RMTA v2.1

The DE Quick Start tutorial provides an introduction to basic DE functionality and navigation.
Please work through the documentation and add your comments on the bottom of this page, or email comments to support@cyverse.org. Thank you.

Rationale and background:

RMTA is a workflow that can rapidly process raw RNA-seq Illumina data by mapping reads using HiSat2 and then assemble transcripts using either Cufflinks or Stringtie. RMTA can process Fastq files containing paired-end or single-end reads. Alternatively, RMTA can directly process one or more sequence read archives (SRA) from NCBI using an SRA ID.

RMTA minimally requires the following input data:

1. Reference Genome (FASTA) or Hisat2 Indexed Reference Genome (in a subdirectory)
2. Reference Transcriptome (GFF3/GTF/GFF)
3. RNA-Seq reads (FASTQ) - Single end or Paired-end or NCBI SRA id or multiple NCBI SRA id's (list in a single column text file).

Pre-Requisites

1. A CyVerse account. (Register for a CyVerse account here - user.cyverse.org)
2. Mandatory arguments
    a. **Hisat2 reference genome:** Select at least one of the below three options for the indexing of the Reference Genome
        i. Custom Reference genome
        ii. Select reference genome from the list
        iii. Hisat2 Indexed folder
    b. **Hisat2 reference annotation**: Select at least one of the below two options for using as annotation
        i. Custom Reference annotation
        ii. Select reference annotation from the list

    Use one of the following three:
    c. **Paired-end reads**
        i. FASTQ Files (Read 1): Input reads 1 file of paired-end data
        ii. FASTQ Files (Read 2): Input reads 2 files of paired-end data
    d. **Single-end reads**
        i. single end FASTQ files
    e. **SRA**
        i. SRA ID: Single SRA id that you want to use
        ii. File containing SRA id's: Multiple SRA's that you want to use
    f. **Cufflinks/Stringtie**: Only one of the below two options needs to be checked. Cannot select both
        i. StringTie
        ii. Cufflinks
        iii. Coverage cut-off threshold: Select from 0-5
        iv. FPKM cut-off threshold: FPKM cut-off you want to use to filter the transcripts
    g. Cuffmerge: Run Cuffmerge for Stringtie/Cufflinks gtfs (Only works with more than one sample files)
3. Advanced options
    a. Phred quality score: encoding for quality score: Phread64 (Default is Phred 33)
    b. Fragment Library Type: specify the format of the library either FR, RF, F, R etc.
    c. Trim bases from 5' end of read: Trim bases from 5' (left) end of each read before alignment
    d. Trim bases from 3' end of read: Trim bases from 3' (right) end of each read before alignment
    e. Minimum intron length: Set minimum intron length
    f. maximum intron length: Set maximum intron length

Test/sample data

The following test data are provided for testing RMTA in here - /iplant/home/shared/iplantcollaborative/example_data/RMTA

1. Reference Genome: Sorghum_bicolor.Sorbi1.20.dna.toplevel_chr8.fa
2. Reference Annotation: Sorghum_bicolor.Sorbi1.20_chr8.gtf
3. left_reads- sample_1_R1.fq.gz
4. right_reads-sample_1_R2.fq.gz
5. Stringtie
6. Fragment Library Type: FR

Leave the rest as default

Results

Successful execution of RMTA will generate two output folders

1. Index: This folder consists of the index of the genome
2. Output: This folder consists of the output from Hisat2, Stringtie and Cuffcompare (Please refer to the manual for the explanation of outputs from these individual programs)