#!/bin/bash

# 1. Создание директории для результатов

mkdir -p terminal\_task/results

# 2. Загрузка файла аннотации генома (если он не загружен)

cd terminal\_task

if [[ ! -f "gencode.v41.basic.annotation.gff3.gz" ]]; then

echo "Скачивание файла аннотации..."

wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode\_human/release\_41/gencode.v41.basic.annotation.gff3.gz

fi

# 3. Разархивация файла

gunzip -kf gencode.v41.basic.annotation.gff3.gz # -k сохраняет .gz, -f перезаписывает

# 4. Обработка файла

echo "Обработка файла аннотации..."

awk '

BEGIN { OFS="\t" }

# Убираем комментарии и оставляем только строки с "gene" в третьей колонке

$3 == "gene" && $9 ~ /gene\_type=unprocessed\_pseudogene/ {

# Извлекаем gene\_name из 9-й колонки

match($9, /gene\_name=([^;]+)/, arr)

gene\_name = arr[1]

chrom = $1

start = $4

end = $5

strand = $7

if (strand == "+") {

adjusted\_start = start

adjusted\_end = start + 1

} else if (strand == "-") {

adjusted\_start = end

adjusted\_end = end + 1

}

print chrom, adjusted\_start, adjusted\_end, strand, gene\_name

}

' gencode.v41.basic.annotation.gff3 > results/result.tsv

echo "Результат записан в: terminal\_task/results/result.tsv"