Step 0

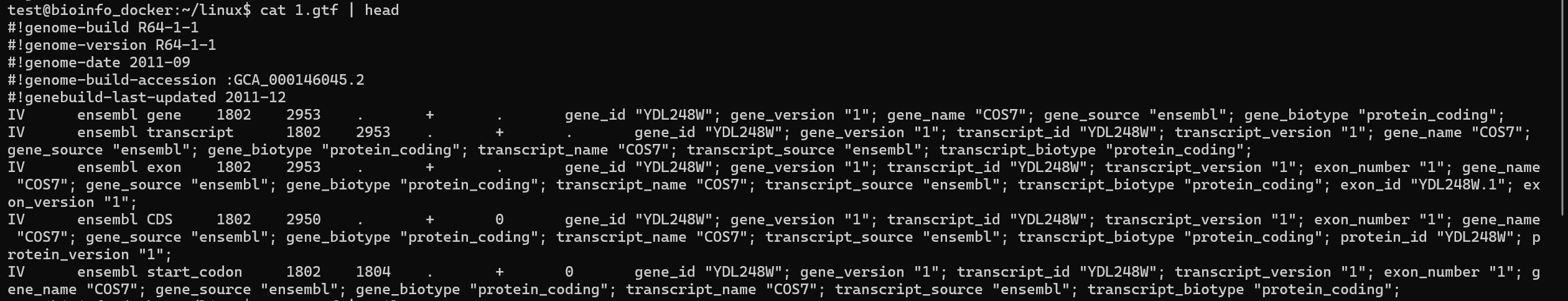
cd linux

ls

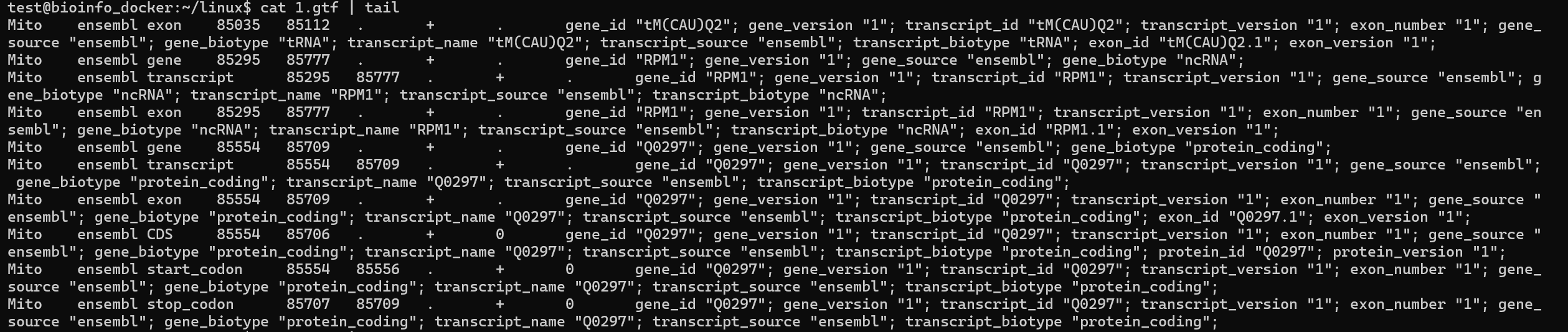
1.gtf file

Step 1

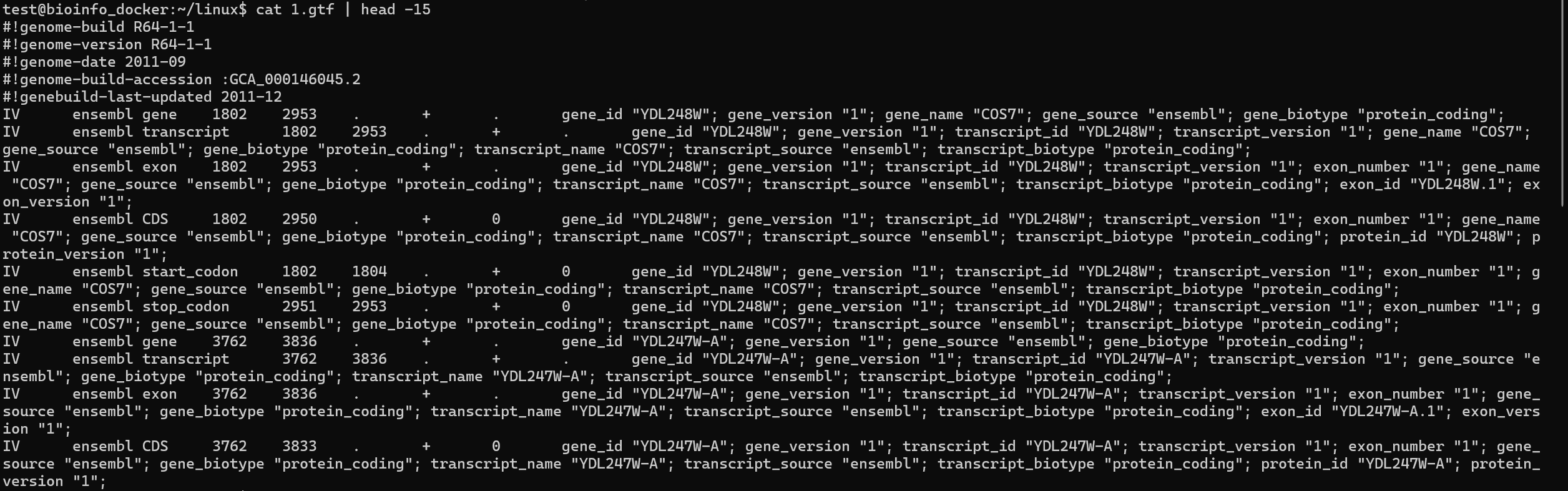
cat 1.gtf | head



cat 1.gtf | tail



cat 1.gtf | head -15



ls -lh 1.gtf

-rw-rw-r-- 1 test test 12M Sep 11 2018 1.gtf

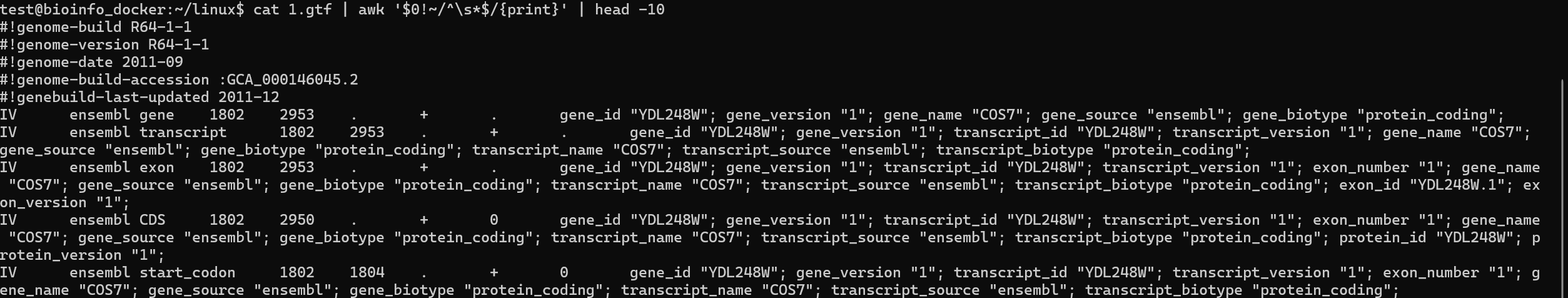
wc -l 1.gtf

42252 1.gtf

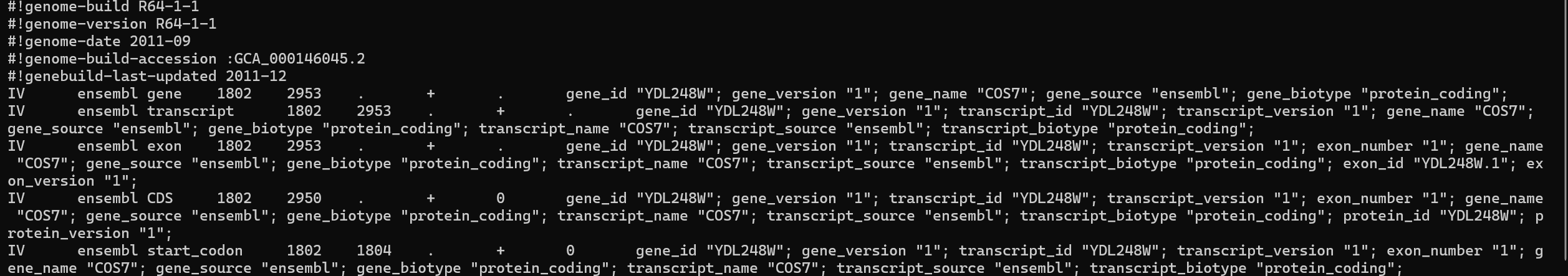
grep -v '^#' 1.gtf | grep -v '^$' | wc -l

42247

cat 1.gtf | awk '$0!~/^\s\*$/{print}' | head -10

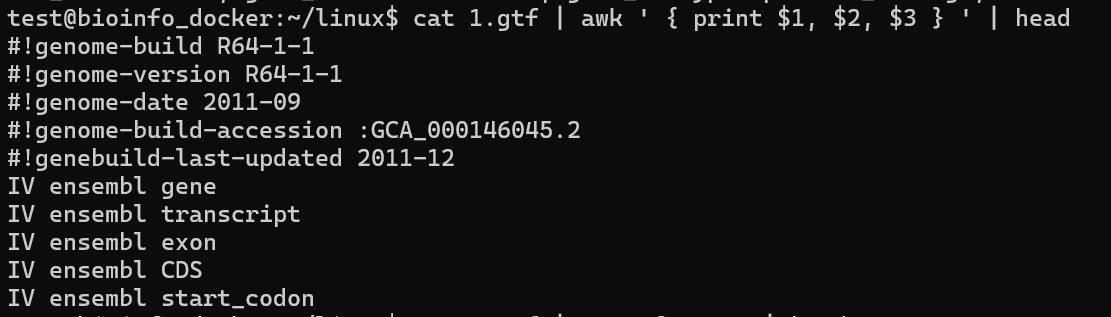


grep -v '^\s\*$' 1.gtf | head -10

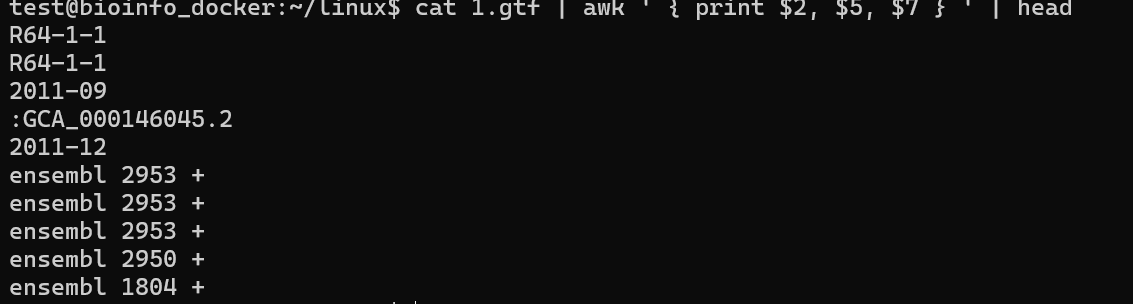


Step 2

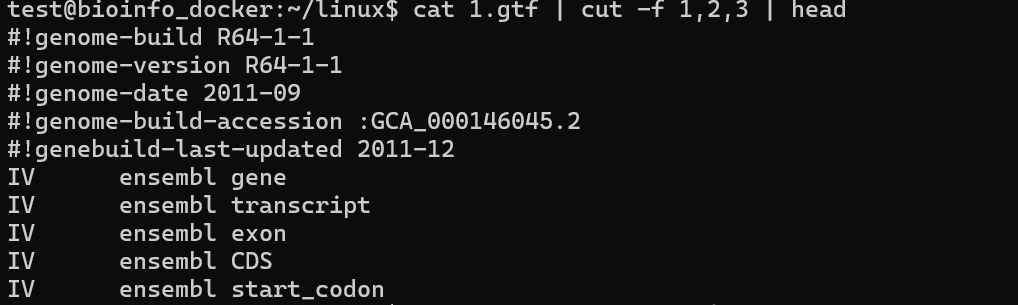
cat 1.gtf | awk ' { print $1, $2, $3 } ' | head



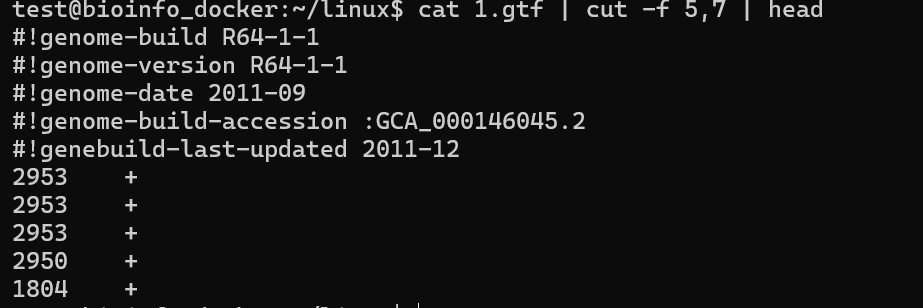
cat 1.gtf | awk ' { print $2, $5, $7 } ' | head



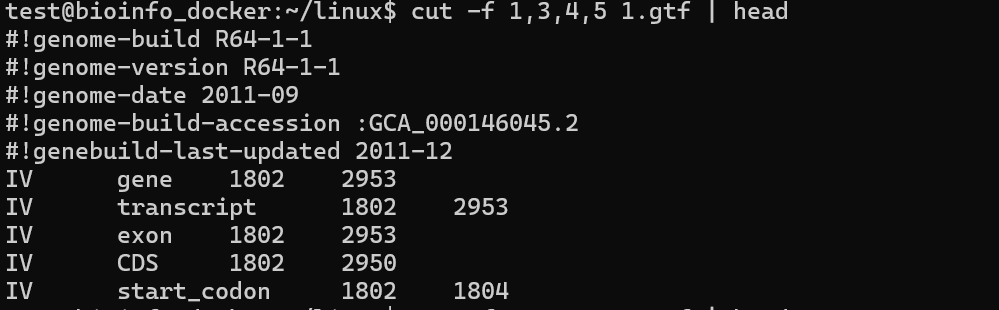
cat 1.gtf | cut -f 1,2,3 | head



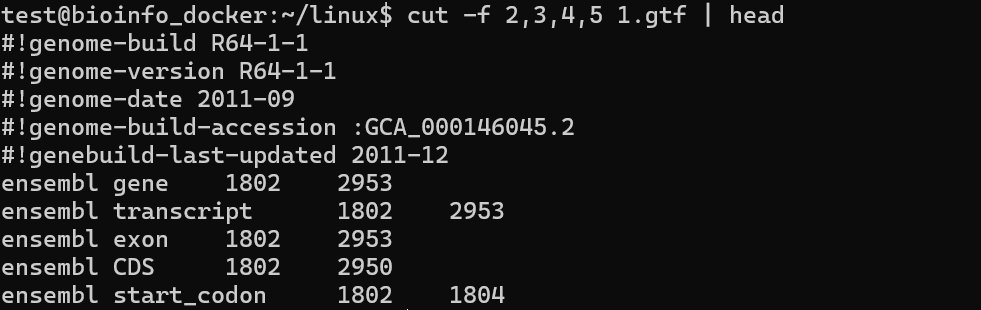
cat 1.gtf | cut -f 5,7 | head



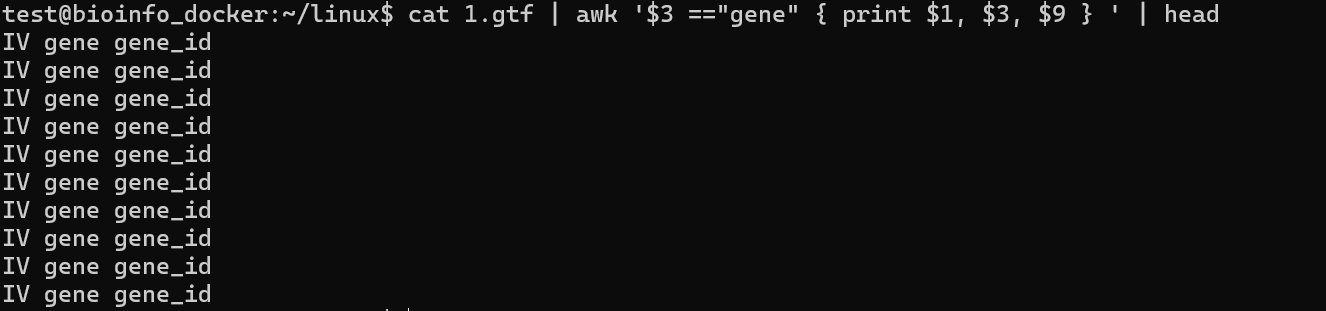
cut -f 1,3,4,5 1.gtf | head



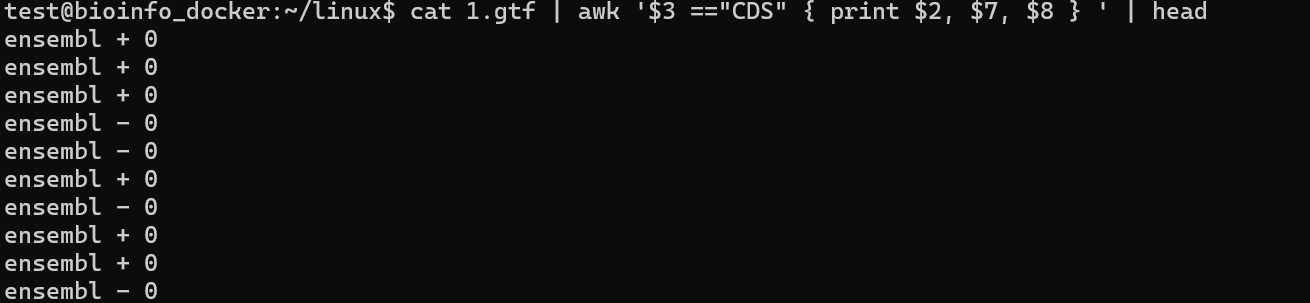
cut -f 2,3,4,5 1.gtf | head



cat 1.gtf | awk '$3 =="gene" { print $1, $3, $9 } ' | head



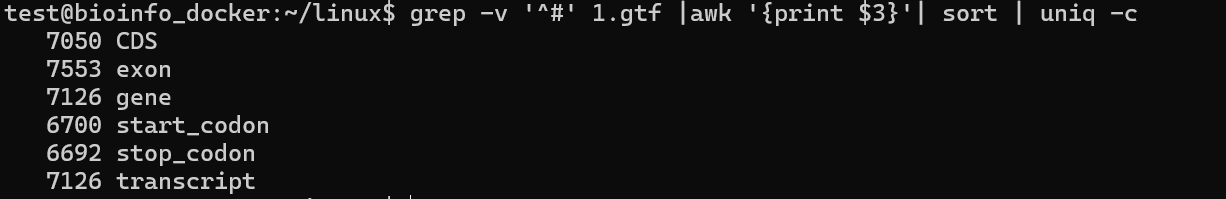
cat 1.gtf | awk '$3 =="CDS" { print $2, $7, $8 } ' | head



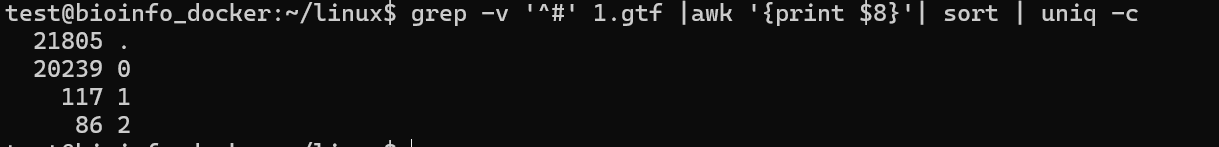
Step 3

Step 3.1

grep -v '^#' 1.gtf |awk '{print $3}'| sort | uniq -c

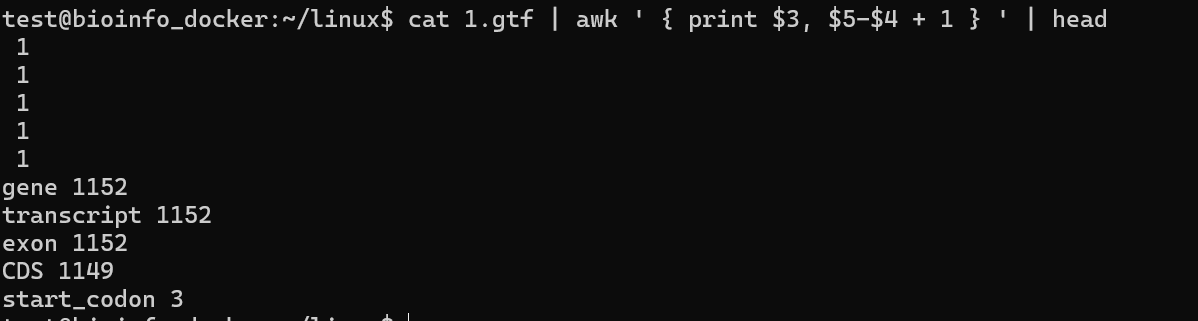


1.gtf |awk '{print $8}'| sort | uniq -c



Step 3.2

cat 1.gtf | awk ' { print $3, $5-$4 + 1 } ' | head



cat 1.gtf | awk 'BEGIN{size=0;}$3 =="CDS"{ len=$5-$4 + 1; size += len; print "Size:", size } ' | tail -n 1

Size: 9030648

cat 1.gtf | awk 'BEGIN{L=0;}$3 =="CDS"{L+=$5-$4 + 1;}END{print L;}'

9030648

cat 1.gtf | awk '$3 =="CDS"{L+=$5-$4 + 1;}END{print L;}'

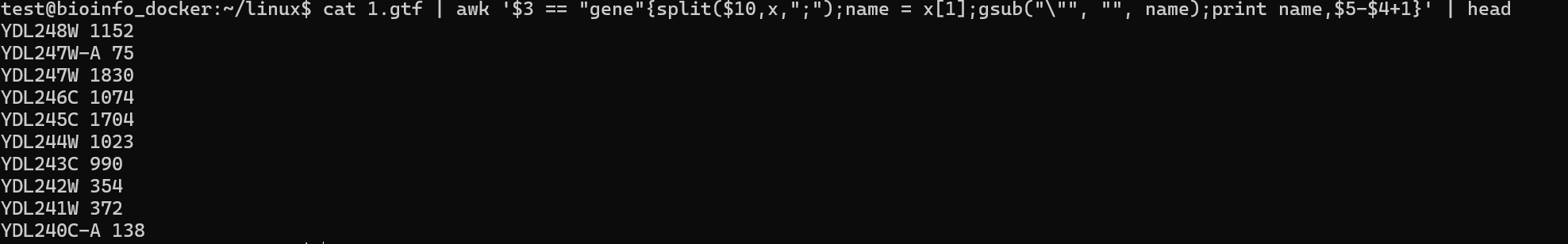
9030648

awk 'BEGIN {s = 0;line = 0;}$3 =="CDS" && $1 =="I"{ s += $5-$4+1;line += 1}END {print "mean="s/line}' 1.gtf

mean=1239.52

Step 3.3

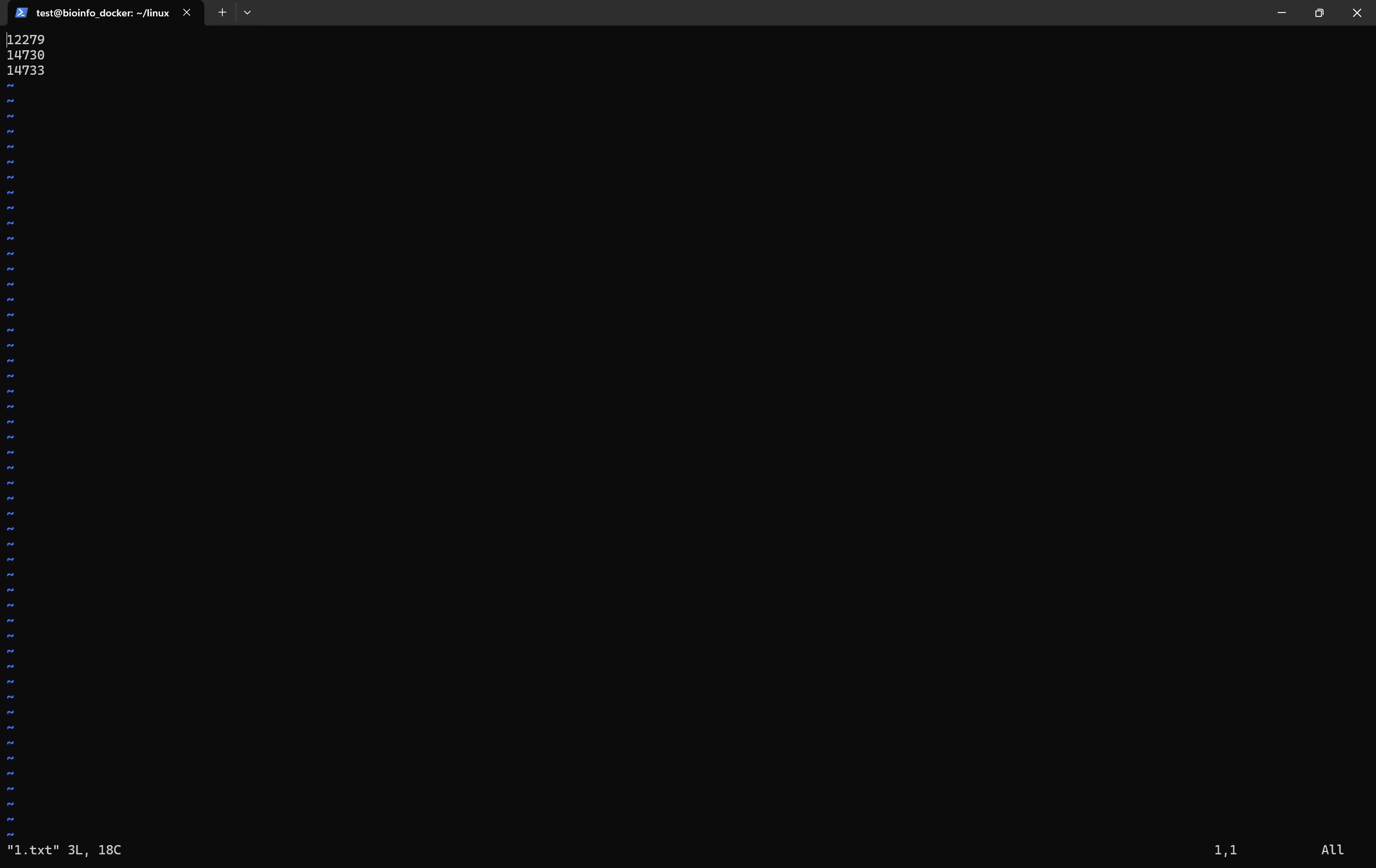
cat 1.gtf | awk '$3 == "gene"{split($10,x,";");name = x[1];gsub("\"", "", name);print name,$5-$4+1}' | head

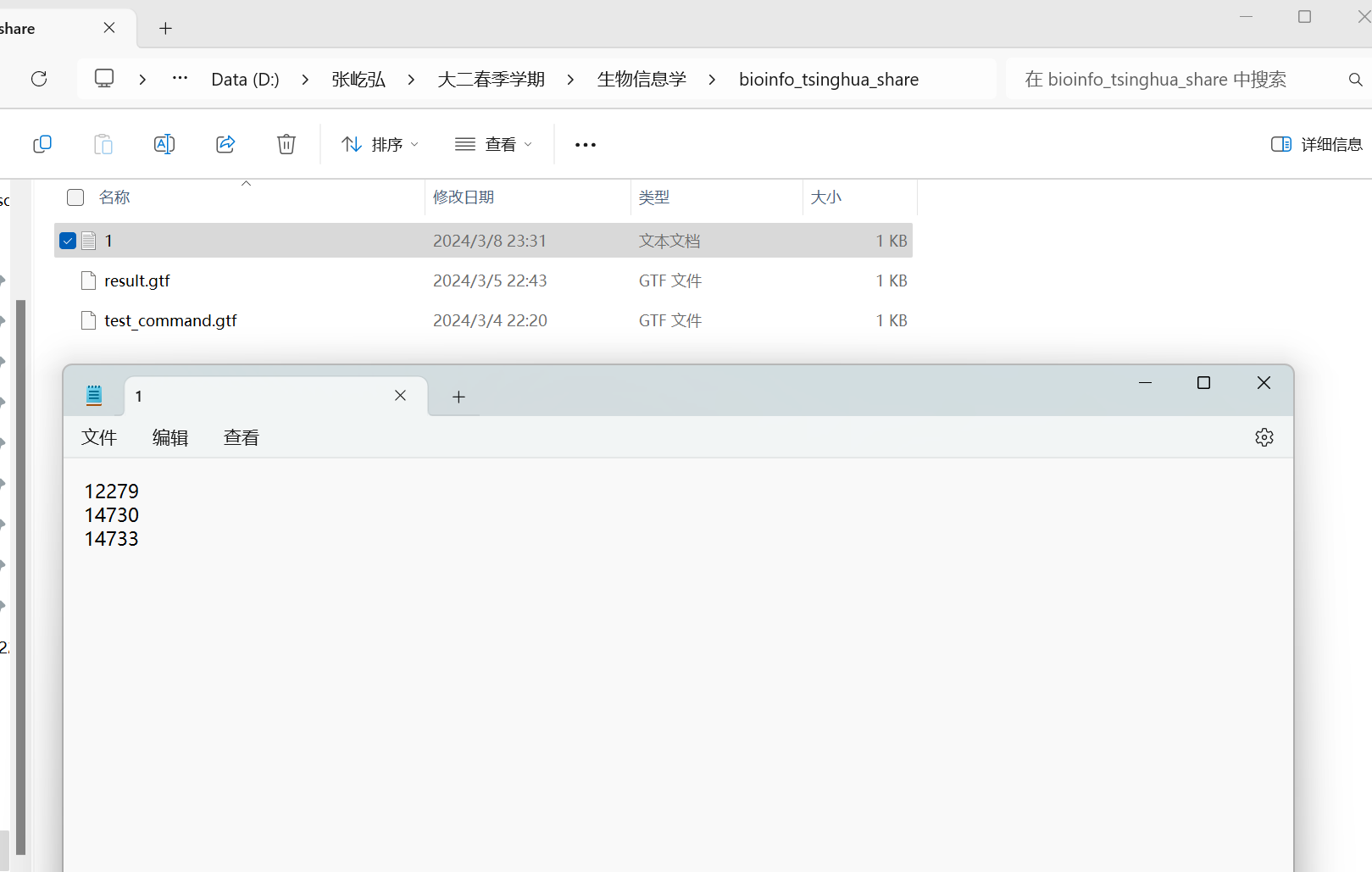


Step 4.1

grep exon 1.gtf | awk '{print $5-$4+1}' | sort -n | tail -3 > 1.txt

cp 1.txt /home/test/share





Step 4.2

vi run.sh

#!/bin/bash

grep exon \*.gtf | awk '{print $5-$4+1}' | sort -n | tail -3

bash run.sh

