

Bulk RNA-seq Cheatsheet

The tables below consist of valuable codes that will help you through this module. Each table represents a different library/tool and its corresponding commands.

Please note that these tables are not intended to tell you all the information you need to know about each command. The hyperlinks found in each piece of code will take you to the documentation for further information on the usage of each command.

Base R

Read the Base R package documentation [here](#).

Library/Package	Piece of Code	What it's called	What it does
Base R	<code>list.files()</code>	List files	Produces a character vector of files or directories in the specified directory
Base R	<code>names()</code>	Names	Gets or sets the names of an object
Base R	<code>colnames()</code>	Column names	Gets or sets the column names of a matrix or data frame
Base R	<code>all.equal()</code>	All equal	Checks if two R objects are nearly equal
Base R	<code>attr()</code>	Object Attributes	Gets or sets the attributes of an object
Base R	<code>rowSums()</code>	Row Sums	Returns the sum of the rows in a numeric matrix-like object (i.e., a matrix, data.frame, etc.)
Base R	<code>relevel()</code>	Relevel	Reorders the levels of a factor as specified
Base R	<code>summary()</code>	Object Summary	Returns a result summary of an object
Base R	<code>as.data.frame()</code>	Data Frame	Checks if an object is a data.frame, and transforms the object into one, if possible

DESeq2

Read the DESeq2 package documentation [here](#), and the package vignette by Love, Anders, and Huber [here](#).

Library/Package	Piece of Code	What it's called	What it does
DESeq2	<code>rlog()</code>	Regularized Log Transformation	Returns log2 transformed count data
DESeq2	<code>vst()</code>	Variance Stabilizing Transformation	Estimates the dispersion trend of count data
DESeq2	<code>DESeqDataSetFromTximport()</code>	DESeqDataSet constructor	Creates a DESeqDataSet object

Library/Package	Piece of Code	What it's called	What it does
DESeq2	DESeq()	Differential Expression Analysis Based on the Negative Binomial Distribution	Estimates size factors, estimates dispersion, and performs negative binomial fitting and Wald statistics as steps in the default differential expression analysis
DESeq2	plotPCA()	PCA plot	Produces a principal component analysis plot for transformed data, while checking for batch effects
DESeq2	counts()	Counts	Saves the count data as a matrix
DESeq2	results()	Results	Returns the results table from a DESeq2 analysis
DESeq2	assay()	Assay	Returns count values in a matrix

FastQC and fastp

Read the **FastQC** documentation [here](#) and the **fastp** documentation [here](#).

Library/Package	Piece of Code	What it's called	What it does
fastp	fastp	FASTQ preprocessor	Preprocesses FASTQ files through adapter trimming, quality filtering, length filtering, and a number of additional options
FastQC	fastqc	FASTQC (Quality Control)	Performs quality control checks on raw sequence data and outputs a QC(quality control) report

ggplot2

Read the **ggplot2** package documentation [here](#). A vignette on the usage of the **ggplot2** package can be found [here](#).

Library/Package	Piece of Code	What it's called	What it does
ggplot2	ggsave()	GG Save	Saves the last plot in working directory
ggplot2	last_plot()	Last plot	Returns the last plot produced
ggplot2	geom_point()	Geom point	Creates a scatterplot (when added to the ggplot() function)
ggplot2	xlab(); ylab()	X Axis Labels; Y Axis Labels	Modifies the labels on the x axis and on the y axis, respectively
ggplot2	coord_fixed()	Cartesian Coordinates with Fixed Aspect Ratio	Coerces the coordinates on the plot to represent a fixed specified ratio

stringr, readr, tximport, dplyr, pheatmap

Documentation for each of these packages can be accessed by clicking the package name in the table below.

Library/Package	Piece of Code	What it's called	What it does
stringr	word()	Word	Extracts words from a character vector
readr	write_rds()	Write RDS	Writes data to a .rds output file
tximport	tximport()	Tximport	Imports transcript-level estimates and summarizes abundances, counts, and transcript lengths to the gene-level by default
dplyr	pull()	Pull	Extracts a variable as a vector
pheatmap	pheatmap()	Pheatmap	Plots clustered heatmaps

Salmon

Read the Salmon tool documentation [here](#).

Tool	Piece of Code	What it's called	What it does
Salmon	salmon index	Salmon index	Builds a transcriptome index which is required for Salmon quantification (from the command line)
Salmon	salmon quant	Salmon quantification	Runs Salmon's quantification of transcript expression (from the command line)

Useful command line commands:

```
rstudio@e549475e8bdb:~/kitematic$ ls
ccd1-training-waiver.DOCX  docker-install  intro-to-R-tidyverse  LICENSE.md  machine-learning  README.md  RNA-seq  scrRNA-seq
rstudio@e549475e8bdb:~/kitematic$ cd RNA-seq
rstudio@e549475e8bdb:~/kitematic/RNA-seq$ cd data
rstudio@e549475e8bdb:~/kitematic/RNA-seq/data$ mkdir command_line_example
rstudio@e549475e8bdb:~/kitematic/RNA-seq/data$ cd command_line_example/
rstudio@e549475e8bdb:~/kitematic/RNA-seq/data/command_line_example$ wget -O bash_example.sh https://raw.githubusercontent.com/GameServerManagers/LinuxGSM/master/linuxgsm.sh
--2019-07-24 20:46:35-- https://raw.githubusercontent.com/GameServerManagers/LinuxGSM/master/linuxgsm.sh
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.200.133
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|151.101.200.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 11826 (12K) [text/plain]
Saving to: 'bash_example.sh'
```

← Command to list the files or subdirectories in the current directory
← Command to change into the subdirectory "RNA-seq"
← Make a new directory named "command_line_example" in the current directory
← Now 'cd' into that new directory
← 'wget' downloads a file from the internet (in this case a bash script) and saves it in the current directory
← The '-O' flag in our previous command allows us to name the file we retrieve

```
bash_example.sh      100%[=====>] 11.55K  --.-KB/s   in 0.007s

2019-07-24 20:46:35 (1.62 MB/s) - 'bash_example.sh' saved [11826/11826]

rstudio@e549475e8bdb:~/kitematic/RNA-seq/data/command_line_example$ bash bash_example.sh
fetching serverlist.csv...OK
Usage: bash_example.sh [option]

Installer - Linux Game Server Managers - Version v19.9.0
https://linuxgsm.com

Commands
install      | Select server to install.
servername   | Enter name of game server to install. e.g bash_example.sh csgoserver.
list         | List all servers available for install.
rstudio@e549475e8bdb:~/kitematic/RNA-seq/data/command_line_example$ echo This is the end of the command line demonstration
This is the end of the command line demonstration
rstudio@e549475e8bdb:~/kitematic/RNA-seq/data/command_line_example$ >
```

← bash command runs a bash script
← echo command prints the string that follows

Feel free to give these commands a try on your own! (Note that our example begins in the training-module directory.)