reduces statistical power, to the point that an enrichment test is not meaningful. The results of the analysis are presented as observations, rather than as comparisons to any null model. **We have clarified this point on Page 10.**

- KEGG analysis: given the challenges in annotation, well discussed by the authors, a table with the metabolites which mapped to pathways and the ones which did not map should be provided.

We now provide mapping tables in **Supplementary Table 1: Metabolite ID Dictionary** so that readers will be able to see metabolites which we were able to map to KEGG pathways. We have also added text pointing to this resource at the end of the introduction on **Page 3.**

- Line 194 "we were able to 194 identify 70 enzymes (involved in 35 KEGG pathways) producing or consuming substrates in the 195 current study that are targetable by 57 approved or investigational compounds". This data are not provided.

We now provide this data as **Supplementary Table KEGG/DrugBank Pharmacology Coverage** .

- Line 219 "tumor/adjacent-normal tissue samples (i.e. each such pair was derived from the same patient). 219 Despite elevation of lactate in 6/9 studies, 82/343 (24%) paired tumor/normal samples contained 220 lower levels of lactate than their matched normal tissue counterpart." Does this imply that the paired analyses should always be done and the unpaired analysis should not be trusted? This needs to be discussed for clarity.

Thank you for the comment. No, both paired and unpaired analyses are informative, and in some ways give complementary information. In particular, the paired analysis normalizes for patient-to-patient variation when comparing tumor tissues to normal tissues. We have clarified this on **Page 12.** The purpose of the quoted text above was simply to highlight that while lactate is commonly elevated in many cancers, it is *not true* that each and every sample from a lactate-elevated cancer type is elevated in lactate.

Instead, elevation of lactate in a cancer type is a summarization of many data points, and it may be the case that the patients depleted of lactate (relative to normal tissue) are interesting in their own right.

- Line 237: is grading information/classification comparable between cancers. Is a meta-analysis looking at grade across cancer justified? This would need to be discussed

Thank you for the question. Tumor grade is a measure of how abnormal/malignant tumor cells appear histologically and/or morphologically. While the precise details of grading systems vary from one cancer type to the next, it is universally true that the grade of a tumor connotes how aggressive it looks, *e.g.* higher grade tumors are considered more aggressive. Our meta-analysis is focused on identifying metabolites correlated with features of aggressive disease (*i.e.* correlated/anti-correlated to grade or stage). In no case do we ever equate grades between cancer types (*i.e.* we do not imply that grade 3 breast cancers are in any way equivalent to grade 3 kidney cancers). Our approach is therefore suitable for comparison across cancer types. We have now included this information in the text on **Page 13.** We also describe in the **Methods** on **Page 21** that for one tumor type (prostate cancers), we converted Gleason scores to a numeric tumor grade (so as to make prostate cancers comparable with all the other cancer types in the meta-analysis).

- Linked to the above point, it is not clear why were other clinical variables not tested.

Thank you for the question. Manual curation of the studies indicated that the two most frequently reported/available clinical variables on tumor progression were stage and grade, with a larger number of data points for grade. We have now included this information in the text on **Page 13,** and have also included the results on the clinical analysis for stage on the website and in our results.

- Figure S7 is a very nice figure but KEGG ID would help in searches and comparison with other studies.

We have added the **Supplementary Table 1: Differential Abundance Score** which contains the raw information used to produce the figure with the added column identifying the KEGG ID for each pathway shown.