Controlling the False Discovery Rate

A Practical and Powerful Approach to Multiple Testing

Benjamini Y and Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 1, pp. 289-300

Acknowledgement: Lecture10_MultipleTesting.ppt (washington.ehttps://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdfdu)

Background and Setting

Assume we are testing $H_1, H_2, \ldots, H_m$
m_0 = number of true hypotheses
R = number of rejected hypotheses
V = number of type I errors [false positives]

$$Q = \begin{cases} V/R & R>0 \\ 0 & \text{otherwise} \end{cases}$$

We further assume:

- m is known
- R is an observable random variable
- U, V, S and T are unobservable random variables

Number of errors committed when testing m null hypotheses

	declared non-significant	declared significant	total
True null hypotheses	**U** **Correct**	**V** **Type I error**	m_0
Non-true null hypotheses	**T** **Type II error**	**S** **Correct**	$m - m_0$
	m − **R**	**R**	m

Simple example for the hapo_metabolomics_2020:

We want to test the correlation of **all** metabolites in the data set with the value of FPG.

Conducting m= 51 tests independently we get R=28 significant ones (at a 0.05 level). We can expect by chance 2-3 wrong discoveries.

How we "weed out" false discoveries depends on the type of error rate we choose and the approach to control it.

BH-demo/BH-FDR-control-implementation.pdf at main · amuzikansky/BH-demo (github.com)

Multiple Testing Type 1 Error Rates

- **Per comparison error rate (PCER):**
  PCER = E(**V**/m)

- **Per-family error rate (PFER):**
  PFER = E(**V**).

- **Family-wise error rate (FWER):**
  FEWR = P(**V** ≥ 1)

- **Generalized FWER (k-FWER):**
  gFEWR(k) = P(**V** > k)

- **False discovery rate (FDR)**
  FDR =E(**V/R**) = E(**V/R** | **R**>0) P(**R**>0)

- **Positive false discovery rate (pFDR)**
  pFDR = E(**V/R** | **R** > 0) - next week.

	declared non-significant	declared significant	total
True null hypotheses	**U**	**V**	m_0
Non-true null hypotheses	**T**	**S**	$m - m_0$
	m − **R**	**R**	m

Controlling the Error Rate: Multiple Comparison Procedures (MCP)

- FWER ≤ $1 - (1 - \alpha)^m$ (strictly equal for independent tests)
- For m=100, FWER = 99% ($\alpha = 0.05$)
- MCP goal is to “control the probability of any type I error in families of comparisons under simultaneous consideration”.
- Many procedures have been developed to control FWER (Sidak, Dunnet, Scheffe, Tukey’s HSD, Hsu’s Best, etc.)
- Two general types of FWER correction:
  - Single-step procedure: each p-value gets the same adjustment - e.g. Bonferroni correction
    {if each hypothesis is tested at α/m, FWER ≤ α}
  - Sequential (step-up/step-down): broadly speaking, these are order based adjustments - e.g. Hochberg (1988), Holm-Bonferroni (2018), uniformly more powerful than single step procedures.
- Problems with MCP for FWER:
  - mostly assume test statistics are MVN (or t)
  - controlling FWER can be too stringent - depending on context, some false positive tests can be acceptable.
  - **can result in dramatic loss of power**.

Important properties of FDR

1. If all null hypothesis are true: $m_0 = m$

   $\rightarrow$ FDR is equivalent to FWER

   $\{v = 0 \rightarrow v/r = 0$
   $v > 0 \rightarrow v/r = 1$
   $E(V/R) = 0*P(V=0) + 1*P(V \geq 1) = P(V \geq 1) = FWER\}$

   Therefore controlling the FDR implies control of FWER

1. If $m_0 < m$

   $\rightarrow$ FDR $\leq$ FWER

   $\{v > 0 \rightarrow v/r < 1$
   $E(V/R) = E(V/R \mid v = 0) * P(V=0) + E(V/R \mid V \geq 1) * P(V \geq 1) \leq P(V \geq 1)\}$

   Procedures controlling the FWER also controls FDR

	declared non-significant	declared significant	total
True null hypotheses	**U**	**V**	m_0
Non-true null hypotheses	**T**	**S**	$m - m_0$
	m − **R**	**R**	m

BH proposed approach to multiple testing: control the FDR

Let $p_1, p_2, \ldots p_m$ be the set of p-values corresponding to the hypotheses $H_1, H_2, \ldots H_m$

To control the FDR at level q^* follow this procedure:

i. order all p-values for the hypothesis tested from smallest to largest, that is, $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$

ii. Let k be the largest value of i for which $p_{(i)} \leq (i / m)q^*$

iii. reject all $H_{(i)}$, $i = 1, 2, \ldots, k$

BH procedure control levels assumes **independent** test statistics.

Benjamini and Yekutieli expanded on the approach for cases of dependency

[The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics, Aug., 2001, Vol. 29, No. 4 (Aug., 2001), pp. 1165-1188]

BH Example:

- A group of m=15 tests.
- Controlling FWER with Bonferroni - reject all hypotheses < 0.05/15=0.0033
- Controlling the FDR at q* =0.05 →

```
> pvec <- c(0.0001,0.0004,0.0019,0.0095,0.0201,0.0278,0.0298,0.0344,0.0459,
+           0.3240,0.4262,0.5719,0.6528,0.7590,1.000)
> round(cbind("pval"=pvec,
+             'BH'=p.adjust(pvec,method='BH'),
+             'bon'=p.adjust(pvec,method='bonferroni')),4)
        pval     BH    bon
 [1,] 0.0001 0.0015 0.0015
 [2,] 0.0004 0.0030 0.0060
 [3,] 0.0019 0.0095 0.0285
 [4,] 0.0095 0.0356 0.1425
 [5,] 0.0201 0.0603 0.3015
 [6,] 0.0278 0.0639 0.4170
 [7,] 0.0298 0.0639 0.4470
 [8,] 0.0344 0.0645 0.5160
 [9,] 0.0459 0.0765 0.6885
[10,] 0.3240 0.4860 1.0000
[11,] 0.4262 0.5812 1.0000
[12,] 0.5719 0.7149 1.0000
[13,] 0.6528 0.7532 1.0000
[14,] 0.7590 0.8132 1.0000
[15,] 1.0000 1.0000 1.0000
>
```

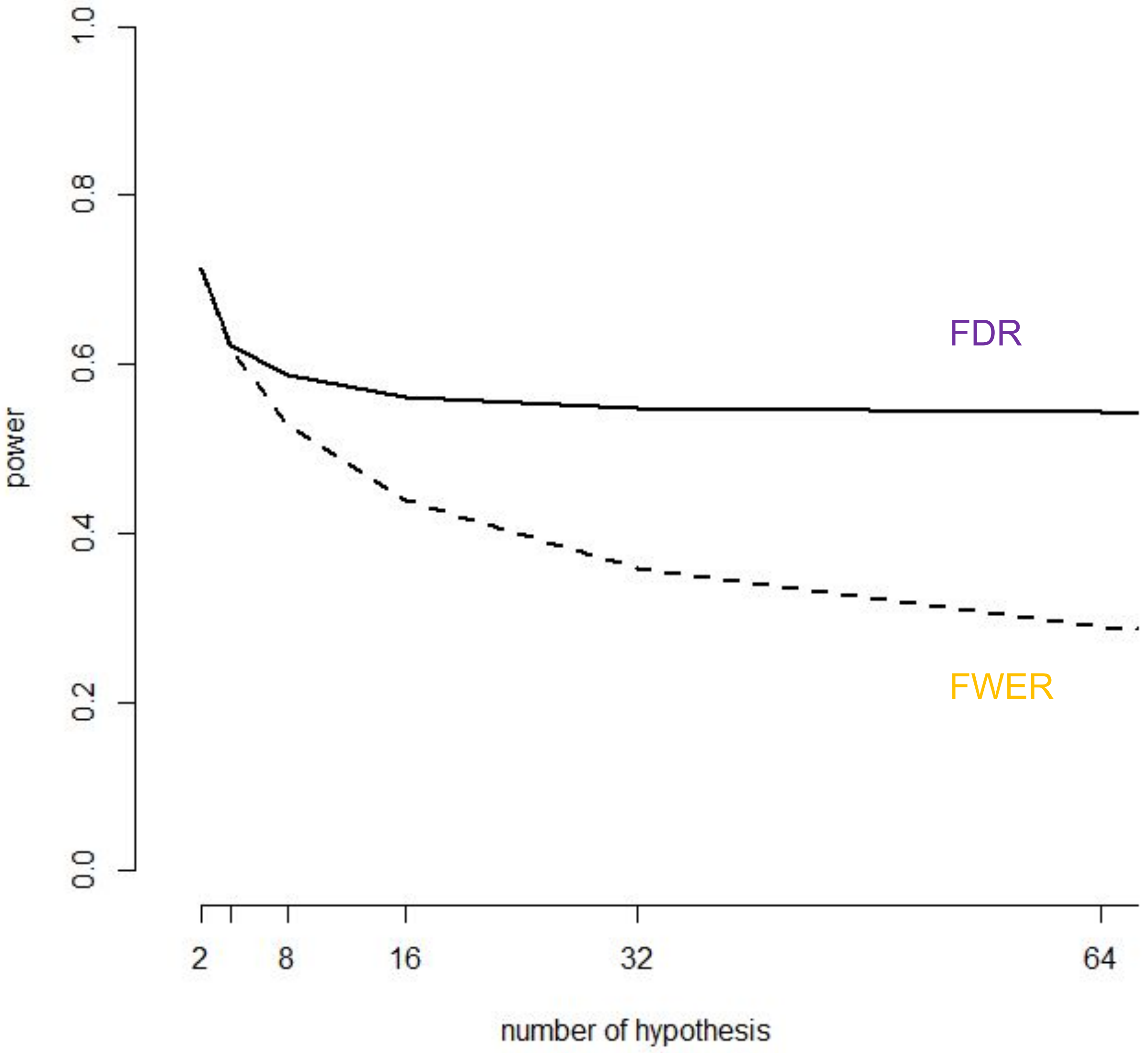
[BH-demo/Demo2.pdf at main · amuzikansky/BH-demo (github.com)]

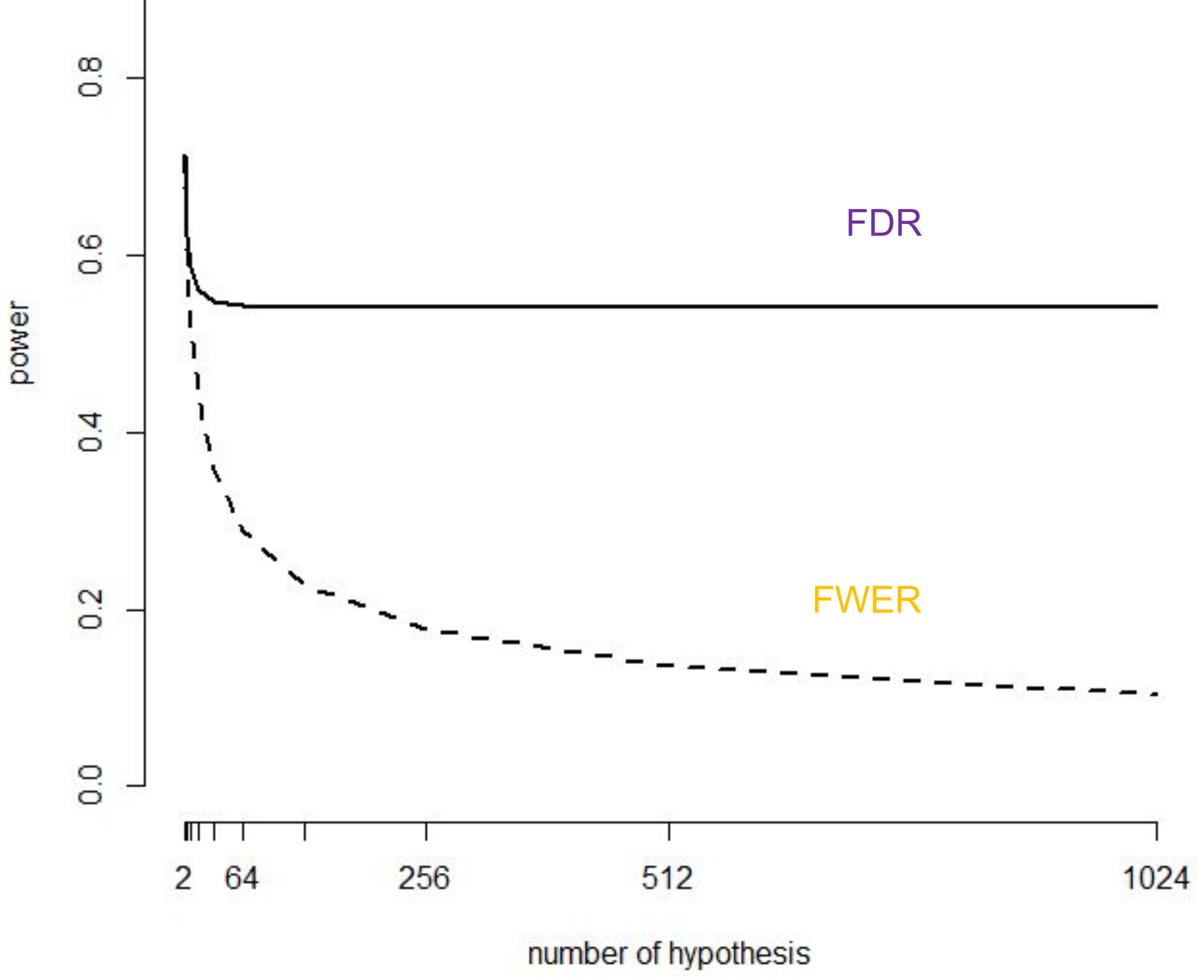
[control-implementation.pdf at main · amuzikansky/BH-demo (github.com)]

rank = i	$p_{(i)}$	(i/m) q*	q = $p_{(i)}$* m/i
1	**0.0001**	0.0033	0.0015
2	**0.0004**	0.0067	0.003
3	**0.0019**	0.01	0.0095
4	**0.0095**	0.0133	0.0356
5	0.0201	0.0167	0.0603
6	0.0278		0.0695
7	0.0298		0.0639
8	0.0344		0.0645
9	0.0459		0.0765
10	0.3240		0.486
11	0.4262		0.5811
12	0.5719		0.714
13	0.6528		0.722
14	0.7590		0.813
15	1.000		1.000

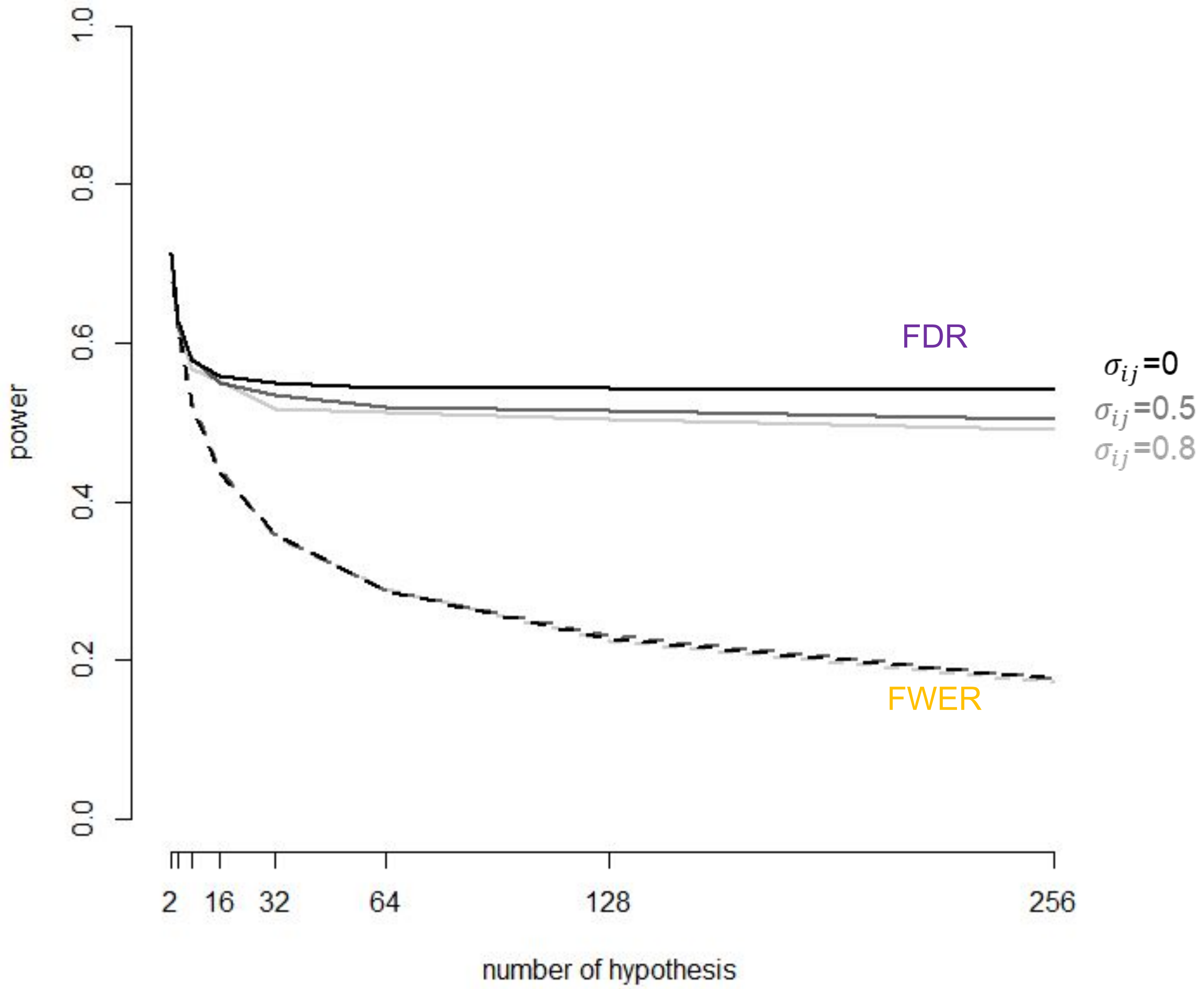
Simulation-based estimates of power
Independent hypothesis

3/4 X ~ n(0,1)
1/4 X ~ n(2.8,1)
H_0: $\mu = 0$; H_0: $\mu \neq 0$
sim =10^5





Simulation-based estimates of power
Correlated hypothesis



$X \sim \mathrm{mvn}\ (\mu, \Sigma)$

$3/4\ \mu = 0$

$1/4\ \mu = 2.8$

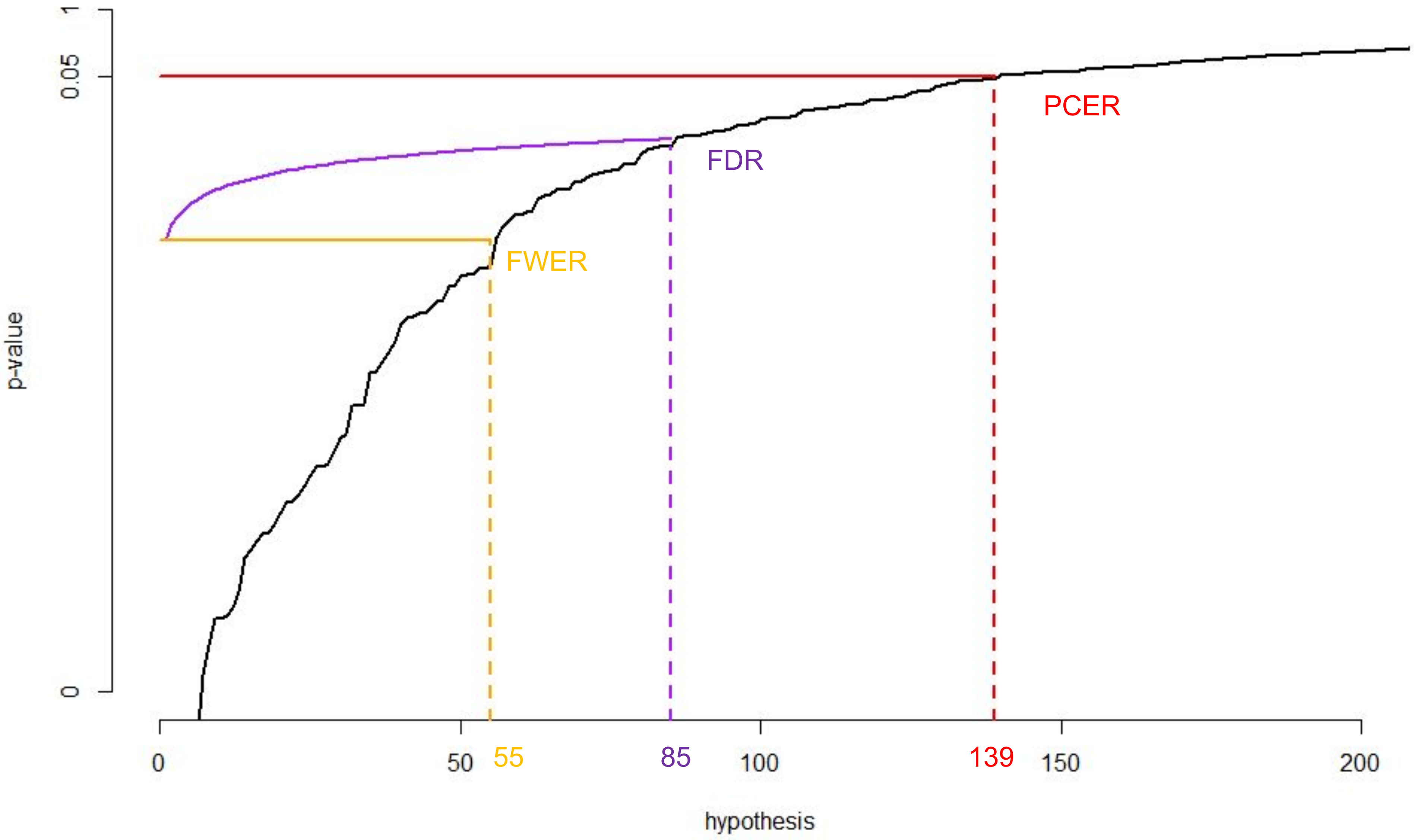
$\sigma_{ij} = 1$ for $i = j$

$\sigma_{ij} = \{0, .5, .8\}$ for $i \neq j$

H_0: $\mu = 0$; H_0: $\mu \neq 0$

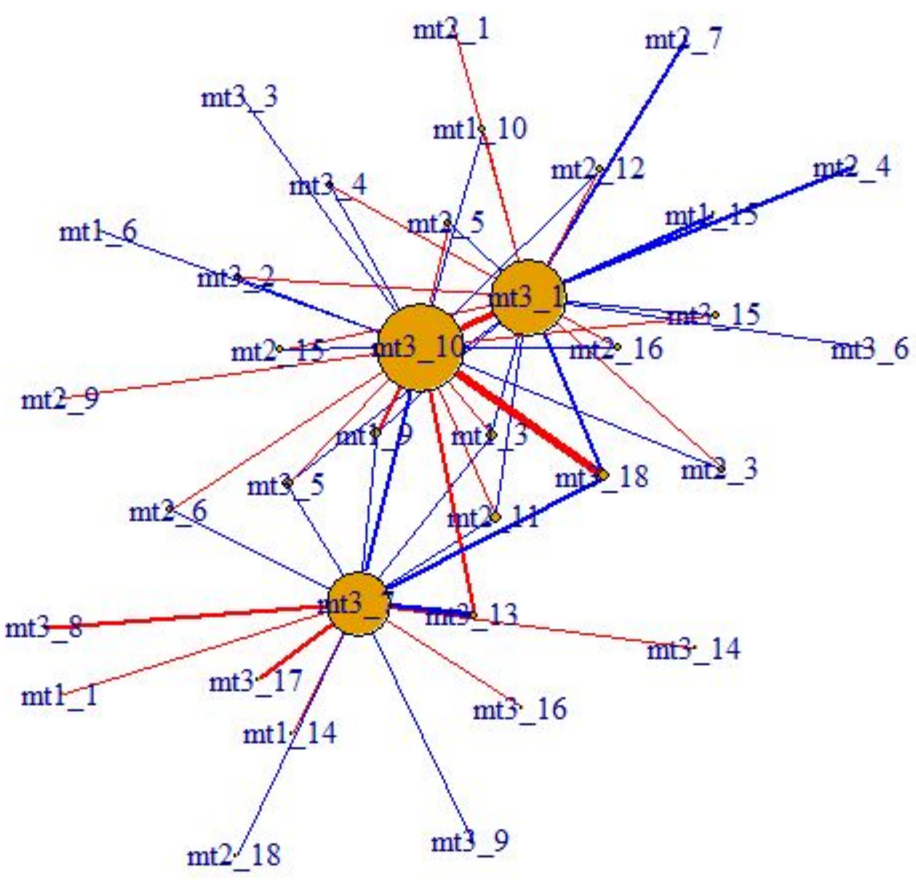
sim $= 10$^4

Differential Network Analysis (Kate's example)

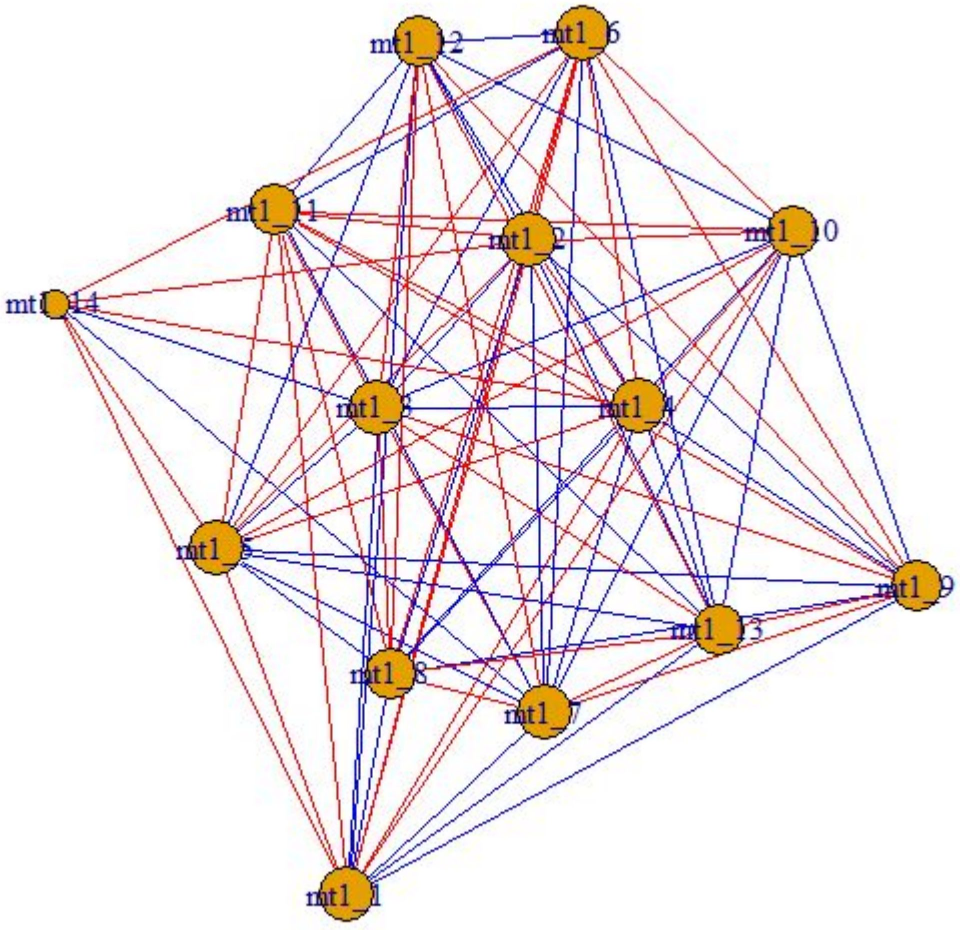


Determine which of the 1275 possible edges (51 choose 2) differs significantly between ancestry group 1 and ancestry group 2

Differential Network Analysis (Kate's example)



FWER (55)



FDR (85)

Conclusions

- FWER maybe ill-suited to
  - Large number of hypothesis
  - The null is unlikely to be true in many instances
- FDR is less stringent and more powerful
- There are improved FDR versions