Analysis of GTP dataset

Nils Kurzawa[1]

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

02 April, 2020

Package

TPP2D 1.3.10

Contents

1 Step-by-step walk through the anlysis

```
# This script uses the development version of TPP2D
if(require("BiocManager"))
  install.packages("BiocManager")
BiocManager::install("nkurzaw/TPP2D")
```

Load required libraries

```
library(TPP2D)
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
library(dplyr)
library(tidyr)
library(ggplot2)
library(readxl)
library(UpSetR)
```

Define plot style

```
theme_paper <- theme_bw(base_size = 6) +
  theme(legend.background = element_blank(),
        legend.key = element_blank(),
        panel.background = element_blank(),
        panel.grid.major = element_line(colour = "grey92", size = 0.25),
        panel.grid.minor = element_line(colour = "grey92", size = 0.15),
        panel.border = element_blank(),
        strip.background = element_blank(),
        plot.background = element_blank(),
        complete = TRUE,
        axis.line = element_line(color = "black", size = 0.25),
        text = element_text(size = 7),
        axis.ticks = element_line(color = "black", size = 0.25),
        axis.title = element_text(size = 8),
        axis.text = element_text(size = 6))
```

Annotate GTP and ATP binding proteins

```
all_atp_binder <- AnnotationDbi::select(
  org.Hs.eg.db::org.Hs.eg.db,
  keys = "GO:0005524",
  columns = c("SYMBOL", "IPI"),
  keytype = "GOALL")
##
```

```
## 'select()' returned 1:many mapping between keys and columns

all_gtp_binder <- AnnotationDbi::select(
  org.Hs.eg.db::org.Hs.eg.db,
  keys = "GO:0005525",
  columns = c("SYMBOL", "IPI"),
  keytype = "GOALL")
## 'select()' returned 1:many mapping between keys and columns
```

Download the supplementary table from the journal's website

```
# still needs to be added when supplementary table is available online
```

Read in the data and reformat to a data frame as would be obtained after import of the raw data:

```
gtp_raw <- read_xlsx("Supplementary_Data_3.xlsx", sheet = "GTP") %>%
  dplyr::select(representative,
                clustername,
                qupm,
                qusm,
                experiment,
                temperature,
                matches("sumionarea"),
                matches("rel_fc_protein"))  %>%
  gather(key, value, matches("sumionarea"), matches("rel_fc_protein")) %>%
  mutate(conc = as.numeric(gsub("uM", "", gsub(".+_protein_[0-9,H,L]+_[0-9,H,L]+_", "", key))),
         temperature = as.numeric(gsub("C", "", temperature)),
         key = case_when(grepl("sumionarea", key) ~ "raw_value",
                         grepl("rel_fc", key) ~ "rel_value")) %>%
  spread(key, value) %>%
  arrange(representative, temperature, conc) %>%
  group_by(clustername, temperature, conc) %>%
  filter(qupm == max(qupm),
         qusm == max(qusm),
         raw_value == max(raw_value)) %>%
  filter(!duplicated(clustername)) %>%
  ungroup %>%
  mutate(log2_value = log2(raw_value),
         log_conc = log10(conc/1e6)) %>%
  filter(qupm > 1)

# resolve ambiguous protein names
gtp_fil <- resolveAmbiguousProteinNames(gtp_raw)

# recompute reporter ion signal from robust Isobarquant fold changes
gtp_df <- recomputeSignalFromRatios(gtp_fil)
```

Compute null and alternative model fits and extract parameters

```
gtp_params_df <- getModelParamsDf(gtp_df, maxit = 500)
saveRDS(gtp_params_df, file = "../pre_run_data/gtp_params_df.rds")
```

Compute F statistics

```
gtp_fstat_df <- computeFStatFromParams(gtp_params_df)
```

Get B datasets expected under the null model and perform model fitting and compute F statistics to obtain a null distribution for FDR calibration:

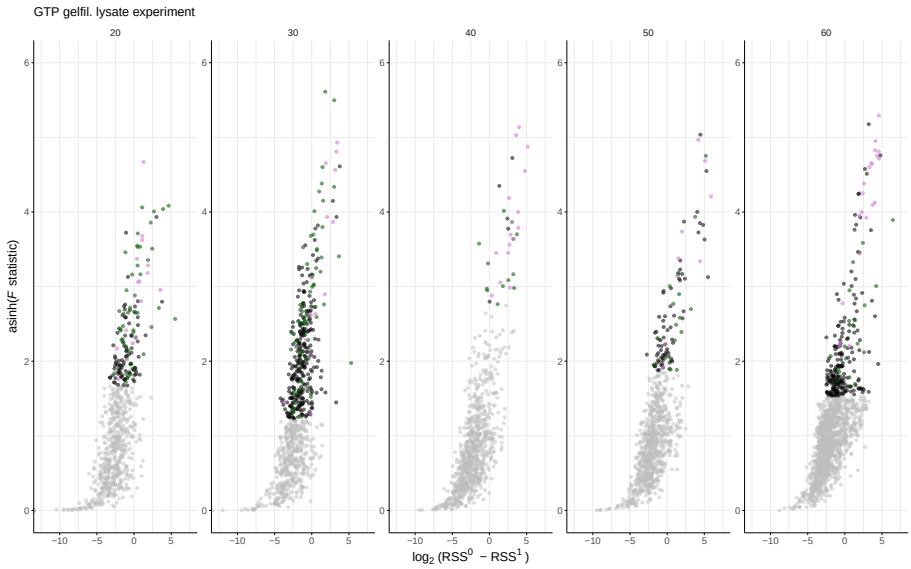
```
set.seed(12, kind = "L'Ecuyer-CMRG")
gtp_null_df <- bootstrapNullAlternativeModel(
  df = gtp_df, params_df = gtp_params_df,
  maxit = 500, B = 100,
  BPPARAM = BiocParallel::MulticoreParam(workers = 20, progressbar = TRUE),
  verbose = FALSE)
saveRDS(jq1_null_df, file = "../pre_run_data/jq1_null_df.rds")
```

Compute FDR and find hits:

```
gtp_fdr_df <- getFDR(df_out = gtp_fstat_df,
                     df_null = gtp_null_df)

gtp_hits_df <- findHits(gtp_fdr_df, alpha = 0.1)
```

```
ggplot(gtp_fdr_df %>%
         filter(dataset == "true"),
       aes(log2(rssH0 - rssH1), asinh(F_statistic))) +
  geom_point(color = "gray", alpha = 0.5, size = 0.5) +
  geom_point(color = "black", alpha = 0.5,
             size = 0.5,
             data = gtp_hits_df %>%
              filter(!clustername %in%
                       all_atp_binder$SYMBOL,
                     !clustername %in%
                       all_gtp_binder$SYMBOL)) +
  geom_point(color = "darkgreen", alpha = 0.5,
             size = 0.5,
             data = filter(gtp_hits_df, clustername %in%
                             all_atp_binder$SYMBOL)) +
  geom_point(color = "violet", alpha = 0.5,
             size = 0.5,
             data = filter(gtp_hits_df, clustername %in%
                             all_gtp_binder$SYMBOL)) +
  facet_wrap(~nObsRound, scales = "free", ncol = 5) +
   labs(x = expression('log'[2]~'(RSS'^0~' - RSS'^1~')'),
        y = expression('asinh('*italic(F)*' statistic)')) +
  coord_cartesian(xlim = c(-12.5, 7.5), ylim = c(0, 6)) +
  ggtitle("GTP gelfil. lysate experiment") +
  theme_paper
```



2 Compare sets

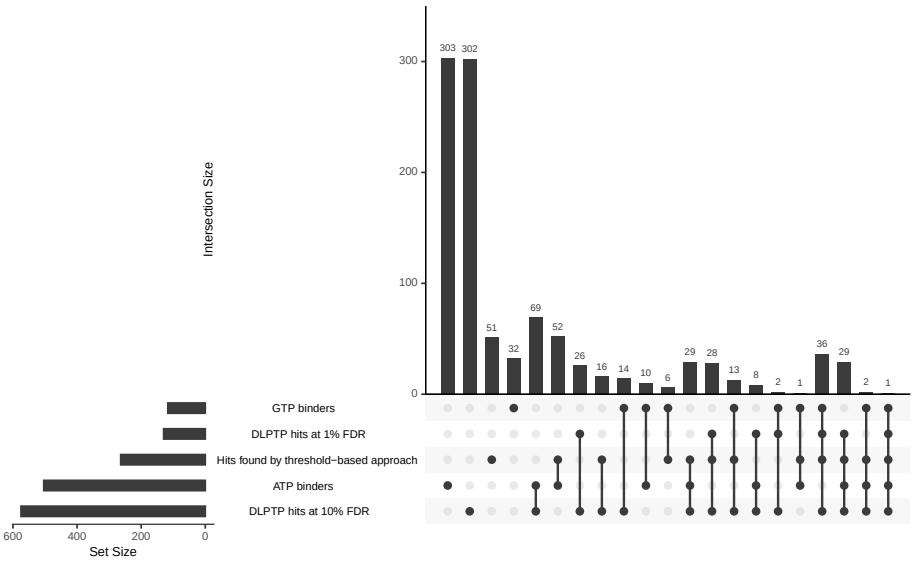
```
gtp_threshold_df <- read_xlsx("Supplementary_Data_3.xlsx", sheet = "GTP") %>%
  dplyr::select(clustername, stabilized_hits_found_by_threshold) %>%
  filter(stabilized_hits_found_by_threshold)

gtp_hits_10per <- findHits(gtp_fdr_df, 0.1) %>%
  filter(detected_effectH1 == "stability", slopeH1 > 0)
gtp_hits_1per <- findHits(gtp_fdr_df, 0.01) %>%
  filter(detected_effectH1 == "stability", slopeH1 > 0)


set_intersect_df <- data.frame(
  gene_name = filter(gtp_fdr_df, !duplicated(clustername))$clustername) %>%
  mutate(`ATP binders` = as.numeric(gene_name %in% all_atp_binder$SYMBOL),
         `GTP binders` = as.numeric(gene_name %in% all_gtp_binder$SYMBOL),
         `DLPTP hits at 10% FDR` =
           as.numeric(gene_name %in% gtp_hits_10per$clustername),
         `DLPTP hits at 1% FDR` =
           as.numeric(gene_name %in% gtp_hits_1per$clustername),
         `Hits found by threshold-based approach` =
           as.numeric(gene_name %in% gtp_threshold_df$clustername))

upset(set_intersect_df, nsets = 5)
```

Analysis of GTP dataset




```
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] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] UpSetR_1.4.0     readxl_1.3.1     ggplot2_3.2.1    tidyr_1.0.0
## [5] TPP2D_1.3.10     dplyr_0.8.3      BiocStyle_2.12.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_0.2.5     xfun_0.10            purrr_0.3.3
##  [4] colorspace_1.4-1     vctrs_0.2.0          htmltools_0.4.0
##  [7] stats4_3.6.1         yaml_2.2.0           blob_1.2.0
## [10] rlang_0.4.1          pillar_1.4.2         glue_1.3.1
## [13] withr_2.1.2          DBI_1.0.0            BiocParallel_1.18.1
## [16] BiocGenerics_0.30.0  bit64_0.9-7          foreach_1.4.7
## [19] lifecycle_0.1.0      plyr_1.8.4           stringr_1.4.0
## [22] munsell_0.5.0        gtable_0.3.0         cellranger_1.1.0
## [25] zip_2.0.4            memoise_1.1.0        codetools_0.2-16
## [28] evaluate_0.14        labeling_0.3         Biobase_2.44.0
## [31] knitr_1.25           IRanges_2.18.3       doParallel_1.0.15
## [34] parallel_3.6.1       AnnotationDbi_1.46.1 Rcpp_1.0.2
## [37] scales_1.0.0         backports_1.1.5      BiocManager_1.30.9
```

```
## [40] org.Hs.eg.db_3.8.2   S4Vectors_0.22.1     bit_1.1-14
## [43] gridExtra_2.3        digest_0.6.22        stringi_1.4.3
## [46] openxlsx_4.1.0.1     bookdown_0.14        grid_3.6.1
## [49] tools_3.6.1          bitops_1.0-6         magrittr_1.5
## [52] RSQLite_2.1.2        lazyeval_0.2.2       RCurl_1.95-4.12
## [55] tibble_2.1.3         crayon_1.3.4         pkgconfig_2.0.3
## [58] zeallot_0.1.0        ellipsis_0.3.0       MASS_7.3-51.4
## [61] assertthat_0.2.1     rmarkdown_1.16       iterators_1.0.12
## [64] R6_2.4.0             compiler_3.6.1
```