

Mayday 1.1 beta 1

Tutorial

Contents

1	Introduction	3
2	Installation	3
2.1	System requirements	3
2.2	Installation	4
2.3	Setting up preferences	5
3	Quick start tutorial	7
3.1	<i>Mayday</i> data organization	7
3.2	Open a data set	8
3.3	Global information	10
3.4	Context menu	11
3.5	Analyzing data	12
3.5.1	<i>k-Means</i> plug-in	12

3.5.2	Load analyzed data	15
3.6	Visualization of data	17
3.6.1	Expression image	19
3.6.2	Profile Plot	24
3.6.3	Box plot	28
A	Probe list file format - Example	30
B	Glossary	31

1 Introduction

Each microarray experiment requires comprehensive and careful analysis of the obtained data. Particularly in the field of gene expression data analysis, a number of software applications exists, supporting the experimenter or data analyst in conducting this task.

Many different methods, ranging from statistical tests to clustering algorithms, data visualization tools and highly sophisticated techniques, have been and are constantly newly developed for microarray data analysis. Especially interactive visualizations, that support the data analyst in exploring the data, can be crucial to the successful interpretation of a microarray experiment. Easy and user-friendly access to a multitude of methods is of high significance to the outcome of the data analysis.

Mayday is a freely available **microarray data analysis** platform and is designed to be a flexible solution for processing microarray data. *Mayday* features interactive data visualization as well as a very generalized plug-in framework to support analytical tools. The intended audience of *Mayday* is on the one hand researchers performing and analyzing microarray experiments and, on the other hand, researchers developing new methods for microarray data analysis.

2 Installation

2.1 System requirements

Mayday is based on the Java programming language, thus you will need at least the Java runtime environment 1.4.1 for your operating system. Please see <http://java.sun.com/> for further information, how to install Java on your machine.

Mayday offers the possibility to export visualized data into pixel and vector based picture formats like PNG, JPEG, TIFF or SVG. To do so you will need the Batik SVG Toolkit. For download and further information see <http://xml.apache.org/batik/>. If you do not want to export picture files you will not need the Batik SVG Toolkit.

2.2 Installation

- First, please get the files listed below from <http://www.zbit.uni-tuebingen.de/pas/mayday/mayday.html>.
 - *Mayday*-program files (zip file)
 - Sample data set (zip file)
 - plug-ins:
 - * `Mayday-plugin_k-means.zip`
(necessary for this tutorial!)
 - * `Mayday-plugin_simple-profile.zip`
- Unzip the *Mayday*-program files. A main directory will be created. Inside you find 3 further directories.
 - `Mayday`
 - + `Mayday-1.1-beta` (program files)
 - + `plugins`
 - + `sample`

Unzip the sample data set and the plugin files into the specific directories.

- Use the file `mayday.bat` (Windows) or `mayday.sh` (Linux/Unix/MacOS) from the `Mayday-1.1-beta`-directory to start *Mayday*. Before the first start, you need to change the file with respect to your directories as described below.
- Open the file according to your operating system with your favorite editor (e.g. notepad, emacs, vi, ...). The environment variables `MAYDAY_HOME` and `BATIK_LIB` have to be set to the related directories, e.g.

```
SET MAYDAY_HOME=C:\Mayday
SET BATIK_LIB=C:\Batik-1.5\lib
```

Listing 1: `mayday.bat`

```
MAYDAY_HOME=/home/user/Mayday
BATIK_LIB=/opt/batik/batik-1.5/lib
```

Listing 2: mayday.sh

- If you do not intend to use the Batik SVG Toolkit leave the corresponding directory name blank.
- After saving the changes, you can start *Mayday*. To do so open a command prompt (e.g. bash or Windows command prompt), change into the *Mayday* directory, and call `mayday.bat` (Windows) or `mayday.sh` (Linux/Unix/MacOS), respectively.

If you experience any problems during installation please report them to dietzsch@informatik.uni-tuebingen.de. Please mention name and version of your operating system, the Java runtime environment you use and the *Mayday* release and give a short description of your problems.

2.3 Setting up preferences

If you start *Mayday* the first time you should set up the preferences.

-
- Open *Mayday* via `mayday.bat` or `mayday.sh`, respectively.
 - Select the menu item **File** → **Preferences...**
 - On the rider **Browser**, type in your favorite browser. Probably your browser needs an URL switch. This is a command line option some browsers need to determine that a given string has to be interpreted as URL.
 - Change to the rider **Plug-ins** and set up your Plug-in directory.

(see *Figure 1*).

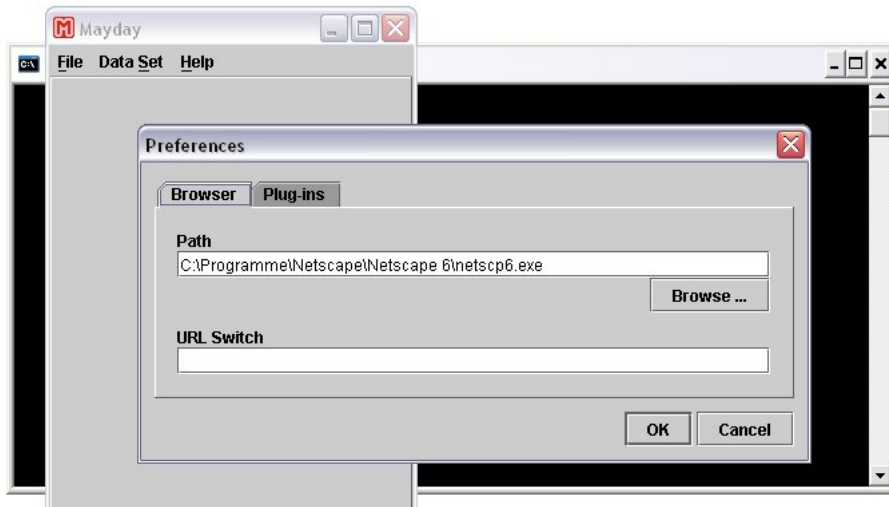


Figure 1: Edit preferences

3 Quick start tutorial

If you use *Mayday* the first time we suggest to work through the following sample session. From now on you need the sample data set (see *Section 2.2*).

3.1 *Mayday* data organization

To understand the handling of *Mayday* you need to know how the data is organized. *Mayday* is an application intended to analyze microarray data, so the underlying data set is an expression matrix. This matrix contains the expression values of microarray experiments. The values of a row belong to one probe and the values of a column belong to one experiment. Every probe (or gene profile) has an identifier. The identifiers of probes and experiments are taken from the expression matrix. They are expected in the first column (probe identifiers) and the headline (experiment identifiers).

The data structure representing the expression matrix is called master table. The results of the analysis tools working on a master table are subsets of this master table. These subsets are called probe lists. A probe list contains only the identifiers of the included probes and is internally connected to the master table. Probe lists are sets in a mathematical sense, so every probe identifier is contained only once in one probe list. However, a probe identifier can be contained in several probe lists.

The interface between *Mayday* and the analysis tools use the probe lists to refer to the expression matrix. To guarantee that at least one probe list exists, a global probe list will be created automatically. This global probe list consists of all probes.

3.2 Open a data set

- Select the menu item **Data Set** → **Open...**
- Find the directory where the sample data set has been stored.
- Open the file `Spellman_alpha_25.dat`.

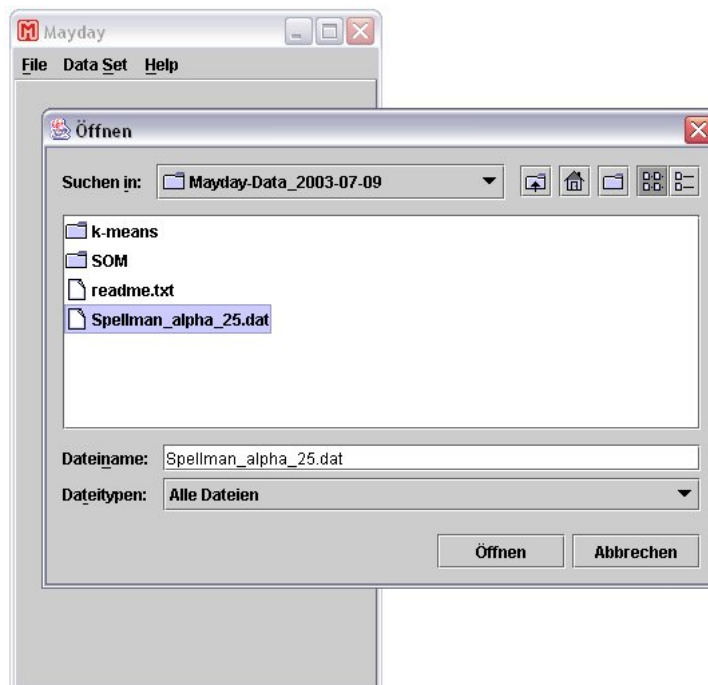


Figure 2: Open a data set

The file contains a tab-separated matrix of expression values from 528 yeast genes. It was extracted from Spellman's experiment to identify cell-cycle-regulated genes of the yeast *Saccharomyces cerevisiae* [3].

- Type in a name for the data set, or confirm by pressing **<Enter>**.
- Choose the data mode *log₂ ratio*.

The data mode is a parameter that *Mayday* needs to interpret the data in order to identify allowed and forbidden operations. It determines whether the file contains absolute, logarithmic, or ratio values. What data mode you choose depends on the loaded data set.

Now a global probe list has been created.

If you want to change the name of the data set and give some explaining information, do the following:

- Open the menu item **Data Set** → **Properties...**

Quick Info is meant to contain a short description of the data, e.g. one short phrase or sentence.

Info is meant to contain some further information, maybe a whole article formatted in HTML.

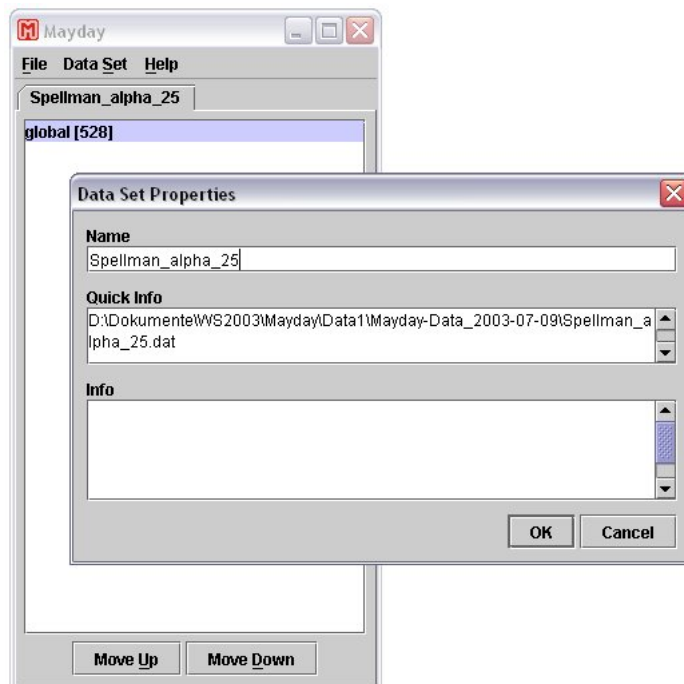


Figure 3: Data Set Properties

3.3 Global information

- Double click on the *Spellman_alpha_25* rider to get some information about the data set. You will see minimum, maximum etc.

For example, you see that the *Spellman*-data consists of 528 probes (genes) and for every probe there are 17 experiments.

In *Mayday* there is a difference between explicit and implicit probes. Explicit probes are those read from an input file. Implicit probes are implicitly contained in the expression matrix such as the mean over all explicit probes or the centers of a k-means cluster. In *Figure 4* you see an explicit global maximum/minimum which means the maximum/minimum of the whole expression matrix. Here there is no implicit probe yet, so there is no implicit maximum/minimum.

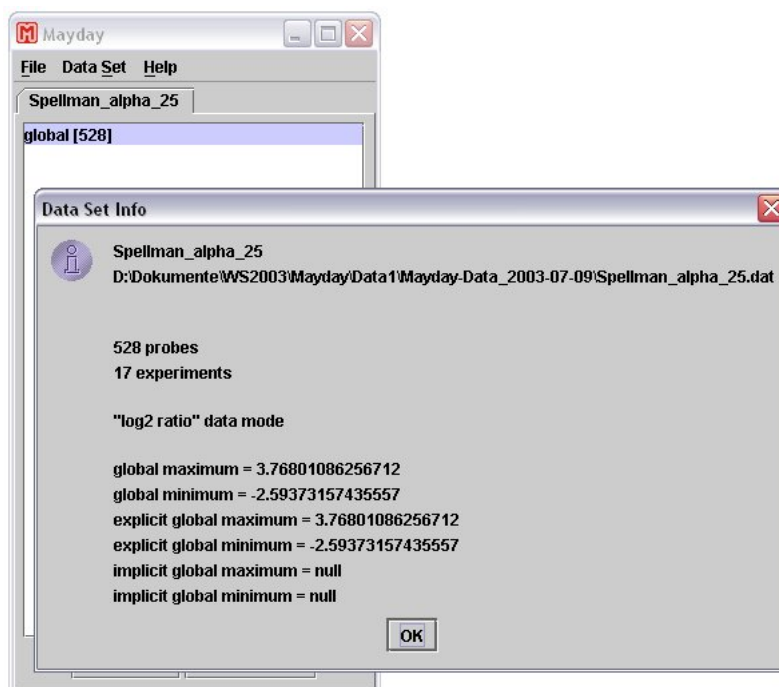


Figure 4: Data Set Info

3.4 Context menu

An important concept of *Mayday* is the context menu which can be opened by clicking the **right mouse button**. The context menu offers almost the whole functionality of *Mayday*.

- On the *global*-entry, click the **right mouse button**.

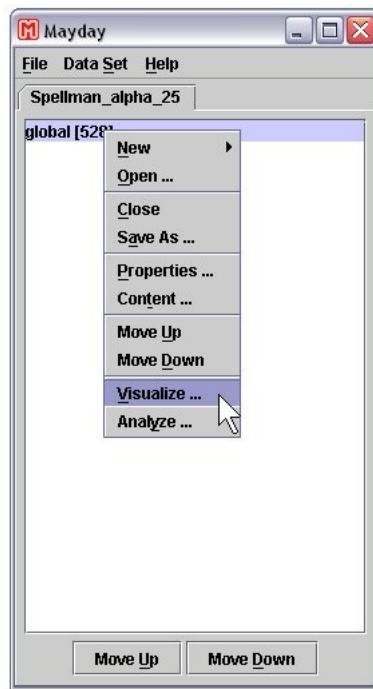


Figure 5: The context menu

3.5 Analyzing data

Mayday has two possibilities to get analyzed data. The first is to analyze the data via plug-ins, for example the *k-means* cluster plug-in. The second possibility is to load pre-analyzed data from a file (see *Section 3.5.2*).

Mayday provides a flexible mechanism to integrate a multitude of established and new data analysis methods. Plug-ins are a concept to capture distinct functional units in interchangeable software modules. The collaboration of these software modules is the basis for the functionality of the whole platform.

On the one hand, the plug-in interface allows *power users* to customize *Mayday* to their needs and on the other hand, it is possible for experts to test their new methods within an existing infrastructure for handling and visualization of data. For available plug-ins check the URL <http://www.zbit.uni-tuebingen.de/pas/mayday/mayday.html>.

3.5.1 *k-Means* plug-in

To apply the *k-means* cluster algorithm:

-
- Click the right mouse button (context menu) over the *global [528]* entry.
 - Open the menu item **Analyze...**

There you will find the Analyzer (see *Figure 6*) which contains all usable plug-ins of your *Mayday*-installation, sorted by categories.

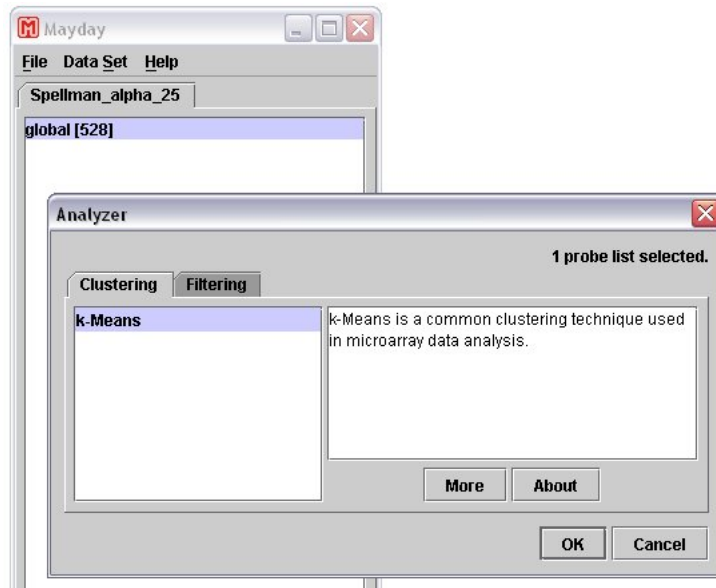
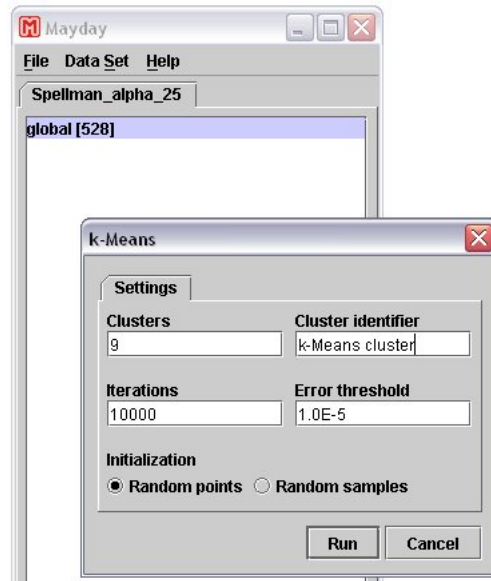


Figure 6: Analyzer

- Select the rider **Clustering** (see *Figure 6*).
- Choose the entry *k-Means* and press **OK**.
- Set the number of clusters to 9 (see *Figure 7*).
- Press **Run** to confirm.

Other parameters are the identifier for storing the resulting clusters in the master table, the iterations and the error threshold. At the bottom you can select which kind of method should be used to generate the initial cluster centers. *Random samples* means that the centers are randomly chosen from the given data set. *Random points* computes virtual centers.

Figure 7: *k-Means* parameters

The resulting 9 clusters will be shown in the master table with different colors. These colors are used for the visualizations. To change the color by hand you can use the probe list properties which are available via the context menu (*right mouse button*).

Notice that almost every functionality is accessible via the right mouse button context menu.

3.5.2 Load analyzed data

Mayday provides the possibility to load pre-analyzed data from a file. Imagine an analysis procedure that is only available with a third party software. *Mayday* allows to visualize the results of this procedure, given that it is saved in a format that *Mayday* can read. The input file contains several clusters given by a grouping of the probe identifiers which must be stored in the XML-based probe list file format. Since probe lists (that only contain probe identifiers that occur in the expression matrix) are the central data concept in *Mayday*, you have to guarantee that only those identifiers occur in the probe list file that can also be found in the expression matrix. An example for a probe list file is given in *Appendix A*.

- Select the **context menu**.
- Select the item **Open...** (*NOT* the **Data Set** → **Open...**-menu).
- Change to the Spellman directory, there you can find the SOM directory.
- Open all 9 files.

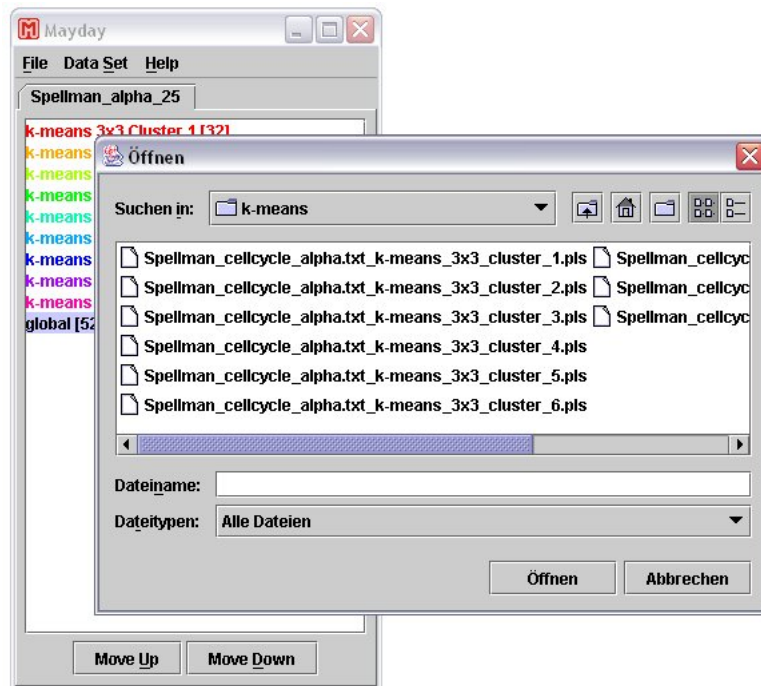


Figure 8: Load pre-analyzed data

3.6 Visualization of data

So far, three different graphical viewers are implemented in *Mayday*: profile plot, box plot, and expression image (heatmap). They are available via the menu item **Viewers** → **New**.

- Select all clusters.
- Open the **context** menu.
- Select the item **Visualize...**

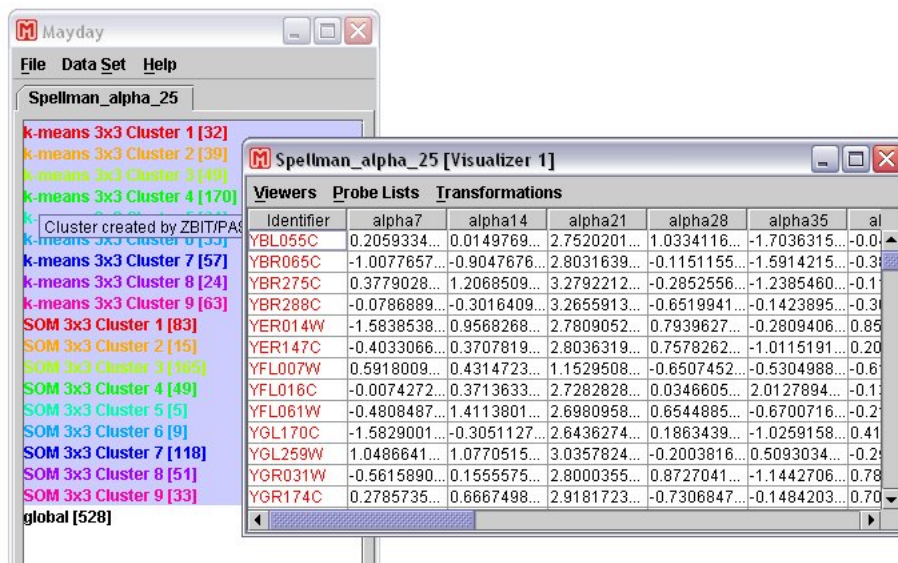


Figure 9: The visualizer

You can see the expression matrix in a tabular view. You see only those probes which you chose in the master table. For example, if the first cluster of the *k-means* analysis is selected, you will only see the 32 probes of this cluster in the tabular view (see *Figure 9*). Additionally you should know that every probe will occur only once in this table, no matter how often it is contained in different clusters.

An important feature of all views is the export to different file formats. The tabular view of the expression matrix can be exported to a plain-text file, so that it can be opened with Excel e.g.

- Select the menu item **Viewers** → **Table** → **Export...**

The graphical views can be exported to several graphic formats (see *Section 3.6.1*).

3.6.1 Expression image

The expression image is often called *heatmap*. The heatmap visualizes the expression matrix by coding the expression values of a probe with a given color palette (see *Figure 10*).

- In the *Visualizer* select the menu item **Viewers** → **New** → **Expression Image**.

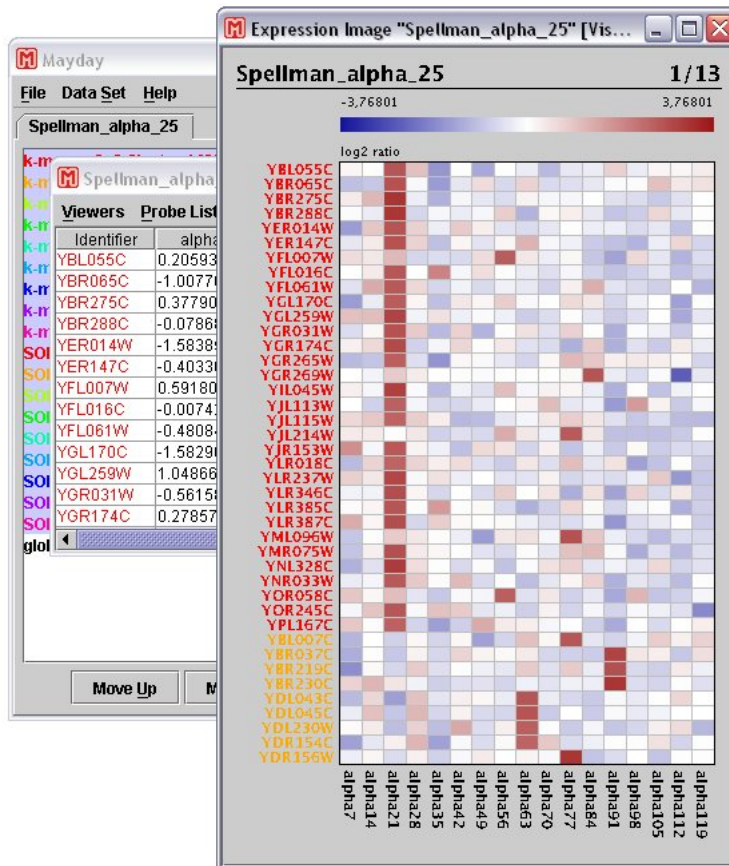


Figure 10: Expression Image

The heatmap will be scaled automatically (see *Figure 10*).

- Press the hotkeys <**CTRL**><**+**> or <**CTRL**><**-**> to zoom in or out.
- Double click on the expression image to get further information about a specific probe.

You will see the expression value and the probe lists containing this probe, such as the cluster names, etc.

- Open the menu **context menu** → **Settings**.

Color Via the **Color**-menu item you can change the color range of the expression image, maybe you prefer the widely-used green/black/red palette.

Probes Per Page Via the **Probes Per Page** → **User-defined...**-item it is possible to modify the number of probes shown on one page.

You will get the heatmap of the whole expression matrix on the current page, if you set the number to 528 (or above).

An important feature is to export this image to a file. To export an image from *Mayday* to a picture file format you need an installed Batik SVG Toolkit (see *Section 2.2*). **Notice that only the shown page will be exported.**

- Open the menu item **context menu** → **Export**.

You are able to choose between different file formats, SVG as a vector based format, the others are pixel based.

A feature in all views allows to create new probe lists by selecting probes in the expression image.

- Hold the <**CTRL**> key down and click on the image. You can select a number of probes.
- Apply the **context menu** → **Probe List from Selection...**-item to get a new probe list.

Again, you can edit the name of the probe list, provide a short description or change the color of the newly created probe list (see *Figure 11*).

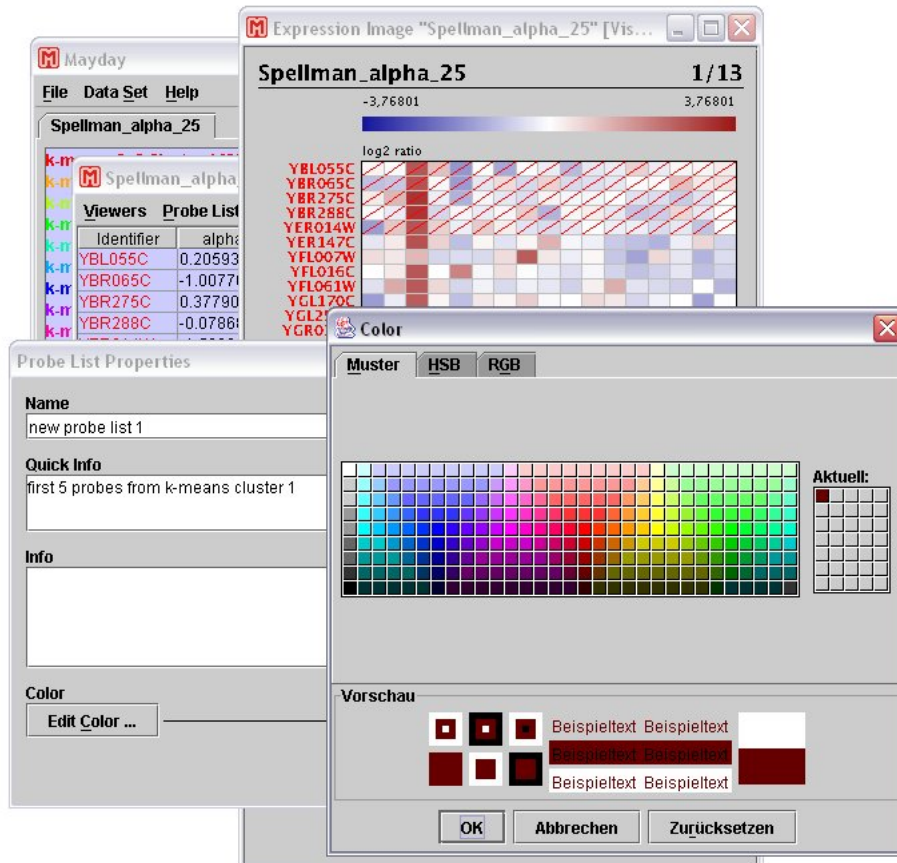


Figure 11: Choosing color

The new probe list is immediately added to the master table. The color of the selected probes has changed to the color chosen in the step before. This change has an effect in the tabular view of the visualizer and in the viewer.

The color in which the probe identifiers are displayed depends on its membership in the probe lists and the position of the related probe lists in the master table. One probe can be a member of more than one probe list.

For example, the second probe in the expression image with the identifier *YBR065C* is a member in 4 probe lists (*new probe list 1*, *k-means cluster 1*, *SOM 3 × 3 cluster 7*, and *global*). The assigned colors for this probe are dark red, red, blue, and black, respectively. The order is important, because one probe gets the color of the highest priority probe list. The order is taken from the master table. That is why the color of the first 5 probe identifiers in the heatmap has changed.

- Bring the main frame of *Mayday* to the front.
- Select *new Probe List 1* and use the **Move Down** button to bring the selected probe list down to one before the global probe list.

You will notice that the color of the first five probes will turn back to red, because now the highest probe list is *k-Means cluster 1*. This ordering is important not only for coloring the probe identifiers, but also for the order in the *Visualizer* and in the viewers.

- Select all *SOM clusters*.
- Move them on top of the main frame (**Move Up**-button).

For the result see *Figure 12*.

- Close the expression image viewer.

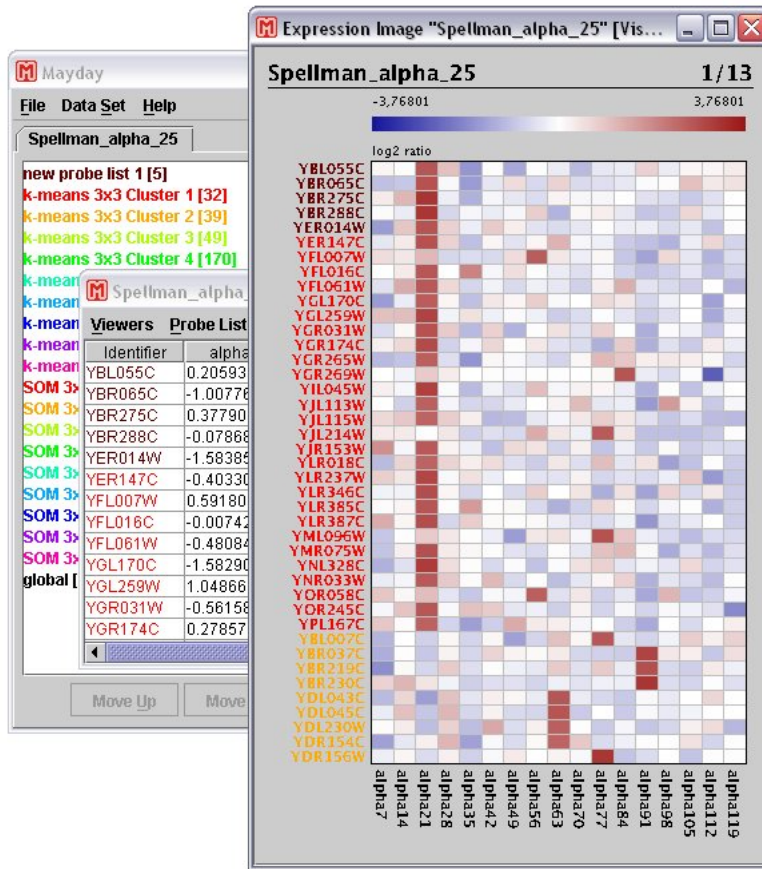


Figure 12: Expression Image, changed identifier's color

3.6.2 Profile Plot

Another important kind of visualization method is the profile plot. You are able to open a single or a multiple profile plot. *Single* means all probes are plotted in the same diagram. The multiple profile plot can show several plots simultaneously.

- Open the menu item **Viewers** → **New** → **Profile Plot** → **Single**.

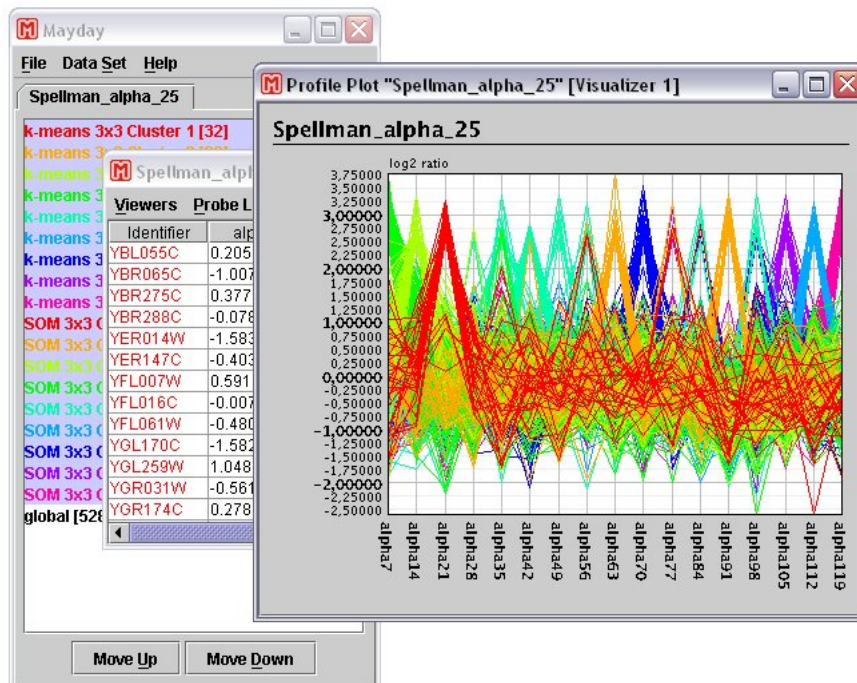


Figure 13: Single Profile Plot

- On this view, apply the **context menu** → **Export...** item to Export the view.
- Close the single profile plot.
- In the Visualizer, open the menu item **Viewers** → **New** → **Profile Plot** → **Multi**.

- Type in the number of diagrams (grid-dimensions) to plot in. Here type 3×3 to plot the 9 profile plots of the clusters computed by *k-means*.
- Make sure that all 9 *k-Means* clusters are spread over the 9 grid cells (see *Figure 14*).

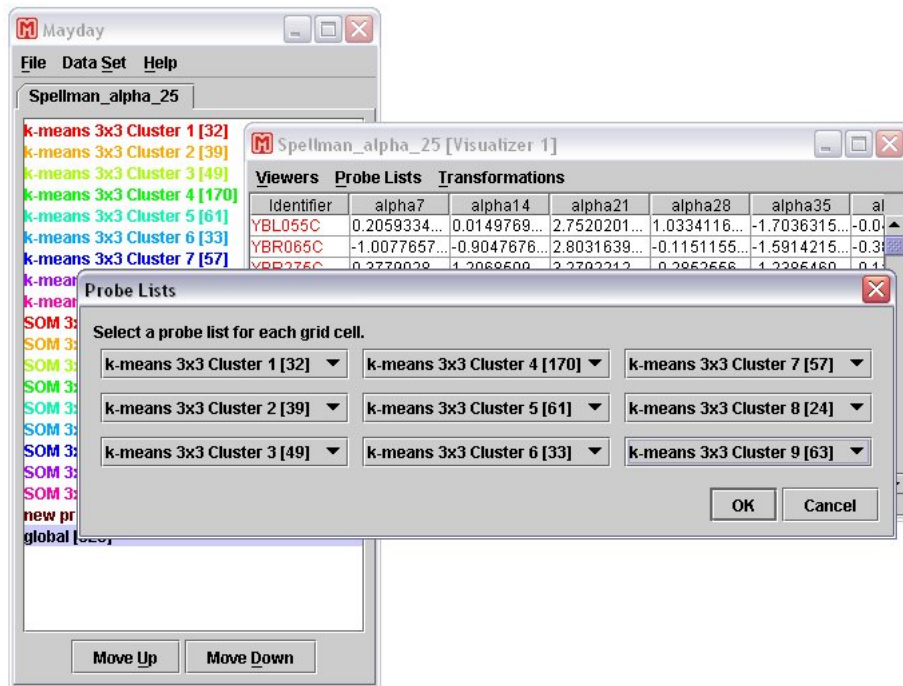


Figure 14: Select a probe list for each grid cell

The result is a 3×3 grid with 9 profile plots (see *Figure 15*).

- To zoom in press **<CTRL><+>**.
- Click on a profile to to select a probe.

The selected probe will be marked in red color.

- Or, open the **context menu** \rightarrow **Go To** \rightarrow **Probe...** and type in the probe identifier which you are interested in, e.g. *YBR065C* (see *last section*).

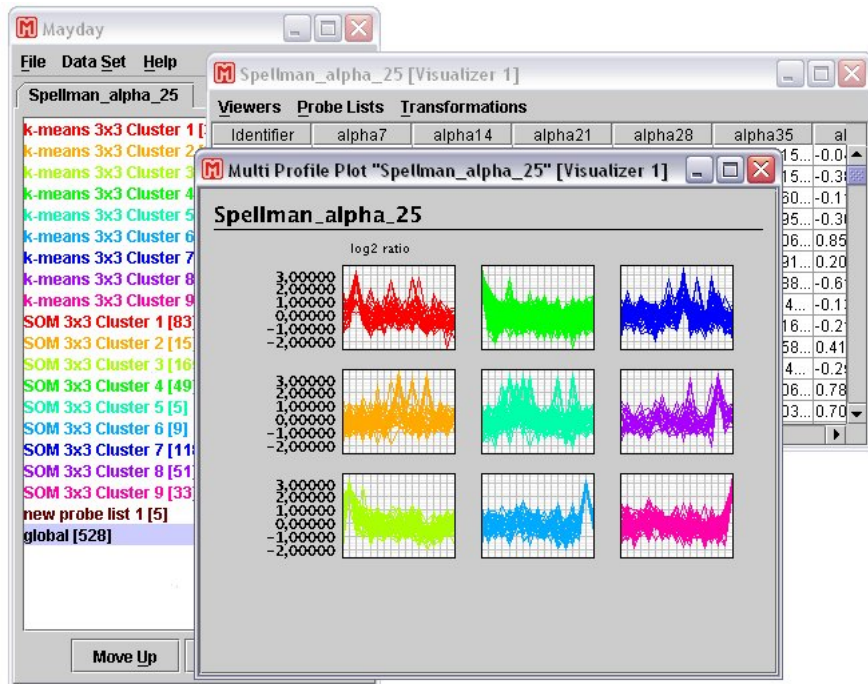


Figure 15: Multi Profile Plot

Remember the color priority ordering from the last section. This allows to compare the results of two different clusterings, for example.

- Click on the main frame of *Mayday* and bring the *SOM* cluster probe lists to the top of the master table (**Move Up**-button).

The colors of the plotted lines change immediately. Subplots with only few different colors show that the two clusters of the different algorithms are very similar to each other, while many colors in one subplot represent a large diversity between the two algorithms (see *Figure 15*). Notice that the movement of probe lists can take some time. The reason is that the plots are recalculated.

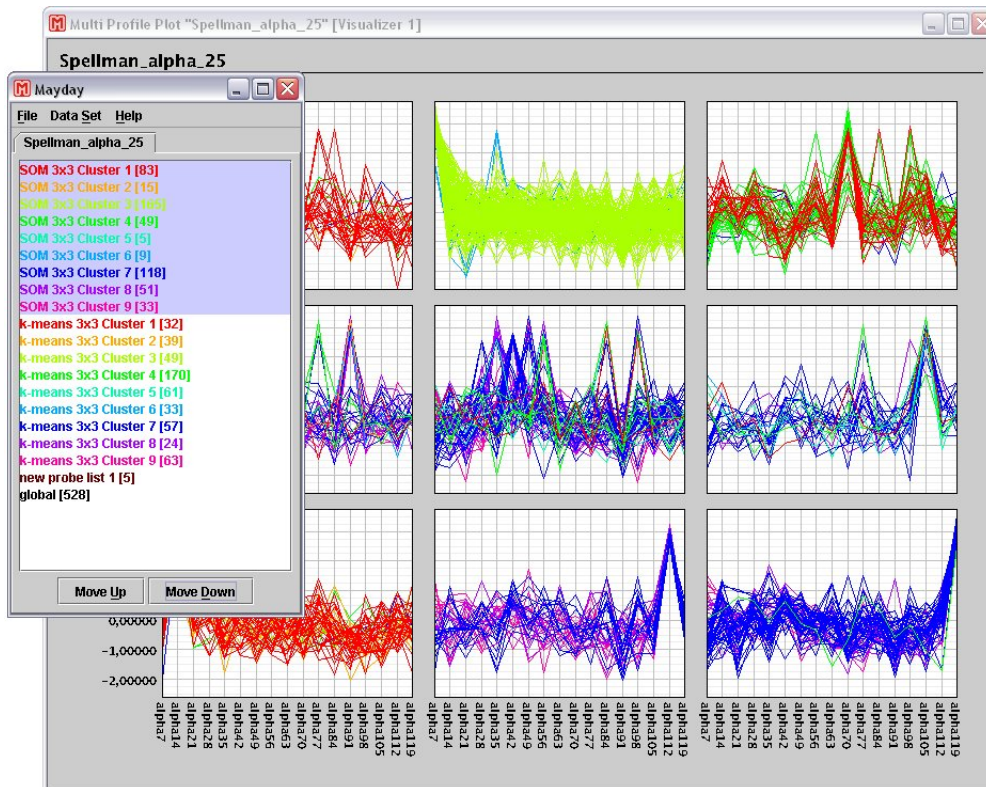


Figure 16: Layers

Have a look at the grid in the middle of the top row. There you can see profiles of two different colors. The blue curves are somewhat hidden. This is the result of the layer concept realized in *Mayday*.

- In order to bring them to the front open the **context menu** (on this subplot) → **Layers** → **SOM 3×3 cluster 6** → **Bring To Front**.

Now the blue curves are on top of the green ones.

3.6.3 Box plot

The box plot is a method, often used in statistics, to investigate data variation. For every experiment there is a bar chart representing the minimum, maximum, median, 1st quartile and 3rd quartile over all probes of a specific probe list.

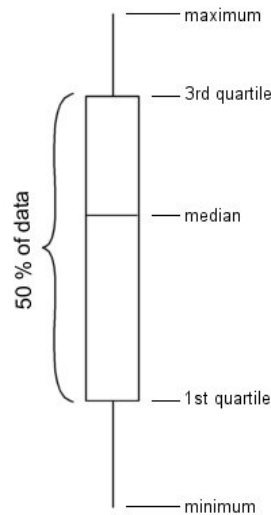


Figure 17: The box plot

The box plot is the third implemented viewer in *Mayday*. It gives a visual overview of the complete data set and offers an easier identification of the difference between several probe lists. The box plots can be shown in single and multiple mode.

-
- Use the menu item **Viewer** → **New** → **Box Plot** → **Multi**
 - Open an 1×2 box plot.
 - Choose *SOM cluster 1* and *SOM cluster 6* to discover the differences of these two clusters. (see *Figure 18*)

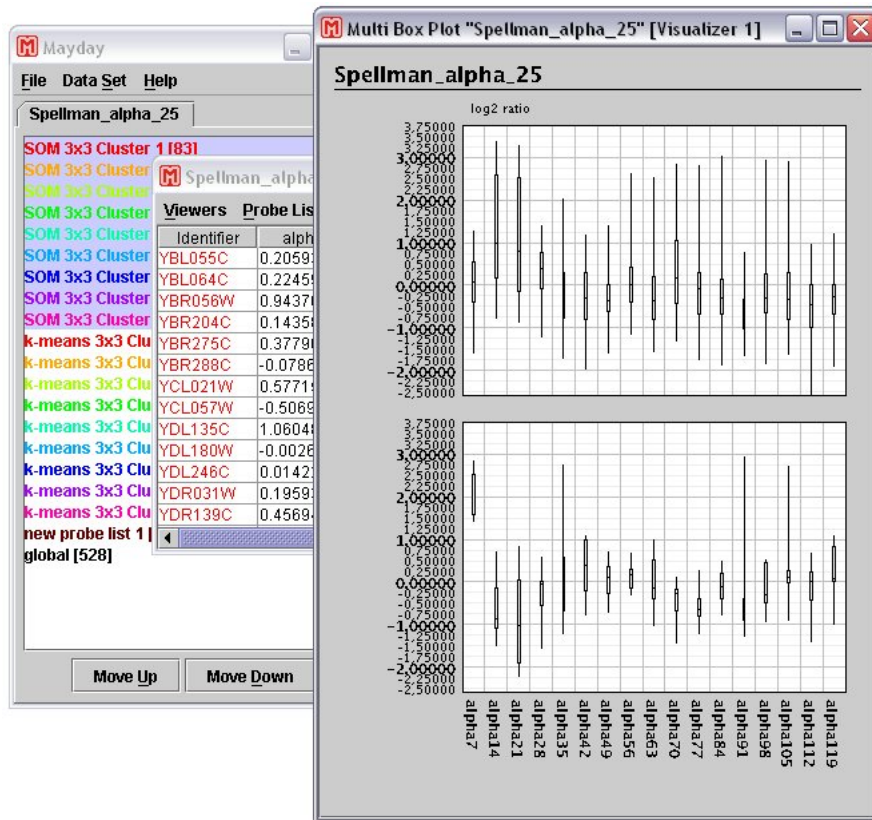


Figure 18: Multi Box Plot

A Probe list file format - Example

In the following you can see an example for a probe list file. If you want to import analyzed data, e.g. clusterings, it is necessary to bring your results in this XML-based format.

```
<?xml version="1.0"?>

<!-- Created by ZBIT/PAS clustering tool -->

<!DOCTYPE probelist SYSTEM "probelist.dtd">

<probelist>
  <annotation>
    <name>
      SOM 3x3 Cluster 2
    </name>
    <quickinfo>
      Cluster created by ZBIT/PAS clustering tool.
    </quickinfo>
    <info>
      <![CDATA[ ]]>
    </info>
  </annotation>

  <layout>
    <color>
      #FFAA00
    </color>
  </layout>

  <probe>YBR230C</probe>
  <probe>YBR298C</probe>
  <probe>YBR299W</probe>
  <probe>YER037W</probe>
  <probe>YER150W</probe>
  <probe>YGL117W</probe>
  <probe>YGR248W</probe>
  <probe>YJR153W</probe>
  <probe>YKL148C</probe>
  <probe>YLR387C</probe>
  <probe>YNL241C</probe>
  <probe>YNL274C</probe>
  <probe>YOL016C</probe>
  <probe>YOR347C</probe>
  <probe>YPL222W</probe>
</probelist>
```

Listing 3: Spellman_cellcycle_alpha.txt_SOM_3x3_cluster_2.pls

B Glossary

Box Plot The box plot is a method, often used in statistics, to investigate data variation. For every experiment there is a bar chart representing the minimum, maximum, median, 1st quartile and 3rd quartile over all probes of a specific probe list.

Color Priority Probes can be contained in several probe lists. Every probe list has its representing color. The color priority defines which color will be used to print the related probe. It depends on the probe's membership in the probe lists and on their position in the main frame of *Mayday*. The probe will be displayed in the color of the topmost probe list in the main frame.

Context Menu Will be opened by clicking the right mouse button over a certain object. In *Mayday* almost the whole functionality is accessible via the context menu.

Data Export You have the possibility to export parts of the expression matrix to plain-text, probe lists to probe list files and graphical views to picture formats like SVG (scalable vector graphics [1]), JPEG / TIFF / PNG (pixel-based formats).

Data Import You have two possibilities to import data. The first is to load an expression matrix as a new data set. *Mayday* assumes that this file has a headline, a first column with the probe identifiers, and a tab-separation of the expression values. The second possibility is to load analyzed data from probe list files (*.pls). The probe list file format is an XML-based file format, an example is shown in *Appendix A*.

Data Set The data set is the topmost organizational unit in *Mayday*. Each master table belongs to exactly one data set. It is possible to open more than one data set, but they are completely independent of each other and strictly separated.

Expression image The expression image is a visualization method that plots the expression values in an expression matrix-like style. The values are color-coded. The rows represent the probes and the columns represent the experiments.

Heatmap See *Expression image*.

Layer Every plot in a viewer consists of several layers. Each probe list defines one layer. It is possible to bring a specific layer to front or to hide layers.

Master Table The master table is the basic data structure representing the expression matrix. It contains all probes with their identifiers and expression values. (see *Section 3.1*)

Plug-ins *Mayday* provides a flexible mechanism to integrate analysis methods. Modules following the plug-in interface are called plug-ins. For available plug-ins see <http://www.zbit.informatik.uni-tuebingen.de/pas/mayday/mayday.html>

Probe Formally spoken, probes are the rows of the expression matrix. A probe represents a gene or an EST of a microarray experiment.

Probe List A probe list is a data structure representing subsets of the master table, e.g. clusters.

Probe lists contain only the probe identifiers and are internally linked to the master table. They are the important data structure for plug-ins to interact with the master table. (see *Section 3.1*)

A probe list can only contain probes that are present in the master table.

Profile Plot A profile plot of a probe is a two-dimensional plot of its expression values as a function of the experiment. The several points in this graph are connected with each other by lines.

Short-cuts

Zoom in	<CTRL><+>
Zoom out	<CTRL><->
Adjust window size to content size	<CTRL><F>
In the Expression image window	
Next page	<CTRL><PicDown>
Previous page	<CTRL><PicUp>
First page	<CTRL><Pos1>
Last page	<CTRL><End>

SVG See *Data Export*.

Viewer Viewers are structures managing the graphical display of the data. In *Mayday 1.1* three viewers are available: profile plot, expression image and box plot. (see *Section 3.6*)

Visualizer The visualizer is a structure managing the visualization of data. The visualizer window contains the expression values in a tabular view. Different viewers are accessible from it.

References

- [1] Batik SVG Toolkit; <http://xml.apache.org/batik>
- [2] *Mayday 1.1*; <http://www.zbit.uni-tuebingen.de/pas/mayday/mayday.html>
- [3] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher, *Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization*, *Molecular Biology of the Cell* 9 (1998), 3273-3297.
- [4] Sun Microsystems Inc. 1995-2003; <http://java.sun.com>