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 interacting 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://github.com/free1234hm/tsminer-tool.git.

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 Train Model.

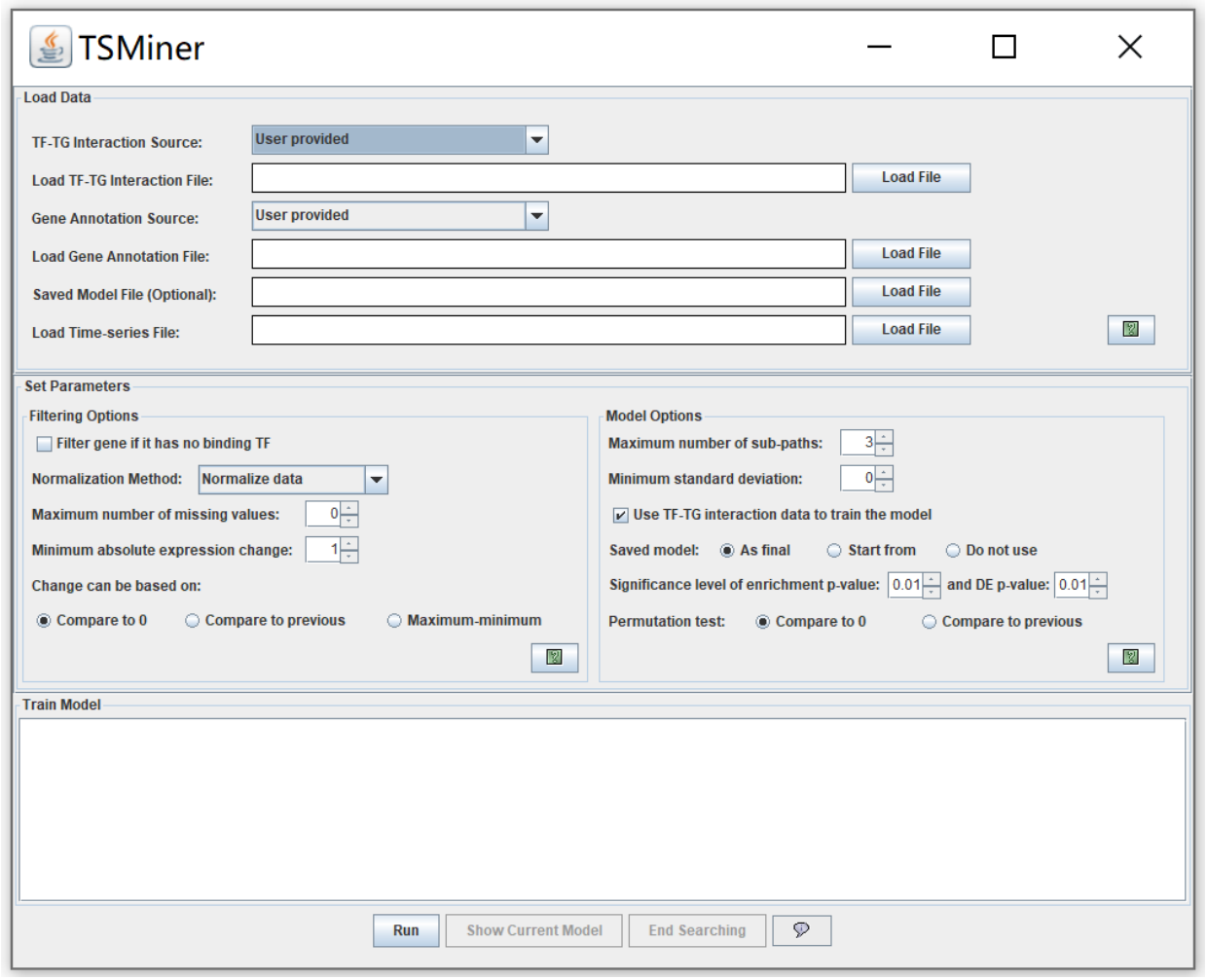


Figure 1. The main interface of TSMiner. This is the first screen that appears when TSMiner is launched. From this screen a user input the data sets and various execution options. Pressing the "Run" button at the bottom of the interface starts the model training procedure.

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 expression 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. TSMiner provides multiple known TF–TG Interaction files (see 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 format of the TF–TG interaction file can either be an ASCII text file or a GNU zip file of an ASCII text file. The TF–TG interaction file is necessary for running TSMiner. TSMiner will report errors when the field is null or a wrong directory. The file contains three columns: 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 has positive correlation interaction, while a value of "-1" represents that the TF–TG pair has negative correlation interaction. (Fig. 2). Importantly, when a TF 'tf1' has both positively and negatively correlated target genes, it will be transformed to two TFs, one only has the positively correlated targets and another only has the negatively correlated targets. They are analyzed separately and represented as 'tf1_posi' and 'tf1_nega', respectively, in the result TF table (see Figures 9 and 10).

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 2. A sample of TF–TG data file 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–TG 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 train 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 3.

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 3. 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 4.

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

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–TG 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–TG interaction file. If this box is unchecked then genes not included in the TF–TG regulation input, are not filtered and are assumed to have a "0" for every entry of the TF–TG regulation predictions.
- Normalization method – Before time-series data is analyzed, it must be transformed to start at 0. The normalization method 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:
    1) Log normalize data – transforms the vector to $\{0, \log_2(v_1/v_0), \dots, \log_2(v_n/v_{n-1})\}$;
    2) Normalize data – transforms the vector to $\{0, (v_1 - v_0), \dots, (v_n - v_{n-1})\}$;
    3) 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.
- Maximum Number of Missing Values – a gene will be filtered if the number of missing values in its time-series exceeds 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–TG 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–TG 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–TG interaction predictions are only used in the enrichment analysis and differential expression analysis. Using the TF–TG interaction data to infer the model gives a more biologically coherent model. The model training is faster when using the TF–TG 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 p-value and DE p-value – the enrichment p-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 p-value specifies the threshold below which the TGs binding to a TF are seen as significantly up or down-regulated.
- Permutation Test – after the TFs significantly associated with each sub-path are identified, TSMiner further studies the expression changes of their target genes for all the time points from the bifurcation point to the last time point. This parameter is used to define the comparison gene set in the permutation test. If "Compare to Previous" is selected, the permutation test is used to study the expression changes between the time points and the next one. If "Compare to Time 0" is selected, the permutation test is used to study the expression changes between the time points and the control group.

3.3 Train 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" button 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.

4 Display Interface

After the model learning executes, the main output window appears. The main window displays the time-series expression heat map of all the genes that were not filtered. An example of such a window is shown in Figure 5. The heat map features the tree structure model generated by TSMiner, in which genes are continuously divided as time increases. For example, genes are divided into three groups at 1 h, then divided into 9 groups at 4 h. Finally, genes are divided into 41 groups at 4 weeks (Figure 6).

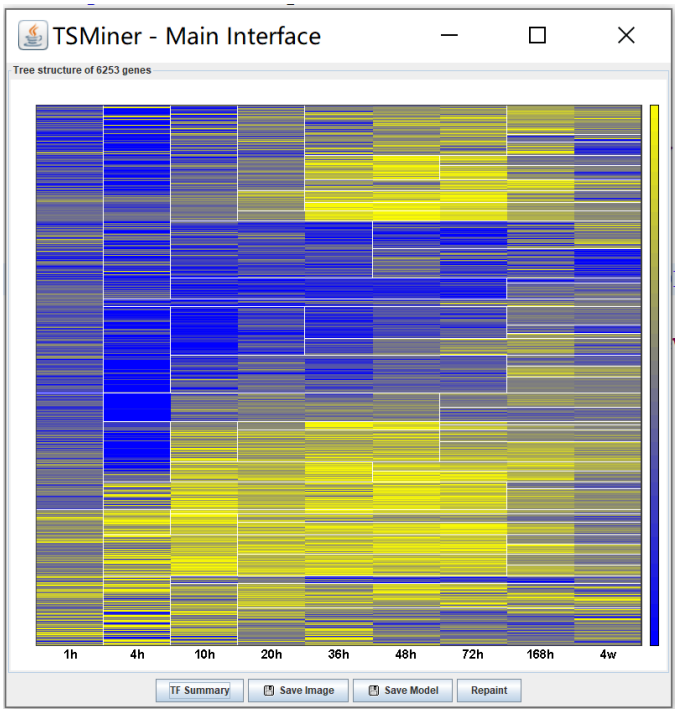


Figure 5. An example of the Display Interface. Along the bottom are buttons with various options.

The interface displays only genes assigned to a specific sub-path as the mouse moves on the heat map (Figure 6). Hovering the mouse cursor over a specific gene expression plot displays the name of the gene and the current time point. Left clicking on a sub-path brings up a "Gene Table" as described in Section 4.1. Right clicking on a sub-path brings up a "TF Table" as described in Section 4.2.

In addition, users can zoom in or out the heat map by scrolling mouse wheel, and press the 'Repaint' button repaints the heat map according to the initial size after zooming in or out it.

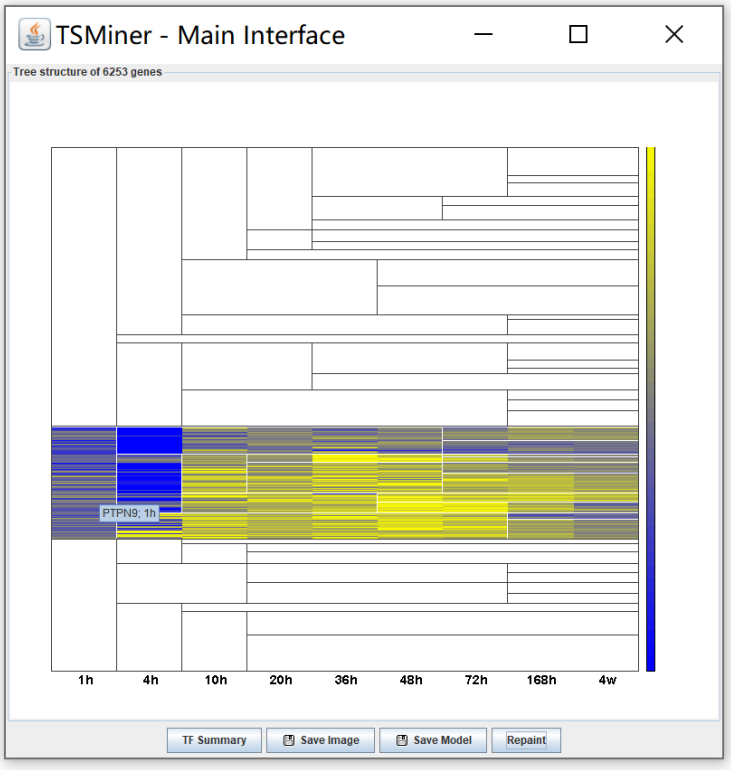


Figure 6. The Display interface window when move the mouse. When hover the mouse cursor can display the current gene name and time point.

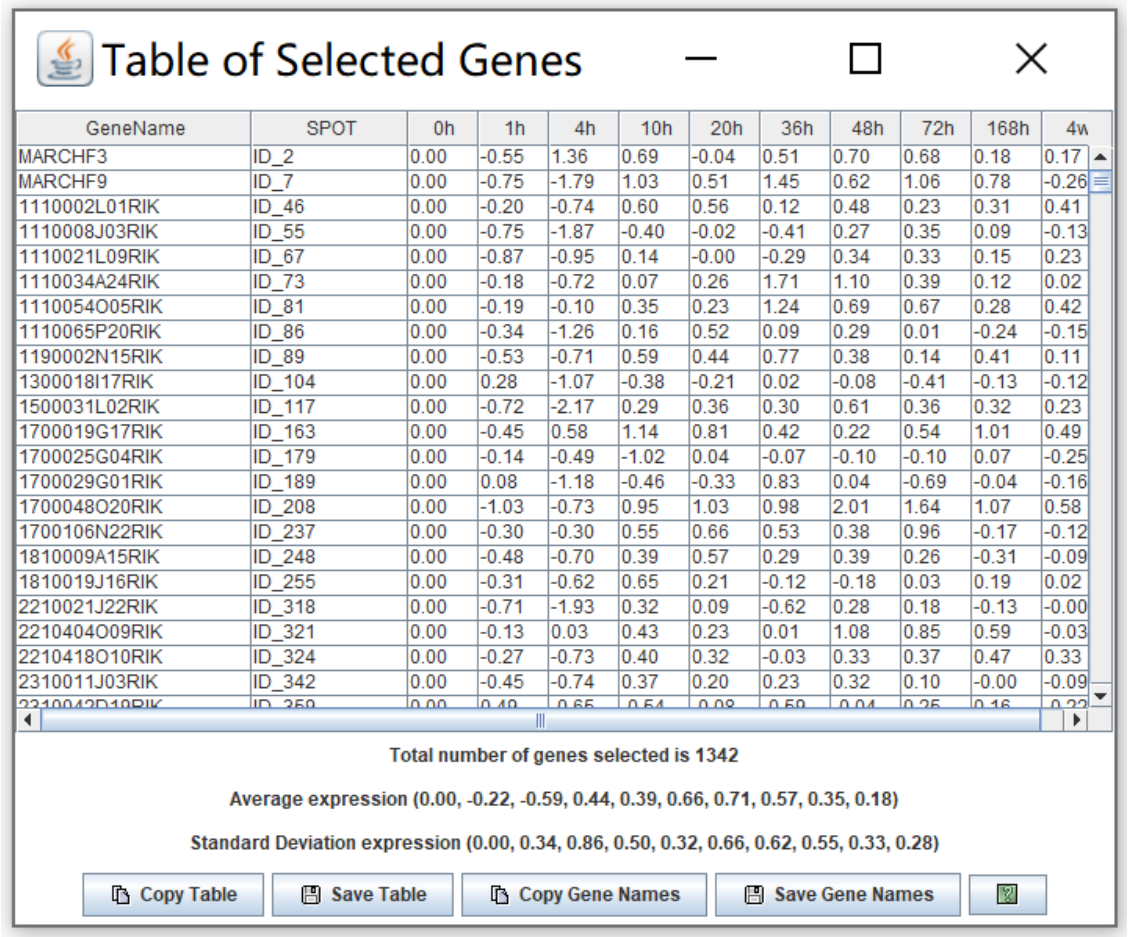
4.1 Save Model

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.

4.2 Gene Table

The Gene Table displays the genes assigned to the currently selected sub-path in the tree structure. The table includes the genes’ expression values after transformation. On the bottom of the table are the average value and the standard deviation of the expression values. An example of such a table is shown in Figure 7. 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 gene expression levels 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.



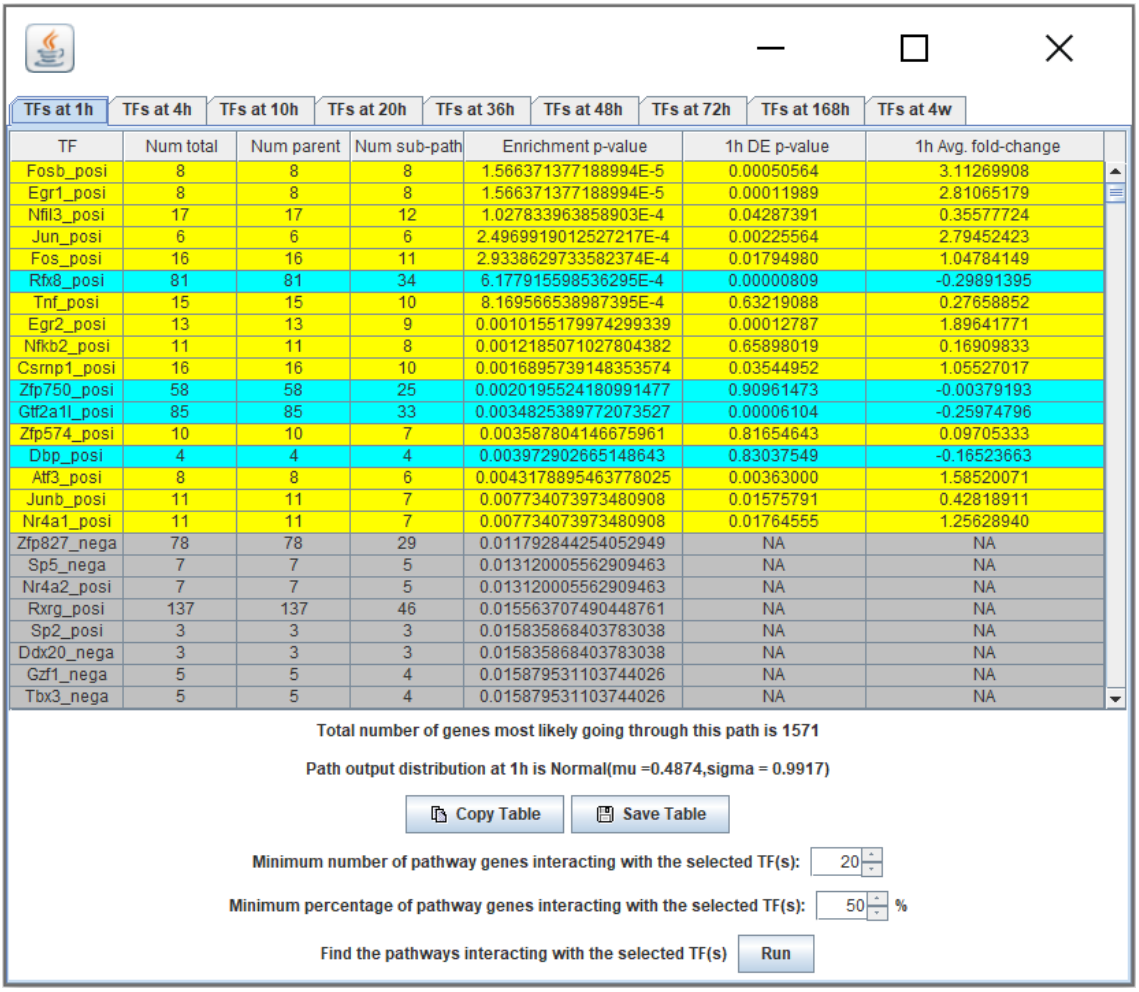
GeneName	SPOT	0h	1h	4h	10h	20h	36h	48h	72h	168h	4w
MARCHF3	ID_2	0.00	-0.55	1.36	0.69	-0.04	0.51	0.70	0.68	0.18	0.17
MARCHF9	ID_7	0.00	-0.75	-1.79	1.03	0.51	1.45	0.62	1.06	0.78	-0.26
1110002L01RIK	ID_46	0.00	-0.20	-0.74	0.60	0.56	0.12	0.48	0.23	0.31	0.41
1110008J03RIK	ID_55	0.00	-0.75	-1.87	-0.40	-0.02	-0.41	0.27	0.35	0.09	-0.13
1110021L09RIK	ID_67	0.00	-0.87	-0.95	0.14	-0.00	-0.29	0.34	0.33	0.15	0.23
1110034A24RIK	ID_73	0.00	-0.18	-0.72	0.07	0.26	1.71	1.10	0.39	0.12	0.02
1110054O05RIK	ID_81	0.00	-0.19	-0.10	0.35	0.23	1.24	0.69	0.67	0.28	0.42
1110065P20RIK	ID_86	0.00	-0.34	-1.26	0.16	0.52	0.09	0.29	0.01	-0.24	-0.15
1190002N15RIK	ID_89	0.00	-0.53	-0.71	0.59	0.44	0.77	0.38	0.14	0.41	0.11
1300018I17RIK	ID_104	0.00	0.28	-1.07	-0.38	-0.21	0.02	-0.08	-0.41	-0.13	-0.12
1500031L02RIK	ID_117	0.00	-0.72	-2.17	0.29	0.36	0.30	0.61	0.36	0.32	0.23
1700019G17RIK	ID_163	0.00	-0.45	0.58	1.14	0.81	0.42	0.22	0.54	1.01	0.49
1700025G04RIK	ID_179	0.00	-0.14	-0.49	-1.02	0.04	-0.07	-0.10	-0.10	0.07	-0.25
1700029G01RIK	ID_189	0.00	0.08	-1.18	-0.46	-0.33	0.83	0.04	-0.69	-0.04	-0.16
1700048O20RIK	ID_208	0.00	-1.03	-0.73	0.95	1.03	0.98	2.01	1.64	1.07	0.58
1700106N22RIK	ID_237	0.00	-0.30	-0.30	0.55	0.66	0.53	0.38	0.96	-0.17	-0.12
1810009A15RIK	ID_248	0.00	-0.48	-0.70	0.39	0.57	0.29	0.39	0.26	-0.31	-0.09
1810019J16RIK	ID_255	0.00	-0.31	-0.62	0.65	0.21	-0.12	-0.18	0.03	0.19	0.02
2210021J22RIK	ID_318	0.00	-0.71	-1.93	0.32	0.09	-0.62	0.28	0.18	-0.13	-0.00
2210404O09RIK	ID_321	0.00	-0.13	0.03	0.43	0.23	0.01	1.08	0.85	0.59	-0.03
2210418O10RIK	ID_324	0.00	-0.27	-0.73	0.40	0.32	-0.03	0.33	0.37	0.47	0.33
2310011J03RIK	ID_342	0.00	-0.45	-0.74	0.37	0.20	0.23	0.32	0.10	-0.00	-0.09

Figure 7. An example of a gene table. The table shows all genes assigned to the currently selected edge.

4.3 TF Table

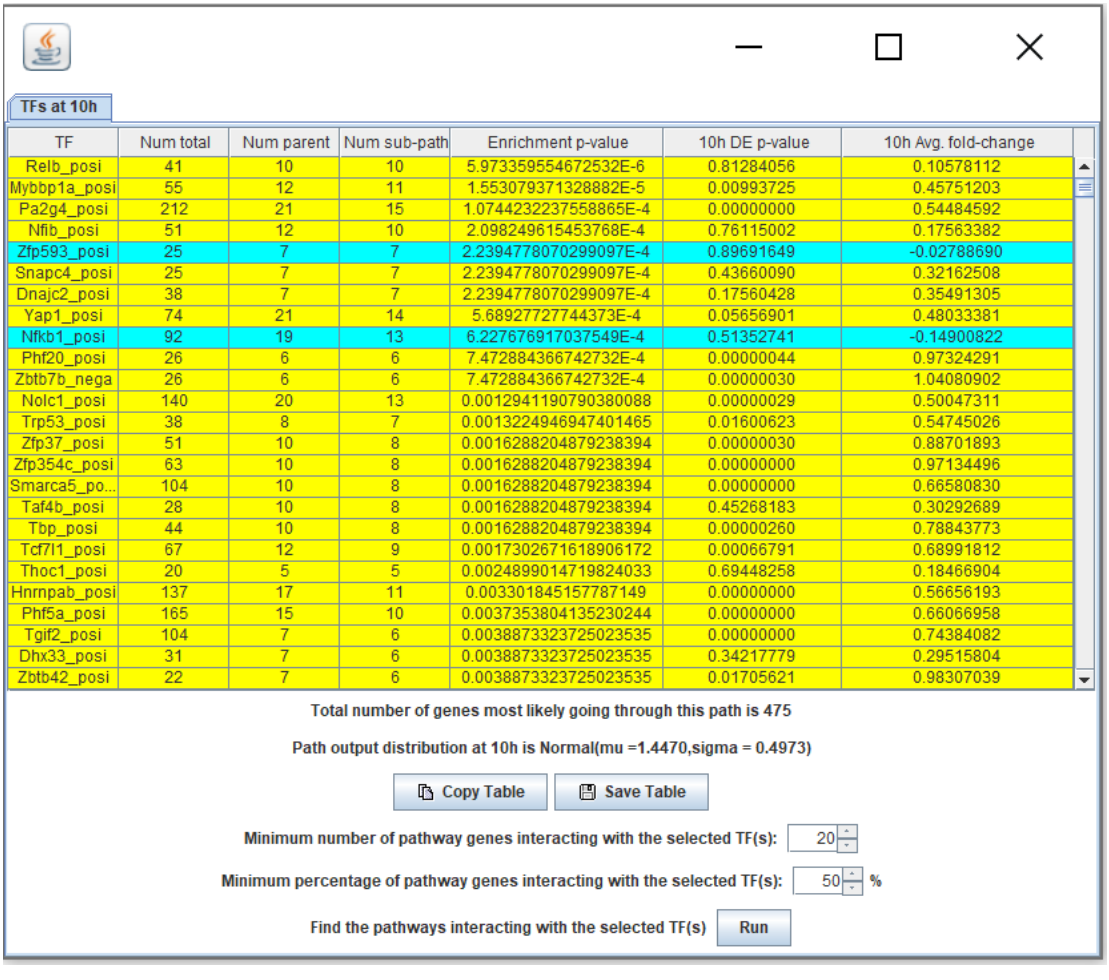
A series of TF tables appears when right clicking on a sub-path out of a bifurcation node (Figure 8). Each of them shows the TFs activated or repressed at specific time point from the bifurcation time point to the last one. In addition, one of these tables appears when right clicking on an edge whose left edge is not a bifurcation node (see Figure 9).

- 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 p-value – the enrichment significance of the targets binding to a TF to this sub-path.
- DE p-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 interacting with the selected TFs. We calculate the sum of pathway genes that directly and indirect with the selected TFs as S. One pathway is identified as significant when S is higher than the parameter "Minimum number of pathway genes interacting with the selected TF(s)", while S divided by the total number of pathway genes is higher than the parameter "Minimum percentage of pathway genes interacting with the selected TF(s)".



TF	Num total	Num parent	Num sub-path	Enrichment p-value	1h DE p-value	1h Avg. fold-change
Fosb_posi	8	8	8	1.566371377188994E-5	0.00050564	3.11269908
Egr1_posi	8	8	8	1.566371377188994E-5	0.00011989	2.81065179
Nfil3_posi	17	17	12	1.027833963858903E-4	0.04287391	0.35577724
Jun_posi	6	6	6	2.4969919012527217E-4	0.00225564	2.79452423
Fos_posi	16	16	11	2.9338629733582374E-4	0.01794980	1.04784149
Rfx8_posi	81	81	34	6.177915598536295E-4	0.00000809	-0.29891395
Tnf_posi	15	15	10	8.169566538987395E-4	0.63219088	0.27658852
Egr2_posi	13	13	9	0.0010155179974299339	0.00012787	1.89641771
Nfkb2_posi	11	11	8	0.0012185071027804382	0.65898019	0.16909833
Csrnp1_posi	16	16	10	0.0016895739148353574	0.03544952	1.05527017
Zfp750_posi	58	58	25	0.0020195524180991477	0.90961473	-0.00379193
Gtf2a1l_posi	85	85	33	0.0034825389772073527	0.00006104	-0.25974796
Zfp574_posi	10	10	7	0.003587804146675961	0.81654643	0.09705333
Dbp_posi	4	4	4	0.003972902665148643	0.83037549	-0.16523663
Atf3_posi	8	8	6	0.0043178895463778025	0.00363000	1.58520071
Junb_posi	11	11	7	0.007734073973480908	0.01575791	0.42818911
Nr4a1_posi	11	11	7	0.007734073973480908	0.01764555	1.25628940
Zfp827_nega	78	78	29	0.011792844254052949	NA	NA
Sp5_nega	7	7	5	0.013120005562909463	NA	NA
Nr4a2_posi	7	7	5	0.013120005562909463	NA	NA
Rxrg_posi	137	137	46	0.015563707490448761	NA	NA
Sp2_posi	3	3	3	0.015835868403783038	NA	NA
Ddx20_nega	3	3	3	0.015835868403783038	NA	NA
Gzf1_nega	5	5	4	0.015879531103744026	NA	NA
Tbx3_nega	5	5	4	0.015879531103744026	NA	NA

Figure 8. An example of multiple TF tables when right clicking on a sub-path out of a bifurcation node.



TF	Num total	Num parent	Num sub-path	Enrichment p-value	10h DE p-value	10h Avg. fold-change
Relb_posi	41	10	10	5.973359554672532E-6	0.81284056	0.10578112
Mybbp1a_posi	55	12	11	1.553079371328882E-5	0.00993725	0.45751203
Pa2g4_posi	212	21	15	1.0744232237558865E-4	0.00000000	0.54484592
Nfib_posi	51	12	10	2.098249615453768E-4	0.76115002	0.17563382
Zfp593_posi	25	7	7	2.2394778070299097E-4	0.89691649	-0.02788690
Snapc4_posi	25	7	7	2.2394778070299097E-4	0.43660090	0.32162508
Dnajc2_posi	38	7	7	2.2394778070299097E-4	0.17560428	0.35491305
Yap1_posi	74	21	14	5.689277277443 73E-4	0.05656901	0.48033381
Nfkb1_posi	92	19	13	6.227676917037549E-4	0.51352741	-0.14900822
Phf20_posi	26	6	6	7.472884366742732E-4	0.00000044	0.97324291
Zbtb7b_nega	26	6	6	7.472884366742732E-4	0.00000030	1.04080902
Nolc1_posi	140	20	13	0.0012941190790380088	0.00000029	0.50047311
Trp53_posi	38	8	7	0.0013224946947401465	0.01600623	0.54745026
Zfp37_posi	51	10	8	0.0016288204879238394	0.00000030	0.88701893
Zfp354c_posi	63	10	8	0.0016288204879238394	0.00000000	0.97134496
Smarca5_po...	104	10	8	0.0016288204879238394	0.00000000	0.66580830
Taf4b_posi	28	10	8	0.0016288204879238394	0.45268183	0.30292689
Tbp_posi	44	10	8	0.0016288204879238394	0.00000260	0.78843773
Tcf7l1_posi	67	12	9	0.0017302671618906172	0.00066791	0.68991812
Thoc1_posi	20	5	5	0.0024899014719824033	0.69448258	0.18466904
Hnrnpab_posi	137	17	11	0.003301845157787149	0.00000000	0.56656193
Phf5a_posi	165	15	10	0.0037353804135230244	0.00000000	0.66066958
Tgif2_posi	104	7	6	0.0038873323725023535	0.00000000	0.74384082
Dhx33_posi	31	7	6	0.0038873323725023535	0.34217779	0.29515804
Zbtb42_posi	22	7	6	0.0038873323725023535	0.01705621	0.98307039

Figure 9. An example of one TF Table when right clicking on an edge whose left edge is not a bifurcation node.

4.4 TF Summary Table

A series of TF tables that summarize the TFs activated or repressed at different 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 10), while the repressed ones are displayed in blue background (Figure 11). The column "TF", "Enrichment p-value", "DE p-value", and "Avg. fold-change", together with the procedures for inferring pathways are the same as those of the TF Tables.

Importantly, a transcriptional activator ('xx_posi') is identified to be activated (in yellow background) and repressed (in blue background) when its target genes are up-regulated and down-regulated, respectively (e.g., Figure 10 and the first table in Figure 11). And a transcriptional repressor ('xx_nega') is identified to be activated (in yellow background) and repressed (in blue background) when its target genes are down-regulated and up-regulated, respectively (e.g., the second table in Figure 11).

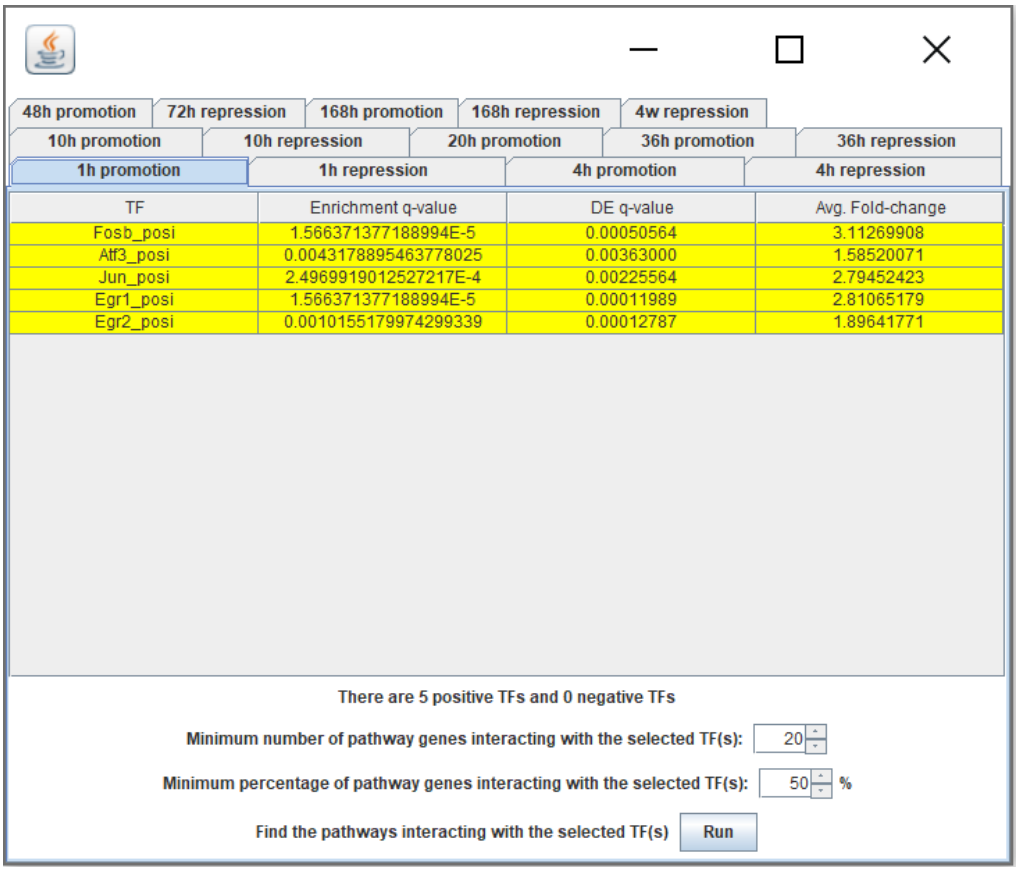


Figure 10. A summary TF table showing the TFs activated at 1 h.

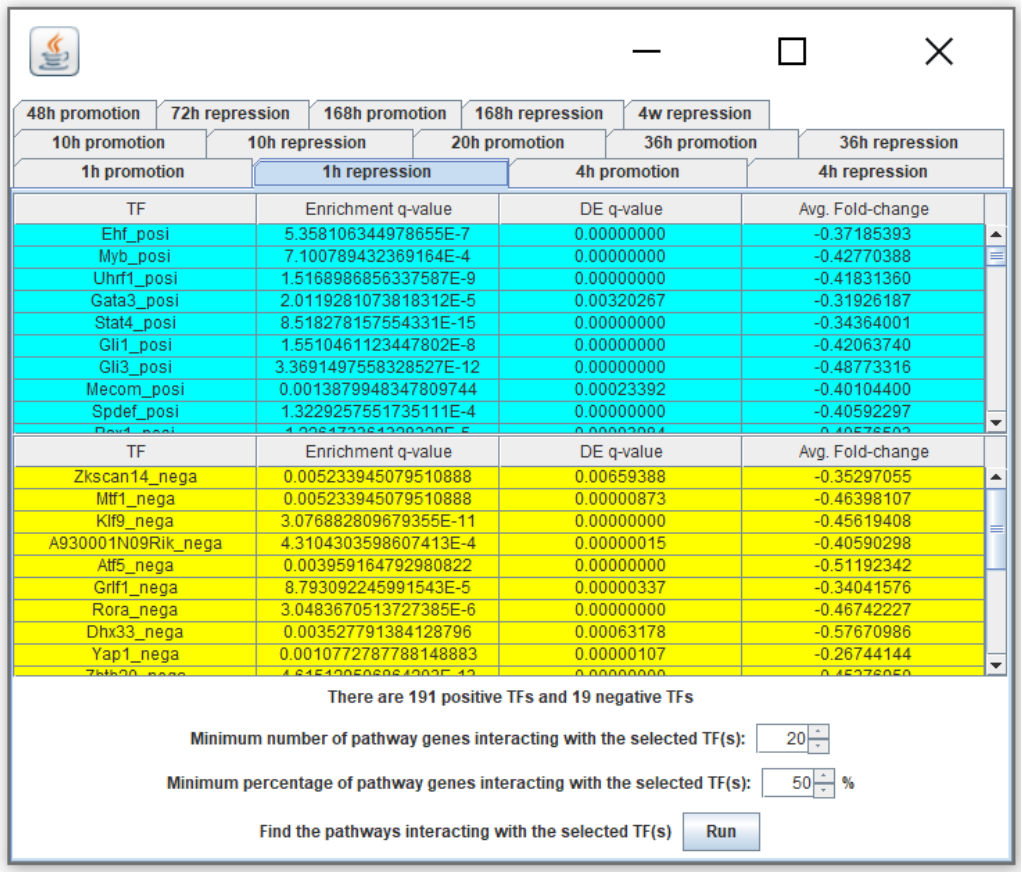


Figure 11. A summary TF table showing the TFs repressed at 1 h.

4.5 Pathway Table

Pathway tables displays the pathways interacting 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 12 shows an example of pathway table, in which the pathways are activated at 1 h.

- Pathway – the name of biological pathways;
- Gene Count – the number of pathway genes directly and indirectly interacting with the selected TFs;
- Percent – the percentage of 'Gene Count' among all the genes involved in the pathway;
- DE p-value – the differential expression significance of the pathway genes;
- DE q-value – transform the differential expression p-value to q-value according to Benjamini–Hochberg (BH) correction;
- Avg. FC – the average fold-changes of the pathway genes 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").

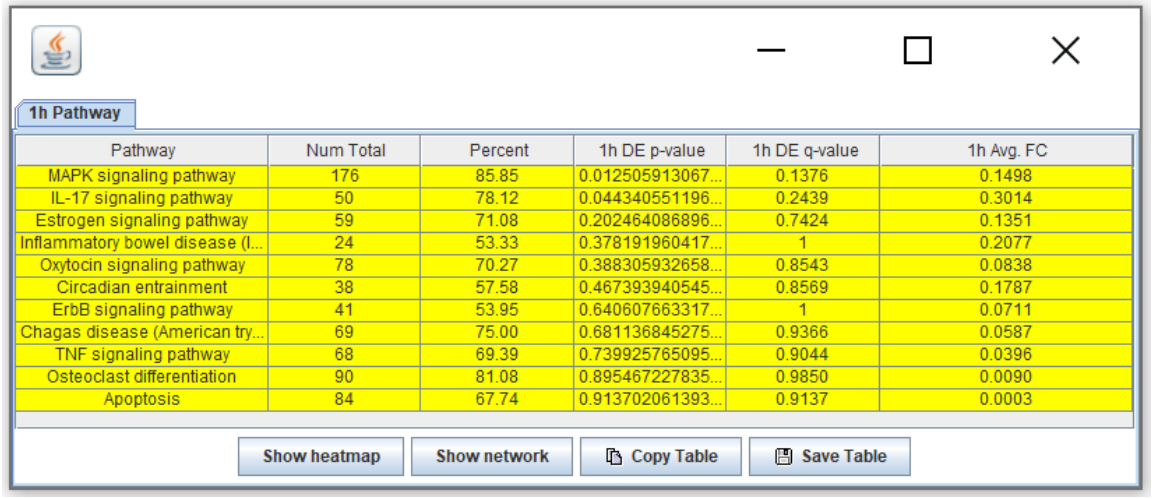


Figure 12. An example of pathway table showing the pathways activated at 1 h.

4.6 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 groups of genes (Figure 13):

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

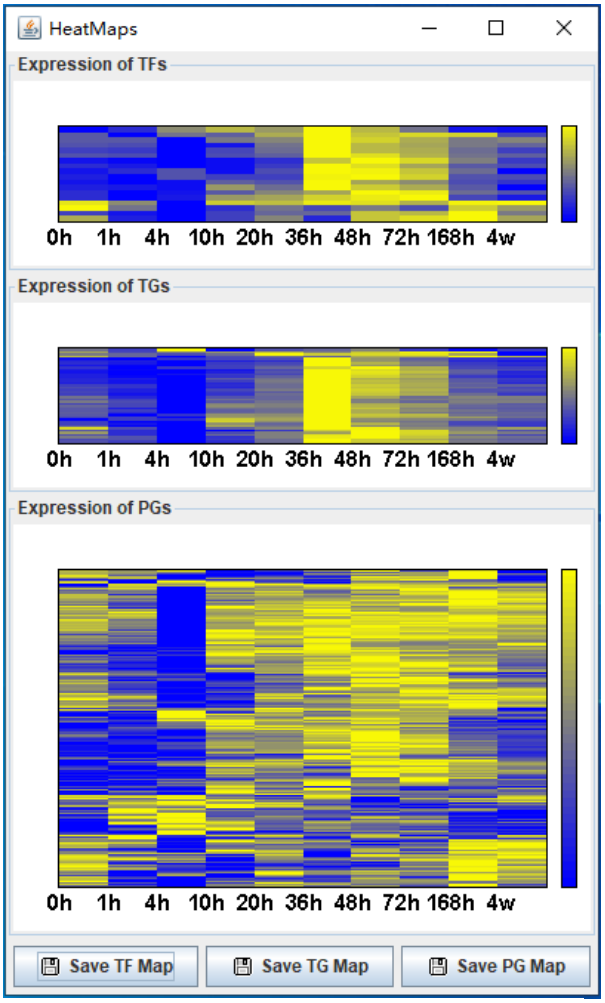


Figure 13. An example of heat maps.

4.7 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 14), in which the TFs interacting 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 zoom 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.

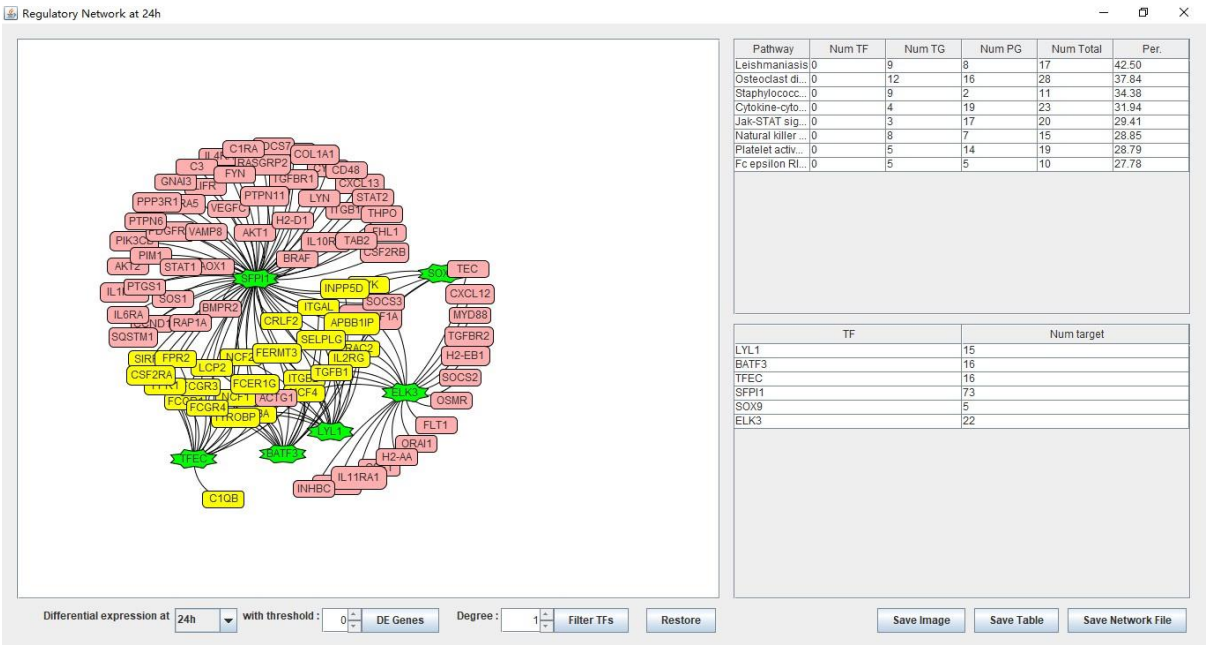


Figure 14. An example of network map.

A. TF–TG Interaction Files

Here we list the TF–TG interaction file used in the case analysis of TSMiner.

File	Description
Cellnet_tftg.txt	TF–TG 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.