

# Advanced scRNA-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.

You may also be interested in the following additional cheatsheets:

- Download the PDF for the [Introduction to R and Tidyverse cheatsheet](#)
- Download the PDF for the [Introduction to Single-Cell RNA sequencing cheatsheet](#)
- Download the PDF for the [SCE](#) and [Seurat](#) [object comparison cheatsheet](#)

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

- [scater](#)
- [miQC](#)
- [SingleR](#)
- [pheatmap](#) and [EnhancedVolcano](#)
- [DESeq2](#) and pseudo-bulking functions
- [tidyverse](#) functions
  - [purrr](#) functions
  - [ggplot2](#) functions
  - [dplyr](#), [tidyr](#), [stringr](#), and [tibble](#) functions

## scater

Read the [scater package documentation](#), and a [vignette on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
scater	<code>plotReducedDim()</code>	Plot reduced dimensions	Plot a given reduced dimension slot from a <code>SingleCellExperiment</code> object by its name
scater	<code>plotUMAP()</code>	Plot UMAP	Plot the "UMAP"-named reduced dimension slot from a <code>SingleCellExperiment</code> object
scater	<code>plotExpression()</code>	Plot expression	Plot expression values for all cells in a <code>SingleCellExperiment</code> object, using the <code>logcounts</code> assay by default

## miQC

Read the [miQC package documentation](#), and a [vignette on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
miQC	<code>mixtureModel()</code>	Mixture model	Fit a <code>miQC</code> mixture model to a <code>SingleCellExperiment</code> object for use in filtering
miQC	<code>filterCells()</code>	Filter cells	Filter cells from a <code>SingleCellExperiment</code> object based on a <code>miQC</code> model, returning a filtered <code>SingleCellExperiment</code> object
miQC	<code>plotMetrics()</code>	Plot metrics	Plot percent of mitochondrial reads against the number of unique genes found for each cell
miQC	<code>plotModel()</code>	Plot model	<code>miQC::plotMetrics()</code> with the <code>miQC</code> fitted model overlaid
miQC	<code>plotFiltering()</code>	Plot filtering	Plot percent of mitochondrial reads against the number of unique genes found, coloring points based on whether they will be filtered out or not

## batchelor and harmony

Read the [batchelor](#) package documentation, and a [vignette on its usage](#).

Read the [harmony](#) package documentation, and a [vignette on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
batchelor	<code>MultiBatchPCA()</code>	Multi-batch PCA	Perform PCA across multiple gene expression matrices, weighted by batch size
batchelor	<code>fastMNN()</code>	Fast mutual nearest neighbors correction	Perform integration on an SCE object with mutual nearest neighbors using the <code>fastMNN</code> algorithm, returning an SCE object with batch-corrected principal components
harmony	<code>HarmonyMatrix()</code>	Perform <code>harmony</code> integration on a matrix	Perform integration with <code>harmony</code> on either a matrix of principle components or gene expression, returning a matrix of batch-corrected principal components

## SingleR

Read the [SingleR](#) package documentation, and an [e-book on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
SingleR	<code>trainSingleR()</code>	Train the SingleR classifier	Build a <code>SingleR</code> classifier model object from an annotated reference dataset
SingleR	<code>classifySingleR()</code>	Classify cells with SingleR	Use a <code>SingleR</code> model object to assign cell types to the cells in an <code>SCE</code> object
SingleR	<code>SingleR()</code>	Annotate scRNA-seq data	Combines <code>trainSingleR()</code> and <code>classifySingleR()</code> to assign cell types to an <code>SCE</code> object from an annotated reference dataset

## pheatmap and EnhancedVolcano

Read the [pheatmap](#) package documentation.

Read the [EnhancedVolcano](#) package documentation, and [vignette on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
heatmap	<code>heatmap()</code>	Pretty heatmap	Plot a (pretty!) clustered heatmap
EnhancedVolcano	<code>EnhancedVolcano()</code>	Enhanced volcano	Plot a volcano plot to visualize differential expression analysis results

## DESeq2 and pseudo-bulking functions

Read the [DESeq2 package documentation](#), and a [vignette on its usage](#).

Library/Package	Piece of Code	What it's called	What it does
scuttle	<code>aggregateAcrossCells()</code>	Aggregate data across groups of cells	Sum counts for each combination of features across groups of cells, commonly used to <i>pseudo-bulk</i> SCE counts
DESeq2	<code>DESeqDataSet()</code>	DESeq Dataset	Establish a DESeq object from a pseudo-bulked SingleCellExperiment object or a bulk SummarizedExperiment object
DESeq2	<code>estimateSizeFactors()</code>	Estimate size factors	Estimate size factors which are used to normalize counts for differential expression analysis
DESeq2	<code>rlog()</code>	Apply a regularized log transformation	Log2-transform counts in a DESeq object for differential expression analysis
DESeq2	<code>plotPCA()</code>	Sample PCA plot for transformed data	Plot sample PCA from a log-transformed DESeq object to check for batch effects
DESeq2	<code>DESeq()</code>	Perform differential expression analysis	Perform differential expression: Estimate size factors, transform data, estimate dispersions, and perform testing.
DESeq2	<code>plotDispEsts()</code>	Plot dispersion estimates	Plot dispersion estimates from a fitted DESeq object to evaluate model fit
DESeq2	<code>results()</code>	Extract results from a DESeq analysis	Extract results from a fitted DESeq object into a data frame
DESeq2	<code>resultsNames()</code>	Extract results names	Return coefficient names from a fitted DESeq object

Library/Package	Piece of Code	What it's called	What it does
DESeq2	<code>lfcShrink()</code>	Shrink log2 fold changes	Add shrunken log2-fold changes to a results table produced by <code>DESeq2::results()</code>

## tidyverse functions

### purrr functions

Read the [purrr package documentation](#) and a [vignette on its usage](#), and download the [purrr package cheatsheet](#).

Library/Package	Piece of Code	What it's called	What it does
purrr	<code>map()</code>	map	Apply a function across each element of list; return a list
purrr	<code>imap()</code>	imap	Apply a function across each element of list and its index/names; return a list
purrr	<code>map2()</code>	map2	Apply a function across each element of two lists at a time; return a list
purrr	<code>reduce()</code>	Reduce	Reduce a list to a single value by applying a given function

Note that `purrr::map()` functions can take advantage of R's new (as of version 4.1.0) [anonymous function syntax](#):

```
# One-line syntax:
\(x) # function code goes here #

# Multi-line syntax:
\(x) {
  # function code goes      #
  # inside the curly braces #
}

# Example: Use an anonymous function with `purrr::map()`
# to get the colData's rownames for each SCE in `list_of_sce_objects`
purrr::map(
  list_of_sce_objects,
  \(x) rownames(colData(x))
)
```

### ggplot2 functions

Read the [ggplot2 package documentation](#) and an [overall reference for ggplot2 functions](#), and download the [ggplot2 package cheatsheet](#).

Library/Package	Piece of Code	What it's called	What it does
ggplot2	<code>geom_bar()</code>	Barplot	Creates a barplot of counts for a given categorical variable when added as a layer to a <code>ggplot()</code> object

Library/Package	Piece of Code	What it's called	What it does
ggplot2	<code>scale_fill_brewer()</code>	Add brewer fill scale	Apply a Brewer "fill" color palette to a categorical variable in a <code>ggplot()</code> object
ggplot2	<code>guides()</code>	Guides	Function to customize legend ("guide") appearance
ggplot2	<code>facet_grid()</code>	Facet grid	Plot individual panels using specified variables to subset the data across rows and/or columns of a grid
ggplot2	<code>vars()</code>	Vars	Helper function to specify variables to <code>facet_grid()</code> or <code>facet_wrap()</code>
ggplot2	<code>theme_bw()</code>	Black and white theme	Display <code>ggplot</code> with gridlines but a white background
ggplot2	<code>theme()</code>	Theme	Customize elements of a <code>ggplot</code> plot theme
ggplot2	<code>element_text()</code>	Element text	Customize textual elements of a <code>ggplot</code> theme

## dplyr, tidyr, stringr, and tibble functions

Read the full documentation and download cheatsheets (where available) for these `tidyverse` packages at the following links:

- [dplyr documentation](#) and [dplyr cheatsheet](#)
- [tidyr documentation](#) and [tidyr cheatsheet](#)
- [stringr documentation](#) and [stringr cheatsheet](#)
- [tibble documentation](#)

Library/Package	Piece of Code	What it's called	What it does
dplyr	<code>pull()</code>	Pull	Extract a single column from a data frame into a stand-alone vector
dplyr	<code>count()</code>	Count	Count the number of observations in each group of a data frame
dplyr	<code>left_join()</code>	Left join	Joins two data frames together, retaining only rows present in the first ("left") argument to the function
dplyr	<code>relocate()</code>	Relocate	Change column order in a data frame by relocating one or more columns
dplyr	<code>case_when()</code>	Case when	Return a value based on a set of <code>TRUE / FALSE</code> comparisons; a vectorized <code>if-else</code>
tidyr	<code>pivot_longer()</code>	Pivot longer	Convert a "wide" format data frame to a "long" format data frame
tibble	<code>as_tibble()</code>	As tibble	Convert an object to a tibble

Library/Package	Piece of Code	What it's called	What it does
<code>stringr</code>	<code>str_detect()</code>	String detect	Returns <code>TRUE</code> / <code>FALSE</code> if a string contains a given substring
<code>stringr</code>	<code>str_starts()</code>	String starts	Returns <code>TRUE</code> / <code>FALSE</code> if a string starts with a given substring