

## Feedback — Module 4 Exam

[Help Center](#)

Thank you. Your submission for this exam was received.

You submitted this exam on **Sat 3 Oct 2015 12:47 AM PDT**. You got a score of **20.00** out of **20.00**.

For this project, it is recommended that you use the VMBox virtual environment provided with the Course package and the tools therein. You may also use your own system and software, however make sure that appropriate versions are installed. The answers are compatible with the following versions of the software: samtools v.1.2, bowtie2 v.2.2.2, tophat v.2.0.14, and cufflinks/ cuffmerge/ cuffcompare/ cuffdiff v.2.2.1.

You are performing an RNA-seq experiment to determine genes that are differentially expressed at different stages in the development of Arabidopsis thaliana shoot apical meristem. You collected samples at day 8 and day 16 (files “Day8.fastq” and “Day16.fastq”), extracted and sequenced the cellular mRNA, and are now set to perform the bioinformatics analysis. The reference genome you will need for the analysis is “athal\_chr.fa” and the reference gene annotations are in “athal\_genes.gtf”. Use default parameters unless otherwise specified. Sample command files that you can modify to create your own pipeline are provided in the file “commands.tar.gz”. All files are provided in the archive gencommand\_proj4.tar.gz.

[Click here to download the Project 4 Data Files](#)

NOTE: The genome and annotation data were obtained and modified from the Arabidopsis Information Resources (TAIR) Database, and the RNA-seq reads were extracted from GenBank’s Short Read Archive (SRA).

Create a bowtie index of the genome using bowtie2-build, with the prefix ‘athal’. Include a copy of the reference genome with the name “athal.fa” in the index directory.

**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.

## Question 1

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

**You entered:**

63845 58398

Your Answer	Score	Explanation
63845	✓ 0.50	
58398	✓ 0.50	
Total	1.00 / 1.00	

## Question 2

How many reads were mapped in each set?

You entered:

63489 57951

Your Answer		Score	Explanation
63489	✓	0.50	
57951	✓	0.50	
Total		1.00 / 1.00	

## Question 3

How many reads were uniquely aligned in each case?

You entered:

63133 57504

Your Answer		Score	Explanation
63133	✓	0.50	
57504	✓	0.50	
Total		1.00 / 1.00	

## Question 4

How many spliced alignments were reported for each set?

**You entered:**

8596 10695

Your Answer		Score	Explanation
8596	✓	0.50	
10695	✓	0.50	
Total		1.00 / 1.00	

## Question 5

How many reads were left unmapped from each set?

**You entered:**

84 34

Your Answer		Score	Explanation
84	✓	0.50	
34	✓	0.50	
Total		1.00 / 1.00	

## Question 6

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

**You entered:**

186 80

Your Answer		Score	Explanation
-------------	--	-------	-------------

186	✓	0.50
80	✓	0.50
Total		1.00 / 1.00

## Question 7

How many transcripts were reported for each?

**You entered:**

192 92

Your Answer		Score	Explanation
192	✓	0.50	
92	✓	0.50	
Total		1.00 / 1.00	

## Question 8

How many single transcript genes were produced for both sets?

**You entered:**

180 69

Your Answer		Score	Explanation
180	✓	0.50	
69	✓	0.50	
Total		1.00 / 1.00	

## Question 9

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

**You entered:**

119 24

Your Answer		Score	Explanation
119	✓	0.50	
24	✓	0.50	
Total		1.00 / 1.00	

## Question 10

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

**You entered:**

73 68

Your Answer		Score	Explanation
73	✓	0.50	
68	✓	0.50	
Total		1.00 / 1.00	

## Question 11

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

**You entered:**

16 36

Your Answer		Score	Explanation
16	✓	0.50	
36	✓	0.50	
Total		1.00 / 1.00	

## Question 12

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

**You entered:**

2 0

Your Answer		Score	Explanation
2	✓	0.50	
0	✓	0.50	
Total		1.00 / 1.00	

## Question 13

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

**You entered:**

133 21

Your Answer		Score	Explanation
133	✓	0.50	
21	✓	0.50	
Total		1.00 / 1.00	

## Question 14

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

You entered:

14 22

Your Answer		Score	Explanation
14	✓	0.50	
22	✓	0.50	
Total		1.00 / 1.00	

## Question 15

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

You entered:

4 1

Your Answer		Score	Explanation
4	✓	0.50	
1	✓	0.50	



Total

1.00 / 1.00

## Question 16

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

You entered:

Your Answer		Score	Explanation
129	✓	1.00	
Total		1.00 / 1.00	

## Question 17

How many transcripts?

You entered:

Your Answer		Score	Explanation
200	✓	1.00	
Total		1.00 / 1.00	

## Question 18

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

**You entered:**

129

Your Answer		Score	Explanation
129	✓	1.00	
Total		1.00 / 1.00	

## Question 19

How many genes were detected as differentially expressed?

**You entered:**

4

Your Answer		Score	Explanation
4	✓	1.00	
Total		1.00 / 1.00	

## Question 20

How many transcripts were differentially expressed between the two samples?

**You entered:**

5

Your Answer		Score	Explanation
5	✓	1.00	
Total		1.00 / 1.00	

