**High-dimensional analysis with Spectre  
Post-session activities**

1. **Loading packages and reading in data**
   1. Load the *Spectre* package, and run package.check() and package.load() to check all Spectre’s dependencies are installed and loaded
   2. Read the 9 example fcs files into an object called data\_list  
      **Hint:** make sure you set the do.embed.file.names argument equal to TRUE
   3. Use the do.list.summary() function do perform a sanity check. Do all the files have the same parameter names and number of parameters? What is the smallest and largest number of events in a file in the dataset?
   4. Use do.merge.files() to merge the data\_list into a single data table called cell\_dat
2. **Transforming channel values**
   1. Create a vector of channel names you want to transform  
      **Hint:** use the vector c(3:10, 14:17, 22:28, 30, 32:42) to subset the vector names(cell\_dat)
   2. Use the do.asinh() function to add asinh-transformed parameters to cell\_dat. Se the cofactor argument to 5
   3. Create a vector of names for the transformed columns to refer to them later  
      **Hint:** use the paste0() function
   4. Sample 10,000 events from cell\_dat and store this in an object called cell\_sub
   5. Plot the transformed parameters of cell\_sub and check they appear appropriate (pick a suitable parameter for the x axis, such as CD45)  
      **Hint:** there is an example of how to do this in the handout using lapply()
3. **Adding sample metadata**
   1. Read the sample\_details.csv example file and store it as an object
   2. Add the sample metadata to the cell\_dat dataset using do.add.cols()
   3. Check this has been done correctly. How many batches and how many groups are there?
4. **Cluster the events**
   1. Perform flowSOM clustering, using the transformed\_cols as the channels, to partition the data into 28 metaclusters  
      **Hint:** this will take a few minutes, make sure you store its output in cell\_dat!
   2. Use do.subsample(cell\_dat, min.per = TRUE, divide.by = “Sample”) to sample the same number of events from each Sample (based on the smallest Sample). Overwrite the cell\_sub object with this sampled dataset
   3. Run the Fourier-interpolated t-SNE (Fit-SNE) dimension reduction algorithm **on the cell\_sub dataset**, using the transformed\_cols as the channels  
      **Hint:** this will take a few minutes, make sure you store its output in cell\_sub
   4. Plot the FItSNE\_X and FItSNE\_Y variables against each other and colour by FlowSOM\_metacluster  
      **Hint:** set the col.type argument equal to “factor” to have block colouring
   5. Create the same plot as in (d) but create separate subplots for each Genotype of mouse. Can you see any differences in metacluster abundance between the two?
   6. Plot the FItSNE\_X and FItSNE\_Y variables against each other but this time create separate subplots for each channel in transformed\_cols, and colour the points by the channel’s values  
      **Hint:** you could use make.multi.plot() to do this (it may take a few minutes)
   7. Use the do.aggregate() function to calculate the median value of each of transformed\_cols by FlowSOM\_metacluster and store it in an object called medians
   8. Use make.pheatmap() to create a heatmap with FlowSOM\_metacluster on the rows, transformed\_cols on the columns, and the median expression as the colour of the tilesr  
      **NOTE:** at this point you could annotate and merge the clusters into biological cell types, though this will take a considerable amount of work so we will skip this step (you can see how this is done in the handout)
5. **Write out files**Use the write.files() function to save an individual .fcs *and* .csv file for each Sample in cell\_dat
6. **Extracting summary data**
   1. Use the following code to create a vector of simple names and overwrite the long names in cell\_dat  
        
      simple\_names <- c("Tbet", "Tim3", "LAG3\_APC", "IkBa", "CD19", "NKp46", "Ki67",  
       "PDL2", "IRF4", "SigF", "FoxP3", "KLRG1", "CD11c", "GITR",  
       "MHCI", "CD69", "CD8", "CD11b", "CD25", "Gr-1", "p4E-BP1",   
       "CD3", "CTLA4", "PD-1", "Sca1", "CD45", "CD44", "CD4", CD117",   
       "MHCII", "ICOS")  
        
      names(cell\_dat)[47:77] <- simple\_names
   2. Use the following code to create a table of summary statistics:  
        
      sum\_dat <- create.sumtable(dat = cell\_dat,  
       sample.col = "Sample",  
       pop.col = "FlowSOM\_metacluster",  
       use.cols = simple\_names,  
       annot.cols = c("Genotype", "Batch"))  
        
      Do you understand what each argument is doing? Look at the output of sum\_dat[1:5, 1:6] to help you understand.

**If you have any questions or want feedback, please contact** [**hefin.rhys@ucb.com**](mailto:hefin.rhys@ucb.com)