NGS Bioinformatics

Read alignment practical assignment

Module topic: Read alignment
Contact session title: Read alignment
Trainer: Gerrit Botha
Participant: <*write your name here*>
Date: <*write today's date here*>

Alignment, manipulation and visual inspection

Introduction

All information is found in practical material. We will use this document to answer the questions.

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.

Task 1: Q1 – Performing Read Alignment
What is the length of chromosome 7 of the mouse genome? (Hint: Look at the fasta header for chromosome 7)

<*start typing your answer here*>

Task 2: Q2 – Performing Read Alignment
How much space is saved by using a BAM file instead of a SAM file?

<*start typing your answer here*>

Task 3: Q3 – Performing Read Alignment
From looking at the output metrics file - how many reads were marked as duplicates?
What was the percent duplication?

<*start typing your answer here*>

Task 4: Q4 – Performing Read Alignment
What is the total number of reads?

<*start typing your answer here*>

Task 5: Q5 – Performing Read Alignment
What proportion of the reads were mapped?

<start typing your answer here>

Task 6: Q6 – Performing Read Alignment
How many reads were paired correctly/properly?

<start typing your answer here>

Task 7: Q7 – Performing Read Alignment
How many read pairs mapped to a different chromosome?

<start typing your answer here>

Task 8: Q8 – Performing Read Alignment
What is the insert size mean and standard deviation?

<start typing your answer here>

Task 9: Q9 – Performing Read Alignment
In your web browser open the file called md5638_plot.html to view the QC information and answer the following questions:

What is the insert size mean and standard deviation?

<start typing your answer here>

Task 10: Q10 – Performing Read Alignment
In your web browser open the file called md5638_plot.html to view the QC information and answer the following questions:

Which of the first fragments or second fragments are higher base quality on average?

<start typing your answer here>

Task 11: E1 – Alignment Visualisation
Go to chromosome chr7, positions 87,483,625-87,484,330 using the navigation bar across the top. Take in the glorious view of a genome pileup. Stop and smell the roses! Click on stuff! Scroll around, zoom in and out a bit!

<start typing your answer here>

Task 12: E2 – Alignment Visualisation
Go back to chromosome 7:87,483,625-87,484,330. What is the (rough) coverage across this region? (Hint: Look at the coverage track)

<start typing your answer here>

Task 13: E3 – Alignment Visualisation
Can you spot the three mutant variants (two small and one larger) in this region? State what the evidence is for them?

<start typing your answer here>

Task 14: Q1 – NGS Workflows
What do the -M and -R options do?

<start typing your answer here>

Task 15: Q2 – NGS Workflows
What does the -bS option do?

<start typing your answer here>

Task 16: Q2 – NGS Workflows
What does the -bS option do?

<start typing your answer here>

Task 17: Q3 – NGS Workflows
What is the total number of reads?

<start typing your answer here>

Task 18: Q4 – NGS Workflows
What proportion of the reads were mapped?

<start typing your answer here>

Task 19: Q5 – NGS Workflows
How many reads were paired correctly/properly?

<start typing your answer here>

Task 20: Q6 – NGS Workflows

How many reads mapped to a different chromosome?

<start typing your answer here>

Task 21: Q7 – NGS Workflows
What is the insert size mean and standard deviation?

<start typing your answer here>

Task 22: Q8 – NGS Workflows
In a web browser open the file called plots.html to view the QC information.

How many reads have zero mapping quality?

<start typing your answer here>

Task 23: Q9 – NGS Workflows
In a web browser open the file called plots.html to view the QC information.

Which of the first fragments or second fragments are higher base quality on average?

<start typing your answer here>

Task 24: E1 – NGS Workflows
Go to Chromosome IV and position 764,292. (Hint: use the navigation bar across the top)

<start typing your answer here>

Task 25: E2 – NGS Workflows
What is the reference base at this position?

<start typing your answer here>

Task 26: E3 – NGS Workflows
Do the reads agree with the reference base?

<start typing your answer here>

Task 27: E4 – NGS Workflows
What about the adjacent position (IV:764,293)? What is the reference base at this position?
Is it supported by the reads?

<start typing your answer here>

Task 28: E5 – NGS Workflows
Go to Chromosome IV and position 766,589.

<start typing your answer here>

Task 29: E6 – NGS Workflows
What sort of mutation are the alignments indicating might be present?

<start typing your answer here>

Task 30: E7 – NGS Workflows
Go to Chromosome IV and position 770,137 using the navigation bar across the top.

<start typing your answer here>

Task 31: E8 – NGS Workflows
What sort of mutation are the alignments indicating might be present? Is there anything in the flanking sequence of the reference genome that might make you suspicious about this mutation?

<start typing your answer here>

Task 32 E9 – NGS Workflows
Convert the BAM file to a CRAM file

<start typing your answer here>
