

**Статистика множественных сравнений в
ассоциативных исследованиях полиморфизма
ДНК:**

Кошмар Бонферрони

Рубанович А.В.

Институт общей генетики им. Н.И. Вавилова РАН

Опыты по выявлению людей с паранормальными способностями: Joseph Rhine (1950)

Вывод Д. Райна:

«нельзя сообщать подопытным людям
об их паранормальных способностях»

12 человек угадали 9 из 10 карт, двое все 10 карт

Все выявленные «экстрасенсы» в последующих опытах не
подтвердили свои способности

Что произошло?

Вероятность угадать все 10 карт = $\left(\frac{1}{2}\right)^{10} \approx 0.00098$

Вероятность угадать 9 карт = $10\left(\frac{1}{2}\right)^{10} \approx 0.0098$

Вероятность угадать 9 или все 10 карт = $11\left(\frac{1}{2}\right)^{10}$

1 - P(100 раз не найти)

Шансы найти «экстрасенса»

среди 100 человек = $1 - (1 - 0.0107)^{100} \approx 0.66$

Шансы найти «экстрасенса»

среди 1000 человек = $1 - (1 - 0.0107)^{1000} \approx 0.9998$

Точный тест Фишера с помощью монеты

Контроль:
1 дицентрик на 1000 клеток

против

Облучение:
9 дицентриков на 1000 клеток

Эти различия значимы
(в единичном тесте),

Control:
1 мутация на 100 здоровых

против

Case:
9 мутаций на 100 больных

...и не зависит от самих объемов выборок!!!

.0107

в 10 бросаниях монеты

но при проведении 100 тестов
вероятность появления
фальшивых отличий равна

$$p = 1 - (1 - 0.0107)^{100} \approx 0.66$$

Почему в середине 90-ых заговорили о проблеме «multiple comparisons» ?

- В ассоциативных исследованиях полиморфизма ДНК, как правило, фигурируют 3-5 (или 5-20) полиморфных локусов (это число еще нужно умножить на число

Проведение большого количества тестов
(множественных сравнений)
связано с опасностью фальшивых открытий

- Мета-исследования (исследования исследований): как объединять и сравнивать данные, полученные разными авторами

Генерируем две одноклассовые выборки

Наблюдаем соотношение

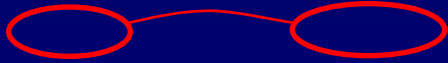
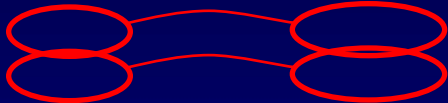
Odd Ratio – отношение шансов, популярная мера сопряженности. При отсутствии связи $OR=1$

Должно быть $OR=1$

Частоты минорных аллелей (в среднем 0.1)

Ген	Больные	Здоровые	OR	p
1	7	8		
2	10	2		
3	17	5		
4	13	12		
5	12	11		
6	7	10		
7	10	12		
8	14	9		
9	14	8		
10	9	12		
11	13	9		
12	9	14		
13	10	13		
14	8	8		
15	14	12		
16	17	7		
17	11	13		
18	10	8		
19	16	10		
20	12	8		

4



Значим

Сразу 3 локуса «ассоциированы» с заболеваемостью!

Как же избежать фальшивых открытий?

При проведении m независимых статистических тестов на уровне значимости α , вероятность хотя бы одного фальшивого результата должна быть

$$1 - (1 - \alpha)^m < 0.05$$



$$\alpha = 1 - (1 - 0.05)^{1/m} \approx \frac{0.05}{m}$$

Правило Карло Бонферрони (1935):

При проведение m независимых статистических тестов значимы только те результаты, для которых

$$p < \frac{0.05}{m}$$

Правило Бонферрони ликвидирует значимость вполне определенных результатов:

Две мутации, ассоциированные с заболеванием

	Контроль (100)	Больные (100)	OR	p
Мутация 1	1	7	7,45	0,0349
Мутация 2	5	15	3,6	0,0253

Однако правило Бонферрони требует:
 $p < 0,05/2=0,025$

1 против 7 при равных по объему выборках:

$$p = \binom{1}{2}^8 + 8 \binom{1}{2}^8 = 9 \binom{1}{2}^8 = \frac{9}{256} \approx 0.035$$

Правило Бонферрони ликвидирует значимость вполне определенных результатов:

Одна мутация + любой, не относящийся к делу признак

	Контроль (100)	Больные (100)	OR	p
Мутация 1	1	7	7,45	0,0349
Фамилия начинается с гласной буквы	36	40	1,19	0,646

Правило Бонферрони требует:
 $p < 0,05/2=0,025$

Увы, это не
шутка!!!

Assessment of individual sensitivity to ionizing radiation and DNA repair efficiency in a healthy population

F. Marcona, C. Andreoli, et al. Mut. Res., 541 (2003)

Table 1

Analysis of chromosome radiosensitivity by selected variables in 31 healthy subjects

Variable				P-value ^b
All				
Age (years) ^c				
<40				
40-44				
>44				0.990
Gender				
Male				
Female				0.953
Smoking habits ^d				
NS	19	1.9 ± 1.6	32.7 ± 8.8	
LS	6	1.6 ± 0.7	32.0 ± 8.9	
HS	6	1.7 ± 1.5	24.3 ± 6.8	0.050
Occupational exposure ^e				
Unexposed	17	1.8 ± 1.2	29.6 ± 9.6	
Exposed		1.8 ± 1.7	32.6 ± 7.9	0.246
Genotype				
GSTM1-positive	20	1.7 ± 1.3	33.4 ± 9.3	
GSTM1-null	11	2.0 ± 1.5	26.4 ± 6.0	0.025
GSTT1-positive	25	2.0 ± 1.4	32.2 ± 9.0	
GSTT1-null	6	0.9 ± 0.9	25.7 ± 6.8	0.174
NQO1-positive	19	1.8 ± 1.6	31.0 ± 7.8	
NQO1 mut	11	1.8 ± 1.0	29.3 ± 9.7	
NQO1 double mut	1	1.0	48.0	0.251

Незначимо!
 Поправка Бонферрони требует:

$$p < \frac{0,05}{3} \approx 0.0167$$

ГЕНОТИПЫ

High-Throughput Detection of GST Polymorphic Alleles in a Pediatric Cancer Population

P. Barnette, R. Scholl, et al. *Cancer Epidemiology, Biomarkers & Prevention* Vol. 13, 304–313, 2004

Контроль

13 генотипов

8 видов заболеваний

OR=6,4
P=0,007

Гомозиготность по делециям GST защищает от рака!

OR=2,3
P=0,018

Незначимо!
Поправка Бонферрони требует:

$$p < \frac{0,05}{13 \cdot 8} \approx 0.0005$$

	All patients		ALL		Ewing's Sarcoma		Glial brain tumors		Medulloblastoma		Neuroblastoma		Osteosarcoma		Retinoblastoma		Rhabdomyosarcoma		
	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	
GSTM1 genotypes																			
<i>GSTM1</i> *0/0	183	56	48	51	6	46	8	47	8	57	3	30	2	17					64
<i>GSTM1</i> *A/unknown	131	41	30	32	4	31	8	47	3	21	3	30	5	42					18
<i>GSTM1</i> *A/B	2	0	2	2	2	15	0	0	0	0	1	10	0	0	0	0	0	0	0
<i>GSTM1</i> *B/unknown	14	4	14	15	1	8	1	6	2	21	2	20	5	42	2	25	2	14	18
Total	300	100	85	100	12	100	13	100	9	100	9	100	12	100	7	100	14	100	
GSTT1 genotypes																			
<i>GSTT1</i> *0/0	66	22	24	15	9	11		7	3	25	4	44	2	17	0	0	2	14	14
<i>GSTT1</i> */unknown	234	78	141	85	72	89		93	9	75	5	56	10	83	7	100	12	86	86
Total	300	100	165	100	81	100	12	100	15	100	9	100	12	100	7	100	14	100	
GSTP1 genotypes																			
<i>GSTP1</i> *A/A	122	42											42	3	38	5	31	31	
<i>GSTP1</i> *A/B	102	35											33	3	38	6	38	38	
<i>GSTP1</i> *A/C	30	10											0	1	12	3	19	19	
<i>GSTP1</i> *A/D	0	0											0	0	0	0	0	0	
<i>GSTP1</i> *B/B	22	8											25	1	12	0	0	0	
<i>GSTP1</i> *B/C	11	4											0	0	0	2	13	13	
<i>GSTP1</i> *C/C	1	0											0	0	0	0	0	0	
<i>GSTP1</i> *C/D	0	0											0	0	0	0	0	0	
Total	288	100											100	8	100	16	101	101	
GSTP1 allele frequencies																			
<i>GSTP1</i> *A	376	65											58	10	63	19	59	59	
<i>GSTP1</i> *B	157	27											42	5	31	8	25	25	
<i>GSTP1</i> *C	43	7.5	31	9	11	7	0	0	4	15	3	15	1	6	0	0	16	16	
<i>GSTP1</i> *D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Total	576	100	340	100	166	100	24	100	32	100	20	100	18	100	24	100	32	100	

Implication of Xenobiotic Metabolizing Enzyme gene (CYP2E1, CYP2C19, CYP2D6, mEH and NAT2) Polymorphisms in Breast Carcinoma

A. Khedhaier, E. Hassen, et. al. *BMC Cancer*, 2008

Genotypes	Controls		All Patients		P	OR (95%CI)
	n	f	n	f		
CYP2E1 (C-1091T)	244		304			
CYP2E1 (C/C)	232	0.95	296	0.97	N.S	1.91(0.72–5.22)
CYP2E1 (C/T)	12	0.05	08	0.03	N.S	0.52(0.19–1.4)
CYP2E1 (T/T)	00	0.00	00	0.00	N.S	-
CYP2C19 (exon5 G-A)	240		304			
CYP2C19 (G/G)	197	0.822	239	0.787	N.S	0.8(0.51–1.26)
CYP2C19 (A/G)	41	0.170	61	0.200	N.S	1.23(0.77–1.97)
CYP2C19 (A/A)	02	0.008	04	0.013	N.S	1.5(0.33–6.8)
CYP2D6 (G1934A)						
CYP2D6 (G/G)	237	0.971	297	0.977	N.S	1.0(0.67–1.49)
CYP2D6 (G/A)	56	0.243	58	0.191	N.S	0.77(0.41–1.44)
CYP2D6 (A/A)	07	0.031	07	0.023	N.S	0.73(0.25–2.07)
mEH (T348C)	244		306			
mEH1 (Tyr/Tyr)	113	0.463	149	0.486	N.S	1.1(0.77–1.56)
mEH2 (Tyr/His)	115	0.471	119	0.388	N.S	0.71(0.5–1.02)
mEH3 (His/His)	16	0.065	38	0.124	0.02	2.02(1.06–3.89)
NAT2	237		290			
Rapid Acetylators	14	0.059	24	0.082	N.S	1.44(0.69–3)
Intermediate Acetylators	105	0.443	92	0.317	N.S	0.58(0.4–0.85)
Slow Acetylators	118	0.497	174	0.600	0.01	1.51(1.05–2.17)

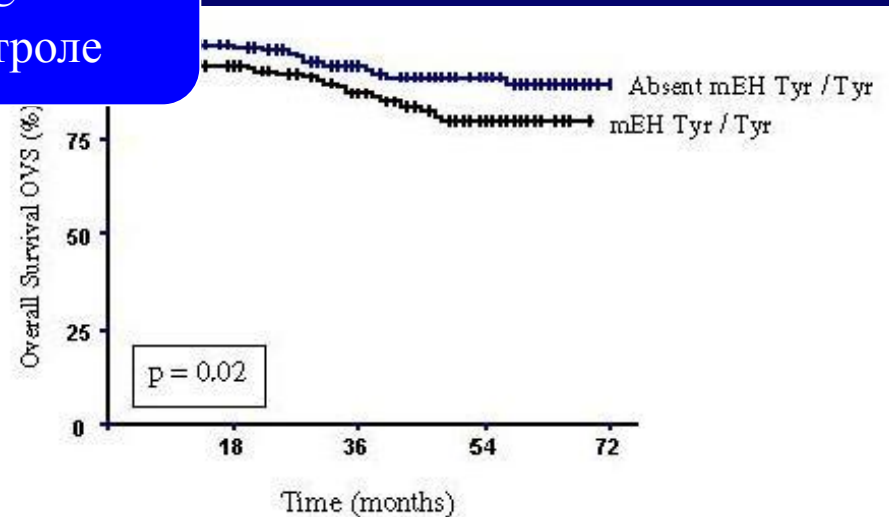
microsomal Epoxide Hydroxylase
Минорный аллель (C)
обуславливает активную

У больных
гомозигот CC
против 6% в контроле

Все стройно,
но Бонферрони требует:

$$p < \frac{0,05}{14} \approx 0.004$$

При этом среднее время выживания гомозигот CC выше, чем у больных с другими генотипами (повышенная эффективность химиотерапии?)



Bonferroni method creates more problems than it solves

Аргументы Томаса Пернежера (Thomas Perneger, 1998):

- Интерпретация данных зависит от числа тестов. Это противно нашей интуиции. Данные не могут терять значимость от того, что их кто-то подтвердил!
- При большом количестве тестов гипотеза о том, что все наблюдаемые различия неслучайны, никому не нужна
- При коррекции Бонферрони вероятность упустить существенные различия столь велика, что
- ...лучше просто перечислить какие тесты дали значимые результаты и, главное, почему



“Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference...”

Из разговоров на форуме molbiol.ru:

- ...мы не будем гробить свои результаты из-за какого-то там Бонферрони
- ...спрашивать диссертанта о Бонферрони – это дурной тон
- ...«бонферронофобия» набирает обороты
- ...разработчики программ не желают вводить кнопочку «Bonferroni»
- ...хуже Бонферрони ничего нет, не считая отсутствия всякой коррекции

Чуть-чуть об ошибках статистических тестов

Нулевая
различия

Традиционно биолог ориентирован на контроль
ошибки I рода (через уровень значимости),
т.е. на гарантии отсутствия ложных открытий,

ии
ости

Ошибка I рода (α)

Вероятность отвергнуть правильную нулевую гипотезу =
Вероятность обнаружить различия там, где их нет = **Вероятность совершить фальшивое открытие**



Ошибка II рода (β)

Вероятность принять неправильную нулевую гипотезу =
Вероятность не обнаружить существующие различия =
Вероятность упустить открытие



Мощно
Вероятн
Вероятн

... и при этом мало заботится о возможности
упустить открытие (ошибка II рода)

Мощность статистических тестов

- ❑ Мощность 80% считается приемлемой
(мощность обычных тестов в реальных ситуациях)
- ❑ Консервативный тест - это тест с низкой мощностью
(напр., критерий Колмогорова-Смирнова)
- ❑ **Метод Бонферрони – архиконсервативен**
(особенно при сравнении частот)
- ❑ Во многих случаях коррекция Бонферрони делает выявление значимых результатов попросту невозможным

От чего зависят ошибки статистических тестов?

❑ От размаха реально существующих отличий и разброса данных

❑ От объемов выборок

Ошибка I рода (вероятность фальшивого открытия)

слабо зависит от объемов выборок,
если они сравнимы по величине

(главная причина маломощности метода Бонферрони)

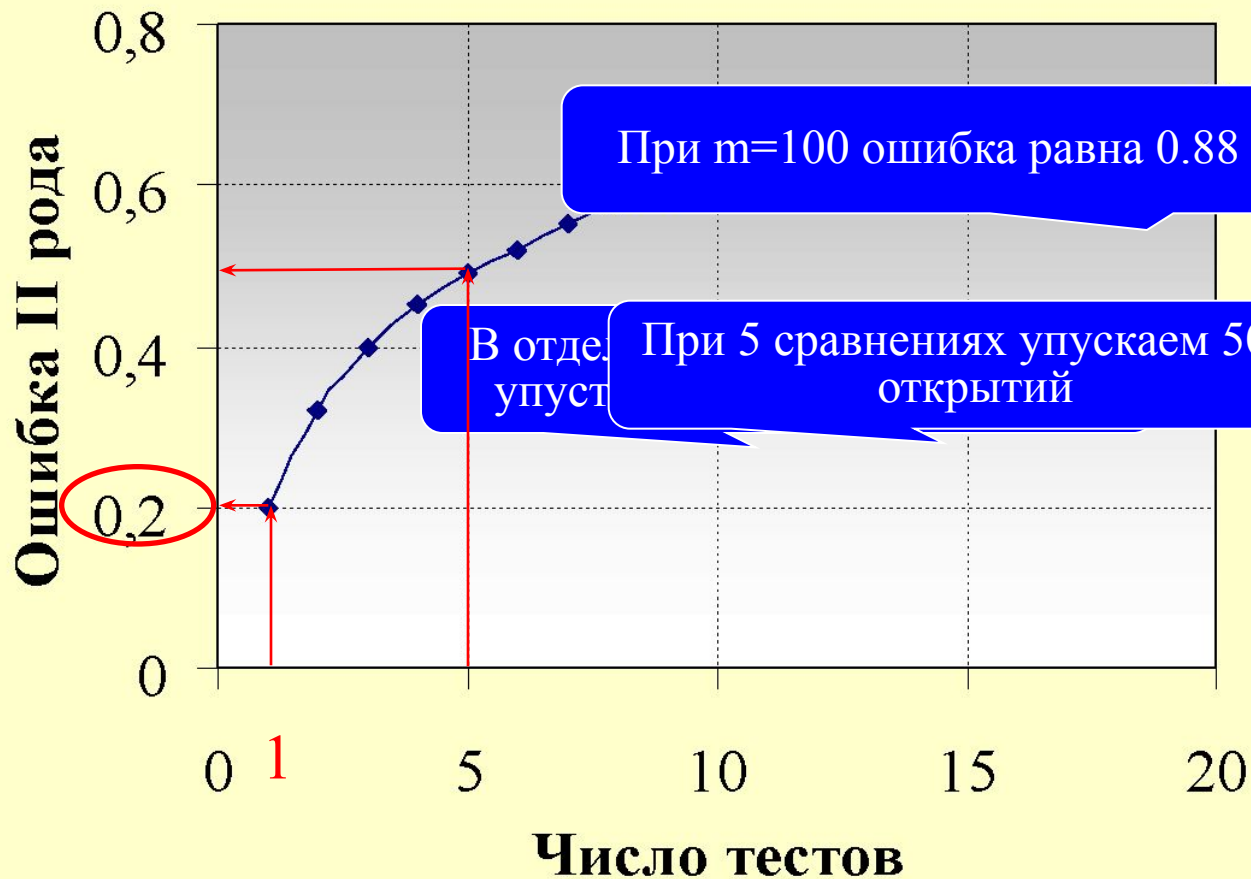
Крайний случай:

«критерий» св. Фомы Неверующего (0033)

Ошибка I рода = 0 \Leftrightarrow Ошибка II рода = 1

Зависимость ошибки II рода от числа тестов при

При 100 сравнениях ради того, чтобы гарантировать отсутствие хотя бы одного ложного результата, мы упускаем 88% открытий!



Новый принцип проверки статистических гипотез: FDR-контроль

False Discovery Rate control: Benjamini, Hochberg (1995)

Вероятность фальшивого открытия < Уровня значимости
Ошибка I рода < 0.05

Традиционный принцип
заменяется на



105 статей в



Средняя доля фальшивых открытий < Выбранный уровень

$$E\left(\frac{\text{Число неправильно отвергнутых нулевых гипотез}}{\text{Число отвергнутых нулевых гипотез}}\right) < 0.05$$

Алгоритм контроля FDR

(Benjamini, Hochberg, 1995)

- Упорядочиваем тесты по уровню p -величин:

$$p_1 \leq p_2 \leq \dots \leq p_m.$$

- Для контроля FDR на уровне α находим

Величина p для j -ого теста (гена)

Порядковый номер

Желательный уровень значимости

$$j^* = \max \left\{ j : p_j \leq \frac{j}{m} \alpha \right\}$$

- Считаем различия значимыми при $j \leq j^*$

Общее число тестов (генов)

- При $j > j^*$ различия считаются незначимыми

Пример: множественные сравнения по 10 тестам

Тест	p_i	Корр Bonf	
1	0,001	0,005	0,005
2	0,0055	0,005	0,010
3			0,015
4			0,020
5			0,025
6			0,030
7	0,3		
8			
9			
10	0,8	0,005	0,050

Располагаем тесты в порядке увеличения p_i

Значимые различия после коррекции по FDR

В первой клетке —
во второй клетке —
втрое больше
и т.д.

Поправка Бонферрони
оставляет значимым лишь
первое сравнение

И это все!!!

Для 6-ого
ЭТОГО значения

НОСТЬ

То же самое в общем случае:

m – число сравнений, $\alpha = 0.05$ (например)

	Bonferroni (1939)	FDR (1995)
$p_1 <$	α/m	α/m

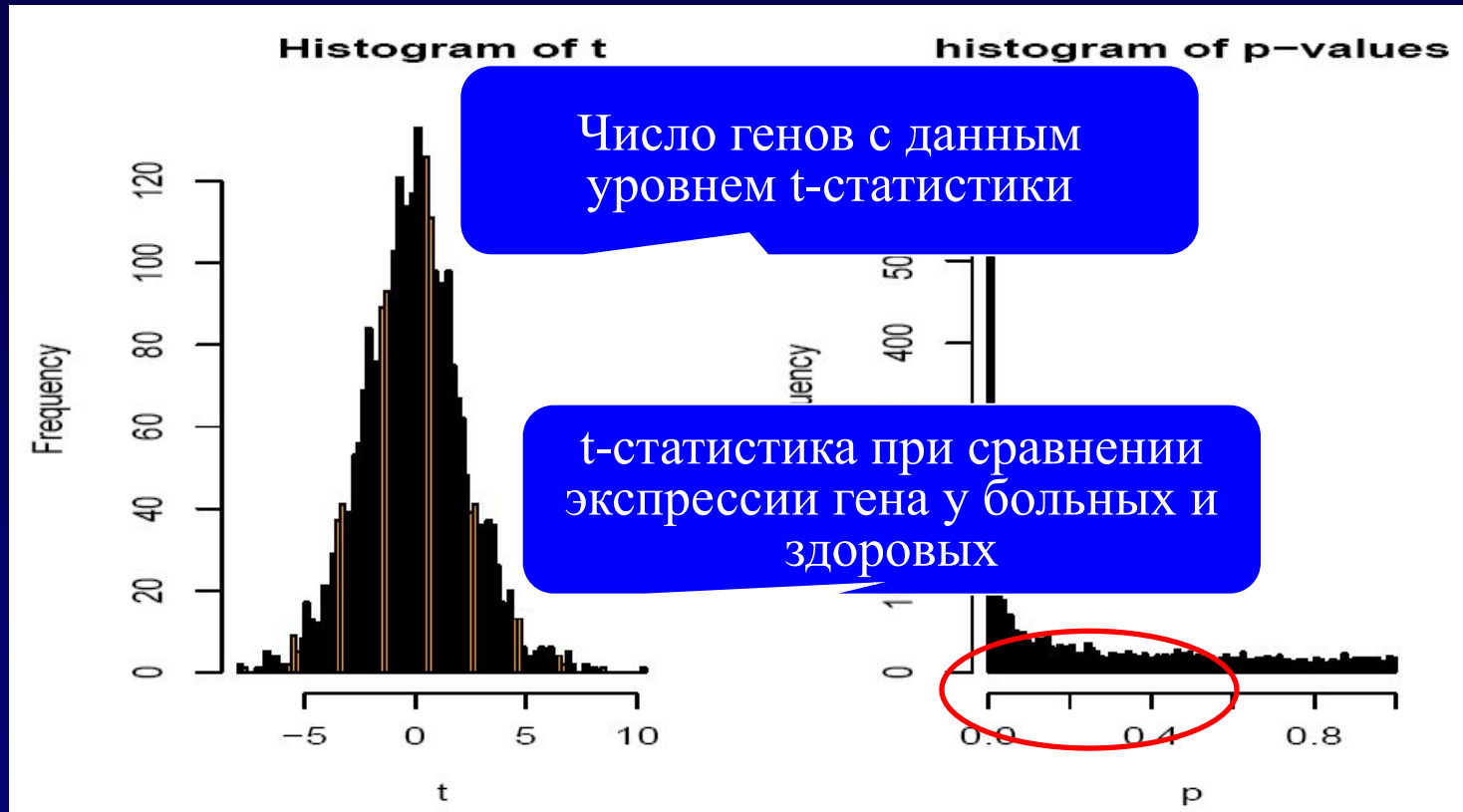
Бонферрони торжествует!

Контроль FDR приносит ощутимые результаты,
если хотя бы один тест удовлетворяет
правилу $p < \alpha/m$

p_i	α/m	α/m
...	α/m	...
p_m	α/m	α

Пример: экспрессия 3051 генов при острой лейкемии

Golub T.R. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. // Science. 2001, v.286.



- t-тест: 1045 генов, для которых $p < 0.05$
- Коррекция Бонферрони : 98 генов для $p' < 0.000016$
- FDR: 681 генов, для которых $FDR < 0.05$

Что делать, если FDR не помогает?

Permutation tests:

случайные перестановки пометок «case-control»
в компьютерных симуляциях по алгоритму:

- В исходной базе данных делаем случайную перестановку лейблов case-control

Точный тест Фишера – это тоже permutation test,
только реализованный аналитически (p
вычисляется
по формулам комбинаторной теории вероятностей)

- Вычисляем откорректированное p как

$$p' = \frac{\text{Число случаев } (p_{perm} \leq p)}{N}$$

Permutation test применительно к данным об ассоциации заболеваемости с 10 SNP

Переставляем отметки «case-control» 10000 раз. В результате получаем коррекцию p

SNP	Частота минорного аллеля		OR	p-value	
	Case (100)	Control (100)		Бонферрони	FDR
1	62	26	4,6	0,0001	0,000
2	31	19	3,7	0,009	0,010
3	21	13	2,8	0,011	0,007
4	20	8	2,9	0,023	0,025
5	19	11	3,0	0,071	0,109
6	17	10	2,0	0,096	0,098
7	44	30	1,8	0,103	0,058
8	54	39	1,8	0,120	0,067
9	59	53	1,3	0,571	0,476
10	40	41	1,0	0,911	1,000

Но так бывает не всегда

Значимо по Бонферрони

Значимо по FDR
коррекция на множественность

Почему Permutation test так либерален?

- ❑ Поправки Бонферрони и FDR предполагают, что все тесты независимы
- ❑ Перестановка лейблов «case-control» сохраняет корреляционные связи между данными (в случае полиморфных генов можно сказать: «учитывает неравновесие по сцеплению»)
- ❑ Учитывает совместные распределения тестовых статистик
- ❑ В результате **Permutation test** значительно менее консервативен, чем **Bonferroni** и **FDR**



Как реализовать permutation tests?

Не предусмотрен в стандартных статпакетах.

На сегодняшний день доступно:

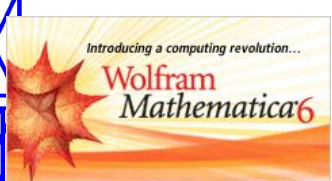
- ❑ **SNPAlyze™** free 30 days trial
<https://www.dynacom.co.jp/en/>

- ❑ **PowerMarker**
<http://statgen.ncsu.edu/powermarker/>

- ❑  Free software environment
встроенные функции
<http://www.r-project.org/index.html>

И все же всегда проще самому!

```
simNum = 10000;
suDif = Table[0, {Length@base}];
Do[RandomPermutation[volSample];
  tot = Join[one.health];
  ill = tot[[Take[1, volSample]]];
  health = tot[[Take[1, volSample]]];
  genCase1 = one.ill;
  genControl = one.health;
  xiSq = genCase1[[genControl]]^2 - 2*genCase1[[genControl]] + 1;
  suDif = suDif + UnitStep[p - p1] * xiSq;
simp = suDif / simNum; NA
```



Борьба с кошмаром Бонферрони продолжается!

- Нейронные сети, случайные деревья, генетические алгоритмы, гибсовские поля и еще Бог знает что
- Типичная ситуация для 3 локусов: 27 возможных генотипов, но лишь 5-7 гаплотипов встречаются с частотой $>1\%$ **аняется!**
- Если число локусов ≥ 3 , то обычно число наблюдаемых гаплотипов \ll числа генотипов

Soft для работы с гаплотипами



<http://www.bios.unc.edu/~lin/hapstat/>
hapstat
statistical analysis of haplotype-disease association



ICO
Institut Català d'Oncologia
Hospital Duran i Reynals

SNPStats: Your web tool for SNP analysis.

X. Sole, E. Guino, J. Valls, R. Iniesta¹, V. Moreno (2006)

<http://bioinfo.iconcologia.net/index.php?module=Snpstats>

Однолокусный анализ

- Allele and genotype frequencies
- Test for Hardy-Weinberg equilibrium
- Analysis of association with a response variable based on linear or logistic regression
- Multiple inheritance models: co-dominant, dominant, recessive, over-dominant and additive
- Analysis of interactions (gene-gene or gene-environment)

Многолокусный анализ

- Linkage disequilibrium statistics
- Haplotype frequency estimation
- Analysis of association of haplotypes with the response
- Analysis of interactions (haplotypes-covariate)

Index

Descriptive statistics

Single-SNP analysis

snp1
snp2

Multiple-SNP analysis

Linkage disequilibrium analysis
Haplotype analysis

Descriptive statistics

Response variable: group Type: ca

All subjects: 212

	group=C	group=Co	OR (95% CI)
A/G Female	38	48	1.00
Male	53	43	0.64 (0.36-1.15)

	group=C	group=Co	OR (95% CI)
A/G Female	6	4	1.00
Male	5	6	1.80 (0.32-10.0)

Percentage of typed samples: 205/212 (96.7%)

Haplotype association with response (n=164, crude analysis)

	<i>COMT</i>	<i>GSTP1</i>	<i>MTHFR</i>	<i>Freq</i>	<i>OR (95% CI)</i>	<i>P-value</i>
1	L	A	C	0.2719	1	---
2	H	A	C	0.1996	2.69 (1.02 - 7.12)	0.048
3						.076
4						.062
5	H	A	T	0.099	2.08 (0.77 - 5.60)	0.15
6	L	G	T	0.0906	1.50 (0.49 - 4.62)	0.48
7	L	A	T	0.0773	3.24 (0.67 - 15.68)	0.15
8	H	G	T	0.0402	13.88 (1.21 - 158.65)	0.036

По отдельности локусы не обнаруживают значимых ассоциаций!

Global haplotype association p-value: **0.033**

Почему не следует пренебрегать коррекцией на множественность сравнений?

- ❑ Чтобы не уподобиться старому Джозефу
- ❑ Чтобы преодолеть соблазн раздувать отчет о проделанной работе
- ❑ Чтобы предупредить выпадения недоброжелателей
- ❑ Это просто и делается вручную.

Если хотя бы одно сравнение проходит через сито Бонферрони, то с Вас хватит **FDR**. В противном случае Вам не обойтись без перестановочного компьютерного теста.



Спасибо
Оргкомитету школы
и Всем присутствующим !!!

*С удовольствием предоставляю копию презентации
заинтересованными коллегам*

rubanovich@vigg.ru

Bonferroni, FWER, FDR и все такое

- Bonferroni контролирует **FWER** (**f**amily-**w**ise **e**rror **r**ate), т.е., вероятность хотя бы одного фальшивого открытия ассоциации гена с заболеваемостью
- **FDR** – это контроль средней доли фальш-ассоциированных генов среди всех генов, для которых отвергнута нулевая гипотеза.

Association between Common Variation in 120 Candidate Genes and Breast Cancer Risk
P. Pharoah, J. Tyrer¹ et al. PLoS Genetics, 2007.

710 SNP (120 genes), Case - 4400, Control – 4400.

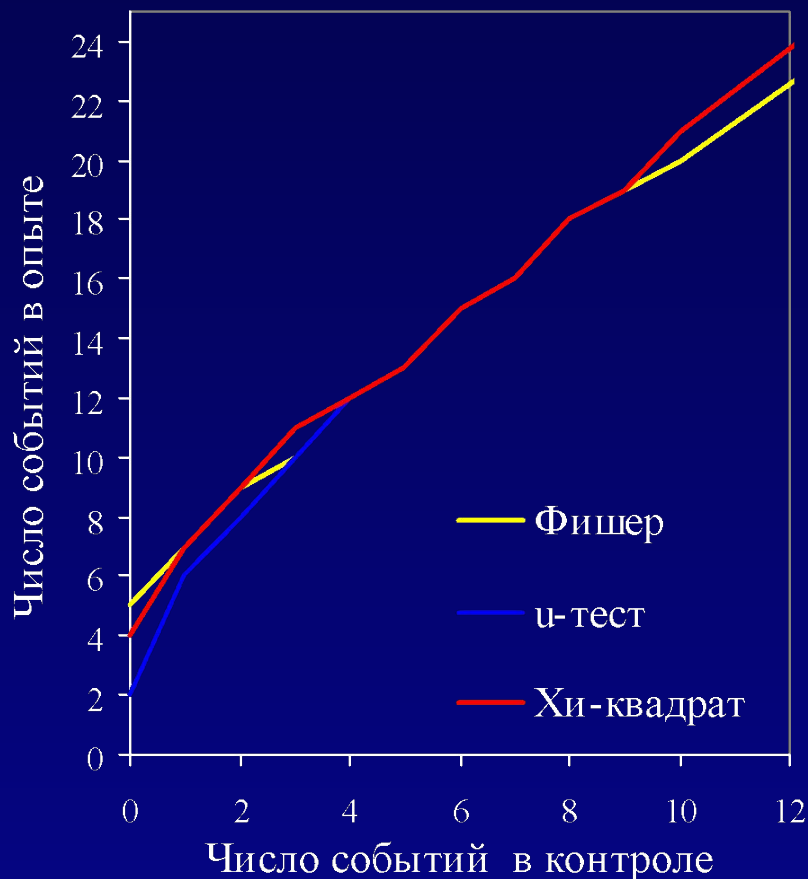
Слабые ассоциации:

- 104 SNP в 8 генах метаболизма стероидных гормонов
- 112 SNP в 18 генах контроля клеточного цикла

- Изменчивость признаков не определяется редкими мутациями отдельных генов, но зависит от большого числа SNP (со слабыми индивидуальными эффектами)
- Ассоциативные исследования полиморфизма ДНК с необходимостью должны быть многолокусными

Сравнение частот при уровне значимости 0.05

Объемы выборок в опыте и контроле одинаковы



Число событий в контроле	Минимальное число событий в опыте при значимом отличии от контроля		
	u-тест	χ^2	Фишер
0	2	4	5
1	6	7	7
2	8	9	9

больше 5 независимо от того, сколько клеток Вы просчитали – 100 или 1000

6	15	15	15
7	16	16	16
8	18	18	18
9	19	19	19
10	21	21	20
20	35	35	33
30	47	47	46