MEMo 1.0 Software: README

Contents

1 Introduction

MEMo is a method for identifying mutually exclusive driver networks in cancer. The method is freely available as a Java-based command line tool. This guide describes the basics of using the command line tool, and accessing the source code.

2 System Requirements

To install MEMo, you must have:

- Java 1.5 or later.
- Python 2.5 or later.

3 Installation

3.1 On Mac OS X / Linux

If you are installing MEMo on Mac OS X or Linux, you probably already have Python and Java installed. Once you have verified that, follow the steps below:

- Unzip/untar memo into a directory of your choosing, e.g. /Users/xxx/memo.
- Add MEMO_HOME as an environment variable, and set it to your memo directory.
- Add MEMO_HOME/bin to your global path.

For example, on Mac OS X, I have added the following to my .bash_profile:

```
export MEMO_HOME="/Users/cerami/dev/sander/memo/"
export PATH=$PATH:$MEMO_HOME/bin
```

3.2 On Windows

To install MEMo on Windows, you may need to install Python first. Download it from: http://www.python.org/download/releases, and follow the installation instructions.

After that, you will need to:

- Unzip/untar memo into a directory of your choosing, e.g. c:\memo.
- Add MEMO_HOME as an environment variable, and set it to your memo directory. In Windows, this is usually done via the Control Panel. Complete instructions are available at: http://www.cs.usask.ca/~wew036/latex/env.html.
- Add MEMO_HOME/bin to your global path. In Windows, this is also done via the Control Panel. See the URL above, if you need detailed instruction.

3.3 Directory Structure

Once you have unzipped and untarred MEMo, you will see the following directory structure:

- **bin**: contains all MEMo scripts, including those required to initialize the database.
- **build**: the build directory, containing compiled classes and required third-party library files.
- **cancer_data**: cancer data for TCGA GBM and TCGA ovarian cancers.
- **config**: configuration directory for configuring logging and the embedded database.
- **data**: data directory containing human gene information from NCBI and pathway data from Pathway Commons and Reactome.

- **lib**: all third-party JARs.
- **src**: all source code.
- **templates**: web templates, used to generate the final MEMo web reports.
- **testData**: test data, used by our unit tests.
- **web**: web resources, including Cytoscape Web, used by the final MEMo web reports.

4 Initializing the Database

Before you can execute MEMo, you must load the embedded database with background information, including information regarding human genes and human interaction networks. You have the option of loading HRN1 or HRN2 (see manuscript by Ciriello, et. el. for details). To load HRN1:

- cd bin
- ./loadHrn1.sh

Loading of the database should take less than 3 minutes.

5 Running MEMo on Sample GBM Data

To run MEMo on the TCGA GBM data set described in the paper by Ciriello, et. el. follow these steps:

- initialize the database (see above)
- cd cancer_data/gbm
- memo.py memo_gbm.props

MEMo will automatically write its results to a directory called html_out and to a text file called: MemoReport.txt. To view the web report, open html_out/index.html in a web browser (Figure 1).

6 Running MEMo on Sample Ovarian Data

To run MEMo on the TCGA Ovarian data set described in the paper by Ciriello, et. el. follow these steps:

- initialize the database (see above)
- cd cancer_data/ovarian

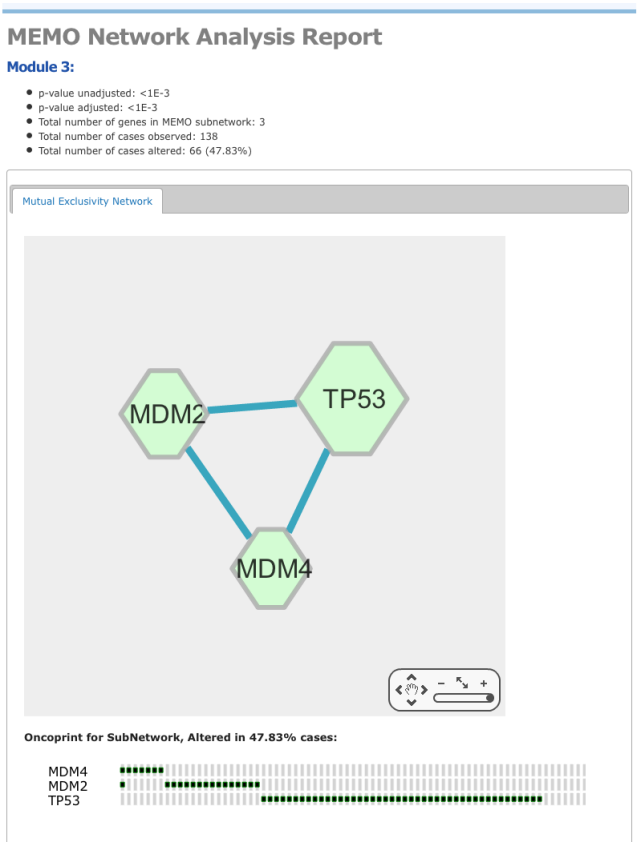

Figure 1. Example Report Generated by MEMo. Report shows one module automatically identified by MEMo in TCGA GBM.

- memo.py memo_ovarian.props

MEMo will automatically write its results to a directory called html_out and to a text file called: MemoReport.txt. To view the web report, open html_out/index.html in a web browser.

7 Running MEMo on Your Own Data

7.1 Getting Started

To run MEMo on your own data, you will need two things:

- a configuration file.
- a set of data files. These data files contain genomic profiling data, including mutations, a list of cases analyzed, results of GISTIC copy number analysis, etc. These files must be formatted correctly, in order to be processed by MEMo.

Warning: Generating these files may require a fair amount of work, as several of them are the results of other algorithms, including GISTIC [1], RAE [2] or MutSig [3]. Sample files for GBM and Ovarian are available as reference, and should be enough to get your started.

7.2 Configuration File

A sample configuration file is available in cancer_data/gbm/memo_gbm.props, and is shown below:

```
mutation_file=data_mutations_MAF.txt
cna_file=data_CNA_RAE.txt
copy_number_driven_genes_file=copy_number_driven_genes.txt
gistic_del_file=GBM_GISTIC_Del.txt
gistic_amp_file=GBM_GISTIC_Amp.txt
case_file=cases_all_three.txt
mut_sig_file=sig_genes_phase_1_2.txt
mut_sig_q_value_threshold=.05
min_number_of_alterations=4
title=TCGA Glioblastoma Multiforme, Based on Phase I and II Sequencing Data
```

Each of the configuration parameters is described in Table 1 below.

7.3 Running memo.py

To run MEMo on your own data set, you must make sure to first initialize the database (see above). To run MEMo on Mac OS X or Linux, type `memo.py` and specify your configuration file. For example:

- follow the steps above to initialize the database
- cd cancer_data/gbm
- run: memo.py memo_gbm.props

On Windows, you would type something like:

```
C:\>cd cancer_data\gbm
C:\>c:\Python31\python.exe c:\memo\bin\memo.py memo_gbm.props
```

Either way, you should then see output like this:

```
-----------------------------------------------
MEMo Network Analysis.
Computational Biology Center, MSKCC.
-----------------------------------------------
```

```
Initializing Database.  This will take a few moments...
....
```

MEMo will automatically write its results to a directory called html_out and to a text file called: MemoReport.txt. To view the web report, open html_out/index.html in a web browser.

8 Software Implementation Notes

MEMo software is written in the Java programming language. It uses Hibernate and the Java HyperSQL embedded database to store the Human Reference Networks (HRNs) and Entrez Gene information, and uses the Java JUNG library for all graph operations. Network visualization is provided by Cytoscape Web [4].

9 Questions?

If you have questions regarding MEMo, please email us at: memo AT cbio.mskcc.org.

Table 1. MEMo Configuration Parameters.

Parameter	**Description**
mutation_file	Name of the file containing mutations observed within your study. File must be formatted in the Mutation Assessment Format (MAF) file format. An example can be found in cancer_data/gbm/data_mutations_MAF.txt
cna_file	Name of the file containing discretized copy number calls within your study. This file can be auto-generated by the GISTIC [1] or RAE algorithms [2]. An example can be found in cancer_data/gbm/data_CNA_RAE.txt.
copy_number_driven_genes_file	Name of the file containing genes which pass an mRNA concordance filter test (see manuscript by Ciriello, et. el. for details). This file simply contains a list of HUGO gene symbols, one per line. An example can be found in cancer_data/gbm/copy_number_driven_genes.txt.
gistic_del_file	Name of the file containing recurrently deleted regions of interest identified by GISTIC or RAE. This file can be auto-generated by the GISTIC [1] or RAE algorithms [2]. An example can be found in cancer_data/gbm/GBM_GISTIC_Del.txt.
gistic_amp_file	Name of the file containing recurrently amplified regions of interest identified by GISTIC or RAE. This file can be auto-generated by the GISTIC [1] or RAE algorithms [2]. An example can be found in cancer_data/gbm/GBM_GISTIC_Amp.txt.
case_file	Name of the file containing identifiers for all cases or patients in your study. This file simply contains a list of case identifiers, one per line. An example can be found in cancer_data/gbm/cases_all_three.txt.
mut_sig_file	Name of the file containing results of MutSig analysis, identifying recurrently altered genes [3]. An example can be found in cancer_data/gbm/sig_genes_phase_1_2.txt.
mut_sig_q_value_threshold	Threshold q-value for the MutSig analysis. Genes with a q-value less than or equal to this threshold are included in the MEMo analysis.
min_number_of_alterations	Minimum number of alterations threshold. Only genes with the minimum number of alteration events (e.g. mutation + copy number events) are included for MEMo analysis.
title	Report title, used in the final web reports generated by MEMo.

References

1. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104: 20007-12.

2. Taylor BS, Barretina J, Socci ND, Decarolis P, Ladanyi M, et al. (2008) Functional copy-number alterations in cancer. *PLoS ONE* 3: e3179.

3. Getz G, Hofling H, Mesirov JP, Golub TR, Meyerson M, et al. (2007) Comment on "the consensus coding sequences of human breast and colorectal cancers". *Science* 317: 1500.

4. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, et al. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26: 2347-8.