Выполненное ДЗ присылать на почту [dmitrii.lv.konovalov@gmail.com](mailto:dmitrii.lv.konovalov@gmail.com)

По вопросам писать в telegram +7-915-455-46-41

Дедлайн: 18:10 .15.02.22

За каждый день после дедлайна штраф -1 балл.

**1.** [**UCSC**](http://genome-euro.ucsc.edu/cgi-bin/hgGateway)

**1.1.** Выберите какой-нибудь белок человека и найдите информацию о гене этого белка в геномном браузере UCSC.

Укажите**[[1]](#footnote-0)**:

* имя гена,
* идентификатор гена в Gencode,
* на какой цепи он закодирован,
* в какой хромосоме находится,
* к каким плечу и полосе принадлежит участок (например, chr7:p14.2, здесь chr7:p – короткое плечо хромосомы 7, 14.2 – номер полосы в этом плече),
* сколько альтернативных продуктов (транскриптов) закодировано в гене; для каждого транскрипта**[[2]](#footnote-1)** укажите идентификатор Gencode, координаты в хромосоме (включая UTR'ы), число экзонов (общее) и длину последовательности белка (если транскриптов больше трёх, то описать нужно только три первых).

**1.2.** Сохраните картинку окрестности гена из Genome Browser c треками[[3]](#footnote-2):

* транскрипты GENCODE и RefSeq,
* консервативность последовательности среди позвоночных (Conservation),
* частые полиморфизмы (Common SNPs) последней версии (151).

Все остальные треки скройте.

**2.** [**Ensembl**](https://www.ensembl.org/index.html)

2.1[[4]](#footnote-3) Постройте выравнивание выбранного гена человека с гомологичным геном шимпанзе (прикрепить к письму)

2.2[[5]](#footnote-4) Оцените процент различий на 100 нуклеотидов.

2.3. Сравните с полногеномной оценкой. (источник указать)

**3\*[[6]](#footnote-5).** [**Genome Data Viewer**](https://www.ncbi.nlm.nih.gov/genome/gdv/)

Рассмотрите тот же фрагмент генома в Genome Data Viewer[[7]](#footnote-6).

Опишите, в чем отличие от других браузеров, в чем он удобнее/неудобнее, какую информацию можно извлечь только из этого браузера.

1. Всю информацию берите из аннотации GENCODE. [↑](#footnote-ref-0)
2. Щелкните на транскрипт, чтобы увидеть его характеристики. Количество альтернативных продуктов – это количество строк в треке Gencode при full-отображении.

   Длину белка можно узнать в разделе "Sequence and Links" на странице транскрипта. [↑](#footnote-ref-1)
3. Чтобы сохранить картинку, нажмите правой кнопкой на трек и выберите пункт "View image". Если не получится, сделайте скриншот. [↑](#footnote-ref-2)
4. Найдите заданный ген в поиске Ensembl. В меню слева выберите Comparative Genomics → Genomic alignments. Нажмите Select another alignment, введите название организма – Chimpanzee → Apply. Сохраните полученное выравнивание (Download) [↑](#footnote-ref-3)
5. Подсчитать число различий можно командой infoalign пакета EMBOSS, или через biopython.

   from Bio import AlignIO

   alignment=AlignIO.read(open("alignment.fa"), 'fasta')

   seq1=str(alignment[0].seq)

   seq2=str(alignment[1].seq)

   matches = sum(nuc1 == nuc2 for nuc1, nuc2 in zip(seq1, seq2))

   identity = 100.0 \* matches / len(seq1) [↑](#footnote-ref-4)
6. Задание опционально [↑](#footnote-ref-5)
7. В два клика из UCSC. Cверху нажать на View->NCBI [↑](#footnote-ref-6)