

A Tidy Transcriptomics Introduction to RNA-seq Analysis

Jian

16/07/2020

Introduction

```
# load libraries

# tidyverse core packages
library(tibble)
library(dplyr)
library(tidyr)
library(readr)
library(stringr)
library(ggplot2)

# tidyverse friendly packages
# (BiocManager::install("tidybulk"))
# (BiocManager::install("ComplexHeatmap"))
library(tidyHeatmap)
library(tidybulk)
library(ggrepel)
library(plotly)

# dataset (BiocManager::install("airway"))
library(airway)
```

Airway RNA-seq dataset

Here we'll perform RNA-seq analyses with data from the *airway* package. It includes 4 cell line samples treated with dex and 4 cell line control samples.

Setting up the data

Convert the data from a *RangedSummarizedExperiment* object into a *tibble* object.

```
# Load airway RNA-seq data
data(airway)

# Conver to tidybulk tibble
counts_airway <- airway %>%
  tidybulk()
```

```
# View data
counts_airway
```

```
## Warning: '...' is not empty.
##
## We detected these problematic arguments:
## * 'needs_dots'
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?

## # A tibble: 512,816 x 12
##   feature sample counts SampleName cell dex albut Run avgLength Experiment
##   <fct>   <fct>   <int> <fct>      <fct> <fct> <fct> <fct>      <int> <fct>
## 1 ENSG00~ SRR10~     679 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 2 ENSG00~ SRR10~      0 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 3 ENSG00~ SRR10~    467 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 4 ENSG00~ SRR10~    260 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 5 ENSG00~ SRR10~     60 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 6 ENSG00~ SRR10~      0 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 7 ENSG00~ SRR10~   3251 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 8 ENSG00~ SRR10~   1433 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 9 ENSG00~ SRR10~    519 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## 10 ENSG00~ SRR10~    394 GSM1275862 N613~ untrt untrt SRR1~     126 SRX384345
## # ... with 512,806 more rows, and 2 more variables: Sample <fct>,
## #   BioSample <fct>
```