

Apply to question 1-5.

Align both RNA-seq data sets to the reference genome using tophat. Analyze the results to answer the following questions. If multiple answers are required for one question, separate the answers with a space (e.g., 1000 2000).

Apply to question 6-10.

Assemble the aligned RNA-seq reads into genes and transcripts using cufflinks. Use the labels 'Day8' and 'Day16', respectively, when creating identifiers. For this portion of the analysis, answer the following questions. Use the same format as above for answers consisting of multiple values or parts.

Apply to question 11-15.

Run cuffcompare on the resulting cufflinks transcripts, using the reference gene annotations provided and selecting the option '-R' to consider only the reference transcripts that overlap some input transfrag. For this step, using the *.tmap files answer the following, for both sets. Use the same format as above for answers consisting of multiple parts.

Apply to question 16-20.

Perform the differential gene expression analysis. For this, in a first stage run cuffmerge using the provided annotation to merge and reconcile the two sets of cufflinks transcripts. Make a note of the resulting file, 'merged.gtf'. In a second step, use cufdiff to perform the differential expression analysis.

NOTE: Note that in general at least three replicates per condition are

required to establish statistical significance. The single replicate example is provided here only to illustrate the analysis.

```
[guest@centos6]$ gunzip gencommand_proj4.tar.gz
```

```
[guest@centos6]$ tar xvc gencommand_proj4.tar
```

```
[guest@centos6]$ ls -tl
```

```
[guest@centos6]$ cd gencommand_proj4
```

```
[guest@centos6 gencommand_proj4]$ mkdir athal
```

```
[guest@centos6 gencommand_proj4]$ bowtie2-build athal_chr.fa athal/  
athal
```

```
[guest@centos6 gencommand_proj4]$ vi commands/com.tophat
```

Note: edit this command file (see commands fold)

```
[guest@centos6 gencommand_proj4]$ sh commands/com.tophat
```

1. How many alignments were produced for the 'Day8' and 'Day16' RNA-seq data sets, respectively?

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day8/  
accepted_hits.bam | cut -f7 | wc -l
```

```
63845
```

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day16/  
accepted_hits.bam | cut -f7 | wc -l
```

```
58398
```

2. How many reads were mapped in each set?
3. How many reads were uniquely aligned in each case?
5. How many reads were left unmapped from each set?

Note: find answer in the summary file

```
[guest@centos6 gencommand_proj4]$ head Tophat/Day8/  
align_summary.txt
```

```
[guest@centos6 gencommand_proj4]$ head Tophat/Day16/  
align_summary.txt
```

4. How many spliced alignments were reported for each set?

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day8/  
accepted_hits.bam | cut -f6 | grep "N" | head
```

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day8/  
accepted_hits.bam | cut -f6 | grep "N" | wc -l
```

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day16/  
accepted_hits.bam | cut -f6 | grep "N" | head
```

```
[guest@centos6 gencommand_proj4]$ samtools view Tophat/Day16/  
accepted_hits.bam | cut -f6 | grep "N" | wc -l
```

For questions 6-10:

```
[guest@centos6 gencommand_proj4]$ vi commands/com.cufflinks
```

note: to edit cufflinks file (see commands)

```
[guest@centos6 gencommand_proj4]$ sh commands/com.cufflinks
```

```
[guest@centos6 gencommand_proj4]$ ls Cufflinks/Day8
```

6. How many genes were generated by cufflinks for each set (Day8 and Day16)?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/transcripts.gtf |  
cut -f9 |cut -d ";" -f1 | sort -u | wc -l
```

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/transcripts.gtf |  
cut -f9 |cut -d ";" -f1 | sort -u | wc -l
```

7. How many transcripts were reported for each?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/transcripts.gtf |  
cut -f9 |cut -d ";" -f2 | sort -u | wc -l
```

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/transcripts.gtf |  
cut -f9 |cut -d ";" -f2 | sort -u | wc -l
```

8. How many single transcript genes were produced for both sets?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/transcripts.gtf |  
cut -f9 |cut -d ";" -f1,2 | uniq | cut -d ";" -f1 | uniq -c | awk '$1 == 1' | wc -l
```

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/transcripts.gtf |  
cut -f9 |cut -d ";" -f1,2 | uniq | cut -d ";" -f1 | uniq -c | awk '$1 == 1' | wc -l
```

9. How many single-exon transcripts were in the two sets?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/transcripts.gtf |  
cut -f9 |grep "exon_number" | cut -d ";" -f2,3 | cut -d ";" -f1 | uniq -c | awk  
'$1 == 1' | wc -l
```

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/transcripts.gtf |  
cut -f9 |grep "exon_number" | cut -d ";" -f2,3 | cut -d ";" -f1 | uniq -c | awk  
'$1 == 1' | wc -l
```

10. How many multi-exon transcripts were reported for each set?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/transcripts.gtf |  
cut -f9 |grep "exon_number" | cut -d ";" -f2,3 | cut -d ";" -f1 | uniq -c | awk  
'$1 > 1' | wc -l
```

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/transcripts.gtf |  
cut -f9 |grep "exon_number" | cut -d ";" -f2,3 | cut -d ";" -f1 | uniq -c | awk  
'$1 > 1' | wc -l
```

Questions 11-15

```
[guest@centos6 gencommand_proj4]$ vi commands/com.cuffcompare
```

Note: edit com.cuffcompare file (see commands)

```
[guest@centos6 gencommand_proj4]$ sh commands/com.cuffcompare
```

```
[guest@centos6 gencommand_proj4]$ ls Cuffcompare
```

```
[guest@centos6 gencommand_proj4]$ ls Cufflinks/Day8 -tl
```

```
[guest@centos6 gencommand_proj4]$ head Cufflinks/Day8/  
Day8.transcripts.gtf.tmap
```

11. How many cufflinks transcripts fully reconstruct annotation transcripts in each sample?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/  
Day8.transcripts.gtf.tmap | cut -f3 | grep "=" | wc -l  
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/  
Day16.transcripts.gtf.tmap | cut -f3 | grep "=" | wc -l
```

12. How many splice variants does the gene AT4G20240 have in the two samples?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/  
Day8.transcripts.gtf.tmap | grep "AT4G20240" | wc -l  
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/  
Day16.transcripts.gtf.tmap | grep "AT4G20240" | wc -l
```

13. How many cufflinks transcripts are partial reconstructions of reference transcripts ('contained')?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/  
Day8.transcripts.gtf.tmap | cut -f3 | grep "c" | wc -l  
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/  
Day16.transcripts.gtf.tmap | cut -f3 | grep "c" | wc -l
```

14. How many cufflinks transcripts are novel splice variants of reference genes?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/  
Day8.transcripts.gtf.tmap | cut -f3 | grep "j" | wc -l  
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/  
Day16.transcripts.gtf.tmap | cut -f3 | grep "j" | wc -l
```

15. How many cufflinks transcripts were formed in the introns of reference genes?

```
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day8/  
Day8.transcripts.gtf.tmap | cut -f3 | grep "i" | wc -l  
[guest@centos6 gencommand_proj4]$ cat Cufflinks/Day16/  
Day16.transcripts.gtf.tmap | cut -f3 | grep "i" | wc -l
```

Questions 16-20

```
[guest@centos6 gencommand_proj4]$ vi commands/com.cuffdiff  
[guest@centos6 gencommand_proj4]$ vi Cuffmerge/GTFs.txt  
Note: edit both files  
[guest@centos6 gencommand_proj4]$ sh commands/com.cuffdiff  
[guest@centos6 gencommand_proj4]$ ls Cuffmerge -tl  
[guest@centos6 gencommand_proj4]$ head Cuffmerge/merged.gtf
```

16. How many genes (loci) were reported in the merged.gtf file?

```
[guest@centos6 gencommand_proj4]$ cat Cuffmerge/merged.gtf | cut -f9 |  
cut -d ";" -f1 | sort -u | wc -l
```

17. How many transcripts?

```
[guest@centos6 gencommand_proj4]$ cat Cuffmerge/merged.gtf | cut -f9 |  
cut -d ";" -f2 | sort -u | wc -l
```

18. How many genes total were included in the gene expression report from cuffdiff?

```
[guest@centos6 gencommand_proj4]$ ls Cuffdiff
```

```
[guest@centos6 gencommand_proj4]$ cat Cuffdiff/gene_exp.diff | head
```

```
[[guest@centos6 gencommand_proj4]$ cat Cuffdiff/gene_exp.diff | cut -f2 |  
sort -u | wc -l
```

19. How many genes were detected as differentially expressed?

```
[guest@centos6 gencommand_proj4]$ cat Cuffdiff/gene_exp.diff | cut -f14  
| grep "yes" | wc -l
```

20. How many transcripts were differentially expressed between the two samples?

```
[guest@centos6 gencommand_proj4]$ cat Cuffdiff/isoform_exp.diff || cut -
```



```
f14 | grep "yes" | wc -l
```