User Manual of TSMiner

Contents

1 Introduction

TSMiner is a software program for reconstructing time-specific regulatory networks from time-series expression profiles. TSMiner has three key modules: first, predicting the time-specific activated or repressed transcription factors (TFs); second, predicting the biological pathways interacted with the predicted TFs; third, merging the TFs and pathways into time-specific regulatory networks. TSMiner provides extensive visualization methods to help users explore the results. For example, in addition to made use of the tree structure map developed by the Dynamic Regulatory Events Miner (DREM) (Ernst et al. 2007), we supply a series of lists to show the TFs and pathways identified for different time points, and supply network maps and heat maps with extensive interactive options to help users explore the TF-pathway regulatory networks.

TSMiner contains one main interface and multiple display interface. The main interface is designed to input data, set parameters and show the model learning process. While the display interface aims to help users explore the interested TFs, pathways, and network maps. Users should agree to the DREM, Piccolo, and Batik licenses before using the TSMiner software. The source code and the complied tool can be downloaded at https://sourceforge.net/projects/tsminer-tool/.

2 Preliminaries

- To use TSMiner a version of Java 1.5 or later must be installed. If Java 1.5 or later is not currently installed, then it can be downloaded from http://www.java.com.
- TSMiner can be executed from a command line change to the TSMiner directory and then type: java -mx1024M -jar TSMiner.jar. If TSMiner reports “Out of Memory Error”, users can increase the -mx parameter.

3 Main Interface

The main interface is divided into three sections: Load Data, Set Parameters, and Search Model.

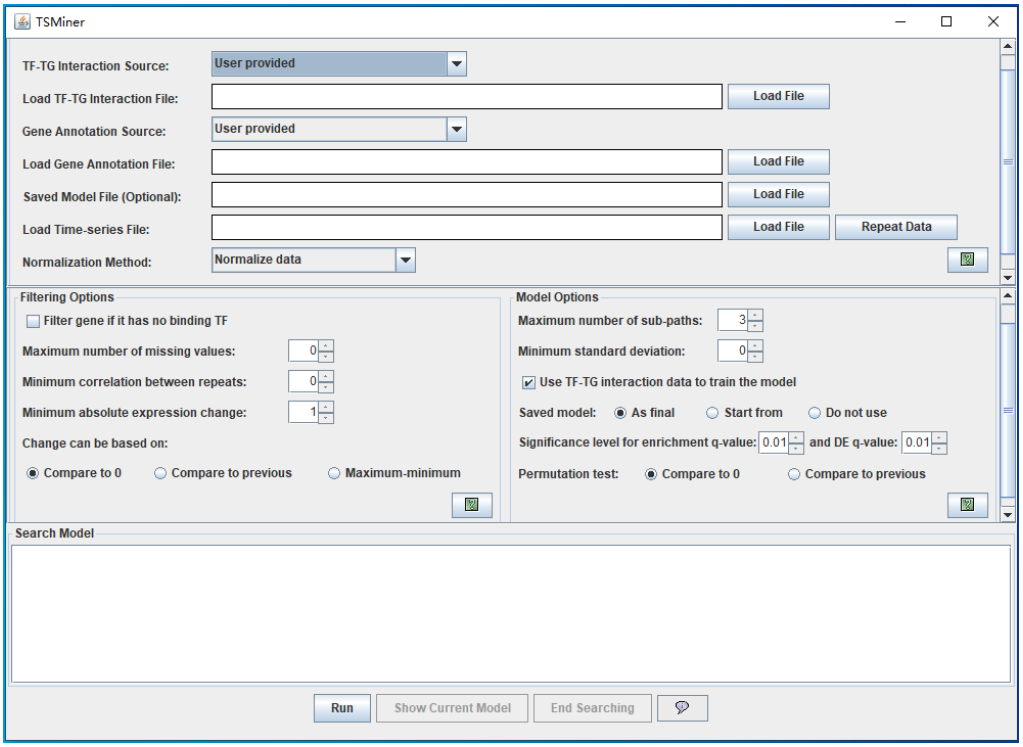


Figure 1. An image of the main input interface of TSMiner. This is the first screen that appears when TSMiner is launched. From this screen a user input the data sets and various execution options. Pressing the "Run" button at the bottom of the interface starts the model learning.

3.1 Load Data

In the "Load Data" section, users can specify the file of TF-target (TG) interaction data (necessary), time-series expression data (necessary), pathway gene sets (optional), and saved model (optional), and choose the normalization method for the time-series data.

3.1.1 "TF-TG interaction Source" and "Load TF-TG interaction File"

Predictions of TF-TG regulatory interactions are input. The source of these predictions can either be user provided or one of the files that currently is present in the "TF-TG Interaction Source" directory of the TSMiner directory. The TF-TG Interaction files provided are described in Appendix A. If User Provided" is selected, then the "Load TF-gene File" field is editable and a user can select the file directory. This regulatory information could come from Chromatin Immunoprecipitation (ChIP) experiments, TF binding site motif information, or databases using other predicting methods. The TF- gene regulation input does not need to be associated with specific time points. Otherwise the "Load TF-gene File" field is not editable. The format of the TF-gene interaction file is the same as that in DREM, which can either be an ASCII text file or a GNU zip file of an ASCII text file. The TF-gene interaction file is necessary for running TSMiner. TSMiner will report errors when the field is null or a wrong directory. The file can be in one of two formats, a grid format or a three column format.

In the grid format the columns correspond to the transcription factors, and the rows correspond to the genes. The first column contains gene symbols. The first row contains the gene symbol column header followed by the names of each transcription factor all delimited by tabs. An entry of 1 corresponds to the prediction that the transcription factor does regulate the gene. While not used in the provided files it also possible to differentiate between positive and negative interactions by using a "1" for positive correlation and "-1" for negative interactions. (Fig. 2)

ID	Sin3a	Nr1i3	Ctnnb1	Cebpb	Cebpa	Ets2	Hif1a
Gamt	0	0	0	0	0	0	0
Adora2b	0	0	0	0	0	0	0
Erbb2	0	0	1	0	0	0	0
Psme3	0	0	0	0	0	0	0
Dqx1	0	1	0	0	0	0	0
Erbb3	0	1	0	0	0	0	0
6330439K1	0	0	0	0	0	0	0
Psme2	0	0	0	0	0	0	0
Luc7l	1	0	0	0	0	0	0
Smc6	0	0	0	0	0	0	1
Vps35	0	0	1	0	0	0	1
Smc3	0	0	0	0	0	0	0
Elk3	0	0	0	0	0	0	0
Smc2	1	0	0	0	0	0	0
Elk4	0	0	0	0	0	0	0
Ipw	0	0	0	0	0	0	0

Figure 2. A sample of TF-gene data file in grid format when viewed in Microsoft Excel.

In the three column format the first column contains the transcription factors, the second column the regulated gene, and the third column input value. The first row is a header row where the header of the first column must be "TF" column, and the second column must have the header "Gene". A value of "1" represents that the TF-TG pair shows positive correlation interaction, while a value of "-1" represents that the TF-TG pair shows negative correlation interaction. (Fig. 3)

TF	Gene	Input
Nr1i3	Otc	1
Creb3l3	Leap2	1
Nr1i3	Ugt2a3	1
Nr1i3	Leap2	1
Nr1i3	Rdh7	1
Creb3l3	Apoa4	1
Nr1i3	Cyp3a25	1
Nr1i3	Cps1	1
Nr1i3	Hsd17b6	1

Figure 3. A sample of TF-gene data file in three column format when viewed in Microsoft Excel.

3.1.2 "Gene Annotation Source" and "Load Gene Annotation File"

Gene annotation file is optional, which is used to input information for biological pathway gene sets. It can be got from the KEGG or Reactome Pathway Database. The source and format of the gene annotation file are the same as the TF-gene interaction file. The gene annotation files provided by TSMiner are described in Appendix B.

3.1.3 Saved Model File

The "Saved Model File" field allows a user to specify a file containing a saved model, thus saving time if the model has already been computed. A saved model file can also be used to initialize from where the search for a model starts. The option controlling how the saved model file is used is determined by the

"Saved Model" option on the Search Options panel described in Section 3.2.2.

3.1.4 Load Time-series File

The "Load Time-series File" field is used to input an expression data file. The first column is gene symbols, and the remaining columns contain the expression values at each time point ordered sequentially based on time. If an expression value is missing, then the field should be left empty. The first row of the data contains column headers. A sample expression data file is shown in Figure 4.

Symbol	0h	0.5h	2h	4h	24h	48h	72h	120h	168h
0610005C1	115.5467	116.1001	80.15296	37.37271	113.9373	78.24521	89.57055	110.9746	110.5581
0610007P1	50.50665	63.44209	45.10752	32.9013	46.76262	47.64302	71.33584	57.91616	70.64555
0610008F0	3.605273	2.627928	3.095434	1.797209	5.265093	1.207709	1.570786	2.678795	3.452512
0610009B2	12.37059	13.38686	12.0852	9.413711	13.58073	12.55957	11.66662	12.53183	12.44654
0610009D0	20.76723	25.93396	22.5382	26.06879	22.54145	23.18622	20.64051	23.21606	24.63465
0610009O2	11.22965	12.84	8.172621	7.076712	12.14311	19.25495	16.03805	13.17682	13.9542
0610010F0	3.352192	2.270783	2.463658	3.149141	3.252627	2.329017	2.593569	1.737018	1.421042
0610010K1	13.86787	17.29463	11.32233	12.1788	16.63747	22.92486	21.2324	16.83939	21.30328
0610011F0	84.03353	91.79236	71.24632	50.19096	59.43466	69.8669	58.35006	71.91464	83.97706
0610012G0	9.546221	11.4466	7.436447	6.932824	10.34509	11.19815	11.40282	11.27216	11.58902
0610012H0	12.84113	9.557958	8.717559	9.570791	18.86755	8.130192	10.28626	14.67633	15.3093
0610030E2	7.047176	5.567177	4.058102	3.558776	6.86119	5.060256	6.512457	6.498618	6.542187
0610031J0	49.53234	54.52473	39.88724	39.68267	43.53963	54.15857	43.34826	54.85014	52.63077
0610031O1	12.54111	14.68027	20.64164	15.23505	8.854553	5.647944	8.021064	8.858592	10.14824
0610037L1	9.481928	12.84251	9.947504	9.47527	12.48761	12.70276	12.09249	11.50828	11.12271
0610040J0	17.83149	18.03284	16.44293	9.91072	18.49357	17.44603	20.21583	16.40713	16.39479

Figure 4. A sample of time-series data file when viewed in Microsoft Excel.

In addition, the input time-series data can optionally show spot IDs at the first column. Spot IDs uniquely identify an entry in the data file, and if they are not included in the data file, then they will be automatically generated. While spot IDs must be unique, the same gene symbol may appear multiple times in the data file corresponding to the same gene appearing on multiple spots on the array. A sample expression data file with Spot IDs is shown in Figure 5.

Spot	Symbol	0h	0.5h	2h	4h	24h	48h	72h	120h	168h
1	0610005C1	115.5467	116.1001	80.15296	37.37271	113.9373	78.24521	89.57055	110.9746	110.5581
2	0610007P1	50.50665	63.44209	45.10752	32.9013	46.76262	47.64302	71.33584	57.91616	70.64555
3	0610008F0	3.605273	2.627928	3.095434	1.797209	5.265093	1.207709	1.570786	2.678795	3.452512
4	0610009B2	12.37059	13.38686	12.0852	9.413711	13.58073	12.55957	11.66662	12.53183	12.44654
5	0610009D0	20.76723	25.93396	22.5382	26.06879	22.54145	23.18622	20.64051	23.21606	24.63465
6	0610009O2	11.22965	12.84	8.172621	7.076712	12.14311	19.25495	16.03805	13.17682	13.9542
7	0610010F0	3.352192	2.270783	2.463658	3.149141	3.252627	2.329017	2.593569	1.737018	1.421042
8	0610010K1	13.86787	17.29463	11.32233	12.1788	16.63747	22.92486	21.2324	16.83939	21.30328
9	0610011F0	84.03353	91.79236	71.24632	50.19096	59.43466	69.8669	58.35006	71.91464	83.97706
10	0610012G0	9.546221	11.4466	7.436447	6.932824	10.34509	11.19815	11.40282	11.27216	11.58902
11	0610012H0	12.84113	9.557958	8.717559	9.570791	18.86755	8.130192	10.28626	14.67633	15.3093
12	0610030E2	7.047176	5.567177	4.058102	3.558776	6.86119	5.060256	6.512457	6.498618	6.542187
13	0610031J0	49.53234	54.52473	39.88724	39.68267	43.53963	54.15857	43.34826	54.85014	52.63077
14	0610031O1	12.54111	14.68027	20.64164	15.23505	8.854553	5.647944	8.021064	8.858592	10.14824
15	0610037L1	9.481928	12.84251	9.947504	9.47527	12.48761	12.70276	12.09249	11.50828	11.12271

Figure 5. A sample of time-series data file when viewed in Microsoft Excel.

Pressing the Repeat Data button brings up an interface as shown in Figure 6. Repeat data files must have the same format as the single time-series data file, including the same number of rows and columns. Repeat data values will be the median of the values from the single data file using. Repeat data can be selected to be from either "Different time periods" or "The same time period". If the data is from Different time periods then data was collected over multiple distinct time-series, but presumably at the same sampling rate. If the data is from the same time period, then this implies multiple measurements were collected at each time point during one time-series. If the repeat data is from Different time periods, the repeat data will be averaged after normalization, while if the repeat data is from The Same Time

Period the repeat data will be averaged before normalization. In the case the repeat data is from Different time periods, the repeat data can be used to filter genes with inconsistent expression patterns as explained in Section 3.2.1.

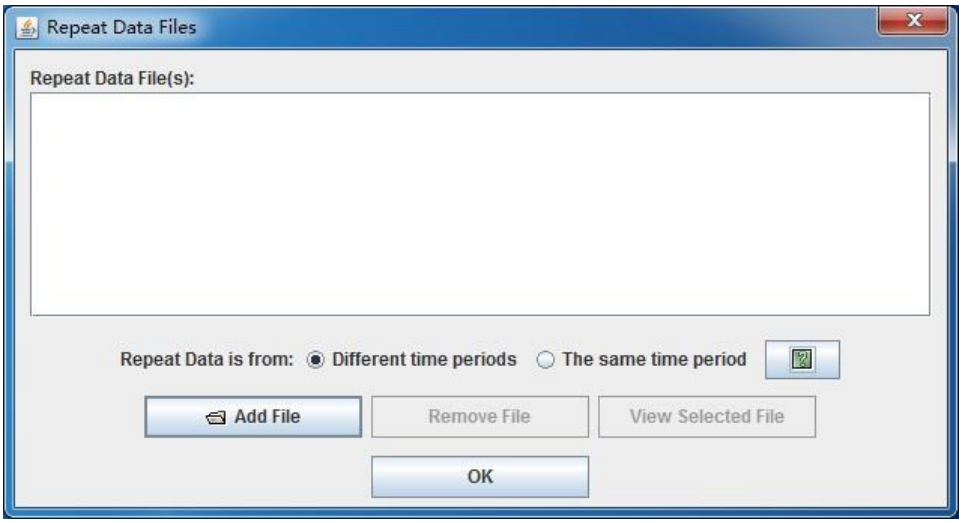


Figure 6. The above window is used to specify repeat data files.

3.1.5 Normalization method

Before time-series data is analyzed, it must be transformed to start at 0. The normalization method is similar to that in DREM, which can be selected to be of one of three types: Log normalize data, Normalize data, or No normalization/add 0. Given a time-series vector of gene expression values $\{v_0, v_1, \dots, v_n\}$ the transformations are as follows:

- Log normalize data – transforms the vector to $\{0, \log_2(v_1/v_0), \dots, \log_2(v_n/v_{n-1})\}$
- Normalize data – transforms the vector to $\{0, (v_1 - v_0), \dots, (v_n - v_{n-1})\}$.
- No normalization/add 0 – transforms the vector to $\{0, v_0, v_1, \dots, v_n\}$.

Time point 0 usually corresponds to a control before the experimental conditions were applied. If the input data file contains raw expression values, then the Log normalize data option should be selected. If the input data file has been computed the log ratios, then the Normalize data option should be selected. If the input data file already contains log ratio data against a control, but no time point 0 experiment was conducted, then the No normalization/add 0 option should be selected.

3.2 Set Parameters

The "Set Parameters" section is divided into two panels: "Filtering Options" and "Model Options", which are discussed in the next subsections.

3.2.1 Filtering Options

Through the parameters on the Filtering panel a user can adjust the criteria for filtering genes. If a gene is filtered, then it will be excluded from further analysis. Genes can be filtered if they do not show a sufficient response to experimental conditions (Minimum Absolute Expression Change), there are too many missing values (Maximum Number of Missing Values), or the gene expression pattern over repeats is too inconsistent (Minimum Correlation between Repeats). A gene can also be filtered if it does appear in the TF-gene interaction input file. If the "Log normalize data" or "Normalize data" options are selected, a gene will automatically be filtered if its expression value at the first time point is missing.

- Filter gene if it has no binding TF – if this box is checked then genes are filtered if they are not

included in the TF-gene interaction file. If this box is unchecked then genes not included in the TF-gene regulation input, are not filtered and are assumed to have a "0" for every entry of the TF- gene regulation predictions.

- Maximum Number of Missing Values – a gene will be filtered if the number of missing values in its time-series exceeds this parameter.
- Minimum Correlation between Repeats – this parameter is only applied if the repeat data files are input. If there are two repeat files, a gene will be filtered if its correlation between the two data sets is below this parameter. If there are more than two repeat files, then the gene will be filtered if the mean of all its pairwise correlations between experiments is below this parameter.
- Minimum Absolute Expression Change – after data transformation (Log normalized data, Normalized data, or No Normalization/add 0), if the absolute value of the gene's expression change is below this threshold, then the gene will be filtered. The expression change can be defined based on three methods (see below).
- Change can be based on – this parameter defines the expression change for gene filtering. If "Maximum-Minimum" option is selected a gene will be filtered if the maximum absolute difference between the values of any two time points is less than the value. If "Compare to 0" is selected a gene will be filtered if the absolute expression changes from all time points to time point 0 is less than the value. If "Compare to Previous" is selected a gene will be filtered if the absolute expression change between each time point I and its previous time point less than the value. If these options are multi-selected, a gene will be filtered if its expression changes is lower than all the selected standards.

3.2.2 Model Options

- Maximum number of sub-paths – this parameter controls the maximum number of sub-paths allowed out of a bifurcation node.
- Minimum Standard Deviation – this parameter controls the minimum standard deviation on the Gaussian distributions. Increasing this parameter is recommended if applying TSMiner to RNA-seq data to avoid potential overfitting of low variance in expression due to the small discrete counts.
- Use TF-gene interaction data to train the model – this option and its application are referred from DREM (Ernst et al. 2007). If this box is checked then the TF-gene interaction data is used jointly with the time-series data to infer the model and then assign genes to the sub-paths. If this box is unchecked then the time-series data alone is used to infer a model, and the TF-gene interaction predictions are only used in the enrichment analysis and differential expression analysis. Using the TF-gene interaction data to infer the model gives a more biologically coherent model. The model learning is faster when using the TF-gene information only as a post-processing step.
- Saved Model – this option is only relevant if a file is specified under Saved Model File. If the parameter is set to "As final" the model in the Saved Model File is used as the final one. If the parameter is set to "Start from", TSMiner will start from the saved model. If the parameter is set to "Do not use" then TSMiner will ignore the Saved Model File field and start a new search.
- Significance level of enrichment q-value and DE q-value – the enrichment q-value specifies the threshold below which the TGs binding to a TF are identified as significantly enriched in a sub-path

out of a bifurcation point. While the DE q-value specifies the threshold below which the TGs binding to a TF are seen as significantly up or down-regulated.

- Permutation Test – after the significantly enriched TFs of each sub-path for all splits are identified, TSMiner will further study the expression changes of their regulating genes for all the time points from their enriched splits to the last time points. This parameter is used to define the comparison set in the permutation test. If “Compare to Previous” is selected, the permutation test is used to study the expression changes of the regulated genes for the time segments. If “Compare to Time 0” is selected, the permutation test is used to study the expression changes of the regulated genes between the time points and the control group.

3.3 Search Model

The text box here displays the running progress of TSMiner after pressing the “Run” button. And there are other two buttons controlling the model learning. The buttons are the “Show Current Model” and “End Searching”. Pressing the “Show Current Model” displays the current best map TSMiner has found so far, but does not end the search. Pressing the “End Searching” Button forces TSMiner to end adding paths. TSMiner then proceeds to the second phase of its search where it considers deleting paths and delaying paths.

4 Display Interface

After the model learning executes, the main output window appears. The main window displays the time-series of all the genes that were not filtered overlaid with a tree structure. An example of such a window is shown in Figure 7. The map features the paths and splits in the time-series data and genes are assigned to paths through the model. All path segments and splits are annotated with associated TFs and gene annotations. Each node is associated with a Gaussian distribution. The Gaussian distribution associated with the node determines its y-axis location on the map. The area of a node is proportional to the Gaussian's standard deviation. Green nodes represent split nodes, these are nodes for which multiple paths exit the node.

Left clicking on an edge displays only genes assigned to a path going through that edge. For instance, Figure 8 shows the interface after clicking on the orange edge of Figure 7. Left clicking on a green split node displays all genes passing through the split node. The genes will be colored based on the colors of the sub-edges to which they were assigned out of the split node. Right clicking on a sub- edge out of a split node brings up a “Path Table” as described in Section 4.1. Right clicking on a node which is not a split node or an edge which is not a sub-edge out of a split, it showed an “Edge Table” as described in Section 4.2. Right clicking on a split node brings up a “Split Table” as described in Section 4.3. Holding the mouse over a specific gene expression plot displays the name of the gene.

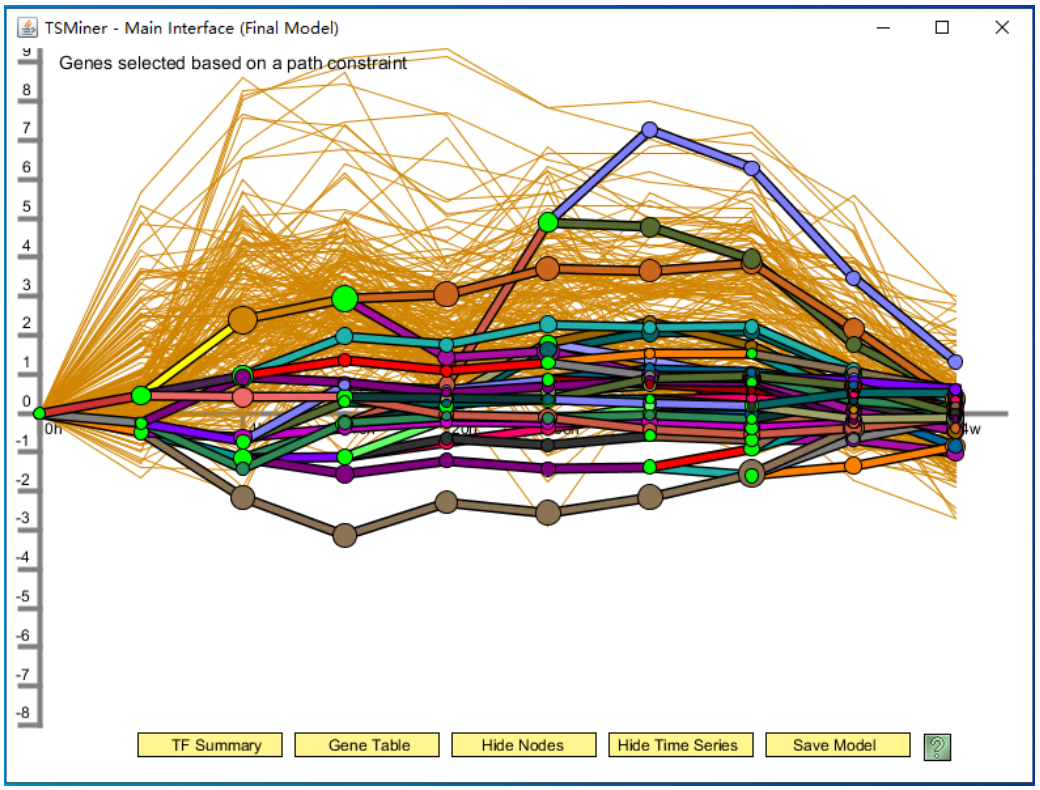


Figure 7. An example of the Display Interface. The interface has a tree structure overlaid on top of the time-series expression profiles. Along the bottom are buttons with various options.

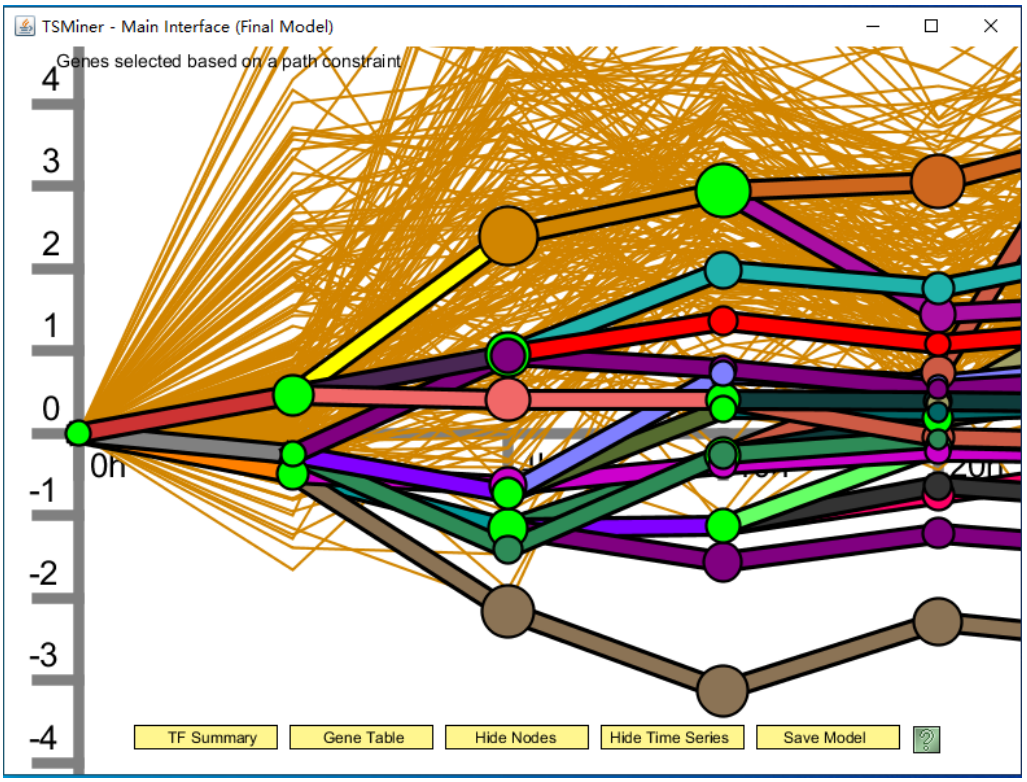


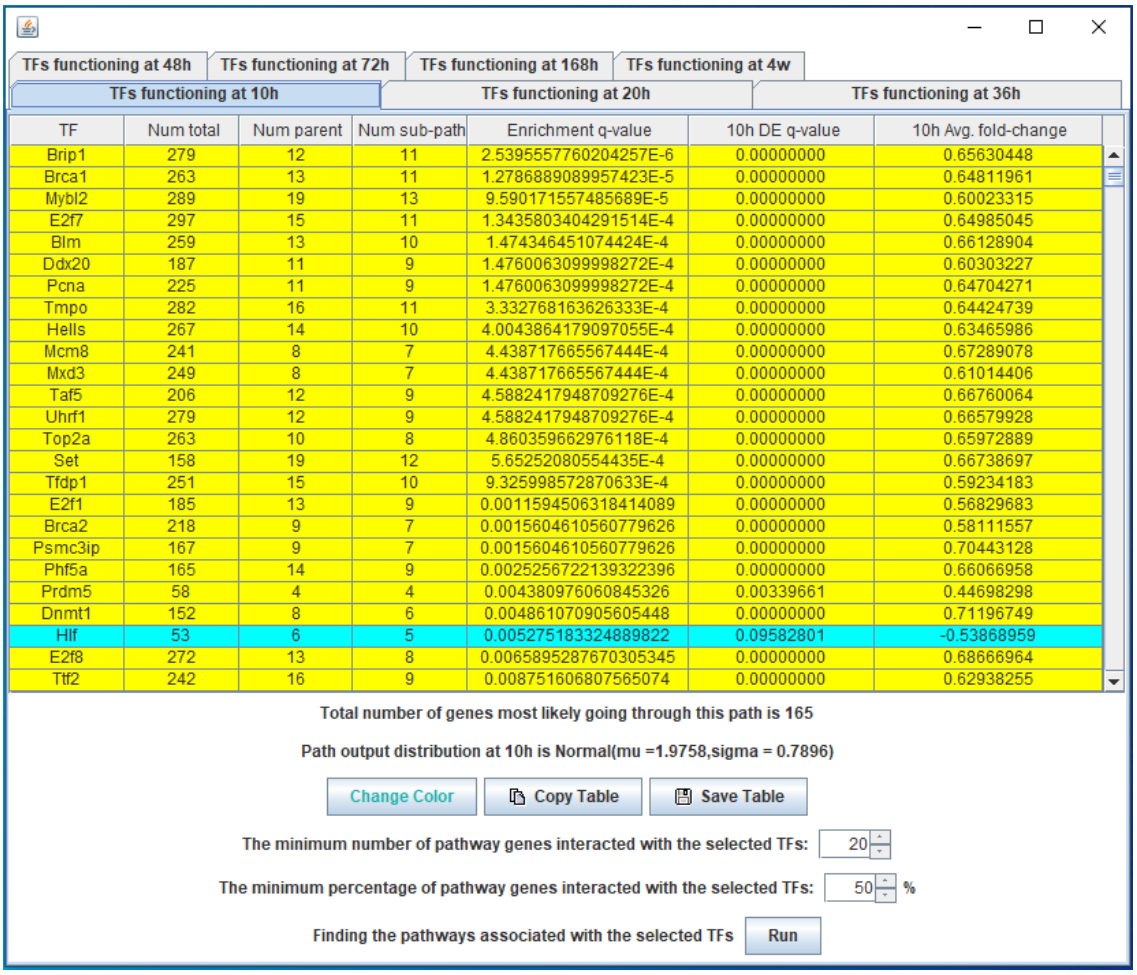
Figure 8. The Display interface window when hold the right mouse button down and move the mouse. When hold the left mouse button down and move the mouse users can move the window.

4.1 TF Table

A series of TF tables such as in Figure 9 appears when right clicking on a sub-path out of a bifurcation node. Each of them shows the TFs activated or repressed at specific time point from the bifurcation time point to the last one. While one of these tables such as in Figure 10 appears when right clicking on a

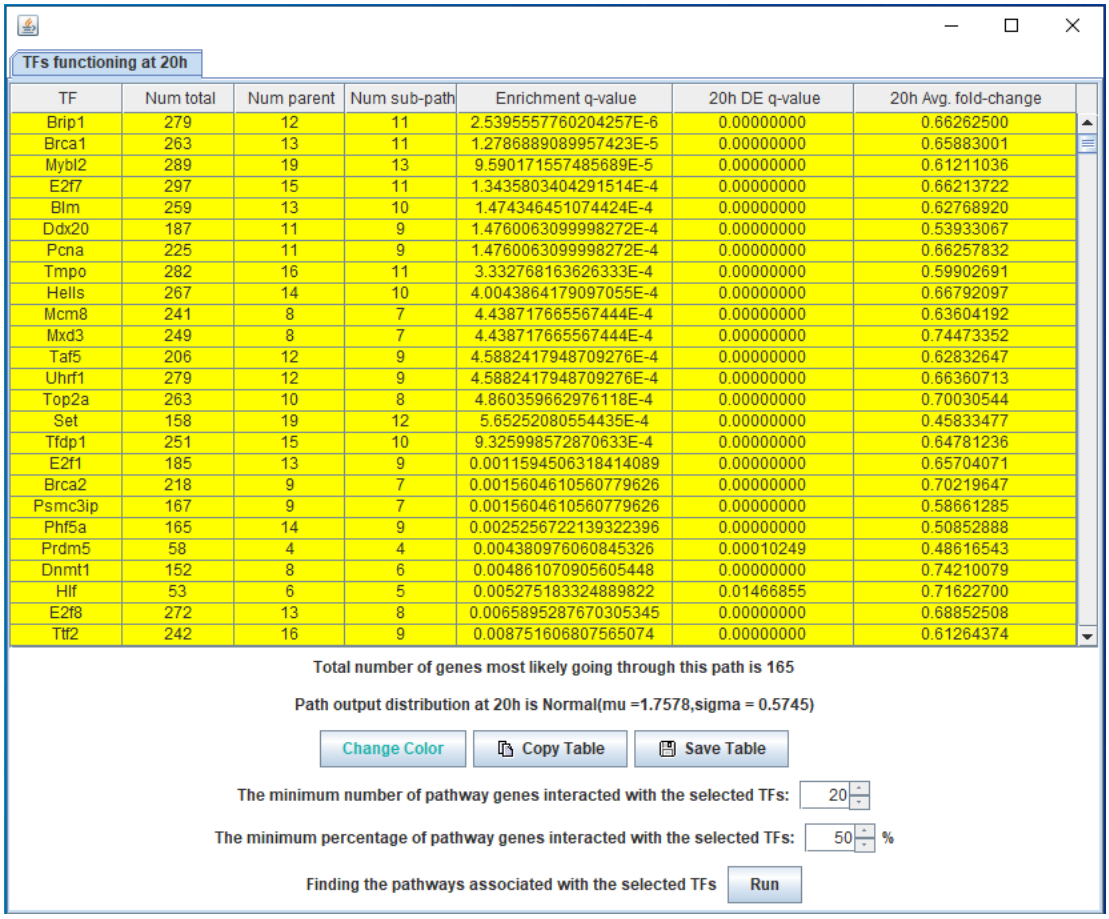
node which is not a split node or an edge whose left node is not a bifurcation node.

- TF – the name of the transcription factor.
- Num Total – the total number of genes in the expression data regulated by the TF.
- Num Parent – the number of genes going into the split node on the left of this path regulated by the TF.
- Num sub-path – the number of genes regulated by the TF assigned to this sub-path.
- Enrichment q-value – the enrichment significance of the targets binding to a TF to this sub-path.
- DE q-value – the differential expression significance of the targets binding to a TF for a specific time point.
- Avg. fold-change – the average fold-changes of the targets binding to a TF comparing a specific time point to the previous one (the option "Permutation test" in the main interface is set as "Compare to previous") or to the first one (the option "Permutation test" in the main interface is set as "Compare to 0"). This value is used to determine whether the TF activate or repress its targets. When a TF positively regulate the targets (interaction score is "1" in the TF-TG interaction data), it is activated when the average fold-change of the targets is higher than zero, and repressed when the average fold-change of the targets is lower than zero. While it is opposite for a TF negatively regulate the targets (interaction score is "-1" in the TF-TG interaction data). The activated and repressed TFs are displayed in yellow and blue background, respectively.
- Infer pathways – users can select one or all items in the TF table and infer the pathways interacted with the selected TFs. We calculate the sum of pathway genes directly and indirectly interacted with the selected TFs (TGs+PGs) as S, one pathway is identified when S is higher than the parameter "The minimum number of pathway genes interacted with the selected TFs (TGs+PGs)", while S divided by the pathway gene number is higher than the parameter "The minimum persentage of pathway genes interacted with the selected TFs (TGs+PGs)".



TF	Num total	Num parent	Num sub-path	Enrichment q-value	10h DE q-value	10h Avg. fold-change
Brip1	279	12	11	2.5395557760204257E-6	0.00000000	0.65630448
Brca1	263	13	11	1.2786889089957423E-5	0.00000000	0.64811961
Mybl2	289	19	13	9.590171557485689E-5	0.00000000	0.60023315
E2f7	297	15	11	1.3435803404291514E-4	0.00000000	0.64985045
Blm	259	13	10	1.474346451074424E-4	0.00000000	0.66128904
Ddx20	187	11	9	1.4760063099998272E-4	0.00000000	0.60303227
Pcna	225	11	9	1.4760063099998272E-4	0.00000000	0.64704271
Tmpo	282	16	11	3.332768163626333E-4	0.00000000	0.64424739
Hells	267	14	10	4.0043864179097055E-4	0.00000000	0.63465986
Mcm8	241	8	7	4.438717665567444E-4	0.00000000	0.67289078
Mxd3	249	8	7	4.438717665567444E-4	0.00000000	0.61014406
Taf5	206	12	9	4.5882417948709276E-4	0.00000000	0.66760064
Uhrf1	279	12	9	4.5882417948709276E-4	0.00000000	0.66579928
Top2a	263	10	8	4.860359662976118E-4	0.00000000	0.65972889
Set	158	19	12	5.65252080554435E-4	0.00000000	0.66738697
Tfdp1	251	15	10	9.325998572870633E-4	0.00000000	0.59234183
E2f1	185	13	9	0.0011594506318414089	0.00000000	0.56829683
Brca2	218	9	7	0.0015604610560779626	0.00000000	0.58111557
Psmc3ip	167	9	7	0.0015604610560779626	0.00000000	0.70443128
Phf5a	165	14	9	0.0025256722139322396	0.00000000	0.66066958
Prdm5	58	4	4	0.004380976060845326	0.00339661	0.44698298
Dnmt1	152	8	6	0.004861070905605448	0.00000000	0.71196749
Hlf	53	6	5	0.005275183324889822	0.09582801	-0.53868959
E2f8	272	13	8	0.0065895287670305345	0.00000000	0.68666964
Ttf2	242	16	9	0.008751606807565074	0.00000000	0.62938255

Figure 9. An example of multiple TF tables.



TF	Num total	Num parent	Num sub-path	Enrichment q-value	20h DE q-value	20h Avg. fold-change
Brip1	279	12	11	2.5395557760204257E-6	0.00000000	0.66262500
Brca1	263	13	11	1.2786889089957423E-5	0.00000000	0.65883001
Mybl2	289	19	13	9.590171557485689E-5	0.00000000	0.61211036
E2f7	297	15	11	1.3435803404291514E-4	0.00000000	0.66213722
Blm	259	13	10	1.474346451074424E-4	0.00000000	0.62768920
Ddx20	187	11	9	1.4760063099998272E-4	0.00000000	0.53933067
Pcna	225	11	9	1.4760063099998272E-4	0.00000000	0.66257832
Tmpo	282	16	11	3.332768163626333E-4	0.00000000	0.59902691
Hells	267	14	10	4.0043864179097055E-4	0.00000000	0.66792097
Mcm8	241	8	7	4.438717665567444E-4	0.00000000	0.63604192
Mxd3	249	8	7	4.438717665567444E-4	0.00000000	0.74473352
Taf5	206	12	9	4.5882417948709276E-4	0.00000000	0.62832647
Uhrf1	279	12	9	4.5882417948709276E-4	0.00000000	0.66360713
Top2a	263	10	8	4.860359662976118E-4	0.00000000	0.70030544
Set	158	19	12	5.65252080554435E-4	0.00000000	0.45833477
Tfdp1	251	15	10	9.325998572870633E-4	0.00000000	0.64781236
E2f1	185	13	9	0.0011594506318414089	0.00000000	0.65704071
Brca2	218	9	7	0.0015604610560779626	0.00000000	0.70219647
Psmc3ip	167	9	7	0.0015604610560779626	0.00000000	0.58661285
Phf5a	165	14	9	0.0025256722139322396	0.00000000	0.50852888
Prdm5	58	4	4	0.004380976060845326	0.00010249	0.48616543
Dnmt1	152	8	6	0.004861070905605448	0.00000000	0.74210079
Hlf	53	6	5	0.005275183324889822	0.01466855	0.71622700
E2f8	272	13	8	0.0065895287670305345	0.00000000	0.68852508
Ttf2	242	16	9	0.008751606807565074	0.00000000	0.61264374

Figure 10. An example of one Table.

4.2 Summary TF Table

A series of TF tables that summarize the TFs activated or repressed at time points. There are two tables that display the activated and repressed TFs for each time point, respectively. The activated TFs are displayed in yellow background (Figure 11), while the repressed ones are displayed in blue background (Figure 12). The column "TF", "Enrichment q-value", "DE q-value", and "Avg. fold-change", together with the procedures for inferring pathways are the same as those of TF Tables.

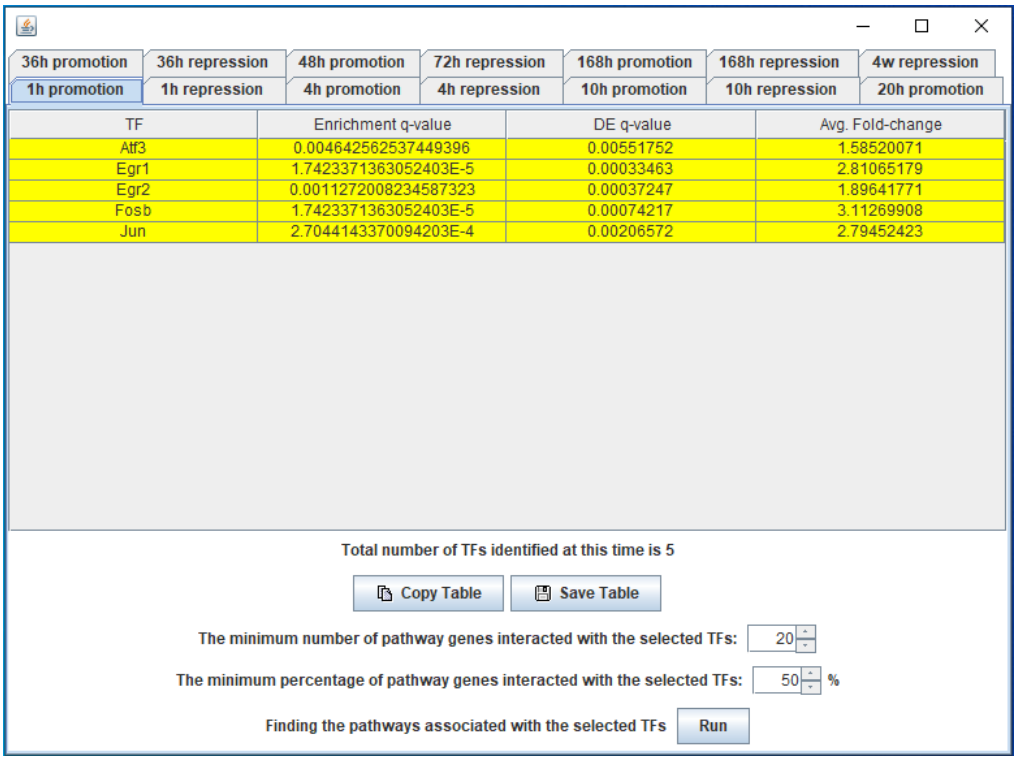


Figure 11. A summary TF table for the TFs activated at 1 h.

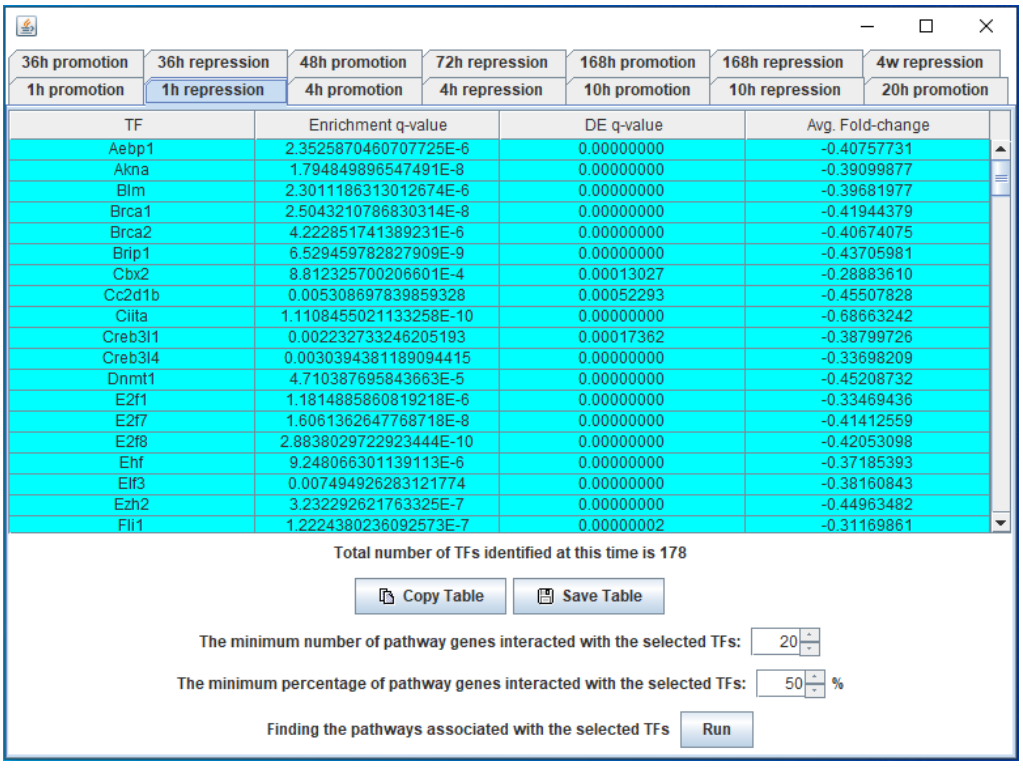


Figure 12. A summary TF table for the TFs repressed at 1 h.

4.3 Pathway Table

Pathway tables displays the pathways interacted with the time-specific TFs. A pathway table appears after users select one or more TFs from the "TF Table" and click on the "Run" button. Figure 13 and 14 show two pathway tables, in which the pathways are activated and repressed at 1 h, respectively.

- Pathway – the name of biological pathways;
- Num TF – the number of the selected TFs that involved in the pathway;
- Num TG – the number of genes that included in the pathway and directly binding to the selected TFs;
- Num PG – the number of genes that included in the pathway and interacted with the TF-TG cascading with the selected TFs;
- Num Total – the sum of "Num TF", "Num TG", and "Num PG";
- Percent – the percentage of "Num Total" among the genes included in pathway.
- Permutation test P – the differential expression significance of the genes included in a pathway for a specific time point.
- Permutation test FDR –The permutation test p-value were transformed to FDR by Benjamini–Hochberg correction.
- Avg.FC –the average fold-changes of the genes included in a pathway comparing a specific time point to the previous one (the option "Permutation test" in the main interface is set as "Compare to previous") or to the first one (the option "Permutation test" in the main interface is set as "Compare to 0").

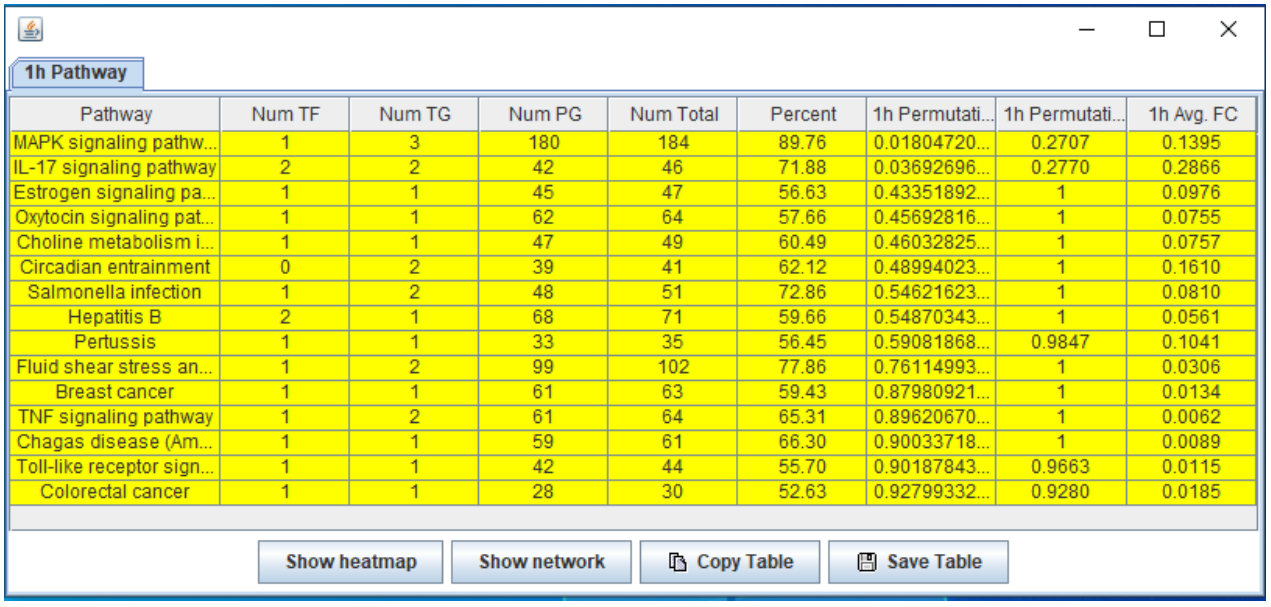


Figure 13. An example of pathways interacting with the TFs activated at 1 h.

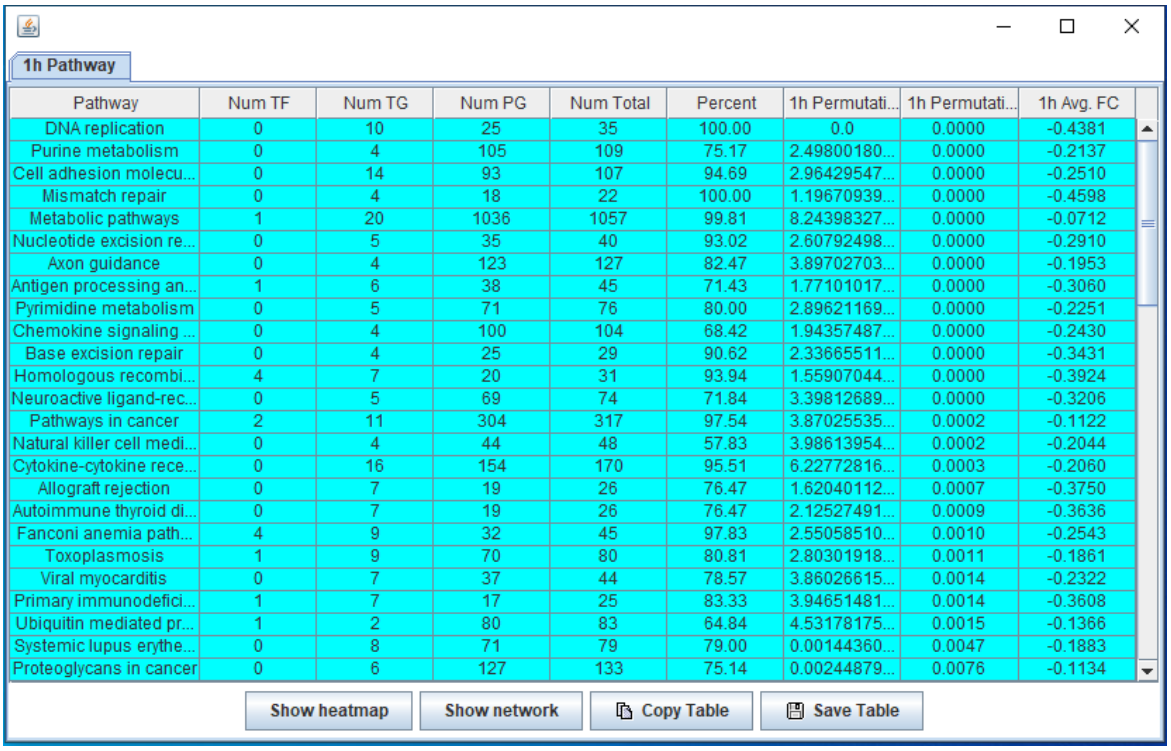


Figure 14. An example of pathways interacting with the TFs repressed at 1 h.

4.4 Heat Map

When users select one or more interested pathways in the "Pathway Table" and click on the "Show heatmap" button, TSMiner provides heat maps to show the expression levels of three sets of genes: (see Figure 15)

- TFs – the TFs interacted with the user-selected pathways.
- TGs – the genes binding to the above TFs and included in the user-selected pathways.
- PGs – the genes included in the user-selected pathways and interacted with the above TF-TG regulatory pairs.

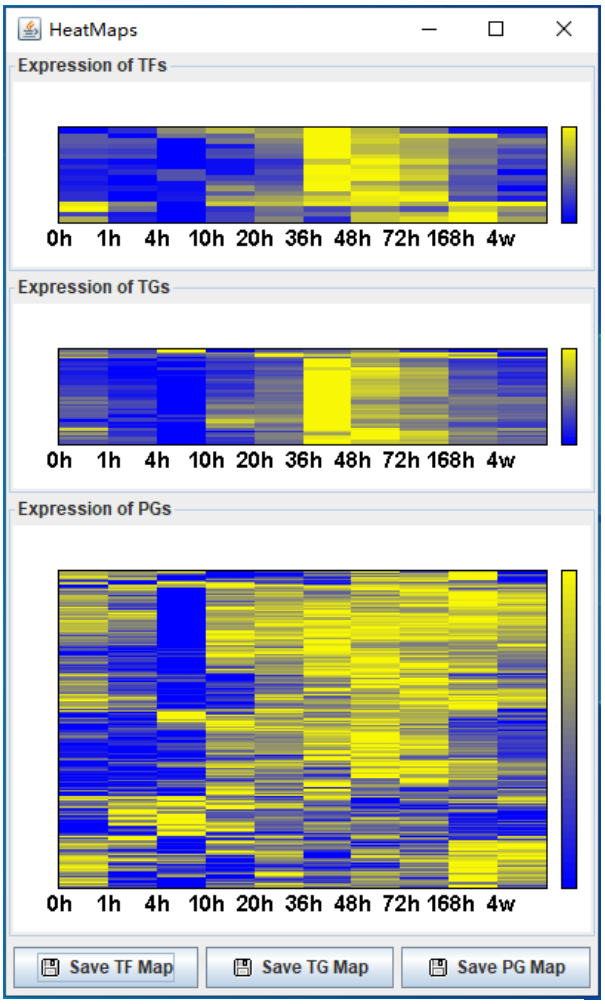


Figure 15. An example of heat maps.

4.5 Network Map

When users select one or more pathways in the “Pathway Table” and click on the “Show network” button, TSMiner provides a network map with multiple interactive visualization operations (see Figure 16), in which the TFs interacted with the user-selected pathways are linked to two types pathway genes—the TGs and PGs as described in the “Heat Map”. We use green, yellow, and pink nodes to represent the TFs, TGs, and PGs, respectively.

- Users can zoomed in or out the network map by scrolling mouse wheel.
- Users can highlight the up-regulated (red) and down-regulated (blue) genes at different time points based on user-defined threshold.
- Users can discover the key TFs based on user-defined degree of connection.
- Users can show the pathway genes associated with a TF selected in the TF list.
- Users can show the TFs associated with a pathway selected in the pathway list.
- After some transformations, users can restore the initial network by clicking on the “Restore” button.

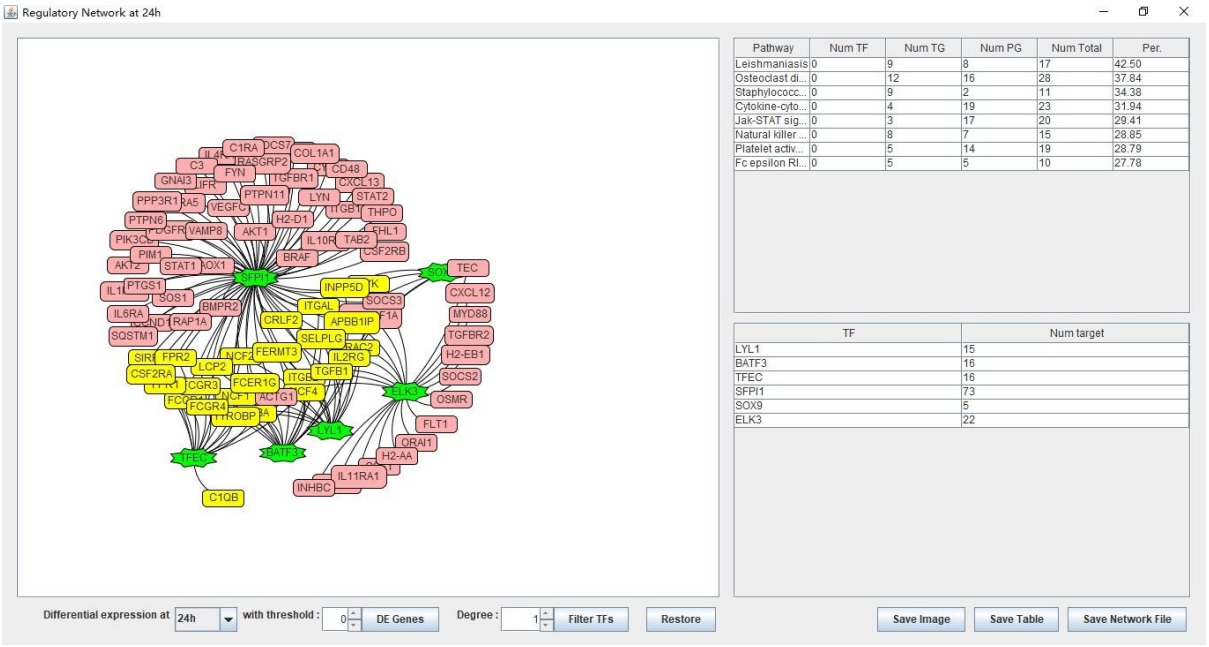


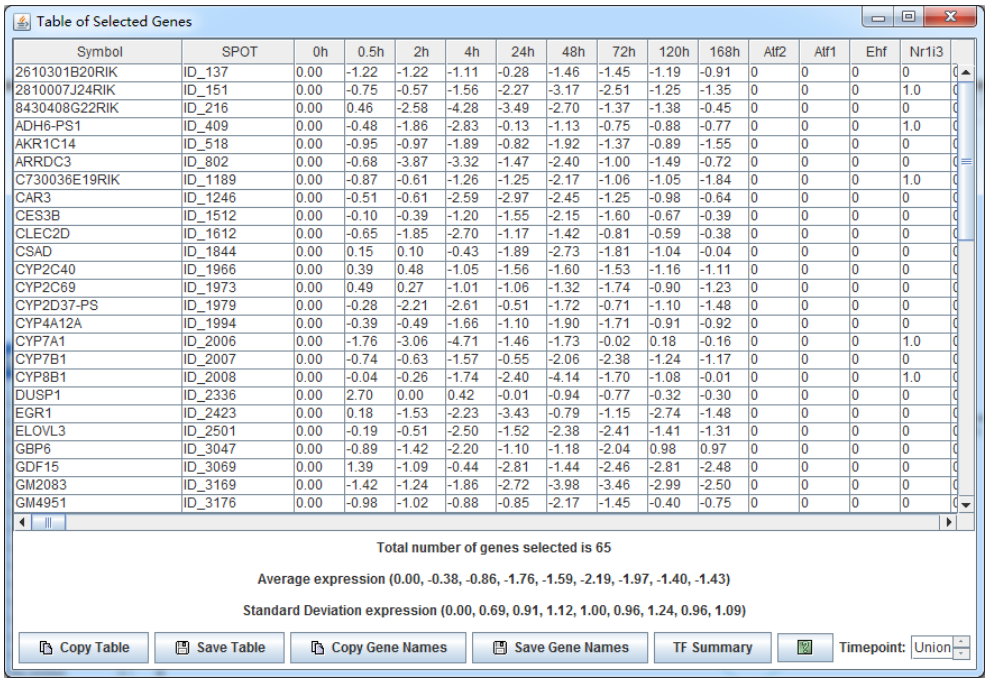
Figure 16. An example of network map.

4.6 Gene Table

Pressing the Gene Table button displays a table which contains the genes that is assigned to the currently selected edge in the tree structure. The table includes the gene's expression values after transformation. On the bottom of the table are the average and standard deviation of the expression values at each time point. An example of such a table is shown in Figure 17. The columns of the table are as follows:

- Gene Symbol – This column contains the gene symbols. The name for this column is read from the header in the data file.
- Spot ID – An entry in this column contains a list of spot IDs of spots which contain the gene of the row. The entries are delimited by a “;”.
- Time Point columns – The time-series of gene expression levels for the gene after any selected normalization method (Log normalize data, Normalize data, or No normalization/add 0). The headers of these columns are read from the time-series data file.

- TF-gene columns – These columns contain the TF-gene regulation interaction inputs.



Symbol	SPOT	0h	0.5h	2h	4h	24h	48h	72h	120h	168h	Atf2	Atf1	Ehf	Nr1i3
2610301B20RIK	ID_137	0.00	-1.22	-1.22	-1.11	-0.28	-1.46	-1.45	-1.19	-0.91	0	0	0	0
2810007J24RIK	ID_151	0.00	-0.75	-0.57	-1.56	-2.27	-3.17	-2.51	-1.25	-1.35	0	0	0	1.0
8430408G22RIK	ID_216	0.00	0.46	-2.58	-4.28	-3.49	-2.70	-1.37	-1.38	-0.45	0	0	0	0
ADH6-PS1	ID_409	0.00	-0.48	-1.86	-2.83	-0.13	-1.13	-0.75	-0.88	-0.77	0	0	0	1.0
AKR1C14	ID_518	0.00	-0.95	-0.97	-1.89	-0.82	-1.92	-1.37	-0.89	-1.55	0	0	0	0
ARRDC3	ID_802	0.00	-0.68	-3.87	-3.32	-1.47	-2.40	-1.00	-1.49	-0.72	0	0	0	0
C730036E19RIK	ID_1189	0.00	-0.87	-0.61	-1.26	-1.25	-2.17	-1.06	-1.05	-1.84	0	0	0	1.0
CAR3	ID_1246	0.00	-0.51	-0.61	-2.59	-2.97	-2.45	-1.25	-0.98	-0.64	0	0	0	0
CES3B	ID_1512	0.00	-0.10	-0.39	-1.20	-1.55	-2.15	-1.60	-0.67	-0.39	0	0	0	0
CLEC2D	ID_1612	0.00	-0.65	-1.85	-2.70	-1.17	-1.42	-0.81	-0.59	-0.38	0	0	0	0
CSAD	ID_1844	0.00	0.15	0.10	-0.43	-1.89	-2.73	-1.81	-1.04	-0.04	0	0	0	0
CYP2C40	ID_1966	0.00	0.39	0.48	-1.05	-1.56	-1.60	-1.53	-1.16	-1.11	0	0	0	0
CYP2C69	ID_1973	0.00	0.49	0.27	-1.01	-1.06	-1.32	-1.74	-0.90	-1.23	0	0	0	0
CYP2D37-PS	ID_1979	0.00	-0.28	-2.21	-2.61	-0.51	-1.72	-0.71	-1.10	-1.48	0	0	0	0
CYP4A12A	ID_1994	0.00	-0.39	-0.49	-1.66	-1.10	-1.90	-1.71	-0.91	-0.92	0	0	0	0
CYP7A1	ID_2006	0.00	-1.76	-3.06	-4.71	-1.46	-1.73	-0.02	0.18	-0.16	0	0	0	1.0
CYP7B1	ID_2007	0.00	-0.74	-0.63	-1.57	-0.55	-2.06	-2.38	-1.24	-1.17	0	0	0	0
CYP8B1	ID_2008	0.00	-0.04	-0.26	-1.74	-2.40	-4.14	-1.70	-1.08	-0.01	0	0	0	1.0
DUSP1	ID_2336	0.00	2.70	0.00	0.42	-0.01	-0.94	-0.77	-0.32	-0.30	0	0	0	0
EGR1	ID_2423	0.00	0.18	-1.53	-2.23	-3.43	-0.79	-1.15	-2.74	-1.48	0	0	0	0
ELOVL3	ID_2501	0.00	-0.19	-0.51	-2.50	-1.52	-2.38	-2.41	-1.41	-1.31	0	0	0	0
GBP6	ID_3047	0.00	-0.89	-1.42	-2.20	-1.10	-1.18	-2.04	0.98	0.97	0	0	0	0
GDF15	ID_3069	0.00	1.39	-1.09	-0.44	-2.81	-1.44	-2.46	-2.81	-2.48	0	0	0	0
GM2083	ID_3169	0.00	-1.42	-1.24	-1.86	-2.72	-3.98	-3.46	-2.99	-2.50	0	0	0	0
GM4951	ID_3176	0.00	-0.98	-1.02	-0.88	-0.85	-2.17	-1.45	-0.40	-0.75	0	0	0	0

Figure 17. An example of a gene table. The table shows all genes assigned to the currently selected edge.

4.7 Hide/Show Nodes

The “Hide Nodes” button is appeared when the interface appears for the first time. When pressing the Hide Nodes button the edges and nodes of the dynamic regulatory map are hidden (see Figure 18). Pressing the Show Nodes button will pull the interface back to its previous state.

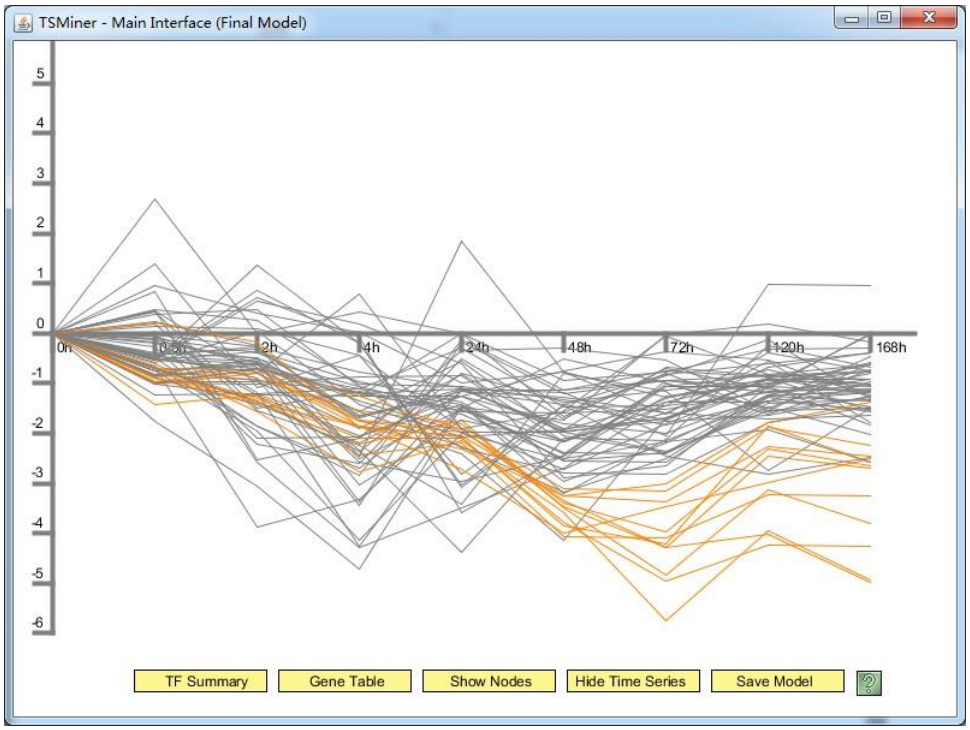


Figure 18. The interface window from Figure 9 after pressing the Hide Nodes button.

4.8 Hide/Show Time Series

The “Hide Time Series” button is appeared when the interface appears for the first time. When pressing the Hide Time Series button, the time-series plots of all the genes are hidden (see Figure 19). Pressing the Show Time Series button will pull the interface back to its previous state.

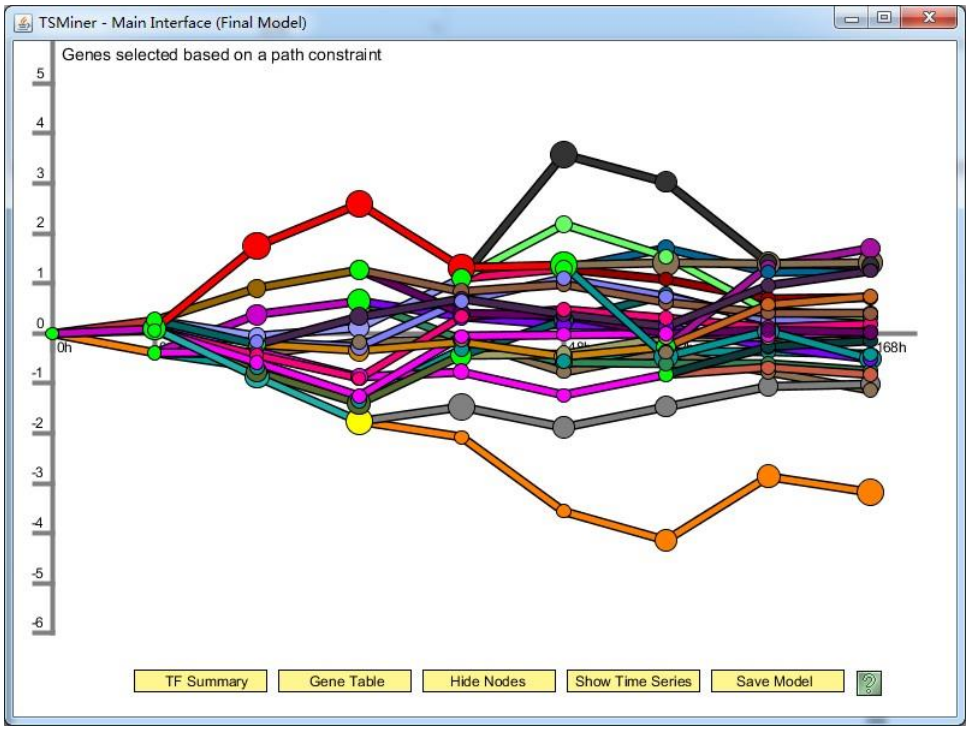


Figure 19. The interface window from Figure 7 after pressing the Hide Time Series button.

4.9 Save Model

Pressing the Save Model button opens a dialog window from which the current model can be saved into a text file. A saved model can then later be input to the “Saved Model File” field in the Main Interface.

A. TF-gene Interaction Files

Here we list the TF-gene interaction file used in the case analysis of TSMiner.

File	Description
Cellnet_tftg.txt	TF-gene interaction file for 1027 TFs that may function in the mouse liver.

B. Gene Annotation Sources

The table below lists all gene annotation data sets that can be selected under Gene Annotation Source. The Gene Ontology annotations were obtained from ftp://ftp.ncbi.nih.gov/gene/, and the KEGG and Reactome Pathway information were obtained from the NCBI BioSystems Database (ftp://ftp.ncbi.nih.gov/pub/biosystems/).

File	Description
Gene_Ontology_bp.txt	Gene Ontology Biological Process
Gene_Ontology_cc.txt	Gene Ontology Cellular Component
Gene_Ontology_mf.txt	Gene Ontology Molecular Function
KEGG_pathway_mus.txt	Pathway gene sets derived from the KEGG pathway database
Reactome_pathway_mus.txt	Pathway gene sets derived from the Reactome pathway database

REFERENCES

Consortium EP Birney E Stamatoyannopoulos JA Dutta A Guigo R Gingeras TR Margulies EH Weng Z Snyder M Dermitzakis ET et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.

Ernst J, Plasterer HL, Simon I, Bar-Joseph Z. 2010. Integrating multiple evidence sources to predict transcription factor binding in the human genome. *Genome Res* **20**: 526-536.

Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. 2007. Reconstructing dynamic regulatory maps. *Mol Syst Biol* **3**: 74.

Hahn JS, Hu Z, Thiele DJ, Iyer VR. 2004. Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Mol Cell Biol* **24**: 5249-5256.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99-104.

Krishnapuram B, Carin L, Figueiredo MA, Hartemink AJ. 2005. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Trans Pattern Anal Mach Intell* **27**: 957-968.

Kurinna S, Barton MC. 2011. Cascades of transcription regulation during liver regeneration. *Int J Biochem Cell Biol* **43**: 189-197.