homework_05

Esteban Jorquera; Marlon Aldair; Diego Ramírez

2022-03-07

Github link: https://github.com/EstebanJorquera/UNAM_BIOinfoII.git

```
# Note installing packages through knitter is not the best idea,
# the following code will likely fail if the package is not already installed, with a missing mirror er
# for this reason there's an included R script just for that
# anyway this meant to be run on a cluster rather than locally, so happy copy pasting :)!
# Installs required packages if not already installed (avoids re installing)
if (!requireNamespace("BiocManager", quietly = TRUE))     install.packages("BiocManager")

if (!requireNamespace("tidyverse", quietly = TRUE))       install.packages("tidyverse")
if (!requireNamespace("dplyr", quietly = TRUE))           install.packages("dplyr")
if (!requireNamespace("tidyr", quietly = TRUE))           install.packages("tidyr")
if (!requireNamespace("ggplot2", quietly = TRUE))         install.packages("ggplot2")
if (!requireNamespace("wordcloud", quietly = TRUE))       install.packages("wordcloud")
if (!requireNamespace("RColorBrewer", quietly = TRUE))    install.packages("RColorBrewer")

# Libraries
library(tidyverse)
library(dplyr)
library(tidyr)
library(ggplot2)
library(wordcloud)
library(RColorBrewer)

BiocManager::install(version = "3.13")
packages = c("DESeq2", "tximport")
BiocManager::install(packages)
```

```
#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
```

```
#$ -N Esteban_kallisto_index

# Send an email after the job has finished
#$ -m e
#$ -M eijorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads deeptools module,
# executes bamCoverage for bam to bw conversion of the mus musculus ChIP-seq alignment data
(module load kallisto/0.45.0 ;
wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M28/gencode.vM28.transcripts.fa.g
kallisto index -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 /mnt/Ci
mkdir /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/out/kallisto ;
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt

#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
#$ -N Esteban_kallisto_quant

# Send an email after the job has finished
#$ -m e
#$ -M eijorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads deeptools module,
# executes bamCoverage for bam to bw conversion of the mus musculus ChIP-seq alignment data
(module load kallisto/0.45.0 ;
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
kallisto quant -i /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/index_kallisto45_gencode-m28 -o /mnt
```

```r
getwd()
setwd("/mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/output")

files <- file.path("/mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_5/output",list.dirs(dir(".")),"abund
names(files) <- str_extract(files,"SRR\\d+") # so that tximport identifies them
```

```
files

# Load table with trx id and gene id correspondence
tx2gene <- read.csv("/mnt/Timina/bioinfoII/rnaseq/resources/gencode/gencode.vM28.basic.trx_id-gene_id-n

# Load table with trx id and gene name correspondence
tx2genename <- read.csv("/mnt/Timina/bioinfoII/rnaseq/resources/gencode/gencode.vM28.basic.trx_id-gene_

# Run tximport
# tx2gene links transcript IDs to gene IDs for summarization
txi.kallisto <- tximport(files, type = "kallisto", tx2gene = tx2gene, ignoreAfterBar=TRUE, ignoreTxVers
txi.kallisto.name <- tximport(files, type = "kallisto", tx2gene = tx2genename, ignoreAfterBar=TRUE, ign

names(txi.kallisto)
head(txi.kallisto$counts)
```