



QUESITI PROVA PRATICA - SELEZIONE 2025N14

PROVA N. 1

Realizzi un container software per un'analisi di allineamento di sequenza, seguendo le specifiche sotto riportate.

Creazione del container

Utilizzando un sistema di containerizzazione (es. Docker, Singularity/Apptainer), crei un container che includa:

- Un sistema operativo base (es. Ubuntu o RHEL).
- Un software di allineamento di sequenze a sua scelta (es. BWA, Bowtie2, HISAT2).
- Tutte le dipendenze necessarie per il funzionamento del software scelto.

Dimostrazione del funzionamento

Dimostri il corretto funzionamento del container eseguendo le seguenti operazioni:

- Avviare il container e verificare la presenza del software di allineamento.
- Montare una directory esterna per testare l'accesso ad un file FASTQ di esempio.
- Recuperare la versione esatta del software presente nel container.



PROVA N. 2

Effettui un'analisi di allineamento di reads rispetto a un genoma noto.

Allineamento

Utilizzando un software di allineamento di sequenze a sua scelta tra BWA, Bowtie2 o HISAT2, esegua le seguenti operazioni:

- Esegua l'allineamento di un file FASTQ fornito con un riferimento genetico specifico (cromosoma 21 del genoma umano).
- Salvi il risultato in un file BAM ordinato e indicizzato. Spieghi perché è importante ordinare e indicizzare il file BAM per analisi successive.

Conteggio delle reads allineate

- Utilizzi un comando per estrarre il numero totale di reads correttamente allineate.
- Generi un output testuale o un file tabellare contenente il numero totale di reads in input e il numero di reads correttamente allineate.



PROVA N. 3

Realizzi un'analisi esplorativa di una matrice di conte geniche, con l'obiettivo di valutare la presenza di gruppi sperimentali o effetti tecnici (batch effect) e di fornire una motivazione per l'applicazione di tecniche di normalizzazione.

Installazione degli strumenti necessari

- Identifichi un ambiente di programmazione appropriato.
- Verifichi la presenza delle librerie necessarie per l'analisi ed installi quelle eventualmente mancanti.

Analisi esplorativa

Dopo aver letto la matrice delle conte fornita in formato CSV, crei una serie di plot per valutare:

- La distribuzione delle espressioni geniche tra i campioni.
- La presenza di clustering o separazione tra gruppi di campioni (es. in base a condizioni sperimentali o a fattori tecnici, noti o ignoti).
- La variabilità tra campioni che potrebbe suggerire la necessità di normalizzazione.



PROVA N. 4

Realizzi uno script per l'analisi di varianti genomiche, in un linguaggio di programmazione a sua scelta. Il programma dovrà svolgere le seguenti operazioni:

Filtraggio delle varianti

Leggere una tabella di varianti genomiche da un file di input fornito.

Filtrare le varianti in base ai seguenti criteri:

- Cromosoma: 5
- Posizione: compresa tra 10200 e 10300 (estremi inclusi)

Scrivere le varianti filtrate in un nuovo file di output, mantenendo tutte le informazioni originali per ciascuna variante selezionata.

Calcolo statistiche

Dopo aver filtrato le varianti, produrre le seguenti statistiche di base:

- Numero totale di varianti filtrate.
- La variante con il punteggio di qualità più alto.



PROVA N. 5

Sviluppare un workflow per l'analisi di dati biologici che esegua una serie di comandi utilizzando un framework di gestione dei workflow. Il workflow dovrà gestire quattro comandi fittizi (A, B, C, D) con relazioni di input/output definite.

Installazione

Installare il framework scelto (es. Snakemake, Nextflow) e le sue dipendenze.

Sviluppo del workflow

Creare un workflow che implementi i seguenti passaggi:

- **Comando A:** Leggere un file di input (input_A.fasta) e generare un file di output (output_A.txt).
- **Comando B:** Leggere un file di input (input_B.fastq) e generare un file di output (output_B.txt).
- **Comando C:** Leggere i file output_A.txt e output_B.txt come input e generare un file (output_C.csv).
- **Comando D:** Leggere output_C.csv come input e generare un file finale (result_final.txt).

Esecuzione

Eseguire il workflow, verificando che il comando C venga eseguito *solamente* dopo il completamento di A e B.