

Multiple testing for genomics

(updated 2022-02-03)

Pierre Neuvial

CNRS & Institut de Mathématiques de Toulouse

License CC BY-NC-ND 4.0

References

- A survey:

Jelle J. Goeman, Aldo Solari. **Tutorial in biostatistics: multiple hypothesis testing in genomics.** *Statistics in Medicine*, 2014

- Lecture notes from Etienne Roquain (more mathematical):

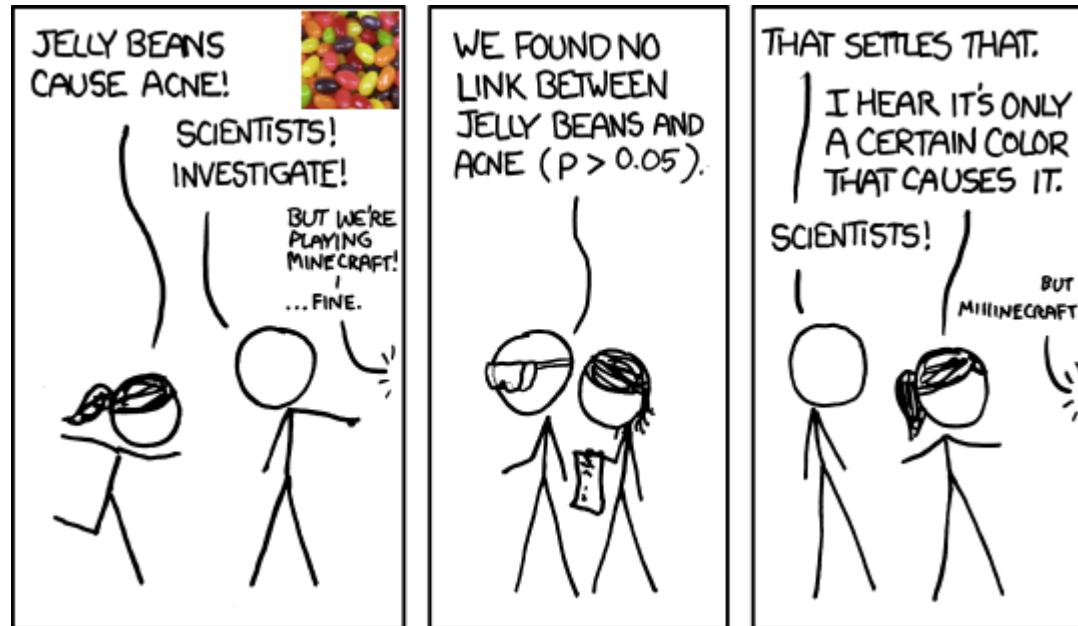
On controlling the amount of false positives when making multiple simultaneous tests. Summer School *Mathematical Methods of Statistics*, Angers, 2016. <http://www.lebesgue.fr/content/sem2016-MMS2016-Roquain>

Introduction

Examples of multiple testing situations

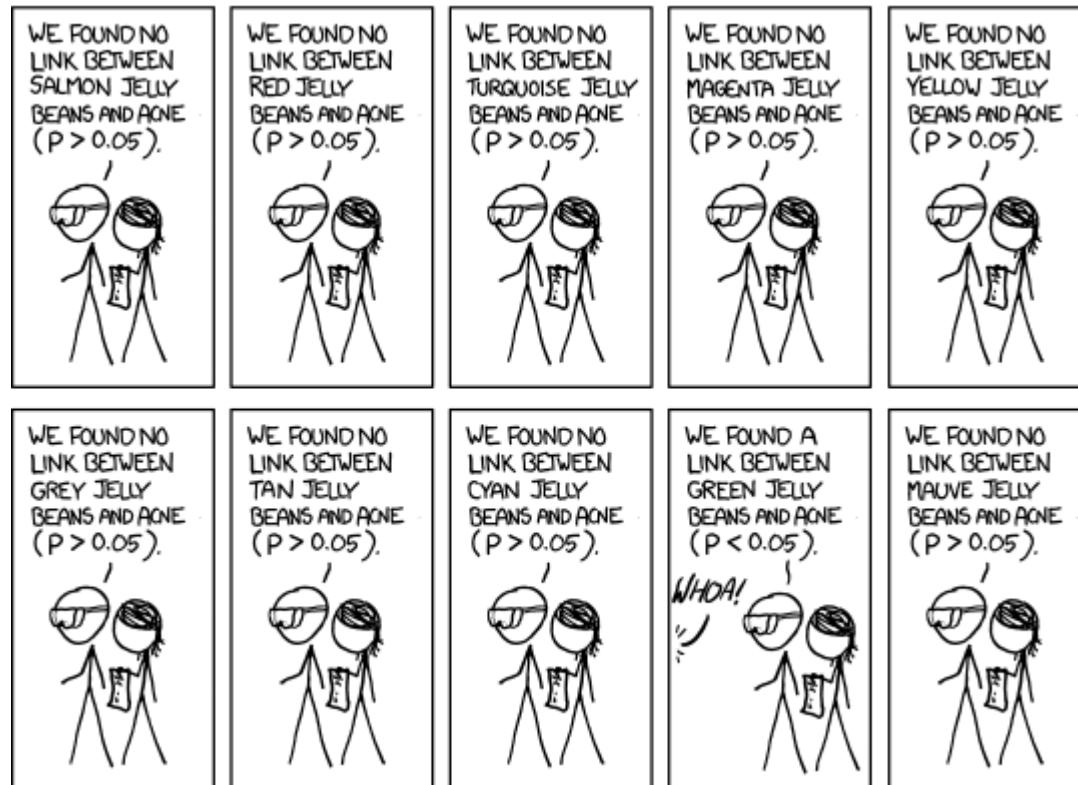
Example 0 - Jelly beans

Source: <http://imgs.xkcd.com/comics/significant.png>



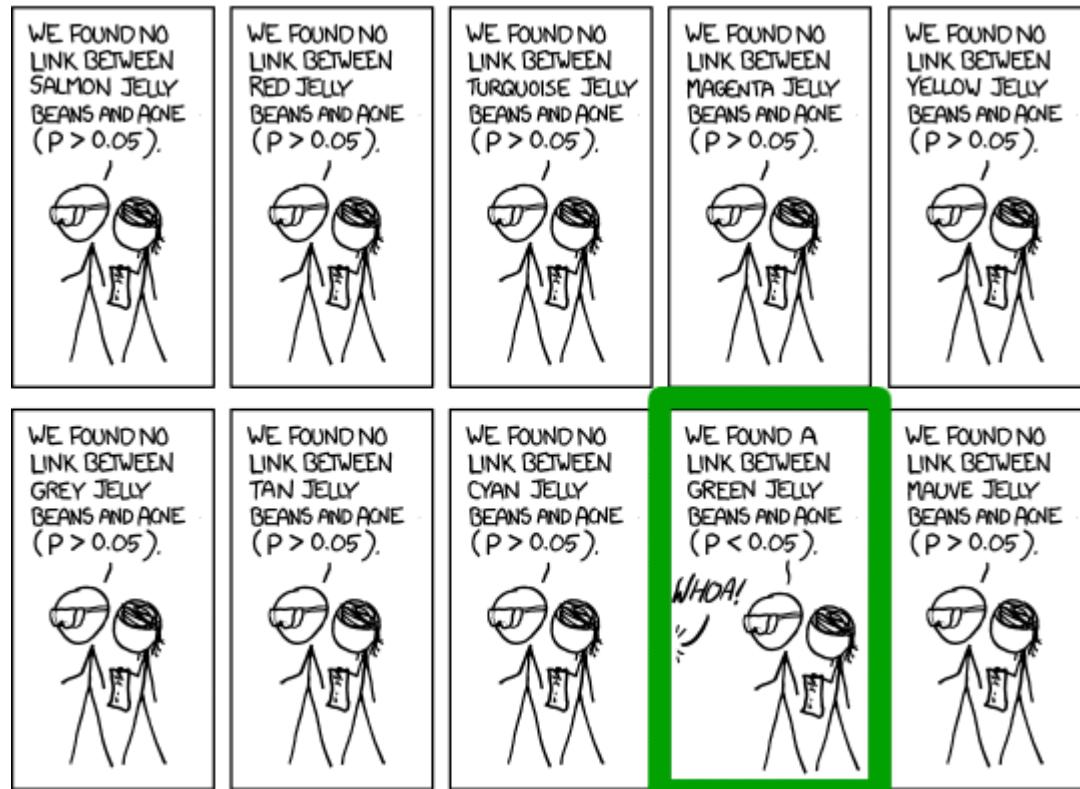
Example 0 - Jelly beans

Source: <http://imgs.xkcd.com/comics/significant.png>



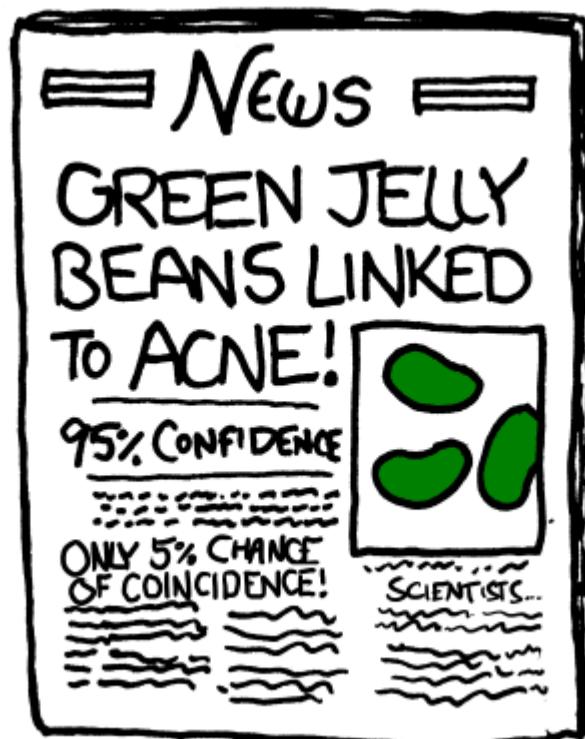
Example 0 - Jelly beans

Source: <http://imgs.xkcd.com/comics/significant.png>

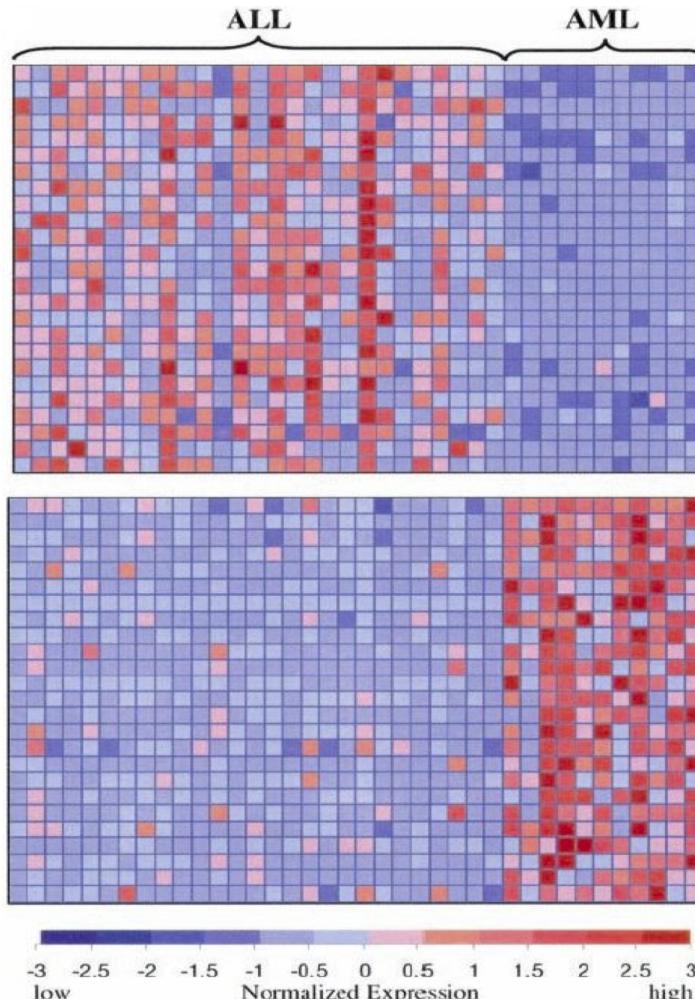


Example 0 - Jelly beans

Source: <http://imgs.xkcd.com/comics/significant.png>



Example 1 - Differential expression analyses

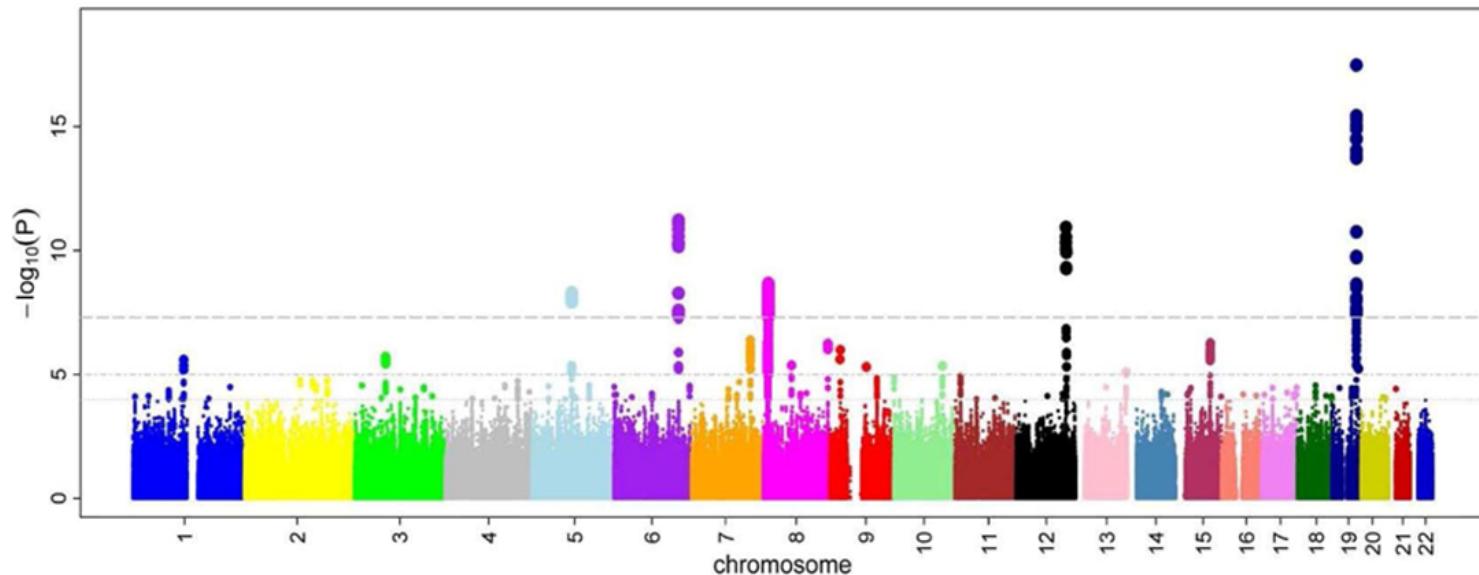


One test for each gene!

- $m \sim 10^4$ tests
(genes/transcripts/splicing variants)
- $n \sim 10^1 - 10^3$ observations
(individuals)

Which genes are differentially expressed?

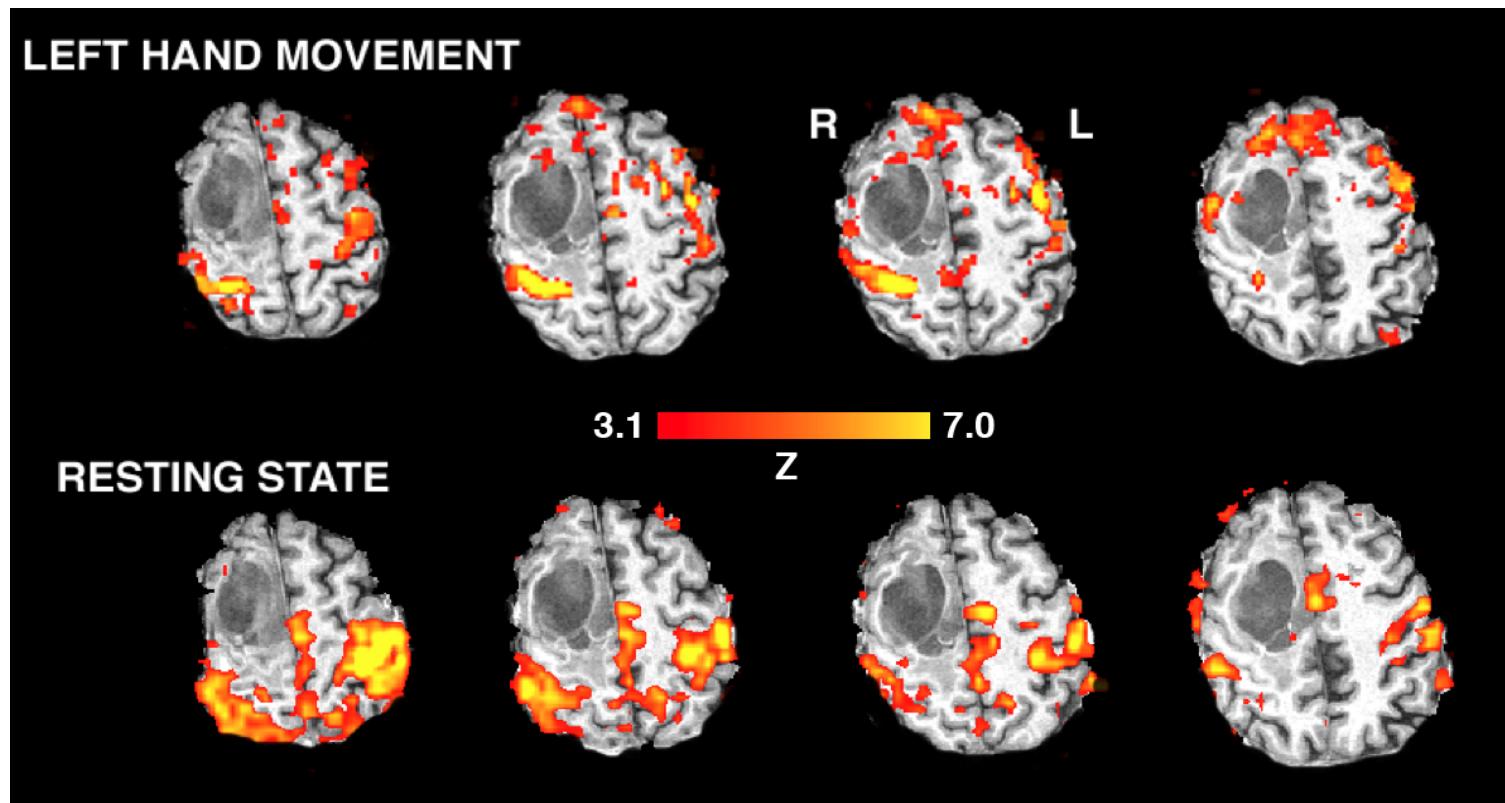
Example 2 - Genome-Wide Association Studies



- $m \sim 10^6$ tests (genomic markers)
- $n \sim 10^3 - 10^4$ observations (individuals)

Which markers are significantly associated with a phenotype of interest?

Example 3 - Neuroimaging



Which voxels are significantly more activated during movement?

Questions

What is a good definition of "significant" when performing many tests simultaneously?

definition of multiple testing risks

How can we control these multiple testing risks?

study of multiple testing procedures

Illustration: a differential expression analysis

Leukemia data set

Chiaretti et. al., *Clinical cancer research*, 11(20):7209–7219, 2005

Data and code available from <https://pneuvial.github.io/sanssouci/>

- Expression measurements (mRNA)
 - $m = 9038$ genes
 - $n = 79$ cancer patients
- Two groups of patients:

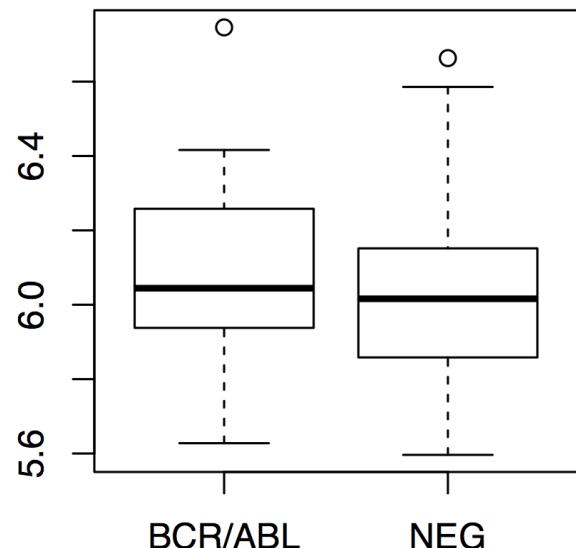
BCR/ABL	NEG
37	42

Question: find genes whose average expression differs between the two groups

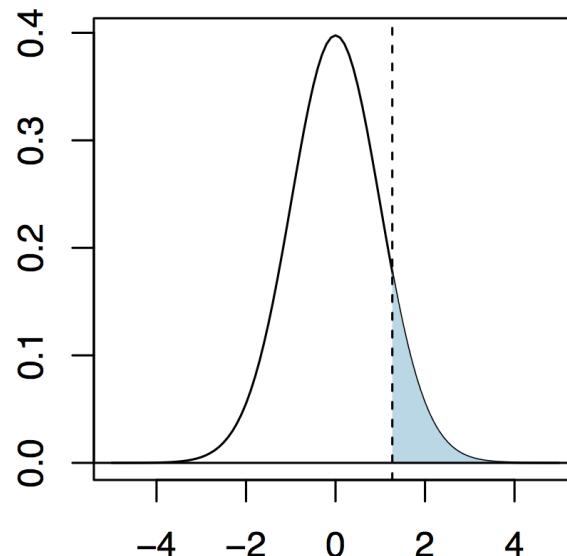
Marginal Testing framework: for each gene, test the null hypothesis of no different expression.

Test for one gene

33231_at:



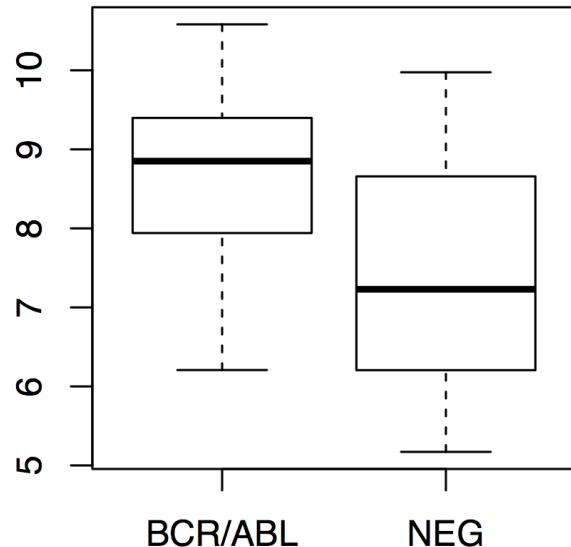
stat = 1.27 ; p = 0.21



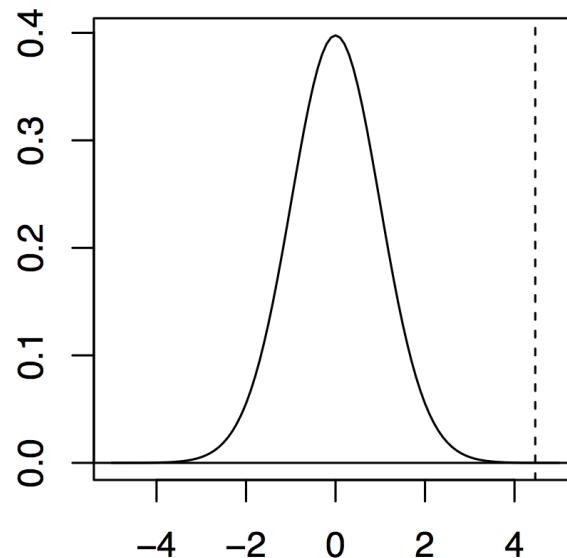
- p -value = blue area under the curve
- Here: No evidence of difference between groups

Test for another gene

33232_at:



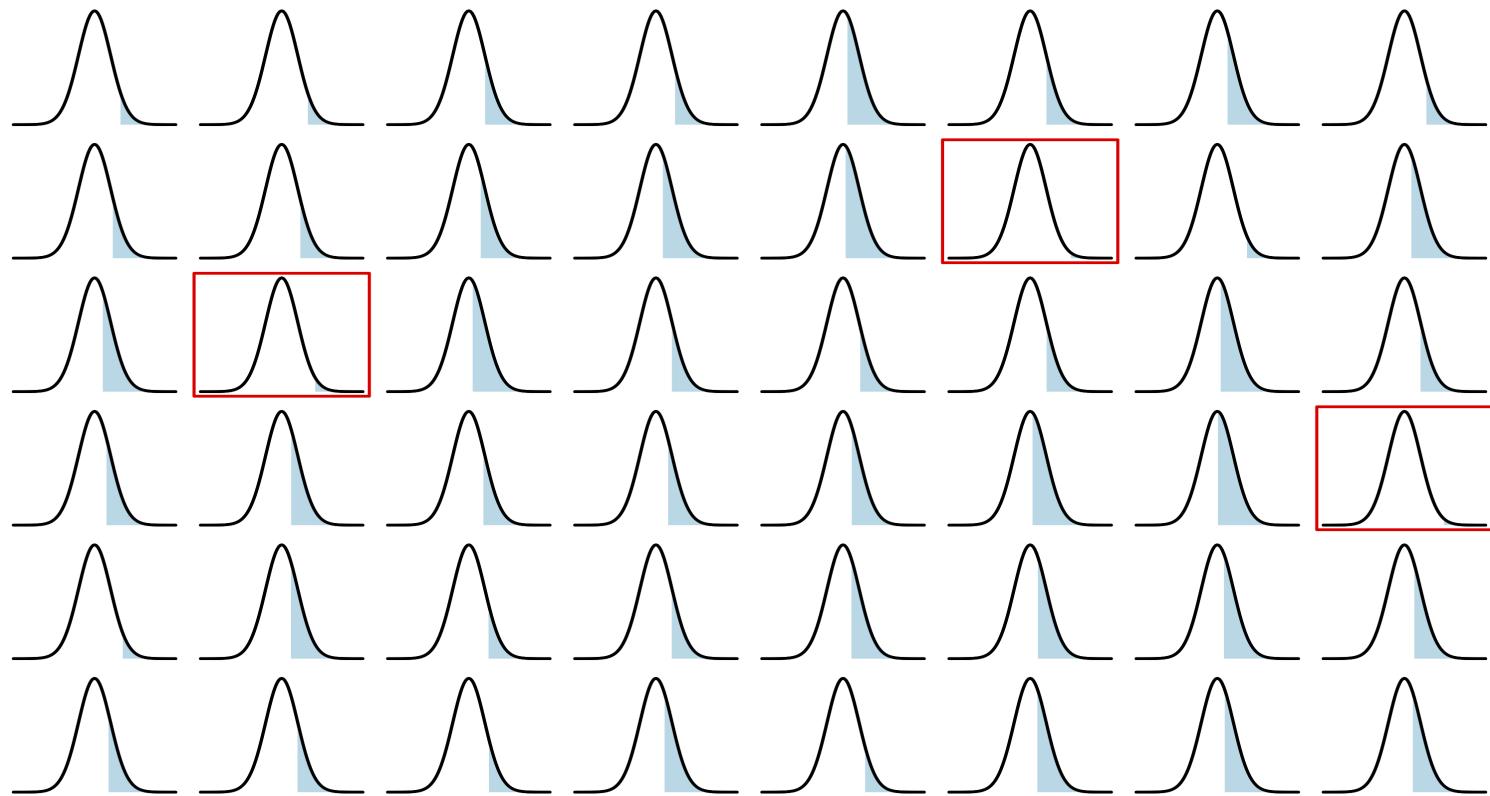
stat = 4.46 ; p = 2.7e-05



- p -value = blue area under the curve
- Here: Some evidence of difference between groups. "Significant"?

Multiple testing ($m = 48$)

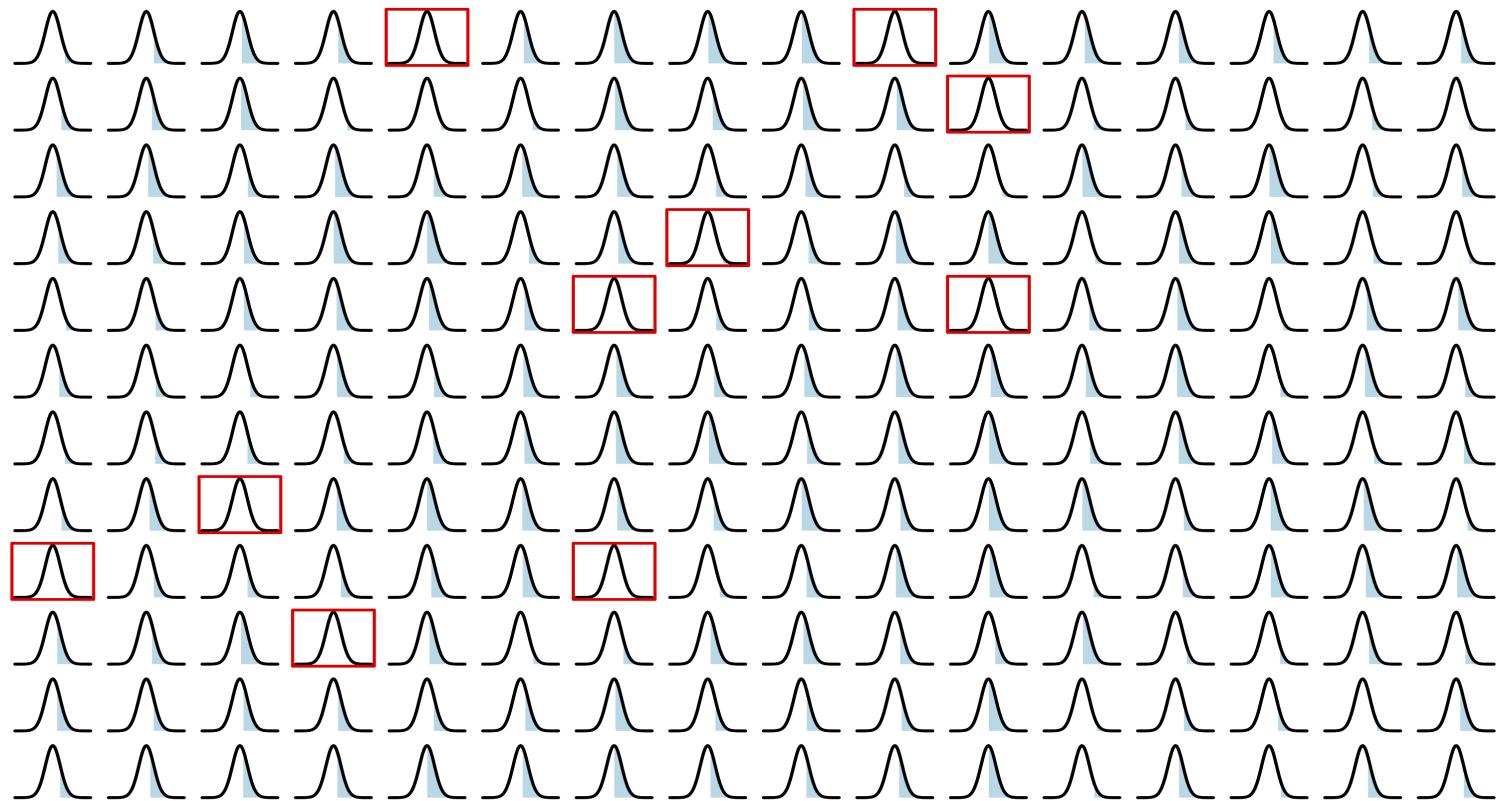
Example of pure (independent) noise:



Genes with p -value < 0.05 are highlighted in red

Multiple testing ($m = 192$)

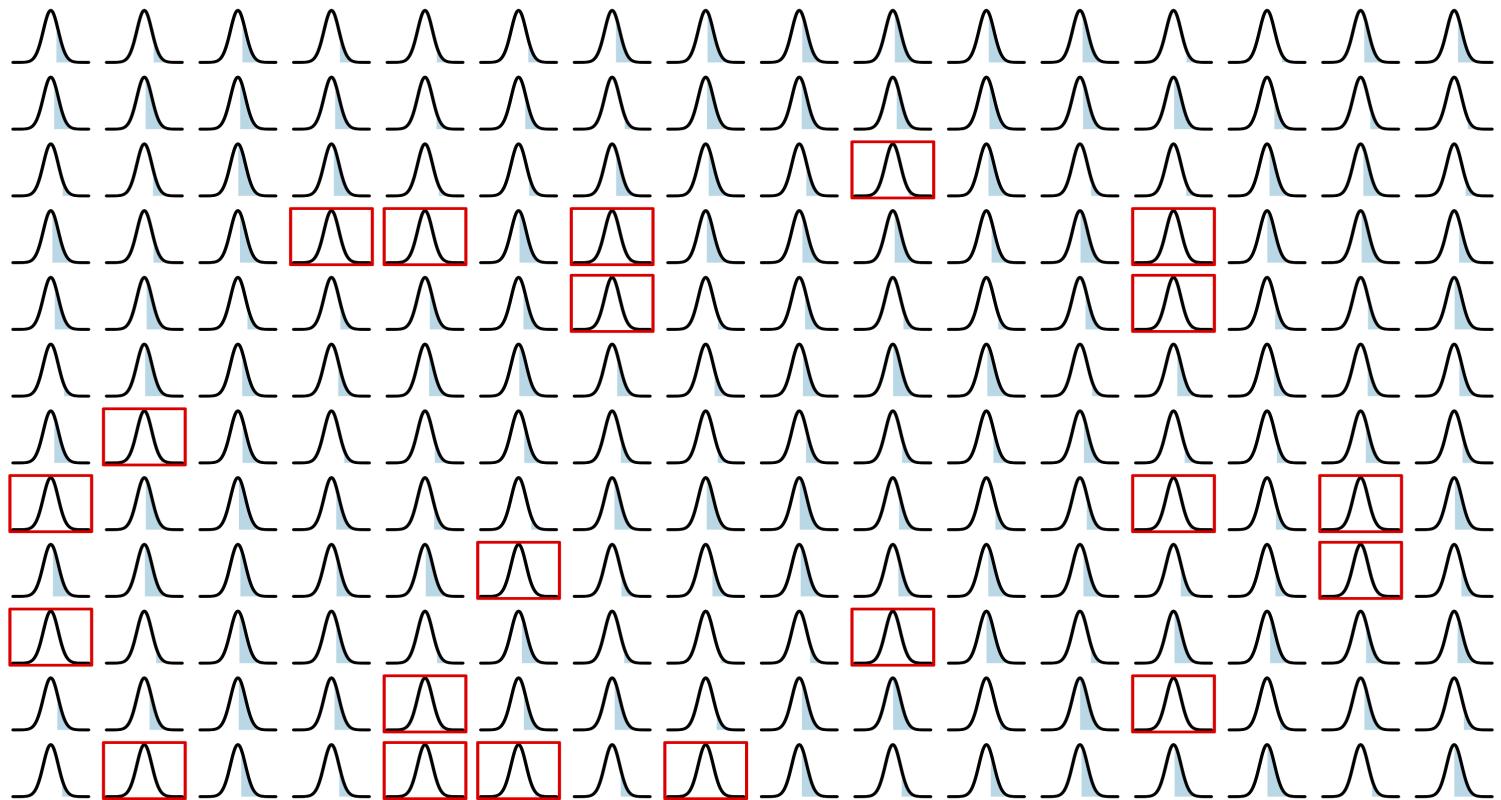
Example of pure (independent) noise:



Genes with $p\text{-value} < 0.05$ are highlighted in red

Multiple testing ($m = 192$)

First 192 genes of the Leukemia data set:



Genes with $p\text{-value} < 0.05$ are highlighted in red

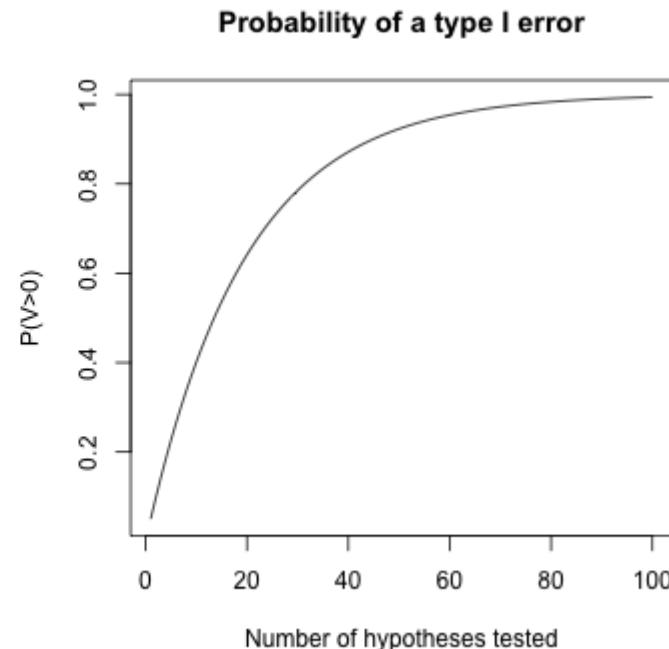
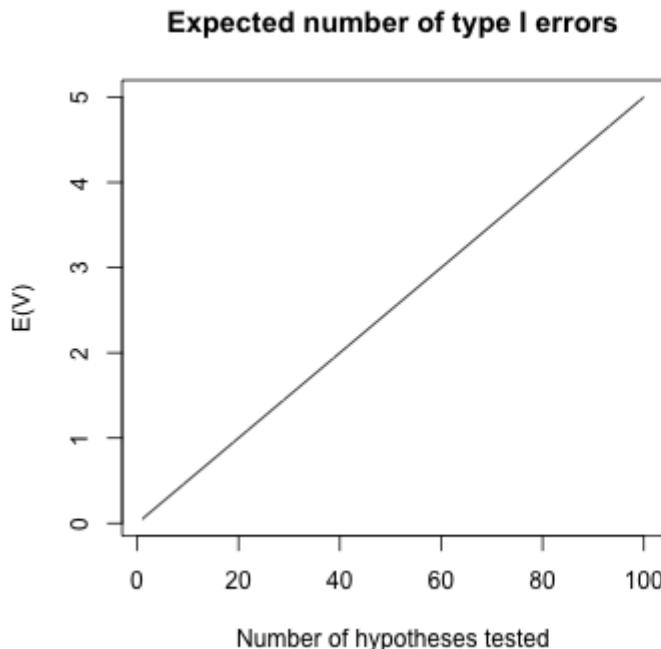
Thresholding p -values at a fixed level is not acceptable

Data: p_1, \dots, p_m : p -values for m tests

Strategy: reject \mathcal{H}_0 for all i such that $p_i \leq \alpha$

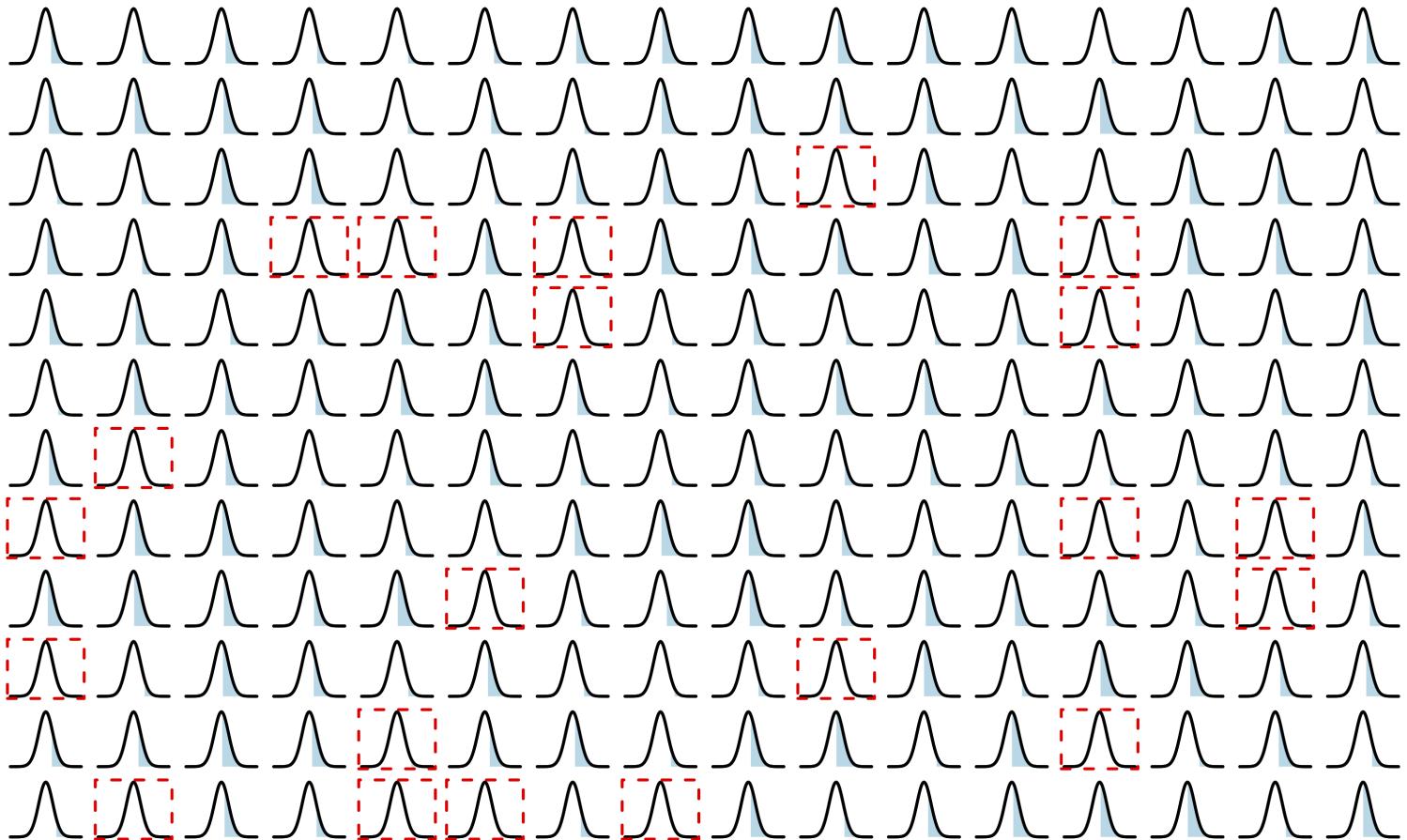
- Expected # of type I errors
- Probability of a type I error

...scales linearly with m [proof]
...quickly grows to 1 [proof]

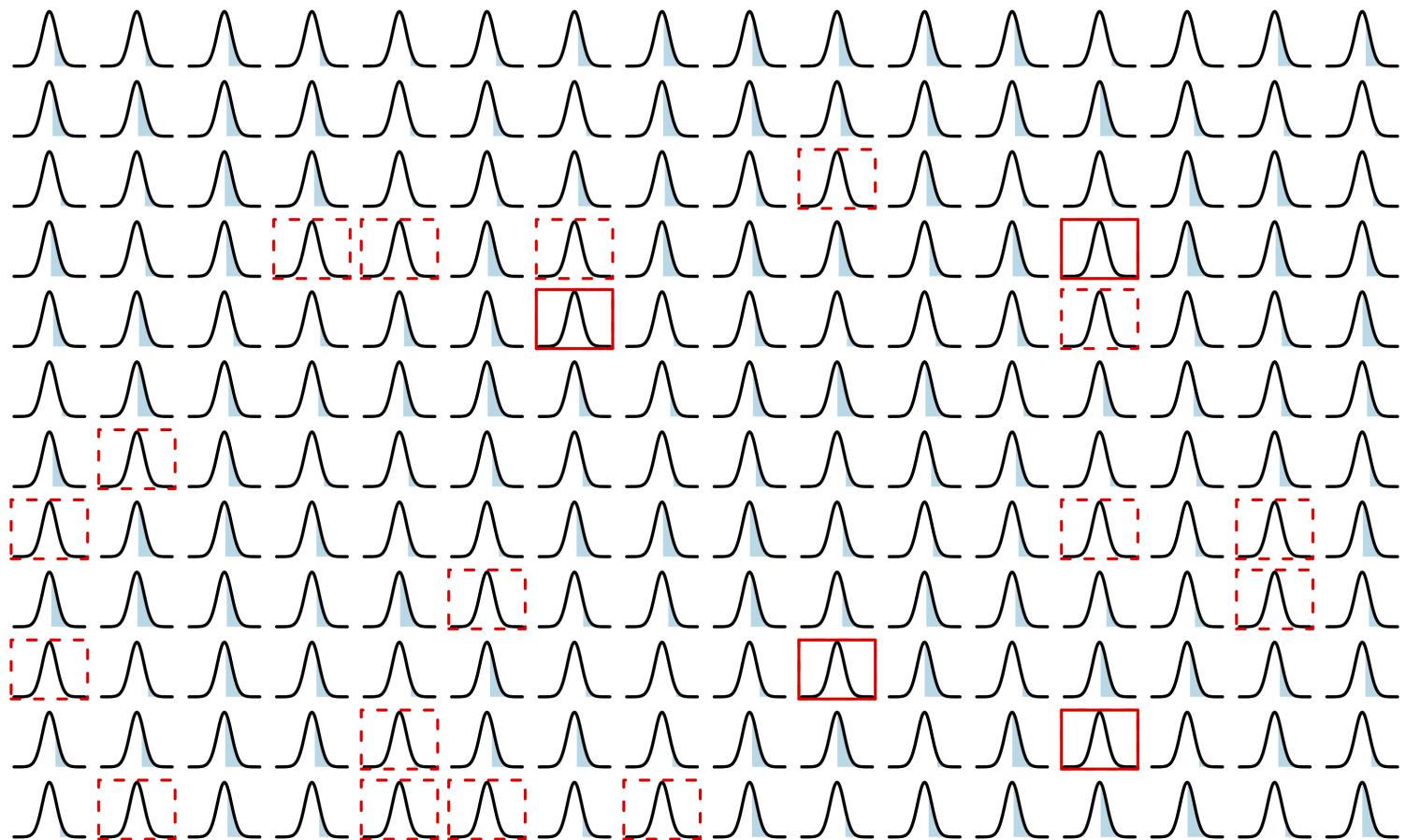


Multiple testing: risks and procedures

Solution 1: FWER thresholding (Holm-Bonferroni)



Solution 2: FDR thresholding (Benjamini-Hochberg)



Questions

What is a good definition of "significant" when performing many tests simultaneously?

definition of multiple testing risks: FDR, FWER

How can we control these multiple testing risks?

study of multiple testing procedures

- dependency assumptions
- power/conservativeness
- algorithms and their implementations

Notation

- $\mathcal{H} = \{1, \dots, m\}$ m null hypotheses to be tested
- $\mathcal{H}_0 \subset \mathcal{H}$: true null hypotheses, $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$
- $m_0 = |\mathcal{H}_0|$, $\pi_0 = m_0/m$
- $(p_i)_{1 \leq i \leq m}$: p -values
- R : a set of rejected hypotheses
- $V = |R \cap \mathcal{H}_0|$: number of "false positives" within R .

Multiple testing risks

- Family-Wise Error Rate:
 - FWER = $\mathbb{P}(V > 0)$
- False Discovery Rate:
 - FDR = $\mathbb{E} \left(\frac{V}{|R| \vee 1} \right)$

Multiple testing procedures

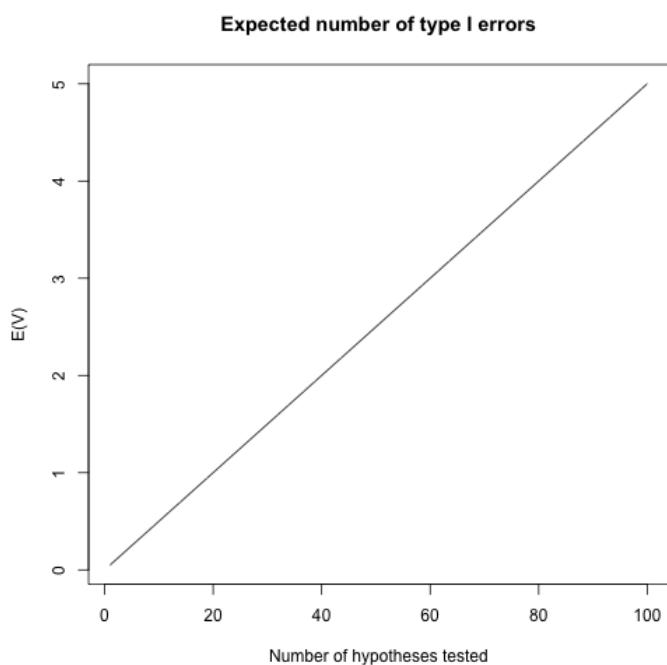
Goal: **build from the data a set R controlling FWER or FDR**

Thresholding procedures:
 $R = \{i \in \mathcal{H}, p_i \leq \hat{t}\}$

Expected number of type I errors scales linearly with m

Data: p_1, \dots, p_m : p -values for m tests

Strategy: reject \mathcal{H}_0 for all i such that $p_i \leq \alpha$



$$\text{Recall: } V = \sum_{i \in \mathcal{H}_0} \mathbf{1}_{p_i \leq \alpha}$$

$$E(V) = \sum_{i \in \mathcal{H}_0} P_{\mathcal{H}_0}(p_i \leq \alpha)$$

$$= \sum_{i \in \mathcal{H}_0} \alpha$$

$$= |\mathcal{H}_0| \alpha$$

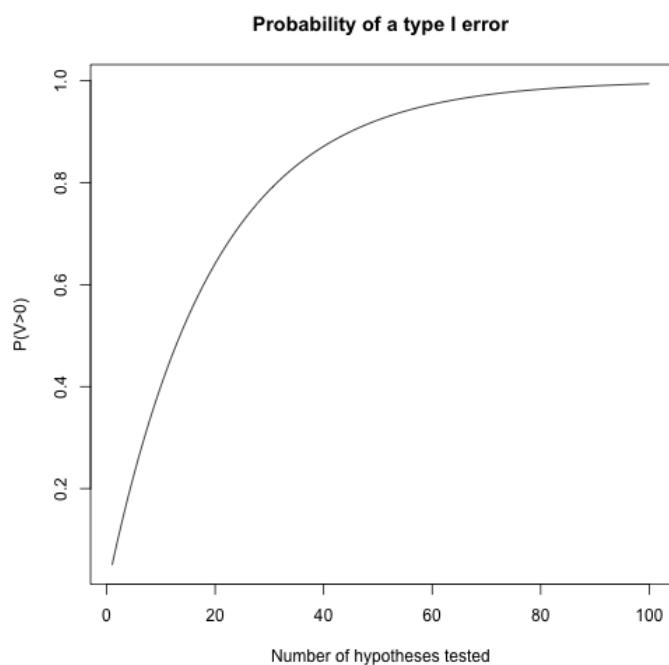
$$= \pi_0 m \alpha$$

Go to [fixed thresholding]

Probability of a type I error quickly grows to 1

Data: p_1, \dots, p_m : p -values for m tests

Strategy: reject \mathcal{H}_0 for all i such that $p_i \leq \alpha$



$$P(V = 0) = \mathbb{P}(\forall i \in \mathcal{H}_0, p_i > \alpha)$$

Assuming independent tests:

$$P(V = 0) = \prod_{i \in \mathcal{H}_0} \mathbb{P}(p_i > \alpha)$$

$$= \prod_{i \in \mathcal{H}_0} (1 - \alpha)$$

$$= (1 - \alpha)^{m_0}$$

$$\text{Hence } P(V > 0) = 1 - (1 - \alpha)^{m_0}$$

Go to: [fixed thresholding] [Sidak procedure]

Family-Wise Error Rate control

General dependence: Bonferroni and Holm procedures

The Bonferroni procedure

Definition

Reject all i such that $p_i \leq \alpha/m$

Properties

FWER control at level $\pi_0\alpha(\leq \alpha)$ under *arbitrary dependence*

[Proof]: union bound on events " i is a false positive" for $i \in \mathcal{H}_0$

Limitation

Conservativeness: α/m can be small!

Directions for increased power

- other dependency assumptions: independence, positive dependence
- estimation of π_0

Proof of FWER control by the Bonferroni procedure

Let $V(t)$ be the number of false positives obtained by rejecting all p -values less than t :

$$V(t) = \sum_{i \in \mathcal{H}_0} 1_{p_i \leq t}$$

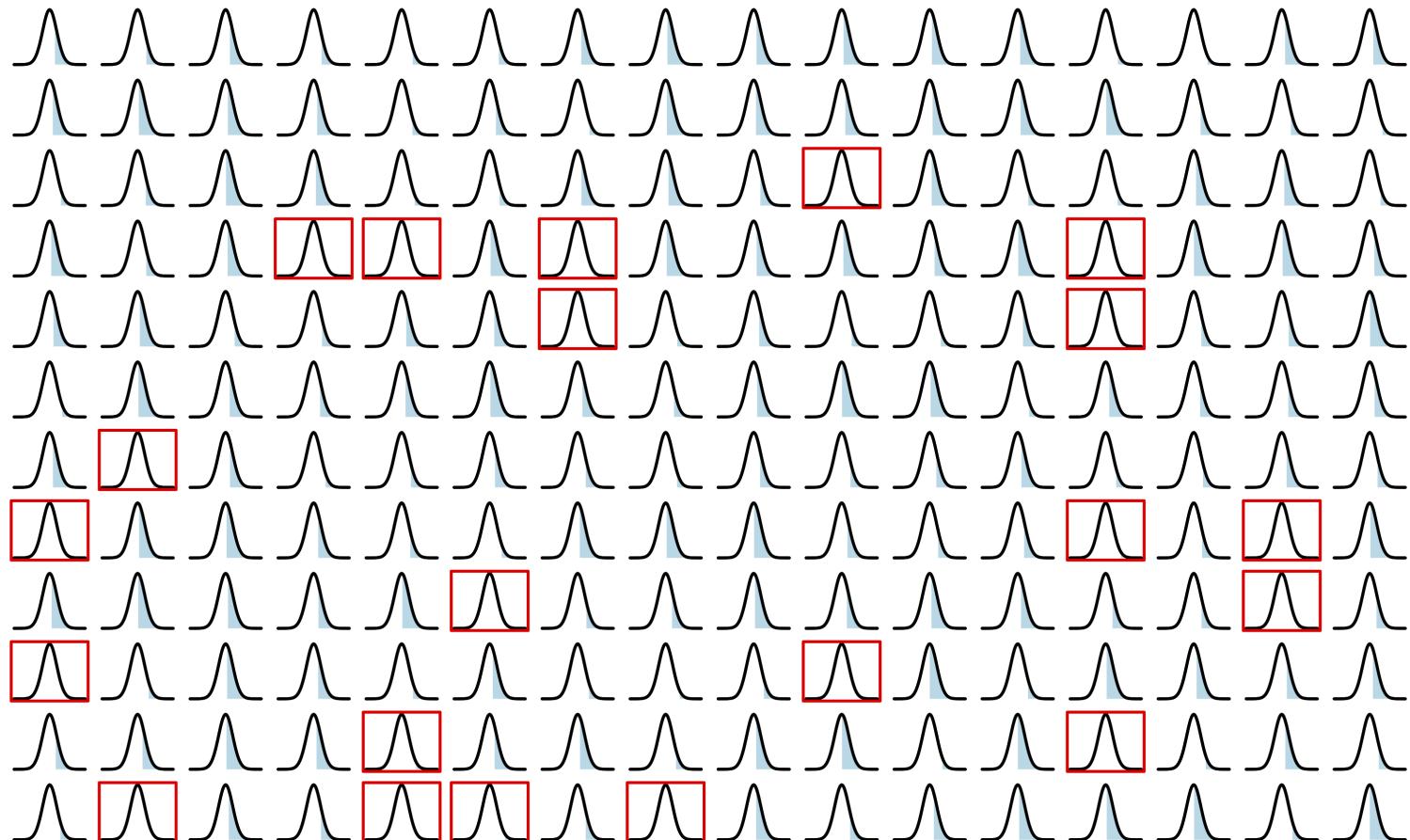
We have:

$$\begin{aligned} P(V(t) > 0) &\leq \sum_{i \in \mathcal{H}_0} \mathbb{P}_{\mathcal{H}_0}(p_i \leq t) \\ &= \sum_{i \in \mathcal{H}_0} t = m_0 t \end{aligned}$$

The Bonferroni procedure at level α rejects all p -values less than $t = \alpha/m$, therefore its FWER is $P(V(\alpha/m) > 0) \leq \pi_0 \alpha$.

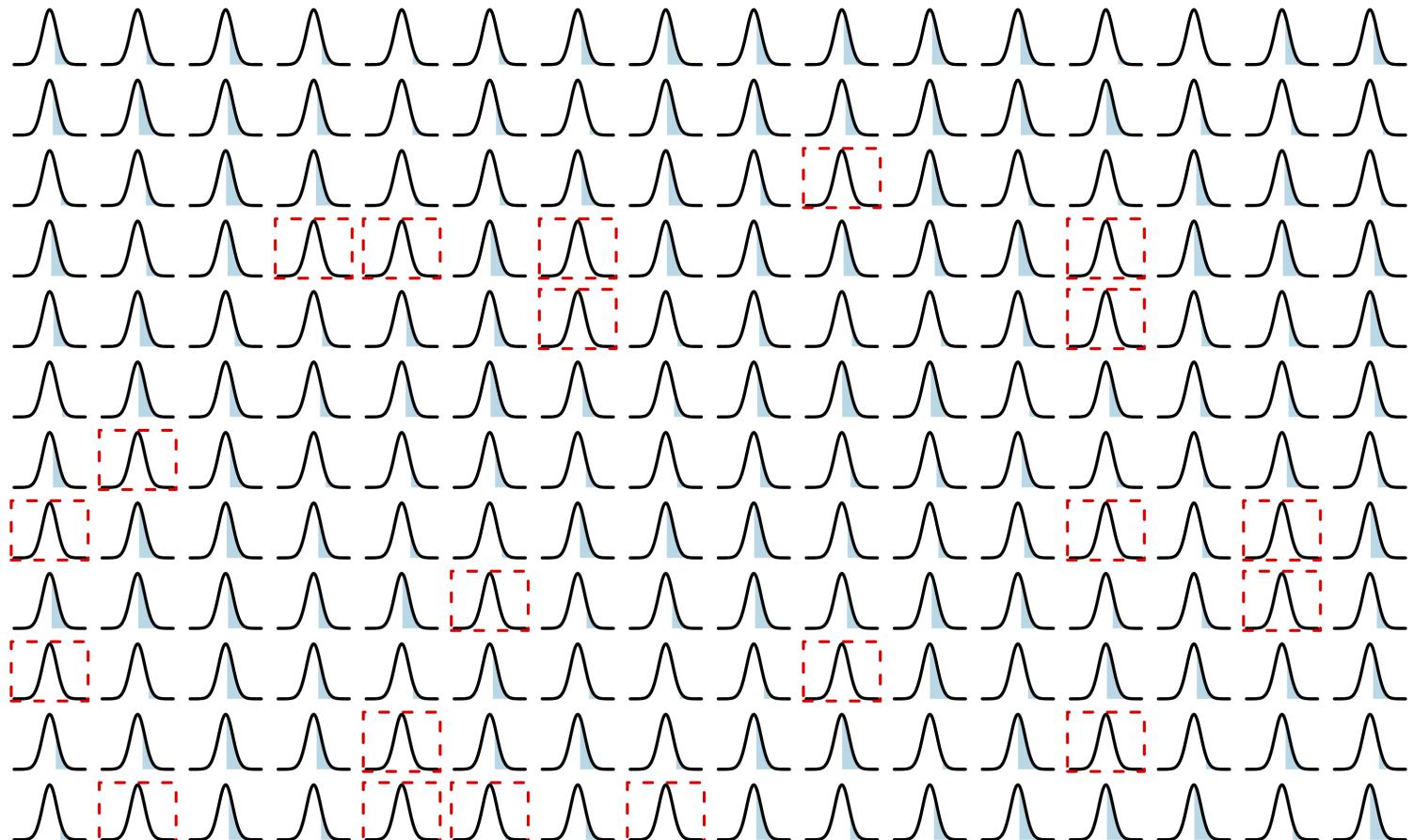
Leukemia data set: no multiple testing correction

971 genes called significant genes at (uncorrected) level $\alpha = 0.05$



Leukemia data set: FWER thresholding by Bonferroni

20 genes called significant at FWER level $\alpha = 0.05$



The Sidak procedure

Definition

Reject all i such that $p_i \leq 1 - (1 - \alpha)^{1/m}$

Property

FWER control at level $1 - (1 - \alpha)^{\pi_0} \leq \alpha$ under *independence*

(Proof: cf "Probability of a type I error")

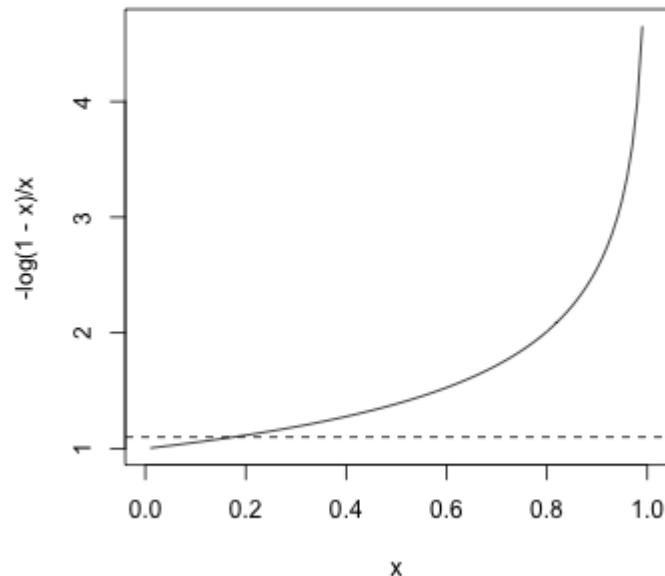
Comments

- Sidak is slightly more powerful than Bonferroni, but at the price of a much narrower applicability

In genomic applications, Bonferroni should be preferred to Sidak

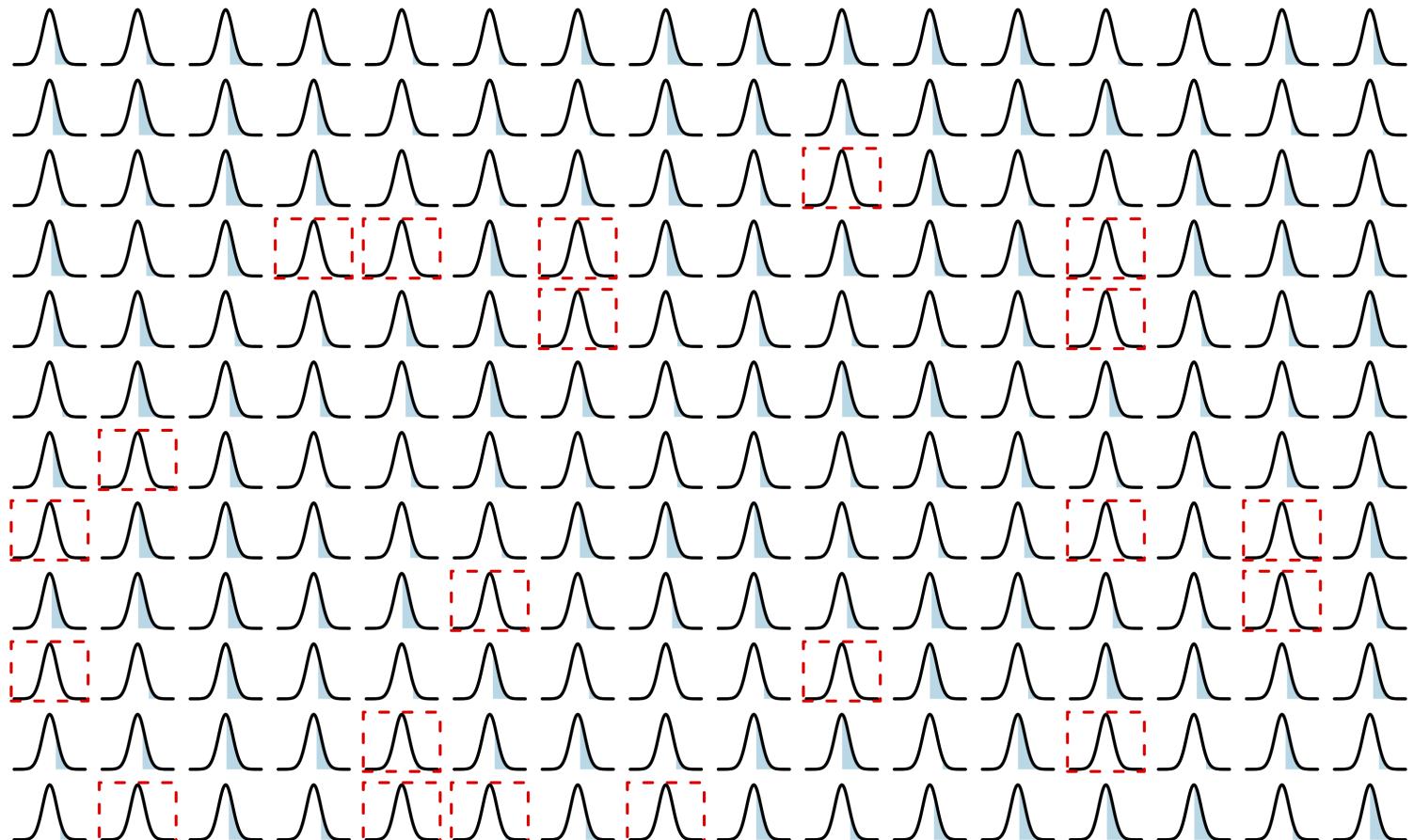
Sidak vs Bonferroni

$$\frac{\alpha}{m} \leq 1 - (1 - \alpha)^{1/m} \leq \frac{-\log(1 - \alpha)}{\alpha} \frac{\alpha}{m}$$



Leukemia data set: FWER thresholding by Sidak

20 genes called significant at FWER level $\alpha = 0.05$



The Holm procedure

Definition

Let $p_{(1)} \leq \dots \leq p_{(m)}$ be the ordered p -values.

Reject all i such that $\forall j \leq i, p_{(j)} \leq \alpha/(m - j + 1)$

Properties

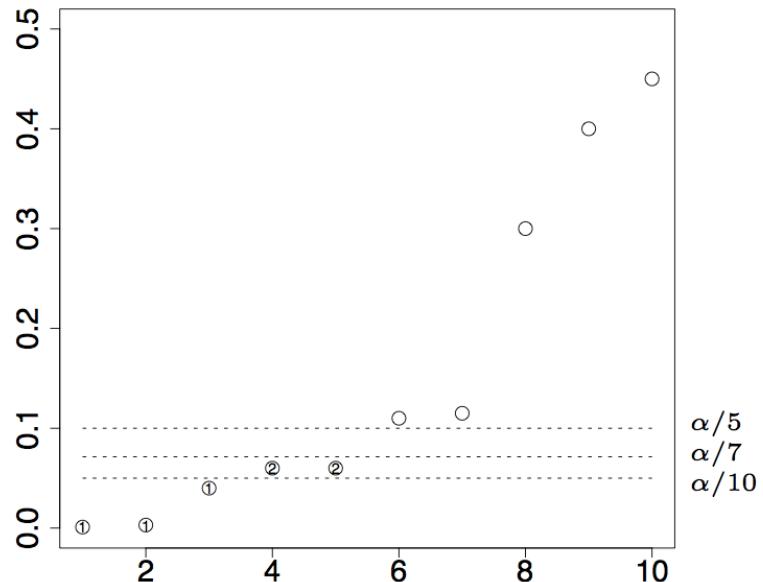
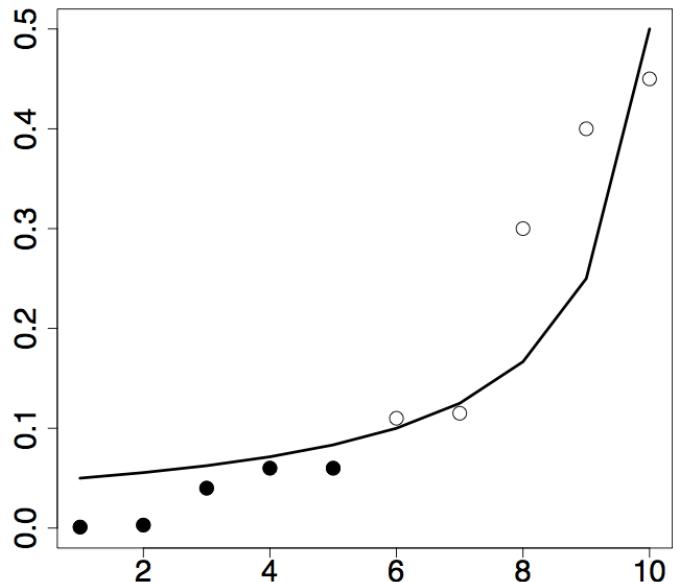
FWER control at level α under arbitrary dependence

Comments

- same guarantees as Bonferroni
- at least as powerful: $\alpha/(m - j + 1) \geq \alpha/m$!

⇒ Holm should be preferred to Bonferroni

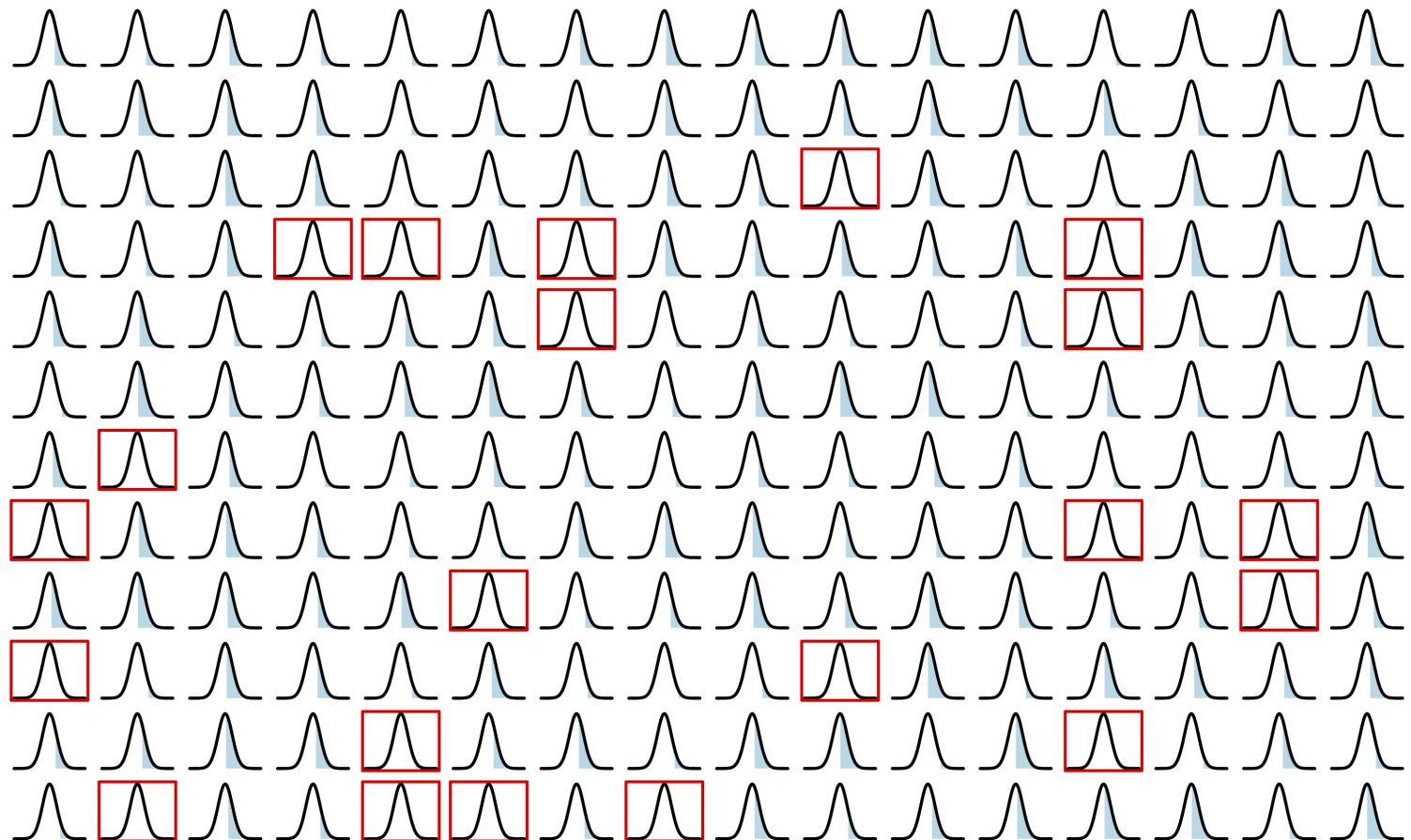
The Holm procedure: illustration and interpretation



Adaptation to π_0 !

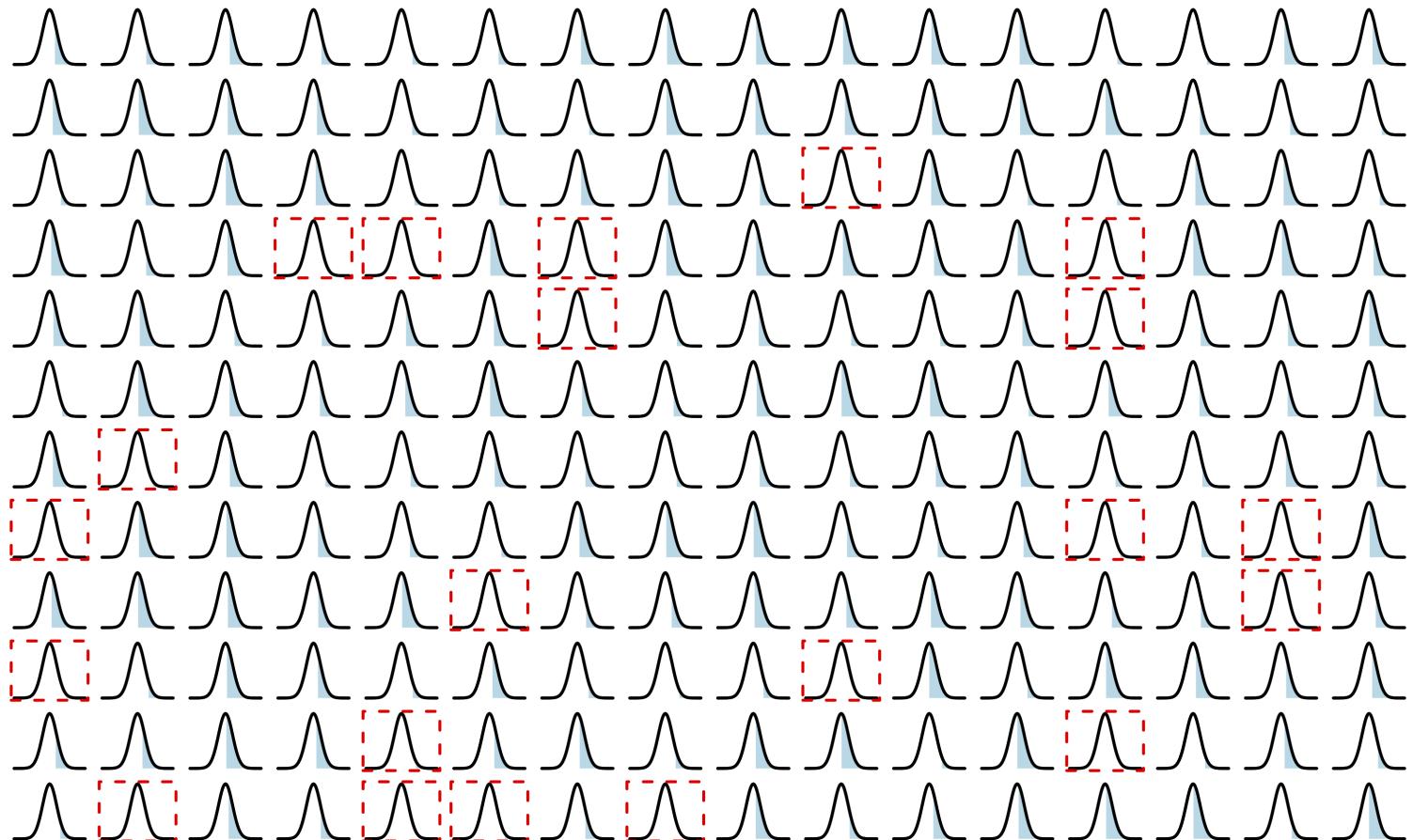
Leukemia data set: no multiple testing correction

971 genes called significant genes at (uncorrected) level $\alpha = 0.05$



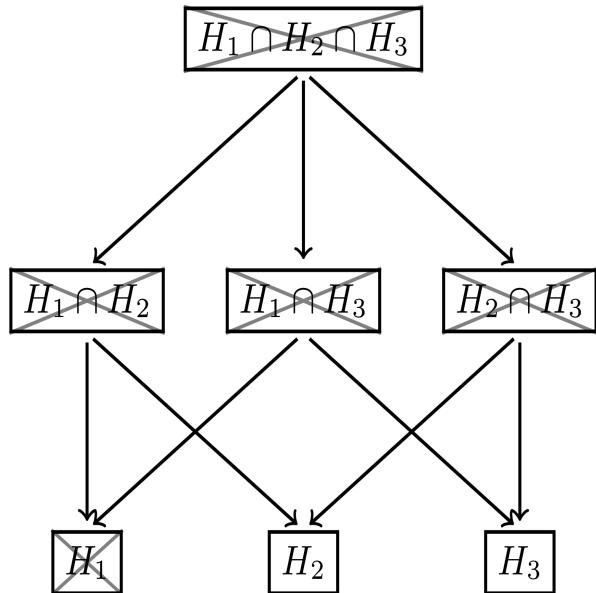
Leukemia data set: FWER thresholding by Holm

20 genes called significant at FWER level $\alpha = 0.05$



Sidetrack: closed testing procedure

Marcus, Peritz and Gabriel, *Biometrika* (1976)



For $I \subset \{1, \dots, m\}$, consider:

- $H_I = \bigcap_{i \in I} H_{0,i}$
- $(\phi_I)_I$ a local test, i.e. ϕ_I is a test of H_I ($\forall I$)

The closed testing procedure rejects H_I iff:

- $\forall J \supset I, H_J$ is rejected at level α by ϕ_J

Figure: Example with $m = 3$, with H_I rejected by ϕ_I marked by X. Closed testing rejects H_1 but not H_2 or H_3 .

FWER control by closed testing

Proposition

If ϕ_I is a α -level test H_I ($\forall I$), then the associated closed testing procedure controls the FWER at level α

Features

- simplicity and versatility
- requires 2^m tests to be performed! Need for "shortcuts"

Closed testing with Bonferroni local tests

The simplest α -level local test of H_I :

$$\phi_I = \mathbf{1} \{ \exists i \in I, p_i \leq \alpha / |I| \}$$

The corresponding closed testing procedure is the *Holm procedure*

FWER control under positive dependence

Simes inequality

R. J. Simes. *Biometrika* 73.3 (1986), pp. 751–754

If the p -values $(p_i)_{1 \leq i \leq m}$, are PRDS (\mathcal{H}_0), then

$$\mathbb{P}(\exists k \in \{1, \dots, m_0\} : p_{(k:\mathcal{H}_0)} \leq \alpha k / m_0) \leq \alpha,$$

where $p_{(1:I)} \leq \dots \leq p_{(1:I)}$ denote the ordered p -values $(p_i)_{i \in I}$

Positive Regression Dependency on a Subset (PRDS)

Non-decreasing set

$S \subset [0, 1]^m$ is non-decreasing iif : for all $(q, q') \in [0, 1]^m$ such that $\forall i \in \{1, \dots, m\}, q_i \leq q'_i, q \in S$ implies $q' \in S$.

PRDS(I) assumption

The p -value family $\mathbf{p} = (p_i)_{1 \leq i \leq m}$ is PRDS on a subset $I \subset \{1, \dots, m\}$ if:

For any measurable non-decreasing set $S \subset [0, 1]^m$, the function $u \mapsto \mathbb{P}(\mathbf{p} \in S | p_i = u)$ is non-decreasing for all $i \in I$.

PRDS is a *technical assumption* that is both:

- widely accepted in genomics
- difficult/impossible to prove in practice

Closed testing with Simes local tests

Simes test

$$\phi_I = \mathbf{1}\{\exists i \in I, p_{(i:I)} \leq \alpha i / |I|\}$$

Simes test is a α -level local test for H_I under PRDS (\mathcal{H}_0)

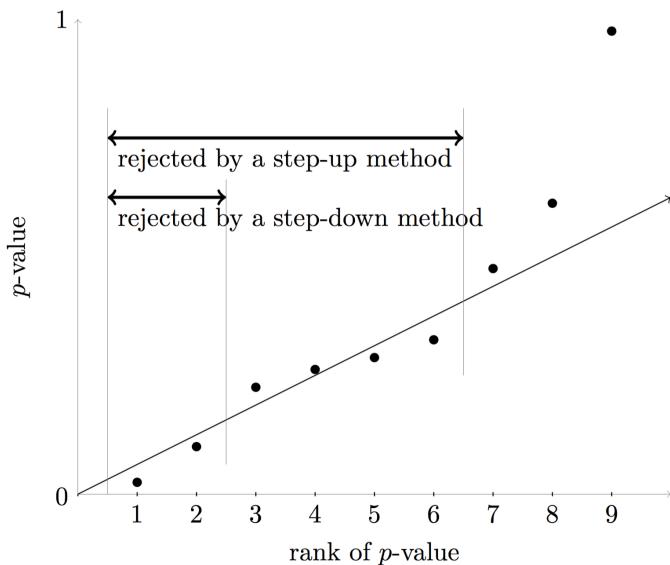
Hochberg/Hommel procedures

- The corresponding closed testing procedures are the Hochberg/Hommel procedures
- This proves that Hochberg/Hommel control FWER under PRDS (\mathcal{H}_0)

Holm vs Hochberg

Holm and Hochberg have the same *critical values*: $\alpha/(m - i + 1)$

- Holm (step-down): Reject all i such that $\forall j \leq i, p_{(j)} \leq \alpha/(m - j + 1)$
- Hochberg (step-up): Reject all i such that $p_{(i)} \leq \alpha/(m - i + 1)$



- Hochberg is slightly more powerful than Holm ...
- ... but at the price of a narrower applicability

NB: Hommel \gg Hochberg

Using the p.adjust function

Holm and Hochberg

```
adjp <- p.adjust(p_values, method = "holm")
sum(adjp <= 0.05)
```

```
## [1] 19
```

```
adjp <- p.adjust(p_values, method = "hochberg")
sum(adjp <= 0.05)
```

```
## [1] 19
```

Hommel (not run [1])

```
adjp <- p.adjust(p_values, method = "hommel")
sum(adjp <= 0.05)
```

[1] This code is quadratic in m . Exact linear shortcut available in Meijer, RJ, Krebs, TJP, Goeman, JJ. **Hommel's procedure in linear time.** *Biometrical Journal*. 2019; 61: 73– 82.

FWER control: a summary

Main procedures for each type of dependency assumption

- (Independence: Sidak)
- Arbitrary dependence: Bonferroni \ll **Holm**
- PRDS (\mathcal{H}_0): Hochberg \ll **Hommel**

Remarks

- Implementation: R function `p.adjust`
- The procedures presented until now are adjusted to a "worst case" dependence structure. Can we do better?

Adaptation to dependence

Westfall and Young: MinP and MaxT methods

Starting point: reformulation of FWER

$$\text{FWER}(t) = \mathbb{P}(\exists i \in \mathcal{H}_0, p_i \leq t) = \mathbb{P}\left(\min_{i \in \mathcal{H}_0} p_i \leq t\right)$$

Choosing t as the α - quantile of $\min_{i \in \mathcal{H}_0} p_i$ ensures $\text{FWER}(t) \leq \alpha$!

Permutation-based estimation of the law of the minimal null p -value

The distribution of this statistic is unknown, but can often be estimated using permutation-based methods, e.g. in:

- simple two-group tests in differential expression analyses
- case-control GWAS studies

Westfall and Young: MinP and MaxT methods

Westfall, P. H., & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment (Vol. 279). John Wiley & Sons.

Implementation using the Bioconductor/R package `multtest`

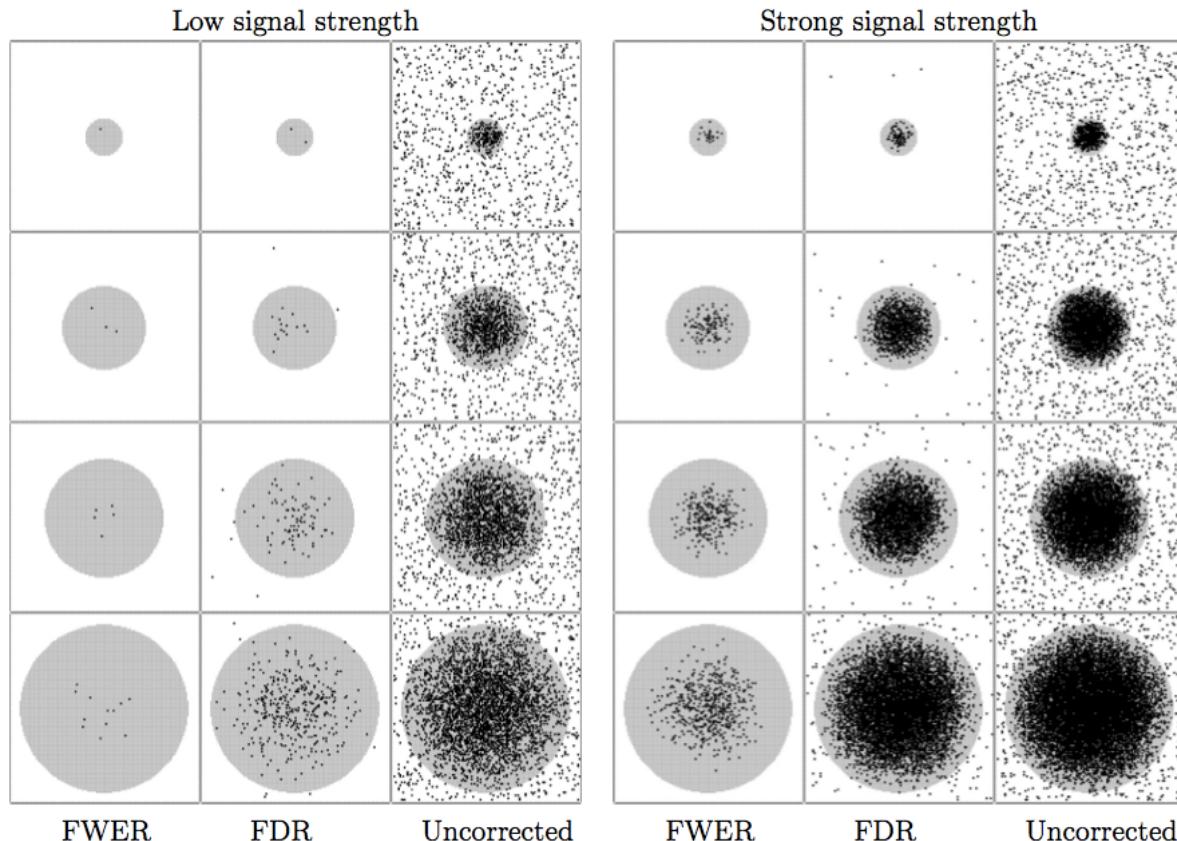
```
library("multtest")
resT <- mt.maxT(dat, categ, B = 1000)
sum(resT$adjp < alpha)
## [1] 31
```

Substantially more discoveries than other FWER controlling methods!

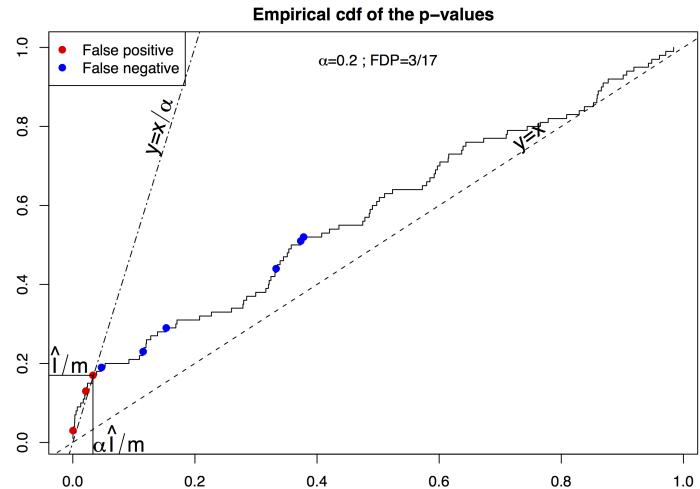
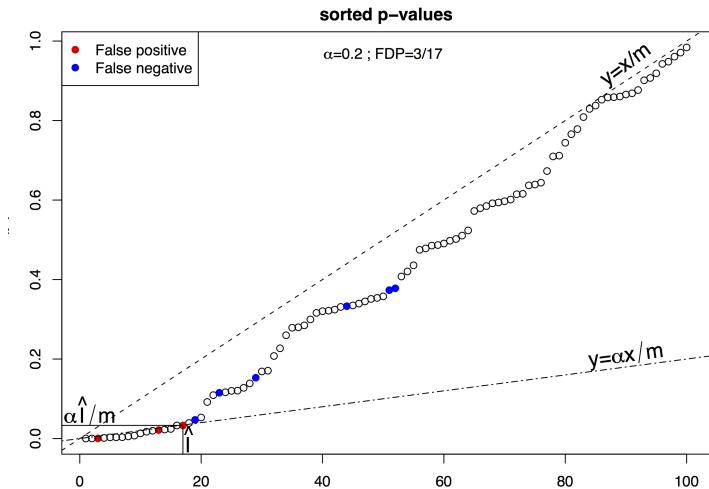
False Discovery Rate control

The Benjamini and Hochberg procedure

Adaptation to unknown sparsity



The Benjamini and Hochberg (1995) procedure



Algorithm

- sort p -values: $p_{(1)} \leq \dots \leq p_{(m)}$
- define $\hat{I} = \max \left\{ k \mid p_{(k)} \leq \alpha \frac{k}{m} \right\}$
- reject all i such that $p_i \leq p_{(\hat{I})} (= \alpha \hat{I} / m)$

The BH procedure

- $\pi_0 = |\mathcal{H}_0|/m$: proportion of true null hypotheses
- $\text{FDR} = \mathbb{E} \left(\frac{V}{|R| \vee 1} \right)$: expected proportion of false positive among rejections

Properties

Benjamini and Hochberg (1995)

BH (α) provides FDR control at level $\pi_0\alpha$ if the p -values under \mathcal{H}_0 are either

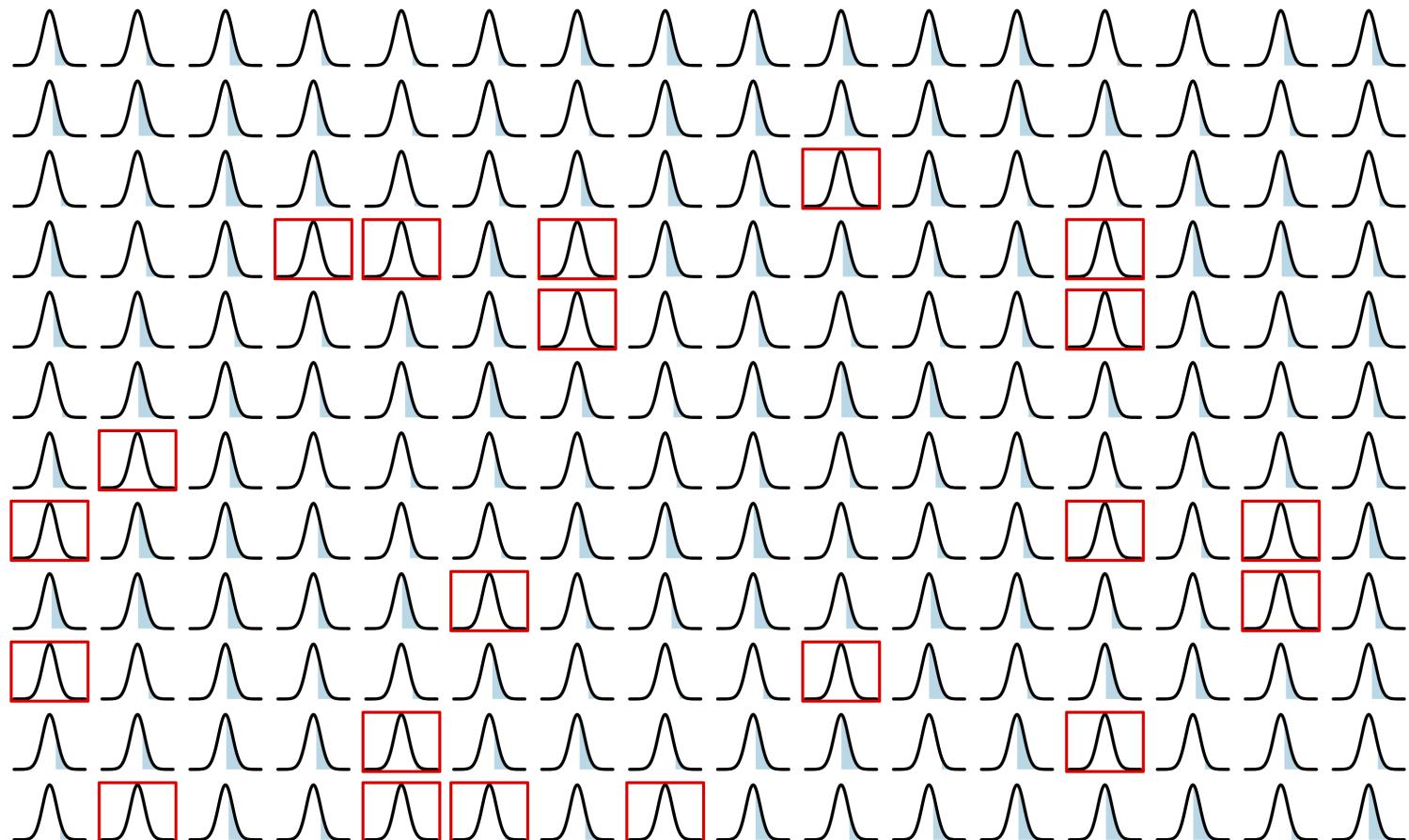
- independent
- positively associated (PRDS (\mathcal{H}_0))

Improvements in the statistical literature

- general dependence: Benjamini and Yekutieli (2001)
- estimation of π_0 , in the hope of a sharper FDR control

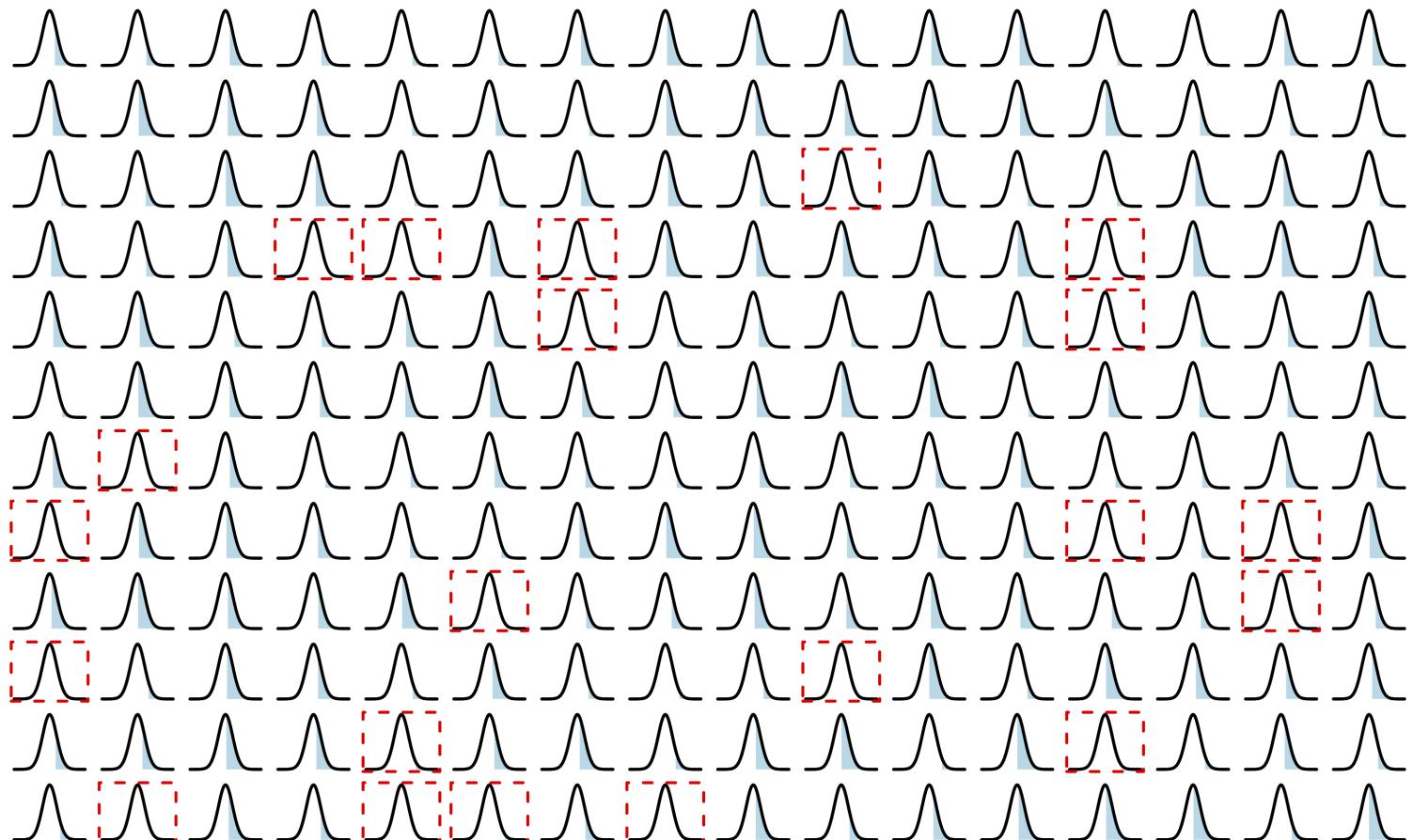
Leukemia data set: no multiple testing correction

971 genes called significant genes at (uncorrected) level $\alpha = 0.05$



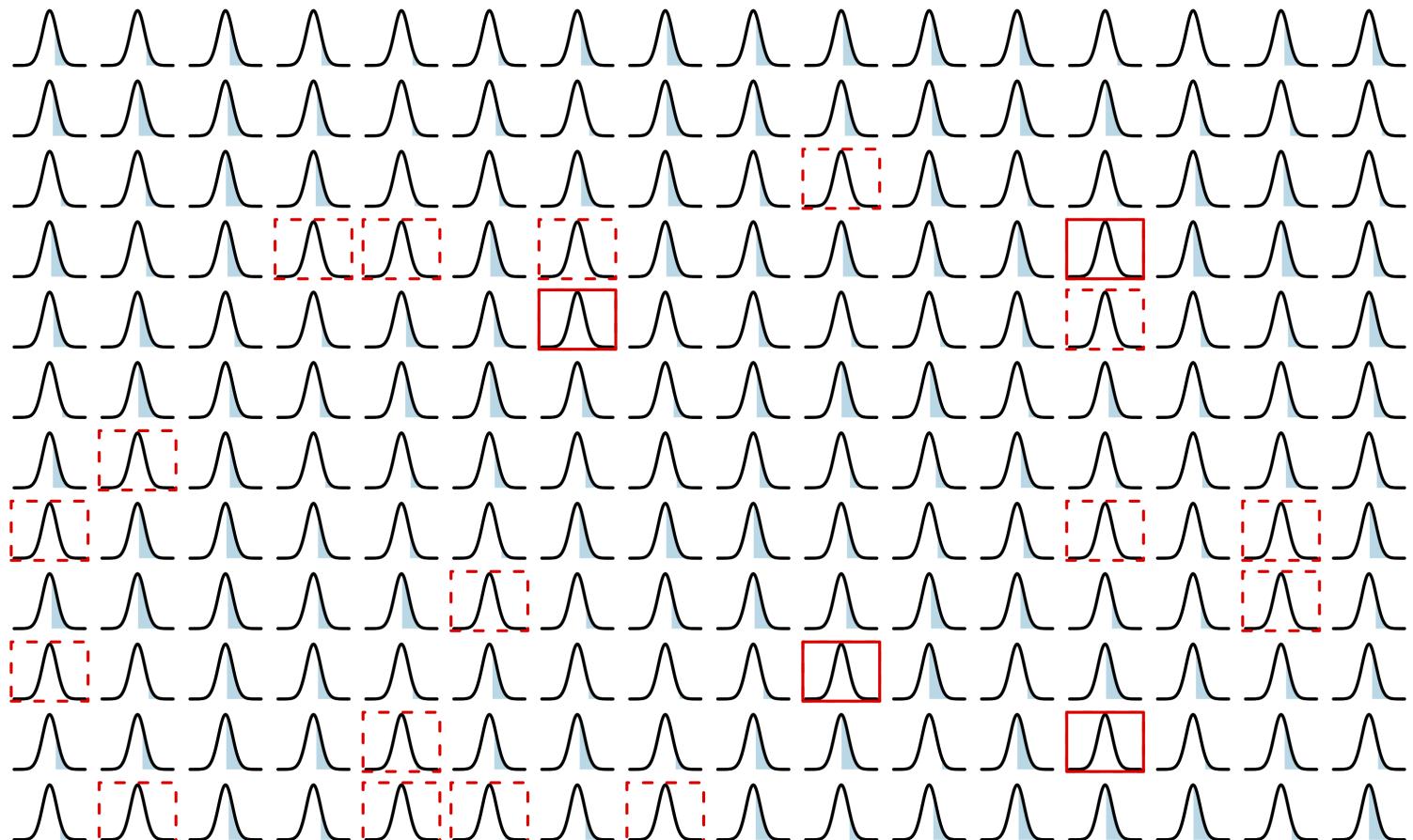
Leukemia data set: FWER thresholding by Holm

20 genes called significant at FWER level $\alpha = 0.05$



Leukemia data set: FDR thresholding by BH

163 genes called significant at FDR level $\alpha = 0.05$



Using the p.adjust function

```
adjp <- p.adjust(p_values, method="BH")  
sum(adjp <= alpha)
```

```
## [1] 150
```

The BH procedure is widely used

Controlling the false discovery rate: A practical and powerful approach to multiple testing Y. Benjamini, Y. Hochberg. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol 57(1), pp. 289–300. 1995.

Citations of the BH 1995 paper

- 6,000 publications in the PubMed database with "False Discovery Rate" in their title or abstract
- 60,000 citations according to scholar.google.com.

Comparison to other highly cited statistics papers

- Benjamini, Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing: 60,000
- Kaplan, Meier. Nonparametric estimation from incomplete observations: 57,000
- Dempster, Laird, Rubin. Maximum likelihood from incomplete data via the EM algorithm (1977): 56,000
- Cox. Regression and life tables (1975): 50,000
- Bland, Altman. Statistical methods for assessing agreement between two methods of clinical measurement: 43,000
- Tibshirani. Regression shrinkage and selection via the lasso (1996): 30,000

FDR itself is prone to false positives

Google FDR citations

All Images News Videos Shopping More Settings Tools

britannica roosevelt medal quotes franklin roosevelt franklin d d roosevelt inspirational

5 Facts about Franklin D. Roosevelt

1. HE HAS A NICKNAME WHICH IS FDR.
2. HE IS THE 32ND PRESIDENT OF THE UNITED STATES.
3. HE IS THE ONLY PRESIDENT TO BE ELECTED FOR FOUR TERMS.
4. HE HAD POLIO.
5. HE HAD THREE VICE PRESIDENTS.

Il est dur d'échouer ; mais il est pire de n'avoir jamais tenté de réussir. — Franklin Roosevelt Pourquoi-entreprendre.fr

THE TEST OF OUR PROGRESS IS NOT WHETHER WE ADD MORE TO THE ABUNDANCE OF THOSE WHO HAVE MUCH; IT IS WHETHER WE PROVIDE ENOUGH FOR THOSE WHO HAVE TOO LITTLE. — Franklin D. Roosevelt The Knowledge Movement Facebook

The liberty of a democracy is not safe if the people tolerate the growth of private power to a point where it comes stronger than their democratic state itself. That, in its essence, is fascism - ownership of government by an individual, by a group. — FDR Another Day, More Republican Dirty Tricks.com

The only thing we have to fear is fear itself. — Franklin D. Roosevelt

Human kindness has never weakened the stamina or softened the fiber of a free people. A nation does not have to be cruel to be tough. — Franklin Delano Roosevelt

No one can make you feel inferior without your consent. — Eleanor Roosevelt

We are a nation of many nationalities, many races, many religions-bound together by a single unity, the unity of freedom and equality. Whoever seeks to set one nationality against another, seeks to degrade all nationalities. — Franklin D. Roosevelt — QUOTES

THE Obam Acc so far Citat

"The only thing we have to fear is fear itself." — Franklin Delano Roosevelt

"WITH A GOOD CONSCIENCE OUR ONLY SURE REWARD, WITH HISTORY THE FINAL JUDGE OF OUR DEEDS; LET US GO FORTH TO LEAD THE LAND WE LOVE, ASKING HIS BLESSING AND HIS HELP, BUT KNOWING THAT HERE ON EARTH GOD'S WORK MUST TRULY BE OUR OWN" — JFK INAUGURAL ADDRESS 1961

"MEN ARE NOT PRISONERS OF FATE BUT ONLY PRISONERS OF THEIR OWN MINDS" — FRANKLIN D. ROOSEVELT

"In the truest sense, freedom cannot be bestowed; it must be achieved." — Franklin D. Roosevelt

"Human kindness has never weakened the stamina or softened the fiber of a free people. A nation does not have to be cruel to be tough." — FDR

FDR control under general dependence?

Hommel's inequality

G Hommel. Biometrische Zeitschrift 25.5 (1983), pp. 423–430

Recall Simes' inequality

If the p -values $(p_i)_{1 \leq i \leq m}$, are PRDS on \mathcal{H}_0 , then

$$\mathbb{P}(\exists k \in \{1, \dots, m_0\} : p_{(k:\mathcal{H}_0)} \leq \alpha k/m_0) \leq \alpha$$

Hommel's inequality

Under arbitrary dependence,

$$\mathbb{P}(\exists k \in \{1, \dots, m_0\} : p_{(k:\mathcal{H}_0)} \leq \alpha k/m_0) \leq \alpha K_{m_0},$$

where $K_m = \sum_{j=1}^m j^{-1} \sim \log(m)$ is Hommel's correction factor for dependency

The Benjamini and Yekutieli procedure

Benjamini and Yekutieli (2001)

Definition

$$\text{BY}(\alpha) := \text{BH}\left(\alpha/K_m\right)$$

Properties

- controls FDR at level $\pi_0\alpha$ under arbitrary dependence
- generally *very* conservative since $K_m \sim \log(m)$. Sometimes even more conservative than Holm!

```
adjp <- p.adjust(p_values, method="BY")
sum(adjp <= alpha)
```

```
## [1] 27
```

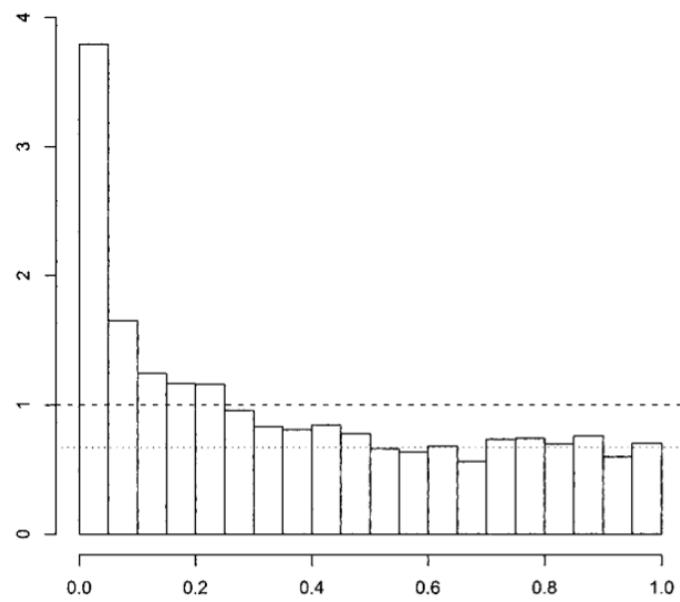
⇒ rarely used in practice due to its conservativeness

Adaptation to π_0 ,
proportion of true nulls

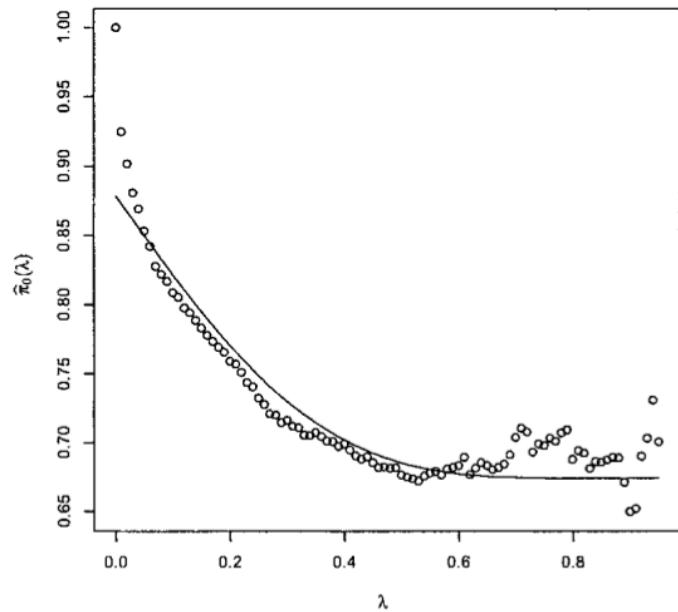
Estimation of π_0

"Storey's method" (Schweder and Spjotvoll, 1982)

Estimation



Bias/variance tradeoff



Illustrations from Storey & Tibshirani, PNAS 2003

Limitations of FDR control

Limitation 1: FDR control is not FDP control

- FDR is the *expected* proportion of false positives among rejections:

$$FDR = E(FDP)$$

- $FDR \leq \alpha$ does not imply that $FDP \leq \alpha$!

The concentration of FDP around FDR depends on the unknown dependency structure between hypotheses

Limitation 1: FDR control is not FDP control

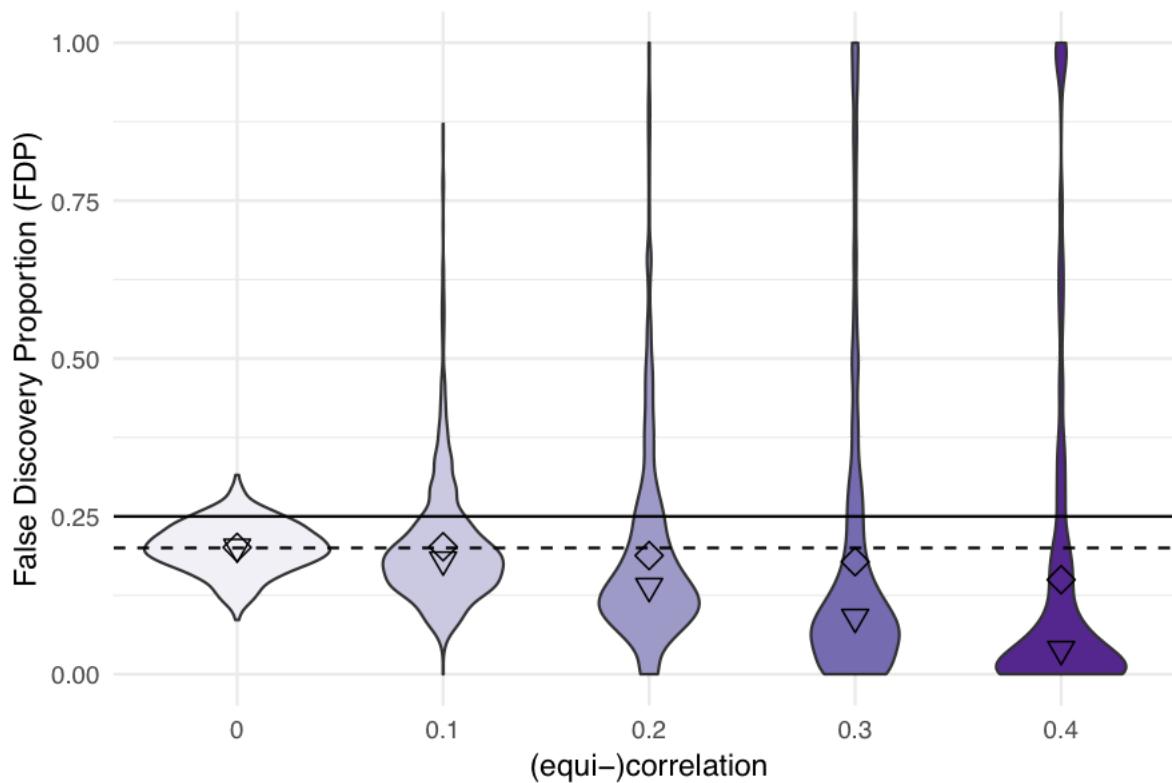
Simulation setting: Gaussian equi-correlation

- $m = 1000, m_0 = 800$
- Gaussian test statistics:
 - $X_i \sim \mathcal{N}(0, 1)$ under \mathcal{H}_0
 - $X_i \sim \mathcal{N}(2, 1)$ under \mathcal{H}_1
- Equi-correlation: $\text{cor}(X_i, X_j) = \rho \in [0, 1]$ for $i \neq j$

Remark

- simple simulation setting with one parameter for dependence
- the PRDS assumption is satisfied in this setting

Limitation 1: FDR control is not FDP control



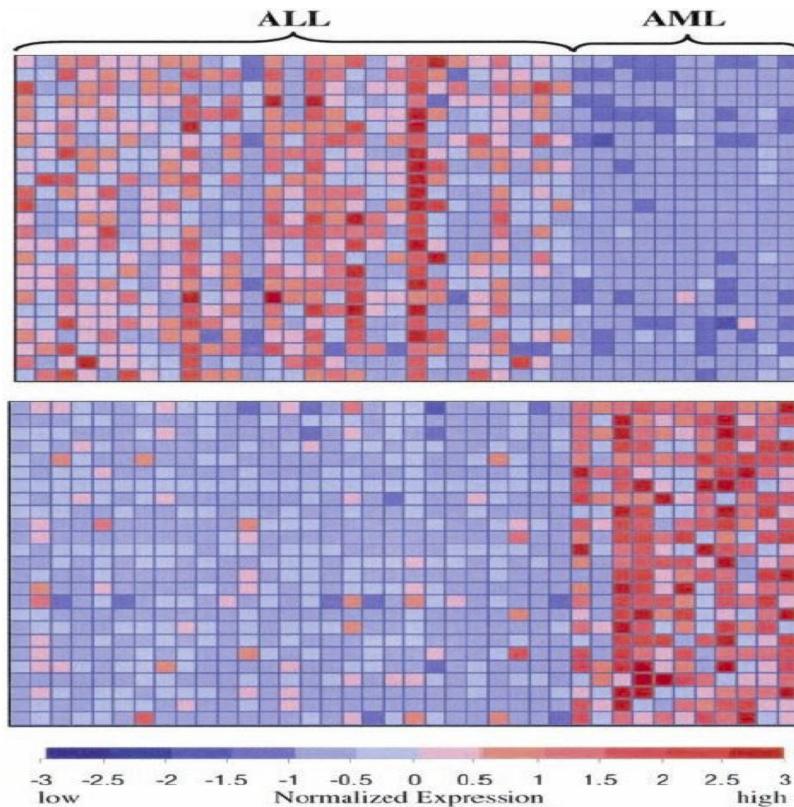
Limitation 2: FDR is not robust to selection

In practice, researchers tend to use FDR for exploratory purposes, by:

1. defining a list of candidates using FDR control
2. refine this list based on *side information*, e.g.:
 - gene pathways for differential expression analyses
 - linkage disequilibrium for genome-wide association studies (GWAS)
 - brain regions in functional magnetic neuro-imaging (fMRI) studies

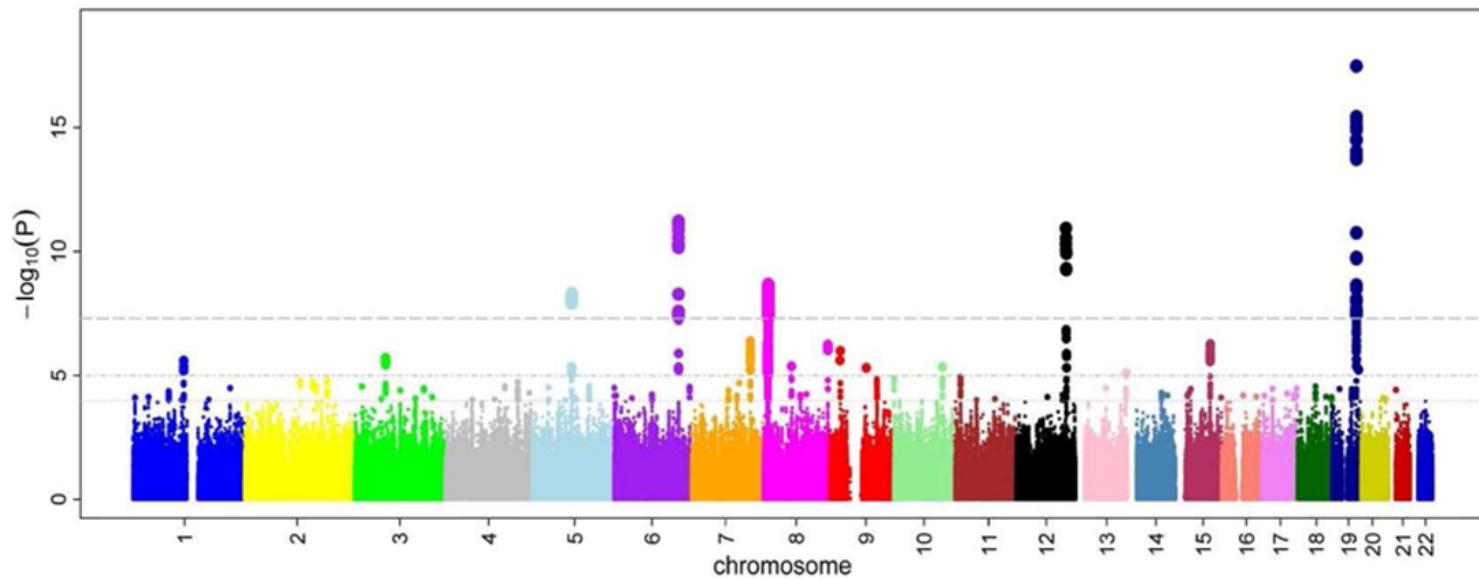
FDR control is lost in such refined lists!

Example: differential expression analyses



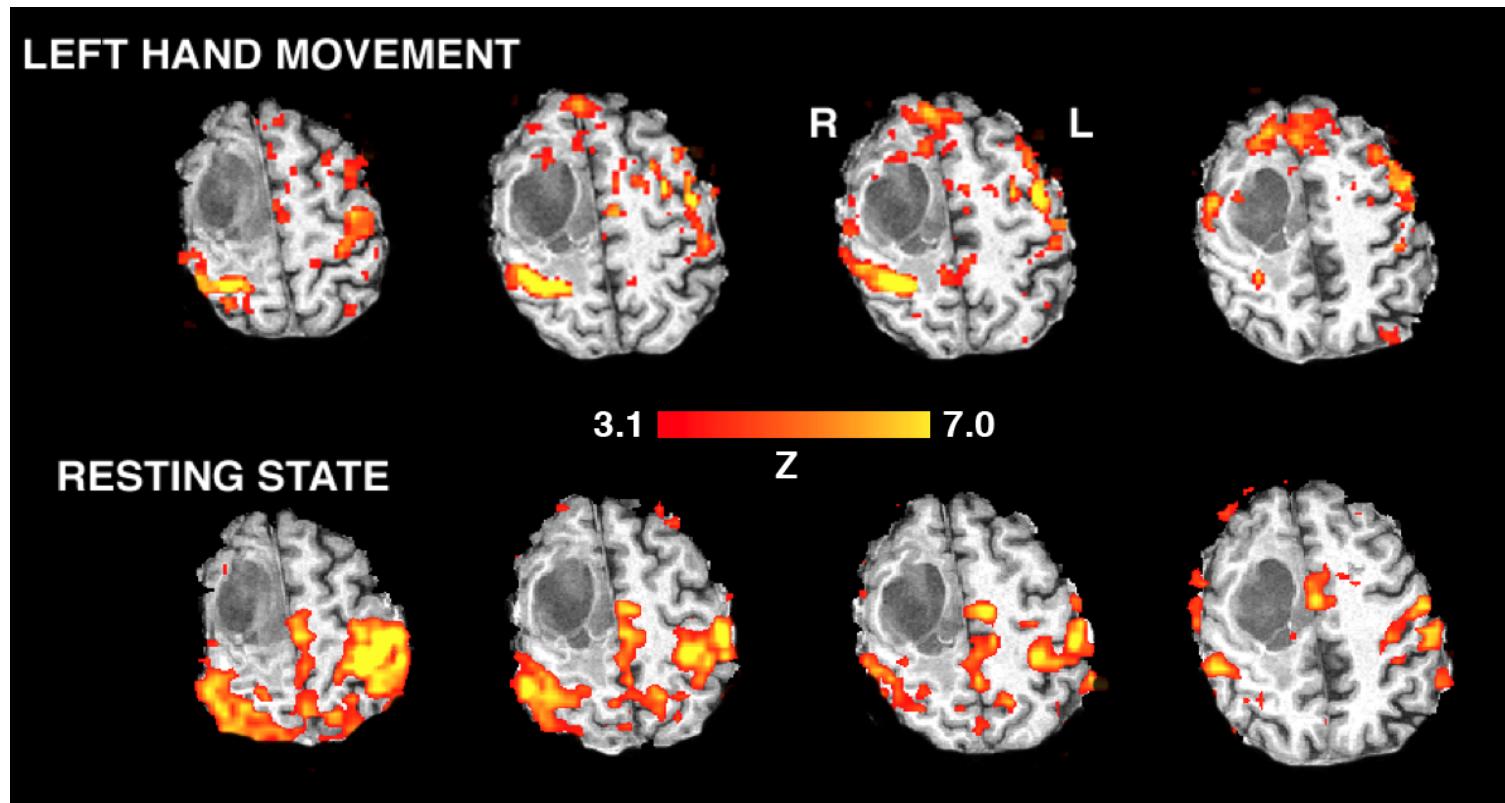
Side information: gene pathways

Example: genome-Wide Association Studies



Side information: genome regions (linkage disequilibrium between markers)

Example: functional magnetic neuro-imaging



Side information: brain regions

Post hoc inference

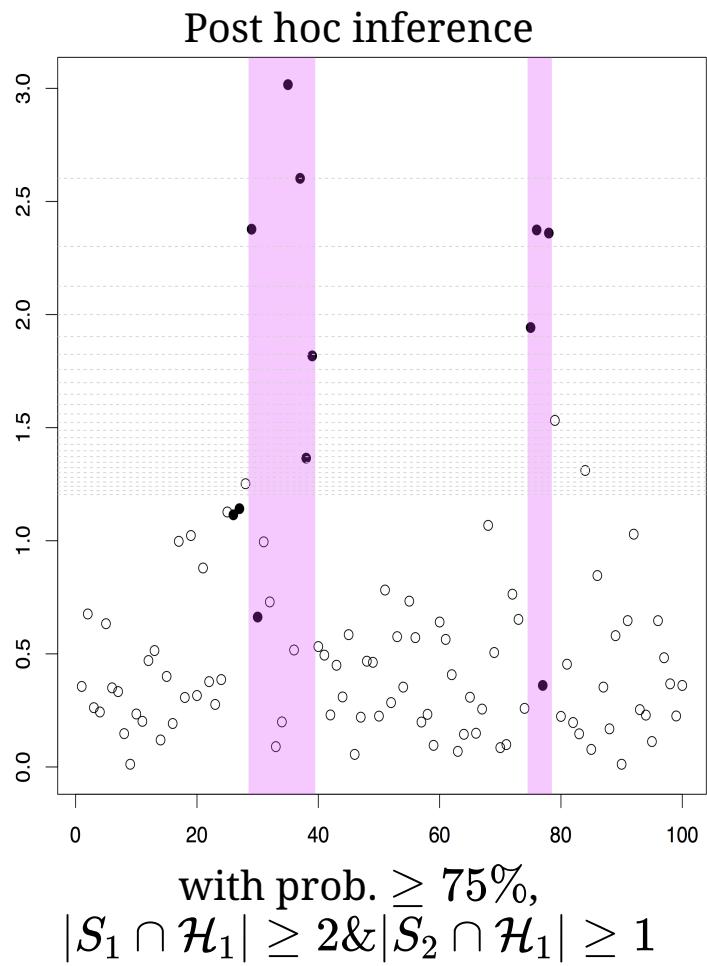
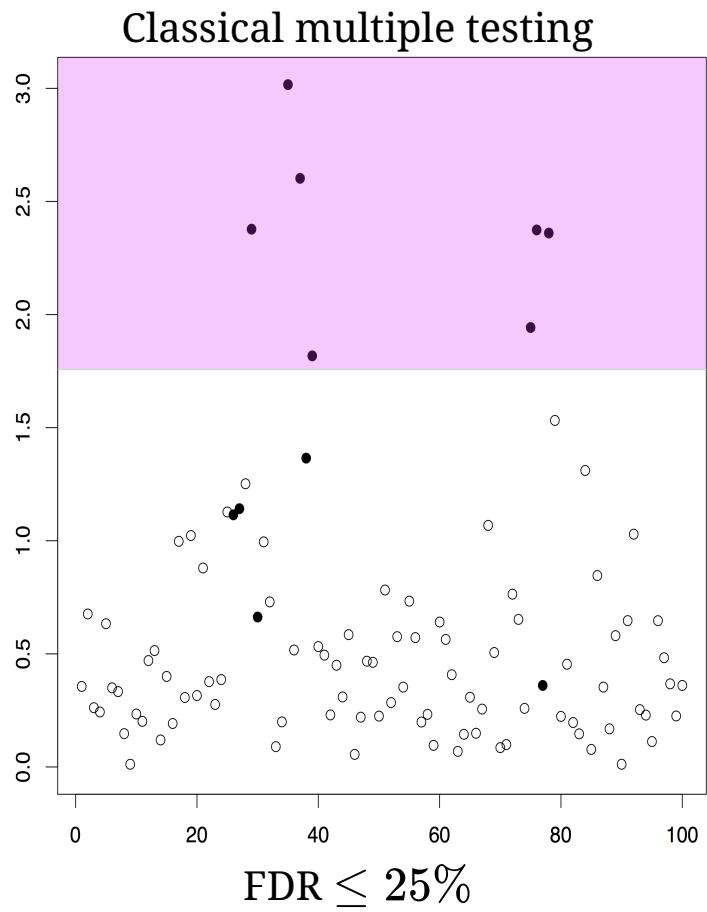
Post hoc inference

Goal: find a *post hoc upper bound* on $V(S) = |S \cap \mathcal{H}_0|$, that is:

find \bar{V}_α such that with probability larger than $1 - \alpha$,

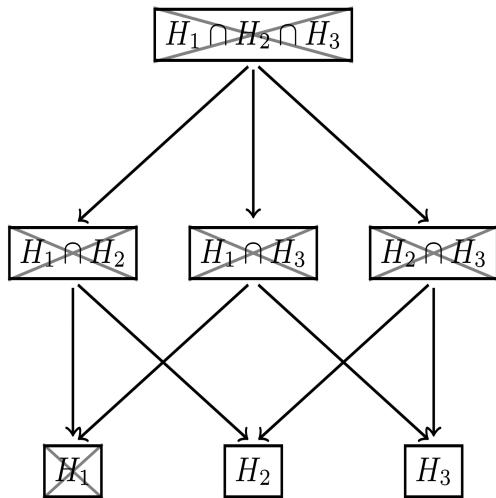
$$\forall S \subset 1 \dots m, \quad V(S) \leq \bar{V}_\alpha(S)$$

Post-hoc inference: illustration



Post hoc inference via closed testing

Goeman and Solari. *Stat. Science* (2011)



Idea: exploit *non-consonant rejections* to provide an upper bound on $V(S) = |S \cap \mathcal{H}_0|$ for any rejection set S

$$\bar{V}_\alpha(S) =$$

$$\max\{|I| : I \subset S, H_I \text{ not rejected}\}$$

Simes shortcut: post hoc bound under PRDS assumption

- $\bar{V}_\alpha(S)$ cannot be calculated for $m \geq 20$ for a generic local test
- For Simes local test (under PRDS): efficient algorithm in $O(m \log m)$

Post hoc bounds using reference families

Blanchard, Neuvial and Roquain. *Annals of Statistics* (2020) R package
sansSouci

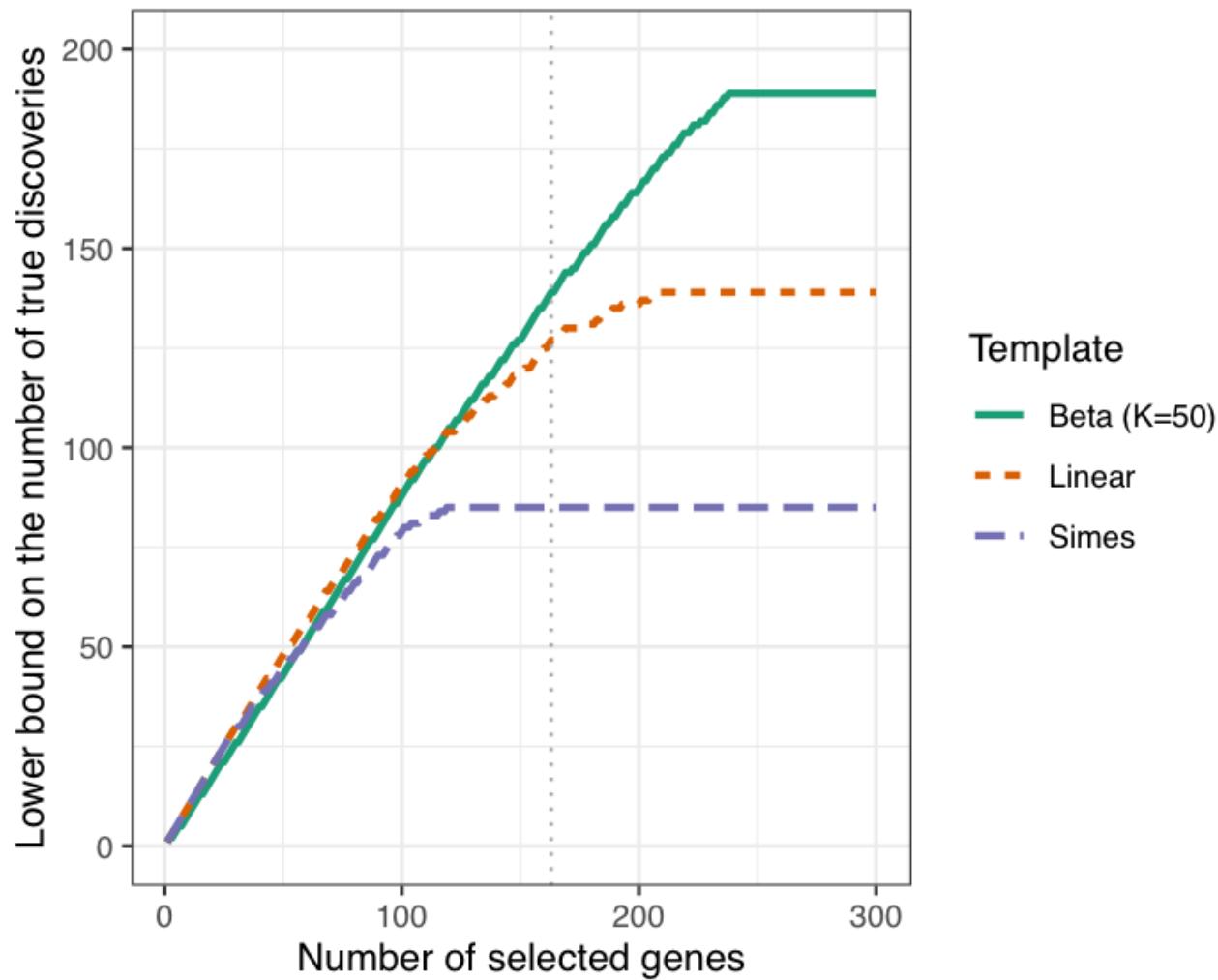
Alternative construction of post hoc bounds

- Recovers the bounds of GS2011 under PRDS:

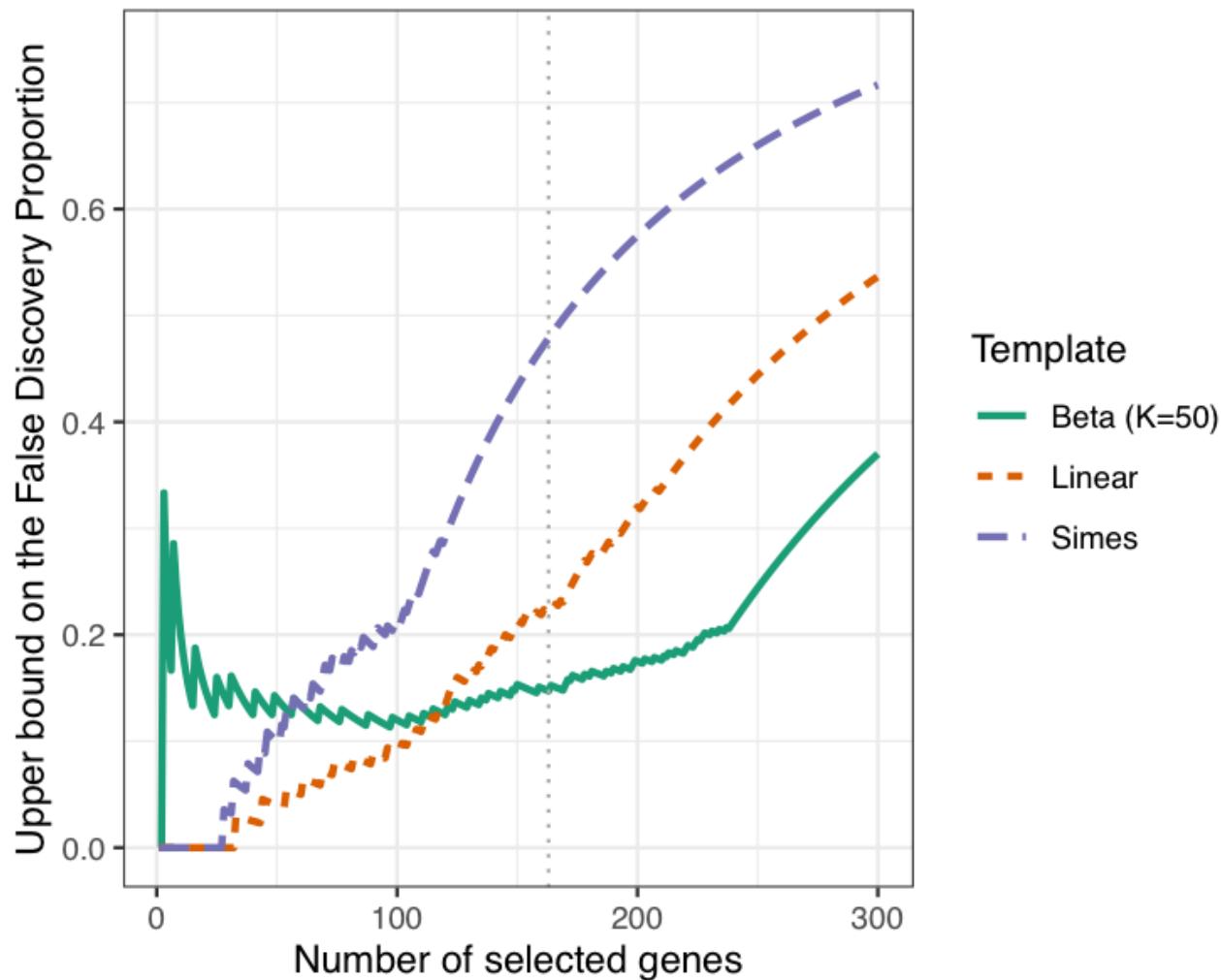
$$|S \cap \mathcal{H}_0| \leq |S| \wedge \min_{1 \leq k \leq |S|} \left\{ \sum_{i \in S} 1\{p_i > \alpha k/m\} + k - 1 \right\}$$

- Can be extended to other dependency settings, e.g. two-group testing by permutations.
- Valid for differential expression studies, GWAS, neuroimaging studies

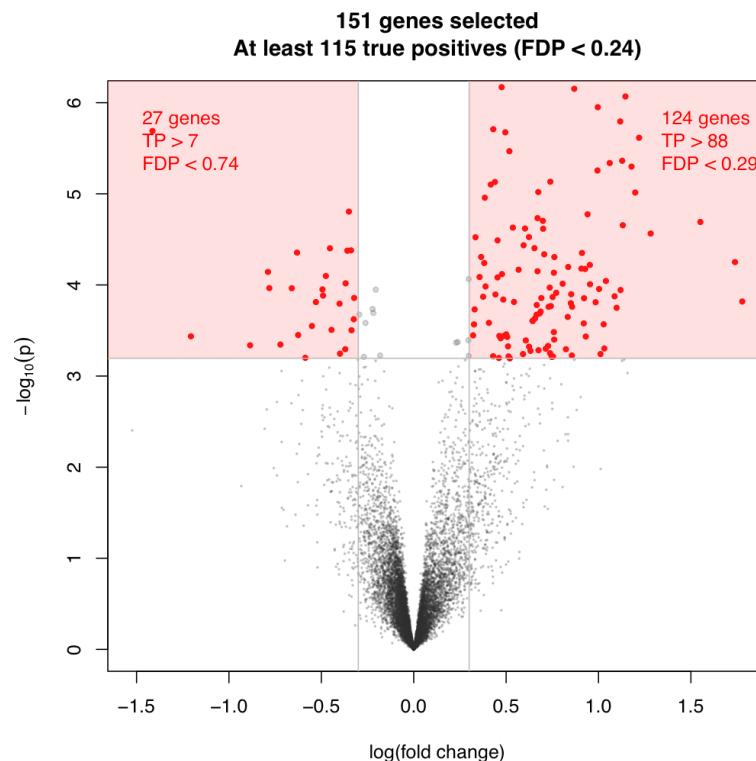
Leukemia data set: TP confidence bounds



Leukemia data set: FDP confidence bounds

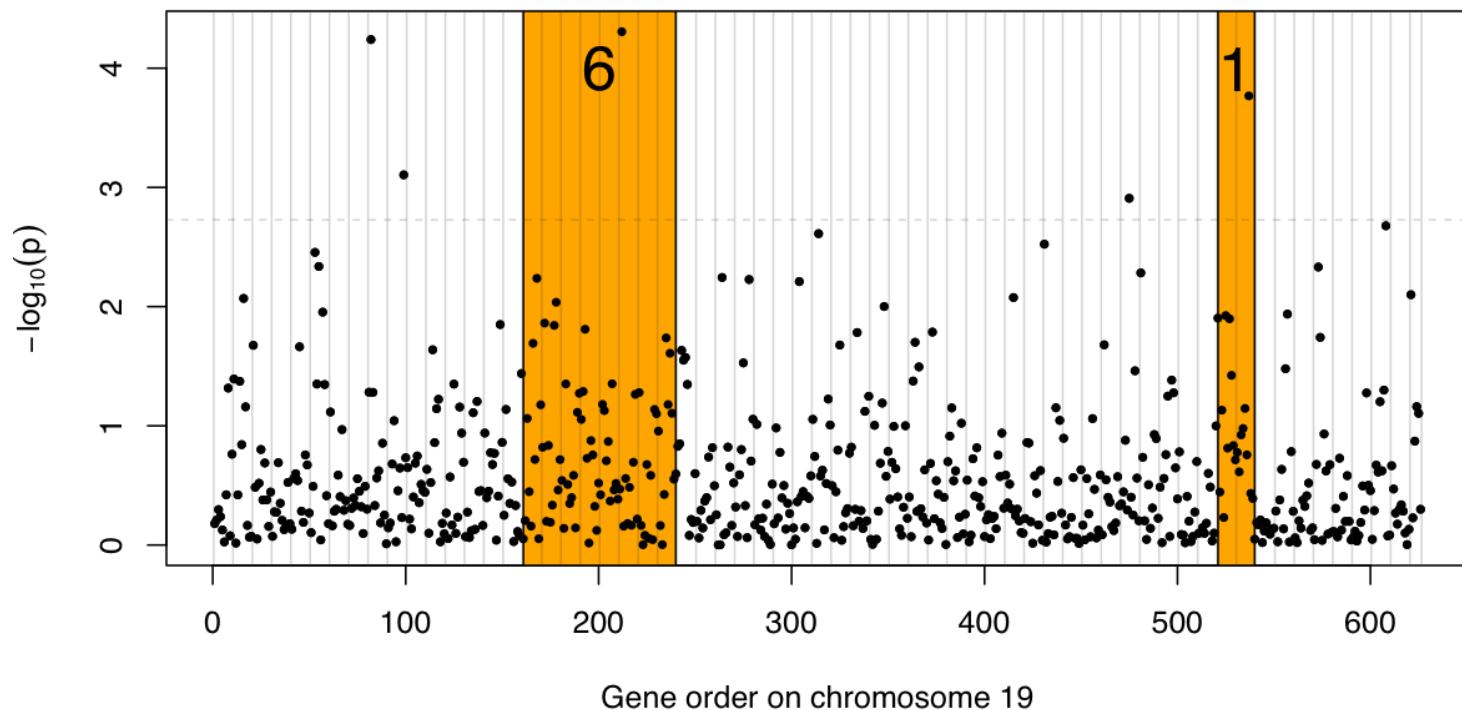


Leukemia data set: volcano plot



Interactive volcano plot: <https://pneuvial.shinyapps.io/volcano-plot/>

Leukemia data set: regional association plot



The selection can be done interactively:

https://pneuvial.shinyapps.io/posthoc-bounds_ordered-hypotheses/

Conclusion

Conclusion

- (large-scale) multiple testing is ubiquitous in biomedical data analysis
- multiple testing risks \neq multiple testing procedures
- FWER and FDR control different risks
 - FWER for confirmatory analyses
 - FDR for "exploratory" analyses
- post hoc inference can bypass some of the limitations of FDR

Some caveats

- interpretation of FDR control: FDR is an expectation!
- applicability conditions (dependence assumptions)

Related topics not explicitly discussed

- scientific reproducibility, hidden multiplicity and selective inference
- online multiple testing