

Advanced NGS Analysis (Day 2)

Session IV

Lyda Hill department of Bioinformatics 2022 Nanocourse Series

Date & Time: June 27-28: 9AM-5PM (NG3.202)

Course Instructors: Bo Li, Daehwan Kim, Christopher Chaney, & Micah Thornton

UTSouthwestern
Medical Center



Day 2: RNA-Seq Analysis Using Pseudo/Quasi-Alignment and Expectation Maximization (Kallisto, Salmon, H2Q).

What you will learn in this Session (2 Parts)

- *Some Theoretical Considerations:*
 - *What is Pseudo/Quasi-Alignment*
 - *What is Alignment*
 - *What is Expectation Maximization*
 - *What is Expectation Maximization for Gene Transcript Quantification*
 - *What is the resolution of Genetic Transcript Data?*
- *Some Practical Considerations:*
 - *What is Kallisto*
 - *What is Salmon*
 - *What is H2Q*

Part 1: Pseudo and Quasi Alignment & Quantification Resolution

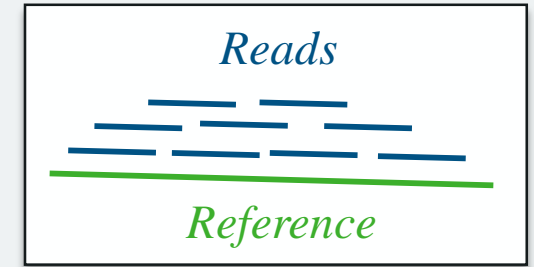
RNA-Seq Experiments

- RNA-Seq experiments are fundamentally distinct from DNA-seq experiments, and seeks to answer a different set of questions.
- Usually we are seeking to determine whether the level of expression of a particular gene is related to a phenotypic characteristic of interest.

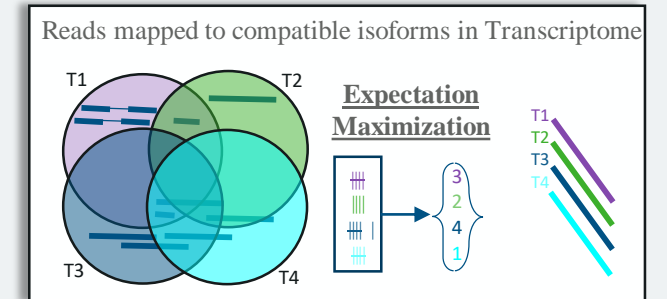
Pseudo/Quasi Alignment in RNA Experiments

- Sometimes the *exact* position of a sequencing read is not of critical import.
 - There are a few approaches for resolving the *approximate* location of a read.
 - Procedures work by determining the subset of *transcript isoforms* compatible with a read.
 - Two such approaches are known as:
 - Pseudo-Alignment
 - The Approach used by **Kallisto**.
 - Uses the De Bruijn ('Deh-Broine') graph procedure.
 - Quasi-Alignment
 - The Approach used by **Salmon**.
 - Uses a *K*-mer Hash table and Suffix Array.

Typical 'DNA-Seq Like' Experiment



Typical 'RNA-Seq Like' Experiment



Recall that in most typical sequencing experiments we are dealing with a large collection of shorter subsequences called reads, which we attempt to map to a larger sequence known as the reference.

Resources – Kallisto (Pseudo-alignment)

1. <https://tinyheero.github.io/2015/09/02/pseudoalignments-kallisto.html> (Higher Level Overview pseudo alignment)
2. <https://www.youtube.com/watch?v=f-ecmECK7lw> (Video Describing how To Build The De Bruijn graph)
3. <https://www.nature.com/articles/nbt.2023> (Nature Primer on Using De Bruijn Graphs for Genomic Alignments).

Resources – Salmon (Quasi-alignment)

1. https://hbctraining.github.io/Intro-to-rnaseq-hpc-salmon-flipped/lessons/08_quasi_alignment_salmon.html (Higher Level Overview Quasi-Alignment)
2. <https://academic.oup.com/bioinformatics/article/32/12/i192/2288985?login=true> (RapMap Paper and Description).

Part 2: Expectation Maximization & Gene Transcript Quantification