RESEARCH PAPER

Accurate error control in high-dimensional association testing using conditional false discovery rates

James Liley[1,2] | **Chris Wallace**[1,2,3]

[1] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[2] Department of Medicine, Addenbrookes Hospital, University of Cambridge, Cambridge, UK

[3] Cambridge Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, Cambridge, UK

Correspondence
James Liley, MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, UK.
Email: james.liley@igmm.ed.ac.uk

Abstract

High-dimensional hypothesis testing is ubiquitous in the biomedical sciences, and informative covariates may be employed to improve power. The conditional false discovery rate (cFDR) is a widely used approach suited to the setting where the covariate is a set of p-values for the equivalent hypotheses for a second trait. Although related to the Benjamini–Hochberg procedure, it does not permit any easy control of type-1 error rate and existing methods are over-conservative. We propose a new method for type-1 error rate control based on identifying mappings from the unit square to the unit interval defined by the estimated cFDR and splitting observations so that each map is independent of the observations it is used to test. We also propose an adjustment to the existing cFDR estimator which further improves power. We show by simulation that the new method more than doubles potential improvement in power over unconditional analyses compared to existing methods. We demonstrate our method on transcriptome-wide association studies and show that the method can be used in an iterative way, enabling the use of multiple covariates successively. Our methods substantially improve the power and applicability of cFDR analysis.

KEYWORDS
conditional false discovery rate, empirical Bayes, false discovery rate, high-dimensional association study, transcriptome-wide association study, unsupervised learning

1 | INTRODUCTION

In the 'omics' approach to biology, a large number n of descriptive variables are considered in the analysis of a biological system, intended to provide a near-exhaustive characterisation of the system under consideration. Typically only a small proportion of the investigated variables are associated with the behaviour of the system, and we seek to identify this subset of variables, along with the magnitude and direction of their associated effect sizes. A first step is generally to test each hypothesis in a frequentist framework, generating a corresponding set of p-values. Often, additional information is available in the form of an external covariate, which assigns a numerical value to each hypothesis which has

different (unknown) distributions amongst associations and non-associations. Information from such covariates can be incorporated into hypothesis testing to improve power in detecting associations.

A range of procedures have been proposed for this type of analysis. An important consideration is the form of the (two-dimensional) rejection rule applied to the p-value-covariate pairs. An optimal procedure (in terms of minimising type 2 error and controlling type 1 error) determines rejection regions on the basis of a ratio of bivariate probability densities (PDFs) of the p-value and covariate under the null and under the alternative. One approach to the problem at hand is to estimate this ratio directly (Alishahi et al., 2016; Du et al., 2014; Lei and Fithian, 2018). Other approaches include 'filtering' on covariate values (Bourgon et al., 2010), weighting hypotheses according to the value of the covariate (Basu et al., 2018; Benjamini et al., 2006; Cai et al., 2016; Genovese et al., 2006; Ignatiadis et al., 2016), modulating a univariate test of p-values in response to the covariate in some other way (Li and Barber, 2016, 2017; Scott et al., 2015) and binning covariates in order to treat each bin separately (Ferkingstad et al., 2008). Since covariates can be of many types (continuous, categorical; univariate, multivariate; known or unknown distributional properties) and can relate to the p-values in a range of ways, this array of methods is necessary to manage the range of problem types.

The conditional false discovery rate (cFDR) circumvents the difficulties of estimating PDFs by approximating the optimal ratio using cumulative density functions (CDFs) (Andreassen et al., 2013c). In this case, the covariate is generally a set of p-values arising from an analogous procedure on the same variables for a second 'conditional' trait with an unknown degree of similarity to the trait giving rise to the primary set of p-values (which we call the 'principal' trait). The method has been extensively used in genomics (Andreassen et al., 2013a, 2013b, 2013c, 2014a, 2014b, 2015; Broce et al., 2018; Desikan et al., 2015a, 2015b; Ferrari et al., 2017; Karch et al., 2018; Le Hellard et al., 2016; Liley and Wallace, 2015; Liu et al., 2013; Lv et al., 2017; McLaughlin et al., 2017; Schork et al., 2016; Shadrin et al., 2018; Smeland et al., 2017a, 2017b, 2017c; Van der Meer et al., 2020; Wang et al., 2016; Witoelar et al., 2017; Zuber et al., 2018). Formally, the cFDR is a posterior probability of non-association with the principal trait given that p-values for the principal and conditional traits fall below p-value thresholds p, q, respectively. It is readily estimated using empirical CDFs (ECDFs) (Andreassen et al., 2013c).

The cFDR is a useful Bayesian quantity in its own right. Generally, the cFDR is used in effectively a frequentist way: roughly, for each observed p-value pair (p_i, q_i), we estimate the cFDR at $(p, q) = (p_i, q_i)$ and reject the null hypothesis if this estimated value is less than some threshold α. This process is nearly analogous to the Benjamini–Hochberg procedure (B-H) (Benjamini and Hochberg, 1995) on a single set of p-values p_i, but unlike B-H, it does not control the false discovery rate (FDR) at α (nor any other conventional measure of type-1 error rate). In a previous paper (Liley and Wallace, 2015), we proposed a rough method to approximately control FDR in this setting, but our method was drastically conservative.

The main contribution of this paper is to propose a much improved type-1 error rate control strategy for cFDR, which improves power relative to previous methods. Our method transforms cFDR estimates into 'v-values', which function analogously to p-values and can be used to control FDR or family-wise error rate (FWER). In four secondary contributions, we (a) propose an improvement to the existing estimator which improves power, (b) show several asymptotic results about the method and demonstrate that the effect of certain troublesome properties is small, (c) enable and demonstrate iterative use of the procedure and (d) compare the general cFDR method with PDF-based, parametric and kernel density estimator (KDE)-based approaches. An R package is provided.

In this paper, we begin by describing a motivating example using transcriptome-wide association studies (TWAS). We then summarise the cFDR and its estimator and describe its relation to the B-H procedure. We then describe our method to transform cFDR estimates into p-value-like quantities and discuss how the cFDR approach relates to similar methods in the field. We evaluate the type-1 error rate control and power of the method and finally describe an iterated form of the procedure for use with multiple sets of covariates.

1.1 | Motivating example

We consider a TWAS (Gusev et al., 2016) of breast cancer BRCA (Michailidou et al., 2017) and ovarian cancer (OCA; Phelan et al., 2017), which are epidemiologically and biologically similar diseases (Greene et al., 1984). TWASs test for association between levels of predicted expression of transcripts (gene products) in various tissues between cases (BRCA or OCA) and controls. For each transcript–tissue pair, the TWAS generates a p-value against the null hypothesis that the predicted mean expression of that transcript in that tissue is the same in case and control populations, according to a transcript-prediction rule learnt from independent data. The TWASs in question test around 10,000 gene transcripts across 45 tissues (though many transcript–tissue pairs are missing), and after we restrict to transcript–tissue pairs common to both studies, we are left with a set of $\approx 10^5$ p-values p_{BRCA}, p_{OCA} for association with BRCA and OCA, respectively (further detail is given in
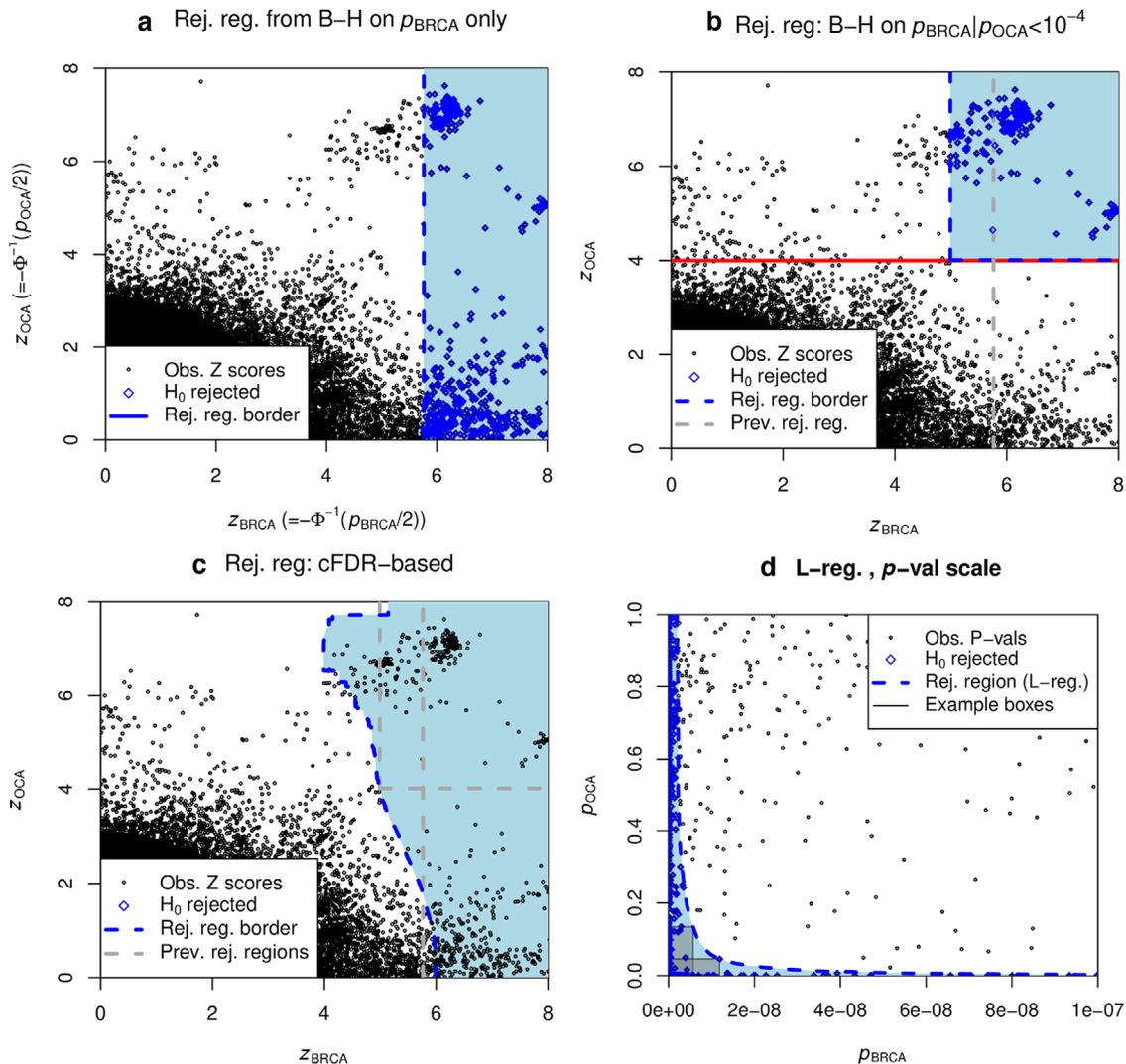
FIGURE 1 Illustration of cFDR approach using example data from the TWAS study of breast cancer (BRCA), conditioning on ovarian cancer (OCA). Each plot shows test statistics from BRCA (x axis) and OCA (y axis) on either the Z (a–c) or p-value (d) scale. All rejection regions use methods to control FDR at $< 1 \times 10^{-6}$. (a) B-H procedure applied to BRCA statistics alone leads to a rejection region to the right of the blue dashed vertical line. (b) B-H applied to those variables for which z_{OCA} exceeds the threshold shown by a solid red line. (c) cFDR procedure: for the ith values $z_{BRCA}(i)z_{OCA}(i)$, a B-H procedure aiming to control the FDR at α is conducted on only the variables for which $z_{OCA} \geq z_{OCA}(i)$, and if the ith null hypothesis is rejected during this procedure, it is rejected overall. We term the rejection region corresponding to this value α an 'L-region' $L(\alpha)$, shown as the shaded region. (d) The exposition that follows using p-values rather than Z scores, and so we reproduce the data and $L(\alpha)$ on the p-value scale. On this scale, the estimated cFDR at a point p_{BRCA}, p_{OCA} can be considered an estimate of the FDR corresponding to a fixed rejection region given by the box with p_{BRCA}, p_{OCA} as its top-right corner, and the L-region $L(\alpha)$ roughly as the locus of top-right corners of boxes with estimated cFDR equal to α. Two such boxes are illustrated in the figure

Supporting Information Section 1.1). We wish to find which of the variables are associated with BRCA, and thus investigate a null hypothesis H_{BRCA}^0 of non-association. Given established genetic correlations BRCA and OCA, we hope to leverage the OCA TWAS results to increase power in this search. We will assume that we have no prior knowledge that any variables are more likely to be BRCA- or OCA- associated, that absolute Z-scores $z_{BRCA} = -\Phi^{-1}(p_{BRCA}/2), z_{OCA} = -\Phi^{-1}(p_{OCA}/2)$ have a block-diagonal correlation structure where block locations are known, and that under a null hypothesis H_{BRCA}^0 of no association with BRCA, z_{BRCA} and z_{OCA} is independent.

A straightforward approach is to apply the B-H procedure to the values p_{BRCA} (Figure 1, panel a). BRCA and OCA tend to have associations at the same variables, suggesting that a rejection region should reflect this to improve power. A natural way to do this is to only consider those variables for which z_{OCA} exceeds some threshold, which allows rejection of H_{BRCA}^0 at a looser z_{BRCA} threshold (Figure 1, panel b). FDR control is maintained under the independence assumption

above (Bourgon et al., 2010). However, this procedure is problematic: a threshold on z_{OCA} must be chosen *a priori* and variables with z_{OCA} falling below the red line have H^0_{BRCA} retained automatically.

The cFDR procedure circumvents this problem (Figure 1, panel c). The associated rejection region, which we term an L-region, 'adapts' to the joint distribution of z_{BRCA} and z_{OCA}. For small α, the L-region approximates an optimal rejection region (Appendix A.1). A major shortcoming is that although a B-H procedure with FDR controlled at α is repeatedly used to generate the L-region, the overall FDR is not controlled at α, nor any straightforward function of α (Liley and Wallace, 2015; Appendix A.2).

In this paper, we demonstrate a straightforward and effective way to control the type-1 error rate (specifically FDR or FWER) in the cFDR procedure. As α varies from 0 to 1, the leftmost boundary of the L-region 'sweeps' across the entire $(+,+)$ quadrant, and for $\alpha_1 < \alpha_2$, we have $L(\alpha_1) \subseteq L(\alpha_2)$. Thus we can associate each point (x,y) in the $(+,+)$ quadrant with the smallest L-region containing it, which will generally have (x,y) on its leftmost border. Loosely, we control FDR by estimating the probability that each point would lie within its associated region under H^0_{BRCA}. We term this the v-value, which has similar properties to a p-value and can be used in the B-H procedure.

Care must be taken when applying rejection rules to the same data on which those rules were determined, so we use a leave-one-out procedure which avoids this problem (Section 3, Appendix A.4). We show that the rejection region generated by the cFDR approximates the best-possible rejection region (Section 2.2, Appendix A.1), and that rejection regions converge reasonably fast as the number of variables under consideration increases (Section A.3). The rejection region is non-parametric, and we show that the cFDR method can outperform parametric methods (Section 5.2). Finally, the v-values may be considered 'adjusted' p-values, which enables straightforward iteration of the method with further sets of p-values at the same variables, discussed in Section 5.5.

2 | REVIEW OF cFDR ESTIMATOR

2.1 | Definitions

Assume that we have results from n pairs of hypothesis tests against two series of null hypotheses $(H^p_0(i), H^q_0(i))$ in the form of a set S of bivariate p-values $S = (p_i, q_i), i = 1 \dots n$. In our motivating example, $H^p_0(i)$ and $H^q_0(i)$ are non-association of the ith tissue–gene pair with BRCA and OCA, respectively. We consider $(H^p_0(i), H^q_0(i))$ to be realisations of independent Bernoulli random variables H^p_0, H^q_0 satisfying $Pr(H^p_0) = \pi_0, Pr(H^q_0) = \pi^q_0$, and p_i, q_i to be independent and identically distributed (IID) realisations of random variables P, Q satisfying:

$$P|H^p_0 \sim U(0,1)$$
$$P \perp\!\!\!\perp Q|H^p_0 \tag{1}$$

although assumption (1) can be relaxed. We denote

$$F_0(p,q) = Pr(P \le p, Q \le q|H^p_0) = pF^q_0(q) \tag{2}$$

$$F(p,q) = Pr(P \le p, Q \le q)$$
$$f_0(p,q) = f(P = p, Q = q|H^p_0) = f^q_0(q) \tag{3}$$
$$f(p,q) = f(P = p, Q = q),$$

where the separability of (2) and (3) is due to assumption (1).

2.2 | Optimal procedure

Under H^p_0, the probability of a random instance of (P, Q) falling in a region R is $\int_R f_0(p,q)dpdq$. To find an ideal two-dimensional rejection region for hypothesis testing, we wish to fix this value at a level α while maximising the probability

$\int_R f(p,q)dpdq$. This optimal region (or one such optimal region) is given by the set of points $\{(p,q) : f_0(p,q)/f(p,q) \geq k_\alpha\}$ for some k_α (a formal statement and proof are given in Appendix A.1, and this is also shown in various forms in Du et al. (2014), Alishahi et al. (2016) and Lei and Fithian (2018). In equivalent terms, an optimal decision rule for the set S would rank p-value pairs according to $f_0(p_i,q_i)/f(p_i,q_i)$ or equivalently $Pr(H_0^p|P = p_i, Q = q_i)$.

A natural approach is to estimate f_0 and f using a parametric approximation (Du et al., 2014; Lei and Fithian, 2018) or local approximations using KDEs (Alishahi et al., 2016) or spline models (Zablocki et al., 2014). However, PDFs are difficult to estimate in general, and there may be little reason to believe parametric assumptions are satisfied; in our motivating example (Figure 1, panels a-c) there is little reason to think that a smooth rejection region would be optimal.

2.3 | Conditional false discovery rate

The cFDR (Andreassen et al., 2013c) takes an alternative approach of instead ranking points by an estimate of $F_0(p,q)/F(p,q)$. This estimate is obtained by estimating the monotonically related quantity:

$$cFDR(p,q) = Pr(H_0^p|P \leq p, Q \leq q) \tag{4}$$

$$= \frac{Pr(P \leq p|H_0^p, Q \leq q)}{Pr(P \leq p|Q \leq q)}Pr(H_0^p|Q \leq q). \tag{5}$$

Suppose we have a multi-set X of p-value pairs (p_i,q_i). If almost all these pairs are IID realisations (p_i,q_i) of (P,Q), then for fixed p,q, the ECDFs

$$\frac{1}{|X|}|\{i : (p_i,q_i) \in X, p_i \leq p, q_i \leq q\}|$$

$$\frac{1}{|X|}|\{i : (p_i,q_i) \in X, q_i \leq q\}|$$

are consistent estimators of $Pr(P \leq p, Q \leq q), Pr(Q \leq q)$, respectively. Given assumption (1), we have $p = Pr(P \leq p|H_0^p) = Pr(P \leq p|Q \leq q, H_0^p)$ and (for the moment) we may conservatively approximate $Pr(H_0^p|Q \leq q) = 1$. Given X, we thus define the estimated cFDR (denoted $\widehat{cFDR}$), as a function of two variables $(p,q) \in (0,1)$:

$$\widehat{cFDR}_X(p,q) = p\frac{\max(|\{i : q_i \leq q, (p_i,q_i) \in X\}|, 1)}{\max(|\{i : p_i \leq p, q_i \leq q, (p_i,q_i) \in X\}|), 1)}. \tag{6}$$

For fixed p,q, $\widehat{cFDR}_X(p,q)$ is a generally biased but consistent estimator of $cFDR(p,q)/Pr(H_0^p|Q \leq q)$, which converges uniformly on fixed regions at a rate of $O(n^{-1/2})$ (see Appendix A.3), and it is usually a downwards-biased (conservative) estimator of $cFDR(p,q)$.

Approximating $Pr(H_0^p|Q \leq q) = 1$ in Equation (6) disregards any variation on $Pr(H_0^p|Q \leq q)$ with q, so we introduce at this stage an estimate of $Pr(H_0^p|Q \leq q)$, which we can multiply with $\widehat{cFDR}_X(p,q)$ to improve the accuracy of approximation of $cFDR(p,q)$. Our estimate is

$$Pr(H_0^p|Q \leq q) = Pr(H_0^p)\frac{Pr(Q \leq q|H_0^p)}{Pr(Q \leq q)}$$

$$\approx \pi_0\frac{Pr(Q \leq q|P > 1/2)}{Pr(Q \leq q)}$$

$$\approx \frac{\min(1, |\{i : (p_i,q_i) \in X, q_i \leq q, p_i > 1/2\}|)}{\min(1, |\{i : (p_i,q_i) \in X, q_i \leq q\}||\{i : (p_i,q_i) \in X, p_i > 1/2\}|)}$$

$$= \widehat{Pr}_X(H_0^p|Q \leq q), \tag{7}$$

where we approximate $\pi_0 = Pr(H_0^P) = 1$. We denote

$$\widehat{cFDR}_X^n(p, q) = \widehat{cFDR}_X(p, q)\widehat{Pr}_X(H_0^p | Q \leq q). \tag{8}$$

Estimating π_0 (rather than setting $\pi_0 = 1$) would uniformly scale all estimates of $cFDR(p, q)$, which has no effect on our rejection procedure.

In the hypothesis testing setting, we aim to use $\widehat{cFDR}$ or $\widehat{cFDR}^n$ to construct a decision rule on our set S of observed p-value pairs (we will forego the n superscript from now, with the understanding that it may be added). A simple approach is to reject $H_0^p(i)$ if $\widehat{cFDR}_S(p_i, q_i) \leq \alpha$, but $\widehat{cFDR}_S(p, q)$ is not monotonically increasing with p and we do not want to reject H_0^p for some (p_i, q_i) and not for some other pair (p_j, q_j) with $q_i = q_j$ but $p_j < p_i$. Hence, we can use the decision rule (as per Andreassen et al., 2013c)

$$\text{Reject } H_0^p \text{ if: } \exists \, p' \geq p_i \, : \, \widehat{cFDR}_S(p', q_i) \leq \alpha. \tag{9}$$

This enables a rejection region with a single rightmost boundary, as shown in panels D in Figure 1. It closely parallels the B-H (Benjamini and Hochberg, 1995) procedure on a set of p-values. Suppose we have a set S_1 of p-values $p_1, p_2, \ldots, p_n$, and define

$$BH_{S_1}(p) = p \frac{|S_1|}{\max(1, |\{i \, : \, p_i \leq p, p_i \in S_1\}|)}.$$

Then the B-H procedure can be written as

$$\text{Reject } H_0^p \text{ if: } \exists \, p' \geq p_i \, : \, BH_{S_1}(p') \leq \alpha. \tag{10}$$

The B-H procedure controls FDR at α, and if it is performed with S_1 as the subset of S for which $q_i \leq \gamma$ (where γ is a threshold chosen independently of S), the FDR will still be controlled at α if assumption (1) is satisfied (Bourgon et al., 2010; Ignatiadis et al., 2016). The rejection procedure (9) is equivalent to repeatedly performing this 'thresholded' B-H at $\gamma = q_i$, and using that decision rule for point (p_i, q_i) (panel c, Figure 1).

When procedure (9) is used, the FDR is no longer controlled at α, and indeed can exceed α by an arbitrary proportion. This is most easily seen in the extreme case

$$P, Q | H_0^p \sim U(0, 1)^2 \tag{11}$$

$$P, Q | H_1^p \sim (0, 0), \tag{12}$$

in which we can show that the FDR α_{TRUE} of rejection procedure (9) applied to $\widehat{cFDR}$ satisfies

$$\lim_{n \to \infty} \left(\frac{\alpha_{TRUE}(1 - \alpha)}{\alpha(1 - \alpha_{TRUE})} \right) = \log \left(\frac{1 - \alpha\pi_0}{1 - \pi_0} \right) \tag{13}$$

and when applied to $\widehat{cFDR}^n$ satisfies

$$\lim_{n \to \infty} \left(\frac{FDR}{\alpha} \right) = \frac{1 - \log\left(\frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0} \right)}{1 - \alpha \log\left(\frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0} \right)} \tag{14}$$

with the corollary that the actual FDR when using rejection procedure (9) can be an arbitrarily large multiple of α. These formulae are proved in Theorem A.3 (Appendix A.2).

Previous work using the cFDR generally interprets it in a Bayesian context, without requiring a bound on FDR or FWER. In Liley and Wallace (2015), we introduced a method to choose an α^* such that rejection criterion (9) roughly controlled the FDR at α, but that method was overly conservative, generally controlling FDR at a far lower level than needed.

3 | MAP FROM p-VALUE PAIRS TO v-VALUES

We identify a 'rejection region' associated with $\widehat{cFDR}$ by adding a 'test point' (p, q) to a set of points X and considering the region for which a hypothesis corresponding to (p, q) is rejected under (9) with $S = X + (p, q)$.

The function $\widehat{cFDR}_{X+(p,q)}(p, q)$ is now defined on the unit square. It is difficult to use, however: when considered as a function of p with fixed q, it does not monotonically increase with p. We thus define

$$\widehat{cFDRt}_X(p, q) = \min_{p' \geq p} \widehat{cFDR}_{X+(p',q)}(p', q) \tag{15}$$

and define the 'L-region' $L_X(\alpha)$:

$$L_X(\alpha) = \left\{ (p, q) : \widehat{cFDRt}_X(p, q) \leq \alpha \right\} \tag{16}$$

$$= \{ (p, q) : \exists p' \geq p : \widehat{cFDR}_{X+(p',q)}(p', q) \leq \alpha \} \tag{17}$$

and define the 'L-curve' as the rightmost boundary of this region. We note that

$$\alpha \leq \beta \Rightarrow L_X(\alpha) \subseteq L_X(\beta). \tag{18}$$

We now show the following, and include the brief proof:

Theorem 3.1. *Assume that $P, Q|H_0^P$ is a bivariate continuous random variable with support $[0, 1]^2$, and set μ_0 as its induced measure. Suppose $X = (p_i, q_i) \in [0, 1]^2$, $i \in 1 \dots n$ is a fixed finite set of points. Define $L_X(\alpha)$ as per Equation (16), with $\widehat{cFDR}$ defined as per either Equation (6) or Equation (8), and define the 'v-value' as a function of $p, q \in (0, 1)^2$*

$$v_X(p, q) = \inf_{\gamma : (p,q) \in L_X(\gamma)} (\mu_0[L_X(\gamma)]) \quad \left(= \min_{\gamma : (p,q) \in L_X(\gamma)} \left(\int_{L_X(\gamma)} f_0(x, y) dx dy \right) \right) \tag{19}$$

the second definition being valid if $P, Q|H_0^P$ admits a PDF (as in Section 2). Then for $\alpha \in (0, 1)$

$$Pr\left(v_X(P, Q) \leq \alpha | H_0^P \right) \leq \alpha. \tag{20}$$

Proof. Since X is finite $L_X(\alpha)$ is Lesbegue-measurable, so the integral in (19) is well-defined. Note that since L-regions are closed regions inside contours of a function monotonic in p, we could replace 'inf' with 'min' in definition (19).

Given $\alpha \in (0, 1)$, let

$$\Gamma(\alpha) = \{ \gamma : \mu_0[L_X(\gamma)] \leq \alpha \}. \tag{21}$$

Suppose there exists $\gamma(\alpha) \in \Gamma(\alpha)$ with $(p, q) \in L_X(\gamma(\alpha))$. Then from definition (19), we have

$$v_X(p, q) \leq \mu_0[L_X(\gamma(\alpha))] \leq \alpha \tag{22}$$

so, using property (18)

$$Pr(v_X(P, Q) \leq \alpha | H_0^P) \leq Pr\left[(P, Q) \in \bigcup_{\gamma(\alpha) \in \Gamma(\alpha)} L_X(\gamma(\alpha)) | H_0^P \right]$$

$$= \sup_{\gamma(\alpha) \in \Gamma(\alpha)} (\mu_0[L_X(\gamma(\alpha))])$$

$$\leq \alpha.$$

If there exists γ such that $\alpha = \mu_0[L(\gamma)]$ then $(p, q) \in L_X(\gamma) \Leftrightarrow v_X(p, q) \leq \alpha$, so equality is achieved in (20). $\qquad\square$

Remark 3.2. With Γ defined as per (21), and setting μ_1 as the induced measure of $P, Q|H_1^P$, the power of the rejection procedure

$$\{\text{Reject } H_0^P \text{ if } v_X(p, q) \leq \alpha\} \tag{23}$$

is

$$Pr(\text{reject } H_0^P | H_1^P) = \sup_{\gamma(\alpha) \in \Gamma(\alpha)} (\mu_1(L_X(\gamma(\alpha)))), \tag{24}$$

and the type-1 error rate is $\leq \alpha$.

Theorem 3.1 assumes the probability measure F_0 of $P, Q|H_0^P$ is known. In practice, it must be estimated. This can be readily done given assumption (1), as will be shown in Section 3.2. We can think of the function $v_X(p, q)$ as a map from the unit square to the unit interval, where the map is defined by the points X.

We note that property (20) indicates that the value $v_X(p, q)$ is interpretable as a p-value against H_0^P, using $\widehat{cFDRt}(p, q)$ as a test statistic (it may even be thought of as a definition of a p-value; Storey et al., 2003). In this sense, it is slightly conservative due to the '$\leq \alpha$' rather than '$= \alpha$' in (20). Resultant v-values may thus be used in the B-H procedure to control FDR, or used with a Šidák correction to control FWER. For the remainder of this paper, we will seek to control FDR by the B-H procedure.

In order to use Theorem 3.1 to generate a p-value for a test point (p, q). we must assume X is 'fixed'. In practice, this means (p_i, q_i) must be independent of values (p, q); hence, not in X.

Given our set of points $S = (p_i, q_i)$, this is easily managed: to test (p_i, q_i), we simply leave (p_i, q_i) out of the points used to define the set of L-regions we use on (p_i, q_i) itself. That is, given our set S of datapoints as above, we define the 'leave-one-out' v-value

$$v(p_i, q_i) = v_{S-(p_i, q_i)}(p_i, q_i). \tag{25}$$

The problem can also be managed by leaving out blocks of points; for a partition of $1 \ldots n$ into blocks $1, 2, \ldots, k$, supposing i is in block $b(i)$, the 'block-out' v-value is defined as

$$v(p_i, q_i) = v_{S-\text{block } b(i)}(p_i, q_i). \tag{26}$$

If observations (p_i, q_i) are not independent but have a block-diagonal correlation structure, then this procedure is necessary in order to ensure property (20) holds for $(p, q) = (p_i, q_i)$: since each observation (p_i, q_i) carries information about the other p-value pairs it is correlated with, removing it will not remove the influence of point (p_i, q_i) on the map. In this case, blocks should be chosen so that p-value pairs are independent between blocks, but possibly dependent within blocks. Such structure arises often in -omics experiments; in genetics, independence of allele counts may be assumed between chromosomes, but not generally within.

For comparison, we also define the 'naive' v-value

$$v(p_i, q_i) = v_S(p_i, q_i), \tag{27}$$

where p_i, q_i is in S.

In the subsequent section, we note several asymptotic properties of L-regions and v-values. We note that consistency of L-region estimation is not necessary for type-1 error rate control: there is no requirement in Theorem 3.1 for the values in X to have the same distribution as P, Q. L-regions are also identical under monotonic transformations of the function $\widehat{cFDR}(p, q)$, so the method is unaffected by the approximation $Pr(H_0^P) \approx 1$ in Equation (7).

3.1 | Asymptotic properties of L-regions and v-values

We describe several asymptotic properties of L-regions and v-values. Estimator (7) is not generally consistent, so we focus our attention on properties of regions defined using $\widehat{cFDR}$ rather than $\widehat{cFDR}^n$. We divide out the quantity estimated in (7), write $F(q) = Pr(Q \leq q)$ and define

$$C(p, q) = \frac{cFDR(p, q)}{Pr(H_0^p | Q \leq q)} = p \frac{F(q)}{F(p, q)},\tag{28}$$

which we generally assumed to be differentiable on the unit square. We recall definition (15) and note the following (proved in Appendix A.3):

Theorem 3.3. *Let R be the region of the unit square for which $F(p, q) \geq \gamma > 0$ and $F(q) > 0$. Then on R, $\widehat{cFDR}(p, q)$ converges uniformly to $C(p, q)$, and if $\partial C(p, q)/\partial p \geq 0$, then so does $\widehat{cFDRt}(p, q)$.*

This theorem indicates that the empirical L-curve converges to a contour of $C(p, q)$, where the contour exists. Although the region R does not cover the entire unit square, in practice it usually has Lesbegue measure 0: if $Pr(H_0^Q, H_0^P) > 0$ and $P, Q | H_0^Q, H_0^P \sim U(0, 1)^2$ then $F(p, q)$ is bounded away from 0 on $(0, 1)^2$. We note that $C(p, q)$ is meaningful only when $F(p, q) > 0$.

Because of the uniform convergence, this can be translated into a statement about v-values. Given an L-region $L(\alpha)$, we define the M-region as the 'expected' L-region:

$$M(\alpha) = \{(p, q) : C(p, q) \leq \alpha\}\tag{29}$$

and the 'error' on the v-value $v = \int_{L(\alpha)} f_0(p, q) dp dq$ as

$$|\Delta v| = \left| \int_{L(\alpha)} f_0(p, q) dp dq - \int_{M(\alpha)} f_0(p, q) dp dq \right|.\tag{30}$$

Now we have

Theorem 3.4. *Define R as in Theorem 3.3, and further assume that $f_0(p, q) = f(P = p, Q = q | H_0^P)$ is known and on R we have $\partial C(p, q)/\partial p \geq \gamma_2 > 0$. Write $R^c = [0, 1]^2 \setminus R$. Then the maximum error on any v-value is*

$$\int_{R^c} f_0(p, q) dp dq + O\left(\frac{1}{\sqrt{n}}\right)\tag{31}$$

thus, as $n \to \infty$, v-values based on $\widehat{cFDR}$ converge at a rate $O(n^{-1/2})$ to those that would be obtained using $C(p, q)$, plus a fixed error. We also note that under the conditions in both theorems, if R has negligible Lesbegue measure, and there exists γ such that

$$\int_{M(\gamma)} f_0(p, q) dp dq = 1$$

then as $n \to \infty$ the power of rejection procedure (23) satisfies

$$Pr(\text{reject } H_0^P | H_1^P) \to \mu_1[M(\gamma)],\tag{32}$$

where μ_1 is defined as in Equation (23). Finally, we note that neither consistency nor unbiasedness of the $\widehat{cFDR}$ estimator is necessary for the p-value property in Theorem 3.1 to hold. Proofs of Theorems 3.3 and 3.4 are given in Appendix A.3.

3.2 | Estimation of $P, Q|H_0^p$

Recalling Equation (3), we may write $f_0(p,q) = f_0^q(q)$. To estimate f_0^q, we assume that $(Q|H_0^P) \sim (Q|P \geq 1/2)$, and approximate the latter with a mixture-Gaussian distribution

$$-\Phi^{-1}\left(\frac{Q}{2}\right)\Big|P > \frac{1}{2} \sim \begin{cases} |N(0,1)| & \text{prob} = \pi_0 \\ |N(0,\sigma_0^2)| & \text{prob} = 1 - \pi_0, \end{cases} \tag{33}$$

where $N(\mu, \sigma)$ is the normal distribution with mean μ and variance σ^2. Estimates $\widehat{\pi_0}, \widehat{\sigma_0}$ of π_0 and σ_0 can be readily made using an expectation-maximisation algorithm (Dempster et al., 1977), using the values q_i for which the corresponding p_i is $\geq 1/2$. We then estimate the density $f_0(p,q)$ of $P, Q|H_0^P$ as

$$\widehat{f_0}(p,q) = \mathbf{1}f_0^q(q) = \widehat{\pi_0} + (1 - \widehat{\pi_0})\frac{\phi(\Phi^{-1}(q/2), \sigma = \widehat{\sigma_0})}{\phi(\Phi^{-1}(q/2), \sigma = 1)}, \tag{34}$$

where ϕ is the normal density function with SD σ. If P, Q have a known dependence under H_0^P, an alternative distribution can be used for computing $v(L)$ (see Supporting Information Section 1.2). The PDF f_0^q could be estimated in other ways; for example, a kernel density estimate (Sheather and Jones, 1991).

Type-1 error control is maintained under a relaxation of the assumption that $P|H_0^P \sim U(0,1)$ if the distribution of P dominates $U(0,1)$; that is, if $Pr(P \leq p|H_0^P) \leq p$ for all p, since we will overestimate f_0 inside L-regions in this case and hence v-values will be conservative.

3.3 | Correlation between v-values

Decision rules based on multiple p-values generally require adjustment if p-values are dependent (e.g. Benjamini and Hochberg, 1995). If v-values are obtained by the leave-one-out procedure (25) they are slightly pairwise dependent. The dependence is small; if $X' = X - (p_i, q_i) - (p_j, q_j)$ the values $v_{X'}(p_i, q_i)$, $v_{X'}(p_j, q_j)$ are independent, so the pairwise dependence between v-values corresponding to (p_i, q_i), (p_j/q_j) only arises from the differences $v_{X'+(p_j,q_j)}(p_i, q_i) - v_{X'}(p_i, q_i), v_{X'+(p_i,q_i)}(p_j, q_j) - v_{X'}(p_j, q_j)$; that is, the effect of a single point $((p_j, q_j), (p_i, q_i)$, respectively) on the map $v_{X'}$ defined by $|X'| = |X| - 2$ points. Indeed, we show that the expected change to v-values on adding a single new point is small:

Theorem 3.5. *Suppose we add a point (p^*, q^*) to a set of n points (p_i, q_i), considered as realisations of P, Q and conditions are satisfied for convergence of v-values as above. Let $\Delta v(L(\alpha))$ be the shift in a v-value corresponding to an L-curve $L(\alpha)$ after adding (p^*, q^*). Then*

$$E_{\alpha \sim U(0,1)}(|\Delta v(L(\alpha))|) = O\left(\frac{1}{n^2}\right). \tag{35}$$

The proof is given in Appendix A.4. When v-values are defined using block-out as in (26), v-values are independent within-block but dependent between blocks. The B-H procedure is also sensitive to higher order (non-pairwise) dependence between v-values, but we show by simulation in Section 5 residual dependence does not generally lead to failure of FDR control, even when we increase dependence by enforcing correlation between observations p_i, p_j and between q_i, q_j.

3.4 | Algorithm

We can now present our final algorithm.

We can interpret v_i as 'the probability that a randomly chosen (p, q) pair has a more extreme $\widehat{cFDR}$ value than $\widehat{cFDR}(p_i, q_i)$'; that is, as a p-value. This allows straightforward FWER or FDR control, especially as v-values are almost independent. The v-values order hypotheses such that a rejection rule {reject $H_0^P r(i)$ if $v_i \leq \alpha$} has near-optimal power, in

ALGORITHM 1 Controlling type-1 error rate in cFDR

	Input: 'principal' p-values $p_1, p_2, \dots p_n$; 'conditional' p-values $q_1, q_2, \dots q_n$; optionally fold assignment $b : 1 \dots n \to 1 \dots k$ such that $(p_i, q_i) \perp\!\!\!\perp (p_j, q_j) | b(i) = b(j)$
	Output: v-values $v_1, v_2 \dots v_n$
1:	Identify the set $\{q_i : p_i > 1/2\}$ and make estimates $\widehat{\pi_0}, \widehat{\sigma_0}$ of π_0, σ_0 as per (33)
2:	Set $\widehat{f}_0(p, q)$ as per Equation (34)
3:	**for** $i \in 1 \dots n$ **do**
4:	Set $S' = \{(p_j, q_j) : j \neq i\}$ (leave-one-out) or $S' = \{(p_j, q_j) : b(j) \neq b(i)\}$ (block-out)
5:	Find $c_i = \min\{c : (p_i, q_i) \in L_{S'}(c)\}$
6:	Set $v_i = \int_{L_{S'}(c_i)} \widehat{f}_0(p, q) dp dq$
7:	Return $(v_1, v_2, \dots, v_i)$

terms of corresponding to near-optimal forms for rejection regions. Much of our methodology can also be used if values q_i are not p-values for some second trait, as long as they fall in (0,1). However, approximation (33) may be inappropriate if this is not the case.

4 | RELATION TO OTHER METHODS

A wide range of approaches have been proposed for the problem of high-dimensional association testing using an informative covariate. Given the correspondingly wide variation in problems of this type, the optimal method is likely to depend on circumstance. In general, we will take P, p, H_0^p to refer to p-values and hypotheses for the trait of primary interest, and Q, q to refer to the covariate.

4.1 | Determination of rejection region form

The simplest approach to covariate-based testing is 'independent filtering' (Bourgon et al., 2010) in which attention is restricted to the set $\{(p_i, q_i) : q_i \geq q_0\}$, with the B-H procedure then applied to the corresponding subset of values of p_i. This procedure is equivalent to rejection regions which are a series of rectangles with upper border at $q = q_0$. Independent filtering is clearly non-optimal, but is well-suited to some problem types (Bourgon et al., 2010).

As discussed above, a range of approaches aim to approximate the optimal rejection regions based on f_0/f. In Du et al. (2014) and Lei and Fithian (2018), parameterisation leads to rejection regions constricted to a particular parametric class; in Du et al. (2014) that of oracle rejection regions under mixture-Gaussian forms of f_0 and f. In Alishahi et al. (2016) and Zablocki et al. (2014), boundaries of rejection regions are necessarily smooth at a scale corresponding to the smoothing kernel width, but can take otherwise arbitrary forms. An alternative approach is to 'bin' covariates (Ferkingstad et al., 2008; Ignatiadis et al., 2016) which leads to L-curves which are step functions with steps spaced according to the resolution of the bins.

An approach in Li and Barber (2016) estimates $Pr(H_0^p | Q = q)$ for each q to modulate a B-H type test for each observation. The entire effect of the covariate in this method is encompassed through the value of $Pr(H_0^p | Q = q)$, which necessarily relies on point-estimates of the PDF $f(Q = q | H_0^p)$, and hence the method is dependent on the accuracy of this estimate.

Another common approach to covariate modulation is the weighted B-H procedure (Benjamini et al., 2006), in which each p-value p_i is reweighted to a value p_i/w_i (where $\sum w_i = 1$) and the standard B-H procedure is then applied to the values p_i/w_i. Our method can be interpreted in these terms, setting $w_i = v_i/p_i$, but this is rather unnatural; there is no clear way to interpret what the ratio v_i/p_i means, and this approach does not make use of the 'p-value property' in Equation (20).

The use of ECDFs to generate rejection regions has the advantage of making use of the global distribution of P, Q, while spline- and kernel-density based estimates can generally only use local observations. The cFDR-based method has the obvious disadvantage of not converging to the optimum rejection region, and it can be less powerful than parametric approaches if parametric assumptions hold. However, using CDFs rather than PDFs allows faster convergence of rejection

regions with n, and this favours the cFDR approach if n is small, particularly if CDF- and PDF- based regions are similar and PDFs are difficult to model well.

Under certain circumstances, contours of CDF- and PDF-based methods are similar. A precise statement, proof and demonstration is given in Appendix A.5.

4.2 | Censoring of points

In general terms, the process determining a decision rule to be used on observation (p_i, q_i) cannot easily make use of the datapoint (p_i, q_i) itself, since the use of the point biases the choice of decision rule in some way. Approaches by Du et al. (2014) and Alishahi et al. (2016) censor the points used in the decision rule to those already rejected in a stepwise approach, and a method in Lei and Fithian (2018) masks the information available for the decision rule by effectively adding the point $(1 - p_i, q_i)$ to the dataset.

Since cFDR uses the entire dataset to estimate ECDFs, complex censoring can require that the cFDR estimator be changed in a non-trivial way. In particular, there is no obvious way to apply the methods proposed by Alishahi et al. (2016) or Lei and Fithian (2018). We propose avoiding the problem by leaving out the point (p_i, q_i) directly (Equations (25) and (26)), at the cost of residual correlation in resultant v-values. While crude, this corresponds to a near-minimum censorship of points, and the resultant correlation tends to be small enough to ignore (see Appendix A.4).

4.3 | Asymmetry and management of extreme outliers

An important property of the cFDR-based method is asymmetry, in that H_0^p cannot generally be rejected based on a low q_i alone (this can be seen by noting that $\widehat{cFDR}(p_i, q_i) \geq p_i$, and p_i can only exceed $\widehat{cFDR}^n$ in rare circumstances). Parametric approaches such as those in Du et al. (2014) and Lei and Fithian (2018) are not generally robust to this; for example, in Du et al. (2014), an extremely low q_i could lead to rejecting H_0^p even if the corresponding p_i were close to 1 and P, Q were independent (given that the degree of dependence is estimated). This property of the cFDR is very important when p_i and q_i are derived from genome-wide association studies (GWAS) on different diseases; it is entirely possible and even expected that a very strong association in the conditional trait is not an association with the principal trait. This property also differentiates our approach from meta-analysis of two sets of p-values.

4.4 | Relation to original FDR-controlling method for cFDR

In a paper in 2015 (Liley and Wallace, 2015), we identified the problem of failure of FDR control at α when using a rejection rule $\widehat{cFDR} \leq \alpha$ and proposed a rough solution. We proposed identifying L-curves and estimating f_0 as above, and for each L-region $L_S(\alpha^*)$, identifying a rectangle $R(\alpha^*)$ contained within it with vertices $(0,0)$, $(0, q_r)$, $(p_r, 0)$, (p_r, q_r). Since $R(\alpha^*) \subseteq L_S(\alpha^*)$, we have

$$Pr((P,Q) \in L_S(\alpha^*)) \geq Pr((P,Q) \in R(\alpha^*)) \tag{36}$$

and $\widehat{(p_r, q_r)} \leq \alpha^*$, so the FDR associated with rejecting any (p, q) pairs falling in $L_S(\alpha^*)$ was approximately

$$E\left(\frac{|\{i \,:\, p_i, q_i \in L_S(\alpha^*), H_0^p\}|}{\min(|\{i \,:\, p_i, q_i \in L_S(\alpha^*)\}|, 1)}\right) \approx \frac{Pr((P,Q) \in L_S(\alpha^*)|H_0^p)}{Pr((P,Q) \in L_S(\alpha^*))}$$

$$\leq \frac{Pr((P,Q) \in L_S(\alpha^*)|H_0^p)}{Pr((P,Q) \in R(\alpha^*)|H_0^p)} \frac{Pr((P,Q) \in R(\alpha^*)|H_0^p)}{Pr((P,Q) \in R(\alpha^*))}$$

$$\approx \frac{Pr((P,Q) \in L_S(\alpha^*)|H_0^p)}{Pr((P,Q) \in R(\alpha^*)|H_0^p)} \widehat{cFDR}_S(p_r, q_r) \tag{37}$$

TABLE 1 Variables used in simulations

Variable	Description	Sampling distribution
n	Total number of variables	$10^{U(3,4)}$ (rounded)
n_1^{pq}	Number of variables assoc. with P, Q	$U(0, 200)$ (rounded)
n_1^p	Number of variables associated with P	$U(0, 200)$ (rounded)
n_1^q	Number of variables associated with Q	$U(0, 200)$ (rounded)
s_p	Scale for distribution of P (see below)	$U(\frac{3}{2}, 3)$
s_q	Scale for distribution of Q	$U(\frac{3}{2}, 3)$
d	Form of alternative distributions	1: Normal, 2: t (3df), 3: Cauchy, eq. prob.

$$\leq \frac{\int_{L_S(\alpha^*)} f_0 \, dp \, dq}{\int_{R(\alpha^*)} f_0 \, dp \, dq} \alpha^*. \tag{38}$$

To approximately control FDR at α, our procedure found α^* so that expression (38) was $\leq \alpha$ and rejection H_0^p whenever $(p_i, q_i) \in L_S(\alpha^*)$.

As well as being approximate, this procedure was conservative due to inequality (36). Our new method avoids this conservative assumption and is on firmer theoretical ground. Furthermore, our old method precluded use of $\widehat{cFDR}^n$ given approximation (37). We show by simulation below that this results in substantial improvement in power in our new method.

5 | ASSESSMENT OF PERFORMANCE

In this section, we address five main points. Firstly, we demonstrate that our new method controls type-1 error rate (FDR) appropriately and that the censoring approach of (25) and (26) is necessary. Secondly, we demonstrate that power is substantially improved relative to our previous method for fixed level of FDR control, and that use of $\widehat{cFDR}^n$ over $\widehat{cFDR}$ improves power further. We then demonstrate that in settings where parametric assumptions are not satisfied, rejection regions based on $\widehat{cFDR}^n$ can correspond to a more powerful procedure than rejection regions based on alternative CDF or PDF estimators. We examine the effect of correlation between observations p_i, q_i on our main FDR-controlling methods and demonstrate that the disadvantage of using our leave-one-out method (Equation (25)) instead of the leave-out-block method (Equation (26)) out method in the presence of correlation is loss of power rather than loss of FDR control. Finally, we assess the degree of shared association between P and Q which is necessary for our method to give an advantage over p-values alone.

In each simulation, we sampled a set of values $S = (p_i, q_i)$, $i \in 1 \dots n$. The sampling schema we used itself depended on a series of underlying parameters, which were themselves sampled from a joint distribution specified in Table 1, or a conditional distribution of it. We also separately considered several fixed values of parameters.

We first chose a fixed total number of hypotheses n, then split these into four classes of fixed size: C_1 of size n_1^{pq} associated in both P and Q, C_2 of size n_1^p associated only with P, C_3 of size n_1^q associated only with Q and C_4 associated with neither P nor Q. Within each class, samples (p_i, q_i) were identically distributed.

For $i \in C_1, C_2$, we sampled p_i (determined by d, s_p) by first simulating Z scores:

$d = 1: \ -\Phi^{-1}(\frac{p_i}{2})\frac{1}{s_p} \sim N(0, 1)$

$d = 2: \ -\Phi^{-1}(\frac{p_i}{2})\frac{1}{s_p} \sim t(\text{df} = 3, \text{ncp} = 0)$

$d = 3: \ -\Phi^{-1}(\frac{p_i}{2})\frac{1}{s_p} \sim \text{Cauchy}(\text{location} = 0, \text{scale} = 1),$

where $-\Phi^{-1}(\frac{p_i}{2})$ can be considered a Z-score corresponding to p_i, and s_p a scaling factor for the distribution. We set the distribution of $q_i \in C_1, C_3$ similarly, with s_q in place of s_p. The values p_i, q_i for $i \in C_4$ were sampled from $U(0, 1)$.

Although effect sizes are often assumed to follow normal distributions, real data are often noisier, with longer tails, and recent work suggests non-normal distributions may be a better fit in the case of GWAS data (Walters et al., 2019). We chose the alternative distributions (normal, t (3df), and Cauchy) to span behaviours from 'well-behaved' (normal) to 'long-tailed' (t) to 'chaotic' (Cauchy) to survey a wider range of possibilities.

Samples p_i, p_j and q_i, q_j were generally independent (unless otherwise specified), but we also sampled under two patterns of dependence. Firstly, we simulated a 'block' correlation structure in which we divided samples into three blocks, and within each block sampled z-scores $z_{p_i}, z_{p_j}, z_{q_i}, z_{q_j}$ corresponding to p_i, p_j and q_i, q_j such that $\mathrm{cor}(z_{p_i}, z_{p_j}) = \mathrm{cor}(z_{q_i}, z_{q_j}) = \rho$ if i, j were in the same block and class, and $\mathrm{cor}(z_{p_i}, z_{p_j}) = \mathrm{cor}(z_{q_i}, z_{q_j}) = 0$ otherwise. Secondly, we simulated an equicorrelation structure in which $\mathrm{cor}(z_{p_i}, z_{p_j}) = \mathrm{cor}(z_{q_i}, z_{q_j}) = \rho$ whenever i and j were in the same class. When $d \in 2, 3$, we used the off-diagonal elements of the normalised dependence matrix in the multivariate T distribution in place of correlation.

When relevant, we also sampled parameters from the distribution specified in Table 1 conditional on $n_1^p + n_1^{pq} = 0$; that is, no associations with P. We plotted results from these simulations separately to those with parameters drawn from the unconditional distribution.

Given a rejection procedure, we defined

$$
FDP = \begin{cases} 0 & \text{if no rejections} \\ \dfrac{\text{number of falsely rejected null hypotheses}}{\text{total number of rejections}} & \text{if} \geq 1 \text{ rejection} \end{cases}
$$

$$
TDP = \begin{cases} 0 & \text{if no rejections} \\ \dfrac{\text{number of correctly rejected null hypotheses}}{\text{true number of associations}} & \text{if} \geq 1 \text{ rejection.} \end{cases}
$$

We analysed type 1 error in terms of the estimated FDR, $FDR = E(FDP) \approx \overline{FDP}$ and power in terms of the corresponding true-discovery rate $TDR = E(TDP) \approx \overline{TDP}$. We compared FDP and TDP between samples by estimating them via a Gaussian-weighted moving average across the independent variable (usually $n_1^p + n_1^{pq}$). We show 95% pointwise confidence envelopes derived as per Gatz and Smith (1995), except in cases where such envelopes obstruct viewability of the plot. In these cases, we state that values $TDR(A)$ for one method A 'exceed' paired values $TDR(B)$ of another method B if in at least six of eight equal subdivisions of the x-axis range the following three conditions hold: $TDR(A) > TDR(B)$ more than $TDR(B) > TDR(A)$, the mean $\overline{TDR(A) - TDR(B)}$ is positive and a Wilcoxon test of ranks on $TDR(A) - TDR(B)$ rejects the null hypothesis of a symmetric distribution around 0 with $p < 5 \times 10^{-3}$ (or the equivalent with FDR in place of TDR). Each line on each plot corresponds to > 5000 simulation runs. All raw simulation results and analytic code are publically available at https://github.com/jamesliley/cfdr_pipeline.

5.1 | New FDR-controlling procedure leads to greater power than previous method, and adjustment improves power further

We first compared FDR control amongst five methods, aiming to control the FDR at either $\alpha = 0.1$ or $\alpha = 0.01$:

1. the B-H method applied to the values p_i, labelled 'P-val'
2. the B-H method applied to 'naive' v-values $v(p_i, q_i) = v_S(p_i, q_i)$ as per Equation (27) for reference, labelled 'Naive'
3. the B-H method applied to 'leave-one-out' v-values $v(p_i, q_i) = v_{S-(p_i,q_i)}(p_i, q_i)$ as per Equation (25), labelled 'LOO'
4. the B-H method applied to block-out v-values (after randomly separating observations into three equally sized subdivisions, so (p_i, q_i) is in subdivision $b(i)$, defining v-values $v_{S-b(i)}(p_i, q_i)$) as per Equation (26), labelled 'LOB'
5. our previous method for FDR control applied to (p_i, q_i), labelled 'Orig.'

We sampled simulation parameters according to Table 1 or the corresponding conditional distribution of Table 1 with $n_1^p + n_1^{pq} = 0$.

Expected FDP was consistent with the FDR control level when using leave-one-out v-values or 'block-out v-values (rejection procedures 3,4). When using the 'naive' v-values $v_S(p_i, q_i)$ (rejection procedure 2), FDR was not controlled at the requisite level. The FDR using methods 3 and 4 exceeded the FDR of our original method (rejection procedure 5),
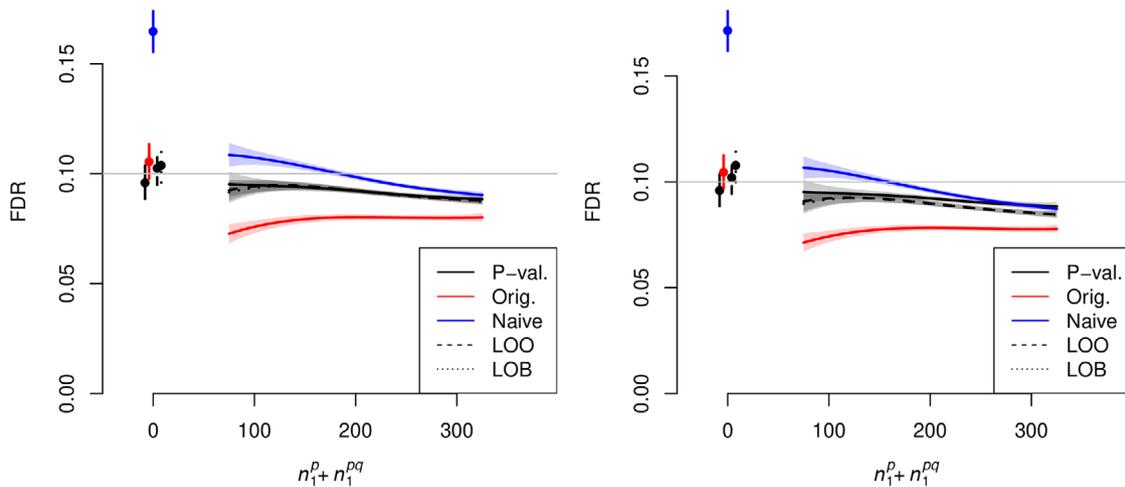
FIGURE 2 FDR control of various methods against $n_1^p + n_1^{pq}$, the total number of variables associated with P (the primary study under consideration). The horizontal line shows $\alpha = 0.1$, the desired FDR control level. Simulations in the left panel integrate L-regions over the true distribution f_0; simulations in the right panel integrate over the estimated distribution as per Equation (33). Shaded regions indicate 95% confidence envelopes. Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range. Lines on the left indicate FDR control with $n_1^p + n_1^{pq} = 0$. A corresponding plot with $\alpha = 0.01$ is shown in Supplementary Figure 7
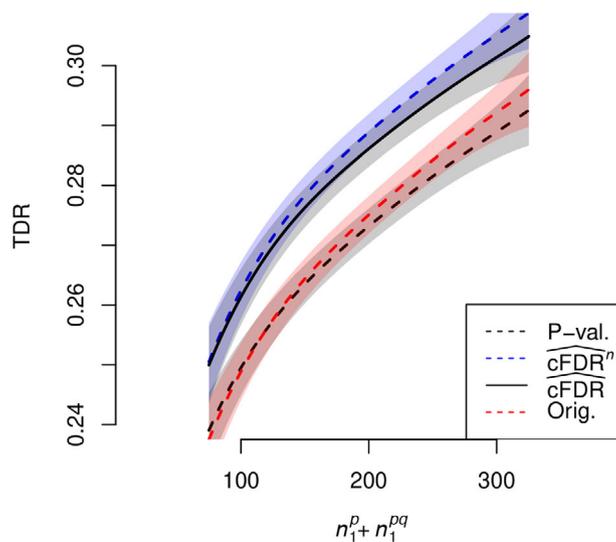
FIGURE 3 TDR of various methods against $n_1^p + n_1^{pq}$, the total number of variables associated with P (the primary study under consideration), at FDR control level $\alpha = 0.1$. Shaded areas show 95% pointwise confidence envelopes. A corresponding plot with $\alpha = 0.01$ is shown in Supplementary Figure 8. Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range

indicating that our original method was conservative. FDR control was maintained when using the approximation of f_0 in Equation (33). Results are shown in Figure 2.

Having established the validity of rejection methods 3 and 4, we compared the power of 'adjusted' cFDR ($\widehat{cFDR}^n$) and non-adjusted cFDR ($\widehat{cFDR}$) using the leave-one-out v-value (rejection procedure 3) and the power of our previous method, rejection procedure 5, applied to $\widehat{cFDR}$ (labelled 'Orig'). The TDR of $\widehat{cFDR}^n$ exceeded the power of $\widehat{cFDR}$, which in turn exceeded the power of our previous rejection procedure on $\widehat{cFDR}$ (Figure 3).

We report FDR and TDR for p-values, $\widehat{cFDR}^n$ and oracle cfdr for a range of fixed simulation parameters in Table 2

TABLE 2 FDR and TDR of p-value, $\widehat{cFDR}^n$, and oracle cfdr (best possible procedure) using leave-one-out v-values (Equation (25)) for a range of simulation parameters, controlling FDR at $\alpha = 0.1$. All descriptions are relative to 'Reference,' which has the following parameter values: $n = 5000$, $n_1^p = n_1^q = n_1^{pq} = 100$, $s_p = s_q = 2$, $d = 2$. TDR is undefined if $n_1^p + n_1^{pq} = 0$. 'Negative information' means *fewer*-than random shared associations, rather than more ($n_1^p = n_1^q = 2000$, $n_1^{pq} = 0$). Complete parameter values, confidence intervals and more detailed results are shown in Supplementary Table 1

Description	TDR(P)	FDR(cFDR)	TDR(cFDR)	TDR(oracle)
Reference	0.194	0.0955	0.208	0.212
No effects		0.0973		
Weak effects	0.00803	0.0814	0.00795	0.00904
Large variance in effect sizes	0.493	0.0957	0.517	0.527
Larger n	0.173	0.0995	0.189	0.193
Smaller n	0.26	0.0795	0.265	0.269
No non-null shared hypotheses	0.188	0.102	0.178	0.187
All non-null hypotheses shared	0.195	0.0963	0.26	0.309
Negative information	0.302	0.0585	0.314	0.319
Block correlation	0.2	0.0974	0.213	0.217
Equicorrelation	0.191	0.0897	0.205	0.209
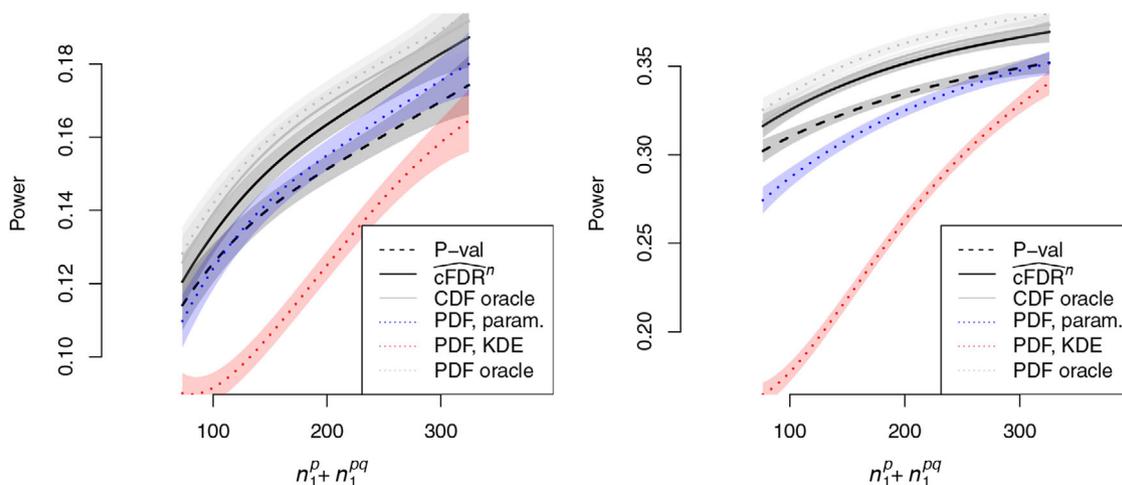
FIGURE 4 TDR of PDF-based methods against $n_1^p + n_1^{pq}$, the total number of variables associated with P (the primary study under consideration), controlling FDR at $\alpha = 0.1$. In the left panel, parametric assumptions were satisfied (i.e. $d = 1$ in Table 1) and in the right panel, they are not ($d = 2, 3$). Shaded regions show pointwise 95% confidence intervals. A corresponding plot with $\alpha = 0.01$ is shown in Supplementary Figure 8. Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range

5.2 | PDF-based estimation leads to a less powerful procedure than CDF-based estimation

As $n \to \infty$, consistent estimators of $Pr(H_0^P | P = p, Q = q)$ will converge to optimal rejection regions while estimators of $Pr(H_0^P | P \leq p, Q \leq q)$ will not, and hence the former will ultimately be more powerful. However, we found that under the distribution of simulation parameters in Table 1, the ECDF-based estimator $\widehat{cFDR}^n$ is considerably more powerful than two PDF-based estimators of $Pr(H_0^P | P = p, Q = q)$.

Results are shown in Figure 4. We considered parametric (labelled 'PDF param') and KDE-based (labelled 'PDF KDE') estimators of $Pr(H_0^P | P = p, Q = q)$. The parametric model was based on a four-Gaussian model detailed in Supporting Information, Section 1.4. We separated cases in which parametric assumptions were satisfied (i.e. $d = 1$ in Table 1) and in which they were not ($d = 2, 3$). The TDR of $\widehat{cFDR}^n$ exceeded the TDR of both estimators of $Pr(H_0^P | P = p, Q = q)$. The performance of an oracle CDF procedure (using exact contours of F_0/F as rejection regions, labelled 'CDF oracle') and an oracle PDF procedure (using exact contours of f_0/f as rejection regions, labelled 'PDF oracle') are shown for comparison.
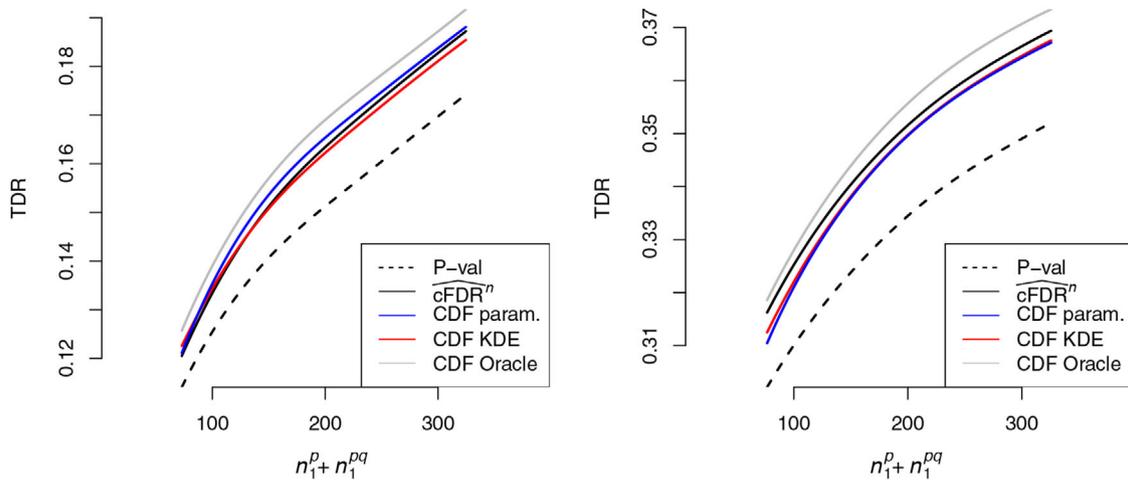
FIGURE 5 TDR of various methods against $n_1^p + n_1^{pq}$, the total number of variables associated with P (the primary study under consideration), restricting to simulations in which parametric assumptions were satisfied (left panel) or were not satisfied (right panel), at FDR control level $\alpha = 0.1$. A corresponding plot with $\alpha = 0.01$ is shown in Supplementary Figure 9. Confidence intervals are omitted for visual clarity Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range

5.2.1 | Parametric- and KDE-based cFDR estimators are less powerful than the ECDF-based estimator

We also examined PDF- and KDE-based estimates of the cFDR rather than the cfdr. Details of the alternative estimators are given in Supporting Information Section 1.4.

When parametric assumptions were satisfied (Figure 5, left panel), performance of the ECDF-based $\widehat{cFDR}^n$, parametric (labelled 'CDF param') and KDE-based (labelled 'CDF KDE') cFDR estimators was equivocal. When parametric assumptions were not satisfied ($d = 2, 3$ as per Table 1; Figure 5, right panel), the TDR of the ECDF estimator exceeded the TDR of the parametric and KDE estimators. The performance of an oracle CDF procedure (using exact contours of F_0/F as rejection regions) is shown for comparison.

5.3 | Correlated samples lead to lower TDR but FDR control is maintained

When p_i, q_i had either block correlation or equicorrelation, we found that FDR control was maintained when using leave-one-out v-values (Equation (25)) and when using leave-out-block v-values (Equation (26)). Under equicorrelation, the TDR of leave-out-block exceeded the TDR of leave-one-out v-values. Under block correlation, although TDR of leave-out-block did not formally exceed the TDR of leave-one-out, a paired Wilcoxon rank-sum test on TDR values rejected the null hypothesis of a symmetric distribution around 0 with $p < 1 \times 10^{-6}$ in favour of leave-out-block.

Maintenance of FDR control is expected, as the B-H procedure controls FDR more conservatively when p-values are positively correlated than when independent. Figure 6 shows FDR and TDR controlling at $\alpha = 0.1$ in the case $\rho = 0.01$, including performance of p-values p_i under the B-H procedure for comparison. The case $\rho = 0.1$ is similar and is shown in Supplementary Figure 11.

5.4 | TDR of $\widehat{cFDR}^n$ becomes higher than TDR of p-value alone when $\approx 20\%$ of hypotheses are shared

Finally, we assessed the proportion of non-null hypotheses for P which needed to be shared with Q in order for $\widehat{cFDR}^n$ to have an advantage in TDR over only considering P. When no non-null hypotheses are shared, the values q_i confer no information on H_0^P, so we expect that use of v-values will add only noise and the TDR of the p-value will be larger than that of the v-value. We plotted the difference in TDR between $\widehat{cFDR}^n$ (using leave-one-out v-values) and p-values p_i against
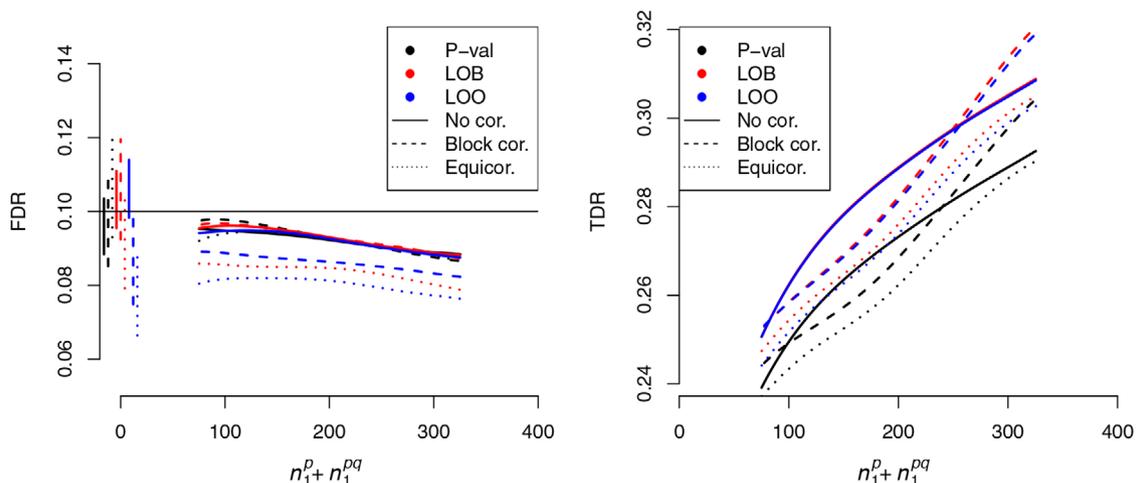
FIGURE 6 FDR (left) and TDR (right) of FDR-controlling methods leave-out-block (Equation (26)) and leave-one-out (Equation (25)) applied to $\widehat{cFDR}^n$, and the B-H procedure applied to p-values, under several models of correlation between observations ($\rho = 0.01$). Confidence envelopes are omitted for visual clarity. Vertical lines show FDR with 95% confidence intervals at $n_1^p + n_1^{pq} = 0$ (the p-value appears not to control FDR under equicorrelation, but it is well-known to do so theoretically). A corresponding figure with $\rho = 0.1$ is shown in Supplementary Figure 11. Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range
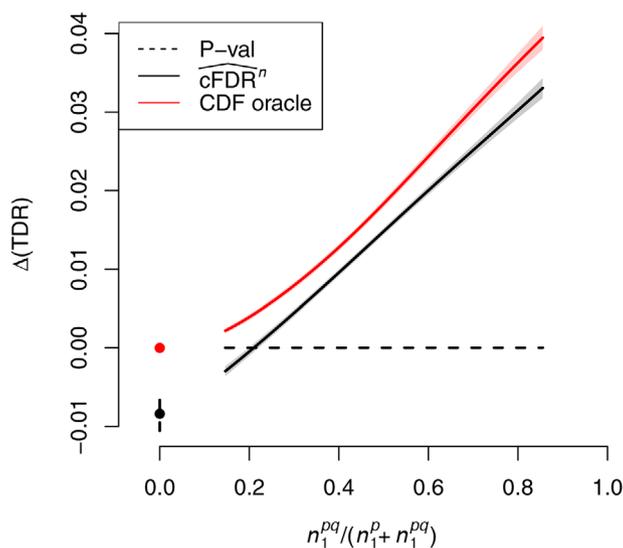
FIGURE 7 Difference in TDR between $\widehat{cFDR}^n$ (assessed by leave-one-out v-values) and p-values, controlling FDR at $\alpha = 0.1$, against $n_1^{pq}/(n_1^{pq} + n_1^p)$ (proportion of non-null hypotheses for P which are shared with Q). The performance of the oracle CDF method is shown for comparison. Shaded areas show pointwise 95% confidence intervals. Points and lines at the leftmost edge show TDR values and 95% confidence intervals when $n_1^{pq}/(n_1^{pq} + n_1^p) = 0$. A corresponding figure with $\alpha = 0.01$ is shown in Supplementary Figure 12. Curves show moving weighted averages using a Gaussian kernel with SD 15% of the x axis range

$n_1^{pq}/(n_1^{pq} + n_1^p)$ and found that this difference became positive when around 20% of hypotheses were shared (Figure 7). This figure is dependent on our simulation parameters: a smaller percentage of hypotheses may lead to an advantage of $\widehat{cFDR}^n$ if, for instance, effect sizes were very large.

5.5 | Iterated cFDR

Since our proposed method for type-1 error rate control maps p-value/covariate pairs to v-values preserving the p-value property, we are free to use the resultant v-values in a second cFDR-based analysis against a second covariate. This enables
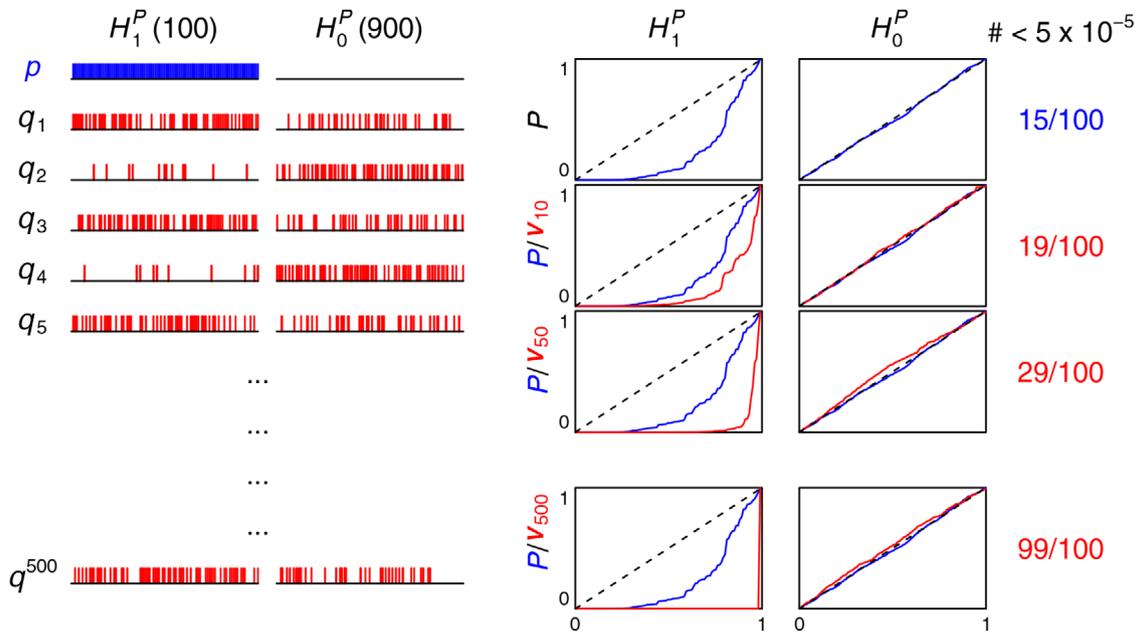
FIGURE 8 Iterated cFDR. The left part of plot shows where non-null hypotheses fall in p/q^j (blue vertical lines for p, red for q_j). Non-null hypotheses are shared more-than-randomly in only every second set q^j. The right part shows p/v_j values (blue/red lines, respectively) plotted in an ascending order under H_1^p, H_0^p, and the number of p/v_j values which reach the Bonferroni-corrected significance

immediate and simple adaptation to a setting in which more than one set of covariates are available. In our motivating example, this would allow us to subsequently 'condition' on other potentially related diseases as well as OCA.

We simulated a set of p-values $\{p\} = \{p_i, i \in 1 \dots 1000\}$, with 100 true associations (H_1^p) in which p-values were sampled from $2\Phi(-|N(0, 3^2)|)$ (where Φ is the normal CDF) and 900 non-associations (H_0^p) in which p-values were sampled from $U(0,1)$. We then similarly simulated sets of covariates $\{q^j\} = \{q_i^j, i \in 1 \dots 1000\}$ with 100 true associations (H_1^q), which for even j were randomly spaced amongst the 1000 variables (uninformative covariates) and for odd j overlapped more-than-randomly with associations with principal p-values (informative covariates), with around 54 shared associations on average (strictly, such that $Pr(H_1^q|H_1^p) = 15Pr(H_1^q|H_0^p)$).

Starting with $v_0 = p$, we conditioned on each set of $\{q^j\}$ in succession, so $v_{i+1} = v(v_i, q^i)$. We used $\widehat{cFDR}^n$ as an estimator and used leave-one-out v-values. Originally 19 of 100 null hypotheses were correctly rejected using p alone ($p < 5 \times 10^{-5} = 0.05/1000$). On repeated conditioning, almost all v-values when $H^p = 1$ tended to 0: 99 null hypotheses were correctly rejected using v_{500}. Under H_0^p, v-values remained uniform on (0,1) (Figure 8). This indicated the potential to greatly strengthen the power of a high-dimensional association analysis by repeated conditioning in this manner, even when only half of the sets of covariates are informative.

5.6 | Summary of BRCA analysis

Finally, we return to the motivating example. cFDR rejects more null hypotheses for BRCA (724) than B-H on BRCA data alone (678, Figure 1a) or the subset of variables with OCA association (280, Figure 1b). The procedure is asymmetrical in that it will not reject a BRCA null hypothesis for a low OCA p-value alone and can readily be reversed: Supplementary Figure 6 shows a similar analysis analysing association with OCA.

6 | DISCUSSION

We present an improvement to the cFDR method, a widely used procedure in genetic discovery. Our new methods essentially involve computing an analogy of the p-value corresponding to the ranking of hypotheses defined by the cFDR estimator. Our method enables the cFDR to be used definitively in the discovery phase of -omics studies with control of a

type-1 error rate. The general procedure of multiple p-value testing with a covariate has wide scientific application; see Du et al. (2014), Alishahi et al. (2016) and Li and Barber (2016) for example.

The $\widehat{cFDR}$ and $\widehat{cFDR}^n$ estimators make no distributional assumptions on P, Q. The type-1 error rate controlling method requires modelling of the PDF of $P, Q|H_0^p$, but this requires approximating a univariate PDF $Q|H_0^p$. Furthermore, this PDF is only used as an integrand rather than for direct point estimates. It is reasonable to expect that for approximations to complex PDFs, relative average errors over intervals will be smaller than relative errors at individual points; parametric approximations tend to be smoother than the true distribution at a fine scale, and KDE-based approximations rougher. An obvious shortcoming of cFDR-based methods is the lack of asymptotic optimality. Methods based on consistent estimators of f_0/f will eventually outperform any estimator of F_0/F for large enough n (see Supporting Information, Section 1.5). However, the ECDF-based cFDR estimator was far stronger than PDF-based estimators at the values of n we simulated at ($10^3 - 10^4$). In practical terms, it is important to note that n, being the number of variables, cannot generally be increased indefinitely, as opposed to, for instance, sample size. Essentially, the fitting of L-curves corresponds to a procedure by which the similarity between P, Q is assessed, and the degree of modulation when moving from p to v values corresponds to this similarity. Moreover, this assessment of similarity occurs intrinsically on the basis of the joint CDF rather than relying on a parametric description.

L-curves may not change monotonically with Q; that is, it may be possible to reject a null hypothesis with p-values (p, q_1) and not reject a null hypothesis with p-values (p, q_2), $q_2 < q_1$ (see the lower-left panel of Figure 1). It would be possible within our framework of FDR control (Theorem 3.1) to force L-curves to be monotonic with Q, and indeed since this would result in straight-up-and-down segments on L-curves, the loss of power due to noise when P and Q are unrelated (Figure 7) may be reduced in this case. However, non-monotonicity of L-curves is potentially advantageous in a biostatistical setting. Between TWAS or GWAS for similar diseases, it may be the case that shared non-null hypotheses have 'moderately' small p-values, corresponding to common general shared medium-risk pathological causes, but non-shared non-null hypotheses have 'extremely' small p-values, corresponding to specific high-risk pathologies. Non-monotonic L curves allow this effect to be modelled. Unrestricted L-curves also allow use to be made of q-values such that $Pr(H_0^P|Q = q)$ is lower with low q, rather than higher: we show this in Table 2, row 'negative information'.

Our proposed 'iterated cFDR' procedure can be thought of as a meta-analysis of a series of experiments $E_P, E_{Q_1}, E_{Q_2}, \ldots$ giving rise to p-value sets $\{p_i\}, \{q_i^1\}, \{q_i^2\}, \ldots$ when only the first set ($\{p_i\}$) is known to test the correct hypotheses; that is, be $U(0, 1)$ for null hypotheses. It enables us to find the set of non-null hypotheses corresponding to E_P (denoted $\{H_1^P\}$), even though the set of non-null hypotheses corresponding to E_{Q_j} (denoted $\{H_1^{Q_j}\}$) may only partly overlap $\{H_1^P\}$, may contain hypotheses not in $\{H_1^P\}$ and (half the time) may even carry no information about H_1^P at all. This could be used to refine the set of association statistics $\{p_i\}$ for a disease of interest by using sets of association statistics $\{q_i^1\}, \{q_i^2\}, \ldots$ at the same variables for a range of separate traits. It could also be used to improve power when repeating an -omics study in a new ethnic group by levering on previous studies in different ethnicities.

In summary, our method improves the power of cFDR analyses and allows it to be used confidently in the setting of multiple hypothesis testing. This can enable more efficient use of data, and more information to be gained from the same datasets. Our method contributes to a set of tools for high-dimensional statistical analysis and has wide application across a range of fields in biomedicine and elsewhere.

7 | CODE AVAILABILITY

All functions necessary to apply the methods detailed in this work are available in the R package
https://github.com/jamesliley/cfdr
A full pipeline to generate the results in this paper is available in the git repository
https://github.com/jamesliley/cfdr_pipeline

CONFLICT OF INTEREST
None declared.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in TWAS-HUB at http://twas-hub.org/traits/.

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID
James Liley https://orcid.org/0000-0002-0049-8238
Chris Wallace https://orcid.org/0000-0001-9755-1703

REFERENCES
Alishahi, K., Ehyaei, A. R., & Shojaie, A. (2016). A generalized Benjamini–Hochberg procedure for multivariate hypothesis testing. arXiv. arXiv:1606.02386.

Andreassen, O. A., Djurovic, S., Thompson, W. K., Schork, A. J., Kendler, K. S., O'Donovan, M. C., Rujescu, D., Werge, T., van de Bunt, M., Morris, A. P., et al. (2013a). Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics*, *92*(2), 197–209.

Andreassen, O. A., Thompson, W. K., & Dale, A. M. (2013b). Boosting the power of schizophrenia genetics by leveraging new statistical tools. *Schizophrenia Bulletin*, *40*(1), 13–17.

Andreassen, O. A., Thompson, W. K., Schork, A. J., Ripke, S., Mattingsdal, M., Kelsoe, J. R., Kendler, K. S., O'Donovan, M. C., Rujescu, D., Werge, T., and Sklar, P. (2013). Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLOS Genetics*, *9*(4), e1005544.

Andreassen, O. A., Harbo, H. F., Wang, Y., Thompson, W., Schork, A., Mattingsdal, M., Zuber, V., Bettella, F., Ripke, S., Kelsoe, J., et al. (2015). Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Molecular Psychiatry*, *20*(2), 207.

Andreassen, O. A., McEvoy, L. K., Thompson, W. K., Wang, Y., Reppe, S., Schork, A. J., Zuber, V., Barrett-Connor, E., Gautvik, K., Aukrust, P., et al. (2014a). Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. *Hypertension*, *63*(4), 819–826.

Andreassen, O. A., Zuber, V., Thompson, W. K., Schork, A. J., Bettella, F., Consortium, P., GWAS, C., Djurovic, S., Desikan, R. S., Mills, I. G., et al. (2014b). Shared common variants in prostate cancer and blood lipids. *International Journal of Epidemiology*, *43*(4), 1205–1214.

Basu, P., Cai, T. T., Das, K., & Sun, W. (2018). Weighted false discovery rate control in large-scale multiple testing. *Journal of the American Statistical Association*, *113*(523), 1172–1183.

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 (Methodological)*, *57*(1), 289–300.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507.

Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*, *107*(21), 9546–9551.

Broce, I., Karch, C. M., Wen, N., Fan, C. C., Wang, Y., Tan, C. H., Kouri, N., Ross, O. A., Höglinger, G. U., Muller, U., et al. (2018). Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies. *PLoS Medicine*, *15*(1), e1002487.

Cai, T. T., Sun, W., & Wang, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. In Royal Statistical Society (Vol. 81, No. 2).

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, *39*(1), 1–38.

Desikan, R. S., Schork, A. J., Wang, Y., Thompson, W. K., Dehghan, A., Ridker, P. M., Chasman, D. I., McEvoy, L. K., Holland, D., Chen, C.-H., et al. (2015a). Polygenic overlap between c-reactive protein, plasma lipids, and Alzheimer disease. *Circulation*, *131*(23), 2061–2069.

Desikan, R. S., Schork, A. J., Wang, Y., Witoelar, A., Sharma, M., McEvoy, L. K., Holland, D., Brewer, J. B., Chen, C.-H., Thompson, W. K., et al. (2015b). Genetic overlap between Alzheimer's disease and Parkinson's disease at the mapt locus. *Molecular Psychiatry*, *20*(12), 1588.

Du, L., & Zhang, C. (2014). Single-index modulated multiple testing. *The Annals of Statistics*, *42*(4), 1262–1311.

Ferkingstad, E., Frigessi, A., Rue, H., Thorleifsson, G., & Kong, A. (2008). Unsupervised empirical Bayesian multiple testing with external covariates. *Annals of Applied Statistics*, *2*(2), 714–735.

Ferrari, R., Wang, Y., Vandrovcova, J., Guelfi, S., Witeolar, A., Karch, C. M., Schork, A. J., Fan, C. C., Brewer, J. B., Momeni, P., et al. (2017). Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *Journal of Neurology, Neurosurgery and Psychiatry*, *88*(2), 152–164.

Gatz, D. F., & Smith, L. (1995). The standard error of a weighted mean concentration–'I. Bootstrapping vs other methods. *Atmospheric Environment*, *29*(11), 1185–1193.

Genovese, C. R., Roeder, K., & Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika*, *93*(3), 509–524.

Greene, M. H., Clark, J. W., & Blayney, D. W., (1984). The epidemiology of ovarian cancer. *Seminars in Oncology*, *11*, 209–226.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Peninx, B. W., Jansen, R., De Geus, E. J., Boomsma, D. I., Wright, F. A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, *48*(3), 245.

Ignatiadis, N., Klaus, B., Zaugg, J. B., & Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, *13*(7), 577.

Karch, C. M., Wen, N., Fan, C. C., Yokoyama, J. S., Kouri, N., Ross, O. A., Höglinger, G., Müller, U., Ferrari, R., Hardy, J., et al. (2018). Selective genetic overlap between amyotrophic lateral sclerosis and diseases of the frontotemporal dementia spectrum. *JAMA Neurology*, *75*(7), 860–875.

Kiefer, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific Journal of Mathematics*, *11*(2), 649–660.

Le Hellard, S., Wang, Y., Witoelar, A., Zuber, V., Bettella, F., Hugdahl, K., Espeseth, T., Steen, V. M., Melle, I., Desikan, R., et al. (2016). Identification of gene loci that overlap between schizophrenia and educational attainment. *Schizophrenia Bulletin*, *43*(3), 654–664.

Lei, L., & Fithian, W. (2018). Adapt: An interactive procedure for multiple testing with side information. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(4), 649–679.

Li, A., & Barber, R. F. (2016). Multiple testing with the structure adaptive Benjamini–Hochberg algorithm. arXiv. arXiv:1606.07926.

Li, A., & Barber, R. F. (2017). Accumulation tests for FDR control in ordered hypothesis testing. *Journal of the American Statistical Association*, *112*(518), 837–849.

Liley, J., & Wallace, C. (2015). A pleiotropy-informed Bayesian false discovery rate adapted to a shared control design finds new disease associations from GWAS summary statistics. *PLOS Genetics*, *11*(2), e1004926.

Liu, J. Z., Hov, J. R., Folseraas, T., Ellinghaus, E., Rushbrook, S. M., Doncheva, N. T., Andreassen, O. A., Weersma, R. K., Weismüller, T. J., Eksteen, B., et al. (2013). Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nature Genetics*, *45*(6), 670.

Lv, W.-Q., Zhang, X., Zhang, Q., He, J.-Y., Liu, H.-M., Xia, X., Fan, K., Zhao, Q., Shi, X.-Z., Zhang, W.-D., et al., (2017). Novel common variants associated with body mass index and coronary artery disease detected using a pleiotropic cFDR method. *Journal of Molecular and Cellular Cardiology*, *112*, 1–7.

McLaughlin, R. L., Schijven, D., Van Rheenen, W., Van Eijk, K. R., O'Brien, M., Kahn, R. S., Ophoff, R. A., Goris, A., Bradley, D. G., Al-Chalabi, A., et al., (2017). Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nature Communications*, *8*, 14774.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92.

Phelan, C. M., Kuchenbaecker, K. B., Tyrer, J. P., Kar, S. P., Lawrenson, K., Winham, S. J., Dennis, J., Pirie, A., Riggan, M. J., Chornokur, G., et al. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nature Genetics*, *49*(5), 680.

Schork, A. J., Wang, Y., Thompson, W. K., Dale, A. M., & Andreassen, O. A., (2016). New statistical approaches exploit the polygenic architecture of schizophrenia 'implications for the underlying neurobiology. *Current Opinion in Neurobiology*, *36*, 89–98.

Scott, J. G., Kelly, R. C., Smith, M. A., Zhou, P., & Kass, R. E. (2015). False discovery rate regression: An application to neural synchrony detection in primary visual cortex. *Journal of the American Statistical Association*, *110*(510), 459–471.

Shadrin, A. A., Smeland, O. B., Zayats, T., Schork, A. J., Frei, O., Bettella, F., Witoelar, A., Li, W., Eriksen, J. A., Krull, F., et al. (2018). Novel loci associated with attention-deficit/hyperactivity disorder are revealed by leveraging polygenic overlap with educational attainment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(2), 86–95.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 683–690.

Smeland, O. B., Frei, O., Kauppi, K., Hill, W. D., Li, W., Wang, Y., Krull, F., Bettella, F., Eriksen, J. A., Witoelar, A., et al. (2017a). Identification of genetic loci jointly influencing schizophrenia risk and the cognitive traits of verbal-numerical reasoning, reaction time, and general cognitive function. *JAMA Psychiatry*, *74*(10), 1065–1075.

Smeland, O. B., Wang, Y., Frei, O., Li, W., Hibar, D. P., Franke, B., Bettella, F., Witoelar, A., Djurovic, S., Chen, C.-H., et al. (2017b). Genetic overlap between schizophrenia and volumes of hippocampus, putamen, and intracranial volume indicates shared molecular genetic mechanisms. *Schizophrenia Bulletin*, *44*(4), 854–864.

Smeland, O. B., Wang, Y., Lo, M.-T., Li, W., Frei, O., Witoelar, A., Tesli, M., Hinds, D. A., Tung, J. Y., Djurovic, S., et al. (2017c). Identification of genetic loci shared between schizophrenia and the big five personality traits. *Scientific Reports*, *7*(1), 2222.

Storey, J. D. et al. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics*, *31*(6), 2013–2035.

Van der Meer, D., Rokicki, J., Kaufmann, T., Córdova-Palomera, A., Moberget, T., Alnæs, D., Bettella, F., Frei, O., Doan, N. T., Sønderby, I. E., et al. (2020). Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Molecular Psychiatry*, *25*(11), 3053–3065.

Walters, K., Cox, A., & Yaacob, H. (2019). Using GWAS top hits to inform priors in Bayesian fine-mapping association studies. *Genetic Epidemiology*, *43*(6), 675–689.

Wang, Y., Bos, S. D., Harbo, H. F., Thompson, W. K., Schork, A. J., Bettella, F., Witoelar, A., Lie, B. A., Li, W., McEvoy, L. K., et al. (2016). Genetic overlap between multiple sclerosis and several cardiovascular disease risk factors. *Multiple Sclerosis Journal*, *22*(14), 1783–1793.

Witoelar, A., Jansen, I. E., Wang, Y., Desikan, R. S., Gibbs, J. R., Blauwendraat, C., Thompson, W. K., Hernandez, D. G., Djurovic, S., Schork, A. J., et al. (2017). Genome-wide pleiotropy between Parkinson disease and autoimmune diseases. *JAMA Neurology*, *74*(7), 780–792.

Zablocki, R. W., Schork, A. J., Levine, R. A., Andreassen, O. A., Dale, A. M., and Thompson, W. K. (2014). Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics*, *30*(15), 2098–2104.

Zuber, V., Jönsson, E. G., Frei, O., Witoelar, A., Thompson, W. K., Schork, A. J., Bettella, F., Wang, Y., Djurovic, S., Smeland, O. B., et al. (2018). Identification of shared genetic variants between schizophrenia and lung cancer. *Scientific Reports*, *8*(1), 674.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.

APPENDIX

A.1 | Optimal procedure

In this section, we show the following result. This is not original; it is shown in various forms in (at least) Alishahi et al. (2016), Du et al. (2014) and Lei and Fithian (2018).

Theorem A.1. *Let f_0 and f_1 be positive Lesbegue-integrable functions of (p, q) on some region Ω. Suppose a Lesbegue-measurable region R_0 satisfies:*

1. $R_0 = \{(p, q) : f_0(p, q)/f_1(p, q) \leq k, (p, q) \in \Omega\}$
2. $\int_{R_0} f_0(p, q) dp dq = \alpha$
3. $\int_{R_0} f_1(p, q) dp dq = 1 - \beta.$

Then no Lesbegue-measurable region $R \subset \Omega$ satisfies both

$$\int_R f_0(p, q) dp dq = \alpha \tag{A.1}$$

and

$$\int_R f_1(p, q) dp dq > 1 - \beta. \tag{A.2}$$

Proof. Suppose such a region existed. Then given condition (A.1), we must have $f_0(p, q)/f_1(p, q) > k$ in $R \setminus R_0$, and since the integral of f_0 over R is equal to its integral over R_0,

$$\int_{R \setminus R_0} f_0(p, q) dp dq = \int_{R_0 \setminus R} f_0(p, q) dp dq = \alpha - \int_{R_0 \cap R} f_0(p, q) dp dq. \tag{A.3}$$

Hence

$$\int_R f_1(p, q) dp dq = \int_{R \setminus R_0} f_1(p, q) dp dq + \int_{R \cap R_0} f_1(p, q) dp dq$$

$$\leq k \int_{R \setminus R_0} f_0(p,q) dp dq + \int_{R \cap R_0} f_1(p,q) dp dq$$

$$= k \int_{R_0 \setminus R} f_0(p,q) dp dq + \int_{R \cap R_0} f_1(p,q) dp dq$$

$$\leq \int_{R_0 \setminus R} f_1(p,q) dp dq + \int_{R \cap R_0} f_1(p,q) dp dq$$

$$= \int_{R_0} f_1(p,q) dp dq = 1 - \beta \tag{A.4}$$

a contradiction of (A.2). Regions $R \neq R_0$ can satisfy (A.1) and

$$\int_R f_1(p,q) dp dq = 1 - \beta \tag{A.5}$$

if and only if $R_0 \setminus R$ and $R \setminus R_0$ have Lesbegue measure 0. $\qquad\square$

Corollary A.2. *If $f_0(p,q) = f(P = p, Q = q | H_0^P)$ and $f_1(p,q) = f(P = p, Q = q | H_1^P)$ (where H_1^P is the alternative) then amongst all rejection regions R with fixed type-1 error rate $\alpha = \int_R f_0(p,q) dp dq$, power is maximised on a region inside a contour of f_0/f_1, if such a region exists.*

Denoting $f(p,q) = \pi_0 f_0(p,q) + (1 - \pi_0) f_1(p,q)$, it is clear that a contour of f_0/f_1 is also a contour of f_0/f and of $Pr(H_0^P | P = p, Q = q)$, so an optimal rejection region is given by

$$\{(p,q) : Pr(H_0^P | P = p, Q = q) \leq k_\alpha\} \tag{A.6}$$

for some k_α.

A.2 | Failure of FDR control with $\widehat{cFDR} < \alpha$

As described in Section 2, rejection procedure (9) is similar to the B-H procedure, and it may be naively thought that it also controls the FDR at α. This is not the case, and indeed the FDR of such a procedure (and the corresponding procedure with $\widehat{cFDR}^n$) may exceed α by an arbitrary factor depending on α and π_0.

This is most easily seen by considering the extreme case in which

$$P, Q | H_0^p \sim U(0,1)^2 \tag{A.7}$$

$$P, Q | H_1^P \sim (0,0), \tag{A.8}$$

where $\pi_0 = Pr(H_0^P)$ as usual. In this case we show:

Theorem A.3. *Under the above distribution of P, Q, as $n \to \infty$, the FDR α_{TRUE} of rejection procedure (9) for $\widehat{cFDR}$ satisfies*

$$\frac{\alpha_{TRUE}(1 - \alpha)}{\alpha(1 - \alpha_{TRUE})} = OR(\alpha_{TRUE}, \alpha) \to \log\left(\frac{1 - \alpha\pi_0}{1 - \pi_0}\right) \tag{A.9}$$

and the corresponding procedure for $\widehat{cFDR}^n$ satisfies

$$\frac{\alpha_{TRUE}}{\alpha} \to \frac{1 - \log\left(\frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0}\right)}{1 - \alpha \log\left(\frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0}\right)}. \tag{A.10}$$
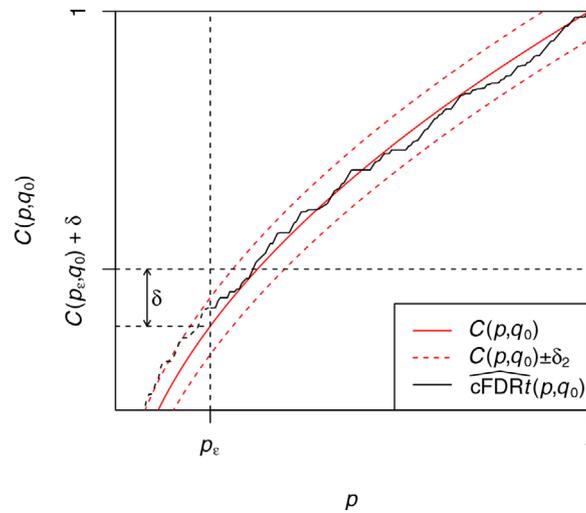
FIGURE A.1 Rejection regions for rejection procedure (9) under assumptions in Section A.2. FDR is not controlled for either of the cFDR-based rejection regions. For reference, the B-H procedure applied to the set of (p, q) pairs with $q \leq q_0$ would reject everything in the dark grey rectangle, which includes all true positives, and this would control FDR at α. The cFDR-based rejection regions reject the same number of true positives, but far more false positives, so FDR control is lost

Corollary A.4. *For $\widehat{cFDR}$, the relative error in FDR (relative to α) can grow arbitrarily large as $\pi_0 \to 1, \alpha \to 0$. For $\widehat{cFDR}^n$, the error can grow arbitrarily large as $\alpha \to 0$, regardless of π_0.*

Proof. Suppose that we have a dataset $S = \{(p_i, q_i)\}, i \in 1 \ldots n$ of draws from P, Q under (A.7) and (A.8). Due to assumption (A.7) we have $Pr(P \leq p, Q \leq q | H_0^p) = pq$, and due to (A.8) we have $Pr(P \leq p, Q \leq q | H_1^p) = 1$. Now

$$cFDR(p, q) = Pr(H_0^p | P \leq p, Q \leq q)$$

$$= \frac{\pi_0 pq}{(1 - \pi_0) + \pi_0 pq}.$$

Now

$$Pr(H_0^p | Q \leq q) = \frac{Pr(Q \leq q | H_0^p) Pr(H_0^p)}{Pr(Q \leq q | H_1^p) Pr(H_1^p) + Pr(Q \leq q | H_0^p) Pr(H_0^p)}$$

$$= \frac{\pi_0 q}{(1 - \pi_0) + \pi_0 q}.$$

The estimate $\widehat{cFDR}(p, q)$ is proportional to a consistent estimator of

$$\frac{cFDR(p, q)}{Pr(H_0^p | Q \leq q)} = p \frac{(1 - \pi_0) + \pi_0 q}{(1 - \pi_0) + \pi_0 pq} \tag{A.11}$$

and since $P > 1/2 \Rightarrow H_0^p$, approximation (7) in the main paper is consistent and $\widehat{cFDR}^n(p, q)$ is a generally consistent estimator of $cFDR(p, q)$.

The FDR of the rejection procedure $\widehat{cFDR}(p, q) \leq \alpha$ converges to the FDR of the rejection region $R_\alpha = \{(p, q) : cFDR(p, q)/Pr(H_0^p | Q) \leq q) < \alpha\}$ as $n \to \infty$ (see diagram in Figure A.1). Since this rejection region contains $(0,0)$, all $(1 - \pi_0)n$ non-null hypotheses will be rejected, and the proportion of the total null hypotheses rejected will converge

by the law of large numbers to

$$\int_{R_\alpha} f(P, Q|H_0^p)dpdq = \int_{R_\alpha} dpdq$$

$$= \frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0} \log\left(\frac{1-\alpha\pi_0}{1-\pi_0}\right), \tag{A.12}$$

and thus the FDR α_{TRUE} converges to

$$\alpha_{TRUE} \to \frac{\text{Number of null } (p_i, q_i) \text{ in } R_\alpha}{\text{Total number of } (p_i, q_i) \text{ in } R_\alpha}$$

$$= \frac{\pi_0 n \int_{R_\alpha} dpdq}{(1-\pi_0)n + \pi_0 n \int_{R_\alpha} dpdq}$$

$$= \alpha \frac{\log\left(\frac{1-\alpha\pi_0}{1-\pi_0}\right)}{1 - \alpha + \alpha \log\left(\frac{1-\alpha\pi_0}{1-\pi_0}\right)}, \tag{A.13}$$

which can be written as

$$\log\left(\frac{1-\alpha\pi_0}{1-\pi_0}\right) = \frac{\alpha_{TRUE}(1-\alpha)}{\alpha(1-\alpha_{TRUE})} = OR(\alpha_{TRUE}, \alpha), \tag{A.14}$$

where OR is the odds ratio. Hence α_{TRUE} can exceed α by an arbitrary degree when pi_0 is close to 1.

The second part of the proof can be shown similarly. The RHS of Equation (A.10) rises as $-\log(\alpha)$ as $\alpha \to 0$, whatever the value of π_0. □

A.3 | Convergence results

In these appendices, we omit the X from $\widehat{cFDR}_X(p, q)$ and other functions when it is clear. We consider p, q to be in $(0, 1)^2$.

Set $n = |X|$, $F_n(q) = \min(1, |\{i : q_i \le q, (p_i, q_i) \in X\}|)$, $F_n(p, q) = \min(1, |\{i : p_i \le p, q_i \le q, (p_i, q_i \in X)\}|)$, $F(q) = Pr(Q \le q)$, $F(p, q) = Pr(P \le p, Q \le q)$, and

$$C(p, q) = \frac{cFDR(p, q)}{Pr(H_0^p|Q \le q)} = \frac{pF(q)}{F(p, q)}. \tag{A.15}$$

We will assume $\partial F(p, q)/\partial p$ exists on $(0, 1)^2$.

We define $\widehat{cFDRt}_X(p, q) = \min_{p' \ge p} \widehat{cFDR}_{X+(p',q)}(p', q)$, and $\widehat{cFDRt}_X^n(p, q)$ similarly for $\widehat{cFDR}_X^n$ ('t' for 'truncated'), so

$$L_X(\alpha) = \{(p, q) : \widehat{cFDRt}_X(p, q) \le \alpha\}. \tag{A.16}$$

The boundary of $L_X(\alpha)$ is continuous and piecewise differentiable.

In this section, we show a series of results relating to convergence of cFDR estimates. We show results relating to convergence of $\widehat{cFDR}$ and $\widehat{cFDRt}$ on a line $q = q_0$, along with convergence of the co-ordinates of L-curves on such lines. We then show slightly weaker results regarding convergence across two-dimensional regions of the unit square.

Theorem A.5. *Suppose that on a line segment $q = q_0$, $p_\gamma < p < 1$, we have $F(p, q) \ge \gamma > 0$ and $F(q_0) > 0$. Then on this segment, $\widehat{cFDR}(p, q)$ converges uniformly to $C(p, q)$ as $n \to \infty$. If additionally we have $\partial C(p, q)/\partial p \ge 0$, then $\widehat{cFDRt}(p, q)$ converges uniformly to $C(p, q)$ also.*

Proof. Condition on $q = q_0$, and (for the moment) $F_n(q) = m$. Set $\epsilon < \delta$ and let

$$g^-(p, \epsilon) = p\frac{\frac{m}{n}}{F(p, q) + \epsilon} \qquad g^+(p, \epsilon) = p\frac{\frac{m}{n}}{F(p, q) - \epsilon}. \tag{A.17}$$

From the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality, we have

$$Pr\left(F(p, q) - \epsilon \leq \frac{F_n(p, q)}{m} \leq F(p, q) + \epsilon \,\Big|\, q = q_0, F_n(q_0) = m\right) \qquad \geq 1 - e^{-2m\epsilon^2}$$

$$\Rightarrow Pr\left(g^-(p, \epsilon) \leq \widehat{cFDR}(p, q)) \leq g^+(p, \epsilon) \,\Big|\, q = q_0, F_n(q_0) = m\right) \qquad \geq 1 - e^{-2m\epsilon^2}. \tag{A.18}$$

If $\partial C(p, q)/\partial p \geq 0$, then (A.18) also holds for $\widehat{cFDRt}$. To see this, note $\widehat{cFDR}(p, q) \geq \widehat{cFDRt}(p, q)$, so if $\widehat{cFDRt}(p, q) \geq g^+(p, \epsilon)$ then $\widehat{cFDR}(p, q) \geq g^+(p, \epsilon)$ also. Now

$$\frac{\partial}{\partial p}C(p, q) \geq 0 \Rightarrow F(p, q) \geq p\frac{\partial}{\partial p}F(p, q)$$

$$\Rightarrow F(p, q) + \epsilon \geq p\frac{\partial}{\partial p}F(p, q)$$

$$\Rightarrow g^-(p, \epsilon) > 0.$$

Suppose that for some p we had $\widehat{cFDRt}(p, q_0) \leq g^-(p)$. Then either $\widehat{cFDR}(p, q_0) = \widehat{cFDRt}(p, q_0)$ or $\widehat{cFDRt}(p, q_0) = \widehat{cFDR}(p', q_0)$ for some $p' > p$. In the first case $\widehat{cFDR}(p, q_0) \leq g^-(p)$, and in the second, $\widehat{cFDR}(p', q_0) = \widehat{cFDRt}(p, q_0) \leq g^-(p) \leq g^-(p')$; in either case, $\widehat{cFDR}(p, q)$ escapes the bound $g^-(p)$ somewhere. Thus the probability on the LHS of (A.18) can only increase if $\widehat{cFDRt}$ replaces $\widehat{cFDR}$, and $\widehat{cFDRt}(p, q)$ is contained within the bounds $g^-(p), g^+(p)$ with probability at least $1 - \exp(-2m\epsilon^2)$.

We now move to remove the condition $F_n(q_0) = m$. Denote the events

$$A : \left\{g^-(p, \epsilon) \leq \widehat{cFDR}(p, q) \leq g^+(p, \epsilon)\right\} \tag{A.19}$$

$$B : \{q = q_0\}$$

and, for some $\epsilon_2 < F(q_0)$

$$C : \left\{p\frac{F(q) - \epsilon_2}{F(p, q) + \epsilon} \leq \widehat{cFDR}(p, q) \leq p\frac{F(q) + \epsilon_2}{F(p, q) - \epsilon}\right\}. \tag{A.20}$$

Denote by $S(\epsilon_2)$, the set of integers in $[n(F(q_0) - \epsilon_2), n(F(q_0) + \epsilon_2)]$ (and assume n is large enough that $S(\epsilon_2)$ is nonempty). If $m = F_n(q_0) \in S(\epsilon_2)$, the interval in event A is a subinterval of that in event C. Thus

$$Pr(C|B) = \sum_m Pr(C|B, F_n(q_0) = m)Pr(F_n(q_0) = m)$$

$$\geq \sum_{m \in S(\epsilon_2)} Pr(C|B, F_n(q_0) = m)Pr(F_n(q_0) = m)$$

$$\geq \sum_{m \in S(\epsilon_2)} Pr(A|B, F_n(q_0) = m)Pr(F_n(q_0) = m)$$

$$\geq (1 - e^{-2\min\{S(\epsilon_2)\}\epsilon^2})Pr(m \in S(\epsilon_2))$$

$$\geq (1 - e^{-2n(F(q_0) - \epsilon_2)\epsilon^2})(1 - e^{-2n\epsilon_2^2}), \tag{A.21}$$

where the last inequality comes from the DKW inequality on $F_n(q)$. Since $p \geq p_\epsilon$ and $F(p, q) \geq \gamma$ the widest part of the interval in event C can be made arbitrarily small on the interval $(p_\epsilon, 1)$ and $\widehat{cFDR}(p, q)$ converges uniformly to $C(p, q)$. If $\partial C(p, q)/\partial p \geq 0$, then so does $\widehat{cFDRt}(p, q)$. □

Corollary A.6. *Under the assumptions in Theorem A.5, $\widehat{cFDR}(p, q)$ and $\widehat{cFDRt}(p, q)$ are bound with fixed probability on the line segment $q = q_0$, $p_\gamma < p < 1$ in intervals of width $O(n^{-1/2})$.*

Proof. In inequality (A.21), set

$$\epsilon = \frac{r}{\sqrt{F(q_0) - \epsilon_2}} \qquad \epsilon_2 = \frac{r_2}{\sqrt{n}}. \tag{A.22}$$

Then the RHS is $(1 - \exp(-2r^2))(1 - \exp(-2r_2^2))$, which may be made arbitrarily small by varying r, r_2, and the difference between the upper and lower bounds in event $C|B$ is

$$p\frac{F(q_0) + \epsilon_2}{F(p, q_0) - \epsilon} - p\frac{F(q_0) - \epsilon_2}{F(p, q_0) + \epsilon} = 2p\frac{\sqrt{F(q_0)}r + F(p, q_0)r_2}{F(p, q_0)^2} \frac{1}{\sqrt{n}} + O\left(\frac{1}{n}\right). \tag{A.23}$$

□

Theorem A.7. *Suppose that on a line segment $q = q_0$, $p_\gamma < p < 1$, we have $F(p, q) \geq \gamma > 0$, $F(q_0) > 0$, and $\partial C(p, q)/\partial p \geq \gamma_2 > 0$. Denote by $l(\alpha)$ the value of p at the intersection of the L-curve $L(\alpha)$ with the line $q = q_0$, so*

$$l(\alpha) = \sup\{p : \widehat{cFDRt}(p, q_0) \leq \alpha\}, \tag{A.24}$$

and $c(\alpha)$ the value of p such that $C(p, q_0) = \alpha$ (unique if it exists). For any $\delta > 0$, the function $|l(\alpha) - c(\alpha)|$ converges uniformly to 0 for $\alpha \in [C(p_\epsilon, q_0) + \delta, 1]$.

Proof. Since $C(p, q_0)$ is continuous and increasing on $[p_\epsilon, 1]$, the value $c(\alpha)$ exists for $\alpha \in [C(p_\epsilon, q_0), C(1, q_0)] \supset [C(p_\epsilon, q_0) + \delta, 1]$ by the intermediate value theorem. The function $\widehat{cFDRt}(p, q)$ is continuous and nondecreasing on [0,1] and hence $l(\alpha)$ exists for $\alpha \in [\widehat{cFDRt}(0, q_0), \widehat{cFDRt}(1, q_0)] = [0, 1]$.

Given arbitrarily small positive ϵ_3, $\delta_2 < \delta$ choose n large enough that $\widehat{cFDRt}(p, q_0)$ is contained in $[C(p, q_0) - \delta_2, C(p, q_0) + \delta_2]$ for $p \in [p_\epsilon, 1]$ with probability at least $1 - \epsilon_3$. Then with probability $\geq 1 - \epsilon_3$, whenever the curve $\widehat{cFDRt}(p, q)$ is in the region bounded by the rectangle $p_\epsilon \leq p \leq 1$, $C(p_\epsilon, q_0) + \delta \leq q \leq 1$, it is bounded by the curves $C(p, q_0) - \delta_2$, $C(p, q_0) + \delta_2$. The distance between the two curves in the q-direction is at most $2\gamma_2\delta_2$. Thus, if for some $\alpha \in [C(p_\epsilon, q_0) + \delta, 1]$, we have $|l(\alpha) - c(\alpha)| > \gamma_2\delta_2$, the curve $\widehat{cFDRt}(p, q_0)$ must escape the region bounded the curves $C(p, q_0) - \delta_2$, $C(p, q_0) + \delta_2$.

So with probability at least $1 - \epsilon_3$, we have

$$\forall \alpha \in [C(p_\epsilon, q_0) + \delta, 1] : |l(\alpha) - c(\alpha)| \leq \gamma_2\delta_2, \tag{A.25}$$

which proves the statement. This is illustrated in Figure A.2. □

We now proceed to the proof of Theorem 3.3, restated here:

Theorem 3.3. *Let R be the region of the unit square for which $F(p, q) \geq \gamma > 0$ and $F(q) > 0$. Then on R, $\widehat{cFDR}(p, q)$ converges uniformly to $C(p, q)$, and if $\partial C(p, q)/\partial p \geq 0$, then so does $\widehat{cFDRt}(p, q)$.*

Proof. We proceed very similarly to Theorem A.5. We employ a result from Kiefer (1961) that for any $\epsilon > 0$

$$Pr\left(\sup |F_n(p, q) - F(p, q)| \geq \frac{r}{\sqrt{n}}\right) \leq c(\epsilon)e^{(2-\epsilon)r^2} \tag{A.26}$$
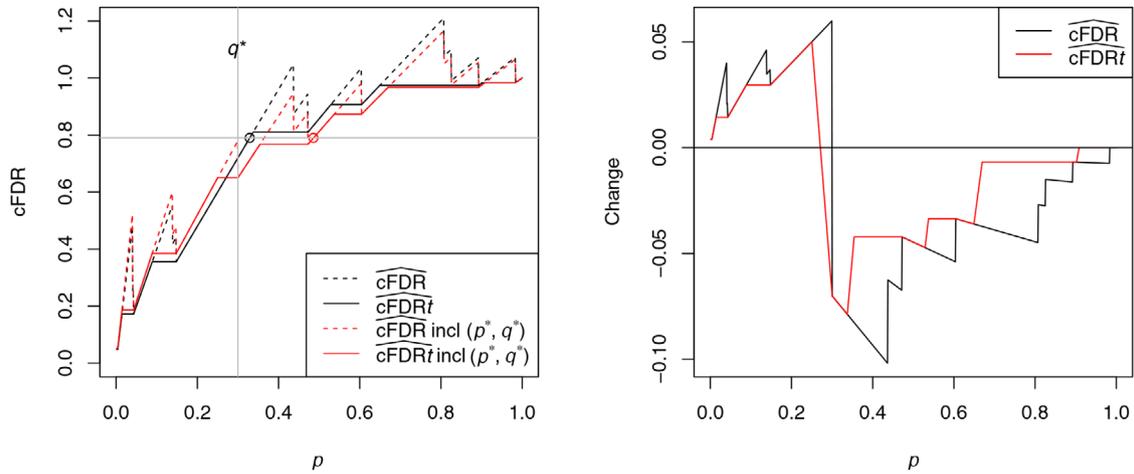
FIGURE A.2 Convergence of intersections of L-curves with a line $q = q_0$. Functions $\widehat{cFDRt}(p, q_0)$, $C(p, q_0)$ and $C(p, q_0) \pm \delta_2$ are shown. The vertical distance between dashed red lines is $2\delta_2$, and since $\partial C(p, q0)/\partial p \geq \gamma_2$ the horizontal distance is at most $2\delta_2\gamma_2$. We must restrict the proof to $\alpha > C(p, q_0) + \delta$ because we cannot assert the behaviour of $\widehat{cFDRt}(p, q)$ left of the line $p = p_\epsilon$

from which, wherever $F(p, q) > f_{min} > \frac{r}{\sqrt{n}}$

$$Pr\left(\frac{1}{F(p,q) + \frac{r}{\sqrt{n}}} \leq \frac{1}{F_n(p,q)} \leq \frac{1}{F(p,q) - \frac{r}{\sqrt{n}}}\right) \geq 1 - c(\epsilon)e^{(2-\epsilon)r^2}$$

$$Pr\left(\frac{p\frac{F_n(q)}{n}}{F(p,q) + \frac{r}{\sqrt{n}}} \leq \widehat{cFDR}(p,q) \leq \frac{p\frac{F_n(q)}{n}}{F(p,q) - \frac{r}{\sqrt{n}}}\right) \geq 1 - c(\epsilon)e^{(2-\epsilon)r^2}. \tag{A.27}$$

The values $F_n(p, q)$ and $F_n(q)$ are dependent. However, given some $r_2 > 0$, we have for all q (by the DKW inequality)

$$Pr\left(F(q) - \frac{r_2}{\sqrt{n}} \leq \frac{F_n(q)}{n} \leq F(q) + \frac{r_2}{\sqrt{n}}\right) \geq 1 - 2e^{-2r_2^2}. \tag{A.28}$$

Denoting the event in (A.27) by A, the event in (A.28) by B, and C as

$$\frac{p(F(q) - \frac{r_2}{\sqrt{n}})}{F(p,q) + \frac{r}{\sqrt{n}}} \leq \widehat{cFDR}(p,q) \leq \frac{p(F(q) + \frac{r_2}{\sqrt{n}})}{F(p,q) - \frac{r}{\sqrt{n}}} \tag{A.29}$$

we have, since the interval in A is a subinterval of that in C when conditioning on B:

$$Pr(C) = Pr(C|B)Pr(B) + Pr(C|\neg B)Pr(\neg B) \tag{A.30}$$

$$\geq Pr(C|B)(1 - 2e^{-2r_2^2}) \tag{A.31}$$

$$\geq Pr(A|B)(1 - 2e^{-2r_2^2}) \tag{A.32}$$

$$\geq (1 - c(\epsilon)e^{(2-\epsilon)r^2})(1 - 2e^{-2r_2^2}). \tag{A.33}$$

As before, this bound also holds for $\widehat{cFDRt}$ as long as $\partial C(p, q)/\partial p > 0$. $\qquad \square$

Corollary A.8. *Under the assumptions in 3.3, $\widehat{cFDR}$ and $\widehat{cFDR}t$ are bound with fixed probability in R in intervals of width $O(n^{-1/2})$*

Proof. The difference between the upper and lower bounds in (A.29) is

$$\frac{2p(rF(q) + r_2F(p,q))}{F(p,q)^2}\frac{1}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}^3}\right). \tag{A.34}$$

$\square$

Our final result describes errors on v-values. Given an L-region $L(\alpha)$, we define the M-region as the 'expected' L-region:

$$M(\alpha) = \{(p,q) : C(p,q) \le \alpha\}, \tag{A.35}$$

and the 'error' on the v-value $v = \int_{L(\alpha)} f_0(p,q)dpdq$ as

$$|\Delta v| = \left|\int_{L(\alpha)} f_0(p,q)dpdq - \int_{M(\alpha)} f_0(p,q)dpdq\right|. \tag{A.36}$$

We now are now in a position to prove Theorem 3.4.

Theorem 3.4. *Define R as in Theorem 3.3, and further assume that $f_0(p,q) = f(P = p, Q = q|H_0^P)$ is known and on R we have $\partial C(p,q)/\partial p \ge \gamma_2$. Write $R^c = [0,1]^2 \setminus R$. Then the maximum error on any v-value is*

$$\int_{R^c} f_0(p,q)dpdq + O\left(\frac{1}{\sqrt{n}}\right). \tag{A.37}$$

Proof. Using Theorem 3.3, bound $\widehat{cFDR}t(p,q)$ between $C(p,q) - \delta$, $C(p,q) + \delta$ with probability $\ge 1 - \epsilon_3$, where $\delta = O(1/\sqrt{n})$.

Since $F(p,q)$ is non-decreasing with p, we can describe $R = \{(p,q) : F(p,q) \ge \gamma\}$ as the union of line segments $q = q_0$, $p_\epsilon(q_0) \le p \le 1$. We now define R_1 as the union of all line segments $q = q_0, p_\epsilon(q_0) + \delta\gamma_2 \le p \le 1$; that is, R with the leftmost border shifted $\delta\gamma_2$ to the right.

We show the result by firstly noting that if an L-curve intersects a line segment $q = q_0$ at $l(\alpha) > p_\epsilon(q_0) + \delta\gamma_2$, and we have that $|l(\alpha) - c(\alpha)| > \delta\gamma_2$ (where $c(\alpha)$ is the intersection of the border of $M(\alpha)$ with $q = q_0$), then event C (Equation (A.29)) must have occurred in R, by the same argument as for Theorem A.7. Thus with probability at least $1 - \epsilon_3$, every segment of a right-most border of an L-region $L(\alpha)$ in R_1 is at a horizontal distance from the corresponding rightmost-border of $M(\alpha)$ of at most $\delta\gamma_2$

We now write

$$\Delta v = \left(\int_{L(\alpha)\cap R^c} f_0(p,q)dpdq - \int_{M(\alpha)\cap R^c} f_0(p,q)dpdq\right)$$

$$+ \left(\int_{L(\alpha)\cap(R\setminus R_1)} f_0(p,q)dpdq - \int_{M(\alpha)\cap(R\setminus R_1)} f_0(p,q)dpdq\right)$$

$$+ \left(\int_{L(\alpha)\cap R_1} f_0(p,q)dpdq - \int_{M(\alpha)\cap R_1} f_0(p,q)dpdq\right). \tag{A.38}$$

The first term is at most $\int_{R^c} f_0(p,q)dpdq$. The region $R \setminus R_1$ has constant width $\delta\gamma_2$, and since f_0 only varies with q, hence the second term is at most $\int_{R\setminus R_1} f_0(p,q)dpdq = \delta\gamma_2 = O(n^{-1/2})$. Within R_1, if the horizontal separation between curves

at the rightmost border of $L(\alpha)$ and $M(\alpha)$ is greater than $\delta\gamma_2$, then C has occurred, so this can happen with probability at most ϵ_3. Thus with probability $1 - \epsilon_3$, the third term is also bounded by $\delta\gamma_2 = O(n^{-1/2})$, establishing the result. $\qquad\square$

A.4 | Influence of a single point

Intuitively, adding a single point to a map defined by n other points should have a small effect on that map, and hence on the resultant v-values. We show the following:

Theorem 3.5. *Suppose we add a point (p^*, q^*) to a set of n points (p_i, q_i), considered as realisations of P, Q, and conditions are satisfied for convergence of v-values as above. Let $\Delta v(L(\alpha))$ be the shift in a v-value corresponding to an L-curve $L(\alpha)$ after adding (p^*, q^*). Then*

$$E_{\alpha \sim U(0,1)}(|\Delta v(L(\alpha))|) = O\left(\frac{1}{n^2}.\right). \tag{A.39}$$

Proof. Consider the profile of $\widehat{cFDR}t(p, q)$ on a line $q = q_0$, and how this changes with the addition of (p^*, q^*). The functions $F_n(q), F_n(p, q)$ will be taken to be with respect to the n points (p_i, q_i) but not (p^*, q^*).

For $q_0 < q^*$, the addition of (p^*, q^*) changes neither $F_n(q_0)$ nor $F_n(p, q_0)$, so on lines $q = q_0 < q^*$ the profile of $\widehat{cFDR}t$ will remain the same.

Denote

$$c^+(p) = p\frac{F_n(q) + 1}{F_n(p, q)} \qquad c^-(p) = p\frac{F_n(q) + 1}{F_n(p, q) + 1}. \tag{A.40}$$

For $q_0 > q^*$, $p < p^*$, the value of $\widehat{cFDR}(p, q_0)$ will increase by

$$c^+(p) - p\frac{F_n(q)}{F_n(p, q)} = \frac{p}{F_n(p, q_0)}, \tag{A.41}$$

and for $q_0 > q^*$, $p > p^*$, it will decrease by

$$p\frac{F_n(q)}{F_n(p, q)} - c^-(p) = p\frac{F_n(q_0) - F_n(p, q_0)}{F_n(p, q_0)(F_n(p, q_0) + 1)}. \tag{A.42}$$

In either case, $\widehat{cFDR}(p, q_0)$ changes by $O(\frac{1}{n^2})$. The behaviour of $\widehat{cFDR}t$ is a little more complex. If we define $c_t^+(p)$ and $c_t^-(p)$ analogously to $\widehat{cFDR}t(p, q_0)$, then for $p > p^*$ $\widehat{cFDR}t(p, q_0)$ shifts to $c_t^-(p)$, and for $p < p^*$ it shifts to $\min(c_t^+(p), c_t^-(p^*))$ (see example in Figure A.3).

We can show that the absolute difference in $\widehat{cFDR}t(p, q_0)$ is always less than the absolute difference in $\widehat{cFDR}(p, q_0)$ after adding (p^*, q^*). Denote these differences $\Delta\widehat{cFDR}(p, q_0)$ and $\Delta\widehat{cFDR}t(p, q_0)$. Since $\widehat{cFDR}t(p, q_0)$ always shifts to between $c_t^-(p)$ and $c_t^+(p)$, it suffices to show that

$$c_t^+(p) - \widehat{cFDR}t(p, q_0) \le \Delta\widehat{cFDR}(p, q_0) \tag{A.43}$$

$$\widehat{cFDR}t(p, q_0) - c_t^-(p) \le \Delta\widehat{cFDR}(p, q_0). \tag{A.44}$$

Inequality (A.43) follows from the observation that $c^+(p) \propto \widehat{cFDR}(p, q_0)$, so order relations between $\widehat{cFDR}(p, q_0)$ and $c^+(p, q_0)$ are preserved. Thus

$$|\Delta\widehat{cFDR}t(p, q_0)| = \min_{p' \ge p} \Delta|\widehat{cFDR}(p', q_0)| \le |\Delta\widehat{cFDR}(p', q_0)|. \tag{A.45}$$

Order relations are not preserved between $\widehat{cFDR}(p, q_0)$ and $c^-(p)$, but the denominators increment at the same values of p. The functions $\widehat{cFDR}(p, q_0)$ and $c^-(p)$ both rise linearly in p between successive increment points p_a, p_d of $F_n(p, q_0)$,
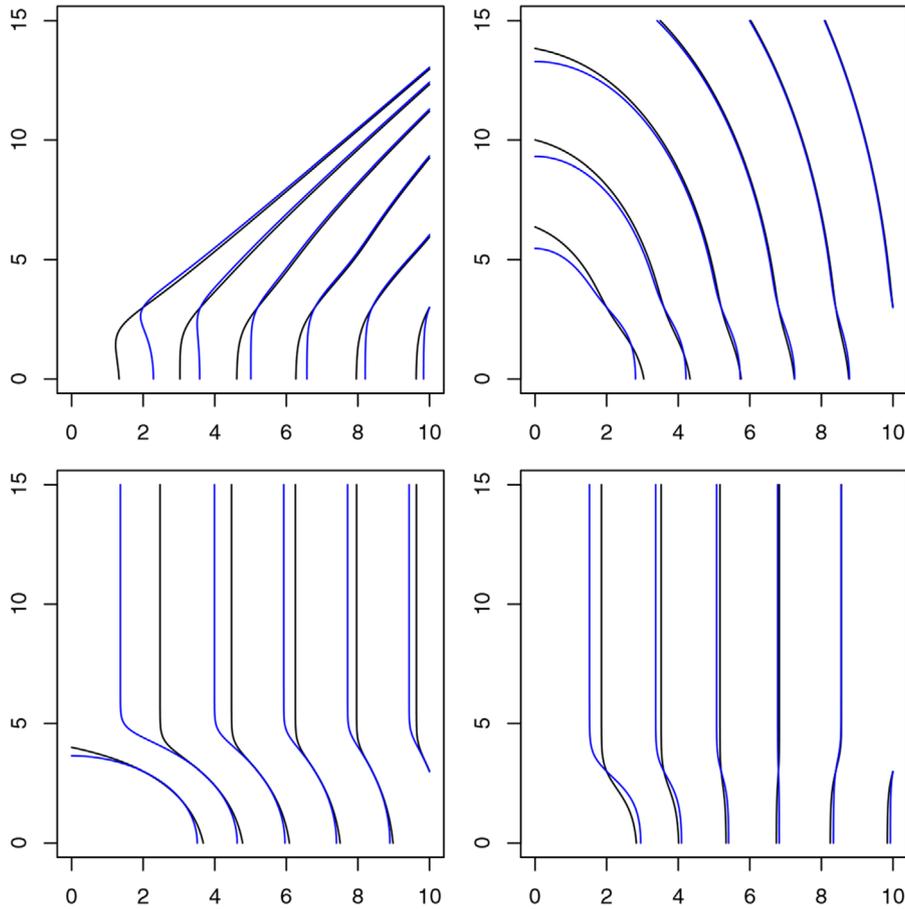
FIGURE A.3   Behaviour of $\widehat{cFDR}$ and $\widehat{cFDRt}$ on a line $q = q_0$ after adding a point (p^*, q^*) to a set of n points (n is considerably smaller in this example than in Figure A.2). In the left panel, curves of $\widehat{cFDR}$ and $\widehat{cFDRt}$ before and after adding (p^*, q^*) are shown. Adding (p^*, q^*) may have a substantial impact on the intersection of an L-curve with $q = q_0$, such as that in the horizontal line: the black and red points show the intersection points of a curve before and after adding (p^*, q^*). However, the average effect across all curves is limited to the integral of the difference between the black and red lines, which is $O(1/n^2)$. The right panel demonstrates that $|\Delta\widehat{cFDRt}(p, q)| \leq |\Delta\widehat{cFDR}(p, q)|$

with $\widehat{cFDR}(p, q_0)$ having the higher gradient, since

$$\frac{F_n(q)}{F_n(p, q)} > \frac{F_n(q) + 1}{F_n(p, q) + 1}. \tag{A.46}$$

At p_d, both functions are discontinuous and drop in value. On (p_a, p_d), the values of $c_t^-(p)$ and $\widehat{cFDRt}(p, q_0)$ are either equal to $c^-(p)$, $\widehat{cFDR}(p, q_0)$, or 'censored' at some values $c^-(p')$, $\widehat{cFDR}(p', q_0)$ with $p' > p$ (see the right-hand part of Figure A.3 for an example of this). We note that $c^-(p) > c^-(p')$, $p' > p \Rightarrow \widehat{cFDR}(p, q_0) > \widehat{cFDR}(p, q_0)$, so the first point at which $c_t^-(p)$ is censored on (p_1, p_2) is further right than the first point at which $\widehat{cFDRt}(p, q_0)$ is censored. Denote the leftmost point at which $\widehat{cFDRt}(p, q_0)$ is censored as p_b and the leftmost point at which $c_t^-(p)$ is censored as p_c, so $p_a \leq p_b \leq p_c \leq p_d$.

On (p_a, p_b), where neither are censored, $\widehat{cFDR}(p, q_0) - c^+(p) = \widehat{cFDRt}(p, q_0) - c_t^-(p)$ and $\Delta\widehat{cFDR}(p, q_0) = \Delta\widehat{cFDR}(p, q_0)$. On (p_b, p_c), when only $\widehat{cFDRt}(p, q_0)$ is censored, $\widehat{cFDRt}(p, q_0) - c_t^-(p) = \widehat{cFDR}(p_b, q_0) - c^-(p) < \widehat{cFDR}(p, q_0) - c^-(p)$, so $\Delta\widehat{cFDR}(p, q_0) \leq \Delta\widehat{cFDR}(p, q_0)$. On (p_c, p_d), we have $\widehat{cFDRt}(p, q_0) - c^-(p) = \widehat{cFDR}(p_b, q_0) - c^-(p_c) \leq \widehat{cFDR}(p_c, q_0) - c^-(p_c) \leq \widehat{cFDR}(p, q_0) - c^-(p)$, so again, $\Delta\widehat{cFDRt}(p, q_0) \leq \Delta\widehat{cFDR}(p, q_0)$. Thus, for all p,

$$|\Delta\widehat{cFDRt}(p, q_0)| \leq |\Delta\widehat{cFDR}(p, q_0)| = O\left(\frac{1}{n^2}\right), \tag{A.47}$$

where the multiplicative factor in $O(1/n^2)$ is independent of q_0. This inequality is demonstrated in the right panel of Figure A.3.

Denote by l_α the value of p at the intersection of an L-curve corresponding to $\widehat{cFDRt}(p,q) \leq \alpha$ with the line $q = q_0$. We have $l_\alpha = \max\{p : \widehat{cFDRt}(p, q_0) = \alpha\}$. The value l_α may shift substantially when adding p^*, q^*, as shown in Figure A.3

However, the effect is small on average. The plot of the function $l(\alpha)$ before and after adding (p^*, q^*) is identical to the plot of the function of p given by $\widehat{cFDRt}(p, q_0)$ before and after adding (p^*, q^*) rotated by $\pi/2$. The average difference in movement of l_α is the integral of the difference in l_α with and without (p^*, q^*). However, this is simply the area between the two curves, which is invariant under rotating $\pi/2$. Hence

$$\int_0^1 \Delta l_\alpha d\alpha = \int_0^1 \Delta\widehat{cFDRt}(p, q_0)dp = O\left(\frac{1}{n^2}\right). \tag{A.48}$$

Denote the region $L(\alpha) : \widehat{cFDRt}(p,q) \leq \alpha$ and the co-ordinates of its rightmost border (L-curve) $(q, l_\alpha(q))$, $q \in (0, 1)$. Then, denoting the indicator function by I

$$v(L(\alpha)) = \int_0^1 \int_0^1 I((p,q) \in L(\alpha))f_0(p,q)dp\,dq$$

$$= \int_0^1 \int_0^1 I((p,q) \in L(\alpha))f_0^q(q)dp\,dq$$

$$= \int_0^1 f_0^q(q) \int_0^1 I((p,q) \in L(\alpha))dp\,dq$$

$$= \int_0^1 f_0^q(q)l_\alpha(q)dq, \tag{A.49}$$

and the average error in v-values $v(L(\alpha))$ over $\alpha \sim U(0, 1)$ is

$$E_{\alpha \sim U(0,1)}(|\Delta v(L(\alpha))|) = \int_0^1 \Delta v(L(\alpha))d\alpha$$

$$= \int_0^1 \int_0^1 f_0^q(q)\Delta l_\alpha(q)dq\,d\alpha$$

$$= \int_0^1 f_0^q(q) \int_0^1 \Delta l_\alpha(q)\,d\alpha dq$$

$$= O\left(\frac{1}{n^2}\right) \int_0^1 f_0^q(q)dq$$

$$= O\left(\frac{1}{n^2}\right) \tag{A.50}$$

as required. □

A.5 | Asymptotic equivalence of PDF- and CDF- based L-regions

We show in this section that under a fairly common condition L-regions based on the PDF of p, q are similar to L-regions based on the CDF. In this section, we generally work on the Z-scale rather than the p-value scale for convenience.

Denote a 'fast-decreasing' function as a function g such that for each $\epsilon_1, \epsilon_2 > 0$, there exists δ such that for all X, Y of distance at least δ from the origin, we have

$$\iint_{\substack{x \leq X, y \leq Y \\ (x-X)^2 + (y-Y)^2 \leq \epsilon_1^2}} g(x,y)dxdy > (1 - \epsilon_2) \iint_{x \leq X, y \leq Y} g(x,y)dxdy \tag{A.51}$$
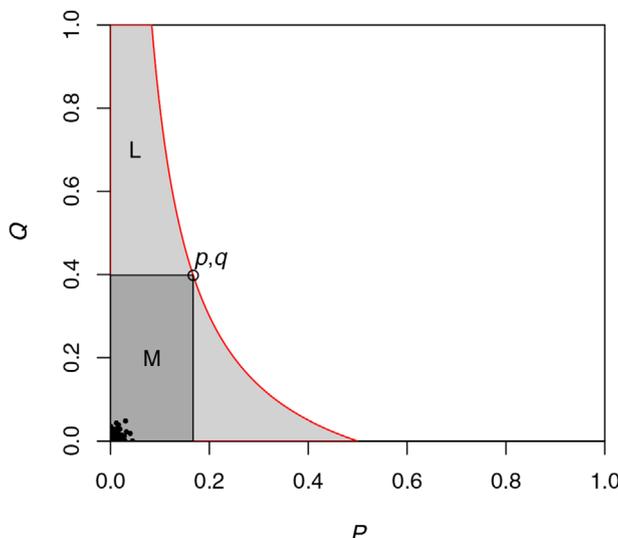
FIGURE A.4 These plots show contours of CDF- based and PDF-based L-regions for a range of distributions of P, Q. Plots are on the Z-score scale (e.g., rejection regions in terms of Z_P, Z_Q). The distributions are parameterised in terms of the mixture-Gaussian distribution detailed in Supporting Information, Section 1.4.1; parameters $(\pi_0, \pi_1, \pi_2, \tau_1, \tau_2, \sigma_1, \sigma_2)$ were $(0.7,0.1,0.1,2,3,1.5,1.5)$, $(0.7,0.1,0.1,2,2,3,3)$, $(0.99,0.0005,0.0003,2,3,4,3)$ and $(0.7,0.1,0.05,3,2,2,2)$, respectively. Curves are generated passing through the points $Z_P, Z_Q = (3, 2 \ldots 6)$, with curves further to the right corresponding to smaller α. As α gets smaller, contours $F_0(x, y)/F(x, y) = \alpha$ (black lines) become closer to contours of $f_0(x, y)/f(x, y) = \alpha$ (blue lines) under reasonably general circumstances

so for $x \leq X, y \leq Y$, the function g falls off rapidly enough as x, y decrease that we can disregard its value except when it is close to X, Y.

We show the following:

Theorem A.9. *Denote $c(x, y) = f_0(x, y)/f(x, y)$ and $C(x, y) = F_0(x, y)/F(x, y)$. Given a region of the $(-, -)$ quadrant $A_\epsilon = (-\infty, 0] \times (I_1 - \epsilon, I_2 + \epsilon)$ (where $\epsilon > 0$ is arbitrarily small), suppose that for $x, y \in A_\epsilon$ and for sufficiently small α we have*

1. *f_0 and f are fast-decreasing continuous positive functions,*
2. *Along horizontal rays in A, $c(x, y)$ satisfies $\partial^2 \log(c(x, y))/\partial x^2 > 0$ and*
3. *The contour $c(x, y) = \alpha$ is continuous and bounded, and the rightmost bound increases to ∞ as $\alpha \to 0$.*

Then for each $\epsilon_3 > 0$, there exists an ϵ_1 as above and an α_1 such that whenever $\alpha < \alpha_1$, there is a contour of $C(x, y)$ is never further than ϵ_3 from the contour $c(x, y) = \alpha$ in the region A_0.

Proof. Set R_3 as the region defined by the union of all circles of radius ϵ_3 with centres on points $y, l_\alpha(y)$. Choose $\epsilon_1 = \epsilon_3/2$ (supposing that $\epsilon_1 < \epsilon$), and define R_1 similarly to R_3 with radii ϵ_1. Let α^+ be the minimum value of f_0/f on the rightmost border of R_1, and α^- the maximum value on the leftmost border so $\alpha^+ > \alpha > \alpha^-$.

Condition 2 implies that for fixed y

$$\frac{d}{dx}\left(\frac{c(x + \epsilon_1, y)}{c(x, y)}\right) < 0. \tag{A.52}$$

Since the horizontal distance between the rightmost border of R_3 and the curve is at least $2\epsilon_1$ and similarly from the leftmost border of R_3, the values $\alpha^+ - \alpha, \alpha - \alpha^-$ must increase for fixed ϵ_1 as we move left. Thus, for some fixed $\epsilon_2 > 0$, choose δ_2 large enough that $\alpha^+/\alpha^- > 1/(1 - \epsilon_2)^2$ and larger than the δ corresponding to ϵ_1, ϵ_2 by assumption, and α_1 large enough that the contour $c(x, y)$ is entirely left of the line $x = -\delta_2$.

Let X, Y be a point in A_0 to the right of R_3, so a circle of radius ϵ_1 centred at X, Y is in A_ϵ but does not intersect R_1. Thus across such a circle, the value of $c(x, y)$ is at least α^+. Similarly, across a circle of radius ϵ_1 centred to the left of R_3, the value of $c(x, y)$ is at most α^-

For x, y to the right of R_3, denote by H the circle of radius ϵ_1 centred at x, y. Now by the fast-decreasing property of f_0 and f, we have

$$F_0(x, y) > \int_H f_0(x, y) dx dy > \alpha^+ \int_H f(x, y) dx dy > \alpha^+(1 - \epsilon_2) F(x, y) \tag{A.53}$$

so $C(x, y) > \alpha^+(1 - \epsilon_2)$. Similarly for x, y to the left of R_3, we have $C(x, y) < \alpha^-/(1 - \epsilon_2)$. By our choice of α_1, we have $\alpha^+(1 - \epsilon_2) > \alpha^-/(1 - \epsilon_2)$, so any contour of $C(x, y)$ at a level between these values must pass within R_3 through A_0. $\quad\square$

Contours of F_0/F correspond to contours of $cFDR$, and contours of f_0/f correspond to contours of $Pr(H_0^p | P = p, Q = q)$. Theorem A.9 has obvious analogies in other quadrants, and for the p-value rather than z-score scale.

The conditions in the theorem may seem restrictive, but they are satisfied by many distributions; for instance, when f_0 and f are mixture Gaussian, and f dominates f_0 as $|x| \to \infty$. Figure A.4 shows the similarity of a range of shapes of contours of C and c.