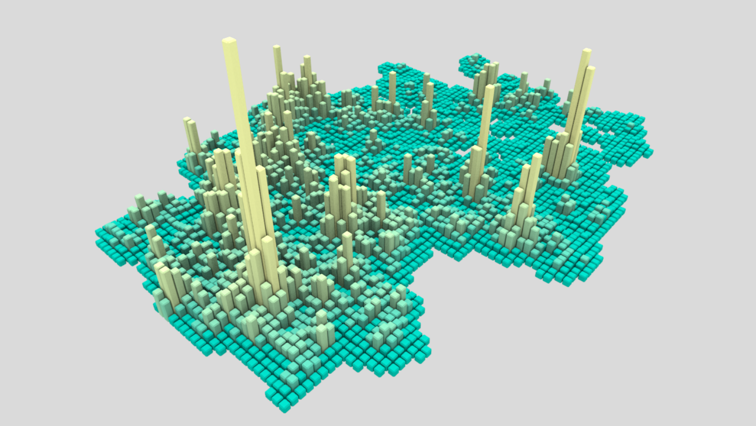
Bringing Closure to False Discovery Rate control: A General Principle for Multiple Testing

Ziyu Xu, Aldo Solari, Lasse Fischer,
Rianne de Heide, Aaditya Ramdas, Jelle Goeman

Leiden University Medical Center
The Netherlands

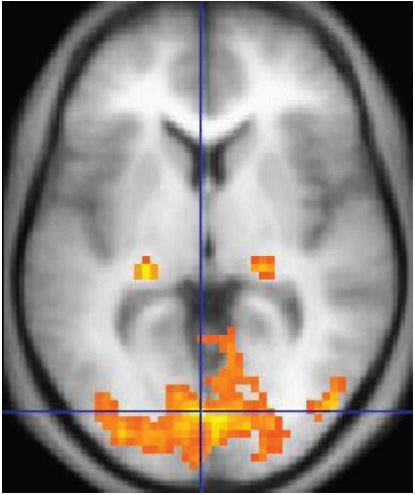
SMPGD Grenoble, 2026-01-30

High-resolution data: spatial statistics

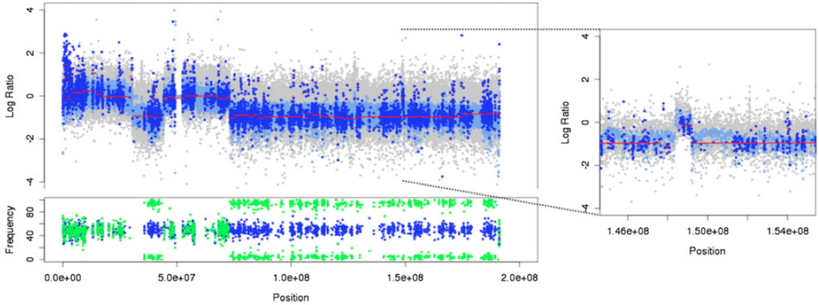


https://janakiev.com/blog/openstreetmap-with-python-and-overpass-api/

High-resolution data: neuroimaging



High-resolution data: omics



http://massgenomics.org/2012/02/exome-based-copy-number-analysis-with-varscan-2.html

High-resolution data

Unit level: which unit has signal?

Low power per unit

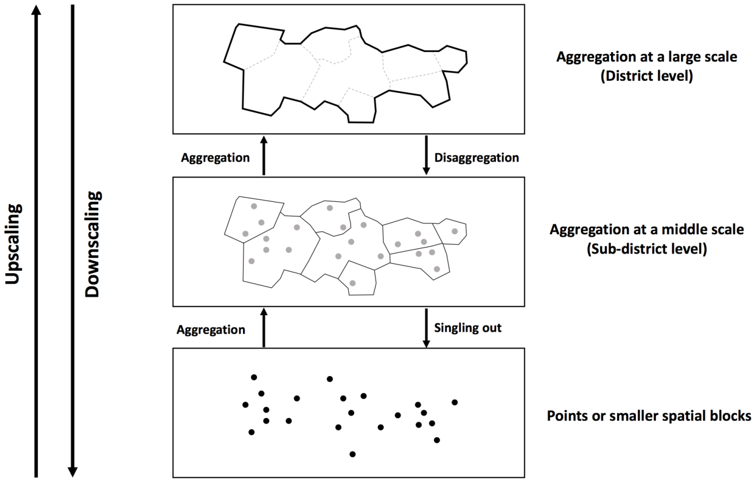
- Low signal/noise
- Huge multiple testing burden
- Precise localization of signal

Solution: aggregate to regions

At region level

- More meaningful, more interesting
- More powerful
- Less precise localization of signal

Aggregation



https://saarelgroup.wordpress.com/maps/#jp-carousel-277

Aggregation strategies

Data or knowledge?

- Atlas-based (e.g. gene ontology)
- Data-driven (e.g. clustering)
- Mixed

How much to aggregate?

- Too little: low power
- Too much: vague statements

Ideal: interactive aggregation

Find the most precise statements that the data allow

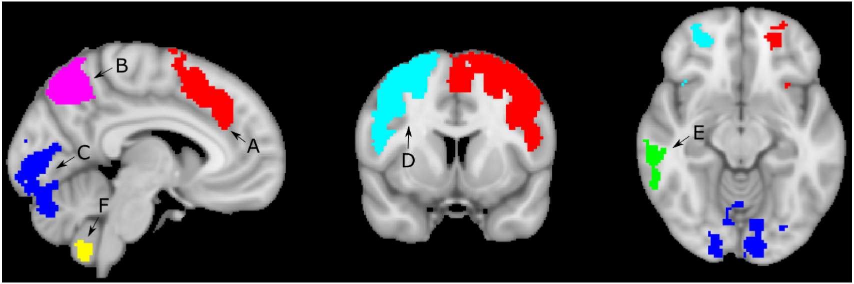
Motivating example: brain of ≈200,000 voxels

Multiplicity 1

There are $\approx 2^{200,000}$ brain areas we could look at!

Multiplicity 2

Often we are interested in more than one region



Subsetting, Enlarging or Splitting

Benjamini-Hochberg

Finds a random set R of hypotheses such that FDR(R) $\leq \alpha$.

High-dimensional data

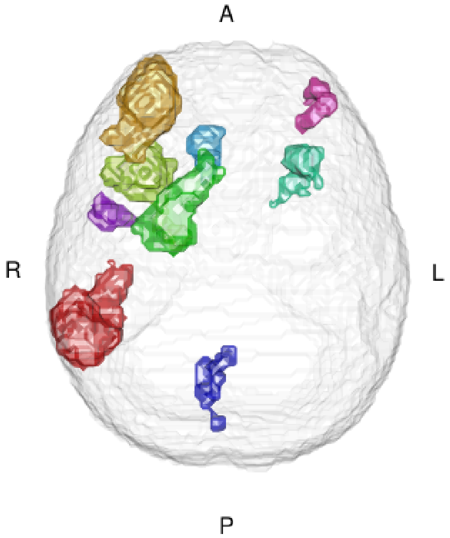
Set is often not satisfactory

- Too small
- Too large
- Not biologically meaningful

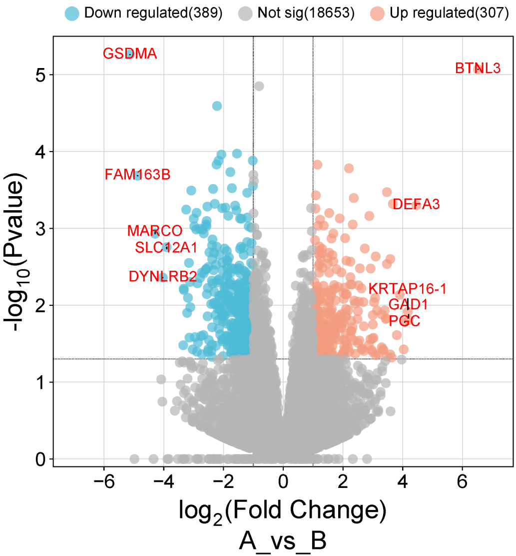
Problem

Any change to the set R destroys FDR guarantee

Simultaneity: neuroimaging



Simultaneity: genomics



A multiple testing problem

Many hypotheses

$H_1, \ldots, H_m$

Multiple testing goal

1. Many discoveries
2. Control some error rate
3. Post hoc flexibility

How to design a multiple testing procedure?

e-Closure

A general recipe for making multiple testing methods

Building blocks

Intersection hypotheses and *e*-values

Contributions

- Recovers the Closure Principle for familywise error rate
- Extends to False Discovery Rate
- Uniformly improves eBH and Benjamini-Yekutieli
- Introduces unprecedented flexibility in multiple testing

A general view of error rates

True and false hypotheses

$N \subseteq [m]$ hypotheses are null; the rest are potential discoveries

Famous error rates

- Familywise error: $\mathrm{P}(|R \cap N| > 0) \leq \alpha$
- Per-family error rate: $\mathrm{E}(|R \cap N|) \leq \alpha$
- False Discovery Rate (FDR): $\mathrm{E}(\frac{|R \cap N|}{|R|}) \leq \alpha$

General form

Control some expected loss: $\mathrm{E}(\mathrm{f}_N(R)) \leq \alpha$

The e-value

e-value

- $e \geq 0$; Large e indicates evidence against H_0
- If H_0 is true: $\mathrm{E}(e) \leq 1$
- Reject when $e \geq 1/\alpha$ for Type I error control

$$\mathrm{P}(e \geq 1/\alpha) \leq \alpha \mathrm{E}(e) \leq \alpha \qquad \text{(Markov)}$$

Convert p-value to e-value (calibrator)

Many possibilities: decreasing $g \geq 0$ such that $\int_0^1 g(p)\mathrm{d}p = 1$

All-or-nothing p-to-e calibrator

$g(p) = \frac{1\{p \leq \alpha\}}{\alpha}$ reproduces usual hypothesis test result

Intersection hypotheses

Intersection hypothesis

For $S \subseteq [m]$, $H_S = \bigcap_{i \in S} H_i$, which is true iff all H_i, $i \in S$ true

The e-collection

$E = (e_S)_{S \subseteq [m]}$: local e-values such that $\mathrm{E}(e_N) \leq 1$

Sufficient

Each e_S is an e-value for H_S, $S \subseteq [m]$

The e-Closure procedure

The e-Closed Procedure

$$\mathcal{R}_\alpha(E) = \{R \subseteq [m] : \alpha e_S \geq \mathrm{f}_S(R) \quad \forall S \subseteq [m]\}$$

The e-Closure principle

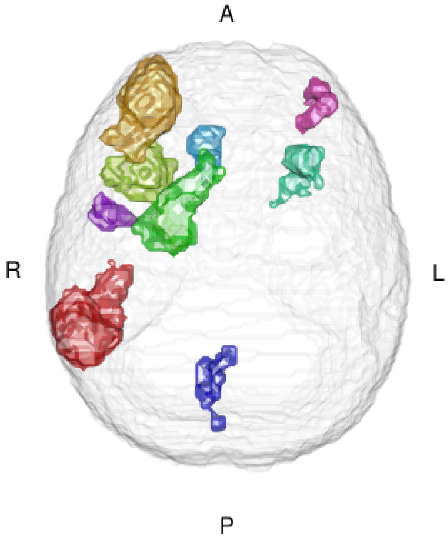
R controls $\mathrm{E}(\mathrm{f}_N(R)) \leq \alpha$ iff $R \in \mathcal{R}_\alpha(E)$ for an e-collection E

Simultaneous control

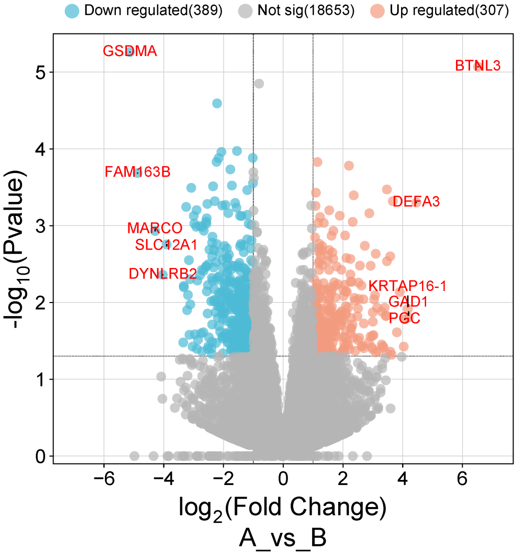
$\mathrm{E}(\mathrm{f}_N(R)) \leq \alpha$ simultaneously over $R \in \mathcal{R}_\alpha(E)$

$$\mathrm{E}\Big(\max_{R \in \mathcal{R}_\alpha(E)} \mathrm{f}_N(R)\Big) \leq \alpha$$

Simultaneity: neuroimaging



Simultaneity: genomics



Post hoc error rate

All error rates

$$\mathcal{F} = \{\text{all functions } \mathrm{f}_N(R)\}$$

Simultaneous (= post hoc) choice of error

$$\mathrm{E}\Big(\sup_{\mathrm{f}\in\mathcal{F}} \max_{R\in\mathcal{R}_\alpha^{\mathrm{f}}(E)} \mathrm{f}_N(R)\Big) \leq \alpha$$

Use case

Switch from FWER to FDR if not much signal present

State of the art: eBH

Per hypothesis *e*-values

$e_1 \geq \ldots \geq e_m$ for $H_1 \ldots, H_m$

eBH = e-Benjamini Hochberg

Rejects $R = [r]$ with $r = \max\{i : me_i/i \geq 1/\alpha\}$

Compare

Benjamini-Hochberg on $1/e$

Validity

For all joint distributions of *e*-values (unlike BH)

FDR via *e*-Closure

Local *e*-values

Only admissible way of combining *e*-values (Vovk & Wang)

$$e_S = \frac{1}{|S|} \sum_{i \in S} e_i$$

e-Closure result

$$\mathcal{R}_\alpha(E) = \left\{ R \subseteq [m] : \frac{\alpha}{|S|} \sum_{i \in S} e_i \geq \frac{|R \cap S|}{|R|} \quad \forall S \subseteq [m] \right\}$$

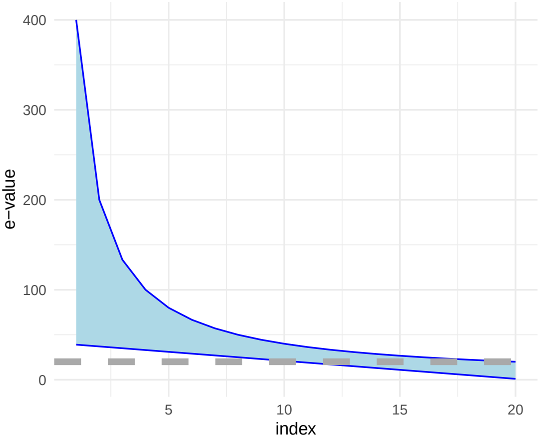
Theorem

$\mathcal{R}_\alpha(E)$ uniformly improves eBH

Improvement illustration

If all *e*-values in this region

eBH rejects nothing; closed eBH rejects all hypotheses



Computation

An exponential nightmare?

$$\mathcal{R}_\alpha(E) = \left\{ R \subseteq [m] : \frac{\alpha}{|S|} \sum_{i \in S} e_i \geq \frac{|R \cap S|}{|R|} \quad \forall S \subseteq [m] \right\}$$

Use

- Worst case S given $|S|$ and $|R \cap S|$ is known
- Sums can be pre-calculated
- Optimization is convex in some parameters

Result

Find largest R in $O(m^2 \log m)$, in practice $m \log m$ time.

Benjamini-Yekutieli

Start from p-values

$p_1 \leq \ldots \leq p_m$ with unknown joint distribution ($\rightarrow$ no BH)

Benjamini-Yekutieli procedure

Reject $R = [r]$ with $r = \max\{i : ip_i/mh_m \geq \alpha\}$

Added harmonic number correction

Conservative by $h_m = \sum_{j=1}^{m} 1/j$ compared to BH

Can we improve?

We use e-Closure

Closed BY

A p-to-e calibrator (Xu et al)

$$g(p) = \frac{k1\{h_k p \leq \alpha\}}{\alpha \lceil k h_k p / \alpha \rceil}$$

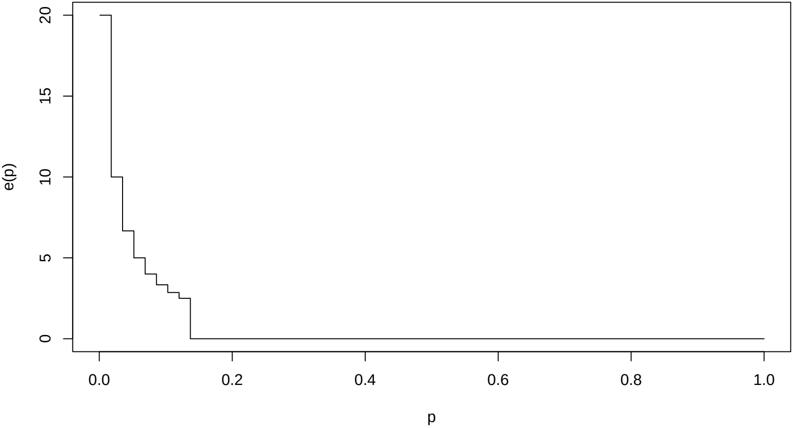
Resulting local e-value

$$e_S = \sum_{i \in S} \frac{1\{h_{|S|} p_i \leq \alpha\}}{\alpha \lceil |S| h_{|S|} p_i / \alpha \rceil}$$

Apply e-Closure

Resulting Closed BY uniformly improves BY

Harmonic Calibrator



The harmonic calibrator for $k = 8$ and $\alpha = 0.4$

BY vs Closed BY in standard data sets

Dataset	m	BY / $\overline{BY}$ rejections $\alpha = 5\%$	BY / $\overline{BY}$ rejections $\alpha = 10\%$	source
APSAC	15	3 / 3	3 / 5	BH ’95
NAEP	34	6 / 8	8 / 11	BH ’00
PADJUST	50	12 / 15	17 / 20	`p.adjust`
PVALUES	4289	129 / 145	225 / 275	`fdrtool`
VANDEVIJVER	4919	614 / 677	779 / 866	Goeman Solari ’14
GOLUB	7128	617 / 648	743 / 799	Efron Hastie ’16

Post hoc α

Flexibility

- Choose rejected set post hoc
- Choose error loss post hoc

Go one step further

Choose α post hoc (Koning 2023)

Simultaneous over α

$$\mathrm{E}\left(\sup_{\alpha \in (0,1)} \sup_{\mathrm{f} \in \mathcal{F}} \max_{R \in \mathcal{R}_\alpha^{\mathrm{f}}(E)} \frac{\mathrm{f}_N(R)}{\alpha} \right) \leq 1$$

Requires

e-values do not depend on α

Benjamini-Hochberg

Can regular Benjamini-Hochberg be improved?

Reject $R = [r]$ with $r = \max\{i : ip_i/m \geq \alpha\}$

Fits the e-Closure principle but with a weird e-collection

$$e_S = \frac{m}{|S|} \sum_{i \in S} \frac{1}{\alpha |R|} \mathbf{1}\left\{ p_i \leq \frac{\alpha |R|}{m} \right\}$$

No improvement or simultaneity.

Weird facts

1. BH is known to be inadmissible $\rightarrow$ Minimally Adaptive BH
2. Needs PRDS: assumption on distribution of p's of false nulls

e-Closure

A general recipe for making multiple testing methods

Unites "all" multiple testing methods

Simplifies multiple testing

Choose how to summarize evidence against H_S; rest is computation

Flexibility

Simultaneous over rejected sets, error rates, α

Power

Uniformly improves known methods: eBH, BY, not (yet?) BH

References

Wang, Ramdas (2022)
False discovery rate control with e-values
JRSSB 84 (3) 822-–852

Z Xu, A Solari, L Fischer, R de Heide, A Ramdas, J Goeman (2025)
Bringing closure to FDR control: a general principle for multiple testing
arxiv 2509.02517