Non-parametric analysis of thermal proteome profiles

Dorothee Childsh

08 August 2018, 19:09:27

Contents

1 Introduction

This workflow shows how to reproduce the analysis described by Childs, Bach, Franken et al. (2018): Non-parametric analysis of thermal proteome profiles reveals novel drug-binding proteins.

2 Preparation

Load necessary packages:

```
library(tidyverse)
```

This workflow shows how to reproduce the analysis described by Childs, Bach, Franken et al. (2018): Non-parametric analysis of thermal proteome profiles reveals novel drug-binding proteins.

3 Data import

First we load the data from the different TPP experiments. All data have been downloaded from the supplements of the respective publications (M. M. Savitski et al. 2014, Franken et al. (2015), Reinhard et al. (2015)) converted into tidy format, and concatenated into one table. This table will be available as supplementary material to the paper in the future.

```
tppData <- readRDS("tppData.Rds")
```

Let's take a look at the imported data:

```
tppData %>%
  mutate(molarDrugConcentration = factor(molarDrugConcentration),
         replicate = factor(replicate),
         dataset = factor(dataset)) %>%
  summary %>%
  knitr::kable()
```

	dataset	uniqueID	relAbundance	temperature	molarDrugConcentration	replicate	uni
	ATP :268000	Length:1432280	Min. : 0.0	Min. :25.00	0 :716140	1:716140	Min
	Dasatinib 0.5:308520	Class :character	1st Qu.: 0.1	1st Qu.:44.00	5e-07:154260	2:716140	1st
	Dasatinib 5 :308520	Mode :character	Median : 0.6	Median :52.50	1e-06:120080		Med
	Panobinostat :240160		Mean : 0.6	Mean :51.86	5e-06:154260		Mea
	Staurosporine:307080		3rd Qu.: 1.0	3rd Qu.:59.00	2e-05:153540		3rd
			Max. :577.6	Max. :67.00	0.002:134000		Max
			NA's :372809				NA'

4 Data preprocessing

First, we remove all decoy proteins remaining in the panobinostat data. They can be recognized by the prefix `###`, which was assigned by the quantification software `isobarQuant`.

```
tppData <- tppData %>% filter(!grepl("###[[:alnum:]]*###", uniqueID))
```

Next, we remove all proteins that were not found with at least one unique peptide

```
tppData <- filter(tppData, uniquePeptideMatches >= 1)
```

Next, we remove all proteins that only contain missing values

```
tppData <- tppData %>% filter(!is.na(relAbundance))
```

Only leave proteins reproducibly observed with full melting curves in both replicates and treatment groups per dataset. A full melting curve is defined by the presence of measurements at all 10 temperatures for the protein.

```
tppData <- tppData %>%
  group_by(dataset, uniqueID) %>%
  mutate(n = n()) %>%
  group_by(dataset) %>%
  mutate(max_n = max(n)) %>%
  filter(n == max_n)
```

Count the numbers of proteins remaining in each dataset. They coincide with the values reported in Table 1 of the paper.

```
tppData %>%
  distinct(dataset, uniqueID) %>%
  distinct %>%
  group_by(dataset) %>%
  tally %>%
  knitr::kable()
```

dataset	n
ATP	4177
Dasatinib 0.5	4625
Dasatinib 5	4154
Panobinostat	3649
Staurosporine	4505

Bibliography

Franken, Holger, Toby Mathieson, Dorothee Childs, Gavain M A Sweetman, Thilo Werner, Ina Tögel, Carola Doce, et al. 2015. "Thermal Proteome Profiling for Unbiased Identification of Direct and Indirect Drug Targets Using Multiplexed Quantitative Mass Spectrometry." *Nat. Protoc.* 10 (10): 1567–93.

Reinhard, Friedrich B M, Dirk Eberhard, Thilo Werner, Holger Franken, Dorothee Childs, Carola Doce, Maria Fälth Savitski, et al. 2015. "Thermal Proteome Profiling Monitors Ligand Interactions with Cellular Membrane Proteins." *Nat. Methods* 12 (12): 1129–31.

Savitski, Mikhail M, Friedrich B M Reinhard, Holger Franken, Thilo Werner, Maria Fälth Savitski, Dirk Eberhard, Daniel Martinez Molina, et al. 2014. "Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome." *Science* 346 (6205): 1255784.