

# Tutorial: High-Throughput Screening Network analysis with the HTS-Net Pipeline

Ghislain Bidaut, Claire Rioualen

December 29, 2016

Copyright: Ghislain Bidaut, Claire Rioualen

This document has been placed under the CeCILL Licence.

## Introduction

HTS-Net, a network-based analysis program, was developed to identify gene regulatory modules impacted in high-throughput screenings by integrating regulatory data (regulome), protein-protein interactions (interactome) and RNAi screening z-scores.

HTS-Net works by discovering subnetworks in parallel on an interactome and a regulome network. RNAi-based gene scores are superimposed on the interaction and the regulation maps separately. Each one of them is then searched for high scoring areas. The identified regions of interest, so-called subnetworks, are extracted and reported. The obtained modules are merged to form meta-subnetworks that incorporate regulation information and PPIs.

The program takes normalized data as an input, under the form of a list of genes with associated score values. Two organisms are supported so far (*Homo sapiens* and *Mus musculus*). The data is superimposed on 2 networks:

- A protein-protein interactions network (interactome)
- A transcription factor-target gene network (regulome)

These networks were aggregated from the following databases:

- Interactome: DIP, HPRD, I2D, Intact, Mint, Proteinpedia, Mysickova et al., Ravasi et al., Yu et al.
- Regulome: ITFP, TRED, Pazaar, ORegAnno

After detection and statistical validation by data shuffling, subnetworks detected in the interactome and in the regulome are integrated by measuring their identity (number of common nodes).

Note that non free data (TRANSFAC) is not available through this implementation. It is also possible to use HTS-Net through the commandline by downloading source code available at <http://htsnet.marseille.inserm.fr/distribution.html>

## Usage

The HTS-Net pipeline is accessible through our Mobyale portal. Mobyale allows querying the input data, verifying data and arguments integrity and tracing your run. It will also send you an email when the HTS-Net run is done.

## Data Formatting

The input file containing the dataset (screening scores) must be correctly formatted according to the format accepted by HTS-Net:

- 2-column tab-delimited file
- first column contains EntrezGene identifiers
- second column contains the associated normalized score measured from the scening

If you need to convert gene IDs from EntrezGene identifiers to Gene Symbols or the reverse, you can use the `convertGID_GS` utility from our Mobyale portal: [http://mobyalehome.marseille.inserm.fr/cgi-bin/portal.py#forms::convertGID\\_GS](http://mobyalehome.marseille.inserm.fr/cgi-bin/portal.py#forms::convertGID_GS)

This is especially important as HTS-Net works internally with Entrez Gene identifiers and does not accept any other type of identifier.

## Access to the system

HTS-Net is accessible from our Mobyale portail at the following address: <http://htsnet.marseille.inserm.fr/index.html> by clicking on "HTS-Net RUN page on Mobyale portal" link.

You don't need to possess an account on our Mobyale portal and can directly use it as a guest user. Once you access the system, you are welcomed with a parameter form. This form allows the customization of the HTS-Net analysis. Help can be obtained for all parameters by clicking on the '?'. All parameters marked with an asterisk (\*) are mandatory.

# HTS-Net 2016.08

## Detection of discriminative subnetworks from High Throughput Screening

Run

Reset

Input parameters

\* Screening score file ?

paste upload EDIT CLEAR

Enter your data below:

Seed list ?

paste upload EDIT CLEAR

Enter your data below:

\* Taxon Homo sapiens

\* Subnetwork score threshold ? 0.5

\* Minimum improvement value ? 0.05

\* Maximum Subnetwork Depth ? 5

\* Number of random subnetwork to generate for statistical validation ? 1000

Use absolute value of gene score to calculate subnetwork score ? Yes

Use anticorrelation value to calculate subnetwork score ? No

Fitting model for null distributions of subnetwork scores ? Gaussian mixture

P-value Interactome ? 1e-3

P-value Regulome ? 5e-2

Maximum subnetwork overlap allowed ? 0.8

Nodes to highlight ?

paste upload EDIT CLEAR

Enter your data below:

Regulome-Interactome subnetworks connectivity threshold ? 2

\* General name of your project ? screening-project

Name of your analysis ?

Description file ?

paste upload EDIT CLEAR

Enter your data below:

Author(s): (Alphabetical order) Ghislain Bidaut, Bernard Chetrit, Quentin Da Costa, Claire Rioualen

HTS-Net welcome screen

## List of parameters to set up

- **Screening score file:** Upload your formatted data file (or paste the content). We propose to use the Chia-Fav dataset provided at the HTS-Net Web page: <http://htsnet.marseille.inserm.fr/data-examples.html>.
- **Seed list:** This allows to limit the recursive exploration of regulome and interactome to a set of genes of specific interest. This is optional, and we won't use it in the current example.
- **Taxon:** Here, we specify Homo sapiens.
- **Minimum score threshold:** The minimum score required to keep a subnetwork before statistical validation. Here, we keep the default parameter 0.5.
- **Minimum improvement value *th*:** This is the threshold of minimal score improvement used for subnetwork aggregation. Here, we keep 0.05.
- **Absolute value:** This specifies whether we detect subnetworks with both positive and negative gene z-scores. We will use “No”.
- **Use Anticorrelation:** This specifies that we wish to detect genes with highly negative z-scores. We will set it to “Yes” for our example.
- **Fitting model:** This specifies the model to use for null distribution of score. Again we keep the default value “Gaussian mixture”.
- **P-values Interactome and Regulome:** These are the p-values used for the comparison of selected subnetworks and random subnetworks. They are left to their default for the present tutorial.
- **Overlap:** If two subnetworks overlap (proportion of proteins/genes in common of 2 subnetworks) over this rate, only the subnetwork with the best score is kept. By default, we will keep at 0.8 (80%).
- **Interactome-Regulome subnetworks connectivity threshold:** This specifies the number of common nodes between interactome and regulome subnetworks for the construction of meta-subnetworks. Set to 2.
- **Project name:** Please specify a name for your project in this box (e.g. "project\_screening").
- **Analysis name:** Please specify a name for your analysis in this box (e.g. "chia\_analysis"). This is used throughout the final report.
- Other optional parameters are left to their default. Note that non-default values are displayed in yellow windows.

Once done, your configuration should look as follows.

# HTS-Net 2016.08

## Detection of discriminative subnetworks from High Throughput Screening

Run

Reset

Input parameters

\* Screening score file ?

paste upload EDIT CLEAR

Parcourir... Aucun fichier sélectionné. select

| Entrez | Gene ID       | Fav |
|--------|---------------|-----|
| 5460   | -5.1481605125 |     |
| 79085  | -4.9337946482 |     |
| 375387 | -4.8418470569 |     |
| 10413  | -4.2564149425 |     |
| 5708   | -4.1719019843 |     |
| 400134 | -4.1240057060 |     |

Seed list ?

paste upload EDIT CLEAR

Enter your data below:

\* Taxon Homo sapiens

\* Subnetwork score threshold ? 0.5

\* Minimum improvement value ? 0.05

\* Maximum Subnetwork Depth ? 5

\* Number of random subnetwork to generate for statistical validation ? 1000

Use absolute value of gene score to calculate subnetwork score ? No

Use anticorrelation value to calculate subnetwork score ? Yes

Fitting model for null distributions of subnetwork scores ? Gaussian mixture

P-value Interactome ? 1e-3

P-value Regulome ? 5e-2

Maximum subnetwork overlap allowed ? 0.8

Nodes to highlight ?

paste upload EDIT CLEAR

Enter your data below:

Regulome-Interactome subnetworks connectivity threshold ? 2

\* General name of your project ? screening-project

Name of your analysis ?

Description file ?

paste upload EDIT CLEAR

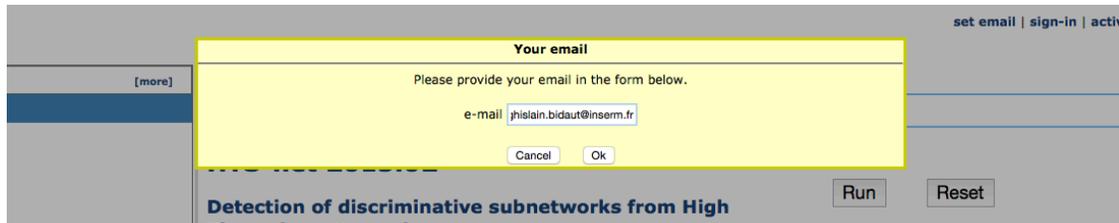
Enter your data below: select

Author(s): (Alphabetical order) Ghislain Bidaut, Bernard Chetrit, Quentin Da Costa, Claire Rioualen

Configuration of HTS-Net

## Start HTS-Net run

To start the analysis, click on the **Run** button at the top of the form.



The screenshot shows a web interface with a grey background. At the top right, there are links for "set email", "sign-in", and "activ". In the center, a yellow dialog box titled "Your email" is displayed. Inside the dialog, it says "Please provide your email in the form below." Below this, there is a label "e-mail" followed by a text input field containing "phislain.bidaut@inserm.fr". At the bottom of the dialog are "Cancel" and "Ok" buttons. In the background, a "Run" button is visible, and below the dialog, the text "Detection of discriminative subnetworks from High" is partially visible.

The system prompts for an email address. Please provide your email in the box.

### *E-mail prompt*



The screenshot shows a yellow dialog box for Captcha confirmation. At the top, there is an image of a sunset over mountains with the word "laugh" written in a stylized font. Below the image, it says "To validate your submission, please type the above text in the field below." There is a text input field containing "laugh". At the bottom of the dialog are "Cancel" and "Ok" buttons. To the right of the dialog, a "Reset" button is visible on the main interface.

Then, there is a Captcha confirmation. Please fill in the text appropriately.

### *Captcha*

After this operation, a Tab containing the HTS-Net run appears. You can verify the parameters from this display.

The screenshot shows a web browser window with a tab titled "HTS-Net - 01/02/17 14:36:35". The address bar contains the URL <http://mobyhome.marseille.inserm.fr/data/jobs/HTS-Net/O07575714993954>. Below the address bar are four buttons: "update", "get help", "back to form", and "remove job". The main content area is titled "parameters" and is divided into two sections. The first section, "Input parameters", contains three rows: "Screening score file (Expression) ?" with a dropdown menu showing "chia-screen-data-Fav.txt (not detected)" and a "save" button; "Use absolute value of gene score to calculate subnetwork score ? False"; and "Use anticorrelation value to calculate subnetwork score ? True". The second section, "job execution", is currently empty.

### *Summary of the parameters*

You can close your browser if necessary, the run will continue on our server.

### **HTS-Net results**

After waiting a *couple of hours*, you will receive a notification in your mailbox, to the address provided above (be careful, this email may be filtered out by spam filter). Results are provided under the form of a URL that links back to a page showing the following form.

## Results

Three files are available from this page:

- **HTS-Net analysis report.**

This file contains analysis report (list of subnetworks detected, list of genes, and annotations) in HTML format.

- **Standard output and Standard error files.**

These files contain the messages produced by HTS-Net in the standard output and standard error. This is mostly useful for developers and advanced users.

You can download the HTML report by clicking on the save button , unzip it on your desktop and open the index.html file.

**chia\_test analysis**

---

*Results*

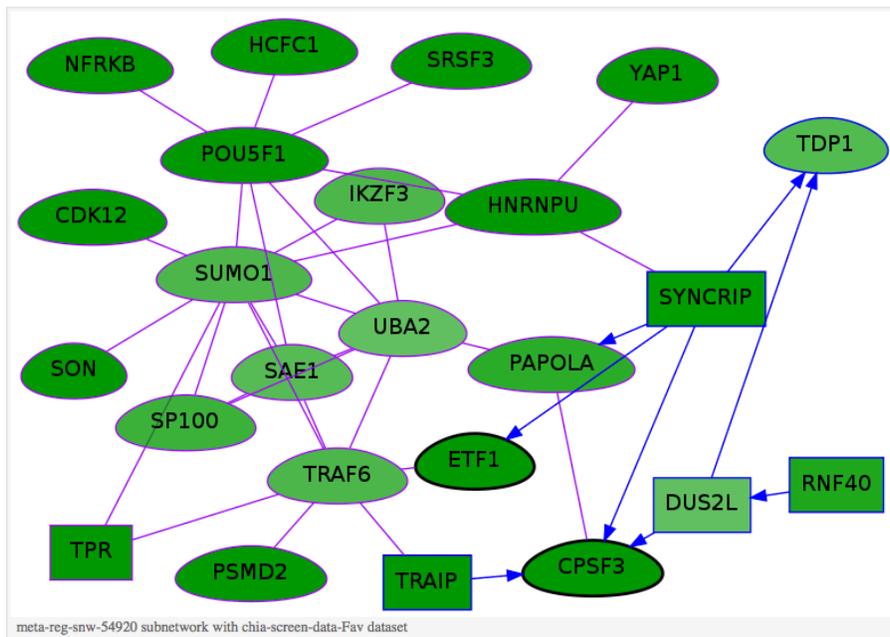
Three tabs are provided: Interactome, Regulome, Integrated Analysis. Each of these give access to a list of subnetworks, a list of genes contained in these subnetworks, and a list of enriched GO terms. Please refer to the publication for details on the analysis provided.

By clicking on the Integrated Analysis link and on the meta-reg-snw-54920 link, you can display the following subnetwork as an exemple.

meta-reg-snw-54920

Interacting subnets

| Subnetwork                    | Dataset              | Score | p-value 1 | p-value 2 | p-value 3 | Size |
|-------------------------------|----------------------|-------|-----------|-----------|-----------|------|
| <a href="#">int-snw-10054</a> | chia-screen-data-Fav | 2.750 | 9.25e-43  | 4.20e-02  | 3.99e-03  | 19   |
| <a href="#">reg-snw-54920</a> | chia-screen-data-Fav | 2.249 | 2.82e-32  | 4.34e-02  | 1.06e-02  | 7    |



Results

Contact

For all questions and queries, please contact ghislain dot bidaut at inserm dot fr.