Analysis of PCI-34051 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 `TPP2D` analysis

```
# 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(clusterProfiler)
##
## Registered S3 method overwritten by 'enrichplot':
##   method               from
##   fortify.enrichResult DOSE
## clusterProfiler v3.12.0  For help: https://guangchuangyu.github.io/software/clusterProfiler
##
## If you use clusterProfiler in published research, please cite:
## Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He. clusterProfiler: an R package for comparing biologica
library(org.Hs.eg.db)
## Loading required package: AnnotationDbi
## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:dplyr':
##
##     combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get,
##     grep, grepl, intersect, is.unsorted, lapply, Map, mapply, match,
##     mget, order, paste, pmax, pmax.int, pmin, pmin.int, Position,
##     rank, rbind, Reduce, rownames, sapply, setdiff, sort, table,
##     tapply, union, unique, unsplit, which, which.max, which.min
## Loading required package: Biobase
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
## Loading required package: IRanges
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:tidyr':
##
##     expand
## The following objects are masked from 'package:dplyr':
##
##     first, rename
## The following object is masked from 'package:base':
##
##     expand.grid
##
## Attaching package: 'IRanges'
## The following objects are masked from 'package:dplyr':
##
##     collapse, desc, slice
##
## Attaching package: 'AnnotationDbi'
## The following object is masked from 'package:dplyr':
##
##     select
##
```

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))
```

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:

```
pci_raw <- read_xlsx("Supplementary_Data_2.xlsx", sheet = "PCI34051") %>%
  dplyr::select(representative,
                clustername,
                qupm,
                qusm,
                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
pci_fil <- resolveAmbiguousProteinNames(pci_raw)

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

Compute null and alternative model fits and extract parameters

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

Compute F statistics

```
pci_fstat_df <- computeFStatFromParams(pci_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")
pci_null_df <- bootstrapNullAlternativeModel(
  df = pci_lys_df, params_df = pci_params_df,
  maxit = 500, B = 100,
  BPPARAM = BiocParallel::MulticoreParam(workers = 20, progressbar = TRUE),
  verbose = FALSE)
saveRDS(pci_null_df, file = "../pre_run_data/pci_null_df.rds")
```

Remove carry-over cases:

```
## manually identified carry-over cases
carry_over_cases <-
c("ALDH1B1", "BTK", "CAMK2G", "CSK", "CSNK2A1", "CSNK2A2", "GAK",
  "CSNK2B", "GSK3A", "LYN", "MAP4K1", "MAPK1", "MAPK9", "NEK9",
  "NQO1", "PDXK", "PRKAA1", "PRKAG1", "RPS6KA1", "ULK3",
  "CDK2", "CDK5")
```

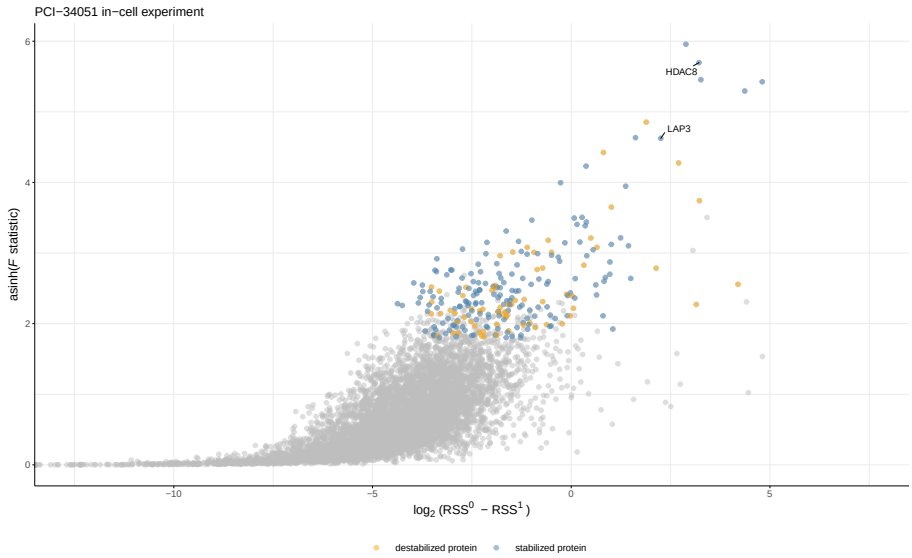
Compute FDR and find hits:

```
pci_fdr_df <- getFDR(df_out = pci_fstat_df %>%
                         filter(!clustername %in% carry_over_cases),
                     df_null = pci_null_df %>%
                         filter(!clustername %in% carry_over_cases))

pci_hits_df <- findHits(pci_fdr_df, alpha = 0.1)
```

```
ggplot(pci_fdr_df %>%
         filter(dataset == "true"),
       aes(log2(rssH0 - rssH1), asinh(F_statistic))) +
  geom_point(color = "gray", alpha = 0.5, size = 1) +
  geom_point(aes(color = group), alpha = 0.5,
             size = 1,
             data = pci_hits_df %>%
                 mutate(group = case_when(
                     slopeH1 > 0 ~ "stabilized protein",
                     slopeH1 < 0 ~ "destabilized protein"))) +
  ggrepel::geom_text_repel(
    aes(label = clustername),
    data = filter(pci_hits_df, clustername %in%
                    c("HDAC8", "LAP3")),
    size = 2, segment.size = 0.2, min.segment.length = unit(1, "pt")) +
  scale_color_manual("", values = c("orange", "steelblue")) +
  coord_cartesian(xlim = c(-12.5, 7.5)) +
  ylab("asinh(F statistic)") +
  labs(x = expression('log'[2]~'(RSS'^0~' - RSS'^1~')'),
       y = expression('asinh('*italic(F)*' statistic)')) +
  ggtitle("PCI-34051 in-cell experiment") +
```

```
  theme_paper +
  theme(legend.position = "bottom")
```



2 Plot example profiles

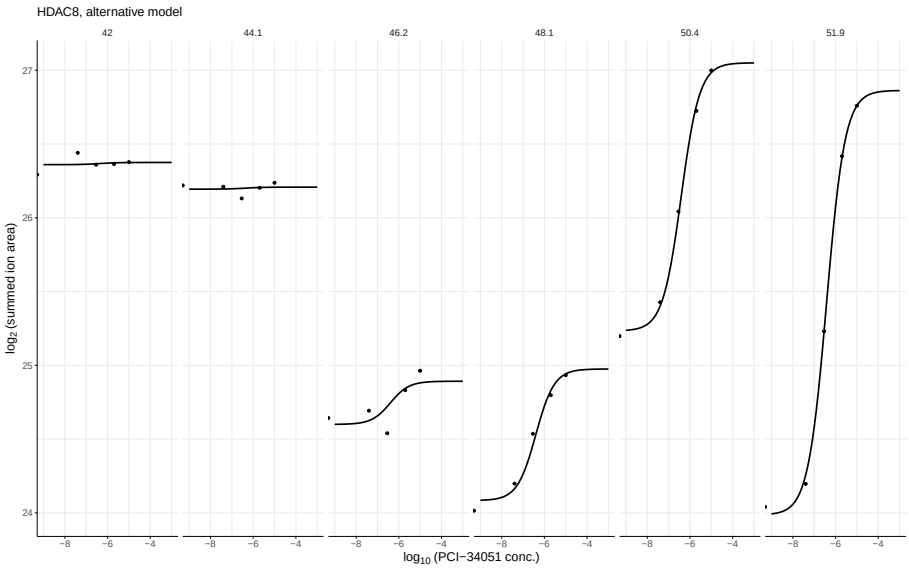
HDAC8

```
hdac8_fit <- plot2dTppFit(pci_df, "HDAC8", "H1")$data

hdac8_df <- filter(pci_df, clustername == "HDAC8")

ggplot(hdac8_fit, aes(log_conc, y_hat)) +
  geom_line() +
  geom_point(aes(log_conc, log2_value),
             data = hdac8_df, size = 0.5) +
  facet_wrap(~temperature, ncol = 6) +
  labs(x = expression('log'[10]~ '(PCI-34051 conc.)'),
       y = expression('log'[2]~ '(summed ion area)')) +
  ggtitle("HDAC8, alternative model") +
  theme_paper
```

Analysis of PCI-34051 dataset

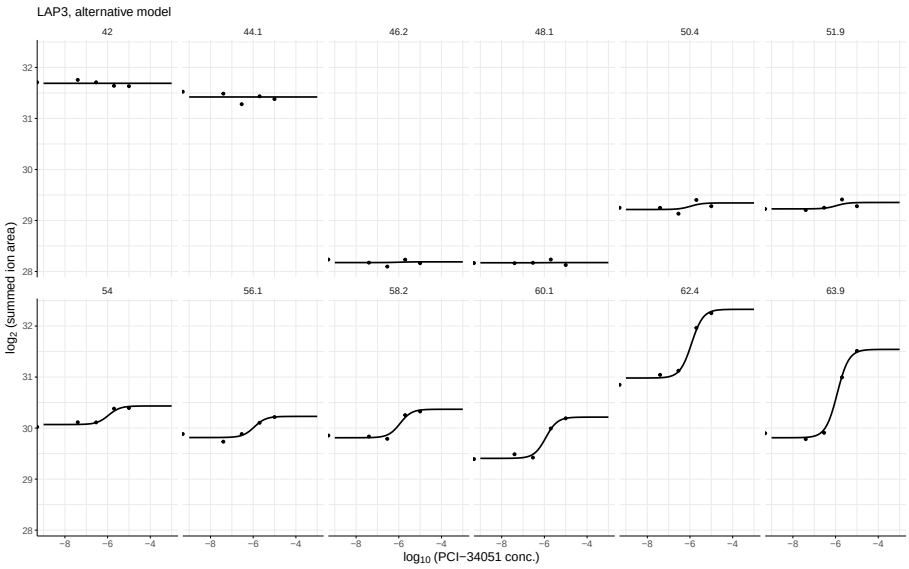


LAP3

```
lap3_fit <- plot2dTppFit(pci_df, "LAP3", "H1")$data

lap3_df <- filter(pci_df, clustername == "LAP3")

ggplot(lap3_fit, aes(log_conc, y_hat)) +
  geom_line() +
  geom_point(aes(log_conc, log2_value),
             data = lap3_df, size = 0.5) +
  facet_wrap(~temperature, ncol = 6) +
  labs(x = expression('log'[10]~ '(PCI-34051 conc.)'),
       y = expression('log'[2]~ '(summed ion area)')) +
  ggtitle("LAP3, alternative model") +
  theme_paper
```



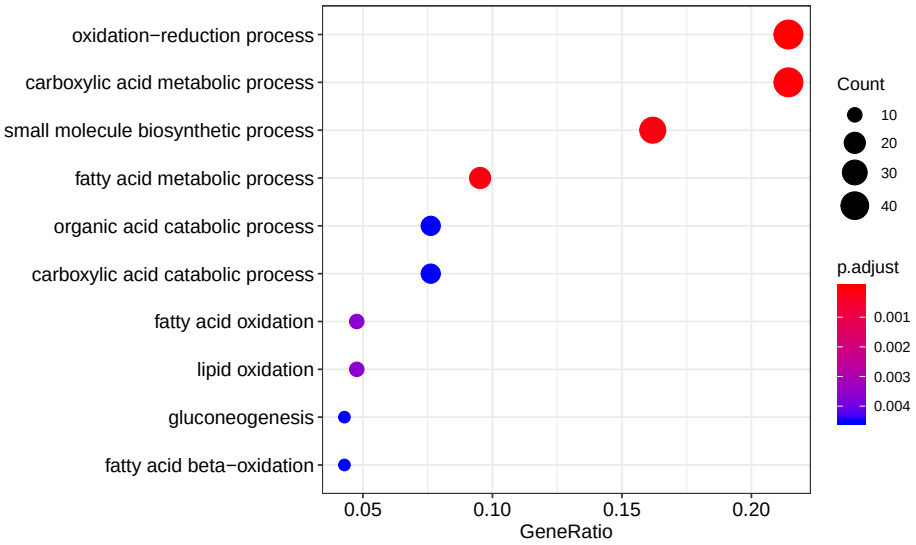
3 GO analysis

```
hits_entrez <- bitr(pci_hits_df$clustername,
                fromType = "SYMBOL",
                toType = c("ENTREZID"),
                OrgDb = org.Hs.eg.db)
## 'select()' returned 1:1 mapping between keys and columns
## Warning in bitr(pci_hits_df$clustername, fromType = "SYMBOL", toType =
## c("ENTREZID"), : 9.21% of input gene IDs are fail to map...

backg_entrez <- bitr(pci_fdr_df$clustername,
                     fromType = "SYMBOL",
                     toType = c("ENTREZID"),
                     OrgDb = org.Hs.eg.db)
## 'select()' returned 1:many mapping between keys and columns
## Warning in bitr(pci_fdr_df$clustername, fromType = "SYMBOL", toType =
## c("ENTREZID"), : 13.42% of input gene IDs are fail to map...


ego <- enrichGO(gene = hits_entrez$ENTREZID,
                universe = backg_entrez$ENTREZID,
                OrgDb = org.Hs.eg.db,
                ont = "BP",
                pAdjustMethod = "BH",
                pvalueCutoff = 0.01,
                qvalueCutoff = 0.05,
                readable = TRUE)

dotplot(ego)
## wrong orderBy parameter; set to default `orderBy = "x"`
```



```
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  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] org.Hs.eg.db_3.8.2     AnnotationDbi_1.46.1   IRanges_2.18.3
##  [4] S4Vectors_0.22.1       Biobase_2.44.0         BiocGenerics_0.30.0
##  [7] clusterProfiler_3.12.0 readxl_1.3.1           ggplot2_3.2.1
## [10] tidyr_1.0.0            TPP2D_1.3.10           dplyr_0.8.3
## [13] BiocStyle_2.12.0
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-6        enrichplot_1.4.0    bit64_0.9-7
##  [4] progress_1.2.2      httr_1.4.1          doParallel_1.0.15
##  [7] RColorBrewer_1.1-2  UpSetR_1.4.0        tools_3.6.1
## [10] backports_1.1.5     R6_2.4.0            DBI_1.0.0
## [13] lazyeval_0.2.2      colorspace_1.4-1    withr_2.1.2
## [16] prettyunits_1.0.2   tidyselect_0.2.5    gridExtra_2.3
## [19] bit_1.1-14          compiler_3.6.1      xml2_1.2.2
## [22] labeling_0.3        triebeard_0.3.0     bookdown_0.14
## [25] scales_1.0.0        ggridges_0.5.1      stringr_1.4.0
## [28] digest_0.6.22       rmarkdown_1.16      DOSE_3.10.2
## [31] pkgconfig_2.0.3     htmltools_0.4.0     rlang_0.4.1
## [34] RSQLite_2.1.2       gridGraphics_0.4-1  farver_1.1.0
## [37] jsonlite_1.6        BiocParallel_1.18.1 GOSemSim_2.10.0
## [40] zip_2.0.4           RCurl_1.95-4.12     magrittr_1.5
## [43] ggplotify_0.0.4     GO.db_3.8.2         Matrix_1.2-17
## [46] Rcpp_1.0.2          munsell_0.5.0       viridis_0.5.1
## [49] lifecycle_0.1.0     stringi_1.4.3       yaml_2.2.0
## [52] ggraph_2.0.0        MASS_7.3-51.4       plyr_1.8.4
## [55] qvalue_2.16.0       grid_3.6.1          blob_1.2.0
## [58] ggrepel_0.8.1       DO.db_2.9           crayon_1.3.4
## [61] lattice_0.20-38     cowplot_1.0.0       graphlayouts_0.5.0
## [64] splines_3.6.1       hms_0.5.1           zeallot_0.1.0
## [67] knitr_1.25          pillar_1.4.2        fgsea_1.10.1
## [70] igraph_1.2.4.1      reshape2_1.4.3      codetools_0.2-16
## [73] fastmatch_1.1-0     glue_1.3.1          evaluate_0.14
## [76] data.table_1.12.6   BiocManager_1.30.9  urltools_1.7.3
## [79] vctrs_0.2.0         tweenr_1.0.1        foreach_1.4.7
```

```
## [82] cellranger_1.1.0    gtable_0.3.0        purrr_0.3.3
## [85] polyclip_1.10-0     assertthat_0.2.1    xfun_0.10
## [88] ggforce_0.3.1       openxlsx_4.1.0.1    europepmc_0.3
## [91] tidygraph_1.1.2     viridisLite_0.3.0   tibble_2.1.3
## [94] rvcheck_0.1.6       iterators_1.0.12    memoise_1.1.0
## [97] ellipsis_0.3.0
```