Introduction

Gene4x is a pipeline for the identification of communities on a gene multiplex composed of four layers: co-expression network, transcriptional and post-transcriptional co-targeting network and protein-protein interaction network.

The pipeline performs the following operations:

1. Multiplex reconstruction
2. Layers filtering
3. Community detection on the multiplex

ATTENTION! The alpha thresholds for step 2. has to be chosen from the user. We suggest to follow the three criteria proposed in the paper.
ATTENTION! The community detection algorithm for step 3. has to be chosen from the user, the suggested one is OSLOM.

Installation

The following software components are required to run **Gene4x:**

- **Gene4x**
- R (http://www.r-project.org/)
- R packages: igraph and BioPhysConnectoR
- package Ganet [REF] (https://sites.google.com/site/gokmenaltay/ganet). To install the package download the file Ganet_1.0.tar.gz at the above link. Set the working directory in the same folder that ganet is placed. Then, in the R console, write the following: install.packages("GAnet_1.0.tar.gz", type="source").
- Python

- Python modules: networkx and scipy

In principle, **Gene4x** should run under Linux and Mac.

Usage

To run **Gene4x**:
- download the net_mirna_and_tf.zip file at the link
  http://personalpages.to.infn.it/~caselle/BioPhys/Gene4x/
- unzip net_mirna_and_tf.zip -d /path_to_Gene4x/data/
- Set the working directory in Gene4x (cd path_to_Gene4x/)
- ./ Gene4x_run.sh

Inputs

While running, the program requires the following inputs:

- The path to the file containing the mRNA expression matrix in log2 with the first column containing genes names, the second column containing gene annotations (or gene names) and the first row containing sample names.

- **I**nteger corrisponding to the optimal alpha value, choosing among (0=0.005,1=0.01,2=0.02,3=0.03,4=0.05,5=0.1,6=0.2,7=0.3,8=0.4,9=0.5) for the espression network, transcription factor co-taregting network and microRNA co-taregting network. To choose the optimal alpha value you can use the three criteria suggested in the paper taking advantage of the files:
  - Gene4x/temp/selection_links_EXP.txt for the co-expression layer.

- Gene4x/temp/selection_links_TF.txt for the transcription factor co-taregting layer.
- Gene4x/temp/selection_links_MIRNA.txt for the microRNA co-taregting layer.

- The integer corrisponding to the algorithm you want to use for network clustering, choosing among:

  - 0: oslom undirected,
  - 2: infomap undirected
  - 4: louvain
  - 5: label propagation method,
  - 8: modularity optimization

Outputs

During **Gene4x** running some intermidiate outputs are created in Gene4x/temp/ the folder will be delated when thre execution is completed.

The final output comm_final_genesymbol.txt is in the folder Gene4x. The file reports in the first column an integer used as identifier for the communities. Then in each row the list of genes belonging to that community is reported. Communities are ordered in respect to thei dimension.

Data

The datasets to reproduce the results of our paper are available at:
- gastric GSE13911

- lung GSE10072,
- pancreas GSE15471
- colon GSE44076

Contact

Please feel free to contact us at

cantinilaura88@gmail.com

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

For every use of the original or modified pipeline cite us:

Cantini et al., Community detection on genomic multiplex networks reveals cancer drivers