Module 7: RNA-Seq Human
Module 8: RNA-Seq Pathogen

MCQ Assessment
Authors: *Nyasha Chambwe, Jon Ambler, Phelelani Mpangase*

1. The transcriptome is the entire set of expressed RNA molecules in a cell or a population of cells.
    a. False
    b. True

2. What does the 4th line of the FASTQ file format specify?
    a. header
    b. separator
    c. sequencing quality scores
    d. read sequence

3. RNA-sequencing (RNA-seq) is a technique to sequence
    a. siRNAs
    b. DNA
    c. rRNAs
    d. complementary CDNA derived from RNAs

4. RNA-seq is a flexible assay that allows you to assess all of the following ***except***
    a. differential expression
    b. expression quantification
    c. transcription factor binding
    d. differential splicing

5. Which is the ***third*** step in a typical RNA-seq library preparation protocol?
    a. adapter ligation
    b. sequencing
    c. fragmentation
    d. RNA extraction

6. Our hypothesis indicates that the expression of our gene of interest *MYFAVE* is **relatively low** but it has a large functional impact. We are going to perform RNA-seq in WT and

MYFAVE KO cell lines to determine the functional effect of *MYFAVE* loss. What is the most optimal experimental design that will best position us for this investigation?

a. perform single-cell RNA-seq on the MYFAVE KO cells
b. **sequence 5 biological replicates each of WT and KO cells at 40X depth**
c. sequence 10 biological replicates each of WT and KO cells at 5X depth
d. sequence 3 biological replicates each of WT and KO cells at 100X depth

Normalisation

7. Normalisation deals with bias that that originates from:
    a. sequence GC content
    b. differences in sequencing depth
    c. differences in gene length
    d. both B and C

8. Before normalisation, longer genes produce:
    a. less read fragments than shorter genes
    b. more read fragments than shorter genes
    c. the same amount of read fragments than shorter genes

9. Principal Component Analysis (PCA) is:
    a. useful for detecting problematic samples
    b. a dimensionality reduction method
    c. requires transformed data using something like rlog to be most effective
    d. all of the above

10. The underlying assumption in most tools for differential expression analysis is that most genes are not differentially expressed.
    a. True
    b. False

11. Genes with lower expression have:
    a. Lower levels of absolute variance compared to highly expressed genes
    b. Higher levels of absolute variance compared to highly expressed genes
    c. The same levels of absolute variance compared to highly expressed genes

Mapping

12. The main purpose(s) of read mapping/alignment in the RNA-seq data analysis workflow is to:
    a. create a *".bam"* file that can be indexed with samtools
    b. exclude low quality reads from the RNA-seq data
    c. **identify the location in the reference genome/transcriptome where the RNA-seq data originated from.**
    d. trim low quality reads from the RNA-seq data

13. RNA-seq reads can only be mapped/aligned to a reference genome.
    a. True
    b. **False**

14. Eukaryotic genomes contain introns as part of their gene structure. It is because of these introns that "splice-aware" alignment algorithms should be used when aligning to RNA-seq reads from eukaryotic genomes.
    a. **True**
    b. False

15. Which of the following does the HISAT2 aligner use in its alignment algorithm?
    a. global index
    b. local index
    c. foreign index
    d. **both (a) and (b)**

16. Pseudoaligners (e.g., Kallisto) can be used to discover novel transcripts when used to map RNA-seq reads to a reference transcriptome.
    a. True
    b. **False**

17. Mapping RNA-seq reads to a reference transcriptome is faster because
    a. transcriptome aligners use super-fast machine learning algorithms
    b. **there is less target sequence to map to since the reference transcriptome is spliced**
    c. they do not need to create a "*.bam"* file with the alignments
    d. the nucleobase thymine (T) is replaced with uracil (U) in RNA, and also RNA is single stranded

18. If your study organism has no reference genome or transcriptome to align your RNA-seq reads to, your other option is to
    a. Skip the alignment step and use Kallisto for quantification since it does not require a reference genome or transcriptome.
    b. Give up and wait until another lab has sequenced your study organism.
    c. **Reconstruct your study organism transcriptome using your RNA-seq reads and map your RNA-seq reads to the reconstructed transcriptome.**
    d. Use HISAT2 to reconstruct the reference transcriptome for your study organism

19. A splice-aware aligner is critical for RNA-seq alignments because it can handle the case where a read is split and maps to multiple exons.
    a. True
    b. False

Downstream Analysis

20. Gene set enrichment analysis is a good way to determine the main pathways or biological processes involved for your list of differentially expressed genes.
    a. False
    b. True