Introduction

In this vignette, we demonstrate the application of StarBioTrek as tool for pathways analysis integrating different data types. For basic use of the StarBioTrek package, please refer to the vignette Working with StarBioTrek package.

StarBioTrek is used as tool to measure pathway activity and pathway cross-talk integrating TCGA data.

Case Study n 1 Relationship between metabolism and cell growth and death in cancer

The aim is the study of ratio among metabolism and cellular processes as cell growth and death in the cancer. According to KEGG pathway there are different pathways involved in the metabolism that can be grouped in six big sets as: Carbohydrate metabolism, Energy metabolism, Lipid metabolism, Aminoacid metabolism, Glycan biosynthesis and metabolism and Metabolism of cofactors and vitamins.

In this case we want to see if there is a correlation between lipid metabolism and cellular processes in cancer. First of all we download the set of pathways for the analyses.

For lipid metabolism:

```r
path_lip<-getKEGGdata(KEGG_path="Lip_met")
```

The set of pathways for lipid metabolism includes:Fatty acid biosynthesis, Fatty acid elongation, Fatty acid degradation, Synthesis and degradation of ketone bodies, Cutin, suberin and wax biosynthesis, Steroid biosynthesis, Primary bile acid biosynthesis, Secondary bile acid biosynthesis, Steroid hormone biosynthesis, Glycerolipid metabolism, Glycerophospholipid metabolism, Ether lipid metabolism, Sphingolipid metabolism, Arachidonic acid metabolism, Linoleic acid metabolism, alpha-Linolenic acid metabolism and Biosynthesis of unsaturated fatty acids.

For cellular processes:

```r
pathcell_grow_d<-getKEGGdata(KEGG_path="cell_grow_d")
```

The set of pathways for cellular processes includes:Cell cycle, Apoptosis and p53 signaling pathway.

Then, we use the function dev_std_crtlk to create a measure of pathway cross-talk (pairwise pathway measure) using TCGA data (e.g. Data_CANCER_normUQ_filt).

```r
score_euc_dist_Lip_met<-dev_std_crtlk(dataFilt=Data_CANCER_normUQ_filt,path_lip)
```

The function svm_classification is used to obtain the best pairwise of pathway able to classify normal vs breast cancer. The training dataset was 60/100 of the data while the testing 40/100. In this analysis we considered the two classes from TCGA: normal and tumour. The output will be a list of AUC value for each pairwise measure of pathway.
We considered the pairwise of pathways that obtained a performance of AUC major 0.80.

```r
tumo <- SelectedSample(Dataset = Data_CANCER_normUQ_filt, typesample = "tumor")[, 1:100]

norm <- SelectedSample(Dataset = Data_CANCER_normUQ_filt, typesample = "normal")[, 1:100]

nf <- 60

res_class <- svm_classification(TCGA_matrix = score_euc_dist_Lip_met, nfs = nf,
                                normal = colnames(norm), tumour = colnames(tumo))
```

The function `process_matrix` creates a TCGA matrix with the measure of cross-talk previously used, only for the pairwise pathway obtained by `select_class`.

```r
better_perf <- select_class(auc.df = res_class, cutoff = 0.80)

matrix_best_perf <- process_matrix(measure = score_euc_dist_Lip_met, list_perf = better_perf)

score_bestlipd <- colMeans(tumo_bestlipd)
```

Now we want to create a pathawy cross-talk also for the pathways of cellular processes.

First of all we select the tumour samples and then create a matrix of distance using `dev.std.crlk`.

```r
tumo_cell_grow_d <- SelectedSample(Dataset = Data_CANCER_normUQ_filt, typesample = "tumor")[, 1:100]

score_euc_dist_cell_grow_d <- dev.std.crlk(dataFilt = tumo_cell_grow_d, pathcell_grow_d)

score__cell_grow_d <- process_matrix_cell_process(score_euc_dist_cell_grow_d)

score__cell_grow_d_mean <- colMeans(score__cell_grow_d)
```

Now we want to see if there is a correlation among cellular processes and the lipid metabolism in breast cancer.

```r
correlazione <- cor(score__cell_grow_d_mean, score_bestlipd)

plot_matrix <- cbind(score__cell_grow_d_mean, score_bestlipd)
```

**References**

