illumına®

RNA Sequencing Analysis Using TopHat

The Tuxedo Suite, comprising Bowtie, TopHat, and Cufflinks, is widely adopted for RNA sequencing analysis, and can be run in multiple modes.

Tuxedo Suite

The Tuxedo suite is a free, open-source collection of applications that has been widely adopted as analysis toolset for differential gene expression, discovery of splice forms, and fast alignment of short reads. Some of the applications used in RNA sequencing analysis are the following:

- **Bowtie**[1] is an ultrafast, memory-efficient aligner for large sets of short reads.
- **TopHat**[2] uses using Bowtie to align RNA-Seq reads to mammalian-sized genomes, and then analyzes the mapping results to identify splice junctions between exons.
- **Cufflinks**[3] assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples.

The RNA Sequencing analysis can either be run using the command line or the workflow manager Galaxy.

Command Line

The various applications in the Tuxedo suite all have their instructions for using the command-line interface. You can use these, but you will need to understand where one application ends and the other starts. To make it easy for you, we have generated a guide that covers the entire workflow for RNA Sequencing, which can be downloaded from Illumina's website:

http://www.illumina.com/documents/products/technotes/RNASeqAnalysisTopHat.pdf

Before running TopHat from the command line, you need to install the various applications on one of your servers. The same TopHat User Guide also provides installation instructions for all the applications in one place.

Galaxy

Galaxy (galaxy.psu.edu) is a web-based, open-source workflow manager with a thriving user base, documentation videos, and public forums for community questions[4,5]. Galaxy provides a workflow for running TopHat and associated applications from the Tuxedo suite:

https://main.g2.bx.psu.edu/root?tool_id=tophat

A good tutorial running TopHat through Galaxy is available here:

https://main.g2.bx.psu.edu/u/jeremy/p/galaxy-rna-seq-analysis-exercise

Galaxy can be run using the free public server, but if you plan on analyzing large datasets or sensitive samples, it may be better to install your own Galaxy instance, either locally or on the cloud. This is explained on the following pages:

http://wiki.g2.bx.psu.edu/Admin/Get Galaxy

References

1. Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25.
2. Trapnell C, Pachter L, Salzberg SL.(2009) TopHat: discovering splice junctions with RNA-Seq. Trapnell C, Pachter L, Salzberg SL. Bioinformatics 25(9):1105-11.
3. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28(5):511-5.
4. Langmead B, Trapnell C, and S Salzberg. (2009) Ultrafast and memoryefficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.
5. Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. 2010 Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11(8):R86.

FOR RESEARCH USE ONLY