

Progetto corso bioinformatica

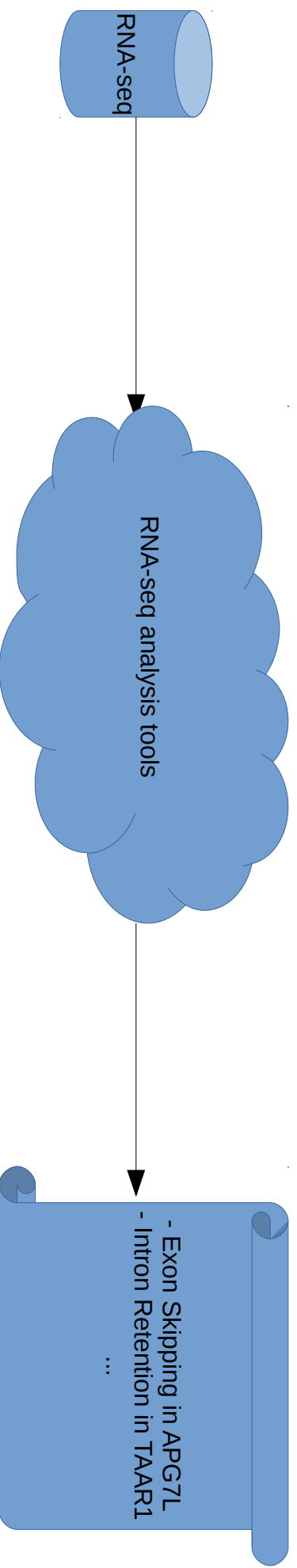
Proposta: implementazione di un software che, dato un dataset di RNA-seq in input, produca in output una lista di geni che potrebbero essere interessati da tale dataset.

Possible approccio: utilizzare le tecniche apprese durante il corso per implementare un filtro basato sull'analisi dei k-mer presenti nel dataset e nei vari geni.



Progetto corso bioinformatica

Problema: dato un dataset di un esperimento RNA-seq, si vogliono descrivere gli eventi di splicing espressi



Progetto corso bioinformatica

Problema: dato un dataset di un esperimento RNA-seq, si vogliono descrivere gli eventi di splicing espressi



Progetto corso bioinformatica

Problema: dato un dataset di un esperimento RNA-seq, si vogliono descrivere gli eventi di splicing espressi

Note:

- * La maggior parte dei geni non è espressa
- * Un buon numero di analisi sono inutili
- * Eseguire una singola esecuzione richiede tempo



Progetto corso bioinformatica

Problema: dato un dataset di un esperimento RNA-seq, si vogliono descrivere gli eventi di splicing espressi

Note:

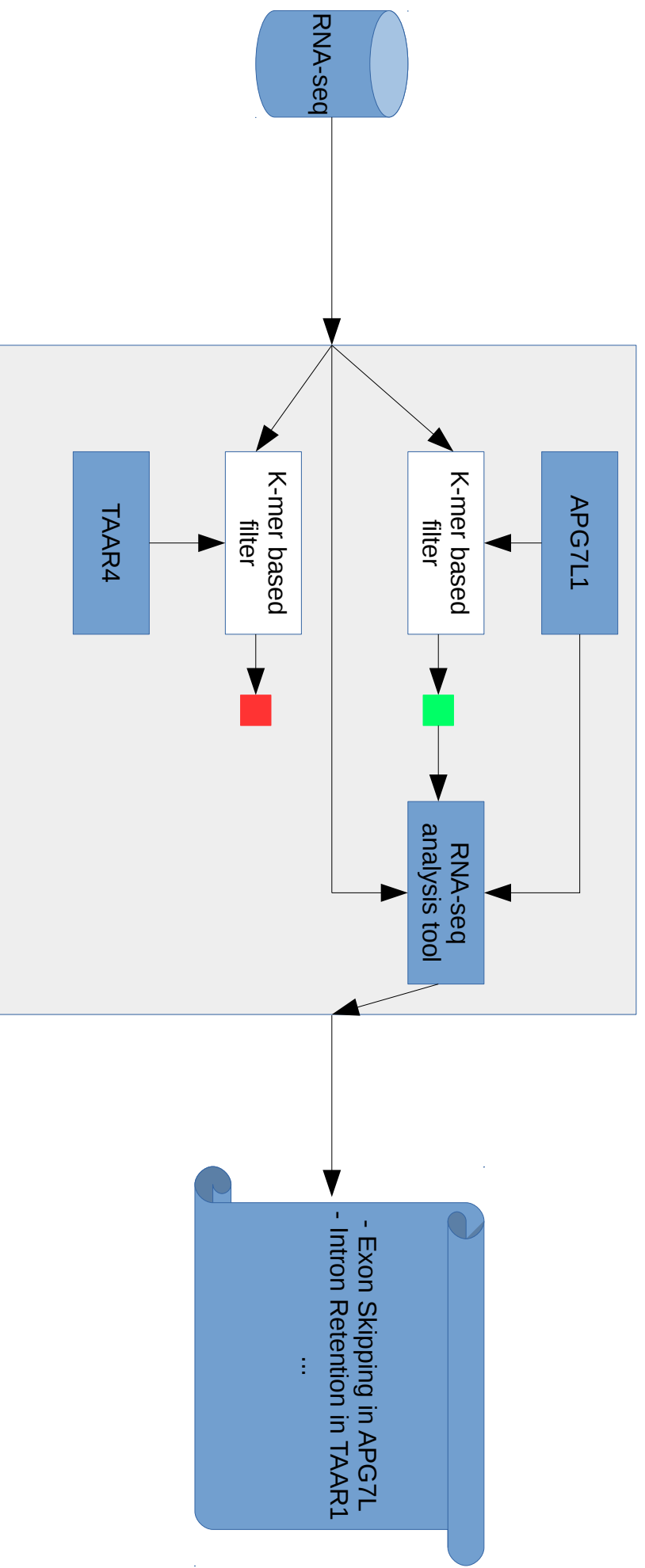
- * La maggior parte dei geni non è espressa
- * Un buon numero di analisi sono inutili
- * Eseguire una singola esecuzione richiede tempo

Proposta: implementazione di un filtro che permetta di rimuovere dall'analisi i geni sicuramente non interessanti



Progetto corso bioinformatica

Proposta: implementazione di un filtro che permetta di rimuovere dall'analisi i geni sicuramente non interessanti



Progetto corso bioinformatica

Proposta: implementazione di un filtro che permetta di rimuovere dall'analisi i geni sicuramente non interessanti

Possibile approccio:

- * Creare un dizionario basato sui k-mer che associ k-mers e geni (indice statico, creato una volta sola)
- * Dato un esperimento RNA-seq, fare un'analisi dei k-mer contenuti per poter filtrare i geni che sicuramente non sono espressi dall'esperimento (no falsi negativi)

Direzioni:

- * Esistono software che utilizzano indici basati sui k-mer (Sailfish? Salmon? Sequence bloom trees? Altro?)
- * Creare l'indice una sola volta offline
- * Effettuare query sull'indice in parallelo

Progetto corso bioinformatica

Note

- * Documentate il codice
- * Scrivete un manuale per l'utente
- * Sviluppate usando sistemi di controllo versione (preferibilmente github utilizzando repository del gruppo di ricerca AlgoLab*)

* <https://github.com/AlgoLab/>

