Package 'ProData'

November 10, 2026

Title SELDI-TOF data of Breast cancer samples

Version 1.50.0

Author Xiaochun Li

Description A data package of SELDI-TOF protein mass spectrometry data of 167 breast cancer and normal samples.

Maintainer Xiaochun Li <xiaochun@jimmy.harvard.edu>

Depends R (>= 2.4.0), Biobase (>= 2.5.5)

License GPL

biocViews ExperimentData, CancerData, BreastCancerData, MassSpectrometryData, NCI

git_url https://git.bioconductor.org/packages/ProData

git_branch RELEASE_3_23

git_last_commit 40150ca

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Contents

f45cbmk 2
ProData 3

Index **4**

f45cbmk *A SELDI-TOF data of 167 breast cancer and normal samples from the Miron Laboratory*

Description

The data consist of 167 SELDI-TOF spectra from either individuals with breast cancer or without (normal controls). Two covariates are available. The data have been pre-processed by Ciphergen's software and the specific details are as follows. The baselines of the spectra were subtracted using the default parameter settings of the Ciphergen software, and the spectra were normalized by the total ion current normalization function of the Ciphergen software. Peaks were detected with signal to noise ratio (S/N) above 10. Proto-biomarkers, i.e., peaks in each spectrum that have the same M/Z value, were detected with the following parameters: first pass, 10 S/N, min peak threshold 10 pass, 5 S/N, and add estimated peaks to complete clusters and it is the jointly normalized data that are available here. The data are presented in the form of an `exprSet` object.

Usage

```
data(f45cbmk)
```

Format

The different covariates are:

GROUP a character vector with four possible values, "A", "B", "C" and "D". "A" represents HER2 positive patients, "B" normal healthy patients, "C" ER/PR positive patients and "D" samples from a single healthy woman.

SPECTRUM a character vector, the spectrum id of each sample.

Source

Alex Miron's Lab at the Dana-Farber Cancer Institute.

References

Shi, Q. et al. Biomarker Discovery in Plasma Using Proteomics Analysis for Early Detection of Breast Cancer. 2004, manuscript.

Examples

```
data(f45cbmk)
```

ProData *A SELDI-TOF data of 167 breast cancer and normal spectra*

Description

This library contains both the raw spectra and the proto-biomarker data pre-processed by Ciphergen's software of the 167 samples. Those one hundred and sixty seven samples are collected from 155 subjects in CPT tubes with plasma isolated and stored in -80C until needed. Among the 167 samples, 55 are HER2 positive (A), 64 are normal healthy women (B), 35 are ER/PR positive (mostly) (C) and 13 samples are from a single healthy woman. Samples in group D are the only ones from a single subject, all the other samples represent draws from individual subjects. Samples were thawed and aliquoted into 100ul vials. The samples were fractionated to simplify the proteome into sub-proteomes. The fractions 4 and 5 (f45) were processed by the Ciphergen IMAC protocol with EAM of CHCA.

Information on the spectrum ID and the pheotype information is stored in the `exprSet` object "f45cbmk", the proto-biomakers pre-processed by Ciphergen's software.

As an alternative, package `PROcess` may be used for pre-processing of the raw spectra to get a set of proto-biomarkers.

Source

Alex Miron's Lab at the Dana-Farber Cancer Institute.

Examples

```
# plot a raw spectrum
f45c <- system.file("f45c", package="ProData")
fs <- dir(f45c,full.names=TRUE)
plot(read.csv(gzfile(fs[1])), type="l")

# find out sizes of phenotype groups:
library(Biobase)
data(f45cbmk)
SpecGrp <- pData(f45cbmk)
table(SpecGrp[,1])
```

Index

∗ **datasets**
f45cbmk, 2
ProData, 3

f45cbmk, 2

ProData, 3