TimeClust: a clustering tool for gene expression time series
User's guide (Ver. 1.3)

Paolo Magni, Fulvia Ferrazzi, Lucia Sacchi, Riccardo Bellazzi
Dipartimento di Informatica e Sistemistica,
Università degli Studi di Pavia,
Via Ferrata 1, I-27100 Pavia, Italy

November 15, 2007

Abstract

TimeClust is an user-friendly software package to cluster genes according to their temporal expression profiles. It can be conveniently used to analyze data obtained from DNA microarray time-course experiments. It implements two original algorithms expressed designed for clustering short time series together with hierarchical clustering and self-organizing maps.

TimeClust executable files for Windows and LINUX platforms can be downloaded free of charge for non-profit institutions from the following web site: http://aimed11.unipv.it/TimeClust.

For software support please contact timeclust@aimed11.unipv.it.

TimeClust has been described in Magni et al., Bioinformatics, 2007.

TimeClust at work: a simple example

The data analyzed in this section to describe the functionality of the software tool TimeClust refer to genes involved in the human cell cycle. The data were collected by Whitfield et al. and are described in [1]. Original data are available for download at the site http://genome-www.stanford.edu/Human-CellCycle/Hela. Whitfield et al. performed different experiments in which they used cDNA microarrays to measure gene expression values at different time points along the cell cycle. In one of these, they measured the $log_2\frac{Cy5}{Cy3}$ of gene expression at 26 time points over 44 hours using microarrays containing about 42000 clones. Moreover, in [1] each clone has been associated to a specific phase of the cell cycle on the basis of the temporal location of the peak of expression.

Two subsets of clones identified by Whitfiled et al. was selected and data are distributed as an example with TimeClust. In particular, one refers to a small group of 20 genes (toy data set) which are ”markers” of the different cell cycle phases and the other one to a large group of about 800 periodically expressed genes (typical data set), which include the previous 20 markers. The first data set is distributed with the software tool in the file expressiondata.txt together with the corresponding gene label files, i.e. genenames.txt, clonenames.txt, cellcyclephases.txt containing the names of the genes, names of the clones and the cell cycle phase associated to each gene respectively. Similarly, the second data set is distributed in the files expressiondata2.txt, clonenames2.txt and cellcyclephases2.txt.

For simplicity, in the following we will illustrate the software tool by considering only the toy data set, but similar analysis can be conducted on the typical one. TimeClust requires that the expression data are contained into an ASCII file structured as a spreadsheet or table data of only numbers. In each row all the expression data of a gene are reported. Each column corresponds to a different time instant of the time series. Data on the same line have to be separated by spaces or tabulators.

The gene expression time series of the trial data set are shown in Fig. 1. Following [1], the 20 genes included in this data set can be associated to the cell cycle phases in the following way:

- **Phase G1/S** CCNE1, E2F1, CDC6, PCNA (4 genes)
- **Phase S** RFC4, DHFR, RRM2, RAD51 (4 genes)
- **Phase G2** TOP2A, CDC2, CCNF, CCNA2 (4 genes)
- **Phase G2/M** STK15, BUBB1, CCNB1, PLK1 (4 genes)
- **Phase M/G1** PTTG1, RAD21, VEGFC, CDKN3 (4 genes)

When the program starts the main window reported in Fig. 2 appears. By clicking on the menu item *File/Choose working directory* and browsing the file system it is possible to choose the working directory in which the results of the cluster analysis will be saved. This choice can also be made subsequently in all the TimeClust windows.

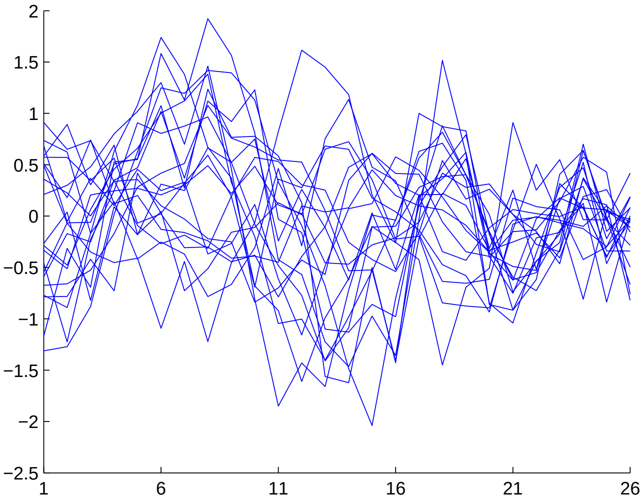


Figure 1: The gene expression time series of the 20 genes of the trial dataset to cluster.

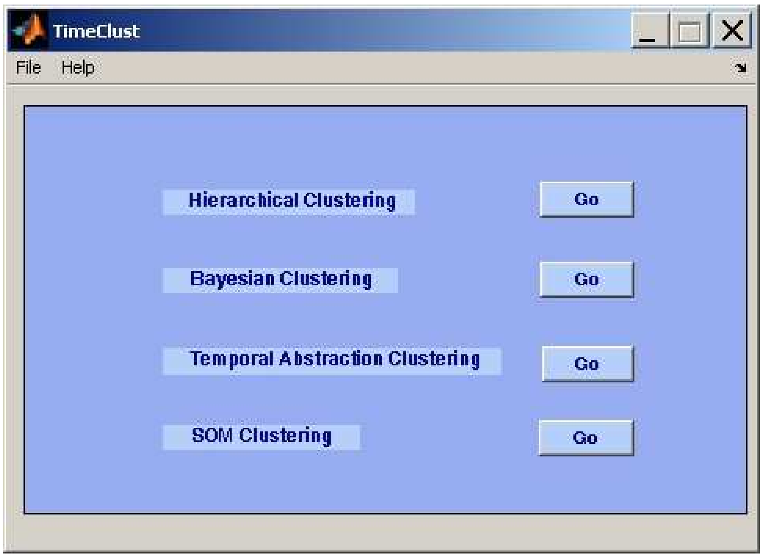
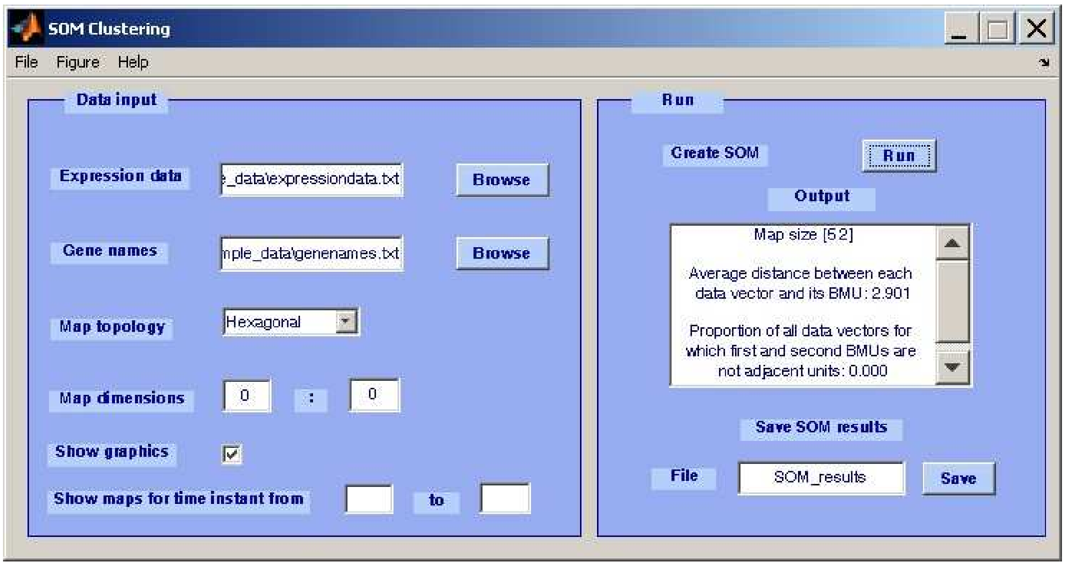


Figure 2: The TimeClust main window. Press one of the *Go* buttons to select the clustering algorithm.

By pushing one of the *Go* buttons it is possible to choose the desired clustering algorithm to process the data. A new window, specifically designed for each of the four implemented algorithms, will be open.

SOM pre-processing or clustering

By choosing the SOM pre-processing/clustering algorithm, the window reported in Fig. 3 appears. In the *Data input* section of this window, the user has to choose, by clicking on the corresponding *Browse* button and browsing the file system, the file containing the expression data to be clustered (e.g. expressiondata.txt). Then he/she can choose in the same way the file containing the gene labels (e.g. genenames.txt, clonename.txt or cellcyclephases.txt). This task is not mandatory and if it is not performed, gene labels are automatically generated in the following way: gene1 for the gene whose expression time series is in the first row of the expression data file, gene2 for the gene whose expression time series is in the second row of the expression data file, etc.



Expression data: _data\expressiondata.txt

Gene names: nple_data\genenames.txt

Map topology: Hexagonal

Map dimensions: 0 : 0

Show graphics ☑

Show maps for time instant from ___ to ___

```
Map size [5 2]

Average distance between each
data vector and its BMU: 2.901

Proportion of all data vectors for
which first and second BMUs are
not adjacent units: 0.000
```

File: SOM_results

Figure 3: The SOM clustering/pre-clustering window.

The topology of the map can be selected to be either *Hexagonal* or *Rectangular* and the dimension of the map can be fixed. If a 0:0 map is introduced (this is the default value) the map dimensions will be automatically defined on the basis of the characteristics of the dataset. If SOM are used to reduce the dimensionality of the clustering problem the total number of map-units should be less than the number of gene time series.

Finally, the results will be graphically shown or not in accordance to the choice made by the user through the *Show Graphics* check box. If it has been checked, it is possible to further select the time instants for which the corresponding SOM map will be shown (see later). If nothing is introduced, all the time instants will be considered.

By clicking on the *Run* button the algorithm starts. Several figures are produced when the *Show Graphics* check box is selected. In particular, a figure (Fig. 4) reports the U-matrix[1] and the map itself (on which for each map-unit that is a BMU[2] the name of all the genes for which it is BMU are reported); one or more figures report the maps computed for each of the time instants selected by the user (Fig. 5); one figure shows again the map in which each map-unit is black filled proportionally to the number of genes for which the map-unit is the BMU (Fig. 6); a figure for each cluster reports the expression time series of genes that have the same BMU together with the BMU codebook (Fig. 7).

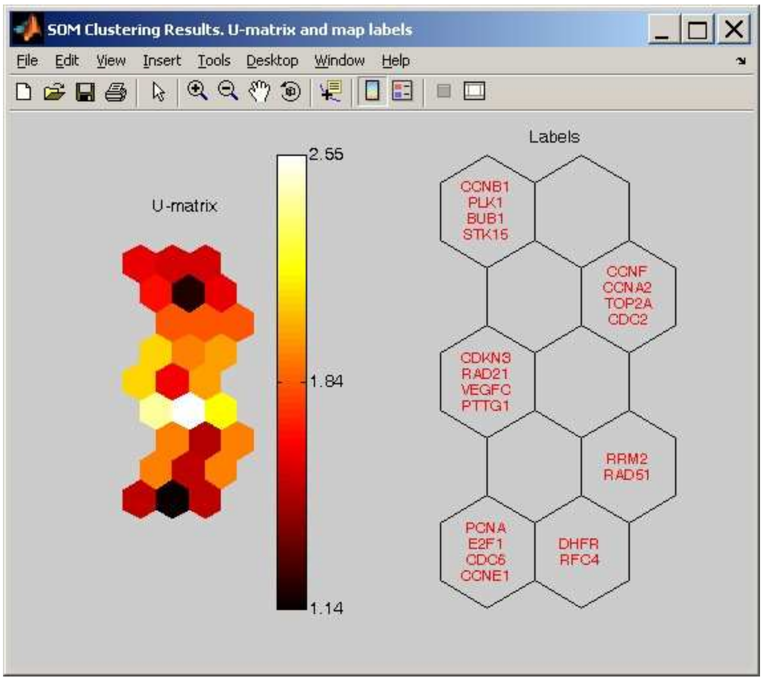


Figure 4: SOM results: the U-matrix and the map with the gene names associated to each BMU.

All the figures can be saved as MATLAB figures (*.fig) or in several other formats such as JPEG or encapsuled-postscript through the menu item *Figure/Save* or *File/Save as* that is present on each figure. MATLAB figures can be subsequently printed or visualized through the *Figure/Print* or *Figure/Load* menu items reported on all the clustering windows.

As reported in Fig. 3, the *Output* box of the SOM clustering window shows the size of the map (specified by the user or computed by the algorithm), the average distance

[1] The unified distance matrix (U-matrix) visualizes distances between neighboring map units adding extra units in the map between two adjacent units. It helps the discovery of cluster structures of the map: high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves.

[2] The Best Matching Unit (BMU) of a gene is the map-unit whose codebook is closer to gene expression time series.

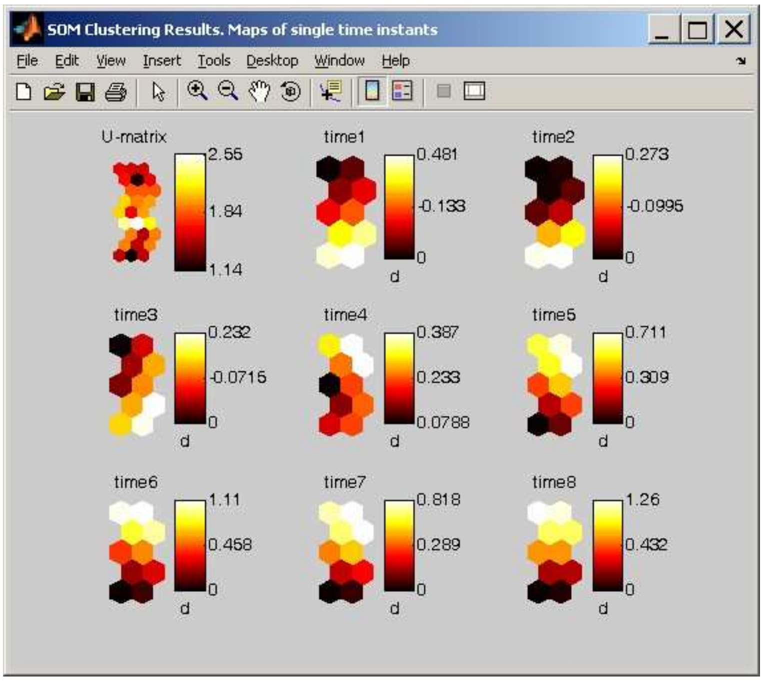


Figure 5: SOM results: the U-matrix and the maps of each time instant.

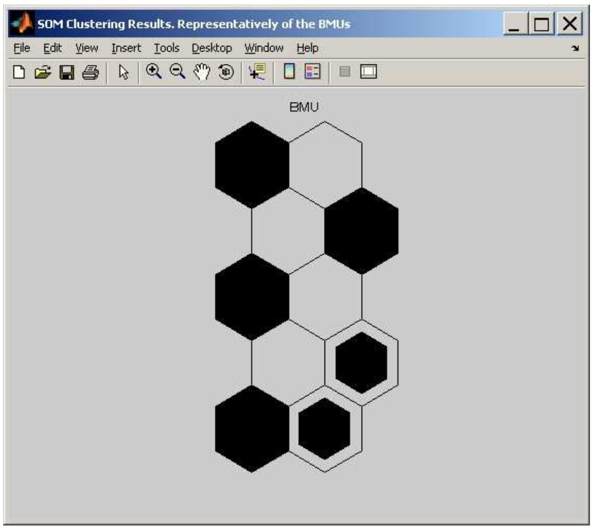

SOM Clustering Results. Representatively of the BMUs
File Edit View Insert Tools Desktop Window Help
BMU

Figure 6: SOM results: each map-unit is black filled proportionally to the number of genes for which the map-unit is the BMU.

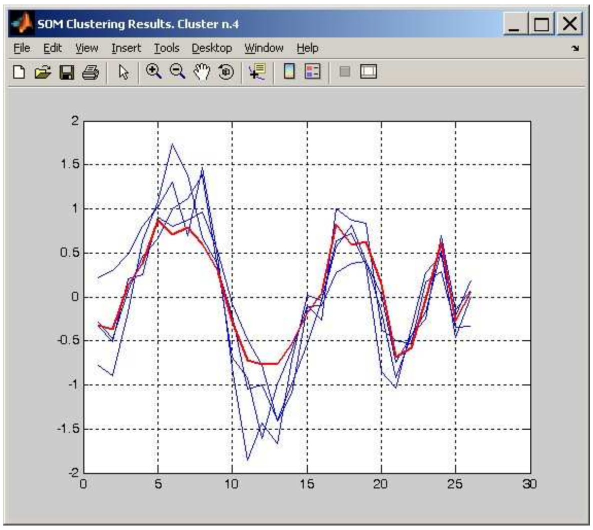
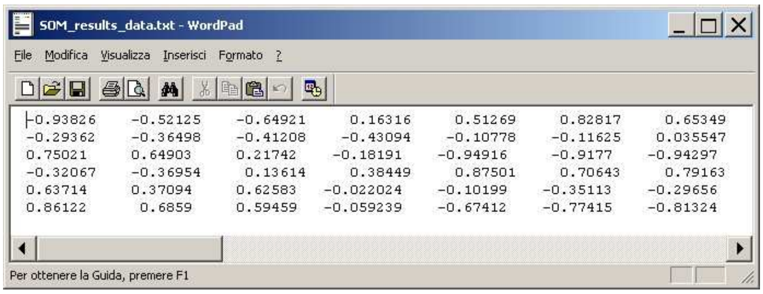


Figure 7: SOM results: the expression time series of genes that have the same BMU, i.e. the ones clustered together. In red the representative codebook profile is reported.

between each gene time series and its BMU and the percentage of genes for which first and second BMUs are not adjacent units.

Finally, by writing a file name (e.g. SOM_results) and by clicking on the *Save* button, the BMU and their associated genes can be saved on two text files: SOM_results_data.txt (Fig. 8) containing the codebook (i.e., the gene expression data) of each cell that is a BMU of at least one gene; SOM_results_name.txt (Fig. 9) containing for each of these map-units the list of gene names that have that unit as their BMU. If SOMs are used as preprocessing tool, these files can be used as the two input files for other clustering methods.

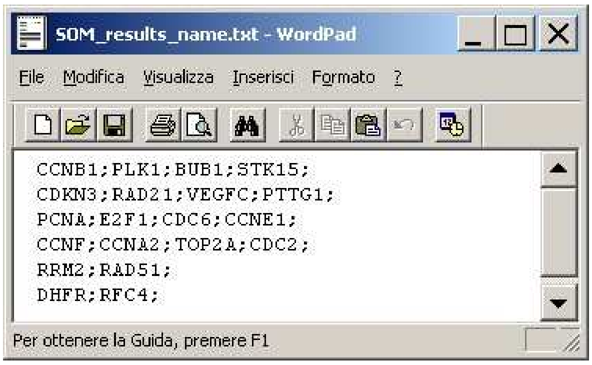

SOM_results_data.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
-0.93826   -0.52125   -0.64921    0.16316    0.51269    0.82817    0.65349
-0.29362   -0.36498   -0.41208   -0.43094   -0.10778   -0.11625   0.035547
0.75021    0.64903    0.21742   -0.18191   -0.94916   -0.9177   -0.94297
-0.32067   -0.36954    0.13614    0.38449    0.87501    0.70643    0.79163
0.63714    0.37094    0.62583  -0.022024  -0.10199   -0.35113   -0.29656
0.86122     0.6859    0.59459  -0.059239  -0.67412   -0.77415   -0.81324
```

Figure 8: SOM results: the data.txt file. In each line the codebook, i.e. the expression data, associated to a BMU is reported.

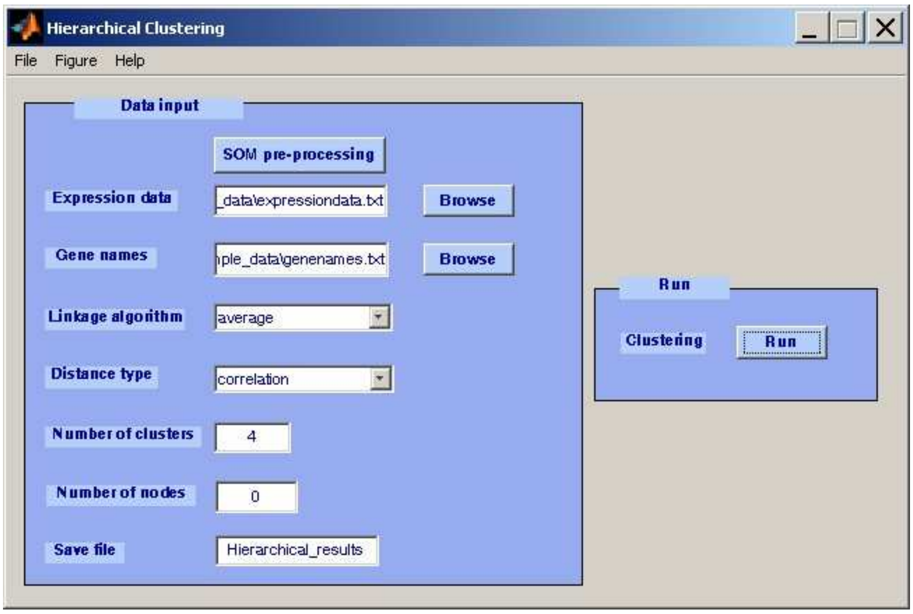
Hierarchical Clustering (HC)

By choosing the HC algorithm from the TimeClust main window (Fig. 2), the window reported in Fig. 10 appears. When the number of genes is very large, it can be useful


SOM_results_name.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
CCNB1;PLK1;BUB1;STK15;
CDKN3;RAD21;VEGFC;PTTG1;
PCNA;E2F1;CDC6;CCNE1;
CCNF;CCNA2;TOP2A;CDC2;
RRM2;RAD51;
DHFR;RFC4;
```

Figure 9: SOM results: the name.txt file. In each line the gene names associated to a BMU are reported.



Expression data: _data\expressiondata.txt
Gene names: ple_data\genenames.txt
Linkage algorithm: average
Distance type: correlation
Number of clusters: 4
Number of nodes: 0
Save file: Hierarchical_results
Clustering: Run

Figure 10: The Hierarchical Clustering window.

to reduce the dimensionality of the clustering problem by using the SOM as a pre-processing tool. For this reason, from the HC window it is possible to directly open the SOM clustering window in order to pre-process the data, save the results of the pre-processing step and continue with the analysis.

As in the other clustering windows, in the *Data input* section the user has to choose the file containing the expression data to cluster (e.g. expressiondata.txt) by clicking on the corresponding *Browse* button and browsing the file system. Then he/she can choose in the same way the file containing the gene labels/names. This task is not mandatory and if it is not performed, gene labels are again automatically generated as explained in the previous subsection (gene1 for the gene whose expression time series is in the first line of the expression data file, gene2 for the gene whose expression time series is in the second line of the expression data file, etc.).

The linkage algorithm, used to evaluate the distance between two clusters during the agglomerative steps, can be selected among *single, complete, average, centroid* and *ward*. The *centroid* or the *ward* algorithms should be used only with the Euclidean distance.

The pairwise distance measure between gene time series, fundamental ingredient of this similarity-based clustering algorithm, can be selected among *correlation, cross-correlation* and *Euclidean distance*. For what concerns the correlation, the pairwise distance is computed as one minus the sample correlation between points of the two time series, whereas for what concerns the crosscorrelation function, it is computed as one minus the sample crosscorrelation between points of the two time series, evaluated in correspondence of the time lag that maximizes the crosscorrelation.

Moreover, it is possible to specify the number of clusters to build and the number of nodes to show in the dendrogram (Fig. 11).

By clicking on the *Run* button the algorithm starts. Several figures are produced. In particular, a figure depicts the dendrogram (Fig. 11), a figure shows the heat map (Fig. 12) and a figure for each cluster reports the expression time series of genes that belong to the cluster (Fig. 13).

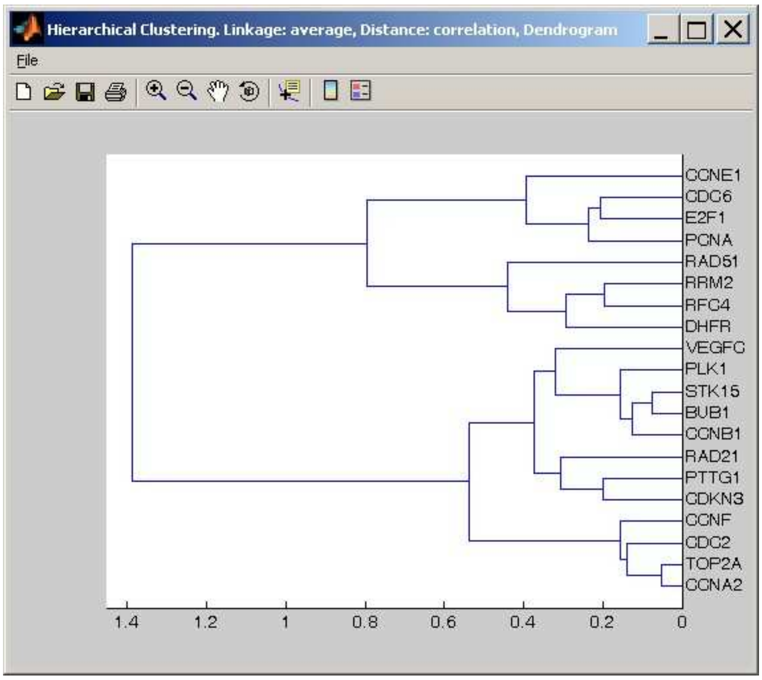


Figure 11: HC results: the dendrogram.

Finally, if a file name is inserted in the appropriate text box, the results of the clustering procedure will be automatically saved into a text file containing the list of the genes of each cluster(Fig. 14).

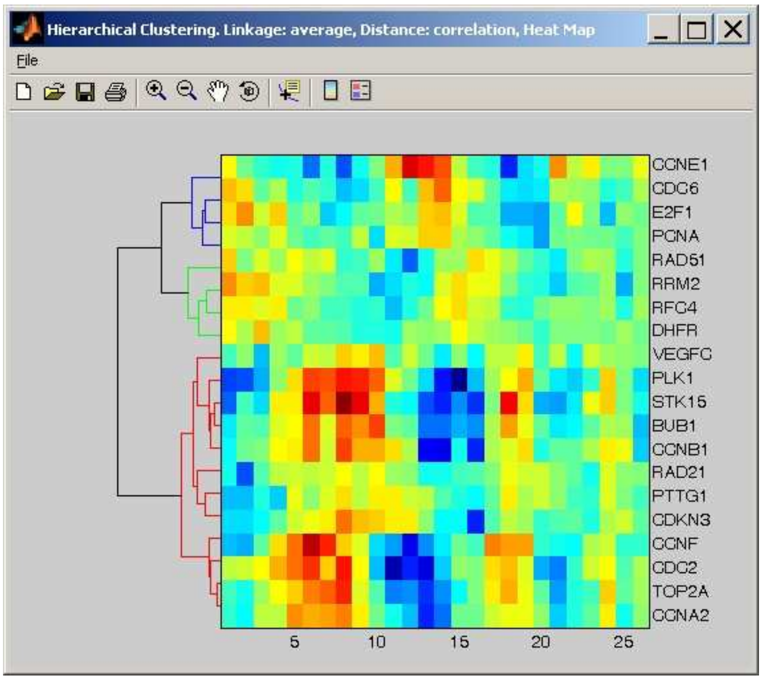


Figure 12: HC results: the heat map.

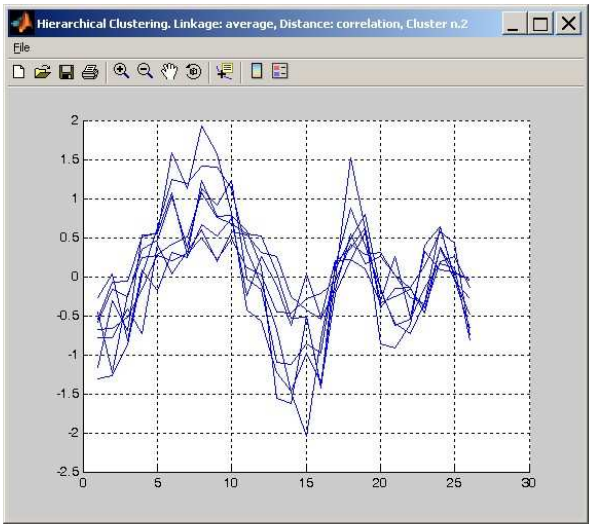
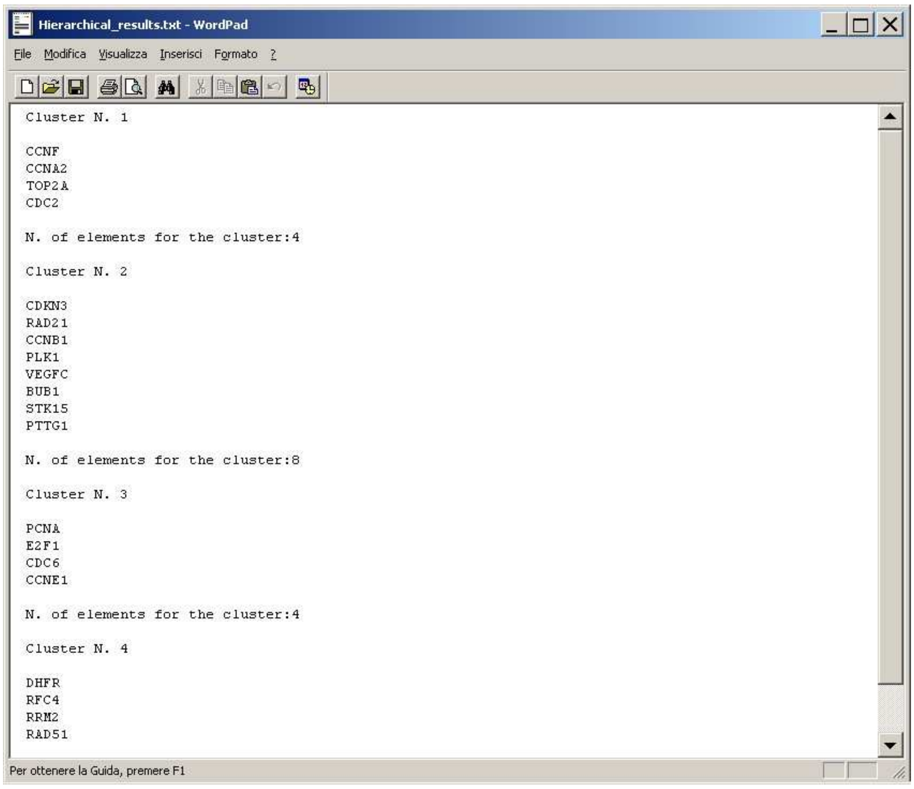

Hierarchical Clustering. Linkage: average, Distance: correlation, Cluster n.2
File
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
5
10
15
20
25
30

Figure 13: HC results: the expression time series of genes that are in the cluster.


Hierarchical_results.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
Cluster N. 1

CCNF
CCNA2
TOP2A
CDC2

N. of elements for the cluster:4

Cluster N. 2

CDKN3
RAD21
CCNB1
PLK1
VEGFC
BUB1
STK15
PTTG1

N. of elements for the cluster:8

Cluster N. 3

PCNA
E2F1
CDC6
CCNE1

N. of elements for the cluster:4

Cluster N. 4

DHFR
RFC4
RRM2
RAD51
```

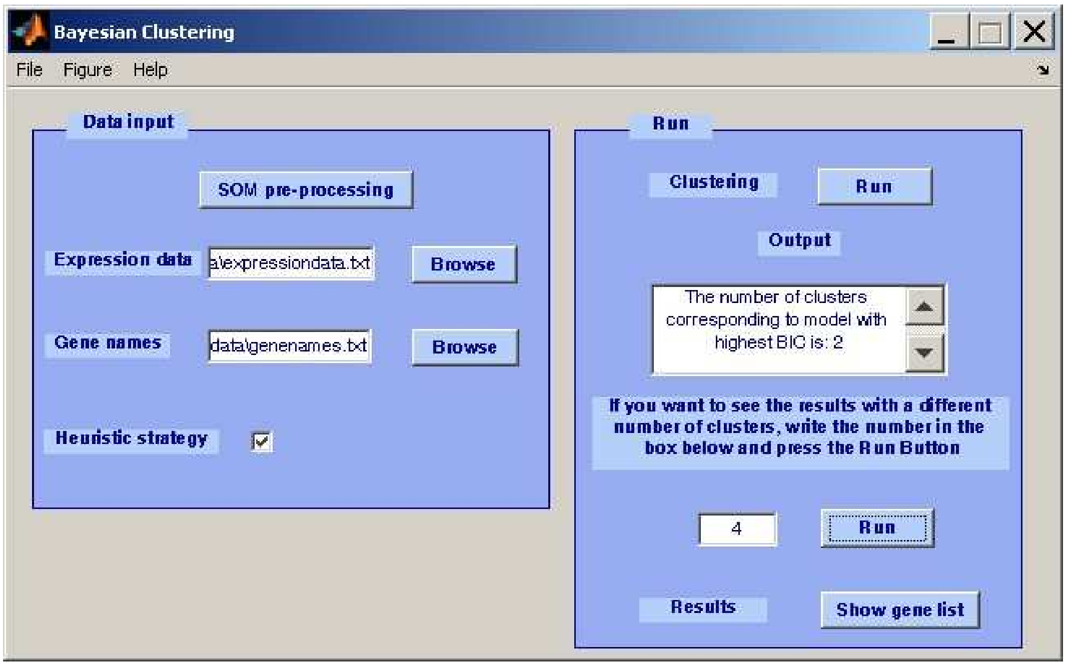
Figure 14: HC results: list of genes in each cluster.

Bayesian Clustering (BC)

By choosing the BC algorithm from the TimeClust main window (Fig. 2), the window reported in Fig. 15 appears. In this case, due to the computational burden of the BC procedure, when the number of genes is very large it is very useful to reduce the dimensionality of the clustering problem by using the SOM approach. From the BC window it is possible to directly open the SOM clustering window in order to pre-process the data, save the results of the pre-processing step and continue with the analysis.

As in the other clustering windows, in the *Data input* section of this window, the user has to choose by clicking on the corresponding *Browse* button and browsing the file system the file containing the expression data to cluster (e.g. expressiondata.txt). Then he/she can choose, in the same way, the file containing the gene names/labels. This task is not mandatory and if it is not performed, gene labels are again automatically generated as explained in the previous subsection (gene1 for the gene whose expression time series is in the first line of the expression data file, gene2 for the gene whose expression time series is in the second line of the expression data file, etc.).

The only choice that has to be made in this window, checking the appropriate check box, is related to the adoption of the heuristic distance-based search strategy. As described in [2], the heuristic strategy looks for a suboptimal solution avoiding the comparison of all possible models in the agglomerative steps. This strategy is useful when the number of genes is very high.



Expression data: a\expressiondata.txt

Gene names: data\genenames.txt

Heuristic strategy ☑

The number of clusters corresponding to model with highest BIC is: 2

If you want to see the results with a different number of clusters, write the number in the box below and press the Run Button

4

Figure 15: The BC window.

By clicking on the *Run* button the algorithm starts and results are automatically saved in a sub-directory called *BC_Results*, created in the current directory. The output box of the BC window (Fig. 15) shows the number of clusters corresponding to the model with the highest BIC. Several figures are produced. In particular, a figure depicts the dendrogram and the relationship between the number of clusters and the BIC score (Fig. 16) and a figure for each cluster reports the expression time series of genes that belong to the cluster together with the typical profile (Fig. 17). The number of clusters in which the dataset is divided is automatically found as the one that maximizes the BIC score. The list of gene names in each cluster is saved in a text file, whose name is *#clusters.txt*, where # is the number of clusters in which the dataset has been grouped. Only the file corresponding to the best subdivision is automatically generated. However, it is possible to visualize any different cluster subdivision (and produce the corresponding file) by simply writing in the right text box the desired number of clusters and clicking the *Run* button (Fig. 17), without running again the algorithm. All the #clusters files can be visualized within the editor specified in the config.txt file[3] by clicking on the *Show gene list* button (Fig. 18).

Temporal abstraction clustering (TAC)

Choosing the TAC algorithm from the TimeClust main window (Fig. 2), the window reported in Fig. 19 appears. As previously mentioned, when the number of genes is

[3] For more details about the config.txt file and its configuration/modification see the README file on the TimeClust web site.

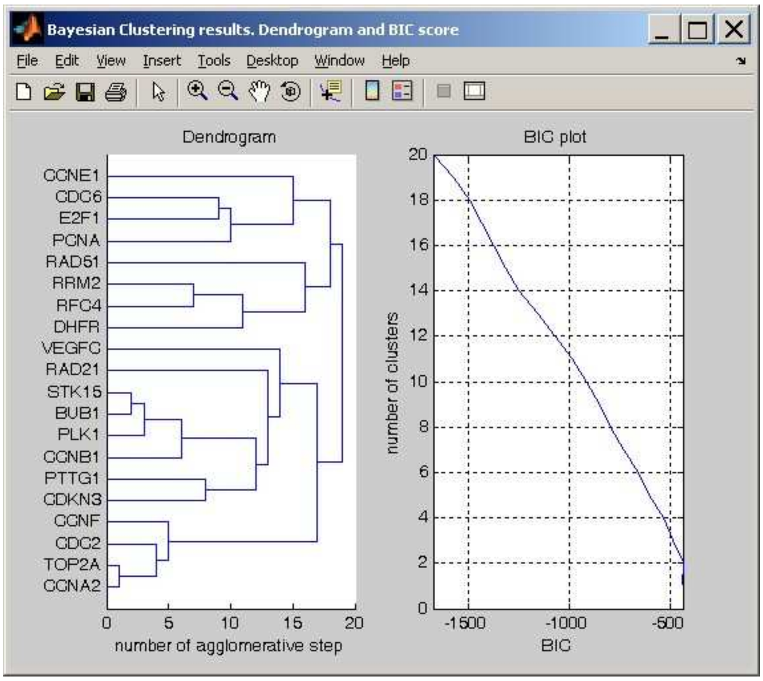


Figure 16: BC results: the dendrogram and the relationship between the BIC score and the number of clusters.

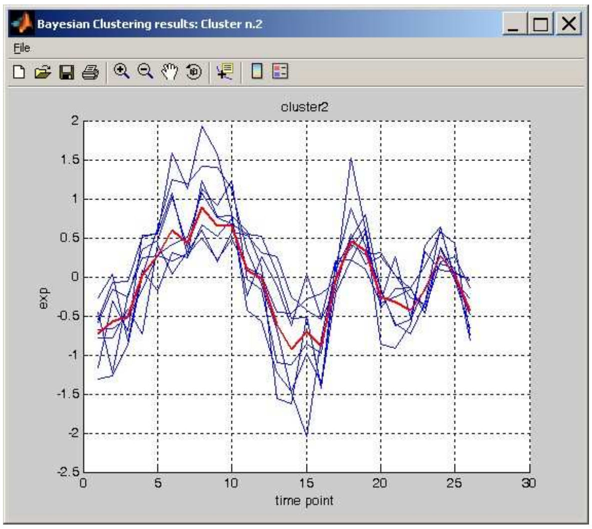
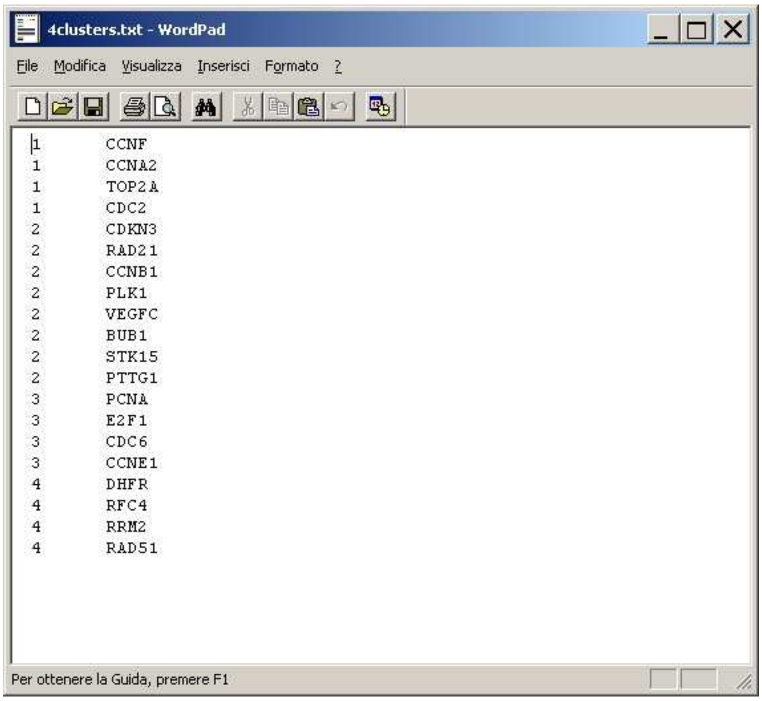
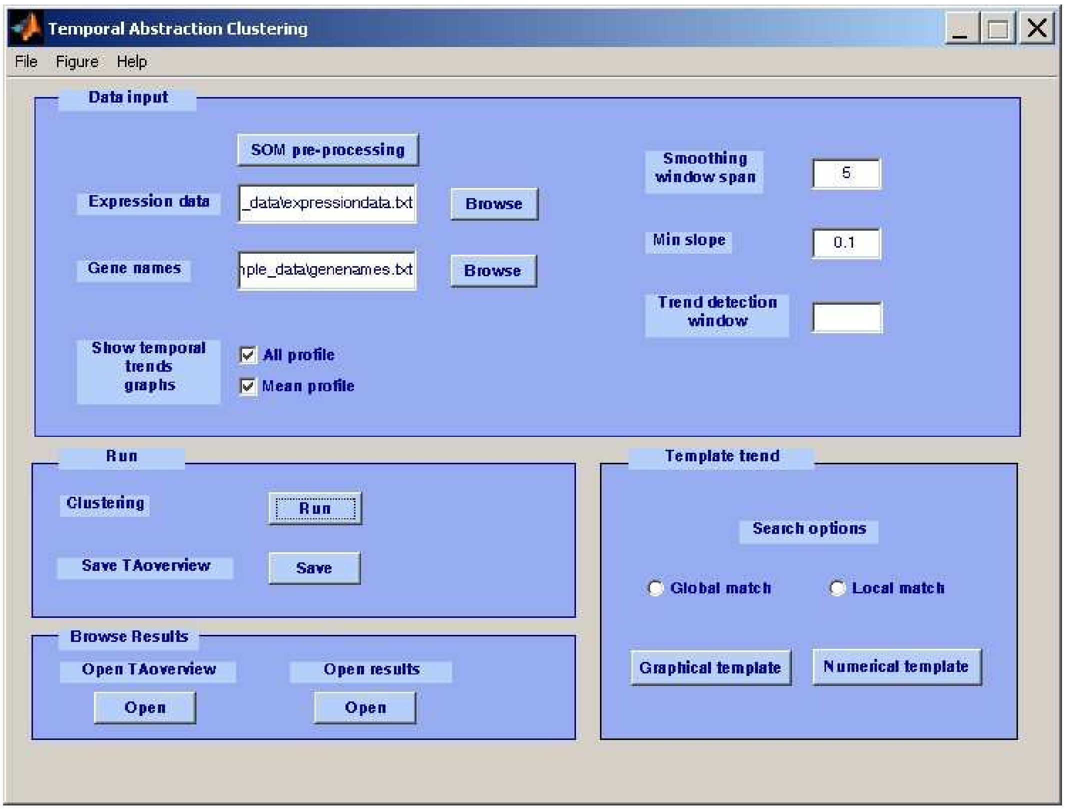


Figure 17: BC results: the expression time series of genes that are in the cluster. In red the typical profile is reported.


4clusters.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
1       CCNF
1       CCNA2
1       TOP2A
1       CDC2
2       CDKN3
2       RAD21
2       CCNB1
2       PLK1
2       VEGFC
2       BUB1
2       STK15
2       PTTG1
3       PCNA
3       E2F1
3       CDC6
3       CCNE1
4       DHFR
4       RFC4
4       RRM2
4       RAD51
```

Figure 18: BC results: the 4cluster file. In the fist column there is the number of the cluster and in the second column the gene name.



Expression data: _data\expressiondata.txt

Gene names: nple_data\genenames.txt

Show temporal trends graphs: ☑ All profile ☑ Mean profile

Smoothing window span: 5

Min slope: 0.1

Trend detection window:

○ Global match ○ Local match

Figure 19: The TAC window.

very large it can be useful to reduce the dimensionality of the clustering problem. For this reason, from the TAC window it is possible to directly open the SOM clustering window in order to pre-process the data, save the results of the pre-processing step and continue with the analysis.

As in the other clustering windows, in the *Data input* section the user has to choose the file containing the expression data to cluster (e.g. the expressiondata.txt) by clicking on the corresponding *Browse* button and browsing the file system. Then he/she can choose, in the same way, the file containing the gene names/labels. This task is not mandatory and if it is not performed, gene labels are automatically generated as already explained.

Before starting the Temporal Abstraction algorithm [3], it is often useful to smooth the data in order to make the algorithm more robust to noise. To this end, the moving average smoother has been implemented in the TimeClust tool. It computes a new time series from the original one by substituting each value of the time series with the average value computed in a window of width specified by the user and centered in the point itself. The larger the smoothing window is, the more data will be smoothed. The desired width of the smoothing window has to be specified in the corresponding text box of the TAC window. If no smoothing is desired, it is necessary to introduce a window width of length 0.

Two other important parameters of the TAC algorithms have to be specified: the minimum absolute value of the slope and the minimum length of the segment on which it has to be determined. The last parameter in general depends on the length of the time series. For this reason the default value is 3 if the time series is shorter than 10 points, 5 if the time series is longer than 10 but shorter than 50 points and the 10% of the total length for longer time series.

By selecting the appropriate check box it is possible to plot for each cluster the gene expression profiles (Fig. 20) or only the mean profile (Fig. 21) or both. For each cluster the symbolic label of the profile (-1=Decresing, 0=Steady,1=Increasing) is indicated together with the number of genes into the cluster. If one figure is few readable (labels appear to be superimposed the one on the others) to solve this problem the user can simply maximize the figure. Moreover, when clicking on these figures the list of genes in the cluster appears into the desired editor as specified in the config.txt file (Fig. 22).

If required by the user, results are saved in the *TA_Results* subdirectory of the current directory. Results are stored in a directory-tree structure. For each L1 cluster a directory is created. Each directory contains: two figures (one representing profiles of all gene belonging to that cluster and the other representing the mean expression profile), a file with the name of the genes in the cluster and as many directories as the number of L2 clusters found for that L1 label. The same structure is replicated for each L2 and L3 cluster.

If the *Save* button of the *Run* section of the TAC window is clicked, an overview of the clustering subdivision, containing a short description of the clustering results in terms of number of clusters and corresponding labels is stored in a text file. It can be subsequently viewed by clicking the *Open* TAoverview button (Fig. 23).

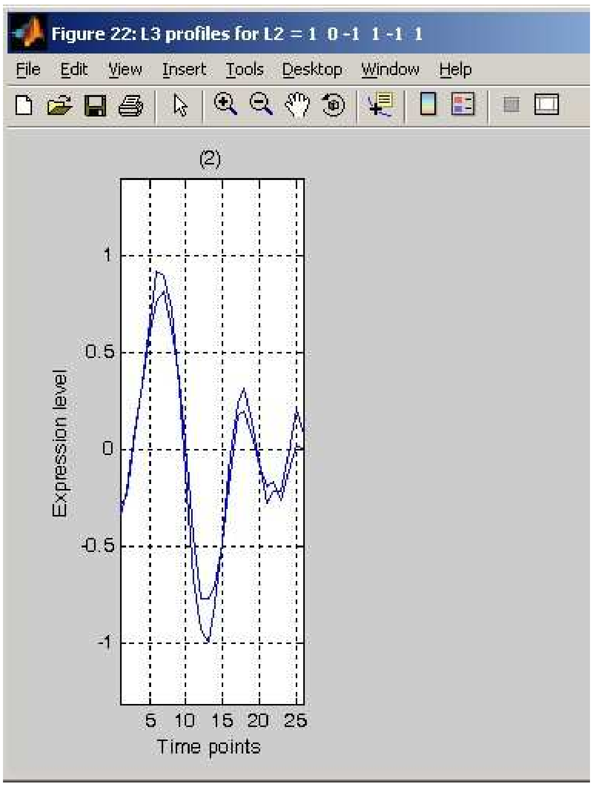


Figure 22: L3 profiles for L2 = 1 0 -1 1 -1 1
File Edit View Insert Tools Desktop Window Help
(2)
Expression level
1
0.5
0
-0.5
-1
5 10 15 20 25
Time points

Figure 20: TAC results: the expression time series of genes that are in the cluster. The label of the profile is reported in the figure title, whereas the number of genes into the cluster is indicated over the subplot between brackets.

If the results are saved in the directory-tree structure, they can be subsequently browsed by clicking on *Open* results button (Fig. 24).

The *Template trend* section of the TAC window is devoted to the search of clusters similar to a user-defined template. Thanks to this function it is possible to look for particular gene expression profiles. The algorithm searches for occurrences of the template pattern in the L2 clustering labels. Two different search methods are available and can be selected through a radio button: the *Global match* and the *Local match*. The first looks for clusters that have exactly the same label as the template, whereas the second one looks for clusters with labels containing a substring which matches the template. The template can be defined in two ways: graphically, by drawing the desired profile in a window (Fig. 25), or directly by putting in a different window the label corresponding to the desired profile (Fig. 26). In both cases, the list of clusters matching the template is shown in a new window like the one of Fig. 27. Again, by clicking on this window with the right button of the mouse the list of genes included in each matching cluster is shown (Fig. 28).

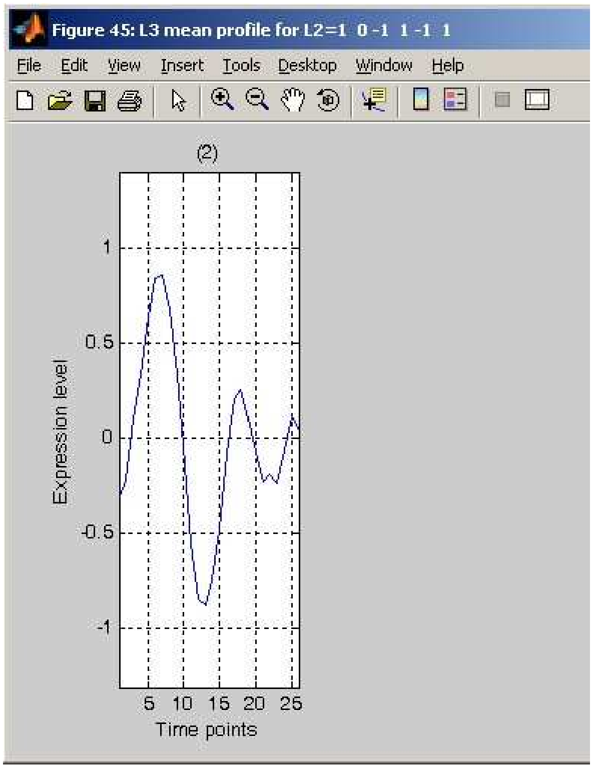


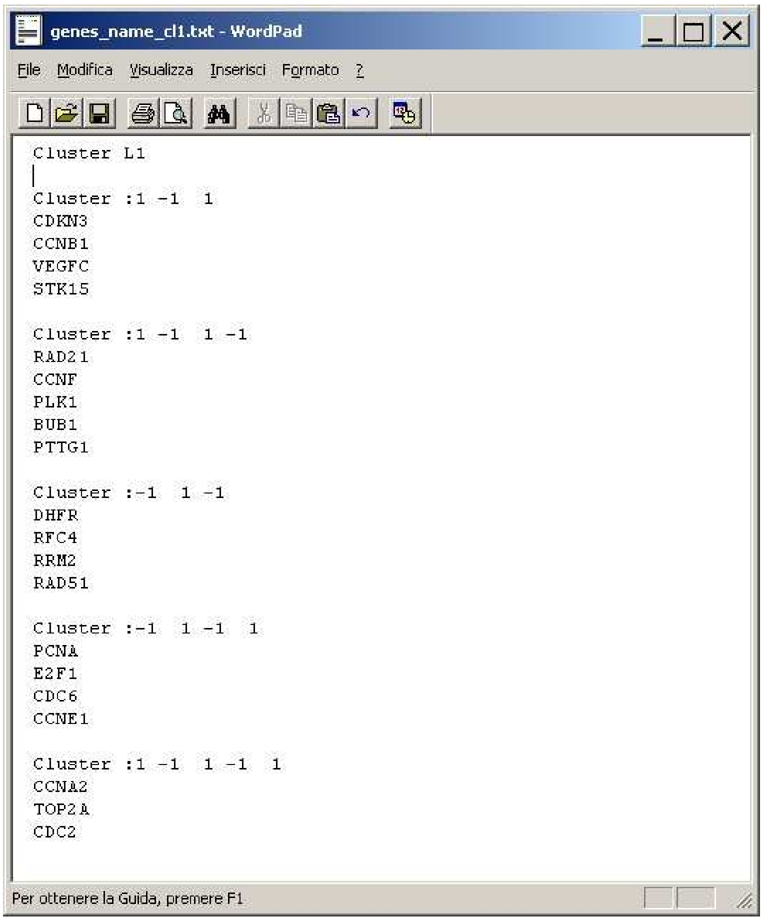
Figure 21: TAC results: the mean expression profile of a cluster. The label of the profile is reported in the figure title, whereas the number of genes into the cluster is indicated over the subplot between brackets.

References

[1] M. L. Whitfield, G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, J. C. Matese, C. M. Perou, M. M. Hurt, P. O. Brown, and D. Botstein. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell*, 13:1977–2000, 2002.

[2] F. Ferrazzi, P. Magni, and R. Bellazzi. Random walk models for Bayesian clustering of gene expression profiles. *Applied Bioinformatics*, 4:263–276, 2005.

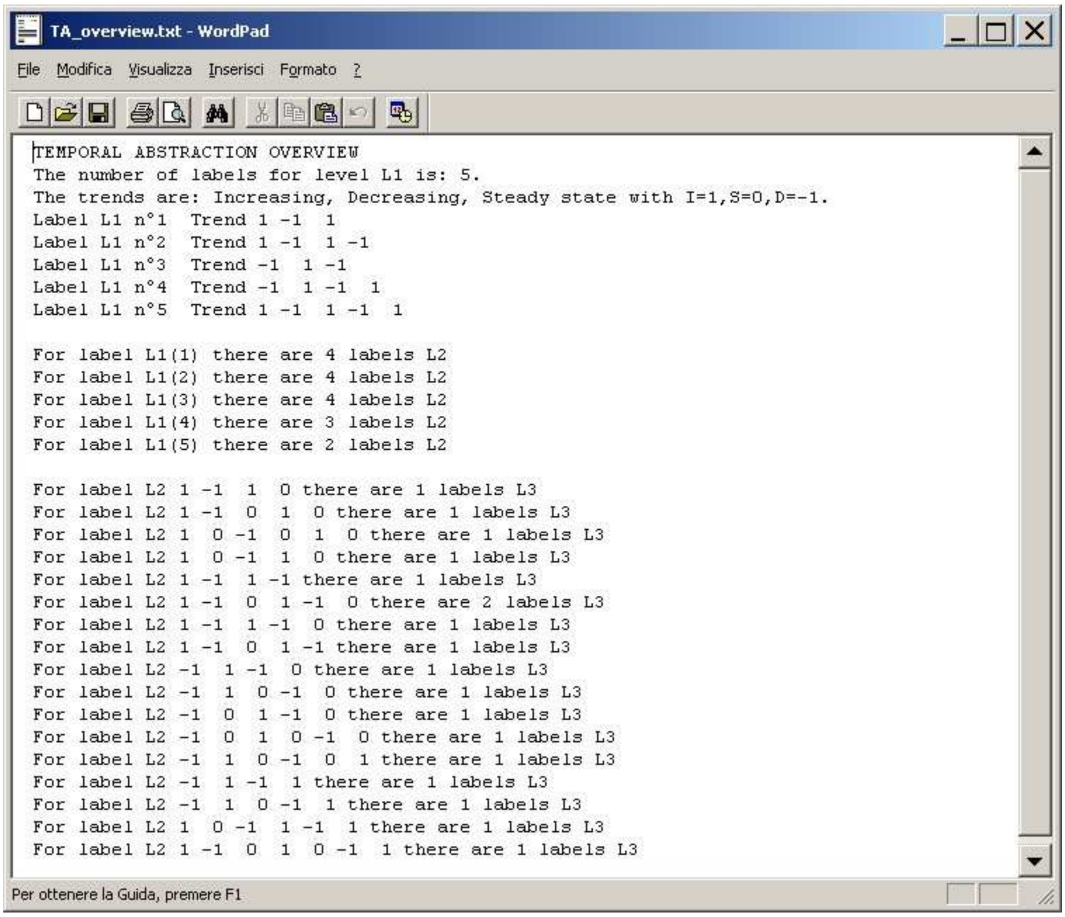
[3] L. Sacchi, R. Bellazzi, C. Larizza, P. Magni, T. Curk, U. Petrovic, and B. Zupan. TA-Clustering: cluster analysis of gene expression profiles through temporal abstractions. *International Journal of Medical Informatics*, 74:505–517, 2005. doi:10.1016/j.ijmedinf.2005.03.014.



genes_name_cl1.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
Cluster L1

Cluster :1 -1  1
CDKN3
CCNB1
VEGFC
STK15

Cluster :1 -1  1 -1
RAD21
CCNF
PLK1
BUB1
PTTG1

Cluster :-1  1 -1
DHFR
RFC4
RRM2
RAD51

Cluster :-1  1 -1  1
PCNA
E2F1
CDC6
CCNE1

Cluster :1 -1  1 -1  1
CCNA2
TOP2A
CDC2
```

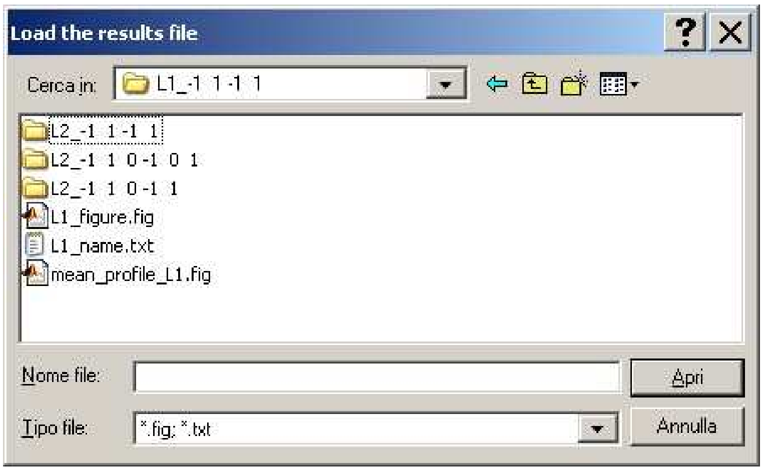
Figure 22: TAC results: the list of gene names of a cluster



TA_overview.txt - WordPad
File Modifica Visualizza Inserisci Formato ?

```
TEMPORAL ABSTRACTION OVERVIEW
The number of labels for level L1 is: 5.
The trends are: Increasing, Decreasing, Steady state with I=1,S=0,D=-1.
Label L1 n°1  Trend 1 -1  1
Label L1 n°2  Trend 1 -1  1 -1
Label L1 n°3  Trend -1  1 -1
Label L1 n°4  Trend -1  1 -1  1
Label L1 n°5  Trend 1 -1  1 -1  1

For label L1(1) there are 4 labels L2
For label L1(2) there are 4 labels L2
For label L1(3) there are 4 labels L2
For label L1(4) there are 3 labels L2
For label L1(5) there are 2 labels L2

For label L2 1 -1  1  0 there are 1 labels L3
For label L2 1 -1  0  1  0 there are 1 labels L3
For label L2 1  0 -1  0  1  0 there are 1 labels L3
For label L2 1  0 -1  1  0 there are 1 labels L3
For label L2 1 -1  1 -1 there are 1 labels L3
For label L2 1 -1  0  1 -1  0 there are 2 labels L3
For label L2 1 -1  1 -1  0 there are 1 labels L3
For label L2 1 -1  0  1 -1 there are 1 labels L3
For label L2 -1  1 -1  0 there are 1 labels L3
For label L2 -1  1  0 -1  0 there are 1 labels L3
For label L2 -1  0  1 -1  0 there are 1 labels L3
For label L2 -1  0  1  0 -1  0 there are 1 labels L3
For label L2 -1  1  0 -1  0  1 there are 1 labels L3
For label L2 -1  1 -1  1 there are 1 labels L3
For label L2 -1  1  0 -1  1 there are 1 labels L3
For label L2 1  0 -1  1 -1  1 there are 1 labels L3
For label L2 1 -1  0  1  0 -1  1 there are 1 labels L3
```

Figure 23: TAC results: the TAoverview file.



```
L2_-1 1 -1 1
L2_-1 1 0 -1 0 1
L2_-1 1 0 -1 1
L1_figure.fig
L1_name.txt
mean_profile_L1.fig
```

Nome file:
Tipo file: *.fig; *.txt

Figure 24: TAC results: the directory-tree structure in which results are stored.

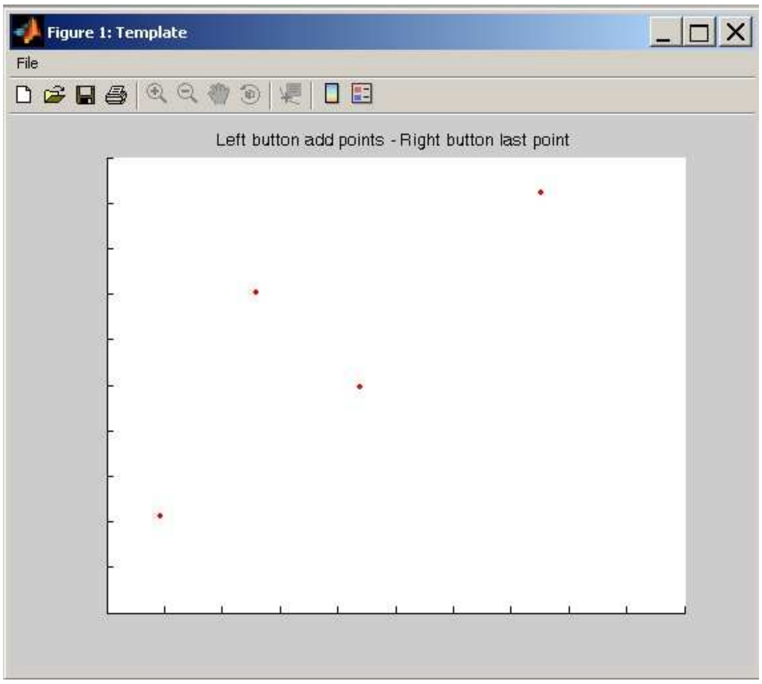
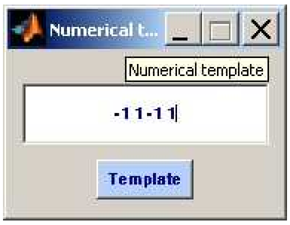
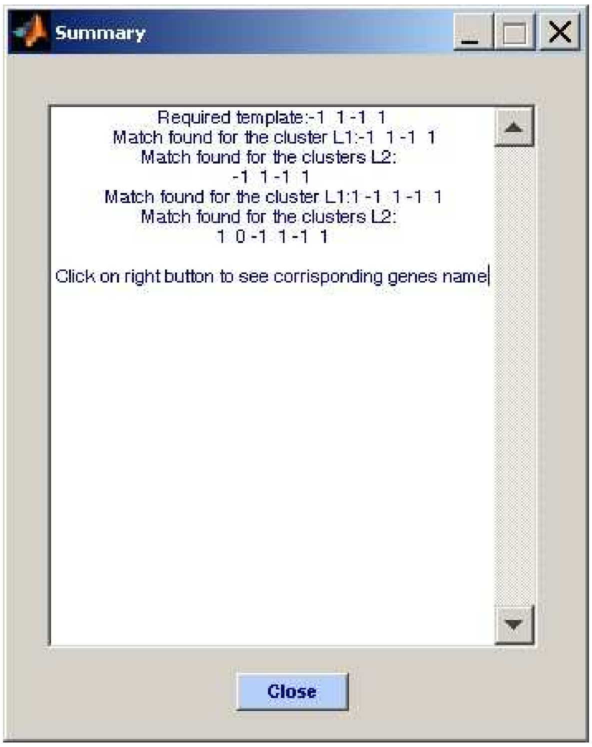


Figure 25: TAC. The window for the graphical definition of the template to be searched: 1) Build the profile adding points with the left button of the mouse; 2) Press the right button of the mouse to introduce the last point of the template: Points have to be added following the temporal order.



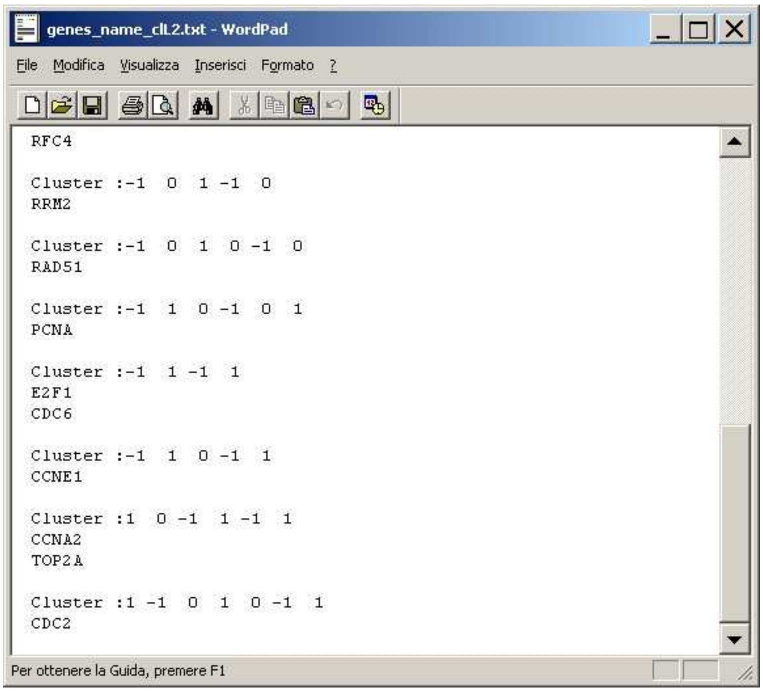
-1 1 -1 1

Figure 26: TAC. The window for the introduction of the label of the template to search. 1 indicates Increasing, 0 Steady and -1 Decreasing. Elements of the label are separated by white spaces.



```
Required template:-1  1 -1  1
Match found for the cluster L1:-1  1 -1  1
Match found for the clusters L2:
-1  1 -1  1
Match found for the cluster L1:1 -1  1 -1  1
Match found for the clusters L2:
1  0 -1  1 -1  1

Click on right button to see corrisponding genes name
```

Figure 27: The window showing the results of a template matching.


genes_name_clL2.txt - WordPad
File Modifica Visualizza Inserisci Formato ?
Per ottenere la Guida, premere F1

```
RFC4

Cluster :-1  0  1 -1  0
RRM2

Cluster :-1  0  1  0 -1  0
RAD51

Cluster :-1  1  0 -1  0  1
PCNA

Cluster :-1  1 -1  1
E2F1
CDC6

Cluster :-1  1  0 -1  1
CCNE1

Cluster :1  0 -1  1 -1  1
CCNA2
TOP2A

Cluster :1 -1  0  1  0 -1  1
CDC2
```

Figure 28: TAC results: the list of a gene names included in each matching cluster.