**Frequently Asked Questions (FAQs)**

This document serves to answer additional questions that might arise when reading our work, entitled “A Causal Model of Ion Interference enables Assessment and Correction of Ratio Compression in Multiplex Proteomics” (Madern *et al.).* Most of the questions herein arerephrased questions that have previously been asked by reviewers during the peer review process. Consequently, a few words of due acknowledgement: We are grateful to all the reviewers (and editors) for their insightful comments, valuable suggestions and more generally, for their time in reviewing our work. With their help, we could further improve our manuscript and resolve remaining unclarities.

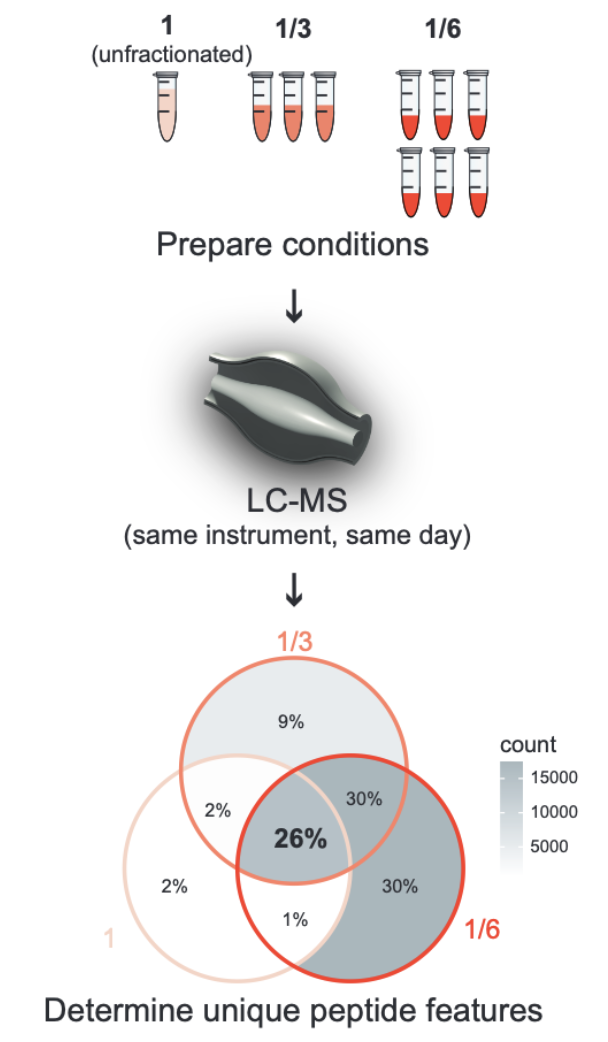
Note that additional FAQs might be added to this document in the future ☺

**FAQ 1:**In Figure 2 and corresponding Figure S2, why are the OIL distributions, ROC-curves and fold change distributions so different between the various panels (e.g. Figure 2A vs Figure 2B; or Figure 2G vs Figure 2H)?

RESPONSE:   
The main reason for these differences lies in the way the data was filtered prior to the analysis, as illustrated in Figure S1. To enable a fair comparison between all conditions within an experiment (panel), the data was filtered for unique peptide features that were independently quantified in all conditions of an experiment. Had we not done so, the results would have been heavily influenced by the varying depth of quantification, stemming from differences in measurement speed and measurement time, which would have hindered our ability to isolate the effect solely caused by variation of the parameter of interest (e.g. sample complexity). In other words, this filtering strategy allowed us to monitor changes brought about by variations in the parameter of interest with no other confounding effects present.

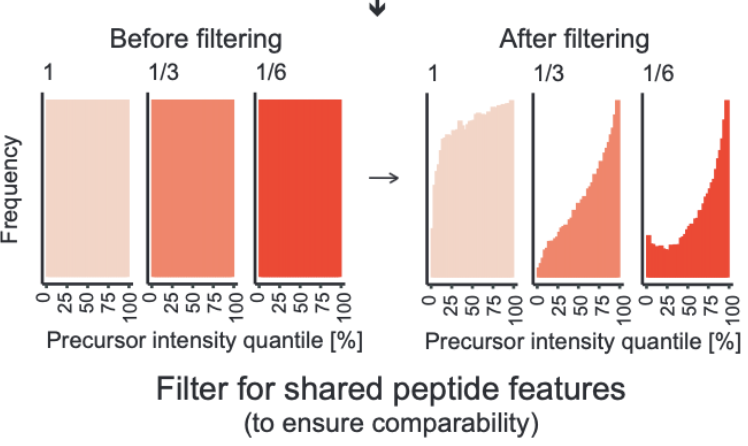
However, the same filtering strategy for shared peptide features resulted in OIL values (and in extension, fold changes and ROC-curves) diverging between experiments (panels). To illustrate this, let us take the experiment comparing different sample complexities (as illustrated in Figure S1) as an example:

**Figure S1 (sub-panel)**

**

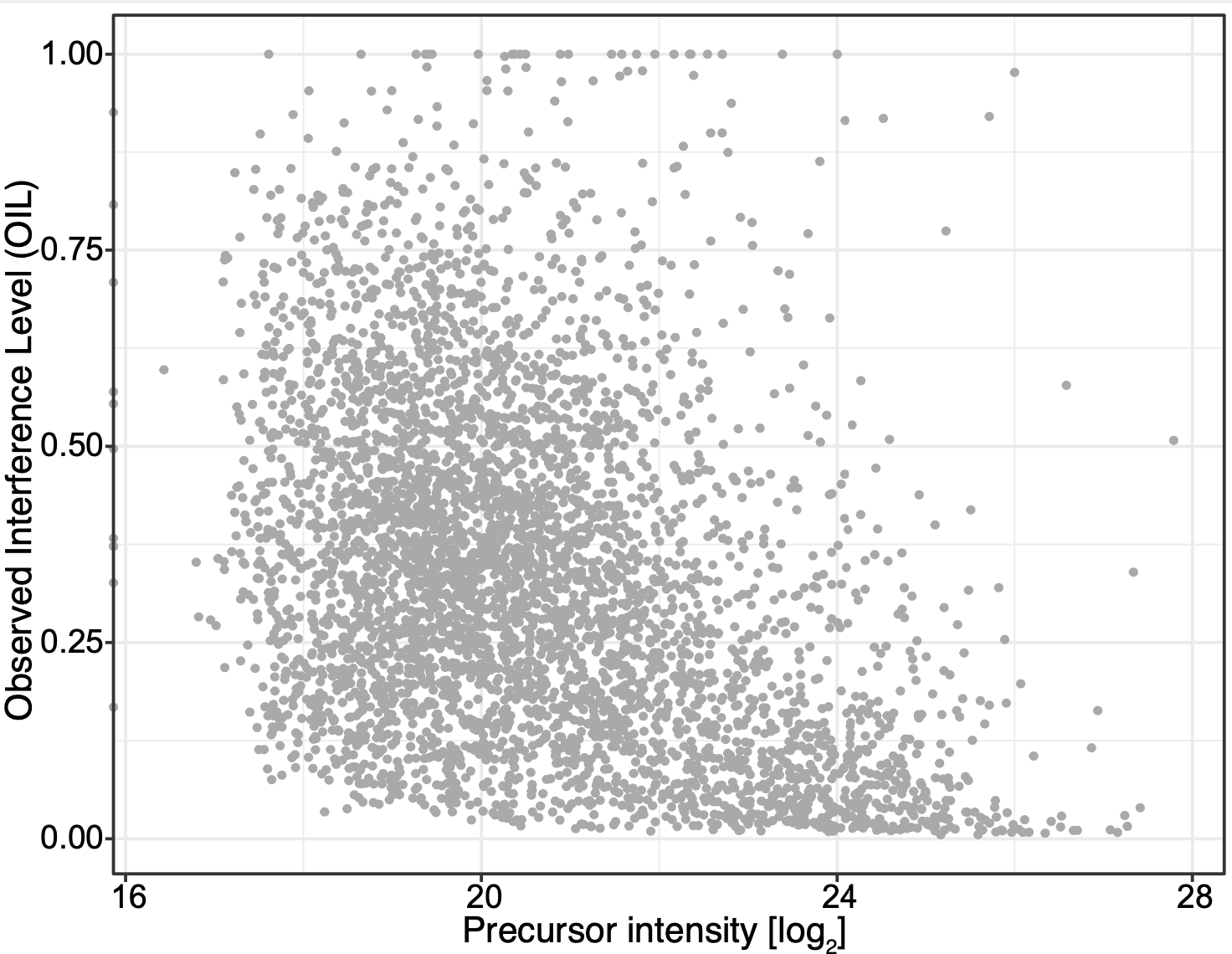
The Venn diagram shows that only 26% of all unique peptide features identified in this experiment were actually quantified throughout all compared conditions. Most peptide features quantified in the low-complexity samples (complexity ⅙) were therefore discarded during filtering, because they were not quantified in the unfractionated sample (complexity 1). As can be seen, the peptide features that pass the filtering almost make up the entirety of features that have been quantified in the unfractionated sample. Due to the nature of DDA in LC-MS, when measuring a high-complexity sample (like the unfractionated sample) with an overabundance of MS1 precursor ion features to select for MS2 quantification, the mass spectrometer prioritizes the most abundant precursor ions for subsequent MS2 quantification. Because of this behavior and the limit in measurement time, many lower-intensity precursors may go unquantified. We can therefore assume that these 26% contain peptide features with high MS1 precursor ion intensities. Importantly, when filtering for those 26% in the low-complexity data, peptide features with above-average precursor intensities are enriched at the same time. The resulting quantile distributions (Figure S1) confirm this selection bias for high-intensity precursor intensities from the filtering:

**Figure S1 (sub-panel)**

**

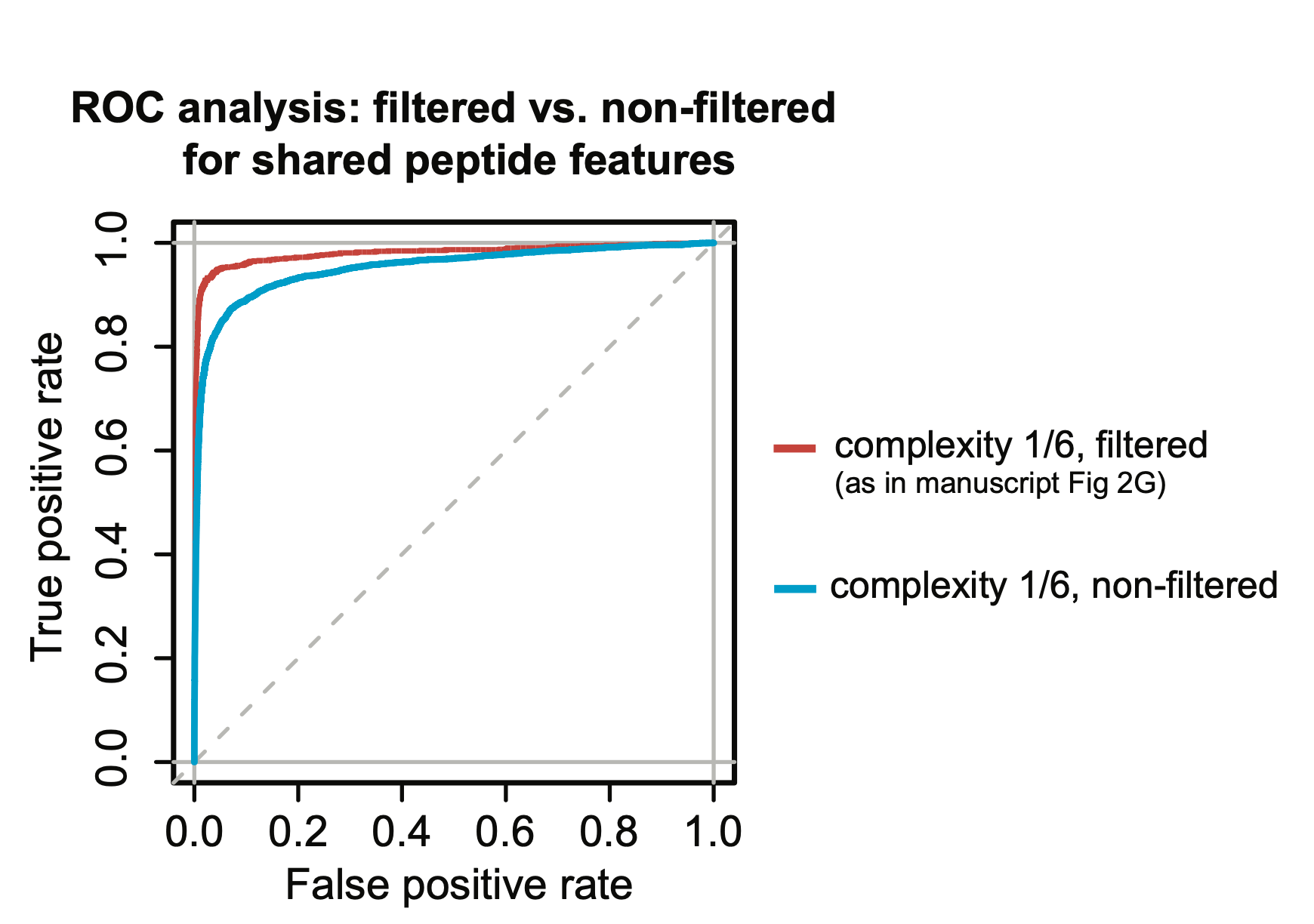
Please note that, as previously mentioned, this “bias” was intentional to ensure a fair comparison between the different conditions within an experiment. However, from an external viewpoint, this constitutes a "genuine bias” which has some consequences: By indirectly selecting for high-intensity peptide features, the analysis also selects for below-average OIL values. This is supported by the data; the following plot depicts the relationship between precursor intensities and OIL values, calculated for yeast PSMs in the low-complexity samples (complexity ⅙):

**FAQ 1 - Extra Figure 1**

**

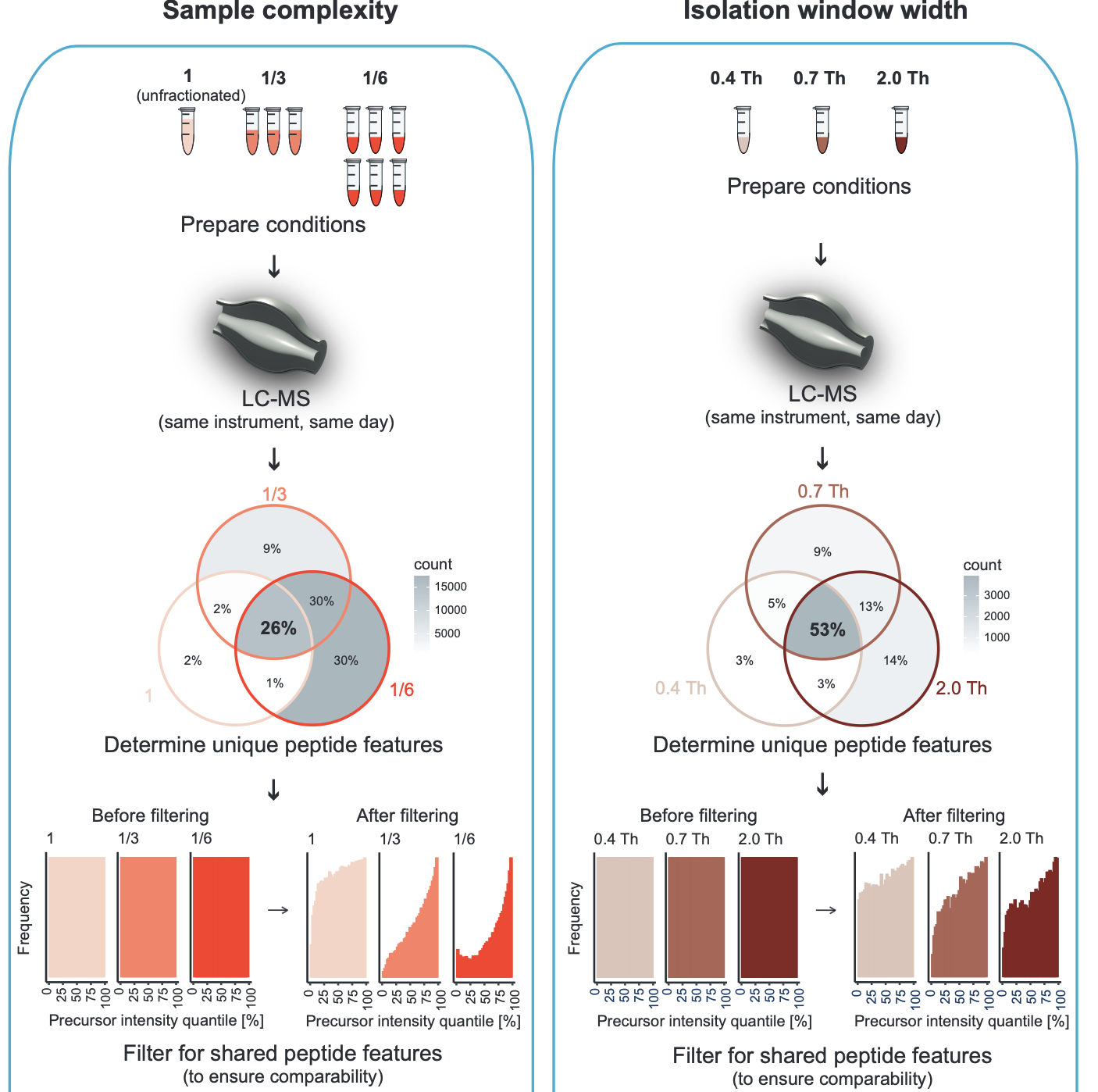
Naturally, the lower the OIL values, the less-compressed the corresponding yeast fold changes will be, which in turn improves the overall classification results in the ROC analyses. Moreover, high intensity also leads to better precision (as shown in Figure S3 in the revised manuscript), which in turn further improves classification. To illustrate these filtering effects on the ROC analysis, the following plot depicts the classification results for the low sample complexity condition, performed for the filtered data (in red, as already shown in Figure 2G) as well as for the corresponding unfiltered data (shown in blue):

**FAQ 1 - Extra Figure 2**

**

It can be seen that the unfiltered data exhibits much worse classification. Now the crucial point is that the aforementioned selection bias towards high-intensity peptide features may not be equal in all the experiments (i.e. all panels in Figure 2) we conducted. Our data indicates that this filtering bias is much less pronounced in comparisons with similar depth of quantification in the compared conditions, e.g. for comparison of different isolation window widths, as illustrated in Figure S1 in the revised manuscript (see histograms after filtering at the very bottom):

**Figure S1 (sub-panel)**

**

Therefore, the implemented analysis strategy, while ensuring a setting for fair comparisons between all conditions within an experiment (panel), simultaneously prevents a fair comparison between experiments (panels).

Additionally, we would like to note that we can not rule out experimental batch-effects. All measurements within an experiment (panel) have, for the sake of comparability, been performed on the same day on the exact same LC-MS system using the same peptide solution(s) injected from the same tube(s). However, the entirety of measurements was conducted on different days (sometimes with months apart), on different Orbitrap instruments and possibly with varying measurement quality. Also, to prevent using up any one sample completely (in case more was needed later), the sample material required for all measurements in Figure 2 was taken from different sample pools of low sample complexity (i.e. the six samples P1-P6 with sample complexity ⅙). All of this may lead to batch effects. If comparability between experiments was desired, it would have been ideal to measure everything on the same LC-MS system, on the same day, using the exact same peptide solution. Nevertheless, the filtering strategy described above would have made comparisons between experiments (panels) meaningless.

**FAQ 2:**Why not use another metric for MS1 isolation purity like PIF instead of PPF? What is done to ensure that the peaks matched by mass windows belong to the right isotopic distribution?

RESPONSE:   
The algorithm for calculating the “Precursor Purity Fraction” (PPF) does not include any extra measures to ensure that all supposed isotope peaks in an isolation window belong to the precursor ion cluster, other than the stringency of the implemented selection criteria itself. For the expected +1 and -1 isotopes, the corresponding selection ranges are determined by a mass error tolerance of ± 0.00125 Th with respect to the predicted m/z, calculated on the basis of the observed m/z of the precursor ion in the MS1 scan. We think that any co-occurrence of non-precursor ions in these ranges, given these rather narrow margins, is exceedingly rare. Additionally, our modeling workflow (as it is available on GitHub) allows for custom mass error tolerances that can be set even more stringently. It is further noteworthy that, with typical isolation window ranges in modern DDA quantification (here assumed to be smaller than 1 Th), this potential issue only concerns precursor ions of charge states z=3 or higher.

Still, we consider that this could be a potential angle for improvement, although possibly not easy to implement. Alternatively, PPF can easily be replaced by any other MS1 isolation window purity metric within the modeling workflow, in case another metric should lead to a better model prediction result. It should be mentioned that the biggest problem with other metrics was that they would result in NA values (or 0) when the precursor ion peak was not found in adjacent MS1 scans. This automatically excludes all those PSMs (sometimes more than 5% of all) from the modeling and thus EIL estimation. PPF successfully circumvents this problem by imputation of precursor ion peak intensities “at noise level” during the calculation, instead of automatically assigning a missing value. This ensures that all PSMs are amenable to our modeling workflow, meaning that we do not have to discard PSMs. In fact, this was the main reason why we considered it necessary to create our own metric (PPF).

**FAQ 3:**Is the modeling workflow for interference estimation also applicable to other sample types with different proteomes or peptide characteristics, for example samples featuring specific modifications on peptides that might behave differently during HCD fragmentation?

RESPONSE:   
Yes, this is actually a strength of our method and algorithm. We found the method to be applicable to any TMT sample, measured via MS2-based quantification on Thermo Orbitrap instruments. This includes samples with different ranges of sample complexity, and varying peptide characteristics. We tried to illustrate the general applicability of the method in Figure S9 wherein its use is showcased for (already published) yeast TKO datasets, FAIMS-MS2 quantified samples as well as acetyl and phospho peptide-enriched samples. Further, we have already successfully applied the method to datasets from murine T-cell PTM studies created in collaboration with other labs (data not shown in this study; separate manuscript in preparation).

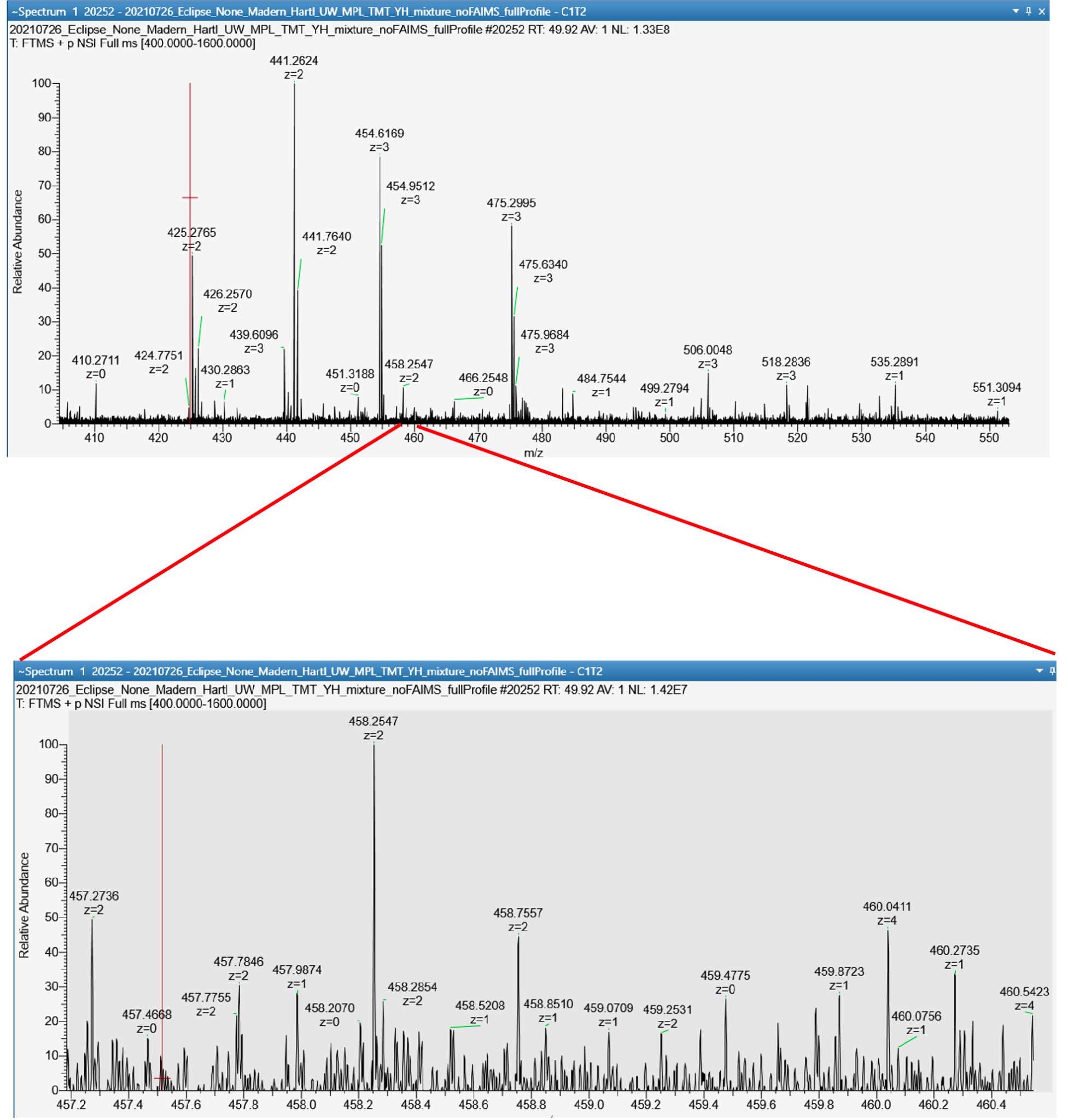
Concerning the point of other characteristics”: While we have not observed it for phospho (STY) and acetyl (K) peptide-enriched samples in our data, we can imagine that other PTMs or chemical modifications might strongly affect a peptide’s fragmentation efficiency. This would certainly have to be taken into account in the modeling (in order to determine adequate peptide classes with distinct fragmentation efficiencies; see Figure 4C-H). Fortunately, our method can easily adjust to this demand: The workflow (as it is available on GitHub) already contains an optional parameter that lets users specify potential PTMs or modifications that might influence the fragmentation efficiency of peptides in their sample. These modifications are then considered by the decision tree algorithm for determining empirical peptide classes with distinct fragmentation efficiencies. The respective parameter in the workflow (on GitHub) is called ptm\_pattern, and requires a regular expression pattern that matches the modification of interest, as contained in the column “modifiedSequence” (or suchlike) in the PSM-table used as input.

**FAQ 4:**Orbitrap instruments also allow for measurements with the noise filter completely turned off. If turned off, are the interfering ions there that would otherwise be filtered out with the noise filter turned on?

RESPONSE:

We were wondering about the same question, which is why we performed a measurement without the noise filter on (i.e. in “full-profile mode”, as opposed to “reduced-profile mode”). We learned that the noise is extremely complex. The file size per measurement matched this observation - the raw file was 26 GB (as opposed to 2 GB with noise filter on). The figure below shows two screenshots of this measurement from the same MS1 scan with different zoom levels.

**FAQ 4 - Extra Figure 1**



Due to the overabundance of peaks along the whole scan range, we found it difficult to discern the “true” interference ions in the noise from “electrical” noise. Even the charge state of peaks at lower intensity could often not be determined or were annotated as 1 or 0. This made it especially hard to figure out if any single peak originated from “electrical” noise or “real” peptide ions. We therefore cannot fully answer this question but the observed noise pattern suggests that interfering peptide ions could be present.

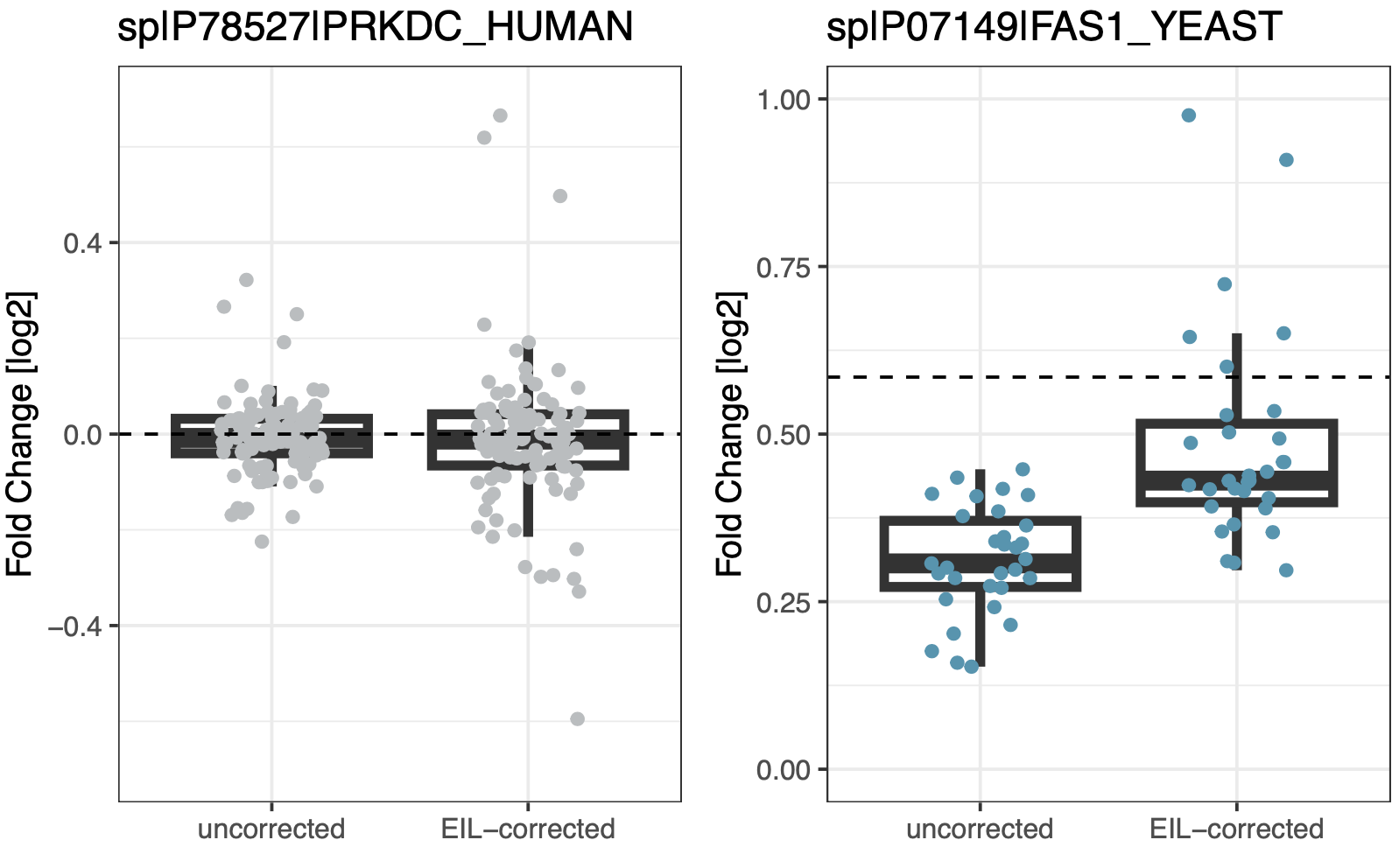
**FAQ 5:**The authors have investigated the effect of interference correction on protein level. What about the effect of this procedure on peptide level, especially with regards to accuracy and precision?

RESPONSE:

First, a short summary on what we observed on protein level: In short, we saw an increase in fold change accuracy, accompanied by a decrease in fold change precision, especially in the human (non-DE) protein population. These two effects seemed to balance out in their combined effect on correct classification of DE and non-DE proteins.

On peptide level, we see the same behavior. We have investigated the distribution of estimated fold changes for peptides from the same protein, before and after interference correction. We looked at the yeast and human proteins with the respective highest number of peptides quantified in the samples of reduced sample complexity (P1-P6). The following two plots compare the distribution of uncorrected and corrected peptide level fold changes for those two proteins (left: human protein; right: yeast protein). Fold changes [log2] were calculated as the differences in average log2-intensities of groups 100:9 and 100:6. The dashed line represents the ground-truth fold change. Each data point corresponds to a single peptide with a unique sequence. Note that, in order to obtain peptide level information, interference-corrected as well as uncorrected PSM intensities were first aggregated to peptide level by summation (only relevant if multiple PSMs per unique peptide sequence were quantified).

**FAQ 5 – Extra Figure 1**

**

As can be seen, the EIL-based interference-correction approach improves accuracy, but leads to higher variance in the calculated log2-fold changes (i.e. reduced precision) for peptides from the same protein. We speculate that there might be better ways to calculate interference-corrected fold changes utilizing EIL values - especially to curb the effect of the correction on data points with already high within-group variance prior to correction, which are more likely to become outliers after correction. For example, something similar to DeSeq2’s shrinkage of effect size of fold change estimates for highly variable features could be promising. This is something that could be further explored in the future.

**FAQ 6:**In Figure 4H, the addition of the final parameter noiseEstimate to the model seems to only moderately affect the overall prediction results. It also appears to affect one empirical peptide class more than the others (i.e. the light-blue points in Figure 4G on the lower-left quadrant). Could the noise be omitted in the modeling procedure and still produce similar results in estimating interference at MS2 level?

RESPONSE:   
First, here is Figure 4E-H for reference:

**Figure 4E-H**

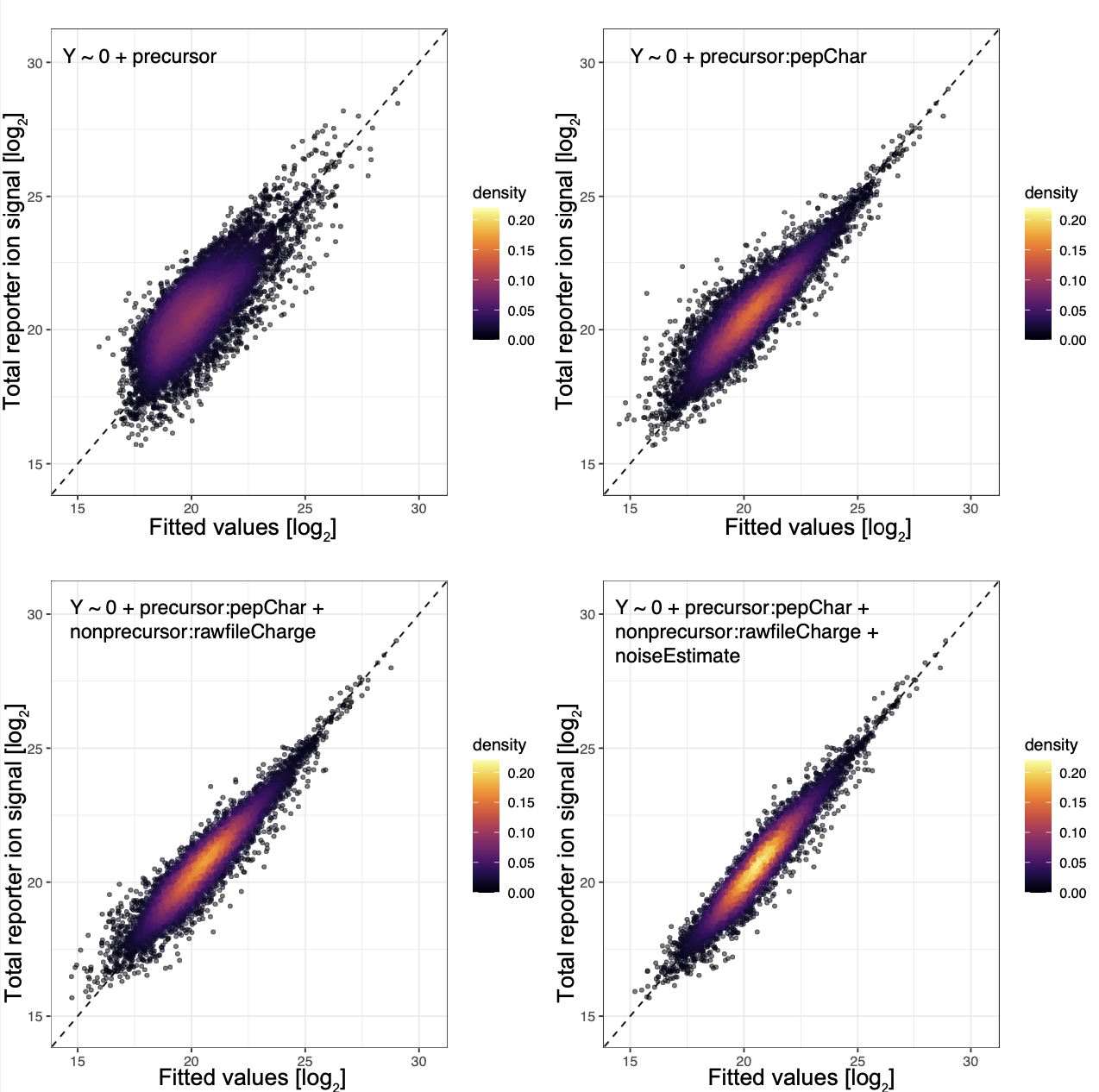
A diagram of a signal

Description automatically generated with medium confidence

To answer this question, we need to address several points regarding the results of Figure 4E-H:

A) We acknowledge that the style of plotting used in Figure 4E-H made it rather hard to see differences, especially between the last two panels (Figure 4G and Figure 4H). Therefore, we have plotted the same figures again, this time colored by point density instead of peptide class:

**FAQ 6 – Extra Figure 1 (recoloring of Figure 4E-H)**

**

Please note that the color scale for point density has been held constant in all four plots, and the number of points in each plot is the same. While the overall shape of all data points does not change much, we think that one can now better appreciate that the points are closer distributed along the diagonal (i.e. perfect prediction) in the final model prediction (bottom-right).

B) Indeed, the final parameter “noiseEstimate” seems to affect some data points, particularly those of the light-blue colored peptide class/classes, more than others. We ascribe this to the fact that some peptide classes are closer to the lower end of the intensity range, and therefore much more susceptible to “noise”, than peptide classes with generally high precursor abundance. For the light-blue class/classes in Figure 4E-H, this appears to be the case. Based on an additional analysis which was ultimately added as supplemental figure Figure S8 (showing a comparison of the modeling results for both high-intensity precursor peptides as well as low-intensity precursor peptides) we could show that the contribution of “noise” to the reporter ion signal is the strongest (i.e. more than 20% of total signal) for the lowest 25% PSMs ranked by precursor intensity. We think that this extra analysis illustrates why PSMs on the lower end of the intensity range are more susceptible to the addition of the model parameter “noiseEstimate” than others.

C) Further, we think there is a valid reason as to why the difference between prediction results of the last two models could be less than initially expected. We believe that the “visible interference” in the MS1 window, described by the parameter named “nonprecursor:rawfileCharge”, is presumably well correlated with the variable “noiseEstimate”. This stands to reason, as “noise” is nothing else but interfering ions that did not pass the “noise filter” due to low intensity. Importantly, when explanatory variables are correlated in linear modeling, and one is entirely omitted, the one remaining will cover part of the explanatory power that would otherwise (in the full model) be ascribed to the other. This phenomenon is called “omitted variable bias”, and we think that it might also have played a role here.

Finally, to the question regarding the omission of the variable “noiseEstimate” in the model: While this would simplify things, we do not think that this is desirable, as “noise” appears to contribute substantially to the total reporter ion signal (please see Figure S7 and new figure Figure S8), especially for low-intensity PSMs. Even if the parameter for visible interference in the isolation window (named “nonprecursor:rawfileCharge”) can explain some of the variance caused by noise, the omission of “noiseEstimate” would automatically neglect all PSMs with PPF=1, which would then always be declared interference-free in such a reduced model. Our data shows well that PPF=1 does only rarely imply zero reporter ion interference at MS2 level (please see Figure S5). We think that the greatest strength of our interference prediction approach is that it can actually estimate the degree of interference for precursors that initially appear interference-free based on the isolation window.

**FAQ 7:**How does the computational interference correction approach compare against the “ideal” FAIMS-MS3 quantification in terms of accuracy, precision and correct classification in statistical testing?

RESPONSE:   
We think that this question should be explored in a follow-up investigation using an experimental design that contains only the minimum required number of peptides that differ in relative abundance across groups (e.g. provided by a spike-in of a handful of peptides with differential abundance, and not a whole proteome). This would ensure that there are zero unwanted effects from between-sample normalization. As mentioned in our results section we reason that some unwanted interference caused by yeast peptides in our study (although only one-tenth that of the human-derived interference) affects the between-sample normalization in a way such that differences in normalized yeast intensities are artificially diminished between groups. This bias, although small, makes the MS2-recorded data in our study not ideal for comparison with FAIMS-MS3 in terms of accuracy.

However, regarding the challenge of DE analysis and correct classification of DE proteins, the results presented in our manuscript already provide some valuable insights. From Figure S10 we can deduce that interference-corrected (MS2-quantified) reporter intensities perform comparable to uncorrected (MS2-quantified) intensities in differential expression testing due to the discussed tradeoff between accuracy and precision. On the other hand, from Figure 2I we can deduce that FAIMS-MS3 outperforms MS2 in the classification of differentially expressed unique peptide features quantified in all conditions. Combining these results, we can conclude that FAIMS-MS3 outperforms interference-corrected MS2 quantified intensities in differential expression testing and correct classification. We think that this further extends to the protein level (although the additional measurement speed in MS2 quantification could provide some boost in precision due to a potential increase of peptide numbers per protein measured).

**FAQ 8:**Regarding the metric “Observed Interference Level” (OIL) , why not use the “Interference Free Index” (IFI) from the Gygi lab?

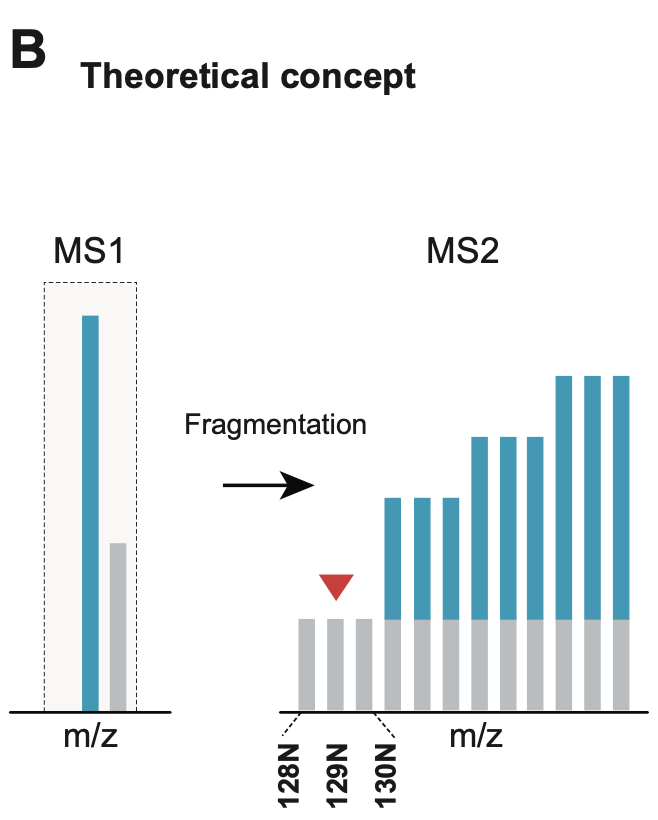
RESPONSE:   
The “Interference Free Index” (IFI) was defined for signal-to-noise ratios (<https://pubs.acs.org/doi/10.1007/s13361-016-1434-9>). We have defined the “Observed Interference Level” (OIL) using “raw” intensities as the mass spectrometer reports them. While signal-to-noise ratios and raw intensities correlate well (as signal-to-noise ratios use raw intensities in the numerator during calculation), they are not identical, and there is no single scaling factor that will convert one to the other, even for peaks from the same spectrum. Thus, (1-OIL)-values will slightly differ from IFI-values. We think that this warrants the definition of a new metric as it prevents any confusion in the future.

**FAQ 9:**The experimental design features more than 10 times the relative human peptide abundance over yeast peptide abundance (w/w). This will inevitably lead to a reduced number of quantified yeast peptides. Why not increase the relative amount of yeast in the experimental design?

RESPONSE:   
The design of our artificial two-proteome multiplexing experiment was guided by two primary aims: First, we needed a dataset where reporter ion interference was behaving in a consistent and controlled manner, e.g. as uniform as possible, across all samples. Second, we required a sufficiently large number of peptides for which we could study this consistently-behaving interference, i.e. measure it directly and gauge its effect on differential expression testing and ratio compression - yet without compromising the first assumption. These two aims culminated in the experimental design showcased in our study.

We want to stress that, if the yeast peptide amount in our experiment would have been much higher (or even as high as human), reporter ion interference would not have behaved in a consistent and predictable manner. Instead, it would have been an arbitrary mixture of yeast and human signal, depending on which ions were co-fragmented - sometimes resulting in a more uniform interference background if more of human peptides were co-fragmented, and sometimes resulting in an interference background with differential abundance if more yeast peptides were co-fragmented. This was undesirable, because we needed a system where we could directly assess the degree of reporter ion interference by knowing exactly how interference behaves. In our experimental design, that was possible for any quantified yeast peptide (below in turquoise):

**Figure 1B**

**

Thanks to the overabundance of human peptides, we could read out the reporter ion interference in channels 128N, 129N and 130N, and then extend this knowledge to any other channel in the experiment, as any interference signal was very unlikely to come from yeast. The same concept makes the commercially available TKO standards (e.g. TKO6, TKO11, etc., published by the Gygi lab) a popular system where the degree of reporter ion interference of the whole PSM can be assessed completely by the level of TMT signal detected in channels where a specific protein should be absent.

Now we want to clarify why it was so important for our study to reliably assess the degree of interference in MS2-scans (for quantified yeast peptides). The two main reasons are listed here:

1) This circumstance made the “Observed Interference Level” (OIL) values in our study a reliable metric for the degree of interference at MS2 level. The authors of the TKO studies (Gygi lab, <https://pubs.acs.org/doi/10.1007/s13361-016-1434-9>) also make use of this concept in their system to calculate the so-called “Interference-Free-Index” (IFI), which is highly related to OIL.

2) By knowing that any interfering ion in the MS1 isolation window is (most likely) human and therefore causes uniform interference, we could actually model the observed reporter ion interference at MS2 level in dependence of the characteristics of the MS1 isolation window. This linear regression analysis made it possible for a causal exploration of how interference works. The gained knowledge was ultimately used to extend the linear model approach in order to accurately predict the degree of reporter ion interference at PSM-level in any MS2-based multiplex proteomics experiment.

**FAQ 10:**Does the enrichment efficiency (here defined as the relative abundance of modified peptides in PTM-enriched samples, ranging from 0 to 100%) of phospho peptides, acetylated peptides or any other modified peptides matter for the interference modeling approach?

RESPONSE:   
We ensured that the enrichment efficiency does not have an impact on the general applicability of the interference modeling workflow. That is because any precursor peptide’s unique fragmentation efficiency, regardless if modified or not, is adequately accounted for in the model: The workflow, as it is available on GitHub, contains an optional parameter that lets users specify potential PTMs or modifications that might influence the fragmentation efficiency of peptides in their sample. These modifications are then considered by the decision tree algorithm for determining empirical peptide classes with distinct fragmentation efficiencies. The respective parameter in the workflow (on GitHub) is called ptm\_pattern, and requires a regular expression pattern that matches the modification of interest, as contained in the column “modifiedSequence” (or suchlike) in the PSM-table used as input. In the code, the usage of this optional parameter and its purpose are extensively described.