Introduction

Spectre provides a number of options for performing quantitative, differential and statistical analysis of cytometry data after the initial analysis using clustering or similar methods. Here we provide a demonstration of a number these options.

For this tutorial, we will use one of the demo datasets included in Spectre: a dataset of cells isolated from murine brains, 7 days following mock infection, or infection with West Nile virus (WNV). The. **demo.clustered** dataset has already been subject to arcsinh transformation, clustering, and population annotation..

```
library('Spectre')
package.load()

cell.dat <- Spectre::demo.clustered
cell.dat
```

```
            FileName      NK11       CD3      CD45      Ly6G     CD11b      B220      CD8a
Ly6C      CD4  NK11_asinh   CD3_asinh  CD45_asinh  Ly6G_asinh ...
     1:  CNS_Mock_01.csv   42.3719   40.098700   6885.08  -344.7830 14787.30  -40.2399   83.7175
958.7000  711.0720  0.04235923  0.040087962   2.627736 -0.33829345 ...
     2:  CNS_Mock_01.csv   42.9586  119.014000   1780.29  -429.6650  5665.73   86.6673   34.7219
448.2590  307.2720  0.04294540  0.118734817   1.340828 -0.41743573 ...
     3:  CNS_Mock_01.csv   59.2366  206.238000  10248.30 -1603.8400 19894.30  427.8310  285.8800
1008.8300  707.0940  0.05920201  0.204803270   3.022631 -1.25101677 ...
     4:  CNS_Mock_01.csv  364.9480   -0.233878   3740.04  -815.9800  9509.43  182.4200  333.6050
440.0710  249.7840  0.35729716 -0.000233878   2.029655 -0.74509796 ...
     5:  CNS_Mock_01.csv  440.2470   40.035200   9191.38    40.5055  5745.82 -211.6940  149.2200
87.4815  867.5700  0.42713953  0.040024513   2.914359  0.04049443 ...
    ---
169000: CNS_WNV_D7_06.csv  910.8890   72.856100  31466.20  -316.5570 28467.80   -7.7972 -271.8040
12023.7000 1103.0500  0.81693878  0.072791800   4.142314 -0.31149515 ...
169001: CNS_WNV_D7_06.csv  -10.2642   64.188700  45188.00  -540.5140 22734.00  202.4110 -936.4920
4188.3300  315.9400 -0.01026402  0.064144703   4.504101 -0.51715205 ...
169002: CNS_WNV_D7_06.csv -184.2910   -9.445650  11842.60   -97.9383 17237.00  123.4760 -219.9320
8923.4000 -453.4640 -0.18326344 -0.009445510   3.166628 -0.09778240 ...
169003: CNS_WNV_D7_06.csv  248.3860  229.986000  32288.20  -681.1630 19255.80 -656.0540 -201.5880
10365.7000   61.6765  0.24590035  0.228005328   4.168089 -0.63716643 ...
169004: CNS_WNV_D7_06.csv  738.9810   95.470300  46185.10 -1004.6000 22957.80 -661.6280   72.3356
9704.4700  -31.8532  0.68430866  0.095325863   4.525922 -0.88462254 ...
```

We will also provide some 'cell count' data for each sample (i.e. number of total leukocytes in each sample). We expsect

```
counts.dt <- data.frame('Sample' = unique(cell.dat[['Sample']]),
                        'Counts' = c(4.20E+05, 2.40E+05, 2.56E+05, 2.52E+05, 3.45E+05, 7.02E+05,
                                     5.07E+06, 2.94E+06, 2.12E+06, 4.32E+06, 4.08E+06, 1.83E+06)
                        )
counts.dt
```

```
    Sample Counts
1  01_Mock_01  4.20E+05
2  02_Mock_02  2.40E+05
3  03_Mock_03  2.56E+05
4  04_Mock_04  2.52E+05
5  05_Mock_05  3.45E+05
6  06_Mock_06  7.02E+05
7   07_WNV_01  5.07E+06
8   08_WNV_02  2.94E+06
```

```
 9   09_WNV_03  2.12E+06
10   10_WNV_04  4.32E+06
11   11_WNV_05  4.08E+06
12   12_WNV_06  1.83E+06
```

Key to the comparison of populations across samples, is the generation of 'summary' data. Where 'cellular' data consists of cells (rows) vs cell features (columns: e.g. CD4 expression, CD8 expression etc); 'summary' data consists of samples (rows) vs sample features (number of monocytes per sample, expression level of Ly6C on CD8 T cells, etc). This summary data can then be used to generate plots that compare these metrics between experimental groups.

First, let's examine the columns in the cellular dataset.

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

```
      [,1]
 [1,] "FileName"
 [2,] "NK11"
 [3,] "CD3"
 [4,] "CD45"
 [5,] "Ly6G"
 [6,] "CD11b"
 [7,] "B220"
 [8,] "CD8a"
 [9,] "Ly6C"
[10,] "CD4"
[11,] "NK11_asinh"
[12,] "CD3_asinh"
[13,] "CD45_asinh"
[14,] "Ly6G_asinh"
[15,] "CD11b_asinh"
[16,] "B220_asinh"
[17,] "CD8a_asinh"
[18,] "Ly6C_asinh"
[19,] "CD4_asinh"
[20,] "Sample"
[21,] "Group"
[22,] "Batch"
[23,] "FlowSOM_cluster"
[24,] "FlowSOM_metacluster"
[25,] "Population"
[26,] "UMAP_X"
[27,] "UMAP_Y"
```

We can choose any number of these to be measured as 'dynamic' colums (**dyn.cols**), where we will measure the median expression of these markers on each population in each sample. In this case we will choose CD11b (#15) and Ly6C (#18).

```
dyn.cols <- names(cell.dat)[c(15,18)]
dyn.cols
```

```
[1] "CD11b_asinh" "Ly6C_asinh"
```

To create the summary data, we can use the **create.sumtable** function.

```
sum.dat <- create.sumtable(dat = cell.dat, # The dataset to be summarised
                           sample.col = 'Sample', # The column that denotes the sample name/ID
                           pop.col = 'Population', # The column that denotes the population name/ID
                           use.cols = dyn.cols, # Columns (markers) whose expression we will measure on
  each population in each sample
```

```
                    annot.cols = c('Group', 'Batch'), # Additional columns we would like to
    include for annotation purposes (e.g. group names, batch names, etc)
                    counts = counts.dt) # A data.frame or data.table of the total cells per
    sample, to generate counts of each cell type per sample.
```

Once the function is complete, we can review the data.

```
sum.dat
```

Each row represents a sample, and each column a feature of that sample (e.g. Percent of sample -- CD4 T cells, etc).

```
         Sample Group Batch Percent of sample -- CD4 T cells Percent of sample -- CD8 T cells    ...
 1: 01_Mock_01  Mock     A                        0.7346282                        2.6969910    ...
 2: 02_Mock_02  Mock     B                        0.5060006                        1.0379500    ...
 3: 03_Mock_03  Mock     B                        0.5405026                        1.0037905    ...
 4: 04_Mock_04  Mock     A                        0.2522882                        0.4048345    ...
 5: 05_Mock_05  Mock     A                        0.2015021                        0.5495512    ...
 6: 06_Mock_06  Mock     B                        0.5315886                        1.3085259    ...
 7:  07_WNV_01   WNV     A                        5.4094708                        5.0529248    ...
 8:  08_WNV_02   WNV     B                        2.0636974                        2.3901928    ...
 9:  09_WNV_03   WNV     A                        2.6314145                        3.4814676    ...
10:  10_WNV_04   WNV     A                        2.6910280                        3.2397408    ...
11:  11_WNV_05   WNV     B                        3.1520784                        3.6292854    ...
12:  12_WNV_06   WNV     A                        3.4251318                        3.7500666    ...
```

Review all of the sample 'features' that we have calculated.

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

```
      [,1]
 [1,] "Sample"
 [2,] "Group"
 [3,] "Batch"
 [4,] "Percent of sample -- CD4 T cells"
 [5,] "Percent of sample -- CD8 T cells"
 [6,] "Percent of sample -- Infil Macrophages"
 [7,] "Percent of sample -- Microglia"
 [8,] "Percent of sample -- Neutrophils"
 [9,] "Percent of sample -- NK cells"
[10,] "Cells per sample -- CD4 T cells"
[11,] "Cells per sample -- CD8 T cells"
[12,] "Cells per sample -- Infil Macrophages"
[13,] "Cells per sample -- Microglia"
[14,] "Cells per sample -- Neutrophils"
[15,] "Cells per sample -- NK cells"
[16,] "MFI of CD11b_asinh -- CD4 T cells"
[17,] "MFI of CD11b_asinh -- CD8 T cells"
[18,] "MFI of CD11b_asinh -- Infil Macrophages"
[19,] "MFI of CD11b_asinh -- Microglia"
[20,] "MFI of CD11b_asinh -- Neutrophils"
[21,] "MFI of CD11b_asinh -- NK cells"
[22,] "MFI of Ly6C_asinh -- CD4 T cells"
[23,] "MFI of Ly6C_asinh -- CD8 T cells"
[24,] "MFI of Ly6C_asinh -- Infil Macrophages"
[25,] "MFI of Ly6C_asinh -- Microglia"
[26,] "MFI of Ly6C_asinh -- Neutrophils"
[27,] "MFI of Ly6C_asinh -- NK cells"
```