Iterative network guided cMapping and validation

Supplementary Material and Methods - Supplementary Code: ITERATIVE_CMAPPING

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

April 30, 2014

`DNquery` *Querying the drug network with a seed compound*

Description

This function queries the drug network described in [1] for compounds whose consensual trascriptional response is similar to that of a given one (i.e. the seed compound).
This set of compounds is called the seed neighborhood. Once the seed neighborhood is computed an enrichment analysis for over-represented drug communities (identified in [1]) is also performed

Usage

```
DNquery(seed = "paclitaxel", distTh = 0.8065, printToFile = FALSE)
```

Arguments

`seed`	String specifying the name of the compound to be used as seed (default = 'paclitaxel')
`distTh`	The distance threshold below which the transcriptional response of two compounds should be considered significantly similar (default = 0.8065 as heuristically determined in [1])
`printToFile`	A boolean parameter specifying if the output of this function should be stored in a tab delimited txt file (default = FALSE). If TRUE then a file (whose name is `$$_DN_neighborhood`, wher `$$` is the name of the seed compound) is created in the `~/OUTPUT` directory (where ~ is the working directory)

Value

A data frame with a row for each of the identified seed neighbors, with the following columns:

`D`	Distance between the seed compound and the compound specified by the row
`quantile perc`	Percentile where the drug distance falls when sorting all the distances in decreasing order
`Drug`	Name of the seed neighboring drug
`C id`	Numerical identifier of the community containing the drug under consideration

order	The neighborhood order (i.e. a neighbor of order K contains the K closest to the seed neighbors according the distance specified in D)
C occ	Community occurrence = how many drugs belonging to the community whose identifier is specified in id are observed in the neighborhood of order K, where K is the row number
C card	Community cardinality = how many drugs are contained in the community in the drug network described in [1]
Total #drugs	Total number of drugs contained in the drug network described in [1]
C Overrep p-val	Probability of observing by chance the number of drugs specified in C Occ, in a neighborhood whose order is specified in order, given the background populations specified in C card and Total #drugs
Adj p-val	The p-value described above, after correction for multiple hypothesis testing
MOAs	The modes of action (or drug features) enriched in the drug community specified in C id

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

Examples

```
## querying the drug network for the neighbors of daunorubicin (a topoisomerase inhibitor)
NN<-DNquery('daunorubicin')

## visualising the first 10 neighbors of daunorubicin
print(NN[1:10,])
```

DeriveConsistentSignature

Computing consistent signatures

Description

This function computes signatures of genes that are consistently up- (resp. down-) regulated when considering the optimal signature [1] of a seed compound and the prototype ranked lists of other user defined connectivity map [2] compounds

Usage

```
DeriveConsistentSignature(seed = "paclitaxel",
                          otherCompounds = c("MG-132", "celastrol", "5224221"),
                          PTH = 30, FUZZYNESS = 2, printToFile = FALSE)
```

Arguments

seed	A string specifing the name of the connectivity map [2] drug that should be used as seed
otherCompounds	A list of strings specifying the names of the other compounds whose prototype ranked list [1] should be checked for consistency with the optimal signature of the seed
PTH	The expression percentile that should be considered when building the consistent signature (see the material and methods of our manuscript for further details)
FUZZYNESS	The number of other compounds that should satisfy the consistency
printToFile	A boolean parameter specifying if the output of this function should be stored in two tab delimited txt file (default = FALSE), respectively for the up- and the down-regulated part of the signature. If TRUE then two files, whose name will be `$_$$_consistentSignatureUP` (resp. `$_$$_consistentSignatureDOWN`), where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, are created in the `~/OUTPUT` directory (where ~ is the working directory)

Value

A list containing two data frames (`seedUPreg` and `seedDOWNreg`). The first column of these data frame contains the probe-set identifiers in the up-regulated (resp. down-regulated) part of the consistent signature. The following columns contain the percentile in which each probe-set falls along the prototype ranked list [1] of the seed and the other compounds.

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[2] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

Examples

```
DeriveConsistentSignature(seed='paclitaxel',otherCompounds=c('MG-132','celastrol'))
```

`DeriveInConsistentSignature` *Computing inconsistent signatures*

Description

This function computes signatures of genes that are inconsistently up- (resp. down-) regulated when considering the optimal signature [1] of a seed compound and the prototype ranked lists of other user defined connectivity map [2] compounds

Usage

```
DeriveInConsistentSignature(seed = "paclitaxel",
                            otherCompounds = c("MG-132", "celastrol", "5224221"),
                            PTH = 30, FUZZYNESS = 2, printToFile = FALSE)
```

Arguments

`seed`	A string specifing the name of the connectivity map [2] drug that should be used as seed
`otherCompounds`	A list of strings specifying the names of the other compounds whose prototype ranked list [1] should be checked for inconsistency with the optimal signature of the seed
`PTH`	The expression percentile that should be considered when building the inconsistent signature (see the material and methods of our manuscript for further details)
`FUZZYNESS`	The number of other compounds that should satisfy the inconsistency
`printToFile`	A boolean parameter specifying if the output of this function should be stored in two tab delimited txt file (default = FALSE), respectively for the up- and the down-regulated part of the signature. If TRUE then two files, whose name will be `$_$$_inconsistentSignatureUP` (resp. `$_$$_inconsistentSignatureDOWN`), where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, are created in the `~/OUTPUT` directory (where ~ is the working directory)

Value

A list containing two data frames (`seedUPreg` and `seedDOWNreg`). The first column of these data frame contains the probe-set identifiers in the up-regulated (resp. down-regulated) part of the consistent signature. The following columns contain the percentile in which each probe-set falls along the prototype ranked list [1] of the seed and the other compounds.

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[2] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

Examples

```
DeriveInConsistentSignature(seed='paclitaxel',otherCompounds=c('MG-132','celastrol'))
```

`DeriveMSTSignature` *Computing inconsistent signatures (less stringently)*

Description

This function computes signatures of genes that are inconsistently up- (resp. down-) regulated when considering the prototype ranked lists [1] of a seed compound and those of other user defined connectivity map [2] compounds

Usage

```
DeriveMSTSignature(seed = "paclitaxel",
                   otherCompounds = c("albendazole", "fenbendazole",
                                      "nocodazole", "parbendazole"),
                   PTH = 25, FUZZYNESS = 4, printToFile = FALSE)
```

Arguments

`seed`	A string specifing the name of the connectivity map [2] drug that should be used as seed
`otherCompounds`	A list of strings specifying the names of the other compounds whose prototype ranked list [1] should be checked for inconsistency with that of the seed
`PTH`	The expression percentile that should be considered when building the inconsistent signature (see the material and methods of our manuscript for further details)
`FUZZYNESS`	The number of other compounds that should satisfy the inconsistency
`printToFile`	A boolean parameter specifying if the output of this function should be stored in two tab delimited txt file (default = FALSE), respectively for the up- and the down-regulated part of the signature. If TRUE then two files, whose name will be `$_$$_MST_UP` (resp. `$_$$_MST_DOWN`), where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, are created in the `~/OUTPUT` directory (where ~ is the working directory)

Value

A list containing two data frames (`seedUPreg` and `seedDOWNreg`). The first column of these data frame contains the probe-set identifiers in the up-regulated (resp. down-regulated) part of the inconsisten signature. The following columns contain the percentile in which each probe-set falls along the prototype ranked list [1] of the seed and the other compounds.

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[2] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

Examples

```
DeriveMSTSignature(seed='paclitaxel',otherCompounds=c('albendazole','nocodazole'),FUZZYNESS=2)
```

DeriveSingleSignature *Computing single drug optimal signatures*

Description

This function computes the optimal signature (as defined in [1]) for a compound contained in the connectivity map dataset [2]

Usage

```
DeriveSingleSignature(seed = "paclitaxel")
```

Arguments

seed	A string specifying the name of the connectivity map compound whose optimal signature should be computed (default = 'paclitaxel')

Value

A list containing two data frames (seedUPreg and seedDOWNreg). The first column of these data frame contains the probe-set identifiers in the up-regulated (resp. down-regulated) part of the optimal signature. The second column contains the percentile in which each probe-set falls along the prototype ranked list [1] of the compound under consideration.

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[2] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

Examples

```
DeriveSingleSignature(seed='metformin')
```

`percHeatMaps` *Visualising heatmaps of expression percentiles*

Description

This function visualise the expression percentiles of a set of genes along the prototype ranked lists [1] of a seed drug and those of other user defined compounds

Usage

```
percHeatMaps(probes, seed, otherCompounds, printToFile = FALSE)
```

Arguments

`probes`	A string vector containing the probe-set identifiers whose percentile should be visualised
`seed`	A string specifying the name of the seed compound
`otherCompounds`	A string vector containing the names of the other compounds in the connectivity map [2] dataset
`printToFile`	A boolean parameter specifying if the heatmap produced by this function should be stored in a png file (default = FALSE). If TRUE then a file, whose name will be `$_$$_percHeatMap.png`, where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, will be created in the `~/OUTPUT` directory (where ~ is the working directory)

Details

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

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[2] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

`plotRunningSums` *Visualising running sums for consistent/inconsistent signatures*

Description

This function visualise the enrichment score [1] running sums of two different signatures along the prototype ranked lists [2] of a seed compound and those of other user defined connectivity map drugs [3]

Usage

```
plotRunningSums(consistentSigTable,
                inconsistentSigTable,
                seed = "paclitaxel",
                otherCompounds = c("MG-132", "celastrol", "5224221"),
                printToFile = FALSE)
```

Arguments

`consistentSigTable`	A signature generated by the function `DeriveConsistentSignature`
`inconsistentSigTable`	A signature generated by the function `DeriveInConsistentSignature`
`seed`	A string specifying the name of the seed compound
`otherCompounds`	A string vector containing the names of the other compounds in the connectivity map [3] dataset
`printToFile`	A boolean parameter specifying if the plots generated by this function should be stored in a png file (default = FALSE). If TRUE then a file, whose name will be `$_$$_RS.png`, where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, will be created in the `~/OUTPUT` directory (where ~ is the working directory)

Details

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

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Subramanian,A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102, 15545.

[2] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[3] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

`plotRunningSumsMST` *Visualising running sums for (less stringently) inconsistent signatures*

Description

This function visualise the enrichment score [1] running sums of a given signature along the prototype ranked lists [2] of a seed compound and those of other user defined connectivity map drugs [3]

Usage

```
plotRunningSumsMST(MSTsignatureTable,
                   seed = "paclitaxel",
                   otherCompounds = c("albendazole", "fenbendazole",
                                      "nocodazole", "parbendazole"),
                   printToFile = FALSE)
```

Arguments

`MSTsignatureTable`	A signature generated by the function `DeriveMSTSignature`
`seed`	A string specifying the name of the seed compound
`otherCompounds`	A string vector containing the names of the other compounds in the connectivity map [3] dataset
`printToFile`	A boolean parameter specifying if the plots generated by this function should be stored in a png file (default = FALSE). If TRUE then a file, whose name will be `$_$$_RS.png`, where `$` is the name of the seed compound and `$$` is a string composed by the other compound names, will be created in the `~/OUTPUT` directory (where ~ is the working directory)

Details

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

Author(s)

Francesco Iorio (iorio@ebi.ac.uk)

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

References

[1] Subramanian,A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102, 15545.

[2] Iorio,F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107, 14621.

[3] Lamb,J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929.

Index