

Мы секвенировали небольшой участок с какой-то хромосомы человека и получили следующую последовательность. Каждые две последовательности для каждого человека соответствуют паре гомологичных хромосом.

Человек 1

ATCAATTAATGTATTAGCA
ATCAATTAATGTATTAGCA

Человек 2

ATCAATTAATGTATTAGCA
TTCAATTAATGAATTAGCA

Человек 3

ATCAATAAATCTATTAGCA
TTCAATTAATGTATTAGCA

Человек 4

ATCAATAAATGTATTAGCA
ATCAATTAATGTATTAGCA

1. Отметьте локусы, содержащие SNP в наборе ДНК-последовательностей. (перекопируйте последовательности в текстовый редактор и далее работайте с ними в текстовом редакторе)

2. Мы секвенировали одну и ту же геномную область нового человека и получили следующую последовательность. Мы обнаружили новый SNP? Если да, то где он?

Человек 5

ATCAATTAATGTATTTGCA
ATCAATAAATCTATTAGCT

3. Какое число аллелей в каждом локусе?

4. Исходя из данных последовательностей 5 людей выше, какова частота минорной аллели в каждом SNP?

5. Исходя из списка SNPs, обнаруженных в результате секвенирования геномной области 5 людей, представленных выше, какой генотип человека 3 в каждом из геномных локусов?

6) Переведите последовательности букв А, С, Т, и G в последовательности 0 и 1, используя 1 для обозначения минорной аллели и 0 для обозначения мажорной аллели для каждого SNP.

7) Используя конвертированные данные из 0 и 1, посчитайте корреляцию (r) между первым и вторым локусом SNP.

Коэффициент корреляции для бинарных данных есть коэффициент сопряженности Phi

Обозначим ячейки бинарной таблицы:

		Признак X		Итого
		0	1	
Признак Y	0	a	b	a+b
	1	c	d	c+d
Итого		a+c	b+d	a+b+c+d

Тогда формула коэффициента имеет вид:

$$\varphi = \frac{a * d - b * c}{\sqrt{(a + b) * (c + d) * (a + c) * (b + d)}}$$

8) В исследованиях GWAS случай-контроль, для одного из генетических локусов были получены следующие результаты

	Случай (Case)	Контроль (Control)
Число минорных аллелей (a)	4	2
Число мажорных аллелей (A)	102	99

Какова частота минорной аллели среди всех генотипированных людей?

Эта аллель - общий вариант, низко-частотный вариант или редкий вариант?

Посчитайте статистику хи-квадрат и p-value.

Онлайн калькулятор

<http://medstatistic.ru/calculators/calchi.html>

```
можно в python'e
from scipy.stats import chi2_contingency
stat, p, dof, expected = chi2_contingency([[4, 102], [2, 99]],
correction=False)
print(p)
```

```
#точный тест Фишера
from scipy import stats
table=[[2,99],[4,102]]
oddsratio, pvalue=stats.fisher_exact(table)
print(pvalue)
print(oddsratio)
```

Исходя из значения p-value, вы считаете, что этот локус статистически значимо ассоциируется с заболеванием?

GWAS_catalog

<https://www.ebi.ac.uk/gwas/diagram#>

https://en.wikipedia.org/wiki/GWAS_catalog

Можно по нему походить, если останется время...