Supplementary Information: *Rtpca: an R package for differential thermal coaggregation analysis*

***Nils Kurzawa*[1], *André Mateus*[1], *and Mikhail M. Savitski*[1]**

[1]European Molecular Biology Laboratory (EMBL), Genome Biology Unit

18 March, 2020

Package

Rtpca 0.0.99

Contents

1 Introduction

Thermal proteome profiling (TPP) (Mateus et al., 2020; Savitski et al., 2014) is a mass spectrometry-based, proteom-wide implemention of the cellular thermal shift assay (Molina et al., 2013). It was originally developed to study drug-(off-)target engagement. However, it was realized that profiles of interacting protein pairs appeared more similar than by change (Tan et al., 2018, Becher et al. (2018)) which was coined as 'thermal proximity co-aggregation' (TPCA) (Tan et al., 2018). The R package `Rtpca` enables analysis of TPP datasets in the light of protein-protein interactions or protein complex associations. Here, we exemplify the analysis based on a dataset by Hashimoto et al. (Hashimoto et al., 2020) studying the dynamics of protein-protein interactions in human cells during Cytomegalovirus infection.

2 Step-by-step walk through the data analysis

First, we need to load the required libraries (these need to be installed as specified in the comments):

```
library(dplyr) # install.packages("dplyr")
library(readxl) # install.packages("readxl")
library(Rtpca) # require(devtools); devtools::install_github("nkurzaw/Rtpca")
library(ggplot2) # install.packages("ggplot2")
```

Then, we download the supplementary data from Hashimoto et al. which contains the TPP data:

```
if(!file.exists("41467_2020_14586_MOESM4_ESM.xlsx")){
    download.file(
    url = "https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-020-14586-5/MediaObjects/41467_202
    destfile = "41467_2020_14586_MOESM4_ESM.xlsx")
}
```

We then read in the excel table:

```
suppl1_df <- read_xlsx("41467_2020_14586_MOESM4_ESM.xlsx", sheet = "TableS1B")
```

From the table we create a matrix of the median relative fold changes for the Mock dataset:

```
mock_df <- suppl1_df %>%
    dplyr::select(Uniprot_ID, Gene_symbol, dplyr::matches("Mock")) %>%
    na.omit()
mock_mat <- as.matrix(mock_df[,-c(1,2)])
rownames(mock_mat) <- mock_df$Gene_symbol
id_dup <- which(duplicated(rownames(mock_mat)))
mock_mat <- mock_mat[-id_dup,]
```

The same is done for the 24h post infection data:

```
inf24h_df <- suppl1_df %>%
    dplyr::select(Uniprot_ID, Gene_symbol, dplyr::matches("24")) %>%
    na.omit()
```

```
inf24h_mat <- as.matrix(inf24h_df[,-c(1,2)])
rownames(inf24h_mat) <- inf24h_df$Gene_symbol
id_dup_24h <- which(duplicated(rownames(inf24h_mat)))
inf24h_mat <- inf24h_mat[-id_dup_24h,]
```

We now load our annotation of protein pairs within the same protein complexes (this dataset comes with the `Rtpca` package, please make sure to cite it whenever you use it Ori et al. (2016)):

```
data("ori_et_al_complex_ppis")
```

We extend this annotation by one of the significantly changing protein-protein interactions that Hashimoto et al. reported that is not present in our annotation:

```
ori_et_al_complex_ppis_extended <-
    bind_rows(ori_et_al_complex_ppis,
           tibble(
               complex_name = "PI3K Pathway",
               x = "CRK",
               y = "PIK3R1",
               pair = "CRK:PIK3R1"))
```

2.1 Running Rtpca for a single condition

To run Rtpca for the mock dataset:

```
mockTPCA <- runTPCA(
    objList = list(mock_mat),
    ppiAnno = ori_et_al_complex_ppis
)
## Checking function arguments.
##
## Creating distance matrices.
##
## Testing for complex co-aggregation.
##
## Performning PPi ROC analysis.
```

We can then look at the ROC curve with:

```
plotPPiRoc(mockTPCA, computeAUC = TRUE)
```

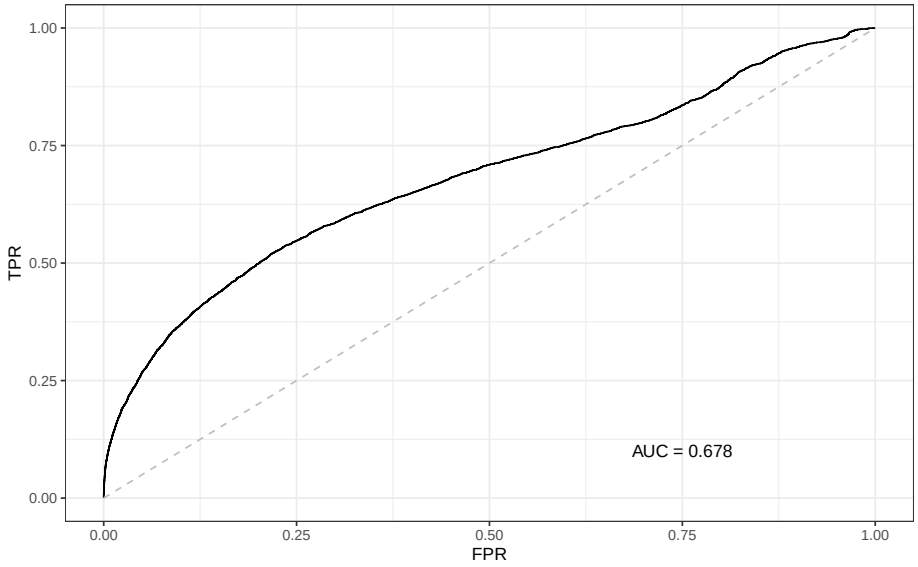

Figure 1: **ROC curve for identification of protein pairs part of the same complex in the Mock dataset**

and the result table of significantly co-melting protein pairs:

```
head(mockTPCA@tpcaResultTable %>% filter(p_adj < 0.01) %>% arrange(p_adj))
## # A tibble: 6 x 5
##   complex_name count mean_dist p_value p_adj
##   <chr>        <int>     <dbl>   <dbl> <dbl>
## 1 ACTR3:ARPC2      2    0.0406       0     0
## 2 AIMP1:AIMP2      2    0.0427       0     0
## 3 AIMP1:QARS       2    0.0511       0     0
## 4 ARCN1:COPB1      2    0.0292       0     0
## 5 CCT2:CCT6A       2    0.0520       0     0
## 6 COPS4:GPS1       2    0.0518       0     0
```

The same can be done for the 24h post infection dataset:

```
inf24hTPCA <- runTPCA(
    objList = list(inf24h_mat),
    ppiAnno = ori_et_al_complex_ppis
)
## Checking function arguments.
##
## Creating distance matrices.
##
## Testing for complex co-aggregation.
##
## Performning PPi ROC analysis.
```

```
plotPPiRoc(inf24hTPCA, computeAUC = TRUE)
```

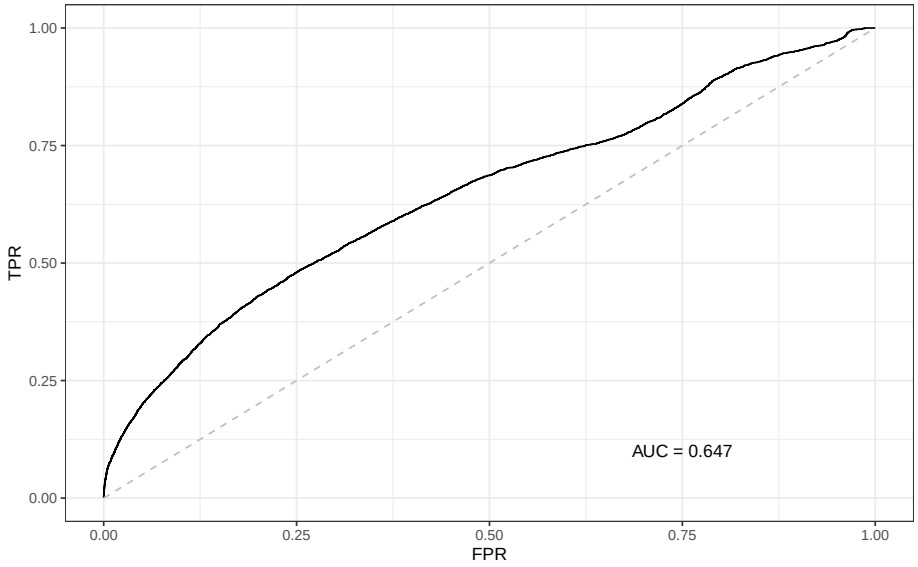

Figure 2: **ROC curve for identification of protein pairs part of the same complex in the 24h post infection dataset**

```
head(inf24hTPCA@tpcaResultTable %>% filter(p_adj < 0.01) %>% arrange(p_adj))
## # A tibble: 6 x 5
##   complex_name      count mean_dist p_value p_adj
##   <chr>             <int>     <dbl>   <dbl> <dbl>
## 1 AP2B1:AP2M1           2    0.0348       0     0
## 2 HSP90AA1:HSP90AB1     2    0.0245       0     0
## 3 PSMA3:PSMA5           2    0.0333       0     0
## 4 PSMC1:PSMC2           2    0.0361       0     0
## 5 TUBA1A:TUBA1C         2    0.0265       0     0
## 6 TUBA1A:TUBA4A         2    0.0263       0     0
```

2.2 Running a differential Rtpca analysis

To run the differential TPCA test between the mock and the 24h post infection datasets we run:

```
mockInf24hDiffTPCA <- runDiffTPCA(
    objList = list(mock_mat),
    contrastList = list(inf24h_mat),
    ppiAnno = ori_et_al_complex_ppis_extended
)
```

And we can inspect the result with a volcano plot:

```
plotDiffTpcaVolcano(mockInf24hDiffTPCA)
```

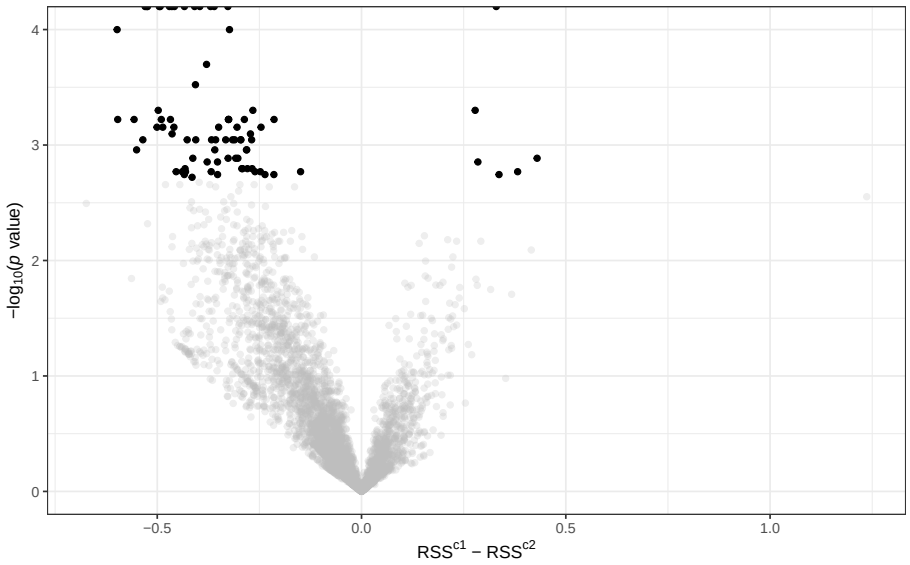

Figure 3: **Volcanoplot of differential TPCA results**
Colored in black are the significantly differentially coaggregating protein pairs betwee both datasets.

The interaction between CRK and PIK3R1 is one of the significantly changing interations, which we can see by labeling it in the volcano plot:

```
plotDiffTpcaVolcano(mockInf24hDiffTPCA) +
    geom_point(color = "red",
               data = filter(mockInf24hDiffTPCA@diffTpcaResultTable,
                             pair == "CRK:PIK3R1"))
```

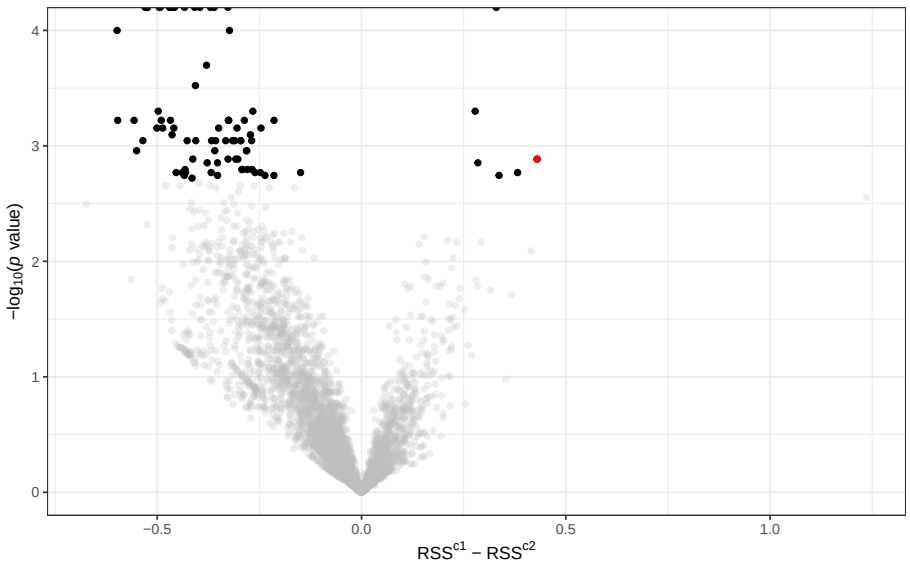

Figure 4: **Volcanoplot of differential TPCA results**
Colored in red is the protein pair CRK-PIK3R1 which is significantly coaggregating more strongly in the 24h post infection data.

As we can see from figure 5 many of the protein-protein interactions that appear to losen up in the post infection dataset compared to Mock are between members of ribosome. This could reflect translational stalling in response to viral infection.

```
plotDiffTpcaVolcano(mockInf24hDiffTPCA) +
    geom_point(color = "steelblue",
               data = filter(mockInf24hDiffTPCA@diffTpcaResultTable,
                             p_adj < 0.1,
                             grepl("^RP[L,S]", pair)))
```

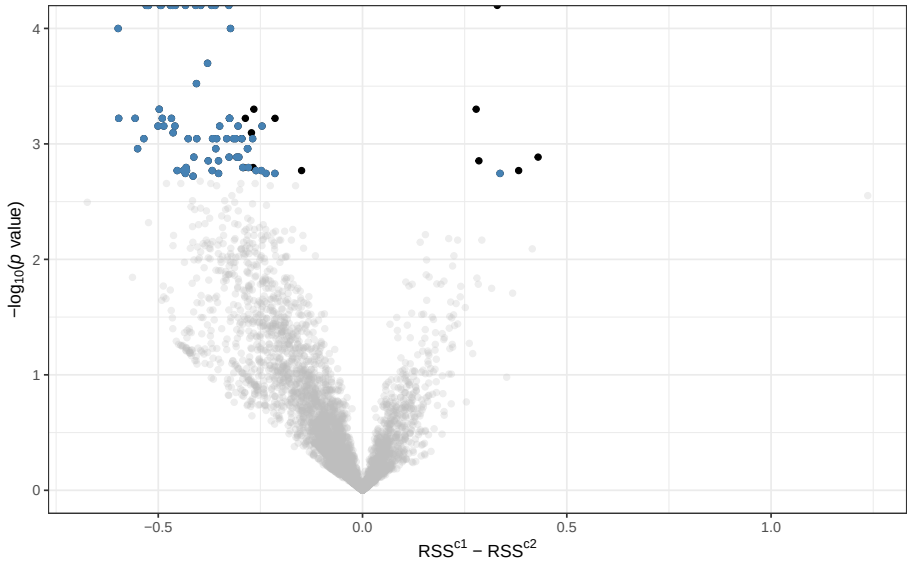

Figure 5: Volcanoplot of differential TPCA results
Colored in blue are significantly changing protein pairs including at least one ribosomal protein.

```
sessionInfo()
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel  stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
## [1] ggplot2_3.2.1       Rtpca_0.0.99        tidyr_1.0.0
## [4] Biobase_2.44.0      BiocGenerics_0.30.0 readxl_1.3.1
## [7] dplyr_0.8.3         BiocStyle_2.12.0
```

```
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.2         plyr_1.8.4         cellranger_1.1.0
##  [4] pillar_1.4.2       compiler_3.6.1     BiocManager_1.30.9
##  [7] tools_3.6.1        zeallot_0.1.0      digest_0.6.22
## [10] evaluate_0.14      tibble_2.1.3       lifecycle_0.1.0
## [13] gtable_0.3.0       pkgconfig_2.0.3    rlang_0.4.1
## [16] cli_1.1.0          yaml_2.2.0         xfun_0.10
## [19] withr_2.1.2        stringr_1.4.0      knitr_1.25
## [22] pROC_1.15.3        vctrs_0.2.0        grid_3.6.1
## [25] tidyselect_0.2.5   glue_1.3.1         R6_2.4.0
## [28] fansi_0.4.0        rmarkdown_1.16     bookdown_0.14
## [31] purrr_0.3.3        magrittr_1.5       backports_1.1.5
## [34] scales_1.0.0       htmltools_0.4.0    assertthat_0.2.1
## [37] colorspace_1.4-1   labeling_0.3       utf8_1.1.4
## [40] stringi_1.4.3      lazyeval_0.2.2     munsell_0.5.0
## [43] crayon_1.3.4
```

References

Becher, I., Andrés-Pons, A., Romanov, N., Stein, F., Schramm, M., Baudin, F., Helm, D., Kurzawa, N., Mateus, A., Mackmull, M.-T., et al. (2018). Pervasive Protein Thermal Stability Variation during the Cell Cycle. Cell 1–13.

Hashimoto, Y., Sheng, X., Murray-Nerger, L.A., and Cristea, I.M. (2020). Temporal dynamics of protein complex formation and dissociation during human cytomegalovirus infection. Nature Communications.

Mateus, A., Kurzawa, N., Becher, I., Sridharan, S., Helm, D., Stein, F., Typas, A., and Savitski, M.M. (2020). Thermal proteome profiling for interrogating protein interactions. Molecular Systems Biology 1–11.

Molina, D.M., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E.A., Dan, C., Sreekumar, L., Cao, Y., and Nordlund, P. (2013). Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. Science *341*, 84–88.

Ori, A., Iskar, M., Buczak, K., Kastritis, P., Parca, L., Andrés-pons, A., Singer, S., Bork, P., and Beck, M. (2016). Spatiotemporal variation of mammalian protein complex stoichiometries. Genome Biology 1–15.

Savitski, M.M., Reinhard, F.B.M., Franken, H., Werner, T., Savitski, M.F., Eberhard, D., Martinez Molina, D., Jafari, R., Dovega, R.B., Klaeger, S., et al. (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. Science *346*, 1255784.

Tan, C.S.H., Go, K.D., Bisteau, X., Dai, L., Yong, C.H., Prabhu, N., Ozturk, M.B., Lim, Y.T., Sreekumar, L., Lengqvist, J., et al. (2018). Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. Science *0346*, 1–12.