AdaPT: Interactive Multiple Testing with Side Information

Lihua Lei

Department of Statistics, Stanford University

JRSS Webinar, 2021

Collaborator

Will Fithian (UC Berkeley)

Multiple Hypothesis Testing

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

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 $\mathrm{FDP} = \frac{V}{R \vee 1}$

Goal: control False Discovery Rate [Benjamini and Hochberg, '95]

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

Side Information

Observe side information $x_i \in \mathcal{X}$ for each H_i

$x_1, \ldots, x_n$ treated as fixed

Ordered multiple testing

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

Side Information

Observe side information $x_i \in \mathcal{X}$ for each H_i

$x_1, \ldots, x_n$ treated as fixed

Ordered multiple testing

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

Other examples:

- Data from a similar experiment
- Spatiotemporal location e.g. $H_i : \ f(t_i) \leq 0$

Side Information

Observe side information $x_i \in \mathcal{X}$ for each H_i

$x_1, \ldots, x_n$ treated as fixed

Ordered multiple testing

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

Other examples:

- Data from a similar experiment
- Spatiotemporal location e.g. $H_i : \ f(t_i) \leq 0$

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

Side Information in Biology

RESEARCH Open Access

A practical guide to methods controlling false discoveries in computational biology

Keegan Korthauer[1,2†], Patrick K. Kimes[1,2†], Claire Duvallet[3,4†], Alejandro Reyes[1,2†], Ayshwarya Subramanian[5†], Mingxiang Teng[6], Chinmay Shukla[7], Eric J. Alm[3,4,5] and Stephanie C. Hicks[8*]

Table 1 Independent and informative covariates used in case studies

Case study	Covariates found to be independent and informative
Microbiome	**Ubiquity**: the proportion of samples in which the feature is present. In microbiome data, it is common for many features to go undetected in many samples.
	Mean nonzero abundance: the average abundance of a feature among those samples in which it was detected. We note that this did not seem as informative as ubiquity in our case studies.
GWAS	**Minor allele frequency**: the proportion of the population which exhibits the less common allele (ranges from 0 to 0.5) represents the rarity of a particular variant.
	Sample size (for meta-analyses): the number of samples for which the particular variant was measured.
Gene set analyses	**Gene set size**: the number of genes included in the particular set. Note that this is not independent under the null for over-representation tests, however (see Additional file 1: Supplementary Results).
Bulk RNA-seq	**Mean gene expression**: the average expression level (calculated from normalized read counts) for a particular gene.
Single-Cell RNA-seq	**Mean nonzero gene expression**: the average expression level (calculated from normalized read counts) for a particular gene, excluding zero counts.
	Detection rate: the proportion of samples in which the gene is detected. In single-cell RNA-seq it is common for many genes to go undetected in many samples.
ChIP-seq	**Mean read depth**: the average coverage (calculated from normalized read counts) for the region
	Window Size: the length of the region

AdaPT in a Nutshell

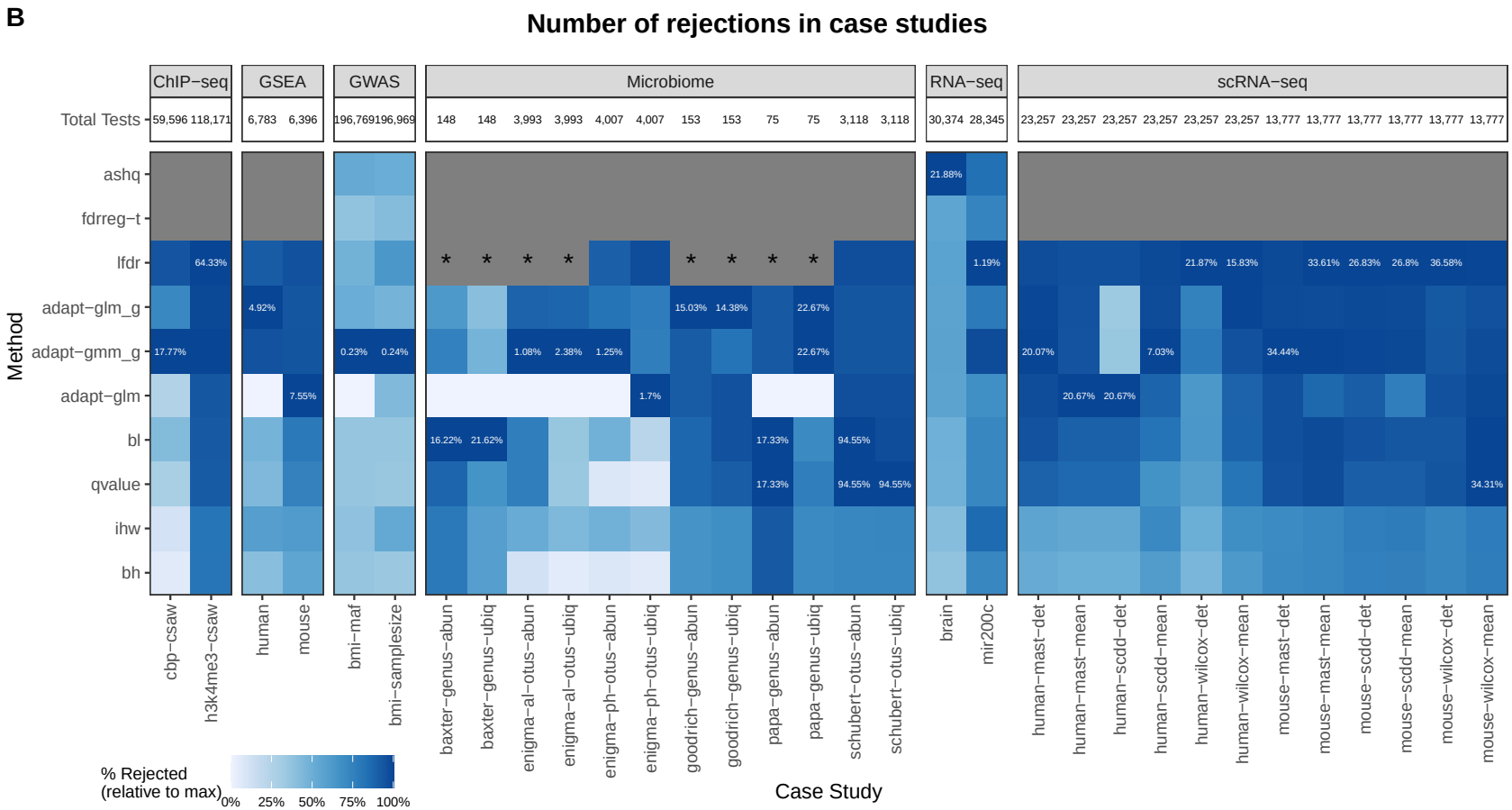
We proposed Adaptive P-value Thresholding (AdaPT)

- controls FDR in **finite samples**
- robust to **arbitrary misspecification** of non-nulls
- can wrap around any black-box algorithm
- able to deal with any type of covariates

AdaPT works well in a range of applications

Chao and Fithian '21:

AdaPT-GMM: Powerful and robust covariate-assisted multiple testing

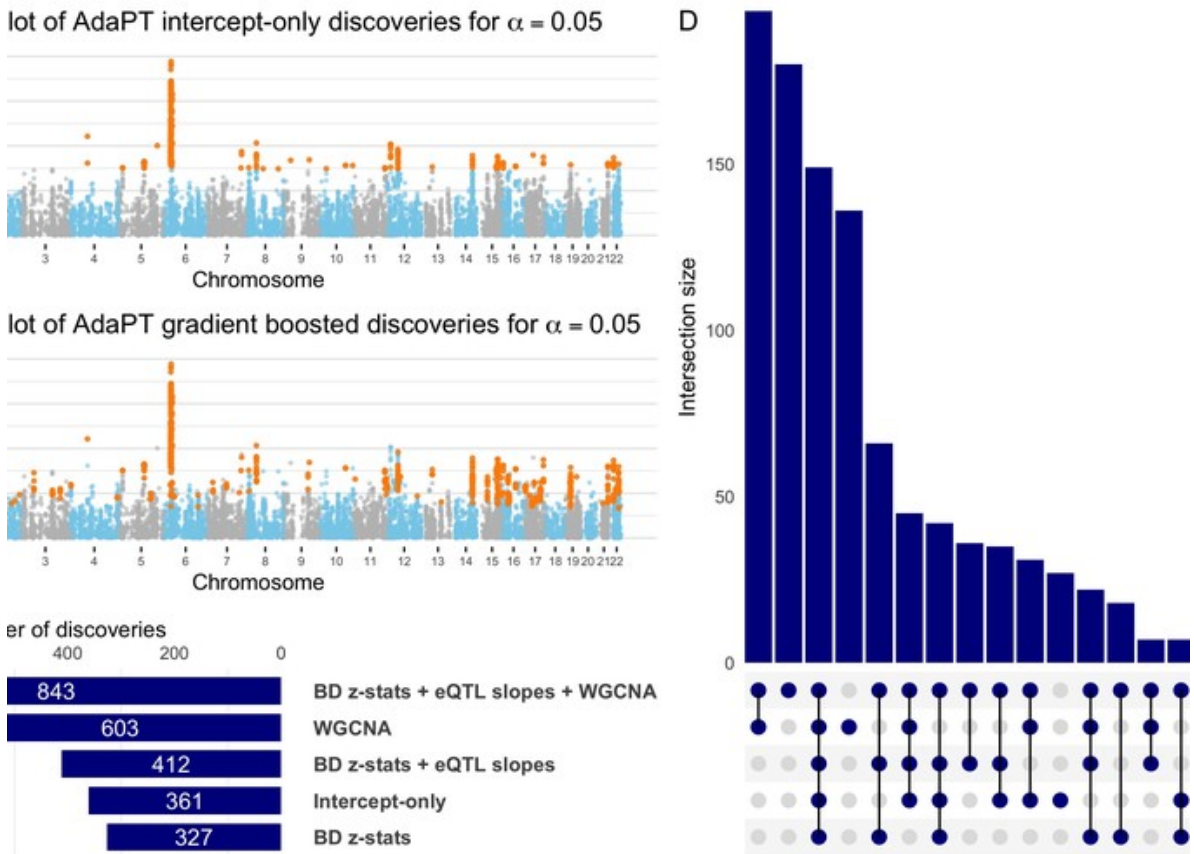


Examples from Korthauer et al. ('19) Genome Biology.

AdaPT works well in a range of applications

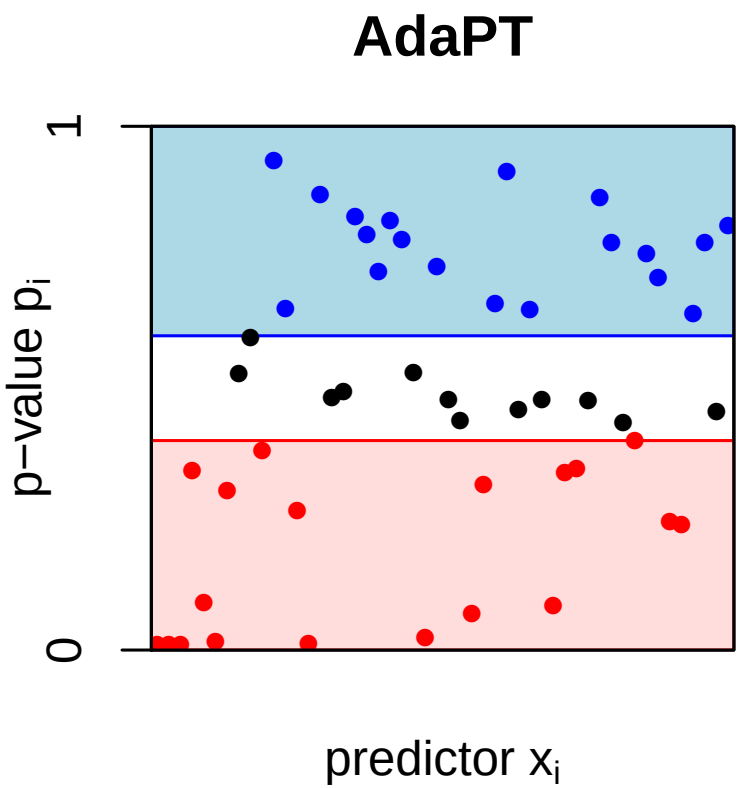
Application of post-selection inference to multi-omics data yields insights into the etiologies of human diseases

Ronald Yurko, Max G'Sell, Kathryn Roeder, Bernie Devlin



AdaPT, Visualized

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{blue points} + 1}{\#\text{red points} \vee 1} \qquad (\#\text{blue points} \approx \#\text{red nulls})$$



Covariate-dependent threshold $s_t(x)$

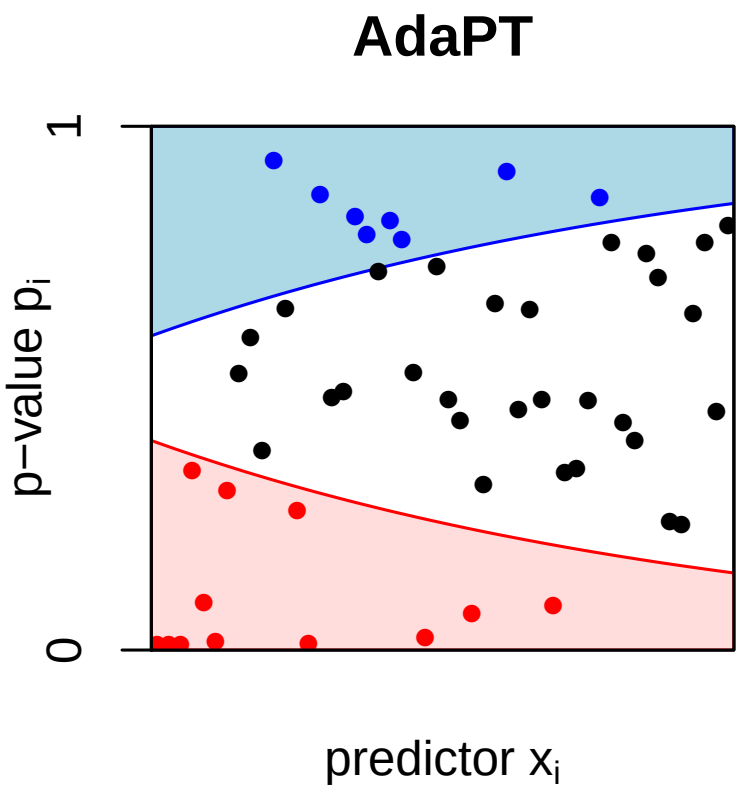
Mirror image $1 - s_t(x)$

Update $s_t(x)$ until $\widehat{\mathrm{FDP}}_t \leq \alpha$

Reject all red points

AdaPT, Visualized

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{blue points} + 1}{\#\text{red points} \vee 1} \qquad (\#\text{blue points} \approx \#\text{red nulls})$$



Covariate-dependent threshold $s_t(x)$

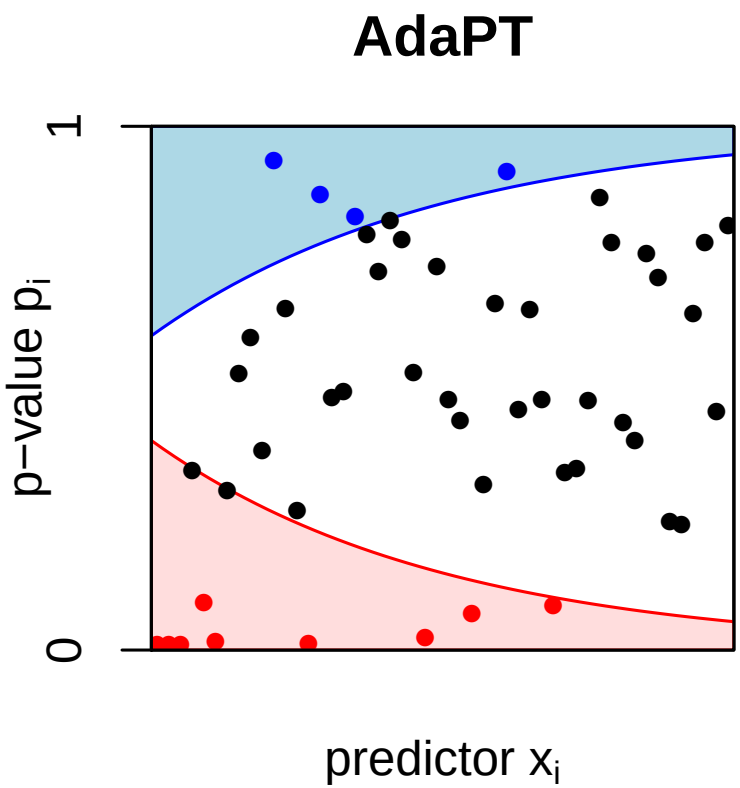
Mirror image $1 - s_t(x)$

Update $s_t(x)$ until $\widehat{\mathrm{FDP}}_t \leq \alpha$

Reject all red points

AdaPT, Visualized

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{blue points} + 1}{\#\text{red points} \vee 1} \qquad (\#\text{blue points} \approx \#\text{red nulls})$$



Covariate-dependent threshold $s_t(x)$

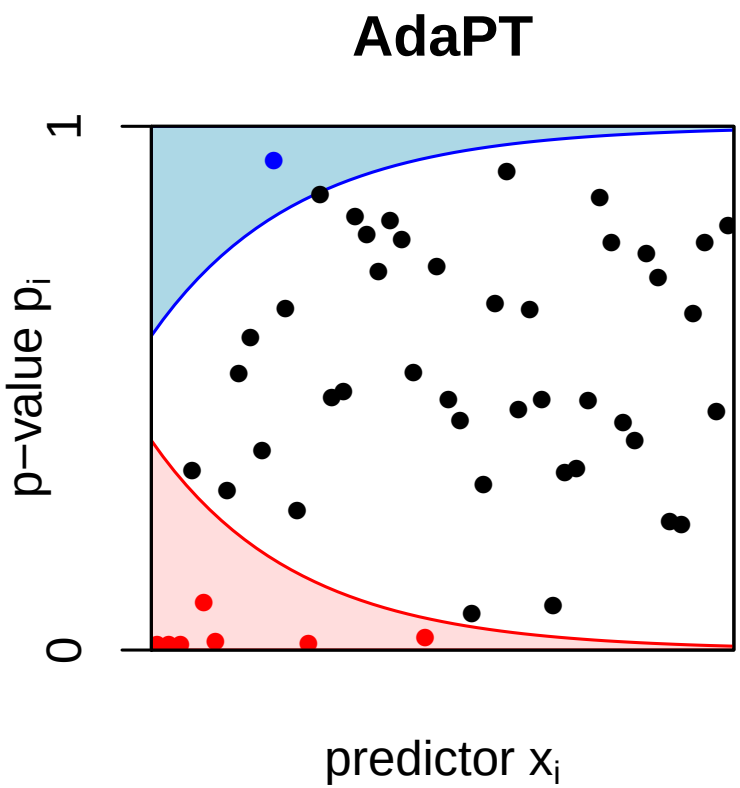
Mirror image $1 - s_t(x)$

Update $s_t(x)$ until $\widehat{\mathrm{FDP}}_t \leq \alpha$

Reject all red points

AdaPT, Visualized

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{blue points} + 1}{\#\text{red points} \vee 1} \qquad (\#\text{blue points} \approx \#\text{red nulls})$$



Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

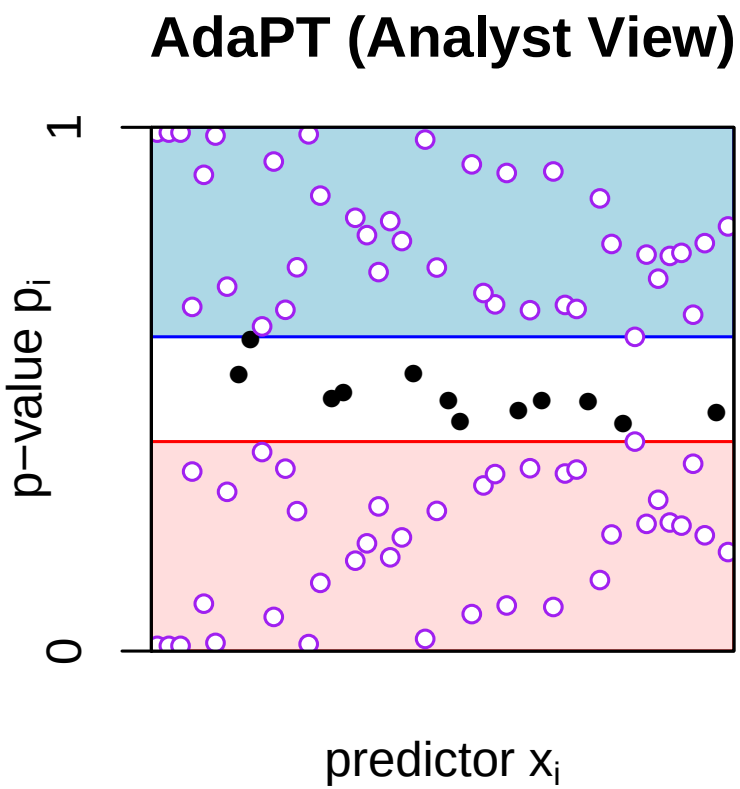
Update $s_t(x)$ until $\widehat{\mathrm{FDP}}_t \leq \alpha$

Reject all red points

AdaPT, "Analyst View"

Define partially masked p-values:

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



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

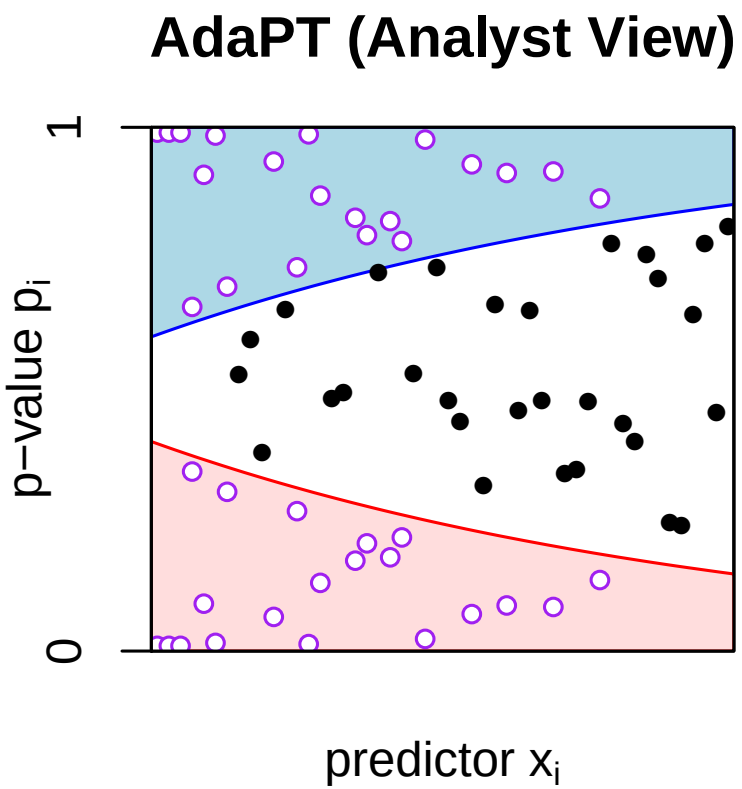
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

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



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

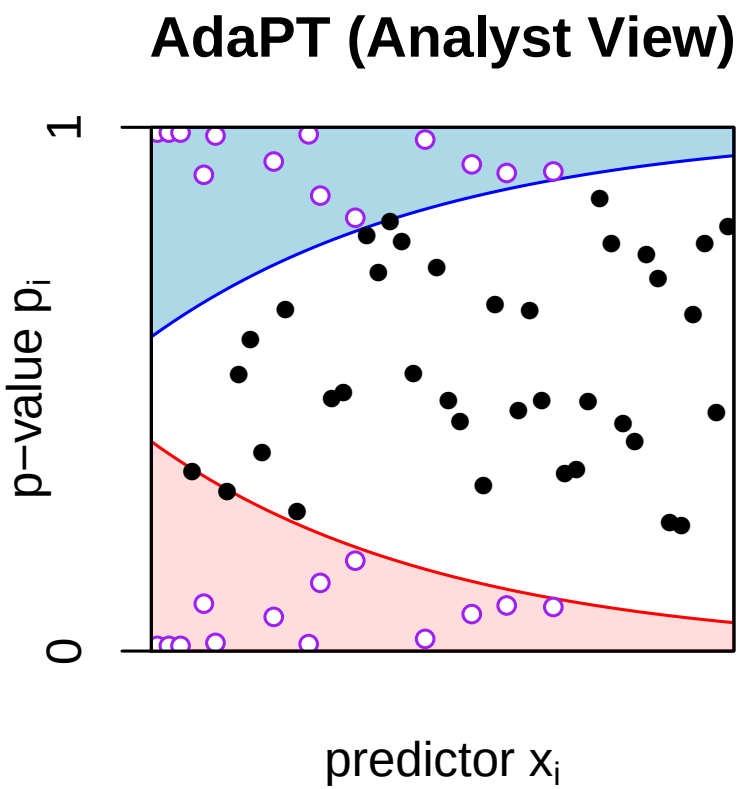
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

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



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

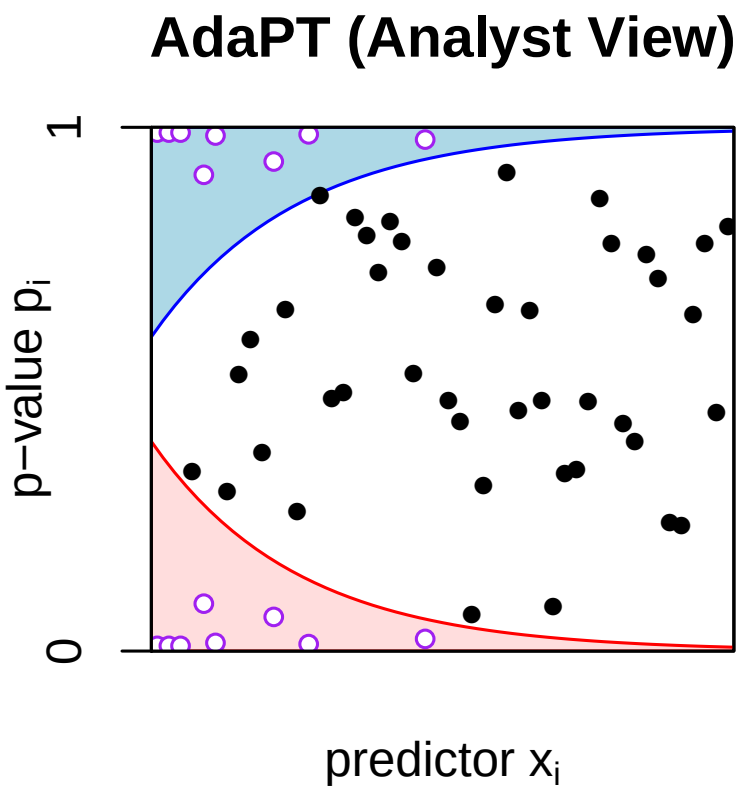
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

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



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

- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT: Finite-Sample FDR Control

Theorem 1 (L. and Fithian, '18).

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 (e.g. uniform). Then AdaPT controls FDR at level α.

- Non-null p-values can be arbitrarily dependent
- Most p-values in practice are mirror conservative
- AdaPT controls FDR with arbitrary update rule

AdaPT is flexible!

AdaPT is unusually flexible in that the FDR is controlled

AdaPT is flexible!

AdaPT is unusually flexible in that the FDR is controlled

- no matter how misspecified our model is

AdaPT is flexible!

AdaPT is unusually flexible in that the FDR is controlled

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

AdaPT is flexible!

AdaPT is unusually flexible in that the FDR is controlled

- 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 unusually flexible in that the FDR is controlled

- 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 unusually flexible in that the FDR is controlled

- 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 unusually flexible in that the FDR is controlled

- 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

Updating the Threshold: Guiding Principle

Theorem 2 (L. and Fithian, '18).

Under the two-group model and mild assumptions, the optimal threshold $s(x)$ is a level curve of local FDR.

Updating the Threshold: Guiding Principle

Theorem 2 (L. and Fithian, '18).

Under the two-group model and mild assumptions, the optimal threshold $s(x)$ is a level curve of local FDR.

1. Propose a **working model**;
2. Use **your favorite method** to fit the model;
3. Estimate **level curves of local FDR**;

Updating the Threshold: Guiding Principle

Theorem 2 (L. and Fithian, '18).

Under the two-group model and mild assumptions, the optimal threshold $s(x)$ is a level curve of local FDR.

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

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

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

Idea:

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

Rejection Thresholds and Local fdr

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

Idea:

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

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

Rejection Thresholds and Local fdr

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

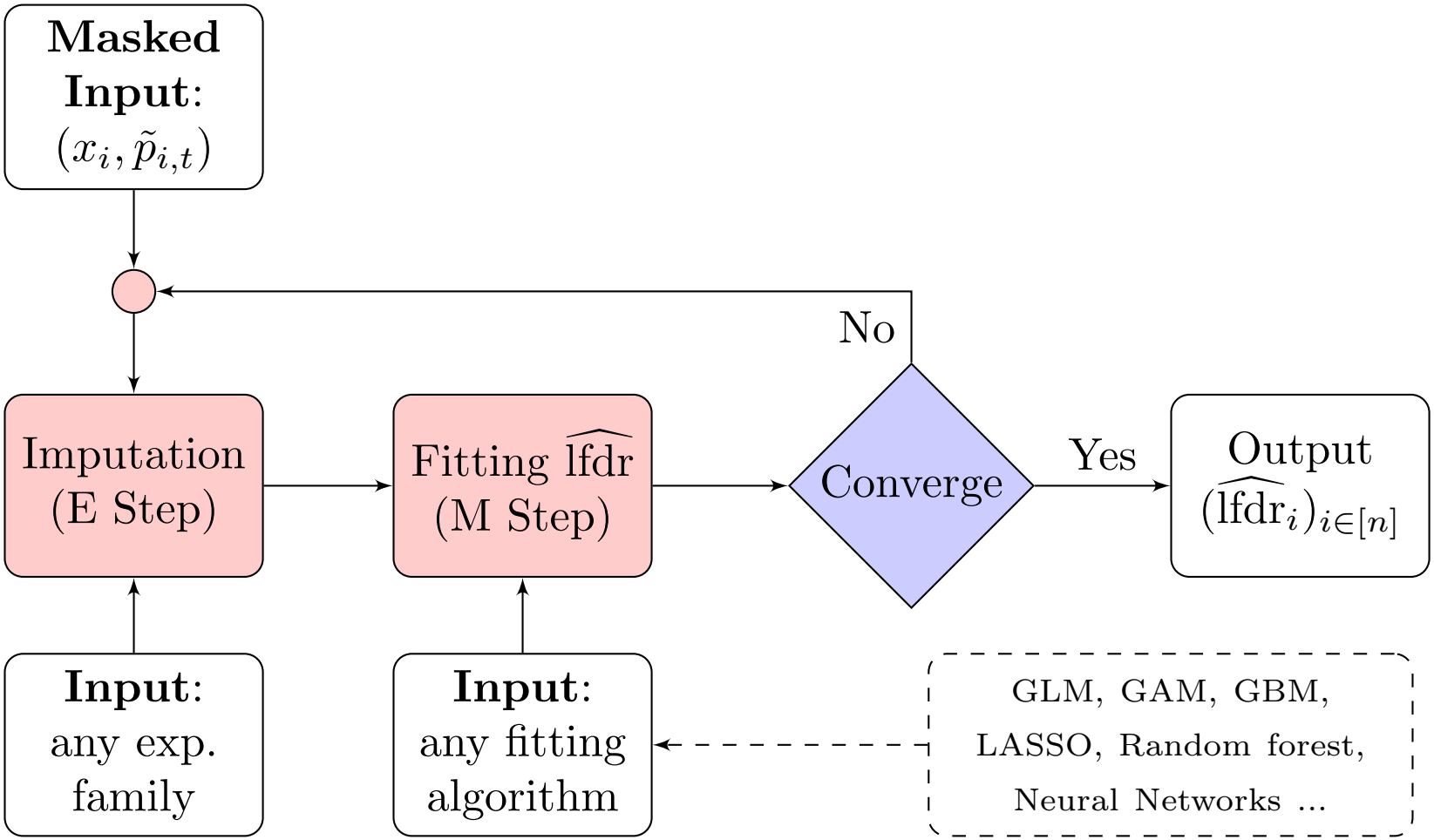
Idea:

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

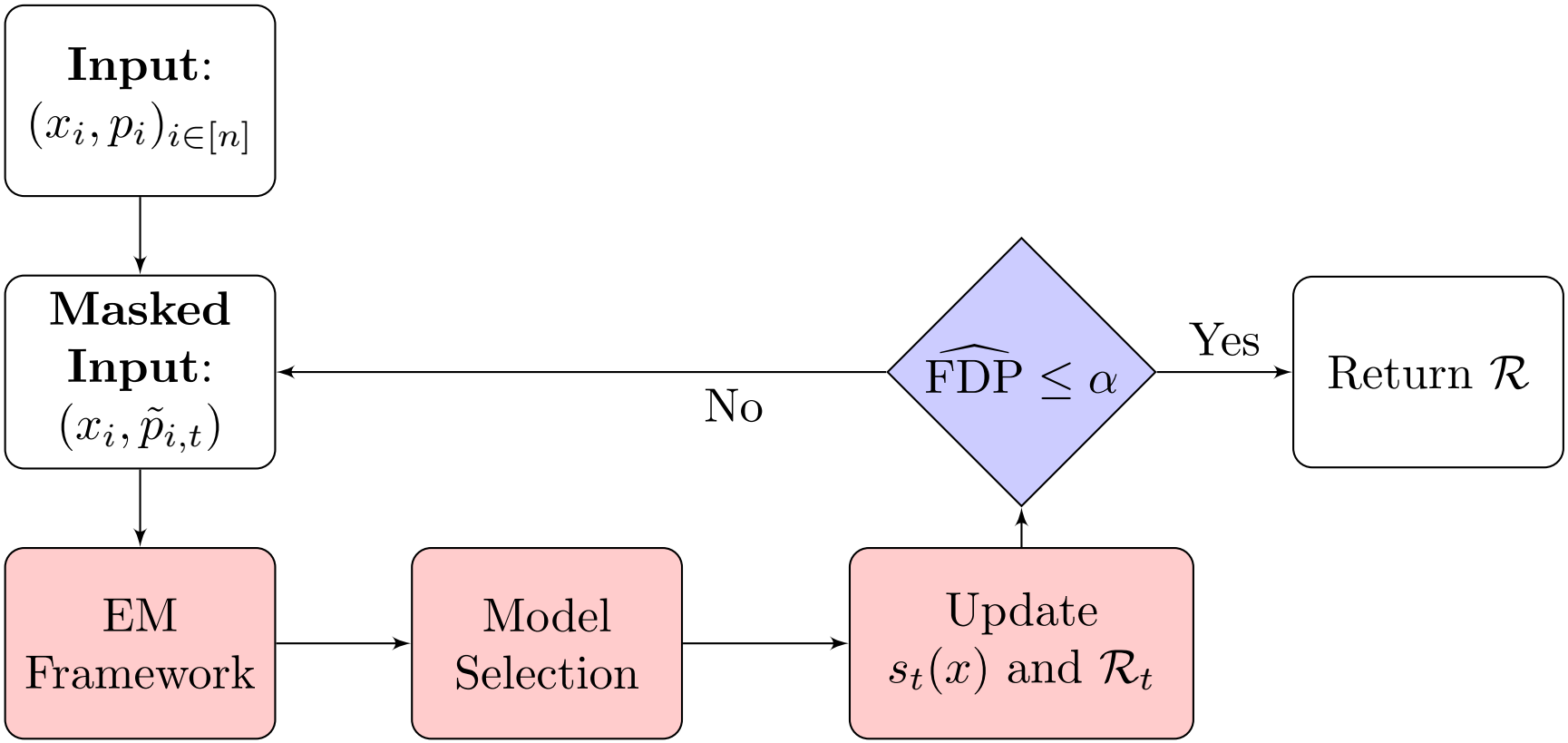
Note: Need to impute H_i and p_i (only know $\min\{p_i, 1 - p_i\}$)

EM update is easy!

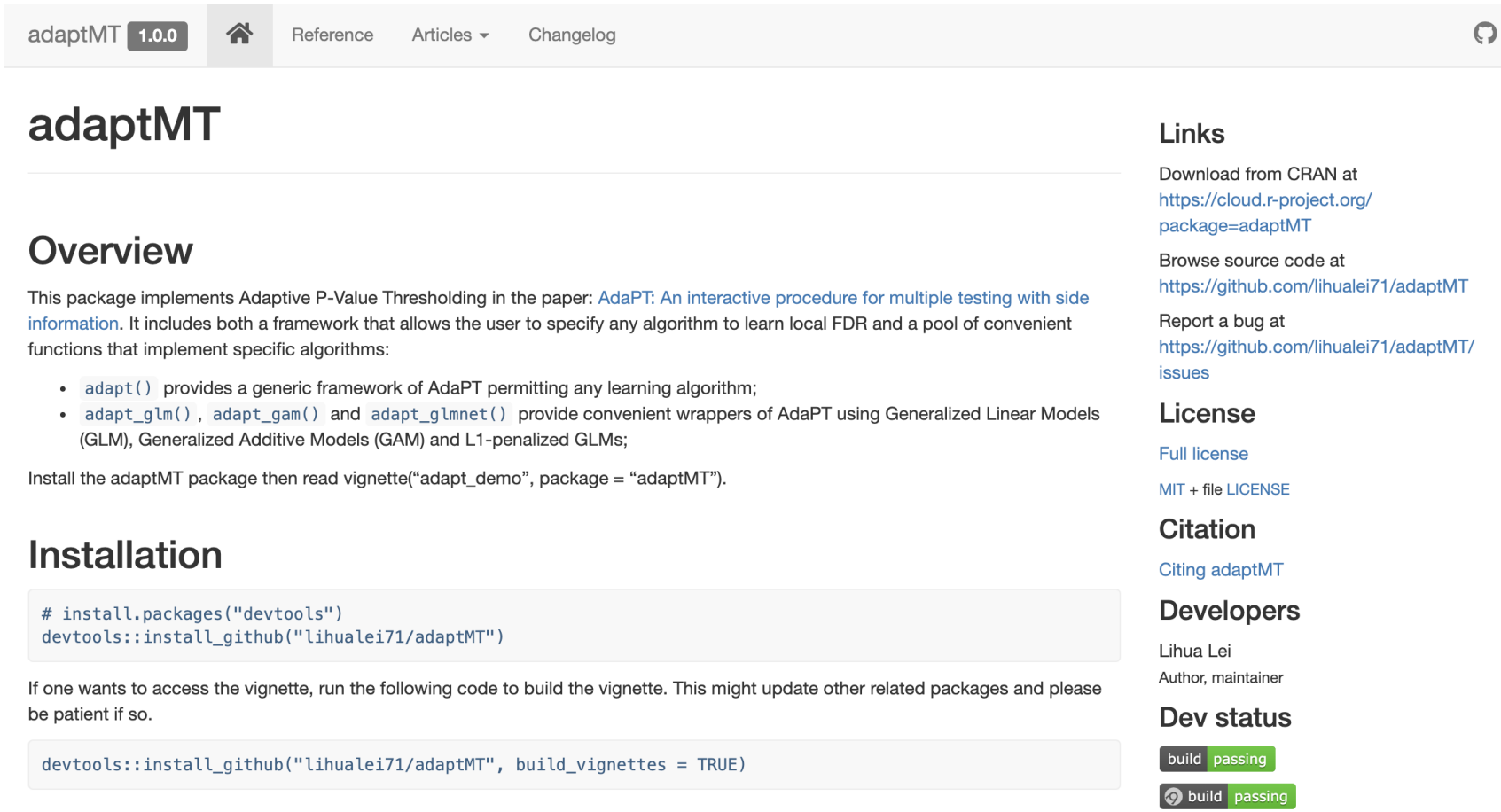
Generalized EM Framework



AdaPT Pipeline



R package: adaptMT



adaptMT 1.0.0 | Reference | Articles | Changelog

adaptMT

Overview

This package implements Adaptive P-Value Thresholding in the paper: AdaPT: An interactive procedure for multiple testing with side information. It includes both a framework that allows the user to specify any algorithm to learn local FDR and a pool of convenient functions that implement specific algorithms:

- `adapt()` provides a generic framework of AdaPT permitting any learning algorithm;
- `adapt_glm()`, `adapt_gam()` and `adapt_glmnet()` provide convenient wrappers of AdaPT using Generalized Linear Models (GLM), Generalized Additive Models (GAM) and L1-penalized GLMs;

Install the adaptMT package then read vignette("adapt_demo", package = "adaptMT").

Installation

```
# install.packages("devtools")
devtools::install_github("lihualei71/adaptMT")
```

If one wants to access the vignette, run the following code to build the vignette. This might update other related packages and please be patient if so.

```
devtools::install_github("lihualei71/adaptMT", build_vignettes = TRUE)
```

Links

Download from CRAN at https://cloud.r-project.org/package=adaptMT

Browse source code at https://github.com/lihualei71/adaptMT

Report a bug at https://github.com/lihualei71/adaptMT/issues

License

Full license

MIT + file LICENSE

Citation

Citing adaptMT

Developers

Lihua Lei

Author, maintainer

Dev status

build passing

build passing

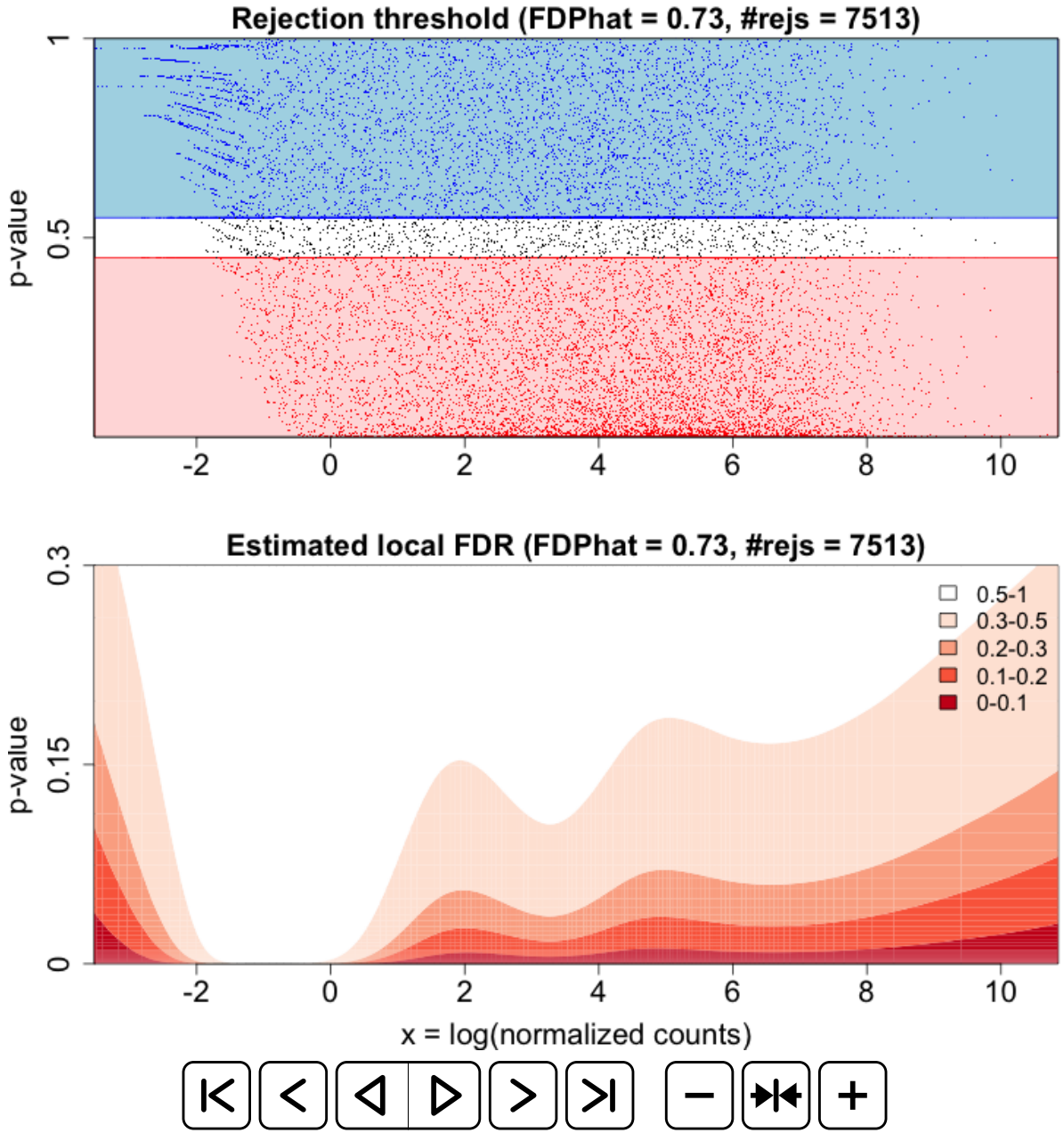
- Convenient wrappers: `adapt_glm/gam/glmnet/gbm`
- Generic interface: allows any user-specified learning algorithm

Example: RNA-seq Data

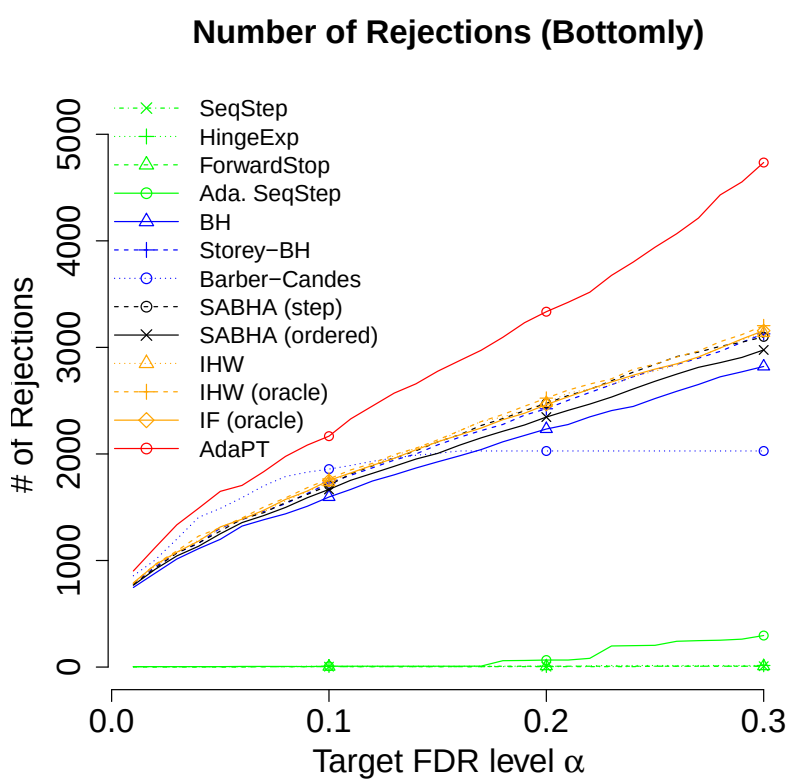
Expression in two mouse strains C57BL/6J (B6) and DBA/2J (D2)

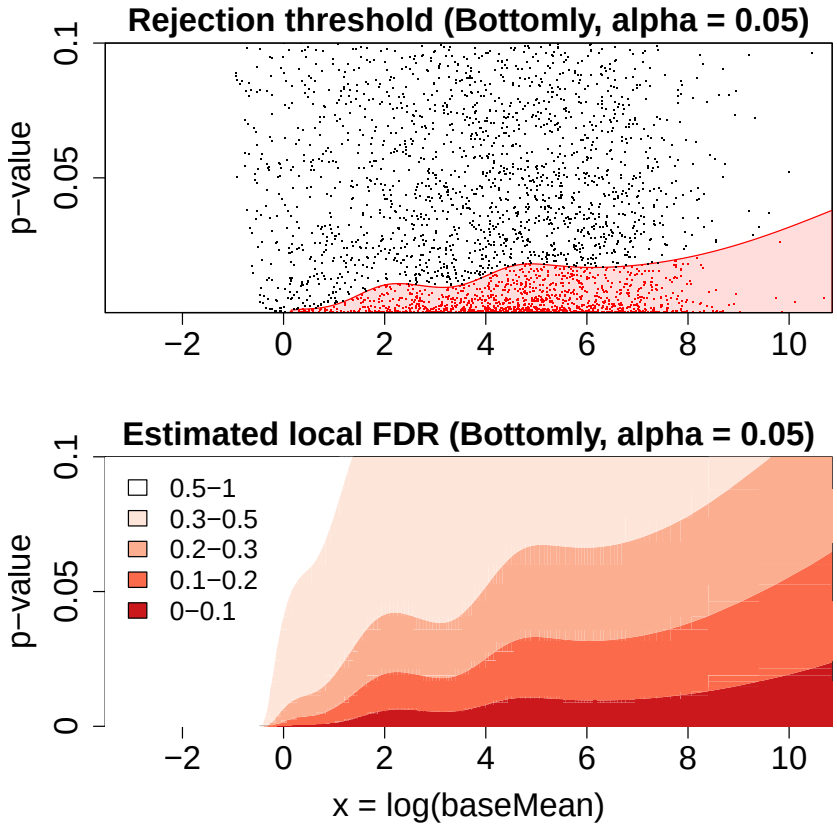
- $n = 13932$ genes, 21 samples (10 B6 and 11 D2)
- H_i : no differential response in gene i
- p_i : computed via `DEseq2` package
- x_i : logarithmic normalized count via `DEseq2` package
- Data by [Bottomly et al. '11]; studied also by [Ignatiadis et al. '16]

Animation: RNA-seq Data



RNA-seq Data: Power Comparison





Learnability of AdaPT

- Even the initial step contains a lot of information

$$\mathscr{F}_{-1} = \{(x_i, \min\{p_i, 1 - p_i\})_{i=1}^{n}\}$$

- Can be highly informative about which hypotheses are non-null
- $\min\{p_i, 1 - p_i\} = 10^{-8} \Longrightarrow H_i$ is likely false

Learnability of AdaPT

- Even the initial step contains a lot of information

$$\mathscr{F}_{-1} = \{(x_i, \min\{p_i, 1 - p_i\})_{i=1}^{n}\}$$

- Can be highly informative about which hypotheses are non-null
- $\min\{p_i, 1 - p_i\} = 10^{-8} \Longrightarrow H_i$ is likely false
- For large t, $\mathscr{F}_t \approx$ whole data set: nearly no information loss

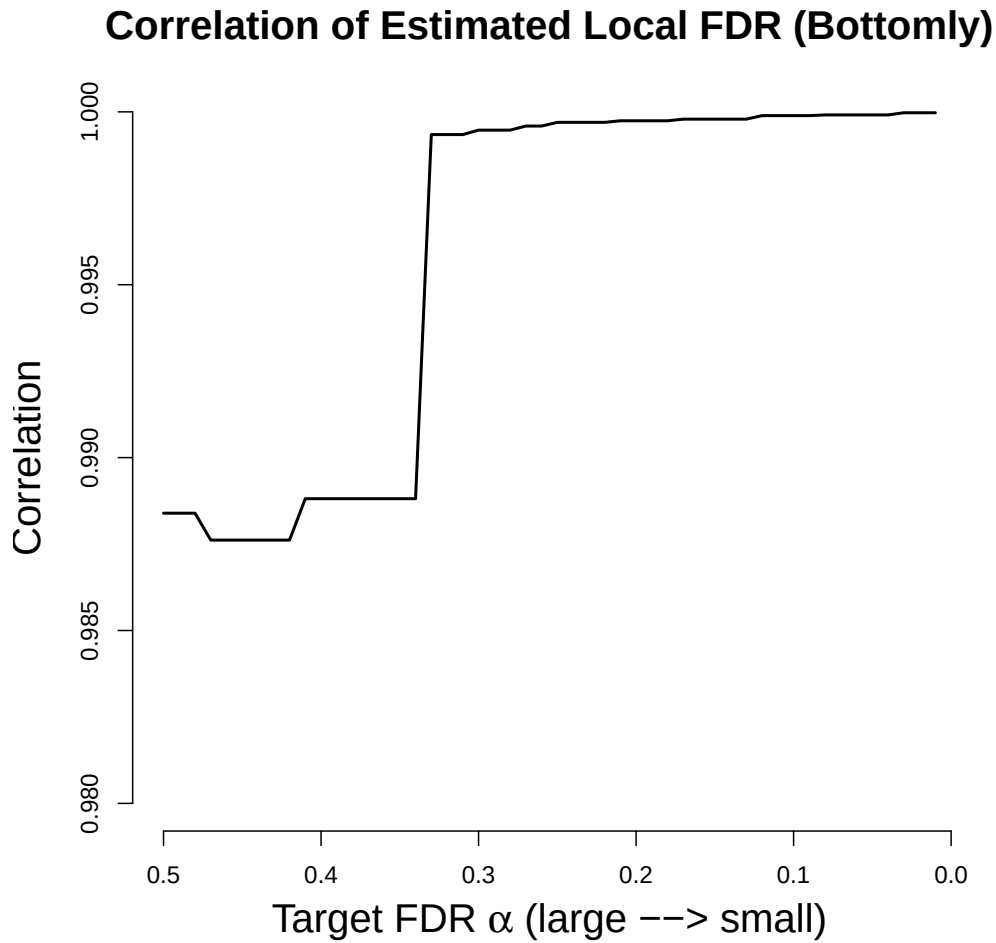
Learnability of AdaPT

- Even the initial step contains a lot of information

$$\mathscr{F}_{-1} = \{(x_i, \min\{p_i, 1 - p_i\})_{i=1}^n\}$$

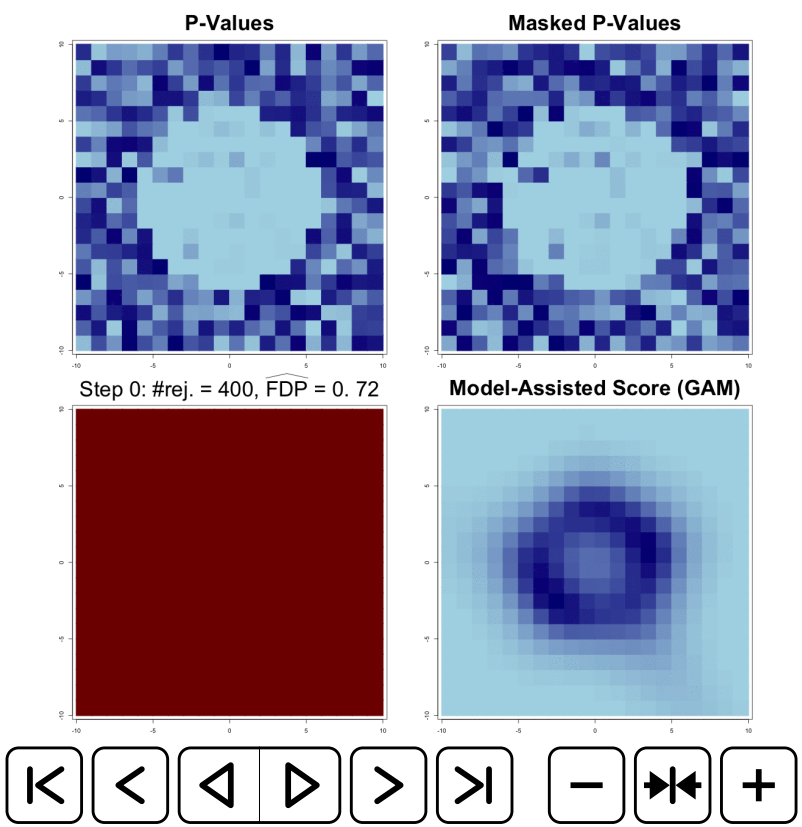
- Can be highly informative about which hypotheses are non-null
- $\min\{p_i, 1 - p_i\} = 10^{-8} \Longrightarrow H_i$ is likely false
- For large t, $\mathscr{F}_t \approx$ whole data set: nearly no information loss
- More informative than just using large p-values $\{p_i : p_i \geq \lambda\}$

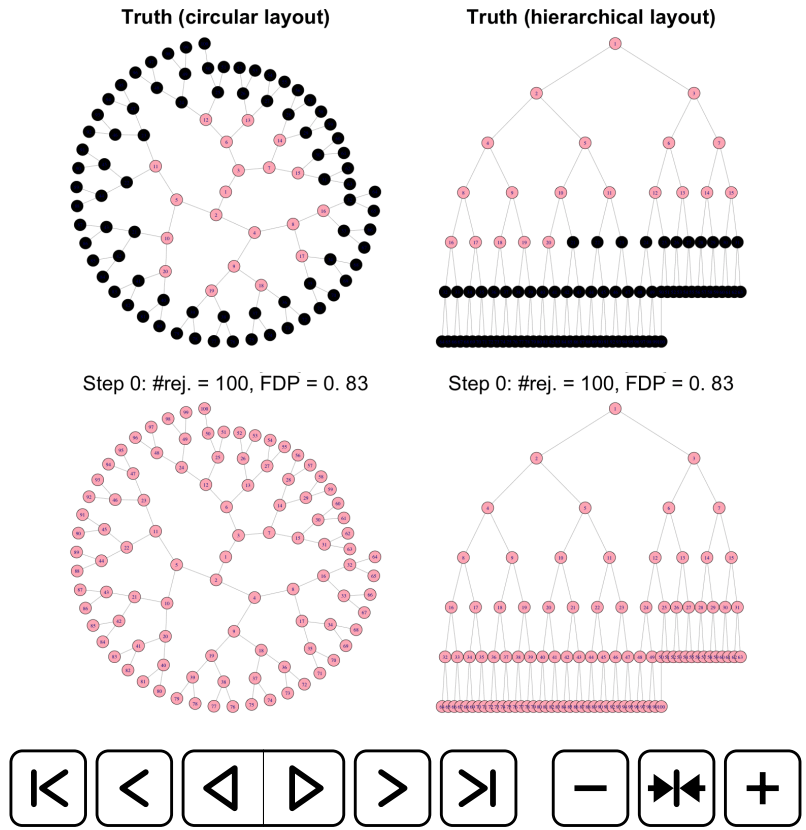
Information Loss by Partial Masking



Extension: FDR Control With Structural Constraints

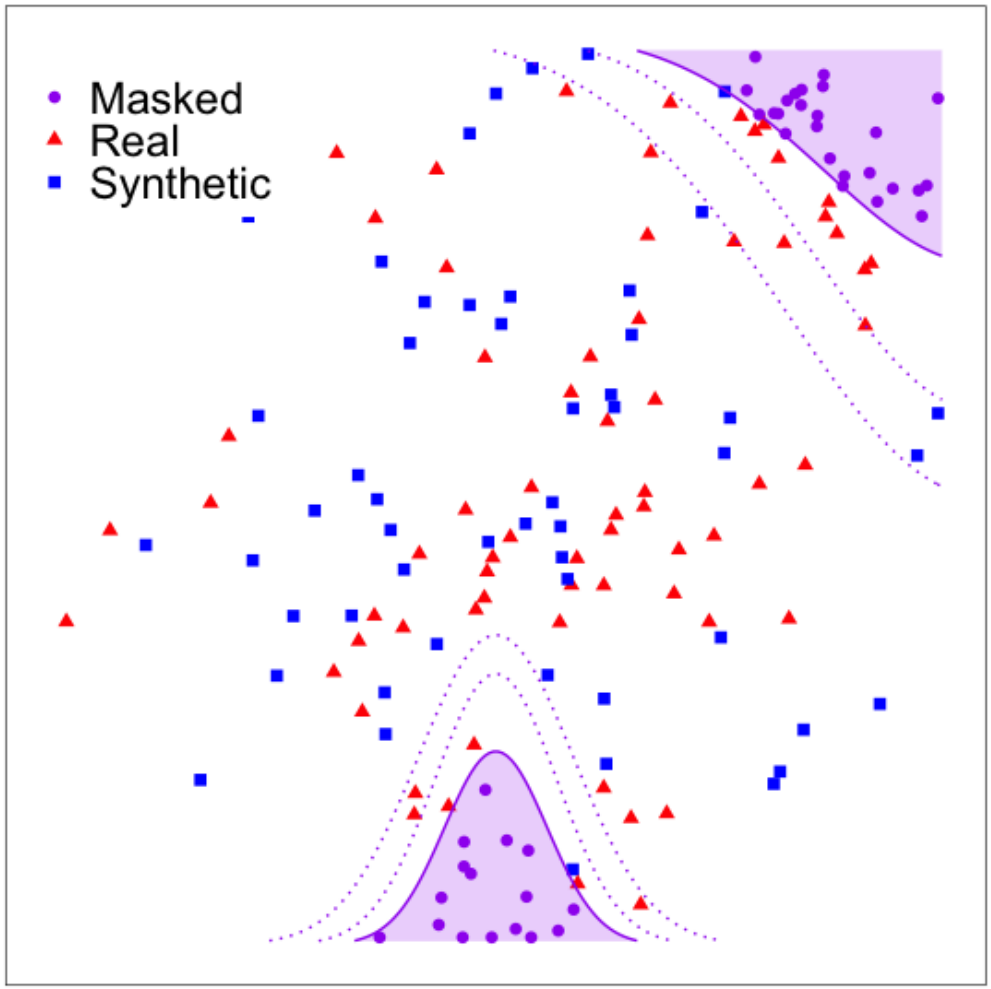
Selectively Traversed Accumulation Rules (STAR) L., Ramdas, Fithian '21





Extension: FDR Control With Multivariate Test Statistics

Bags of Null Statistics (BONuS) Yang, L., Ho, Fithian '21



Summary

We proposed Adaptive P-value Thresholding (AdaPT)

- controls FDR in **finite samples**
- robust to **arbitrary misspecification** of non-nulls
- can wrap around any black-box algorithm
- able to deal with any type of covariates

Thank you!