Re-analysis of Panobinostat in-cell dataset by Becher et al, 2016

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(gplots)
##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##     lowess
```

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 excel table (Supplementary Dataset S1) by Becher et al. (2016)

```
if(!file.exists("41589_2016_BFnchembio2185_MOESM254_ESM.xlsx")){
download.file(
  url = "https://static-content.springer.com/esm/art%3A10.1038%2Fnchembio.2185/MediaObjects/41589_2016_BFnche
  destfile = "41589_2016_BFnchembio2185_MOESM254_ESM.xlsx")
}
```

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

```
pano_cell_raw <- read_xlsx("41589_2016_BFnchembio2185_MOESM254_ESM.xlsx",
                           sheet = 1, skip = 1) %>%
  dplyr::select(representative,
                clustername,
                experiment = ms_experiment,
                qupm,
                qusm,
                temperature,
                matches("sumionarea"),
                -matches("total"),
                matches("rel_fc_protein"),
                -matches("transformed"),
                -matches("orig"),
                -matches("log2rel"))  %>%
  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
pano_cell_fil <- resolveAmbiguousProteinNames(pano_cell_raw)

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

Compute null and alternative model fits and extract parameters

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

Compute F statistics

```
pano_fstat_df <- computeFStatFromParams(pano_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:

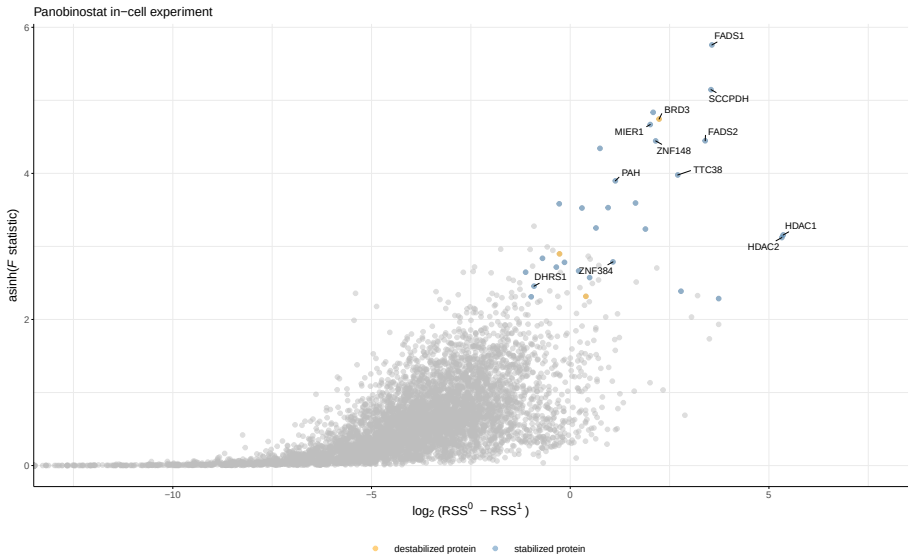
```
set.seed(12, kind = "L'Ecuyer-CMRG")
pano_null_df <- bootstrapNullAlternativeModel(
  df = pano_cell_df, params_df = pano_params_df,
  maxit = 500, B = 100,
  BPPARAM = BiocParallel::MulticoreParam(workers = 20, progressbar = TRUE),
  verbose = FALSE)
saveRDS(pano_null_df, file = "../pre_run_data/pano_null_df.rds")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

Compute FDR and find hits:

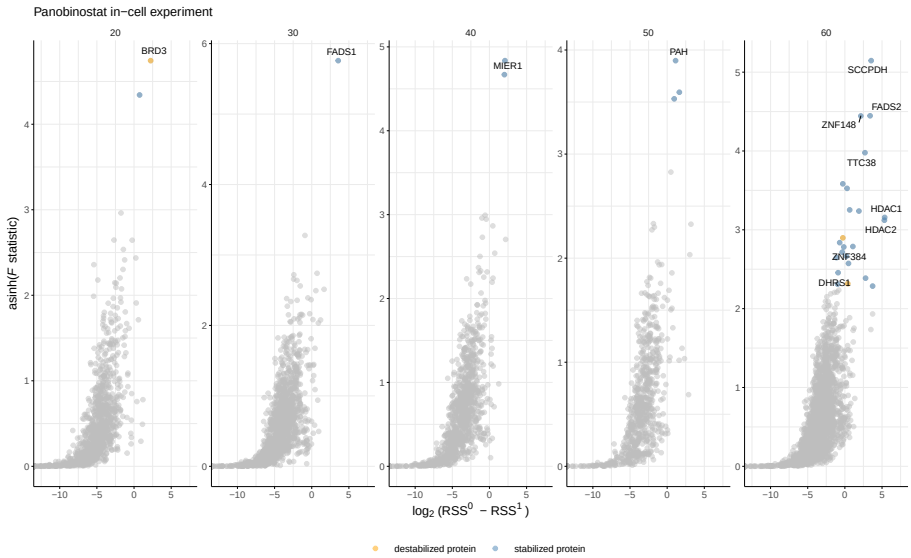
```
pano_fdr_df <- getFDR(df_out = pano_fstat_df,
                      df_null = pano_null_df)

pano_hits_df <- findHits(pano_fdr_df, alpha = 0.1)
```

```
ggplot(pano_fdr_df %>%
           filter(dataset == "true") %>%
           mutate(group = case_when(slopeH1 > 0 ~ "stabilized protein",
                                    slopeH1 < 0 ~ "destabilized protein")),
       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 = pano_hits_df %>%
                 mutate(group = case_when(
                     slopeH1 > 0 ~ "stabilized protein",
                     slopeH1 < 0 ~ "destabilized protein"))) +
  ggrepel::geom_text_repel(
    aes(label = clustername),
    data = filter(pano_hits_df, clustername %in%
                    c("HDAC1", "HDAC2",
                      "HDAC6", "PAH",
                      "TTC38", "FADS1",
                      "FADS2", "MIER1",
                      "BRD3", "SCCPDH",
                      "ZNF148", "DHRS1",
                      "ZNF384")),
    size = 2, segment.size = 0.2, min.segment.length = unit(2, "pt")) +
  scale_color_manual("", values = c("orange", "steelblue")) +
  labs(x = expression('log'[2]~'(RSS'^0~' - RSS'^1~')'),
       y = expression('asinh('*italic(F)*' statistic)')) +
  ggtitle("Panobinostat in-cell experiment") +
  coord_cartesian(xlim = c(-12.5, 7.5)) +
  theme_paper +
  theme(legend.position = "bottom")
```

Re-analysis of Panobinostat in-cell dataset by Becher et al, 2016

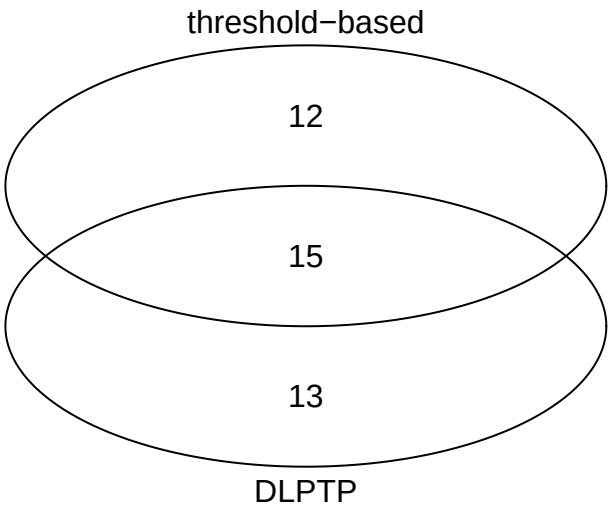



```
ggplot(pano_fdr_df %>%
           filter(dataset == "true") %>%
           mutate(group = case_when(slopeH1 > 0 ~ "stabilized protein",
                                    slopeH1 < 0 ~ "destabilized protein")),
       aes(log2(rssH0 - rssH1), asinh(F_statistic))) +
  geom_point(color = "gray", alpha = 0.5, size = 1) +
  geom_point(aes(color = group), alpha = 0.5,
             size = 1,
             data = pano_hits_df %>%
                 mutate(group = case_when(
                     slopeH1 > 0 ~ "stabilized protein",
                     slopeH1 < 0 ~ "destabilized protein"))) +
  ggrepel::geom_text_repel(
    aes(label = clustername),
    data = filter(pano_hits_df, clustername %in%
                    c("HDAC1", "HDAC2",
                      "HDAC6", "PAH",
                      "TTC38", "FADS1",
                      "FADS2", "MIER1",
                      "BRD3", "SCCPDH",
                      "ZNF148", "DHRS1",
                      "ZNF384")),
    size = 2, segment.size = 0.2, min.segment.length = unit(2, "pt")) +
  scale_color_manual("", values = c("orange", "steelblue")) +
  facet_wrap(~nObsRound, scales = "free", ncol = 5) +
  labs(x = expression('log'[2]~'(RSS'^0~' - RSS'^1~')'),
       y = expression('asinh('*italic(F)*' statistic)')) +
  ggtitle("Panobinostat in-cell experiment") +
  coord_cartesian(xlim = c(-12.5, 7.5)) +
  theme_paper +
  theme(legend.position = "bottom")
```

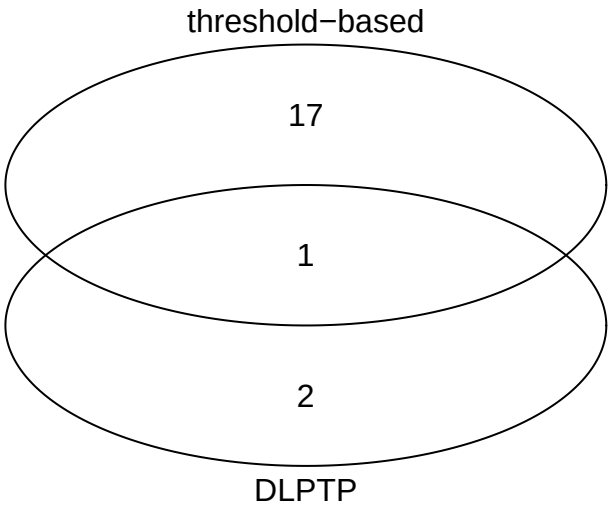


2 Compare to previous analysis

```
pano_thres_df <- read_xlsx("41589_2016_BFnchembio2185_MOESM254_ESM.xlsx",
                           sheet = 1, skip = 1) %>%
  filter(qupm > 1)

#stabilization
venn(list("DLPTP" = (pano_hits_df %>% filter(slopeH1 > 0))$clustername,
          "threshold-based" = (pano_thres_df %>% filter(protein_stabilized_neighb_temp_good_curves_count > 1
```



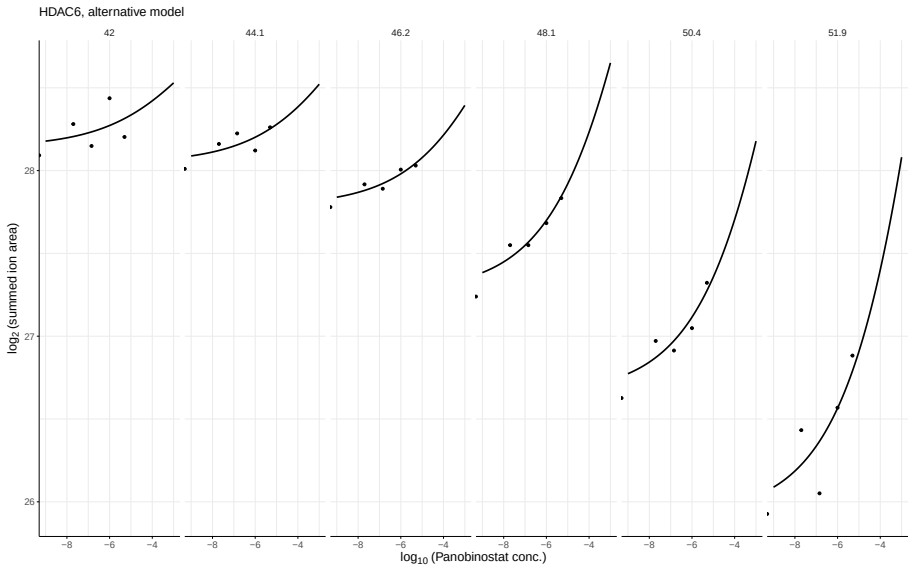

```
#destabilization
venn(list("DLPTP" = (pano_hits_df %>% filter(slopeH1 < 0))$clustername,
          "threshold-based" = (pano_thres_df %>% filter(protein_destabilized_neighb_temp_good_curves_count >
```



3 Plot example profiles

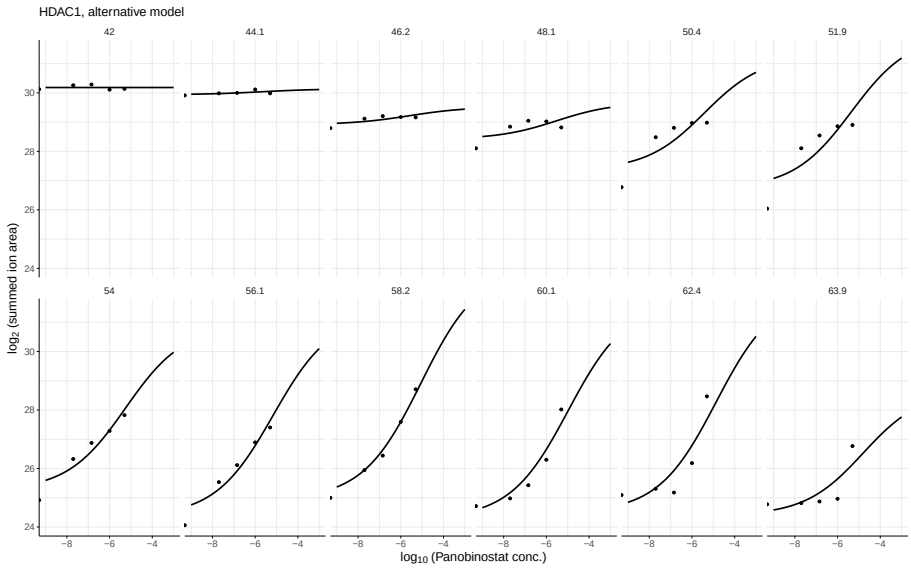
HDAC6

```
hdac6_fit <- plot2dTppFit(pano_cell_df, "HDAC6", "H1")$data

hdac6_df <- filter(pano_cell_df, clustername == "HDAC6")

ggplot(hdac6_fit, aes(log_conc, y_hat)) +
  geom_line() +
  geom_point(aes(log_conc, log2_value),
             data = hdac6_df, size = 0.5) +
  facet_wrap(~temperature, ncol = 6) +
  labs(x = expression('log'[10]~ '(Panobinostat conc.)'),
       y = expression('log'[2]~ '(summed ion area)')) +
  ggtitle("HDAC6, alternative model") +
  theme_paper
```



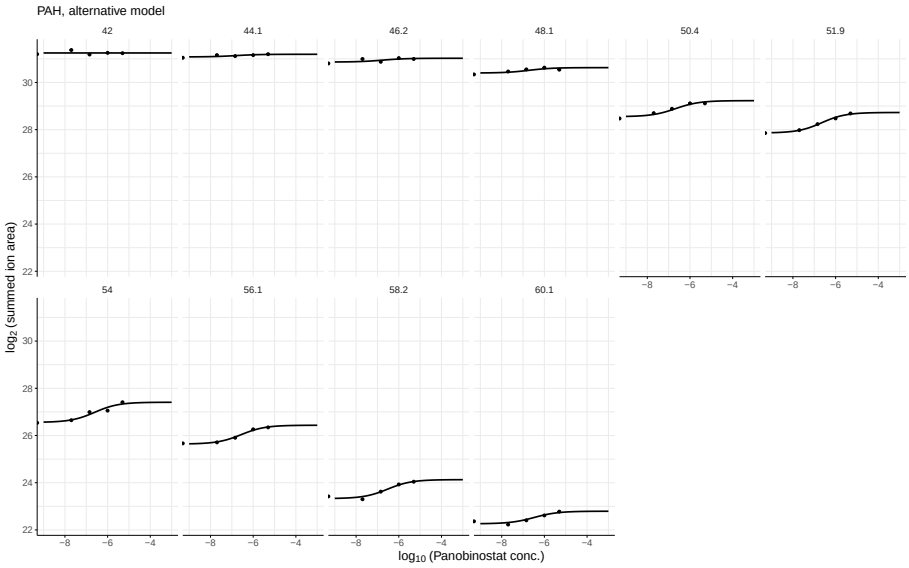
Other profiles for comparison:

```
hdac1_fit <- plot2dTppFit(pano_cell_df, "HDAC1", "H1")$data

hdac1_df <- filter(pano_cell_df, clustername == "HDAC1")

ggplot(hdac1_fit, aes(log_conc, y_hat)) +
  geom_line() +
  geom_point(aes(log_conc, log2_value),
             data = hdac1_df, size = 0.5) +
  facet_wrap(~temperature, ncol = 6) +
  labs(x = expression('log'[10]~ '(Panobinostat conc.)'),
       y = expression('log'[2]~ '(summed ion area)')) +
  ggtitle("HDAC1, alternative model") +
  theme_paper
```

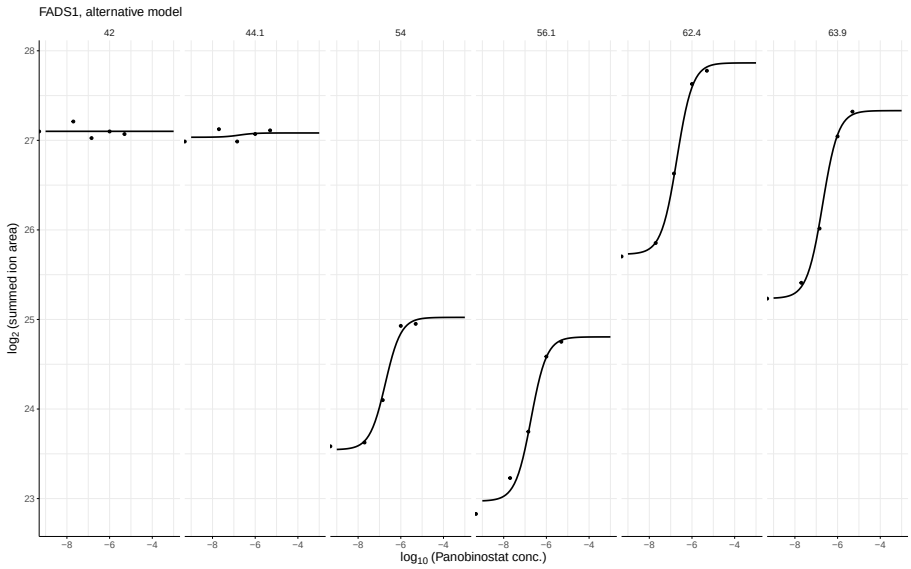


PAH

```
pah_fit <- plot2dTppFit(pano_cell_df, "PAH", "H1")$data

pah_df <- filter(pano_cell_df, clustername == "PAH")

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




```
fads1_fit <- plot2dTppFit(pano_cell_df, "FADS1", "H1")$data

fads1_df <- filter(pano_cell_df, clustername == "FADS1")

ggplot(fads1_fit, aes(log_conc, y_hat)) +
  geom_line() +
  geom_point(aes(log_conc, log2_value),
             data = fads1_df, size = 0.5) +
  facet_wrap(~temperature, ncol = 6) +
  labs(x = expression('log'[10]~ '(Panobinostat conc.)'),
       y = expression('log'[2]~ '(summed ion area)')) +
  ggtitle("FADS1, alternative model") +
  theme_paper
```




```
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] gplots_3.0.1.2   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] gtools_3.8.1        tidyselect_0.2.5    xfun_0.10
##  [4] purrr_0.3.3         colorspace_1.4-1    vctrs_0.2.0
##  [7] htmltools_0.4.0     yaml_2.2.0          rlang_0.4.1
## [10] pillar_1.4.2        glue_1.3.1          withr_2.1.2
## [13] BiocParallel_1.18.1 foreach_1.4.7       lifecycle_0.1.0
## [16] stringr_1.4.0       munsell_0.5.0       gtable_0.3.0
## [19] cellranger_1.1.0    zip_2.0.4           caTools_1.17.1.2
## [22] codetools_0.2-16    evaluate_0.14       labeling_0.3
## [25] knitr_1.25          doParallel_1.0.15   parallel_3.6.1
## [28] Rcpp_1.0.2          KernSmooth_2.23-16  scales_1.0.0
## [31] backports_1.1.5     BiocManager_1.30.9  gdata_2.18.0
## [34] digest_0.6.22       stringi_1.4.3       openxlsx_4.1.0.1
## [37] ggrepel_0.8.1       bookdown_0.14       grid_3.6.1
```

```
## [40] tools_3.6.1         bitops_1.0-6        magrittr_1.5
## [43] lazyeval_0.2.2      RCurl_1.95-4.12     tibble_2.1.3
## [46] crayon_1.3.4        pkgconfig_2.0.3     zeallot_0.1.0
## [49] MASS_7.3-51.4       ellipsis_0.3.0      assertthat_0.2.1
## [52] rmarkdown_1.16      iterators_1.0.12    R6_2.4.0
## [55] compiler_3.6.1
```

References

Becher, I., Werner, T., Doce, C., Zaal, E.A., Tögel, I., Khan, C.A., Rueger, A., Muelbaier, M., Salzer, E., Berkers, C.R., et al. (2016). Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat. Nature Chemical Biology *12*, 908–910.