# Module 4 Quiz

**8/10 points (80%)**

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**

- [ ] A codon is a nucleotide triplet that is translated into one amino acid.
- [ ] The coding region with a protein-coding gene is used as the template for forming a protein.
- [ ] Alternative splicing is a common phenomenon in both animals and plants.

---

1 / 1 points

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

- [ ] Some eukaryotic gene transcripts can consist of a single exon.
- [ ] Genes that have only one exon are not alternatively spliced.
- [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.

# Module 4 Quiz

**8/10 points (80%)**

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, igv

tophat, bedtools

tophat, bwa

**Correct**

tophat, bcftools

1 / 1 points

4.
Which of the following is FALSE:

RNA-seq can be used to quantify the expression levels of proteins.

**Correct**

'Transfrag' stands for 'transcript fragments', a reference to the fact than transcript assemblers cannot always reconstruct full-length splice variants.

RNA-seq analyses can reveal known genes and their splice variants, as well as novel genes.

Spliced reads can be used to determine the introns in a gene.

0 / 1 points

# Module 4 Quiz

Quiz, 10 questions

**8/10 points (80%)**

5.
What programs could be used to: i) assemble transcripts from RNA-seq reads, and ii) identify potentially novel transcripts and genes?

- cufflinks, cuffcompare
- cufflinks, cuffmerge
- tophat, cufflinks
- cufflinks, cufflinks

**This should not be selected**

---

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 ```
```

- The gene spans the interval chr1:3413609-3671498.
- The 3’ UTR of transcript MT070533 is located at positions chr1:3421702-3421791
- Transcript MT162897 is located on the reverse strand.
- Transcript MT070533 has 4 exons.

**Correct**

---

# Module 4 Quiz

Quiz, 10 questions

**8/10 points (80%)**

1 / 1 points

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
```
```

- Align paired-end data
- Take compressed input data
- Report only the top 10 alignments for each read

**Correct**

- Take input data stored in the /home/me directory

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
```
```

- Use the default reference transcript annotation to guide assembly

**Correct**

- Generate a file named transcripts.gtf in directory Test1, containing all assembled transfrags
- Use 0.10 as the minimum level for intra-intronic transcripts

Report all isoforms whose abundance is 0.05 or more of the most abundant isoform's for that gene

# Module 4 Quiz

**8/10 points (80%)**

Quiz, 10 questions

0 / 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

**This should not be selected**

The library was strand-specific

Of the mapped mate 1 reads, 11.7% had multiple matches on the genome

94.0% of the mate 2 reads were mapped

1 / 1 points

10.

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
```
```

Locus XLOC_000042 is novel

# Module 4 Quiz

**Correct**

**8/10 points (80%)**

Quiz, 10 questions

Locus XLOC_000002 is not significantly differentially expressed between the conditions

The p-value for the t-test at locus XLOC_000002 is 0.5277

The q-value, or false discovery rate (FDR), at locus XLOC_000002 is 0.998846