Iterative network guided cMapping and validation

Supplementary Material and Methods - Supplementary Code: SIG_REVERSION

This document describes functions, scripts and data objects used in the software enclosed to the paper entitled *A semi-supervised approach for refining transcriptional signatures of drug response and repositioning predictions*, by Francesco Iorio et al, submitted as research paper to PLoS ONE.

Author: Francesco Iorio (iorio@ebi.ac.uk)

Paper website: http://www.ebi.ac.uk/~iorio/PLoS_ONE_Submission

May 1, 2014

`cell_lines_connected_to_mult_sig`
Identifying cancer cell lines connected to multiple signatures

---

Description

This function identifies cancer cell lines in the GDSC [1] panel whose post-processed basal expression profile is connected to multiple signatures

Usage

```
cell_lines_connected_to_mult_sig(multiple_sig_cs, th = 0.3)
```

Arguments

`multiple_sig_cs`	Connection scores basal expression profiles of the GDSC [1] cell lines to multiple signatures. A list of connection scores data frames obtained by using the `CS` function
`th`	False discovery rate threshold. A cell line is connected simoultaneously to the multiple signatures if the false discovery rate of all the connection scores is below this threshold

Details

For usage examples see the pipeline described at http://
www.ebi.ac.uk/~iorio/PLoS_CB_Submission/sigRevPL/
SigRevPL.html

Value

A list containing two string vectors: `POS` and `NEG`. The former contains COSMIC [2] identifiers of cell lines positively connected simoultaneously to the multiple signatures, the latter those of the cell lines negatively connected simoulteneously to the multiple signatures.

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

Paper website: http://www.ebi.ac.uk/~iorio/PLoS_ONE_Submission

References

[1] Garnett,M.J. et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature, 483, 570-575

[2] Forbes,S.A. et al. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res, 39, D945-50.

`genericTtest` *Generic t-test for drug response differences*

Description

This function implements a simple t-test to assess the extent of difference (and its statistical significance) in drug response across two user defined population of cell lines of the GDSC [1] screening

Usage

```
genericTtest(IC50s,
             ALLscreenedCellLines,
             specific_cell_lines,
             drug = drug_id,
             display = TRUE,
             labels = NULL)
```

Arguments

`IC50s`	A matrix of IC50 values contained in the `SCREENING` object
`ALLscreenedCellLines`	A string vector containing the COSMIC [2] identifiers of all the cell lines to be included in the test
`specific_cell_lines`	A string vector containing the COSMIC [2] identifiers of a subset of cell lines to tested for differences in drug response
`drug`	The internal identifiers of a drug
`display`	A boolean parameter specifying if a box plot should be plotted (default = TRUE)
`labels`	A string vector with the labels to be plotted below the two groups of cell lines. If `display = FALSE` then this parameter is ignored

Value

A list containing the following items:

PVAL	The p-value of the performed t-test
deltaMEAN	Difference of the mean IC50 values across the two groups of cell lines
effectSize	Cohen's d quantifying the effect size of the group/drug-response association
N1	Number of cell lines in the first group
N2	Number of cell lines in the second group

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

Paper website: http://www.ebi.ac.uk/~iorio/PLoS_ONE_Submission

References

[1] Garnett,M.J. et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature, 483, 570-575

[2] Forbes,S.A. et al. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res, 39, D945<e2><80><93>50.

Examples

```
## loading functions and objects needed to retrieve drug names and targets
source("CODE/CONNECTION_SCORES_library.R")

## loading drug annotations
load('DATA/GDSC_DRUG_ANNOTATIONS.ro')

## loading cell line annotations
load('DATA/GDSC_CELL_LINE_ANNOTATIONS.ro')

## loading drug screening data
load("DATA/GDSC_DRUG_SCREENING_DATA.ro")


## selecting areo-digestive-tract cancer cell lines
aero_dig_tract_cancer_cell_lines<-
as.character(MASTER_LIST$COSMIC.ID
[which(MASTER_LIST$GDSC.description_1=="aero_dig_tract")])

## assessing the difference in response to paclitaxel
## across two group of cell lines (aero-digestive-tract cancer vs others)
genericTtest(SCREENING$IC50s,
             rownames(SCREENING$IC50s),
             aero_dig_tract_cancer_cell_lines,
             drug='11',labels=c('aerodig tract','others'))
```

`test_pred_ability` *Testing the predictive ability of the signatures*

Description

This function evaluates the difference in drug response across two groups of cell lines in the GDSC [1] panel. The first one is composed by cell lines negatively connected to a set of signatures (simoultaneously). The second one contains all the other cell lines in the panel

Usage

```
test_pred_ability(multiple_sig_cs, DRUGS, th = 0.3, mainTitle, display = TRUE)
```

Arguments

`multiple_sig_cs`	Connection scores of the basal expression profiles of the GDSC [1] cell lines to multiple signatures. A list of connection scores data frames obtained by using the `CS` function
`DRUGS`	The internal identifiers of the drugs to be tested
`th`	False discovery rate threshold. A cell line is connected simoultaneously to the multiple signatures if the false discovery rate of all the connection scores is below this threshold
`mainTitle`	Main title of the resulting figure (if `display = TRUE`))
`display`	A boolean parameter specifying if a box plot should be plotted (default = TRUE)

Details

For usage examples see the pipeline described at
http://www.ebi.ac.uk/~iorio/PLoS_ONE_Submission/sigRevPL/SigRevPL.html

Value

A data frame with a row for each tested drug, and the following columns:

`used signature(s)`	Name of the signature(s) tested
`drug id`	Internal identifier of the tested drug
`drug`	Name of the tested drug
`target`	Target of the tested drug
`p-value`	P-value of a t-test assessing the extent of difference in drug response when dichotomising the set of cell lines in the GDSC [1] panel in two groups: those negatively connected to the tested signatures and all the others
`deltaMean`	The difference in average IC50s across the two groups of samples described above
`effectSize`	The effect size of the t-test
`N2`	Number of cell lines negatively connected to the tested signatures
`N1`	Number of cell lines in the rest of the GDSC panel

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

Paper website: http://www.ebi.ac.uk/~iorio/PLoS_ONE_Submission

References

[1] Garnett,M.J. et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature, 483, 570-575

Index