RapidRNAseq - Kallisto wrapper, developed by Kevin Blighe (k.blighe@qmul.ac.uk) at Queen Mary University of London

Release date: Thursday 7th June 2016

Version: Version 3

Introduction

RapidRNAseq is a wrapper for Kallisto, for use on the MATURA eMedLab virtual environment. Kallisto rapidly quantifies RNA-seq reads over a reference annotation FASTA file.

Program syntax

Paired-end

```
RapidRNAseqv3.sh PAIRED [Rep1MatePair1,Rep1MatePair2,Rep2MatePair1,Rep2MatePair2,...,] [Reference FASTA/FASTA.gz] [Bootstrap quantification value] [Output dir] [Output prefix]
```

Single-end

```
RapidRNAseqv3.sh SINGLE [Fragment (read) length] [Fragment (read) length standard deviation] [[Rep1,Rep2,...,] [Reference FASTA/FASTA.gz] [Bootstrap quantification value] [Output dir] [Output prefix]
```

Command-line parameters

Paired-end

1. PAIRED

2. Comma-separated list of mate-pair 1 and 2 reads, with support for replicate sample analysis
   (FASTQ or FASTQ.gz format)

3. Reference FASTA against which quantification is made. Two are currently available:
   a. EBI: /reference/gencode.v24.transcripts.fa
   b. NCBI: /reference/Homo_sapiens.GRCh38.rel79.cdna.all.fa
   (both contain most up-to-date cDNA FASTA sequence for coding and non-coding transcripts)

4. Sample bootstrap value for quantification
   (typically 100)

5. Output directory
   (you'll require write-access to this directory, i.e., your home directory)

6. A prefix to add to the output files

Single-end

1. SINGLE

2. Estimated fragment/read length
   (check library preparation method to infer this – usually 150-200 for Illumina paired-end)

3. Fragment/read length standard deviation
   (can be inferred from quality control checking of reads prior to quantification - generally 1-20)

4. Sample reads, or comma-separated list of replicate sample reads
   (FASTQ or FASTQ.gz format)

5. Reference FASTA against which quantification is made. Two are currently available:
   a. EBI: /reference/gencode.v24.transcripts.fa
   b. NCBI: /reference/Homo_sapiens.GRCh38.rel79.cdna.all.fa
   (both contain most up-to-date cDNA FASTA sequence for coding and non-coding transcripts)

6. Sample bootstrap value for quantification
   (typically 100)

7. Output directory
   (you'll require write-access to this directory, i.e., your home directory)

8. A prefix to add to the output files

Example

Paired-end, single sample

```
RapidRNAseqv3.sh paired QMUL001_1.fastq.gz,QMUL001_2.fastq.gz /reference/gencode.v24.transcripts.fa
100 /home/kevinb/RNAseq/ QMULL001
```

Paired-end, replicate sample

```
RapidRNAseqv3.sh paired Rep1QMUL001_1.fastq.gz,Rep1QMUL001_2.fastq.gz,
Rep2QMUL001_1.fastq.gz,Rep2QMUL001_2.fastq.gz /reference/gencode.v24.transcripts.fa 100
/home/kevinb/RNAseq/ QMUL001
```

Single-end, single sample

```
RapidRNAseqv3.sh single 200 5 QMUL001.fastq.gz /reference/gencode.v24.transcripts.fa 100
/home/kevinb/RNAseq/ QMUL001
```

Single-end, replicate sample

```
RapidRNAseqv3.sh single 200 5 Rep1QMUL001.fastq.gz,Rep2QMUL001.fastq.gz
/reference/gencode.v24.transcripts.fa 100 /home/kevinb/RNAseq/ QMUL001
```

Further information

http://pachterlab.github.io/kallisto/