Competition-based control of the false discovery proportion

Dong Luo | Arya Ebadi | Kristen Emery | Yilun He | William Stafford Noble | Uri Keich

1 School of Mathematics and Statistics, University of Sydney, New South Wales, Australia
2 Departments of Genome Sciences, University of Washington, Seattle, Washington, USA

Correspondence
Uri Keich, School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia.
Email: uri.keich@sydney.edu.au

Abstract
Recently, Barber and Candès laid the theoretical foundation for a general framework for false discovery rate (FDR) control based on the notion of “knockoffs.” A closely related FDR control methodology has long been employed in the analysis of mass spectrometry data, referred to there as “target–decoy competition” (TDC). However, any approach that aims to control the FDR, which is defined as the expected value of the false discovery proportion (FDP), suffers from a problem. Specifically, even when successfully controlling the FDR at level \( \alpha \), the FDP in the list of discoveries can significantly exceed \( \alpha \). We offer FDP-SD, a new procedure that rigorously controls the FDP in the knockoff/TDC competition setup by guaranteeing that the FDP is bounded by \( \alpha \) at a desired confidence level. Compared with the recently published framework of Katsevich and Ramdas, FDP-SD generally delivers more power and often substantially so in simulated and real data.

KEYWORDS
FDP control, knockoffs, peptide detection, target–decoy competition

1 INTRODUCTION

1.1 Competition-based approach to multiple hypothesis testing

Competition-based approach to multiple hypothesis testing offers an alternative way of tackling the problem in the absence of canonically computed \( p \)-values. Instead, this fairly new approach derives its analysis using one or, more generally, a small number of competing null scores. Specifically, 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 testing setup, we further assume that we cannot compute informative \( p \)-values for the observed scores. Instead, we can only generate for each \( H_i \) one competing null score, \( \tilde{Z}_i \), and the analysis relies on the competition between \( Z_i \) (referred to as the “target” or “observed/original” score) and its corresponding \( \tilde{Z}_i \) (the “decoy” or “knockoff” score).

In the computational mass spectrometry community such target–decoy competition (TDC) has been widely practiced to control the false discovery rate (FDR) since it was first proposed by Elias and Gygi (2007), Cerqueira et al. (2010), and The et al. (2016). More recently, Barber and Candès (2015) used the same principle in their knockoff+ procedure to control the FDR in feature selection in a classical linear regression model. While Barber and Candès’ knockoff construction is significantly more elaborate than that of the analogous decoys in the mass spectrometry context (more on that below), knockoffs and decoys serve the same purpose in competition-based FDR control.
Indeed, the procedures to control the FDR, known as TDC (‘target–decoy competition’ or target–decoy analysis) in the mass spectrometry community, and SequentialStep+ (with \(c = 1/2\)) in the knockoff setting proceed along the same lines. In both cases, we only reject (report as discoveries) hypotheses \(H_i\) which are target/original wins: \(Z_i > Z_i\). Additionally, the number of decoy/knockoff wins \((Z_i > Z_i)\) in the top \(k\) scoring hypotheses is used to estimate the number of false discoveries in the target wins among the same top \(k\) hypotheses. Thus, the ratio between the number of decoy wins and the number of target wins yields an estimate of the FDR among the target wins in the top \(k\) hypotheses. To control the FDR at level \(\alpha\), the procedures choose the largest \(k\) for which the estimated FDR is still \(\leq \alpha\), and it reports all target wins among those top \(k\) hypotheses (e.g., TDC in Supplementary Section 7.2). It was recently shown that, assuming that true nulls are independently equally likely to end up as a target or a decoy win, 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 (Barber & Candès, 2015; He et al., 2015).

### 1.2 Controlling the FDR in variable selection

Barber and Candès (2015) used the above competition-based principle in their knockoff+ procedure to control the FDR in feature selection in a classical linear regression model, \(\mathbf{y} = \mathbf{X}\beta + \varepsilon\), where \(\mathbf{y} \in \mathbb{R}^n\) is the response vector, \(\mathbf{X}\) is the \(n \times p\) known, real-valued design matrix, \(\beta \in \mathbb{R}^p\) is the unknown vector of coefficients, and \(\varepsilon \sim \mathcal{N}(0, \sigma^2 I)\) is Gaussian noise. Briefly, knockoff+ relies on introducing an \(n \times p\) knockoff design matrix \(\tilde{\mathbf{X}}\), where each column consists of a knockoff copy of the corresponding original variable. These knockoff variables are constructed so that in terms of the underlying regression problem the true null features (the ones that are not included in the model) are in some sense indistinguishable from their knockoff copies. The procedure then assigns to each null hypothesis \(H_i : \beta_i = 0\) two test statistics \(Z_i, \tilde{Z}_i\) which correspond to the point \(\lambda\) on the Lasso path (Tibshirani, 1996) at which feature \(X_i\), respectively, its knockoff competition \(\tilde{X}_i\), first enters the model. The intuition here is that generally \(Z_i > \tilde{Z}_i\) for true model features, whereas for null features \(Z_i\) and \(\tilde{Z}_i\) are identically distributed. The procedure then continues with the competition, reporting only features where the original scores are higher than the knockoff as described above.

Following that work and the introduction of a more flexible formulation of the variable selection problem in the model-X framework of Candès et al. (2018), competition-based FDR control has gained a lot of interest in the statistical and machine learning communities, where it has been applied to various applications in biomedical research (Xiao et al., 2017) and has been extended to work in conjunction with deep neural networks (Lu et al., 2018) and time series data (Fan et al., 2018), as well as to work in a likelihood setting without requiring the use of latent variables (Sudarshan et al., 2020).

### 1.3 Controlling the FDR and the false discovery proportion

FDR control is a popular approach to the analysis of multiple testing. However, it should not be confused with controlling the false discovery proportion (FDP). The latter is the proportion of true nulls among all the rejected hypotheses (discoveries), and the FDR is its expectation (taken with respect to the true null hypotheses). In particular, while controlling the FDR at level \(\alpha\), the FDP in any given sample can exceed \(\alpha\).

Thus, in practice, controlling the FDP is arguably more relevant than the FDR in many cases. Figure 1 provides examples from both the mass spectrometry (left) and feature selection from linear regression (right) domains showing that while the FDR is controlled (left: \(\alpha = 0.05 > 0.047 = \text{FDP} \), right: \(\alpha = 0.2 > 0.198 = \text{FDP}\)) the FDP can significantly exceed \(\alpha\).

When introducing the notion of FDR, Benjamini and Hochberg (1995) noted that, strictly speaking, the FDP cannot be controlled at any nontrivial level. Indeed, imagine that all the hypotheses are true nulls: rejecting any hypothesis would then imply the FDP is 1. Nevertheless, some notions of controlling the FDP, or false discovery exceedance (FDX) control, have been extensively studied in the canonical setup of multiple hypothesis testing where \(p\)-values are available.

One such approach is to construct confidence bands, or envelopes, that can be used to simultaneously control the FDP at any threshold \(\alpha \in (0, 1)\) at the prescribed confidence level \(1 - \gamma\). That is, the procedure provides a discovery list for each \(\alpha \in (0, 1)\) at the prescribed confidence level \(1 - \gamma\). This simultaneous approach, which can be used in post hoc analysis, was recently further advanced by Gemen et al. (2021) who demonstrated that the closed testing principle can be used to improve such procedures.

An alternative approach is to control the FDP at a predetermined threshold \(\alpha\) with the same confidence level \(1 - \gamma\), that is, the probability that the FDP in the procedure’s discovery list exceeds \(\alpha\) is \(\leq \gamma\). A representative of this approach, which inspired the work presented here, is a step-down procedure—a principle explained in Section 3.2—developed by Guo and Romano (2007).
Returning to our competition setup, while focusing on $k$-FWER control (i.e., no more than $k$ false discoveries) in the competition setup, Janson and Su (2016) suggested how one can use their approach to control the FDP at a single threshold $\alpha$ (FDX-control). More recently, Katsevich and Ramdas (2019) developed a general framework for simultaneous control of the FDP that applies in both the canonical $p$-values setup as well as in both the canonical $p$-values setup as well as in both the canonical $p$-values setup as well as in our analysis relies on (see Supplementary Table 1 for a summary of our notation). Let $H_i$ ($i = 1, \ldots, m$) denote our $m$ null hypotheses, for example, in the linear regression problem $H_i$ is “the coefficient of the $i$th feature is 0.” Associated with each $H_i$ are two competing scores: a target/observed score $Z_i$ (the higher the score the less likely $H_i$ is) and a decoy/knockoff score $\tilde{Z}_i$. For example, in linear regression $Z_i$ ($\tilde{Z}_i$) correspond to the point $\lambda$ on the Lasso path at which the feature (its knockoff) entered the model. Adopting the notation of Emery et al. (2020), we associate with each hypothesis a score $W_i$ and a target/decoy-win label $L_i$. By default $W_i = Z_i \vee \tilde{Z}_i$ (where $x \vee y$ denotes $\max(x, y)$) but as Barber and Candès pointed out, other functions such as $W_i = |Z_i - \tilde{Z}_i|$ can be considered as well.

As for $L_i$:

$$L_i = \begin{cases} 1 & Z_i > \tilde{Z}_i \text{ (} H_i \text{ corresponds to a target/original feature win)} \\ 0 & Z_i = \tilde{Z}_i \text{ (tie,} H_i \text{ is ignored)} \\ -1 & Z_i < \tilde{Z}_i \text{ (} H_i \text{ corresponds to a decoy/knockoff win)} \end{cases}.$$  

Because $H_i$ is ignored if $L_i = 0$ without loss of generality, we assume that $L_i \neq 0$ for all $i$.

Let $N = \{i : H_i \text{ is a true null}\}$ and note that while typically in the context of hypotheses testing $N$ is a constant, albeit unknown set, it is beneficial here to allow $N$ to be a random set as well. Our fundamental assumption is the following:
**Assumption 1.** Conditional on all the scores $\{W_i\}_i$ and all the false null labels $\{L_i : i \notin N\}$, the true nulls are independently equally likely to be a target or a decoy win, that is, the random variables (RVs) $\{L_i : i \in N\}$ are conditionally independent uniform $\pm 1$ RVs.

Clearly, if the assumption holds then $\{L_i : i \in N\}$ are still independent uniform $\pm 1$ RVs after ordering the hypotheses in decreasing order so that $W_1 \geq W_2 \geq \cdots \geq W_m$.

Specific competition paradigms that satisfy Assumption 1 include the theoretical model of TDC introduced by He et al. (2015) to study the peptide-detection problem (Section 5.4). Indeed, their assumptions of “equal chance” and “independence” are an equivalent formulation of our assumption. Similarly, Assumption 1 is satisfied in the regression context for both the original FX (fixed design matrix $X$) knockoff scores construction of Barber and Candès (2015, Lemma 1.1 and its ensuing discussion), and the MX (random design matrix) knockoffs of Candès et al. (2018, Lemma 2). Note that in both cases our $W_i$ is their $|W_i|$ and $L_i$ is the sign of their $W_i$.

Our list of reported discoveries consists of all target wins among the top $k$ scores for some $k$. Therefore, without loss of generality, we assume our hypotheses are ordered in decreasing order of $W_i$, and our goal is to analyze the following random variables/processes (for $i = 0$ we set all counts to 0): $D_i = \sum_{j=1}^{i} 1_{\{L_j = −1\}}$ (the number of decoy wins in the top $i$ scores), $T_i = \sum_{j=1}^{i} 1_{\{L_j = 1\}}$ (the corresponding number of target wins); no ties imply $D_i + T_i = i$, and $N_i^T := \sum_{j=1}^{i} 1_{\{L_j = 1, j \in N\}}$ (the number of true null target wins in the top $i$ scores). With this notation, the FDP among all target wins in the top $i$ scores is $Q_i = N_i^T / (T_i \lor 1)$.

Note that above we assumed that all target–decoy ties ($L_i = 0$) are thrown out, but if instead we randomly break ties then Assumption 1 still holds. In our practical analysis we randomly broke ties. Similarly, how the $W_i$ are sorted in case of ties should not affect the theoretical analysis as long as that ordering is independent of the corresponding labels.

## 3 CONTROLLING THE FDP

### 3.1 Katsevich and Ramdas’ approach to FDP control

A stochastic process $\{\xi_i\}_i$ is a $1 − \gamma$ upper prediction band for the random process $Z_i$ with $i \in I \subset \mathbb{N}$ if $P(\exists i \in I : Z_i > \xi_i) \leq \gamma$. Katsevich and Ramdas recently developed a general framework for constructing such bands that, as they showed, can be specialized to construct an upper prediction band $\{\xi_i\}$ on $V_d := N_i^T$ (the number of true null target wins before the $d$th decoy win). As pointed out by Katsevich and Ramdas, this band can be used to control the FDP by reporting all target wins among the top $k_{KR}$ scores, where

$$k_{KR} = \max \{k \leq m : \frac{\xi_{D_k+1}}{T_k} \leq \alpha \lor k = 0\}.$$  

We refer to this procedure as FDP-KRB (Algorithm 2, Supplementary Section 7.2).

### 3.2 FDP-SD: A novel approach to FDP control via stepdown

Originating in the canonical context where $p$-values are available, step-down procedures work by sequentially comparing the $i$th smallest $p$-value, $p_{(i)}$, against a precomputed bound $\delta_i$. Specifically, the procedure looks for $k_{SD} = \max\{i : p_{(i)} \leq \delta_i \text{ for } j = 1, 2, \ldots, i\}$ and rejects the corresponding $k_{SD}$ hypotheses (Lehmann & Romano, 2005).

FDP-SD is inspired by a step-down procedure developed by Guo and Romano (2007) to control the FDP when $p$-values are available. Because we have no $p$-values in our competition context, we instead use the number of decoy wins: FDP-SD sequentially goes through the hypotheses sorted in order of decreasing scores, comparing the observed number of decoy wins $D_i$ with precomputed bounds $\delta_{\alpha, \gamma}(i)$ that depend on the desired FDP threshold $\alpha$ and the confidence level $1-\gamma$.

The bounds $\delta_{\alpha, \gamma}(i)$ are set to allow us to control the FDP when rejecting all target wins in the top $i$ scores for a fixed $i$. Specifically, imagine we report all target wins in the top $i$ scores if $D_i \leq \delta_{\alpha, \gamma}(i)$, and otherwise we report none. Then $\delta_{\alpha, \gamma}(i)$ should be sufficiently small so that regardless of the number of true nulls among the top $i$ scores, the probability that the FDP among our reported discoveries exceeds $\alpha$ is $\leq \gamma$. To ensure optimality of the bound, we also require that the same cannot be guaranteed for any bound greater than $\delta_{\alpha, \gamma}(i)$. It is not difficult to show that this requires us to define $\delta_{\alpha, \gamma}(i)$ as

$$\delta_{\alpha, \gamma}(i) := \max\{d \in \{-1, 0, 1, \ldots, i - 1\} : F_{B(k(d)+d,1/2)}(d) \leq \gamma\},$$  

where $k(d) = k(i, d) := \lfloor (i - d)\alpha \rfloor + 1$ and $F_{B(n,p)}(-)$ denotes the cumulative distribution function (CDF) of a binomial $B(n, p)$ RV so $F_{B(k(d)+d,1/2)}(d) = P[B(k(d)+d, 1/2) \leq d]$.

The intuition here is that if there are $k(d)$ or more false (true null) discoveries, then the FDP exceeds $\alpha$ so we make
sure that the probability there were \( k(d) \) or more true null target wins was bounded by \( \gamma \). The reason we can do this is that, without loss of generality all decoy wins are true nulls, and hence the total number of true nulls in the top \( i \) scores is bounded from below by \( k(d) + d \), and each true null is independently equally likely to be a target or a decoy win. Hence, the unobserved number of false target wins is stochastically bounded by a \( B(k(d) + d, 1/2) \). \( RV \).

Typically, \( \delta_{\alpha, \gamma}(i) = -1 \) for small values of \( i \). Indeed, with \( i = 1 \) it is impossible to get any confidence \( 1 - \gamma > 1/2 \), that the corresponding hypothesis is not a true null target win. Therefore, we should only compare \( D_i \) with \( \delta_{\alpha, \gamma}(i) \) when the latter is \( \geq 0 \). Using Lemma 2 in Supplementary Section 7.3, which shows that \( F_{B(k(d) + d, 1/2)}(d) \) is increasing in \( d \leq i \), it is easy to see that with

\[
i_o = \delta_{\alpha, \gamma}(\alpha, \gamma) := \max\{1, \lceil \lceil \log_2 (1/\gamma) \rceil - 1 \rceil / \alpha \}, \quad (4)\]

\( i \geq i_o \) if and only if \( \delta_{\alpha, \gamma}(i) \geq 0 \). Note also that for a fixed \( \alpha \) and \( \gamma \), \( \delta_{\alpha, \gamma}(i) \) is nondecreasing in \( i \).

After computing \( i_o \) FDP-SD finds

\[
k_{FDP-SD} = \max \left\{ i : \prod_{j=0}^{i} 1_{D_j \leq \delta_{\alpha, \gamma}(j)} = 1 \text{ or } i = 0 \right\}, \quad (5)\]

where \( 1_A \) is the indicator of the event \( A \), and it reports the \( T_{FDP-SD} \) target discoveries (wins) among the top \( k_{FDP-SD} \) scores. The following theorem guarantees that the FDP-SD procedure, which is summarized in Supplementary Section 7.2, controls the FDP.

**Theorem 1.** With \( k_{FDP-SD} \) defined as in (5) let \( Q_{FDP-SD} \) be the FDP among the \( T_{FDP-SD} \) target wins in the top \( k_{FDP-SD} \) scores. Then \( P(Q_{FDP-SD} > \alpha) \leq \gamma \).

The proof, inspired by that of Theorem 3.2 of Guo and Romano (2007), is given in Supplementary Section 7.3.

The bounds \( \delta_{\alpha, \gamma} \) that FDP-SD relies on are computed in (3) using binomial CDFs. Because the binomial distribution is discrete, it is typically impossible to find a \( d \) for which \( F_{B(k(d) + d, 1/2)}(d) \leq \gamma \) holds with equality. As a result, FDP-SD typically attains a higher confidence level than required: \( P(Q_{FDP-SD} > \alpha) < \gamma \). We address this issue by introducing a more powerful, randomized version of FDP-SD in Supplementary Section 7.2. The proof that the randomized version still rigorously controls the FDP is similar to the proof of Theorem 1 so it is skipped here.

## 4 Extending FDP-SD to Utilize Multiple Decoys

Emery et al. (2020) recently developed FDR-controlling procedures for the setup where we have \( d > 1 \) decoys for each hypothesis. Using their framework, which is applicable when the decoys are independently generated, as well as when they satisfy a weaker exchangeability condition (Emery et al., 2020, Supplementary Section 6.13 of that paper), we can extend FDP-SD to take advantage of multiple decoys in a fairly straightforward manner.

Indeed, assume that associated with each of the \( m \) hypotheses are \( d \) decoys. Let \( d_i := d + 1 \) and let \( c = i_c/d_i \) and \( \lambda = i_l/d_i \) with \( i_c, i_l \in \{1, 2, ..., d\} \) be the target and decoy win thresholds (here we regard these thresholds as predetermined tuning parameters and reserve the question of how to set them for future research).

Let \( r_i \in \{1, ..., d_i\} \) be the rank of the target score in the combined list of the target and all decoy scores associated with hypothesis \( H_i \) (with higher ranks corresponding to larger scores). As usual, we break any ties among the scores at random. Define the label \( L_i \) associated with \( H_i \) by

\[
L_i := \begin{cases} 1 \quad \text{if } r_i \geq d_i - i_c + 1 \\ 0 \quad \text{if } r_i \in (d_i - i_l, d_i - i_c + 1) \\ -1 \quad \text{if } r_i \leq d_i - i_l \\ 
\end{cases} \quad (6)
\]

In words, if the rank of the target score is among the top \( i_c \) ranks (top \( 100 \cdot c \)% ranks) we label \( H_i \) as a target win, whereas if the target rank is among the bottom \( d_i - i_l \) ranks (bottom \( 100 \cdot (1 - \lambda) \)% ranks) we label \( H_i \) as a decoy win. Otherwise, we ignore \( H_i \) for the rest of the procedure, labeling it with \( L_i = 0 \).

Define the winning score \( W_i \) to be the \( s_i \)th highest ranked score for hypothesis \( H_i \), where

\[
s_i := \begin{cases} r_i \quad \text{if } L_i = 1 \\ u_i \quad \text{if } L_i = 0 \\ \varphi(r_i) \quad \text{if } L_i = -1 \\ \end{cases} \quad (7)
\]

Here, \( u_i \) is a (uniformly chosen) random element of \( \{d_1 - i_c + 1, ..., d_i\} \), and \( \varphi : \{1, ..., d_i - i_l\} \rightarrow \{d_1 - i_c + 1, ..., d_i\} \) is a map of losing ranks (those for which \( L_i = -1 \) into winning ranks (those for which \( L_i = 1 \)). In words, (7) says that if we have a target-winning hypothesis (i.e., \( L_i = 1 \)), the winning score is the target score; otherwise, the winning score is one of the decoy scores among the winning ranks. The mapping \( \varphi \) is constructed so that assuming, for example, that the decoys are independently generated, the rank \( r_i \) of a true null target score is distributed uniformly in \( \{1, ..., d_i\} \). The formal definition of this mapping is given in Emery et al. (2020), but two common choices are the max mapping, \( \varphi \equiv d_1 \), and the mirror mapping, \( \varphi(j) = d_i - j + 1 \), which combined with setting \( c = \lambda = 1/2 \) gives us the mirror method that we use in this work. Note that this construction extends the single decoy case, where a truly null hypothesis is required to be a target or decoy.
win with equal probability (in this case, there is only one possible mapping function).

Once we labeled the hypotheses and computed the winning scores, we apply a slightly generalized version of FDP-SD that is adapted to make use of the multiple decoys (see Algorithm 5 and its randomized version, Algorithm 6, in the supplement).

The proof that, for a predetermined choice of c and λ, both these procedures control the FDP in the resulting list of discoveries is almost identical to that of Theorem 1. The key difference is that the probability of observing a decoy-winning true null, given that it was counted, is no longer 1/2, as in the single decoy case, but instead

\[ R = R(c, \lambda) := \frac{1 - \lambda}{c + 1 - \lambda} . \]  

(8)

5 Applications to Real and Simulated Data

To evaluate the procedures presented here we looked at their performance on simulated and real data where competition-based FDR control is already an established practice. In each case our model, and specifically Assumption 1, either explicitly holds or is believed to be a reasonable approximation. Hence, we can apply and compare the competition-based FDR/FDP-controlling procedures presented here: TDC, FDP-KRB, and the novel FDP-SD (here we used the randomized version described in Supplementary Section 7.2).

5.1 Feature selection in linear regression: Simulated and GWAS data

Starting with controlling the FDR/FDP in variable selection via knockoffs, we looked at the first example of Tutorial 1 of “Controlled variable Selection with Model-X Knockoffs” (http://web.stanford.edu/group/candes/knockoffs/software/knockoffs/tutorial-1-r.html “Variable Selection with Knockoffs”) (Candès et al., 2018). Specifically, we repeated the following sequence of operations 1000 times: we randomly drew a normally distributed 1000 × 1000 design matrix and generated a response vector using only 60 of the 1000 variables while keeping all other parameters the same as in the online example (amplitude = 4.5, \( \rho = 0.25 \), \( \Sigma \) is a Toeplitz matrix whose dth diagonal is \( \rho^{d-1} \)). We then computed the model-X knockoff scores (taking a negative score as a decoy win and a positive score as a target win) and applied all the procedures at FDR/FDP levels \( \alpha \in \{0.1, 0.2\} \) and confidence levels of \( 1 - \gamma \in \{0.90, 0.95\} \).

Our GWAS (genome-wide association studies) example is taken from Katsevich and Ramdas (2019), which in turn is based on data made publicly available by Sesia et al. (2020). The goal of this analysis was to identify genomic loci (the features) that are significant factors in the expression of each of the eight traits that were analyzed (the dependent variables). The raw data were taken from the UK Biobank (Bycroft et al., 2018) and transformed to a regression problem by Sesia et al. (2020), who then created knockoff statistics. We downloaded the scores using the functions download_kz_data and read_kz_data defined in Katsevich and Ramdas’ (https://raw.githubusercontent.com/ekatsevi/simultaneous-fdp/master/UKBB_utils.R) UKBB_utils.R. Consistent with the latter, we applied TDC with \( \alpha = 0.1 \), whereas the FDP controlling procedures used \( \gamma = 0.05 \) and \( \alpha \in \{0.05, 0.1, 0.2\} \).

Considering the GWAS data, Figure 2 (top-left) shows that for \( \alpha = 0.05 \) FDP-SD yields the larger number of discoveries for all eight traits with FDP-KRB typically yielding only 0–50% of the number reported by FDP-SD. For \( \alpha = 0.1, 0.2 \) (middle and bottom left panels), the results are a little more mixed: for three of the 16 trait–parameters combinations FDP-SD loses to FDP-KRB but for the other 13 FDP-SD yields more discoveries and typically by a wide margin. Similarly, in our simulated linear regression data for all parameter combinations FDP-SD again reports more discoveries than FDP-KRB (Figure 3).

In terms of how much power is given up when controlling the FDP using FDP-SD versus controlling the FDR using TDC, in the GWAS example we see wide variations in terms of power loss, depending on the trait–parameters combination: Figure 2 (right column). A similar variability is observed in the linear regression simulation (Figure 3, this figure appears in color in the electronic version of this article, and any mention of color refers to that version): compare the “*” mark in the TDC column with the blue (\( \gamma = 0.05 \)) and red (\( \gamma = 0.1 \)) “*” marks in FDP-SD’s column, as well as the TDC’s “+” mark with the corresponding blue (\( \gamma = 0.05 \)) and red (\( \gamma = 0.1 \)) “+” marks of FDP-SD.

5.2 Normal mixtures model

We next used the same mixture of normals model as in Emery et al. (2020) to compare the performance of the competition-based procedures across a wider variety of controlled setups. Briefly, we drew decoy scores (\( Z_i \)) as well as true null scores (\( Z_i, i \in N \)) from a hypothesis-specific \( N(\mu_i, \sigma_i) \) distribution and false null target scores (\( Z_i, i \notin N \)) from a shifted \( N(\mu_i + \rho_i, \sigma_i^2) \).

We distinguish between simulating calibrated scores, where the null distribution does not vary with the
hypothesis, that is, $\mu_i = \mu$ and $\sigma_i = \sigma$ for all $i$ (in this paper, we used $\mu = 0$, $\sigma = 1$ as well as fixing $\rho_i = 3$, except when noted otherwise), and uncalibrated scores. In the latter case, $\mu_i$ is sampled from an $N(0, 1)$ distribution, $\sigma_i = 1 + \xi_i$, where $\xi_i$ is sampled from $\exp(1)$ (the exponential distribution with rate 1), and $\rho_i$ is sampled from a $1 + \exp(\nu)$ distribution, where $\nu$ is a hyperparameter that determines the degree of separation between the false and true null target scores (we used $\nu = 0.075$).

We used this model to randomly draw 40K sets of paired target and decoy scores for each of the following 18 data–parameter combinations: varying the number of hypotheses $m \in \{500, 2k, 10k\}$, the proportion of true nulls $\pi_0 \in \{0.2, 0.5, 0.8\}$, and using calibrated or uncalibrated scores.

**FIGURE 2** Power of FDP controlling procedures in the GWAS example. For each of the eight investigated traits the left column panels describe the relative power of FDP-KRB and FDP-SD defined as the ratio of the number of genomic loci the method discovers (averaged over 1K runs for FDP-SD), while controlling the FDP, over the number reported by the optimal method for that trait–parameters combination (indicated below the trait). The confidence level was fixed at $1 - \gamma = 0.95$, and we varied the FDP/FDR threshold: $\alpha = 0.05$ (top row), $\alpha = 0.1$ (middle row), and $\alpha = 0.2$ (bottom row). The right column panels show the power loss when controlling the FDP (confidence $1 - \gamma = 0.95$) using FDP-SD versus controlling the FDR (same $\alpha$) using TDC, where we varied the FDP/FDR threshold: $\alpha = 0.05$ (top row), $\alpha = 0.1$ (middle row), and $\alpha = 0.2$ (bottom row). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.
We then applied the considered FDR/FDP-controlling procedures to each simulated competition set with FDR/FDP thresholds of \( \alpha = 1\% \), 5\%, and 10\%, and confidence levels \( 100(1 - \gamma) = 95\% \) and 99\%.

The left panel of Figure 4 shows that in the normal mixtures simulation FDP-KRB’s median power never exceeds that of FDP-SD with a typical power loss of 14\% compared with the latter. At the same time, the right panel of the figure shows that for these data the median power loss when controlling the FDP using FDP-SD versus controlling the FDR using TDC is 6.5\% when using \( \gamma = 0.05 \), and it is 11.2\% when using \( \gamma = 0.01 \).

To gain further insight on how the three procedures differ, we compared their results when we vary a single parameter of the mixtures model at a time. Again, for each parameter combination of our model, we applied the procedures to 40K sets of paired target–decoy calibrated scores and examined their reported discoveries. Starting with comparing FDP-SD and FDP-KRB, we applied both while controlling the FDP at \( \alpha \in \{0.01, 0.02, \ldots , 0.1\} \) and with confidence of \( 100(1 - \gamma) = 95\% \) respectively. Supplementary Figure 6 shows the median power (over 40K sets) of each procedure as we first vary \( m \), and then the “signal-to-noise ratio” parameters, \( \pi_0 \) and the separation parameter \( \rho \). In all considered cases, the median power of FDP-SD was at least as high as FDP-KRB’s and often by a wide margin.

We similarly compared FDP-SD with the FDR-controlling TDC using a common FDP/FDR threshold of \( \alpha = 0.05 \), and we set \( \gamma = 0.05 \) for FDP-SD (calibrated scores, \( \pi_0 = 0.5 \)). Supplementary Figure 7 shows that, as expected, increasing \( m \) (keeping the other parameters constant) yields diminished variability in TDC’s FDP (top row). At the same time, the power loss associated with FDP-SD’s increased confidence also diminishes (middle row) as its power increases (bottom row). A similar evolution is observed in Supplementary Figure 8 as we decrease \( \pi_0 \) while keeping all other parameters the same (\( \alpha = 0.05, \gamma = 0.05, m = 2K \)). Increasing the signal-to-noise ratio by increasing the true–false nulls separation parameter \( \rho \) rather than by decreasing \( \pi_0 \) has a similar effect on FDP-SD’s power (increases, Supplementary Figure 9 bottom row) and its power loss to TDC (decreases, Supplementary Figure 9 middle row), but its effect of TDC’s FDP seems to be reversed (increases, Supplementary Figure 9 top row).

5.3 Multidecoy control of the FDP

We used the above normal mixtures model to study the performance of the multi-decoy version of FDP-SD with \( c = \lambda = 1/2 \) and the mirror map for \( \varphi \) in (7). Supplementary Figure 10 shows that with the same parameter combinations we used above, there is a substantial power gain when increasing the number of decoys from one (red) to three (green). Increasing the number of decoys further from three to seven typically shows a mild to no visible increase in the median power (over 40K drawn sets).

Focusing on one parameter combination, Figure 5 shows the box plots (each made of 40K points) of the power of FDP-SD using one, three, and seven decoys as well as of TDC. Again we observe a substantial increase in power going from one to three decoys and a milder one going from three to seven. Interestingly, we see that using FDP-SD to control the FDP at level \( \alpha = 0.05 \) (and 95\% confidence) with three or seven decoys yields more correct discoveries than using TDC to control the FDR at the same level of \( \alpha \) (Figure 5, left). This is even more pronounced for \( \alpha = 0.1 \) (right).

We also examined how conservative is FDP-SD by comparing the empirical 0.95 quantile of the actual FDP (in each of the 40K reported discovery lists per parameter combination) to the nominal FDP threshold \( \alpha \). Given that we used a 95\% confidence level when applying FDP-SD, the aforementioned 0.95 quantile should not exceed \( \alpha \), and the closer it is to \( \alpha \) the more powerful/less conservative the procedure is. Thus, it is not surprising that Supplementary Figure 11 shows that in the same setup as that of Supplementary Figure 10 using a single decoy (red) is quite conservative. What the former figure did not show is that using three or seven decoys seems to be close to optimal in these settings, as suggested by the small gaps between the green (three) and the blue (seven) lines and the diagonal.
**FIGURE 4** Simulated normal mixtures: power loss relative to FDP-SD and relative to TDC. **Left:** for each of the 108 combinations of calibrated/uncalibrated scores with \( m \in \{500, 2k, 10k\} \), \( \pi_0 \in \{0.2, 0.5, 0.8\} \), \( \alpha \in \{0.01, 0.05, 0.1\} \), and \( \gamma \in \{0.01, 0.05\} \), we noted the median of the loss in power when using FDP-KRB compared with using FDP-SD. The median was taken over 40K samples, and the relative loss is defined as \( 1 - (T'_{\text{FDP-KRB}} + 10^{-12})/(T'_{\text{FDP-SD}} + 10^{-12}) \), where \( T'_{\text{FDP-*}} \) is the number of true discoveries reported by the method. Notably, FDP-SD’s median number of discoveries is never smaller than that of FDP-KRB across all 108 data–parameters combinations. The median of the 108 median power losses of FDP-KRB is 14.0%. **Right:** using the same randomly generated data we noted the median of the loss in power when using FDP-SD (with confidence \( \gamma \in \{0.01, 0.05\} \)) to control the FDP compared with using TDC to control the FDR. The medians of the two sets of 54 median power losses (108 combinations split according to the confidence parameter \( \gamma \)) are 11.2% (\( \gamma = 0.01 \)) and 6.5% (\( \gamma = 0.05 \)).

**FIGURE 5** Simulated experiments: FDP-SD with multiple decoys can outperform TDC. **Left:** using \( m = 10k \), \( \pi_0 = 0.5 \), \( \gamma = 0.05 \), and \( \alpha = 0.05 \), we plot the power of TDC and of FDP-SD with one, three, and seven decoys. Each boxplot consists of the same 40K simulated samples. For each number of decoys, FDP-SD reported a small collection of cases where the power is small (< 0.2), the remaining points being shown in the figure. The number of such cases is written above the x-axis. **Right:** uses the same randomly generated data but with \( \alpha = 0.1 \) instead. Note that only nine out of 40K cases reported a small power for FDP-SD with seven decoys. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.
5.4 Controlling the FDR in shotgun proteomics

Finally, we report some results using real mass spectrometry data but first we need to provide some further context to the problem we study here. Tandem mass spectrometry (MS/MS) currently provides the most efficient means of studying proteins in a high-throughput fashion. As such, MS/MS is the driving technology for much of the rapidly growing field of proteomics—the large-scale study of proteins. Proteins are the primary functional molecules in living cells, and 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.

In a “shotgun proteomics” MS/MS experiment, the proteins that are extracted from a complex biological sample are not measured directly. For technical reasons, the proteins are first digested into shorter chains of amino acids called “peptides.” The peptides are then run through the mass spectrometer, in which distinct peptide sequences generate corresponding spectra. A typical 30-min MS/MS experiment will generate approximately 18,000 such spectra. Canonically, each observed spectrum is generated by a single kind of peptide. Thus, the first goals of the downstream analysis are to identify which peptide generated each of the observed spectra and to determine which peptides and which proteins were present in the sample.

In each of those three problems, the canonical approach to determine the list of discoveries is by controlling the FDR through some form of target–decoy competition. Consider, for example, the problem of peptide detection that we focus on here, where the goal is to infer from the, typically, tens of thousands of generated spectra which of the peptides in an appropriate reference (“target”) database were present in the sample.

The process is typically initiated by scanning each input spectrum against the target database for its best matching peptide. Pioneered by SEQUEST (Eng et al., 1994), the search engine uses an elaborate score function to quantify the quality of the match between each of the database peptides and the observed spectrum, recording the optimal PSM for the given spectrum along with its score (Nesvizhskii, 2010). We then assign the ith database peptide a score $Z_i$ which is the maximal score of all PSMs associated with it (assume for simplicity that a PSM score is $> 0$ so we assign $Z_i = 0$ if no PSM is associated with the ith peptide).

The problem with those PSMs is that 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 (Noble & MacCoss, 2012). Hence, sometimes the 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. This uncertainty carries over to the peptide level: a peptide with a positive score $Z_i > 0$ was not necessarily present in the sample.

Ideally, we would report all the database peptides that were present in the sample but all we have are the scores. We therefore formulate the problem as a multiple hypothesis testing one: $H_1$ corresponds to the assumption that the ith database peptide was not present in the sample. We control the FDR using TDC by competing the ith target score, $Z_i$ (the score of the maximally matching PSM associated with the ith target peptide) with a corresponding decoy score, $Z_i$, generated by using a decoy database. Specifically, each target peptide is paired with an artificial decoy peptide (the analog of the knockoff) that is obtained by randomly shuffling or simply reversing the target peptide. The spectra are then searched against the database of decoy peptides, and a score $Z_i$ is assigned to the ith decoy peptide by taking the maximally scoring PSM associated with that decoy peptide.

Finally, exactly as in the knockoff+ procedure, we then report all top scoring target winning peptides ($Z_i > Z_j$) with the score cutoff determined by TDC. One reason this competition-based approach was adopted, rather than relying on standard methods for control of the FDR such as the procedures by Benjamini and Hochberg (1995) or Storey (2002), is that the latter require sufficiently informative $p$-values and, initially, no such $p$-values were computed in this context (using the decoys we can always assign a “1-bit $p$-value” to the hypotheses but those are not informative enough to obtain effective results using the latter procedures).

We examined, using real data, the performance of FDP-SD in the peptide detection problem. Specifically, we used the same methodology as described in Emery et al. (2020)—recapped here in Supplementary Section 7.4—for detecting peptides in the ISB18 dataset (Klimek et al., 2008). This process generated 900 sets of paired target and decoy scores assigned to each peptide in our database. We then applied TDC and FDP-SD to each of these 900 sets using an FDR/FDP threshold of $\alpha = 5\%$, and confidence level of $100(1 - \gamma) = 95\%$. We relied on the controlled nature of the experiment that generated the ISB18 data to estimate the FDP in each case (Supplementary Section 7.4).

Even though our model is just an approximation of the real peptide detection problem, FDP-SD’s FDP exceeded $\alpha$ in only 36/900, or 4% of the samples, which is less than the allowed error rate of $\gamma = 0.05$. Additionally, Figure 6 shows how the relative power loss associated with using FDP-SD is distributed across the 900 samples (median power loss is 6.7%). For reference, the distribution of TDC’s FDP in this experiment is given in Figure 1 (left).
6 | COMPARISON WITH JANSON AND SU

In this paper, we focused on comparing FDP-SD with FDP-KRB but there is an alternative approach we should discuss: while focusing on \( k \)-FWER control (i.e., no more than \( k \) false discoveries) in the competition-based setup, Janson and Su (2016) suggest three ways of using their \( k \)-FWER control to control the FDP (FXD-control): the second is based on an inversion principle and is computationally impractical, while the third is based on the Romano and Wolf (2007) heuristic rather than rigorously proved. Interestingly, we believe that a simple variation on that heuristic yields FDP-SD, which we propose and rigorously establish the validity of Section 3.2.

This leaves us with their first approach (referred here as FDP-JS) that is based on an augmentation procedure originally due to van der Laan et al. (2004). Briefly, FDP-JS applies their \( k \)-FWER control and then checks if \( (k-1)/R \leq \alpha \), and if so it makes \( r \) more rejections, where \( r \) satisfies \( (k-1+r)/(R+r) \leq \gamma \). While FDP-JS rigorously controls the FDP, there is an obvious practical difficulty when trying to implement it: how does one set \( k \), keeping in mind that control of the FDP is only guaranteed for a single predetermined value of \( k \)?

Regardless of the problem of choosing \( k \), Supplementary Figure 12 shows that in the context of our normal mixtures model FDP-SD often uniformly delivers more power than FDP-JS even when trying to optimize \( k \). For the remaining three of the nine parameter combinations using FDP-JS with the optimal value of \( k \) will deliver more power than FDP-SD but for almost all other values of \( k \) FDP-SD will be significantly more powerful. Finally, Supplementary Figure 13 shows that we can adjust FDP-SD’s \( i_0 \) in those last three cases so that again, it uniformly outperforms FDP-JS. Specifically, we set \( i_0 = \min\{i : \delta_i = \max[0, \nu_k]\} \), where \( \nu_k \) is the parameter \( v \) determined by (3.1) of Janson and Su (2016). It is not hard to see that FDP-SD rigorously controls the FDP for any predetermined value of \( i_0 \) so varying \( i_0 \) is similar to varying \( k \) in FDP-JS.

7 | DISCUSSION

FDP-SD was developed to address the gap between controlling the FDR and the FDP in a competition-based setup. In practice, this difference can be substantial, particularly when the list of discoveries is not very large. Complexitywise, FDP-SD requires sorted data, but beyond that it is linear; hence, its runtime complexity is \( O(m \log m) \).

Our procedure was developed independently of the recent work of Katsevich and Ramdas (2019). The latter provides simultaneous FDP control that allows post hoc analysis that FDP-SD is not suitable for. However, when controlling the FDP at a predetermined level our more focused approach provides a nontrivial advantage. Notably, in the canonical \( p \)-value setup, Goeman et al. (2021) were able to improve upon the Katsevich and Ramdas’ framework by applying a protocol they developed that is based on the closed testing principle. It would be interesting to see if a similar improvement to Katsevich and Ramdas’ framework can be applied in our competition setup, in which case it would also improve FDP-KR.

In the multidecoy version of FDP-SD, we implicitly used a relaxation of Assumption 1 that allows for the true null target-decoy win probability to be different than 1/2. Notably, the same relaxation can be applied even in the single decoy case to allow for a null target win probability of \( c \), and a decoy win probability of \( 1-\lambda \), so that \( c+(1-\lambda) \leq 1 \), and then apply Supplementary Algorithms 5 and 6 as they are, with \( R \) in (8) replacing the 1/2 of FDP-SD.

Finally, we demonstrated that the multidecoy version of FDP-SD with \( c=\lambda=1/2 \) (mirror) can significantly outperform the single decoy version (and even outperform the FDR-controlling TDC). Notably, Emery et al. showed, in the related context of competition-based FDR control, that an optimal choice of \( c \) and \( \lambda \) can potentially increase the power significantly. This begs the question of whether a similar increase can be achieved while still rigorously controlling the FDP. Similarly, optimizing the choice of \( i_0 \), currently defined by (4), can improve power but it is not clear how to do it while maintaining FDP control.
ACKNOWLEDGMENTS

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

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

The knockoff statistics used for our GWAS example were generated by Sesia et al. (2020) and can be downloaded directly at https://msesia.github.io/knockoffzoom/ukbiobank.html, but we downloaded it using the UKBB_utils.R available at https://raw.githubusercontent.com/ekatsevi/simultaneous-fdp/master/UKBB_utils.R.

The ISB18 dataset (Klimek et al., 2008) was downloaded from https://regis-web.systemsbiology.net/Public Datasets.

ORCID

Uri Keich 🄐 https://orcid.org/0000-0002-3209-5011

REFERENCES

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

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57, 289–300.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., et al. (2018) The UK biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209.

Candès, E.J., Fan, Y., Janson, L. & Lv, J. (2018) Panning for gold: model-X knockoffs for high-dimensional controlled variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(3), 551–577.

Cerqueira, F.R., Graber, A., Schwikowski, B. & Baumgartner, C. (2010) 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.

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

Emery, K., Hasam, S., Noble, W.S. & Keich, U. (2020) Multiple competition-based FDR control and its application to peptide detection. In International Conference on Research in Computational Molecular Biology. Berlin: Springer, pp. 54–71.

Eng, J.K., McCormack, A.L. & Yates, III, J.R. (1994) 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.

Fan, Y., Lv, J., Sharifvaghefi, M. & Uematsu, Y. (2018) IPAD: stable interpretable forecasting with knockoffs inference. SSRN 3245137. https://ssrn.com/abstract=3245137

Genovese, C. & Wasserman, L. (2006) Exceedance control of the false discovery proportion. Journal of the American Statistical Association, 101(476), 1408–1417.

Goeman, J.J., Hemerik, J. & Solari, A. (2021) Only closed testing procedures are admissible for controlling false discovery proportions. The Annals of Statistics, 49(2), 1218–1238.

Guo, W. & Romano, J. (2007) A generalized Sidak-Holm procedure and control of generalized error rates under independence. Statistical Applications in Genetics and Molecular Biology, 6(1), 3.

He, K., Fu, Y., Zeng, W.F., Luo, L., Chi, H., Liu, C., et al. (2015) A theoretical foundation of the target-decoy search strategy for false discovery rate control in proteomics. arXiv [Preprint]. Available from: https://arxiv.org/abs/1501.00537 [Accessed: 1st July, 2019].

Janson, L. & Su, W. (2016) Familywise error rate control via knockoffs. Electronic Journal of Statistics, 10(1), 960–975.

Katsevich, E. & Ramdas, A. (2019) Simultaneous high-probability bounds on the false discovery proportion in structured, regression, and online settings. arXiv [Preprint]. Available from: https://doi.org/10.48550/arXiv.1803.06790 [Accessed: 10th January, 2021].

Klimek, J., Eddes, J.S., Hohmann, L., Jackson, J., Peterson, A., Letarte, S., et al. (2008) The standard protein mix database: a diverse dataset to assist in the production of improved peptide and protein identification software tools. Journal of Proteome Research, 7(1), 96–1003.

Lehmann, E.L. & Romano, J.P. (2005) Generalizations of the familywise error rate. Annals of Statistics, 33(3), 1138–1154.

Lu, Y.Y., Fan, Y., Lv, J. & Noble, W.S. (2018) DeepPINK: reproducible feature selection in deep neural networks. In: NIPS18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates, pp. 8690–8700.

Nesvizhskii, A.I. (2010) 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.

Noble, W.S. & MacCoss, M.J. (2012) Computational and statistical analysis of protein mass spectrometry data. PLOS Computational Biology, 8(1), e1002296.

Romano, J.P. & Wolf, M. (2007) Control of generalized error rates in multiple testing. Annals of Statistics, 35(4), 1378–1408.

Sesia, M., Katsevich, E., Bates, S., Candès, E. & Sabatti, C. (2020) Multi-resolution localization of causal variants across the genome. Nature Communications, 11(1), 1093.

Storey, J.D. (2002) A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B, 64, 479–498.

Sudarshan, M., Tansey, W. & Ranganath, R. (2020) Deep direct likelihood knockoffs. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F. & Lin, H. (Eds), Advances in neural information processing systems. Red Hook, NY: Curran Associates, Inc., vol. 33, pp. 5036–5046.

Th, M., Tasnim, A. & Kill, L. (2016) How to talk about protein-level false discovery rates in shotgun proteomics. Proteomics, 16(18), 2461–2469.

Tibshirani, R.J. (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B, 58(1), 267–288.

van der Laan, M.J., Dudoit, S. & Pollard, K.S. (2004) Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. Statistical Applications in Genetics and Molecular Biology, 3(1), 15.

Xiao, Y., Angulo, M.T., Friedman, J., Waldor, M.K., Weiss, S.T. & Liu, Y.Y. (2017) Mapping the ecological networks of microbial communities. Nature Communications, 8(1), 2042.
SUPPORTING INFORMATION
Web Appendices, Tables, and Figures referenced in Sections 1-7 are available with this paper at the Biometrics website on Wiley Online Library.
An R-package implementing FDP-SD is available for download from the same Biometrics website or from CRAN https://cran.r-project.org/web/packages/stepdownfdp/index.html.

How to cite this article: Luo, D., Ebadi, A., Emery, K., He, Y., Noble, W.S. & Keich, U. (2023) Competition-based control of the false discovery proportion. Biometrics, 79, 3472–3484. https://doi.org/10.1111/biom.13830