Joint Mirror Procedure
Controlling FDR for Simultaneous Signals

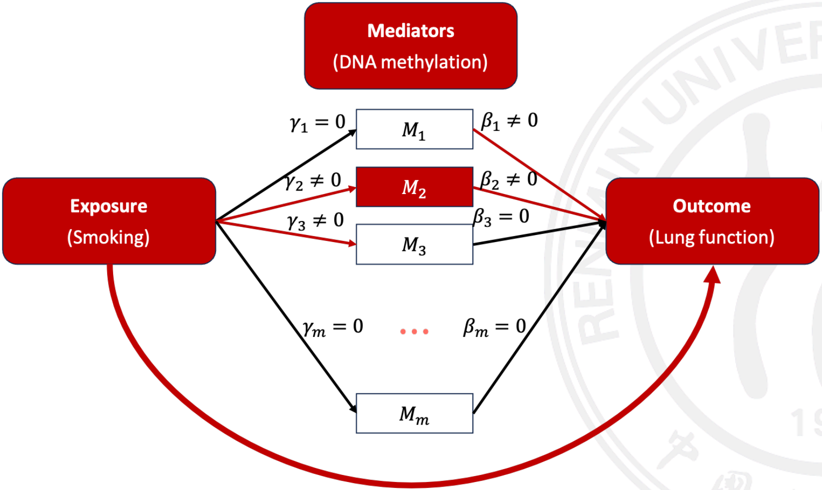
Linsui Deng

joint-work with
Kejun He and Xianyang Zhang

2025 年 8 月 9 日

Content

Content



BackGround 1. Which **DNA methylation sites** are simultaneously influenced by **smoking** and contribute to **lung function** deficiency?

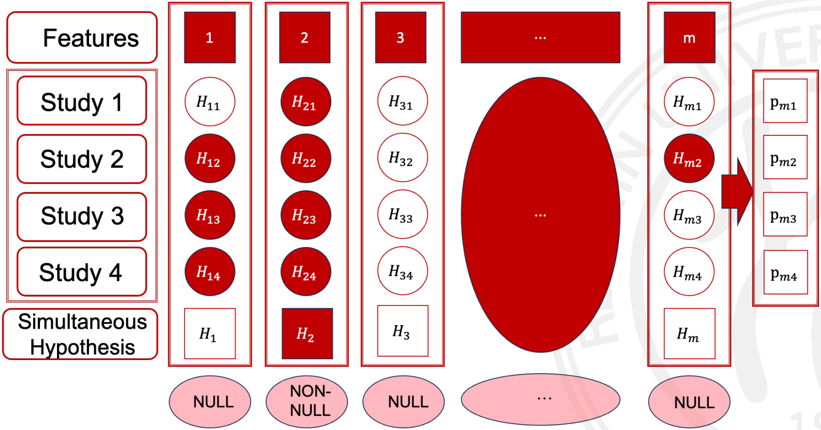
Replicacibility Study

Crohn's disease SNP		1	2	3	⋯	m
adolescent	Study 1	p_{11}	p_{21}	p_{31}		p_{m1}
belge	Study 2	p_{12}	p_{22}	p_{32}	⋯	p_{m2}
cedar 1	Study 3	p_{13}	p_{23}	p_{33}		p_{m3}
cedar 2	Study 4					
german	Study 5		⋯		⋯	⋯
niddkj	Study 6					
niddknj	Study 7					
wtccc	Study 8	p_{18}	p_{28}	p_{38}	⋯	p_{m8}

BackGround 2. Which **SNPs** are simultaneously associated with **Crohn's disease** in all eight studies?

Content

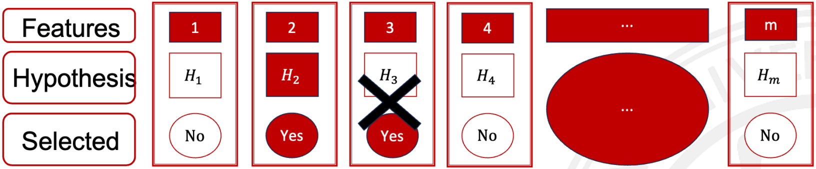
Setup: Simultaneous Hypothesis



$\Downarrow$

- m hypotheses to be tested $\Rightarrow$ multiple testing;
- Null hypotheses have special structures $\Rightarrow$ Composite Null.

Multiple Hypothesis Testing

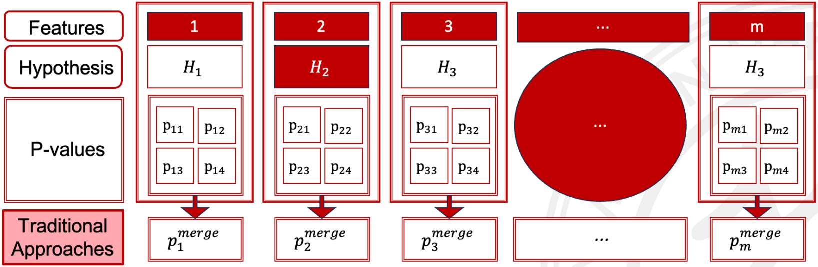


- ⋆ Why **multiple testing**? Type-I error rate control is not adequate
  - $\mathbb{P}$ (Rejecting the null hypothesis|the null hypothesis is true)
- Given selection set $\mathcal{S} \subset \{1, \cdots, m\}$, the number of false discoveries:

$$\mathrm{FD} = \sum_{i=1}^{m} \mathbf{1}\{H_i = 0, i \in \mathcal{S}\}$$

- Two error criterion:
  - Family-wise error rate (Tukey, 1953): $\mathrm{FWER} = \mathbb{P}(\mathrm{FD} \geq 1)$.
  - ✓ False discovery rate (Benjamini and Hochberg, 1995)

$$\mathrm{FDR} = \mathbb{E}\{\mathrm{FDP}\} = \mathbb{E}\left\{\frac{\mathrm{FD}}{|\mathcal{S}| \vee 1}\right\}.$$



- ⋆ For each i, calculate a p-value by merging $(p_{i1}, \cdots, p_{iK})$:
- Ensure validity: Uniform conservatism $\mathbb{P}(p_i^{merge} \leq t \mid H_i = 0) \leq t$
  - e.g., $\mathbb{P}(\max(p_{i1}, \cdots, p_{iK}) \leq t \mid H_i = 0) \leq \mathbb{P}(p_{ik} \leq t \mid H_{ik} = 0) \leq t$
- Improve power:
  - Est. the FDP by approx. the dist. of p^{merge} for null (Dai et al., 2020).
  - Reconstruct $p_i^{merge} \mid H_i = 0 \approx U[0, 1]$ (Liu et al., 2021; Dickhaus et al., 2021; Wang et al., 2022).
    - Apply standard multiple testing approaches (Benjamini and Hochberg, 1995).

Content

- Introduction
  - Motivation
  - Setup
  - Target and Challenges
- Joint Mirror Procedure
  - 1-dim Mirror
  - k-Dim Mirror and Composite FDR
  - Joint Mirror Procedure
  - Composite FDR
- Result
  - Theory
  - Simulation
  - RealData
- Summary

Merging may lose information

We have $(p_{i1}, p_{i2}, \cdots, p_{iK})$ for feature i.

Old: Obtain p_i^{merge} by merging $(p_{i1}, p_{i2}, \cdots, p_{iK})$;

☹ p_i^{merge} does not consider the *signal strengths* across different studies.

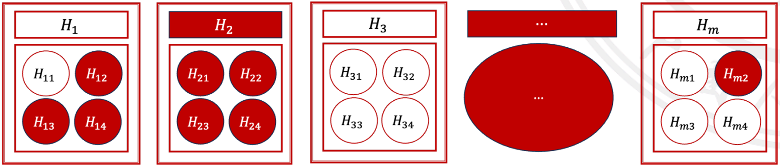
? Can we work on $(p_{i1}, p_{i2}, \cdots, p_{iK})$ directly?

Error Criterion

The null components of $H_1, \cdots, H_m$ may be different.

Old: Simple FDR.

? Should we treat $H_1, \cdots, H_m$ equally?



Content

⋆ Let us examine the number of false discoveries (FD) firstly:

$$\begin{aligned}\mathrm{FD} &= \sum_{i=1}^{m} \mathbf{1}\left\{H_i = 0, i \in \mathcal{S}\right\} \\ &\leq \sum_{i=1}^{m} \sum_{k=1}^{K} \mathbf{1}\left\{H_{ik} = 0, i \in \mathcal{S}\right\} = \sum_{k=1}^{K} \boxed{\sum_{i=1}^{m} \mathbf{1}\left\{H_{ik} = 0, i \in \mathcal{S}\right\}}\end{aligned}$$

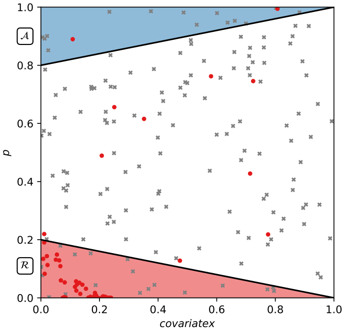
⋆ Let $\mathcal{R}$ be the rejection region, then

$$\mathrm{FD} = \sum_{k=1}^{K} \boxed{\sum_{i=1}^{m} \mathbf{1}\left\{H_{ik} = 0, (p_{i1}, p_{i2}, \cdots, p_{iK}) \in \mathcal{R}\right\}}$$

⋆ (a) Estimate the FD within each study + (b) Sum up the FDs.

Content

- ⋆ Consider only one study first.
- What is 1-dim mirror procedure (Lei and Fithian, 2016) and how it works?
  - p-value p_i, covariate x_i and hypothesis H_i.
  - $p_i \mid x_i \sim U[0,1]$, if $H_i = 0$.



- $\mathcal{R}$: Rejection region
- $\mathcal{A}$: Mirror/ Control region

$\Downarrow$

Upper Bound

$$\begin{aligned} \mathrm{FD} =& \#\{i : H_i = 0, p_i \in \mathcal{R}(x_i)\} \\ \approx& \#\{i : H_i = 0, p_i \in \mathcal{A}(x_i)\} \\ \lesssim& \#\{i : p_i \in \mathcal{A}(x_i)\} \end{aligned}$$

Mirror Conservatism

Univariate mirror conservatism (Lei and Fithian, 2016)

If $H_i = 0$, $\mathbb{P}\left(p_i \in [a_1, a_2] \mid x_i\right) \leqslant \mathbb{P}\left(p_i \in [1 - a_2, 1 - a_1] \mid x_i\right)$

(a) Permutation p-values;

(b) One-sided tests for univariate parameters with monotone likelihood ratio.

Conditional mirror Conservatism

If $H_{ik} = 0$, for p_{ik}, we have

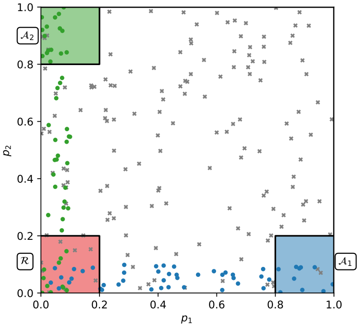
$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\theta},k|-k}\left(p_{ik} \in [a_1, a_2] \mid \mathbf{p}_{i,-k} = \mathbf{t}\right) \\ &\qquad \leqslant \mathbb{P}_{\boldsymbol{\theta},k|-k}\left(p_{ik} \in [1 - a_2, 1 - a_1] \mid \mathbf{p}_{i,-k} = \mathbf{t}\right) \end{aligned}$$

- Univariate mirror conservatism + independent study.

1-dim Mirror $\Rightarrow$ 2-dim Mirror

Consider two studies:

- Paired p-value (p_{i1}, p_{i2});
- $p_{i1} \overset{H_{i1}=0}{\sim} U[0,1] + p_{i2} \overset{H_{i2}=0}{\sim} U[0,1]$



- $\mathcal{R}$: Rejection region
- $\mathcal{A}_1$: Mirror region for 1st study
- $\mathcal{A}_2$: Mirror region for 2st study
- $\mathcal{A}_1 \cup \mathcal{A}_2$: Control region

$$\Downarrow$$

Upper Bound

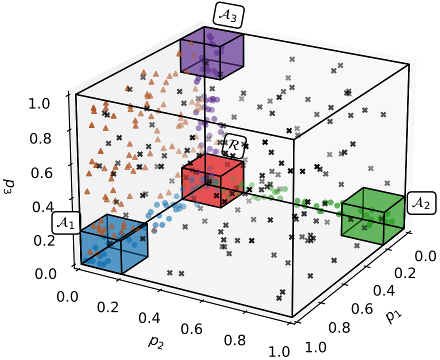
$$\begin{aligned}\text{1st} :&\#\{i : H_{i1} = 0, (p_{i1}, p_{i2}) \in \mathcal{R}\}\\ &\lesssim \#\{i : (p_{i1}, p_{i2}) \in \mathcal{A}_1\}\end{aligned}$$

$$\begin{aligned}&\#\{i : H_i = 0, (p_{i1}, p_{i2}) \in \mathcal{R}\}\\ \le&\#\{i : H_{i1} = 0, (p_{i1}, p_{i2}) \in \mathcal{R}\} + \#\{i : H_{i2} = 0, (p_{i1}, p_{i2}) \in \mathcal{R}\}\\ \lesssim&\#\{i : (p_{i1}, p_{i2}) \in \mathcal{A}_1\} + \#\{i : (p_{i1}, p_{i2}) \in \mathcal{A}_2\} = \#\{i : (p_{i1}, p_{i2}) \in \mathcal{A}_1 \cup \mathcal{A}_2\}\end{aligned}$$

Content

Define $\mathcal{A}_k = \{(t_1, \cdots, t_{k-1}, 1 - t_k, t_{k+1}, \cdots, t_K) : (t_1, \cdots, t_K) \in \mathcal{R}\}$

- $$\begin{aligned}\mathrm{FDP}(\mathcal{R}) &= \frac{|\#\{i : H_i = 0, (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}\}|}{\#\{i : (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}\} \vee 1} \\ &\lesssim \frac{|\#\{i : (p_{i1}, \cdots, p_{iK}) \in \cup_{k=1}^{K} \mathcal{A}_k\}|}{\#\{i : (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}\} \vee 1} := \widehat{\mathrm{FDP}}(\mathcal{R})\end{aligned}$$



- $K = 3$
- $\mathcal{R}$: Rejection region
- $\cup_{k=1}^{K} \mathcal{A}_k$: Control region

Content

Joint Mirror Procedure under FDR level α

1. Initialize $\mathcal{R}_0$, $\mathcal{A}_0^k$, $k = 1, \cdots, K$;
2. While $\widehat{\mathrm{FDP}}(\mathcal{R}_t) > \alpha$ and $\mathcal{R}_t \neq \emptyset$:
   1. Update $\mathcal{R}_{t+1}$ ($\mathcal{A}_{t+1}^k$) satisfying
      - P.I Shrinkage principle

        $$\mathcal{R}_{t+1} \subset \mathcal{R}_t$$

      - P.II Partial masking principle:

        $$\mathcal{R}_{t+1} \in \sigma\left\{(\tilde{\mathbf{p}}_{t,i})_{i=1}^m, A_t, R_t, \mathcal{M}_t\right\} := \mathcal{F}_t$$

        - $\tilde{\mathbf{p}}_{t,i} = \begin{cases} (\min(p_{i1}, 1-p_{i1}), \cdots, \min(p_{iK}, 1-p_{iK})), & \text{if } i \in \mathcal{M}_t \\ (p_{i1}, \cdots, p_{iK}), & \text{otherwise} \end{cases}$
        - Mask Set: $\mathcal{M}_t = \{i : (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}_t \cup \cup_{k=1}^K \mathcal{A}_t\}$.
        - Num of Rej: $R_t = \#\{i : (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}_t\}$.
        - Num of Control: $A_t = \#\{i : (p_{i1}, \cdots, p_{iK}) \in \cup_{k=1}^K \mathcal{A}_t^k\}$.
   2. Update $t = t + 1$.
3. Output: Select $\mathbf{p}_i \in \mathcal{R}_t$.

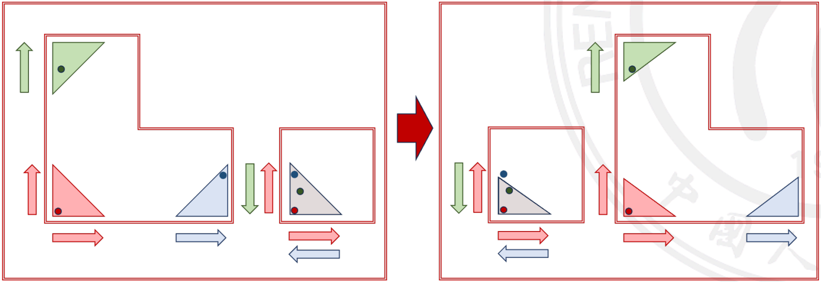
P.I Shrinkage $\Leftrightarrow$ Remove one from mask set $\mathcal{M}_t$ (in $\mathcal{R}_t$ or in $\cup_{k=1}^K \mathcal{A}_t^k$).

P.II Partial masking $\Leftrightarrow$ Selection subject to partial information $\mathcal{F}_t$.

$$\Uparrow$$

Given $\mathcal{F}_t$, select i that is most possible to be in $\cup_{k=1}^K \mathcal{A}_t^k$

$$\widehat{\mathrm{FDP}}(\mathcal{R})\downarrow = \frac{|\#\{i : (p_{i1}, \cdots, p_{iK}) \in \cup_{k=1}^K \mathcal{A}_k\}|\downarrow}{\#\{i : (p_{i1}, \cdots, p_{iK}) \in \mathcal{R}\} \vee 1}$$



Kernel Regression for Choosing i

- Any method: kernel regression $\mathbf{1}\{(p_{i1}, \cdots, p_{ik}) \in \mathcal{R}_t\}$.
  - Interpretable + Computational Efficient.
- Target: Let $\tilde{\mathbf{p}} = \mathrm{Proj}(\mathbf{p}) = (\min(p_1, 1-p_1), \cdots, \min(p_K, 1-p_K))$

$$\text{Select } i^\star = \arg\min_{i \in \mathcal{M}_t} q_i \quad \text{with} \quad q_i = \mathbb{P}\left(\mathbf{p} = \tilde{\mathbf{p}}_i \mid \tilde{\mathbf{p}} = \tilde{\mathbf{p}}_i\right)$$

- A kernel-based estimator for q_i at step t defined as

$$\hat{q}_{t,i} = \frac{\sum_{i' \in \mathcal{U}_t \cap \mathcal{M}_{-1}} \mathbf{1}\left\{\mathbf{p}_{i'} \in \mathcal{R}_{-1}\right\} v_H\left(\tilde{\mathbf{p}}_i, \tilde{\mathbf{p}}_{i'}\right)}{\sum_{i' \in \mathcal{U}_t \cap \mathcal{M}_{-1}} v_H\left(\tilde{\mathbf{p}}_i, \tilde{\mathbf{p}}_{i'}\right)}, i = 1, \ldots, m$$

  - $\mathcal{U}_t = [m] \backslash \mathcal{M}_t$, $\mathcal{M}_{-1} = \left\{i : \sum_{k=1}^K \mathbf{1}(p_{ki} \leq 1/2) \leq 1\right\}$.
  - $v_H(\mathbf{x}, \mathbf{x}') = \mathcal{K}_H(\mathbf{x} - \mathbf{x}') / \mathcal{K}_H(\mathbf{0})$
    - $\mathcal{K}_H(\mathbf{t}) = \det(H)^{-1/2} \mathcal{K}\left(H^{-1/2}\mathbf{t}\right)$
    - $H \in \mathbb{R}^{K \times K}$ is a positive definite bandwidth matrix
    - $\mathcal{K} : \mathbb{R}^K \to \mathbb{R}$ is a positive, bounded and symmetric kernel function.
- Complexity: Update the denominator & enumerator separately.

Partial Order and 3 Realizations

- Intuition: When $K = 1$, $\mathcal{R} = [0, t]$.
- ⋆ Partial order + Kernel regression.
  - P JM.Max: $(t_1, \cdots, t_K) \prec (s_1, \cdots, s_K)$ iff $\max t_k < \max s_k$.
  - K+P JM.Product: $(t_1, \cdots, t_K) \prec (s_1, \cdots, s_K)$ iff $t_1 < s_1, \cdots, t_K < s_K$.
  - K JM.EmptyPoset: non-comparable.

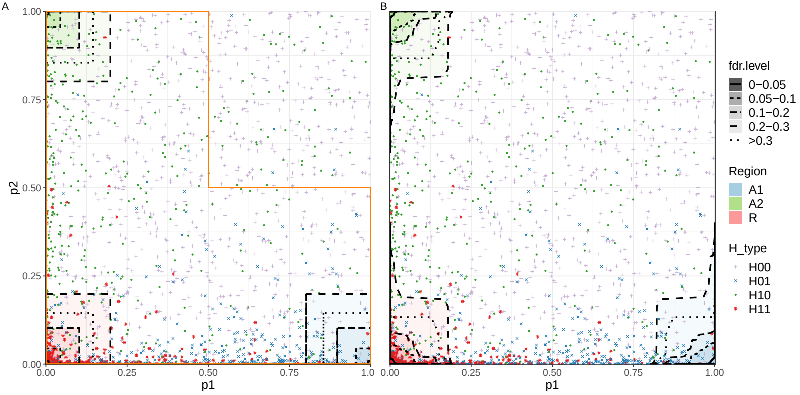
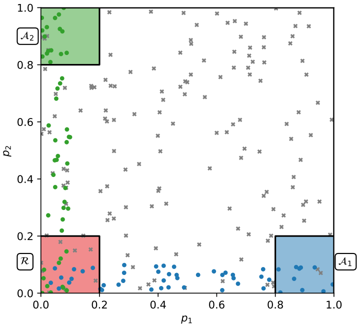


图: Left: JM.Max Right: JM.Product

Content

⋆ How conservative our estimator is?



$$\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 0, H_{i2} = 1, \mathbf{p}_i \in \mathcal{A}_1 \cup \mathcal{A}_2\right\}$$

$$+\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 1, H_{i2} = 0, \mathbf{p}_i \in \mathcal{A}_1 \cup \mathcal{A}_2\right\}$$

$$+\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 0, H_{i2} = 0, \mathbf{p}_i \in \mathcal{A}_1 \cup \mathcal{A}_2\right\}$$

$$\approx\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 0, H_{i2} = 1, \mathbf{p}_i \in \mathcal{R}\right\}$$

$$+\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 1, H_{i2} = 0, \mathbf{p}_i \in \mathcal{R}\right\}$$

$$+2\times\sum_{i=1}^{m} \mathbf{1}\left\{H_{i1} = 0, H_{i2} = 0, \mathbf{p}_i \in \mathcal{R}\right\}$$

⋆ If $(H_{i1}, H_{i2}) = (0, 0)$

⇒ The estimated FD approx. counted twice.

Composite FDR: Formal Definition

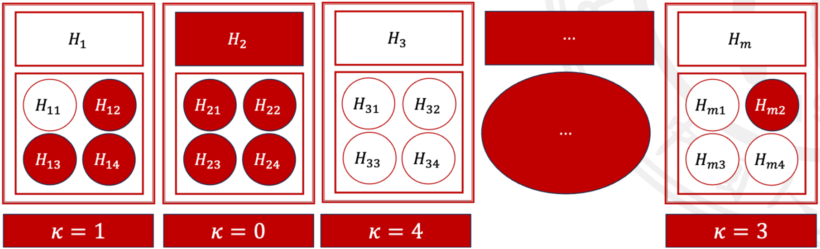
Define $\mathcal{H}^{(\kappa)} = \left\{ i \in [m] : \sum_{k=1}^{K} \mathbf{1}\{H_{ik} = 0\} = \kappa \right\}$ for $\kappa = 0, 1, \ldots, K$.

定义 1 (Composite FDR)

The composite FDR is defined as

$$\mathrm{cFDR}(\widehat{\mathcal{S}}) = \mathbb{E}\{\mathrm{cFDP}(\widehat{\mathcal{S}})\} = \mathbb{E}\left\{ \frac{\sum_{\kappa=1}^{K} \kappa \left| \widehat{\mathcal{S}} \cap \mathcal{H}^{(\kappa)} \right|}{|\widehat{\mathcal{S}}| \vee 1} \right\}$$

where the numerator is the number of weighted false discoveries.



Content

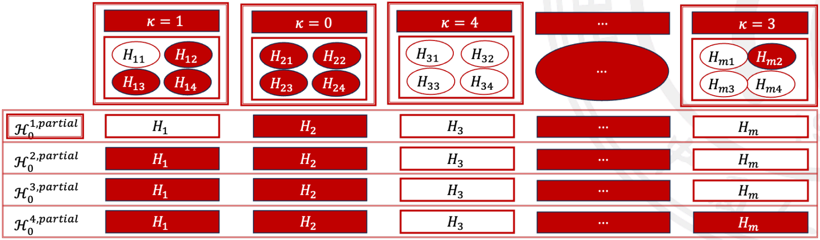
Content

Beyond Simultaneous Signals

Partial Conjunction Hypothesis (Friston et al., 2005, PCH)

Null hypotheses defined as $\mathcal{H}_0^{\kappa,\text{partial}} = \cup_{k=\kappa}^{K} \mathcal{H}^{(k)}$,

- where $\mathcal{H}^{(\kappa)} = \left\{ i \in [m] : \sum_{k=1}^{K} \mathbf{1}\{H_{ik} = 0\} = \kappa \right\}$ for $\kappa = 0, 1, \ldots, K$.
- **Special cases**: $\mathcal{H}_0 = \mathcal{H}_0^{1,\text{partial}}$ and $\mathcal{H}_1 = \mathcal{H}_1^{1,\text{partial}} = \mathcal{H}^{(0)}$.
- **Intuition**: $i \in \mathcal{H}_0^{\kappa,\text{partial}} \Rightarrow \#\{\text{null components of } \{H_{ik}\}_{k=1}^{K}\} \geq \kappa$.



定理 2 (Finite Sample FDR Control for PCH)

Suppose that (a) Independence. Suppose

$$\mathbf{p}_i \perp \mathbf{p}_j \text{ for } i, j \in \mathcal{H}_0^{\kappa,partial} \;+\; \{\mathbf{p}_i\}_{i\in\mathcal{H}_0^{\kappa,\text{partial}}} \perp \{\mathbf{p}_i\}_{i\in\mathcal{H}_1^{\kappa,\text{partial}}}$$

(b) $\mathbf{p}_i$ *is conditionally mirror conservative for all* $i \in \mathcal{H}_0^{\kappa,partial}$.

$\Rightarrow$ *Under* $\mathcal{H}_0^{\kappa,partial}$, $\mathrm{FDR} \leq q/\kappa$ $\Rightarrow$ *Under* $\mathcal{H}_0$, $\mathrm{FDR} \leq q$.

Martingale Technique.(Barber and Candès, 2015) Notice that

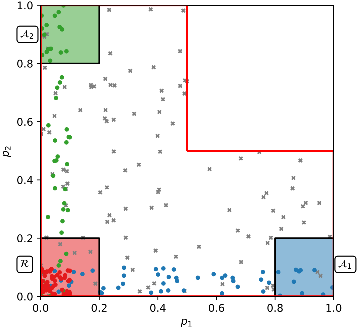
$$\begin{aligned}
\mathbb{E}\{\mathrm{FDP}(\mathcal{R}_\tau)\} &= \mathbb{E}\left\{\widehat{\mathrm{FDP}}(\mathcal{R}_\tau)\frac{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\mathbf{1}(\mathbf{p}_i\in\mathcal{R}_\tau)}{\sum_{k=1}^{K}\sum_{i=1}^{m}\mathbf{1}(\mathbf{p}_i\in\mathcal{A}_\tau^k)+1}\right\} \\
&\overset{1}{\leq} q\mathbb{E}\left\{\frac{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\mathbf{1}(\mathbf{p}_i\in\mathcal{R}_\tau)}{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\sum_{k=1}^{K}\mathbf{1}(\mathbf{p}_i\in\mathcal{A}_\tau{}^k)+1}\right\} \\
&\leq q\mathbb{E}\left\{\frac{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\mathbf{1}(\mathbf{p}_i\in\mathcal{R}_0)}{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\sum_{k=1}^{K}\mathbf{1}(\mathbf{p}_i\in\mathcal{A}_0{}^k)+1}\right\} \leq q/\kappa
\end{aligned}$$

[1] $\tau := \inf\{t : \widehat{\mathrm{FDP}}(\mathcal{R}_t) \leq q\}$ is a stopping time for a backwards super-martingale.

FDR Control

$$\mathbb{E}\left\{\frac{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\mathbf{1}\left(\mathbf{p}_i\in\mathcal{R}_0\right)}{\sum_{i\in\mathcal{H}_0^{\kappa,\text{partial}}}\sum_{k=1}^{K}\mathbf{1}\left(\mathbf{p}_i\in\mathcal{A}_0^k\right)+1}\right\}\le 1/\kappa$$

- $\mathbf{1}\left(\mathbf{p}_i\in\mathcal{R}_0\right)\mid\tilde{\mathbf{p}}_i\sim$ Bernoulli$(1/(\kappa+1))$.



Test $\mathcal{H}_0^{1,partial}$ / $\mathcal{H}_0$

- $\mathcal{H}_0^{1,partial}$: gray + blue + green
- $\mathcal{H}_1^{1,partial}$: red

Test $\mathcal{H}_0^{2,partial}$

- $\mathcal{H}_0^{2,partial}$: gray
- $\mathcal{H}_1^{2,partial}$: blue + green+red

What about cFDR? We require a stringent principle for selecting $\mathcal{R}$.

定理 3 (Finite Sample cFDR Control)

Under the same condition in Theorem 2,

- *If additionally* $\mathcal{R}_t \in \sigma\left\{(\tilde{\mathbf{p}}_i)_{i=1}^m\right\}$ *with* $\tilde{\mathbf{p}}_i = \mathrm{Proj}(\mathbf{p}_i)$, *e.g., JM.Max,*

$\Rightarrow$ $\mathrm{cFDR} \leq q$.

Leave-one-out Technique.(Benjamini and Hochberg, 1995; Barber et al., 2020) Notice that

$$\mathrm{cFDP} = \frac{\sum_{\kappa=1}^{K} \kappa \left|\widehat{\mathcal{S}} \cap \mathcal{H}^{(\kappa)}\right|}{|\widehat{\mathcal{S}}| \vee 1} = \frac{\sum_{i\in\mathcal{H}_0} \kappa_i \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{R}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right\}}{\sum_{i=1}^m \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{R}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right\} \vee 1},$$

$$\leq q \frac{\sum_{i\in\mathcal{H}_0} \kappa_i \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{R}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right\}}{1 + \sum_{i\in\mathcal{H}_0} \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right\}} := q\mathcal{E}\left(\mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right)$$

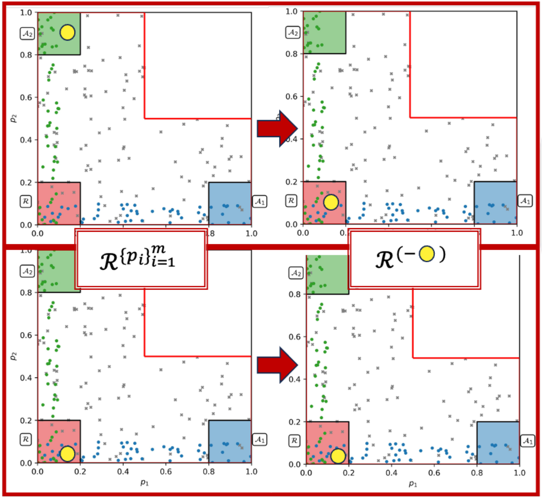
- $\mathcal{A}_{-1} = \cup_{k=1}^K \mathcal{A}_{-1}^k$ is the control side.
- $\mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}$ is the final rejection region with the p-values set $\{\mathbf{p}_i\}_{i=1}^m$.

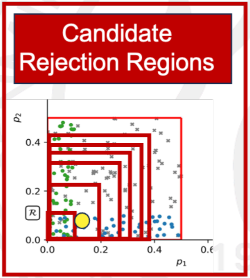
- Prove $\mathbb{E}\left\{\mathcal{E}\left(\mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right)\right\} \leq 1 \Rightarrow \mathrm{cFDR} \leq q$.

Proof of cFDR Control

Key 1 The leave-one-out p-values $\{\mathbf{p}_{(-i)}, \tilde{\mathbf{p}}_i\} = \{\mathbf{p}_1, \cdots, \mathbf{p}_{i-1}, \tilde{\mathbf{p}}_i, \cdots, \mathbf{p}_m\}$.

Key 2 $\mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}$ and $\mathcal{R}^{(-i)} = \mathcal{R}^{\mathbf{p}_{(-i)}, \tilde{\mathbf{p}}_i}$ should be one of the candidate rejection regions.





- $\mathbb{E}\left\{\mathcal{E}\left(\mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right)\right\}$ can be bounded as follows:

$$\xlongequal[\Rightarrow \mathbf{p}_i \notin \mathcal{A}_{-1}]{\mathbf{p}_i \in \mathcal{R}_{-1}} \sum_{i \in \mathcal{H}_0} \mathbb{E}\left(\frac{\kappa_i \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{R}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{\{\mathbf{p}_i\}_{i=1}^m}\right\}}{1+\sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{\{\mathbf{p}_j\}_{i=1}^m}\right\}}\right)$$

$$\xlongequal[\Rightarrow \mathbf{p}_i = \tilde{\mathbf{p}}_i]{\mathbf{p}_i \in \mathcal{R}_{-1}} \sum_{i \in \mathcal{H}_0} \mathbb{E}\left(\frac{\kappa_i \mathbf{1}\left\{\mathbf{p}_i \in \mathcal{R}_{-1}\right\} \mathbf{1}\left\{\tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}\right\}}{1+\sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-i)}\right\}}\right)$$

$$\xlongequal[law]{tower} \sum_{i \in \mathcal{H}_0} \mathbb{E}\left(\frac{\kappa_i \mathbb{P}\left\{\mathbf{p}_i \in \mathcal{R}_{-1} \mid \tilde{\mathbf{p}}_i, \mathbf{p}_{(-i)}\right\} \mathbf{1}\left\{\tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}\right\}}{1+\sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-i)}\right\}}\right)$$

$$\overset{mirror}{\leq} \sum_{i \in \mathcal{H}_0} \mathbb{E}\left(\frac{\mathbb{P}\left\{\mathbf{p}_i \in \mathcal{A}_{-1} \mid \tilde{\mathbf{p}}_i, \mathbf{p}_{(-i)}\right\} \mathbf{1}\left\{\tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}\right\}}{1+\sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-i)}\right\}}\right)$$

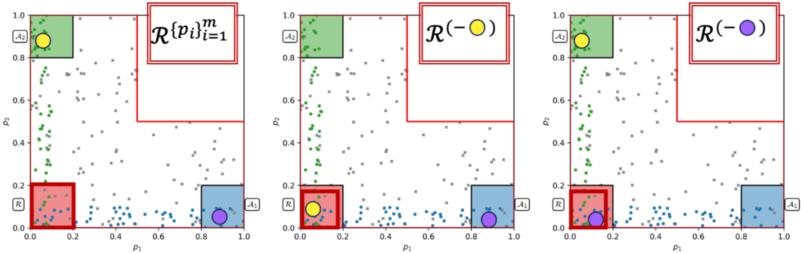
$$\xlongequal[law]{tower} \mathbb{E}\left(\sum_{i \in \mathcal{H}_0} \frac{\mathbf{1}\left\{\mathbf{p}_i \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}\right\}}{1+\sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-i)}\right\}}\right)$$

Proof of cFDR Control

$$
\sum_{i \in \mathcal{H}_0} \frac{\mathbf{1}\left\{\boxed{\mathbf{p}_i \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}}\right\}}{1 + \sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\boxed{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-i)}}\right\}}
$$

$$
\overset{(\star)}{=} \sum_{i \in \mathcal{H}_0} \frac{\mathbf{1}\left\{\boxed{\mathbf{p}_i \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_i \in \mathcal{R}^{(-i)}}\right\}}{1 + \sum_{j \in \mathcal{H}_0, j \neq i} \mathbf{1}\left\{\boxed{\mathbf{p}_j \in \mathcal{A}_{-1}, \tilde{\mathbf{p}}_j \in \mathcal{R}^{(-j)}}\right\}} \leq 1,
$$

- $(\star)$ holds because: under the condition $\square$, $\square \iff \square$.



Power Improvement

- Easy to notice $K\mathrm{FDR} \geq \mathrm{cFDR} \geq \mathrm{FDR}$.
- Will the JM procedure be too conservative?

An example with a general K.

Suppose $H_i = 0$ iff $H_{ki} = 0$ for all k. Let $m_0 = \#\{i : H_i = 0\}$

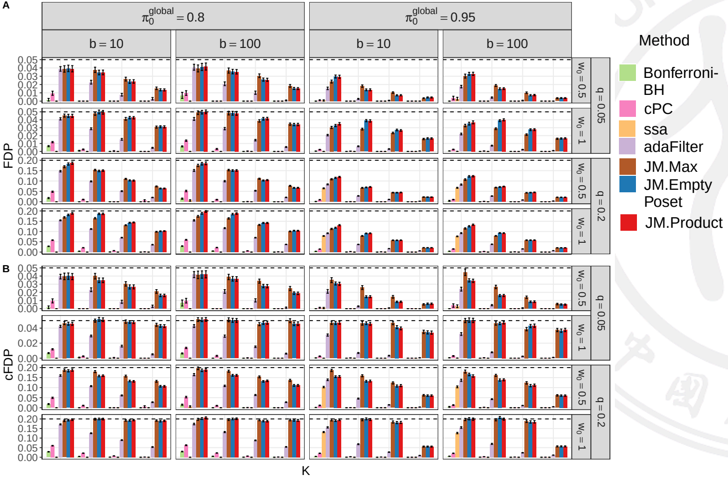
- $\mathbb{E}(\mathrm{FD}) = \sum_{i:H_i=0} \mathbb{P}\{\max(p_{i1}, \cdots, p_{iK}) \leq t\} = m_0 t^K$
- JM procedure: $\widehat{\mathrm{FD}}_{JM} \approx K m_0 t^K$
- Traditionally: $\mathbb{E}(\mathrm{FD}) \lesssim \sum_{i:H_i=0} \mathbb{P}\{p_{ik} \leq t \text{ for some } k\} = m_0 t$
- Recently: either (a) need to est m_0 precisely (b) no comprehensive analysis.

JM capture the correct power of t!

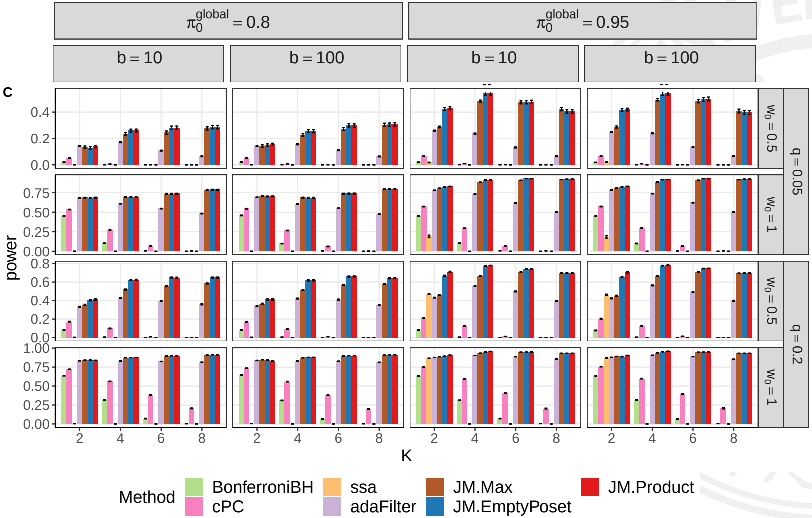
Content

97% null: FDR& cFDR

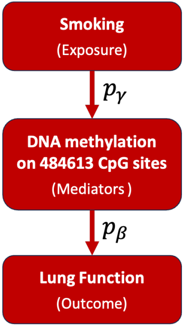
- K: num of study; π_0^{global}: prop. of K studies are null; b: strength of correlation; w_0: signal strength disequilibrium; q: FDR level.
- All methods can control FDR & cFDR;



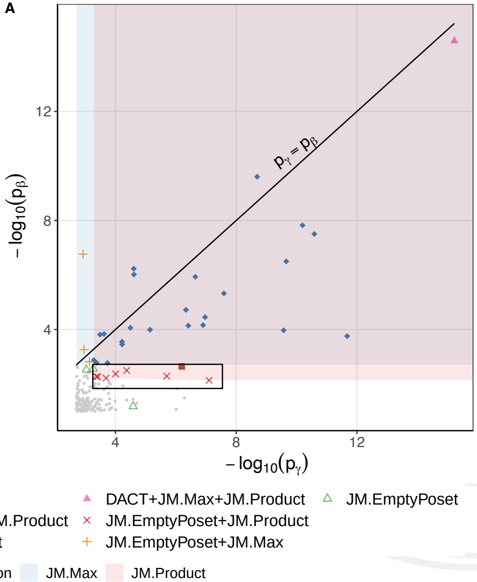
- Three JM methods deliver the highest power, esp, $K \geq 4$ or $w_0 = 0.5$.



Content



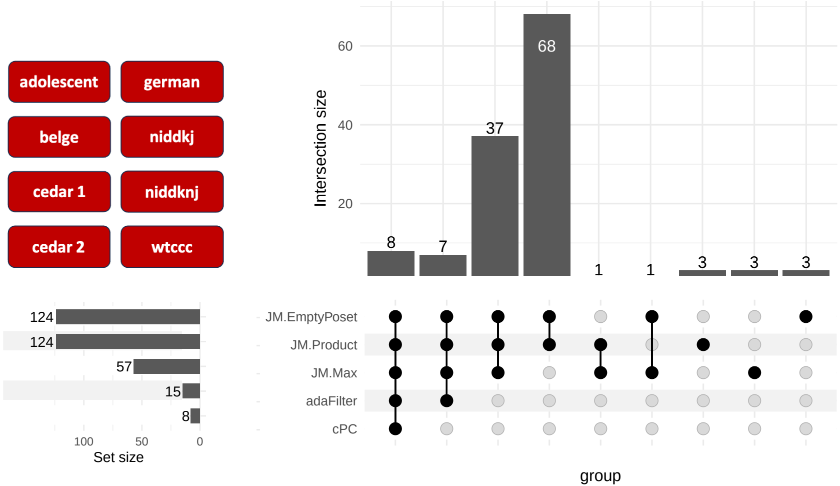
- FDR level: 0.2
- Used by DACT (Liu et al., 2021)
- DACT (21), JM.Max (27), JM.Product (32), JM.EmptyPoset (37)



GWAS for Crohn's Disease

- 8 studies + 953,241 SNPs + FDR level: 0.05.
- Collected by Franke et al. (2010), analyzed by cPC (Dickhaus et al., 2021) in PCH framework.
- Find SNPs that are statistically significant in all 8 studies.

adolescent	german
belge	niddkj
cedar 1	niddknj
cedar 2	wtccc



Content

Contribution

Targets

- ? Can we work on $(p_{i1}, p_{i2}, \cdots, p_{iK})$ directly?
- ☝ Construct K-dimensional rejection region.
- ? Should we treat $H_1, \cdots, H_m$ equally?
- ☝ Counting the number of null components – Composite FDR.

Contributions

- A new FDP estimator for simultaneous signals.
- A new error criterion – Composite FDR.
- Partial-order oriented data-driven rejection region. (Indeed reflect the data pattern.)
- FDR (marginal technique) and cFDR (leave-one-out technique) control in finite samples.

☺ Easy interpretation + Significant power improvement.

Future Directions

⋆ The following points highlight our future directions:

- Asymmetric mirror region;
- Incorporate side information;
- Apply on test statistics directly – Directional FDR;
  - Using Z-values can be more efficient (Leung and Sun, 2021; Leung, 2022).
  - Data splitting strategy (Dai et al., 2022; Guo et al., 2021)
  - Knockoff strategy (Dai and Zheng, 2023)
- Consider the dependency of Stat. in mediation analysis.
- Proper rejection region and less conservative procedure to control PCH.

Thank You!

Q & A

Reference I

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

Barber, R. F., Candes, E. J., and Samworth, R. J. (2020), "Robust inference with knockoffs," *The Annals of Statistics*, 48, 1409–1431.

Benjamini, Y. and 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, 289–300.

Dai, C., Lin, B., Xing, X., and Liu, J. S. (2022), "False Discovery Rate Control via Data Splitting," *Journal of the American Statistical Association*, 1–18.

Dai, J. Y., Stanford, J. L., and LeBlanc, M. (2020), "A multiple-testing procedure for high-dimensional mediation hypotheses," *Journal of the American Statistical Association*, 117, 198–213.

Dai, R. and Zheng, C. (2023), "False discovery rate-controlled multiple testing for union null hypotheses: a knockoff-based approach," *Biometrics*, 0, 1– 13.

Dickhaus, T., Heller, R., and Hoang, A.-T. (2021), "Multiple testing of partial conjunction null hypotheses with conditional p-values based on combination test statistics," *arXiv preprint arXiv:2110.06692*.

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., Lees, C. W., Balschun, T., Lee, J., and Roberts, R. (2010), "Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci," *Nature Genetics*, 42, 1118–1125.

Friston, K. J., Penny, W. D., and Glaser, D. E. (2005), "Conjunction revisited," *Neuroimage*, 25, 661–667.

Guo, X., Ren, H., Zou, C., and Li, R. (2021), "Threshold selection in feature screening for error rate control," *Journal of the American Statistical Association*, 1–37.

Reference II

Lei, L. and Fithian, W. (2016), "AdaPT: An interactive procedure for multiple testing with side information," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 649–679.

Leung, D. (2022), "*Z*-value directional false discovery rate control with data masking," *arXiv preprint arXiv:2201.05828*.

Leung, D. and Sun, W. (2021), "ZAP: *Z*-value adaptive procedures for false discovery rate control with side information," *arXiv preprint arXiv:2108.12623*.

Liu, Z., Shen, J., Barfield, R., Schwartz, J., Baccarelli, A. A., and Lin, X. (2021), "Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies," *Journal of the American Statistical Association*, 117, 67–81.

Tukey, J. W. (1953), "The problem of multiple comparisons." .

Wang, J., Gui, L., Su, W. J., Sabatti, C., and Owen, A. B. (2022), "Detecting multiple replicating signals using adaptive filtering procedures," *The Annals of Statistics*, 50, 1890–1909.