Introduction to the *TPP* package for analyzing Thermal Proteome Profiling data: Temperature range (TR) or concentration compound range (CCR) experiments

Dorothee Childs, Nils Kurzawa
European Molecular Biology Laboratory (EMBL),
Heidelberg, Germany
dorothee.childs@embl.de

TPP version 3.0.3 (Last revision 2016-09-28)

Abstract

Detecting the binding partners of a drug is one of the biggest challenges in drug research. Thermal Proteome Profiling (TPP) addresses this question by combining the cellular thermal shift assay concept [**?**] with mass spectrometry based proteome-wide protein quantitation [**?**]. Thereby, drug-target interactions can be inferred from changes in the thermal stability of a protein upon drug binding, or upon downstream cellular regulatory events, in an unbiased manner.

The analysis of TPP experiments requires several data analytic and statistical modeling steps [**?**]. The package *TPP* facilitates this process by providing exectuable workflows that conduct all necessary steps. This vignette explains the use of the package. For details about the statistical methods, please refer to the papers [**?**, **?**].

Contents

1 Installation

To install the package, type the following commands into the *R* console

```
source("http://bioconductor.org/biocLite.R")
biocLite("TPP")
```

The installed package can be loaded by

```
library("TPP")
```

For the data manipulations in this vignette, we also load the *dplyr* and *magrittr* packages:

```
library("dplyr", quietly = TRUE)
library("magrittr", quietly = TRUE)
```

1.1 Special note for Windows users

The *TPP* package uses the *openxlsx* package to produce Excel output [**?**]. *openxlsx* requires a zip application to be installed on your system and to be included in the path. On Windows, such a zip application ist not installed by default, but is available, for example, via Rtools. Without the zip application, you can still use the 'TPP' package and access its results via the dataframes produced by the main functions.

1.2 TPP-TR and TPP-CCR analysis

The *TPP* package performs two analysis workflows:

1. **Analysis of temperature range (TR) experiments:** TPP-TR experiments combine the cellular thermal shift assay (CETSA) approach with high-throughput mass spectrometry (MS). They provide protein abundance measurements at increasing temperatures for different treatment conditions. The data analysis comprises cross-experiment normalization, melting curve fitting, and the statistical evaluation of the estimated melting points in order to detect shifts induced by drug binding.
2. **Analysis of compound concentration range (CCR) experiments:** TPP-CCR experiments combine the isothermal dose-response (ITDR) approach with high-throughput MS. The CCR workflow of the package performs median normalization, fits dose response curves, and determines the pEC50 values for proteins showing dose dependent changes in thermal stability upon drug treatment.

The following sections describe both functionalities in detail.

2 Analyzing TPP-TR experiments

2.1 Overview

The function `analyzeTPPTR` executes the whole workflow from data import through normalization and curve fitting to statistical analysis. Nevertheless, all of these steps can be invoked separately by the user. The corresponding functions can be recognized by their suffix `tpptr`. Here, we first show how to start the whole analysis using `analyzeTPPTR`. Afterwards, we demonstrate how to carry out single steps individually.

Before you can start your analysis, you need to specify information about your experiments:

- The mandatory information comprises a unique experiment name, as well as the isobaric labels and corresponding temperature values for each experiment.
- Optionally, you can also specify a condition for each experiment (treatment or vehicle), as well as an arbitrary number of comparisons. Comparisons are pairs of experiments whose melting points will be compared in order to detect significant shifts.

The package retrieves this information from a configuration table that you need to specify before starting the analysis. This table can either be a data frame that you define in your R session, or a spreadsheet in .xlsx or .csv format. In a similar manner, the measurements themselves can either be provided as a list of data frames, or imported directly from files during runtime.

We demonstrate the functionality of the package using the dataset `hdacTR_smallExample`. It contains an illustrative subset of a larger dataset which was obtained by TPP-TR experiments on K562 cells treated with the histone deacetylase (HDAC) inhibitor panobinostat in the treatment groups and with vehicle in the control groups. The experiments were performed for two conditions (vehicle and treatment), with two biological replicates each. The raw MS data were processed with the Python package *isobarQuant*, which provides protein fold changes relative to the protein abundance at the lowest temperature as input for the TPP package [**?**].

First, we load the data:

```
data("hdacTR_smallExample")
ls()
```

```
## [1] "TRresults"     "hdacTR_config" "hdacTR_data"   "resultPath"
```

This command loads two objects:

1. `hdacTR_data`: a list of data frames that contain the measurements to be analyzed,
2. `hdacTR_config`: a configuration table with details about each experiment.

2.2 The configuration table

`hdacTR_config` is an example of a configuration table in data frame format. We also provide a .xlsx version of this table. It is stored in the folder `example_data/TR_example_data` in your package installation path. You can locate

the example_data folder on your system by typing

```
system.file('example_data', package = 'TPP')
```

```
## [1] "/private/tmp/RtmpwI3GKc/Rinst10bf31d038e12/TPP/example_data"
```

You can use both versions as a template for your own analysis

Let's take a closer look at the content of the configuration table we just loaded:

```
print(hdacTR_config)
```

```
##       Experiment Condition ComparisonVT1 ComparisonVT2 126 127L 127H 128L 128H 129L 129H
## 1      Vehicle_1   Vehicle             x               67   63   59   56   53   50   47
## 2      Vehicle_2   Vehicle                           x  67   63   59   56   53   50   47
## 3 Panobinostat_1 Treatment             x               67   63   59   56   53   50   47
## 4 Panobinostat_2 Treatment                           x  67   63   59   56   53   50   47
##   130L 130H 131L
## 1   44   41   37
## 2   44   41   37
## 3   44   41   37
## 4   44   41   37
```

It contains the following columns:

- Experiment: name of each experiment.
- Condition: experimental conditions (Vehicle or Treatment).
- Comparisons: comparisons to be performed.
- Label columns: each isobaric label names a column that contains the temperature the label corresponds to in the individual experiments.

An additional Path column must be added to the table if the data should be imported from files instead of data frames.

2.3 The data tables

hdacTR_data is a list of data frames containing the measurements for each experimental condition and replicate:

```
summary(hdacTR_data)
```

```
##                Length Class      Mode
## Vehicle_1      13     data.frame list
## Vehicle_2      13     data.frame list
## Panobinostat_1 13     data.frame list
## Panobinostat_2 13     data.frame list
```

They contain between 508 and 509 proteins each:

```
data.frame(Proteins = sapply(hdacTR_data, nrow))
```

```
##                Proteins
## Vehicle_1           508
## Vehicle_2           509
## Panobinostat_1      508
## Panobinostat_2      509
```

Each of the four data frames in hdacTR_data stores protein measurements in a row wise manner. For illustration, let's look at some example rows of the first vehicle group.

```
hdacVehicle1 <- hdacTR_data[["Vehicle_1"]]
head(hdacVehicle1)
```

```
##      gene_name qssm qupm rel_fc_126 rel_fc_127L rel_fc_127H rel_fc_128L rel_fc_128H
## 3286     HDAC1    5    4 0.00510359   0.0207088   0.0512665   0.0840443    0.158568
## 3584    HDAC10    2    1 0.00000000   0.0180900   0.2511430   0.4034580    0.582994
```

```
## 3000     HDAC2    7    5 0.02006570   0.0589077   0.0718648   0.1011260    0.556456
## 2089     HDAC3    2    2 0.08706000   0.0891621   0.2103700   0.3226950    0.459124
## 1602     HDAC4    4    4 0.04371190   0.1069100   0.1630480   0.2411050    0.382980
## 607      HDAC6    5    4 0.00176507   0.0260307   0.0449839   0.0759111    0.202110
##      rel_fc_129L rel_fc_129H rel_fc_130L rel_fc_130H rel_fc_131L
## 3286    0.410777    0.622789    0.750158    0.866156           1
## 3584    0.631114    0.769128    1.013330    1.093940           1
## 3000    0.850373    0.842952    0.885415    0.972225           1
## 2089    0.651561    0.626848    0.785872    0.740518           1
## 1602    0.596979    0.831169    0.936279    0.955209           1
## 607     0.366981    0.638239    0.891903    0.932266           1
```

The columns can be grouped into three categories:

- a column with a protein identifier. Called `gene_name` in the current dataset,
- the ten fold change columns all start with the prefix `rel_fc_`, followed by the isobaric labels 126 to 131L,
- other columns that contain additional information. In the given example, the columns `qssm` and `qupm` were produced by the python package *isobarQuant* when analyzing the raw MS data. This metadata will be included in the package's output table. Additionally, it can be filtered according to pre-specified quality criteria for construction of the normalization set. The original results of the *isobarQuant* package contain more columns of this type. They are omitted here to keep the size of the example data within reasonable limits.

2.4 Starting the whole workflow by `analyzeTPPTR`

The default settings of the *TPP* package are configured to work with the output of the python package *isobarQuant*, but you can adjust it for your own data, if desired. When analyzing data from *isobarQuant*, all you need to provide is:

- the configuration table,
- the experimental data, either as a list of data frames, or as tab-delimited `.txt` files,
- a desired output location, for example

```
resultPath = file.path(getwd(), 'Panobinostat_Vignette_Example')
```

If you want to use data from other sources than *isobarQuant*, see section **??** for instructions.

If you import the data directly from tab-delimited `.txt` files, please make sure that the entries are not encapsulated by quotes (for example, `"entry1"` or `'entry2'`). All quotes will be ignored during data import in order to robustly handle entries containing single quotes (for example protein annotation by `5'` or `3'`).

By default, plots for the fitted melting curves are produced and stored in pdf format for each protein during runtime and we highly recommend that you do this when you analyze your data. However, producing plots for all 510 proteins in our dataset can be time consuming and would slow down the execution of the current example. Thus, we first disable plotting by setting the argument `plotCurves=FALSE`. Afterwards, we will produce plots for individual proteins of interest. Note that, in practice, you will only be able to examine the results in an unbiased manner if you allow the production of all plots.

We start the workflow by typing

```
TRresults <- analyzeTPPTR(configTable = hdacTR_config,
                          methods = "meltcurvefit",
                          data = hdacTR_data,
                          nCores = 2,
                          resultPath = resultPath,
                          plotCurves = FALSE)
```

This performs the melting curve fitting procedure in parallel on a maximum of two CPUs (requirement for package vignettes). Without specifying the `nCores` argument, fitting is performed by default on the maximum number of CPUs on your device.

`analyzeTPPTR` produces a table that summarizes the results for each protein. It is returned as a data frame and exported to an Excel spreadsheet at the specified output location. It contains the following information for each

experiment:

- normalized fold changes,
- melting curve parameters,
- statistical test results,
- quality checks on the curve parameters and p-values,
- additional columns from the original input data.

The quality of the result for each protein is determined by four filters. Currently, these criteria are checked only when the experimental setup includes exactly two replicates:

Filter	Column name in result table
1. Is the minimum slope in each of the control vs. treatment experiments < -0.06?	`minSlopes_less_than_0.06`
2. Are both the melting point differences in the control vs treatment experiments greater than the melting point difference between the two untreated controls?	`meltP_diffs_T_vs_V_greater_V1_vs_V2`
3. Is one of the p values for the two replicate experiments < 0.05 and the other one < 0.1?	`min_pVals_less_0.05_and_max_pVals_less_0.1`
4. Do the melting point shifts in the two control vs treatment experiments have the same sign (i.e. protein was either stabilized or destabilized in both cases)?	`meltP_diffs_have_same_sign`

The current example revealed 7 out of 510 proteins that fulfilled all four requirements:

```
tr_targets <- subset(TRresults, fulfills_all_4_requirements)$Protein_ID
print(tr_targets)
```

```
## [1] "BAG2"   "DDB2"   "HDAC10" "HDAC6"  "HDAC8"  "IQSEC2" "STX4"
```

3 of the detected proteins belong to the HDAC family. Because Panobinostat is known to act as an HDAC inhibitor, we select them for further investigation.

```
hdac_targets <- grep("HDAC", tr_targets, value=TRUE)
print(hdac_targets)
```

```
## [1] "HDAC10" "HDAC6"  "HDAC8"
```

We next investigate these proteins by estimating their melting curves for the different treatment conditions. However, we can only reproduce the same curves as before if the data is normalized by the same normalization procedure. Although we only want to fit and plot melting curves for a few proteins, the normalization therefore needs to incorporate all proteins in order to obtain the same normalizaton coefficients as before. The following section explains how to invoke these and other steps of the workflow independently of each other.

2.5 Starting individual steps of the workflow

2.5.1 Data import

Currently, the *TPP* package stores the data in *ExpressionSet*s, and so we convert the data that we have into the needed format. An advantage of the *ExpressionSet* container is its consistent and standardized handling of metadata for the rows and columns of the data matrix. This ability is useful for the given data, because it enables the annotation of each fold change column by temperature values as well as the corresponding isobaric labels. Furthermore, each protein can be annotated with several additional properties which can be used for normalization or processing of the package output.

The function `tpptrImport` imports the data and converts it into *ExpressionSets*:

```
trData <- tpptrImport(configTable = hdacTR_config, data = hdacTR_data)
```

```
## Importing data...
```

```
## Comparisons will be performed between the following experiments:
```

```
## Panobinostat_1_vs_Vehicle_1
```

```
## Panobinostat_2_vs_Vehicle_2

##

## The following label columns were detected:
## 126, 127L, 127H, 128L, 128H, 129L, 129H, 130L, 130H, 131L.

## Importing TR dataset:  Vehicle_1

## Removing duplicate identifiers using quality column 'qupm'...

## 508 out of 508 rows kept for further analysis.

##  -> Vehicle_1 contains 508 proteins.

##  -> 504 out of 508 proteins (99.21%) suitable for curve fit (criterion:  > 2 valid fold changes
per protein).

## Importing TR dataset:  Vehicle_2

## Removing duplicate identifiers using quality column 'qupm'...

## 509 out of 509 rows kept for further analysis.

##  -> Vehicle_2 contains 509 proteins.

##  -> 504 out of 509 proteins (99.02%) suitable for curve fit (criterion:  > 2 valid fold changes
per protein).

## Importing TR dataset:  Panobinostat_1

## Removing duplicate identifiers using quality column 'qupm'...

## 508 out of 508 rows kept for further analysis.

##  -> Panobinostat_1 contains 508 proteins.

##  -> 504 out of 508 proteins (99.21%) suitable for curve fit (criterion:  > 2 valid fold changes
per protein).

## Importing TR dataset:  Panobinostat_2

## Removing duplicate identifiers using quality column 'qupm'...

## 509 out of 509 rows kept for further analysis.

##  -> Panobinostat_2 contains 509 proteins.

##  -> 499 out of 509 proteins (98.04%) suitable for curve fit (criterion:  > 2 valid fold changes
per protein).

##
```

The resulting object `trData` is a list of *ExpressionSet*s for each experimental condition and replicate. Going back to the example data shown above (vehicle group 1), the corresponding object looks as follows:

```
trData[["Vehicle_1"]]

## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 508 features, 10 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: rel_fc_131L rel_fc_130H ... rel_fc_126 (10 total)
##   varLabels: label temperature normCoeff
##   varMetadata: labelDescription
## featureData
##   featureNames: AAK1 AAMDC ... ZFYVE20 (508 total)
##   fvarLabels: qssm qupm ... plot (12 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
```

```
## Annotation: Vehicle_1 Vehicle Panobinostat_1_vs_Vehicle_1 Panobinostat_2_vs_Vehicle_2
```

Each *ExpressionSet* S_i contains the fold change measurements (accessible by `exprs`(S_i)), column annotation for isobaric labels and temperatures (accessible by `phenoData`(S_i)), additional measurements obtained for each protein (accessible by `featureData`(S_i)), and the protein names (accessible by `featureNames`(S_i)).

2.5.2 Data normalization

Whether normalization needs to be performed and what method is best suited depends on the experiment. Currently, the *TPP* package offers the normalization procedure described by Savitski (2014)[**?**]. It comprises the following steps:

1. In each experiment, filter proteins according to predefined quality criteria.
2. Among the remaining proteins, identify those that were quantified in *all* experiments (jointP).
3. In each experiment, extract the proteins belonging to jointP. Subselect those proteins that pass the predefined fold change filters.
4. Select the biggest remaining set among all experiments (normP).
5. For each experiment, compute median fold changes over the proteins in normP and fit a sigmoidal melting curves through the medians.
6. Use the melting curve with the best R^2 value to normalize all proteins in each experiment.

The function `tpptrNormalize` performs all described steps. It requires a list of filtering criteria for construction of the normalization set. We distinguish between conditions on fold changes and on additional data columns. The function `tpptrDefaultNormReqs` offers an example object with default criteria for both categories:

```
print(tpptrDefaultNormReqs())
```

```
## $fcRequirements
##   fcColumn thresholdLower thresholdUpper
## 1        7            0.4            0.6
## 2        9            0.0            0.3
## 3       10            0.0            0.2
##
## $otherRequirements
##   colName thresholdLower thresholdUpper
## 1    qssm              4            Inf
```

By default, `tpptrNormalize` applies the filtering criteria in `tpptrDefaultNormReqs`. If you want to normalize a dataset in which the column indicating measurement quality has a different name than 'qssm', you have to change the column name and threshold accordingly. Because our example data was produced by *isobarQuant*, we can use the default settings here.

We normalize the imported data as follows:

```
normResults <- tpptrNormalize(data=trData)
```

```
## Creating normalization set:
## 1.  Filtering by non fold change columns:
## Filtering by annotation columns qssm in treatment group:  Vehicle_1
##  Column qssm between 4 and Inf-> 312 out of 508 proteins passed
## 312 out of 508 proteins passed in total.
## Filtering by annotation columns qssm in treatment group:  Vehicle_2
##  Column qssm between 4 and Inf-> 362 out of 509 proteins passed
## 362 out of 509 proteins passed in total.
## Filtering by annotation columns qssm in treatment group:  Panobinostat_1
##  Column qssm between 4 and Inf-> 333 out of 508 proteins passed
## 333 out of 508 proteins passed in total.
```

```
## Filtering by annotation columns qssm in treatment group:  Panobinostat_2
##  Column qssm between 4 and Inf-> 364 out of 509 proteins passed
## 364 out of 509 proteins passed in total.
## 2.  Find jointP:
## Detecting intersect between treatment groups (jointP).
## -> JointP contains 261 proteins.
## 3.  Filtering fold changes:
## Filtering fold changes in treatment group:  Vehicle_1
##  Column 7 between 0.4 and 0.6 -> 30 out of 261 proteins passed
##  Column 9 between 0 and 0.3 -> 223 out of 261 proteins passed
##  Column 10 between 0 and 0.2 -> 233 out of 261 proteins passed
## 22 out of 261 proteins passed in total.
## Filtering fold changes in treatment group:  Vehicle_2
##  Column 7 between 0.4 and 0.6 -> 21 out of 261 proteins passed
##  Column 9 between 0 and 0.3 -> 215 out of 261 proteins passed
##  Column 10 between 0 and 0.2 -> 227 out of 261 proteins passed
## 14 out of 261 proteins passed in total.
## Filtering fold changes in treatment group:  Panobinostat_1
##  Column 7 between 0.4 and 0.6 -> 34 out of 261 proteins passed
##  Column 9 between 0 and 0.3 -> 217 out of 261 proteins passed
##  Column 10 between 0 and 0.2 -> 224 out of 261 proteins passed
## 21 out of 261 proteins passed in total.
## Filtering fold changes in treatment group:  Panobinostat_2
##  Column 7 between 0.4 and 0.6 -> 15 out of 261 proteins passed
##  Column 9 between 0 and 0.3 -> 221 out of 261 proteins passed
##  Column 10 between 0 and 0.2 -> 225 out of 261 proteins passed
## 10 out of 261 proteins passed in total.
## Experiment with most remaining proteins after filtering:  Vehicle_1
## -> NormP contains 22 proteins.
## ------------------------------------
## Computing normalization coefficients:
## 1.  Computing fold change medians for proteins in normP.
## 2.  Fitting melting curves to medians.
## -> Experiment with best model fit:  Vehicle_1 (R2:  0.9919)
## 3.  Computing normalization coefficients
## Creating QC plots to illustrate median curve fits.
## ------------------------------------
## Normalizing all proteins in all experiments.
## Normalization successfully completed!
```

```
trDataNormalized <- normResults[["normData"]]
```

2.5.3 Melting curve fitting

Next we fit and plot melting curves for the detected HDAC targets. We first select the corresponding rows from the imported data. The data are stored as `expressionSet` objects from the *Biobase* package. It provides a range of functions to access and manipulate texttexpressionSet objects. Examples are the functions `featureNames`, `pData`, or `featureData`. In order to use them outside of the TPP package namespace, we first import the *Biobase* package:

```
library(Biobase, quietly = TRUE)
trDataHDAC <- lapply(trDataNormalized, function(d)
  d[featureNames(d) %in% hdac_targets,])
```

We fit melting curves for these proteins using the function `tpptrCurveFit`:

```
trDataHDAC <- tpptrCurveFit(data = trDataHDAC, resultPath = resultPath, nCores = 1)
```

```
## Fitting melting curves to 3 proteins.
```

```
## Runtime (1 CPUs used):  1.49 secs
```

```
## Melting curves fitted sucessfully!
```

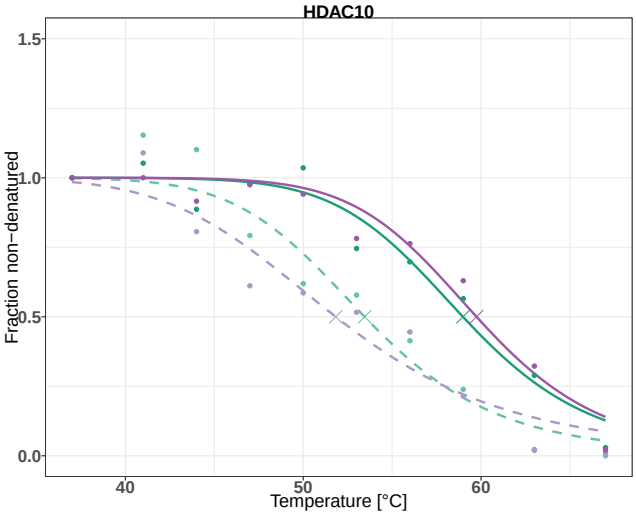
```
## 12 out of 12 models with sufficient data points converged (100 %).
```

The melting curve parameters are now stored within the `featureData` of the *ExpressionSets*. For example, the melting curves estimated for the Vehicle group have the following parameters:

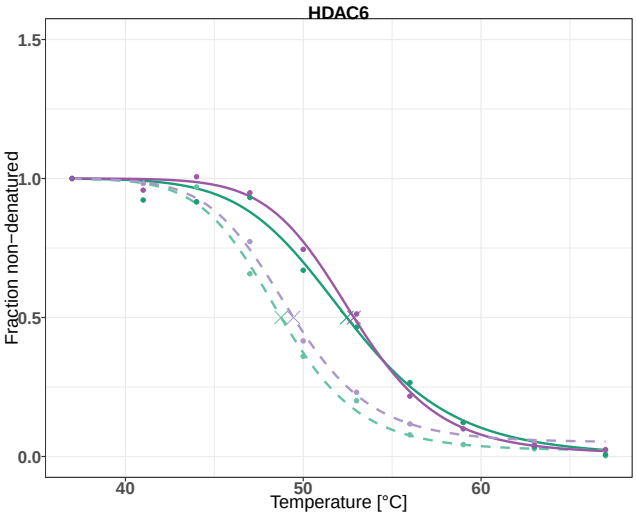
```
pData(Biobase::featureData(trDataHDAC[["Vehicle_1"]]))[,1:5]
```

```
##        qssm qupm         a        b meltPoint
## HDAC10    2    1  757.5239 14.17157  53.45376
## HDAC6     5    4 1049.4633 21.56724  48.75631
## HDAC8     2    1  779.6680 16.77087  46.48942
```

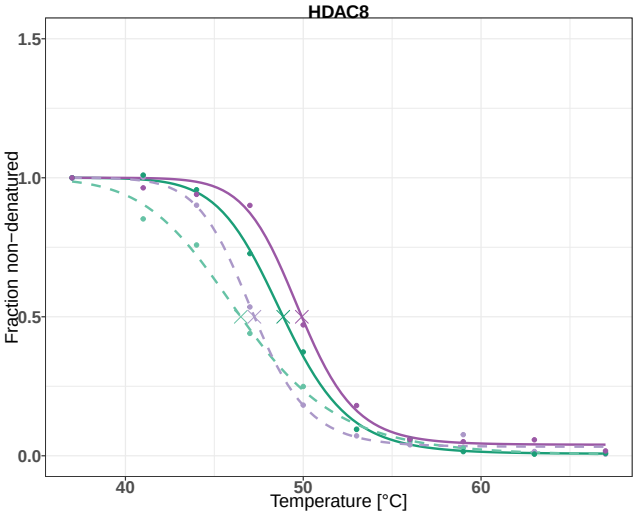
The melting curve plots were stored in subdirectory `Melting_Curves` in `resultPath`. You can browse this directory and inspect the melting curves and their parameters. In the following, you can see the plots that were placed in this directory for the 3 detected targets:



	meltPoint	slope	plateau	R2
Panobinostat_1	58.98	−0.069	0	0.95
Vehicle_1	53.45	−0.068	0	0.95
Panobinostat_2	59.75	−0.071	0	0.96
Vehicle_2	51.83	−0.052	0	0.93



	meltPoint	slope	plateau	R2
Panobinostat_1	52.45	−0.083	0	0.99
Vehicle_1	48.76	−0.11	0.02	1
Panobinostat_2	52.85	−0.1	0.01	1
Vehicle_2	49.46	−0.1	0.05	1



	meltPoint	slope	plateau	R2
Panobinostat_1	48.87	−0.13	0.01	1
Vehicle_1	46.49	−0.091	0	1
Panobinostat_2	49.92	−0.15	0.04	1
Vehicle_2	47.25	−0.15	0.03	1

2.5.4 Significance assessment of melting point shifts

Similar to the normalization explained earlier, significance assessment of melting point shifts has to be performed on the whole dataset due to the binning procedure used for p-value computation. For the given dataset, we have already analyzed all curve parameters by the function `analyzeTPPTR`. Here we show how you can start this procedure independently of the other steps. This can be useful when you only need to re-compute the p-values (for example with a different binning parameter) without the runtime intense curve fitting before.

Melting curve parameter analysis is performed by the function `tpptrAnalyzeMeltingCurves`. It requires a list of *ExpressionSet*s with melting curve parameters stored in the featureData. To avoid runtime intensive repetitions of the curve fitting procedure, `analyzeTPPTR` saved these objects as an intermediate result after curve fitting in the subdirectory `/dataObj`. We can access them by the command:

```
load(file.path(resultPath, "dataObj", "fittedData.RData"), verbose=TRUE)

## Loading objects:
##   trDataFitted
```

This loaded the object `trDataFitted`, which is a list of *ExpressionSet*s in which the melting curve parameters have already been stored in the `featureData` by `tpptrCurveFit`.

Now we start the curve parameter evaluation, this time applying less rigorous filters on the quality of the fitted curves (defined by the minimum R^2), and their long-term melting behavior (constrained by the maximum plateu paramter):

```
minR2New <- 0.5 # instead of 0.8
maxPlateauNew <- 0.7 # instead of 0.3
newFilters <- list(minR2 = minR2New,
                   maxPlateau = maxPlateauNew)
TRresultsNew <- tpptrAnalyzeMeltingCurves(data = trDataFitted,
                                          pValFilter = newFilters)

## Starting melting curve analysis.

## Computing p-values for comparison Panobinostat_1_vs_Vehicle_1 ...
```

```
## Performing quality check on the melting curves of both experiments.
## 1.  R2 > 0.5 (both Experiment):  471 out of 510 models passed.
## 2.  Pl < 0.7 (Vehicle group only):  482 out of 510 models passed.
## => 468 out of 510 models passed in total and will be used for p-value computation.
## Computing p-values for comparison Panobinostat_2_vs_Vehicle_2 ...
## Performing quality check on the melting curves of both experiments.
## 1.  R2 > 0.5 (both Experiment):  477 out of 510 models passed.
## 2.  Pl < 0.7 (Vehicle group only):  487 out of 510 models passed.
## => 475 out of 510 models passed in total and will be used for p-value computation.
## Results table created successfully.
```

We can then compare the outcome to the results that we previously obtained with `minR2 = 0.8` and `maxPlateau = 0.3`:

```
tr_targetsNew <- subset(TRresultsNew, fulfills_all_4_requirements)$Protein_ID
targetsGained <- setdiff(tr_targetsNew, tr_targets)
targetsLost <- setdiff(tr_targets, tr_targetsNew)
print(targetsGained)
## [1] "HDAC1" "HDAC2" "KIF22"
print(targetsLost)
## [1] "BAG2"
```

We observe that relaxing the filters before p-value calculation leads to the detection of 3 new target proteins, while 1 of the previous targets is now omitted. This illustrates the trade-off we are facing when defining the filters: on the one hand, relaxing the filters enables more proteins, and hence, more true targets, to be considered for testing. On the other hand, inlcuding poor fits increases the variability in the data and reduces the power during hypothesis testing. This observation motivated the implementation of an alternative spline-based approach for fitting and testing ("non-parametric analysis of response curves", (NPARC)), which was newly introduced in version 3.0.0 of the `TPP` package. For more information, please check the vignette `NPARC_analysis_of_TPP_TR_data`:

```
browseVignettes("TPP")
```

2.5.5 Output table

Finally, we export the new results to an Excel spreadsheet: Because this process can be time consuming for a large dataset, we only export the detected potential targets:

```
tppExport(tab = TRresultsNew,
          file = file.path(resultPath, "targets_newFilters.xlsx"))
```

2.6 Analyzing data not produced by the accompanying isobarQuant package

2.6.1 Specifying customized column names for data import

By default, `analyzeTPPTR` looks for a protein ID column named `gene_name`, and a quality control column named `qupm` to assist in the decision between proteins with the same identifier. If these columns have different names in your own dataset, you have to define the new names using the arguments `idVar` and `qualColName`. Similarly, the argument `fcStr` has to be set to the new prefix of the fold change columns.

2.6.2 Specifying customized filtering criteria for normalization

You can set the filtering criteria for normalization set construction by modifying the supplied default settings. Remember to adjust the fold change column numbers in case you have more/ less than ten fold changes per experiment.

```
trNewReqs <- tpptrDefaultNormReqs()
print(trNewReqs)
```

```
## $fcRequirements
##   fcColumn thresholdLower thresholdUpper
## 1        7            0.4            0.6
## 2        9            0.0            0.3
## 3       10            0.0            0.2
##
## $otherRequirements
##   colName thresholdLower thresholdUpper
## 1    qssm              4            Inf
```

```
trNewReqs$otherRequirements[1,"colName"] <- "mycolName"
trNewReqs$fcRequirements[,"fcColumn"] <- c(6,8,9)
print(trNewReqs)
```

```
## $fcRequirements
##   fcColumn thresholdLower thresholdUpper
## 1        6            0.4            0.6
## 2        8            0.0            0.3
## 3        9            0.0            0.2
##
## $otherRequirements
##     colName thresholdLower thresholdUpper
## 1 mycolName              4            Inf
```

2.7 Specifying the experiments to compare

You can specify an arbitrary number of comparisons in the configuration table. For each comparison, you add a separate column. The column name needs to start with the prefix 'Comparison'. The column needs to contain exactly two alpha-numerical characters (in our example, we used 'x').

If conditions are specified in the 'Condition' column, comparisons between melting points will always be performed in the direction $Tm_{Treatment} - Tm_{Vehicle}$.

3 Analyzing TPP-CCR experiments

First, we load the data:

```
data("hdacCCR_smallExample")
```

This command loads two objects: the configuration tables for two replicates (`hdacCCR_config_repl1/2`) and two data frames that contain the measurements of both TPP-CCR experiments to be analyzed (`hdacCCR_data_repl1/2`).

3.1 Starting the whole workflow by `analyzeTPPCCR`

We start the workflow for replicate 1 by typing

```
CCRresults <- analyzeTPPCCR(configTable = hdacCCR_config[1,],
                            data = hdacCCR_data[[1]],
                            resultPath = resultPath,
```

```
                                plotCurves = FALSE,
                                nCores = 2)
```

The following proteins passed the criteria of displaying a clear response to the treatment, and enabling curve fitting with $R^2 > 0.8$:

```
ccr_targets <- subset(CCRresults, passed_filter_Panobinostat_1)$Protein_ID
print(ccr_targets)
```

```
##  [1] "ALKBH1" "CHMP5"  "ECH1"   "HDAC1"  "HDAC10" "HDAC2"  "HDAC6"  "HSPB11" "TTC38"
## [10] "ZNF384"
```

4 of the selected proteins belong to the HDAC family. Because Panobinostat is known to act as an HDAC inhibitor, we select them for further investigation.

```
hdac_targets <- grep("HDAC", ccr_targets, value = TRUE)
print(hdac_targets)
```

```
## [1] "HDAC1"  "HDAC10" "HDAC2"  "HDAC6"
```

The following section explains how to invoke the individual steps of the workflow separately.

3.2 Starting individual steps of the workflow

3.2.1 Data import

The function `tppccrImport` imports the data and converts it into an *ExpressionSet*:

```
ccrData <- tppccrImport(configTable = hdacCCR_config[1,], data = hdacCCR_data[[1]])
```

```
## Importing data...
## The following label columns were detected:
## 126, 127L, 127H, 128L, 128H, 129L, 129H, 130L, 130H, 131L.
## Importing CCR dataset:  Panobinostat_1
## Removing duplicate identifiers using quality column 'qupm'...
## 507 out of 507 rows kept for further analysis.
##  -> Panobinostat_1 contains 507 proteins.
##  -> 494 out of 507 proteins (97.44%) suitable for curve fit (criterion:  > 2 valid fold changes
per protein).
##
## Filtering CCR dataset:  Panobinostat_1
## Removed proteins with zero values in column(s) 'qssm':
## 494 out of 507 proteins remaining.
```

3.2.2 Data normalization

Currently, the *TPP* package offers normalization by fold change medians for TPP-CCR experiments. We normalize the imported data by

```
ccrDataNormalized <- tppccrNormalize(data = ccrData)
```

```
## Normalizing dataset:  Panobinostat_1
## Normalization complete.
```

3.2.3 Data transformation

We next have to specify the type of response for each protein, and transform the data accordingly:

```
ccrDataTransformed <- tppccrTransform(data = ccrDataNormalized)[[1]]
```

```
## Transforming dataset:  Panobinostat_1
```

```
## Transformation complete.
```

3.2.4 Dose response curve fitting

Next we fit and plot dose response curves for the detected HDAC targets. We first select the corresponding rows from the imported data:

```
ccrDataHDAC <- ccrDataTransformed[match(hdac_targets, featureNames(ccrDataTransformed)),]
```

We fit dose response curves for these proteins using the function `tppccrCurveFit`:

```
ccrDataFittedHDAC <- tppccrCurveFit(data=list(Panobinostat_1 = ccrDataHDAC), nCores = 1)
```

```
## Fitting 4 individual dose response curves to 4 proteins.
```

```
## Runtime (1 CPUs used):  0.12 secs
```

```
## Dose response curves fitted sucessfully!
```

```
## 4 out of 4 models with sufficient data points converged (100 %).
```

```
tppccrPlotCurves(ccrDataFittedHDAC, resultPath = resultPath, nCores = 1)
```

```
## Plotting dose response curves for 4 proteins.
```

```
## Runtime (1 CPUs used):  1.3 secs
```

```
## Dose response curves plotted sucessfully!
```

```
## $Panobinostat_1
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 4 features, 10 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: rel_fc_131L rel_fc_130H ... rel_fc_126 (10 total)
##   varLabels: label concentration normCoeff
##   varMetadata: labelDescription
## featureData
##   featureNames: HDAC1 HDAC10 HDAC2 HDAC6
##   fvarLabels: qssm qupm ... plot (53 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation: Panobinostat_1 NA
```

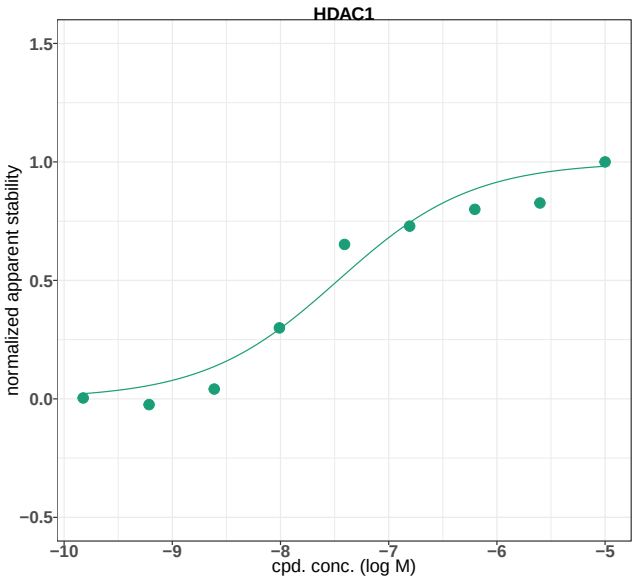
This function produces a table that contains the dose response curve parameters and additional information about each protein:

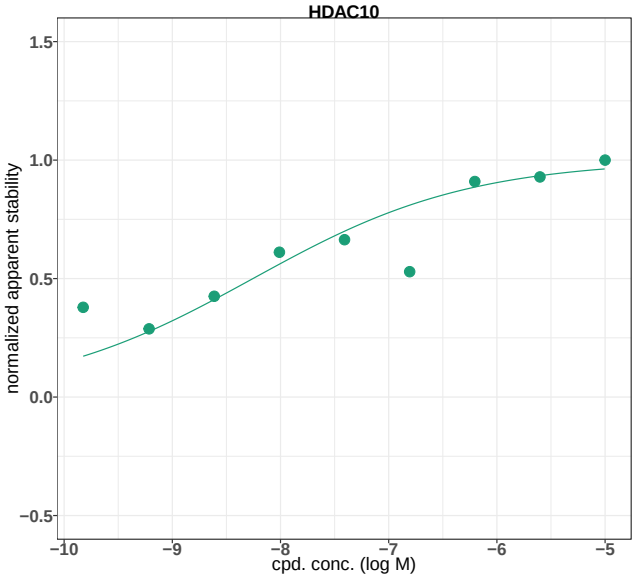
```
ccrResultsHDAC <- tppccrResultTable(ccrDataFittedHDAC)
print(ccrResultsHDAC[,c(1, 22:25)])
```

```
##   Protein_ID rel_fc_131L_transformed_Panobinostat_1
## 1      HDAC1                                     0
## 2     HDAC10                                     0
## 3      HDAC2                                     0
## 4      HDAC6                                     0
##   rel_fc_130H_transformed_Panobinostat_1 rel_fc_130L_transformed_Panobinostat_1
## 1                            0.003808506                            -0.02430098
```

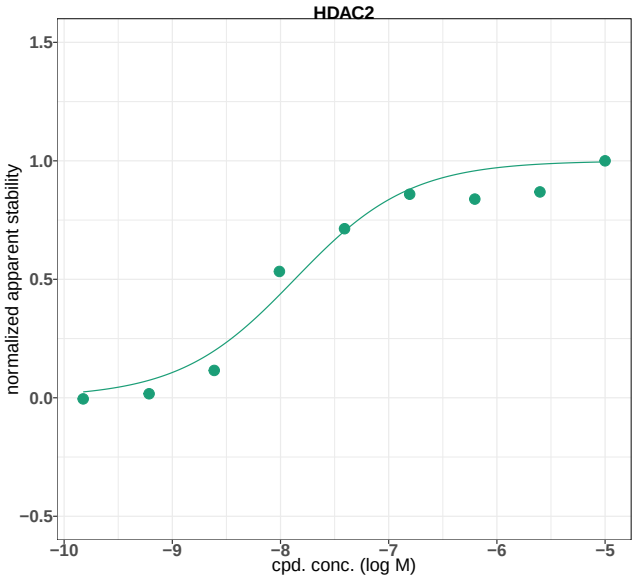
```
## 2                            0.378751669                            0.28785273
## 3                           -0.004674148                            0.01692643
## 4                            0.015491047                           -0.08185915
##   rel_fc_129H_transformed_Panobinostat_1
## 1                            0.04143994
## 2                            0.42538135
## 3                            0.11541646
## 4                            0.04849781
```

The dose response curve plots were stored in subdirectory `DoseResponse_Curves` in `resultPath`. You can browse this directory and inspect the fits and melting curve parameters. In the following, you can see the plot that were placed in this directory for the 4 detected targets:

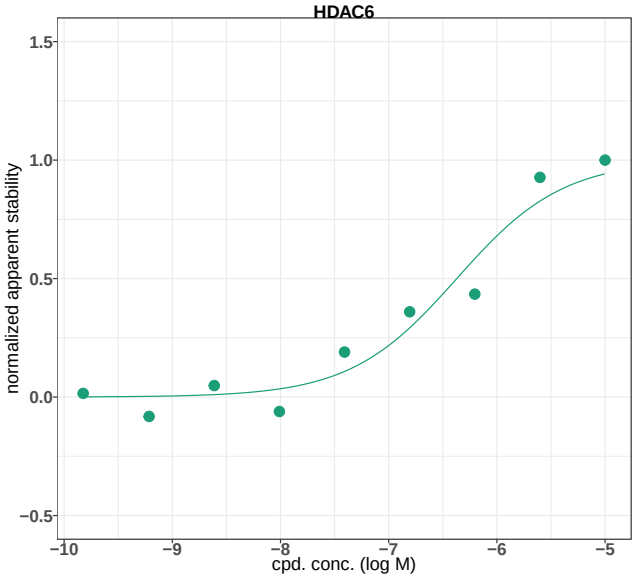




	pEC50	slope	R2
Panobinostat_1	8.25	1	0.859



	pEC50	slope	R2
Panobinostat_1	7.87	1.88	0.968



	pEC50	slope	R2
Panobinostat_1	6.37	2.03	0.954