TMM0.4

USER GUIDE

Lilit Nersisyan and Arsen Arakelyan,

Institute of Molecular Biology NAS, Armenia

September, 2017

TMM User Guide

Table of Contents

Introduction

There are two mechanisms that maintain the length of telomeres in healthy and cancerous tissues: telomerase-dependent and alternative (ALT) mechanisms. This project is aimed at fostering research on telomere maintenance mechanisms (TMM).

- We provide Telomerase and ALT TMM pathways that we've curated based on literature and gene expression data.
- We've created a Cytoscape app for assessment of TMM activity from gene expression data.
- You may use these pathways and our app to assess which TMM mechanism is active in your samples, if you've got gene expression data for them.
- Most importantly, we believe that many researchers may contribute their knowledge or data to improve these pathways and to deepen current understanding of molecular mechanisms leading to TMM activation.

Development

The app has been developed by Lilit Nersisyan and Arsen Arakelyan, from the Bioinformatics Group at the Institute of Molecular Biology of the National Academy of Sciences of the Republic of Armenia (IMB NAS RA).

Citation

When using the app in your research, please link to the app's webpage at http://apps.cytoscape.org/apps/tmm and cite our paper as:

The paper is under revision

License

Copyright © 2017

Lilit Nersisyan, IMB NAS RA

Arsen Arakelyan, IMB NAS RA

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License version 3. The license can be found at http://www.gnu.org/licenses/gpl.html .

Release history

Version 0.4

This is the initial beta release.

Installation

TMM has been tested to work with Cytoscape version 3.4.0 and higher on 64 bit systems.

Dependencies

Install Cytoscape

The TMM app works with Cytoscape. If you don't have it on your system proceed to Cytoscape website (http://cytoscape.org/) and follow installation instructions.

Install TMM 0.1

Go to *Apps -> App Manager ->* choose TMM from the list of available apps and click on the install button.

TMM 0.1 works with PSFC 1.1.6 (or higher) Cytoscape app as its dependency. If you don't have it, Cytoscape will install it automatically during TMM installation. In case of successful installation, the "TMM" app tab should appear in the west panel of Cytoscape.

User flow

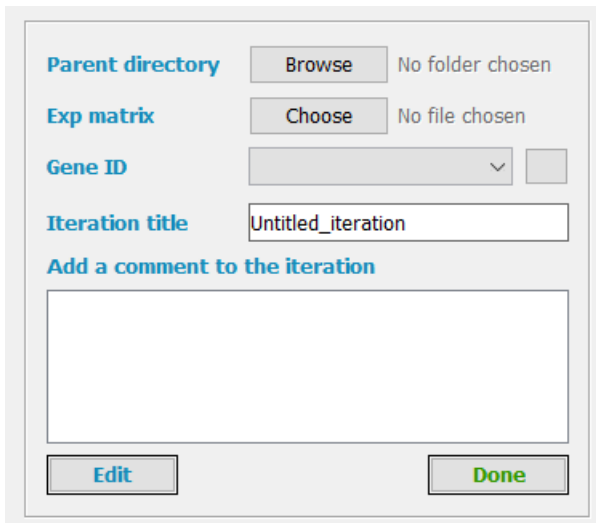## Load the network

To load a network into Cytoscape,

Go to *File -> Import -> Network -> File ->* select the TMM network you'd like to load -> click *Open*

Download the TMM network:

Literature and data curated TMM networks, as well as an example archive containing needed input files are available at: http://big.sci.am/Software/TMM#Downloads

Supply input data

In the TMM tab, the first panel is for supplying input files and directories to TMM.

Parent directory

The parent directory is the folder where you'd like the iteration directory to be created.

Click *Browse* -> in the new window select an existing directory -> click *Open*

Expression matrix

The expression matrix file contains gene expression data in a tab delimited format:

GeneIDs	Sample1	Sample2	Sample3
GeneID1	ExpG1S1	ExpG1S2	ExpG1S3
GeneID2	ExpG2S1	ExpG2S2	ExpG2S3
GeneID3	ExpG3S1	ExpG3S2	ExpG3S3

The column GeneIDs should contain gene identifiers that are also available in the Node Table of the Network loaded in Cytoscape. The networks downloaded from the TMM webpage identify genes by their

Entrez IDs (the "entrez" column in the network node table). Thus, in case of using those networks the user should have Entrez IDs in the GeneIDs column.

ExpG1S1 is the expression value for Gene 1 in sample 1.

To load the expression matrix file:

Click *Choose* -> in the new window select the expression matrix file -> click *Open*

Example gene expression files are available for download from http://big.sci.am/Software/TMM#Downloads.

Gene ID

The Gene ID specifies which column in the Cytoscape Network Node Table corresponds to gene identifiers also present in the expression matrix file. The networks downloaded from the TMM webpage identify genes by their Entrez IDs. Therefore, the user should choose the "entrez" column. The user will be notified In case if no node from the selected column is found in the first column of the gene expression matrix file.

Iteration title

The pathways may be modified by adding nodes, changing the locations of the nodes in the pathway, and deleting nodes. We name one such change an iteration. You may try different iterations to see which one performs the best in terms of prediction accuracy. The TMM app will create a separate directory for each iteration with the supplied iteration name, and will export all the files to that directory.
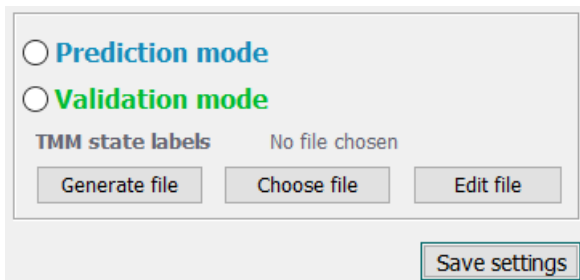
Add a comment to the iteration

You may add a description to the current iteration for further reference (*e.g.* in this iteration, the node FEN1 was removed). The comments will appear in the final report file.

Done and Edit

After submitting all the input files, click the "Done" button. The TMM app will create a directory for the iteration, will export the network file, and will create a file containing fold change values from the supplied expression matrix file. If any of the inputs is not valid, TMM will notify you. If all the files are valid, the input buttons will get disabled, and the rest of the TMM tab functionality will be enabled. To modify the input the you may click the "Edit" button.

Indicate TMM annotation labels

If the samples have been experimentally tested to detect either ALT or Telomerase activity you may supply the results to validate the pathways. In this case, press the "**Validation mode**" radiobutton, otherwise choose the "**Prediction mode**" option to simply assess the relative activity of ALT and Telomerase pathways in your samples.

In case of selecting the Validation mode option, you should supply the experimentally annotated TMM labels. You may click the "**Generate file**" button, which will create a text file in the correct format, where, however, all the samples will have ALT$^-$/Teloemerase$^-$ annotations. If your samples are ALT positive, you should change the "-" sign with "+" under the ALT column. If your samples are Telomerase positive, then those should have "-" replaced with "+" under the Telomerase column. The file is in the following tab delimited format:

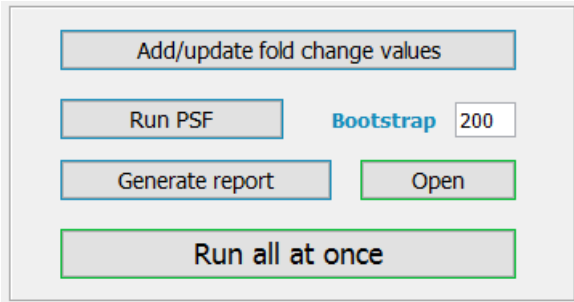| Sample name	ALT	Telomerase
ALT_sample	+	-
Telomerase_sample	-	+
Sample_with_no_active_TMM	-	-
Sample_with_both_TMMs_active	+	+

If you already have the TMM labels file, you may specify it by hitting the "**Choose file**" button. If you'd like to modify the newly generated or selected file, you may click the "**Edit file**" button, which will open the file in your system's default text file editor (usually Notepad).

Example data with TMM labels is available for download from http://big.sci.am/Software/TMM#Downloads.

Save settings

The supplied input files, folder names, comments, as well as the TMM annotation options may be saved by clicking the "**Save settings**" button, and will be reloaded next time the TMM app is launched.

Run TMM

After all the inputs are supplied, TMM does the following steps in order:

Add/update fold change values

Imports the fold change values of genes into the Network Node Table in Cytoscape. The fold change values are computed from the gene expression data. For each gene, the mean expression value is computed across the samples, and each expression value is then divided by this mean value. If the mean is 0 (in case all the samples are zero), the fold change for this gene is assigned a value of 1 for all the samples. Fold changes of 1 are also assigned in case of NA's. The generated fold change values are stored in the iteration directory with the name "fc_iterationTitle.txt". When clicking the "**Add/update fold change values**" button, these fold changes are imported into Cytoscape and can be seen in the Node Table with separate columns for each sample.

Run PSF

Performs a call to PSFC (Pathway Signal Flow Calculator) app to compute pathway PSF activity values of ALT and Telomerase pathways. The PSF algorithm computes the signal propagation from pathway inputs to the outputs (ALT and Telomerase nodes, in our case) based on fold change values and the topology of the pathway. Shortly, for each source -> target interaction, it multiplies the fold change of source to the target and assings the product to the target, if the interaction is of "activation" type. The inverse ($1/Source_{FC}$) of source fold change is multiplied by the fold change of the target node if the interaction is of type "inhibition". The interactions of type "activation" have a delta shaped edge target in the network, while the "inhibition" type has a T-shaped edge target. If the target node has multiple incoming edges, the edge signals are usually summed. Some nodes, however, represent complexes, where presence of all the subunits is necessary for it to function. In this case the minimum value of incoming signals is taken. If the at least one component is enough for the target node to function, the maximum of all the incoming signals is assigned to the target node. Please, see the citation for more details.

Click the "**Run PSF**" button to perform a call to PSFC. The PSF activity values at each node will be computed and kept in summary files in the iteration directory, as well as will be imported to the node table. Of interest are the PSF values at the "ALT" and "Telomerase" nodes, since they reflect the activity of the whole pathways. PSFC also computes significance p values for PSFs by bootstrapping. You may increase the number of bootstrap cycles to increase the power, or you may decrease the number of bootstrap cycles to make the computations faster, but at the expense of having less statistical power. The number of bootstrap cycles can be set in the "**Bootstrap**" textfield.
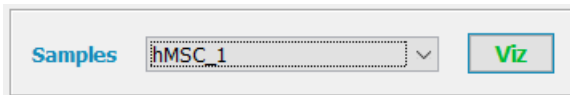
Generate report

Generates a report describing the pathway activity statistics. TMM generated two types of reports in a "report" folder residing in the iteration directory: the PDF file containing statistics and charts, and the ".xls" formatted file containing PSF values of ALT and Telomerase pathways for all the samples. The first page of the PDF report contains the input options, the iteration comment and the obtained statistics. The other pages contain volcano plots and boxplots (if TMM labels are supplied) for ALT and Telomerase pathways, and a 2D-plot where the Telomerase the ALT pathway PSF scores are on the x and y axis, respectively. If TMM labels were supplied, the TMM app also computes the SVM-based thresholds separating ALT from non-ALT and Telomerase from non-Telomerase samples and draws horizontal and vertical lines on the 2D plot. To open the report file, click the "**Open**" button.

Run all at once

All of the three steps above may be run with one click to the button "**Run all at once**".

Visualize PSF results

The PSF values are visualized in Cytoscape with node colors on a red-white-blue scale, where red and blue represent high and low PSF values, and white represents PSF values equal to 1. The sample is chosen with the drop-down menu, and visualized by clicking the "**Viz**" button.

The Telomere Maintenance Mechanisms project

For more information, visit the project webpage: http://big.sci.am/software/tmm/.

© 2017, Lilit Nersisyan and Arsen Arakelyan.

Licensed under: GNU General Public License version 3.

Correspondence:

Lilit Nersisyan,

l_nersisyan@mb.sci.am

Group of Bioinformatics,

Institute of Molecular Biology NAS RA

7 Hasratyan str. 0014, Yerevan, Armenia

September, 2017