Package 'GEOsubmission'

November 14, 2025

Version 1.63.0

Date 2010-04-19

Title Prepares microarray data for submission to GEO

Author Alexandre Kuhn <alexandre.m.kuhn@gmail.com>

Maintainer Alexandre Kuhn <alexandre.m.kuhn@gmail.com>

Imports affy, Biobase, utils

Description Helps to easily submit a microarray dataset and the associated sample information to GEO by preparing a single file for upload (direct deposit).

biocViews Microarray

License GPL (>= 2)

ZipData no

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

git_branch devel

git_last_commit 6bd8c28

git_last_commit_date 2025-10-29

Repository Bioconductor 3.23

Date/Publication 2025-11-13

Contents

expressionNormalized 2
microarray2soft 2
sample1 5
sample2 5
sampleInfo 5
seriesInfo 5

Index **6**

expressionNormalized *expressionNormalized.txt*

Description

Example file with normalized expression values.

Format

tab-delimited text file

microarray2soft *Prepares microarray data for submission to GEO*

Description

Converts a microarray dataset and the corresponding sample information into a SOFT file to be used for GEO submission.

Usage

```
microarray2soft(samplenames, sampleinfo, seriesnames, seriesinfo,
                datadir=NULL, infodir=NULL, writedir=NULL, softname=NULL,
                expressionmatrix=NULL, verbose=TRUE)
```

Arguments

samplenames	character vector. Sample names.
sampleinfo	character string. Name of a tab-delimited file with sample information. Each line in the file corresponds to a sample. Sample names are stored under column header SAMPLE. Other column headers will be looked up for a match against the sample labels required by the SOFT format.
seriesnames	charcter vector. Series name. Works only with 1 series at present time.
seriesinfo	character string. Name of a tab-delimited file with series information. Only 1 line (ie. 1 series) allowed at present time (in addition to the header). Series names are stored under column header SERIES. Other column headers will be looked up for a match against the series labels required by the SOFT format.
datadir	character string. Name of the directory (with associated path) containing the microarray data files. Defaults to the working directory.
infodir	character string. (Full path to) Directory containing the text file with sample info (sampleinfo) and series info (seriesinfo). Defaults to datadir.
writedir	character string. (Full path to) Directory to write the SOFT file to. Defaults to datadir.

`softname` character string. SOFT file name. Defaults to the name of the series info file, with suffix .soft. If "", print to the standard output connection.

`expressionmatrix` character string. Name of a tab-delimited text file with normalized expression values, by default in datadir (but if the string contains "/", it is considered a full path and it is read or written at the corresponding location). If expressionmatrix is non-null and a corresponding file exists, it is used as the source of gene expression values to be written in the SOFT file. Column names in this file should match the names of raw data files given in sampleinfo. If expressionmatrix is non-null but no corresponding file is found, normalized expression values calculated from the microarray data files (only Affymetrix CEL files at the moment) are written to this file.

`verbose` logical. If FALSE suppress all messages (useful to write the SOFT file to standard output). Defaults to TRUE.

Details

Converts a microarray dataset and the corresponding sample information into a SOFT file that is ready to be used for direct deposit into GEO.

Sample information and series information are taken from two separate plain text files. The column names in these two files should match the labels defined by the SOFT format (see http://www.ncbi.nlm.nih.gov/geo/info/soft2. This allows for the extraction of the information needed to write the SOFT format file.

Currently only a subset of SOFT keywords are accepted but they are expected to suffice for standard submissions (in particular for 1-color microarrays).

Normalized expression values are read from a tab-delimited file. Alternatively, in the case of an Affymetrix dataset, raw microarray data files (CEL files) can be processed directly (they are RMA-normalized) and written to the SOFT file.

See the example below or the vignette for the format of sampleinfo, seriesinfo and expressionmatrix.

Note

The SOFT file format is defined at http://www.ncbi.nlm.nih.gov/geo/info/soft2.html

Author(s)

Alexandre Kuhn, kuhnam@mail.nih.gov

Examples

```
##location of example files for sampleinfo and seriesinfo (i.e. provided
##with this vignette)

dataDirectory <- system.file(package='GEOsubmission', 'extdata')

## Define a directory and a file to write the generated example SOFT
## file out to (for the sake of this demo)
soft_example_fullpath <- tempfile(pattern='soft_example')
soft_example_name <- basename(soft_example_fullpath)
```

```
soft_example_dir <- dirname(soft_example_fullpath)

## Example with a dataset (series) named "neuronalCultures", comprised
## of 2 samples (named "1" and "2")

## This example is not run since the CEL files corresponding to sample
## "1" and "2" (and indicated in "sampleInfo.txt") do not exist (which
## would be purposedly detected by microarray2soft and cause it to
## stop).
## Not run:
microarray2soft(c('1','2'), 'sampleInfo.txt', 'neuronalCultures',
                'seriesInfo.txt', datadir=dataDirectory,
                writedir=soft_example_dir, softname=soft_example_name)


## End(Not run)

## The files "sampleInfo.txt" (as sampleInfo) and "seriesInfo.txt" (as
## seriesInfo) are provided as example files in the "extdata" directory of
## the package installation directory. Their content can be printed to the
## R console with
## Not run:
dataDirectory<-system.file(package='GEOsubmission','extdata')
read.delim(file.path(dataDirectory,'sampleInfo.txt'))
read.delim(file.path(dataDirectory,'seriesInfo.txt'))

## End(Not run)

## Example specifying a separate file containing normalized expression
## values (here the example file "expressionNormalized.txt")
microarray2soft(c('1','2'), 'sampleInfo.txt', 'neuronalCultures',
                'seriesInfo.txt', datadir=dataDirectory,
                writedir=soft_example_dir, softname=soft_example_name,
                expressionmatrix='expressionNormalized.txt')


## The file "expressionNormalized.txt" (as expressionmatrix) is provided
## as example file in the "extdata" directory of the package installation
## directory. It can be printed to the R console with
## Not run:
dataDirectory<-system.file(package='GEOsubmission','extdata')
read.delim(file.path(dataDirectory,'expressionNormalized.txt')

## End(Not run)

## Same as the preceding example but write the example SOFT file to the
##console (note the use of verbose=FALSE to suppress messages)
microarray2soft(c('1','2'), 'sampleInfo.txt', 'neuronalCultures',
                'seriesInfo.txt', datadir=dataDirectory,
                softname='', expressionmatrix='expressionNormalized.txt',
                verbose=FALSE)
```

`sample1` *sample1.CEL*

Description

Dummy Affymetrix CEL file.

Format

text file

`sample2` *sample2.CEL*

Description

Dummy Affymetrix CEL file.

Format

text file

`sampleInfo` *sampleInfo.txt*

Description

Example file with sample information.

Format

tab-delimited file

`seriesInfo` *seriesInfo.txt*

Description

Example file with series information.

Format

tab-delimited text file

Index

expressionNormalized, 2

microarray2soft, 2

sample1, 5
sample2, 5
sampleInfo, 5
seriesInfo, 5