Multiple competition-based FDR control and its application to peptide detection*

Kristen Emery[1], Syamand Hasam[1], William Stafford Noble[2 [0000−−0001−7283−4715]], and Uri Keich[1 [0000−−0002−3209−5011]]

[1] School of Mathematics and Statistics F07
University of Sydney
{keme6477@,shas8442@uni.,uri.keich@}sydney.edu.au
[2] Department of Genome Sciences and of Computer Science and Engineering
University of Washington
william-noble@uw.edu

Abstract. Competition-based FDR control has been commonly used for over a decade in the computational mass spectrometry community [5]. Recently, the approach has gained significant popularity in other fields after Barber and Candés laid its theoretical foundation in a more general setting that included the feature selection problem [1]. In both cases, the competition is based on a head-to-head comparison between an observed score and a corresponding decoy / knockoff. We recently demonstrated some advantages of using multiple rather than a single decoy when addressing the problem of assigning peptide sequences to observed mass spectra [17]. In this work, we consider a related problem — detecting peptides based on a collection of mass spectra — and we develop a new framework for competition-based FDR control using multiple null scores. Within this framework, we offer several methods, all of which are based on a novel procedure that rigorously controls the FDR in the finite sample setting. Using real data to study the peptide detection problem we show that, relative to existing single-decoy methods, our approach can increase the number of discovered peptides by up to 50% at small FDR thresholds.

Keywords: multiple hypothesis testing · peptide detection · tandem mass spectrometry · false discovery rate.

1 Introduction

Proteins are the primary functional molecules in living cells, and tandem mass spectrometry (MS/MS) currently provides the most efficient means of studying proteins in a high-throughput fashion. Knowledge of the protein complement in a cellular population provides insight into the functional state of the cells. Thus, MS/MS can be used to functionally characterize cell types, differentiation stages, disease states, or species-specific differences. For this reason, MS/MS is the driving technology for much of the rapidly growing field of proteomics.

* Supplementary Material: https://noble.gs.washington.edu/proj/multidecoy

Paradoxically, MS/MS does not measure proteins directly. Because proteins themselves are difficult to separate and manipulate biochemically, an MS/MS experiment involves first digesting proteins into smaller pieces, called "peptides." The peptides are then measured directly. A typical MS/MS experiment generates $\sim$10 observations ("spectra") per second, so a single 30-minute MS/MS experiment will generate approximately 18,000 spectra. Canonically, each observed spectrum is generated by a single peptide. Thus, the first goal of the downstream analysis is to infer which peptide was responsible for generating each observed spectrum. The resulting set of detected peptides can then be used, in a second analysis stage, to infer what proteins are present in the sample.

In this work, we are interested in the first problem — peptide detection. This problem is important not only as a stepping stone toward the downstream goal of detecting proteins; in many proteomics studies, the peptides themselves are of primary interest. For example, MS/MS is being increasingly applied to complex samples, ranging from the microbiome in the human gut [21] to microbial communities in environmental samples such as soil or ocean water [30]. In these settings, the genome sequences of the species in the community are only partially characterized, so protein inference is problematic. Nonetheless, observation of a particular peptide can often be used to infer the presence of a group of closely related species (a taxonomic clade) or closely related proteins (a homology group). Peptide detection is also of primary interest in studies that aim to detect so-called "proteoforms" — variants of the same protein that arise due to differential splicing of the mature RNA or due to post-translational modifications of the translated protein. Identifying proteoforms can be critically important, for example, in the study of diseases like Alzheimer's or Parkinson's disease, in which the disease is hypothesized to arise in part due to the presence of deviant proteoforms [23, 28, 37].

Specifically, we focus on the task of assigning confidence estimates to peptides that have been identified by MS/MS. As is common in many molecular biology contexts, these confidence estimates are typically reported in terms of the false discovery rate (FDR), i.e., the expected value of the proportion of false discoveries among a set of detected peptides. For reasons that will be explained below, rather than relying on standard methods for control of the FDR such as the Benjamini-Hochberg (BH) procedure [2] the proteomics field employs a strategy known as "target-decoy competition" (TDC) to control the FDR in the reported list of detected peptides [5]. TDC works by comparing the list of peptides detected with a list of artificial peptides, called "decoys," detected using the same spectra set. The decoys are created by reversing or randomly shuffling the letters of the real ("target") peptides. The TDC protocol, which is described in detail in Section 2.2, estimates the FDR by counting the number of detected decoy peptides and using this count as an estimate for the number of incorrectly detected target peptides.

One clear deficiency of TDC is its reliance on a single set of decoy peptides to estimate the FDR. Thus, with ever increasing computational resources one can ask whether we can gain something by exploiting multiple randomly drawn decoys for each target peptide. We recently described such a procedure, called "average target-decoy competition" (aTDC), that, in the context of the related spectrum identification problem (described in Section 2.1), reduces the variability associated with TDC and can provide a modest boost in power [17, 18].

In this paper we propose a new approach to using multiple decoy scores. The proposed procedure relies on a direct competition between the target and its corresponding decoy scores, rather than on averaging single competitions. We formulate our approach in the following more general setting. Suppose that we can compute a test statistic Z_i for each null hypothesis H_i, so that the larger Z_i is, the less likely is the null. However, departing from the standard multiple hypotheses setup, we further assume that we cannot compute p-values for the observed scores. Instead, we can only generate a *small* sample of independent decoys or competing null scores for each hypothesis H_i: $\tilde{Z}_i^j \; j = 1, \ldots, d$ (Definition 2). Note that the case $d = 1$ corresponds to the TDC setup described above. We will show using both simulated and real data that the novel method we propose yields more power (more discoveries) than our aforementioned averaging procedure.

In addition to the peptide detection problem, our proposed procedure is applicable in several other bioinformatics applications. For example, the procedure could be used when analyzing a large number of motifs reported by a motif finder, e.g., [12], where creating competing null scores can require the time consuming task of running the finder on randomized versions of the input sets, e.g., [25]. In addition, our procedure is applicable to controlling the FDR in selecting differentially expressed genes in microarray experiments where a small number of permutations is used to generate competing null scores [36].

Our proposed method can also be viewed as a generalization of Barber and Candés' "knockoff" procedure [1], which is a competition-based FDR control method that was initially developed for feature selection in a classical linear regression model. The procedure has gained a lot of interest in the statistical and machine learning communities, where it has has been applied to various applications in biomedical research [38, 10, 29] and has been extended to work in conjunction with deep neural networks [22] and with time series data [9]. Despite the different terminology, both knockoffs and decoys serve the same purpose in competition-based FDR control; hence, for the ideas presented in this paper the two are interchangeable. A significant part of Barber and Candés' work is the sophisticated construction of their knockoff scores; controlling the FDR then follows exactly the same competition that TDC uses. Indeed, their Selective SeqStep+ (SSS+) procedure rigorously formalizes in a much more general setting the same procedure described above in the context of TDC. Note that Barber and Candés suggested that using multiple knockoffs could improve the power of their procedure so the methods we propose here could provide a stepping stone toward that. However, we would still need to figure out how to generalize their construction from one to multiple knockoffs.

2 Background

2.1 Shotgun proteomics and spectrum identification

In a "shotgun proteomics" MS/MS experiment, proteins in a complex biological sample are extracted and digested into peptides, each with an associated charge. These charged peptides, called "precursors," are measured by the mass spectrometer, and a subset is then selected for further fragmentation into charged ions, which are detected

and recorded by a second round of mass spectrometry [27, 26]. The recorded tandem fragmentation spectra, or spectra for short, are then subjected to computational analysis.

This analysis typically begins with the spectrum identification problem, which involves inferring which peptide was responsible for generating each observed fragmentation spectrum. The most common solution to this problem is peptide database search. Pioneered by SEQUEST [8], the search engine extracts from the peptide database all "candidate peptides," defined by having their mass lie within a pre-specified tolerance of the measured precursor mass. The quality of the match between each one of these candidate peptides and the observed fragmentation spectrum is then evaluated using a score function. Finally, the optimal peptide-spectrum match (PSM) for the given spectrum is reported, along with its score [24].

In practice, many expected fragment ions will fail to be observed for any given spectrum, and the spectrum is also likely to contain a variety of additional, unexplained peaks [26]. Hence, sometimes the reported PSM is correct — the peptide assigned to the spectrum was present in the mass spectrometer when the spectrum was generated — and sometimes the PSM is incorrect. Therefore, we report a thresholded list of top-scoring PSMs, together with the critical estimate of the fraction of incorrect PSMs in our reported list.

2.2 False discovery rate control in spectrum identification

The general problem of controlling the proportion of false discoveries has been studied extensively in the context of multiple hypotheses testing (MHT). We briefly review this setup in Supplementary Section 6.1; however, these techniques cannot be applied directly to the spectrum identification problem. A major reason for that is the presence in any shotgun proteomics dataset of both "native spectra" (those for which their generating peptide is in the target database) and "foreign spectra" (those for which it is not). These create different types of false positives, implying that we typically cannot apply FDR controlling procedures that were designed for the general MHT context to the spectrum identification problem [16].

Instead, the mass spectrometry community uses TDC to control the FDR in the reported list of PSMs [5, 3, 15, 6]. TDC works by comparing searches against a target peptide database with searches against a decoy database of peptides. More precisely, let Z_i be the score of the optimal match (PSM) to the ith spectrum in the target database, and let $\tilde{Z}_i$ be the corresponding optimal match in the decoy database. Each decoy score $\tilde{Z}_i$ directly competes with its corresponding target score Z_i for determining the reported list of discoveries. Specifically, for each score threshold T we only report target PSMs that won their competition: $Z_i > \max\{T, \tilde{Z}_i\}$. Subsequently, the number of decoy wins ($\tilde{Z}_i > \max\{T, Z_i\}$) is used to estimate the number of false discoveries in the list of target wins. Thus, the ratio between that estimate and the number of target wins yields an estimate of the FDR among the target wins. To control the FDR at level α we choose the smallest threshold $T = T(\alpha)$ for which the estimated FDR is still $\leq \alpha$. It was recently shown that, assuming that incorrect PSMs are independently equally likely to come from a target or a decoy match and provided we add 1 to the number of decoy wins before dividing by the number of target wins, this procedure rigorously controls the FDR [13, 20].

2.3 The peptide detection problem

The spectrum identification is largely used as the first step in addressing the peptide identification problem that motivates the research presented here. Indeed, to identify the peptides we begin, just like we do in spectrum identification, by assigning each spectrum to the unique target/decoy peptide which offers the best match to this spectrum in the corresponding database. We then assign to each target peptide a score Z_j which is the maximum of all PSM scores of spectra that were assigned to this peptide in the first phase. Similarly, we assign to the corresponding decoy peptide a score $\tilde{Z}_j$, which again is the maximum of all PSM scores involving spectra that were assigned to that decoy peptide. The rest continues using the same TDC protocol we outlined above for the spectrum identification problem [11, 31].

3 Controlling the FDR using multiple decoys

3.1 Why do we need a new approach?

A key feature of our problem is that due to computational costs the number of decoys, d, is small. Indeed, if we are able to generate a large number of independent decoys for each hypothesis, then we can simply apply the standard FDR controlling procedures (Supplementary Section 6.1) to the empirical p-values. These p-values are estimated from the empirical null distributions, which are constructed for each hypothesis H_i using its corresponding decoys. Specifically, these empirical p-values take values of the form $(d_1 - r_i + 1)/d_1$, where $d_1 = d + 1$, and $r_i \in \{1, \ldots, d_1\}$ is the rank of the originally observed score ("original score" for short) Z_i in the combined list of d_1 scores: $\left(\tilde{Z}_i^0 = Z_i, \tilde{Z}_i^1, \ldots, \tilde{Z}_i^d\right)$ ($r_i = 1$ is the lowest rank). Using these p-values the BH procedure [2] rigorously controls the FDR, and Storey's method [32] will asymptotically control the FDR as the number of hypotheses $m \to \infty$.

Unfortunately, because d is small, applying those standard FDR control procedures to the rather coarse empirical p-values may yield very low power. For example, if $d = 1$, each empirical p-value is either 1/2 or 1, and therefore for many practical examples both methods will not be able to make any discoveries at usable FDR thresholds.

Alternatively, one might consider pooling all the decoys regardless of which hypothesis generated them. The pooled empirical p-values attain values of the form $i/\left(m \cdot d + 1\right)$ for $i = 1, \ldots, md + 1$; hence, particularly when m is large, the p-values generally no longer suffer from being too coarse. However, other significant problems arise when pooling the decoys. These issues — discussed in Supplementary Section 6.2 — imply that in general, applying BH or Storey's procedure to p-values that are estimated by pooling the competing null scores can be problematic both in terms of power and control of the FDR.

3.2 A novel meta-procedure for FDR control using multiple decoys

The main technical contribution of this paper is the introduction of several procedures that effectively control the FDR in our multiple competition-based setup and that rely on the following meta-procedure.

Input: an original/target score Z_i and d competing null scores $\tilde{Z}_i^j$ for each null hypothesis H_i.
Parameters: an FDR threshold $\alpha \in (0,1)$, two tuning parameters $c = i_c/d_1$ ($d_1 = d+1$), the "original/target win" threshold, and $\lambda = i_\lambda/d_1$, the "decoy win" threshold where $i_\lambda, i_c \in \{1,\dots,d\}$ and $c \le \lambda$, as well as a (possibly randomized) mapping function $\varphi : \{1,\dots,d_1 - i_\lambda\} \mapsto \{d_1 - i_c + 1,\dots,d_1\}$.
Procedure:

1. Each hypothesis H_i is assigned an original/decoy win label:

$$L_i = \begin{cases} 1 & r_i \ge d_1 - i_c + 1 \qquad \text{(original win)} \\ 0 & r_i \in (d_1 - i_\lambda, d_1 - i_c + 1) \qquad \text{(ignored hypothesis)} \\ -1 & r_i \le d_1 - i_\lambda \qquad \text{(decoy win)} \end{cases}, \tag{1}$$

   where $r_i \in \{1,\dots,d_1\}$ is the rank of the original score when added to the list of its d decoy scores.
2. Each hypothesis H_i is assigned a score $W_i = \tilde{Z}_i^{(s_i)}$, where $\tilde{Z}_i^{(j)}$ is the jth order statistic or the jth largest score among $\left(\tilde{Z}_i^0 = Z_i, \tilde{Z}_i^1, \dots, \tilde{Z}_i^d\right)$, and the "selected rank", s_i, is defined as

$$s_i = \begin{cases} r_i & L_i = 1 \text{ (so } W_i = Z_i \text{ in an original win)} \\ u_i & L_i = 0 \text{ (where } u_i \text{ is randomly chosen uniformly in } \{d_1 - i_c + 1,\dots,d_1\}) \\ \varphi(r_i) & L_i = -1 \text{ (so } W_i \text{ coincides with a decoy score in a decoy win)} \end{cases} \tag{2}$$

3. The hypotheses are reordered so that W_i are decreasing, and the list of discoveries is defined as the subset of original wins $D(\alpha, c, \lambda) \coloneqq \{i : i \le i_{\alpha c \lambda}, L_i = 1\}$, where

$$i_{\alpha c \lambda} \coloneqq \max\left\{ i : \frac{1 + \#\{j \le i : L_j = -1\}}{\#\{j \le i : L_j = 1\} \vee 1} \cdot \frac{c}{1-\lambda} \le \alpha \right\}. \tag{3}$$

We assume above that all ties in determining the ranks r_i, as well as the order of W_i, are broken randomly, although other ways to handle ties are possible (e.g., Section 8.3 in our technical report [7]).

Note that the hypotheses for which $L_i = 0$ can effectively be ignored as they cannot be considered discoveries nor do they factor in the numerator of (3).

Our procedures vary in how they define the (generally randomized) mapping function φ (and hence s_i in (2)), as well as in how they set the tuning parameters c, λ. For example, in the case $d = 1$ setting $c = \lambda = 1/2$ and $\varphi(1) \coloneqq 2$ our meta-procedure coincides with TDC. For $d > 1$ we have increasing flexibility with d, but one obvious generalization of TDC is to set $c = \lambda = 1/d_1$. In this case $L_i = 1$ if the original score is larger than all its competing decoys and otherwise $L_i = -1$. Thus, by definition, φ is constrained to the constant value d_1 so $s_i \equiv d_1$ and W_i is always set to $Z_i^{(d_1)} = \max\left\{\tilde{Z}_i^0, \dots, \tilde{Z}_i^d\right\}$. Hence we refer to this as the "max method." As we will see, the max method controls the FDR, but this does not hold for any choice of c, λ and φ. The following section specifies a sufficient condition on c, λ and φ that guarantees FDR control.

3.3 Null labels conditional probabilities property

Definition 1. *Let N be the indices of all true null hypotheses. We say the null labels conditional probabilities property (NLCP) is satisfied if conditional on all the scores $\mathcal{W} = (W_1, \dots W_m)$ the random labels $\{L_i : i \in N\}$ are (i) independent and identically distributed (iid) with $P(L_i = 1 \mid \mathcal{W}) = c$ and $P(L_i = -1 \mid \mathcal{W}) = 1 - \lambda$, and (ii) independent of the false null labels $\{L_i : i \notin N\}$.*

Note that in claiming that TDC controls the FDR we implicitly assume that a false match is equally likely to arise from a target win as it is from a decoy win independently of all other scores [13]. This property coincides with the NLCP with $d = 1$ and $c = \lambda = 1/2$. Our next theorem shows that the NLCP generally guarantees the FDR control of our meta-procedure. Specifically, we argue that with NLCP established step 3 of our meta-procedure can be viewed as a special case of Barber and Candés' SSS+ procedure [1] and its extension by Lei and Fithian's Adaptive SeqStep (AS) [19]. Both procedures are designed for sequential hypothesis testing where the order of the hypotheses is predetermined – by the scores W_i in our case.

Theorem 1. *If the NLCP holds then our meta-procedure controls the FDR in a finite-sample setting, that is, $E\left(|D(\alpha, c, \lambda) \cap N| / |D(\alpha, c, \lambda)|\right) \leq \alpha$, where the expectation is taken with respect to* all *the decoy draws.*

Why does Theorem 1 make sense? If the NLCP holds then a true null H_i is an original win ($L_i = 1$) with probability c and is a decoy win with probability $1 - \lambda$. Hence, the factor $\frac{c}{1-\lambda}$ that appears in (3) adjusts the observed number of decoy wins, $\#\{j \leq i : L_j = -1\}$, to estimate the number of (unobserved) false original wins (those for which the corresponding H_i is a true null). Ignoring the +1 correction, the adjusted ratio of (3) therefore estimates the FDR in the list of the first i original wins. The procedure simply takes the largest such list for which the estimated FDR is $\leq \alpha$.

Proof. To see the connection with SSS+ and AS we assign each hypothesis H_i a p-value $p_i \coloneqq P\left(L_i \geq l\right)$. Clearly, if the NLCP holds then

$$p_i = \begin{cases} c & l = 1 \\ \lambda & l = 0 \\ 1 & l = -1 \end{cases}. \tag{4}$$

Moreover, the NLCP further implies that for any $u \in (0, 1)$ and $i \in N$, $P\left(p_i \leq u \mid \mathcal{W}\right) \leq u$, and that the true null labels L_i, and hence the true null p-values, p_i, are independent conditionally on $\mathcal{W}$. It follows that, even after sorting the hypotheses by the decreasing order of the scores W_i, the p-values of the true null hypotheses are still iid valid p-values that are independent from the false nulls. Hence our result follows from Theorem 3 (SSS+) of [1] for $c = \lambda$, and more generally for $c \leq \lambda$ from Theorem 1 (AS) of [19] (with $s = c$).

Remark 1. With the risk of stating the obvious we note that one cannot simply apply SSS+ or AS by selecting $W_i = Z_i$ for all i with the corresponding empirical p-values $(d_1 - r_i + 1)/d_1$. Indeed, in this case the order of the hypotheses (by W_i) is not independent of the true null p-values.

3.4 When does the NLCP hold for our meta-procedure?

To further analyze the NLCP we make the following assumption on our decoys.

Definition 2 (formalizing the multiple-decoy problem). *If the d_1 (original and decoy) scores corresponding to each true null hypothesis are iid independently of all other scores then we say we have "iid decoys".*

It is clear that if we have iid decoys then for each fixed $i \in N$ the rank r_i is uniformly distributed on $1, \ldots, d_1$, and hence $P(L_i = 1) = c$ and $P(L_i = -1) = 1 - \lambda$. However, to determine whether or not r_i is still uniformly distributed when conditioning on $\mathcal{W}$ we need to look at the mapping function φ as well.

More specifically, in the iid decoys case the conditional distribution of $\{L_i : i \in N\}$ given $\mathcal{W}$ clearly factors into the product of the conditional distribution of each true null L_i given W_i: a true null's L_i is independent of all $\{L_j, W_j : j \neq i\}$. Thus, it suffices to show that L_i is independent of W_i for each $i \in N$. Moreover, because W_i is determined in terms of s_i and the *set* of scores $\left\{\tilde{Z}_i^0, \ldots, \tilde{Z}_i^d\right\}$, and because a true null's label L_i and s_i are independent of the last set (a set is unordered), it suffices to show that L_i is independent of s_i. Of course, s_i is determined by φ as specified in (2).

For example, consider the max method where $s_i \equiv d_1$ (equivalently $\varphi \equiv d_1$): in this case, L_i is trivially independent of s_i and hence by the above discussion the method controls the FDR. In contrast, assuming d_1 is even and choosing $\varphi \equiv d_1$ with $c = \lambda = 1/2$ we see that the scores $\{W_i : i \in N, L_i = -1\}$ will generally be larger than the corresponding $\{W_i : i \in N, L_i = 1\}$. Indeed, when $L_i = -1$ we always choose the maximal score $W_i = Z_i^{(d_1)}$, whereas W_i is one of the top half scores when $L_i = 1$. Hence, $P(L_i = -1 \mid \text{ higher } W_i) > 1/2$.

So how can we guarantee that the NLCP holds for pre-determined values of $c = i_c/d_1$ and $\lambda = i_\lambda/d_1$? The next theorem provides a sufficient condition on φ (equivalently on s_i) to ensure the property holds.

Theorem 2. *If the iid decoys assumption holds, and if for any $i \in N$ and $j \in \{d_1 - i_c + 1, \ldots, d_1\}$*

$$P\left(s_i = j, r_i \leq d_1 - i_\lambda\right) = P\left(s_i = j, L_i = -1\right) = \frac{d_1 - i_\lambda}{d_1 \cdot i_c}, \tag{5}$$

then the NLCP holds and hence our meta-procedure with those values of c, λ and φ controls the FDR.

Proof. By (5), for any $i \in N$ and $j \in \{d_1 - i_c + 1, \ldots, d_1\}$,

$$\begin{aligned} P\left(L_i = 1 \mid s_i = j\right) &= \frac{P\left(s_i = j, L_i = 1\right)}{\sum_{l \in \{-1,0,1\}} P\left(s_i = j, L_i = l\right)} \\ &= \frac{1/d_1}{(d_1 - i_\lambda)/(d_1 \cdot i_c) + (i_\lambda - i_c)/d_1 \cdot 1/i_c + 1/d_1} = \frac{i_c}{d_1} = c, \\ P\left(L_i = -1 \mid s_i = j\right) &= \frac{(d_1 - i_\lambda)/(d_1 \cdot i_c)}{(d_1 - i_\lambda)/(d_i \cdot i_c) + (i_\lambda - i_c)/d_1 \cdot 1/i_c + 1/d_1} = \frac{d_1 - i_\lambda}{d_1} = 1 - \lambda. \end{aligned}$$

At the same time $P\left(L_i = 1 \mid s_i = j\right) = 1$ for $j \in \{1, \ldots, i_c\}$ always holds; therefore, L_i is independent of s_i and by the above discussion the NLCP holds. Theorem 1 completes the proof.

For any fixed values of c, λ we can readily define a randomized $\varphi = \varphi_u$ so that the NLCP holds: randomly and uniformly map $\{1, \ldots, d_1 - i_\lambda\}$ onto $\{d_1 - i_c + 1, \ldots, d_1\}$. Indeed, in this case (5) holds:

$$P\left(s_i = j, s_i \neq r_i\right) = P\left(r_i \in \{1, \ldots, d_1 - i_\lambda\}\right) \cdot P\left(s_i = j \mid r_i \in \{1, \ldots, d_1 - i_\lambda\}\right) = \frac{d_1 - i_\lambda}{d_1} \cdot \frac{1}{i_c}. \tag{6}$$

3.5 Mirroring and Mirandom

Using the above randomized uniform map φ_u we have a way to define an FDR-controlling variant of our meta-procedure for any pre-determined c, λ. However, we can design more powerful procedures using alternative definitions of φ (for the same values of c, λ).

For example, with $c = \lambda = 1/2$ and an even d_1 we can consider, in addition to φ_u, the mirror map: $\varphi_m(j) \coloneqq d_1 - j + 1$. It is easy to see that under the conditions of Theorem 2, $P\left(s_i = j, r_i \leq d_1 - i_\lambda\right) = 1/d_1$ hence (5) holds and the resulting method, which we refer to as the "mirror method"' (because when $L_i = -1$, s_i is the rank symmetrically across the median to r_i), controls the FDR. Similarly, we can choose to use a shift map φ_s: $\varphi_s(j) = j + d_1/2$, which will result in a third FDR-controlling variant of our meta-procedure for $c = \lambda = 1/2$.

Comparing the shift and the mirror maps we note that when $L_i = -1$, φ_s replaces middling target scores with high decoy scores, whereas φ_m replaces low target scores with high decoy scores. Of course, the high decoy scores are the ones more likely to appear in the numerator of (3), and generally we expect the density of the target scores to monotonically decrease with the quality of the score. Taken together, it follows that the estimated FDR will generally be higher when using φ_s than when using φ_m, and hence the variant that uses φ_s will be weaker than the mirror. By extension the randomized φ_u will fall somewhere between the other two maps, as can be partly verified by the comparison of the power using φ_m and φ_u in panel A of Supplementary Figure 1.

We can readily extend the mirroring principle to other values of c and λ where i_c divides $d_1 - i_\lambda$, however when $i_c \nmid d_1 - i_\lambda$ we need to introduce some randomization into the map. Basically, we accomplish this by respecting the mirror principle as much as we can while using the randomization to ensure that (5) holds — hence the name *mirandom* for this map/procedure. It is best described by an example.

Suppose $d = 7$. Then for $i_c = 3$ ($c = 3/8$) and $i_\lambda = 4$ ($\lambda = 1/2$) the mirandom map, φ_{md}, is defined as

$$\varphi_{md}(j) = \begin{cases} 8 & j = 1 \\ 8 \text{ (with probability } 1/3\text{), or } 7 \text{ (with probability } 2/3\text{)} & j = 2 \\ 7 \text{ (with probability } 2/3\text{), or } 6 \text{ (with probability } 1/3\text{)} & j = 3 \\ 6 & j = 4 \end{cases}$$

Note the uniform coverage ($4/3$) of each value in the range, implying that if j is randomly and uniformly chosen in the domain then $\varphi_{md}(j)$ is uniformly distributed over $\{6, 7, 8\}$.

More generally the mirandom map φ_{md} for a given $c \leq \lambda$ is defined in two steps. In the first step it defines a sequence of $d_1 - i_\lambda$ distributions $F_1, \ldots, F_{d_1 - i_\lambda}$ on the range $\{d_1 - i_c + 1, \ldots, d_1\}$ so that

- each F_l is defined on a contiguous sequence of natural numbers, and
- if $j < l$ then F_j stochastically dominates F_l and $\min \text{support}\{F_j\} \geq \max \text{support}\{F_l\}$.

In practice, it is straightforward to construct this sequence of distributions and to see that, when combined, they necessarily satisfy the following equal coverage property: for each $j \in \{d_1 - i_c + 1, \ldots, d_1\}$, $\sum_{l=1}^{d_1 - i_\lambda} F_l(j) = \frac{d_1 - i_\lambda}{i_c}$. In the second step, mirandom defines s_i for any i with $r_i \in \{1, \ldots, d_1 - i_c\}$ by randomly drawing a number from F_{r_i} (independently of everything else).

It follows from the equal coverage property that for any $i \in N$ and $j \in \{d_1 - i_c + 1, \ldots, d_1\}$ (5) holds for φ_{md} for essentially the same reason it held for φ_u in (6). Hence, the mirandom map allows us to controls the FDR for any pre-determined values of c, λ.

3.6 Data-driven setting of the tuning parameters c, λ

All the procedures we consider henceforth are based on the mirandom map. Where they differ is in how they set c, λ. For example, choosing $c = \lambda = 1/2$ gives us the mirror (assuming d_1 is even), $c = \lambda = 1/d_1$ yields the max, while choosing $\lambda = 1/2$ and $c = \alpha \leq 1/2$ coincides with Lei and Fithian's recommendation in the related context of sequential hypothesis testing (technically we set $c = \lfloor \alpha \cdot d_1 \rfloor / d_1$ and refer to this method as "LF"). All of these seem plausible; however, our extensive simulations (Supplementary Section 6.3) show that none dominates the others with substantial power to be gained/lost for any particular problem (Supplementary Figure 1, panels B-D). As the optimal values of c, λ seem to vary in a non-trivial way with the nature of the data, as well as with d and α, we turned to data-driven approaches to setting c, λ.

Lei and Fithian pointed out the connection between the (c, λ) (they refer to c as s) parameters of their AS procedure and the corresponding parameters in Storey's procedure. Specifically, AS's λ is analogous to the parameter λ of [33] that determines the interval $(\lambda, 1]$ from which π_0, the fraction of true null hypotheses, is estimated, and AS's c is Storey's rejection threshold (Supplementary Section 6.4).

We take this analogy one step further and essentially use the procedure of [33] to determine c by applying it to the empirical p-values, $\tilde{p}_i \coloneqq (d_1 - r_i + 1)/d_1$. However, to do that, we first need to determine λ.

We could have determined λ by applying the bootstrap approach of [33] to $\tilde{p}_i$. However, in practice we found that using the bootstrap option of the qvalue package [35] in our setup can significantly compromise our FDR control. Therefore, instead we devised an alternative approach inspired by the spline-based method of [34] for estimating π_0, where we look for the flattening of the tail of the p-value histogram as we approach 1. Because our p-values, $\tilde{p}_i$, lie on the lattice i/d_1 for $i = 1, \ldots, d_1$, instead of threading a spline as in [34], we repeatedly test whether the number of p-values in the first half of the considered tail interval $(\lambda, 1]$ is significantly larger than their number in the second half of this interval (Supplementary Section 6.5).

Our finite-decoy Storey (FDS) procedure starts with determining λ as above then essentially applies the methodology of [33] to $\tilde{p}_i$ to set $c = t_\alpha$ before applying mirandom with the chosen c, λ (Supplementary Section 6.6). We defined FDS as close as possible to Storey, Taylor and Siegmund's recommended procedure for guaranteed FDR control in the finite setting [33]. Indeed, as we argue in Supplementary Section 6.7, FDS converges to a variant of Storey's procedure once we let $d \longrightarrow \infty$ (the mirror and mirandom maps in general have an interesting limit in that setup). However, we found that a variant of FDS that we denote as FDS_1 (Supplementary Section 6.6), often yields better power in our setting, so we considered both variants.

FDS and FDS_1 peek at the data to set c, λ hence they no longer fall under mirandom's guaranteed FDR control. Still, our extensive simulations show they essentially control the FDR: their empirical violations of FDR control are roughly in line with that of the max method, which provably controls the FDR (Supplementary Figure 2). Importantly, FDS_1 seems to deliver overall more power than the mirror, max, LF, FDS and TDC, and often substantially more (Supplementary Figure 3). We note, however, that at times FDS_1 has 10-20% less power than the optimal method, and we observe similar issues with the examples mentioned in Supplementary Section 6.2 where BH has no power (Supplementary Section 6.10). These issues motivate our next procedure.

3.7 A bootstrap procedure for selecting an optimal method

Our final, and ultimately our recommended multi-decoy procedure, uses a novel resampling approach to choose the optimal procedure among several of the above candidates. Our optimization strategy is indirect: rather than using the resamples to choose the method that maximizes the number of discoveries, we use the resamples to advise us whether or not such a direct maximization approach is likely to control the FDR.

Clearly, a direct maximization would have been ideal had we been able to sample more instances of the data. In reality, that is rarely possible all the more so with our underlying assumption that the decoys are given and that it is forbiddingly expensive to generate additional ones. Hence, when a hypothesis is resampled it comes with its original, as well as its decoy scores, thus further limiting the variability of our resamples. In particular, direct maximization will occasionally fail to control the FDR. Our Labeled Bootstrap monitored Maximization (LBM) procedure tries to identify those cases.

In order to gauge the rate of false discoveries we need labeled samples. To this end, we propose a segmented resampling procedure that makes informed guesses (described below) about which of the hypotheses are false nulls before resampling the indices. The scores $\left\{\tilde{Z}_i^j\right\}_{j=0}^{d}$ associated with each resampled conjectured *true null* index are then randomly permuted, which effectively boils down to randomly sampling $j \in \{0, 1, \ldots, d\}$ and swapping the corresponding original score $\tilde{Z}_i^0 = Z_i$ with $\tilde{Z}_i^j$.

The effectiveness of our resampling scheme hinges on how informed are our guesses of the false nulls. To try and increase the overlap between our guesses and the true false nulls we introduced two modifications to the naive approach of estimating the number of false nulls in our sample and then uniformly drawing that many conjectured false nulls. First, we consider increasing sets of hypotheses $\mathcal{H}_j \subset \mathcal{H}_{j+1}$ and verify that the number of conjectured false nulls we draw from each $\mathcal{H}_j$ agrees with our estimate of the

number of false nulls in $\mathcal{H}_j$. Second, rather than being uniform, our draws within each set $\mathcal{H}_j$ are weighted according to the empirical p-values so that hypotheses with more significant empirical p-values are more likely to be drawn as conjectured false nulls. Our segmented resampling procedure is described in detail in Supplementary Section 6.8.

In summary, LBM relies on the labeled resamples of our segmented resampling approach to estimate whether we are likely to control the FDR when using direct maximization (we chose FDS, mirror, and FDS_1 as the candidate methods). If so, then LBM uses the maximizing method; otherwise, LBM chooses a pre-determined fall-back method (here we consistently use FDS_1, see Supplementary Section 6.9 for details).

Our simulations suggest that LBM's control of the FDR is on-par with that of the, provably FDR-controlling, max: the overall maximal observed violation is 5.0% for LBM while it is 6.7% for max, and the number of curves (out of 1200) in which the maximal violation exceeds 2% is 21 for LBM, and 24 for the max (panels A and D, Supplementary Figure 2). Power-wise LBM arguably offers the best balance among our proposed procedures, offering substantially more power in many of the experiments while never giving up too much power when it is not optimal (Supplementary Figure 4). Finally, going back to the examples where BH and Storey's procedure applied to the empirical p-values fail we find that all our methods, including LBM, essentially control the FDR where Storey's procedure substantially failed to do so, and similarly that LBM delivers substantial power where BH had none (Supplementary Section 6.10).

4 The peptide detection problem

Our peptide detection procedure starts with a generalization of the WOTE procedure of [11]. We use Tide [4] to find for each spectrum its best matching peptide in the target database as well as in the d decoy peptide databases. We then assign to the ith target peptide the observed score, Z_i, which is the maximum of all the PSM scores that were optimally matched to this peptide. We similarly define the maximal scores of each of that peptide's d randomly shuffled copies as the corresponding decoy scores: $\tilde{Z}_i^1, \ldots, \tilde{Z}_i^d$. If no spectrum was optimally matched to a peptide then that peptide's score is $-\infty$.

We then applied to the above scores TDC ($d = 1$, with the $+1$ finite sample correction) — representing a peptide-level analogue of the picked target-decoy strategy of [31] — as well as the mirror, LBM and the averaging-based aTDC[3] each using $d \in \{3, 5, 7, 9\}$. Note that to ameliorate the effect of decoy-induced variability on our comparative analysis we report the average of our analysis over 100 applications of each method using that many randomly drawn decoy sets (Supplementary Section 6.11).

We used three datasets in our analysis: "human", "yeast" and "ISB18" (Supplementary Section 6.11). Panel D of Supplementary Figure 5 suggests that when applied to the ISB18 dataset all our procedures seem to control the FDR:[4] the empirically estimated FDR is always below the selected threshold. In terms of power, again we see that LBM is the overall winner: it typically delivers the largest number of discoveries, and even in

[3] We used the version named aTDC_1^+, which was empirically shown to control the FDR even for small thresholds / datasets [18].

[4] Being a controlled experiment, the ISB18 dataset allows us to empirically gauge the FDR.

the couple of cases where it fails to do so it is only marginally behind the top method (panels A–C). In contrast, each of the other methods has some cases where it delivers noticeably fewer discoveries.

More specifically, for $\alpha = 0.01$ LBM's average of 142.0 ISB18 discoveries ($d = 3$) represents an 8.0% increase over TDC's average of 131.5 ISB18 discoveries, and we see a 9.4% increase over TDC when using $d = 5$ (143.3 discoveries). In the human dataset and for the same $\alpha = 0.01$ we see a 2.8% increase in power going from TDC to LBM with $d = 3$ (532.4 vs. 547.1 discoveries), and a 4.2% increase when using LBM with $d = 5$ (555.0 discoveries). LBM offers the biggest gains in the yeast dataset where we see (again $\alpha = 0.01$) a 45.5% increase in power going from TDC to LBM with $d = 3$ (76.3 vs. 111.0 discoveries), and a 46.7% increase when using LBM with $d = 5$ (111.9 discoveries). Moreover, we note that for this $\alpha = 0.01$ TDC reported 0 yeast discoveries in 33 of the 100 runs (each using a different decoy database), whereas LBM reported a positive number of discoveries in all 100 runs for each $d > 1$ we considered.

At the higher FDR thresholds of 0.05 and 0.1 LBM offers a much smaller power advantage over TDC and is marginally behind for $\alpha = 0.1$ and $d = 3$ in the human and yeast datasets. Also, consistent with our simulations, we find that the mirror lags behind LBM, and in fact in these real datasets it is roughly on par with TDC.

Finally, although aTDC was designed for the spectrum identification problem and in practice was never applied to the peptide detection problem, it was instructive to add aTDC to this comparison. LBM consistently delivered more detected peptides than aTDC did, although in some cases the difference is marginal. Still, in the human dataset for $\alpha = 0.01$ with $d = 3$ we see a 4.4% increase in power going from aTDC to LBM (524.2 vs. 547.1 discoveries), and with $d = 5$ a 4.6% increase when using LBM (530.8 vs. 555.0 discoveries). Similarly, in the ISB18 dataset for $\alpha = 0.01$ with $d = 3$ we see a 7.3% increase in power going from aTDC to LBM (132.3 vs. 142.0 discoveries), and with $d = 5$ a 6.4% increase when using LBM (134.7 vs. 143.3 discoveries).

In Supplementary Section 6.12 we discuss a further analysis where we added two more spectra runs to the yeast dataset representing a higher budget experiment. In this case at 1% FDR the average number of TDC discoveries was 275.9 and for LBM using $d = 5$ decoys it was 294. Subsequent Gene Ontology enrichment test of the 54 proteins imputed from the peptide discovered by LBM yielded two overrepresented biological process terms that were not present in the 50 proteins imputed from TDC. The two missing terms—"cellular protein localization" and "cellular macromolecule localization"—are closely related and imply that the sample under investigation is enriched for proteins responsible in shuttling or maintaining other proteins in their proper cellular compartments. Critically, an analysis based solely on the traditional TDC approach would entirely miss this property of the sample being analyzed.

5 Discussion

We consider a new perspective on the peptide detection problem which can be framed more broadly as multiple-competition based FDR control. The problem we pose and the tools we offer can be viewed as bridging the gap between the canonical FDR controlling procedures of BH and Storey and the single-decoy approach of the popular TDC used in

spectrum identification (ID). Indeed, our proposed FDS converges to Storey's method as the number of decoys $d \longrightarrow \infty$ (Supplementary Section 6.7).

The methods we propose here rely on our novel mirandom procedure, which guarantees FDR control in the finite sample case for any pre-determined values of the tuning parameters c, λ. Our extensive simulations show that which of our methods delivers the maximal power varies with the properties of the experiment, as well as with the FDR threshold α. This variation motivates our introduction of LBM. LBM relies on a novel labeled resampling technique, which allows it to select its preferred method after testing whether a direct maximization approach seems to control the FDR. Our simulations, as well as our analysis of peptide detection using real datasets, suggest that LBM largely controls the FDR and seems to offer the best balance among our multi-decoy methods as well as a significant power advantage over the single-decoy TDC.

Finally, as mentioned, our approach is applicable beyond peptide detection. Moreover, while we stated our results assuming iid decoys, the results hold in a more general setting of *"conditional null exchangeability"* (Supplementary Section 6.13). This exchangeability is particularly relevant for future work on generalizing the construction of [1] to multiple knockoffs, where the iid decoys assumption is unlikely to hold.

Related work. We recently developed aTDC in the context of spectrum ID. The goal of aTDC was to reduce the decoy-induced variability associated with TDC by averaging a number of single-decoy competitions [17, 18]. As such, aTDC fundamentally differs from the methods of this paper which simultaneously use all the decoys in a single competition; hence, the methods proposed here can deliver a significant power advantage over aTDC (panel F, Supplementary Figure 4 and Supplementary Figure 5). Our new methods are designed for the iid (or exchangeable) decoys case, which is a reasonable assumption for the peptide detection problem studied here but does not hold for the spectrum ID for which aTDC was devised. Indeed, as pointed out in [16], due to the different nature of native/foreign false discoveries, the spectrum ID problem fundamentally differs from the setup of this paper and even the above, weaker, null exchangeability property does not hold in this case. Thus, LBM cannot replace aTDC entirely; indeed, LBM is too liberal in the context of the spectrum ID problem. Note that in practice aTDC has not previously been applied to the peptide detection problem.

While working on this manuscript we became aware of a related Arxiv submission [14]. The initial version of that paper had just the mirror method, which as we show is quite limited in power. A later version that essentially showed up simultaneously with the submission of our technical report [7] extended their approach to a more general case; however, the method still consists of a subset of our independently developed research in that: (a) they do not consider the λ tuning parameter, (b) they use the uniform random map φ_u which, as we show, is inferior to mirandom, and (c) they do not offer either a general deterministic (FDS) or bootstrap based (LBM) data-driven selection of the tuning parameter(s), relying instead on a method that works only in the limited case-control scenario they consider.

References

1. Barber, R.F., Candès, E.J.: Controlling the false discovery rate via knockoffs. The Annals of Statistics **43**(5), 2055–2085 (2015)

2. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B **57**, 289–300 (1995)
3. Cerqueira, F.R., Graber, A., Schwikowski, B., Baumgartner, C.: MUDE: a new approach for optimizing sensitivity in the target-decoy search strategy for large-scale peptide/protein identification. Journal of Proteome Research **9**(5), 2265–2277 (2010)
4. Diament, B., Noble, W.S.: Faster SEQUEST searching for peptide identification from tandem mass spectra. Journal of Proteome Research **10**(9), 3871–3879 (2011)
5. Elias, J.E., Gygi, S.P.: Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods **4**(3), 207–214 (2007)
6. Elias, J.E., Gygi, S.P.: Target-decoy search strategy for mass spectrometry-based proteomics. Methods in Molecular Biology **604**(55–71) (2010)
7. Emery, K., Hasam, S., Noble, W.S., Keich, U.: Multiple competition based FDR control. arXiv (2019), arXiv:1907.01458
8. Eng, J.K., McCormack, A.L., Yates, III, J.R.: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry **5**, 976–989 (1994)
9. Fan, Y., Lv, J., Sharifvaghefi, M., Uematsu, Y.: IPAD: stable interpretable forecasting with knockoffs inference. Available at SSRN 3245137 (2018)
10. Gao, C., Sun, H., Wang, T., Tang, M., Bohnen, N.I., Müller, M.L., Herman, T., Giladi, N., Kalinin, A., Spino, C., et al.: Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in parkinsons disease. Scientific Reports **8**(1), 7129 (2018)
11. Granholm, V., Navarro, J.F., Noble, W.S., Käll, L.: Determining the calibration of confidence estimation procedures for unique peptides in shotgun proteomics. Journal of Proteomics **80**(27), 123–131 (2013)
12. Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N., Macisaac, K.D., Danford, T.D., Hannett, N.M., Tagne, J., Reynolds, D.B., Yoo, J., Jennings, E., Zeitlinger, J., Pokholok, D., Kellis, M., Rolfe, P.A., Takusagawa, K.T., Lander, E.S., Gifford, D.K., Fraenkel, E., Young, R.A.: Transcriptional regulatory code of a eukaryotic genome. Nature **431**, 99–104 (2004)
13. He, K., Fu, Y., Zeng, W.F., Luo, L., Chi, H., Liu, C., Qing, L.Y., Sun, R.X., He, S.M.: A theoretical foundation of the target-decoy search strategy for false discovery rate control in proteomics. arXiv (2015), `https://arxiv.org/abs/1501.00537`
14. He, K., Li, M., Fu, Y., Gong, F., Sun, X.: A direct approach to false discovery rates by decoy permutations. arXiv preprint arXiv:1804.08222 (2018)
15. Jeong, K., Kim, S., Bandeira, N.: False discovery rates in spectral identification. BMC Bioinformatics **13**(Suppl. 16), S2 (2012)
16. Keich, U., Noble, W.S.: Controlling the FDR in imperfect database matches applied to tandem mass spectrum identification. Journal of the American Statistical Association (2017), https://doi.org/10.1080/01621459.2017.1375931
17. Keich, U., Noble, W.S.: Progressive calibration and averaging for tandem mass spectrometry statistical confidence estimation: Why settle for a single decoy. In: Sahinalp, S. (ed.) Proceedings of the International Conference on Research in Computational Biology (RECOMB). Lecture Notes in Computer Science, vol. 10229, pp. 99–116. Springer (2017)
18. Keich, U., Tamura, K., Noble, W.S.: Averaging strategy to reduce variability in target-decoy estimates of false discovery rate. Journal of Proteome Research **18**(2), 585–593 (2018)
19. Lei, L., Fithian, W.: Power of ordered hypothesis testing. In: International Conference on Machine Learning. pp. 2924–2932 (2016)
20. Levitsky, L.I., Ivanov, M.V., Lobas, A.A., Gorshkov, M.V.: Unbiased false discovery rate estimation for shotgun proteomics based on the target-decoy approach. Journal of Proteome Research **16**(2), 393–397 (2017)

21. Lin, H., He, Q.Y., Shi, L., Sleeman, M., Baker, M.S., Nice, E.C.: Proteomics and the microbiome: pitfalls and potential. Expert Reviews in Proteomics **16**(6), 501–511 (2019)
22. Lu, Y.Y., Fan, Y., Lv, J., Noble, W.S.: DeepPINK: reproducible feature selection in deep neural networks. In: Advances in Neural Information Processing Systems (2018)
23. Morris, M., Knudsen, G.M., Maeda, S., Trinidad, J.C., Ioanoviciu, A., Burlingame, A.L., Mucke, L.: Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. Nature Neuroscience **18**, 1183–1189 (2015)
24. Nesvizhskii, A.I.: 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)
25. Ng, P., Keich, U.: Gimsan: a gibbs motif finder with significance analysis. Bioinformatics **24**(19), 2256–2257 (2008)
26. Noble, W.S., MacCoss, M.J.: Computational and statistical analysis of protein mass spectrometry data. PLOS Computational Biology **8**(1), e1002296 (2012)
27. P. Hernandez, M.M., Appel, R.D.: Automated protein identification by tandem mass spectrometry: Issues and strategies. Mass Spectrometry Reviews **25**, 235–254 (2006)
28. Ping, L., Duong, D.M., Yin, L., Gearing, M., Lah, J.J., Levey, A.I., Seyfried, N.T.: Global quantitative analysis of the human brain proteome in Alzheimers and Parkinsons disease. Scientific Data **5**, 180036 (2018)
29. Read, D.F., Cook, K., Lu, Y.Y., Le Roch, K., Noble, W.S.: Predicting gene expression in the human malaria parasite *plasmodium falciparum*. Journal of Proteome Research (2019), in press
30. Saito, M.A., Bertrand, E.M., Duffy, M.E., Gaylord, D.A., Held, N.A., Hervey, W.J., Hettich, R.L., Jagtap, P.D., Janech, M.G., Kinkade, D.B., Leary, D.H., McIlvin, M.R., Moore, E.K., Morris, R.M., Neely, B.A., Nunn, B.L., Saunders, J.K., Shepherd, A.I., Symmonds, N.I., Walsh, D.A.: Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. Journal of Proteome Research **18**(4), 1461–1476 (2019)
31. Savitski, M.M., Wilhelm, M., Hahne, H., Kuster, B., Bantscheff, M.: A scalable approach for protein false discovery rate estimation in large proteomic data sets. Molecular & Cellular Proteomics **14**(9), 2394–2404 (2015)
32. Storey, J.D.: A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B **64**, 479–498 (2002)
33. Storey, J.D., Taylor, J.E., Siegmund, D.: Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. Journal of the Royal Statistical Society, Series B **66**, 187–205 (2004)
34. Storey, J.D., Tibshirani, R.: Statistical significance for genome-wide studies. Proceedings of the National Academy of Sciences of the United States of America **100**, 9440–9445 (2003)
35. Storey, J.D., Bass, A.J., Dabney, A., Robinson, D.: qvalue: Q-value estimation for false discovery rate control (2019), `http://github.com/jdstorey/qvalue`, r package version 2.14.1
36. Tusher, V.G., Tibshirani, R., Chu, G.: Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America **98**, 5116–5121 (Apr 2001). https://doi.org/10.1073/pnas.091062498
37. Wildburger, N.C., Esparza, T.J., LeDuc, R.D., Fellers, R.T., Thomas, P.M., Cairns, N.J., Kelleher, N.L., Bateman, R.J., Brody, D.L.: Diversity of amyloid-beta proteoforms in the Alzheimers disease brain. Scientific Reports **7**, 9520 (2017)
38. Xiao, Y., Angulo, M.T., Friedman, J., Waldor, M.K., WeissT, S.T., Liu, Y.Y.: Mapping the ecological networks of microbial communities. Nature Communications **8**(1), 2042 (2017)