False discovery rate-controlled multiple testing for union null hypotheses: a knockoff-based approach

Ran Dai | Cheng Zheng

Department of Biostatistics, University of Nebraska Medical Center, Omaha, Nebraska, USA

Correspondence
Cheng Zheng, Department of Biostatistics, University of Nebraska Medical Center, Omaha, Nebraska, USA. Email: cheng.zheng@unmc.edu

Funding information
National Cancer Institute, Grant/Award Number: CA119171; National Institute of General Medical Sciences, Grant/Award Number: GM115458

Abstract
False discovery rate (FDR) controlling procedures provide important statistical guarantees for replicability in signal identification based on multiple hypotheses testing. In many fields of study, FDR controlling procedures are used in high-dimensional (HD) analyses to discover features that are truly associated with the outcome. In some recent applications, data on the same set of candidate features are independently collected in multiple different studies. For example, gene expression data are collected at different facilities and with different cohorts, to identify the genetic biomarkers of multiple types of cancers. These studies provide us with opportunities to identify signals by considering information from different sources (with potential heterogeneity) jointly. This paper is about how to provide FDR control guarantees for the tests of union null hypotheses of conditional independence. We present a knockoff-based variable selection method (Simultaneous knockoffs) to identify mutual signals from multiple independent datasets, providing exact FDR control guarantees under finite sample settings. This method can work with very general model settings and test statistics. We demonstrate the performance of this method with extensive numerical studies and two real-data examples.

KEYWORDS
FDR control, heterogeneity, replicability, reproducibility, variable selection

1 INTRODUCTION

There is a pressing need in making discoveries by analyzing information from multiple sources jointly. With recent advances in scientific research, data on the same set of candidate features are often collected independently from multiple sources. For example, social scientists collect data on the economic and socioeconomic status of people from different community groups. In genome-wide association studies (GWAS), associations of genome features with multiple different outcomes of interest are studied in multiple experiments (Uffelmann et al., 2021). These data motivate us to identify mutual signals from multiple experiments for purposes like reproducibility research or mediator identification. This paper focuses on how to identify mutual signals from multiple independent studies and provide variable selection accuracy guarantees with mild design and model assumptions.

Now, we formulate the mutual signal identification problem in statistical terms. Assume we have data from \( K \) independent experiments and denote \( [K] = \{1, \ldots, K\} \). Within the \( k \)th experiment, \((Y_i^k, X_{i1}^k, \ldots, X_{ip}^k) \sim D_k, i = 1, \ldots, n_k\)
1, ..., n_k. In our setting, the outcome variables Y_1, ..., Y^K can be of different data types and (X^{k_1}_1, ..., X^{k_2}_i) can have different distributions among the different experiments. We denote Y^k's and X^k's as continuous variables throughout the paper for the simplicity of notation. In practice, they can be of other data types (continuous/count/nominal/ordinal/mixed). For example, Y^k's can be different disease outcomes and X^k's can be gene expression data measured on different scales. Define H_{0j}^k as the null hypothesis indicating the jth feature not being a signal in the kth experiment (i.e., X^k_j \perp Y^k | X_{-j}^k), where X_j := \{X^k_1, ..., X^k_p \setminus X^k_j\}, and denote H^k = \{j \in [p] : H_{0j}^k \text{ is true}\}, where [p] := \{1, ..., p\}. Instead of testing the H_{0j}^k's, we are interested in testing the union null hypotheses

\[ H_{0j} = \cup_{k=1}^{K} H_{0j}^k, \text{ for } j \in [p]. \tag{1} \]

We define S = \{j \in [p] : H_{0j} is false\} and H = S^c = \cup_{k=1}^{K} H^k = \{j \in [p] : H_{0j} is true\}. \tag{2}

We aim at developing a selection procedure returning a selection set \( \hat{S} \subseteq [p] \) with a controlled false discovery rate (FDR), which is the expected false discovery proportion (FDP):

\[ \text{FDR}(\hat{S}) = \mathbb{E}[\text{FDP}(\hat{S})] = \mathbb{E}\left[\frac{|\hat{S} \cap H|}{|\hat{S}| + 1}\right]. \tag{3} \]

To begin, we give some examples to motivate our method.

1.1 Examples

The problem of testing multiple union null hypotheses is related to many important scientific areas, for example, the reproducibility analysis in GWAS (Bogomolov & Heller, 2013; Heller et al., 2014; Heller & Yekutieli, 2014), comparative research in genomics studies (Rittschof et al., 2014), and mediation analysis (Sampson et al., 2018). Below, we give several motivating examples that can be considered as problems of identifying the mutual signal set \( \hat{S} \).

1.1.1 Repeatability research

In some fields of biology, experimental results are required to agree with each other under conditions that include the same measurement procedure, same operators, same measuring system, same operating conditions, same location, and replicate measurements on the same or similar objects (Plant & Hanisch, 2020; Ioannidis et al., 2009). It aims at identifying signals in repeated experiments. Mathematically, K independent datasets \((Y^1, X^1), ..., (Y^K, X^K)\) are collected, where \(Y^k \in \mathbb{R}^{n_k}\) and \(X^k \in \mathbb{R}^{n_k \times p}\) for \(k \in [K]\), and \((Y^1, X^{11}_{i_1}, ..., X^{1p}_{i_p}) \overset{iid}{\sim} D_k\), for \(i \in [n_k]\), for each \(k \in [K]\), and we test the conditional independence \(Y^k \perp X_{-j}^k | X_{j}^k\), for \(k \in [K]\) and \(j \in [p]\). The jth feature is a mutual signal if and only if the union null hypothesis \(H_{0j}\) does not hold.

As a remark, Heller et al. (2014) proposed a repFDR method, which also provides the FDR control guarantees on testing multiple union null hypotheses. This method is based on the Benjamini–Hochberg (BHq) procedure (Benjamini & Hochberg, 1995); and it assumes that the vector of test statistics for hypotheses in each study are jointly independent or are positive regression dependent (PRDS) on the subset of true null hypotheses. This assumption does not hold in general in our settings. There is a modification of this method that allows for an arbitrary dependence, however, it is known to be very conservative (Benjamini & Yekutieli, 2001).

1.1.3 High-dimensional mediator selection

In many scientific fields, it is important to identify features that are associated with multiple responses. In
particular, mediators can be discovered from simultaneous feature-treatment and feature-outcome associations. For example, suppose we aim at identifying gene expression mediators that are both associated with the treatment and the risk of a certain disease. To do this, we jointly use information from two independent studies, one on the associations between the gene expressions and the treatment, the other on the association between the gene expressions and the risk of the disease with the treatment being fixed. The selection of mediators from high-dimensional gene expression features can be framed as a problem of testing the union null hypotheses with $K = 2$. In particular, $(Y^k_i, X^k_{i1}, \ldots, X^k_{ip}) \overset{\text{iid}}{\sim} D_k$, for $k = 1, 2$, where $Y^1$ and $Y^2$ are the treatment and the outcome, respectively. Note that in this example the true signal sets for $Y^1$ and $Y^2$ are not necessarily identical. We test the conditional independence $Y^k \perp X^k_j | X^k_{-j}$, for $k = 1, 2$, and $j \in [p]$. The $j$th feature is a mediator if and only if the union null hypothesis $H_{0j}$ does not hold.

### 1.2 Prior work

#### 1.2.1 Current advance in FDR control for identifying simultaneous signals

For reproducibility research, Bogomolov et al. proposed methods based on the BH procedure by selecting features that are commonly selected among all the experiments (Heller et al., 2014; Bogomolov & Heller, 2013, 2018). There are multiple works based on computing the local FDR as the optimal scalar summary of the multivariate test statistics (Chi, 2008; Heller & Yekutieli, 2014). Recently, Xiang et al. (2019) presented the signal classification problem for multiple sequences of multiple tests, where the identification of the simultaneous signal is a special case, and Zhao and Nguyen (2020) proposed a nonparametric method for asymptotic FDR control in identifying simultaneous signals. However, all methods above assume not only the independence of the experiments but also the independence (or PRDS) of the $p$-values for the features within each experiment, which is not realistic in many complex high-dimensional data applications, such as the GWAS and other omics data.

#### 1.2.2 Knockoff-based methods

For multiple testing problems within a single experiment, there are recent advances in relaxing the assumption of independence among the features. Powerful knockoff-based methods have been developed for exact FDR control in selecting features with conditional associations with the response (Barber & Candés, 2015; Candés et al., 2018). The original knockoff filter proposed by Barber and Candés (2015) works on linear models assuming no knowledge of the design of covariates, the signal amplitude, or the noise level. It achieves exact FDR control under finite sample settings. It is also extended to work with high-dimensional settings (Barber & Candés, 2019). Later Candés et al. (2018) proposed the Model-X knockoff method, extending the knockoff filter to achieve exact FDR control for nonlinear models. This method allows the conditional distribution of the response to be arbitrary and completely unknown but requires the distribution of $X$ to be known. Barber et al. (2020) further showed that the Model-X knockoff method is robust against errors in the estimation of the distribution of $X$. In addition, Huang and Janson (2020) relaxed the assumptions of the Model-X knockoff method so that the FDR can be controlled as long as the distribution of $X$ is known up to a parametric model. There are also abundant publications on the construction of knockoffs with an approximated distribution of $X$. Liu and Zheng (2019) developed a Deep knockoff machine using deep generative models. More recently, Bates et al. (2021) proposed an efficient general metropolized knockoff sampler. Spector and Janson (2022) proposed to construct knockoffs by minimizing the reconstructability of the features. Knockoff construction for non-continuous data poses its own challenges. Kormaksson et al. (2021) proposed the sequential knockoffs for $X$ containing both continuous and categorical variables. Knockoff-based methods have also been extended to test the intersection of null hypotheses. In this direction, group and multitask knockoff methods (Dai & Barber, 2016), and prototype group knockoff methods (Chen et al., 2019) have been proposed. Variants of knockoff methods have become useful tools in scientific research. For example, to identify the variations across the whole genome associated with a disease, Sesia et al. (2018) developed a hidden Markov model knockoff method for FDR control in GWAS.

### 1.3 Our contributions

In this paper, we propose a knockoff-based procedure to establish exact FDR control in selecting mutual signals from multiple conditional independence tests, assuming very general conditional models. The main contributions of this paper are summarized below:

1. We construct a knockoff-based procedure for testing the union null hypotheses for feature selection, namely the Simultaneous knockoffs. This procedure can work on
general conditional dependence models $Y|X$ and data structures in $X$.

2. We prove that the Simultaneous knockoff method can lead to exact FDR control in testing multiple union null hypotheses for feature selection under finite sample settings.

3. We show that a broad class of filter statistics can be used for this method, and give general recipes for generating different powerful statistics.

4. We demonstrate the FDR control property and the power of our method with extensive simulation settings. We also illustrate the application with two real-data examples.

The rest of the paper is organized as follows. In Section 2, we present the Simultaneous knockoff framework. In Section 3, we give the theoretical guarantees for exact FDR control of the Simultaneous knockoff method in finite sample settings and the robustness result for the potential misspecification of the distribution of $X$. In Section 4, we show the empirical performance of the Simultaneous knockoff method under different model assumptions and data structures. Finally, in Section 5, we apply the Simultaneous knockoff procedure to two real-data examples.

2 | METHODS

In this section, we present the Simultaneous knockoff procedure. This procedure can be paired with both the Fixed-X knockoffs (Barber & Candés, 2015) and the Model-X knockoffs (Candés et al., 2018) to allow for very general model settings and various data structures in real-data applications. Before presenting the Simultaneous knockoff method, we briefly review the Fixed-X and the Model-X knockoff methods.

2.1 | The Fixed-X and the Model-X knockoff procedures

The high-level idea behind the knockoff methods is to construct a “knockoff” copy of the covariates, retaining their inner structures. Unlike the “true” covariates, the knockoff copies are created independent of the response. These knockoff variables are then mixed into the model to monitor the FDP during the selection. Heuristically speaking, if one variable is a true signal, it is more likely to be selected than its knockoff copy, otherwise, it is equally likely to be selected as its knockoff copy. Therefore, by counting the number of knockoff variables entering the selected set, the FDP can be (over) estimated.

2.1.1 | Model settings

For the Fixed-X knockoff method, the setup is a decentralized linear model, $Y = X\beta + \varepsilon$, where $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ and $\varepsilon \in \mathbb{R}^n \sim \mathcal{N}(0, \sigma^2 I_n)$. This method has weak assumptions on the covariates $X$, the amplitudes of the unknown regression coefficients $\beta$, and does not require the noise level ($\sigma^2$) to be known. The Model-X knockoff method works on more general conditional model settings. It does not require the dependence of $Y|X$ to be known by assuming the knowledge of the distribution of $X$ (or if the distribution of $X$ can be well approximated). Therefore, it can work with many more models such as generalized linear models (GLMs) or nonlinear models.

2.1.2 | Algorithm

There are four main steps in the knockoff procedure listed as below.

- **Knockoff construction.** A set of knockoff features $\tilde{X} = [\tilde{X}_1 \ldots \tilde{X}_p]$ are constructed in this step. For the Fixed-X knockoff construction, $\tilde{X}$ needs to satisfy that, for some vector $s \geq 0$,

$$\tilde{X}^\top \tilde{X} = X^\top X, \quad \tilde{X}^\top X = X^\top X - \text{diag}(s).$$

(4)

For the Model-X knockoff construction, $\tilde{X}$ needs to satisfy the pairwise exchangeability condition:

$$[X \tilde{X}]_{\text{Swap}(j)} \overset{d}{=} [X \tilde{X}] \quad \text{and} \quad \tilde{X} \perp Y|X \quad \text{for all} \ j \in [p],$$

(5)

where $\text{Swap}(j)$ stands for exchanging the $j$th column and the $(j + p)$th column of $[X \tilde{X}]$, and $A \overset{d}{=} B$ indicates $A$ and $B$ are identical in distribution. $\tilde{X}$ can be generated using various algorithms (Barber & Candés, 2015; Romano et al., 2020; Liu & Zheng, 2019; Bates et al., 2021; Spector & Janson, 2022). More knockoff construction details are reviewed in Web Appendix A.1.

- **Test statistics calculation.** Appropriate test statistics need to be calculated for the features $X$ and their knockoff copies $\tilde{X}$. For the Fixed-X knockoffs, the statistics $[Z, \tilde{Z}] \in \mathbb{R}^{2p}$ needs to be a function of $(X\tilde{X})^\top [XX], [XX]^\top Y)$. For the Model-X knockoffs, $[Z, \tilde{Z}]$ needs to be a function of $(X \tilde{X}, Y)$, such that if we swap features $X_j$, with its corresponding knockoffs $\tilde{X}_j$, then the statistics $Z_j$ and $\tilde{Z}_j$ get swapped. Examples of the $[Z, \tilde{Z}]$ statistics are provided in Web Appendix A.2.
• **Filter statistics calculation.** We construct the filter statistics \( \mathbf{W} \in \mathbb{R}^p \) such that \( W_j = f(Z_j, \tilde{Z}_j) \), where \( f \) is an antisymmetric function, that is, \( f(x, y) = -f(y, x) \). Without loss of generality, we further let \( f(x, y) > 0 \) when \( x > y \). If \( X_j \) is a signal, we would expect \( \mathbb{P}[Z_j > \tilde{Z}_j] > 0.5 \), while if \( X_j \) is not a signal, we would expect \( Z_j \) and \( \tilde{Z}_j \) to have the same distribution. Thus, we expect \( W_j \) to have a positive sign with \( > 0.5 \) probability if \( X_j \) is a signal and with \( 0.5 \) probability if \( X_j \) is not a signal. This allows us to estimate the FDP in \( \hat{S}(t) := \{j : W_j \geq t\} \) as

\[
\hat{\text{FDP}}(t) = \frac{\# W_j \leq -t}{\# W_j \geq t} \lor 1.
\]  

(6)

• **Threshold calculation and feature selection.** With the knockoff filter, we select

\[ S = \{j : W_j \geq t\}, \text{ where } t = \min \left\{ \min \left\{ \frac{\# \{j : W_j \leq -t\}}{\# \{j : W_j \geq t\}} \lor 1, q \right\} : q \in W_+ \right\}. \]  

(7)

With a more conservative knockoff+ filter, we select

\[ S_+ = \{j : W_j \geq \tau_+\}, \text{ where } \tau_+ = \min \left\{ \min \left\{ \frac{1 + \# \{j : W_j \leq -t\}}{\# \{j : W_j \geq t\}} \lor 1, q \right\} : q \in W_+ \right\}. \]  

(8)

Here, \( q \) is the target FDR level and \( W_+ = \{ |W_j| : |W_j| > 0 \} \).

### 2.2 Simultaneous knockoff framework

In this section, we propose the general **simultaneous knockoff framework**, which enables us to use the knockoff approach for FDR control in testing the union null hypotheses of conditional independence. This approach enjoys very general model assumptions and exact FDR control guarantees in finite sample settings.

#### 2.2.1 Preliminaries

One naïve idea to identify mutual signals in \( K \) experiments is to select the intersection set of the variables selected from the individual experiments. However, this method cannot control the FDR (see more details in Section 4). Therefore, we alternatively aim at constructing valid filter statistics \( \mathbf{W} \) to allow the estimation of the FDP in our multiple testing of the union null hypotheses using Equation (6). We establish a general recipe to construct such \( \mathbf{W} \)s with only the summary statistics that can be calculated using the Fixed-X and the Model-X knockoff methods from single experiments. To begin, we give several definitions.

**Definition 1.** (Swapping) For a set \( S \subseteq [p] \), and for a vector \( \mathbf{V} = (V_1, \ldots, V_p) \in \mathbb{R}^{2p} \), \( V_{\text{Swap}(S)} \) indicates the swapping of \( V_j \) with \( V_{j+p} \) for all \( j \in S \).

**Definition 2.** (Flip sign function) A function \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) is called a flip sign function if it satisfies that for all \( S \subseteq [p] \), \( f([Z, \tilde{Z}]_{\text{Swap}(S)}) = f([Z, \tilde{Z}]) \circ \epsilon(S) \) where \( Z, \tilde{Z}, \epsilon(S) \in \mathbb{R}^p \), and \( \epsilon(S) = -1 \) for all \( j \in S \) and \( \epsilon(S) = 1 \) otherwise. Here, \( \circ \) denotes the Hadamard product.

An example of a flip sign function is \( f([Z, \tilde{Z}]) = Z - \tilde{Z} \)

(9)

More examples of flip sign functions and their relationships to antisymmetric functions are discussed in Web Appendix A.4.

**Definition 3.** (One swap combining function (OSCF)). A function \( f : \mathbb{R}^{2pK} \rightarrow \mathbb{R}^p \) is called an OSCF if it satisfies that for all \( k \in [K] \) and all \( S \subseteq [p] \),

\[ f([Z^k, \tilde{Z}^k]_{\text{Swap}(S)}) = \cdots = f([Z^k, \tilde{Z}^k]_{\text{Swap}(S)}) \]  

where \( Z^k, \tilde{Z}^k \in \mathbb{R}^p \) for all \( k \in [K] \).

As a remark, the definition of an OSCF implicitly requires that for any set \( S \subseteq [p] \),

\[ f([Z^1, \tilde{Z}^1]_{\text{Swap}(S)}), \ldots, [Z^k, \tilde{Z}^k]_{\text{Swap}(S)}) \]  

An example of the OSCF can be defined as below: let \( \mathbf{a} = (a_1, \ldots, a_K) \in \mathbf{A} \) where \( \mathbf{A} = \{0, 1\}^K \). We separate \( \mathbf{A} \) to two sets: the even set \( A_e = \{ \mathbf{a} : \text{mod}([|\mathbf{a}|_1], 2) = \text{mod}(K, 2) \} \); and the odd set \( A_o = \{ \mathbf{a} : \text{mod}([|\mathbf{a}|_1], 2) = \text{mod}(K + 1, 2) \} \). Then, we obtain an OSCF function \( [Z, \tilde{Z}] = f([Z^1, \tilde{Z}^1], \ldots, [Z^K, \tilde{Z}^K]) \) as below:

\[
Z_j = \sum_{\mathbf{a} \in A_o} \prod_{k=1}^K Z_{jk}^{a_k} \tilde{Z}_{jk}^{1-a_k} \quad \text{and} \quad \tilde{Z}_j = \sum_{\mathbf{a} \in A_e} \prod_{k=1}^K Z_{jk}^{a_k} \tilde{Z}_{jk}^{1-a_k},
\]

(10)

where \( Z_{jk} \) and \( \tilde{Z}_{jk} \) are the \( j \)th entry of \( Z^k \) and \( \tilde{Z}^k \), respectively.

In particular, when \( K = 2 \), this construction can be written as:

\[ Z_j = Z_1 Z_{j1} + Z_2 Z_{j2} \quad \text{and} \quad \tilde{Z}_j = \tilde{Z}_1 Z_{j1} + \tilde{Z}_2 Z_{j2} \]  

for \( j = 1, \ldots, p \).

More OSCF examples are given in Web Appendix A.3.

**Definition 4.** (One swap flip sign function (OSFF)) A function \( f : \mathbb{R}^{2pK} \rightarrow \mathbb{R}^p \) is called an OSFF if it satisfies
that for all $k \in [K]$ and all $S \subset [p]$,

$$f([Z^1, \bar{Z}^1], \ldots, [Z^k, \bar{Z}^k]) = f([Z^1, \bar{Z}^1], \ldots, [Z^k, \bar{Z}^k]) \cap \epsilon(S),$$

where $Z^k, \bar{Z}^k \in \mathbb{R}^p$ for $k \in [K]$.

There are multiple ways to construct OSFFs. As shown in Lemma A1 in Web Appendix A.5, if $f_1 : \mathbb{R}^{2pK} \to \mathbb{R}^p$ is an OSCF and $f_2 : \mathbb{R}^p \to \mathbb{R}^p$ is a flip-sign function, then $f = f_2 \circ f_1$ is an OSFF, where $\circ$ denotes the composition of functions. When using the OSCF as defined in Equation (10) and using the flip-sign function as defined in Equation (9), we obtain an OSFF $f([Z^1, \bar{Z}^1], \ldots, [Z^k, \bar{Z}^k]) = \bigotimes_{k=1}^K (Z^k - \bar{Z}^k)$. Alternative ways to construct OSFFs and more examples are provided in Web Appendix A.5.

### 2.2.2 Algorithm

The **Simultaneous knockoff** procedure is described below:

- **Step 1: Knockoff construction for the individual experiments.** Denote the knockoff matrices for $X^1, \ldots, X^k$ as $\tilde{X}^1, \ldots, \tilde{X}^k$. For each $k \in [K]$, select a knockoff construction method (either Fixed or Model-X) as described in Web Appendix A.1 that is compatible with the model setting for the experiment $k$ to generate $\tilde{X}^k$.

- **Step 2: Test statistics calculation for the individual experiments.** For each experiment $k \in [K]$, choose and calculate statistics $[Z^k, \bar{Z}^k] \in \mathbb{R}^{2p}$ that are compatible with the knockoff construction method for experiment $k$. Details on the choices of $[Z^k, \bar{Z}^k]$ can be found in Web Appendix A.2.

- **Step 3: Calculation of the filter statistics $W$.** Choose an arbitrary OSFF $f$ as defined in Definition 4 and calculate $W = f([Z^1, \bar{Z}^1], \ldots, [Z^k, \bar{Z}^k])$. Examples of OSFFs can be found in Web Appendix A.5.

- **Step 4: Threshold calculation and feature selection.** Using the filter statistics $W$ from Step 3, we apply the knockoff+ filter (8) to obtain the selection set $\tilde{S}_+$ under the **Simultaneous knockoff+** procedure; or apply the knockoff filter (7) to obtain $\tilde{S}$ under the **Simultaneous knockoff** procedure.

### 3 THEORETICAL RESULTS

The main result for this paper is the theoretical guarantee that **Simultaneous knockoff** and **Simultaneous knockoff+** procedures can control the modified FDR (as defined in Equation (11) in Theorem 1) and FDR, respectively.

**Theorem 1.** With the individual experiments satisfying the Fixed-X or the Model-X knockoff model settings, the Simultaneous knockoff procedure (7) controls the modified FDR defined as

$$mFDR = \mathbb{E}\left[ \frac{|\tilde{S} \cap H|}{|\tilde{S}| + 1/q} \right] \leq q,$$

and the Simultaneous knockoff+ procedure (8) controls the usual FDR

$$\mathbb{E}\left[ \frac{|\tilde{S} \cap H|}{|\tilde{S}| + 1} \right] \leq q,$$

where $H$ is the union null set as defined in Equation (2).

The Fixed-X and the Model-X knockoff model settings can be found in Web Appendix A.1. The definition of $mFDR$ is close to the FDR, especially when the selection set is relatively large. Although the more conservative Simultaneous knockoff+ procedure can achieve exact FDR control, in real-data applications, the knockoff filter is more widely used (Barber & Candès, 2015; Dai & Barber, 2016; Candès et al., 2018; Sesia et al., 2018; Romano et al., 2020).

The key step for the proof of Theorem 1 is to show that the signs of the $W_j$ for the union nulls are i.i.d. following a Bernoulli($\frac{1}{2}$) distribution, and independent of $|W_j|$ for all $j \in H$. As for the knockoff-based methods, this property effectively guarantees that for all $j \in H$, there are equal probabilities of selecting the feature and its knockoff copy, which allows the knockoff procedure to (over) estimate the FDP. We show this in Lemma 1. The details of the proof can be found in Web Appendix B.

**Lemma 1.** Let $W = f([Z^1, \bar{Z}^1], \ldots, [Z^k, \bar{Z}^k])$ where $f$ is an OSFF. Let $\varepsilon \in \{\pm 1\}^p$ be a sign sequence independent of $W$, with $\varepsilon_j = +1$ for all $j \in S$ and $\varepsilon_j \sim \{\pm 1\}$ for all $j \in H$. Then,

$$(W_1, \ldots, W_p) \overset{d}{=} (W_1 \cdot \varepsilon_1, \ldots, W_p \cdot \varepsilon_p).$$

For models beyond the linear models, we need to use the Model-X knockoffs in the individual experiments. In real applications, the distribution of the candidate features might not be known exactly. In Candès et al. (2018), the robustness against the misspecification of the $X$ distribution is shown empirically. Barber et al. (2020) and Huang and Janson (2020) further addressed this question theoretically. For the Simultaneous knockoff procedure, it is also very important to establish the robustness results against the misspecification of the distribution of $X$. The following theorem shows the result.

**Theorem 2.** Under the definitions in Section 2, for any $\varepsilon \geq 0$, consider the null variables for which $\min_{k, j \in H} \tilde{K}_{I_j} \leq \varepsilon_k$. Then,
\( \epsilon \), where \( \tilde{K}_{lj} = \sum_{i=1}^{n_k} \log \left( \frac{P_j(X_{lj}^k | Z_{lj}^k)Q_k(X_{lj}^k | Z_{lj}^k)}{Q_k(X_{lj}^k | Z_{lj}^k)P_j(X_{lj}^k | Z_{lj}^k)} \right) \) where \( P \) denotes the true distribution and \( Q \) denotes the misspecified distribution. If we use the knockoff filter, then the fraction of the rejections corresponding to such nulls obeys
\[
E \left[ \frac{|\{j : j \in S \cap H and \min_{k : j \in H_k} \tilde{K}_{lj}^k \leq \epsilon\}|}{|S| + 1} \right] \leq q \cdot \epsilon^{1/2}.
\]
(12)

In particular, this implies that the FDR is bounded as
\[
\text{FDR} \leq \min_{\epsilon \geq 0} \left\{ q \cdot \epsilon^{1/2} + \mathbb{P} \left( \max_{j \in H_0} \min_{k : j \in H_k} \tilde{K}_{lj}^k > \epsilon \right) \right\}. \tag{13}
\]

Similarly, if we use the knockoff filter, for any \( \epsilon \geq 0 \), a slightly modified fraction of the rejections corresponding to nulls with \( \min_{k : j \in H_k} \tilde{K}_{lj}^k \leq \epsilon \) obeys
\[
E \left[ \frac{|\{j : j \in S \cap H and \min_{k : j \in H_k} \tilde{K}_{lj}^k \leq \epsilon\}|}{|S| + q^{-1}} \right] \leq q \cdot \epsilon^{1/2}
\]
and from this, we obtain a bound on the modified FDR:
\[
E \left[ \frac{|S \cap H|}{|S| + q^{-1}} \right] \leq \min_{\epsilon \geq 0} \left\{ q \cdot \epsilon^{1/2} + \mathbb{P} \left( \max_{j \in H_k} \min_{k : j \in H_k} \tilde{K}_{lj}^k > \epsilon \right) \right\}. \tag{15}
\]

In real applications, when we have additional samples of \( \mathbf{X} \) (for estimating the distribution of \( \mathbf{X} \)), we will be able to achieve a small enough \( \epsilon \). Otherwise, it has been proposed to evaluate the potential inflation of FDR using simulation (Romano et al., 2020). In Theorem 2, we show an FDR upper bound result for \textit{Simultaneous knockoffs} which is similar to the result in Barber et al. (2020) for Model-X knockoffs, to build some statistical foundations for such simulation approach. Below we give an example to demonstrate its application.

Consider the example of the Gaussian knockoffs in Barber et al. (2020), that is, \( \mathbf{X}^k \) is normally distributed with the mean zero and the variance–covariance matrix \((\Theta^k)^{-1}\) and we use the Gaussian knockoff construction method, that is, sample \( \tilde{X}_j^k | \mathbf{X}^k \sim \mathcal{N}(0, 1 - \mathbf{D}^k \tilde{\Theta}^k \mathbf{D}^k, 2 \mathbf{D}^k - \mathbf{D}^k \tilde{\Theta}^k \mathbf{D}^k) \) where \( \tilde{\Theta}^k \) is an estimated version of \( \Theta^k \) and \( \mathbf{D}^k \) is a nonnegative diagonal matrix such that \( 2 \mathbf{D}^k - \mathbf{D}^k \tilde{\Theta}^k \mathbf{D}^k \) is positive definite, then as shown in Barber et al. (2020), we have with probability at least \( 1 - p^{-1} \),
\[
\max_{j=1, \ldots, p} \tilde{K}_{lj}^k \leq 4 \delta_{\Theta^k} \sqrt{n_k \log(p)} (1 + \text{Rem}), \quad \text{where } \delta_{\Theta^k} = \max_{j=1, \ldots, p} (\Theta_{jj})^{-1/2} ||\Theta^{-1/2} (\Theta - \tilde{\Theta})||_2, \quad \text{and } \text{Rem} \text{ is a vanishing term when } n_k \log(p) \to 0.
\]

We denote the sample size for a larger unlabeled dataset that can be used to estimate the distribution of \( \mathbf{X}^k \). The graphical Lasso estimator of \( \Theta^k \) satisfies that \( ||\hat{\Theta}^k - \Theta^k||_{\infty} \leq \sqrt{\log(p)} \) and therefore \( \delta_{\Theta^k} \leq O(\sqrt{n_k \log(p) / N_k}) \), where \( \Theta^k \) denotes the maximum of the columnwise sparsity level of the matrix \( \Theta^k \), that is, \( \delta_{\Theta^k} = \max_{j=1, \ldots, p} \log(\rho) \). So \( 4 \delta_{\Theta^k} \sqrt{n_k \log(p)^2} \) will be small if the unlabeled sample size \( n_k \) for each subsample is large enough in the sense that \( n_k \gg n_k \log \Theta^k \). Under a special setting where there exists a subset \( \Omega \subset [K] \) such that \( H = \cup_{k \in \Omega} H_k \), then we will just need enough unlabeled sample size within those subsamples with index from \( \Omega \).

Our theoretical guarantees focus on the control of FDR. The power is a monotonically decreasing function of \( K \) and a monotonically increasing function of \( n \). Asymptotically, as \( K \) is fixed, and \( \log \frac{p}{n} \to 0 \), the power converges to 1 as \( n \to \infty \) (see details in Web Appendix C.3.5).

Since there are no theoretical results on the choice of \( \mathbf{W} \) for the most powerful test, we compare the power with several choices of \( \mathbf{W} \)s numerically. To understand the power of the proposed statistics, we plot the empirical distributions of the filter statistics \( \hat{W}_j \), for \( j \in H \) and \( j \in S \), respectively, assuming \( Z_j \xrightarrow{iid} \mathcal{N}(0, 1) \) for \( j \in H^k \) and \( Z_j \xrightarrow{iid} \mathcal{N}(3, 1) \) otherwise (Figure 1). We can see that for \( j \in H, \) the filter statistics \( \hat{W}_j \) is symmetric around 0, whereas for \( j \in S, \mathbb{P}(W_j > 0) > 1/2 \).

4 | SIMULATION

4.1 | Simulation settings

We first consider the \( K = 2 \) case with three data settings:

1. \textit{Continuous.} For both experiments, \( Y_i^k \)'s are continuous and \( Y_i^k | X_i^k \)'s follow linear models.
2. \textit{Binary.} For both experiments, \( Y_i^k \)'s are binary and \( Y_i^k | X_i^k \)'s follow logistic models.
3. \textit{Mixed.} \( Y_i^1 \) is continuous and \( Y_i^1 | X_i^1 \) follows a linear model; \( Y_i^2 \) is binary and \( Y_i^2 | X_i^2 \) follows a probit model.

We compare our proposed method \textit{(simultaneous)} with the two alternative methods below:

- \textit{Pooling.} Data are pooled together and tests of the conditional associations are performed using the knockoff methods for a single experiment.
- \textit{Intersection.} Knockoff methods for single experiments are used for the individual experiments and the intersection of the selected sets is used to construct the selection set of the mutual signals.
FIGURE 1  Distributions of the filter statistics $W_j = Z_j - \bar{Z}_j$, where $Z_j$ and $\bar{Z}_j$ are as defined in Equation (10) with $K = 3$ for the cases where feature $j$ is not a signal in any of the experiments (null), $j$ is a signal in one experiment (null), two experiments (null), and three experiments (alternative).

We first study the effect of the signal sparsity level of the mutual signals and the non-mutual signals. We use $s_0$ to denote the number of simultaneous signals among the $K$ experiments, and $s_k$ to denote the number of the signals that are only present in the $k$th experiment. We study the three cases: 1. $s_1 = s_2 = 0$; 2. $s_1 = 0, s_2 \neq 0$; 3. $s_1 = s_2 \neq 0$.

Next, we study the effect of the correlations among the covariates. Third, we study the effect of the difference in signal strengths between the two experiments. We consider two scenarios for the signal strengths: Scenario 1. the directions and the strengths of the mutual signals are identical among the $K$ experiments; Scenario 2. only the directions of the mutual signals are the same, but the signal strengths are independent among the $K$ experiments. Data generation and algorithm implementation details can be found in Web Appendix C. Additional simulations for the $K = 3$ case, the power comparison among different choices of the filter statistics $W$, and the empirical distributions of $W$ to show why the method has power are also provided in Web Appendix C.

4.2 | Results

Figure 2 shows the power and the FDR for the three methods (simultaneous in black, pooling in red, and intersection in blue) on the three data settings (continuous, binary, and mixed) when we vary $s_1 = s_2$. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version. As $s_1 = s_2$ increases, only the simultaneous method controls the FDR. The simultaneous method has slightly lower power than the pooling method, and the power gap is still moderate when the signals in the two experiments have different strengths (Scenario 2, right panel). The simulation results are in agreement with our theoretical expectations. First, in terms of FDR control,
FIGURE 2  The power and the FDR for the continuous (upper), binary (middle), and mixed (lower) settings when varying $s_1 = s_2$. Results for Scenario 1 are on the left and for Scenario 2 are on the right. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.
the simultaneous method we proposed always controls the FDR across all our designed settings. The pooling method only controls the FDR when all samples from the two experiments are i.i.d. The intersection method controls the FDR when \( s_2 = 0 \) but it fails when \( s_1 = s_2 \neq 0 \). In terms of power, there is some gap between the simultaneous and the pooling methods, because the tests of union null hypotheses are more stringent. However, the gap is moderate. More detailed simulation results can be found in Web Appendix C. The simulation results for the \( K = 3 \) case are similar to the \( K = 2 \) case and are consistent with our theoretical expectations (see Figure C6 in Web Appendix C). The simultaneous method controls the FDR and has good power. The pooling method has high power but also has very high FDR, when there are signals that are shown in either one or two of the samples only. The intersection method has similar power to the simultaneous method, but it cannot control the FDR when a large number of features are signals in only two of the three samples.

The comparison among different \( W \) statistics suggests that the Max and Diff (see definitions in Web Appendix C.1.3) \( W \)s have the best performance among the \( W \)s we have explored. More simulation results can be found in Web Appendix C.

## 5 | REAL-DATA ANALYSIS

In this section, we demonstrate the application of the Simultaneous knockoff method on two real-data examples. For the first data example, we use the Fixed-X knockoffs with linear models for the individual experiments; and for the second data example, we use the Model-X knockoffs with a penalized Cox regression model for each gene expression experiment.

### 5.1 | Application to the communities and crime data

In a crime rate study, we aim to identify features that are universally associated with the community crime rate, regardless of race distribution in the community. This is potentially useful in guiding unbiased policy-making based on race-blinded findings. To achieve that, we select features that are simultaneously associated with the crime rate in different race distribution groups.

We use the publicly available Communities and Crime dataset from the University of California Irvine (UCI) machine learning repository. The dataset contains crime information on \( n = 1994 \) communities with different race distributions in the US. For the individual communities, it has information on the crime rate, as well as 122 other variables that are potentially related to the crime rate. All continuous variables are normalized to the 0–1 range. Our primary outcome of interest is the normalized crime rate, our feature candidates are the \( p = 95 \) features with no missing values that are not directly defined by race. We split the data into two subsets with approximately equal sample sizes based on the proportion of the Caucasian population (high/low) within the community. We fit a linear regression model to each subset of the data, aiming to identify mutual signals from both models with FDR control. We compare the three variable selection procedures (simultaneous, pooling, and intersection) using the knockoff filter (7). We also compare our method with the repFDR (Heller et al., 2014). The repFDR is developed for replicability studies, which requires that under the null the \( Z \)-scores are normally distributed. More details can be found in Web Appendix D.1.

Table 1 shows the results of identified features from different algorithms with a targeted FDR level of \( q = 0.1 \). Our proposed simultaneous method selected the following variables: “the percentage of households with public assistance income in 1989,” “the percentage of kids born to never married,” “the percent of persons in dense housing.” The pooling method selected “the percentage of kids born to never married,” “the percent of persons in dense housing,” and “the number of vacant households.” The intersection method selected “the percentage of kids born to never married,” and “the percent of persons in dense housing.” The repFDR method selected “the percentage of males who have never married.”

To verify the robustness of our proposed method, we also added a set of 95 fake features by permutation. The feature selection results are shown in Table 1 (sensitivity). The variable selections with the simultaneous method is relatively stable.

### 5.2 | Application to the The Cancer Genome Atlas data

In this section, we demonstrate the usage of the Simultaneous knockoffs to identify gene expressions that are associated with glioblastoma multiforme (GBM) for both male and female sub-populations. GBM is known as a hallmark of the malignant process, however, the molecular mechanisms that dictate the locally invasive progression remain an active research area. In this example, we use male and female sub-populations to demonstrate variable selection using our proposed simultaneous method to identify mutual signals from heterogeneous datasets. In real applications, the sub-populations can be much more complicated (i.e., from different sources, collected at different places and times, and with different technologies).
**TABLE 1** Feature selection results for the primary analysis of the community crime data and the sensitivity analysis of the community crime data with added fake features from permutations.

| Analysis  | Method     | Number of features selected | Number of fake features selected |
|-----------|------------|-----------------------------|---------------------------------|
| Main      | Simultaneous | 3                           | NA                              |
|           | pooling     | 3                           | NA                              |
|           | intersection | 2                           | NA                              |
|           | repFDR      | 1                           | NA                              |
| Sensitivity | Simultaneous | 3                           | 0                               |
|           | pooling     | 3                           | 0                               |
|           | intersection | 1                           | 0                               |
|           | repFDR      | 0                           | 0                               |

Note: NA indicates not applicable.

Therefore, the fact that the *simultaneous* method does not require the data to be pooled makes it useful.

Our GBM gene expression data are from The Cancer Genome Atlas (TCGA). The data contain 501 subjects with the overall survival outcome (in days) and 17,813 level-3 gene-level expression data. There are 71 censored and 430 death cases. We use the sure independence screening (SIS, see Fan and Lv 2008) for a marginal screening, leaving $d = \lfloor n / \log(n) \rfloor = 79$ genes with the smallest $p$-values. The SIS method allows dimension reduction from exponentially growing $p$ to a relatively large scale $d < n$, while the reduced model still contains all the true signals with high probability. It has been widely used in other studies (Zhang et al., 2021; Luo et al., 2020). We apply the *Simultaneous knockoffs* to identify genes associated with the survival time within both the male and female GBM patient cohorts. We also compare our method with the methods *pooling*, *intersection*, and *repFDR*. We perform sensitivity analysis by relaxing the screening procedure to include all genes with $p$-values smaller than 0.0002, which leads to 111 candidate genes after the pre-screening step. For missing data, a complete case analysis was performed for the main analysis, while a single imputation was performed in the sensitivity analysis.

With the *Simultaneous knockoff* method, three genes (EID3, RNPS1, and VPS72) are selected. All these genes have been frequently studied for their relationships with cancer, including GBM (Kunadis et al., 2021; Goyal et al., 2021; Heiland et al., 2016). The *pooling* method selected two genes (CROCC and FMR1NB), and the *intersection* method selected none. The *repFDR* method selected 2 genes (MAP2K4 and ZNF239). All the three genes selected by *Simultaneous knockoffs* are also selected when using the threshold $p < 0.0002$ for pre-screening, although one additional gene FMR1NB is also selected under the relaxed screening scenario. Sensitivity analysis also shows that EID3 and VPS72 are selected when we use single imputation to treat the missing data.

### 6 | DISCUSSION

The *Simultaneous knockoff* method is a general framework for testing the union null hypotheses on conditional associations between candidate features and outcomes. It can work with very general conditional models and covariate structures, assuming the $K$ experiments are independent. This method provides opportunities to combine information from the experiments with heterogeneous $X$ structures, different dependencies of $Y|X$, and different outcomes $Y$. The FDR control guarantee is exact for finite sample settings.

This method has even broader applications beyond our motivating examples. For example, when working with the electronic health record (EHR) data from multiple data centers, some outcome variables and covariates are recorded differently among the centers (e.g., for obesity, some centers record the body mass index (BMI) of patients, but others use yes/no); and the demographic distributions are different. The *Simultaneous knockoff* method can be used to identify mutual signals to confirm the associations. This method also only requires very limited information (only the test statistics) to be shared among the data centers, which benefits data collaboration under privacy protections.

One big limitation of the current method is that it is hard to work with ultra-high-dimensional data due to the limits of the computer memory for the knockoff construction. We use the SIS pre-screening step in our real-data example to circumvent this problem. Although theoretically, the *Simultaneous knockoff* method does not require the number of variables to be smaller than the number of observations when using the Model-X knockoff construction, the efficient construction of the knockoffs for ultra-high-dimensional features is still challenging and worth further research. Another limitation of the work is the lack of a theoretical analysis of power. This problem is difficult in general and the power of the Model-X knockoff...
method has just been studied (Wang & Janson, 2021) recently. We expect the power of the Simultaneous knockoff method to decrease monotonically as $K$ and $p$ increase. The exact power change with the growth of $n$, $p$, and $K$ is still a challenging open question.

Our current implementation and numerical experiments use the fixed and second-order knockoff constructions for continuous predictors. Constructing valid knockoffs for non-continuous data is another challenge worth pursuing. For example, the sequential knockoff method (Kormaksson et al., 2021) has been proposed to efficiently implement the general Model-X knockoff sampling algorithm for a combination of continuous and categorical predictors. However, to achieve the theoretical results (Theorem 2) on the robustness of the selection using this method with an approximated $X$ distribution, one still needs to obtain an upper bound for the $KL_j^k$ for this knockoff construction method.

There are some extensions of the knockoff methods to work with group-wise variable selection (Chen et al., 2019; Dai & Barber, 2016), where we are interested in testing whether each specific group of variables is associated with the outcome conditioning on other groups of variables. The current version of Simultaneous knockoff method focuses on the selection of individual features. The extension to work on the selection of groups of features is worth future explorations.

There are many open questions in multiple testing that are related to the hypothesis testing for the union null hypotheses. Although our Simultaneous knockoff method provides solutions to the reproducibility of studies, feature selections across heterogeneous populations, and mediation analysis, there are still more challenges from the real applications. For example, we can further explore methods that will allow combining the information from different datasets with unidentified overlapping samples (e.g., a case–cohort study).

**ACKNOWLEDGMENTS**

This research is partly supported by the National Cancer Institute under grant R01 CA119171 and by the National Institute of General Medical Sciences under grant U54 GM115458.

**DATA AVAILABILITY STATEMENT**

The data that support the findings in this paper are available at https://archive.ics.uci.edu/ml/datasets/communities+and+crime (The Community and Crime data) and at http://www.liuzlab.org/TCGA2STAT/#package-archive (TCGA data) Wan et al. (2015).

**ORCID**

Cheng Zheng @ https://orcid.org/0000-0002-6562-870X

**REFERENCES**

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

Barber, R.F. & Candés, E.J. (2019) A knockoff filter for high-dimensional selective inference. *Annals of Statistics*, 47(5), 2504–2537.

Barber, R.F., Candés, E.J. & Samworth, R.J. (2020) Robust inference with knockoffs. *Annals of Statistics*, 48(3), 1409–1431.

Bates, S., Candés, E., Janson, L. & Wang, W. (2021) Metropolized knockoff sampling. *Journal of the American Statistical Association*, 116(535), 1413–1427.

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(1), 289–300.

Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165–1188.

Bogomolov, M. & Heller, R. (2013) Discovering findings that replicate from a primary study of high dimension to a follow-up study. *Journal of the American Statistical Association*, 108(504), 1480–1492.

Bogomolov, M. & Heller, R. (2018) Assessing replicability of findings across two studies of multiple features. *Biometrika*, 105(3), 505–516.

Candés, E., 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*, 80(3), 551–577.

Chen, J., Hou, A. & Hou, T.Y. (2019) A prototype knockoff filter for group selection with FDR control. *Information and Inference: A Journal of the IMA*, 9(2), 271–298.

Chi, Z. (2008) False discovery rate control with multivariate $p$-values. *Electronic Journal of Statistics*, 2, 368–411.

Dai, R. & Barber, R. (2016) The knockoff filter for FDR control in group-sparse and multitask regression. *Proceedings of The 33rd International Conference on Machine Learning*, 48, 1851–1859.

Fan, J. & Lv, J. (2008) Sure independence screening for ultrahigh dimensional featurespace. *Journal of the Royal Statistical Society: Series B*, 70(5), 849–911.

Goyal, B., Yadav, S. R.M., Awasthee, N., Gupta, S., Kunnumakkara, A.B. & Gupta, S.C. (2021) Diagnostic, prognostic, and therapeutic significance of long non-coding RNA malat1 in cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1875(2), 188502.

Heiland, D.H., Staszewski, O., Hirsch, M., Masalha, W., Franco, P., Grauvogel, J., et al. (2016) Malignant transformation of a dysembryoplastic neuroepithelial tumor (dnet) characterized by genome-wide methylation analysis. *Journal of Neuropathology & Experimental Neurology*, 75(4), 358–365.

Heller, R., Bogomolov, M. & Benjamini, Y. (2014) Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. *Proceedings of the National Academy of Sciences*, 111(46), 16262–16267.

Heller, R. & Yekutieli, D. (2014) Replicability analysis for genome-wide association studies. *Annals of Applied Statistics*, 8(1), 481–498.

Huang, D. & Janson, L. (2020) Relaxing the assumptions of knockoffs by conditioning. *Annals of Statistics*, 48(5), 3021–3042.
Ioannidis, J., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Culhane, A.C., et al. (2009) Repeatability of published microarray gene expression analyses. Nature Genetics, 41(2), 149–155.

Kormaksson, M., Kelly, L.J., Zhu, X., Haemmerle, S., Pricop, L. & Ohlssen, D. (2021) Sequential knockoffs for continuous and categorical predictors: with application to a large psoriatic arthritis clinical trial pool. Statistics in Medicine, 40(14), 3313–3328.

Kunadis, E., Lakiotaki, E., Korkolopoulou, P. & Piperi, C. (2021) Targeting post-translational histone modifying enzymes in glioblastoma. Pharmacology & Therapeutics, 220, 107721.

Liu, Y. & Zheng, C. (2019) Deep latent variable models for generating knockoffs. Stat, 8(1), e260.

Luo, C., Fa, B., Yan, Y., Wang, Y., Zhou, Y., Zhang, Y., et al. (2020) High-dimensional mediation analysis in survival models. PLoS Computational Biology, 16(4), e1007768.

NCI-NHGRI, (2007) Replicating genotype–phenotype associations. Nature, 447(7145), 655–660.

Plant, A.L. & Hanisch, R.J. (2020) Reproducibility in science: a metrology perspective. Harvard Data Science Review, 2(4).

Rittschof, C.C., Bukhari, S.A., Sloofman, L.G., Troy, J.M., Caetano-Anollés, D., Cash-Ahmed, A., et al. (2014) Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. Proceedings of the National Academy of Sciences, 111(50), 17929–17934.

Romano, Y., Sesia, M. & Candès, E. (2020) Deep knockoffs. Journal of the American Statistical Association, 115(532), 1861–1872.

Sampson, J.N., Boca, S.M., Moore, S.C. & Heller, R. (2018) FWER and FDR control when testing multiple mediators. Bioinformatics, 34(14), 2418–2424.

Sesia, M., Sabatti, C. & Candès, E.J. (2018) Gene hunting with hidden Markov model knockoffs. Biometrika, 106(1), 1–18.

Spector, A. & Janson, L. (2022) Powerful knockoffs via minimizing reconstructability. Ann. Statist. 50(1), 252–276.

Uffelmann, E., Huang, Q.Q., Munung, N.S., De Vries, J., Okada, Y., Martin, A.R., et al. (2021) Genome-wide association studies. Nature Reviews Methods Primers, 1(1), 1–21.

Wan, Y.-W., Allen, G.I. & Liu, Z. (2015) TCGA2STAT: simple TCGA data access for integrated statistical analysis in R. Bioinformatics, 32(6), 952–954.

Wang, W. & Janson, L. (2021) A high-dimensional power analysis of the conditional randomization test and knockoffs. Biometrika, 109(3), 631–645.

Xiang, D., Zhao, S.D. & Tony Cai, T. (2019) Signal classification for the integrative analysis of multiple sequences of large-scale multiple tests. Journal of the Royal Statistical Society: Series B, 81(4), 707–734.

Zhang, H., Zheng, Y., Hou, L., Zheng, C. & Liu, L. (2021) Mediation analysis for survival data with high-dimensional mediators. Bioinformatics, 37(21), 3815–3821.

Zhao, S.D. & Nguyen, Y.T. (2020) Nonparametric false discovery rate control for identifying simultaneous signals. Electronic Journal of Statistics, 14(1), 110–142.

SUPPORTING INFORMATION
Web Appendices A–D referenced in Sections 2–5, together with the R codes for the analysis of this paper, are available with this paper at the Biometrics website on Wiley Online Library. The R package gsknockoff and an example code for installing and running the package are available at https://github.com/zzhengccheng/gsknockoff.

How to cite this article: Dai, R. & Zheng, C. (2023) False discovery rate-controlled multiple testing for union null hypotheses: a knockoff-based approach. Biometrics, 79, 3497–3509. https://doi.org/10.1111/biom.13848