Module 4 Quiz

10/10 points (100%)

Quiz, 10 questions

Congratulations! You passed!

Next Item

1 / 1 points

1.
Which of the following is FALSE:

- [x] A human gene can express at most 12 splice variants.

Correct

- [ ] Genes form discrete units along the genome.
- [ ] Eukaryotic genes are interrupted, generally consisting of multiple exons.
- [ ] More than 90% of human genes are alternatively spliced.

---

1 / 1 points

2.
Which of the following is FALSE about the organization of a eukaryotic gene:

- [x] The length of an intron cannot be a multiple of 3.

Correct

- [ ] The number of introns in a transcript is one less than the number of exons.
- [ ] Some eukaryotic gene transcripts can consist of a single exon.
- [ ] Genes that have only one exon are not alternatively spliced.

Module 4 Quiz

10/10 points (100%)

Quiz, 10 questions

1 / 1 points

3.

What programs could you use to align RNA-seq reads to: i) a reference genome, and ii) a transcript database?

- tophat, cufflinks
- tophat, split
- tophat, bowtie

Correct

- cufflinks, bowtie

1 / 1 points

4.

Which of the following is FALSE:

- Spliced reads can be used to determine the introns in a gene.
- RNA-seq can be used to quantify the expression levels of proteins.

Correct

- RNA-seq analyses can reveal known genes and their splice variants, as well as novel genes.
- 'Transfrag' stands for 'transcript fragments', a reference to the fact than transcript assemblers cannot always reconstruct full-length splice variants.

1 / 1 points

5.

Module 4 Quiz

Quiz, 10 questions

10/10 points (100%)

What programs could be used to: i) assemble transcripts from RNA-seq reads, and ii) identify potentially novel transcripts and genes

- cuffdiff, cuffdiff
- bowtie, tophat
- cufflinks, cuffcompare

Correct

- igv, cuffdiff

1 / 1 points

6.

Which of the following is FALSE about the gene annotations in the following GTF snippet:

```
1  ```
2  chr1  MGF gene  3413609 3671498 . - . gene_id "MG051951";
3  chr1  MGF transcript  3413609 3416344 . - .gene_id "MG051951"; transcript_id
     "MT162897";
4  chr1  MGF exon  3413609 3416344 . - . gene_id "MG051951"; transcript_id
     "MT162897";
5  chr1  MGF transcript  3421702 3671498 . - . gene_id "MG051951"; transcript_id
     "MT070533";
6  chr1  MGF exon  3670552 3671498 . - . gene_id "MG051951"; transcript_id
     "MT070533";
7  chr1  MGF CDS 3670552 3671348 . - 0 gene_id "MG051951"; transcript_id "MT070533"
     ;
8  chr1  MGF exon  3421702 3421901 . - . gene_id "MG051951"; transcript_id
     "MT070533";
9  chr1  MGF CDS 3421792 3421901 . - 1 gene_id "MG051951"; transcript_id "MT070533"
     ;
10 ```
```

- Both exons of MT70533 contain both coding and non-coding sequences.
- Transcript MT070533 has 4 exons.

Correct

- Transcript MT070533 is designated as coding.
- Gene MG051951 is protein-coding.

1 / 1 points

Module 4 Quiz

Quiz, 10 questions

10/10 points (100%)

7.
What does the following code NOT do:

```
BWT2IDX=/home/me/genomes/hg20/hg20
ANNOT=/home/me/genomes/hg20/myannot.gtf
ANNOTIDX=/home/me/genomes/hg20/myannot/myannot

mkdir -p /home/me/SRR100000
tophat2 -o /home/me/SRR100000 -p 10 --max-multihits 10 \
        -r 26 --mate-std-dev 25 \
        -a 6 \
        -G $ANNOT --transcriptome-index $ANNOTIDX \
        $BWT2IDX \
        /home/me/SRR100000_1.fastq.gz /home/me/SR100000_2.fastq.gz
```

- Use information about the insert size distribution, e.g. its average (26) and standard deviation (25)
- Load the gene annotations index from /home/me/genomes/hg20/myannot/myannot
- Report spliced reads with at most 6 mismatches in the anchor site

Correct

- Use known gene annotations in the first mapping stage

1 / 1 points

8.
What does the following code NOT do:

```
TOPHATDIR=/home/florea/Tophat/

mkdir -p Test1
cd Test1
ln -s $TOPHATDIR/accepted_hits.bam .
cufflinks -L Test1 -p 8 -j 0.10 -F 0.05 accepted_hits.bam
```

- Run cufflinks to assemble transcripts
- Create a soft link to the BAM read alignment file in the Test1 directory
- Label cufflinks transcripts with the prefix ‘Test1’
- Use the default reference transcript annotation to guide assembly

Correct

Module 4 Quiz

10/10 points (100%)

Quiz, 10 questions

1 / 1 points

9.

Which of the following is NOT described in the following summary file produced by tophat:

```
```
Left reads:
          Input   : 60586968
           Mapped   : 58163843 (96.0% of input)
            of these: 6832240 (11.7%) have multiple alignments (359075 have
              >10)
Right reads:
          Input   : 60586968
           Mapped   : 56969290 (94.0% of input)
            of these: 6668479 (11.7%) have multiple alignments (358573 have
              >10)
95.0% overall read mapping rate.

Aligned pairs: 55880048
     of these: 6491876 (11.6%) have multiple alignments
                 2795712 (5.0%) are discordant alignments
87.6% concordant pair alignment rate.
```
```

- Of the mapped mate 2 reads, 5.0% had multiple matches on the genome
- Of the mapped mate 1 reads, 11.7% had multiple matches on the genome
- The library was strand-specific

Correct

- 94.0% of the mate 2 reads were mapped

1 / 1 points

10.

Module 4 Quiz

Quiz, 10 questions

10/10 points (100%)

Which of the following is NOT TRUE about the output below, obtained from a cuffdiff differential expression analysis:

```
```
XLOC_000002 XLOC_000002 AT1G01020 1:5927-8737 q1 q2 OK 1.13032 3.48406 1
  .62404 0.694576 0.5277 0.998846 no
XLOC_000004 XLOC_000004 AT1G01073 1:44676-44787 q1 q2 NOTEST 0 0 0 0 1 1 no
XLOC_000042 XLOC_000042 AT1G01580 1:209394-213041 q1 q2 OK 1.59512 0 -inf
  nan 5e-05 0.0096703 yes
```
```

- [x] Locus XLOC_000042 is novel

Correct

- [ ] Locus XLOC_000042 is not expressed (FPKM is ‘0’) in condition 2 sample(s)
- [ ] Locus XLOC_000002 is differentially expressed between the two conditions
- [ ] The log fold change for locus XLOC_000042 is ‘-inf’