

# AdaPT: Interactive Multiple Testing with Side Information

Lihua Lei & Will Fithian

Department of Statistics, UC Berkeley

February 21, 2018

# Table of Contents

- ① Background and Motivation
- ② Review of Existing Methods
- ③ AdaPT: Adaptive  $p$ -Value Thresholding
- ④ Updating the Threshold
- ⑤ Applications
- ⑥ Extensions

# Table of Contents

- ① Background and Motivation
- ② Review of Existing Methods
- ③ AdaPT: Adaptive  $p$ -Value Thresholding
- ④ Updating the Threshold
- ⑤ Applications
- ⑥ Extensions

# Multiple Hypothesis Testing

Setting: hypotheses  $H_1, \dots, H_n$  with  $p$ -values  $p_1, \dots, p_n$

Notation:

- $\mathcal{H}_0 = \{i : H_i \text{ is true}\}$ : null hypotheses
- $\mathcal{S} = \{i : H_i \text{ is rejected}\}$ : set of rejections (discoveries)
- $R = |\mathcal{S}|$  total rejections
- $V = |\mathcal{S} \cap \mathcal{H}_0|$  incorrect rejections

False Discovery Proportion  $\text{FDP} = \frac{V}{R \vee 1}$

Goal: control False Discovery Rate [Benjamini and Hochberg, 1995]

$$\text{FDR} = \mathbb{E}[\text{FDP}] \leq \alpha$$

# Multiple Hypothesis Testing

Setting: hypotheses  $H_1, \dots, H_n$  with  $p$ -values  $p_1, \dots, p_n$

Notation:

- $\mathcal{H}_0 = \{i : H_i \text{ is true}\}$ : null hypotheses
- $\mathcal{S} = \{i : H_i \text{ is rejected}\}$ : set of rejections (discoveries)
- $R = |\mathcal{S}|$  total rejections
- $V = |\mathcal{S} \cap \mathcal{H}_0|$  incorrect rejections

False Discovery Proportion  $\text{FDP} = \frac{V}{R \vee 1}$

Goal: control False Discovery Rate [Benjamini and Hochberg, 1995]

$$\text{FDR} = \mathbb{E}[\text{FDP}] \leq \alpha$$

Want control in finite samples. No asymptotics.

## Other Criteria

- Familywise error (FWER) control:

$$\text{FWER} \triangleq \mathbb{P}(V \geq 1);$$

- Marginal FDR control:

$$\text{FDR}_{\text{mar}} \triangleq \frac{\mathbb{E}V}{\mathbb{E}R};$$

- Modified FDR control:

$$\text{FDR}_m \triangleq \mathbb{E} \left[ \frac{V}{R + \alpha^{-1}} \right];$$

- FDX control:

$$\text{FDX} \triangleq \mathbb{P}(\text{FDP} \geq \alpha)$$

## Side Information

Observe side information  $x_i \in \mathcal{X}$  for each  $H_i$  [Ferkingstad et al., 2008, Ignatiadis et al., 2016]

$x_1, \dots, x_n$  treated as fixed

Ordered multiple testing [Foster and Stine, 2008, G'Sell et al., 2015]

- $H_1$  most “promising,” then  $H_2, \dots, H_n$  ( $x_i = i$ )
- Focus power on early hypotheses

## Side Information

Observe side information  $x_i \in \mathcal{X}$  for each  $H_i$  [Ferkingstad et al., 2008, Ignatiadis et al., 2016]

$x_1, \dots, x_n$  treated as fixed

Ordered multiple testing [Foster and Stine, 2008, G'Sell et al., 2015]

- $H_1$  most “promising,” then  $H_2, \dots, H_n$  ( $x_i = i$ )
- Focus power on early hypotheses

Other examples:

- Data from a similar experiment
- Spatiotemporal location e.g.  $H_i : f(t_i) \leq 0$
- “Collaborative filtering” e.g.  
 $H_{ij} : \text{gene } i \text{ is associated with disease } j$

## Side Information

Observe **side information**  $x_i \in \mathcal{X}$  for each  $H_i$  [Ferkingstad et al., 2008, Ignatiadis et al., 2016]

$x_1, \dots, x_n$  treated as fixed

**Ordered multiple testing** [Foster and Stine, 2008, G'Sell et al., 2015]

- $H_1$  most “promising,” then  $H_2, \dots, H_n$  ( $x_i = i$ )
- Focus power on early hypotheses

Other examples:

- Data from a similar experiment
- Spatiotemporal location e.g.  $H_i : f(t_i) \leq 0$
- “Collaborative filtering” e.g.  
 $H_{ij} : \text{gene } i \text{ is associated with disease } j$

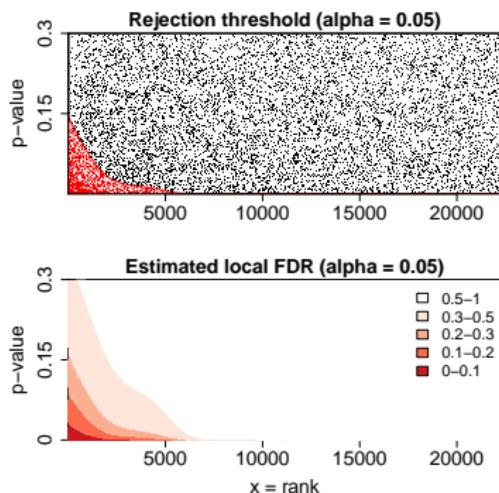
**Idea:** if we **learn** a region of  $\mathcal{X}$  has many non-nulls, can relax multiplicity correction in that region

## Gene/Drug Response Data

- Detect differential expression in breast cancer cells in response to estrogen;
- $n = 22283$  genes, 25 trials at 5 doses incl. control;
- $H_i$  : no differential response in low-dose vs. control;
- $p_i$ : permutation  $t$ -test;
- $x_i$ : rank obtained from auxiliary data.

# Gene/Drug Response Data

- Detect differential expression in breast cancer cells in response to estrogen;
- $n = 22283$  genes, 25 trials at 5 doses incl. control;
- $H_i$  : no differential response in low-dose vs. control;
- $p_i$ : permutation  $t$ -test;
- $x_i$ : rank obtained from auxiliary data.

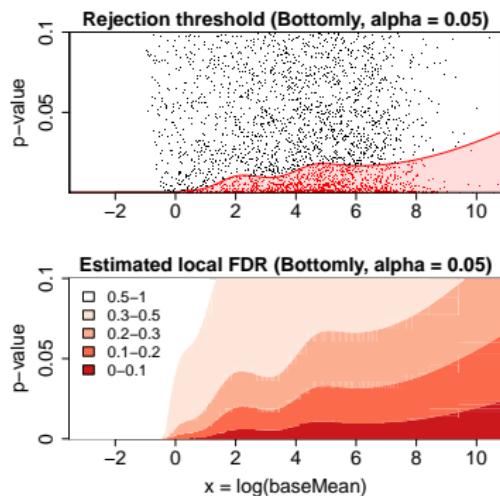


## RNA-seq Data

- Detect differential expression on two mouse strains;
- $n = 13932$  genes;
- $H_i$  : no differential expression of gene  $i$ ;
- $p_i$ : from DEseq2 package;
- $x_i$ : log. normalized count.

# RNA-seq Data

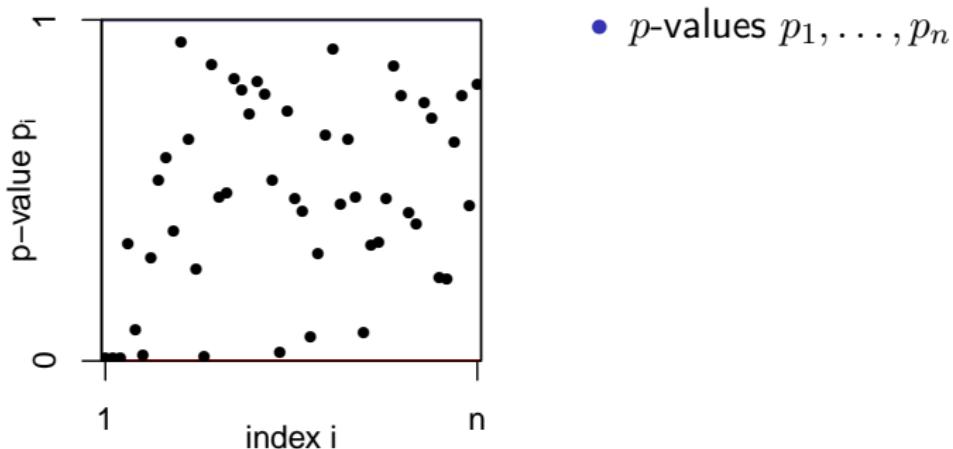
- Detect differential expression on two mouse strains;
- $n = 13932$  genes;
- $H_i$  : no differential expression of gene  $i$ ;
- $p_i$ : from DEseq2 package;
- $x_i$ : log. normalized count.



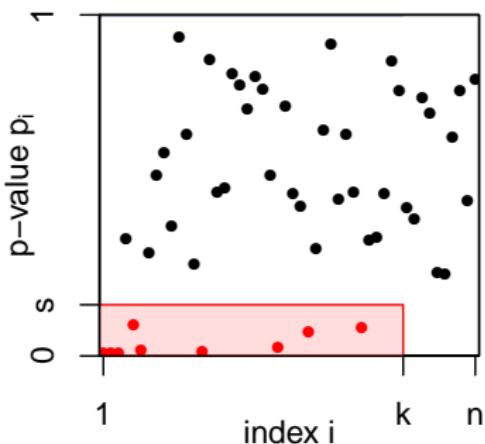
# Table of Contents

- 1 Background and Motivation
- 2 Review of Existing Methods
- 3 AdaPT: Adaptive  $p$ -Value Thresholding
- 4 Updating the Threshold
- 5 Applications
- 6 Extensions

## Review of Existing Methods: General Recipe

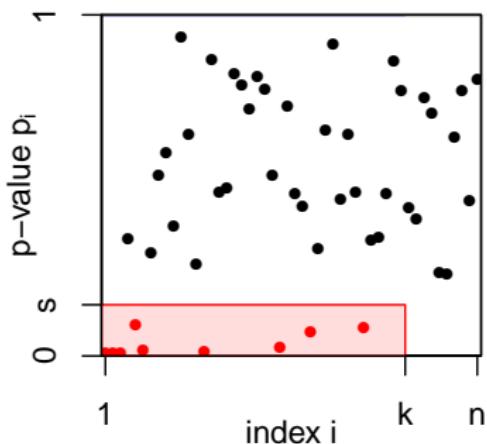


## Review of Existing Methods: General Recipe



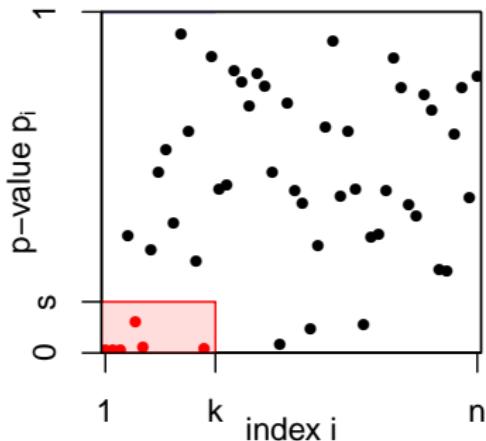
- $p$ -values  $p_1, \dots, p_n$
- **Rejection Set:** a rectangular region indexed by  $s \in [0, 1]$  and  $k \in \{1, \dots, n\}$

## Review of Existing Methods: General Recipe



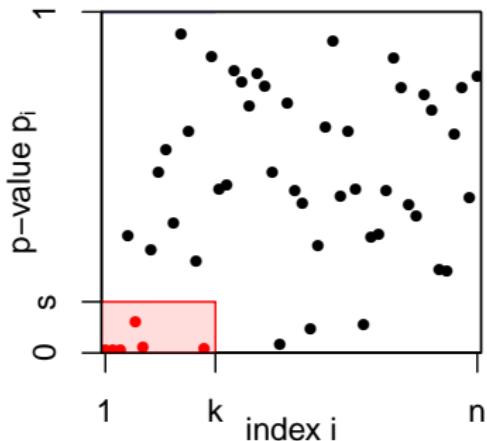
- $p$ -values  $p_1, \dots, p_n$
- **Rejection Set:** a rectangular region indexed by  $s \in [0, 1]$  and  $k \in \{1, \dots, n\}$
- Estimator  $\widehat{\text{FDP}}(k, s)$

## Review of Existing Methods: General Recipe



- $p$ -values  $p_1, \dots, p_n$
- **Rejection Set:** a rectangular region indexed by  $s \in [0, 1]$  and  $k \in \{1, \dots, n\}$
- Estimator  $\widehat{\text{FDP}}(k, s)$
- Gradually reduce  $k$  or  $s$  until  $\widehat{\text{FDP}} \leq \alpha$

## Review of Existing Methods: General Recipe

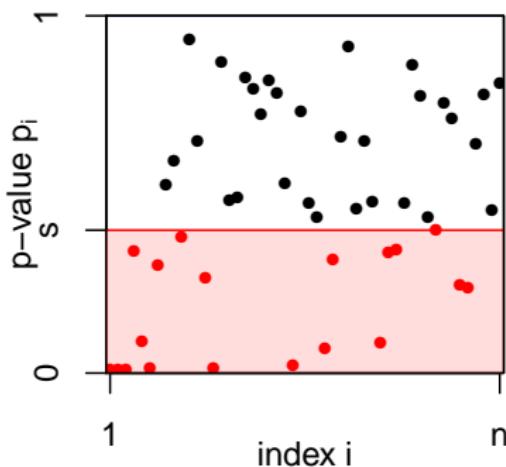


- $p$ -values  $p_1, \dots, p_n$
- **Rejection Set:** a rectangular region indexed by  $s \in [0, 1]$  and  $k \in \{1, \dots, n\}$
- Estimator  $\widehat{\text{FDP}}(k, s)$
- Gradually reduce  $k$  or  $s$  until  $\widehat{\text{FDP}} \leq \alpha$
- Reject  $\{H_i : i \leq \hat{k}, p_i < \hat{s}\}$  (red points)

## BH Procedure [Benjamini and Hochberg, 1995]

$$\widehat{\text{FDP}}_{\text{BH}} = \frac{ns}{R(s)}$$

### Benjamini–Hochberg

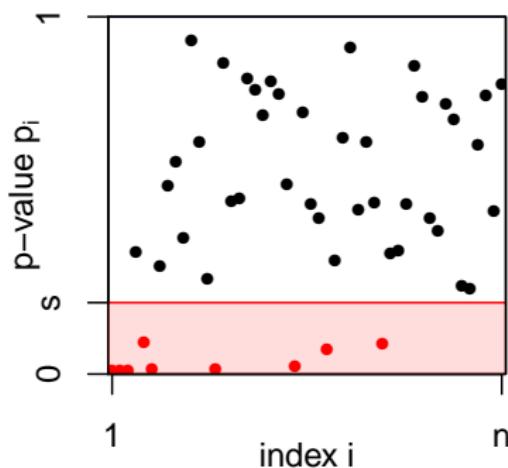


- $R(s) = |\{i : p_i \leq s\}|$
- $\text{FDR} \leq \pi_0 \alpha$
- $\pi_0$  : proportion of nulls

## BH Procedure [Benjamini and Hochberg, 1995]

$$\widehat{\text{FDP}}_{\text{BH}} = \frac{ns}{R(s)}$$

### Benjamini–Hochberg

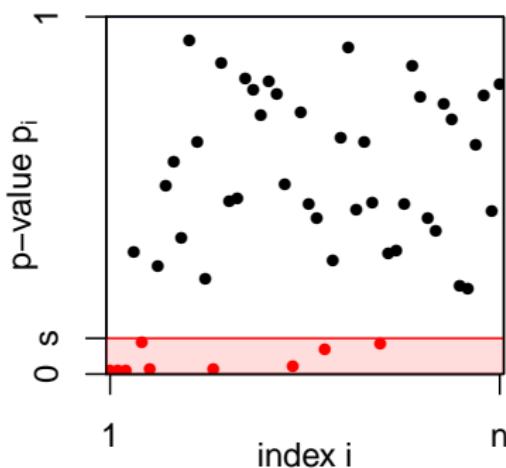


- $R(s) = |\{i : p_i \leq s\}|$
- $\text{FDR} \leq \pi_0 \alpha$
- $\pi_0$  : proportion of nulls

## BH Procedure [Benjamini and Hochberg, 1995]

$$\widehat{\text{FDP}}_{\text{BH}} = \frac{ns}{R(s)}$$

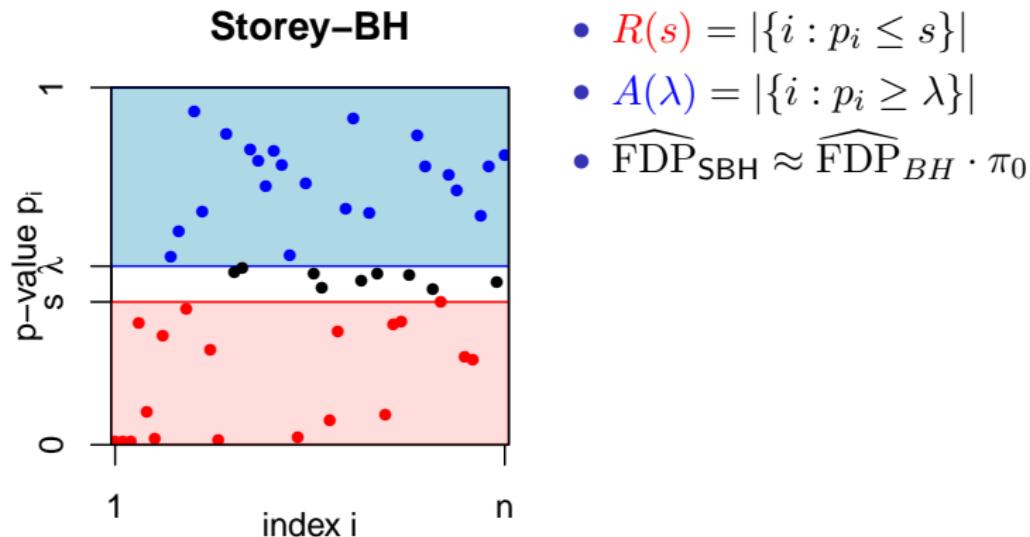
### Benjamini–Hochberg



- $R(s) = |\{i : p_i \leq s\}|$
- $\text{FDR} \leq \pi_0 \alpha$
- $\pi_0$  : proportion of nulls

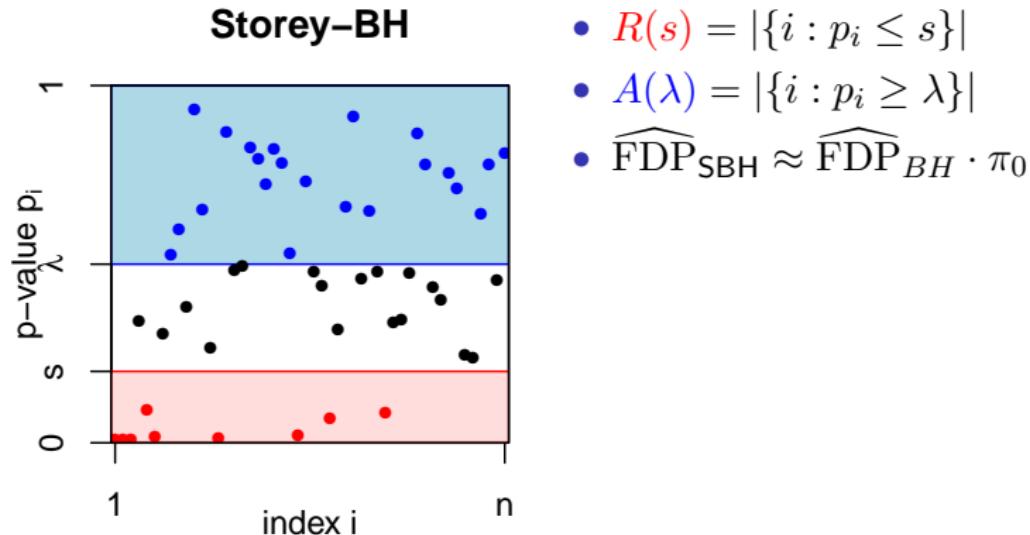
## Storey–BH Procedure [Storey et al., 2004]

$$\widehat{\text{FDP}}_{\text{SBH}} = \frac{ns}{R(s)} \cdot \frac{A(\lambda) + 1}{(1 - \lambda)n}$$



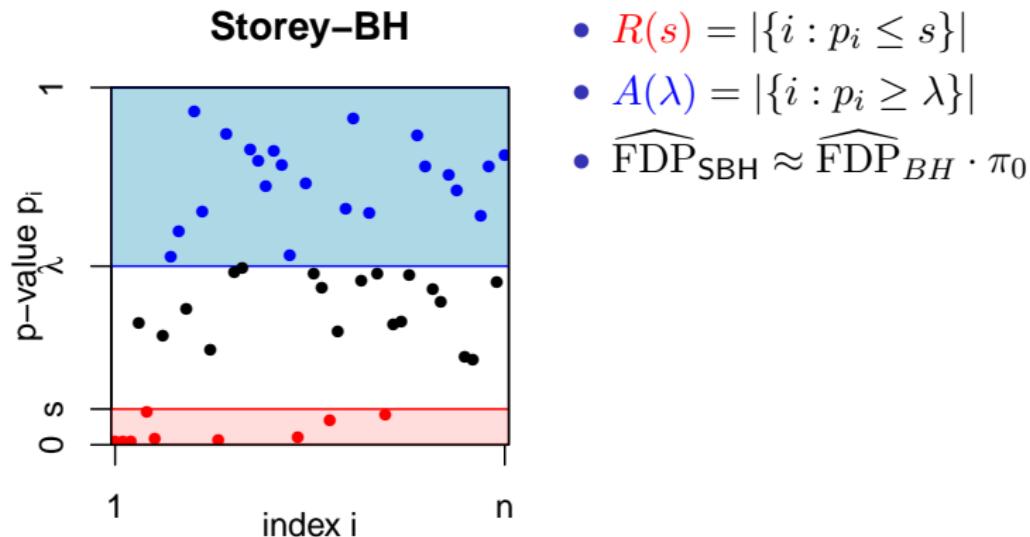
## Storey–BH Procedure [Storey et al., 2004]

$$\widehat{\text{FDP}}_{\text{SBH}} = \frac{ns}{R(s)} \cdot \frac{A(\lambda) + 1}{(1 - \lambda)n}$$



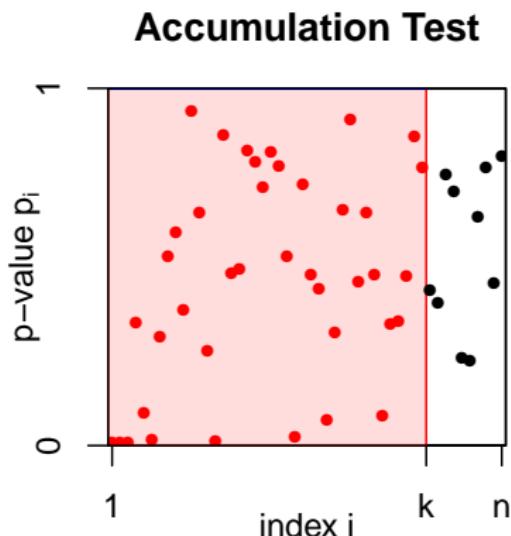
## Storey–BH Procedure [Storey et al., 2004]

$$\widehat{\text{FDP}}_{\text{SBH}} = \frac{ns}{R(s)} \cdot \frac{A(\lambda) + 1}{(1 - \lambda)n}$$



## Accumulation Test [Li and Barber, 2016a]

$$\widehat{\text{FDP}}_{\text{AT}} = \frac{C + \sum_{i=1}^k h(p_i)}{k+1}$$



- $h \geq 0, \int_0^1 h(x)dx = 1;$
- ForwardStop [G'Sell et al., 2015]:

$$h(x) = -\log(1-x);$$

- Seqstep [Barber and Candès, 2015]:

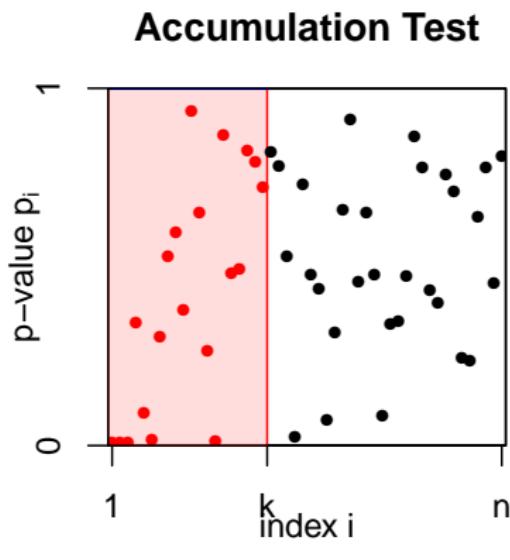
$$h(x) = \frac{I(x > \lambda)}{1-\lambda};$$

- HingeExp [Li and Barber, 2016a]:

$$h(x) = -\frac{I(x > \lambda)}{1-\lambda} \log\left(\frac{1-x}{1-\lambda}\right).$$

## Accumulation Test [Li and Barber, 2016a]

$$\widehat{\text{FDP}}_{\text{AT}} = \frac{C + \sum_{i=1}^k h(p_i)}{k + 1}$$



- $h \geq 0, \int_0^1 h(x)dx = 1;$
- ForwardStop [G'Sell et al., 2015]:

$$h(x) = -\log(1 - x);$$

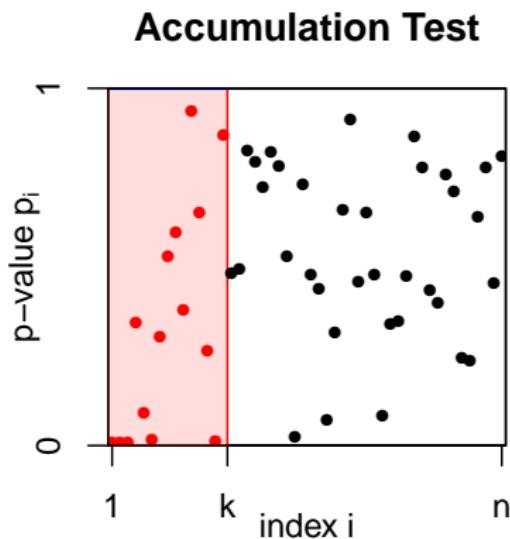
- Seqstep [Barber and Candès, 2015]:
- HingeExp [Li and Barber, 2016a]:

$$h(x) = \frac{I(x > \lambda)}{1 - \lambda};$$

$$h(x) = -\frac{I(x > \lambda)}{1 - \lambda} \log\left(\frac{1 - x}{1 - \lambda}\right).$$

## Accumulation Test [Li and Barber, 2016a]

$$\widehat{\text{FDP}}_{\text{AT}} = \frac{C + \sum_{i=1}^k h(p_i)}{k + 1}$$



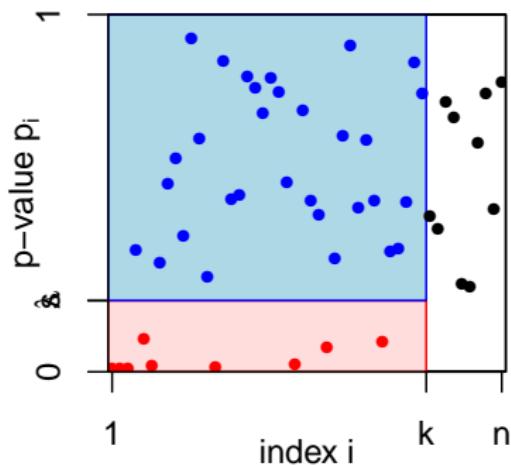
- $h \geq 0, \int_0^1 h(x)dx = 1;$
- ForwardStop [G'Sell et al., 2015]:  
$$h(x) = -\log(1 - x);$$
- Seqstep [Barber and Candès, 2015]:  
$$h(x) = \frac{I(x > \lambda)}{1 - \lambda};$$
- HingeExp [Li and Barber, 2016a]:

$$h(x) = -\frac{I(x > \lambda)}{1 - \lambda} \log\left(\frac{1 - x}{1 - \lambda}\right).$$

## Selective SeqStep [Barber and Candès, 2015]

$$\widehat{\text{FDP}}_{\text{SS}} = \frac{ks}{R(k,s)} \cdot \frac{A(k,s) + 1}{k(1-s)}$$

### Selective SeqStep

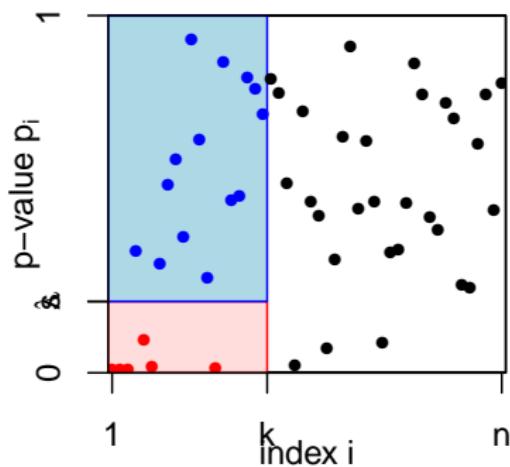


- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

## Selective SeqStep [Barber and Candès, 2015]

$$\widehat{\text{FDP}}_{\text{SS}} = \frac{ks}{R(k, s)} \cdot \frac{A(k, s) + 1}{k(1 - s)}$$

### Selective SeqStep

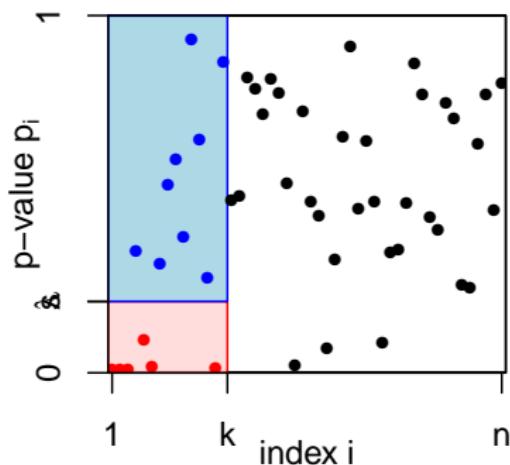


- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

## Selective SeqStep [Barber and Candès, 2015]

$$\widehat{\text{FDP}}_{\text{SS}} = \frac{ks}{R(k, s)} \cdot \frac{A(k, s) + 1}{k(1 - s)}$$

### Selective SeqStep

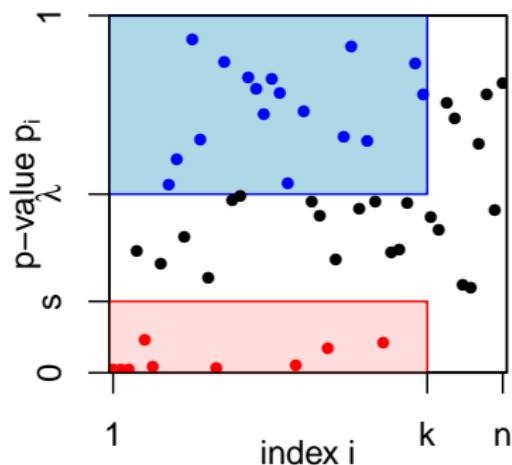


- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

## Adaptive SeqStep [Lei and Fithian, 2016]

$$\widehat{\text{FDP}}_{\text{AS}} = \frac{ks}{R(k, s)} \cdot \frac{A(k, \lambda) + 1}{k(1 - \lambda)}$$

### Adaptive SeqStep



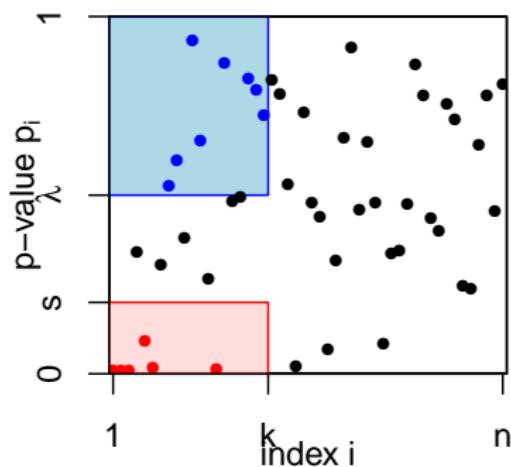
- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

Can be much more powerful if  
 $\pi_0(k) \ll 1$

## Adaptive SeqStep [Lei and Fithian, 2016]

$$\widehat{\text{FDP}}_{\text{AS}} = \frac{ks}{R(k, s)} \cdot \frac{A(k, \lambda) + 1}{k(1 - \lambda)}$$

### Adaptive SeqStep



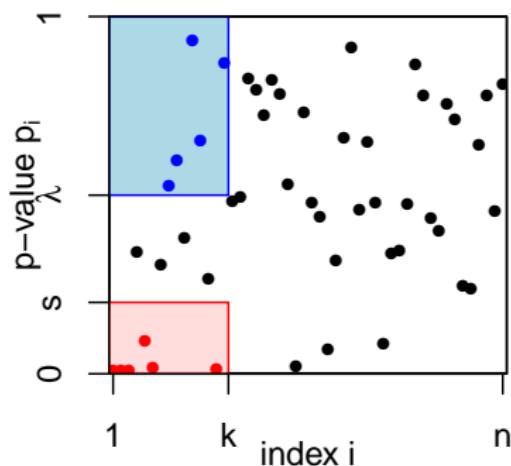
- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

Can be much more powerful if  
 $\pi_0(k) \ll 1$

## Adaptive SeqStep [Lei and Fithian, 2016]

$$\widehat{\text{FDP}}_{\text{AS}} = \frac{ks}{R(k, s)} \cdot \frac{A(k, \lambda) + 1}{k(1 - \lambda)}$$

### Adaptive SeqStep



- $R(k, s) = |\{i \leq k : p_i \leq s\}|$
- $A(k, s) = |\{i \leq k : p_i \geq s\}|$

Can be much more powerful if  
 $\pi_0(k) \ll 1$

## Related Work: Multiple Testing With Side Information

### Independent Hypothesis Weighting (IHW) [Ignatiadis et al., 2016]

- Based on weighted BH (replace  $p_i$  by  $p_i/W_i$  for some weights  $W_i$ );
- Bin  $x_i$ , estimate optimal stepwise rejection thresholds;
- avoid over-fitting by  $k$ -fold cross-fitting;
- **asymptotic FDR control** under **strong modelling assumptions**
  - $p_i$ 's are i.i.d. within each strata;
  - $p_i \mid x_i \sim$  a two-group model;
  - positive alternative density;
- stable binning requires **low-dimensional** covariate.

## Related Work: Multiple Testing With Side Information

IHWc [Ignatiadis and Huber, 2017]

- Learn optimal thresholds using  $\{p_i I(p_i \geq \tau), x_i\}$ ;
- avoid over-fitting by  $k$ -fold cross-fitting;
- finite-sample FDR control (no asymptotics);
- limited learnability using large p-values:
  - are sufficient for learning the proportion of non-nulls;
  - but contain little information of alternative distributions.

## Related Work: Multiple Testing With Side Information

IHWc [Ignatiadis and Huber, 2017]

- Learn optimal thresholds using  $\{p_i I(p_i \geq \tau), x_i\}$ ;
- avoid over-fitting by  $k$ -fold cross-fitting;
- finite-sample FDR control (no asymptotics);
- limited learnability using large p-values:
  - are sufficient for learning the proportion of non-nulls;
  - but contain little information of alternative distributions.

Structure-Adaptive BH Algorithm (SABHA): [Li and Barber, 2016b]

- Similar to IHWc except that no cross-fitting is used;
- finite-sample FDR control (after correcting for Rademacher complexity);
- limited learnability (see above)

## Related Work: Multiple Testing With Side Information

### NeuralFDR [Xia et al., 2017]

- avoid over-fitting by  $k$ -fold cross-fitting;
- smart calibration step to improve power;
- **asymptotic FDX control** under **strong modelling assumptions**
  - $(x_i, p_i)$  are i.i.d.;
  - $p_i \mid x_i \sim$  a two-group model;
  - positive alternative density (implicit in the proof);

## Related Work: Multiple Testing With Side Information

### NeuralFDR [Xia et al., 2017]

- avoid over-fitting by  $k$ -fold cross-fitting;
- smart calibration step to improve power;
- asymptotic FDX control under strong modelling assumptions
  - $(x_i, p_i)$  are i.i.d.;
  - $p_i \mid x_i \sim$  a two-group model;
  - positive alternative density (implicit in the proof);

### Other methods

- Bayesian methods (e.g. Lewinger et al. [2007]): very strong modelling assumptions;
- Other screening methods: require low-dim covariate;
- Other two-group-model based methods: mostly asymptotic;

# Our Goal

Want a procedure

- weak modelling assumption (robust to misspecification);
- controlling FDR in finite samples;
- able to deal with any type of covariate;
- flexible to allow any learning algorithm/data exploration.

# Our Goal

Want a procedure

- weak modelling assumption (robust to misspecification);
- controlling FDR in finite samples;
- able to deal with any type of covariate;
- flexible to allow any learning algorithm/data exploration.

Sounds too ambitious?

# Our Goal

Want a procedure

- weak modelling assumption (robust to misspecification);
- controlling FDR in finite samples;
- able to deal with any type of covariate;
- flexible to allow any learning algorithm/data exploration.

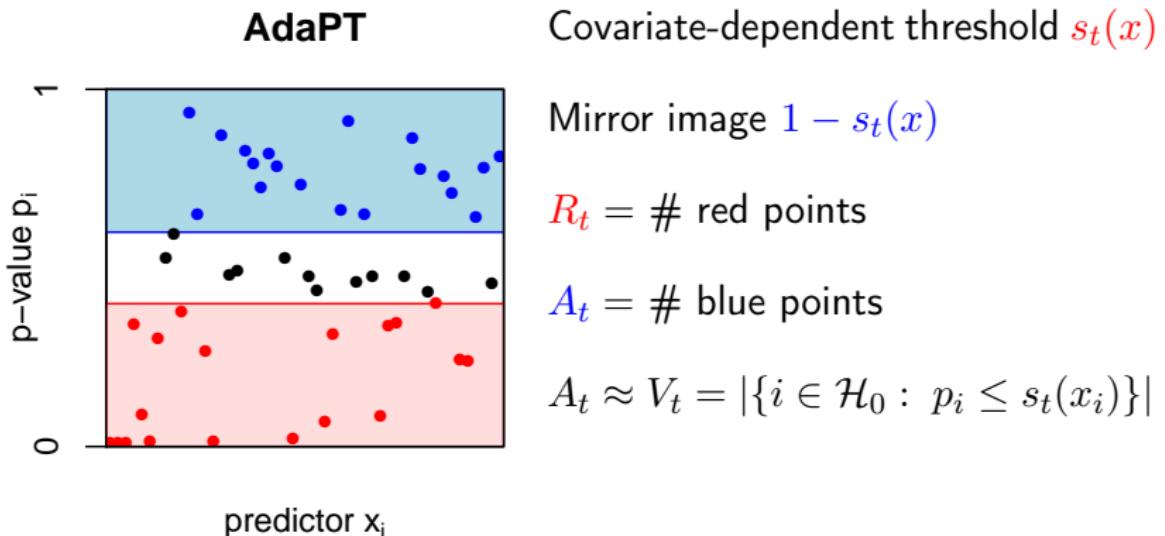
Sounds too ambitious? Not really!

# Table of Contents

- 1 Background and Motivation
- 2 Review of Existing Methods
- 3 AdaPT: Adaptive  $p$ -Value Thresholding
- 4 Updating the Threshold
- 5 Applications
- 6 Extensions

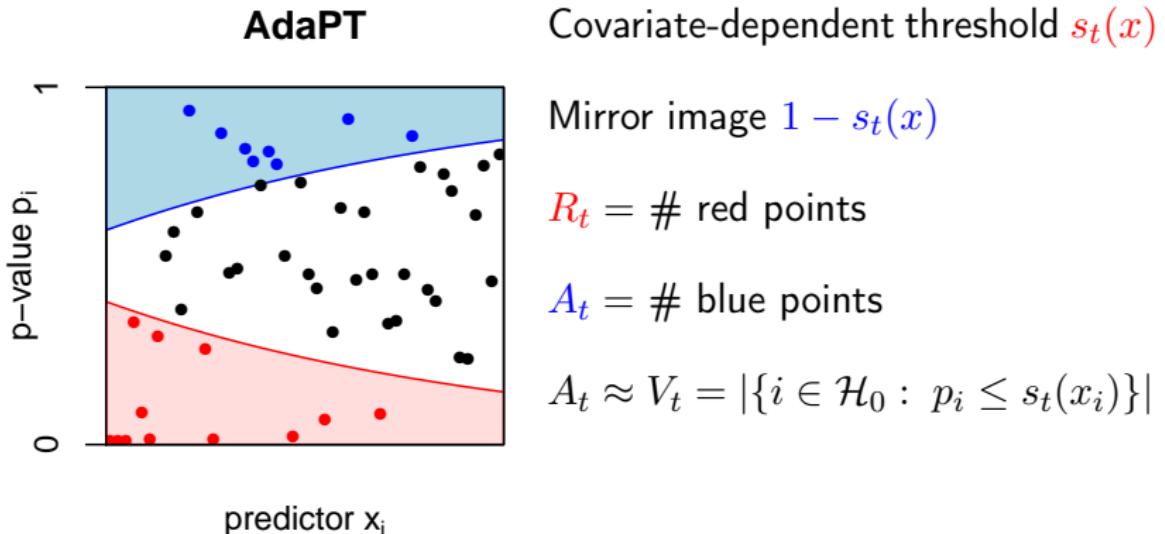
# AdaPT, Visualized

$$\widehat{\text{FDP}}_t = \frac{A_t + 1}{R_t \vee 1}$$



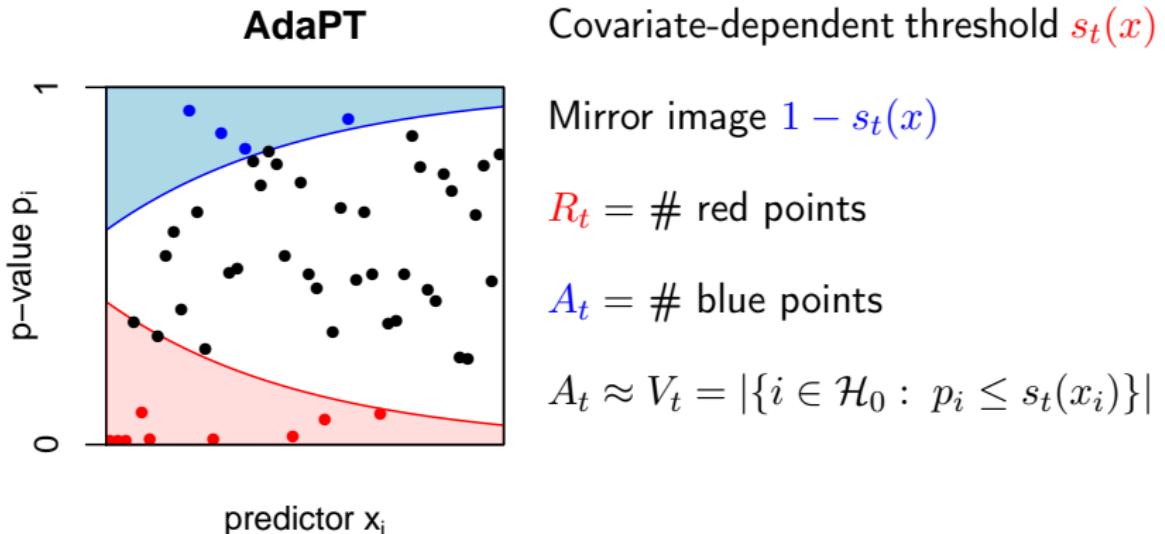
# AdaPT, Visualized

$$\widehat{\text{FDP}}_t = \frac{A_t + 1}{R_t \vee 1}$$



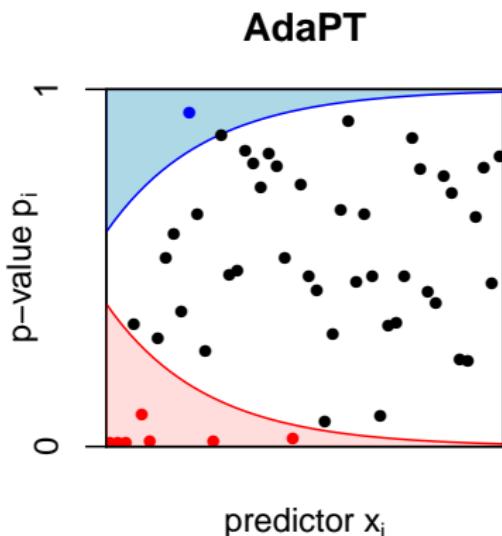
# AdaPT, Visualized

$$\widehat{\text{FDP}}_t = \frac{A_t + 1}{R_t \vee 1}$$



# AdaPT, Visualized

$$\widehat{\text{FDP}}_t = \frac{A_t + 1}{R_t \vee 1}$$



Covariate-dependent threshold  $s_t(x)$

Mirror image  $1 - s_t(x)$

$R_t = \#$  red points

$A_t = \#$  blue points

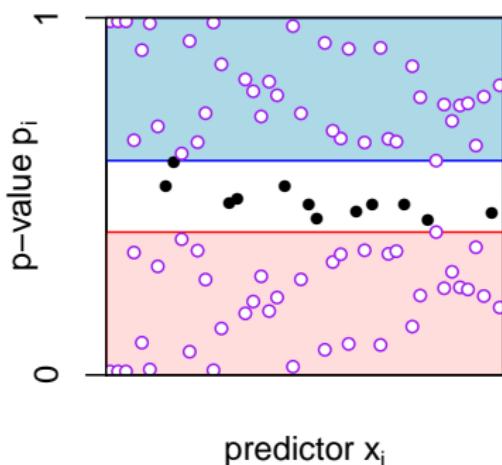
$A_t \approx V_t = |\{i \in \mathcal{H}_0 : p_i \leq s_t(x_i)\}|$

## AdaPT, “Analyst View”

Define partially masked  $p$ -values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \min\{p_i, 1 - p_i\} & \text{otherwise.} \end{cases}$$

### AdaPT (Analyst View)



To select  $s_{t+1}(x)$ , we can only use:

- $x_1, \dots, x_n$
- $\tilde{p}_{t,1}, \dots, \tilde{p}_{t,n}$
- $A_t, R_t$

(and same for  $t' < t$ )

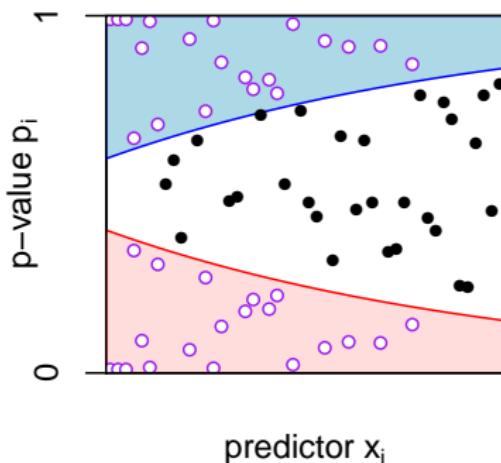
*Any such update rule is OK*

## AdaPT, “Analyst View”

Define partially masked  $p$ -values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \min\{p_i, 1 - p_i\} & \text{otherwise.} \end{cases}$$

### AdaPT (Analyst View)



To select  $s_{t+1}(x)$ , we can only use:

- $x_1, \dots, x_n$
- $\tilde{p}_{t,1}, \dots, \tilde{p}_{t,n}$
- $A_t, R_t$

(and same for  $t' < t$ )

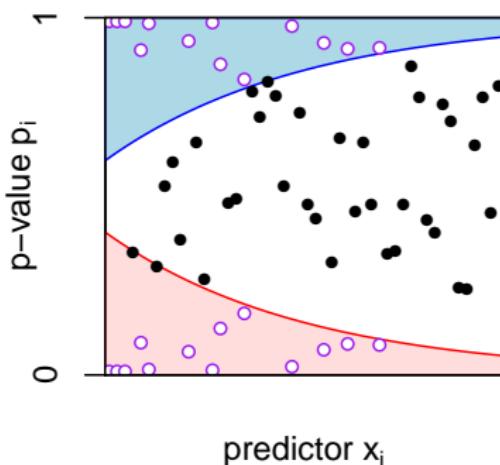
*Any such update rule is OK*

## AdaPT, “Analyst View”

Define partially masked  $p$ -values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \min\{p_i, 1 - p_i\} & \text{otherwise.} \end{cases}$$

### AdaPT (Analyst View)



To select  $s_{t+1}(x)$ , we can only use:

- $x_1, \dots, x_n$
- $\tilde{p}_{t,1}, \dots, \tilde{p}_{t,n}$
- $A_t, R_t$

(and same for  $t' < t$ )

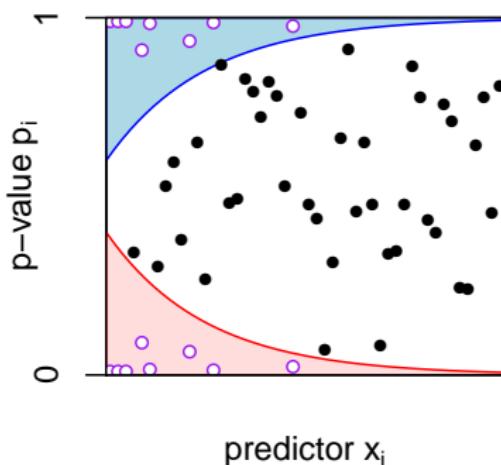
*Any such update rule is OK*

## AdaPT, “Analyst View”

Define partially masked  $p$ -values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \min\{p_i, 1 - p_i\} & \text{otherwise.} \end{cases}$$

### AdaPT (Analyst View)



To select  $s_{t+1}(x)$ , we can only use:

- $x_1, \dots, x_n$
- $\tilde{p}_{t,1}, \dots, \tilde{p}_{t,n}$
- $A_t, R_t$

(and same for  $t' < t$ )

*Any such update rule is OK*

# AdaPT: Finite-Sample FDR Control

## Theorem 1 (Lei and Fithian, 2016).

Assume that, conditional on  $(x_i)_{i=1}^n$  and  $(p_i)_{i \notin \mathcal{H}_0}$ , the null p-values  $(p_i)_{i \in \mathcal{H}_0}$  are independent and mirror-conservative. Then AdaPT controls FDR at level  $\alpha$ .

# Mirror-conservatism

Mirror-conservative:

$$\mathbb{P}(p \in [a, b]) \leq \mathbb{P}(p \in [1 - b, 1 - a]), \forall 0 \leq a \leq b \leq 0.5$$

- Uniform
- Discrete  $p$ -values after randomization
- Permutation test  $p$ -values
- One-sided tests for
  - MLR families (e.g. log-concave location, exponential family)
  - Symmetric unimodal location families

# Mirror-conservatism

Mirror-conservative:

$$\mathbb{P}(p \in [a, b]) \leq \mathbb{P}(p \in [1 - b, 1 - a]), \forall 0 \leq a \leq b \leq 0.5$$

- Uniform
- Discrete  $p$ -values after randomization
- Permutation test  $p$ -values
- One-sided tests for
  - MLR families (e.g. log-concave location, exponential family)
  - Symmetric unimodal location families
- Different from usual conservatism  $\mathbb{P}(p \leq t) \leq t$ :
  - $p = B \sim \text{Ber}(0.9)$  is conservative but not mirror-conservative;
  - $p = 0.1 + 0.9B$  is mirror-conservative but not conservative.

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if Bayesian);

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if Bayesian);
- no matter how we select a model or tuning parameter;

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if Bayesian);
- no matter how we select a model or tuning parameter;
- no matter how crazy we are.

## AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if Bayesian);
- no matter how we select a model or tuning parameter;
- no matter how crazy we are.

Don't worry!

Use your favorite method

Guaranteed FDR nevertheless

# AdaPT is flexible!

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if Bayesian);
- no matter how we select a model or tuning parameter;
- no matter how crazy we are.

Don't worry!

Use your favorite method

Guaranteed FDR nevertheless

Don't be crazy!

Use the best possible method

Degraded power otherwise

# Table of Contents

- 1 Background and Motivation
- 2 Review of Existing Methods
- 3 AdaPT: Adaptive  $p$ -Value Thresholding
- 4 Updating the Threshold
- 5 Applications
- 6 Extensions

## Guiding Principle

**Theorem 2 (Lei and Fithian, 2016).**

*Under mild assumptions, the optimal threshold  $s(x)$  is a level curve of local FDR.*

## Guiding Principle

**Theorem 2 (Lei and Fithian, 2016).**

*Under mild assumptions, the optimal threshold  $s(x)$  is a level curve of local FDR.*

Step 1. Propose a **working model**;

Step 2. Use **your favorite method** to fit the model;

Step 3. Estimate **level curves of local FDR**;

## Guiding Principle

**Theorem 2 (Lei and Fithian, 2016).**

*Under mild assumptions, the optimal threshold  $s(x)$  is a level curve of local FDR.*

- Step 1. Propose a **working model**;
- Step 2. Use **your favorite method** to fit the model;
- Step 3. Estimate **level curves of local FDR**;
- Step 4. Move the threshold towards a “near” level curve;
- Step 5. Repeat Step 2 - Step 4 until  $\widehat{\text{FDP}} \leq \alpha$ .

## Animation: Gene/drug Response Data

## Conditional Two-Groups Model (A Working Model!)

Frame threshold choice in terms of conditional two-groups model:

$$H_i \mid x_i \sim \text{Bernoulli}(\pi_1(x_i))$$

$$p_i \mid H_i, x_i \sim \begin{cases} f_0(p \mid x_i) & \text{if } H_i = 0 \\ f_1(p \mid x_i) & \text{if } H_i = 1 \end{cases}.$$

Assume  $f_0(p \mid x) = 1$ , define conditional mixture density

$$\begin{aligned} f(p \mid x) &= (1 - \pi_1(x)) f_0(p \mid x) + \pi_1(x) f_1(p \mid x) \\ &= 1 - \pi_1(x) + \pi_1(x) f_1(p \mid x), \end{aligned}$$

leading to conditional local fdr

$$\text{fdr}(p \mid x) = \mathbb{P}(H_i \text{ is null} \mid x_i = x, p_i = p) = \frac{1 - \pi_1(x)}{f(p \mid x)}$$

## Rejection Thresholds and Local fdr

$$\text{fdr}(p \mid x) = \mathbb{P}(H_i \text{ is null} \mid x_i = x, p_i = p) = \frac{1 - \pi_1(x)}{f(p \mid x)}$$

Assume  $f_1(p \mid x)$  decreasing  $\Rightarrow \text{fdr}(p \mid x)$  increasing

Best  $s_t(x)$  are **level surfaces** of  $\text{fdr}(p \mid x)$   
(maximizes power subject to  $\text{FDR} \leq \alpha$ )

Idea:

- ① Estimate  $\widehat{\text{fdr}}_t(p \mid x)$  using data at step  $t$
- ② Choose level surface for  $s_{t+1}(x)$

## Modeling and Missing Data

Convenient model is Logistic–Gamma compound GLM:

$$H_i \mid x_i \sim \text{Bernoulli}(\pi_1(x_i)),$$
$$-\log p_i \mid x_i, H_i \sim \begin{cases} \text{Exp}(\mu(x_i)) & \text{if } H_i = 1 \\ \text{Exp}(1) & \text{if } H_i = 0 \end{cases}$$

where  $\log \frac{\pi_1(x)}{1 - \pi_1(x)} = \theta' \phi(x)$ , and  $\mu(x)^{-1} = \beta' \phi(x)$

Note: Need to impute masked  $p_i$  (only know  $\min\{p_i, 1 - p_i\}$ )

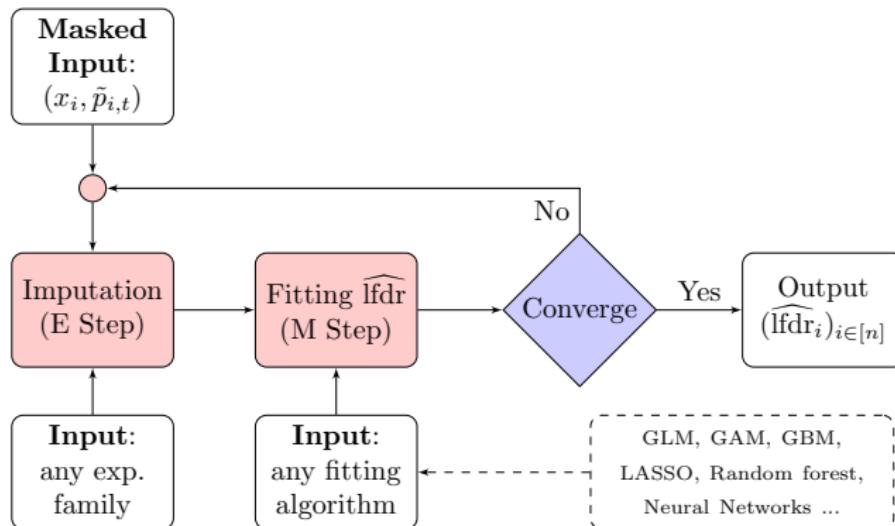
EM update is easy due to GLM structure

## Generalized EM Framework

- E-step imputes the missing data;
- M-step fits the local FDR using the imputed data;
- E-step requires a **posterior distribution**;
- M-step requires a **fitting algorithm**.

# Generalized EM Framework

- E-step imputes the missing data;
- M-step fits the local FDR using the imputed data;
- E-step requires a **posterior distribution**;
- M-step requires a **fitting algorithm**.

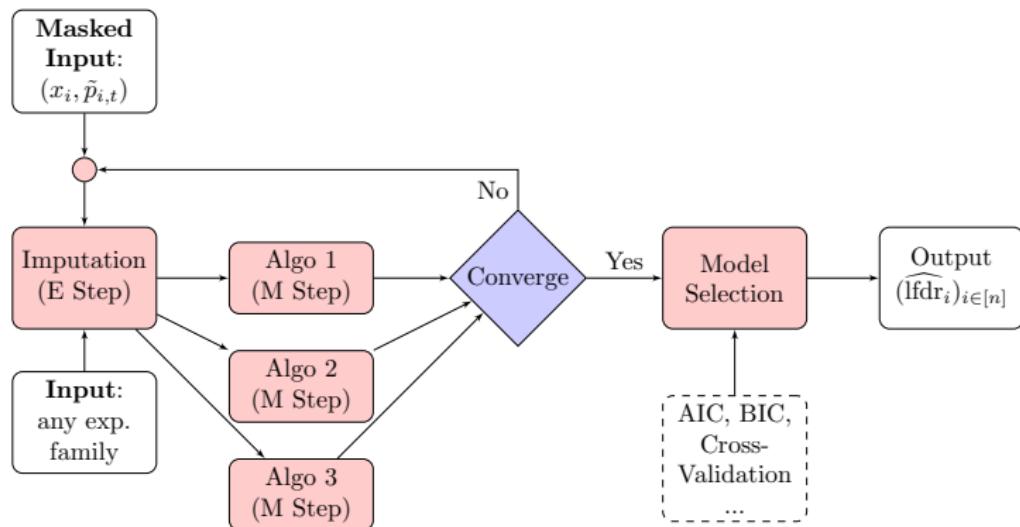


## EM + Model Selection (Parameter Tuning)

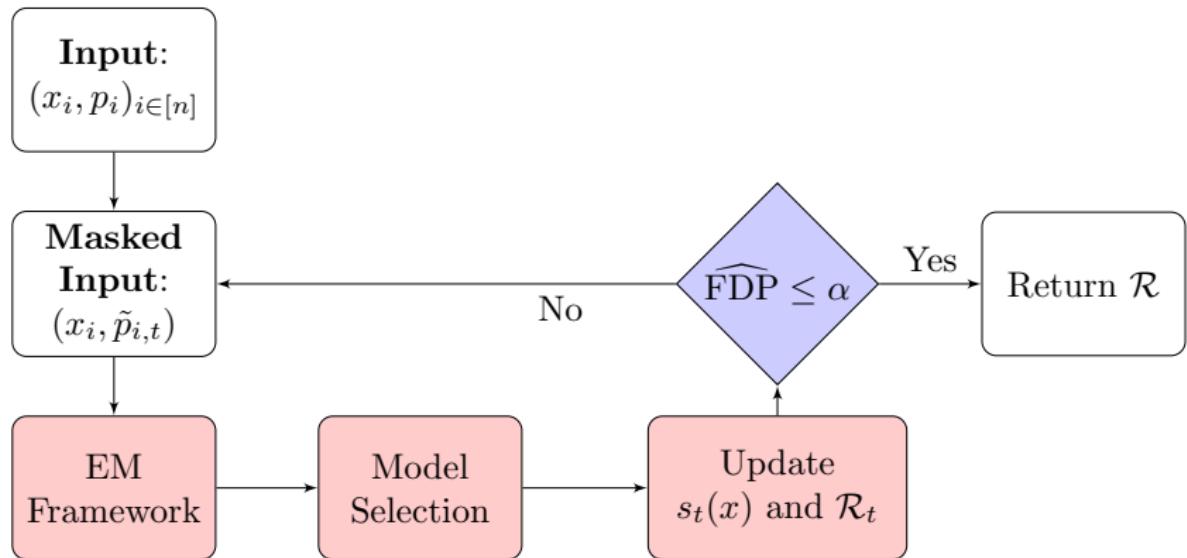
- Provide a list of candidate algorithms (or a single algorithm with a list of candidate tuning parameters);
- Select a model based on pseudo-likelihood given by the exponential family.

## EM + Model Selection (Parameter Tuning)

- Provide a list of candidate algorithms (or a single algorithm with a list of candidate tuning parameters);
- Select a model based on pseudo-likelihood given by the exponential family.



# AdaPT Pipeline



# Table of Contents

- 1 Background and Motivation
- 2 Review of Existing Methods
- 3 AdaPT: Adaptive  $p$ -Value Thresholding
- 4 Updating the Threshold
- 5 Applications
- 6 Extensions

## Gene/drug response Data

Li and Barber [2016a] proposed ordered analysis of gene expression data [Coser et al., 2003, Davis and Meltzer, 2007]

Expression in breast cancer cells in response to estrogen

- $n = 22283$  genes, 25 trials at 5 doses incl. control
- $H_i$  : no differential response in low-dose vs. control
- $p_i$  computed via permutation  $t$ -test

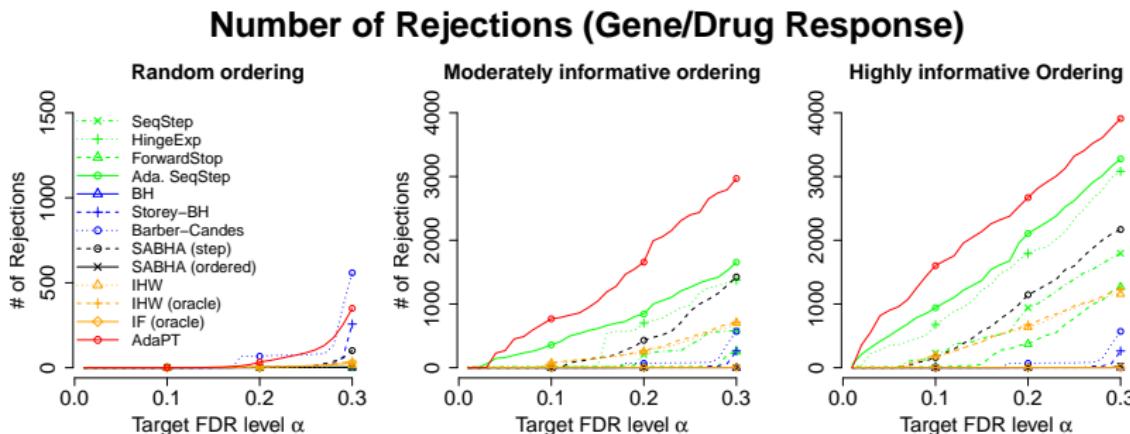
Ordered by  $\tilde{p}_i$ , permutation  $p$ -value comparing high-dose vs. pooled sample of low-dose + control

Can show  $p_i$  independent of  $\tilde{p}_i$  if  $H_i$  true under some conditions

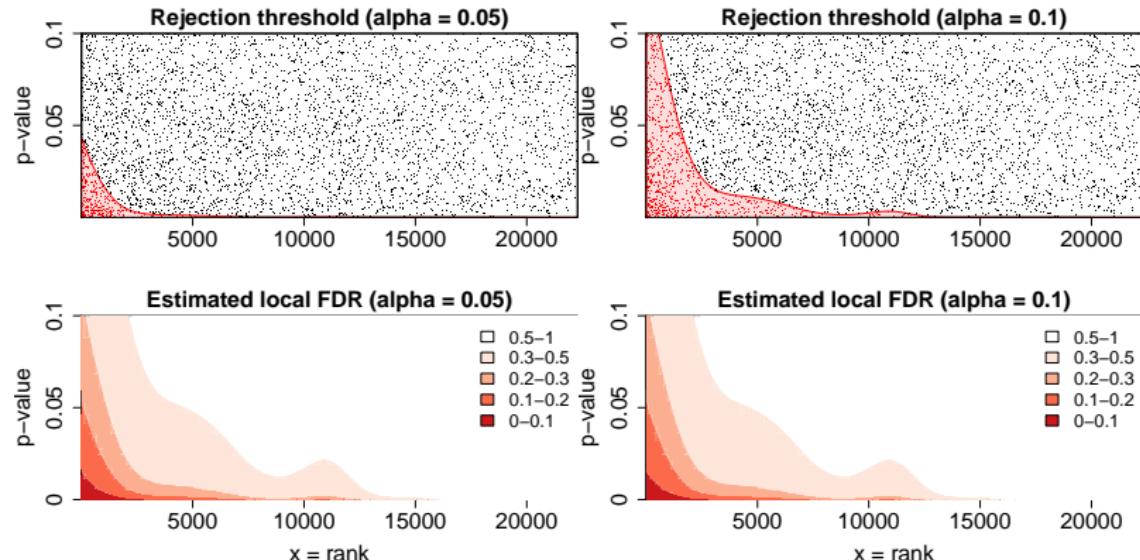
# Gene/drug response Data: Power Comparison

Compared AdaPT to competing methods using three orderings:

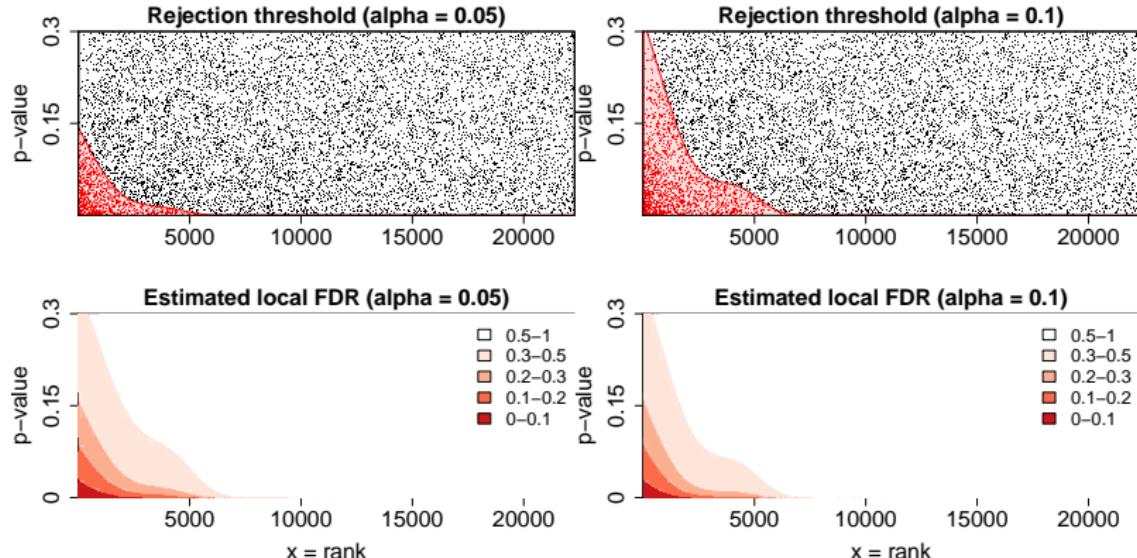
- Original ordering (genome order)
- Moderate dose ordering (dose 2 vs. pooled doses 0 & 1)
- High dose ordering (dose 4 vs. pooled doses 0 & 1)



# Gene/drug response Data: Moderate Dose Ordering



# Gene/drug response Data: High Dose Ordering



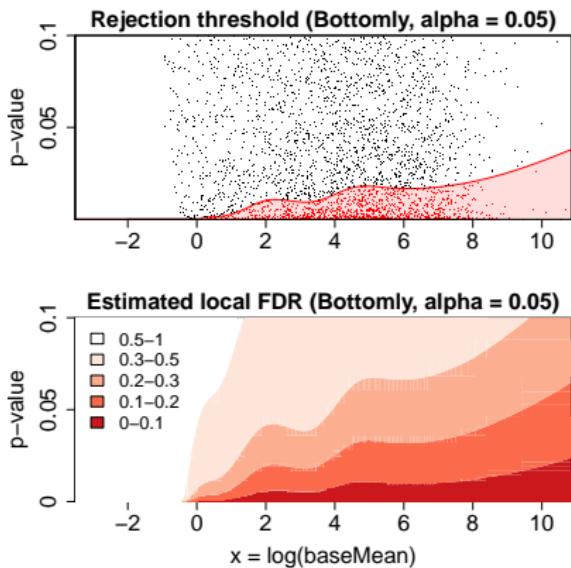
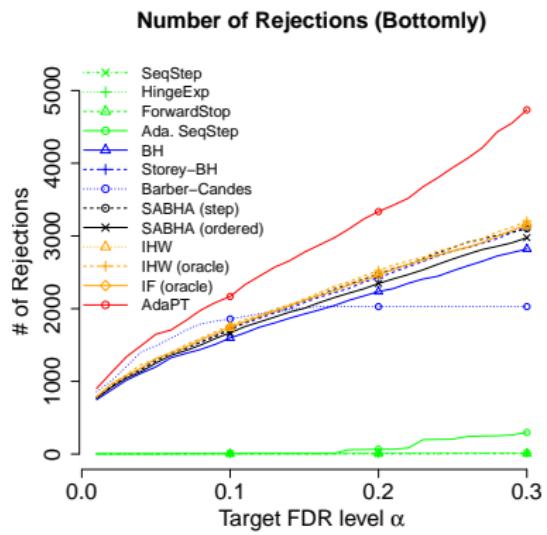
## RNA-seq Data

Ignatiadis et al. [2016] proposed analysis of gene expression data [Bottomly et al., 2011]

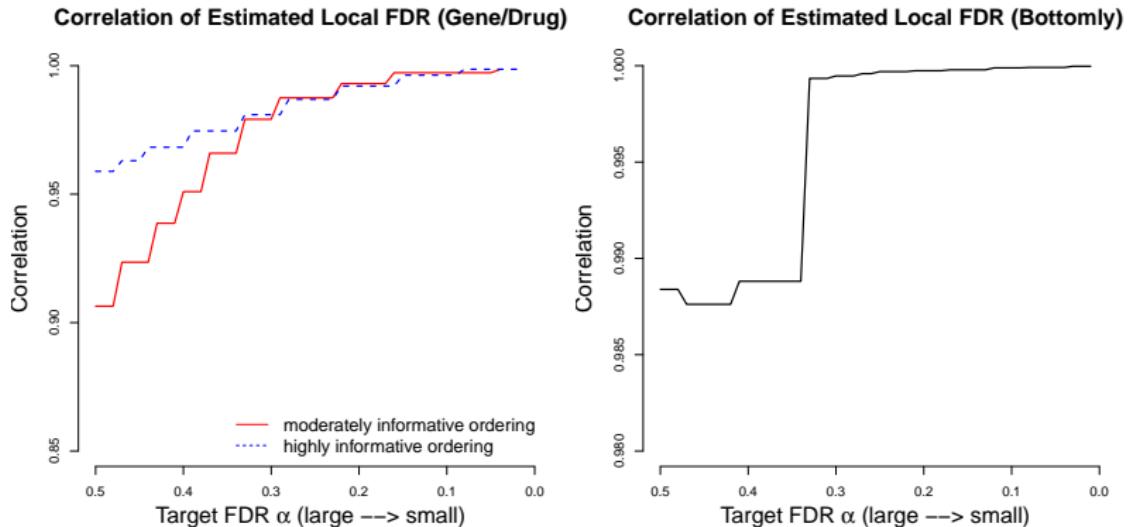
Expression in two mouse strains C57BL/6J (B6) and DBA/2J (D2)

- $n = 13932$  genes, 21 samples (10 B6 and 11 D2)
- $H_i$  : no differential response in gene  $i$
- $p_i$  computed via DEseq2 package
- $x_i$  : logarithmic normalized count via DEseq2 package

# RNA-seq Data: Power Comparison



# Information Loss by Partial Masking

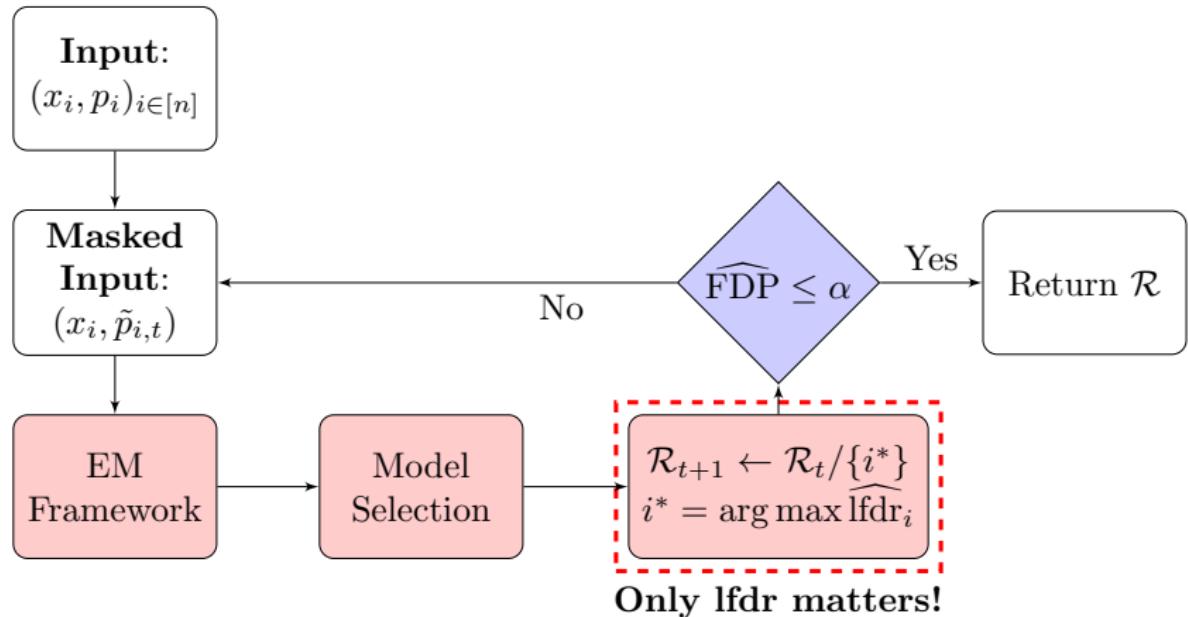


## Animation: RNA-seq Data

# Table of Contents

- ① Background and Motivation
- ② Review of Existing Methods
- ③ AdaPT: Adaptive  $p$ -Value Thresholding
- ④ Updating the Threshold
- ⑤ Applications
- ⑥ Extensions

# AdaPT Without Threshold

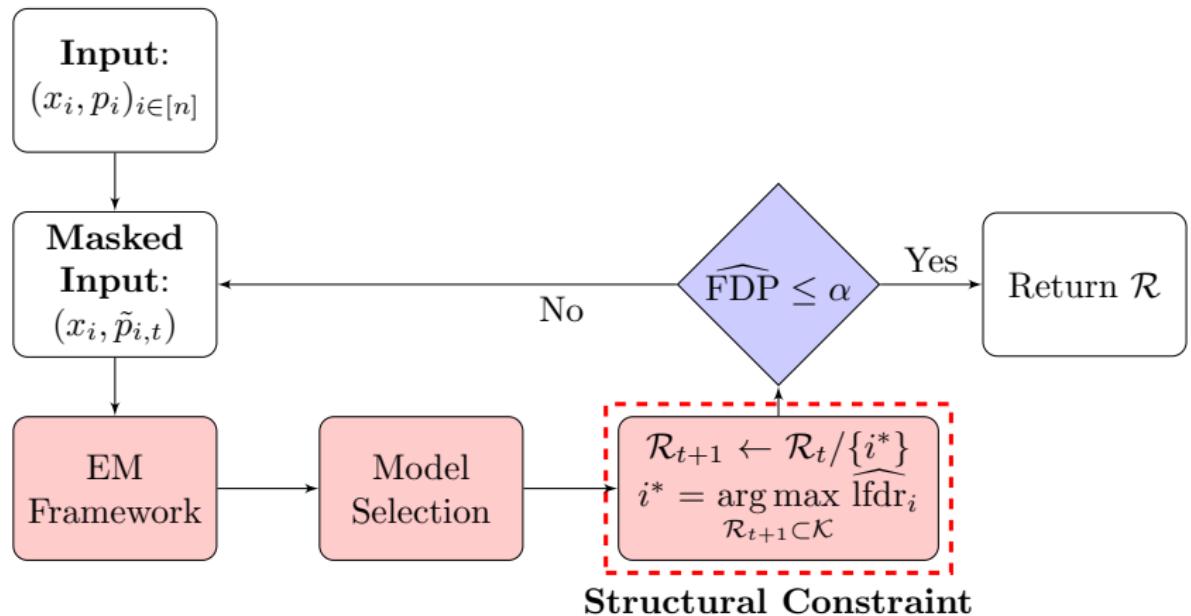


# Multiple Testing With Structural Constraint

Want to detect a subset of hypotheses subjected to certain structural constraint, i.e.  $\{x_i : H_i \text{ rejected}\} \subset \mathcal{K}$ :

- (spatial-temporal multiple testing)  
 $x_i$ : geographic location,  $\mathcal{K}$ : all convex sets;
- (hierarchical testing)  
 $x_i$ : node of a tree,  $\mathcal{K}$ : all subtrees;
- (Selection under strong/weak heredity principles)  
 $x_i$ : node of a DAG,  $\mathcal{K}$ : all subgraphs st. heredity principles

# Selectively Traversed Accumulation Rules (STAR)



## Animation: STAR in Convex Region Detection

## Animation: STAR in Hierarchical Testing

## Animation: STAR in Selection Under Heredity Principle

# References

- Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoff. *The Annals of Statistics*, 43(5):2055–2085, 2015.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
- Daniel Bottomly, Nicole AR Walter, Jessica Ezzell Hunter, Priscila Darakjian, Sunita Kawane, Kari J Buck, Robert P Sculley, Michael Mooney, Shannon K McWeeney, and Robert Hitzemann. Evaluating gene expression in c57bl/6j and dba/2j mouse striatum using rna-seq and microarrays. *PLoS one*, 6(3):e17820, 2011.
- Kathryn R Coser, Jessica Chesnes, Jingyung Hur, Sandip Ray, Kurt J Isselbacher, and Toshi Shioda. Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in mcf7/bus breast cancer cells by dna microarray. *Proceedings of the National Academy of Sciences*, 100(24):13994–13999, 2003.
- Sean Davis and Paul S Meltzer. Geoquery: a bridge between the gene expression omnibus (geo) and bioconductor. *Bioinformatics*, 23(14):1846–1847, 2007.
- Egil Færkingstad, Arnoldo Frigessi, Håvard Rue, Gudmar Thorleifsson, and Augustine Kong. Unsupervised empirical bayesian multiple testing with external covariates. *The Annals of Applied Statistics*, pages 714–735, 2008.
- Dean P Foster and Robert A Stine.  $\alpha$ -investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):429–444, 2008.
- Max Graizer G'Sell, Stefan Wager, Alexandra Chouldechova, and Robert Tibshirani. Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2015.
- Nikolaos Ignatiadis and Wolfgang Huber. Covariate-powered weighted multiple testing with false discovery rate control. *arXiv preprint arXiv:1701.05179*, 2017.
- Nikolaos Ignatiadis, Bernd Klaus, Judith B Zaugg, and Wolfgang Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods*, 2016.
- Juan Pablo Lewinger, David V Conti, James W Baurley, Timothy J Triche, and Duncan C Thomas. Hierarchical bayes prioritization of marker associations from a genome-wide association scan for further investigation. *Genetic epidemiology*, 31(8):871–882, 2007.
- Ang Li and Rina Foygel Barber. Accumulation tests for FDR control in ordered hypothesis testing. *Journal of the American Statistical Association*, (just-accepted):1–38, 2016a.
- Ang Li and Rina Foygel Barber. Multiple testing with the structure adaptive benjamini-hochberg algorithm. *arXiv preprint arXiv:1606.07926*, 2016b.
- John D Storey, Jonathan E Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(1):187–205, 2004.
- Fei Xia, Martin J Zhang, James Y Zou, and David Tse. Neuralfdr: Learning discovery thresholds from hypothesis features. In *Advances in Neural Information Processing Systems*, pages 1540–1549, 2017.

# Thanks!