Re-investigating the correctness of decoy-based false discovery rate control in proteomics tandem mass spectrometry

Jack Freestone[1], William Stafford Noble[2,3], and Uri Keich[1*]

[1]School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Australia
[2]Department of Genome Sciences, University of Washington, Seattle, WA, USA 98195
[3]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA 98195

February 16, 2024

∗ Corresponding author: uri.keich@sydney.edu.au

Abstract

Traditional database search methods for the analysis of bottom-up proteomics tandem mass spectrometry (MS/MS) data are limited in their ability to detect peptides with post-translational modifications (PTMs). Recently, "open modification" database search strategies, in which the requirement that the mass of the database peptide closely matches the observed precursor mass is relaxed, have become popular as a way to find a wider variety of types of PTMs. Indeed, in one study, Kong *et al.* reported that the open modification search tool MSFragger can achieve higher statistical power to detect peptides than a traditional "narrow window" database search. We investigated this claim empirically and, in the process, uncovered a potential general problem with FDR control in the machine learning post-processors Percolator and PeptideProphet. This problem might have contributed to Kong *et al.*'s report that their empirical results suggest that false discovery (FDR) control in the narrow window setting might generally be compromised. Indeed, reanalyzing the same data while using a more standard form of target decoy competition-based FDR control, we found that, after accounting for chimeric spectra as well as for the inherent difference in the number of candidates in open and narrow searches, the data does not provide sufficient evidence that FDR control in proteomics MS/MS database search is inherently problematic.

Key words: FDR control, Percolator, PeptideProphet, peptide analysis, open search, narrow search, group-wise analysis.

1 Introduction

In tandem mass spectrometry, a traditional "narrow window" database search is the *de facto* method for the detection of peptides. In this approach, each spectrum is compared against a relatively small subset of the reference database, which includes all the peptides whose masses lie within a narrow window of tolerance from the observed mass of the precursor associated with the spectrum. The best peptide-spectrum match (PSM) within the list of candidates is found and retained, together with its score[1]. However, this approach is limited in its ability to detect peptides that harbor post-translational modifications (PTMs). Such peptides pose a challenge for a narrow window search

because even partially accounting for a subset of possible PTMs involves augmentations to the database that quickly result in an impractically large database. As an alternative, many database search tools employ an "open modification" strategy. In such a search, each observed spectrum is compared against a substantially larger list of candidate peptides by using a much larger mass tolerance window than in a narrow search, often hundreds of Daltons in width[2]. Recently, these open modification search tools have become very popular, and a great deal of effort has made open searching very efficient[3,4,5,6].

Although an open search allows finding peptides altered by PTMs, the associated increase in the number of peptides that each spectrum is compared against inevitably leads to some loss in statistical power to detect unmodified peptides. We therefore set out to investigate this tradeoff by first comparing the number of peptides we can detect using the scored PSMs from open versus narrow searches executed by three search engines: Comet[7], MSFragger[3], and Tide[8]. As described below, we found that neither strategy clearly discovers more peptides than the other.

In practice, rather than relying on the search engine's own score, scientists often first rescore the PSMs via a machine learning post-processor. Percolator[9] and PeptideProphet[10] are among the most popular such methods, though it is common for these search tools to implement their own, internal rescoring methods[4,6]. As detailed below, redoing the comparison between narrow and open search after the PSMs had been rescored by Percolator and PeptideProphet, we found, consistent with Kong *et al.*[3], that we typically detect more peptides when using the open search PSMs than when using the narrow search ones.

To better understand the results of this comparison, we further investigated the statistical assumptions that these methods make. In general, when comparing the statistical power of competing procedures it is critical to verify that the procedures are equally effective in accounting for type I errors, or false discoveries. By far, the most common approach for handling type I errors in mass spectrometry analysis is through controlling the false discovery rate (FDR) via target-decoy competition (TDC). In TDC each spectrum is searched against a concatenated database for its best PSM. In addition to the real (target) peptides, the concatenated database includes decoy peptides that are generated by randomly shuffling or reversing the target peptides. The number of decoy PSMs is subsequently used to estimate and control the FDR among the reported target peptides[11,12,13].

Thus, the validity of our open-versus-narrow search comparisons hinges on whether TDC is equally effective in controlling the FDR in all the setups we considered. While there is a strong theoretical basis for the validity of TDC[12], the assumptions that TDC relies on might not hold in practice in some cases. In particular, Kong *et al.* recently argued that the "equal chance" assumption, namely, that an incorrect PSM is equally likely to involve a target or a decoy peptide, appears to be invalid in the context of MSFragger narrow searches[3]. If true, this obviously cast doubts on the first open-versus-narrow comparison we reported above, but moreover it casts doubt on the validity of many publications that used TDC to analyze narrow search results. Thus, we were motivated to look more closely at this claim, as well as to examine the validity of TDC when applied in our second setup to PSMs rescored by post-processors.

The results of our analysis are somewhat surprising. First, we used our recently developed tool for Combining Open and Narrow searches with Group-wise Analysis (CONGA)[14] to re-analyze the data used in Kong *et al.*[3] while accounting for chimeric spectra, as well as for the inherent difference in the number of candidates in open and narrow searches. In doing, so we found that the data does not provide sufficient evidence that the FDR in the narrow search is inherently uncontrolled. Second, we uncovered a design flaw in Percolator where its FDR control is compromised in the case where multiple spectra are generated by the same peptide species — a problem that is exacerbated when the analysis is applied to multiple runs from the same experiment. We then demonstrated this problem empirically by applying Percolator to an entrapment setup using data

from a controlled experiment. Third, we found that the same entrapment experiment in which Percolator apparently failed to control the FDR proved even more challenging to PeptideProphet, with the latter substantially underestimating the FDR.

Although the apparent failures of Percolator and PeptideProphet are empirically demonstrated here in the context of a small experiment, we stress that a tool that properly controls the FDR should do that in all setups, and pointing out even one case where it apparently fails to do so should give its users pause. This is especially relevant to Percolator, where we were able to identify a relevant design flaw.

2 Methods

Searches

Entrapment run searches, PRIDE-20 searches, HEK293 searches

Our entrapment experiment searches, PRIDE-20 searches and HEK293 searches followed the same method described by *Freestone et al.*[14].

Target-decoy competition

Because PSM-level FDR control is inherently problematic[12,13], we only consider peptide-level analysis in this work. Specifically, our peptide-level TDC procedure is the "PSM-and-peptide" double competition procedure from Lin *et al.*[13] (summarized here as Supplementary Algorithm S1). This procedure requires that each target peptide is paired with a decoy peptide. Creating this pairing is trivial when no variable modifications are considered: each decoy is paired with the target peptide it was constructed from. When variable modifications are considered the pairing varies with the search engine, as follows:

- In Tide's case, an unmodified decoy is first generated and paired with the unmodified target peptide by random shuffling, e.g., the unmodified target `PEPTIDE` is shuffled to create, say, the unmodified decoy peptide `PTIDEPE`. Any position-specific modifications are jointly applied to the target and its paired decoy, e.g., the +30 N-terminal modified peptide `P[+30]EPTIDE` is paired with `P[+30]TIDEPE`. Any amino acid-specific modification is applied jointly to the target and its paired decoy by considering the relative position of the modified amino acid, e.g., if the first `P` residue in the target `PEPTIDE` is modified to `P[+15]EPTIDE` then the first `P` residue in the paired decoy will be modified the same way to yield `P[+15]TIDEPE`. Similarly, `PEP[+15]TIDE` will be paired with `PTIDEP[+15]E`, where in both cases the second `P` residue is modified, and `P[+15]EP[+15]TIDE` will be paired with `P[+15]TIDEP[+15]E`.

- Comet generates decoys by reversing each target peptide while leaving the C-terminal residue in place. Hence, when pairing a modified target peptide, we simply apply the same reversal process, which includes any amino acid-specific modifications. For example, the modified peptide `P[+15]EPTIDE` will be paired with `DITPEP[+15]E`, and `PEP[+15]TIDE` will be paired with `DITP[+15]EPE`. As in Tide's case, modifications that are position-specific remain attached to that position, e.g., the +2 N-terminal modification `P[+2]EPTIDE` is paired with the decoy `D[+2]ITPEPE`.

- In the case of MSFragger we were able to implement a pairing similar to the one we used in Comet's case by feeding MSFragger our own decoy proteins. These proteins are created by reversing every contiguous block of amino acids that does not contain the residues `K` and `R`.

Because MSFragger uses the strict-trypsin rule (no proline suppression), the resulting target peptides are related to the corresponding decoys by the same reversal as Comet implements (except when the C-terminal of the peptide does not end in `K` or `R`). Hence, the modified peptides are also paired in a similar manner as described above for Comet.

We used Tailor score for Tide, (the negative) E-value for Comet, and Hyperscore for MSFragger when applying TDC to the PSMs scored by the search engine. We used the learned SVM scores for Percolator and the learned linear discriminant analysis scores for PeptideProphet when applying TDC to the PSMs rescored by these post-processors.

Percolator settings

The settings used in Percolator are the same as those described by *Freestone et al.*[14], except we implemented the "PSM-and-peptide" double competition protocol on the learned Percolator scores when doing FDR analysis instead of the dynamic competition protocol that they describe. In Tide searches, Percolator used both XCorr and Tailor scores as features and in MSFragger searches, Percolator used both Hyperscore and E-values as features.

PeptideProphet settings

We used the Philosopher software to apply PeptideProphet to MSFragger searches of the entrapment runs. We followed Philosopher's workflow, using the following commands. To initialize the workspace, we used: `workspace --init`. To store the database in the workspace we used `database ---prefix decoy_ --nodecoys --custom fasta_file` where the `fasta_file` is the combined target-decoy peptide database containing a random subset of the in-sample peptides and the entrapment peptides. Then to run PeptideProphet we used `peptideprophet --database prepared_fasta_file --expectscore --decoyprobs --decoy decoy_ --nonparam --masswidth 1000 search_file` where `prepared_fasta_file` is the database generated in the previous step and `search_file` corresponds to the MSFragger search files used in the entrapment runs (narrow searches showed very little difference regardless of whether `--masswidth` was used). We used mostly the same steps above when applying PeptideProphet to the MSFragger searches of the PRIDE-20 datasets, with the exception that we did not use `--masswidth` for narrow searches and we did not use `--nonparam` for project PXD012611 (since otherwise PeptideProphet failed on all charge states). The FDR analysis was conducted using TDC with PeptideProphet's linear discriminant score.

3 Results

In the absence of a post-processor, open search is not more powerful than narrow search

Motivated by the increasing popularity of open modification searching, we set out to compare the statistical power of open and narrow window searches to detect peptides from MS/MS data. We reasoned that, although open search allows for the discovery of unexpected PTMs, it also necessarily considers many more candidate peptides for each observed spectrum. This large number of candidates increases the risk that a correct match—that is, a match to the peptide that truly generated the observed spectrum—may be out-scored by an incorrect match with a randomly high score. The empirical question boils down to whether the increased rate of discoveries of peptides

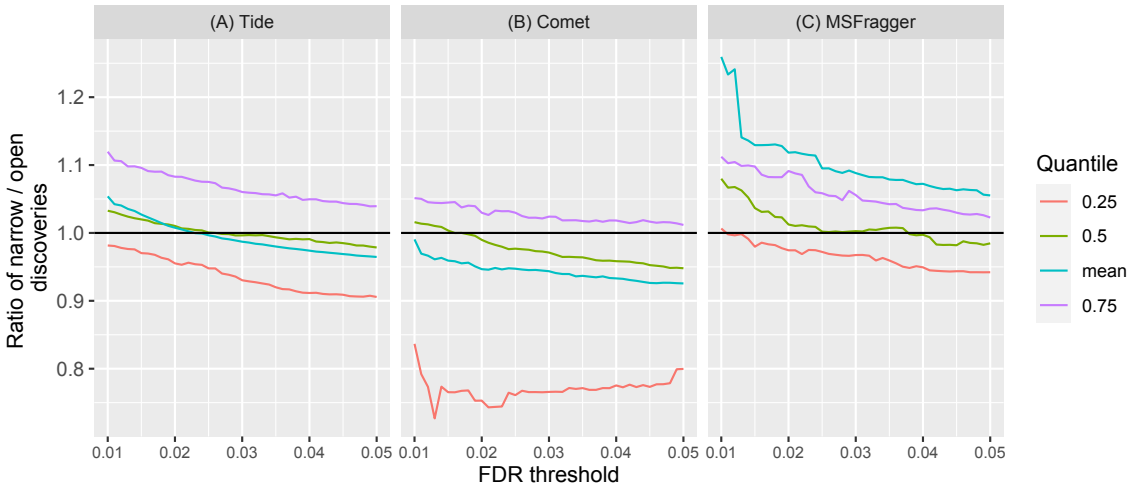


Figure 1: **Quartiles of narrow-to-open search discoveries in real data.** The mean and quartiles over the PRIDE-20 datasets of the ratio of narrow-to-open search discovered peptides using (A) Tide, (B) Comet, and (C) MSFragger, plotted as a function of FDR threshold. In (A), for each of the 20 datasets we average the discovery ratios over 20 randomized decoys generated specifically for that dataset, whereas in (B–C), we use a single reversed decoy database for each dataset. In panel (C) the mean is much higher than the median due to several outlier datasets.

with unexpected PTMs outweighs the loss of power due to the larger number of candidate peptides in the open search.

To investigate this question, we searched a collection of publicly available data (called PRIDE-20) using three different search engines. Specifically, we applied Tide[8], Comet[7], and MSFragger[3] to the PRIDE-20 spectrum files. For each search engine the reported list of discovered peptides was generated using the "PSM-and-peptide" TDC protocol (see Target-decoy competition section).
The results of this experiment show that we often obtain fewer discoveries using open search compared to narrow search (Figure 1). For all three search engines, at an FDR threshold of $\alpha = 1\%$ we overall obtain more discoveries with narrow search than with an open one (i.e., both the mean and the median ratios are > 1), while for higher FDR thresholds the picture is murkier. Notably, there are obvious outliers when using Comet and MSFragger. Table S1 in the supplement reports the ratio of narrow-to-open discoveries for each PRIDE-20 dataset and search engine, along with the size of the search space and the fragment tolerance associated with each dataset. We find that the outlier in the MSFragger analyses, project `PXD023571`, has a larger fragment tolerance. In combination with MSFragger's large open search window (-150 to 500 Da), this difference likely accounts for the greater number of discoveries using a narrow search over an open one. The outlier among Comet's analyses, project `PXD029319`, has the smallest search space and the smallest signal overall (between 130 and 310 discoveries across all search engines). Importantly, regardless of whether these outliers arise due to the nature of the dataset or the search engine's settings, it is still clear that an open search is not consistently more powerful than a narrow search, and moreover this is true across the range of datasets and search engines that we examined.

The results in Figure 1 are in apparent disagreement with results from Kong *et al.*[3], who reported greater power to detect peptides from open compared to narrow searches. One potential source of this discrepancy is that we carried out FDR control directly on the results of the search engine, whereas Kong *et al.* relied on rescoring the PSMs from a machine learning post-processor, namely PeptideProphet[10,15]. To investigate this potential discrepancy, we reanalyzed the PRIDE-20 data by applying TDC to the PSMs that were rescored using the post-processor, Percolator, on the Tide search results and PeptideProphet on the MSFragger search results. Figure 2 shows that,

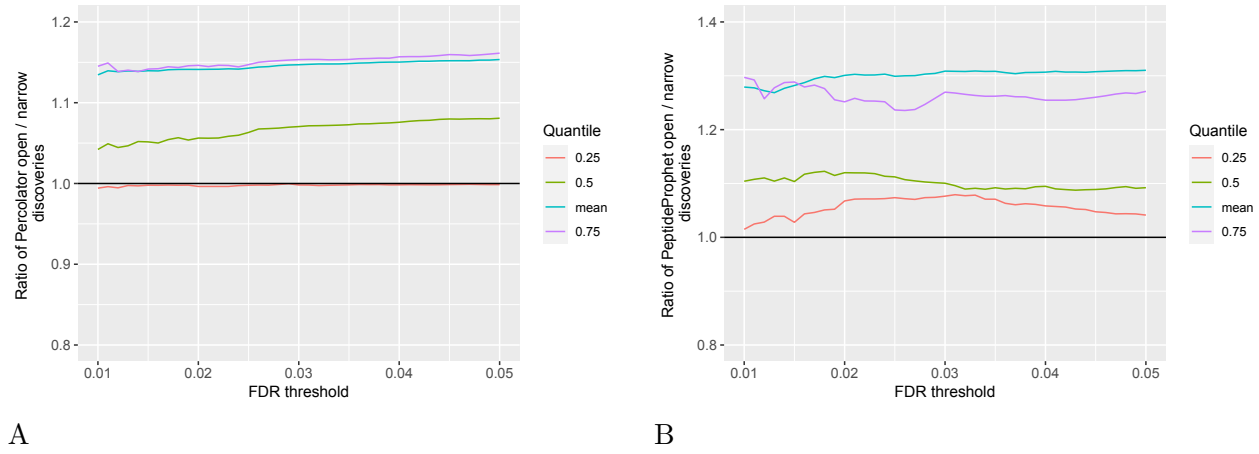


Figure 2: **Quartiles of open-to-narrow search discoveries in real data using Percolator and PeptideProphet.** The mean and quartiles over the PRIDE-20 datasets of the ratio of open-to-narrow search discoveries using (A) Percolator with Tide and (B) PeptideProphet with MSFragger. In (A), for each of the 20 datasets we average the discovery ratios over 20 randomized decoys generated for each dataset while for (B) we use a single reversed decoy database for each dataset.

similar to Kong *et al.*'s results, when using Percolator or PeptideProphet, the boosted open search delivered a greater number of detected peptides than the boosted narrow search.

Revisiting the problem of FDR control in a narrow search

When introducing their open-search software, MSFragger, Kong *et al.* claimed that it uncovered a problem with TDC-based FDR control in narrow search [3]. Specifically, analyzing the 24 spectrum files of the HEK293 dataset [2], they looked for peptides that were discovered in a narrow search with an FDR threshold of 1%, but for which the associated spectrum switched to a different peptide in the open search. In their narrow search results at a 1% FDR threshold, they found 1,139 such "switching" peptides among a total of 101,138 peptides. Assuming that those 1,139 peptides were falsely detected yields a false discovery proportion of 1.13%, which they concluded is too high given the FDR threshold of 1%.

To further support their claim, Kong *et al.* looked for peptides that were identified twice: once in a narrow search without any modifications and once in an open search with a modification mass corresponding to either carbamylation (43.00 Da) or oxidation (15.99 Da). They then collected the experimental spectra responsible for identifying the modified peptides in the open search, and they investigated what peptides were assigned to these spectra in the narrow search results. Because these peptides were identified twice (with and without modifications), we have high confidence that these detections are accurate. Therefore, we expect that if we search these spectra against a database in narrow search mode, while not allowing for the carbamylation or oxidation modifications, then the resulting PSMs should be incorrect. Accordingly, we expect the resulting false PSMs to be equally divided between target and decoy peptides. However, Kong *et al.* report that this set of spectra overwhelmingly matched with more target than decoy peptides: a ratio of 6:1 for carbamylation and of 9:1 for oxidation. Thus, they concluded again that something is wrong with applying TDC to the narrow search. In particular, the abundance of target peptides indicates that the FDR control is liberal.

With its simultaneous consideration of both narrow and open searches, as well as chimeric spectra, CONGA allows us to reevaluate those claims. Specifically, CONGA allows us to factor in peptides that co-generate chimeric spectra, for which both the narrow-search and open-search

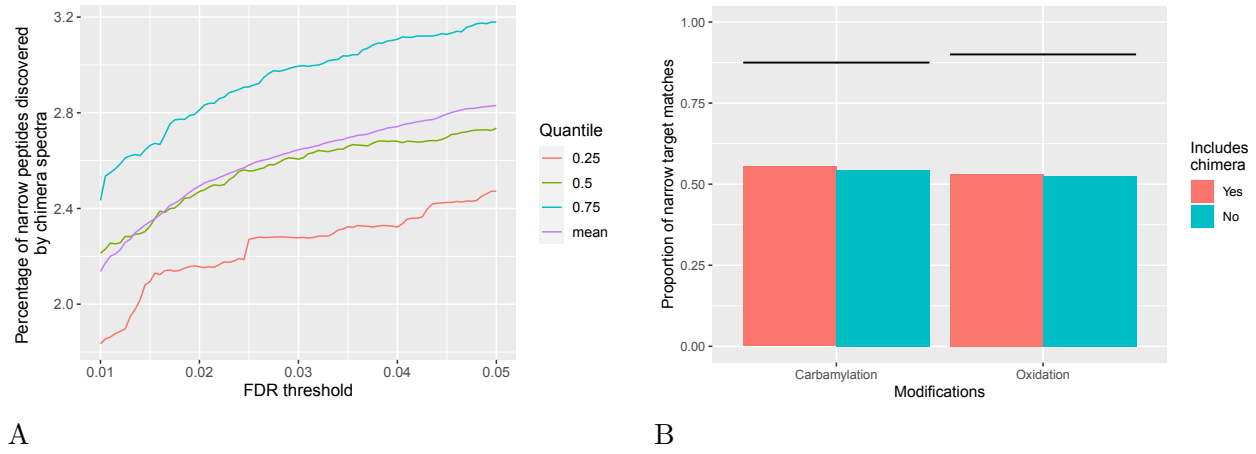


Figure 3: (A) The percentage of narrow-search peptides CONGA discovers through chimeric spectra. The quartiles and mean of the percentage of narrow-search peptides detected by CONGA for which the top-scoring PSM is from a presumed chimeric spectra, plotted as a function of the FDR level. The quartiles and mean are taken over the 24 HEK293 spectrum files, and the percentages are taken with respect to the number of peptides discovered at the same FDR level by CONGA. (B) The median proportion of target matches during a narrow search for spectra that were confidently assigned a match to a peptide with mass modifications coinciding with oxidation and carbamylation during an open search. The median is taken with respect to the 24 HEK293 spectrum files. Spectra that were confidently recognized as generated by co-elluted peptides were removed and the proportions are recalculated in blue. The approximate rate of target matches (6:1 for carbamylation and 9:1 for oxidation) reported by Kong *et al.* are also indicated by the line segments above the bars.

PSMs can be correct. Indeed, Figure 3A suggests that a considerable portion of narrow-discovered peptides that Kong *et al.* considered as suspect might be due to chimeras: at 1% FDR level, the median percentage of narrow-discovered peptides that CONGA attributes to chimeric spectra is 2.21%, and at 5% this median goes up to 2.83%. In addition, CONGA allows us to account for the fact that some peptides detected in the narrow-search are lost when their identifying spectrum switches to a higher scoring but incorrect match in the open search. Indeed, because the open search considers many more candidate peptides than the narrow search, to declare that a narrow match is incorrect we require that the open search peptide that replaced it is itself discovered.

Our analysis involves comparing the results of two procedures applied to the 24 spectrum files of the HEK293 dataset: MSFragger in narrow-search mode, followed by TDC, and CONGA utilizing MSFragger's narrow- and open-searches. We proceed in four steps, summarized in Supplementary Table S2. First, we count the number of narrow-search discovered peptides ("narrow peptides") for each of the 24 spectrum files of the HEK293 dataset at 1% FDR using TDC. The median number of discovered narrow peptides is 10,094. Second, we count the number of the narrow peptides that were lost among CONGA's discovery list using the same 1% FDR threshold. Such a loss could happen because the identifying narrow PSM was lost due to a higher scoring open PSM—the main effect that Kong *et al.* were concerned with—or because the decoy of the narrow peptide outscored it in the open search, or because the narrow PSM did not score high enough to register as a discovery when CONGA determined its joint list of open and narrow discoveries. Regardless of the reason, the median number of narrow peptides that CONGA fails to detect at 1% FDR is 239, or about 2%. Third, we count the number of lost narrow peptides for which the identifying narrow PSM was swapped for at least one higher-scoring open PSM whose peptide was discovered by CONGA at the same 1% FDR threshold. These peptides correspond to what we consider as dubious narrow discoveries: the identified spectrum from the narrow search was reassigned to another peptide in the open search, and moreover this open peptide is very likely to be present in the sample. The median of these dubious narrow discoveries number is 49. Finally, in step four we compute the percentage

of the dubious discoveries among all narrow discoveries. These percentages vary between 0.261% to 0.66% with a median of 0.46%. This is well below the 1% FDR threshold, so we conclude that this analysis does not reveal a problem with the TDC-implemented FDR control of the narrow search of this data.

We also used CONGA to reexamine the second setup that Kong *et al.* analyzed. To do this, we first compiled a list of the spectra that were confidently matched to peptides with mass-modifications coinciding with oxidation and carbamylation (up to ±0.02 Da). Specifically, the relevant PSM had to be the top match for the spectrum, and the peptide had to be discovered at 1% FDR where, to be more inclusive here, we allowed for any peptide in CONGA's augmented list of discoveries (described in Supplementary Algorithm 6 of the CONGA manuscript[14]). Following Kong *et al.*, we then selected the subset of those spectra for which their matched peptides were also discovered, without any modification, at 1% FDR in the MSFragger analysis. Finally, we found the proportion of these selected spectra whose narrow-search optimal matching peptide is a target one. Assuming that the open search match is the correct one, the narrow search should equally likely match a target or a decoy peptide.

Our results, given in Figure 3B, disagree with those of Kong *et al.* They reported that 85.7% of the carbamylated spectra and 90% of the oxidized spectra matched to targets, whereas we observe corresponding values of 55.5% and 53.2%. Some of the observed bias might arise due to chimeric spectra, if some of the open modified peptides share their identifying chimeric spectrum with an unmodified peptide. To account for this possibility, we redid the above analysis after removing those experimental spectra that CONGA identified as chimeric with at least one of the co-generating peptides detected in the narrow search. This brought the target percentages down slightly to 54.3% and 52.3%, respectively. Kong *et al.* mention that in order to annotate the reported mass shifts, they searched for an annotated entry with a mass tolerance of 0.002 Da. Therefore, we tested whether using 0.002 Da instead of 0.02 Da makes a difference, but we found it was only a negligible one. Considering that our chimera analysis is fairly conservative we do not find these biases sufficient to determine that narrow-search based TDC fails to control the FDR.

Post-processors such as Percolator and PeptideProphet can struggle with FDR control

Because Percolator and PeptideProphet are widely used to re-rank PSMs produced by a database search, we sought to empirically assess whether their effectiveness might be associated with compromised FDR control. To do so, we ran a set of entrapment experiments, in which spectra that were generated in a controlled experiment from a known set of peptides are searched against a database containing those peptides plus a set of "entrapment sequences"[16]. We used both Tide and MSFragger to search 20 concatenated databases in both open and narrow modes. Percolator (both search engines) and PeptideProphet (MSFragger only) were then applied to rescore the PSMs to which we applied peptide-level TDC and estimated the FDR.

Both post-processors yield estimated FDRs that are consistently greater than the specified FDR threshold, across a range of thresholds from 0.01 to 0.1, and in both open and narrow searches (Figure 4A–C). Furthermore, as the number of in-sample peptides decreases, the estimated FDR is driven further up for most FDR threshold values. Note that because the ISB18 experimental spectra originated from very few proteins, for lower FDR thresholds and increasingly smaller in-sample portion we often obtain no discoveries. Thus, the larger FDR violations are observed at higher FDR thresholds.

In addition, we also examined the default FDR analysis of the Philosopher workflow[17]. The latter relies on PeptideProphet's posterior error probabilities to estimate and control the FDR

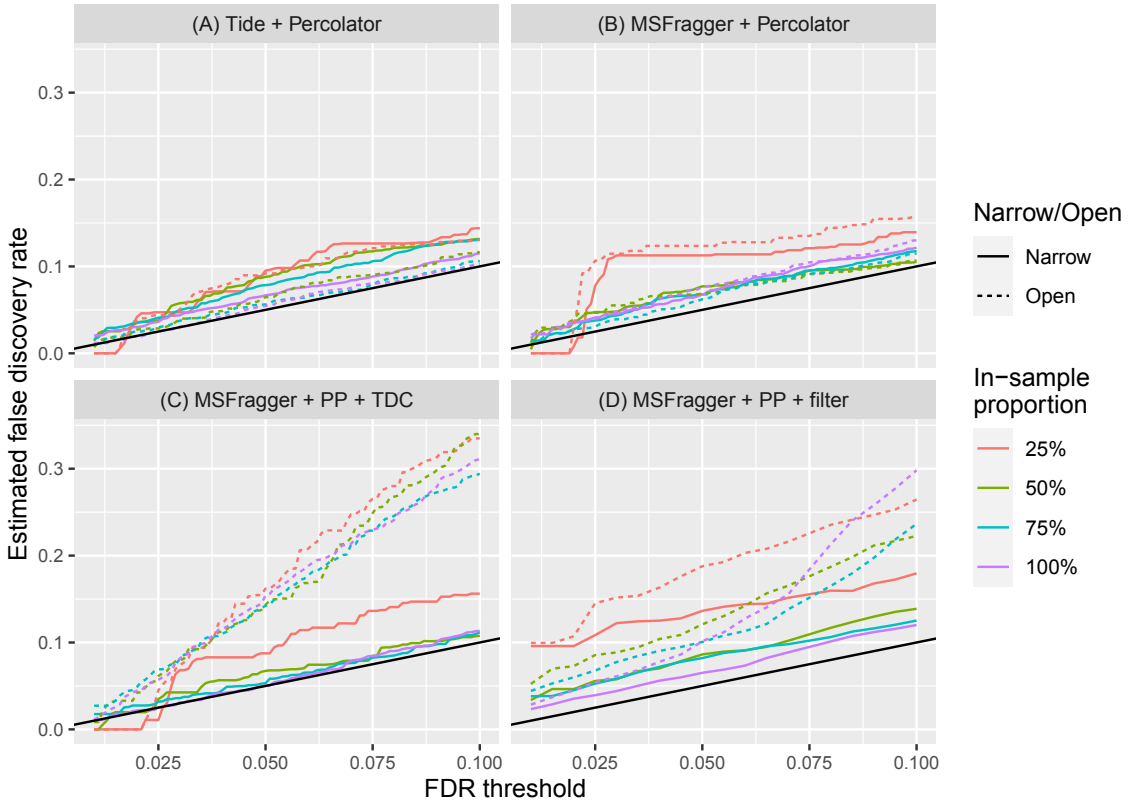


Figure 4: **Percolator and PeptideProphet empirical FDR.** The empirical FDR as estimated from our entrapment experiments on the ISB18 dataset using (A) Percolator with Tide, (B) Percolator with MSFragger, (C) PeptideProphet's discriminant score with MSFragger, and (D) PeptideProphet with Philosopher's default filter function with MSFragger. The FDP is estimated at a range of FDR thresholds ($[0.01, 0.1]$) and its average over 20 randomly generated decoys is the empirical FDR. Each curve represents the empirical FDR using a target database constructed with the specified proportion of the in-sample database, in either narrow- or open-search mode. The estimated FDRs mostly exceed the corresponding threshold (solid black line), indicating that the FDR is not properly controlled.

(see section on PeptideProphet settings). We found that, particularly for lower FDR thresholds, Philosopher makes things even worse (Figure 4D); for example, the (worst-case) FDR at 1% changes from 2.72% for PeptideProphet's discriminant score to 9.95% for the Philosopher approach.

We hypothesized that this failure to control the FDR may arise in part due to multiplicity in the spectra, that is, due to the presence in the dataset of multiple spectra generated from the same peptide species. For example, Percolator trains its machine learning model using a threefold cross-validation scheme, learning the model parameters on 2/3 of the data at a time and applying the trained model to the remaining 1/3. However, the validity of this cross-validation procedure may be compromised if one peptide species generates multiple spectra, with some of those spectra going into the training set and some into the test set. In such a situation, the model may overfit. This spectrum multiplicity problem is exacerbated whenever replicate datasets are analyzed jointly.

To test this hypothesis, we repeated the above entrapment analysis, but instead of analyzing all nine ISB18 runs jointly, we analyzed each run separately (Supplementary Figure S1). The resulting FDR violations greatly subside for the case of Percolator, while for PeptideProphet they become worse. Unlike Percolator, PeptideProphet does not use a cross-validation scheme, so a proper accounting for the apparent FDR violation may be due to factors beyond the multiplicity issue. For example, PeptideProphet estimates model parameters from the data and therefore might struggle with a smaller number of discoveries.

4 Discussion

We set out to examine the tradeoff between open modification searches, which make it possible to detect peptides with unexpected PTMs, and narrow searches, which can do a better job at detecting unmodified peptides. However, in trying to validate our comparisons we found ourselves facing more fundamental questions about the validity of TDC in the setups we considered.

First, we had to revisit Kong *et al.*'s recent claim that they uncovered a problem with TDC-based control of the FDR in narrow searching[3], because this claim obviously had the potential to undermine the validity of our comparisons. Moreover, because of the widespread use of TDC in the analysis of narrow search results, this claim could undermine the validity of a significant number of existing research works. Fortunately, our re-analysis of the data using CONGA did not detect an inherent problem with TDC in the narrow search context. As for the discrepancy between our findings and those of Kong *et al.*, it could be due to a number of reasons: (i) we took into account chimeric spectra; (ii) we required that a higher-scoring match in the open search would also be a reported discovery; and (iii) the original analysis relied on processing the data with PeptideProphet, which as we show here, might struggle with FDR control.

Second, we uncovered a flaw in Percolator that could lead to compromised FDR control when multiple spectra are generated by the same peptide species. Moreover, we empirically showed through entrapment experiments that both Percolator and PeptideProphet can apparently fail to control the FDR in some settings. We therefore urge the community to cautiously evaluate their results when using these tools. In particular, we already prepared a fix to Percolator's FDR control that is implemented from the command line using Python and that will be integrated into Percolator's next release[18].

Returning to our initial goal, we noted that applying peptide-level TDC analysis to the PSMs scored without any post-processing, the narrow search analysis often reports more peptides than the one based on open search PSMs generated from the same data. This result is not surprising, given that (a) when switching from a narrow to an open search a correct narrow PSM now competes with a greater number of incorrect candidate peptides, and (b) for the same reason of a large number of candidates per spectrum, the scores of incorrect peptides are generally higher, regardless of whether or not they out-competed a correct discovery. Because we found no reason to doubt the validity of TDC applied to PSMs scored by the search engine in narrow search mode, this comparison stands.

In a complementary comparison, consistent with Kong *et al.*[3], we found that using Percolator or PeptideProphet to rescore the PSMs prior to applying TDC typically results in more peptides discovered using an open search than when using a narrow search. However, given that we show that both of these post-processors can fail to control the FDR, these comparisons should be taken with a grain of salt.

As mentioned, Percolator's FDR control should be fixed by the time of its next release; however, another solution which is already available is to use CONGA, an alternative post-processor that combines narrow and open searches while rigorously controlling the FDR, and which typically delivers more discovered peptides than either search does on its own[14].

Acknowledgments We are grateful to the anonymous referees for their comments and suggestions which helped to substantially improve this manuscript.

References

[1] A. I. Nesvizhskii. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics*, 73(11): 2092 – 2123, 2010.

[2] J. M. Chick, D. Kolippakkam, D. P. Nusinow, B. Zhai, R. Rad, E. L. Huttlin, and S. P. Gygi. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature Biotechnology*, 33(7):743–749, 2015.

[3] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, and A. I. Nesvizhskii. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods*, 14(5):513–520, 2017.

[4] Hao Chi, Chao Liu, Hao Yang, Wen-Feng Zeng, Long Wu, Wen-Jing Zhou, Rui-Min Wang, Xiu-Nan Niu, Yue-He Ding, Yao Zhang, et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nature biotechnology*, 36(11): 1059–1061, 2018.

[5] S. K. Solntsev, M. R. Shortreed, B. L. Frey, and L. M. Smith. Enhanced global post-translational modification discovery with metamorpheus. *Journal of Proteome Research*, 17 (5):1844–1851, 2018.

[6] M. R. Lazear. Sage: An open-source tool for fast proteomics searching and quantification at scale. *Journal of Proteome Research*, 22(11):3652–3659, 2023.

[7] J. K. Eng, T. A. Jahan, and M. R. Hoopmann. Comet: an open source tandem mass spectrometry sequence database search tool. *Proteomics*, 13(1):22–24, 2012.

[8] B. Diament and W. S. Noble. Faster SEQUEST searching for peptide identification from tandem mass spectra. *Journal of Proteome Research*, 10(9):3871–3879, 2011.

[9] L. Käll, J. Canterbury, J. Weston, W. S. Noble, and M. J. MacCoss. A semi-supervised machine learning technique for peptide identification from shotgun proteomics datasets. *Nature Methods*, 4:923–25, 2007.

[10] A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold. Empirical statistical model to estimate the accuracy of peptide identification made by MS/MS and database search. *Analytical Chemistry*, 74:5383–5392, 2002.

[11] J. E. Elias and S. P. Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, 4(3):207–214, 2007.

[12] K. He, Y. Fu, W.-F. Zeng, L. Luo, H. Chi, C. Liu, L.-Y. Qing, R.-X. Sun, and S.-M. He. A theoretical foundation of the target-decoy search strategy for false discovery rate control in proteomics. *arXiv*, 2015. https://arxiv.org/abs/1501.00537.

[13] A. Lin, T. Short, W. S. Noble, and U. Keich. Improving peptide-level mass spectrometry analysis via double competition. *Journal of Proteome Research*, 21(10):2412–2420, 2022.

[14] J. Freestone, W. S. Noble, and U. Keich. Analysis of tandem mass spectrometry data with CONGA: Combining open and narrow searches with group-wise analysis. *bioRxiv*, 2023. doi: 10.1101/2023.05.02.539167. https://www.biorxiv.org/content/10.1101/2023.05.02.539167v1.

[15] K. Ma, O. Vitek, and A. I. Nesvizhskii. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics*, 13(16):1–17, 2012.

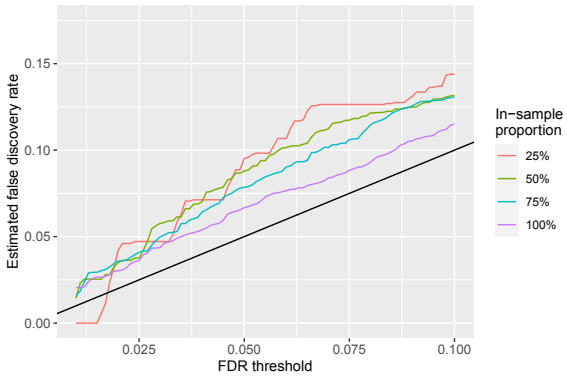
[16] V. Granholm, J. F. Navarro, W. S. Noble, and L. Käll. Determining the calibration of confidence estimation procedures for unique peptides in shotgun proteomics. *Journal of Proteomics*, 80(27):123–131, 2013.

[17] F. da Veiga Leprevost, S. E. Haynes, D. M. Avtonomov, H. Y. Chang, A. K. Shanmugam, D. Mellacheruvu, A. T. Kong, and A. I. Nesvizhskii. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nature Methods*, 17(9):869–870, Sep 2020.

[18] J. Freestone, L. Käll, W. S. Noble, and U. Keich. Semi-supervised learning while controlling the fdr with an application to tandem mass spectrometry analysis. *bioRxiv*, 2023. doi: 10.1101/2023.10.26.564068.

SUPPORTING INFORMATION

The following supporting information is available free of charge at ACS website `https://pubs.acs.org/`.


Estimated false discovery rate
0.15
0.10
0.05
0.00
0.025
0.050
0.075
0.100
FDR threshold
In−sample proportion
25%
50%
75%
100%

For TOC Only