



FDRコンセプトと FDR法についての考察

2004年7月29日

SAS Forum ユーザー会

学術総会 2004

塩野義製薬株式会社 解析センター

佐藤 亜香里・角谷 伸一・田崎 武信



FDRの活躍の場面

False Discovery Rate

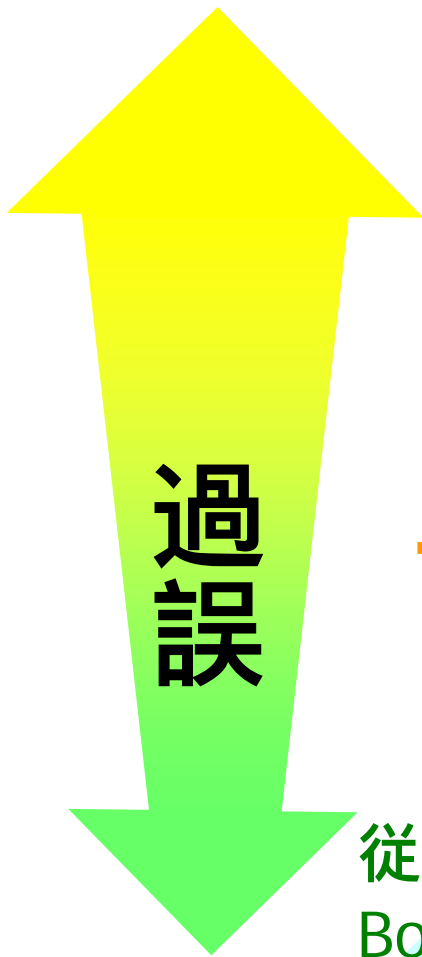
多重比較に対するひとつのアプローチ

Benjamini and Hochberg(1995) : FDRコンセプト, FDR法

Benjamini and Yekutieli(2001) : 相関を考慮したFDR補正法

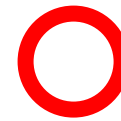
{ 多変数比較
多群比較





□
多重性の
調整なし

相関を考慮したFDR補正法



従来の多重比較法
Bonferroni法, Tukey法

FDR法

保守的

検出力

積極的！



FDRコンセプト



| 仮説 | 検定 | | 計 |
|--------|---------|-----|-----------|
| | 非有意 | 有意 | |
| 帰無仮説が真 | U | V | m_0 |
| 対立仮説が真 | T | S | $m - m_0$ |
| 計 | $m - R$ | R | m |

$$FDR = E\left(\frac{V}{R}\right) \quad \rightarrow \quad FDR = E\left(\frac{V}{R} \mid R > 0\right) \Pr(R > 0)$$

FDR = 棄却した帰無仮説のうち、
間違って棄却した帰無仮説の割合
の期待値



FDR法 (FDR制御法)

m 個の帰無仮説 H_1, H_2, \dots, H_m

p値 p_1, p_2, \dots, p_m



小さい順に並べる

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)} \leq p_{(k+1)} \leq \dots \leq p_{(m)}$$

$$\underbrace{H_{(1)}, H_{(2)}, \dots, H_{(k)}}_{\text{棄却}}, \underbrace{H_{(k+1)}, \dots, H_{(m)}}_{\text{受容}}$$

棄却

受容

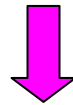
$k : p_{(i)} \leq (i/m) \alpha$ を満たす最大の i

ある固定された α に対して, $FDR \leq \alpha$ を保証する

相関を考慮したFDR補正法

m 個の帰無仮説 H_1, H_2, \dots, H_m

p値 p_1, p_2, \dots, p_m



小さい順に並べる

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)} \leq p_{(k+1)} \leq \dots \leq p_{(m)}$$

$$\underbrace{H_{(1)}, H_{(2)}, \dots, H_{(k)}}_{\text{棄却}}, \underbrace{H_{(k+1)}, \dots, H_{(m)}}_{\text{受容}}$$

棄却

受容

$$\alpha^* \leq \alpha$$

保守的

$k : p_{(i)} \leq (i/m) \alpha^*$ を満たす最大の i

$$\alpha^* = \alpha / \sum_{j=1}^m (1/j)$$

ある固定された α に対して, $FDR \leq \alpha$ を保証する

棄却域の違い

p値

$p_{(i)}$

有意水準

2 /m

/m

1

2

m

i

多重性の
調整なし

FDR法

相関を考慮した
FDR補正法

Bonferroni法

p値の順位

検定の数

多変数比較



健常群 疾患群

遺伝子1

⋮

⋮

⋮

⋮

⋮

⋮

⋮

遺伝子m

| | |
|--|--|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

たくさんの遺伝子ごとに、
疾患群と健常群の2群比較を
くり返す





多変数比較

相関

大 0.7

中 0.5

なし 0

健常群

疾患群

遺伝子1

⋮

遺伝子10

遺伝子11

⋮

遺伝子20

遺伝子21

⋮

遺伝子30

⋮

遺伝子1000

1群10例

t検定で比較

シミュレーション回数
1000回

想定1: 差がない

想定2: 差が1

想定3: 差が2

想定4: 1～10で差が2,
残りで差がない

想定5: 21～30で差が2,
残りで差がない

シミュレーション評価方法



「有意差検出数」の値を
とったくり返しの回数 / 1000回

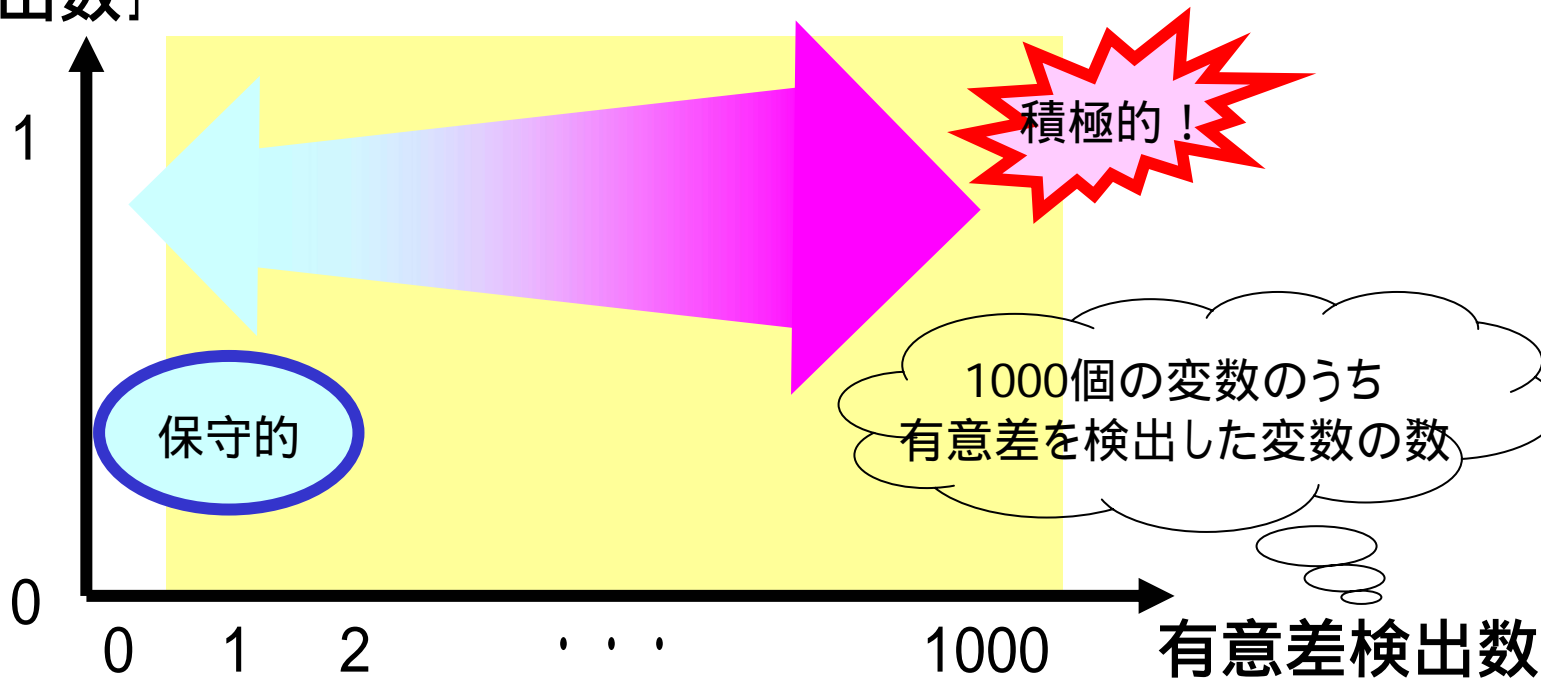
(多重性の)調整なし

FDR法

相関を考慮したFDR補正法

* Bonferroni法

「有意差検出数」
の頻度

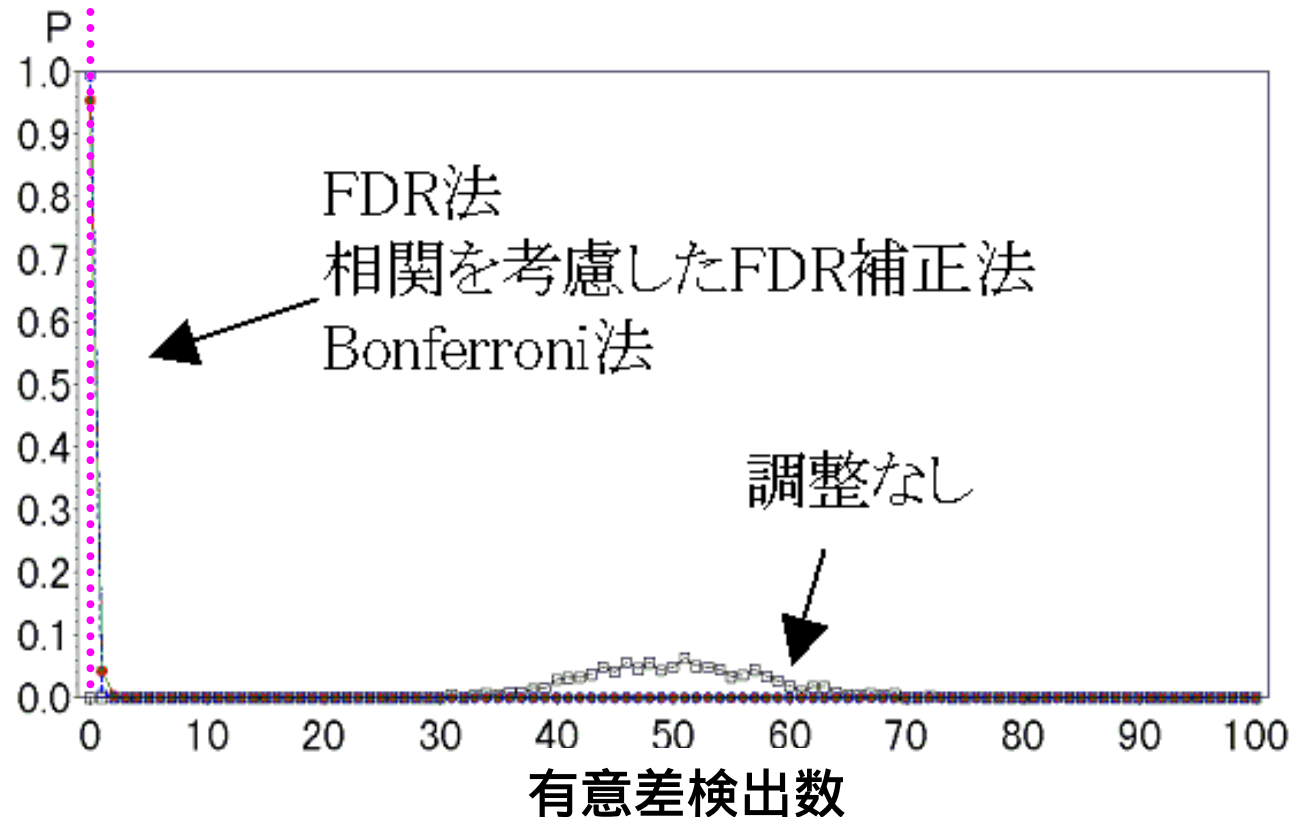


シミュレーション結果 - 想定1 -

差なし

差がある変数

0個



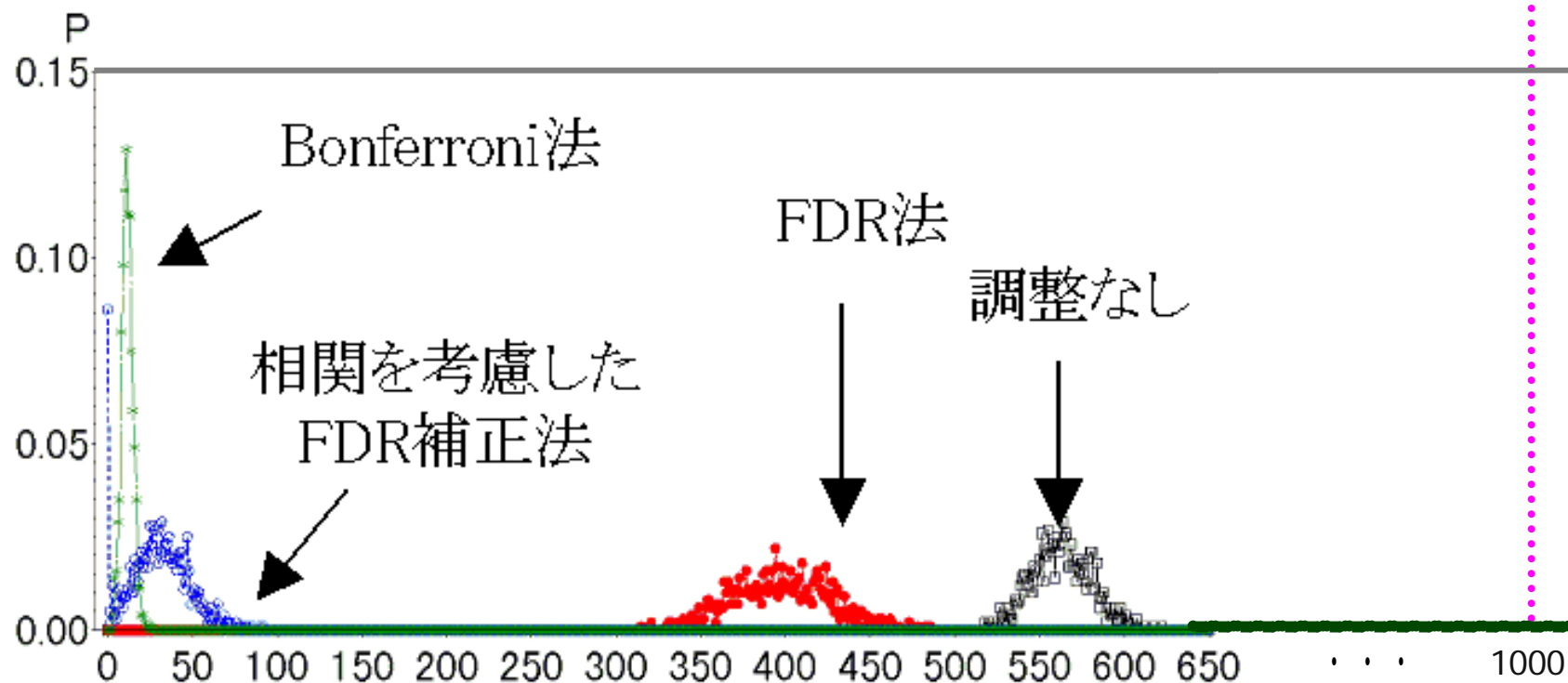
method 1. not Adjusted 2. FDR
3. FDR with correlation 4. Bonferroni

シミュレーション結果 - 想定2 -

差1

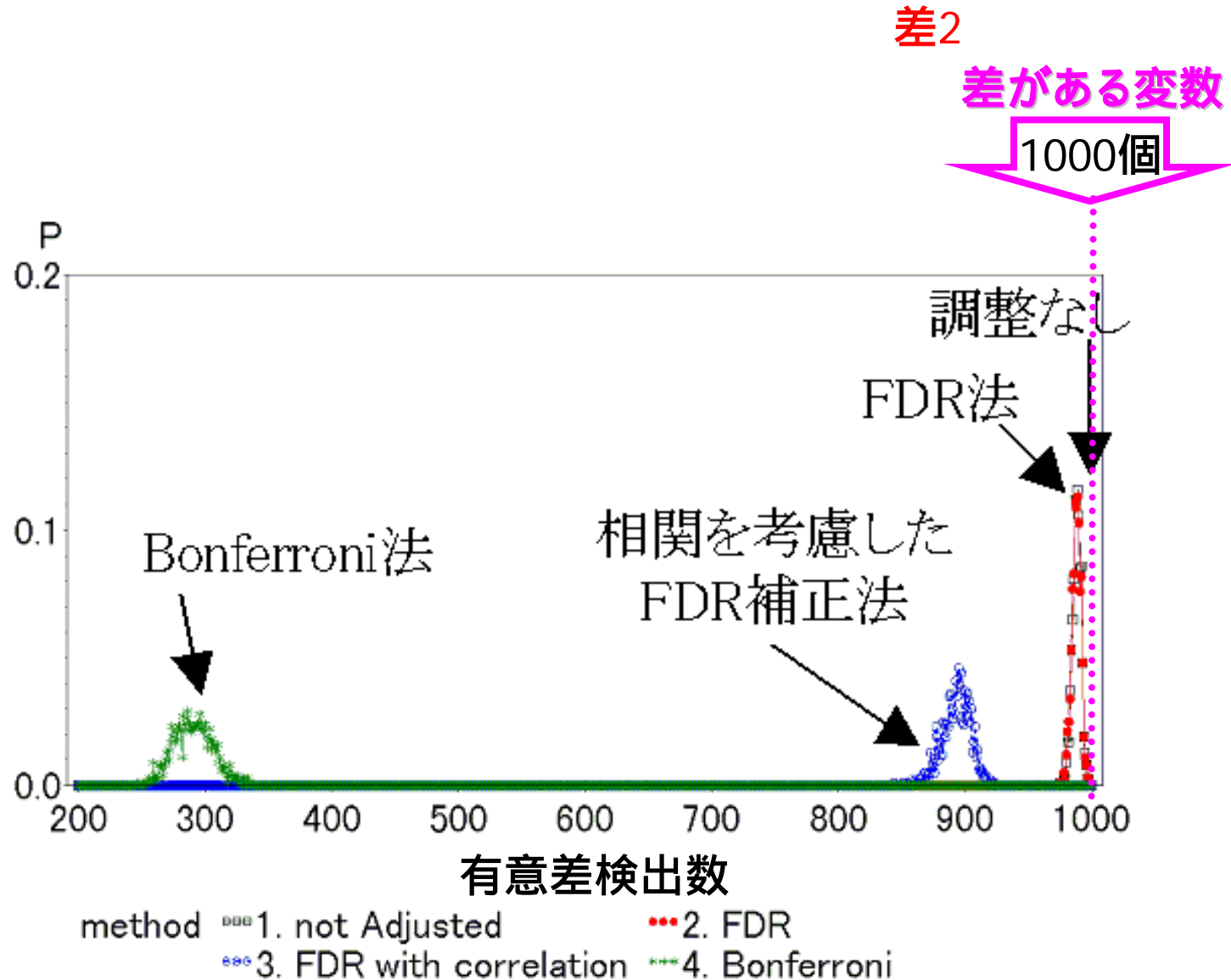
差がある変数

1000個



method 000 1. not Adjusted 2. FDR
000 3. FDR with correlation 4. Bonferroni

シミュレーション結果 - 想定3 -

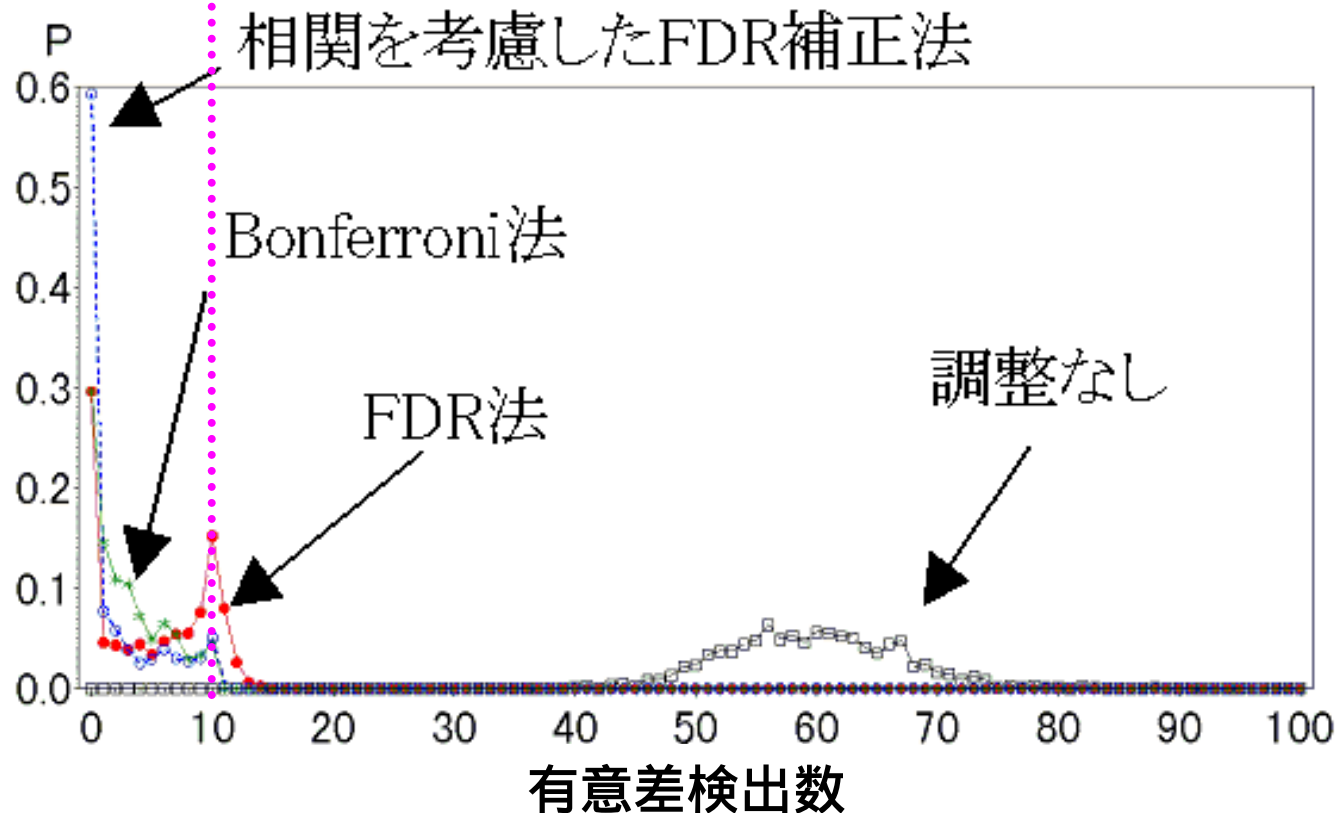


シミュレーション結果 - 想定4 -

10個の遺伝子(相関あり)で差2, 他差なし

差がある変数

10個



method 1. not Adjusted 2. FDR
3. FDR with correlation 4. Bonferroni

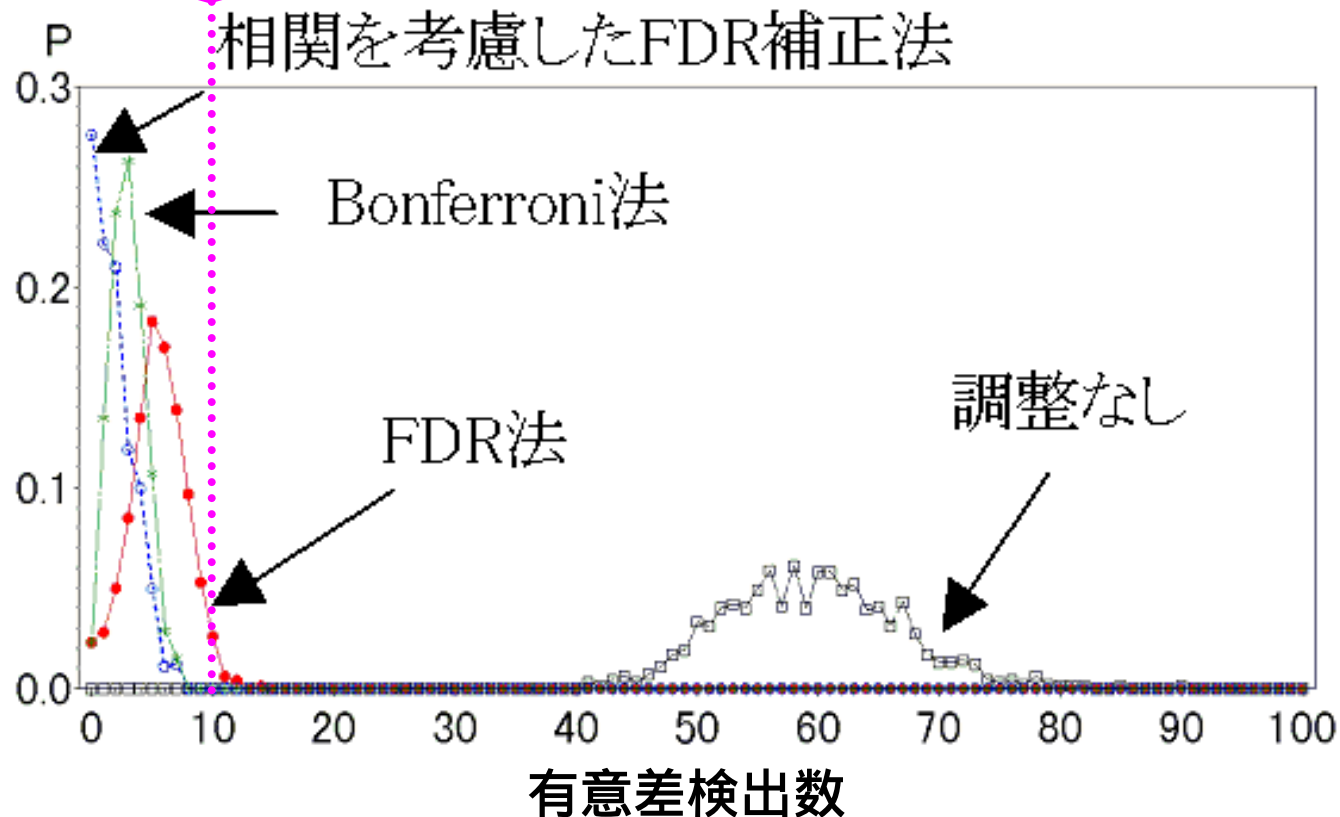


シミュレーション結果 - 想定5 -

10個の遺伝子(相関なし)で差2, 他差なし

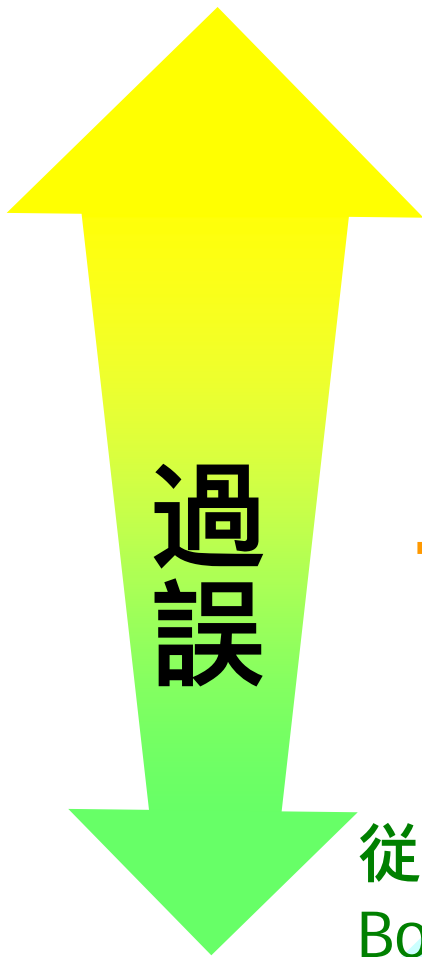
差がある変数

10個



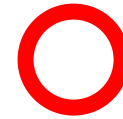
method □ 1. not Adjusted ● 2. FDR
 ○ 3. FDR with correlation × 4. Bonferroni





□
多重性の
調整なし

相関を考慮したFDR補正法



従来の多重比較法

Bonferroni法, Tukey法

FDR法

保守的

検出力

積極的！



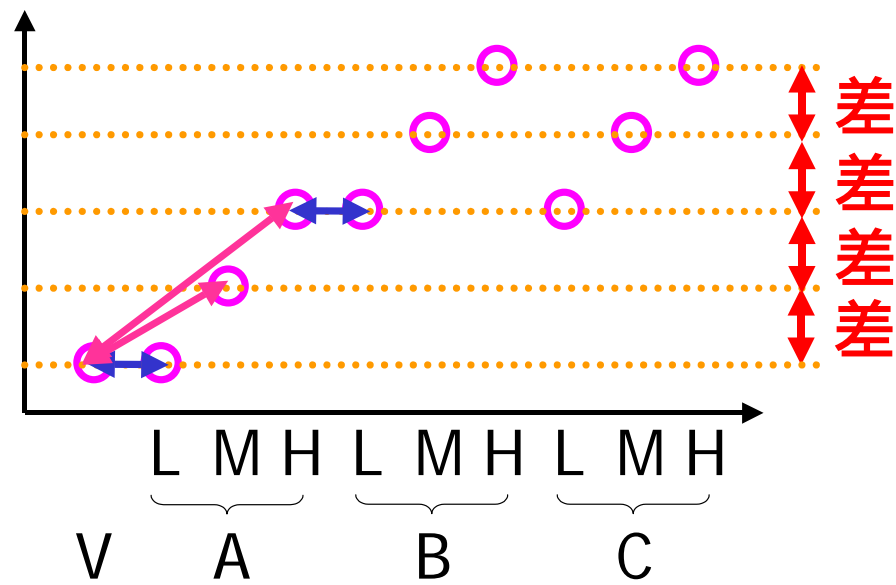
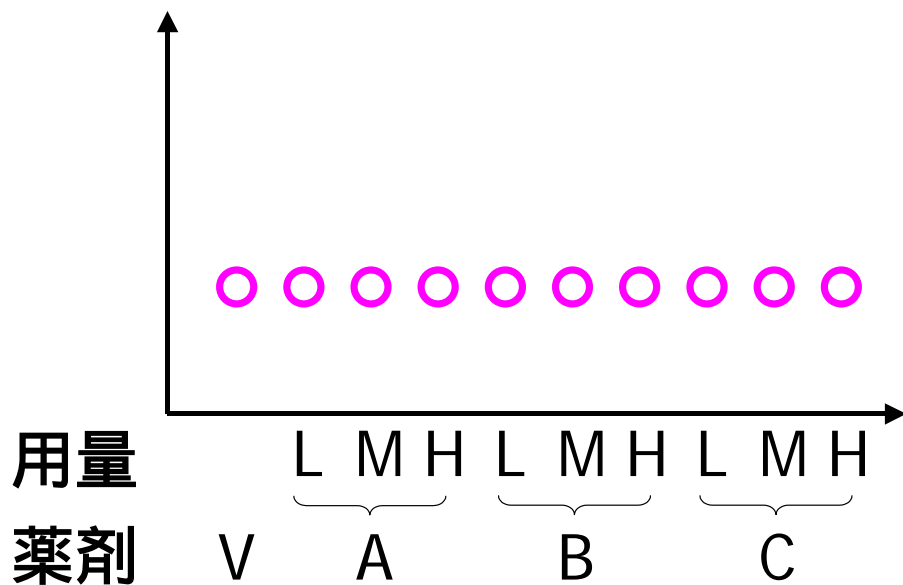
多群比較



想定 : すべての群間で差がない

想定 : 差が1

想定 : 差が1.5



1群8例 (共通分散)

t検定で比較

シミュレーション回数1000回

45個の比較ペア {
 ⇔ 差がある : 39個
 ⇔ 差がない : 6個



シミュレーション評価方法



「有意差検出数」の値を
とったくり返しの回数 / 1000回

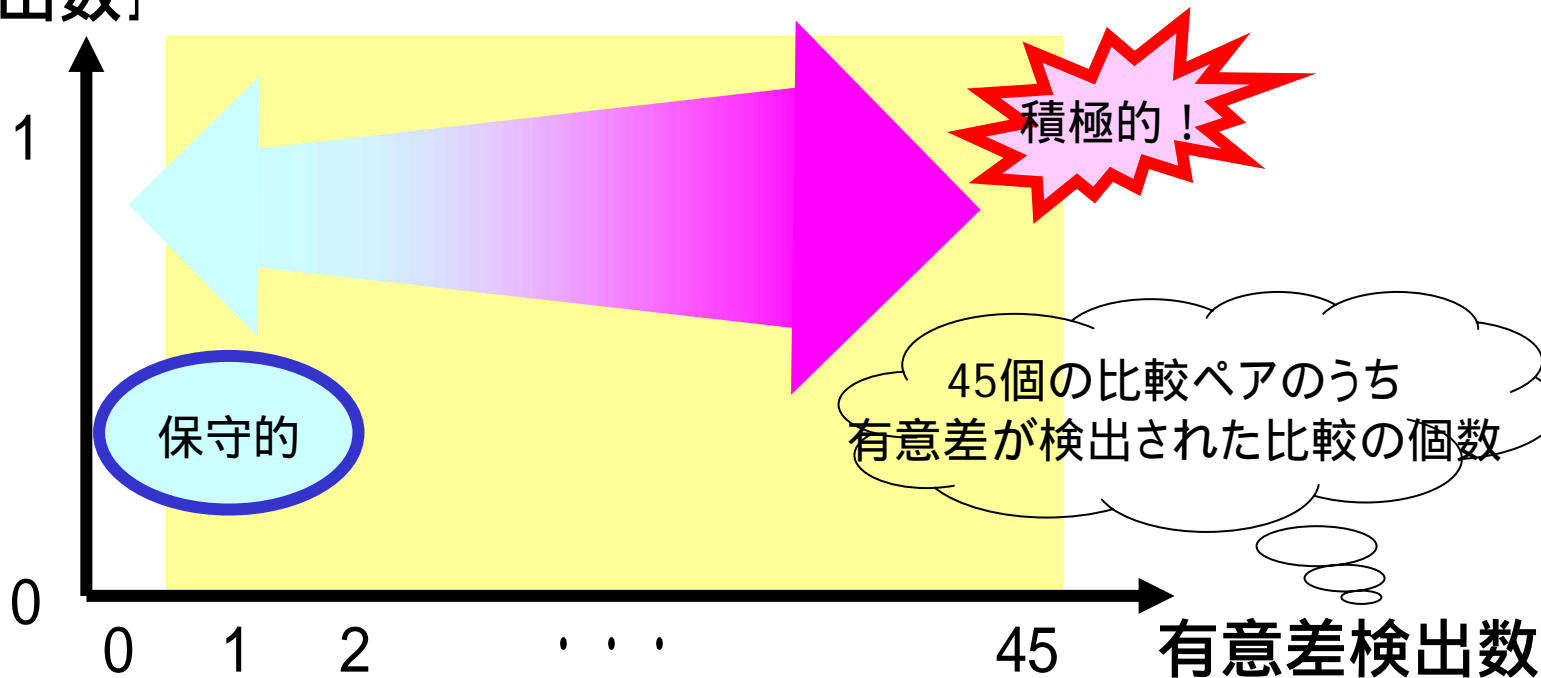
(多重性の)調整なし

FDR法

相関を考慮したFDR補正法

* Tukey法

「有意差検出数」
の頻度



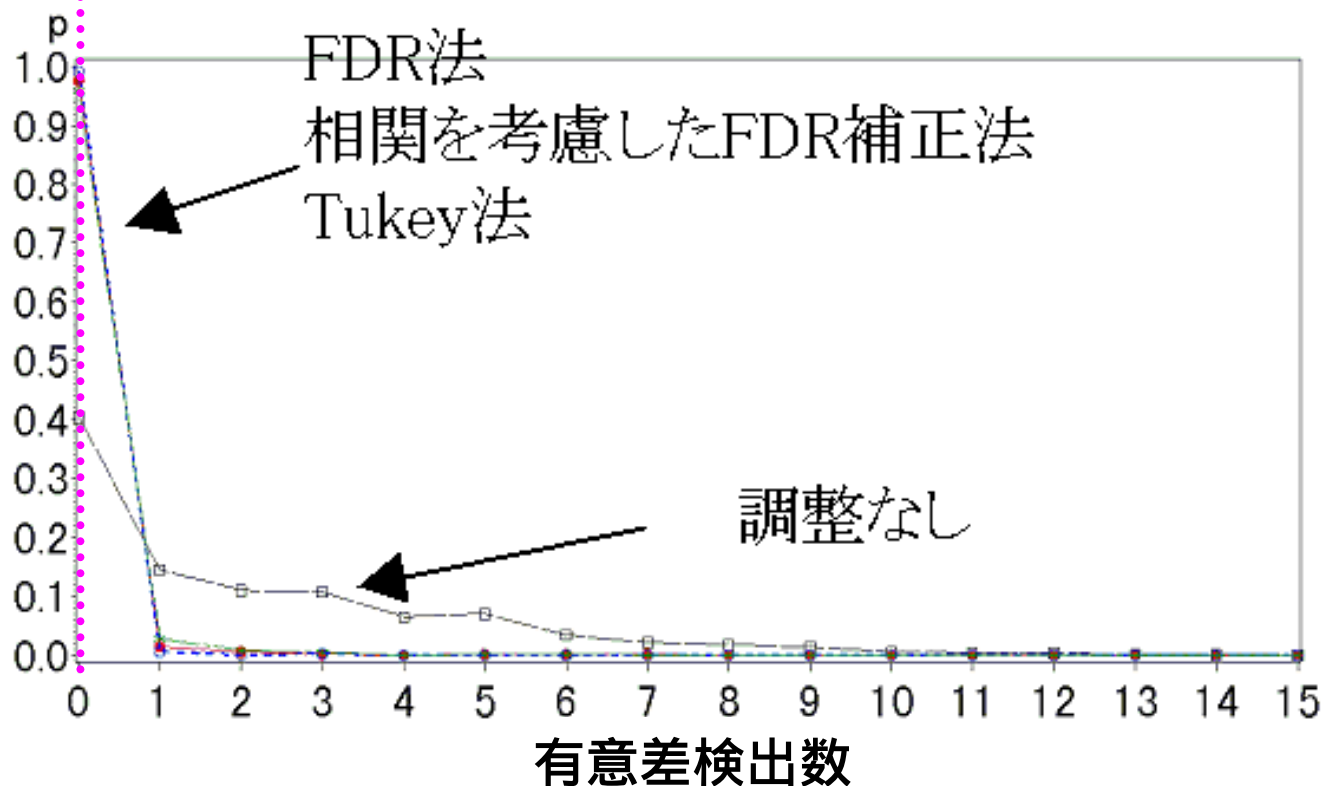
シミュレーション結果 - 想定 -

差なし



差があるペア

0個



method 1. not Adjusted 2. FDR
3. FDR with correlation 4. Tukey



シミュレーション結果 - 想定 -

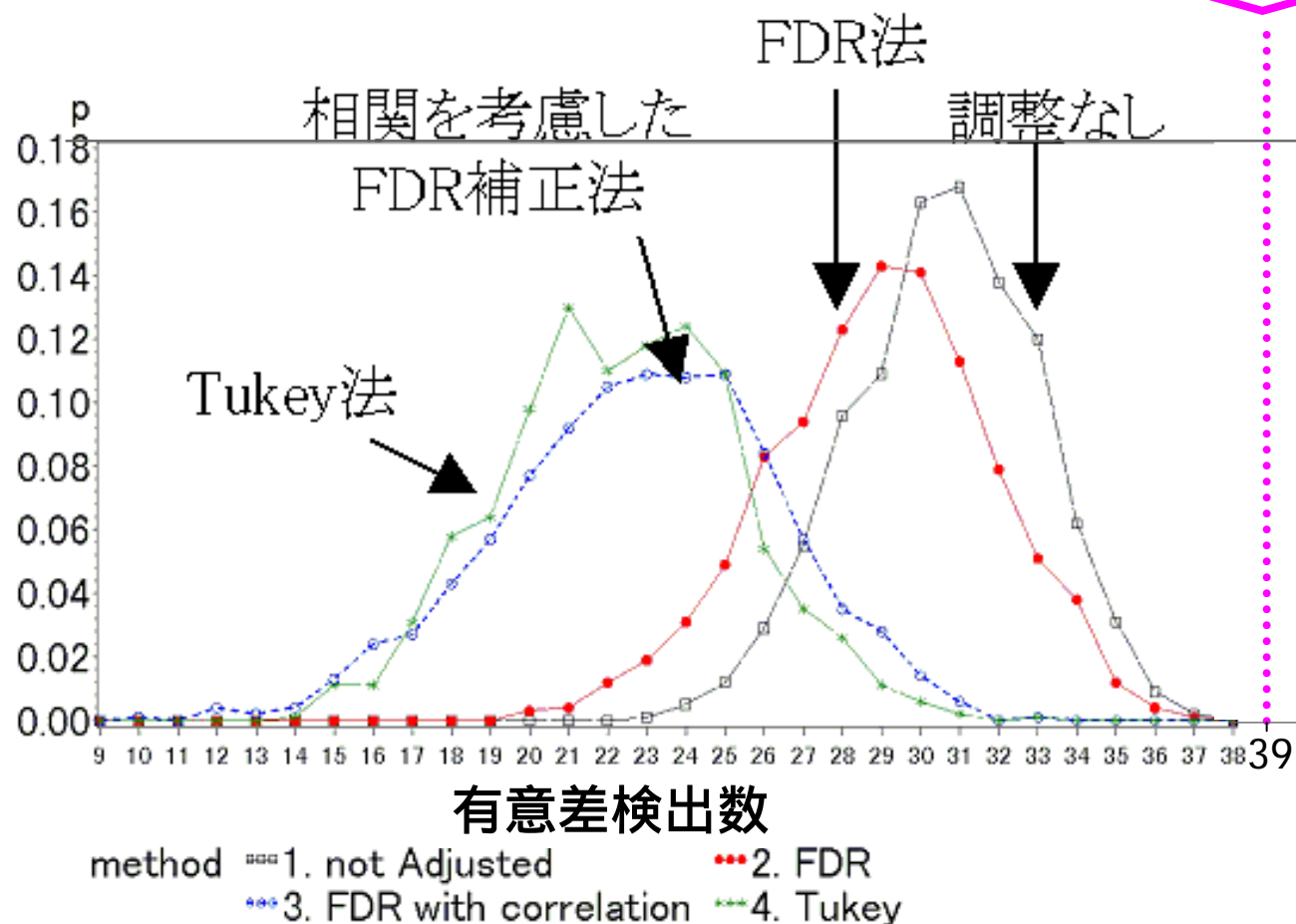


差1



差があるペア

39個



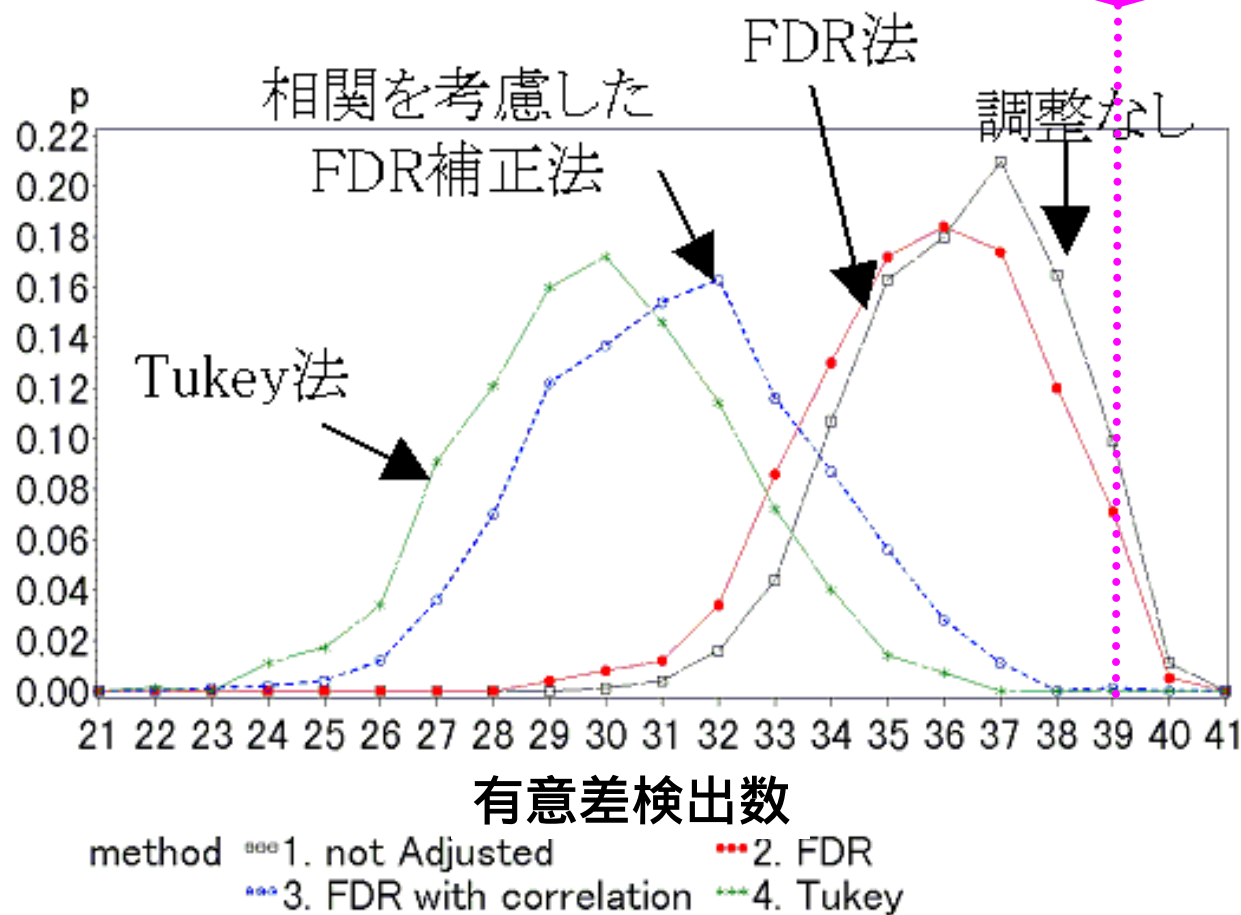
シミュレーション結果 - 想定 -

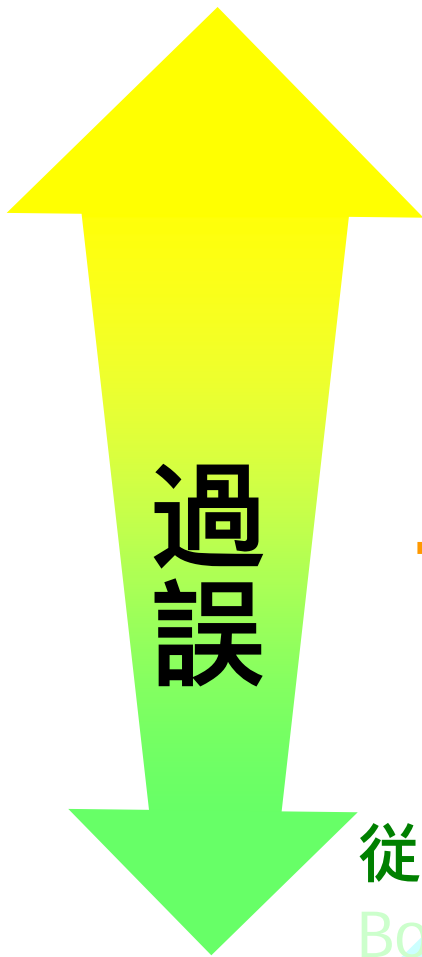


差1.5

差があるペア

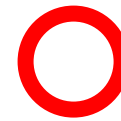
39個





□
多重性の
調整なし

相関を考慮したFDR補正法



従来の多重比較法

Bonferroni法, Tukey法

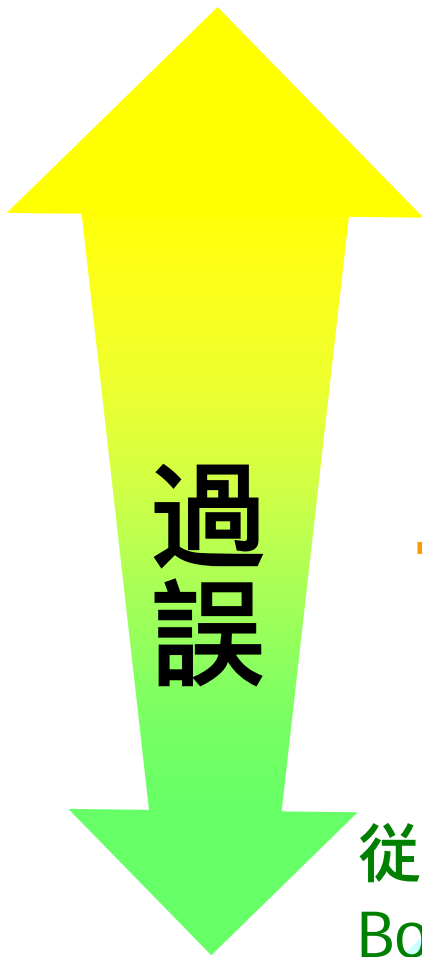
FDR法

保守的

検出力

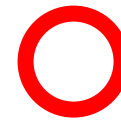
積極的！





□
多重性の
調整なし

相関を考慮したFDR補正法



従来の多重比較法
Bonferroni法, Tukey法

FDR法

保守的

検出力

積極的！





FDR法は、
魅力的な
多重調整法!!





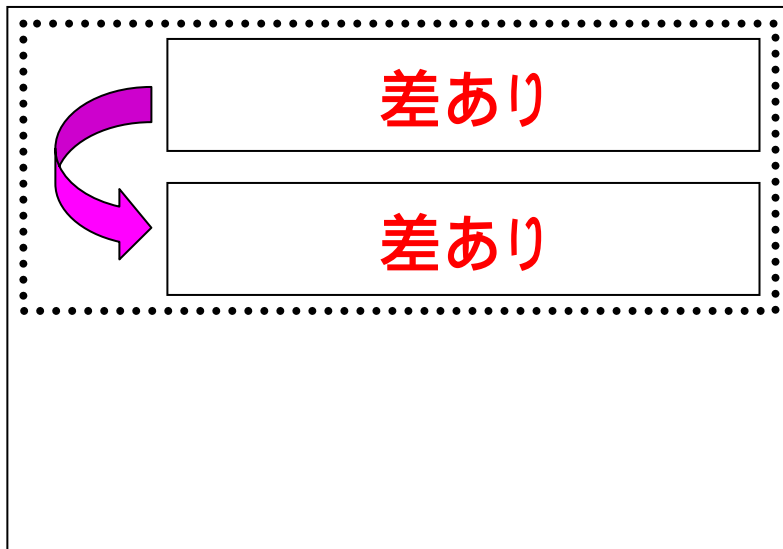
今後の課題

相関があるときの対処

有意差を検出したほうがいいのか、悪いのか

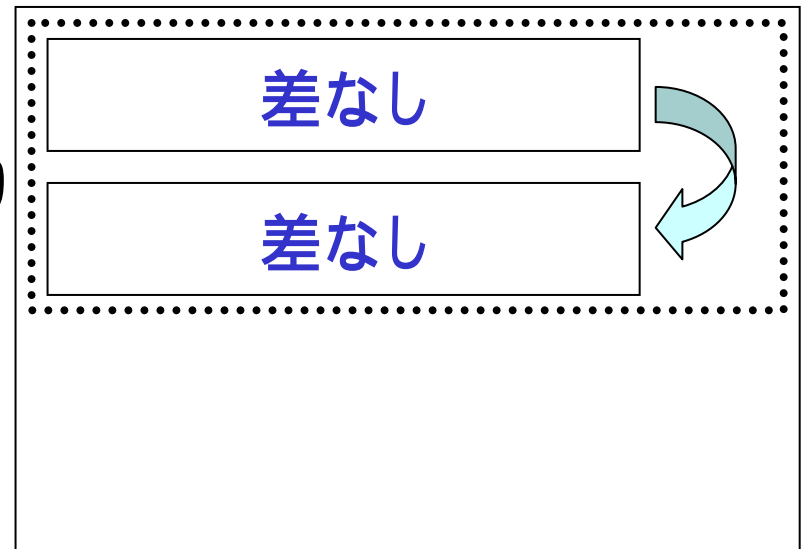


立場によって違うかもしれない



有意差を検出したい

相関あり



有意差を検出たくない

参考文献



- [1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, **57**, 289-300.
- [2] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165-1188.
- [3] Black, M. A. (2004). A note on the adaptive control of false discovery rates. *J. R. Statist. Soc. B*, **66**, 297-304.
- [4] Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 9440-9445.
- [5] Tzeng, J. Y., Byerley, W., Devlin, B., Roeder, K. and Wasserman, L. (2003). Outlier detection and false discovery rates for whole-genome DNA matching. *J. Amer. Statist. Assoc.*, **98**, 236-246.

