Apr 16, 2019

Working

RNA-Seq Data Analysis (Bowtie-TopHat-Cufflinks)

Version 3

PLOS One

Kiichi Hirota[1]

[1]Kansai Medical University

dx.doi.org/10.17504/protocols.io.x9qfr5w

Yoshiyuki Matsuo

EXTERNAL LINK

https://doi.org/10.1371/journal.pone.0215072

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Sumi C, Matsuo Y, Kusunoki M, Shoji T, Uba T, Iwai T, Bono H, Hirota K (2019) Cancerous phenotypes associated with hypoxia-inducible factors are not influenced by the volatile anesthetic isoflurane in renal cell carcinoma. PLoS ONE 14(4): e0215072. doi: 10.1371/journal.pone.0215072

MATERIALS

NAME	CATALOG #	VENDOR
RNeasy® Mini Kit	74104	Qiagen
TruSeq® Stranded mRNA Library Prep Kit	RS-122-2103	illumina

MATERIALS TEXT

Equipments

- Illumina HiSeq 2500 platform
- MacBook Pro (Mid 2012) OS 10.12.2, Processor: 2.7 GHz Intel Core i7, Memory: 16 GB 1600 Mhz DDR3

Applications

- FastQC v 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- fastx_toolkit v 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/)
- Bowtie v.2.2.9
- TopHat v.2.1.1
- Samtools v.1.3.1
- Cufflinks v2.1.1
- TIBCO Spotfire Desktop v7.6.0 with the “Better World” program license (TIBCO Spotfire, Palo Alto, CA, USA, http://spotfire.tibco.com/better-world-donation-program/)
- Metascape (http://metascape.org/)

Library preparation

1 Total RNA was extracted from cells using RNeasy® Mini Kit (Qiagen).

2 Poly(A) RNA libraries were constructed using TruSeq® Stranded mRNA Library Prep Kit (Illumina).

Sequencing

3 The libraries were sequenced at 100 bp paired-ends on an Illumina HiSeq 2500 platform.

4 Sequencing data in FASTQ format were deposited in the DDBJ Sequence Read Archive.

Data analysis

5 The quality of sequence data was evaluated by FastQC after the trimming process by fastx_toolkit.

6 The human reference sequence file (hs37d5.fa) was downloaded from the 1000 genome ftp site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/), and the annotated general feature format (gff) file was downloaded from the Illumina iGenome ftp site (ftp:// igenome:G3nom3s4u@ussd-ftp.illumina.com/Homo_sapiens/NCBI/build37.2/).

7 The human genome index was constructed with bowtie-build in Bowtie.

8 The fastq files were aligned to the reference genomic sequence by TopHat with default parameters. Bowtie2 and Samtools was used with the TopHat program.

9 Estimation of transcript abundance was calculated, and the count values were normalized to the upper quartile of the fragments per kilobase of transcript per million fragments mapped reads (FPKM) using Cufflinks (cuffdiff).

10 Metascape was used for the gene set enrichment analysis. A gene list for metascape analysis was generated using the output from the cuffdiff program.

11 Gene ontology annotations were extracted using Ensembl Biomart and sorted by corresponding values (common logarithms of ([FPKM of RCC4-EV] + 1)/([FPKM of RCC4-VHL] + 1)) calculated from the same cuffdiff output file.

12 The integer one was added to all FPKM values because we cannot calculate the logarithm of 0. The histogram was generated using TIBCO Spotfire Desktop.