

NGS Bioinformatics Course

AFRICA 2021

Module 7: RNA-Seq Expression Analysis (Human)

Practical Assignment

Module topic: RNA-Seq Analysis

Contact session title:

Trainers: Nyasha Chambwe, Phelelani Mpangase, Jon Ambler

Participant: <write your name here>

Date: <write today's date here>

Introduction

This practical illustrates the principles of RNA-seq expression analysis using data derived from human subjects. The practical assignment for this module can be found in the file:

['Module7_RNASeq_Human_Practical.pdf'](#).

Follow the instructions outlined in this manual and perform each analysis step as indicated. Answer the questions at the end of each section and submit your answers by the deadline indicated on Vula.

Tools used in this session

List any tools and appropriate URLs or software download information. Please note: use web-based online applications as far as possible.

[See Section 1.5 Prerequisites in Module7_RNASeq_Human_Practical.pdf](#)

Please note

- **Hand-in information** please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Assigned Tasks

Task 1: Introducing the tutorial dataset

1. Why is there more than one FASTQ file per sample?

<start typing your answer here>

Task 2: Mapping RNA-Seq reads to the genome with HISAT2

2. How many index files were generated when you ran **hisat2-build**?

<start typing your answer here>

3. What was the overall alignment rate for the **PT2** sample to the reference genome?

<start typing your answer here>

4. How does the alignment rate compare with that of the **NP2** sample?

<start typing your answer here>

5. How many **NP2** reads were not aligned to the reference genome?

<start typing your answer here>

Task 3: Visualising transcriptomes with IGV

6. How many CDS features are there in the gene "**SIK1B**"?

<start typing your answer here>

7. Does the RNA-seq mapping agree with the gene model for SIK1B in blue?

<start typing your answer here>

8. Do you think this gene (**SIK1B**) is differentially expressed between prostate cancer and normal adjacent tissue? Is looking at the coverage plots alone a reliable way to assess differential expression?

<start typing your answer here>

Task 4: Transcript quantification with Kallisto

9. What k-mer length was used to build the Kallisto index?

<start typing your answer here>

10. How many transcript sequences are there in the reference transcript annotation ***hsapiens_chr21_transcripts.fa***?

<start typing your answer here>

11. What is the transcripts per million (TPM) value for **ENST00000399975.7 (USP16)** in each of the samples?

<start typing your answer here>

12. Do you think **ENST00000399975.7** is differentially expressed?

<start typing your answer here>

Task 5: Identifying differentially expressed genes with Sleuth

13. What is the most abundantly expressed transcript in the **PT6** sample?

<start typing your answer here>
