Jun 13, 2018

RNA-seq quantification from published data

DOI
dx.doi.org/10.17504/protocols.io.qxhdxj6

Thomas Dugé[1], Franziska Liesecke[1]

[1]Université de Tours

Thomas Dugé

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.qxhdxj6

Protocol Citation: Thomas Dugé, Franziska Liesecke 2018. RNA-seq quantification from published data. **protocols.io https://dx.doi.org/10.17504/protocols.io.qxhdxj6**

Protocol status: Working
We use this protocol and it's working

Created: June 13, 2018

Last Modified: June 13, 2018

Protocol Integer ID: 13001

Abstract

This workflow described how to download RNA-seq raw data from NCBI or EBI and to process them to quantify transcript abundance.

1 Identify RNA-seq runs of interests and get a DRR/ERR/SRR accession number.
In a bash script, assign the accession number to the $acc_sra variable.

Command

use Prozilla to download data from the EBI (Linux)

```
 if [ ${#acc_sra} -eq 9 ]; then proz -k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/${acc_sra}/${acc_sra}_1.fa
stq.gz -P $acc_wd gunzip ${acc_sra}_1.fastq.gz proz -k=8 --force -r --
no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/${acc_sra}/${acc_sra}_2.fa
stq.gz -P $acc_wd gunzip ${acc_sra}_2.fastq.gz else if [ ${#acc_sra} -
eq 10 ]; then proz -k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/00${acc_sra:9:9}/${acc_sra
}/${acc_sra}_1.fastq.gz -P $acc_wd gunzip ${acc_sra}_1.fastq.gz proz -
k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/00${acc_sra:9:9}/${acc_sra
}/${acc_sra}_2.fastq.gz -P $acc_wd gunzip ${acc_sra}_2.fastq.gz else
if [ ${#acc_sra} -eq 11 ]; then proz -k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/0${acc_sra:9:10}/${acc_sra
}/${acc_sra}_1.fastq.gz -P $acc_wd gunzip ${acc_sra}_1.fastq.gz proz -
k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/0${acc_sra:9:10}/${acc_sra
}/${acc_sra}_2.fastq.gz -P $acc_wd gunzip ${acc_sra}_2.fastq.gz else
proz -k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/${acc_sra:9:11}/${acc_sra}
/${acc_sra}_1.fastq.gz -P $acc_wd gunzip ${acc_sra}_1.fastq.gz proz -
k=8 --force -r --no-curses
ftp.sra.ebi.ac.uk/vol1/fastq/${acc_sra:0:6}/${acc_sra:9:11}/${acc_sra}
/${acc_sra}_2.fastq.gz -P $acc_wd gunzip ${acc_sra}_2.fastq.gz fi fi
fi
```

Software

NAME	OS	DEVELOPER	SOURCE LINK
Prozilla	Linux	Kalum Somaratna	https://github.com/totosugito/prozilla-2.0.4

2 Trim reads with Trimmomatic. Assign to $adaptfile variable the appropriate file name containing primer sequences (found in the Trimmomatic folder).

Software

NAME	OS	DEVELOPER	SOURCE LINK
Trimmomatic	Linux	USADELLAB	http://www.usadellab.org/cms/?page=trimmomatic

Command

Trim paired RNA-seq reads with Trimmomatic (Linux)

```
java -jar /home/tduge/trinity/Trimmomatic-0.32/trimmomatic-0.32.jar
PE -threads $thread_number ${acc_sra}_1.fastq ${acc_sra}_2.fastq
${acc_sra}_1P.fq ${acc_sra}_1U.fq ${acc_sra}_2P.fq ${acc_sra}_2U.fq
ILLUMINACLIP:$adaptfile:2:30:10 LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
```

Software

NAME	OS	DEVELOPER	SOURCE LINK
Trimmomatic	Linux	USADELLAB	http://www.usadellab.org/cms/?page=trimmomatic

3 Quantify transcript abundance against a reference transcriptome assembly.
Assign the reference transcriptome to the $sid variable (the fasta assembly may be downloaded in Ensembl for example).

Software

NAME	OS	DEVELOPER	SOURCE LINK
Salmon	Linux	Rob Patro	https://combine-lab.github.io/salmon/

Command

Prepare reference index and quantify transcript abundance (Linux)

```
forward=${acc_sra}_1P.fq
reverse=${acc_sra}_2P.fq
salmon index -t $assembly -i ${assembly}_quasi_index --type quasi -k
31
salmon quant -i $sid -l A -p $thread_number --useVBOpt -1 $forward -2
$reverse -o ${acc_sra}_transcripts_quant_quasi_vbo --seqBias
```