Introduction

Hope4Gene is a Hopfield-like class prediction algorithm for gene expression data. The software reads expression data and a signature and classifies samples of the expression data in respect to the signature.

Installation

The following software components are required to run **Hope4Gene:**
- **Hope4Gene**
- R (http://www.r-project.org/)

In principle, **Hope4Gene** should run under Linux and Mac.

Usage

Hope4Gene can be run in two forms with or without FDR computation.
To run Hope4Gene without FDR computation from shell:

- Go to the downloaded folder
- ./ run_Hope4Gene_noFDR.sh

To run Hope4Gene with FDR computation from shell:

- Go to the downloaded folder
- ./ run_Hope4Gene_FDR.sh

Inputs
The program requires as inputs:
- An expression matrix in log2ratio with the first column containing genes names and the first row containing sample names.
- A signature file composed of two columns:
  - The first containing signature gene names
  - The second containing the class associated to each signature gene
  The first row of the file has to contain a description of the two columns content.
- The total number of classes (greater equal to 2)
- If the total number of classes is greater equal to 3 other two parameters are needed:
  - Number of signature gene resampling to be performed (suggested value 100)
  - Number of signature gene to be choose at random during the resampling (suggested value 200)
If you are running run_Hope4Gene_FDR.sh an other input is needed:
- The number of permutations to be performed in order to estimate the FDR associated to each sample classification.

Outputs
The outputs are:
- final_report_withoutFDR.txt : a report containing on the first column sample class, on the second column the second sample class and on the third column the energy

associated to the sample classification. Each row is one of the samples ordered according to the comparison in the expression matrix.

- permutation_report.txt : a report containing the results of the permmutations with random samples. The report is organized as follows: on the first column sample class, on the second column the energy associated to the sample classification. Each row is one of the samples ordered according to the comparison in the expression matrix.

- final_report_withFDR.txt : a report containing on the first column sample class, on the second column the second sample class, on the third column the energy associated to the sample classification and on the fourth column the FDR associated to sample classification. Each row is one of the samples ordered according to the comparison in the expression matrix.

Data

The datasets used in the paper and created by

Hoshida Y (2010) Nearest Template Prediction: A Single-Sample-Based Flexible Class Prediction with Confidence Assessment. PLoS ONE 5(11): e15543. doi: 10.1371/journal.pone.0015543

and can be downloaded at http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

contact

Please feel free to contact us at

Cantinilaura88@gmail.com

Moreover, feel free to change the code according to your needs.