

1.

Для данного tsv-файла с аннотацией экзонов, который содержит таблицу с 4 столбцами (без имен столбцов): идентификатор экзона, хромосома (1..22, X, Y), начальная позиция и конечная позиция экзона, создать tsv-файл task1.txt с таблицей, отсортированный по столбцу хромосом, начальной и конечной позиции.

Пример ввода:

ENSMUST00000230915_ENSMUSE000001441684 15	66604027	66604240
ENSMUST00000226593_ENSMUSE00000129090 16	52183212	52183358
ENSMUST00000134408_ENSMUSE000001219035 2	126846074	126846245
ENSMUST00000147127_ENSMUSE000001268847 X	73765267	73765372
ENSMUST00000109298_ENSMUSE000001370514 2	165280786	165280954

Вывод:

ENSMUST00000134408_ENSMUSE000001219035 2	126846074	126846245
ENSMUST00000109298_ENSMUSE000001370514 2	165280786	165280954
ENSMUST00000230915_ENSMUSE000001441684 15	66604027	66604240
ENSMUST00000226593_ENSMUSE00000129090 16	52183212	52183358
ENSMUST00000147127_ENSMUSE000001268847 X	73765267	73765372

TSV файл по ссылке: <https://filetransfer.io/data-package/HnP97Imx/download>

2. Дана таблица с путями к .txt файлам:

/here/is/the/path/to/file.txt

/here/is/the/path/to/another/yet_another_file.txt

Получите файл task2.txt с названиями файлов без обозначения формата:

file

yet_another_file

Исходный файл: <https://filetransfer.io/data-package/EvgX6Ezy/download>

3. Дан FASTQ файл.

Оцените качество 2 ряда в нем. Если на каком-то из концов качество ниже, то на каком?

Как это можно объяснить? Напишите соображения в task3.txt

Исходный файл: <https://filetransfer.io/data-package/YteQvi7B/download>

4. Найдите качество 10 позиции 5ого ряда в предыдущем файле, сохраните в task5.txt

Со звездочкой: определите среднее качество 2ого ряда, сохраните в task3extra.txt

5. Оцените FASTQC отчет по ссылке:

https://kodomo.fbb.msu.ru/~asofronova/pr12/Ath_tae_CTTGTA_L003_R2_005_fastqc.html

Напишите аргументированно, как вы считаете, можно ли использовать такие данные для дальнейшей сборки генома в файл task5.txt