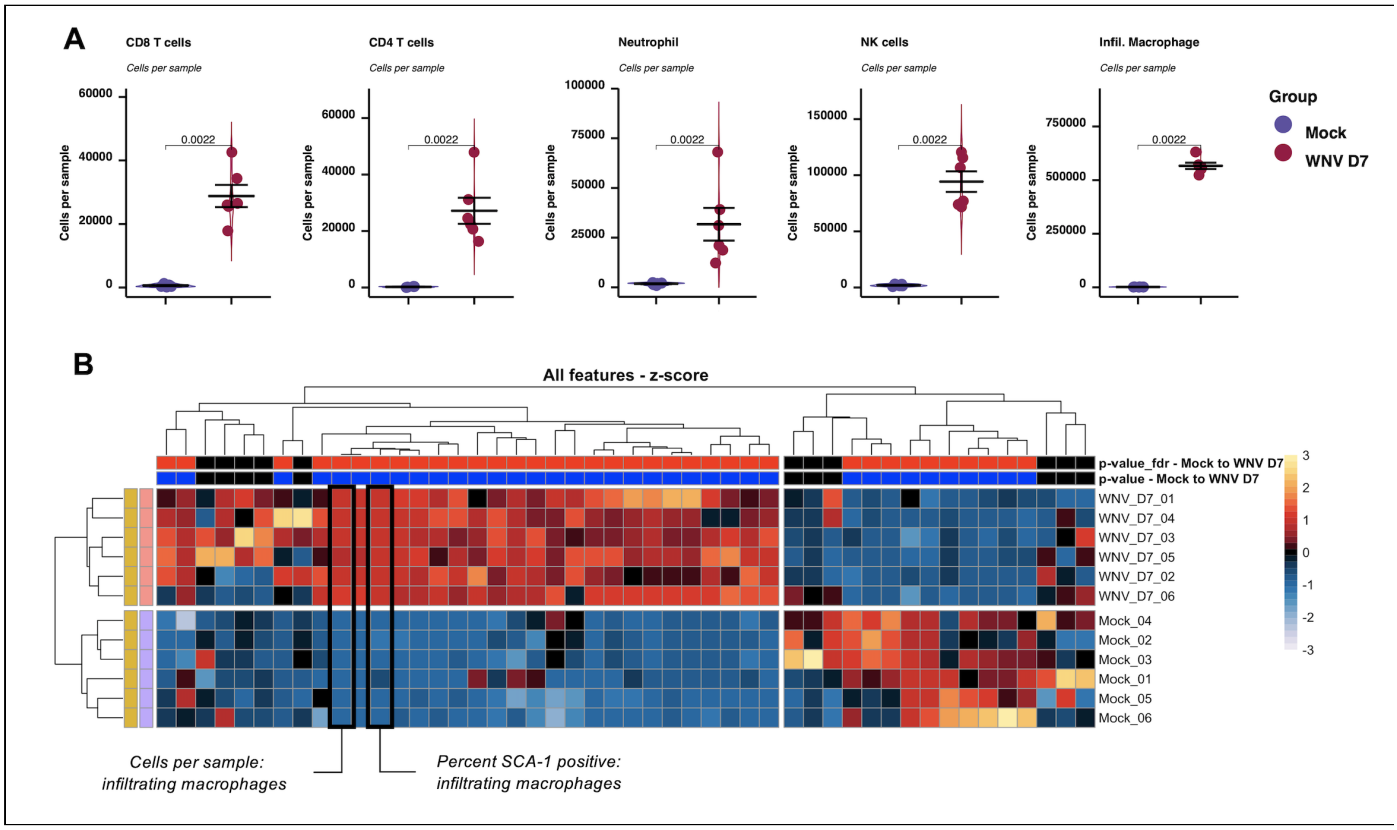
Introduction

Once you have generated summary data (e.g. number of T cells per sample, etc) in Spectre, or other programs such as FlowJo, it can be laborious to manually enter these values into a program to generate graphs for statistical analysis. Here we present a quick workflow script to rapidly generate graphs and heatmaps for quantitative, differential, and statistical analysis. Note, this can be used on summary data generated by any program, including FlowJo.



Citation

If you use Spectre in your work, please consider citing Ashhurst TM, Marsh-Wakefield F, Putri GH et al. (2020). bioRxiv. 2020.10.22.349563. To continue providing open-source tools such as Spectre, it helps us if we can demonstrate that our efforts are contributing to analysis efforts in the community. Please also consider citing the authors of the individual packages or tools (e.g. CytoNorm, FlowSOM, tSNE, UMAP, etc) that are critical elements of your analysis work. We have provided some generic text that you can use for your methods section with each protocol and on the 'about' page.

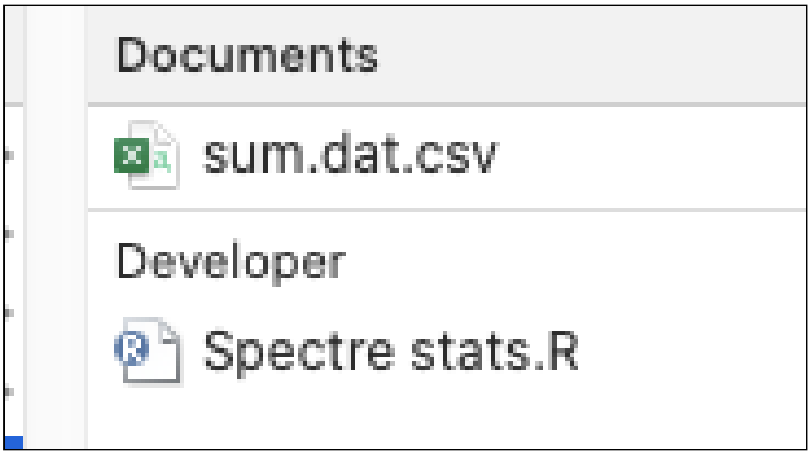
Software and R script preparation

Note

- If you haven't installed Spectre, please visit our Spectre installation page.
- If you aren't familiar with using RStudio or Spectre, please see our RStudio and Spectre basic tutorials.

You can find the 'Spectre stats' script (and some demo data) here:

Copy the script into a folder containing a CSV with your summary data.



Your summary data (in this example, sum.dat.csv) should be saved as a CSV File. You will need at minimum one column that denotes the sample name, and one that denotes the group each sample belongs to. You can have as many additional 'annotation' columns as you like (e.g. batch, timepoint, treatment, etc). Each other column should represent some feature of the samples (CD4 T cells per sample, expression of Ly6C on monocytes in each sample, etc). Tables like this are generated using the 'Table Editor' from FlowJo.

Sample	Group	Batch	Cells per sample -- CD4 T cells	Cells per sample -- CD8 T cells	Cells per sample -- Infil Macrophages	Cells per sample -- Microglia	Cells per sample -- Neutrophils
01_Mock_01	Mock	A	734.628157391567	2696.99104357452	2143.50407567676	88406.9638723961	2858.00543423568
02_Mock_02	Mock	B	506.00064871878	1037.95004865391	2536.49043139799	88861.4985403827	2821.92669477781
03_Mock_03	Mock	B	540.502597220272	1003.79053769479	2470.8690158641	88712.6210866208	3186.85946932472
04_Mock_04	Mock	A	252.288195259329	404.834545881249	1965.50105609012	92419.6198075569	2305.79676132363
05_Mock_05	Mock	A	201.502106612933	549.551199853453	2198.20479941381	90364.5356292361	3352.26231910606
06_Mock_06	Mock	B	531.588632181558	1308.52586383153	4722.96053976692	78736.4547127377	7810.26374974443
07_WNV_01	WNV	A	54094.7075208914	50529.2479108635	605738.161559889	131197.771587744	23899.721448468
08_WNV_02	WNV	B	20636.9740651759	23901.9281710097	674243.824308508	135218.382307645	49775.1493870511
09_WNV_03	WNV	A	26314.1446340396	34814.6759172448	678792.424526533	128070.504406525	39439.9649978124
10_WNV_04	WNV + Rx	A	26910.2796100636	32397.4082073434	652559.687116923	119899.597221412	81197.8284980445
11_WNV_05	WNV + Rx	B	31520.7836242622	36292.8544518398	644041.190506091	124513.374356398	17330.1519527816
12_WNV_06	WNV + Rx	A	34251.3183827838	37500.6658498908	645714.590102807	144516.060299366	21040.8565492995

1. Read in summary data

```
#########################################################################################################
###
#### 1. Load packages, and set working directory
#########################################################################################################
###
```

Load the packages.

```
    ### Load libraries

        library(Spectre)
        Spectre::package.check()    # Check that all required packages are installed
        Spectre::package.load()     # Load required packages
```

Save the folder where this script is located as 'PrimaryDirectory'.

```
    ### Set PrimaryDirectory

        dirname(rstudioapi::getActiveDocumentContext()$path)            # Finds the directory where
this script is located
        setwd(dirname(rstudioapi::getActiveDocumentContext()$path))     # Sets the working directory to
where the script is located
        getwd()
        PrimaryDirectory <- getwd()
```

Create an output directory for the plots. This will be a subfolder of PrimaryDirectory.

```
    ### Create output directory

        dir.create("Output_Spectre_stats", showWarnings = FALSE)
        setwd("Output_Spectre_stats")
        OutputDirectory <- getwd()
        setwd(PrimaryDirectory)
```

2. Import summary data

```
#########################################################################################################
###
#### 2. Import summary data
#########################################################################################################
###
```

Import the summary data into R. Note, the summary data needs to be saved as a '.csv' file (if using excel, try 'Save As' and select CSV).

```
    ### Import data

        setwd(PrimaryDirectory)
        sum.dat <- fread("sum.dat.csv")
```

Preview the data.

```
    ### Columns for analysis

        sum.dat
```

	Sample	Group	Batch	Cells per sample -- CD4 T cells	Cells per sample -- CD8 T cells	Cells per sample -- Infil Macrophages
1:	01_Mock_01	Mock	A	734.6282	2696.9910	2143.504
2:	02_Mock_02	Mock	B	506.0006	1037.9500	2536.490
3:	03_Mock_03	Mock	B	540.5026	1003.7905	2470.869
4:	04_Mock_04	Mock	A	252.2882	404.8345	1965.501
5:	05_Mock_05	Mock	A	201.5021	549.5512	2198.205
6:	06_Mock_06	Mock	B	531.5886	1308.5259	4722.961
7:	07_WNV_01	WNV	A	54094.7075	50529.2479	605738.162
8:	08_WNV_02	WNV	B	20636.9741	23901.9282	674243.824
9:	09_WNV_03	WNV	A	26314.1446	34814.6759	678792.425
10:	10_WNV_04	WNV + Rx	A	26910.2796	32397.4082	652559.687
11:	11_WNV_05	WNV + Rx	B	31520.7836	36292.8545	644041.191
12:	12_WNV_06	WNV + Rx	A	34251.3184	37500.6658	645714.590

Examine the column names.

```
as.matrix(names(sum.dat))
```

```
      [,1]
 [1,] "Sample"
 [2,] "Group"
 [3,] "Batch"
 [4,] "Cells per sample -- CD4 T cells"
 [5,] "Cells per sample -- CD8 T cells"
 [6,] "Cells per sample -- Infil Macrophages"
 [7,] "Cells per sample -- Microglia"
 [8,] "Cells per sample -- Neutrophils"
 [9,] "Cells per sample -- NK cells"
[10,] "Percent Ly6C_asinh positive -- CD4 T cells"
[11,] "Percent Ly6C_asinh positive -- CD8 T cells"
[12,] "Percent Ly6C_asinh positive -- Infil Macrophages"
[13,] "Percent Ly6C_asinh positive -- Microglia"
[14,] "Percent Ly6C_asinh positive -- Neutrophils"
[15,] "Percent Ly6C_asinh positive -- NK cells"
```

Define the columns that denote: sample names, group names, and then any other annotation columns. In this example:

- Column 1 ('Sample') is the sample column
- Column 2 ('Group') is the group column
- Columns 2 ('Group') and 3 ('Batch') are both going to be annotation columns

```
sample.col <- names(sum.dat)[c(1)]
group.col <- names(sum.dat)[c(2)]
annot.cols <- names(sum.dat)[c(2:3)]
```

We can also specify which columns are sample features we wish to plot (in the demo, columns 4 to 15).

```
plot.cols <- names(sum.dat)[c(4:15)]
```

```
as.matrix(plot.cols)
```

```
      [,1]
 [1,] "Cells per sample -- CD4 T cells"
 [2,] "Cells per sample -- CD8 T cells"
 [3,] "Cells per sample -- Infil Macrophages"
 [4,] "Cells per sample -- Microglia"
 [5,] "Cells per sample -- Neutrophils"
```

```
 [6,] "Cells per sample -- NK cells"
 [7,] "Percent Ly6C_asinh positive -- CD4 T cells"
 [8,] "Percent Ly6C_asinh positive -- CD8 T cells"
 [9,] "Percent Ly6C_asinh positive -- Infil Macrophages"
[10,] "Percent Ly6C_asinh positive -- Microglia"
[11,] "Percent Ly6C_asinh positive -- Neutrophils"
[12,] "Percent Ly6C_asinh positive -- NK cells"
```

Here we also need to specify the order in which we want the groups

```
    ### Experimental groups                as.matrix(unique(sum.dat[[group.col]]))
```

```
     [,1]
[1,] "Mock"
[2,] "WNV"
[3,] "WNV + Rx"
```

If we want to change the order, we can do it here (in the demo, the groups are already in the correct order – if they weren't, we could change it: **c("WNV", "Mock", "WNV + Rx")** etc).

```
        grp.order <- c("Mock", "WNV", "WNV + Rx")
        as.matrix(grp.order)
```

```
     [,1]
[1,] "Mock"
[2,] "WNV"
[3,] "WNV + Rx"
```

We also need to define which paired comparisons we want to make for statistics. In the demo we have three: Mock vs WNV, WNV vs WNV + Rx, and Mock vs WNV + Rx.

```
        comparisons <- list(c("Mock", "WNV"),
                            c("WNV", "WNV + Rx"),
                            c("Mock", "WNV + Rx")
                            )
```

Check the entries

```
        comparisons
```

```
[[1]]
[1] "Mock" "WNV"

[[2]]
[1] "WNV"      "WNV + Rx"

[[3]]
[1] "Mock"     "WNV + Rx"
```

Define the statistical tests to use.

```
    ### Setup

        variance.test <- 'kruskal.test'
        pairwise.test <- "wilcox.test"
```

Reorder the rows

```
    ### Reorder summary data and SAVE

        sum.dat <- do.reorder(sum.dat, group.col, grp.order)

        sum.dat
        sum.dat[,c(1:5)]
```

3. Output analysis

```
#########################################################################################################
###
#### 3. Output analysis
#########################################################################################################
###
```

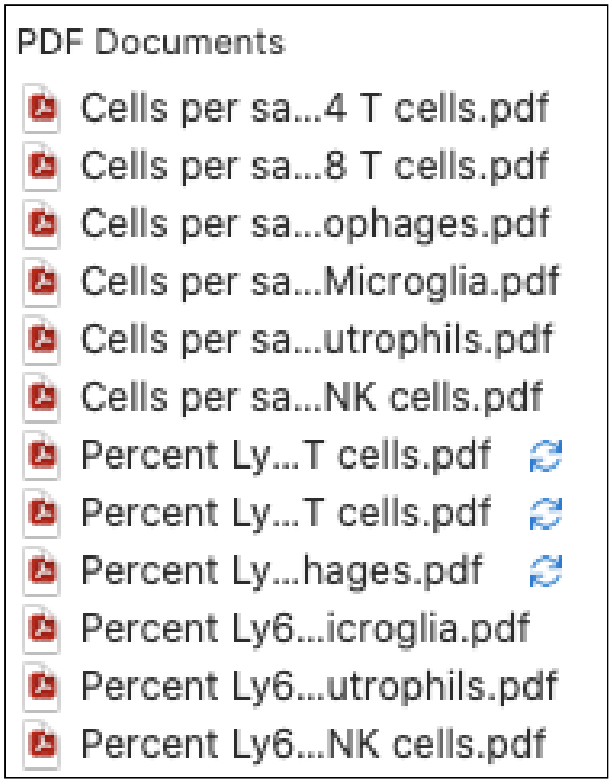
Save the data.

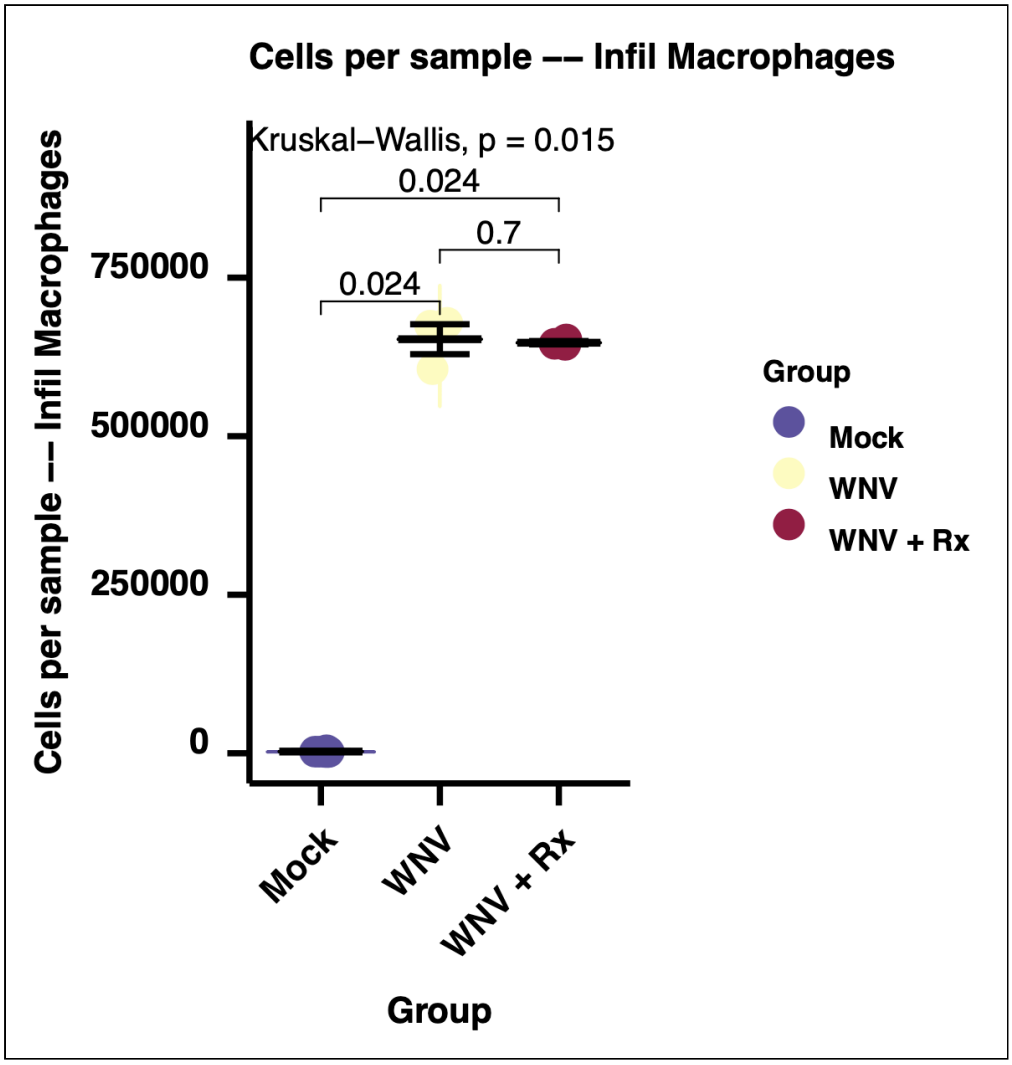
```
    setwd(OutputDirectory)
    fwrite(sum.dat, 'sum.dat.csv')
```

```
#########################################################################################################
###
#### 3. Output analysis
#########################################################################################################
###
```

Loop to create one plot per feature.

```
    ### Autographs

        for(i in plot.cols){

            make.autograph(sum.dat,
                           x.axis = group.col,
                           y.axis = i,
                           y.axis.label = i,

                           grp.order = grp.order,
                           my_comparisons = comparisons,

                           Variance_test = variance.test,
                           Pairwise_test = pairwise.test,

                           title = i,
                           filename = paste0(i, '.pdf'))

        }
```


PDF Documents
Cells per sa...4 T cells.pdf
Cells per sa...8 T cells.pdf
Cells per sa...ophages.pdf
Cells per sa...Microglia.pdf
Cells per sa...utrophils.pdf
Cells per sa...NK cells.pdf
Percent Ly...T cells.pdf
Percent Ly...T cells.pdf
Percent Ly...hages.pdf
Percent Ly6...icroglia.pdf
Percent Ly6...utrophils.pdf
Percent Ly6...NK cells.pdf



Z-score transformation of data.

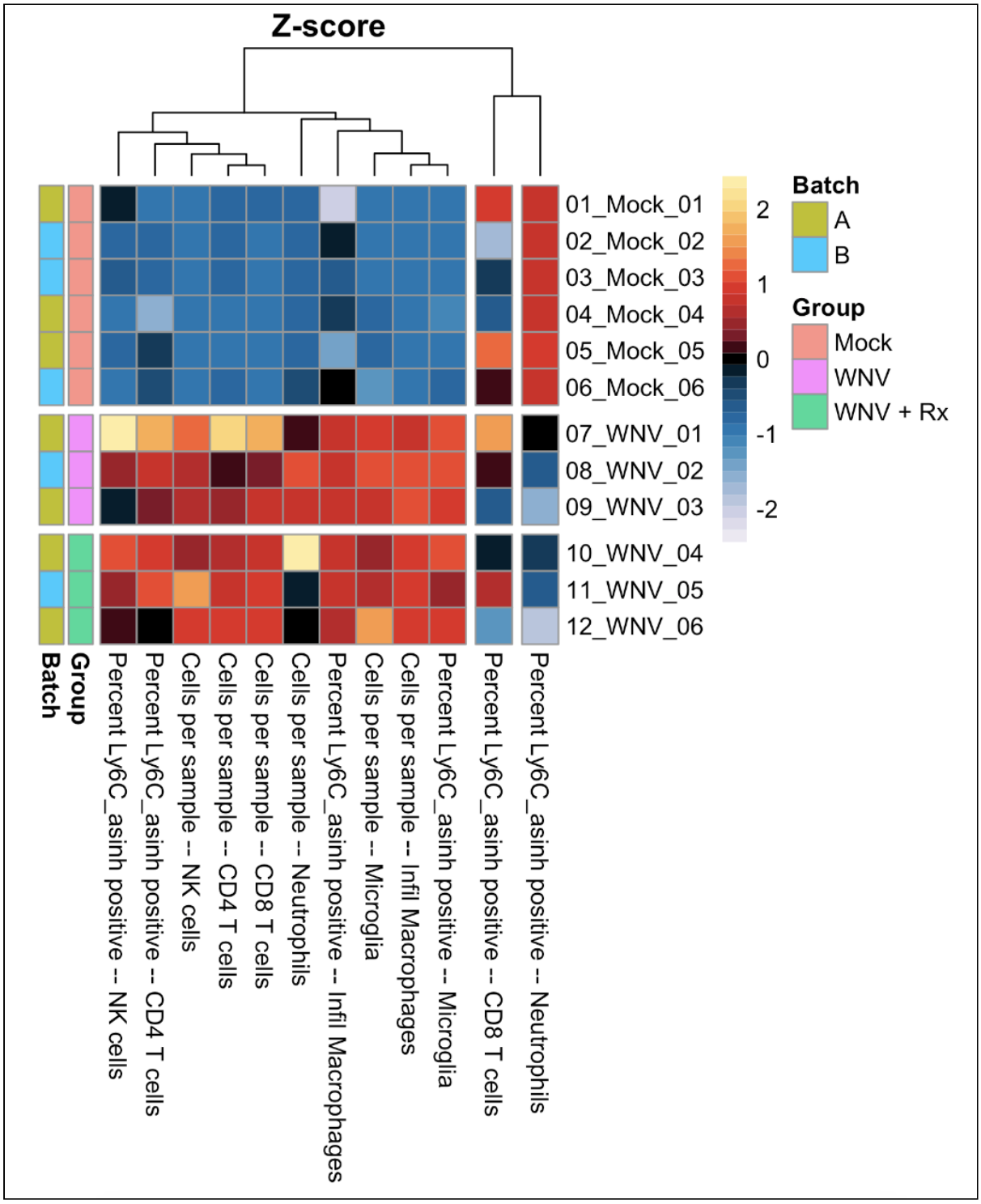
```
### Create a fold change heatmap

    ## Z-score calculation

    sum.dat.z <- do.zscore(sum.dat, plot.cols, replace = TRUE)

    ## Group

    t.first <- match(grp.order, sum.dat.z[[group.col]])
    t.first <- t.first -1
    t.first
```

Make a z-score heatmap.

```
    ## Make heatmap
    make.pheatmap(sum.dat.z,
                  sample.col = sample.col,
                  plot.cols = plot.cols,
                  is.fold = TRUE,
                  plot.title = 'Z-score',
                  annot.cols = annot.cols,
                  dendrograms = 'column',
```

```
                    row.sep = t.first,
                    cutree_cols = 3)
```

Pheatmap...ample.png



4. Output session info

```
#########################################################################################################
###
#### Output session info
#########################################################################################################
###
```

```
        setwd(OutputDirectory)
        dir.create("Output - info", showWarnings = FALSE)
        setwd("Output - info")

        sink(file = "session_info.txt", append=TRUE, split=FALSE, type = c("output", "message"))
        session_info()
        sink()
```