Cytokine Activity Estimation using SCAPE (Single cell transcriptomics-level Cytokine Activity Prediction and Estimation): An Analysis using the Reactome database.

Azka Javaid and H. Robert Frost

```
library(scaper)
library(Seurat)
library(SeuratObject)
library(pheatmap)
```

In this vignette, we perform gene set scoring for 30 cytokines using the Reactome database, detailing genes on the corresponding signaling pathway, for a sample of the Peripheral Blood Mononuclear Cells (PBMC) dataset.

We first show the functionality of our package using an example function call for the `geneReactome` function. Below, we create a gene set for IL6 by parsing the molecules on the IL6 signaling pathway using the associated XML file and the `genesetReactome` function, which operates by identifying the molecules on each signaling pathway, which are retrieved from the diagram file associated with an event pathway hierarchy that can be downloaded/exported in a BioPAX2 format. In the call below, we extract 10 genes that are on the pathway signaling hierarchy related to the IL6 cytokine. This function is currently operational for the IL4 and IL6 cytokines. For all other cytokines, specify the Reactome pathway hierarchy files under each cytokine-specific folder. Lastly, we are displaying the list of all scored cytokines using the Reactome database with the `supportedCytokines` function.

```
genesetReactome(cytokine.eval = "IL6",
                file.name = system.file("extdata",
                                        "IL6_Interleukin6_signaling.xml",
                                        package = "scaper")) %>% head(10)
#>     gene cytokine
#> 1   JAK1      IL6
#> 2  IL6ST      IL6
#> 3    IL6      IL6
#> 4   IL6R      IL6
#> 5   JAK2      IL6
#> 6   TYK2      IL6
#> 7    JAK      IL6
#> 8  STAT1      IL6
#> 9  STAT3      IL6
#> 10  STAT      IL6
supportedCytokines(database = "reactome")
#>  [1] "BDNF"    "BMP2"    "BMP4"    "BMP6"    "CD40LG"  "CXCL12"  "EGF"
#>  [8] "FGF2"    "HGF"     "IFNG"    "IL10"    "IL13"    "IL15"    "IL17A"
#> [15] "IL1A"    "IL1B"    "IL2"     "IL21"    "IL22"    "IL27"    "IL3"
#> [22] "IL4"     "IL6"     "LIF"     "OSM"     "TGFB1"   "TGFB3"   "TNFSF12"
#> [29] "VEGFA"   "WNT3A"
```
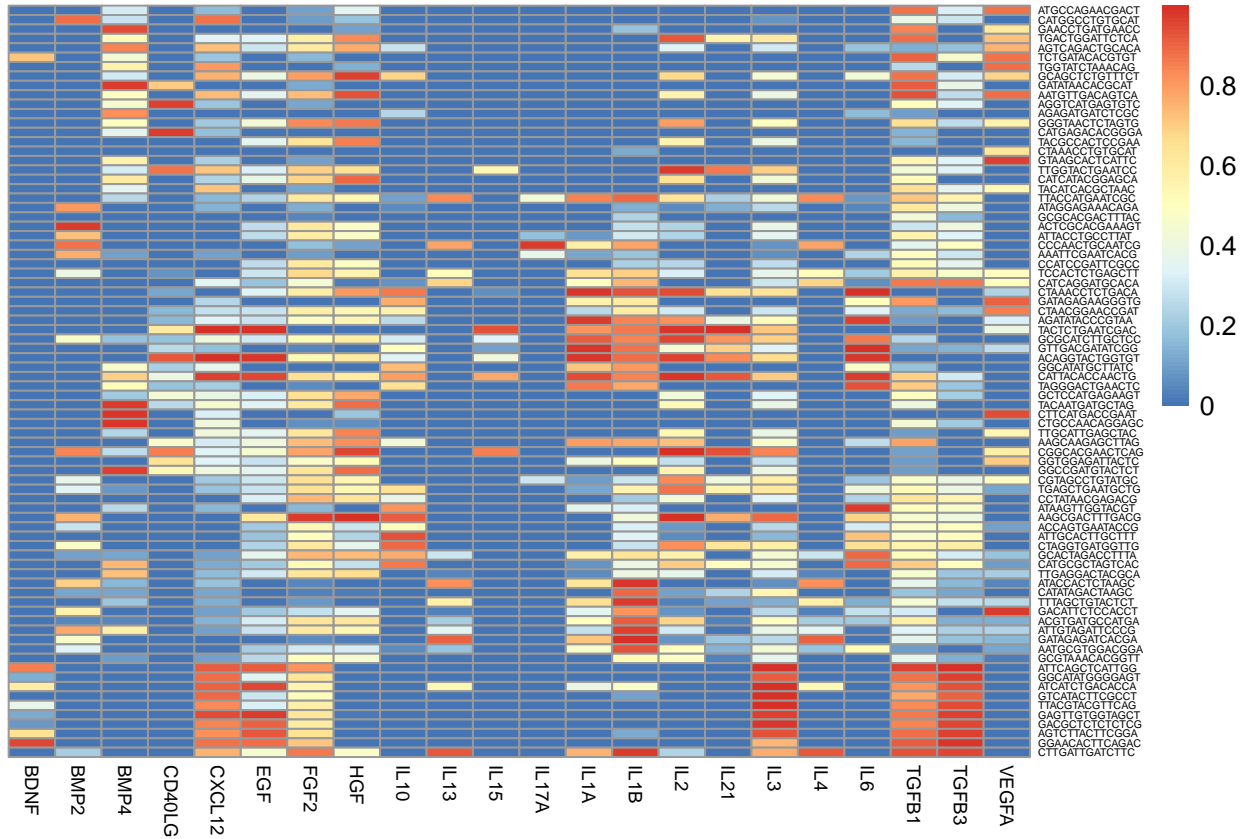
Next, we perform gene set scoring using the `scapeForSeurat` function, which accepts a Seurat object, with

the Reactome gene sets on the `pbmc_small` sampled dataset from the Seurat package, which contains about 230 features/genes assessed over 80 samples/cells. The output is a `VAMcdf` assay in the Reactome.score.output object which is 41 by 80 and contains scored expression for 41 cytokines in 80 cells. We can extract that object and create a heatmap unclustered and clustered visualization of the output scores, as indicated below.

```r
Reactome.score.output <- scapeForSeurat(seurat.object = pbmc_small,
                                        database = "reactome",
                                        normalize = TRUE)
reactome_mat <- as.data.frame(t(as.matrix(Reactome.score.output@assays$VAMcdf@data)))
pheatmap(reactome_mat, fontsize_row = 4, fontsize_col = 7,
         cluster_rows = FALSE, cluster_cols = FALSE)
```



```r
pheatmap(reactome_mat, fontsize_row = 4, fontsize_col = 7)
```