

Progetto 1 (uno studente)

Dato un sample in formato BAM e un insieme di trascritti in formato GTF, produrre uno script Python (non un notebook) che fornisca in output per ogni trascritto quante reads lo supportano. Produrre inoltre un file snakemake con le regole necessarie per lanciare tale script al variare dei file in input.

Tipicamente possono succedere casi che rendono il problema sempre più complesso, ad esempio le reads non coprono l'intero trascritto (e.g., trascritto di 2.000 basi e reads di lunghezza 150); le reads potrebbero non terminare esattamente nelle posizioni di taglio introni/esoni; le reads potrebbero avere qualità di mapping diverse; etc. Non è necessario considerare tutti i casi possibili ma si possono fare delle assunzioni ed ignorare casi che complicano il progetto. Pertanto, è necessario specificare quali assunzioni vengono fatte, ovviamente la valutazione terrà conto anche di quante semplificazioni sono state effettuate.