MetaMap pipeline

Lukas Simon,Sebastian Karg

Abstract

We developed a bioinformatic two-step pipeline, so called "MetaMap" pipeline, to screen human RNA-seq datasets for the presence of microbial and viral reads by re-inspecting the non-human-mapping read fraction.

In the first step, reads are aligned against the human genome with STAR (version 2.5.2) and subsequently the non-human-mapping reads are subjected to metatgenomic classification using CLARK-S (version 1.2.3).

Citation: Lukas Simon,Sebastian Karg MetaMap pipeline. **protocols.io**
dx.doi.org/10.17504/protocols.io.msec6be
Published: 05 Feb 2018

Protocol

Human mapping

Step 1.

cmd COMMAND (STAR - 2.5.2)

```
/naslx/projects/t1172/di36dih/Linux_x86_64_static/STAR \
--readFilesIn ${dir}/${srr}*gz \
--genomeDir /naslx/projects/t1172/di36dih/Genomes/hg38 \
--runThreadN 28 \
--genomeLoad LoadAndKeep \
--readFilesCommand zcat \
--quantMode GeneCounts \
--outReadsUnmapped Fastx \
--outFileNamePrefix ${dir}/$srr.human
```

STAR command, using the parameters: --readFilesIn Takes gz compressed FASTQ files of SRA run ${srr} as input. --genomeDir The hg38 human reference genome --genomeLoad LoadAndKeep This loads the genome into memory once and keeps it there for subsequent STAR runs --runThreadN Using 28 cores. --readFilesCommand zcat For input files in .gz compression. --quantMode GeneCounts This generates the gene count table. --outReadsUnmapped Fastx This saves all unmapped reads in a file in .fastq format. --outFileNamePrefix Saving all Files in the work directory with the prefix ${srr}.human

Metagenomic classification

Step 2.

cmd COMMAND (CLARK - 1.2.3.1)

```
sh estimate_abundance.sh -F ${srr}.metagenome.results.csv -D $db > ${srr}.abundance.tsv
```

Here the ${srr}.metagenome.results.csv files from the step before are evaluated and the abundance of the metagenomes is saved in the files ${srr}.abundance.tsv.