

Кондратюк Александр Николаевич

Группа_3

Организм: Chicken 2006

Хромосома: chr27

1. Открыл хромосому и скачал ее в Colab:

```
✓ 1
cek. !wget https://hgdownload.soe.ucsc.edu/goldenPath/galGal3/chromosomes/chr27.fa.gz

--2023-06-05 14:48:32-- https://hgdownload.soe.ucsc.edu/goldenPath/galGal3/chromosomes/chr27.fa.gz
Resolving hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)... 128.114.119.163
Connecting to hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)|128.114.119.163|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1415075 (1.3M) [application/x-gzip]
Saving to: 'chr27.fa.gz'

chr27.fa.gz          100%[=====] 1.35M  3.17MB/s   in 0.4s

2023-06-05 14:48:33 (3.17 MB/s) - 'chr27.fa.gz' saved [1415075/1415075]

✓ [47] !gunzip chr27.fa.gz
```

2. Проверил наличие материала в хромосоме chr27:

```
!head -2000 chr27.fa

AAGGGAGGGGGGTAATCGTTGAGTAGGTGTGTCTTCCATTCAGCAGGAG
CTGCACATGATGGGATTCTTAGTTCTTGTAGACTAAAGCCCTGACAGATC
CTTGGATACTTAAAGCTGTTGTGATGCTTTACCACATGGTTCAGAACTC
AGACTGGAATATCTCATTGGCCAGTGCTGAACAATGACATTTCAAACCCA
CACGTTCAATTGGGATTTAATTAATTTGGGATTGGATTAAGCTTTTTGGA
ATATGGGCTGTCTCTTTCCATATAGGTGCTCGGCACCTTATGTAGTTGAG
CTTTGCGACATCTTAATAATCCACTGTTGACTGTGTTACTTCTGAAAGAT
CGATTTGAGCAGTCACACAGTAGTAAGTGCCTACGTGGACCAGCAAAGCA
GGGATGGGATCGGTAAACctgatgcaggtaatgcctctcctgggcgctca
ttggttgaatgtctcaagctcacacacttgccgacactgttgctacggt
acagctccacataaggactgttacctgtatgttgctccttaggtatgcttt
gAAATATCTCCCATATTTCAAAGATAGCCCATCCGTTCTGCTTTGTGCTT
GGATATGAGGGCAATATGCCAAAGTTAGCACATCCGTCCTGTGCTGTGCT
TGGCTCTCTTTCTTTACGAATCAATAACATATATAACTGTTTCACAAGAG
AGTCATAAGTTCAGTGAGGGCCTAGGCTACATACAAAAAGAGAGGTATAT
ACTGGAACAGCTGTTTTCACAGGAAGGTTACAGTTTCAGCAAGGGAGAac
```

3. Загрузил файл zhunt3-alan.c и добавил права на исполнение.

```
[5] !gcc zhunt3-alan.c -lm -o zhunt3
```

```
!chmod a+x zhunt3
!./zhunt3 12 8 12 chr27.fa
```

```

dinucleotides 12
min/max 8 12
min/max 8 12
operating on chr27.fa
calculating zscore
opening chr27.fa
inputting sequence
opening chr27.fa.Z-SCORE

run time=1161 sec
use min/max 8 12
analyzing_zscore
opening chr27.fa.Z-SCORE
opening chr27.fa
inputting sequence

```

4. Посчитал количество участков Z-ДНК:

```

[10] cut_file = open('chr27.fa.Z-SCORE.cut', 'w')

z_dna_counter = 0
with open('chr27.fa.Z-SCORE', 'r') as f:
    first_line = True
    for line in f:
        if not first_line:
            if float(line.split()[5]) >= 300:
                cut_file.write(line.split()[0] + '\t' + line.split()[1] + '\t' + line.split()[5] + '\n')
                z_dna_counter += 1
            else:
                first_line = False

cut_file.close()

print(z_dna_counter)

19108

```

5. Посчитал количество квадруплексов:

```

!pip install biopython
from Bio import SeqIO
import re

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting biopython
  Downloading biopython-1.79-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (2.3 MB)
    |#####| 2.3 MB 14.4 MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from biopython) (1.21.6)
Installing collected packages: biopython
Successfully installed biopython-1.79

```

```

[12] input_file="chr27.fa"
    fasta_sequence = SeqIO.parse(input_file, 'fasta')
    for record in SeqIO.parse(input_file, "fasta"):
        print("%s %i" % (record.id, len(record)))
        name, sequence = record.id, str(record.seq)
        pattern="(?:G{3,}[ATGC]{1,7}){3,}G{3,}"
        PQS=[[m.start(),m.end(),m.group(0)] for m in re.finditer(pattern,sequence)]
        len(PQS)

chr27 4841970
646

```

PQS

```
[4512337, 4512368, 'GGGCAGCACTGGGGCAGGATGGGGCCCGGG'],
[4516251, 4516275, 'GGGGGTCCCCAGGGTGGGAAAGGG'],
[4516304, 4516336, 'GGGGGGATGGGATGGGGAACAGCAGGGCTGGG'],
[4522451, 4522470, 'GGGGCATGGGAGGGAAGGG'],
[4523666, 4523681, 'GGGGGGGGGGGGGGG'],
[4527349, 4527373, 'GGGACGGGCCGAGGGGAAGAGGG'],
[4527454, 4527483, 'GGGGCCCCCGCGGGGTCTTAGGGGCCGGG'],
[4528937, 4528958, 'GGGTGAGGGTGGGGGCCAGGG'],
[4532554, 4532587, 'GGGAGAGCTGGGAGAACTGGGCTGGGGATGGGG'],
[4535178, 4535208, 'GGGGGACAATGGGGACAGCGTGGGGAGGGG'],
[4542903, 4542937, 'GGGTGTGGGGAGATGGGGCTGCGGGGCATGGGG'],
[4545429, 4545454, 'GGGCACAGGGTGGGCACAGAGGGGG'],
[4546666, 4546685, 'GGGTGGGTCAGGGATGGGG'],
[4562190, 4562217, 'GGGCTCCAGGGTGGGTTGGGAACAGGG'],
[4565580, 4565614, 'GGGAGCGCGGGGAGCCACGGGCGGGGGCTCCGGG'],
[4565831, 4565854, 'GGGTCCCGGGACGGGGGCGGGGG'],
[4567268, 4567305, 'GGGCACGGGTCAGGGCCGTGGGGTGGGGAGGGCTGGG'],
[4567651, 4567672, 'GGGAAGTGGGAAATGGGAGGG'],
[4613713, 4613732, 'GGGTGGGGAGACGGGAGGG'],
[4614713, 4614736, 'GGGCTGGGGCATGGGAGCTCGGG'],
[4634577, 4634597, 'GGGCAGTGGGAGGGCAAGGG'],
```

6. Составил список генов по файлу “alignment chicken”

```
[16] genes = []
      with open('alignment chicken', 'r') as f:
          for line in f:
              begin, end = map(int, line.split()[1:3])
              genes.append((begin, end))
```

7. Посчитал количество z-днк, которые попали в какой-то ген:

```
.8] intersect_gene = 0
    in_gene = 0
    with open('chr27.fa.Z-SCORE.cut', 'r') as f:
        for line in f:
            zdna_begin, zdna_end = map(int, line.split()[0:2])
            for gene_begin, gene_end in genes:
                if gene_begin <= zdna_begin and zdna_end <= gene_end:
                    in_gene += 1
                    break
                elif gene_begin <= zdna_begin <= gene_end or gene_begin <= zdna_end <= gene_end:
                    intersect_gene += 1
                    break

    print('intersections with gene:', intersect_gene)
    print('in gene:', in_gene)

intersections with gene: 83
in gene: 5196
```

8. Подсчитал количество квадруплексов, лежащих в генах и на пересечении с ними:

```

▶ intersect_gene = 0
in_gene = 0

for g4_begin, g4_end, _ in PQS:
    for gene_begin, gene_end in genes:
        if gene_begin <= g4_begin and g4_end <= gene_end:
            in_gene += 1
            break
        elif gene_begin <= g4_begin <= gene_end or gene_begin <= g4_end <= gene_end:
            intersect_gene += 1
            break

print('intersections with gene:', intersect_gene)
print('in gene:', in_gene)

intersections with gene: 0
in gene: 151

```

9. Выгрузил гены из файла, скаченного из Table Browser:

```

genes = []
with open('alignment chicken', 'r') as f:
    for line in f:
        begin, end = map(int, line.split()[1:3])
        name = line.split()[3]
        strand = line.split()[5]
        genes.append((begin, end, name, strand))

```

10. Сформировал список промоторов этих генов:

```

[23] promoters = []
for gene in genes:
    if gene[3] == '+':
        gene_start = gene[0]
        promoters.append((gene_start - 1000, gene_start, gene[2]))
    else:
        gene_end = gene[1]
        promoters.append((gene_end, gene_end + 1000, gene[2]))

```

11. Подсчитал количество промоутеров, в которые попало Z-ДНК:

```

▶ in_gene = 0
intersect_gene = 0
with open('chr27.fa.Z-SCORE.cut', 'r') as f:
    for promotor in promoters:
        for line in f:
            zdna_begin, zdna_end = map(int, line.split()[0:2])
            if promotor[0] <= zdna_begin and zdna_end <= promotor[1]:
                in_gene += 1
                break
            elif promotor[0] <= zdna_begin <= promotor[1] or promotor[0] <= zdna_end <= promotor[1]:
                intersect_gene += 1
                break
print(in_gene, intersect_gene)

1 0

```

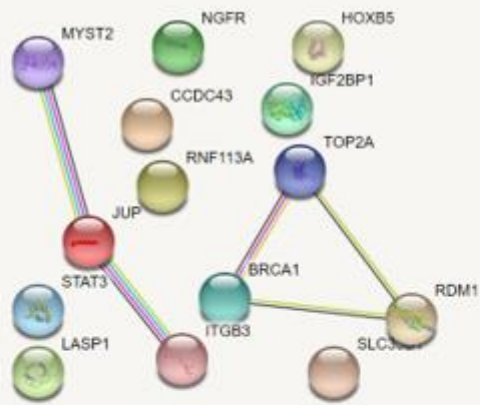
12. Подсчитал количества промоторов, в которые попали квадруплексы и выгрузила имена соответствующих генов в отдельный файл:

```
[26] out = open('genes_with_g4_in_promoters', 'w')
      in_gene = 0
      intersect_gene = 0
      with open('chr27.fa.Z-SCORE.cut', 'r') as f:
          for promotor in promotors:
              for g4_start, g4_end, _ in PQS:
                  if promotor[0] <= g4_start and g4_end <= promotor[1]:
                      in_gene += 1
                      out.write(promotor[2] + '\n')
                      break
                  elif promotor[0] <= g4_start <= promotor[1] or promotor[0] <= g4_end <= promotor[1]:
                      intersect_gene += 1
                      out.write(promotor[2] + '\n')
                      break
```

Список генов в файле genes_with_g4_promoters:

```
NM_001007884
NM_204546
NM_001104958
NM_204315
NM_001004396
NM_204514
NM_001031341
NM_001146133
NM_205071
NM_001025355
NM_001193585
NM_001321599
NM_001177329
NM_204791
NM_001267553
NM_001010842
NR_162022
NM_001030931
NM_001328354
NM_204169
```

Скрин анализа:



[Viewers >](#)
[Legend >](#)
[Settings >](#)
[Analysis ▾](#)
[Exports >](#)
[Clusters >](#)
[More >](#)

Network Stats

number of nodes: 14
 number of edges: 5
 average node degree: 0.714
 avg. local clustering coefficient: 0.357

expected number of edges: 3
 PPI enrichment p-value: 0.218
your network does not have significantly more interactions than expected (what does that mean?)