

Multiples Testen

Daria Salyakina

Max-Planck Institut für Psychiatrie

2004

Multiples Testen

Testhypothesen in Microarray Studien

- um Gene mit unterschiedlicher Expression unter verschiedenen experimentellen Bedingungen aufzufinden
- es ist wichtig eine hohe Wahrscheinlichkeit für eine korrekte Aussage zu haben, dass die Gene signifikant unterschiedlich exprimiert sind

Testhypothesen in Microarray Studien

- Für jedes Gen sind 2 Situationen möglich
 - die Gene sind nicht unterschiedlich exprimiert = H_0 ist wahr
 - die Gene sind unterschiedlich exprimiert = H_A ist wahr
- Testerklärung
 - die Gene haben verschiedene Expressionsniveaus (H_0 wird verworfen)
 - es gibt keinen Unterschied (H_0 wird akzeptiert)

Testhypothesen in Microarray Studien

Zwei Arten von Fehlern können auftreten

Fehler I. Art („Type I error“ oder „false positive“)

- fälschlicherweise deklariert, dass die Gene unterschiedlich exprimiert sind, *verwirft wahre Null-Hypothese*

Fehler II. Art („Type II error“ oder „false negative“)

- wenn man den wahren Effekt nicht aufdecken kann, *es stört uns die falsche Null-Hypothese zu verwerfen*

W	F	W	W	F
t_1	t_2	t_3	t_4	t_5
H_0	H_A	H_0	H_A	H_0

Testhypothesen in Microarray Studien

Zwei Arten von Fehlern können auftreten

Fehler I. Art („Type I error“ oder „false positive“)

- fälschlicherweise deklariert, dass die Gene unterschiedlich exprimiert sind, *verwirft wahre Null-Hypothese*

Fehler II. Art („Type II error“ oder „false negative“)

- wenn man den wahren Effekt nicht aufdecken kann, *es stört uns die falsche Null-Hypothese zu verwerfen*

W	F	W	W	F
t_1	t_2	t_3	t_4	t_5
H_0	H_A	H_0	H_A	H_0

Testhypothesen in Microarray Studien

Microarray Studien umfassen gleichzeitig Tausende von Genen
Dabei steigt die Wahrscheinlichkeit der falschen Schlussfolgerung

- bei 1000 Tests werden 50 falsch positive erwartet die bei Fehler I. Art = 0.05 als signifikant deklariert sind
- die Wahrscheinlichkeit, dass wenigstens ein p-Wert $< \alpha$ für unabhängige Tests wird $1-(1-\alpha)^G$ konvergiert gegen 1 für wachsendes G.

bei $\alpha=0.01$, $G=1000$ ist diese Chance = 0.9999568 !

- individuelle p-Werte von 0.01 korrespondieren nicht mehr mit signifikanten Aussagen.

*Man muß individuelle p-Werte für **multiples Testen** korrigieren!*

Testhypothesen in Microarray Studien

	nicht verworfene Hypothesen	verworfen Hypothesen	
Wahre Null-Hypothesen	U	V Fehler I Art	G_0
Falsche Null-Hypothesen	T Fehler II Art	S	G_1
	G - R	R	G

G - Testhypothesen $H_0 \dots H_G$, G_0 - Anzahl von Wahre Null-Hypothesen

U, V, S, T - unbeobachtete Variablen

beobachtete Variable R = Anzahl von verworfenen Hypothesen

Fehler I. Art

- 1 Per family error rate (PFER) definiert als erwartete Anzahl der Fehlern I. Art

$$\text{PFER} = E(V)$$

- 2 Per comparison error rate (PCER) ist die erwartete Anzahl der Fehler I. Art dividiert durch die Anzahl der Tests

$$\text{PCER} = E(V)/G$$

- 3 Family wise error rate (FWER) ist die Wahrscheinlichkeit für wenigstens einen Fehler I. Art

$$\text{FWER} = P(V > 0)$$

	nicht verworfene Null-Hypothesen	verworfenen Null-Hypothesen	
Wahre Null-Hypothesen	U	V Fehler I Art	G_0
Falsche Null-Hypothesen	T Fehler II Art	S	G_1
	G-R	R	G

- 4 False discovery rate (FDR) Benjamini & Hochberg (1995) haben FDR als erwarteten Anteil Fehlern I. Art an der verworfenen Null-Hypothesen definiert

$$\text{FDR} = E(Q)$$

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

Fehler I. Art

Für eine Multiple Test Prozedur gilt

$$\text{PCER} \leq \text{FWER} \leq \text{FDR}$$

Unter kompletter Null-Hypothese, $G_1 = 0$ gilt

$$\text{FWER} = \text{FDR}$$

Für feste α ist die Kontrolle der Fehler I. Art durch die FWER Prozedur immer konservativer (strenger) als die Kontrolle der Fehler I. Art durch die FDR Prozedur.

Daraus folgt, dass die FDR Methode effektiver für die Entdeckung der wirklich falschen Null-Hypothesen ist.

Kontrolle der FWER

observed p-values	Bonferroni	Holm Step-down	Hochberg Step-up
p_{r_1}	α/G	α/G	α/G
p_{r_2}	α/G	$\alpha/(G-1)$	$\alpha/(G-1)$
:	:	:	:
p_{r_g}	α/G	$\alpha/(G-g+1)$	$\alpha/(G-g+1)$
:	:	:	:
$p_{r_{G-1}}$	α/G	$\alpha/2$	$\alpha/2$
p_{r_G}	α/G	α	α

Kontrolle der FWER

observed p-values	Bonferroni	Holm Step-down	Hochberg Step-up
p_{r_1}	α/G	α/G	α/G
p_{r_2}	α/G	$\alpha/(G-1)$	$\alpha/(G-1)$
:	:	:	:
p_{r_g}	α/G	$\alpha/(G-g+1)$	$\alpha/(G-g+1)$
:	:	:	:
$p_{r_{G-1}}$	α/G	$\alpha/2$	$\alpha/2$
p_{r_G}	α/G	α	α

Bonferroni

$$p \leq \alpha / G$$

$$p_{\text{korr}} = \min(1, p * G)$$

$$p_{\text{korr}} \leq \alpha$$

Holm / Hochberg

$$p \leq \alpha / (G-g+1)$$

$$p_{\text{korr}} = \min(1, p * (G-g+1))$$

$$p_{\text{korr}} \leq \alpha$$

Kontrolle der FWER

	observed p-values	Bonferroni	Holm Step-down	Hochberg Step-up
1	0.00123	0.0123	0.0123	0.0123
2	0.00345	0.0345	0.03105	0.03105
3	0.00356	0.0356	0.02848	0.02848
4	0.0081	0.081	0.0567	0.0567
5	0.0083	0.083	0.0498	0.0498
6	0.0267	0.267	0.1335	0.1335
7	0.0422	0.422	0.1688	0.1688
8	0.0891	0.891	0.2673	0.2673
9	0.1312	1	0.2624	0.2624
10	0.6725	1	0.6725	0.6725

$\alpha = 0.05$
 $G = 10$

Bonferroni
 $p \leq \alpha / G$
 $p_{\text{korr}} = \min(1, p * G)$
 $p_{\text{korr}} \leq \alpha$

Holm / Hochberg
 $p \leq \alpha / (G-g+1)$
 $p_{\text{korr}} = \min(1, p * (G-g+1))$
 $p_{\text{korr}} \leq \alpha$

Kontrolle der FWER

Resampling:

die Abschätzung der gemeinsamen Verteilung der Teststatistiken t_1, \dots, t_G unter der kompletten Null-Hypothese (H_0) bei Permutation der Spalten der Datenmatrix X .

Permutationsalgorithmus (Westfall & Young, 1993) für nicht korrigierte Teststatistiken:
für die Permutation $b=1 \dots B$

- 1) permutieren der n Spalten der Matrix X
- 2) kalkulieren der Teststatistiken $t_{1,b} \dots t_{G,b}$ für jede Hypothese.
- 3) berechnen des sukzessiven Maximums der Teststatistiken

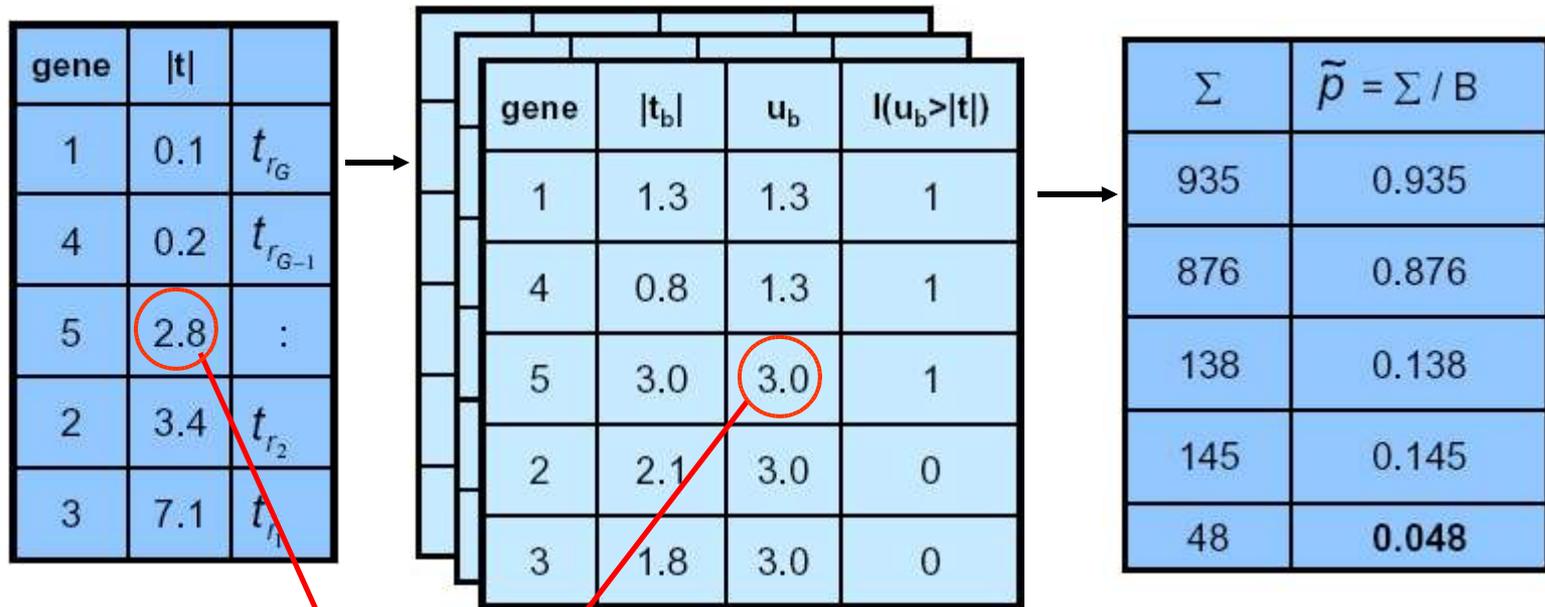
$$u_{G,b} = |t_{r_G,b}|$$
$$u_{g,b} = \max\{u_{g+1,b}, |t_{r_g,b}|\} \quad \text{für } g = G-1, \dots, 1$$

Für zweiseitige alternative Hypothese der Permutation ist der p-Wert für die Hypothese H_g

$$p_g^* = \frac{1}{B} \sum_{b=1}^B I(|t_{g,b}| \geq |t_j|)$$

wobei $I(\cdot)$ eine Indikatorfunktion ist, die 1 zurückgibt, wenn die Gleichung wahr ist und 0 andernfalls.

Permutationsalgorithmus von Westfall & Young (1993)



- Sortieren der beobachtete Werte

B=1000 Permutationen

korrigierte p-Werte

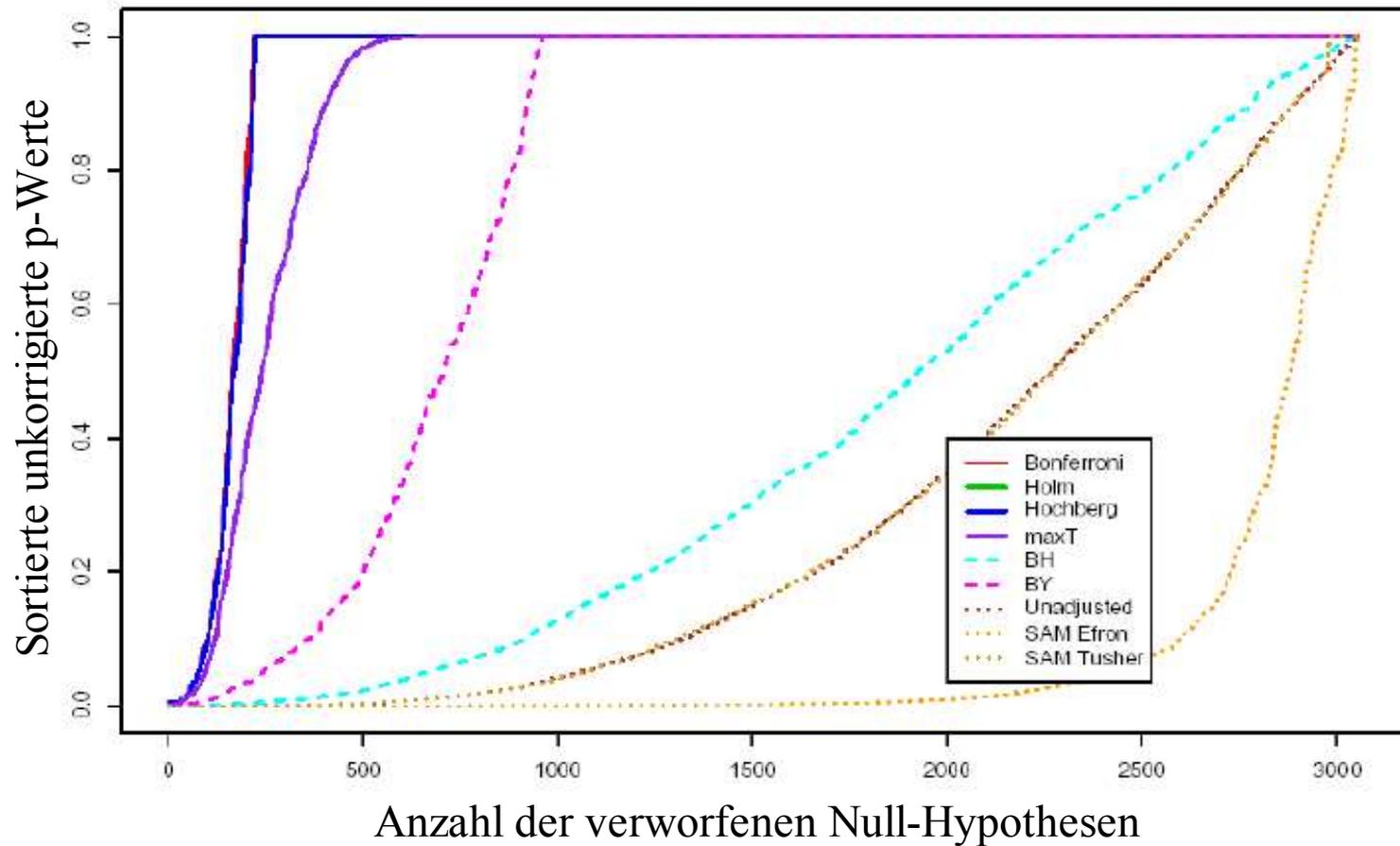
O. Hartmann - NGFN Symposium, 19.11.2002 Berlin

$$\tilde{p}_{r_g}^* = \frac{1}{B} \sum_{b=1}^B I(u_{g,b} \geq |t_{r_g}|)$$

$$u_{G,b} = |t_{r_{G,b}}|$$

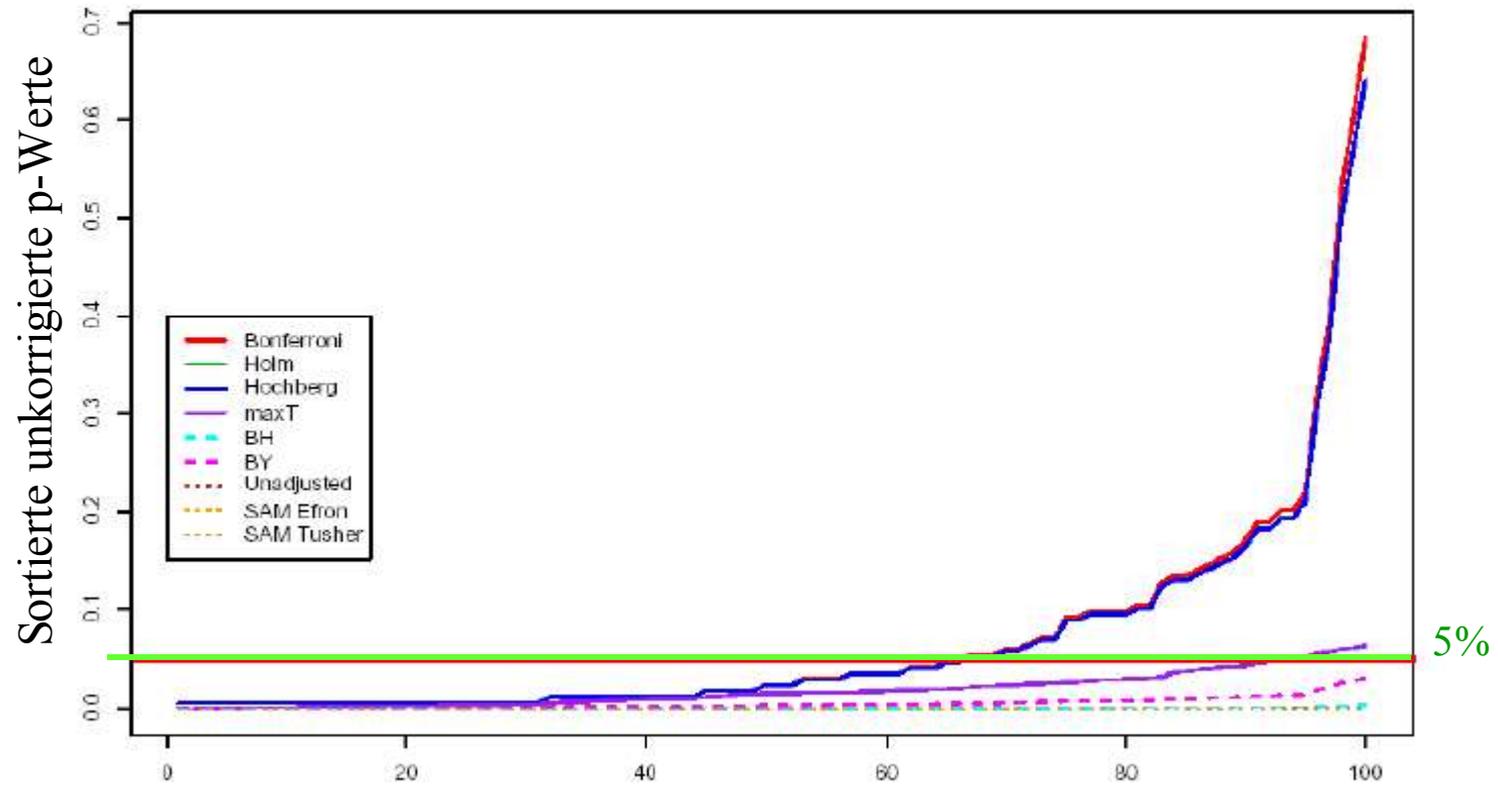
$$u_{g,b} = \max\{u_{g+1,b}, |t_{r_{g,b}}|\} \text{ für } g = G-1, \dots, 1$$

Example: Leukemia study, Golub et al. (1999)



Dudoit et al. (2002)

Example: Leukemia study, Golub et al. (1999)



Anzahl der verworfenen Null-Hypothesen

Dudoit et al. (2002)

Kontrolle der FDR

FDR ist erwarteter Anteil an Fehlern I. Art der verworfenen Null-Hypothesen

$$FDR = E(Q), \quad Q = \begin{cases} V/R & : R > 0 \\ 0 & : R = 0 \end{cases}$$

observed p-values	Benjamini-Hochberg (BH)
p_{r_1}	α/G
p_{r_2}	$2*\alpha/G$
:	:
p_{r_g}	$g*\alpha/G$
:	:
$p_{r_{G-1}}$	$(G-1)*\alpha/G$
p_{r_G}	$G*\alpha/G$

Linear step-up procedure (Benjamini & Hochberg, 1995)

$g^* = \max\{g : p_{r_g} \leq \frac{g}{G}\alpha\}$, verwirft H_g for $g = 1, \dots, g^*$,
 korrigierter p-Werte $\tilde{p}_{r_g} = \min_{k=g, \dots, G} \{\min(\frac{G}{k} p_{r_k}, 1)\}$

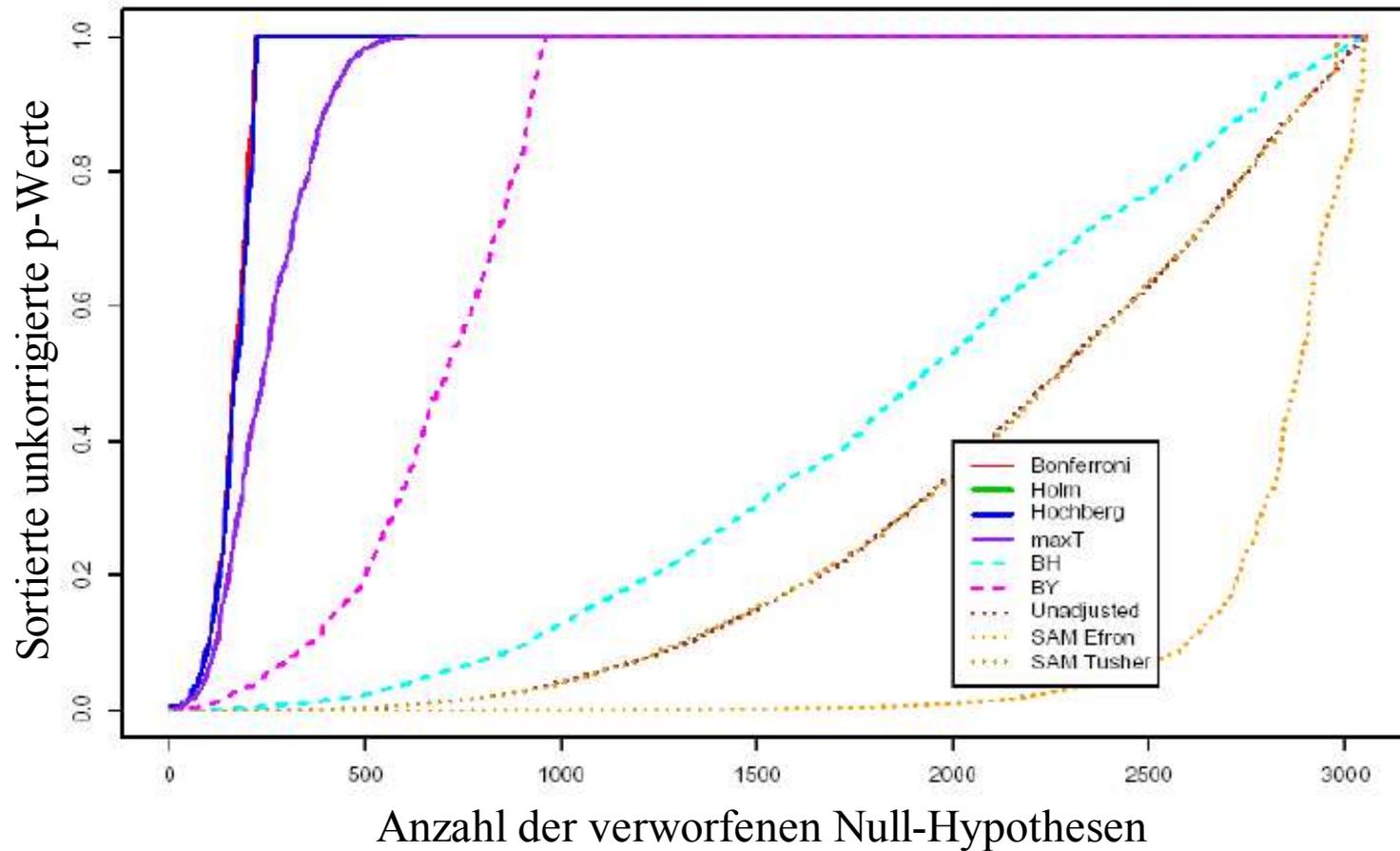
- kontrolliert FDR in Niveau α für unabhängige Tests

$$FDR \leq \alpha * G_0 / G \leq \alpha$$

Resampling FDR adjustments

- Yekutieli & Benjamini (1999) *J. Statist. Plan. Inference* 82, 171-196
- Reiner, Yekutieli & Benjamini (2003) *Bioinformatics* 19, 368-375

Example: Leukemia study, Golub et al. (1999)



Dudoit et al. (2002)

Zusammenfassung

- Beim multiplen Testen gibt es verschiedene Methoden für die Kontrolle der family wise error rate (FWER)
- korrigierte p-Werte bieten flexible Zusammenfassungen der Ergebnisse des multiplen Testens an und erlauben verschiedene Methoden zu vergleichen
- die FDR Methode ist weniger konservativ und hat mehr Power als die FWER.

Empfohlene Software: Bioconductor R multtest package
(<http://www.bioconductor.org/>)

Literature

- **Benjamini, Y. & Hochberg, Y. (1995).** Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Statist. Soc. B* **57**: 289-300.
- **Benjamini, Y. and Hochberg, Y. (2000)** On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.*, **25**, 60–83.
- **Benjamini, Y. and Yekutieli, D. (2001b)** The control of the false discovery rate under dependency. *Ann Stat.* **29**, 1165–1188.
- **Callow, M. J., Dudoit, S., Gong, E. L., Speed, T. P. & Rubin, E. M. (2000).** Microarray expression profiling identifies genes with altered expression in HDL deficient mice, *Genome Research* **10**(12): 2022-2029.
- **S. Dudoit, J. P. Shaffer, and J. C. Boldrick (Submitted).** Multiple hypothesis testing in microarray experiments, Technical Report #110 (<http://stat-www.berkeley.edu/users/sandrine/publications.html>)
- **Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M., Downing, J. R., Caligiuri, M. A., Bloomeld, C. D. & Lander, E. S. (1999).** Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* **286**: 531-537.

Literature

- **Hochberg, Y. (1988).** A sharper bonferroni procedure for multiple tests of significance, *Biometrika* **75**: 800- 802.
- **Holm, S. (1979).** A simple sequentially rejective multiple test procedure, *Scand. J. Statist.* **6**: 65-70.
- **M.-L. T. Lee & G.A. Whitmore (2002)** Power and sample size for DNA microarray studies. *Statistics in Medicine* **21**, 3543-3570.
- **A. Reiner, D. Yekutieli & Y. Benjamini (2003)** Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375
- **Westfall, P. H. & Young, S. S. (1993).** *Resampling-based multiple testing: Examples and methods for p-value adjustment*, John Wiley & Sons.
- **Yekutieli,D. and Benjamini,Y. (1999)** Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan Infer.*, **82**, 171–196.