Идете на сайт UCSC genome browser, Downloads -> Genome Data

<https://hgdownload.soe.ucsc.edu/downloads.html>

Выбрать любой организм КРОМЕ человека и мыши. Любой из вкладок Mammals, Other Vertebrates, Deuterostomes, Insects, Nematodes, Othergenomes. У организма достаточно взять какую-нибудь одну хромосому.

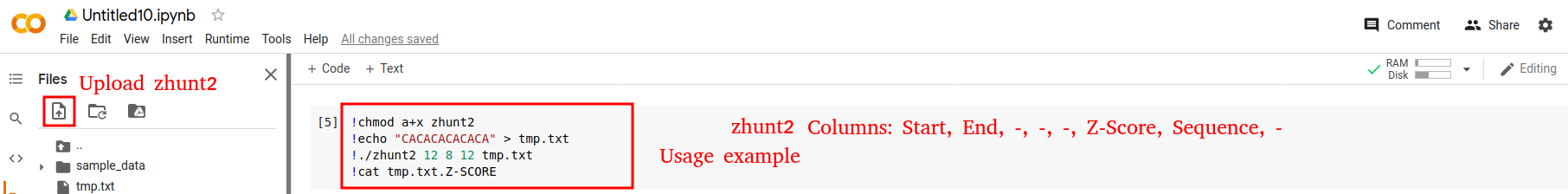
Внести его в таблицу вместе с ФИО и группой

<https://docs.google.com/spreadsheets/d/1OwSkA143pGSei1-N_LgRAoG9ZMuUrwxCTFVKwFOT3TM/edit?usp=sharing>

Найти в геноме участки Z-ДНК (с помощью программы Z-Hunt) и квадруплексы (c помощью паттерна, см семинар)

**Как работать с Z-Hunt?**

Программа Z-Hunt приложена в виде исполняемого файла для OS Linux. Скомпилирована специально для работы в Google Colaboratory(colab.research.google.com). Пример работы:



В целом, необходимо:

(1) создать Colab Notebook;

(2) загрузить zhunt в Colab Notebook и добавить права на исполнение [chmod a+x zhunt];

(3) загрузить геном вируса в формате fasta в Colab Notebook;

(4) запустить zhunt как показано в примере [zhunt2 12 8 12 genome.fasta];

Имена колонок в предсказании zhunt приведены на картинке.

В результате получится аннотация генома Z-ДНК c Z-score. (Использовать порог по Z-score в 300–400. Чем больше Z-score, тем проще формирование Z-ДНК.)

В отчете ответить на вопросы:

1. Сколько участков Z-ДНК найдено в геноме?
2. Сколько квадруплексов найдено в геноме?
3. В какие участки генома (гены или межгеннное пространство или на пересечении) попадают Z-ДНК и квадруплексы. Аннотацию генома генами для выбранного организма скачать из TableBrowser - Genes and Gene Prediction Track – GeneCode annotation tracks.
4. Предоставить список генов, в промоторы которых попадает Z-ДНК. Промоторами считать участки от начала генов и upstream 1000 bp. При этом для генов на положительном и отрицательном стренде по-разному рассчитываются координаты промоторов. Для положительного стренда, если начало гена на координате 5000 bp, промоторный участок будет 4000-5000, а для отрицательного стренда 5000-6000. Имена вытаскиваются из предпоследней колонки таблицы.
5. Предоставить список генов, в промоторы которых попадает квадруплексы.
6. На 10 баллов – сделать Gene Enrichment Analysis – зайти на сайт stringDB <https://string-db.org/>. Нажать на Search. Выбрать слева Multiple Proteins - Найти нужный организм, в окно загрузить список генов и нажать Search. На вкладе Analysis будут результаты обогощения отобранных генов биологическими процессами. Сделать скриншот анализа. Выгрузить таблицу в Excel.