

Bulk RNA-seq Cheatsheet

The tables below consist of valuable functions or commands 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. Please be aware that the documentation will generally provide information about the given function's most current version (or a recent version, depending on how often the documentation site is updated). This will usually (but not always!) match what you have installed on your machine. If you have a different version of R or other R packages, the documentation may differ from what you have installed.

Table of Contents

- [Base R](#)
- [DESeq2](#)
- [FastQC](#) and [fastp](#)
- [ggplot2](#)
- [tximeta](#) and [SummarizedExperiment](#)
- [stringr](#), [readr](#), [dplyr](#), [pheatmap](#)
- [Salmon](#)

Base R

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

| 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](#), and the [package vignette by Love, Anders, and Huber](#).

| Library/Package | Piece of Code | What it's called | What it does |
|-----------------|---------------------------------------|--|--|
| DESeq2 | <code>vst()</code> | Variance Stabilizing Transformation | Applies variance stabilizing transformation to data (log2-like scale) |
| DESeq2 | <code>DESeqDataSet()</code> | DESeqDataSet constructor that can take a <code>SummarizedExperiment</code> | Creates a DESeqDataSet object |
| DESeq2 | <code>DESeqDataSetFromMatrix()</code> | DESeqDataSet constructor | Creates a DESeqDataSet object from a matrix of count data |
| DESeq2 | <code>DESeq()</code> | 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 DESeq2 differential expression analysis |
| DESeq2 | <code>plotPCA()</code> | PCA plot | Produces a principal component analysis plot for transformed data. It can be used to visually inspect the data, which might allow an analyst to identify batch effects. |
| DESeq2 | <code>counts()</code> | Counts | Returns count matrix from <code>DESeqDataSet</code> object |
| DESeq2 | <code>results()</code> | Results | Returns the results table from a DESeq2 analysis |
| DESeq2 | <code>assay()</code> | Assay | Returns matrix from the <code>assay</code> slot of a <code>DESeqDataSet</code> object |

FastQC and fastp

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

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

ggplot2

Read the [ggplot2](#) package documentation, an [overall reference for ggplot2 functions](#), and a [vignette on the usage of the ggplot2 aesthetics](#). Additional vignettes are available from the "Articles" dropdown menu on this webpage.

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

tximeta and SummarizedExperiment

Read the [tximeta package documentation](#), and the [package vignette by Love *et al.*](#). Read the [SummarizedExperiment package documentation](#), and the [package vignette by Morgan *et al.*](#).

| Library/Package | Piece of Code | What it's called | What it does |
|----------------------|---|---------------------------|--|
| tximeta | <code>tximeta()</code> | tximeta | Imports transcript-level estimates, attaches transcriptome annotation, and returns a <code>SummarizedExperiment</code> object |
| tximeta | <code>makeLinkedTxome()</code> | Make Linked Transcriptome | Sets up transcriptome annotation to be used by the <code>tximeta()</code> function (Only necessary if <code>tximeta()</code> fails to find annotation, like for non-human, non-mouse species data) |
| tximeta | <code>summarizeToGene()</code> | Summarize to Gene | Takes a <code>SummarizedExperiment</code> that was set up by <code>tximeta</code> and summarizes transcript data to the gene-level |
| SummarizedExperiment | <code>rowData()</code> <code>colData()</code> | Col/Row Data | Accesses the row or column data from a <code>SummarizedExperiment</code> object |
| SummarizedExperiment | <code>assay()</code> <code>assayNames()</code> | Assay or AssayNames | Accesses the assay data or the names of the assays from <code>SummarizedExperiment</code> object |

stringr, readr, 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 | <code>word()</code> | Word | Extracts words from a character vector |
| readr | <code>write_rds()</code> | Write RDS | Writes data to a .RDS output file |
| dplyr | <code>pull()</code> | Pull | Extracts a variable (column) as a vector |
| pheatmap | <code>pheatmap()</code> | Pretty heatmap | Plots clustered heatmaps |

Salmon

Read the [Salmon tool documentation](#).

| 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) |

| Tool | Piece of Code | What it's called | What it does |
|--------|---|--------------------------|---|
| Salmon | <code>salmon</code> <code>quant</code> | Salmon quantification | Runs Salmon's quantification of transcript expression (from the command line) |

