

## Contents

1. Introduction.....	3
2. General description .....	4
2.1. Main window .....	4
2.2. Main menu commands and toolbar buttons .....	4
2.3. Main window shortcut menu .....	7
2.4. Format of the file with a gene expression table.....	9
2.4.1. Basic format of the SelTag input file.....	9
2.4.2. Groups of fields .....	9
2.5. Format of the file with a gene description .....	10
3. Working with the application.....	11
3.1. Loading data .....	11
3.1.1. File with a gene expression table.....	11
3.1.2. File with gene descriptions .....	12
3.1.3. File with the nucleotide sequences of genes .....	13
3.2. Managing groups of fields (experiments).....	16
3.2.1. Viewing groups.....	16
3.2.2. Adding a new group .....	16
3.2.3. Editing a group .....	18
3.2.4. Deleting a group.....	19
3.3. "Field selection" dialog box .....	20
3.3.1. Dialog box elements.....	21
3.3.1.1. List of data types.....	21
3.3.1.2. List of field groups.....	21
3.3.1.3. List of fields.....	22
3.3.1.4. Control buttons .....	22
3.4. Selecting genes by query.....	22
3.4.1. Dialog box elements .....	22
3.5. Quick search.....	26
3.6. Selecting genes with similar expression profiles .....	27
3.6.1. Specifying gene sets .....	27
3.6.2. Setting up calculation parameters .....	28
3.6.3. Query results.....	30
3.7. Searching for correlation between two sets of genes .....	33
3.8. Searching for correlation between two sets of fields.....	36
3.9. Correlation matrix dialog box .....	39
3.9.1. "Setup for correlations graph" dialog box .....	39
3.10. Building a gene tree.....	42
3.11. Building a field tree .....	44
3.12. "Tree diagram" dialog box.....	46
3.12.1. "Mark nodes" dialog box.....	47
3.12.2. "Pattern setup" dialog box .....	48
3.12.1. "Scalable tree diagram" dialog box.....	50
3.13. Gene clustering.....	53
3.13.1. Clustering genes or fields by the Ben-Dor algorithm .....	55
3.13.2. Clustering by the SOM algorithm.....	57
3.14. Principal component analysis .....	59
3.15. "Graph" dialog box .....	62
3.15.1. "Operations with marked items" dialog box .....	63
3.15.2. "PCA" dialog box .....	63
3.16. Expression map.....	69
3.16.1. "Pattern setup" dialog box.....	70

3.16.2. “Image options” dialog box .....	71
3.17. “Expression profiles” dialog box.....	76
3.17.1. Graph .....	76
3.17.2. Information pane .....	76
3.17.3. Experiments view modes .....	77
3.17.4. Gene list.....	77
3.17.5. List of experiments .....	78
3.17.6. The “Options” dialog box.....	79
3.18. “Select color” dialog box .....	82
3.19. “Select font” dialog box .....	83

## 1. Introduction

SeITag is one of the best-of-breed tools for analyzing gene expression data.

SeITag allows users to:

- Select genes by their expression levels or other parameters obtained in various experiments
- Sort genes by their expression levels determined in various experiments
- Visualize gene expression profiles using advanced graphical tools
- Search for genes with expression profiles similar to one or more genes
- Assess the correlation between expression profiles of two or more genes
- Cluster gene expression data by the similarity of their gene expression profiles and experiments by the similarity of gene expression levels
- Perform hierarchical clustering of genes or tissues and display the results as similarity trees
- Analyze the correlation and covariance matrices of gene expression profiles using the principal component method and visualize the results obtained
- Integrate with external databases to retrieve the data for analysis (for example, UniGene database)

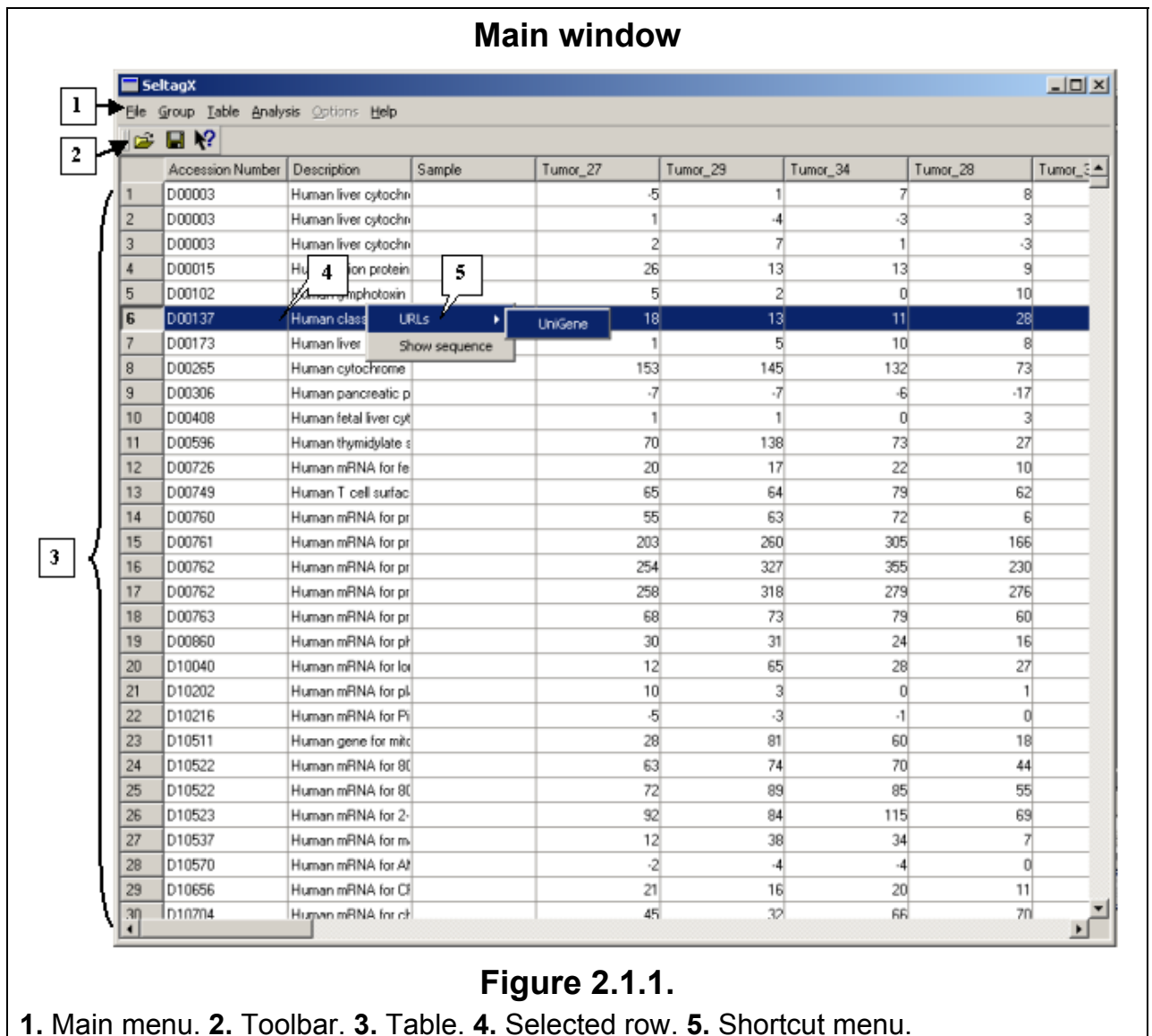
Thus, using SeITag, researchers can analyze all genes and tissues or their marked groups, choose tissue-specific genes basing on integrated criteria, visualize expression data, identify the genes with a correlated expression in a given set of tissues, isolate disease-specific genes with particular characteristics, such as receptors or secreted proteins.

## 2. General description

### 2.1. Main window


The main window (Fig. 2.1.1) presents gene expression data in tabular form. The upper frame of the main window contains the main menu and toolbar buttons.


The table below displays expression data for a set of genes. Each row in this table contains a set of data for one gene. The sets of data are identical for all genes (rows). Table columns show the values of gene expression measured under various conditions. Some columns list other information related to genes, for example, gene names.



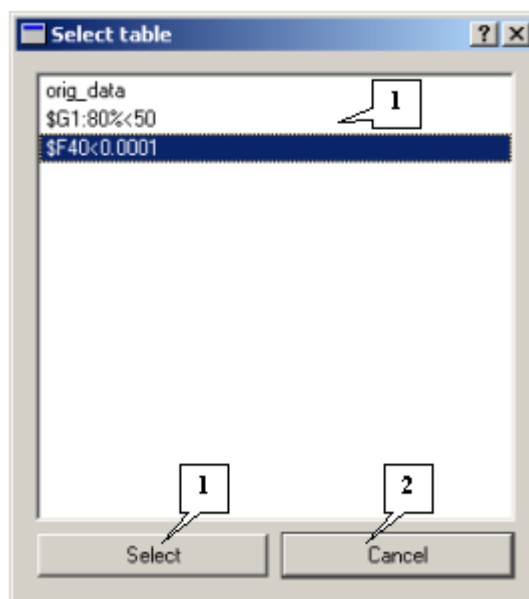
### 2.2. Main menu commands and toolbar buttons

1. The "File" menu contains the following commands for file operations:

- "Open expression data" (Ctrl+O),  button – opens the "Open" dialog box, which allows you to load a file with a gene expression table.
- "Close expression data" (Ctrl+W) – unloads a file with a gene expression table.

- “Save as”,  button – opens the “Save as” dialog box so that you can save the current data under a new name or in a new location.
  - “Link sequence” – downloads a file with the nucleotide sequences of genes. Calls the “Load data” dialog box from where you can load a file. After the file is loaded, the “Show sequence” command on the shortcut menu becomes active.
  - “Link gene data” – links a file with gene descriptions. Calls the “Load data” dialog box from where you can load a file with gene descriptions. After the file is loaded, the “URLs>UniGene” command becomes active on the shortcut menu.
  - “Unlink sequence” – unlinks the selected file with nucleotide sequences from main data.
  - “Unlink gene data” – unlinks the file with gene descriptions from main data.
  - “Quit” – exits the program.
2. The “Group” menu contains commands for managing groups of experiments:
- “View group” – calls the “View group data” dialog box, where you can view groups.
  - “Add group” – calls the “Edit group data” dialog box from where you can create groups.
  - “Edit group” – calls the “Select field group” dialog box. In this dialog box, you can select one group to be edited in the appropriate dialog box.
  - “Delete group” – calls the “Delete field group” dialog box so that you can delete one (or several) groups.
3. The “Table” menu contains commands for working with tables:
- “Select genes by query” – calls the “Make selection” dialog box. In this dialog box, you can select genes that satisfy specified conditions and create a new table for these genes.
  - “Selections list” – calls the “Select table” dialog box, which allows you to select a table. This dialog box displays the main gene expression table that lists all the genes loaded when you started the current project and tables generated as a result of selections (Fig. 2.2.1).
  - “Quick search gene in current selection” – calls the “Quick search” dialog box from where you can search for specific genes in a table by their field value.
  - “Remove all generated selections” – removes all tables generated as a result of selections.

## “Select table” dialog box

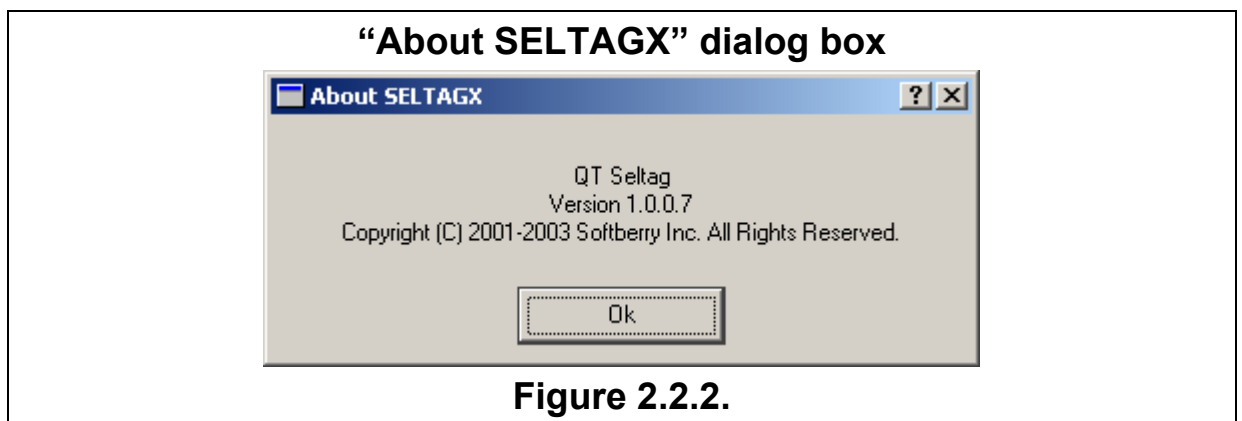


**Figure 2.2.1.**

1. List of tables. 2. The “Select” button. 3. The “Cancel” button.

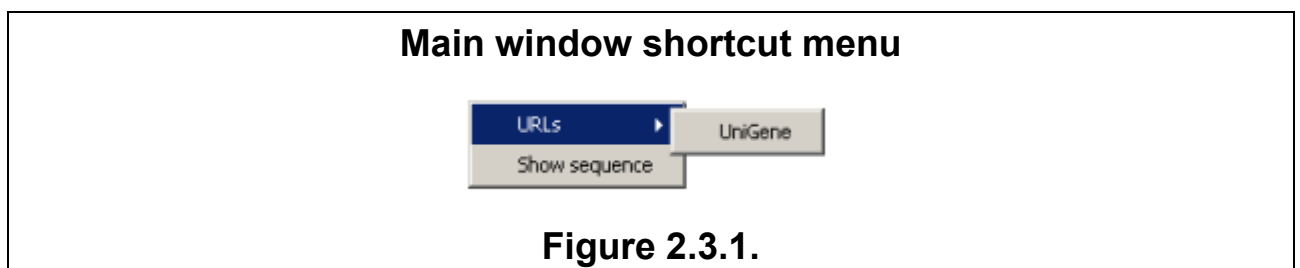
4. The “Analysis” menu contains a set of commands for data analysis:
  - “Correlations”:
    - “Select most correlated genes” – calls the “Select most correlated genes for specified gene set” dialog box designed to search for genes whose expression profiles are best correlated with those of one or more genes.
    - “Get correlations between genes” – calls the “Correlation analysis setup” dialog box, which allows you to calculate the correlation matrix between two sets of genes.
    - “Get correlations between fields” – calls the “Correlation analysis setup” dialog box, which allows you to calculate the correlation matrix between two sets of fields.
  - “Clustering”:
    - “Build gene tree” – calls the “Tree calculation setup” dialog box from where you can build a gene tree.
    - “Find genes cluster” – calls the “Setup for clustering procedure” dialog box, which allows you to set the parameters for gene clustering.
    - “Find gene cluster (Ben-Dor algorithm)” – calls the “Setup clustering procedure (Ben-Dor) for genes” dialog box from where you can set the parameters for gene clustering by the Ben-Dor algorithm.
    - “Find gene cluster (SOM algorithm)” – calls the “Setup for self-generated clustering procedure” dialog box from where you can specify the parameters for gene clustering by the self-organizing maps (SOM) algorithm.
    - “Build field tree” – calls the “Tree calculation setup” dialog box so that you can build a field tree.
    - “Find field cluster (Ben-Dor algorithm)” – calls the “Setup clustering procedure (Ben-Dor) for fields” dialog box, which allows you to set the parameters for field clustering by the Ben-Dor algorithm.

- “Find field cluster (SOM algorithm)” – calls the “Setup for self-generated clustering procedure for fields” dialog box from where you can specify the parameters for field clustering by the self-organizing maps (SOM) algorithm.
  - “Principal component” – calls the “Setup for principal component analysis” dialog box from where you can analyze the correlation and covariance matrices of gene expression profiles using the principal component method and visualize the results obtained.
  - “Expression map” – opens the expression matrix map.
  - “Profiles map” – opens the “Profile dialog” dialog box, which allows you to visualize the profiles of genes from the current table.
5. The “Options” menu is unavailable in this version.
6. The “Help” menu provides access to help topics.
- “About” – calls the “About SELTAGX” dialog box with information about the program (Fig. 2.2.2).



### 2.3. Main window shortcut menu

To open the shortcut menu of the main window (Fig. 2.3.1), click your right mouse button.



#### Shortcut menu commands:

- URLs>UniGene– loads a UniGene database entry corresponding to the selected gene into a new Web browser window (Fig. 2.3.2). This command is active if the file with a gene description has been loaded (see section 3.1.2).

## UniGene database entry

The screenshot shows a web browser window with the title 'seltagx'. The main content area displays the UniGene entry for ADH1B. At the top, there is a navigation bar with links for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and Books. Below this is a search bar and a set of tabs: Limits, Preview/Index, History, Clipboard, and Details. The entry itself includes a 'Summary' section with the text: 'ADH1B: alcohol dehydrogenase 1B (class 1), beta polypeptide Homo sapiens, 319 sequence(s)'. To the left of the entry, there are several links: 'NCBI', 'UniGene', 'Query Tips', 'FAQ', 'DDD', 'Download UniGene', 'Related Resources', 'LocusLink', 'HomoloGene', 'dbEST', 'Trace Archive', 'BLAST', and 'CGAP'. At the bottom of the page, there are links for 'Restrictions on Use' and 'Write to the Help Desk', along with 'NCBI | NLM | NIH'.

Figure 2.3.2.

- Show sequence – displays a window with the nucleotide sequence of the selected gene (Fig. 2.3.3). This command is active if the file with the nucleotide sequences of genes has been loaded (see section 3.1.3).

## Nucleotide sequence

The screenshot shows a window titled 'gi|219427|dbj|D00137.1|HUMADH21C Homo sapiens mRNA for alcohol dehydrog...'. The window contains a large block of nucleotide sequence text, which is a single-line representation of the DNA sequence for the ADH1B gene. The sequence is displayed in a monospaced font and is wrapped across multiple lines. The sequence starts with 'GCAGAGAAGACAGAAACGACATGAGCACAGCAGGAAAAGTAATCAAATGCAAAGCAGCTGTGCTATGGGA' and ends with 'CCAAATCCCTGGGGTAAAAGCTAGGGTAAGGTAAGGATAGACTCACATTTACAAGTAGTGAAGGTCCAA'.

Figure 2.3.3.



## 2.4. Format of the file with a gene expression table

Gene expression data are represented as a table consisting of rows (genes) and columns (fields). Each row in the table contains a set of data for one gene. The fields are the same for all genes (rows). As a rule, columns list gene expression parameters measured under various conditions (in tissues, organs, cell cultures, etc.). In addition, some columns may show additional information (both numerical and textual) related to the genes (e.g. gene names).

In general view, there are four main types of the column (field) contents:

<b>IVALUE</b>	- value – integer
<b>FVALUE</b>	- value – floating-point number
<b>WORD</b>	- value – word (a piece of text without spaces)
<b>STRING</b>	- value – string (a piece of text separated with spaces)

Fields are completely defined by their type and name.

### 2.4.1. Basic format of the SelTag input file

The basic format of the SelTag input file is the following:

; May contain comment in any line of the file

NAME<tab>WORD

GENEID<tab>IVALUE

TISSUECANCER0<tab>FVALUE

TISSUECANCER1<tab>FVALUE

TISSUENORMAL0<tab>FVALUE

TISSUENORMAL1<tab>FVALUE

TISSUENORMAL2<tab>FVALUE

DATA

GENE04675<tab>402<tab>6.00<tab>5.60<tab>5.97<tab>6.00<tab>6.00

GENE46890<tab>794<tab>2.77<tab>3.22<tab>5.65<tab>5.68<tab>5.68

GENE23794<tab>404<tab>5.97<tab>5.97<tab>6.00<tab>5.60<tab>5.97

In this example, <tab> means the tab symbol (the 'Tab' key on the keyboard).

The first lines, above the "DATA" line, contain the description of the data format. In this part of the file, each line contains a description of one field: field name and type.

The lines below the "DATA" line show the gene expression data for each gene. Each line corresponds to a single gene. The data in the fields are separated with the tab. Double tabs mean empty fields.

### 2.4.2. Groups of fields

Fields of the same type can be combined into a **group of fields**. A group can be defined by the group name and the list of group fields. This combination generally represents the functional significance of these fields. For example, the fields that represent gene expression levels in tumor tissues can belong to the "Cancer tissues" group. The same field may join several groups of fields, i.e. some fields in various groups may overlap.

The format description includes a description of the group of fields. The format description begins with the "#GROUP" line, which defines the name of the group. The following lines contain a list of fields included into this group. A group description group should end with the "#ENDGROUP" line.

An example of the data format with defined fields:

; May contain comment in any line of the file

NAME<tab>WORD

GENEID<tab>IVALUE

TISSUECANCER0<tab>FVALUE

TISSUECANCER1<tab>FVALUE

TISSUENORMAL0<tab>FVALUE

TISSUENORMAL1<tab>FVALUE

TISSUENORMAL2<tab>FVALUE

#GROUP<tab>Cancer tissues

TISSUECANCER0

TISSUECANCER1

#ENDGROUP

#GROUP<tab>Arbitrary group

TISSUECANCER1

TISSUECANCER2

TISSUENORMAL0

TISSUENORMAL1

#ENDGROUP

DATA

GENE04675<tab>402<tab>6.00<tab>5.60<tab>5.97<tab>6.00<tab>6.00

GENE46890<tab>794<tab>2.77<tab>3.22<tab>5.65<tab>5.68<tab>5.68

GENE23794<tab>404<tab>5.97<tab>5.97<tab>6.00<tab>5.60<tab>5.97

This data format defines two groups: "Cancer tissues" (includes the TISSUECANCER0 and TISSUECANCER1 fields) and "Arbitrary group" (includes the TISSUECANCER1, TISSUECANCER2, TISSUENORMAL0, and TISSUENORMAL1 fields).

Data can be loaded to SelTag from files containing tables in simpler formats or directly retrieved from tables provided that they match exactly the format of your table (both have the same number of columns with identical names and the columns in the data file are in the same order as the columns in your table).

## 2.5. Format of the file with a gene description

A file of the "Gene Description" type consists of several successive descriptions. Each of these descriptions occupies one or more lines.

The first line is written as:

">" Id "|" Short\_description

where ID is a special identifier for linking to data and Short\_description is a brief description one line long.

Below the first line there can be several lines with URL\_description followed by several lines with Long\_description.

The URL\_description line is written as

"URL:" Menu\_text "|" URL,

Menu\_text is a textual description for URL, and

URL is a normal URL, which is the <http://.../...> line.

Long\_description is a description several lines long.


### 3. Working with the application

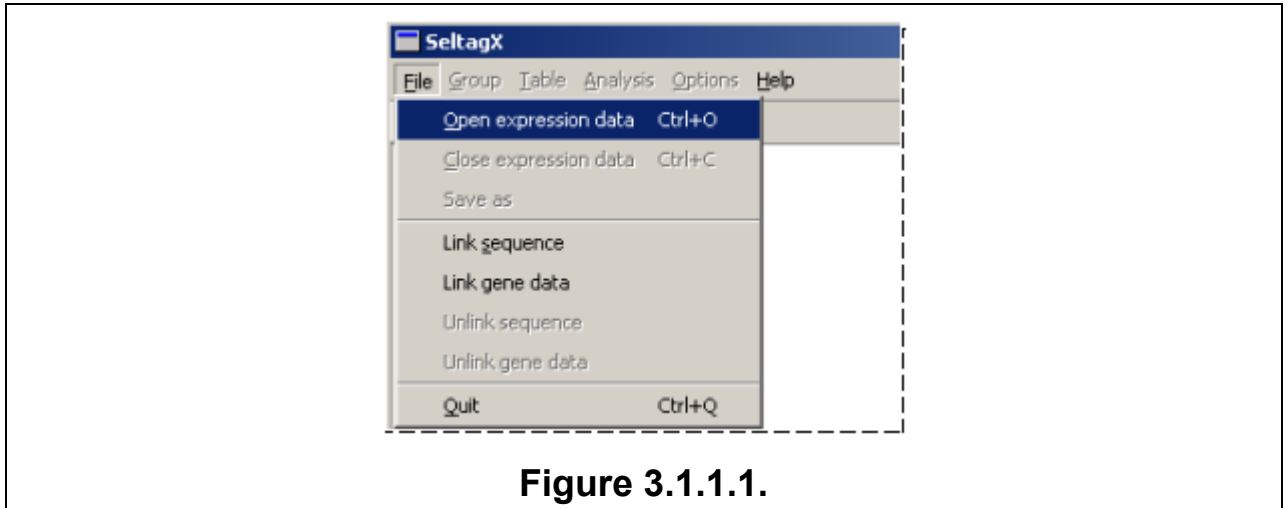
#### 3.1. Loading data

After application startup, the main window opens. To start working with the application, you need to load a file with a gene expression table.

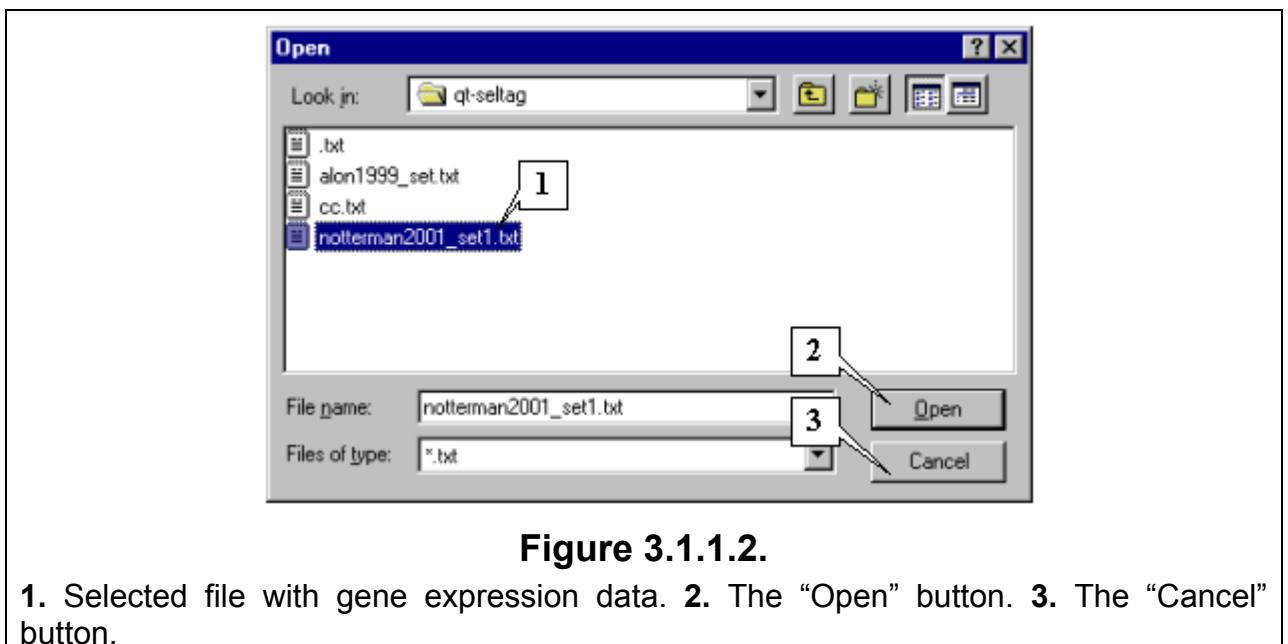
##### 3.1.1. File with a gene expression table

###### Open a file

To load a file with gene expression data, click File>Open expression data, press Ctrl+O, or click the  button (Fig. 3.1.1.1).

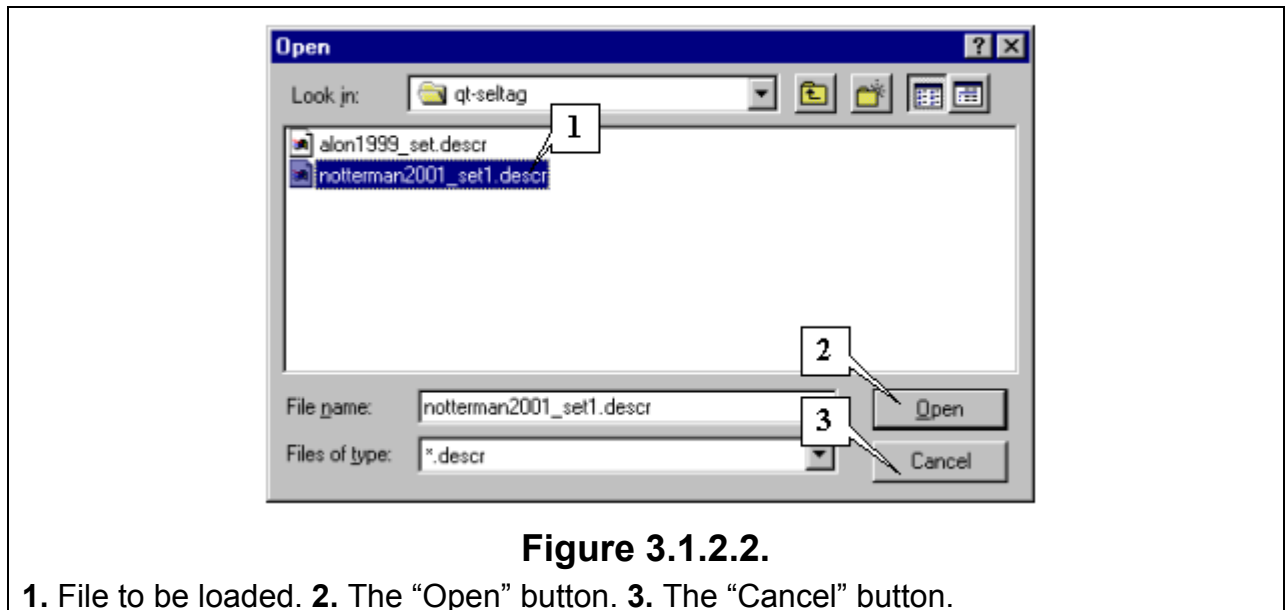


This will open the “Open” dialog box that displays the names of files with data tables (Fig. 3.1.1.2). In this dialog box, select a file and click “Open”. After this, you will see the “Loading data” dialog box, which will be closed after the file is loaded (Fig. 3.1.1.3).





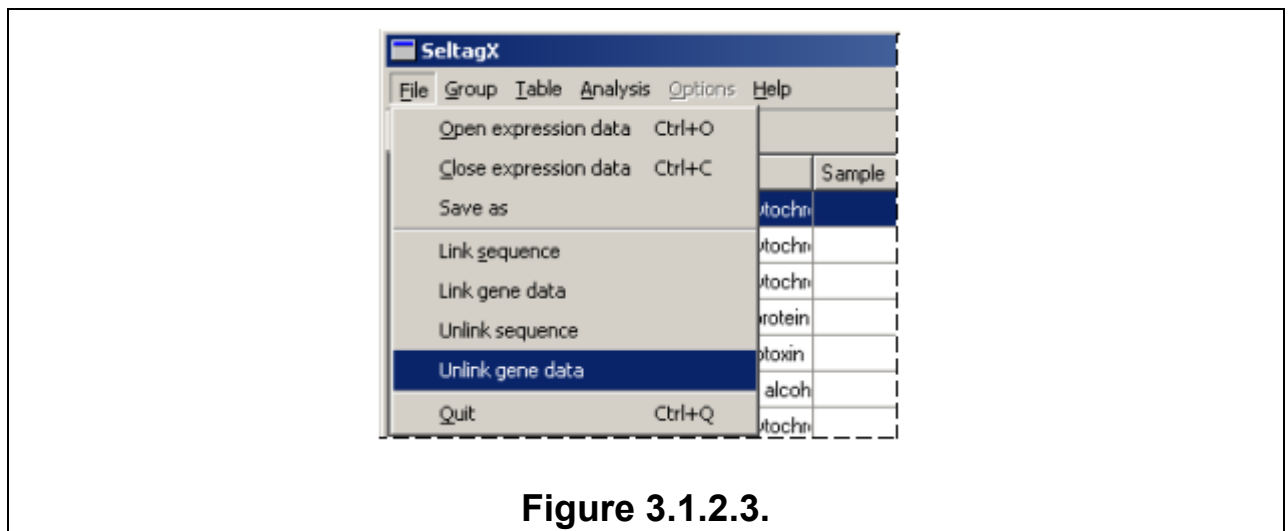
The “Open” dialog box shows the names of files with gene descriptions (Fig. 3.1.2.2). In this dialog box, select a file and click “Open”.



The “URLs>UniGene” command on the shortcut menu of the main table (to open the shortcut menu, click your right mouse button) will become available (Fig. 2.1.1).

#### **Close a file**

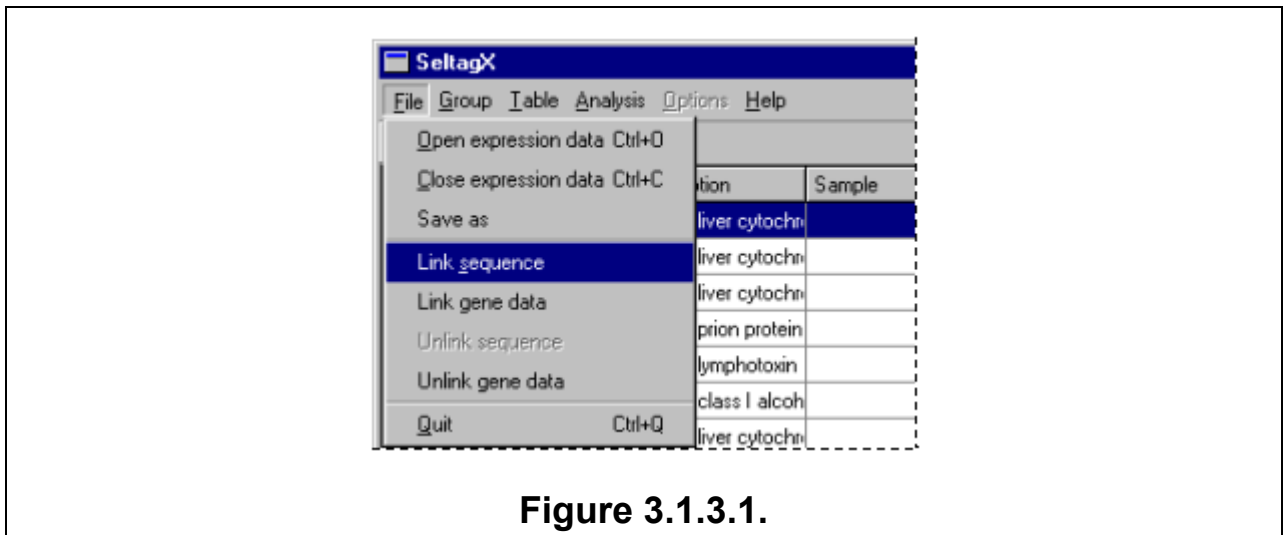
To unlink a file with a gene description from the main data, click the File>Unlink gene data command on the main menu (Fig. 3.1.2.3).



### **3.1.3. File with the nucleotide sequences of genes**

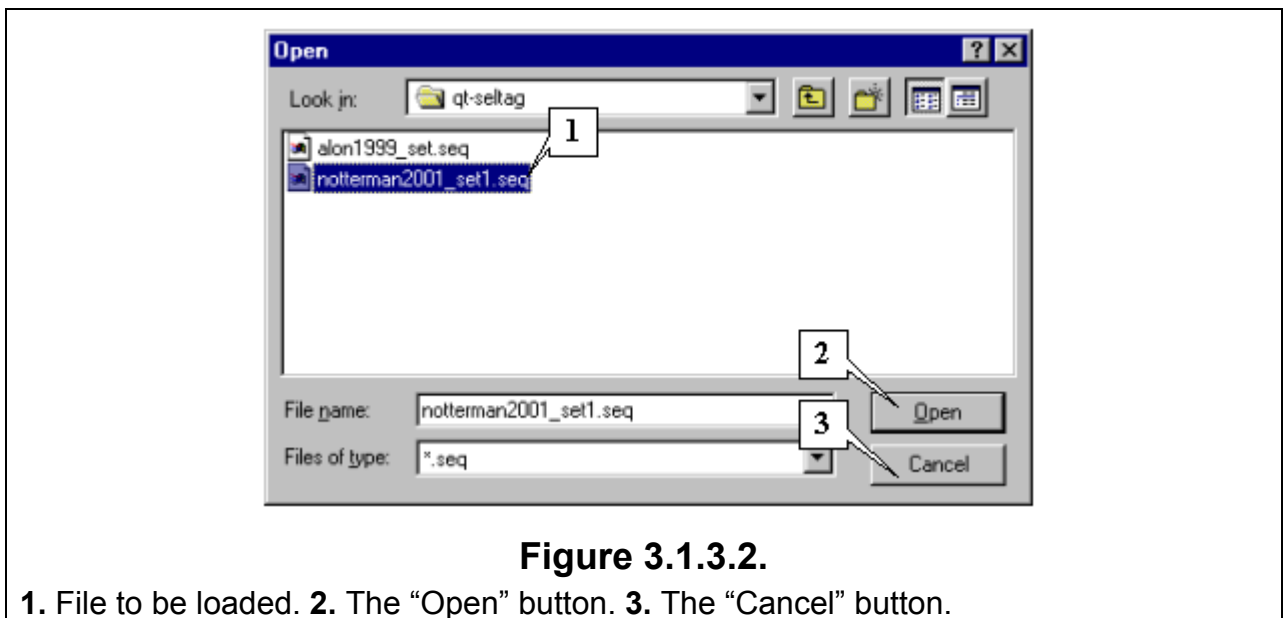
#### **Open a file**

To load a file with the nucleotide sequences of genes, click the File>Link sequence command on the main menu (Fig. 3.1.3.1).



**Figure 3.1.3.1.**

This will open the “Open” dialog box listing the names of files with nucleotide sequences (Fig. 3.1.3.2). In this dialog box, select a file and click “Open”.



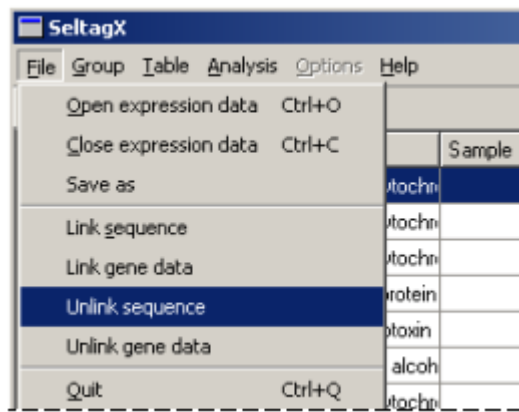
**Figure 3.1.3.2.**

1. File to be loaded. 2. The “Open” button. 3. The “Cancel” button.

The “Show sequence” command on the shortcut menu of the main table will become available (Fig. 2.1.1) (to open the shortcut menu, click your right mouse button). This command evokes a dialog box with the nucleotide sequence of the selected gene (Fig. 2.3.3).

#### Close a file

To unlink a file with nucleotide sequences from the main data, click the File>Unlink sequence command on the main menu (Fig. 3.1.3.4).



**Figure 3.1.3.4.**

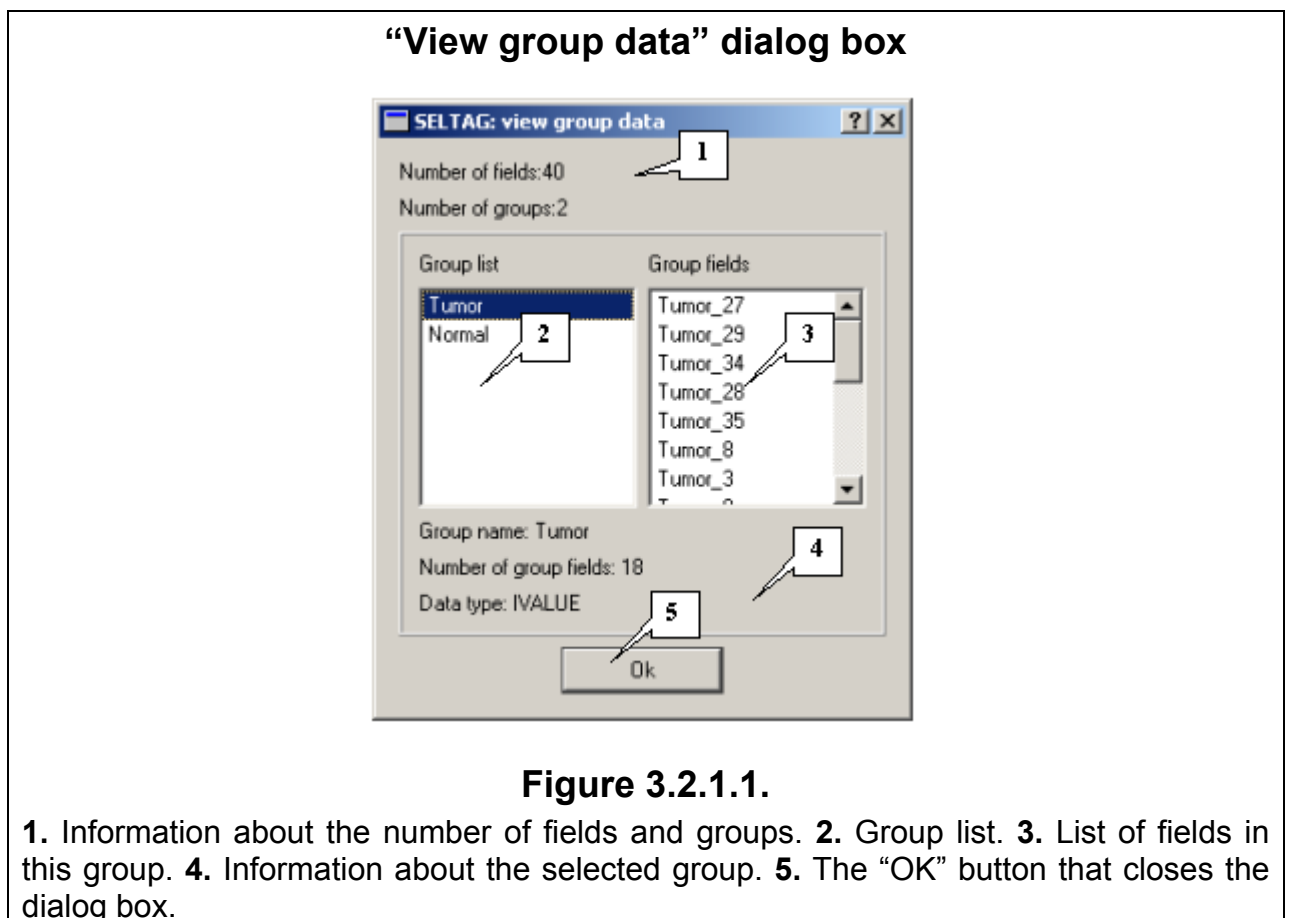
## 3.2. Managing groups of fields (experiments)

To work with groups of fields, use the commands in the “Group” menu.

### 3.2.1. Viewing groups

To view groups, click the Group>View group command. This will open the “View group data” dialog box (Fig. 3.2.1.1).

The “Group list” column contains groups and the “Group Fields” column shows fields available for the selected group. The information about the number of groups and the total number of unique fields in the project is displayed at the top of the dialog box. Below the columns, you can see information about the selected group: group name, the number of fields in this group, and the data type. Click “OK” to close the dialog box.



### 3.2.2. Adding a new group

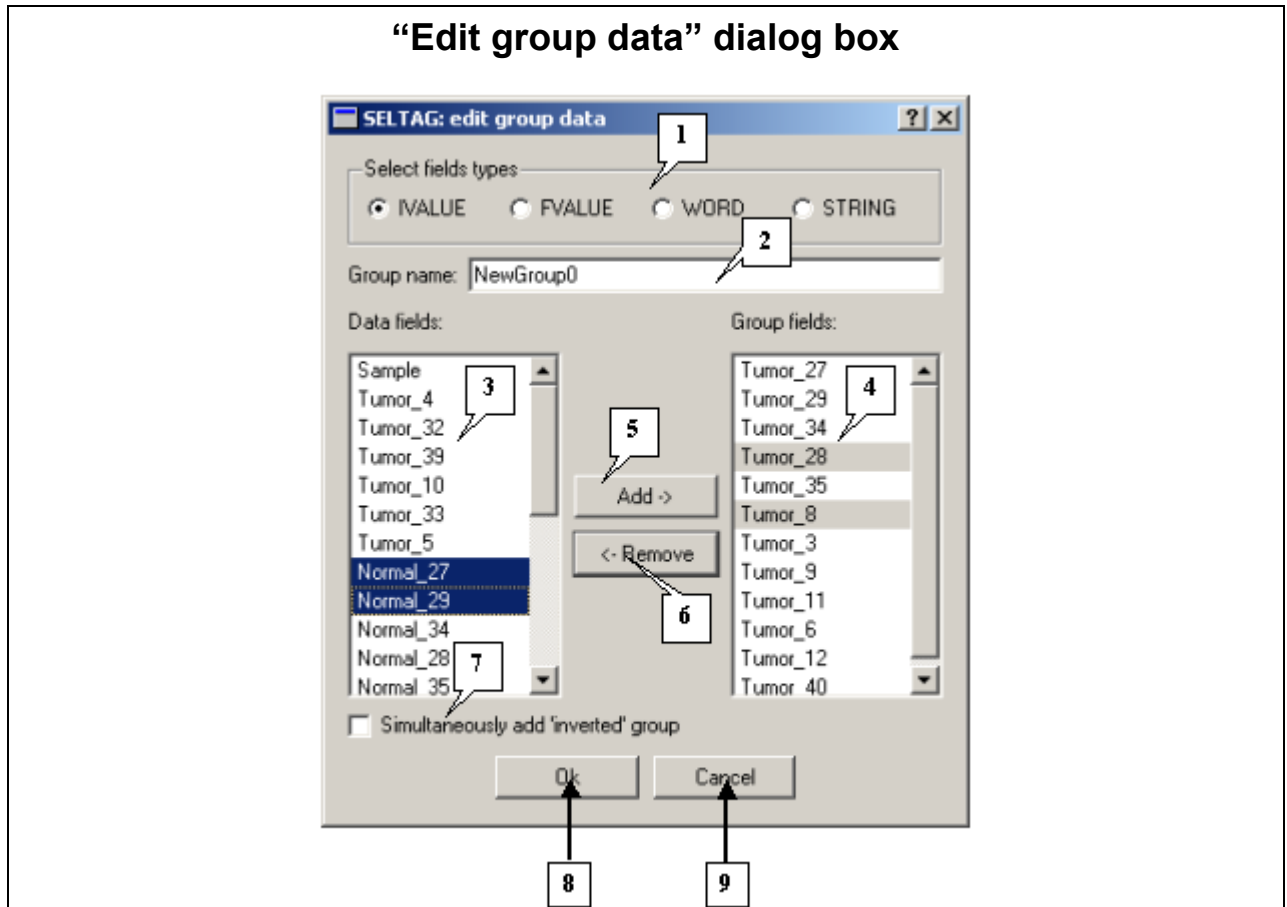
To add new groups to the project, click the Group>Add group command. This will open the “Edit group data” dialog box (Fig. 3.2.2.1).

To create a new group, do the following:

1. Select field types by clicking the radio buttons in the “Select field types” pane. The fields of the selected type will appear in the “Data fields” list. Click the “Add” button to add fields to a group or the “Remove” button to remove fields from a group.
2. Enter the name of the group (for example, NewGroup0) into the “Group Name” field. If the “Simultaneously add ‘inverted’ group” box is checked, the two groups, NewGroup and INV\_NewGroup, will be added to the list (the INV\_NewGroup group includes fields missing in the NewGroup group).



3. Click “OK”. To cancel the operation, click “Cancel”. If a group with the same name already exists, you will see the “Error” dialog box (Fig. 3.2.2.2) where you will be asked to rename the group you have created.



**Figure 3.2.2.1.**

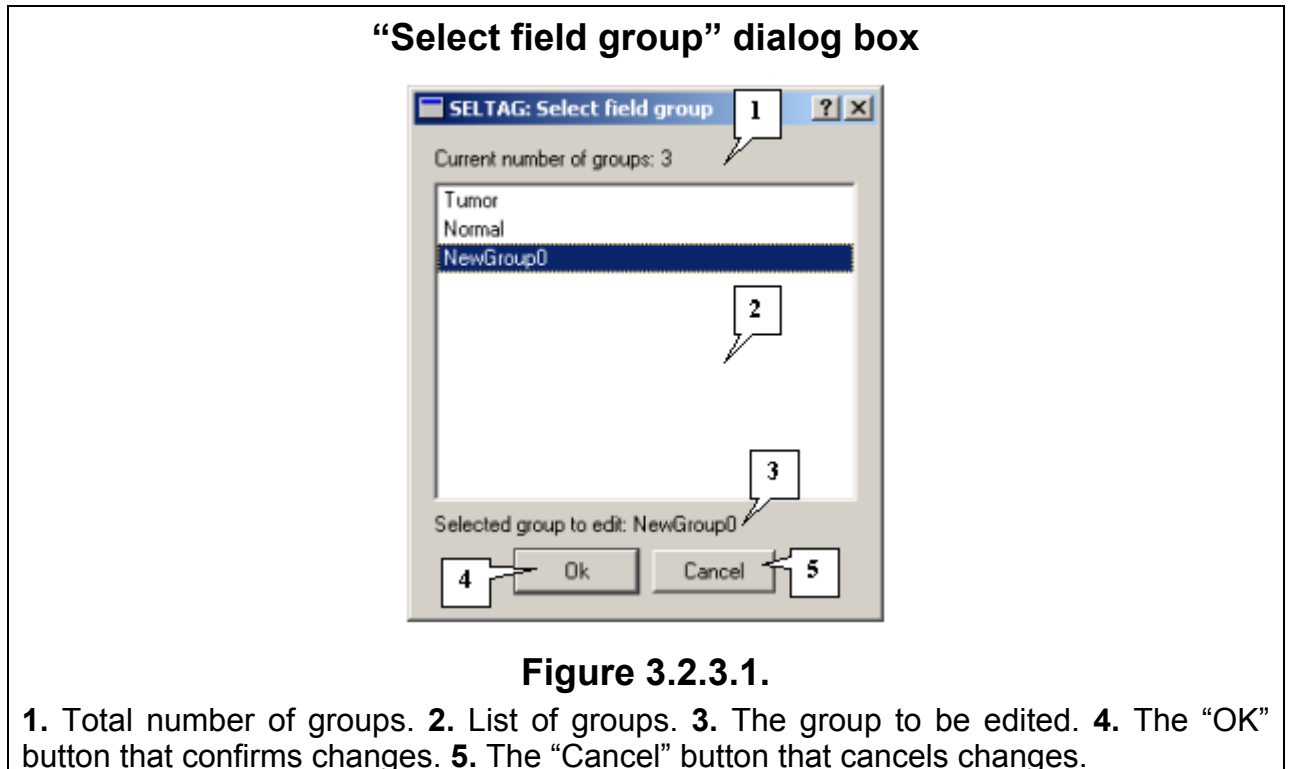
1. The “Select fields type” pane. 2. The “Group name” field. 3. Fields not included into the group. 4. Fields included into the group. 5. The “Add” button for adding fields to the group. 6. The “Remove” button for removing fields from the group. 7. The checkbox that enables creation of an “inverted” group. 8. The “OK” button that applies changes and closes the dialog box. 9. The “Cancel” button that cancels changes and closes the dialog box.



**Figure 3.2.2.2.**

### 3.2.3. Editing a group

To edit the selected group, click the Group>Edit group command. You will see the “Select field group” dialog box (Fig. 3.2.3.1).

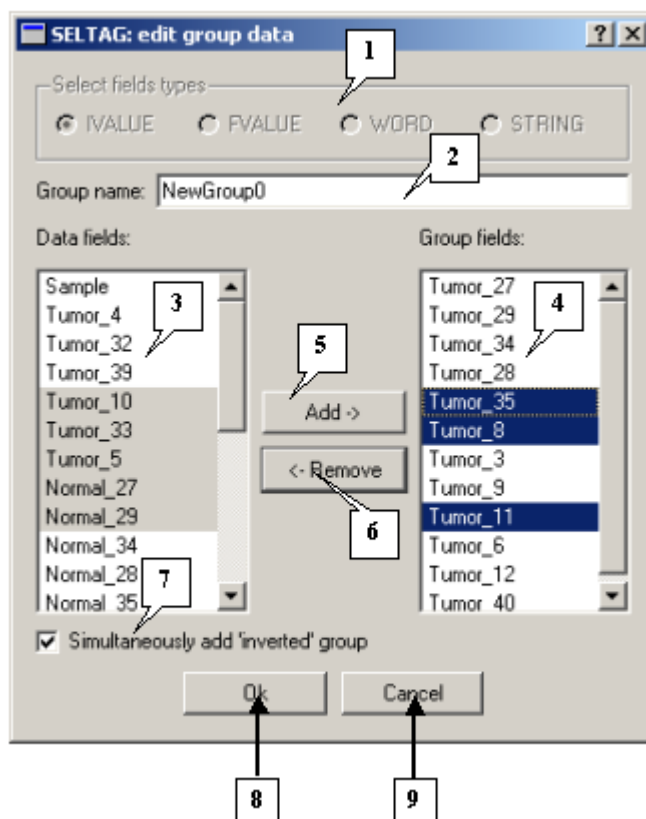


After clicking “OK”, you will see the “Edit group data” dialog box (Fig. 3.2.3.2) from where you can edit the following group parameters:

- a. The list of fields included into a group (displayed in the “Group fields” list). Click “Add” to add fields to a group or “Remove” to remove fields from a group.
- b. Group name. You can edit the name of the group in the “Group Name” field.
- c. Add “inverted” group that will contain all fields from the “Data fields” list. The name of the “inverted” group will have the “INV\_” prefix. To add an inverted group, check the “Simultaneously add ‘inverted’ group” checkbox. If an “inverted” group with this name already exists, the fields in this group will be updated according to the last changes you made.

Click “OK” to apply changes and close the dialog box and “Cancel” to cancel changes and close the dialog box.

## “Edit group data” dialog box



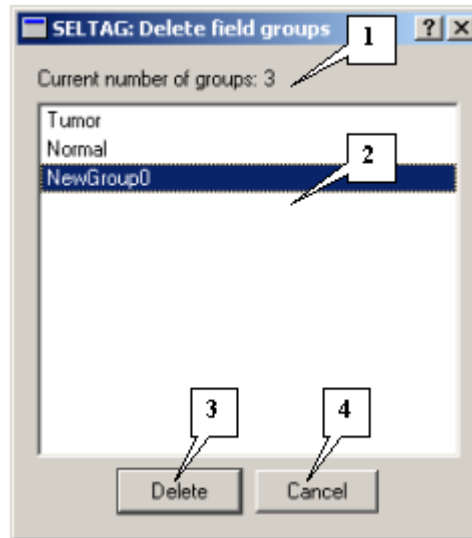
**Figure 3.2.3.2.**

1. Type of field data. 2. The field for editing the group name. 3. Fields not included into the group. 4. Fields included into the group. 5. The “Add” button that adds fields to a group. 6. The “Remove” button for removing fields from a group. 7. The checkbox that enables creation of an “inverted” group. 8. The “OK” button that applies changes and closes the dialog box. 9. The “Cancel” button that cancels changes and closes the dialog box.

### 3.2.4. Deleting a group

Click the Group>Delete group command to open the “Delete field group” dialog box used to delete a group from the current project (see Fig. 3.2.4.1). To delete groups, selected these groups in the list and click “Delete”. To cancel the operation, click “Cancel”. The total number of groups is displayed above the group list. Figure 3.2.4.1 shows that the NewGroup0 group is selected for deleting.

### “Delete field group” dialog box



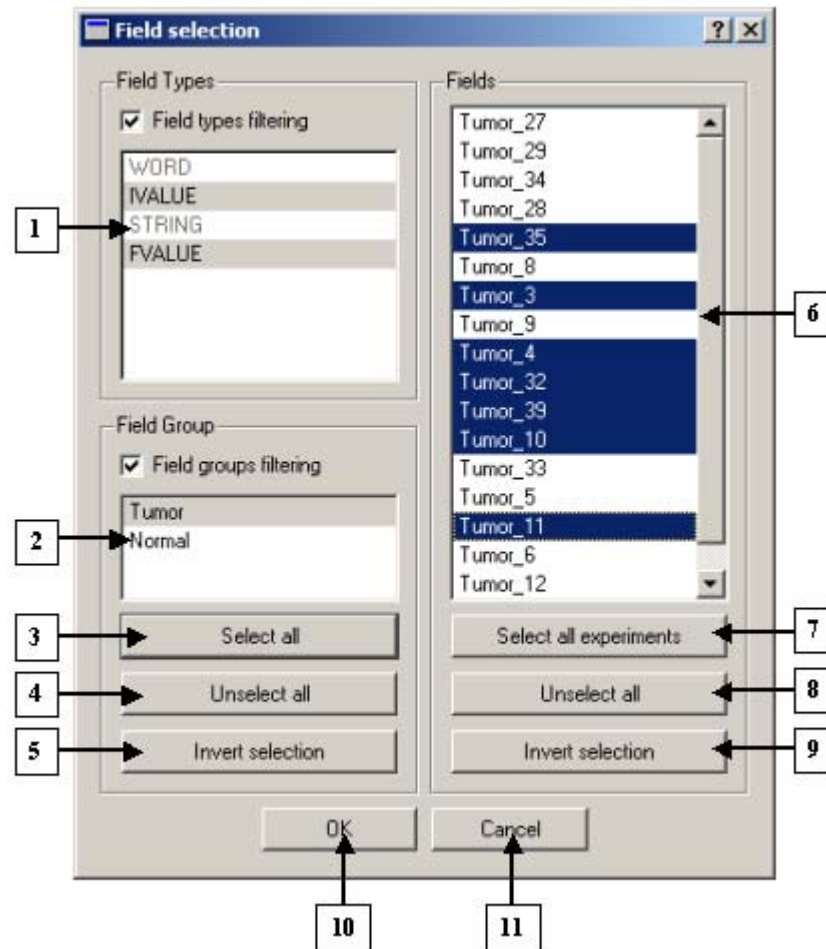
**Figure 3.2.4.1**

1. Total number of groups. 2. List of groups. 3. The “Delete” button. 4. The “Cancel” button.

### 3.3. “Field selection” dialog box

To open the “Field selection” dialog box (Fig. 3.3.1), click the “Fields...” button in the “Select most correlated genes for specified gene set”, “correlation analysis setup”, “Tree calculation setup”, “setup for clustering procedure”, “setup for principal component analysis” dialog boxes or click the Analysis>Profiles map command (if you have no groups yet. This command is used to select fields from the main table.

## “Field selection” dialog box



**Figure 3.3.1.**

**1.** List of data types. **2.** List of field groups (experiments). **3.** The “Select all” button that selects all groups. **4.** The “Unselect all” button that unselects all groups. **5.** The “Invert selection” button that inverts selection of all groups. **6.** List of fields. **7.** The “Select all experiments” button that selects all fields. **8.** The “Unselect all” button that unselects all fields. **9.** The “Invert selection” button that inverts selection of all fields. **10.** The “OK” button that confirms the selection and closes the dialog box. **11.** The “Cancel” button that cancels the selection and closes the dialog box.

### 3.3.1. Dialog box elements

#### 3.3.1.1. List of data types

The list becomes active if the “Field types filtering” box is checked. The list contains the following data types: WORD, IVALUE, STRING, and FVALUE. The data types selected in the list are filters applied to the fields and groups in this project.

If the “Field types filtering” checkbox is unchecked, the list remains inactive, and the groups and fields of all data types will be shown in the lists of groups and fields.

#### 3.3.1.2. List of field groups

The list becomes active if the “Field groups filtering” box is checked. The list contains the names of field groups. The groups selected in the list are filters applied to the fields (experiments).

Click the following buttons to manage groups:

- “Select all” – select all groups.
- “Unselect all” – unselect all selected groups.
- “Invert selection” – invert selection (this button changes the current selection such that all previously selected groups will now be selected and vice versa).  
If the “Field types filtering” checkbox is checked, list 3.3.2 will include the types of groups selected in list 3.3.1. If the “Field types filtering” checkbox is unchecked, list 3.3.2 will show all groups.

### **3.3.1.3. List of fields**

The “Fields” pane displays a list of fields. The field list shows all fields that meet the restrictions defined in lists 3.3.1.1 and 3.3.1.2. If there are no restrictions, the list will show all fields. Click the following buttons to manage the list:

- “Select all experiments” – selects all fields.
- “Unselect all” – unselects all selected fields.
- “Invert selection” – inverts selection.

### **3.3.1.4. Control buttons**

- “Cancel” – cancels selections you made and quits the dialog box.
- “OK” – accepts the fields selected in list 3.3.1.3. In the previous dialog box, from where you called the “Field selection” dialog box, you can see the number of fields selected for further operations.

## **3.4. Selecting genes by query**

You can access the «Make selection» dialog box (Fig. 3.4.1) by clicking the Table>Select genes by query command on the main menu. Using this dialog box, you can select genes from the working table by search expression.

### **3.4.1. Dialog box elements**

**3.4.1.1. Search expression bar** is a text field located at the top of the dialog box. To clear the contents of this text field, click the “Clear Expr.” button.

**3.4.1.2. Status bar** (comments) is located below the search expression field. This field contains additional information.

**3.4.1.3. The “Field” pane – the field index input bar** contains a drop-down list of fields (Fig. 3.4.2.A). Click the “Select” button to view the list. The fields are presented in the following format:

Fn NAME, where n is the number of the field and NAME is its name.

If any field is selected, the status field displays the following information about the field:

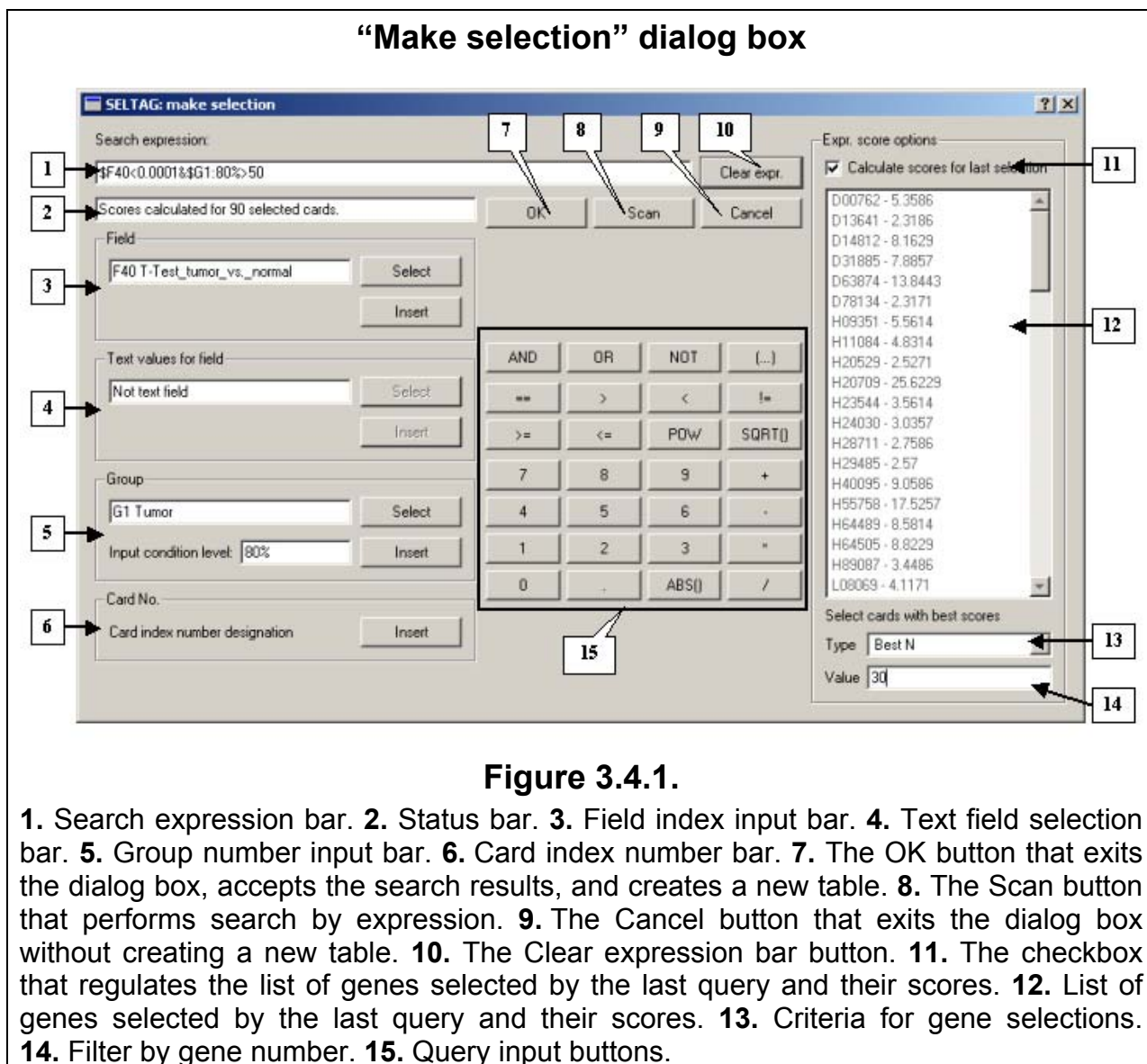
\$Fn: type=FIELDTYPE. Is item of group(s):x, where n is the field number, FIELDTYPE is the name of the data type for this field, and if the field you are searching for is included into other groups, then x lists the numbers of these groups.

The “Insert” button adds the current value of \$Fn to the expression (n is the field number).

**3.4.1.4. Text field selection bar** contains:

- A list of values of text fields (Fig. 3.4.2.B) evoked by the “Select” button. This list becomes active if the selected field is of the string type. If the field type is different from the string one (neither WORD nor STRING), this list remains empty.
- The “Insert” button adds the selected value from the list of current elements in the “TEXT” format to an expression.

## “Make selection” dialog box



**Figure 3.4.1.**

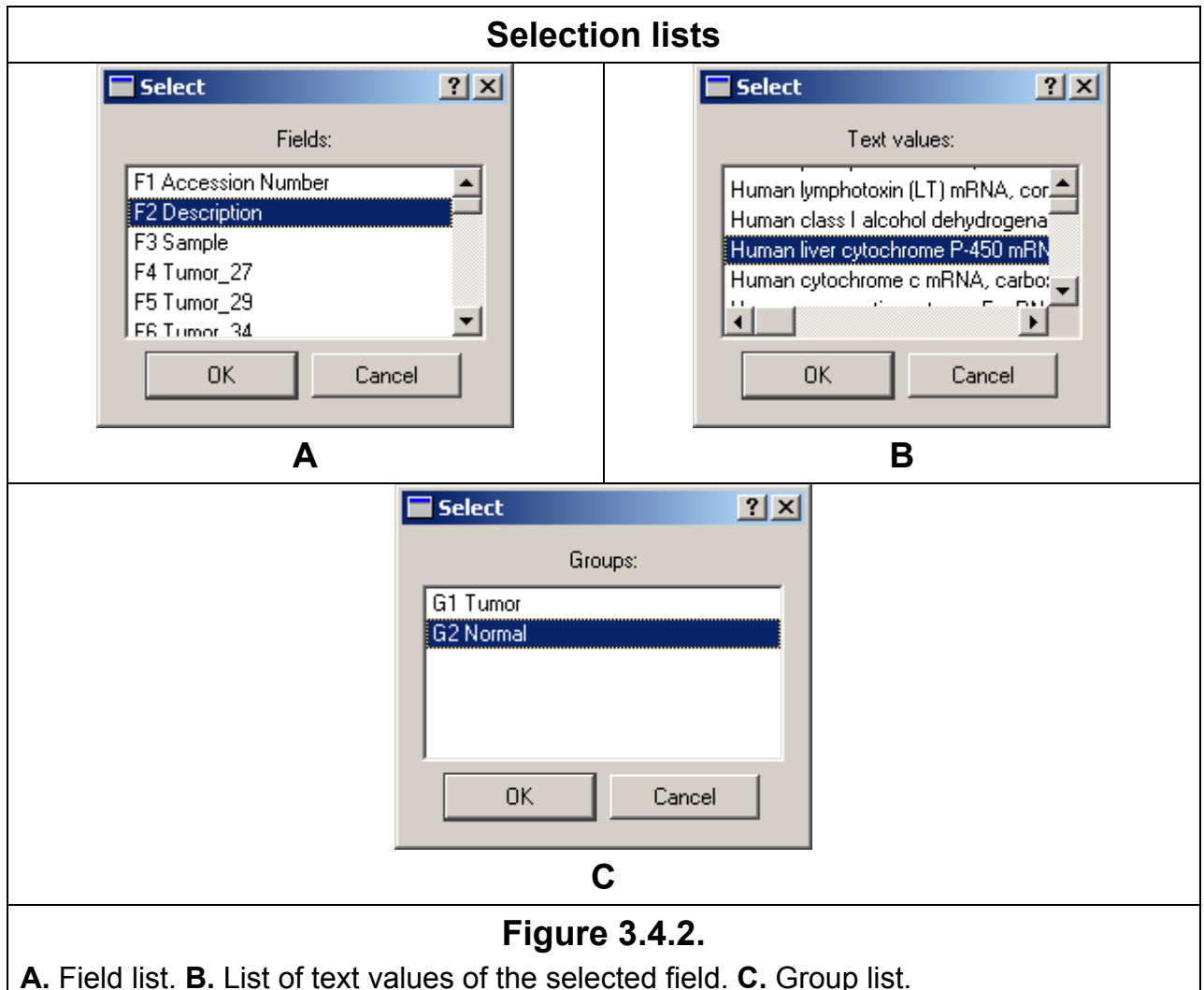
1. Search expression bar. 2. Status bar. 3. Field index input bar. 4. Text field selection bar. 5. Group number input bar. 6. Card index number bar. 7. The OK button that exits the dialog box, accepts the search results, and creates a new table. 8. The Scan button that performs search by expression. 9. The Cancel button that exits the dialog box without creating a new table. 10. The Clear expression bar button. 11. The checkbox that regulates the list of genes selected by the last query and their scores. 12. List of genes selected by the last query and their scores. 13. Criteria for gene selections. 14. Filter by gene number. 15. Query input buttons.

**3.4.1.5. The “Group” pane – Group number input bar** contains the following elements:

- 3.4.1.5.1. The “Group” list shows groups of experiments (Fig. 3.4.2.C). Click the “Select” button to access the list. The groups are in the following format:  
Gn NAME, where n is the group number and NAME is the group name. When the field is selected, the status bar (3.4.1.2.) shows the following information:  
\$Gn: type=FIELDTYPE, fields[x]: y1,y2,y3.....  
where n is the group number, FIELDTYPE is the name of the data type for this group, x is the number of fields included into the group, and y1,y2... are the numbers of fields included into the group.
- 3.4.1.5.2. The “Input condition level” field imposes an additional condition on selections within a group. This condition defines the lower threshold for the number of experiments in the group that satisfy the main condition. The threshold can be written as the number of experiments (in this case, enter an integer number into this field) or their percentage in the group (enter the percentage X in the format X%).
- 3.4.1.5.3. The “Insert” button adds the current value of \$Gn to an expression (n is the field number).

If additional conditions are defined, they are added to the expression:

\$Gn:50%



**3.4.1.6. The “Card No.” pane – the card index number bar** allows you to enter the number of a gene in the table. The “Insert” adds the \$N value to the expression.

**3.4.1.7. Query input buttons** are used to add various operations, figures, and mathematical functions to the expression.

**3.4.1.8. The “Expr. Score options”** contains:

- The “Calculate scores for last selection” checkbox. When this box is checked, a list of genes selected by the last query and their scores will appear.
- The list of genes and their scores.
- The “Type” drop-down list and the “Value” field that impose additional conditions on selecting genes by their scores. The “Type” list has two options:
  - “Not applied” – all genes satisfying the main search condition will be selected.
  - “Best N” – only genes with the highest scores will be selected. The number of genes is defined in the “Value” field.

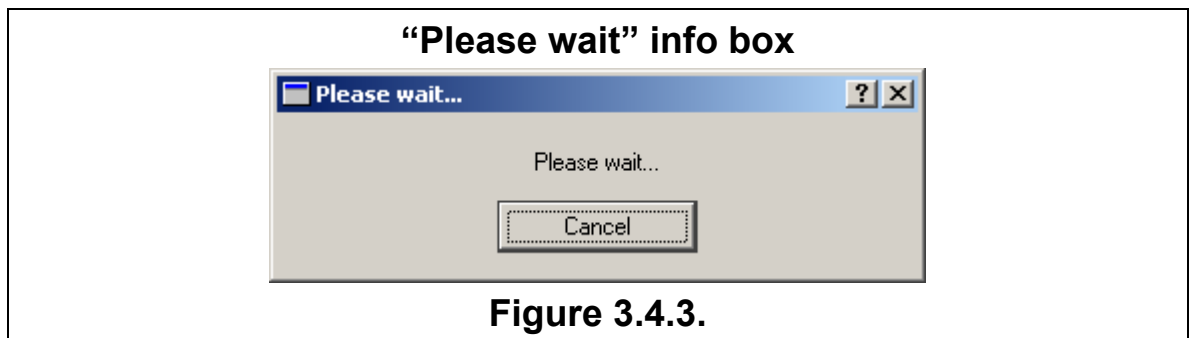
**3.4.1.9. Control buttons:**

- “Cancel” – cancels creating a new table and exits the dialog box.
- “Scan” – performs a search by expression. After the search is complete, the number of selected genes will be displayed in the status bar.



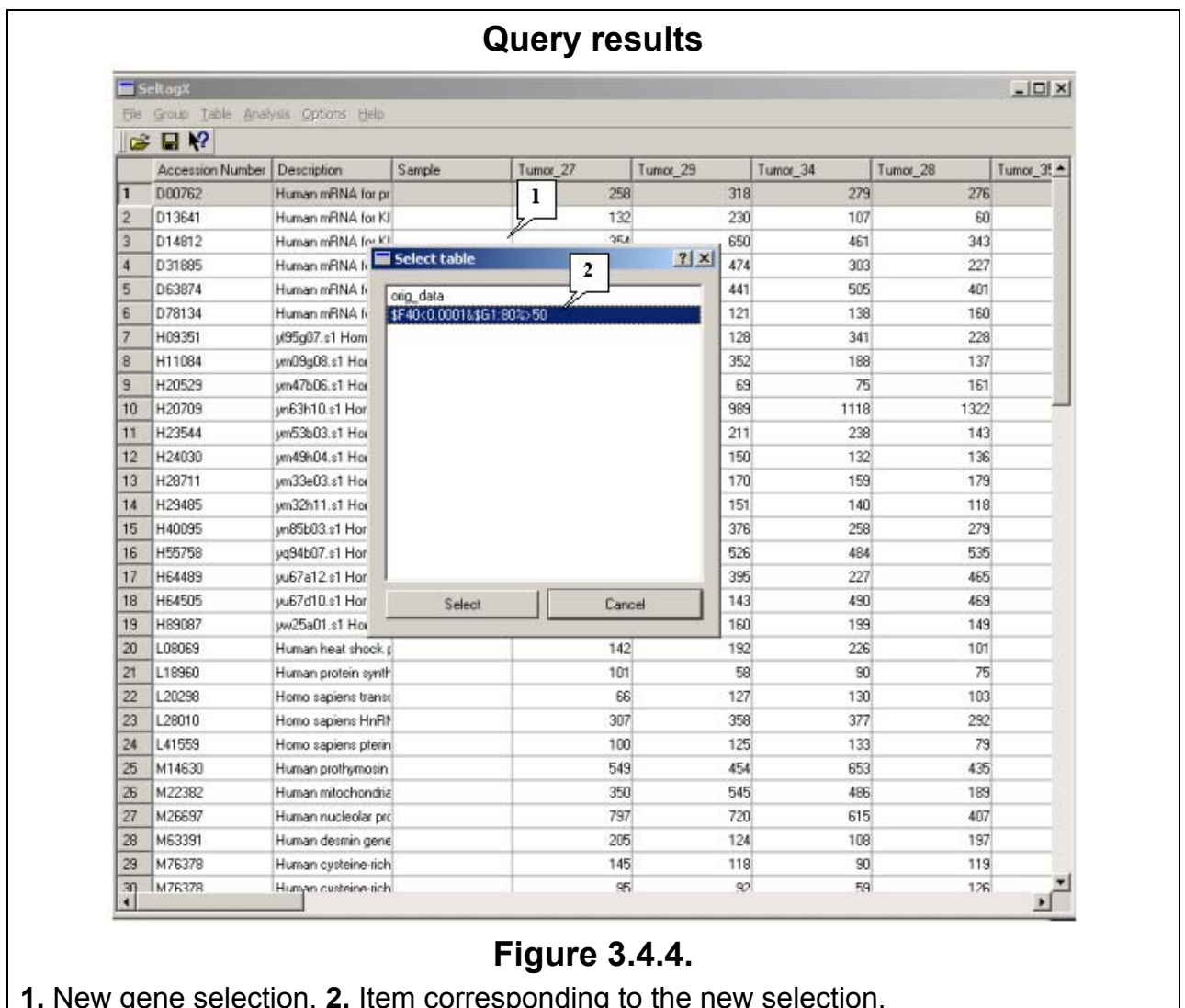
- “OK” – accepts search results, creates a new table of genes, and exits the dialog box.

**Note.** When the search is in progress, you will see the “Please wait” information box (Fig. 3.4.3) that will disappear after the search is complete.



**Figure 3.4.3.**

When the search is complete, an additional item appears in the “Select Table” dialog box (Fig. 3.4.4). The name of this item corresponds to the selected set of genes. You can save all gene selections obtained when working with the program and switch between them. To delete the list of current tables, click the Table>Remove all generated selections command on the main menu. Figure 3.4.4 shows a table of genes selected under the search conditions illustrated in Fig. 3.4.1.



### 3.5. Quick search

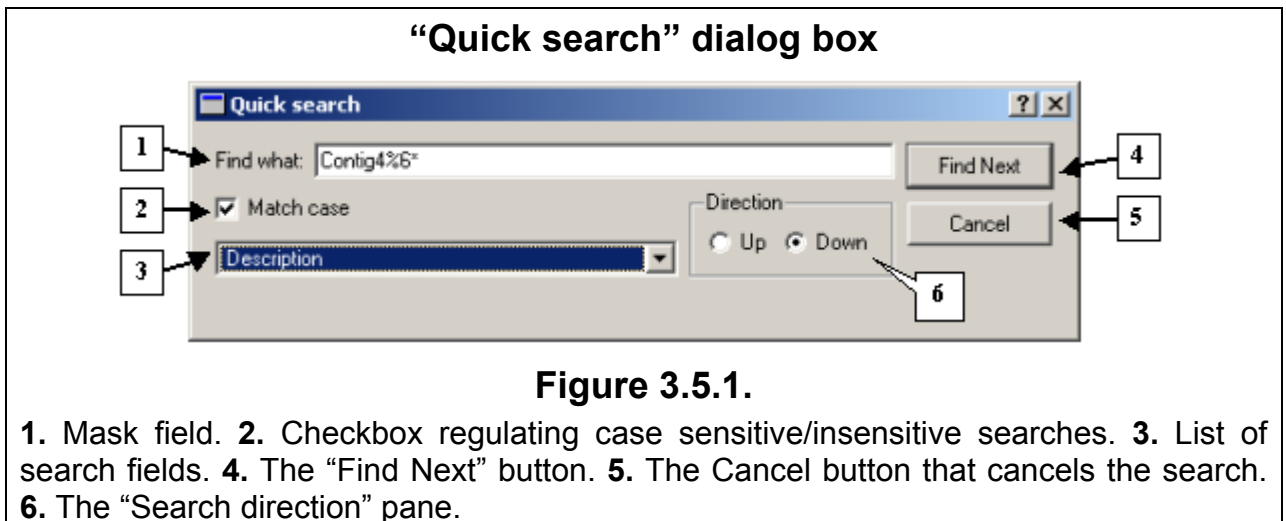
In the “Quick search” dialog box (Fig. 3.5.1), you can perform a quick search for genes. To open this dialog box, click the Table->Quick search gene in current selection command on the main menu.

In the “Find what” text field, enter the search mask. In the mask, the symbol “%” substitutes any single symbol and “\*” substitutes any number of symbols.

The “Match case” checkbox. If the checkbox is checked, the search is case sensitive; otherwise, the search is case insensitive.

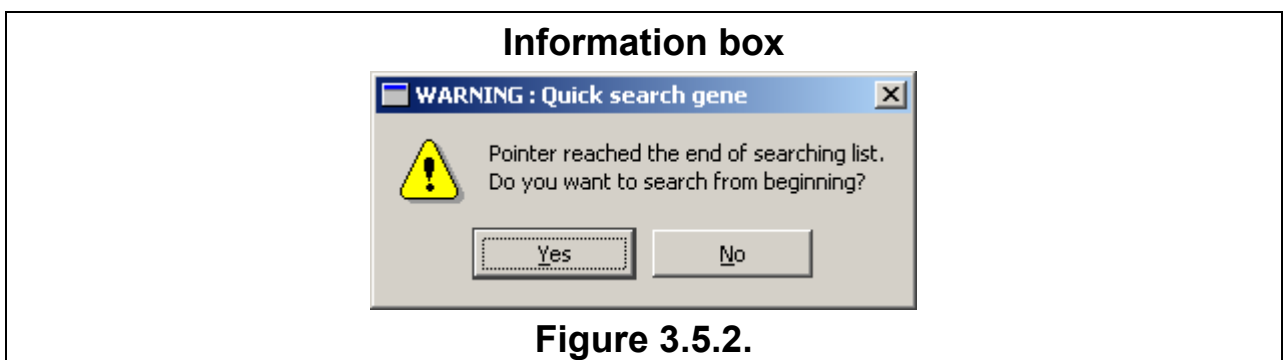
To perform a search by mask only within a specific field, define the name of the field in the list of fields.

You can define the search direction in the “Direction” pane. When the “Up” option is checked, the search will be performed backward from the current line to the top of the table. When the “Down” option is enabled, the program will search forward from the current location to the end of the table.



Click the “Find Next” button to find the next match for the search. A line with the gene found will be highlighted in the table. To cancel the search, click “Cancel”.

When the pointer reaches the end of the table, you will see the corresponding information box (Fig. 3.5.2).



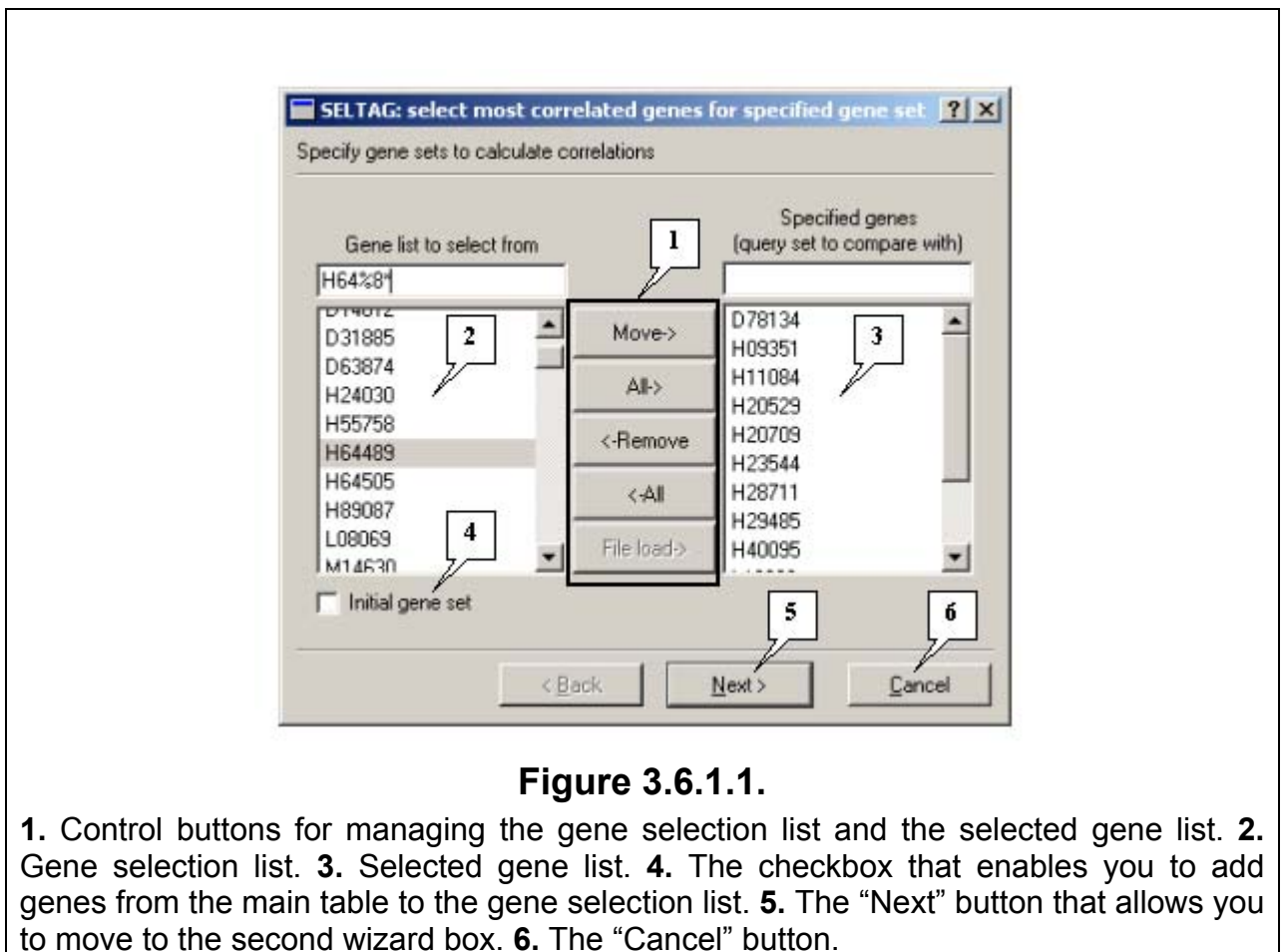
### 3.6. Selecting genes with similar expression profiles

This feature is accessed from the Analysis>Correlations>Select most correlated genes command on the main menu. To specify the parameters of your query, follow the procedure organized as wizard boxes.

#### 3.6.1. Specifying gene sets

The first wizard box (Fig. 3.6.1.1) helps you create selection lists.

3.6.1.1. A list of genes available for selection is titled “Gene list to select from.” The list includes genes from the current table. You can enter a selection mask into the text field above the gene list. In the mask, the “%»” symbol substitutes any single symbol and “\*” substitutes any number of symbols. In Figure 3.6.1, you can see the result of gene selection by mask.



**Figure 3.6.1.1.**

1. Control buttons for managing the gene selection list and the selected gene list. 2. Gene selection list. 3. Selected gene list. 4. The checkbox that enables you to add genes from the main table to the gene selection list. 5. The “Next” button that allows you to move to the second wizard box. 6. The “Cancel” button.

3.6.1.2. A list of selected genes is titled “Specified genes (query set to compare with)”. The list contains genes selected from list 3.6.1.1. You can enter a selection mask into the text field above the list. In the mask, the “%»” symbol substitutes any single symbol and “\*” substitutes any number of symbols.

A set of genes (Set 2) that can contain one or more genes is selected from list 3.6.1.1. Set 2 is displayed in list 3.6.1.2. Set 1 is the genes that remain in list 3.6.1.1. Sets 1 and 2 have no common genes.

3.6.1.3. Control buttons

“Add→” – Move the genes selected in list 1 to list 2.

“All→” – Move all genes from list 1 to list 2.

“←Remove” – Move the genes selected in list 2 to list 1.

“←All” – Move all genes from list 2 to list 1.

“File load→” – Import list 2 from a file. In this version of the program, this option is unavailable.

In Set 1, the program selects genes with expression profiles most similar to those of genes from Set 2. When you exit the dialog box, the program creates the expression profiles for selected genes and writes the correlation coefficients in a separate table.

3.6.1.4. If the “Initial gene set” checkbox is checked, the gene names from the initial table will be added to list 1, regardless of the current table.

To proceed with the operation, click “Next”; to cancel the operation, click “Cancel.”

### 3.6.2. Setting up calculation parameters

In the next wizard box (Fig. 3.6.2.1), you can define values for the following parameters:

3.6.2.1. The “Fields...” button calls the “Field selection” dialog box (Fig. 3.3). The text field displays the number of selected fields.

3.6.2.2. The “Correlation type” list contains three types of correlation coefficients:

- Pearson r – Pearson’s correlation coefficient.
- Spearman r – Spearman’s correlation coefficient.
- Kendall tau – Kendall’s correlation coefficient.

3.6.2.3. The “Threshold type” list contains four types of restrictions on gene selection:

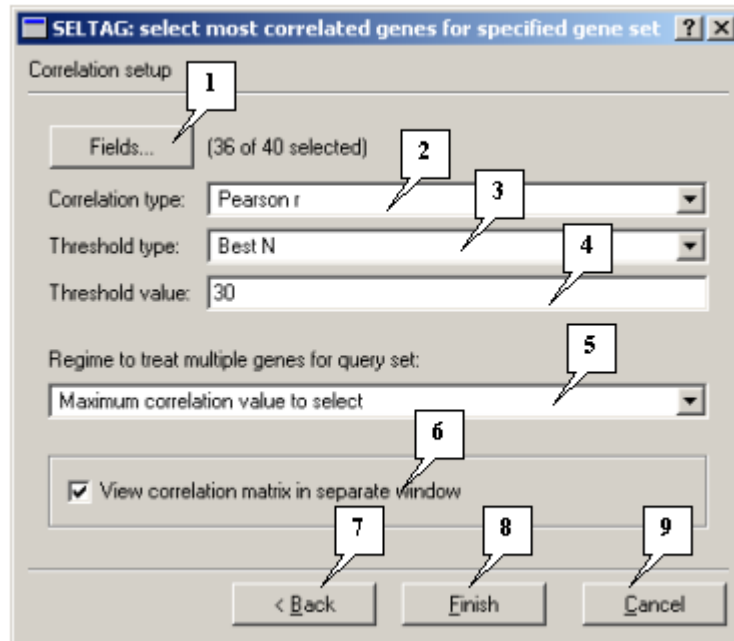
- “Best N” – select N most correlated genes from set 1 (N should be defined in the “and value” field).
- “Best %” – select a portion (in %) of most correlated genes from set 1 (% should be defined in the “and value” field).
- “Value” – select the genes with the absolute correlation value equal or higher than the threshold, which is defined in the “Threshold value” field.
- “All” – select all genes from list 1.

3.6.2.4. In the “Threshold value” field, set the value of the correlation threshold (see section 3.6.2.3.).

3.6.2.5. The “Regime to treat multiple genes for query set” list offers three computation modes:

- “Max. correlation value to select” – when comparing genes from Set 1, the key parameter is the maximum coefficient of correlation of a gene from Set 1 with genes from Set 2.
- “Aver. correlation value to select” – when comparing genes from Set 1, the key parameter is the average coefficient of the correlation of a gene from Set 1 with genes from Set 2.
- “Corr. for aver. field values to select” – when comparing genes from Set 1, the key parameter is the coefficient of correlation of a gene from Set 2 with an average profile of genes from Set 2. This means that the program creates an “imaginary” average gene from Set 2 and uses this average value to calculate the correlation coefficient.

To finish working with the wizard and calculate the correlation, click “Finish”. The program will create expression profiles for the selected genes and write the correlation coefficients in a separate table. To move to the next wizard box, click “Back”; to cancel the operation, click “Cancel”.

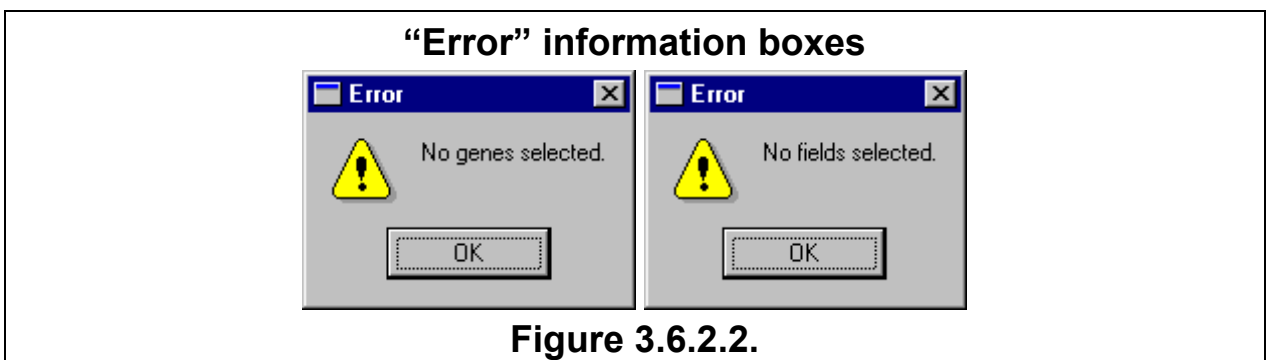


**Figure 3.6.2.1.**

1. The “Field” button that calls the “Field selection” dialog box. 2. List of correlation types. 3. List of threshold types. 4. The threshold value input field. 5. Correlation modes. 6. The checkbox that allows you to view the correlation matrix in a separate window. 7. The “Back” button that takes you to the previous wizard box. 8. The “Finish” button that allows you to finish working with the wizard and perform calculations. 9. The “Cancel” button.

**Notes.**

If the sets of genes have not been formed or calculation fields have not been determined, the “Error” information box with relevant messages will appear (Fig. 3.6.2.2).



**Figure 3.6.2.2.**

When calculation is in progress, you will see the “Please wait” information box (Fig. 3.6.2.3) that will disappear after the calculation is complete.

## “Please wait” information box

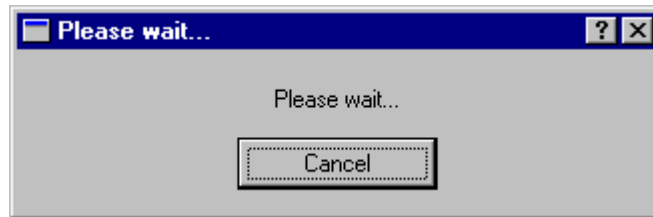


Figure 3.6.2.3.

### 3.6.3. Query results

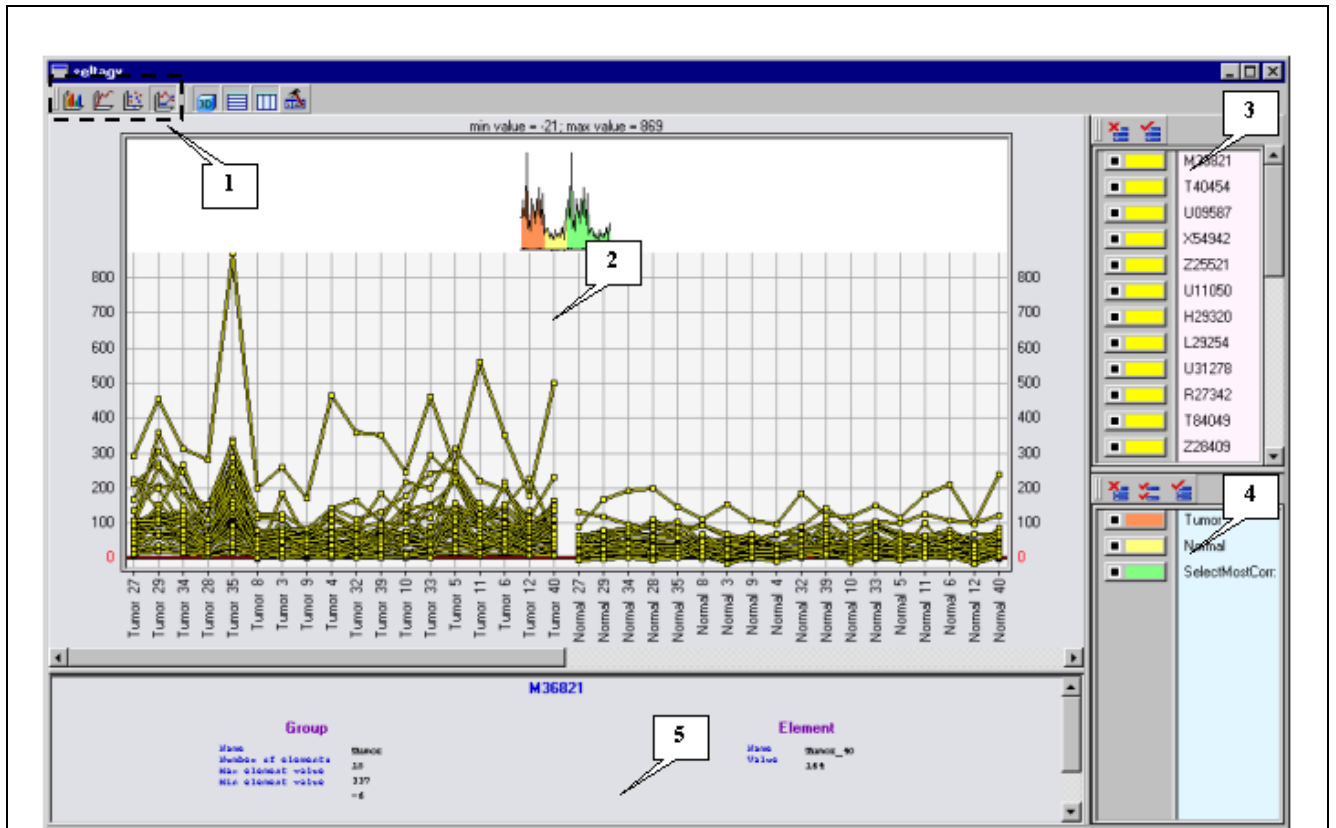
- If the “View corr. matrix in separate window” box is checked, the correlation coefficients will be displayed in a separate window (see section 3.9 about the correlation matrix dialog box). In the matrix, the rows correspond to the genes selected from Set 1 sorted in descending order of their correlation with genes from Set 2. The first column contains the names of genes from Set 1. The second column shows the correlation coefficient that was used in the selection (depending to the selection mode). The remaining columns contain the coefficients of correlation of genes from Set 1 with genes from Set 2.
- The main window will display the table containing the genes selected (Fig. 3.6.3.1). The item corresponding to this new table will be added to the “Select table” dialog box (to open this dialog box, click the Table>Selections list command).

	Accession Number	Description	Sample	Tumor_27	Tumor_29	Tumor_34	Tumor_28	Tumor_35
1	M36821	Human cytokine (GF		137	206	196		119
2	T40454	ya01103.1		167	305	246		110
3	U09587	Human g...		215	271	184		151
4	X54942	H. sapiens ckshs2 m		98	359	193		57
5	Z25521	H. sapiens integrin a		96	261	149		73
6	U11050	Human NIMA-like pr					24	10
7	H29320	ym60e02.s1 Homo s					106	78
8	L29254	Human (clone D21-1					43	20
9	U31278	Human mitotic feedb					56	22
10	R27342	yh53b11.s1 Homo s					75	47
11	T84049	yd76i04.s1 Homo se					112	35
12	Z28409	H. sapiens coprox ge					41	32
13	D13633	Human mRNA for KI					14	5
14	R67999	yu04c12.s1 Homo sa					80	39
15	T54364	ya91g04.s3 Homo s					135	76
16	R46716	yg54i12.s1 Homo se					52	23
17	H80114	yu09g04.s1 Homo s					95	41
18	H87476	yw17g04.s1 Homo s					60	28
19	H81412	yu75g05.s1 Homo s					16	12
20	M61832	Human S-adenosylh					89	154
21	U26312	Human heterochrom					268	65
22	H68165	yu55a12.s1 Homo s					30	15
23	R75843	yl59i12.s1 Homo sa		291	454	313		282
24	T86749	yd77g12.s1 Homo s		100	72	77		30
25	M77836	Human pyroline 5-c		108	106	123		51
26	R87490	ym90i01.s1 Homo s		55	99	81		36
27	H10798	ym04h12.s1 Homo s		94	99	66		56
28	R56401	yg90b09.s1 Homo s		76	77	78		28
29	R71505	yl52h02.s1 Homo sa		44	67	54		17
30	D14657	Human mRNA for KI		44	110	98		54

Figure 3.6.3.1.

1. The table with the genes selected.
2. The name of the new table.

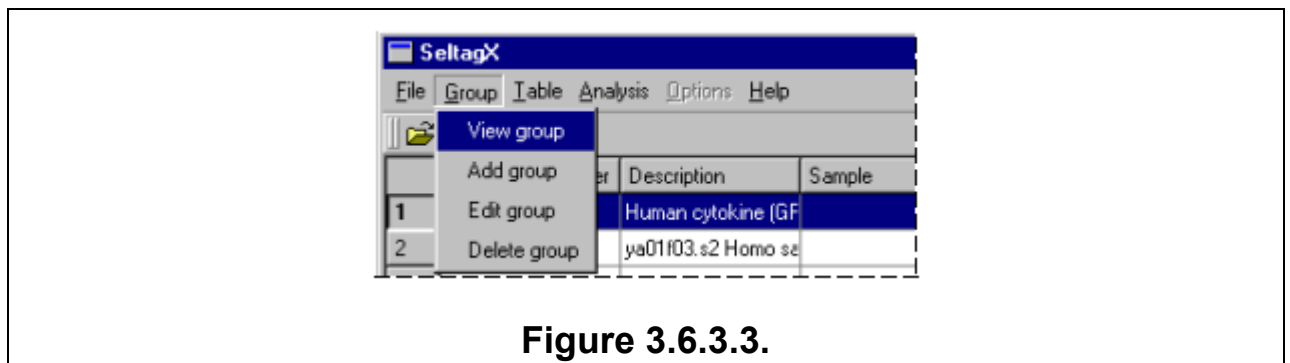
- You will see a dialog box (Fig. 3.6.3.2) with the expression profiles of the selected genes.



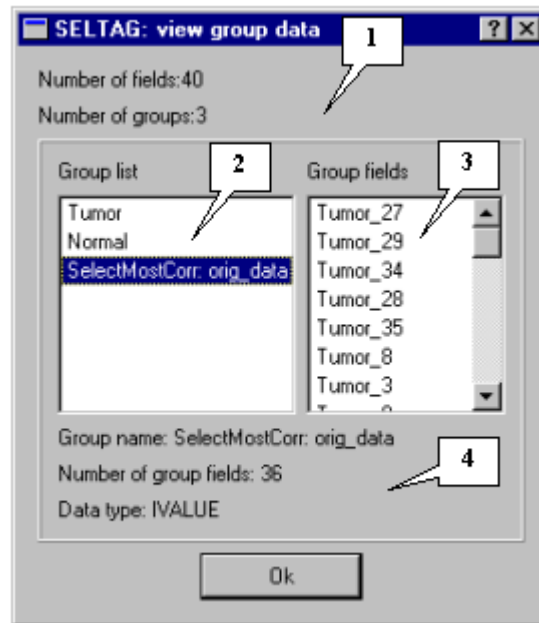
**Figure 3.6.3.2.**

- The buttons for selecting the graphical representation type.
- Profile view area.
- Gene list.
- Group list.
- Information about the gene you are pointing to with your mouse.

- Open the “View group data” dialog box by clicking the Group>View group command on the main menu (Fig. 3.6.3.3). In this dialog box, you can see that a new group titled “SelectMostCorr: orig\_data” was added to the list of groups. This new group contains fields that were used in the computations (Fig. 3.6.3.4).



**Figure 3.6.3.3.**



**Figure 3.6.3.4.**

**1.** Information about the number of fields and groups. **2.** Group list. **3.** The fields included into the selected group. **4.** Information about the selected group.



### 3.7. Searching for correlation between two sets of genes

To initiate a search for correlation between two sets of genes, select the Analysis>Correlations>Get correlations between genes command on the main menu. The Correlation analysis setup wizard will assist you with setting up search parameters.

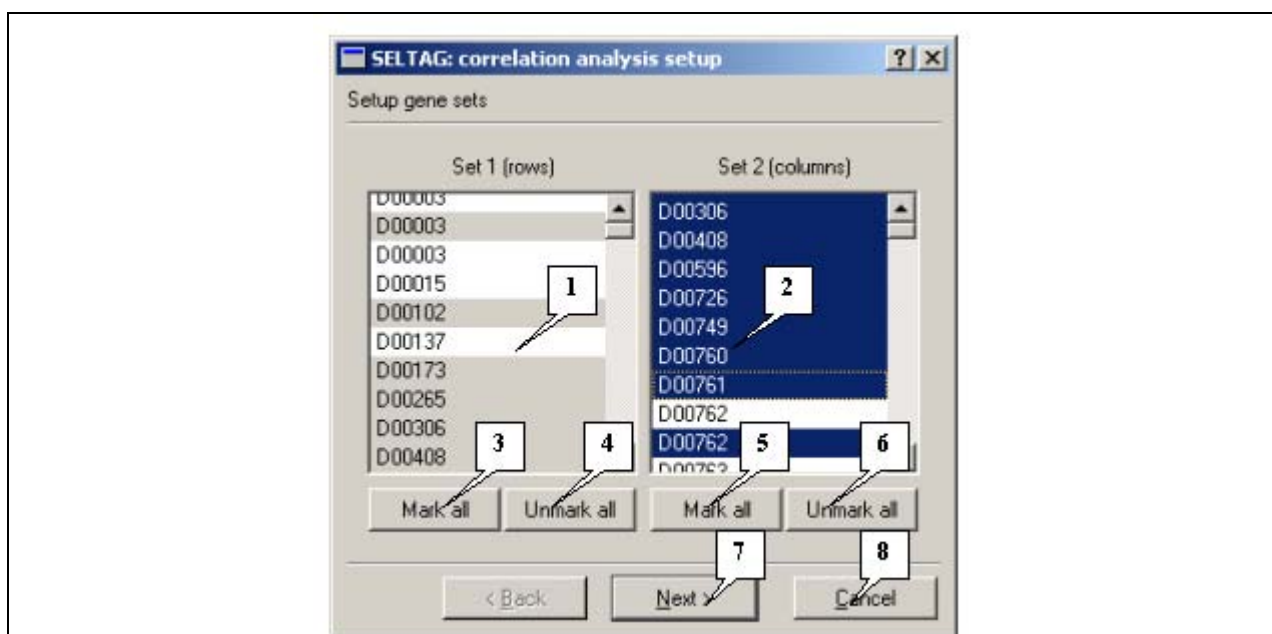
Select two sets from the total number of genes: Set 1 that corresponds to the rows of the correlation matrix and Set 2 that corresponds to the columns of the correlation matrix. These sets of genes may overlap. The pairwise correlation coefficients are calculated for the expression profiles of genes from these sets. As a result, the program creates an  $n \times m$  correlation matrix, where  $n$  is the number of rows and  $m$  is the number of columns. The matrix element  $(i,j)$  corresponds to the correlation coefficient between the expression profiles of the  $i$ th gene from Set 1 and the  $j$ th gene from Set 2.

In the first wizard box (Fig. 3.7.1), you can define the following two sets from the total number of genes:

3.7.1. The “Set 1 (rows)” list contains genes that comprise Set 1. Click “Mark all” to select all genes in the list and “Unmark all” to unselect all genes in the list.

3.7.2. The “Set 2 (columns)” list includes genes that will comprise Set 2. Click “Mark all” to select all genes and “Unmark all” to unselect all genes in the list.

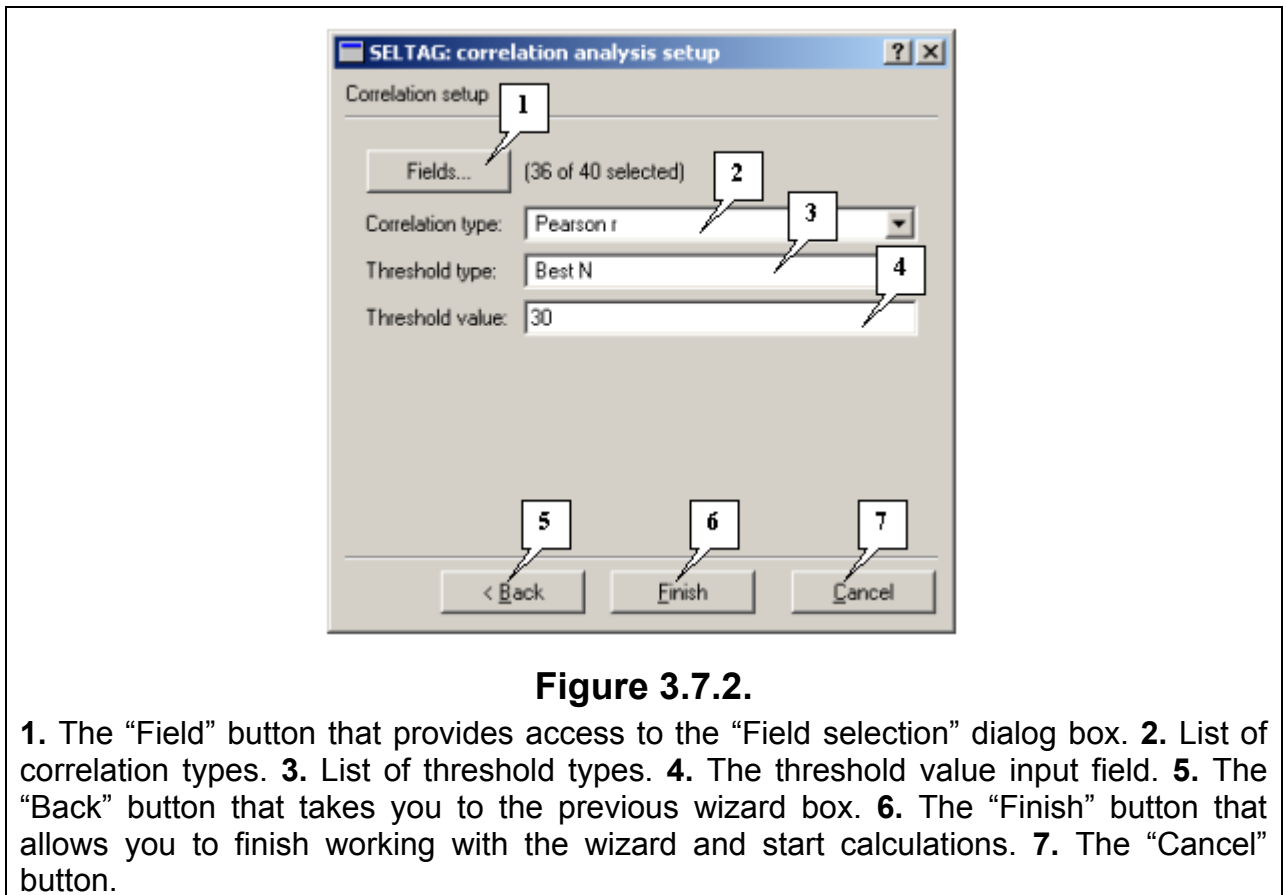
To proceed, click “Next”; to cancel the operation, click “Cancel”.



**Figure 3.7.1.**

1. The Set 1 list of genes. 2. The Set 2 list of genes. 3. The “Mark all” button that selects all genes in the “Set 1” list. 4. The “Unmark all” button that deselects all genes in the “Set 1” list. 5. The “Mark all” button that selects all genes in the “Set 2” list. 6. The “Unmark all” button that deselects all genes in the “Set 2”. 7. The “Next” button that takes you to the next wizard box. 8. The “Cancel” button that cancels the calculations.

The second box of the Correlation analysis setup wizard (Fig. 3.7.2) allows you to define values for the following parameters:



**Figure 3.7.2.**

1. The “Field” button that provides access to the “Field selection” dialog box. 2. List of correlation types. 3. List of threshold types. 4. The threshold value input field. 5. The “Back” button that takes you to the previous wizard box. 6. The “Finish” button that allows you to finish working with the wizard and start calculations. 7. The “Cancel” button.

3.7.3. The “Fields...” button calls the “Field selection” dialog box (see section 3.3 about the “Field selection” dialog box). The text field displays the number of selected fields.

3.7.4. The “Correlation type” list contains the following three types of correlation coefficients:

- Pearson  $r$  – Pearson’s correlation coefficient.
- Spearman  $r$  – Spearman’s correlation coefficient.
- Kendall tau – Kendall’s correlation coefficient.

3.7.5. The “Threshold type” list contains four types of restrictions on gene selection (the matrix elements that exceed the threshold value are shown in red):

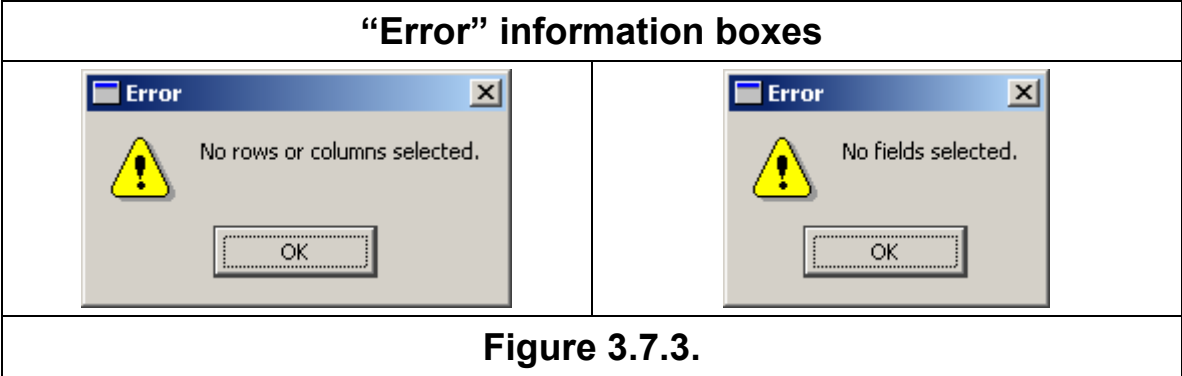
- “Best N” – select N genes with the best correlation value (N should be defined in the “Value” field).
- “Best %” - select a portion (in %) of the genes with the best correlation value (the percentage should be defined in the “Value” field).
- “Value” – select the genes with the absolute correlation value equal or higher than the threshold, which is defined in the “Threshold value” field.
- “All” – select all genes from list 1.

3.7.6. In the “Threshold value” field, set the value of the correlation threshold (see section 3.4.5).

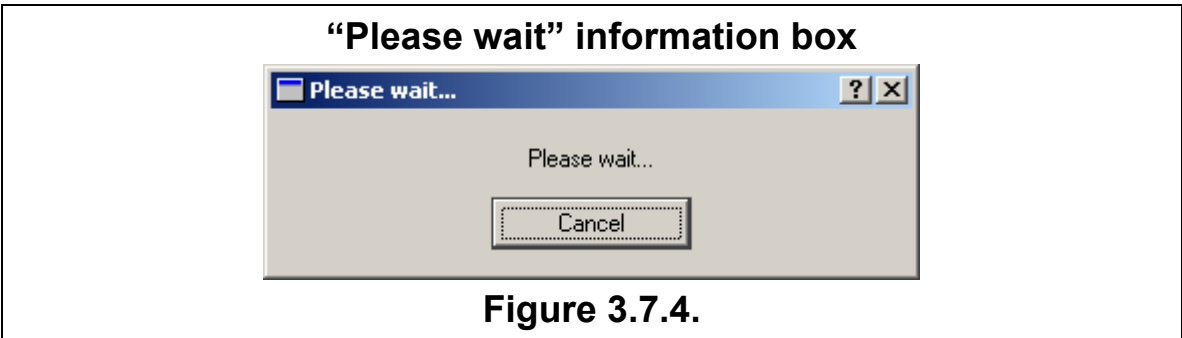
To finish working with the wizard and assess the correlation, click “Finish”. To move to the previous wizard box, click “Back”; to cancel the operation, click “Cancel”.

**Note.**

If the sets of genes have not been formed or calculation fields have not been determined, the “Error” information box with relevant messages will appear (Fig. 3.7.3).



When the calculation is in progress, you will see the “Please wait” information box (Fig. 3.7.4) that will disappear after the calculation is complete.



After the program completes the calculation, it will open a dialog box with the correlation matrix (see section 3.9 about this dialog box).

### 3.8. Searching for correlation between two sets of fields

To initiate a search for correlation between two sets of fields, select the Analysis>Correlations>Get correlations between fields command on the main menu. The Correlation analysis setup wizard will assist you with setting up search parameters.

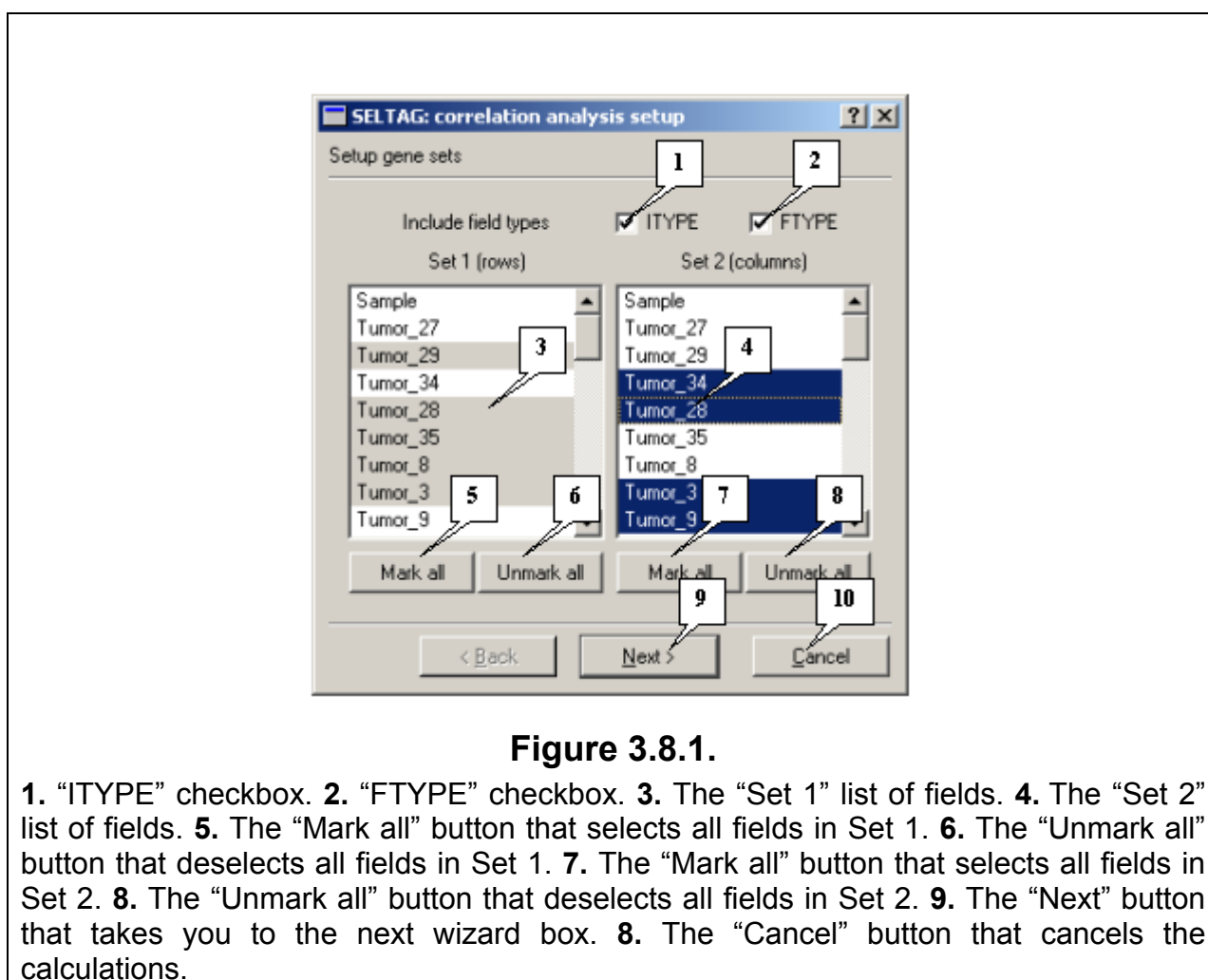
The first box of the wizard (Fig. 3.8.1) has the following elements:

3.8.1. “ITYPE” checkbox. If this checkbox is checked, lists 3.8.3 and 3.8.4 will include all fields of the ITYPE type.

3.8.2. “FTYPE” checkbox. If this checkbox is checked, lists 3.8.3 and 3.8.4 will include all fields of the FTYPE type.

3.8.3. The “Set 1 (rows)” list contains fields that will be included into Set 1. Click “Mark all” to select all fields in the list and “Unmark all” to unselect all fields in the list.

3.8.4. The “Set 2 (columns)” list includes fields that will comprise Set 2. Click “Mark all” to select all fields in the list and “Unmark all” to unselect all fields in the list.



**Figure 3.8.1.**

1. “ITYPE” checkbox. 2. “FTYPE” checkbox. 3. The “Set 1” list of fields. 4. The “Set 2” list of fields. 5. The “Mark all” button that selects all fields in Set 1. 6. The “Unmark all” button that deselects all fields in Set 1. 7. The “Mark all” button that selects all fields in Set 2. 8. The “Unmark all” button that deselects all fields in Set 2. 9. The “Next” button that takes you to the next wizard box. 10. The “Cancel” button that cancels the calculations.

The second box of the Correlation analysis setup wizard (Fig. 3.8.2) has the following elements:

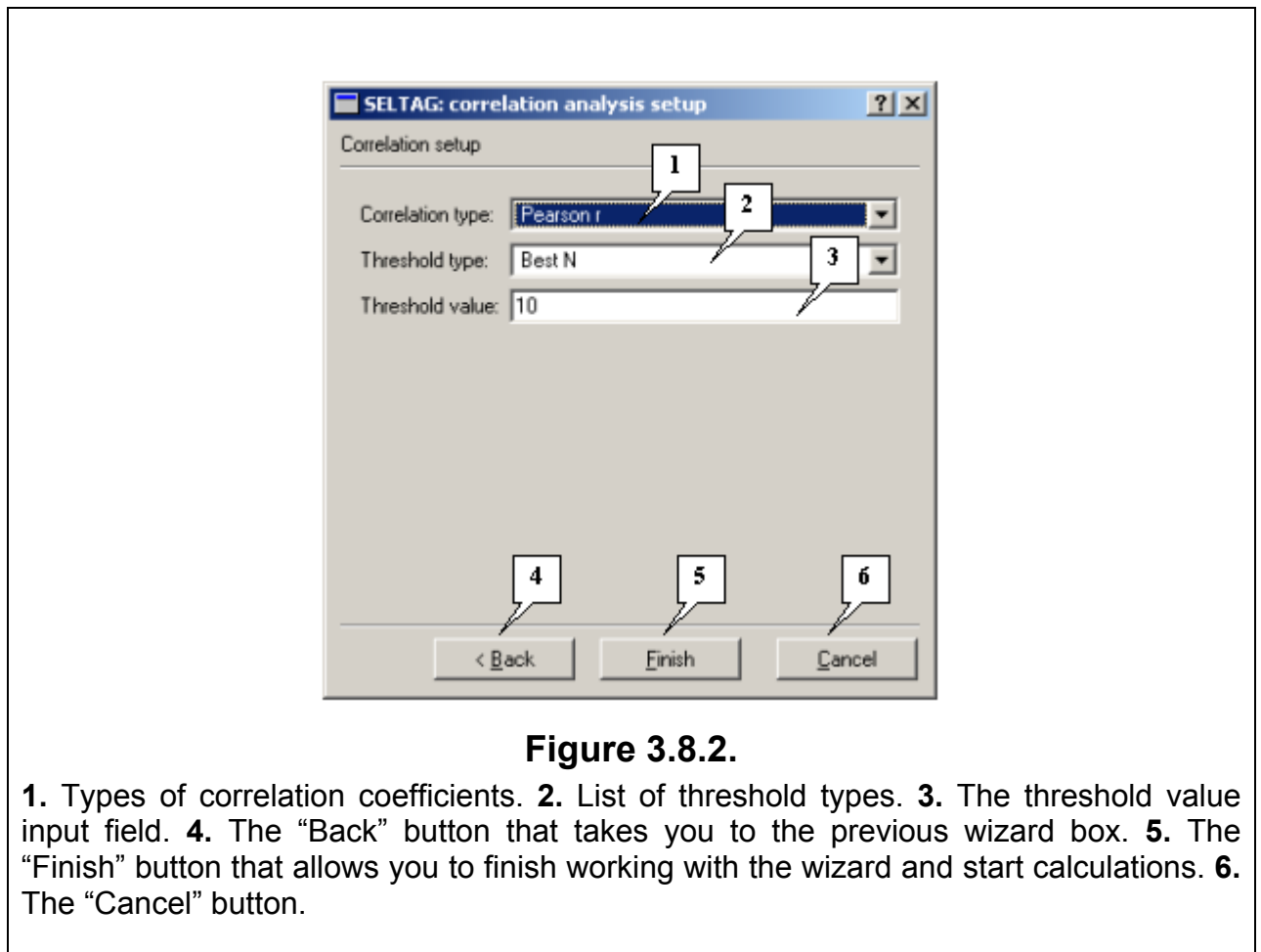
3.8.5. The “Correlation type” list that contains the following three correlation coefficients:

- Pearson r – Pearson’s correlation coefficient.
- Spearman r – Spearman’s correlation coefficient.
- Kendall tau – Kendall’s correlation coefficient.

3.8.6. The “Threshold type” list contains four types of correlation coefficient thresholds (the matrix elements that exceed the threshold value are shown in red):

- “Best N” – select N best correlation coefficients (N should be defined in the “Value” field).
- “Best %” - select a portion (in %) of the best correlation values (the percentage should be defined in the “Value” field).
- «Value» – select the correlation coefficients equal or higher than the threshold, which is defined in the “Threshold value” field.
- «All» – select all fields from list 1.

3.8.7. In the “Threshold value” field, set the value of the correlation threshold (see section 3.8.6).



**Figure 3.8.2.**

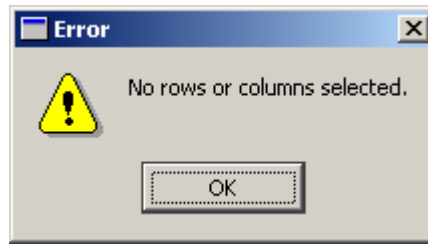
**1.** Types of correlation coefficients. **2.** List of threshold types. **3.** The threshold value input field. **4.** The “Back” button that takes you to the previous wizard box. **5.** The “Finish” button that allows you to finish working with the wizard and start calculations. **6.** The “Cancel” button.

To finish working with the wizard and calculate the correlation, click “Finish”. To move to the previous wizard box, click “Back”; to cancel the operation, click “Cancel”.

**Note.**

If the sets of fields have not been formed, the “Error” information box with the relevant messages will appear (Fig. 3.8.3).

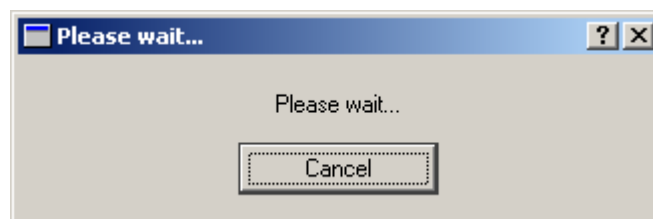
### “Error” information box



**Figure 3.8.3.**

When the calculation is in progress, you will see the “Please wait” information box (Fig. 3.8.4). This information box will disappear after the calculation is complete.

### “Please wait” information box



**Figure 3.8.4.**

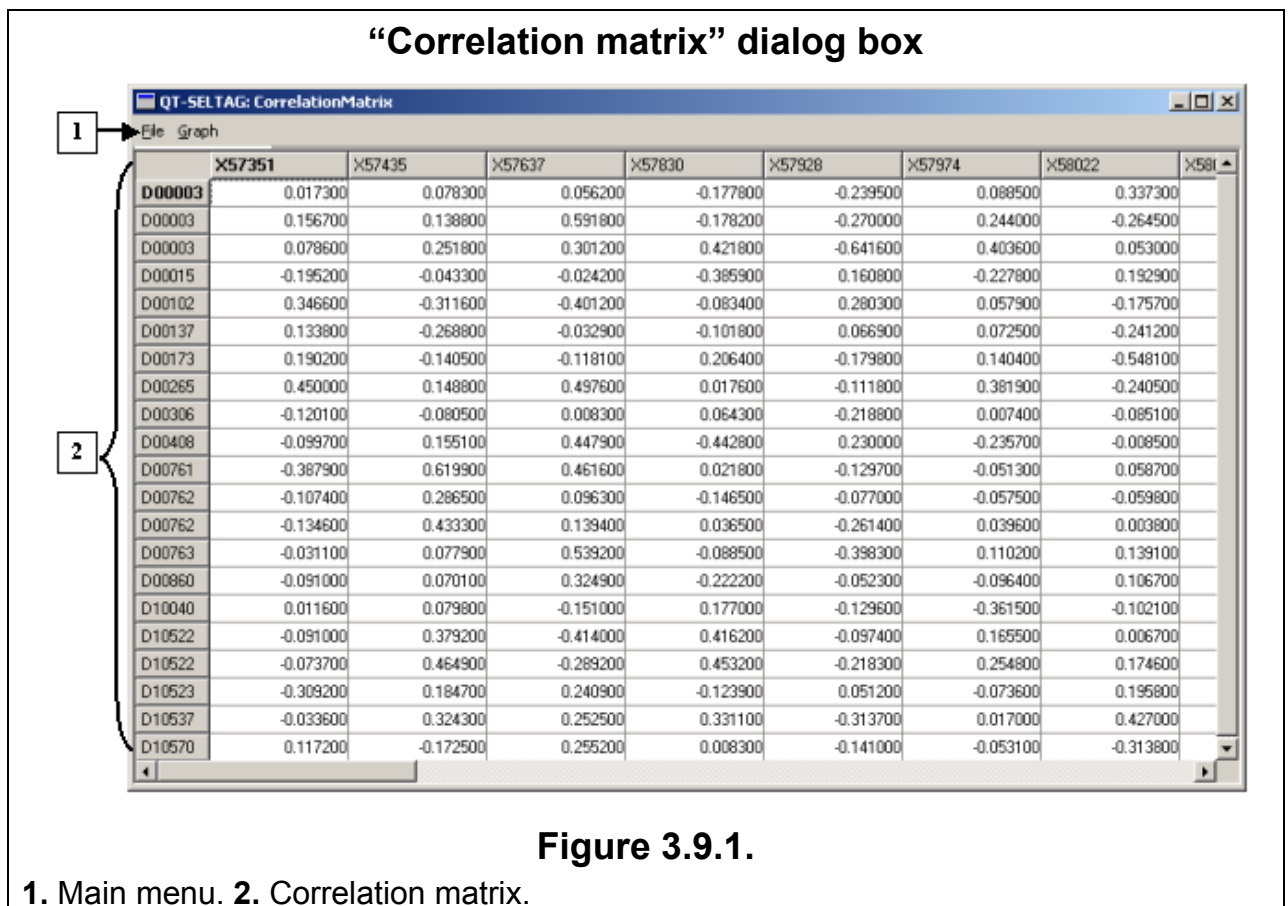
After the program completes the calculation, it will open a dialog box with the correlation matrix (see section 3.9 about this dialog box).

### 3.9. Correlation matrix dialog box

The “Correlation matrix” dialog box (Fig. 3.9.1) can be accessed from the “Select most correlated genes for specified gene set” (if the “View corr. matrix in separate window” box is checked) and the “correlation analysis setup” dialog boxes.

Main menu commands.

1. The “File” menu contains commands for managing files:
  - Save matrix – not available in this version.
  - Save row genes – not available in this version.
  - Quit – exits the dialog box.
2. The “Graph” menu includes a set of commands for operating the matrix.
  - Line plot – calls the “setup for correlations graph” dialog box so that you can map the values of the correlation matrix.

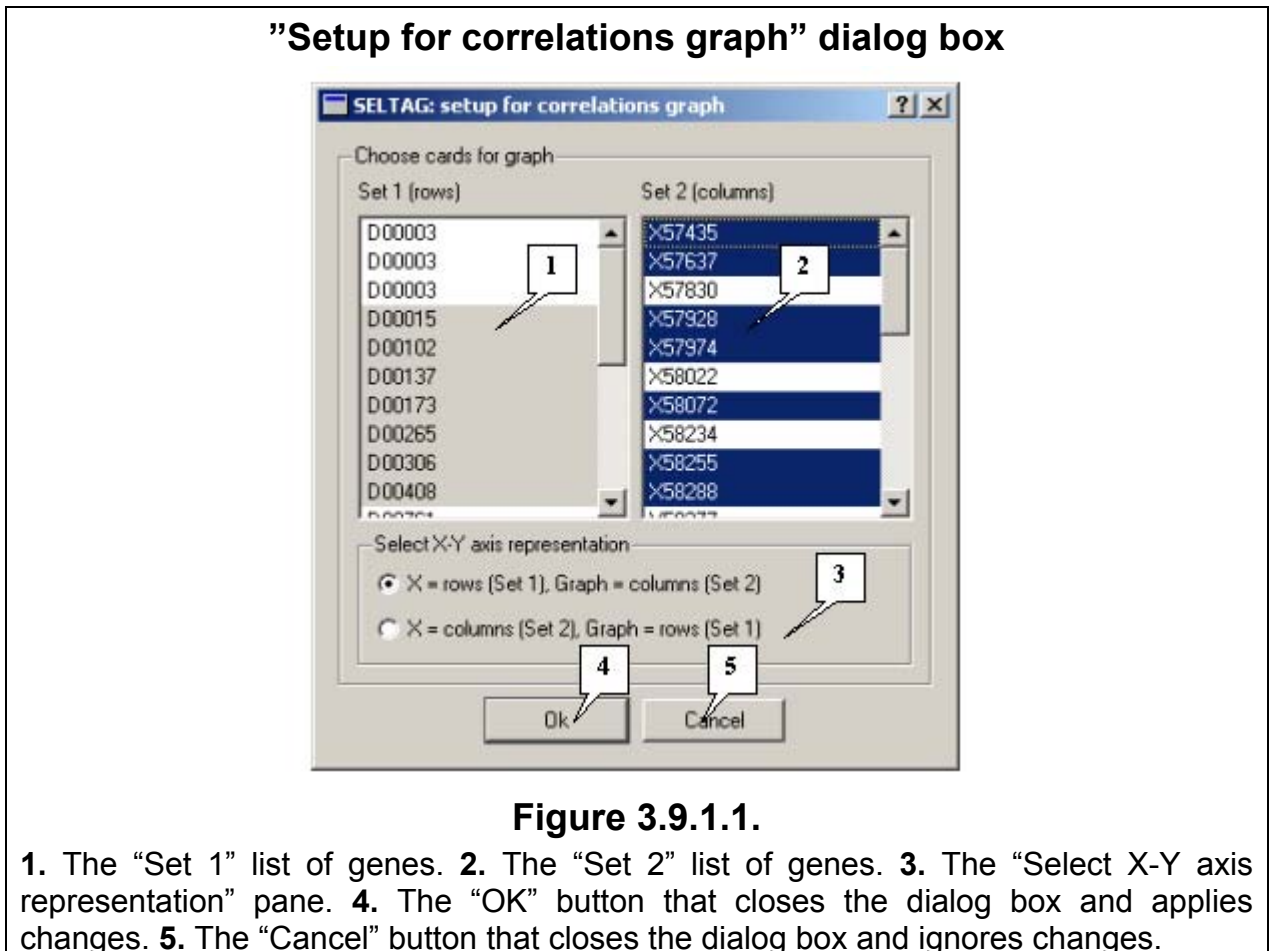


#### 3.9.1. “Setup for correlations graph” dialog box

The “Setup for correlations graph” dialog box (Fig. 3.9.1.1) is called by selecting the Graph>Line plot command on the main menu or pressing the Ctrl+L keys on your keyboard. In this dialog box, you can define the parameters of graphical representation of the correlation matrix in the profile dialog box (about the profile dialog box, see. 3.17):

- 3.9.1.1. The “Set 1 (rows)” list includes genes that will comprise Set 1.
- 3.9.1.2. The “Set 2 (columns)” list includes genes that will comprise Set 2.
- 3.9.1.3. “Select X-Y axis representation” offers two options of axis representation:
  - “X = rows (Set1), Graphs=columns (Set2)” - the X axis represents the genes from Set 1 and the Y axis represents the coefficients of correlation between the genes from Set 1 and Set 2.

- “X = columns (Set2), Graphs=rows (Set1)” – the X axis represents the genes from Set 2 and the Y-axis represents the coefficients of correlation between the genes from Set 1 and Set 2. The list of genes contains genes only from Set 1.
- 3.9.1.4. Control buttons:
- “Cancel” – exits the dialog box; no plot is constructed.
  - “OK” – calls the profile dialog box with the specified representation parameters (Fig. 3.9.1.2).





# Profile dialog box

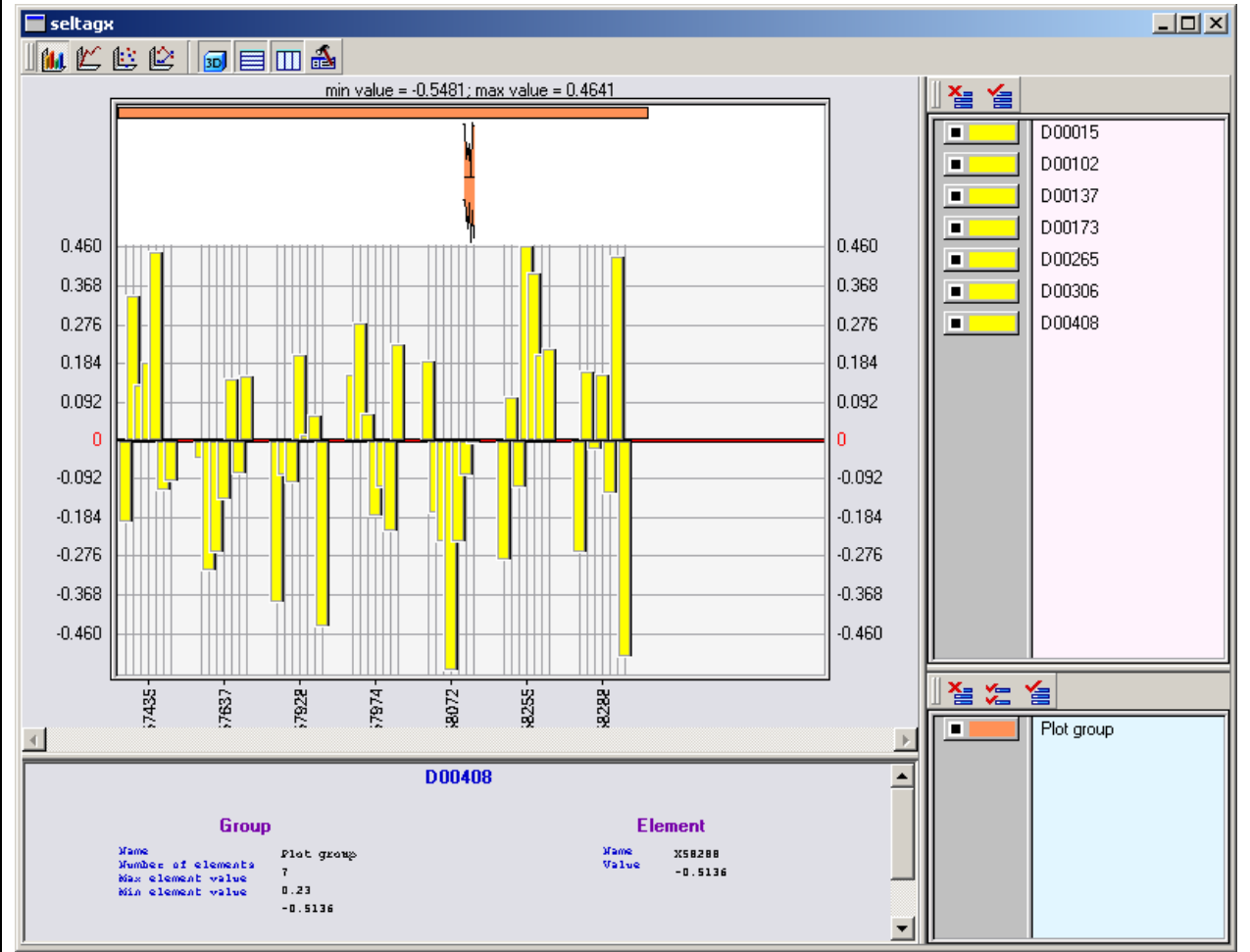
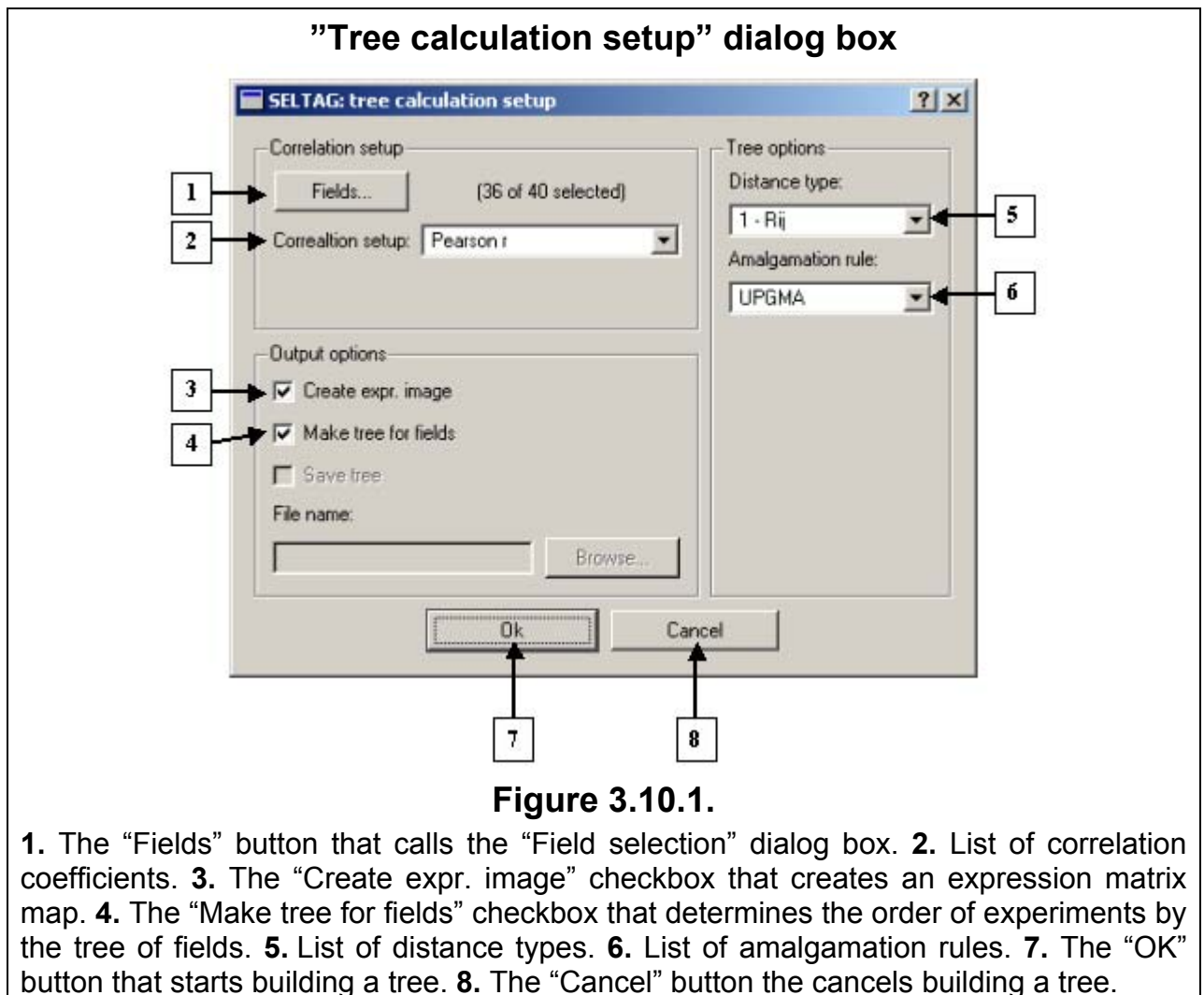


Figure 3.9.1.2.

### 3.10. Building a gene tree

Click the Analysis>Clustering>Build gene tree command to open the “Tree calculation setup” dialog box (Fig. 3.10.1) from where you can define the parameters for building a gene tree for the current table.



3.10.1. The “Fields...” button opens the “Field selection” dialog box (see section 3.3). At the right of this button, you can see the number of selected fields.

3.10.2. The “Correlation type” list contains the following three correlation coefficients:

- Pearson r – Pearson’s correlation coefficient
- Spearman r – Spearman’s correlation coefficient
- Kendall tau – Kendall’s correlation coefficient.

3.10.3. The “Distance type” list includes three types of distance measures calculated from the Rij correlation coefficients:

- 1-Rij
- 1+Rij
- 1-|Rij|

3.10.4. The “Amalgamation rule” list has four types of tree node combination:

- UPGMA
- Nearest neighbor
- Furthest neighbor

- WPGMA

3.10.5. The “Create expr. Image” checkbox. When this checkbox is checked, the program creates an expression matrix profile after it completes calculations.

3.10.6. The “Make tree for fields” checkbox. When this checkbox is checked, a similarity tree for the experimental values of the expression matrix is constructed, and the order of the experiments is determined by the field tree.

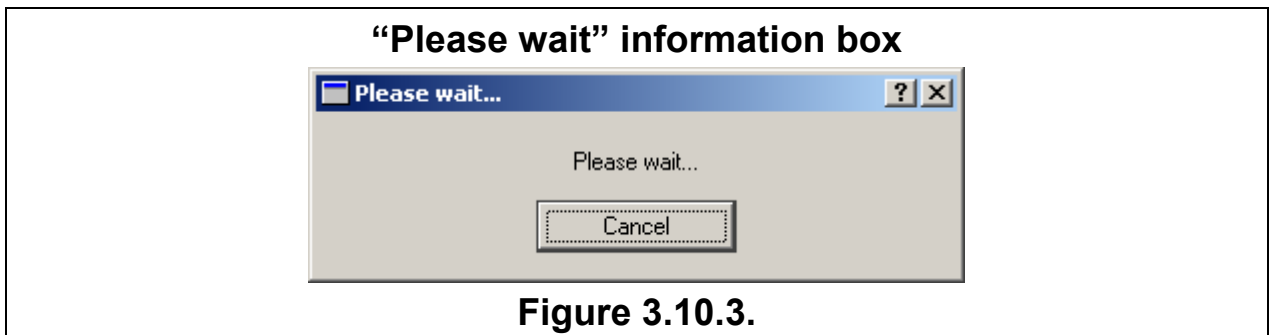
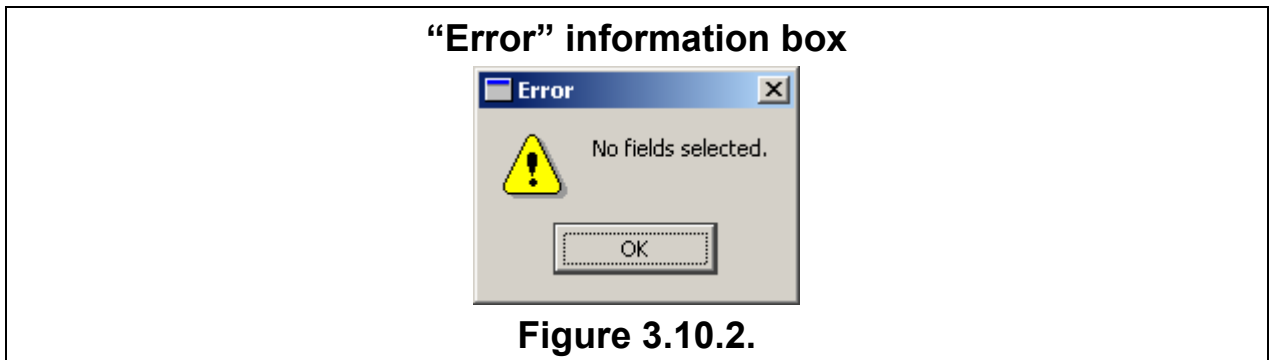
3.10.9. Control buttons:

- “Cancel”- closes the dialog box and cancels the operation.
- “OK” – builds a tree and opens a dialog box with the tree diagram (see section 3.12 about the tree diagram).

**Note.**

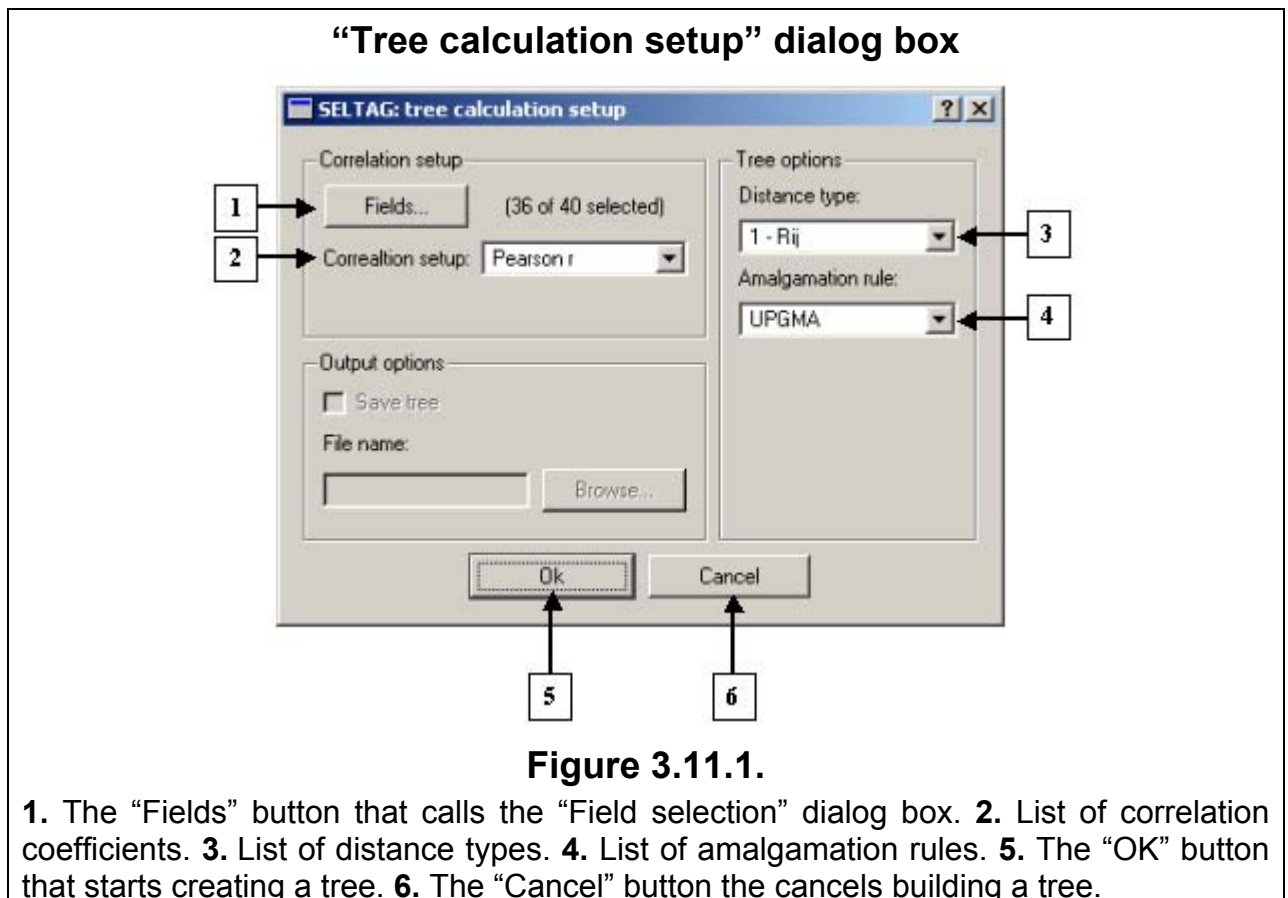
If no fields were selected, the “Error” information box will appear on your screen (Fig. 3.10.2).

When the calculation is in progress, you will see the “Please wait” information box (Fig. 3.10.3). This information box will disappear after the calculation is complete.



### 3.11. Building a field tree

Click the Analysis>Clustering>Build field tree command to open the “Tree calculation setup” dialog box (Fig. 3.11.1) from where you can define the parameters for building a field tree for the current table.



3.11.1. The “Fields...” button opens the “Field selection” dialog box (see section 3.3). At the right of this button, you can see the number of selected fields.

3.11.2. The “Correlation type” list contains the following three correlation coefficients:

- Pearson r – Pearson’s correlation coefficient
- Spearman r – Spearman’s correlation coefficient
- Kendall tau – Kendall’s correlation coefficient.

3.11.3. The “Distance type” list includes three types of distance measures calculated from the Rij correlation coefficients:

- 1-Rij
- 1+Rij
- 1-|Rij|

3.11.4. The “Amalgamation rule” list has the following four types of tree node combination:

- UPGMA
- Nearest neighbor
- Furthest neighbor
- WPGMA

3.11.5. Control buttons

- “Cancel”- closes the dialog box and cancels the operation.

- “OK” – builds a tree and opens a dialog box with the tree diagram (see section 3.12 about the tree diagram).

**Note.**

If no fields were selected, the “Error” information box will appear on your screen (Fig. 3.11.2).

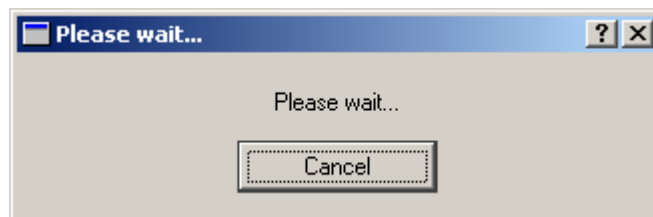
When the calculation is in progress, you will see the “Please wait” information box (Fig. 3.11.3). This information box will disappear after the calculation is complete.

**“Error” information box**



**Figure 3.11.2.**

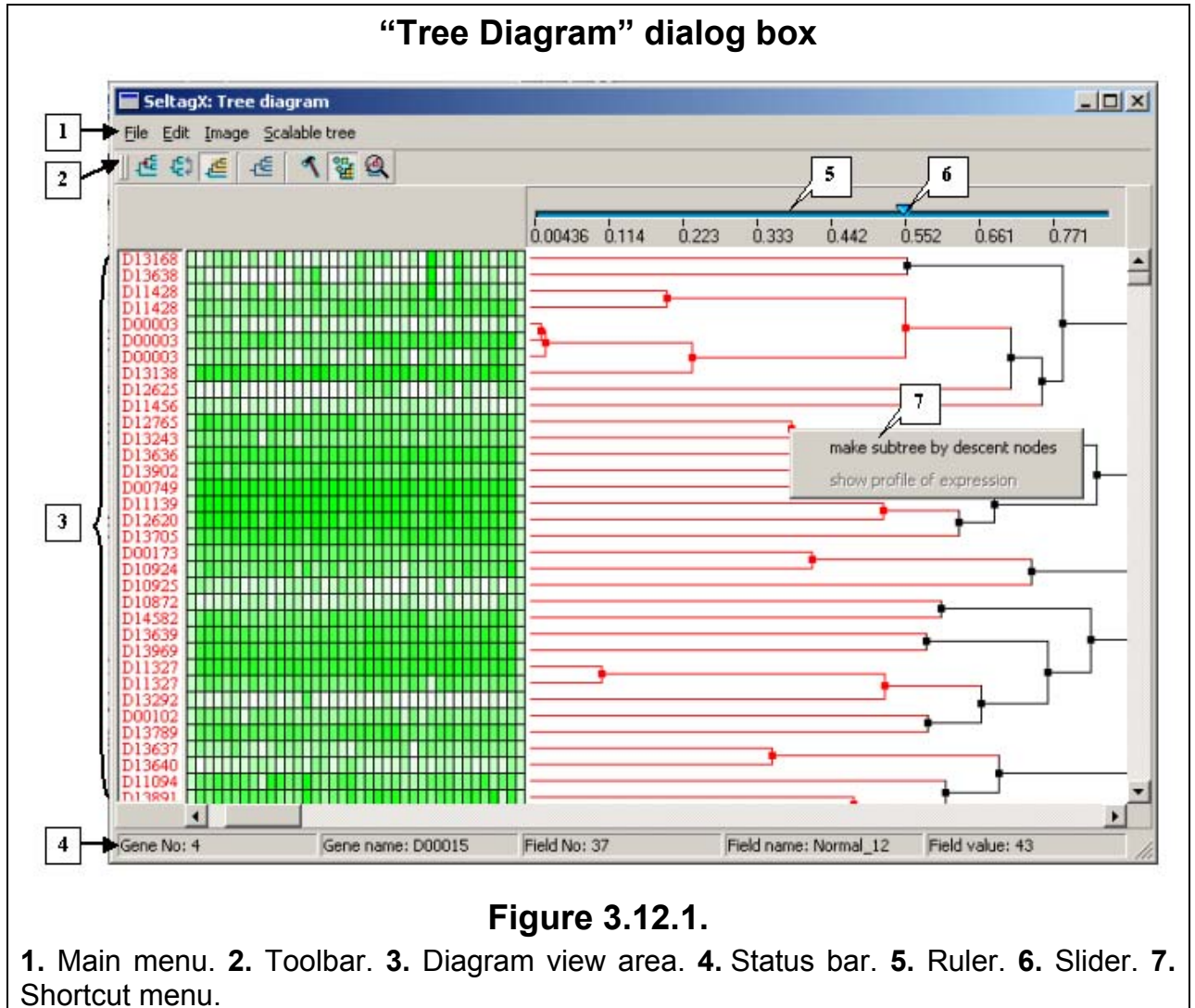
**“Please wait” information box**



**Figure 3.11.3.**

### 3.12. “Tree diagram” dialog box

The “Tree Diagram” dialog box (Fig. 3.12.1) will open after you set up the tree parameters and create a tree in the “Tree calculation setup” dialog box. In the “Tree diagram” dialog box, the tree is presented as a diagram. If you set the program to build a gene tree, the diagram will display the tree and the expression matrix. If you set the program to build a field tree, the diagram will display only the tree.



The “Tree diagram” dialog box has the following elements:








- Main menu
- Toolbar
- Diagram view area
- Status bar
- Ruler

#### Main menu commands

- The “File” menu has one command:
  - Close – closes the dialog box.
- The “Edit” menu offers several modes for editing the tree diagram.
  - Mark – selection commands:

- Mark nodes – opens the “Mark nodes” dialog box from where you can set the parameters for selecting nodes.
    - Unmark all – unselect all nodes.
  - Mouse mode – mouse settings:
    - Mark nodes –you can select nodes on click.
    - Swap child nodes –you can swap child nodes on click.
    - Marking all descent nodes –you can select all child nodes on a single click.
  - Options – calls the “tree options” dialog box.
- The “Image” menu contains commands for working with the expression matrix diagram.
  - “View expr. Image” – if this checkbox is checked, a diagram of the expression matrix will be displayed.
  - “Image setup” – calls the “pattern setup” dialog box from where you can configure parameters of the expression matrix image.
- The “Scalable Tree” menu has one command:
  - “Open Tree” – calls the “Scalable Tree Diagram” dialog box, which allows you to scale the tree obtained.

#### Toolbar buttons:

	“Nodes marking mode” – enables the nodes marking mode. This button corresponds to the Edit>Mouse mode>Mark nodes main menu command.
	“Swap child nodes mode” – enables swapping child nodes. This button corresponds to the Edit>Mouse mode>Swap child nodes main menu command.
	“Marking of all descent nodes mode” – enables marking all child nodes in the tree. This button corresponds to the Edit>Mouse mode>Marking all descent nodes main menu command.
	“Unmark all nodes” – unmarks all selections. This button corresponds to the Edit>Mark>Unmark all main menu command.
	“Options” – calls the “tree options” dialog box. This button corresponds to the Edit>Mark>Options main menu command.
	“Toggle image” – adds an expression matrix diagram (for a field tree this option is not available).
	“Open scalable tree” – opens the “Scalable Tree Diagram” dialog box from where you can magnify or reduce the tree obtained.

#### Status bar

The status bar displays information about the expression matrix cell you are pointing to with your mouse button: the gene number, gene name, field number, field name, and the field value.

#### Ruler

Use the ruler and the slider to select all leaves located to the left of the vertical line, which appears as you move the slider.

#### 3.12.1. “Mark nodes” dialog box

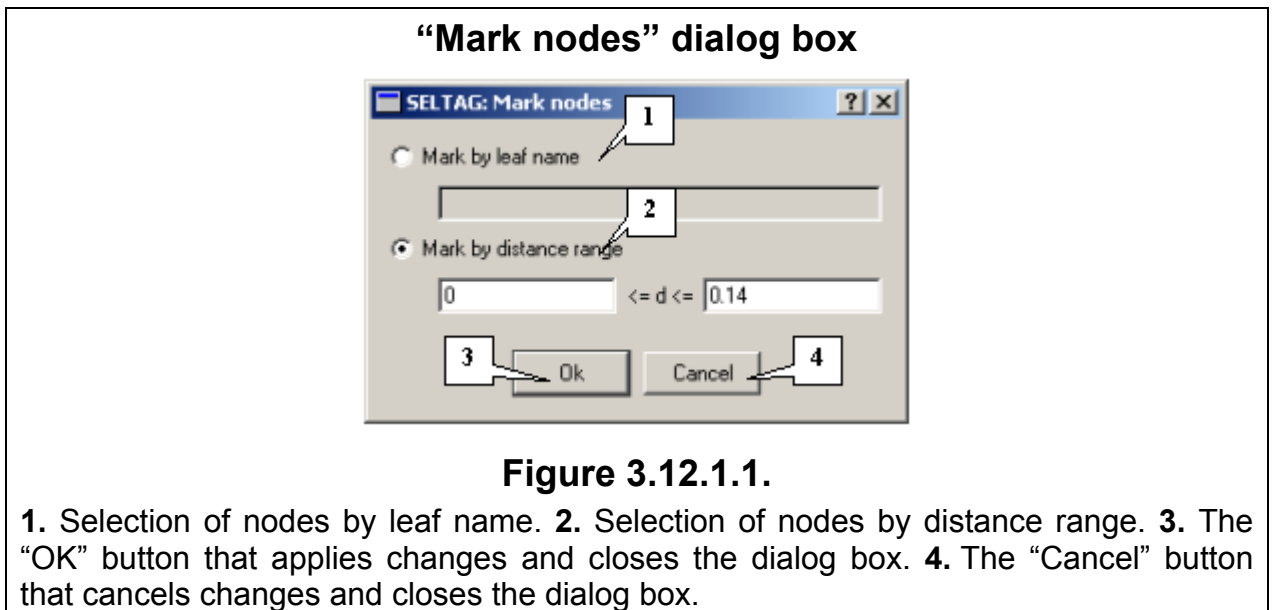
In the “Mark nodes” dialog box (Fig. 3.12.1.1), you can set node selection parameters. To open the dialog box, click Edit>Mark>Mark nodes.

3.13.1.1. The “Mark by leaf name” radio button. If this option is enabled, the nodes are selected by the leaf name specified in the text field below.

3.13.1.2. The “Mark by distance range” radio button. If this option is enabled, the nodes are selected by the specified distance range.

Control buttons:

- “Cancel” – cancels changes and closes the dialog box.
- “OK” – applies changes and closes the dialog box.



### 3.12.2. “Pattern setup” dialog box

In the “Pattern setup” dialog box, you can configure settings for the expression matrix profile. To open the dialog box, click Image>Image setup. See section 3.16.1. about the elements of this dialog box and their functions.

#### Options dialog box

The Options dialog box has the following three tabs:

- The “Common” tab allows you to configure general settings for the image.
- The “Nodes” tab allows you to configure the visualization parameters for tree nodes.
- The “Axis” tab allows you to configure the visualization parameters for axes.

#### “Colors” tab

On the “Colors” tab (Fig. 3.12.2), you can customize colors:

The “**Tree background**” pane allows you to customize the tree background color:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

The “**Tree lines**” pane allows you to customize the color of tree lines:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

The “**Tree nodes**” pane allows you to customize the color of tree nodes:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

The “**Guide line**” pane allows you to customize the color of the vertical line that appears on the screen as you move the slider:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).



- The “Preview” pane shows the color you selected.

The “**Names background**” pane allows you to customize the background color of leaf names:

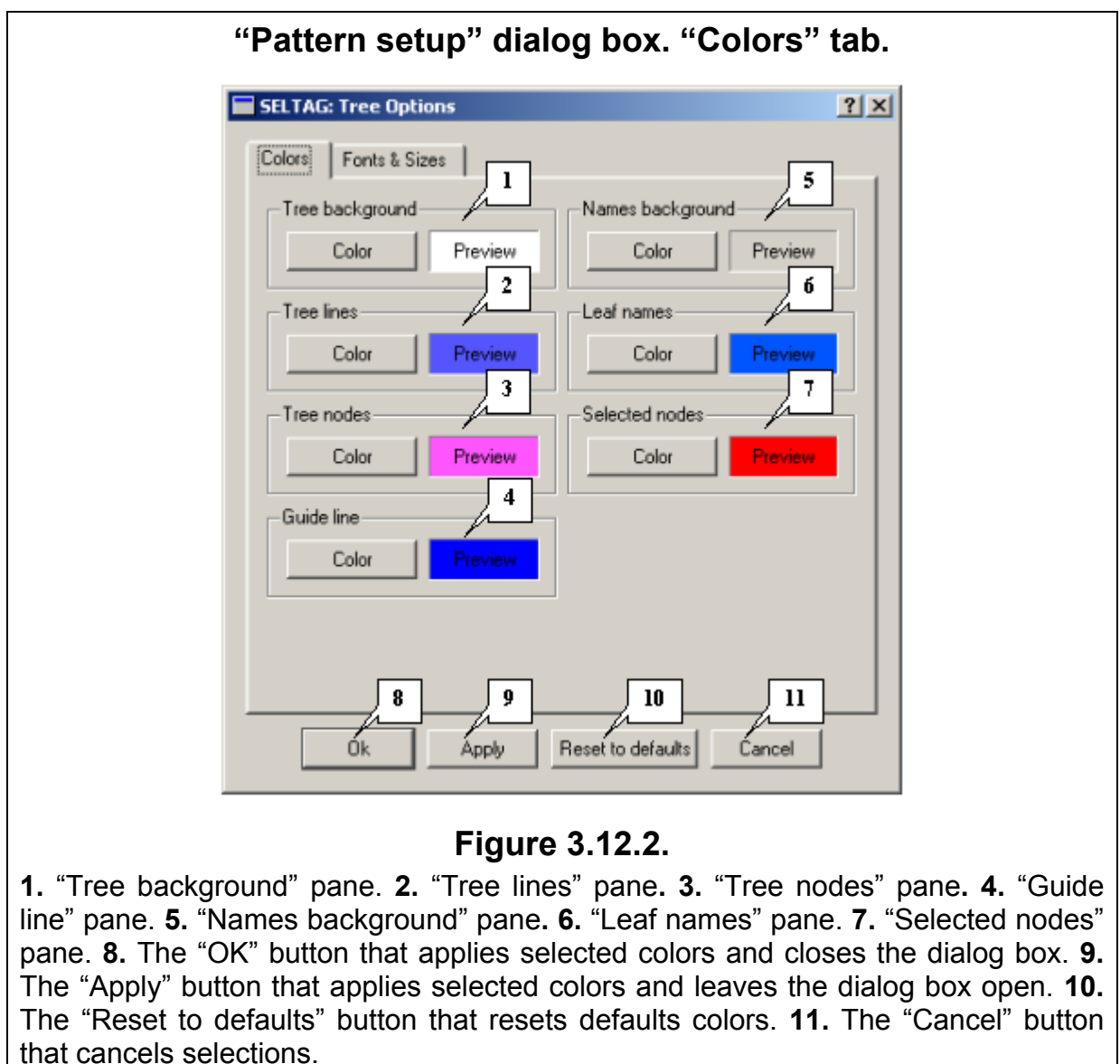
- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

The “**Leaf names**” pane allows you to select the font color for leaf names:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

The “**Selected nodes**” pane allows you to select the color of selected nodes:

- Click the “Color” button to access the “Select color” dialog box (see section 3.18, about selecting colors).
- The “Preview” pane shows the color you selected.

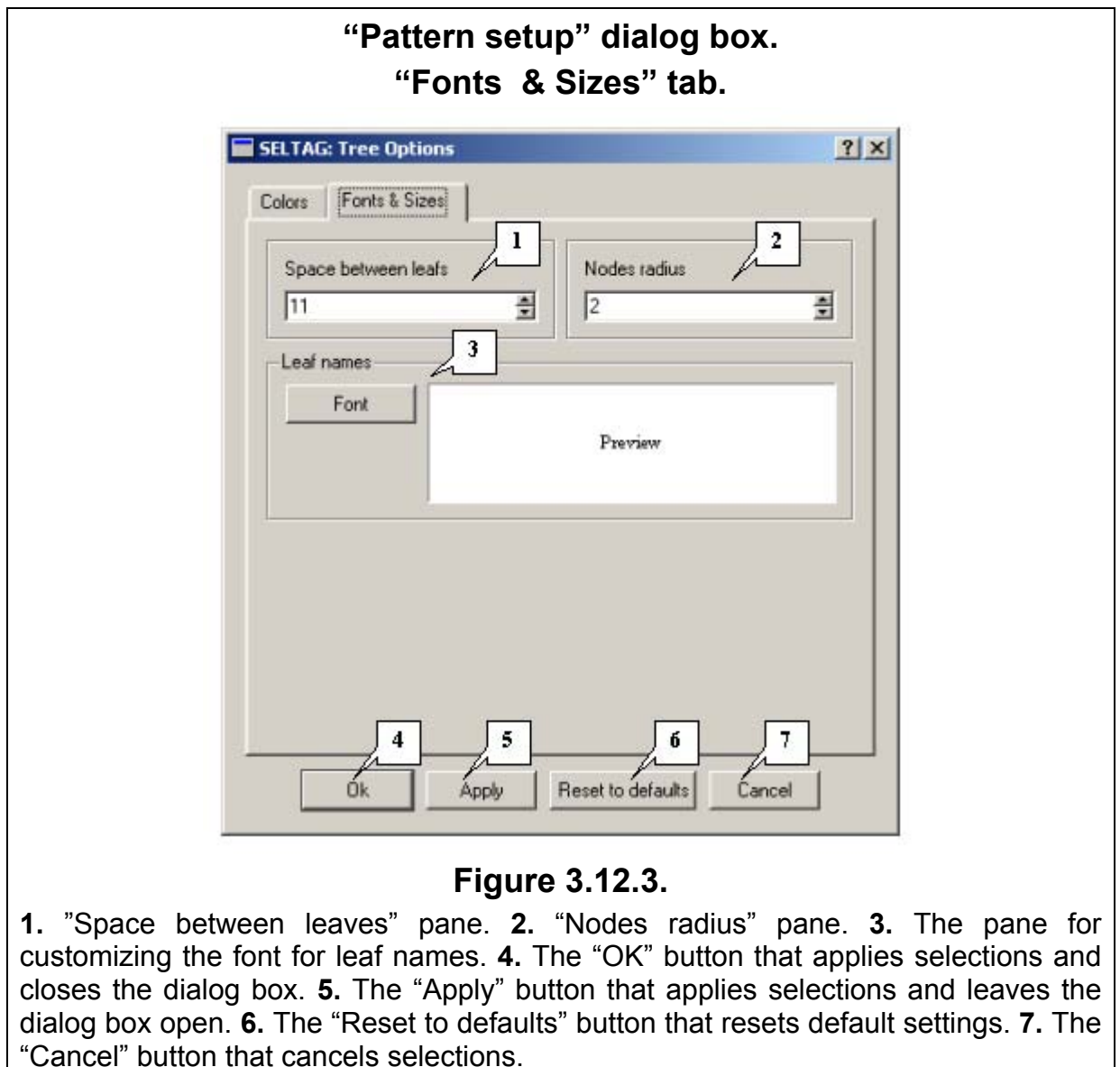


### “Font & Sizes” tab

On the “Font & Sizes” tab (Fig. 3.12.3), change the font and size of tree nodes:  
 In the “Space between leafs” pane, set the distance between leaves.  
 In the “Nodes radius” pane, change the node radius.

In the “Leaf names” pane, you can change the font for leaf names:

- Click the “Font” button to open the “Select font” dialog box (see section 3.19 about selecting fonts).
- The “Preview” pane shows the font you selected.




### Control buttons

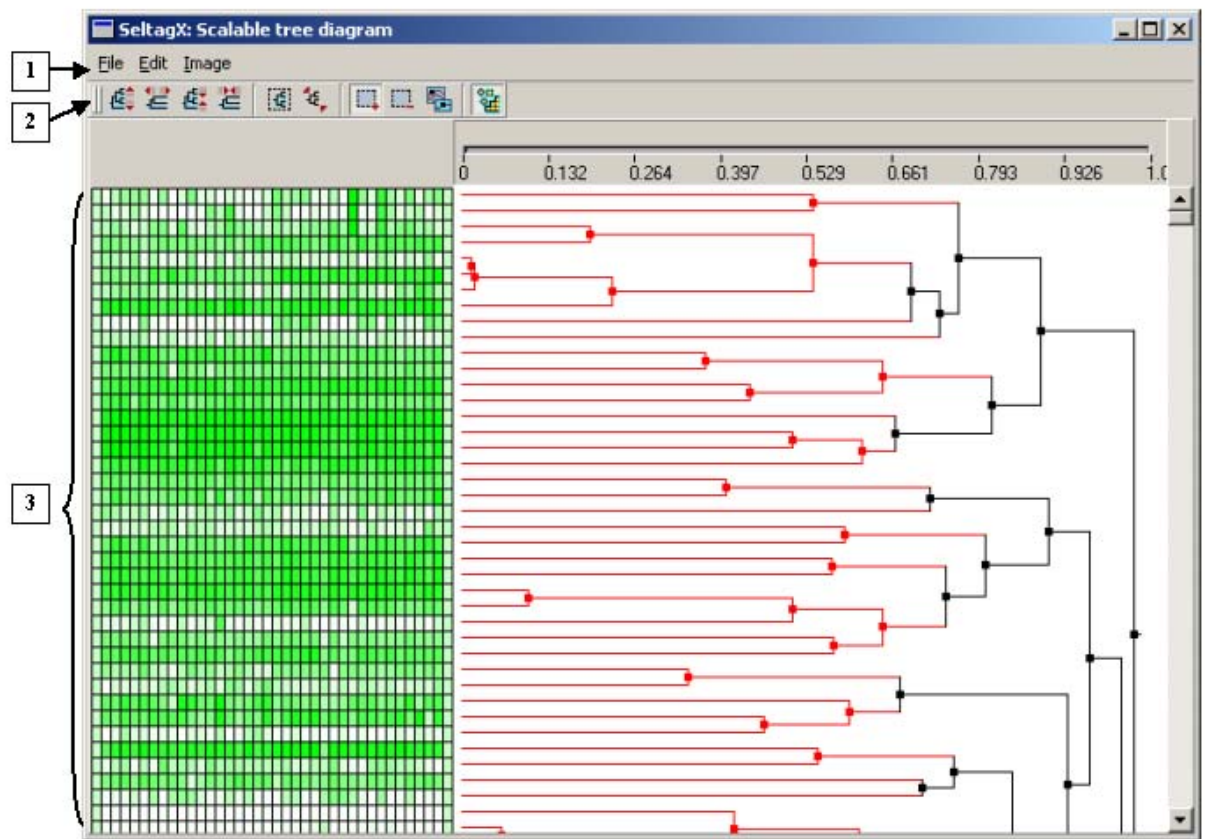
- Click “OK” to apply the current selections and quit the dialog box.
- Click “Apply” to apply the current selections and leave the dialog box open.
- Click “Reset to defaults” to restore default settings.
- Click “Cancel” to cancels selections.

#### 3.12.1. “Scalable tree diagram” dialog box

To open the “Scalable tree diagram” dialog box (Fig. 3.12.1.1):

- Select the Scalable Tree>Open Tree main menu command in the “Tree diagram” dialog box.
- Click the  toolbar button in the “Tree diagram” dialog box.

## “Scalable tree diagram” dialog box













**Figure 3.12.1.1.**

1. Main menu. 2. Toolbar. 3. View and zoom areas.

### Main menu commands

- The “File” menu has one command for managing files:
  - “Close” – closes the dialog box.
- The “Edit” menu contains several commands for magnifying or reducing the image:
  - “Zoom in vertical” – enlarge the image in vertical direction
  - “Zoom in horizontal” – enlarge the image in horizontal direction
  - “Zoom out vertical” – reduce the image in vertical direction
  - “Zoom out horizontal” – reduce the image in horizontal direction
  - “Zoom 100%” – restore the original size
  - “Zoom min” – set the minimum size
  - “Zoom in” – enlarge the selected area of the image
  - “Zoom out” – reduce the selected area of the image
  - “Go to” – select a node to be displayed on the main diagram.
- The “Image” menu contains commands for setting up modes for operating the expression matrix diagram.
  - “View expr. image” – if this checkbox is checked, a diagram of the expression matrix will be displayed.
  - «Image setup» –calls the “pattern setup” dialog box from where you can configure parameters of the expression matrix image (see section 3.16.1, about the “pattern setup” dialog box).

### Toolbar buttons:

	“Zoom in vertical” – enlarge the image in vertical direction
	“Zoom in horizontal” – enlarge the image in horizontal direction
	“Zoom out vertical” – reduce the image in vertical direction
	“Zoom out horizontal” – reduce the image in horizontal direction
	“Zoom 100%” – restore the original size
	“Zoom min” – set the minimum size
	“Zoom in” – enlarge the selected area of the image
	“Zoom out” – reduce the selected area of the image
	“Go to” – select a node to be displayed on the main diagram
	«Image creation» - add an expression matrix diagram (for a field tree this button is not available).

### 3.13. Gene clustering

To perform gene clustering, click the Analysis>Clustering>Find gene cluster menu command. The «setup for clustering procedure» wizard will guide you through the entire process of specifying the gene clustering parameters.

In the first wizard's dialog box (Fig. 3.13.1), set the following clustering parameters:

3.13.1. The “Fields...” button calls the “Field selection” dialog box (see section 3.3 about the “Field selection” dialog box). The text field displays the number of selected fields.

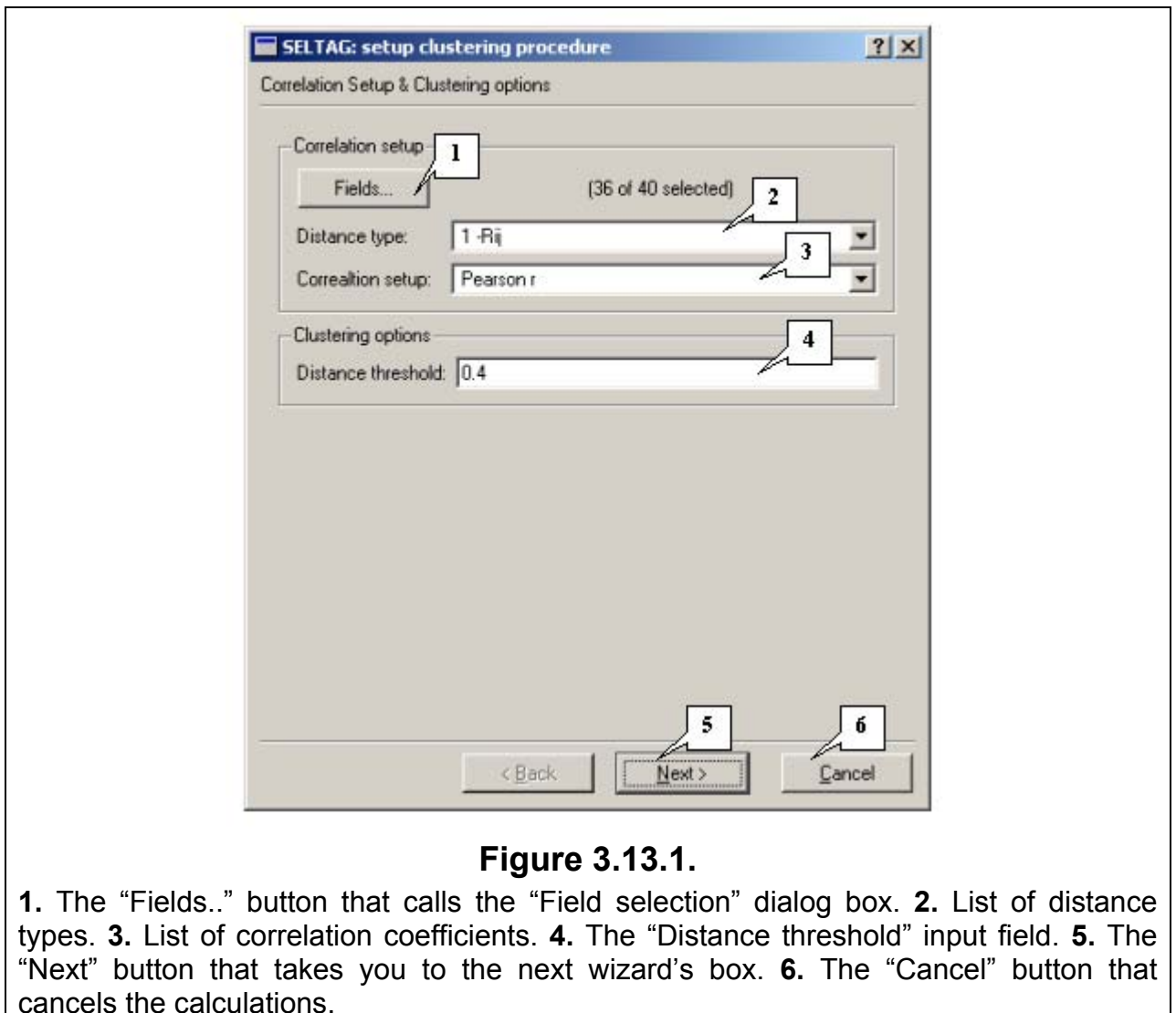
3.13.2. The “Distance type” list includes three types of distance measures calculated from the Rij correlation coefficients:

- 1-Rij
- 1+Rij
- 1-|Rij|

3.13.3. The “Correlation type” list contains the following three correlation coefficients:

- Pearson r – Pearson's correlation coefficient
- Spearman r – Spearman's correlation coefficient
- Kendall tau – Kendall's correlation coefficient.

3.13.4. The “Distance threshold” input field. Genes are fused to form a single cluster if the distance between them is smaller than the specified threshold value.



**Figure 3.13.1.**

1. The “Fields...” button that calls the “Field selection” dialog box. 2. List of distance types. 3. List of correlation coefficients. 4. The “Distance threshold” input field. 5. The “Next” button that takes you to the next wizard's box. 6. The “Cancel” button that cancels the calculations.

To proceed, click “Next”; to cancel the operation, click “Cancel”.

After you click the “Next” button, the program starts a clustering procedure. The clustering results are shown in the second wizard’s box.

In the second wizard’s box (Fig. 3.13.2), you can process clustering results:

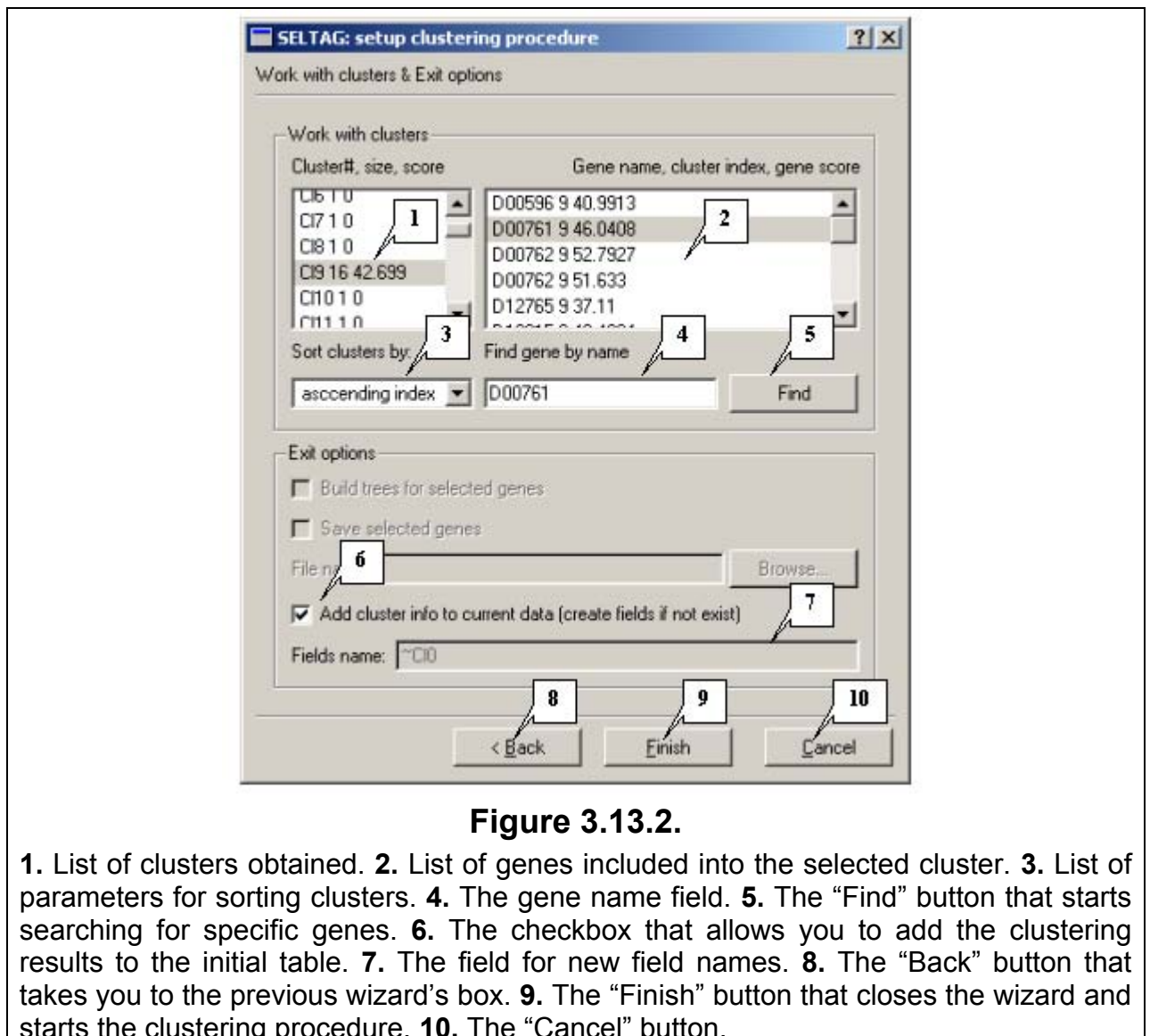
3.13.5. The “Cluster #, size, score” list displays the number, size, and score of the clusters obtained.

3.13.6. The “Gene NAME, cluster index, gene score” list includes genes that belong to the clusters selected in list 3.13.5. This list contains information about the gene name and score.

3.13.7. The “Sort clusters by:” list offers several choices of how to sort the clusters obtained:

- ascending index – sort clusters in ascending order (cluster number)
- r-score – sort clusters by score
- size – sort clusters by size

3.13.8. The “Find gene by name” field allows you to search for a gene in list 3.11.6 by its name. To search for a specific gene, enter the desired gene name and click “Find”. If the gene is in the list, the program will highlight it with another color. If the list contains no such genes, the “Not Clustered!” message will appear on your screen.



**Figure 3.13.2.**

1. List of clusters obtained. 2. List of genes included into the selected cluster. 3. List of parameters for sorting clusters. 4. The gene name field. 5. The “Find” button that starts searching for specific genes. 6. The checkbox that allows you to add the clustering results to the initial table. 7. The field for new field names. 8. The “Back” button that takes you to the previous wizard’s box. 9. The “Finish” button that closes the wizard and starts the clustering procedure. 10. The “Cancel” button.

3.13.9. The “Add cluster info for current data” checkbox. If this box is checked, after the wizard is closed, the clustering results will be recorded into the four new fields titled as follows:

- ~TXT\_n – the cluster number for each gene
- ~TXT\_size – the size of the cluster for each gene
- ~TXT\_Raver – the score of the cluster for each gene
- ~TXT\_Rcard – the gene score,

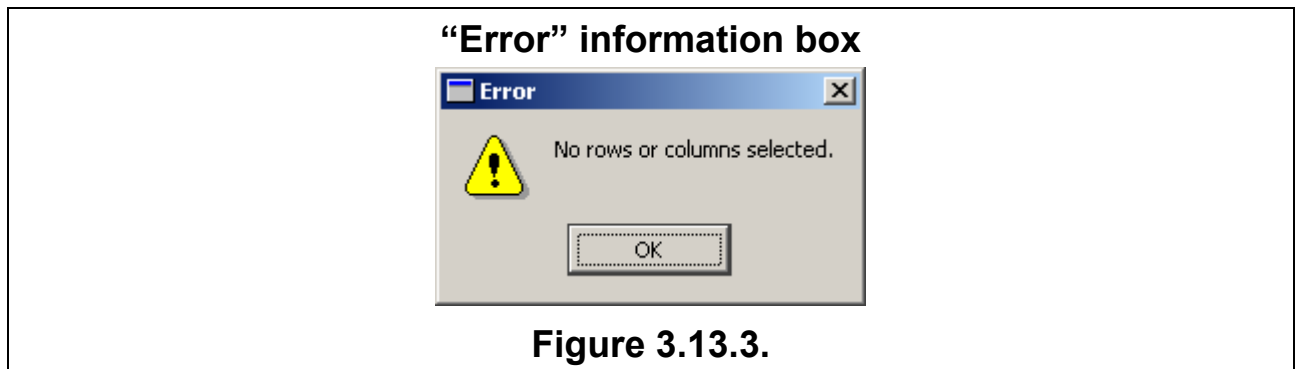
where TXT is a text string you entered into the “Fields name” field.

By default, each time you open the dialog box, TXT is set to Cl#, where # is the number of each dialog box launch. If fields with this name already exist, the program updates them; otherwise, it will create new fields.

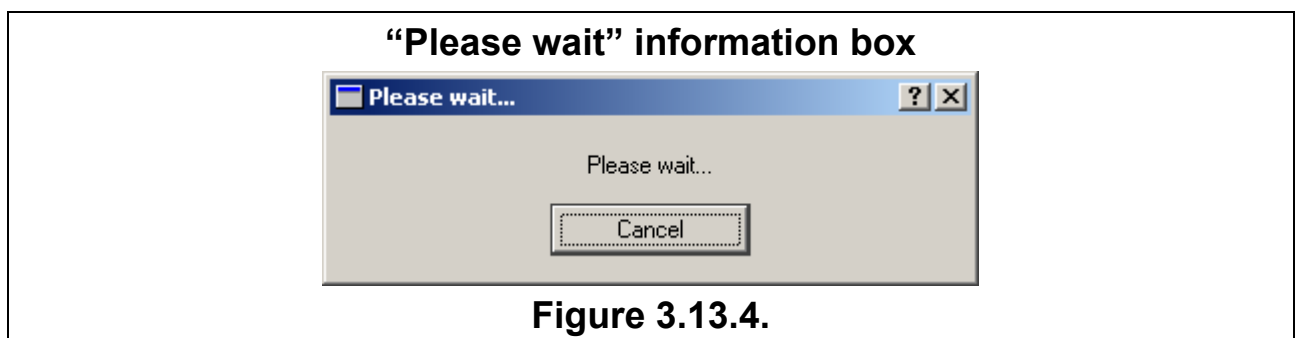
To finish with the wizard and start clustering, click “Finish”. Click “Back” to return to the previous wizard’s box and “Cancel” to cancel the operation.

**Note.**

If no fields were selected, the “Error” information box will appear on your screen (Fig. 3.13.3).



When the calculation is in progress, you will see the “Please wait” information box (Fig. 3.13.4). This information box will disappear after the calculation is complete.



**3.13.1. Clustering genes or fields by the Ben-Dor algorithm**

To cluster genes by the Ben-Dor algorithm, click Analysis>Clustering>Find gene cluster (Ben-Dor algorithm) and to cluster fields using this algorithm, click Analysis>Clustering>Find field cluster (Ben-Dor algorithm). A wizard will help you specify clustering parameters.

For both genes and fields, the clustering setup procedure is the same.



The first wizard's box (Fig. 3.13.1.1) allows you to set the following clustering parameters:

### Specifying clustering parameters by the Ben-Dor algorithm

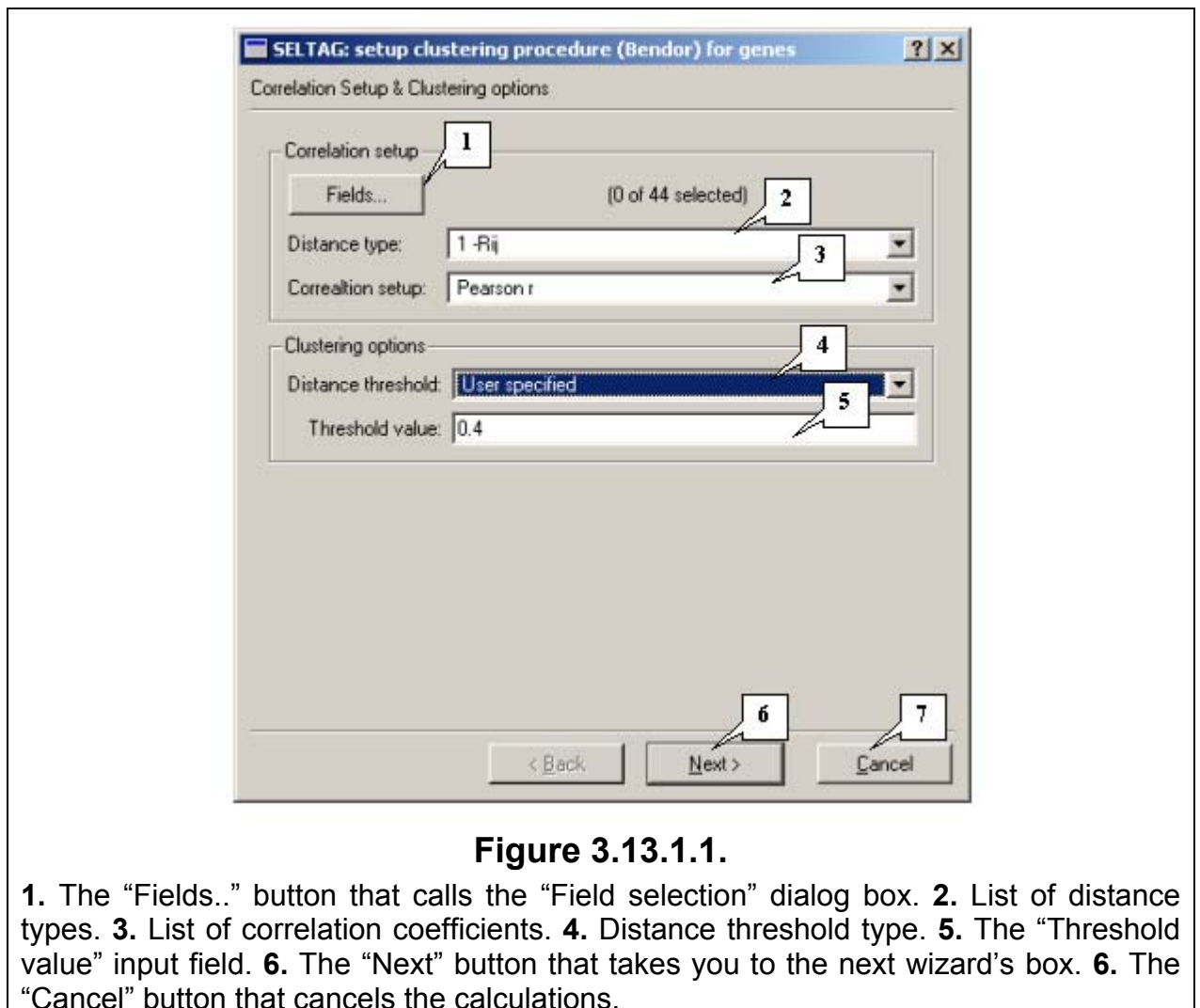
3.13.1. The "Fields..." button calls the "Field selection" dialog box (see section 3.3 about the "Field selection" dialog box). The text field displays the number of selected fields.

3.13.2. The "Distance type" list contains the following seven types of distances:

- 1-Rij
- 1+Rij
- 1-|Rij|

**Note.** The 1-Rij, 1+Rij, and 1-|Rij| distances are calculated by using one of the correlation coefficients (Pearson, Spearman, or Kendall).

- Squared Euclidian – a squared Euclidian distance.
- Euclidian – Euclidian distance
- Manhattan – the distance is calculated as  $\sum_{i=1}^n |x_i - y_i|$
- Chebyshev – the distance is calculated as  $\max_i |x_i - y_i|$



**Figure 3.13.1.1.**

1. The "Fields..." button that calls the "Field selection" dialog box. 2. List of distance types. 3. List of correlation coefficients. 4. Distance threshold type. 5. The "Threshold value" input field. 6. The "Next" button that takes you to the next wizard's box. 6. The "Cancel" button that cancels the calculations.

3.13.3. The "Correlation setup" list contains three types of correlation coefficients:

- Pearson r – Pearson's correlation coefficient



- Spearman r – Spearman’s correlation coefficient
- Kendall tau – Kendall’s correlation coefficient.

3.11.4. The “Distance threshold” drop-down list contains the following options:

- Average – automatically determine the threshold value
- User specified – use user-defined threshold. If this option is selected, the “Threshold value” field becomes active. In this field, enter the clustering threshold value.

To proceed with the wizard, click “Next”; to cancel the operation, click “Cancel”.

The second wizard’s box has the same fields as described in section 3.13.

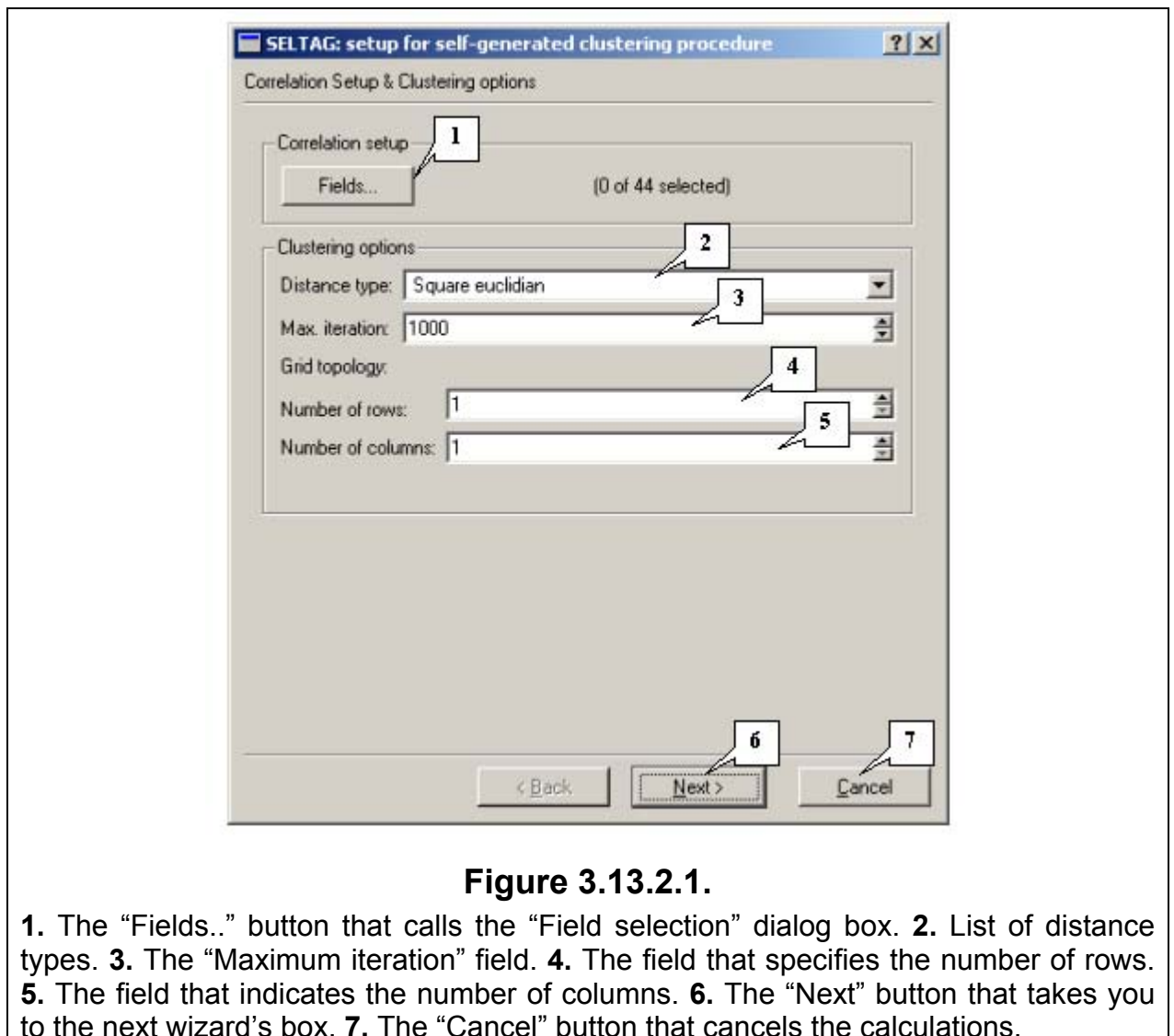
### 3.13.2. Clustering by the SOM algorithm

To cluster genes by the SOM (Self-Organizing Map) algorithm, click Analysis>Clustering>Find gene cluster (SOM algorithm) and to cluster fields using this algorithm, click Analysis>Clustering>Find field cluster (SOM algorithm). A wizard will help you specify clustering parameters.

For both genes and fields, the clustering setup procedure is the same.

The first wizard’s box (Fig. 3.13.2.1) allows you to set the following clustering parameters:

3.11.1. The “Fields...” button calls the “Field selection” dialog box (see section 3.3 about the “Field selection” dialog box). The text field displays the number of selected fields.



**Figure 3.13.2.1.**

1. The “Fields..” button that calls the “Field selection” dialog box.
2. List of distance types.
3. The “Maximum iteration” field.
4. The field that specifies the number of rows.
5. The field that indicates the number of columns.
6. The “Next” button that takes you to the next wizard’s box.
7. The “Cancel” button that cancels the calculations.

3.11.2. The “Distance type” list contains the following four types of distance measures:

- Squared Euclidian – squared Euclidian distance
- Euclidian – Euclidian distance
- Manhattan – the distance is calculated as  $\sum_{i=1}^n |x_i - y_i|$
- Chebyshev – the distance is calculated as  $\max_i |x_i - y_i|$

3.11.3. In the “Max iterations” field, specify the maximum number of iterations. The recommended value is 3000–4000.

3.11.4. In the “Grid topology” pane, specify the following grid options:

- In the “Number of rows” field, enter the number of rows.
- In the “Number of columns” field, specify the number of columns.

To proceed, click “Next”; to cancel the operation, click “Cancel”.

The second wizard’s box has the same fields as described in section 3.13.

### 3.14. Principal component analysis

To perform principal component analysis, click the Analysis>Principal component menu command. You can set the analysis parameters by using the "setup for principal component analysis" wizard.

The wizard allows you to:

- Specify the parameters for the decomposition of the covariance or correlation matrices into eigenvalues and eigenvectors.
- Make plots of eigenvalues and the coefficients of eigenvectors
- Project the values of gene expression onto plains constructed of various pairs of eigenvectors.

For the selected fields, the program calculates the covariance/correlation matrix, its eigenvalues and eigenvectors.

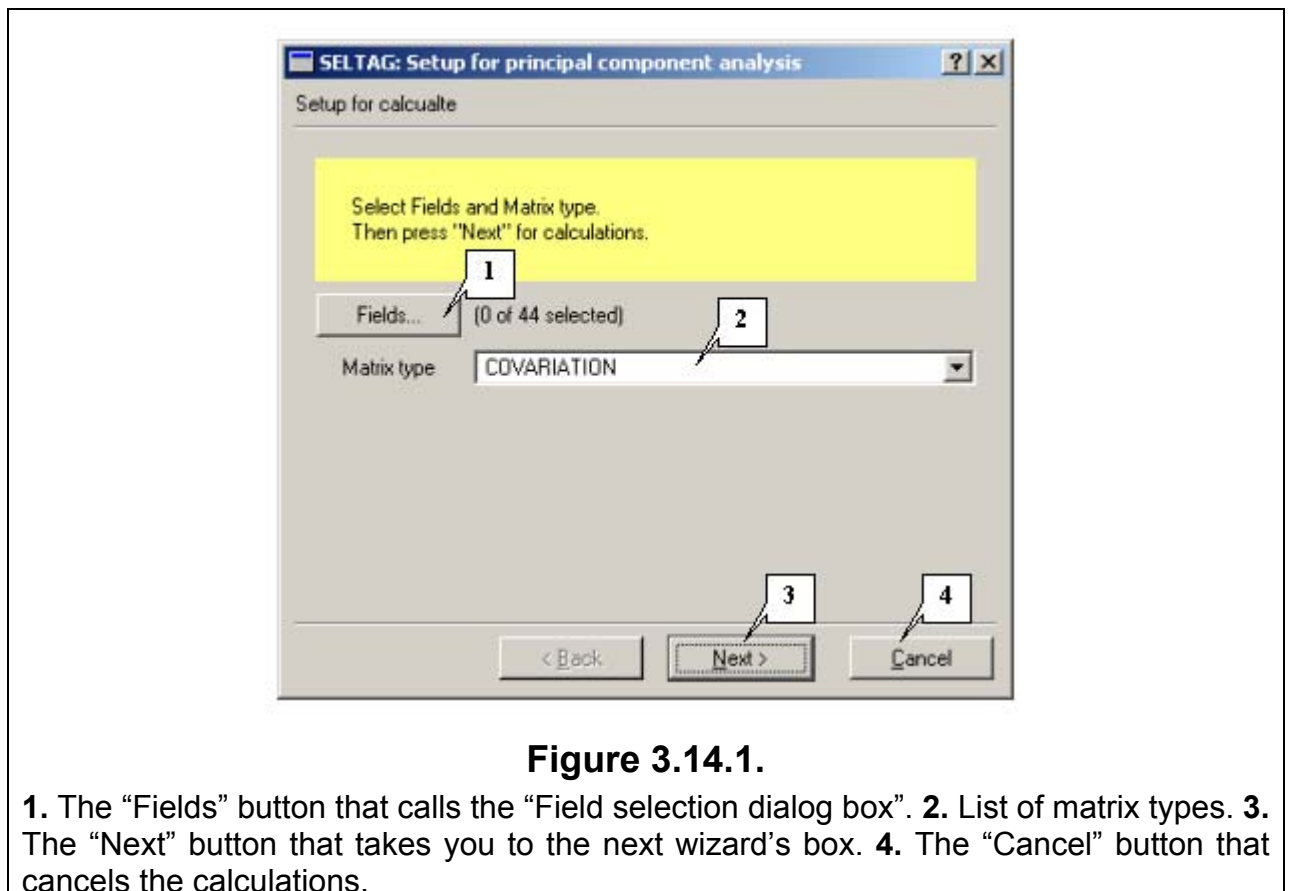
The first wizard's box (Fig. 3.14.1) has the following elements:

3.14.1. The "Fields..." button calls the "Field selection" dialog box. The text field displays the number of selected fields.

3.14.2. The "Matrix type" list that contains the following matrix types:

- Covariance matrix
- Correlation matrix

To proceed with the wizard, click "Next"; to cancel the operation, click "Cancel".



After the computation is complete, the order numbers of the components ranging from 1 to the number of selected fields appear in the "Component plots", "Component X", and "Component Y" lists.

The second wizard's box (Fig. 3.14.2) has the following elements:

3.14.3. The "Component plots" list that contains components (eigenvectors) numbered in descending order of their eigenvalues.

3.14.4. Click "Mark all" to select all components in list 3.14.3.

3.14.5. Click "Unmark all" to unselect all components in list 3.14.4.

3.14.6. The pane that displays the sum of the eigenvalues of the components selected in the "Component plots" list.

- The "Variance" text field that displays the sum of the eigenvalues of the components selected in the "Component plots" list.

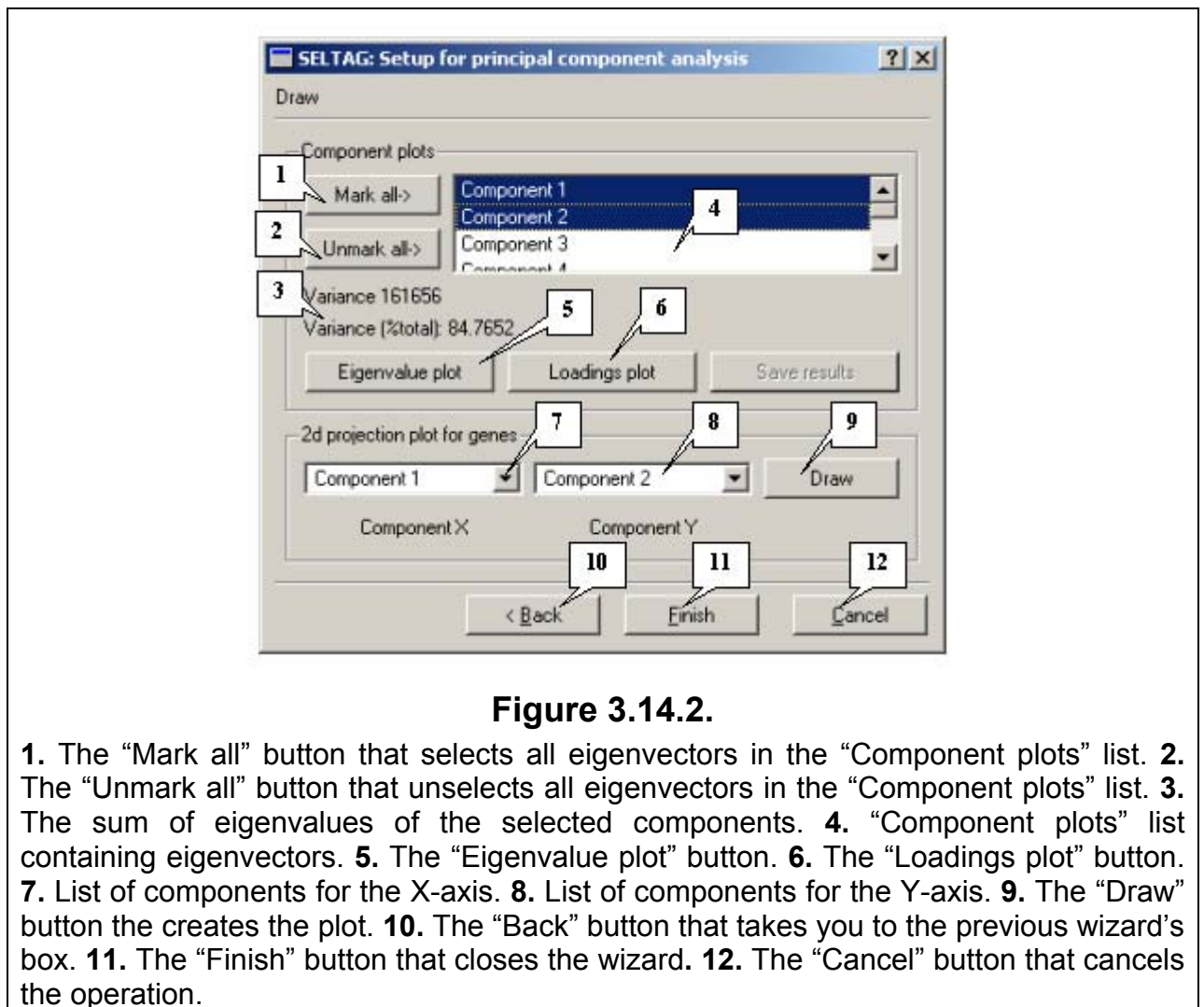
- The "Variance (%total)" text field that displays the sum (in %) of the eigenvalues of the components selected in the "Component plots" list.

3.14.7. The "Eigenvalue plot" calls the "Graph" dialog box, which displays the values of the selected components (see section 3.15 about the "Graph" dialog box).

3.14.8. The "Loadings plot" button calls the "Graph" dialog box, which displays the values of groups for each component (see section 3.15 about the "Graph" dialog box).

### Making projections

3.14.9. "Component X" contains components (eigenvectors) numbered in descending order of the eigenvalues. You can select components for the X-axis from this list.



**Figure 3.14.2.**

1. The "Mark all" button that selects all eigenvectors in the "Component plots" list. 2. The "Unmark all" button that unselects all eigenvectors in the "Component plots" list. 3. The sum of eigenvalues of the selected components. 4. "Component plots" list containing eigenvectors. 5. The "Eigenvalue plot" button. 6. The "Loadings plot" button. 7. List of components for the X-axis. 8. List of components for the Y-axis. 9. The "Draw" button the creates the plot. 10. The "Back" button that takes you to the previous wizard's box. 11. The "Finish" button that closes the wizard. 12. The "Cancel" button that cancels the operation.

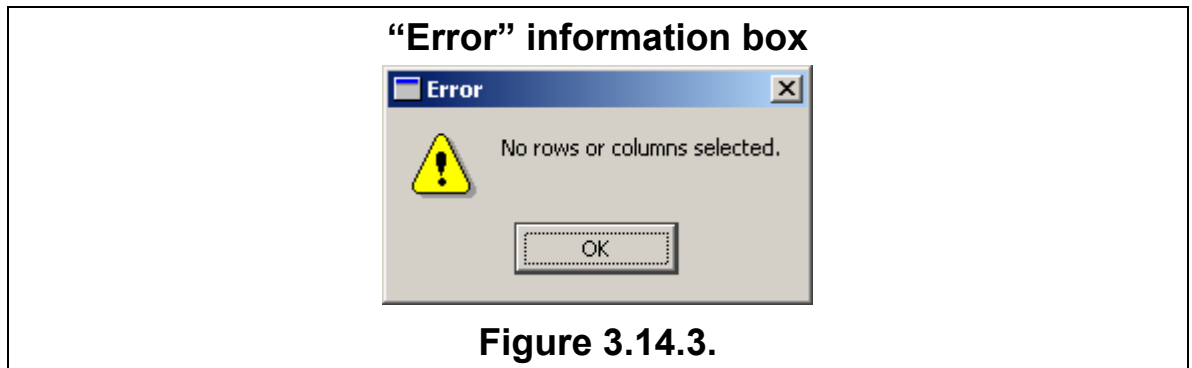
3.14.10. “Component Y” contains components (eigenvectors) numbered in descending order of the eigenvalues. You can select components for the Y-axis from this list.

3.14.11. The “Draw” button makes a plot and calls the “Graph” dialog box (see section 3.15 about the “Graph” dialog box).

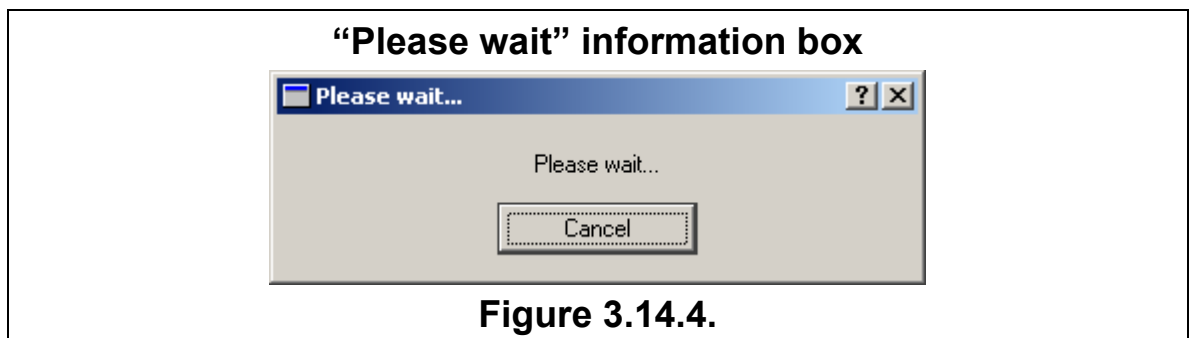
To finish with the wizard and perform computations, click “Finish”. To move to the previous wizard’s box, click “Back”; to cancel the operation, click “Cancel”.

**Note.**

If the sets of fields were not formed, the “Error” information box will appear on your screen (Fig. 3.14.3).







When clustering is in progress, you will see the “Please wait” information box (Fig. 3.14.4). This information box will disappear after the calculation is complete.



### 3.15. “Graph” dialog box

The “Graph” dialog box (Fig. 3.15.1) visualizes the distribution of genes in the space between two main components.

#### Toolbar buttons and fields:

	Calls the “PCA” setup dialog box
	Undo (restore the previous scale of the image)
	Restore the original (100%) scale of the image
	Redo (restore the undone scale of the image).

#### Graph

The graph visualizes the distribution of genes in the space between two main components.

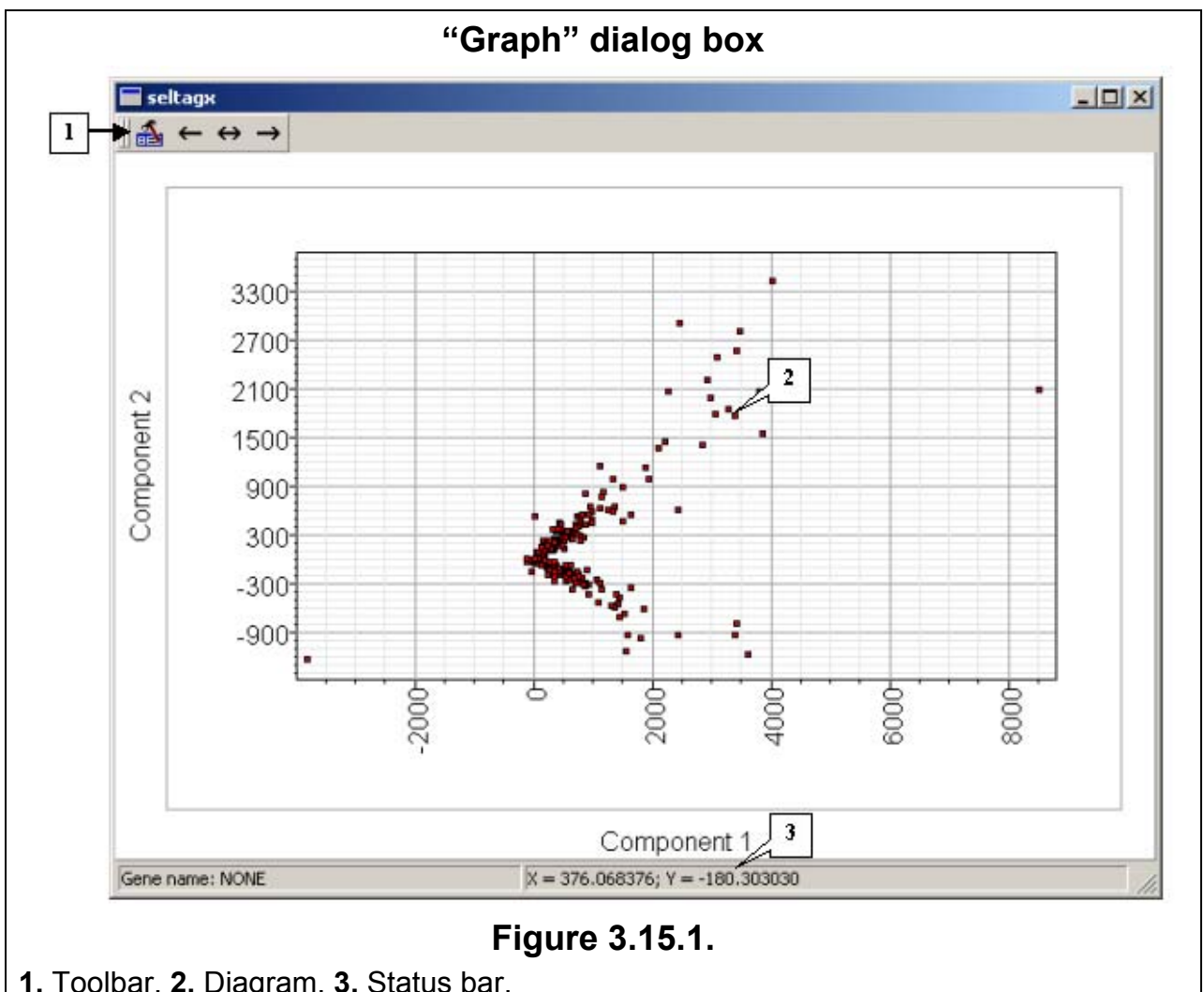
To enlarge the scale of the image, select a rectangle with your left mouse button.

If you select a rectangle with your right mouse button, this will open the “Operations with marked items” dialog box (about this dialog box, see section 3.15.1).

#### Status bar

The status bar displays:

- The name of the gene you are pointing to with your mouse and the current coordinates of the mouse pointer.



### 3.15.1. “Operations with marked items” dialog box

The “Operations with marked items” dialog box (Fig. 3.15.1.1) contains a list of gene names within the selected area and the following control buttons:

**<-Select all** – selects all genes in the list.

**<-Unselect all** – unselects all genes in the list.

**<-Invert** – inverts selection of all genes in the list.

**Color** – calls the “Color dialog” dialog box from where you can customize colors for the graph and the genes selected in the list.

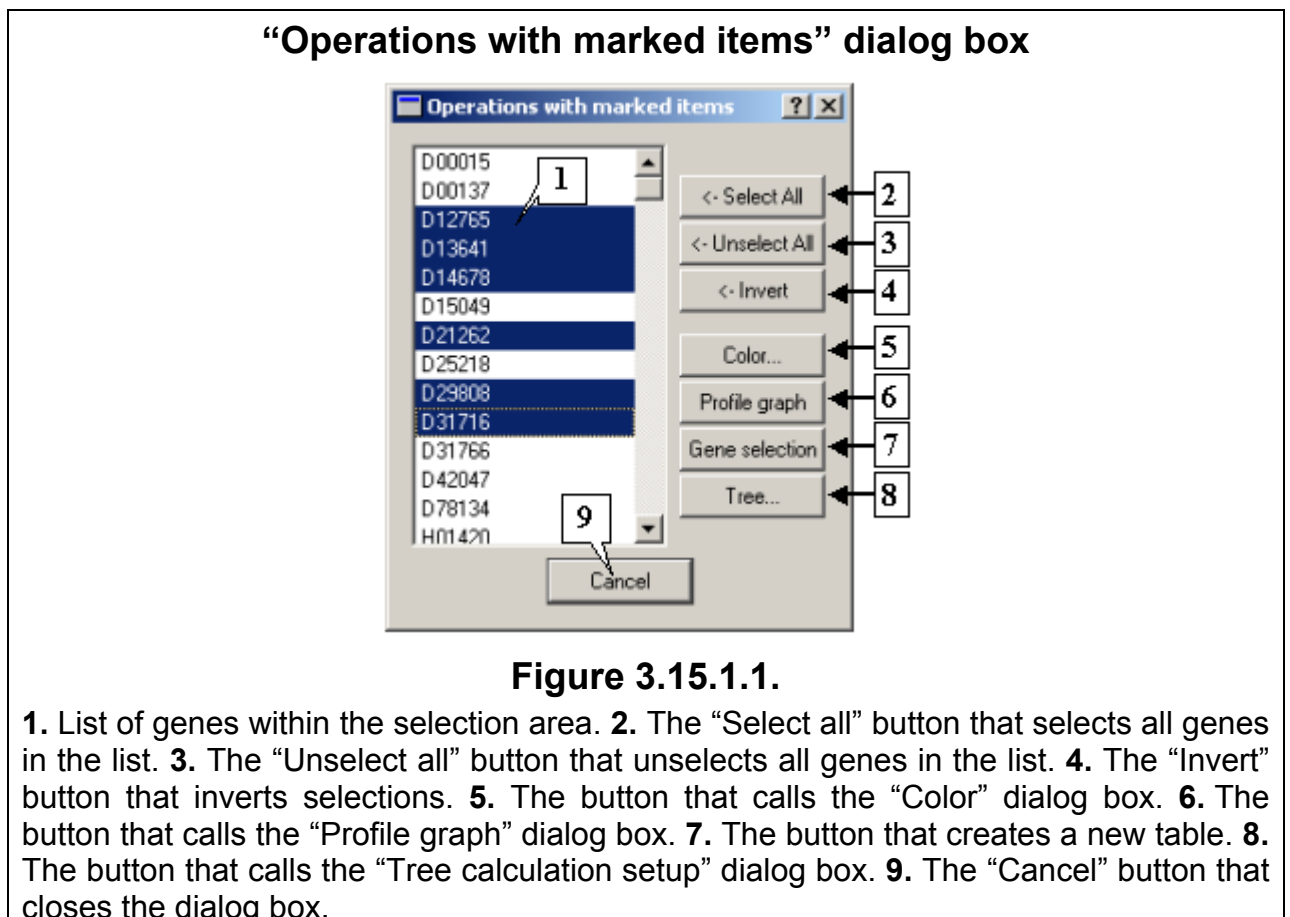
**Profile Graph** – calls the “Profile graph” dialog box that visualizes the expression profiles of genes selected in the list (profiles are drawn by fields, which were taken into account in the PCA computations).

**Gene Selection** – creates a table containing the genes selected in the list.

**Tree** – calls the “Tree calculation setup” dialog box from where you can build a tree of the genes selected in the list.

**Cancel** – quits the dialog box.


When you click the “Profile Graph”, “Gene Selection”, and “Tree” buttons, the program will create new tables containing genes that were subjected to these computations.



**Figure 3.15.1.1.**

1. List of genes within the selection area. 2. The “Select all” button that selects all genes in the list. 3. The “Unselect all” button that unselects all genes in the list. 4. The “Invert” button that inverts selections. 5. The button that calls the “Color” dialog box. 6. The button that calls the “Profile graph” dialog box. 7. The button that creates a new table. 8. The button that calls the “Tree calculation setup” dialog box. 9. The “Cancel” button that closes the dialog box.

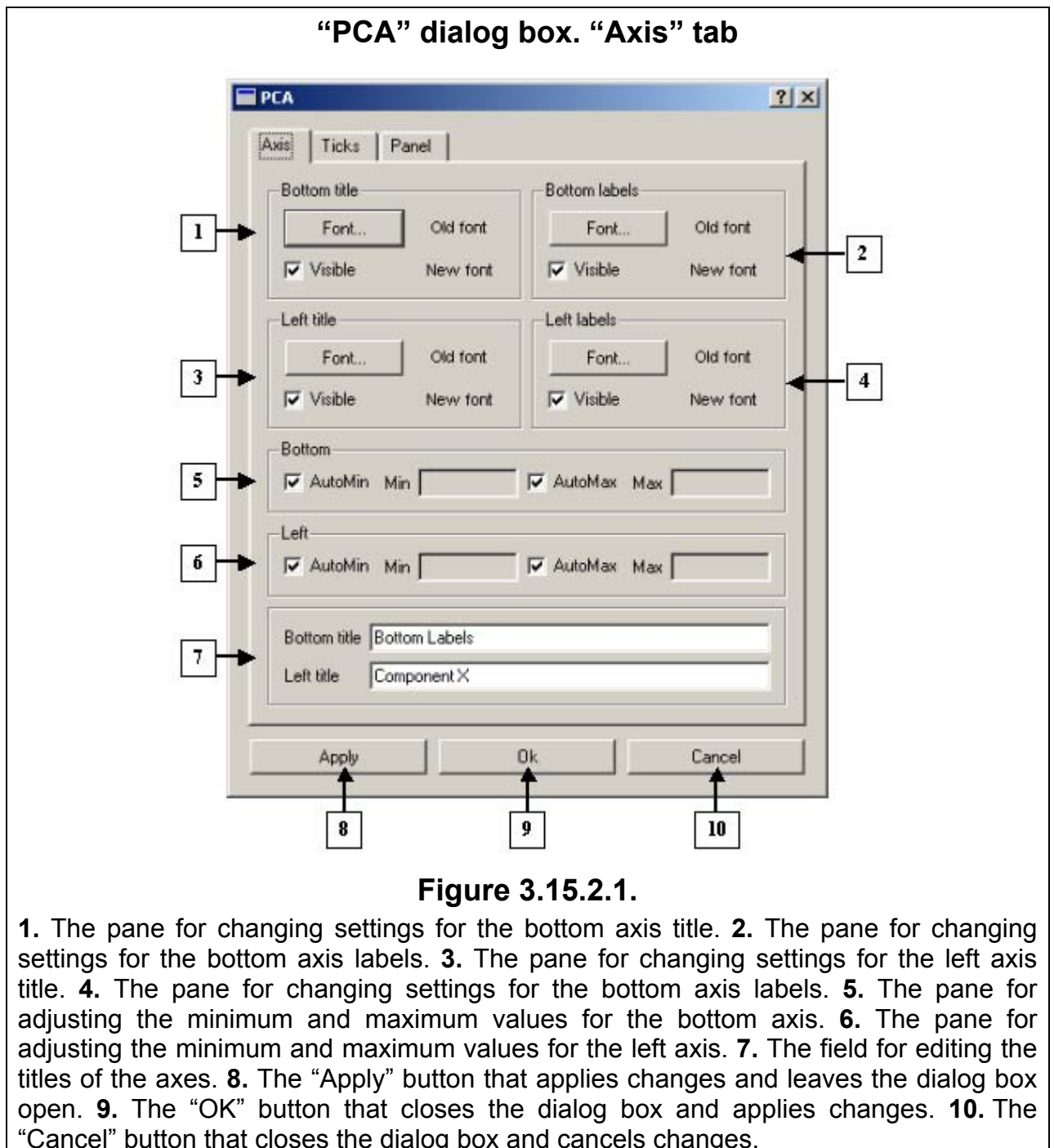
### 3.15.2. “PCA” dialog box

In the “PCA” dialog box, you can change graphics settings for the “Graph” dialog box. To access the “PCA” dialog box, click the  toolbar button. The “PCA” dialog box has three tabs: “Axis”, “Ticks”, and “Panel”.

### “Axis” tab

On the “Axis” tab (Fig. 3.15.2.1), configure the graphics options for axes.

- In the “Bottom title” pane, set the parameters for the bottom axis title:
  - Click the “Font” button to open the “Select font” dialog box (see section 3.19) and select a font for the title.
  - The “Old font” pane shows old font settings.
  - The “New font” pane displays new font settings.
  - Check/uncheck the “Visible” checkbox to display/hide the axis name.



- In the “Bottom labels” pane, set the parameters for bottom axis labels:
  - Click the “Font” button to open the “Select font” dialog box (see section 3.19) and select a font.



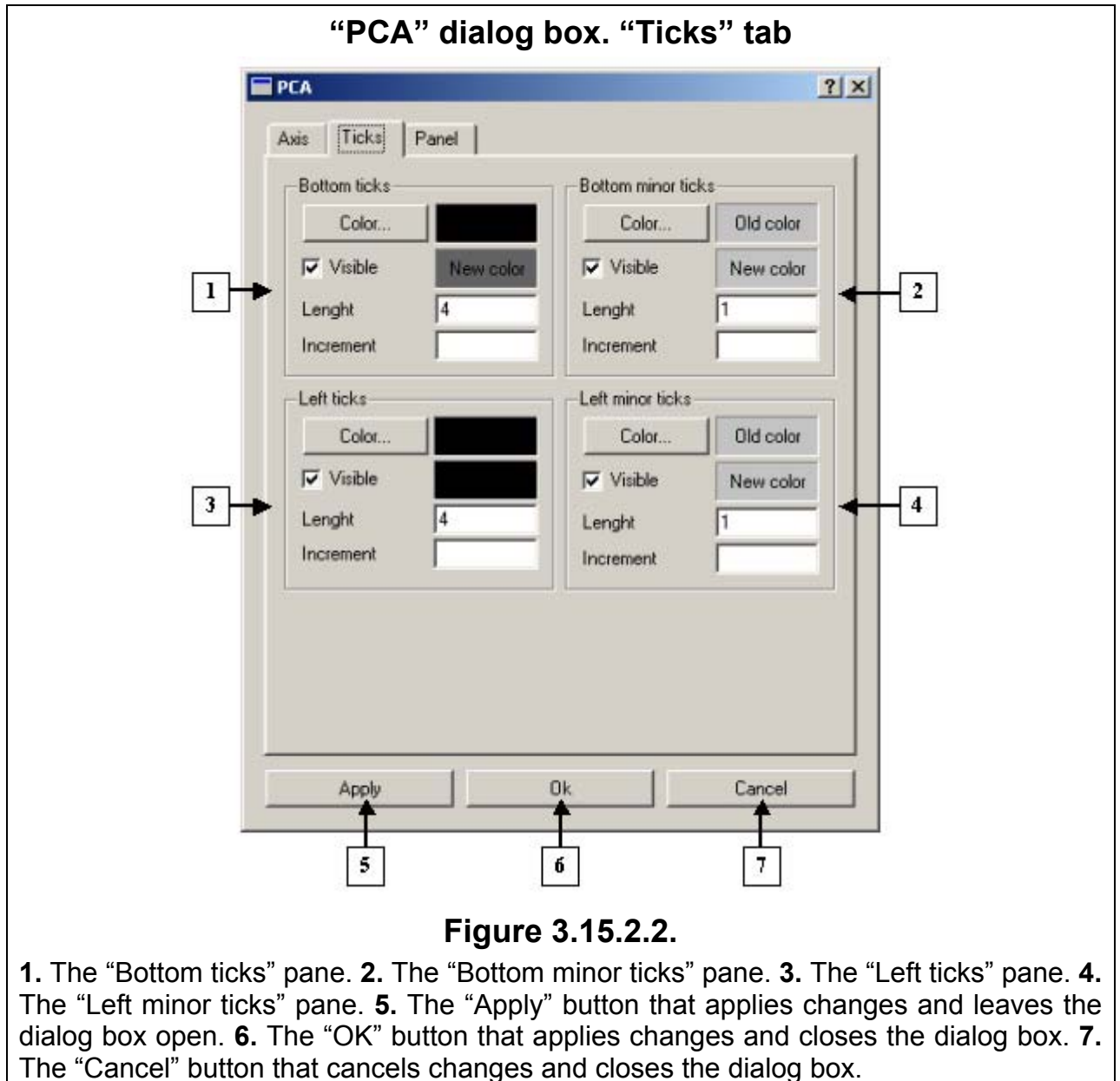
- The “Old font” pane shows old font settings.
- The “New font” pane displays new font settings.
- Check/uncheck the “Visible” checkbox to display/hide the axis labels.
- In the “Left title” pane, configure settings for the left axis title:
  - Click the “Font” button to open the “Select font” dialog box (see section 3.19) and select a font.
  - The “Old font” pane shows old font settings.
  - The “New font” pane displays new font settings.
  - Check/uncheck the “Visible” checkbox to display/hide the axis title.
- In the “Left labels” pane, configure settings for the left axis labels:
  - Click the “Font” button to open the “Select font” dialog box (see section 3.19) and select a font.
  - The “Old font” pane shows old font settings.
  - The “New font” pane displays new font settings.
  - Check/uncheck the “Visible” checkbox to display/hide the axis labels.
- In the “Left” pane, set the minimum and maximum values for the left axis:
  - The “AutoMin” checkbox. If the checkbox is checked, the minimum value will be set automatically. If the checkbox is unchecked, you can type the desired minimum value in the “Min” field.
  - The “AutoMax” checkbox. If the checkbox is checked, the maximum value will be set automatically. If the checkbox is unchecked, you can type the desired maximum value in the “Max” field.
- In the “Bottom” pane, set the minimum and maximum values for the bottom axis:
  - The “AutoMin” checkbox. If the checkbox is checked, the minimum value will be set automatically. If the checkbox is unchecked, you can type the desired minimum value in the “Min” field.
  - The “AutoMax” checkbox. If the checkbox is checked, the maximum value will be set automatically. If the checkbox is unchecked, you can type the desired maximum value in the “Max” field.
- In the pane below, you can edit the titles of the axes:
  - In the “Bottom title” field, edit the title of the bottom axis.
  - In the “Left title” field, edit the title of the left axis.

### **“Ticks” tab**

On the “Ticks” tab (Fig. 3.15.2.2), you can customize graphics options for axis ticks.

- In the “Bottom ticks” pane, customize the major ticks on the bottom axis:
  - The “Color” button calls the “Select color” dialog box (see section 3.18), from where you can change color for the ticks.
  - The “Old color” pane shows the old color of ticks.
  - The “New color” pane shows the new color of ticks.
  - Check/uncheck the “Visible” checkbox to show/hide ticks.
  - The “Length” field specifies the space between ticks.
  - The “Increm.” field specifies the increment.
- In the “Bottom minor ticks” pane, customize the minor ticks on the bottom axis:
  - The “Color” button calls the “Select color” dialog box (see section 3.18), from where you can change color for the ticks.
  - The “Old color” pane shows the old color of ticks.

- The “New color” pane shows the new color of ticks.
- Check/uncheck the “Visible” checkbox to show/hide ticks.
- The “Length” field specifies the space between ticks.
- The “Count.” field specifies the number of minor ticks between the major ones.



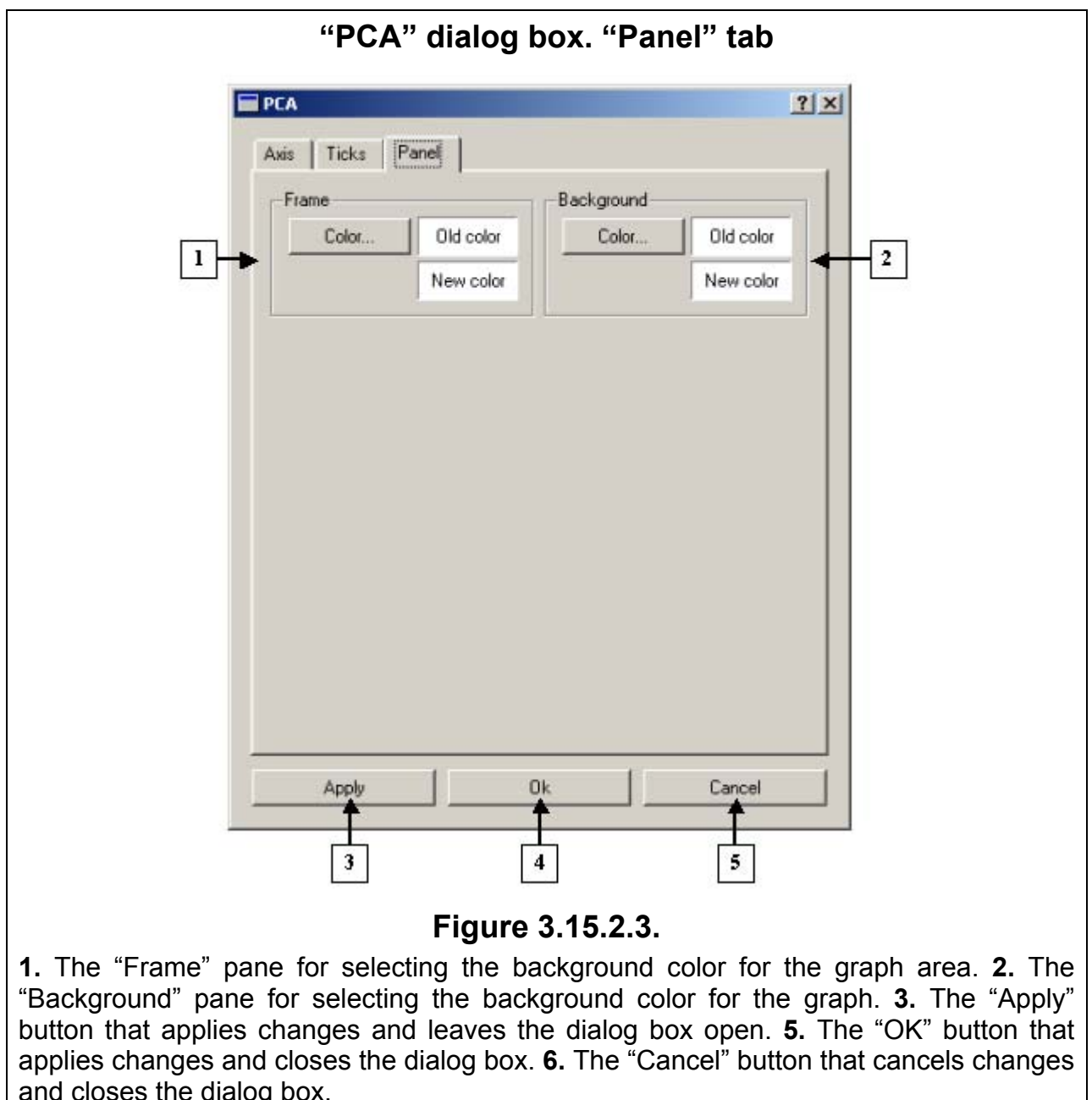
- In the “Left ticks” pane, customize the major ticks on the left axis:
  - The “Color” button calls the “Select color” dialog box (see section 3.18), from where you can change color for the ticks.
  - The “Old color” pane shows the old color of ticks.
  - The “New color” pane shows the new color of ticks.
  - Check/uncheck the “Visible” checkbox to show/hide ticks.
  - The “Length” field specifies the space between ticks.
  - The “Incrim.” field specifies the increment.
- In the “Left minor ticks” pane, customize the minor ticks on the left axis:
  - The “Color” button calls the “Select color” dialog box (see section 3.18), from where you can change color for the ticks.

- The “Old color” pane shows the old color of ticks.
- The “New color” pane shows the new color of ticks.
- Check/uncheck the “Visible” checkbox to show/hide ticks.
- The “Length” field specifies the space between ticks.
- The “Count.” field specifies the number of minor ticks between the major ones.

#### “Panel” tab

On the “**Panel**” tab (Fig. 3.15.2.3), you can select the background color for the graph and graph area.

- In the “Frame” pane, you can select the background color for the graph area:
  - The “Color...” button calls the “Select color” dialog box (see section 3.18), from where you can change the background color.
  - The “Old color” pane shows the old color.
  - The “New color” pane shows the new color.



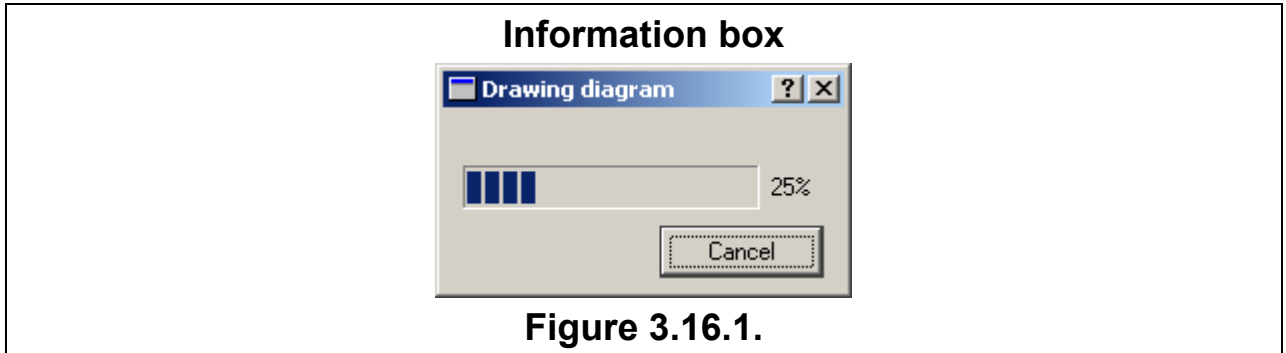
- In the “Background” pane, you can select the background color for the graph:
  - The “Color” button calls the “Select color” dialog box (see section 3.18), from where you can change the background color.
  - The “Old color” pane shows the old color.
  - The “New color” pane shows the new color.

**Control buttons**

- **Apply** – applies changes and leaves the dialog box open
- **OK** – applies changes and closes the dialog box
- **Cancel** - cancels changes and closes the dialog box.

### 3.16. Expression map

The “Expression map” dialog box displays the gene expression matrix. This dialog box is accessed by clicking the Analysis>Expression map main menu command. When the map is being loaded, you can see the “Drawing diagram” information box (Fig 3.16.1) that will disappear after the loading is complete.



**Figure 3.16.1.**

The “Expression map” dialog box has the following elements (Fig. 3.16.2):

- Main (system) menu
- Toolbar
- A pane with column names
- A pane with row names
- Matrix view area
- Status bar


The **main menu** of this dialog box has the following menus:


- The “File” menu has one command:
  - “Close” – closes the dialog box.
- The “Image” menu has two commands:
  - “Options” – calls the “Image options” dialog box (see section 3.16.2), from where you can select image options for this dialog box.
  - “Pattern” – calls the “Pattern setup” dialog box (see section 3.16.1), from where you can select graphics options for the expression matrix.


The **Toolbar** has the following buttons:

 (“Column name”) shows the Column names pane.

 (“Row name”) shows the Row names pane.

 (“Fit mode”) resizes the image to fit as closely as possible within the current view area frame.

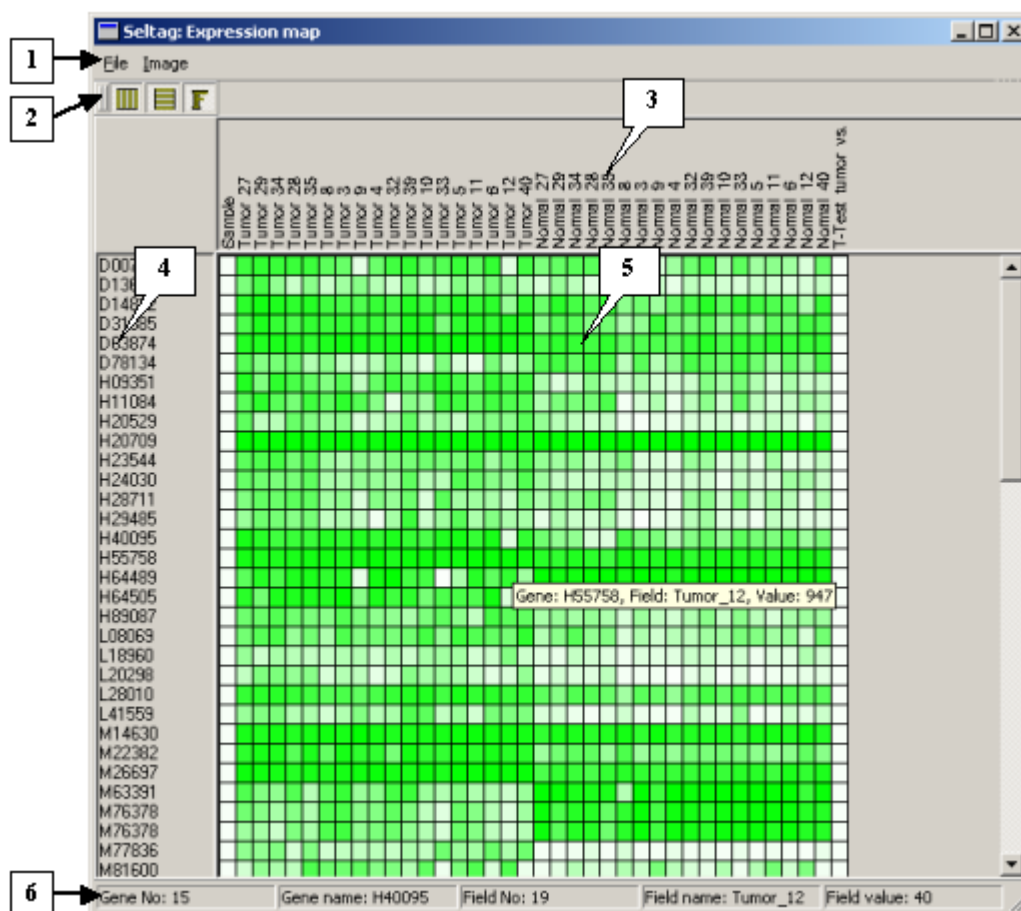
The **Column names pane** shows the names of matrix columns. To display the pane, click .

The **Rows names pane** shows the names of matrix rows. To display the pane, click .

The **View area** displays the expression matrix map.

The **Status bar** displays information about the expression matrix cell you are pointing to with your mouse pointer: gene number, gene name, field number, field name, and the field value.

## “Expression map” dialog box



**Figure 3.16.2.**

1. Main (system) menu. 2. Toolbar. 3. A pane with column names 4. A pane with row names. 5. Matrix view area. 6. Status bar.

### 3.16.1. “Pattern setup” dialog box

3.16.1. In the “Pattern setup” dialog box (Fig. 3.16.1.1), you can set up the parameters of the expression matrix image. To open this dialog box, click the Image>Image setup command. The “Pattern setup” dialog box has the following elements:

3.16.1.1. The “Palette type” list that includes the following options:

3.16.1.1.1 “White(min)->Colormax” –the color of matrix cells changes from white to the most saturated color defined in pane 3.16.1.4.

3.16.1.1.2. “Black(min)->Colormax” –the color of matrix cells changes from black to the most saturated color defined in pane 3.16.1.4.

3.16.1.1.3. “Colormin->White->Colormax” –the color of matrix cells changes from the least saturated color defined in pane 3.16.1.5 to the most saturated color defined in pane 3.16.1.4. The intermediate color is white.

3.16.1.1.4 “Colormin->Black->Colormax”– the color of matrix cells changes from the least saturated color defined in pane 3.16.1.5 to the most saturated color defined in pane 3.16.1.4. The intermediate color is black.

3.16.1.1.5. “Geographic map colors” - the pallet of geographic map colors.

3.16.1.2. The “Range type” list that has the following options:

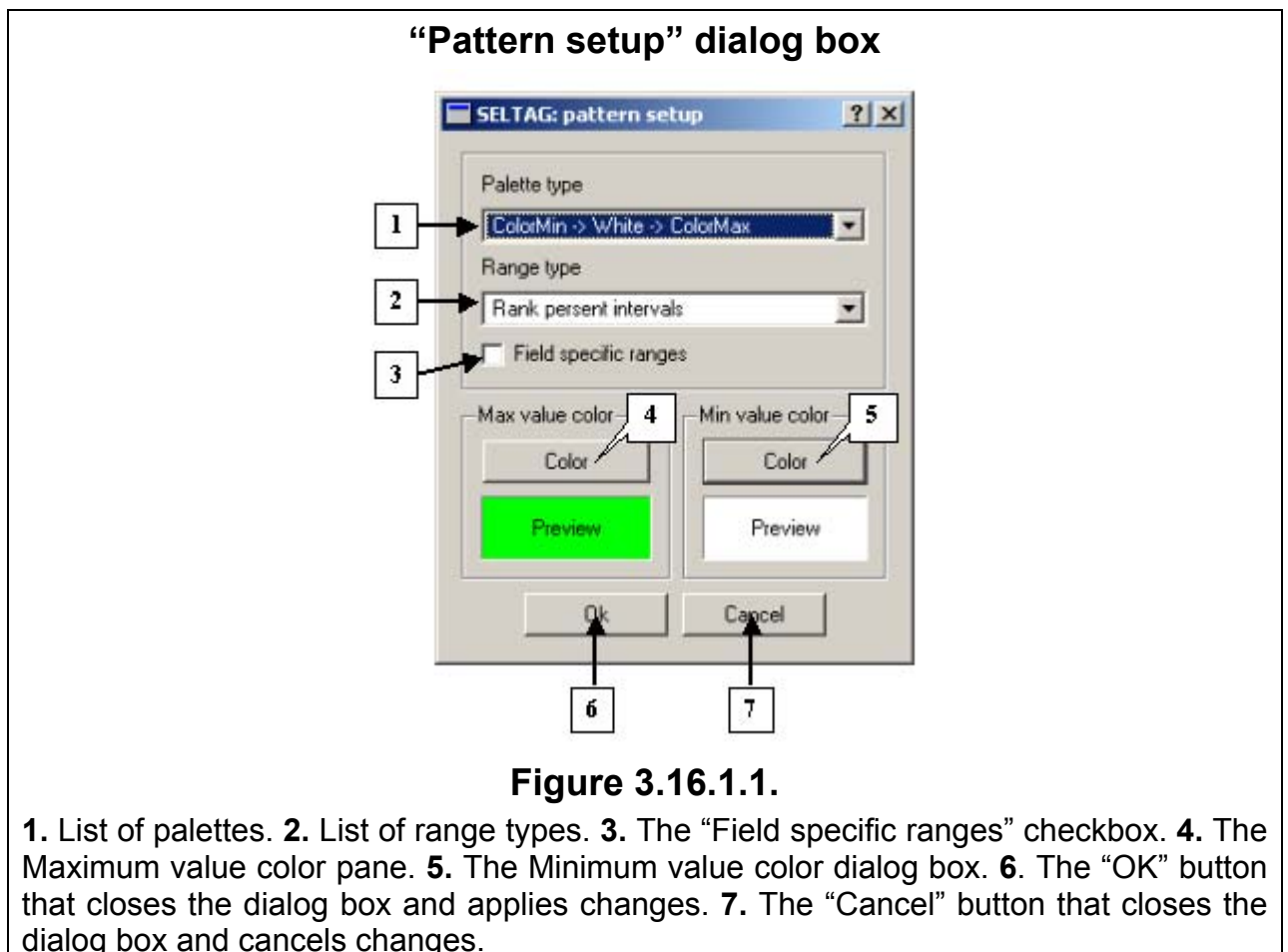
3.16.1.2.1. “Value intervals” – the color range of a matrix element depends on the gene expression value in the current tissue.

3.16.1.2.2. «Rank percent intervals» - the color range of a matrix element depends on the rank of the gene expression value in the current tissue.

3.16.1.3. In the “Max. value color” pane, you can set the color for the maximum value. This option is available for palettes 3.16.1.1.1, 3.16.1.1.2, 3.16.1.1.3, and 3.16.1.1.4. Click the “Color” button to open the “Select color” dialog box where you can select the desired color. The color you selected is shown in the “Preview” pane.

3.16.1.4. In the “Min. value color” pane, you can set the color for the minimum value. This option is available for palettes 3.16.1.1.3 and 3.16.1.1.4. Click the “Color” button to open the “Select color” dialog box where you can select the desired color. The color you selected is shown in the “Preview” pane.

- Control buttons:
- “Cancel” – closes the dialog box and ignores changes.
- “OK” – closes the dialog box and applies changes.



### 3.16.2. “Image options” dialog box

In the “Image options” dialog box (Fig. 3.16.2.1), you can configure the general image parameters. The dialog box has four tabs:

“**Common & Borders**” tab (Fig. 3.16.2.1) sets the general image options:

The “Common” pane has the following checkboxes:

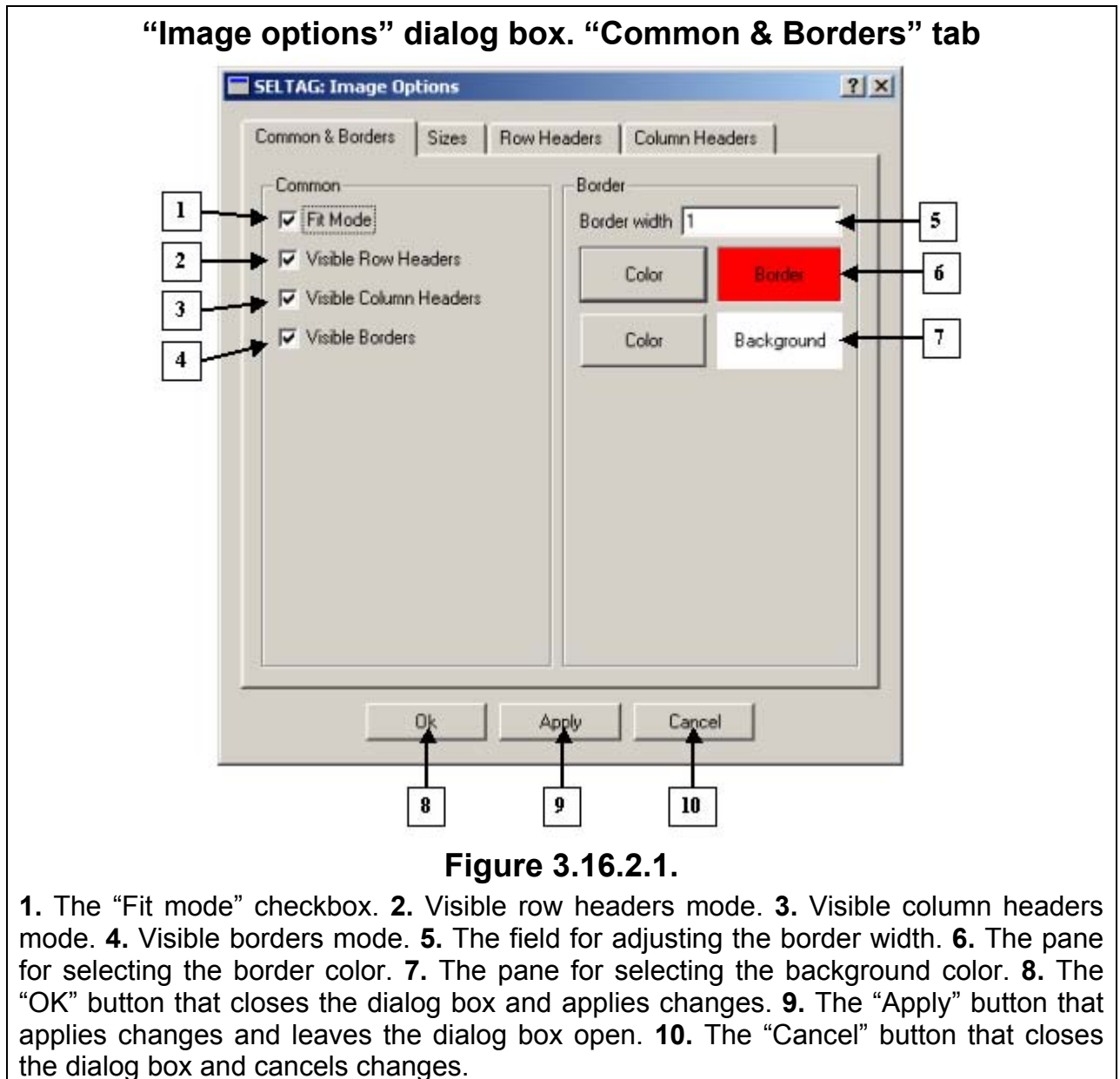
- “Fit Mode” – if checked, the image is resized to fit closely the size of headers.
- “Visible Row Headers” – if checked, the matrix row headers become visible.
- “Visible Column Headers” – if checked, the matrix column headers become visible.
- “Visible Borders” – if checked, the matrix cell borders become visible.

The “Borders” pane has the following options for the cell borders and image background:

In the “Border width” field, you can define the border width in pixels.

Click the “Color” button to open the “Select color” dialog box and select the border color. The current border color is displayed in the “Border” view pane.

Click the “Color” button to open the “Select color” dialog box and select the background color. The current background color is displayed in the “Background” view pane.

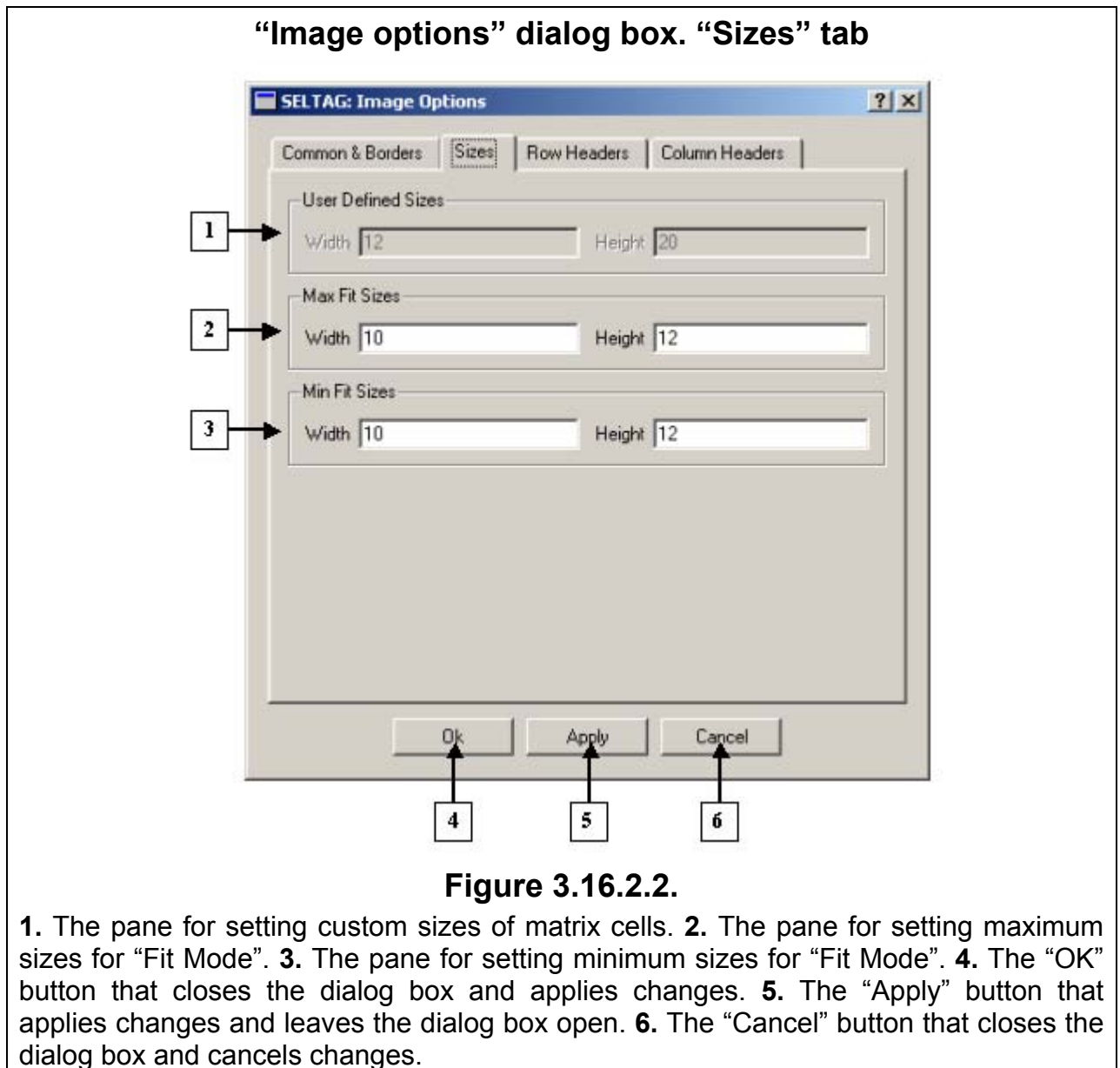


The “**Sizes**” tab (Fig. 3.16.2.2) sets the matrix cell sizes:

- In the “User Defined Sizes” fields, you can set the following dimensions of matrix cells:
  - Width (in the “Width” field)
  - Height (in the “Height” field)
- In the “Max Fit Sizes” fields, you can set maximum cell sizes for the “Fit Mode” mode:



- Width (in the “Width” field)
- Height (in the “Height” field)
- In the “Min Fit Sizes” fields, you can set minimum cell sizes for the “Fit Mode” mode:
  - Width (in the “Width” field)
  - Height (in the “Height” field)



The “**Row Headers**” tab (Fig. 3.16.2.3) has the following options for row headers:

The “**Short Headers**” checkbox. If this box is checked, only short names are displayed as row headers.

In the “**Number of Visible Characters**” field, you can specify the maximum number of characters in short names.

In the “**Interval**” field, you can specify the space between headers (in pixels). This field is active when the “Fit Mode” checkbox is unchecked.

In the “**Row Width**” field, specify the width of row headers.

In the “**Alignment**” list, select the alignment type for the cell contents:

“Left” – to left margin

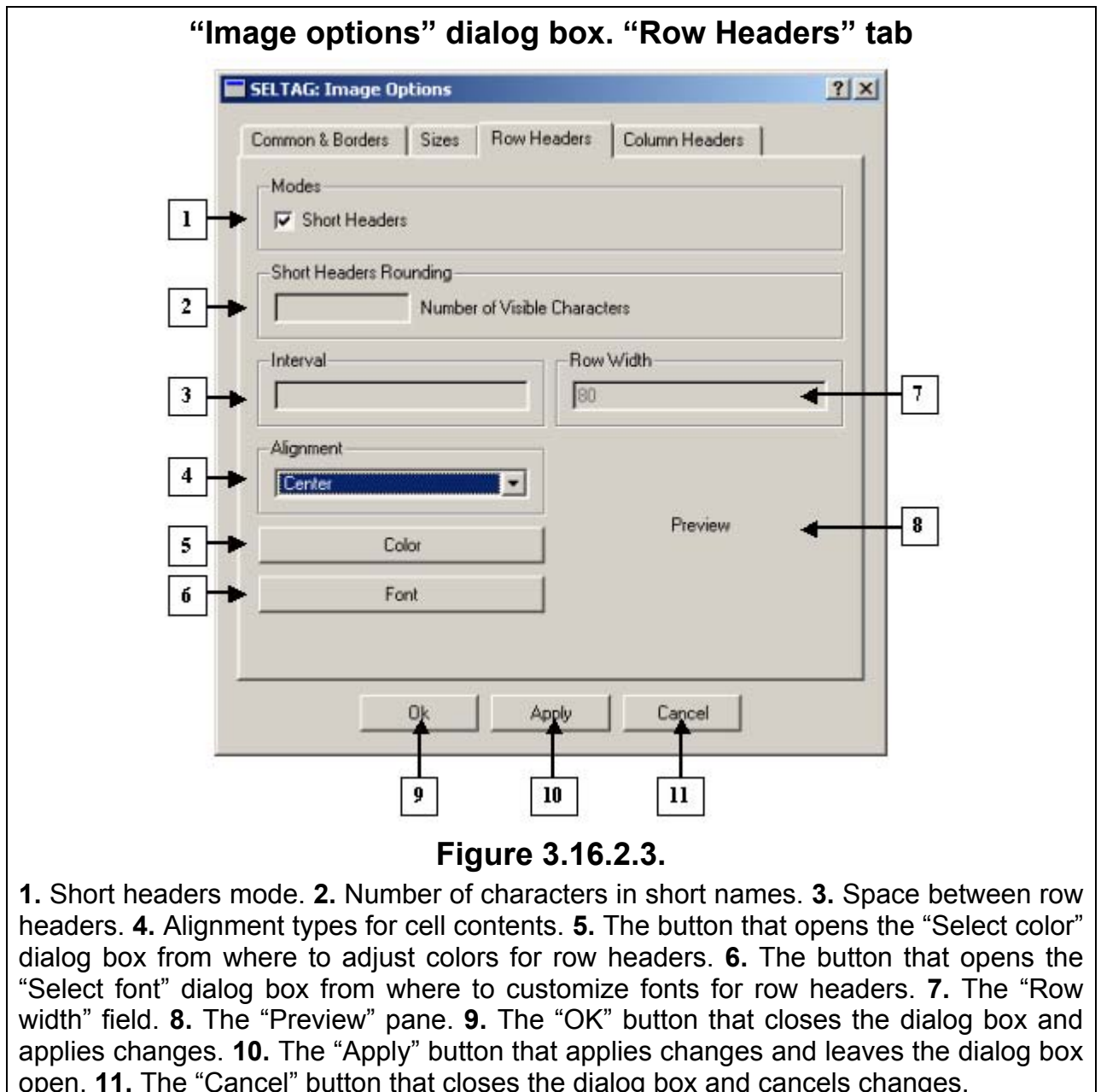
“Center” – centered

“Right” – to right margin

Click the **“Color”** button to open the “Select color” dialog box (see section 3.18) and select a new color for row names.

Click the **“Font”** button to open the “Select font” dialog box (see section 3.19) and select a new font for row names.

The **“Preview”** pane shows the selected font, color and alignment of row names.



The **“Column Headers”** tab (Fig.3.16.2.4) has the following options for column headers:

The **“Short Headers”** checkbox. If this box is checked, only short names are displayed as column headers.

In the **“Number of Visible Characters”** field, you can specify the maximum number of characters in short names.

In the **“Interval”** field, you can specify the space between headers (in pixels). This field is active when the “Fit Mode” checkbox is unchecked.

In the **“Column Height”** field, specify the height of column headers.

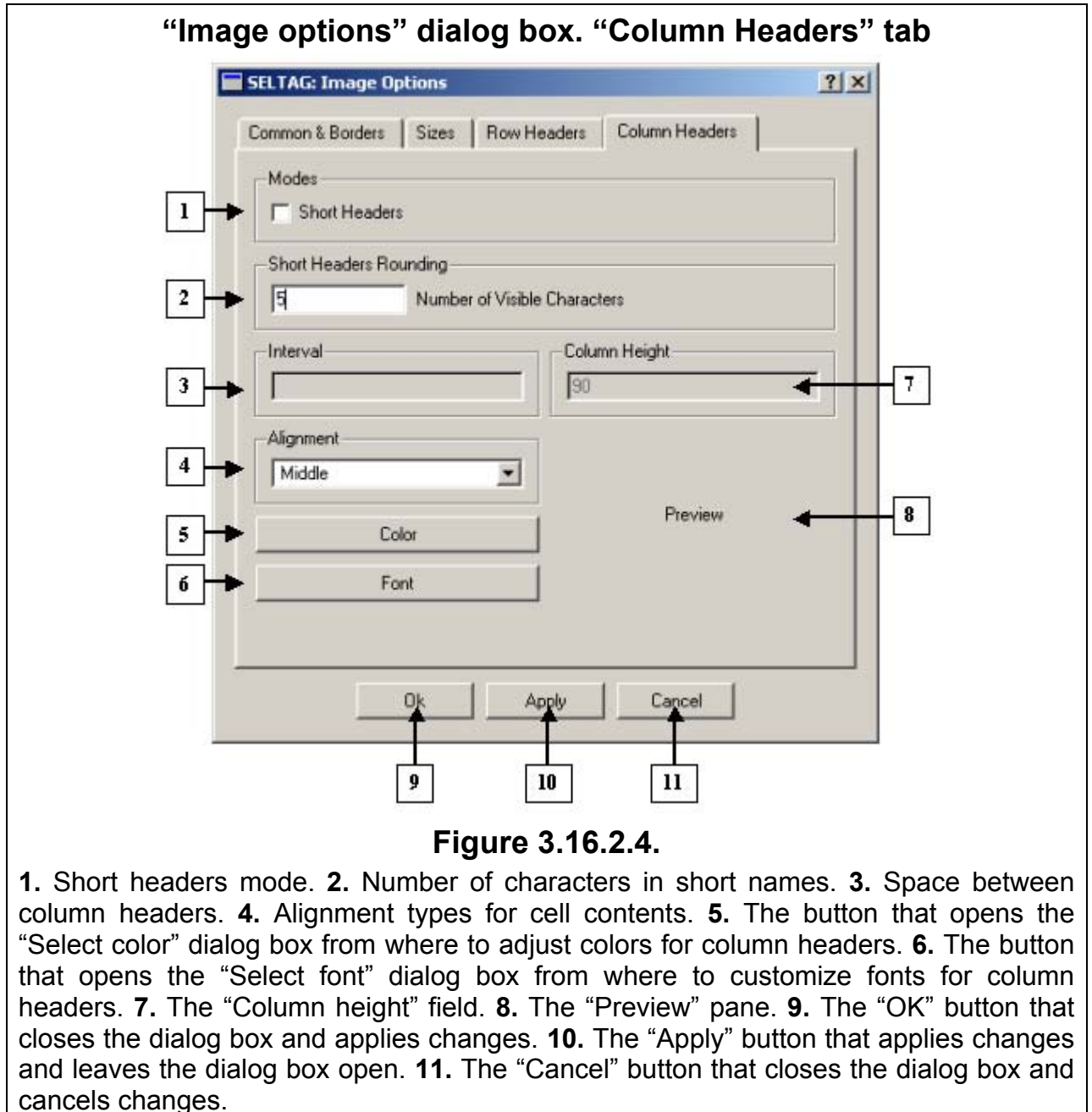
In the **“Alignment”** list, select the alignment type for the cell contents:

- “Top” – to the top of the cell
- “Middle” – to the middle of the cell
- “Bottom” – to the bottom of the cell

Click the **“Color”** button to open the “Select color” dialog box (see section 3.18) and select a new color for column names.

Click the **“Font”** button to open the “Select font” dialog box (see section 3.19) and select a new font for row names.

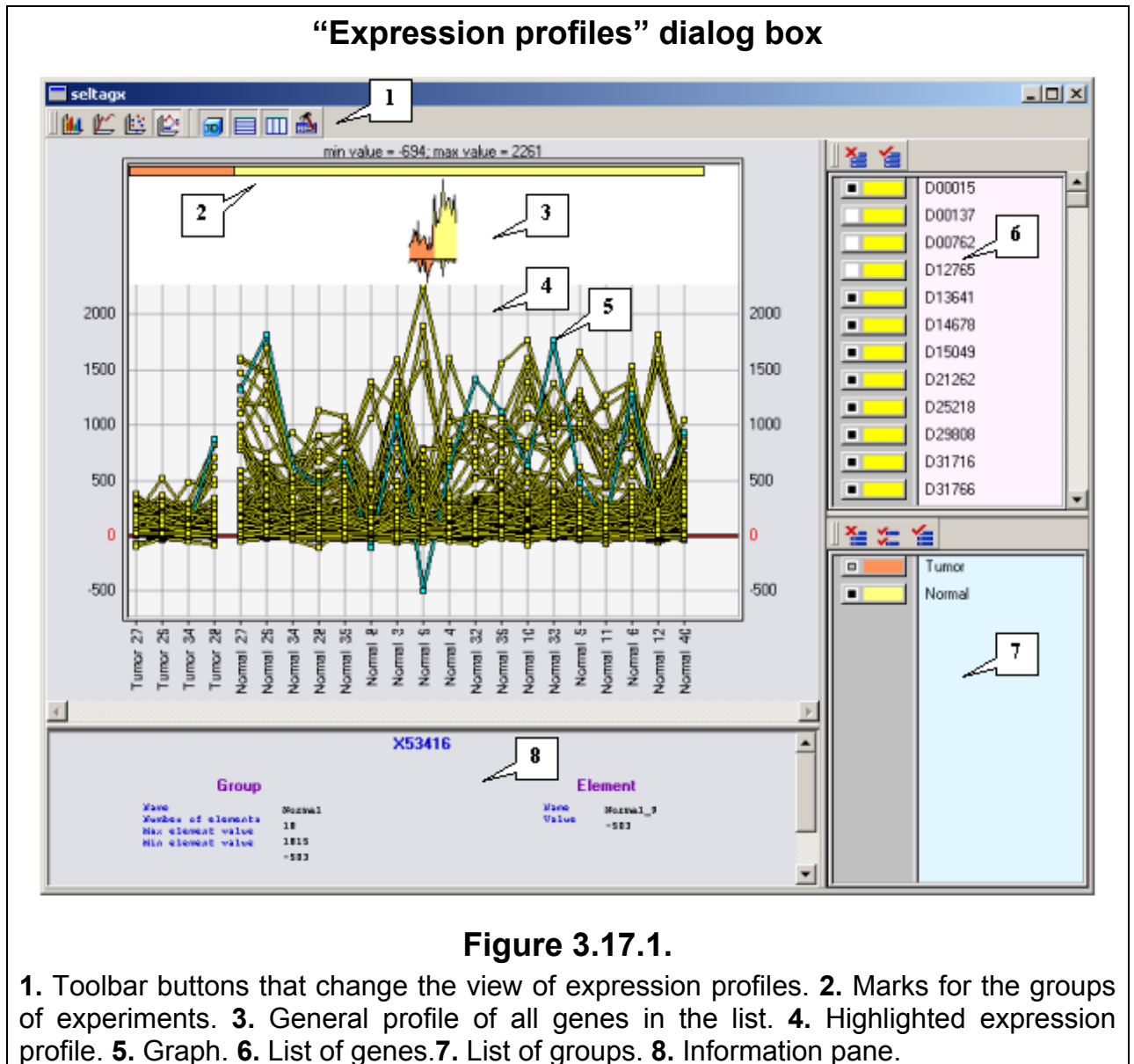
The **“Preview”** pane shows the selected font, color and alignment of row names.



**Note.** The “OK” button that closes the dialog box and applies changes. The “Apply” button that applies changes and leaves the dialog box open. The “Cancel” button that closes the dialog box and cancels changes.






### 3.17. “Expression profiles” dialog box




The “Expression profiles” dialog box is accessed by clicking the Analysis>Expression profiles main menu command. The dialog box displays gene expression profiles. This dialog box contains the following elements: the graph, the list of genes, the list of field groups, and the information pane (Fig. 3.17.1).



#### 3.17.1. Graph

In the upper part of the “Expression profiles” dialog box, there are toolbar buttons that can be used to change the graphical representation of the data:

-  - displays the data as a histogram.
-  - displays the data as a graph with data markers.
-  - displays the data as a graph.
-  - displays the data as a point graph.
-  - displays the horizontal lines of the grid in the diagram view area.

-  - displays the vertical lines of the grid in the diagram view area.
-  - displays the diagram in pseudovolume mode (this mode is available only for bar diagrams).
-  - calls the “Options” dialog box.

If you click on one of experiments with your left mouse button, the gene expression profile corresponding to this experiment will be highlighted in another color in the profile view area (Fig. 3.17.1). The highlighting mode is set in the “Options” dialog box (section 3.17.6).

### 3.17.2. Information pane

You can see the information pane below the profile view area. The information about a graph element is displayed when you point to the element with your mouse cursor (or click your mouse, depending on the settings). In the pane, you can see information about the group of experiments, which is at the current cursor position, and information about the experiment you are pointing to with your mouse cursor. The gene description is displayed below.

### 3.17.3. Experiments view modes

There are two view modes for experiments:

- Single experiment
- Multiple experiments

The modes are switched in the gene list.



In single-experiment mode, a profile is displayed for only one gene. The color of the elements of the expression profile depends on the color of the experimental group. Above the profile, you can see the maximum and minimum values for all groups of experiments. In the upper part of the profile view area, you can see the general profile of this gene in all experiments. The expression value is indicated on the left axis. Short names of the experiments are indicated on the horizontal axis.

In multiple-experiment mode (Fig. 3.17.1), you can see expression profiles for one and more genes. The color of the elements of the expression profiles depends on the color of the gene. Above the profile, you can see the maximum and minimum values for all genes in the list and for all groups of experiments. In the upper part of the profile view area, you can see a line marking the groups of experiments. The color of the marking line corresponds to the color of the group of experiments located above. The general profile for all genes in the list is shown above the expression profiles for separate genes.



### 3.17.4. Gene list

In the right upper corner of the dialog box, you can see a list of genes.

By clicking buttons at the left, you can switch between single-experiment and multiple-experiment modes. These buttons also allow you to show/hide a profile. This option (enabled/disabled) is shown by the black square located on the left side of each button.

If a button corresponding to any gene is pressed, the gene data will be displayed in single-experiment mode; otherwise, it will be displayed in multiple-experiment mode. The black square shows that this profile is displayed (); if the button has no black square in it () , this profile is hidden.


To change the color of the profile, right-click on the corresponding button. This will open a shortcut menu. Select the “Set color” command on the menu and set the color in the corresponding dialog box.


Above the list of genes, you can see buttons for quick selection/deselection of genes ( unselect,  - select).

### 3.17.5. List of experiments




In the right lower corner of the dialog box, you can see a list of experiments. Using the buttons at the left of the names of experiments (the color of the buttons correspond to the color of the group), you can work separately with each group:

 - hide the group

 - show the elements selected in the list for this group. You can open the list from the shortcut menu (see below).

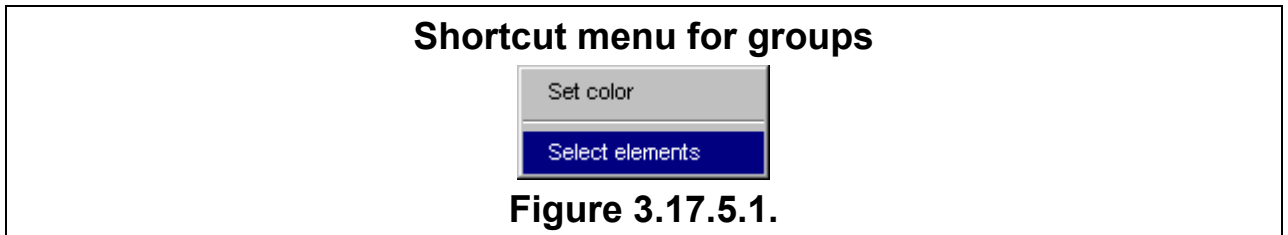
 - display all elements in the group.





If the button for a group is pressed, this group will be displayed in single-experiment mode; otherwise, this group will be displayed in multiple-experiment mode.

Above the list of groups, you can see buttons for quickly changing the view:  - hide groups,  - show experiments selected in the lists of groups,  - show all experiments in all groups.

Click your right mouse button to open a shortcut menu (Fig. 3.17.5.1) that has the following options:

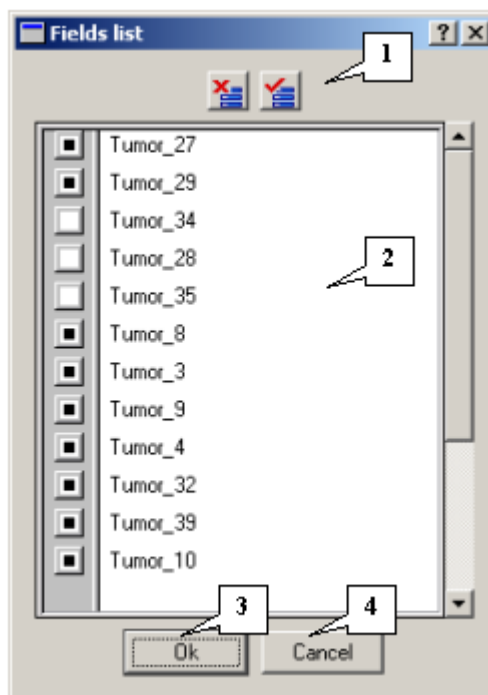
- “Set color” – opens the “Select color” dialog box from where you can select a color for a group.
- “Select elements” – opens the “Fields list” dialog box with the list of experiments in the group (Fig. 3.17.5.2).



In the list, you can select the elements to be displayed with your mouse. To select the elements to be displayed, click the box at the left of the names of the experiments:  - the experiment is displayed and  - the experiment is hidden. Above the list, you can see buttons for quick selection/deselection of experiments ( unselect,  - select).

To confirm your selection, click “OK”. The set of the elements to be displayed is valid only for displaying the experiments from the list.

### “Fields list” dialog box



**Figure 3.17.5.2.**

1. The buttons for quick selection/deselection of experiments. 2. List of experiments. 3. The “OK” button that applies changes and closes the dialog box. 4. The “Cancel” button that closes the dialog box and cancels changes.

### 3.17.6. The “Options” dialog box

#### Selection modes:

- “Highlight mode” – select one profile. If you point with your mouse while pressing the left mouse key to an experiment, the corresponding profile will be highlighted.
- “Many selection mode” – select more than one profile. The profiles are selected by clicking the left mouse key. To select profiles, click the mouse button one more time.
- Click the “Color” button to select the highlighting color.
- In the “Preview” pane, view the highlighting color.

#### Graph options:

- “Global view height”- the height of the general profile (in pixels).
- “Max label area height” - the maximum height of the label area (below the graph, in pixels). Because all labels are positioned vertically, you can look through all short names in groups to find out the longest name. Then, set the length of this longest name as the maximum height. If the lengths of other names are less than the specified value, the height will be automatically chosen to be equal to the maximum name length in all groups. If the name length is greater than the specified value, the name will be truncated.
- “Symbols after dot count” – not available in this version.
- “Always scrollbars” – when this checkbox is checked, the scrollbars are always displayed in the graph, regardless of whether they are active or not.

- “Whole lines only” – when this option is enabled, only whole lines or series of lines (for several genes) are displayed. When this option is disabled, the lines or series of lines are truncated.
- “Moving info” – when this option is enabled, information about an element is automatically displayed in the information pane when you are pointing to this element with your mouse. When this checkbox is unchecked, the information is displayed only after you click on the element.
- Click the «Color» button to set the background color, the “Font” button to select fonts for the graph labels. In the “Diagram” pane, you can view the background color and label font you selected.
- Click the “Color” button to customize the background color for the general profile. In the “Global view background” pane, you can view the background color of the general profile.
- Click the “Color” button to customize the background color for the diagram. In the “Map background” pane, you can view the color of diagram.

#### **Gene list options:**

- “Always scrollbars” - when this checkbox is checked, the scrollbars are always displayed in the graph, regardless of whether they are active or not.
- “Whole lines only” - when this option is enabled, only whole lines or series of lines (for several genes) are displayed. When this option is disabled, the lines or series of lines are truncated.
- Click the «Color» button to set the background color for the list of genes.
- The “Font” button to select the font for the list of genes.
- In the “Preview” pane, you can view the selected background color and font for the list of genes.

#### **Group list options:**

- “Always scrollbars” - when this checkbox is checked, the scrollbars are always displayed in the graph, regardless of whether they are active or not.
- “Whole lines only” - when this option is enabled, only whole lines in the group list are displayed. When this option is disabled, the lines are truncated.
- Click the «Color» button to set the background color for the list of groups.
- The “Font” button to select the font for the list of groups.
- In the “Preview” pane, you can view the selected background color and font for the list of groups.

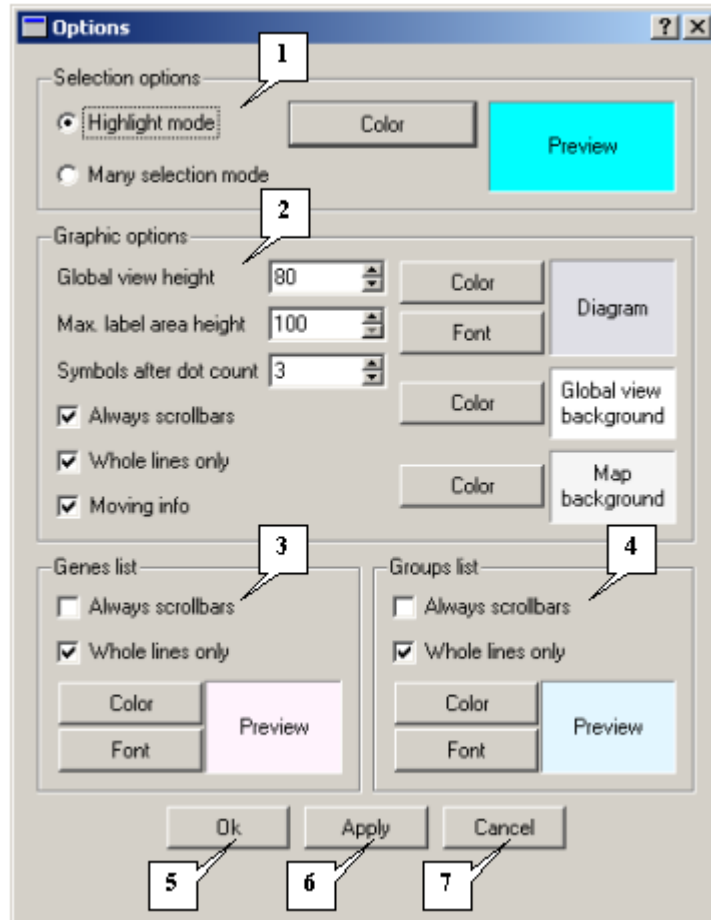
Click the “OK” button to close the dialog box and apply changes.

Click the “Apply” button to apply changes and leave the dialog box open.

Click the “Cancel” button to close the dialog box and cancel changes.



## “Options” dialog box



**Figure 3.17.6.1.**

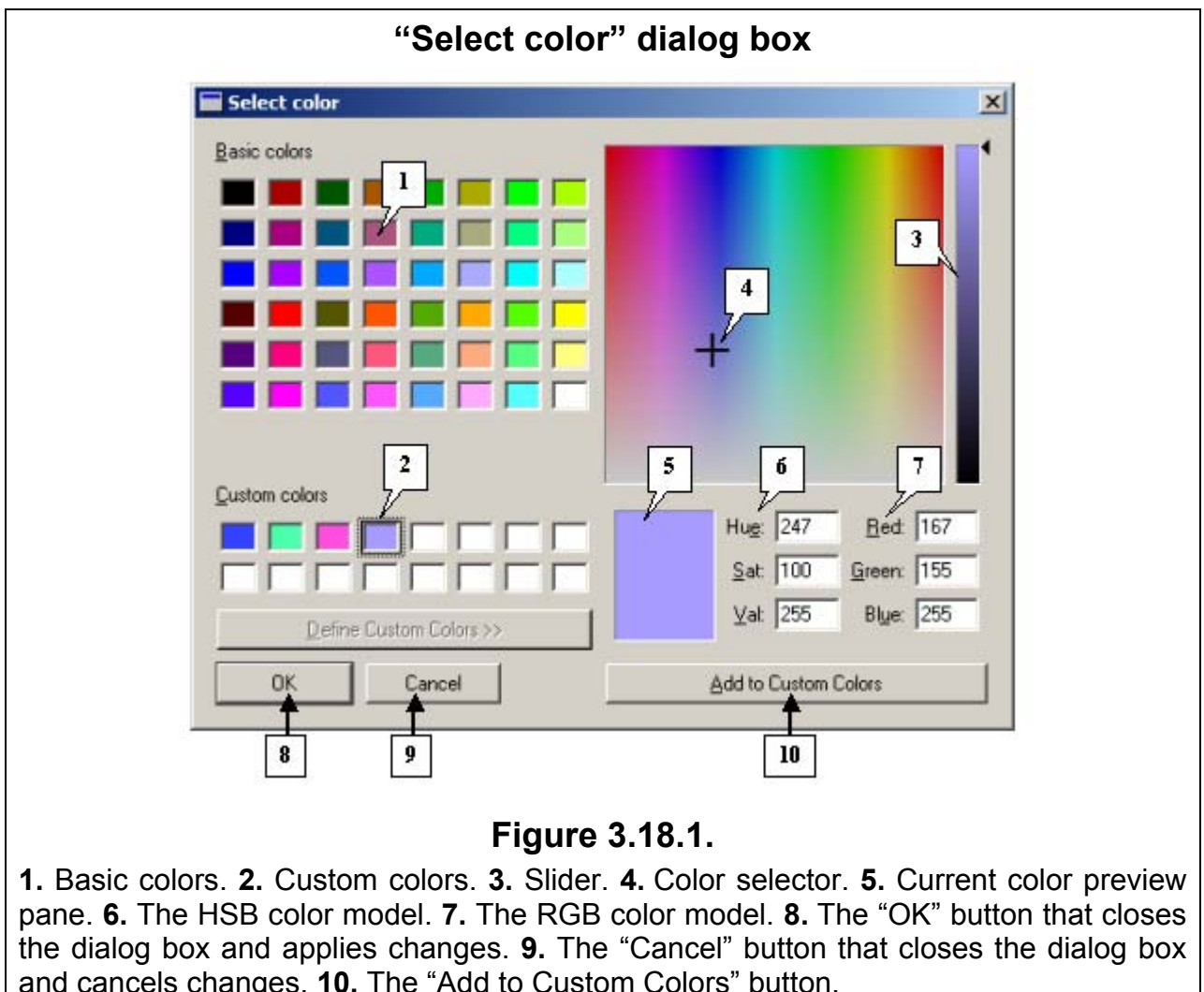
1. The “Selection options” pane. 2. The “Graphic options” pane. 3. The “Genes list options” pane. 4. The “Groups list options” pane. 5. The “OK” button that closes the dialog box and applies changes. 6. The “Apply” button that applies changes and leaves the dialog box open. 7. The “Cancel” button that closes the dialog box and cancels changes.

### 3.18. “Select color” dialog box

In the “Select color” dialog box (Fig. 3.18.1), you can select colors by using several methods:

- Basic colors. Click a color box to change the basic color. You will see the new color in the color preview box. To confirm your selection, click “OK”.
- Custom colors. Click the color selector and pick the required color. You will see the new color in the color preview box. To confirm your selection, click “OK”.
- New color. To add a new color, you can move a slider, color selector, or the HSB or RGB color fields. The new color will appear in the color preview box. To confirm your selection, click “OK”.

**Note.** To change a custom color, click the custom color box and then change the color by moving the slider selector or inserting the required values in the HSB or RGB color fields. After this, click the “Add to Custom Colors” button. The new custom color will appear in the custom color boxes.



**Figure 3.18.1.**

1. Basic colors. 2. Custom colors. 3. Slider. 4. Color selector. 5. Current color preview pane. 6. The HSB color model. 7. The RGB color model. 8. The “OK” button that closes the dialog box and applies changes. 9. The “Cancel” button that closes the dialog box and cancels changes. 10. The “Add to Custom Colors” button.

#### Control buttons

The “OK” button closes the dialog box and applies changes and the “Cancel” closes the dialog box and cancels changes.

### 3.19. “Select font” dialog box

In the “Select Font” dialog box (Fig. 3.19.1) you can select the following font options:

- Font in the “Font” list.
- Font style in the “Font style” list.
- Font size in the “Size” list.
- Font effects in the “Effects” pane.
- Font script in the “Script” list.

You can view the font selected in the “Preview” pane.

#### Control buttons

The “OK” button closes the dialog box and applies changes and the “Cancel” closes the dialog box and cancels changes.

