# Responses

## To Editor

### 1

Use superscript to match authors with institutions on Title page

Done

### 2

Page 18, List of supporting information: 1) As to not confuse the reader, supporting files should be referred to by the same name in the paper as they appear on the file labels.

Done

### 3

References: 1) Ref# 5 and 16 should be checked for completeness

Done

### 4

References: Remove color ink in superscript

Done by removing the hyperref package

### 5

Supporting Information: Suppl.xlsx file: File name and caption should be inserted into the table.

Done

### 6

Cover page for Supporting Information is required: Use heading, Supporting Information, followed by manuscript title, list of authors & affiliations, and a table of contents (list of supplementary components and captions). Number SI pages as S-1, S-2, etc.

Done

### 7

ORCID: Authors submitting manuscript revisions are required to provide their own validated ORCID iDs before completing the submission

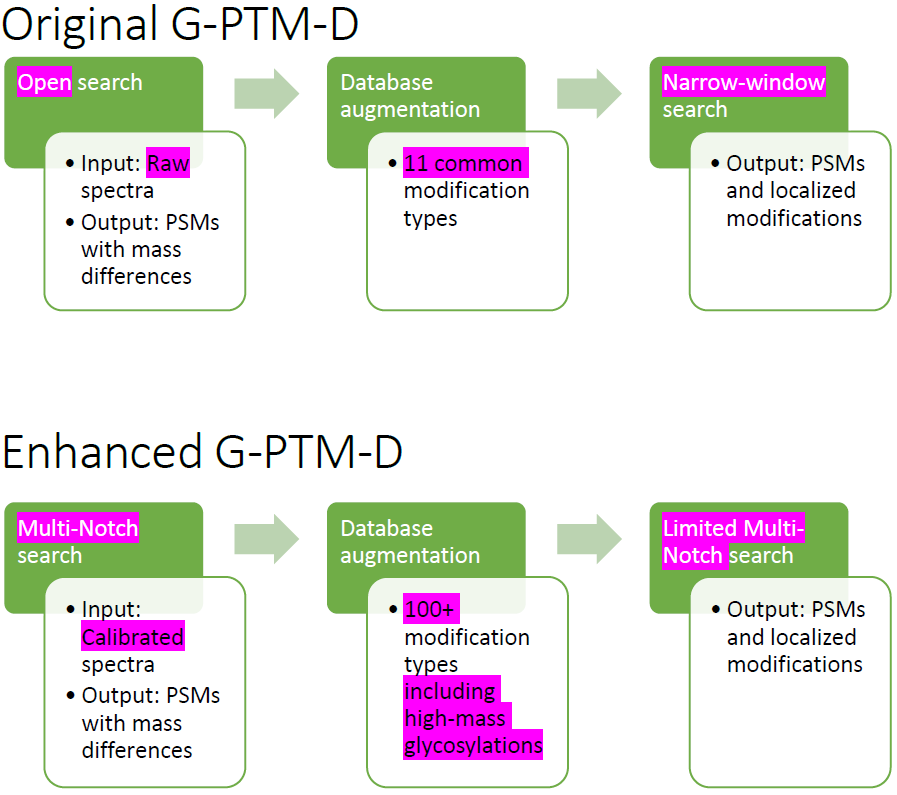
My ORCID is provided in the submission.

## To Reviewer 1

### 1

Fig 1 does not enhance my understanding of the changes in the GPTMD, or its processing steps. Have you considered a parallel figure in which the original algorithm is traced as compared to distinguishing features in the 'enhanced' approach. The figure itself, is also a little sloppy, boxes aligned on one margin are not aligned. For example, none of the boxes align on the right margin the effect is distracting.

Response: We thank the reviewer for this comment. Figure 1 has been replaced with a new figure in this revised edition. This new version addresses all of the reviewers comments. The revised figure is provided below.



### 2

Fig 2. Mass error and mass in figure 2 have units of Th? I am not familiar with this unit, does it mean thousands? Maybe mmu on the y and Da or m/z on X would be more

Response: The Thomson (Th) is a unit of mass-to-charge. To eliminate this possible source of confusion to the reader, we have substitute m/z in the revised figure.

### fig1-calibErrors.png3

What is the effect on a data set that is already well calibrated. The systematic drift in the ideal data set is large and not Gaussian. For a data set of Gaussian distribution or not as wide, the improvement is likely minimal, or potentially worse as it might recalibrate to some local-minima and not improve the results and/or processing. Potentially testable for quick comment by reprocessing data set after calibration (e.g. twice) or taking a random file from your lab you know has better m/z error distribution properties.

Response: The effect on a data set that is already calibrated is non-existent, due to the fact that the calibration procedure only accepts changes if a test database search results show improvement. We tested this and observed that multiple calibration tasks in MetaMorpheus do not change a file following the initial calibration round. This is noted in the calibration supplement, by adding the text:

If a well-calibrated file is given as an input, no changes are made. This is because the calibration procedure only accepts changes if a test database search results in improvement.

### 4

Notch improvement I think the implementation of notch processing is very well presented and timely. I am not satisfied that its utility in this context has been adequately addressed. All other aspects of the improvements in the GPTM are addressed in specific computed detail, but notch improvements are described but not readily verified (e.g. there isn't an accepted number of right and wrong answers for the test dataset). I would like to see a case in which you take a defined or previously published SILAC, dimethyl dataset and simply set the notches for the corresponding mass shift (delta 4, 8) etc. In this way you can very quickly and accurate define the catch rate and other factors from notch search as compared to simply defining light and heavy explicitly. In the case of SILAC data in particular, this may offer important advantages to partial incorporations due to arginine pathways which are hard to suppress

Response: We thank the reviewer for observing the wonderful benefits of the notch search and for the recommendation of the new experiment. We have expanded discussion of the selection of notches in the experiment (see below).

New manuscript text here:

The list of modifications added by G-PTM-D was generated based on a histogram of mass shifts from an open search. Any mass shift that was confidently identified 10 or more times, was considered as a possible modification. If a match was found in Unimod, the modification was added to our list.

## To Reviewer 2

### 1

One of the ‘new’ components described in the text is what the authors call “limited multi-notch”. In fact, almost 3 pages of the manuscript are dedicated to that part, including a sentence “We propose a limited multi-notch search…”. Unless I misunderstand something here, I cannot see how it is different from what has been implemented in most existing search tools for years. Every useful tool I know, Tandem, MSGF+, Comet, MSFragger, Sequest, etc. have an ‘isotope error option’. For example, in X! Tandem, it is specific using the following option: <note type="input" label="spectrum, parent monoisotopic mass isotope error">yes</note> You can read more about this option here: http://www.thegpm.org/tandem/api/spmmie.html In Comet: see here http://comet-ms.sourceforge.net/parameters/parameters\_201701/isotope\_error.php In MSFragger use isotope error option e.g. isotope\_error = /0/1/2 # 0=off, -1/0/1/2/3 (standard C13 error) I suggest that the authors shorten the text related to the limited multi-notch approach. For me it would be enough if they just said that they implemented a common strategy for allowing C12/C13 errors as in other tools.

The limited multi-notch search has functionality identical to the isotope error option in many current search tools. We made sure that the revised version is clear about this, and we removed any wording that may suggest that this is novel, such as "we propose". Still, the property that it eliminates some wrong matches parallels the discussion in the following multi-notch search section. We feel that it adds to the discussion and is necessary to make the manuscript complete.

We amended the manuscript as follows:

A multi-notch search is a remedy to the trade-offs caused by choosing a narrow (few ppm) or wide (few Da) precursor mass tolerance during a traditional narrow-window search. It is not novel, and many useful search tools such as Tandem, MSGF+, Comet, MSFragger, Sequest have this search functionality under names such as "isotope error option".

### 2

The authors do not really address the issue of FDR estimates, even at the PSM level, with respect to modified peptides. While the global (dataset-wide) FDR may be accurate, FDR for different subsets of PSMs is expected to vary widely. When performing searches such as multi-notch searches, open searches, or searches with many variable modifications or proteogenomic searches, it is important to realize that FDR for subsets of peptides with unusual modifications/mass-shifts or novel peptides will be significantly higher than that for the typical unmodified tryptic peptides. Thus, there is a need to perform class-specific (also referred to as group-specific FDR). In MSFragger, for example, this is addressed by modeling PSMs using a probability model that separates peptides into groups with different mass shifts. This ensures that PSM scores (probabilities) are re-calibrated to take into account the differences in the likelihood of observing a peptide with a particular modification (mass-shift) in the dataset. I recommend that the authors discuss this, and if not currently implemented in MetaMorheus, consider this for future work.

Response: The group FDR is implemented in the multi-notch search, by using notch-specific FDRs. Thus, G-PTM-D augmentation is conservative about adding modifications corresponding to exotic/problematic notches. MetaMorpheus .tsv outputs have a dedicated column called QValueNotch which displays the q-value for the specific notch.

We amended the manuscript as follows:

The results of a multi-notch search are assessed using notch-specific FDRs, computed individually for each notch, similarly to MSFragger. Users interested in further downstream analysis using Percolator can use the pre-formatted Percolator output, which is a tab-delimited file with the file extension “\_forPercolator.tsv”.

We will consider having modification-specific FDRs in future software releases! Thank you for the suggestion.

### 3

It is not clear if the authors present MetaMorpheus with G-PTM-D as a tool that can be applied for standard proteomic analysis (and not just as a complementary tool to look for more PTMs). If the former, they need to discuss the protein inference strategy in the context of multi-notch searches. As discussed in the MSFragger manuscript, a conservative option is recommended in which peptides with mass shifts are excluded from the protein inference step except those corresponding to C12/C13 errors, M+16, and N-term +42. While using 208 known modifications in MetaMorheus is not as risky as performing open search in that regard, perhaps not all of those 208 mass shifts should be considered (especially if MetaMorheus does not perform modification-specific modeling, see my point # 2 above). Please comment on this.

Response: We added new text to the manuscript that describes protein inference in MetaMorpheus.

Protein-level inference in MetaMorpheus uses all available observed peptides, including PTM-containing peptides. A minimal list of proteins to explain the observed set of peptides is constructed. Briefly, peptides that could only be a digestion product of one protein in the database (“unique peptides”) are used to construct an initial protein set, and any peptide that could be a digestion product of this initial set is associated with those protein(s) and removed from the pool of unassigned peptides. The remaining unassigned peptides are assigned to proteins using the greedy algorithm. Proteins with identical peptide sets are grouped. Protein group FDR is estimated with the top-picked method. In this approach, each protein is paired with its reversed decoy. The PSM with the higher notch q-value (or lower score, in the case of a tie) is retrieved for each protein in the target-decoy pair. The protein associated with this PSM is considered the lower-quality hit of the pair. The protein group containing this protein is removed from the protein group set. Protein groups are then ranked by the notch q-value of their best PSM, and protein q-value is calculated as the number of decoys divided by the number of targets in the ranked list.

### 4

The authors discuss advantages of multi-notch searches compared to open searches. While I agree with their statement as it relates to ‘bare’ open searching, I want to point out that some of those concerns for open searches can and should be addressed by downstream tools. For example, on p.14, the authors illustrate their case using a peptide ATPARA… that was identified in open search with a score of 17.231, 1 missed cleavage, and large and unexpected negative mass-shift of -416.303. The correct PSM in this case is a shorter peptide, slightly lower scoring (16.246), without missed cleavage, and having an explainable mass shift of 79.9655. As described in the MSFragger manuscript, open search results (and any search results in general) have to be carefully modeled post-database search using downstream statistical tools such as TPP, Percolator, etc. For example, the extended mass model of PeptideProphet (which also uses the number missed cleavages) would likely penalize the incorrect identification (with the unusual -416.303 mass shift and 1 missed cleavage) to the point where it would not pass strict FDR cutoffs. It is true that the post-database search modeling in its simplest form would not rescue the second-best match to this spectrum (correct PSM) if only the top match is considered. However, at least it would not let this false identification to get through. If not this one ID in particular, it would reduce significantly the number of such mistakes dataset-wide. I hope the authors can discuss the importance of post-database search tools in addressing some of the limitations of open searches, or any searches for that matter including multi-notch searches.

Response: We added a note about the PeptideProphet extended mass model to the manuscript.

We amended the manuscript as follows:

It is important to note that there exist post-processing tools that would push identifications with unknown mass shifts and large numbers of missed cleavages down the list, such as the PeptideProphet[REFERENCE] extended mass model.

### 5

The results shown in the manuscript were obtained on a pretty powerful workstation with 24 cores and 128Gb RAM. Are there any RAM limitations for running MetaMorheus on large datasets like the multi-fraction Jurkat dataset?

Response: MetaMorpheus is designed to work on machines with 16 GB RAM. When a large amount of memory is avaiable, a user is able to select parallel searches which would utilize the available memory efficiently, thus speeding up searches of mulitple files at the same time.

We amended the mansucript as follows:

It has attractive properties like Morpheus, such as the ability of running large-scale search tasks on memory-restricted machines.