Content

1. Introduction..1
2. Running the application and loading data..1
3. Hierarchical clustering of experiments..9
4. References 21

1. Introduction

Using SelTag software the alon1999_set.txt data were analyzed. The data contain measurements of expression levels of 2000 human cDNAs and ESTs (including sequences homologous to some known eukaryotic genes) in colon adenocarcinoma tissues from several patients. For some patients, expression of these RNAs was also measured in normal colon tissues. Totally the table contains the measurements of expression in 40 tumor and 22 normal colon tissues. These data are combined into appropriate measurement groups “Tumor” and “Normal”. Analysis consisted in building the hierarchical clustering for tissues. It was obtained the division of tissues (experiments) into two classes. The first class includes predominantly tumor tissues, the second one – normal. Results were compared to the ones obtained in original paper [1].

2. Running the application and loading data.

2.1. On the application startup the “Login” dialog window appears (fig. 2.1). In this window select the “Anonymous” mode and press the “OK” button.

Note. The “Anonymous” mode is intended for working with demo data.

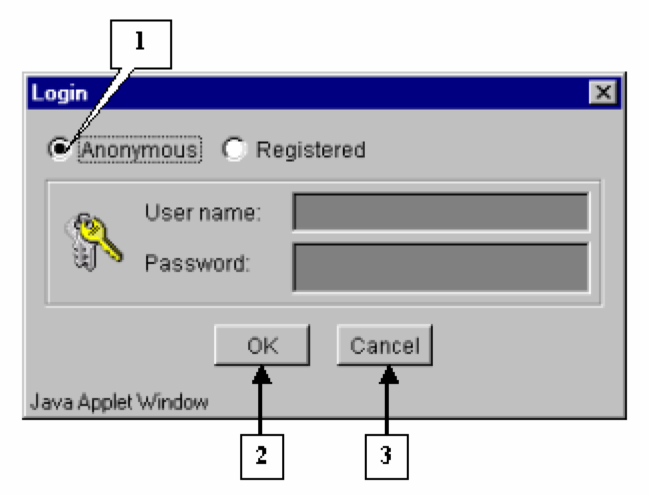


Figure 2.1.

1. Non-registered user mode. **2**. Button to run application. **3.** Button to cancel the application start.

2.2. After the application is started the main application window appears. Select the “File>Open data” command in the main menu (fig. 2.2)

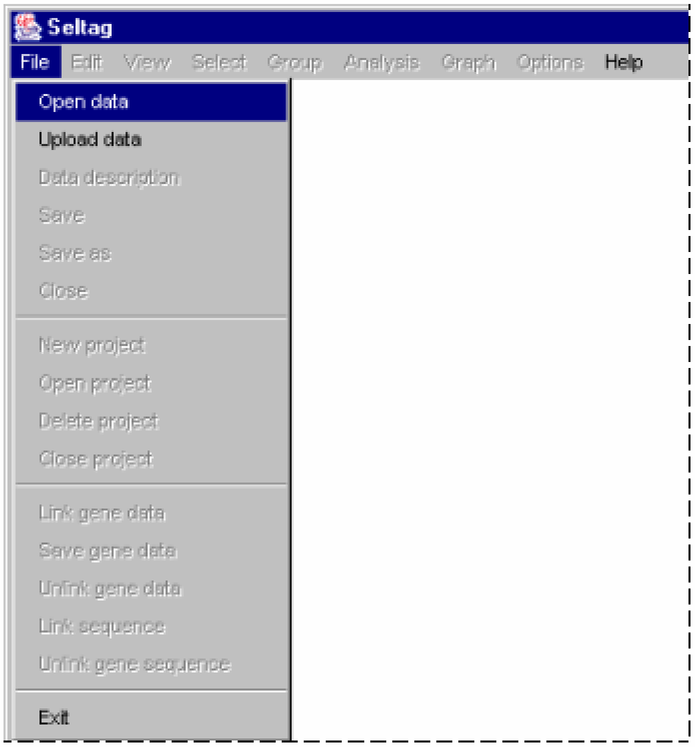
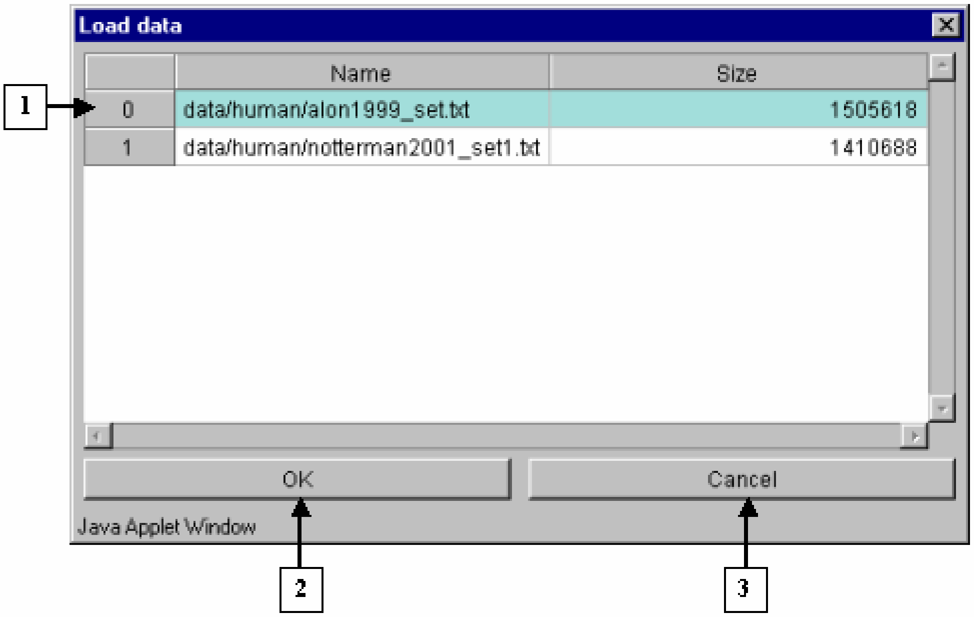


Figure 2.2.

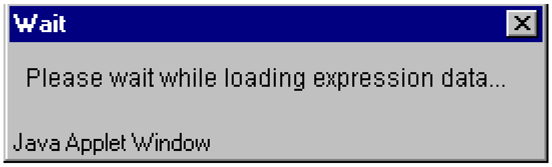
2.3. Once it is done, appears the “Load data” dialog window that contains the names of files with data tables and sizes of these files (fig. 2.3.1). In this window select a file and press the “OK” button. It will cause appearance of the “Wait” message box, which will disappear after finishing of data loading (fig. 2.3.2).



	Name	Size
0	data/human/alon1999_set.txt	1505618
1	data/human/notterman2001_set1.txt	1410688

Figure 2.3.1.

1. Selected file with expression data. **2.** Confirmation button. **3.** Cancel button.

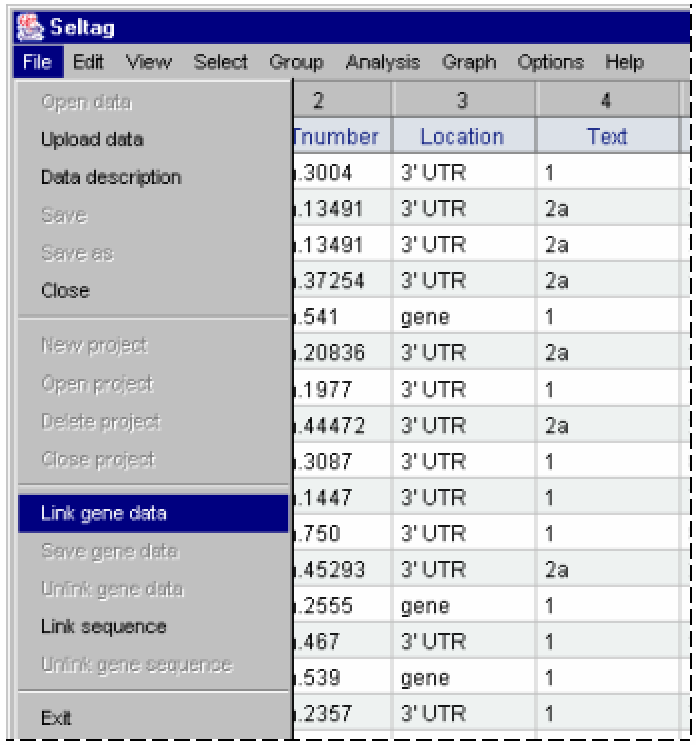


Please wait while loading expression data...

Java Applet Window

Figure 2.3.2.

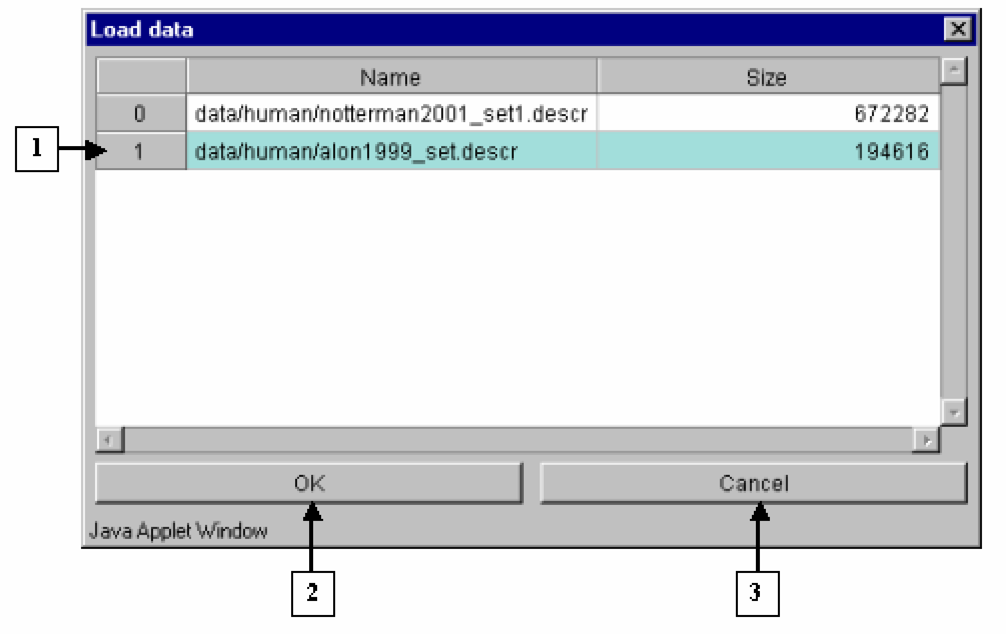
2.4. Table with the selected data will be shown in the main application window.
2.5. To load a file with genes description select the “File>Link gene data” command in the main menu (fig. 2.5).



2	3	4
Tnumber	Location	Text
.3004	3' UTR	1
.13491	3' UTR	2a
.13491	3' UTR	2a
.37254	3' UTR	2a
.541	gene	1
.20836	3' UTR	2a
.1977	3' UTR	1
.44472	3' UTR	2a
.3087	3' UTR	1
.1447	3' UTR	1
.750	3' UTR	1
.45293	3' UTR	2a
.2555	gene	1
.467	3' UTR	1
.539	gene	1
.2357	3' UTR	1

Figure 2.5.

2.6. The “Load data” dialog window that contains the names of files with genes descriptions and sizes of these files will appear (fig. 2.6). In this window select a file and press the “OK” button.

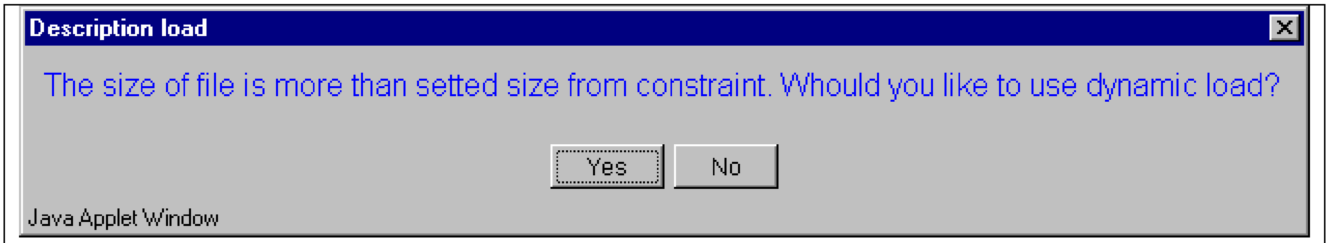


	Name	Size
0	data/human/notterman2001_set1.descr	672282
1	data/human/alon1999_set.descr	194616

Figure 2.6.

1. Selected file with genes description. **2.** Confirmation button. **3.** Cancel button.

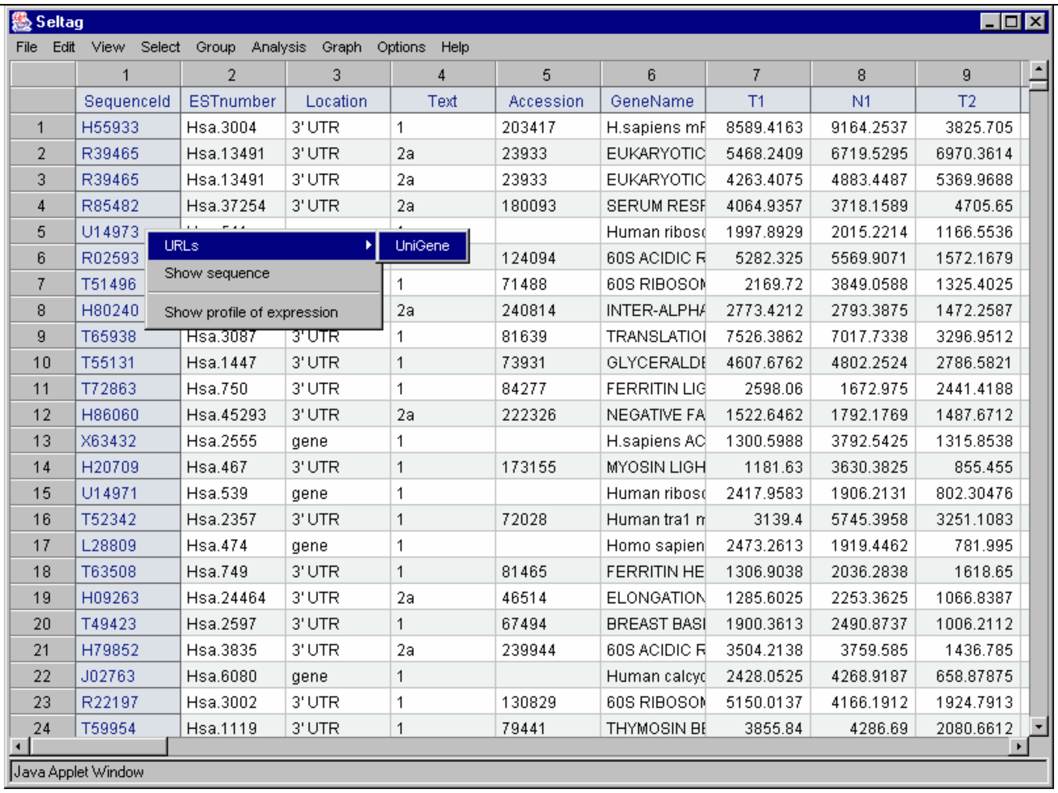
2.7. The “Description load” message box with suggestion to use dynamic file loading mode (fig. 2.7) will appear. Press the “Yes” button.



The size of file is more than setted size from constraint. Whould you like to use dynamic load?

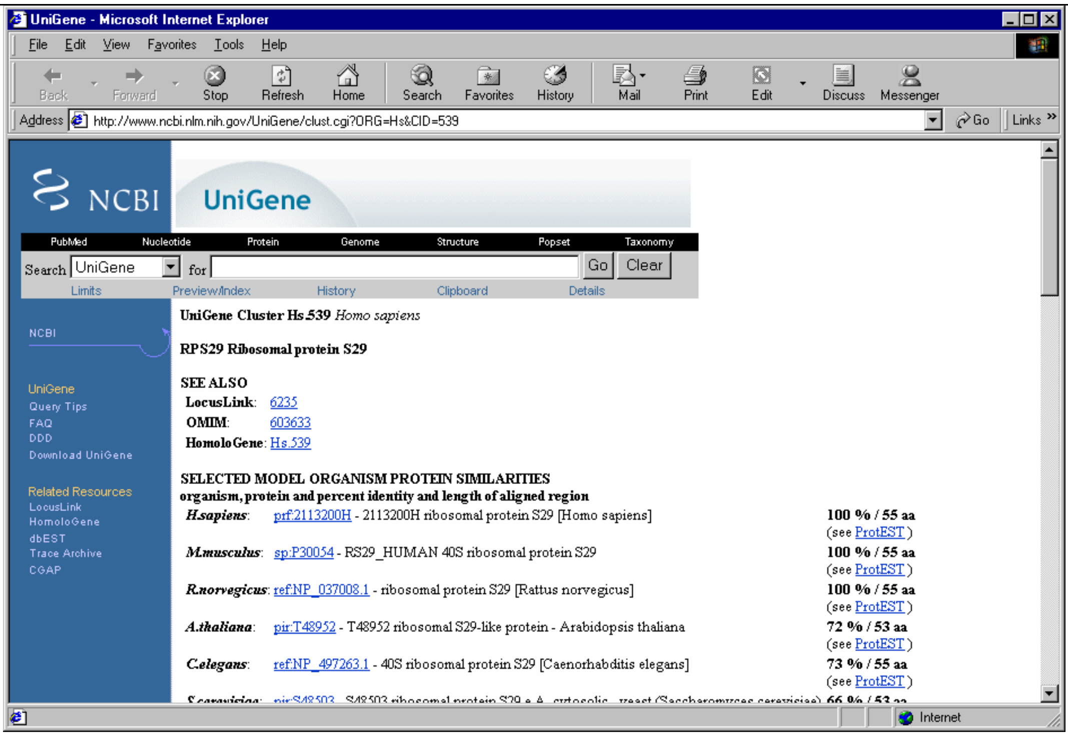
Figure 2.7.

2.8. In contextual menu of the application main table (contextual menu can be called out by the right mouse click) the “URLs>UniGene” command will become active (fig. 2.8.1). This command, using a web link to gene, loads a card from UniGene database for the appropriate gene into new window of your web browser (fig. 2.8.2).


Seltag
File Edit View Select Group Analysis Graph Options Help
URLs
UniGene
Show sequence
Show profile of expression
Java Applet Window

	1	2	3	4	5	6	7	8	9
	SequenceId	ESTnumber	Location	Text	Accession	GeneName	T1	N1	T2
1	H55933	Hsa.3004	3' UTR	1	203417	H.sapiens mF	8589.4163	9164.2537	3825.705
2	R39465	Hsa.13491	3' UTR	2a	23933	EUKARYOTIC	5468.2409	6719.5295	6970.3614
3	R39465	Hsa.13491	3' UTR	2a	23933	EUKARYOTIC	4263.4075	4883.4487	5369.9688
4	R85482	Hsa.37254	3' UTR	2a	180093	SERUM RESF	4064.9357	3718.1589	4705.65
5	U14973			1		Human ribosc	1997.8929	2015.2214	1166.5536
6	R02593				124094	60S ACIDIC R	5282.325	5569.9071	1572.1679
7	T51496			1	71488	60S RIBOSON	2169.72	3849.0588	1325.4025
8	H80240			2a	240814	INTER-ALPHA	2773.4212	2793.3875	1472.2587
9	T65938	Hsa.3087	3' UTR	1	81639	TRANSLATIOI	7526.3862	7017.7338	3296.9512
10	T55131	Hsa.1447	3' UTR	1	73931	GLYCERALDE	4607.6762	4802.2524	2786.5821
11	T72863	Hsa.750	3' UTR	1	84277	FERRITIN LIG	2598.06	1672.975	2441.4188
12	H86060	Hsa.45293	3' UTR	2a	222326	NEGATIVE FA	1522.6462	1792.1769	1487.6712
13	X63432	Hsa.2555	gene	1		H.sapiens AC	1300.5988	3792.5425	1315.8538
14	H20709	Hsa.467	3' UTR	1	173155	MYOSIN LIGH	1181.63	3630.3825	855.455
15	U14971	Hsa.539	gene	1		Human ribosc	2417.9583	1906.2131	802.30476
16	T52342	Hsa.2357	3' UTR	1	72028	Human tra1 m	3139.4	5745.3958	3251.1083
17	L28809	Hsa.474	gene	1		Homo sapien	2473.2613	1919.4462	781.995
18	T63508	Hsa.749	3' UTR	1	81465	FERRITIN HE	1306.9038	2036.2838	1618.65
19	H09263	Hsa.24464	3' UTR	2a	46514	ELONGATION	1285.6025	2253.3625	1066.8387
20	T49423	Hsa.2597	3' UTR	1	67494	BREAST BASI	1900.3613	2490.8737	1006.2112
21	H79852	Hsa.3835	3' UTR	2a	239944	60S ACIDIC R	3504.2138	3759.585	1436.785
22	J02763	Hsa.6080	gene	1		Human calcyc	2428.0525	4268.9187	658.87875
23	R22197	Hsa.3002	3' UTR	1	130829	60S RIBOSON	5150.0137	4166.1912	1924.7913
24	T59954	Hsa.1119	3' UTR	1	79441	THYMOSIN BE	3855.84	4286.69	2080.6612

Figure 2.8.1.


UniGene - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Messenger
Address http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Hs&CID=539
Go Links
NCBI
UniGene
PubMed Nucleotide Protein Genome Structure Popset Taxonomy
Search UniGene for
Go Clear
Limits Preview/Index History Clipboard Details
NCBI
UniGene
Query Tips
FAQ
DDD
Download UniGene
Related Resources
LocusLink
HomoloGene
dbEST
Trace Archive
CGAP
Internet

UniGene Cluster Hs.539 *Homo sapiens*

RPS29 Ribosomal protein S29

SEE ALSO
LocusLink: 6235
OMIM: 603633
HomoloGene: Hs.539

SELECTED MODEL ORGANISM PROTEIN SIMILARITIES
organism, protein and percent identity and length of aligned region

H.sapiens: prf:2113200H - 2113200H ribosomal protein S29 [Homo sapiens] **100 % / 55 aa** (see ProtEST)

M.musculus: sp:P30054 - RS29_HUMAN 40S ribosomal protein S29 **100 % / 55 aa** (see ProtEST)

R.norvegicus: ref:NP_037008.1 - ribosomal protein S29 [Rattus norvegicus] **100 % / 55 aa** (see ProtEST)

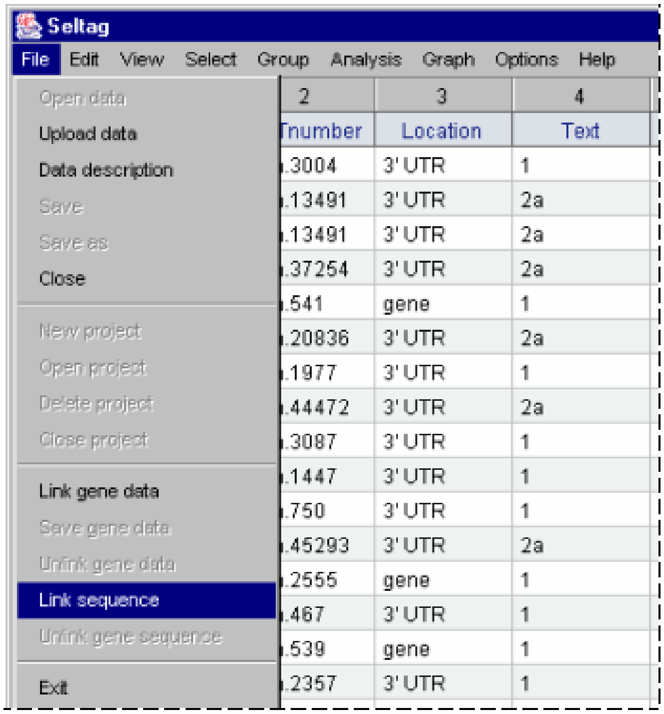
A.thaliana: pir:T48952 - T48952 ribosomal S29-like protein - Arabidopsis thaliana **72 % / 53 aa** (see ProtEST)

C.elegans: ref:NP_497263.1 - 40S ribosomal protein S29 [Caenorhabditis elegans] **73 % / 55 aa** (see ProtEST)

S.cerevisiae: pir:S48503 - S48503 ribosomal protein S29.e.A, cytosolic - yeast (Saccharomyces cerevisiae) **66 % / 53 aa**

Figure 2.8.2.

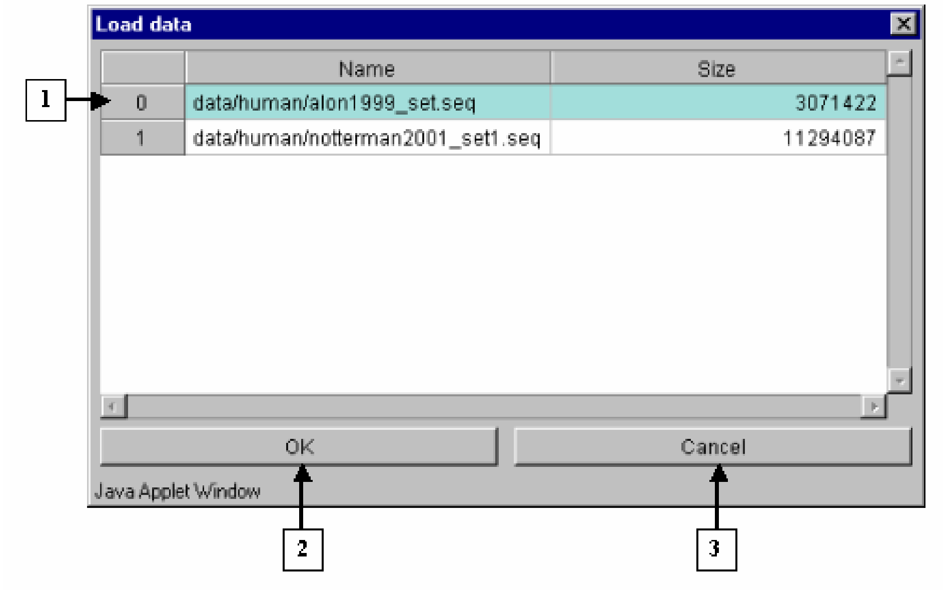
2.9. To load a file with gene's nucleotide sequences, select the "File>Link sequence" command in the main menu (fig. 2.9).



2	3	4
Tnumber	Location	Text
.3004	3' UTR	1
.13491	3' UTR	2a
.13491	3' UTR	2a
.37254	3' UTR	2a
.541	gene	1
.20836	3' UTR	2a
.1977	3' UTR	1
.44472	3' UTR	2a
.3087	3' UTR	1
.1447	3' UTR	1
.750	3' UTR	1
.45293	3' UTR	2a
.2555	gene	1
.467	3' UTR	1
.539	gene	1
.2357	3' UTR	1

Figure 2.9.

2.10. It will result in appearance of the "Load data" dialog window that contains the names of files with genes' sequences and information on sizes of these files (fig. 2.10). In this window select a file and press the "OK" button.

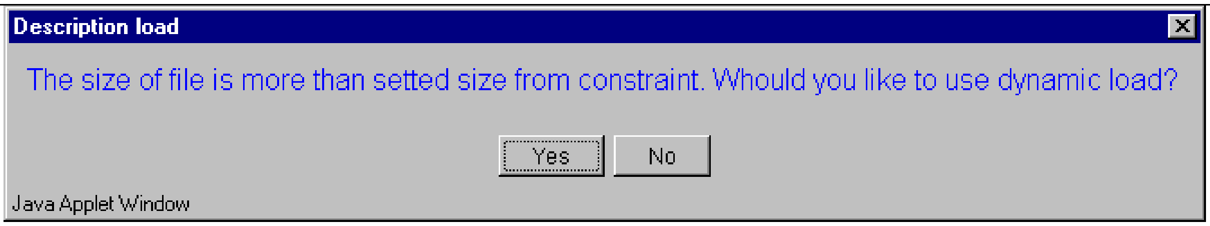


	Name	Size
0	data/human/alon1999_set.seq	3071422
1	data/human/notterman2001_set1.seq	11294087

Figure 2.10.

1. Selected file with genes description. **2.** Confirmation button. **3.** Cancel button.

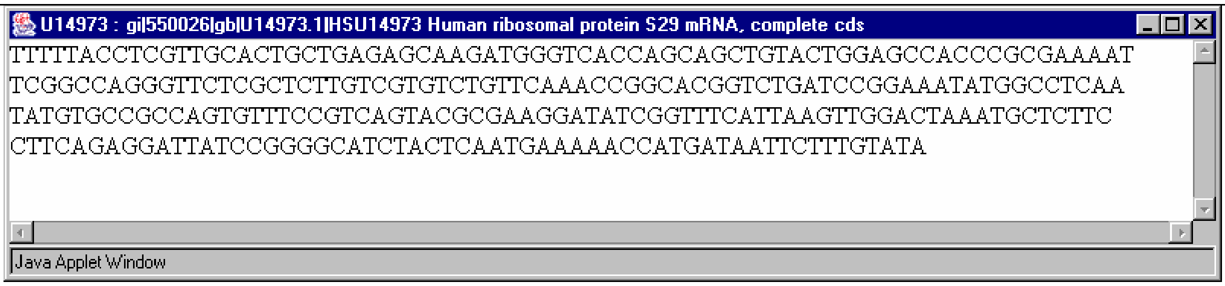
2.11. Further the "Description load" message box with suggestion to use dynamic file loading mode (fig. 2.11) will appear. Press the "Yes" button.



The size of file is more than setted size from constraint. Whould you like to use dynamic load?

Figure 2.11.

2.12. In contextual menu of the application main table (contextual menu can be called out by the right mouse click) the "Show sequence" command will become active (fig. 2.8.1). This command calls out the window with nucleotide sequence of a gene (fig. 2.12).

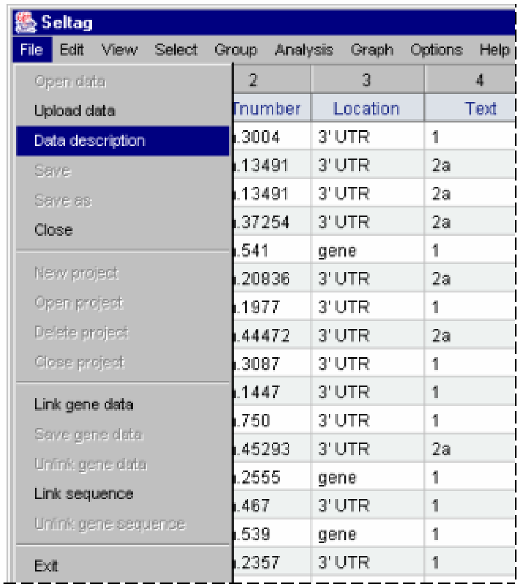


U14973 : gi|550026|gb|U14973.1|HSU14973 Human ribosomal protein S29 mRNA, complete cds

```
TTTTTACCTCGTTGCACTGCTGAGAGCAAGATGGGTCACCAGCAGCTGTACTGGAGCCACCCGCGAAAAT
TCGGCCAGGGTTCTCGCTCTTGTCGTGTCTGTTCAAACCGGCACGGTCTGATCCGGAAATATGGCCTCAA
TATGTGCCGCCAGTGTTTCCGTCAGTACGCGAAGGATATCGGTTTCATTAAGTTGGACTAAATGCTCTTC
CTTCAGAGGATTATCCGGGGCATCTACTCAATGAAAAACCATGATAATTCTTTGTATA
```

Figure 2.12.

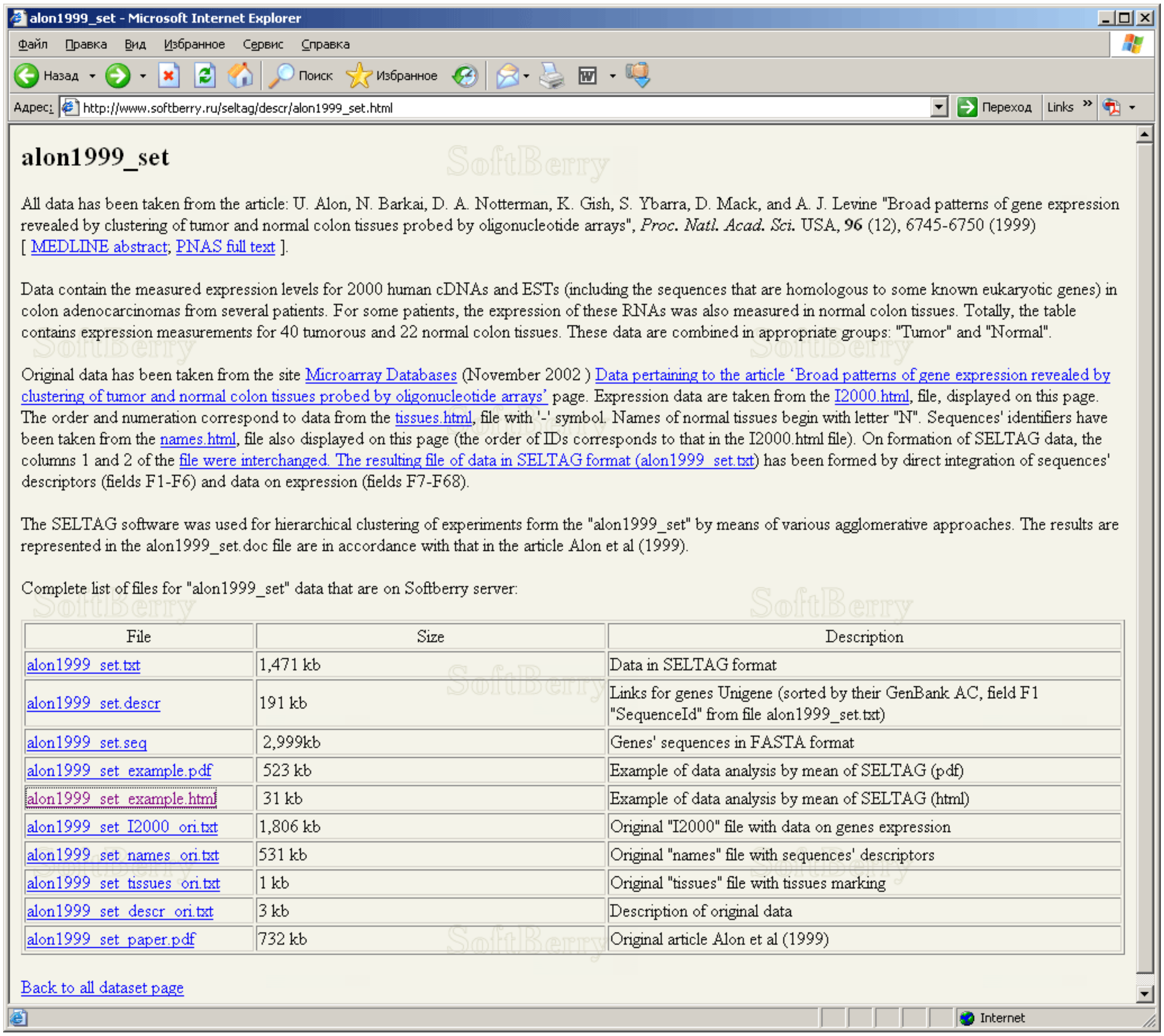
2.13. To retrieve a description of loaded data, select the "File>Data description" command in the main menu (fig. 2.13).



2	3	4
number	Location	Text
.3004	3'UTR	1
.13491	3'UTR	2a
.13491	3'UTR	2a
.37254	3'UTR	2a
.541	gene	1
.20836	3'UTR	2a
.1977	3'UTR	1
.44472	3'UTR	2a
.3087	3'UTR	1
.1447	3'UTR	1
.750	3'UTR	1
.45293	3'UTR	2a
.2555	gene	1
.467	3'UTR	1
.539	gene	1
.2357	3'UTR	1

Figure 2.13.

2.14. It will result in opening of a document with description and list of files for data "alon1999_set" from the Softberry server (fig. 2.14).


alon1999_set - Microsoft Internet Explorer
Файл Правка Вид Избранное Сервис Справка
Назад Поиск Избранное
Адрес: http://www.softberry.ru/seltag/descr/alon1999_set.html
Переход Links
Internet

alon1999_set

All data has been taken from the article: U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine "Broad patterns of gene expression revealed by clustering of tumor and normal colon tissues probed by oligonucleotide arrays", *Proc. Natl. Acad. Sci.* USA, **96** (12), 6745-6750 (1999)
[MEDLINE abstract; PNAS full text].

Data contain the measured expression levels for 2000 human cDNAs and ESTs (including the sequences that are homologous to some known eukaryotic genes) in colon adenocarcinomas from several patients. For some patients, the expression of these RNAs was also measured in normal colon tissues. Totally, the table contains expression measurements for 40 tumorous and 22 normal colon tissues. These data are combined in appropriate groups: "Tumor" and "Normal".

Original data has been taken from the site Microarray Databases (November 2002) Data pertaining to the article ‘Broad patterns of gene expression revealed by clustering of tumor and normal colon tissues probed by oligonucleotide arrays’ page. Expression data are taken from the I2000.html, file, displayed on this page. The order and numeration correspond to data from the tissues.html, file with '-' symbol. Names of normal tissues begin with letter "N". Sequences' identifiers have been taken from the names.html, file also displayed on this page (the order of IDs corresponds to that in the I2000.html file). On formation of SELTAG data, the columns 1 and 2 of the file were interchanged. The resulting file of data in SELTAG format (alon1999_set.txt) has been formed by direct integration of sequences' descriptors (fields F1-F6) and data on expression (fields F7-F68).

The SELTAG software was used for hierarchical clustering of experiments form the "alon1999_set" by means of various agglomerative approaches. The results are represented in the alon1999_set.doc file are in accordance with that in the article Alon et al (1999).

Complete list of files for "alon1999_set" data that are on Softberry server:

File	Size	Description
alon1999_set.txt	1,471 kb	Data in SELTAG format
alon1999_set.descr	191 kb	Links for genes Unigene (sorted by their GenBank AC, field F1 "SequenceId" from file alon1999_set.txt)
alon1999_set.seq	2,999kb	Genes' sequences in FASTA format
alon1999_set_example.pdf	523 kb	Example of data analysis by mean of SELTAG (pdf)
alon1999_set_example.html	31 kb	Example of data analysis by mean of SELTAG (html)
alon1999_set_I2000_ori.txt	1,806 kb	Original "I2000" file with data on genes expression
alon1999_set_names_ori.txt	531 kb	Original "names" file with sequences' descriptors
alon1999_set_tissues_ori.txt	1 kb	Original "tissues" file with tissues marking
alon1999_set_descr_ori.txt	3 kb	Description of original data
alon1999_set_paper.pdf	732 kb	Original article Alon et al (1999)

Back to all dataset page

Figure 2.14.

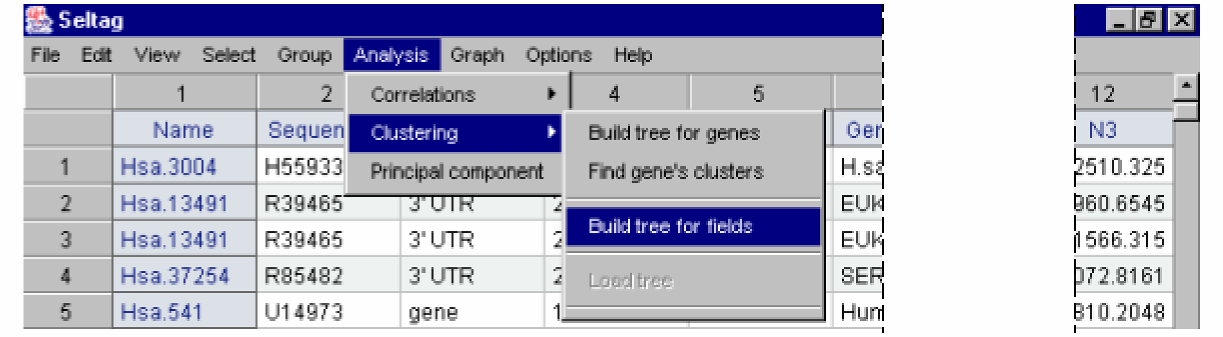
3. Hierarchical clustering of experiments.

One of the approaches for revealing clusters of genes with similar profiles of expression is the method of hierarchical clustering [2]. Such an analysis is based on building of binary tree for experiments by defined metrics of distances between them. Each knot of a tree connects two child knots, lengths of branches correspond to distances between expression profiles in experiments.

This chapter contains description of the building of trees for fields with use of various methods of hierarchical clustering as well as comparison of obtained clustering results with original ones [1].

To perform the claimed task it is necessary to do the following:

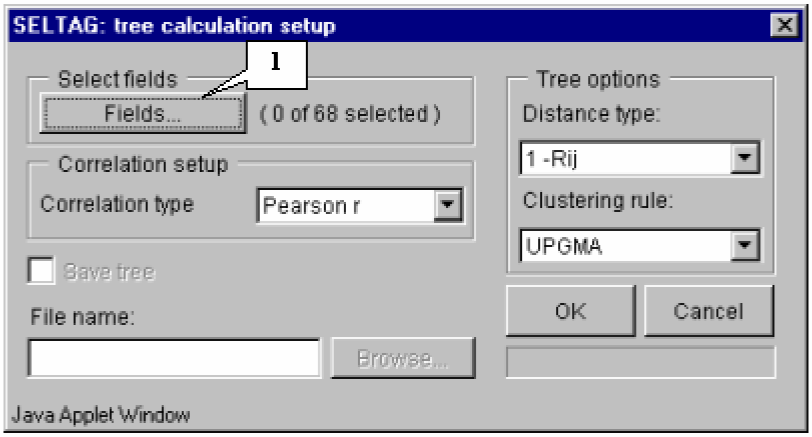
3.1. Select the “Clustering>Build tree for fields” command in the main menu (fig. 3.1).



	1	2		4	5		12
	Name	Sequen				Ger	N3
1	Hsa.3004	H55933				H.sa	2510.325
2	Hsa.13491	R39465	3'UTR	2		EUK	960.6545
3	Hsa.13491	R39465	3'UTR	2		EUK	1566.315
4	Hsa.37254	R85482	3'UTR	2		SER	072.8161
5	Hsa.541	U14973	gene	1		Hum	810.2048

Figure 3.1.

3.2. It will cause opening of the “Tree calculation setup” dialog (fig. 3.2). For the beginning, choose fields that will be used for calculation. To do this press the “Fields” button (fig. 3.2).



(0 of 68 selected)

Correlation type: Pearson r

□ Save tree

File name:

Distance type: 1 -Rij

Clustering rule: UPGMA

Figure 3.2.

1. Button for opening of the “Field selection” dialog window.

3.3. The “Field selection” dialog (fig. 3.3) that is purposed for fields selection will appear. In the current example, all experiments are involved in calculation. Press the “Select all experiments” button.

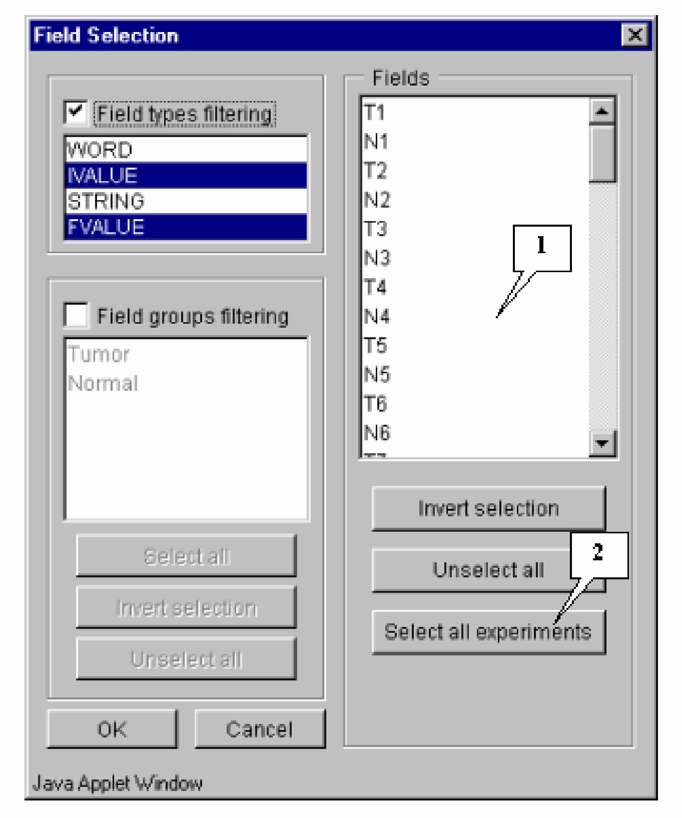
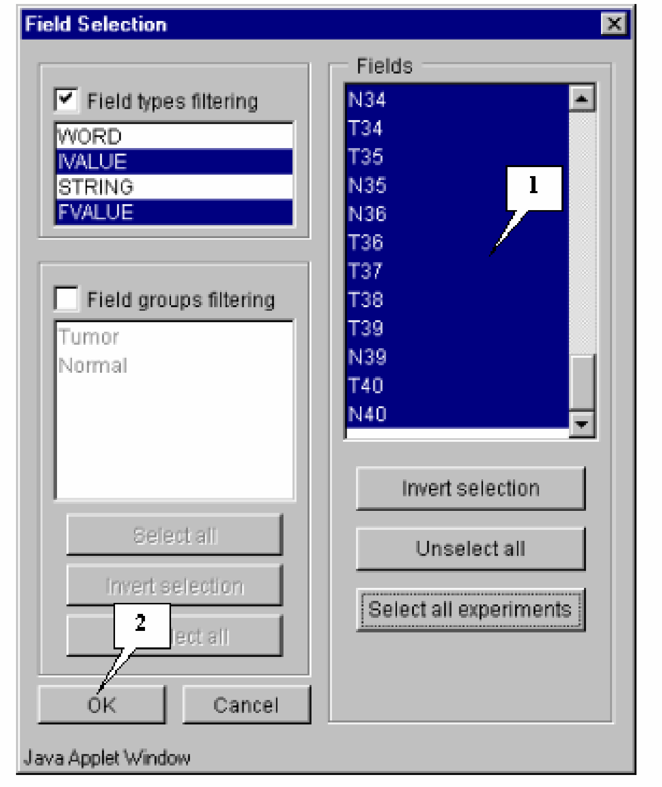


Figure 3.3.

1. Fields selection list. **2.** Button for selecting of all fields with data on experiments.

3.4. It will result in selecting of all fields that contain expression values (fig. 3.4). Press the "OK" button



Field Selection

☑ Field types filtering

WORD
IVALUE
STRING
FVALUE

□ Field groups filtering

Tumor
Normal

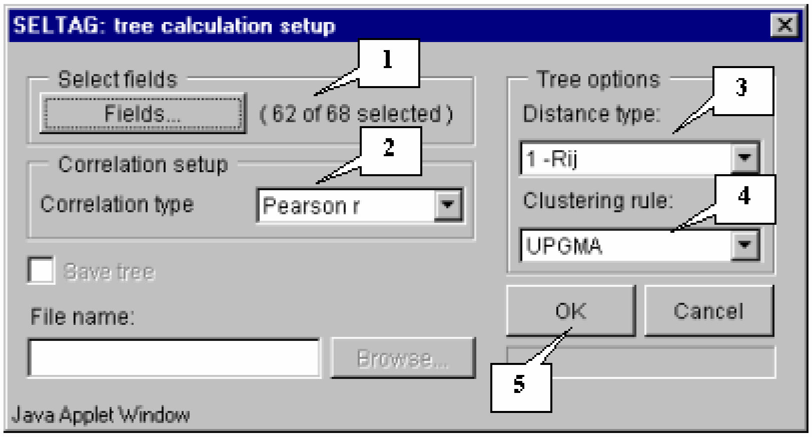
Fields

N34
T34
T35
N35
N36
T36
T37
T38
T39
N39
T40
N40

Figure 3.4.

1. Selected fields with data on experiments. **2.** Confirmation button.

3.5. In the "Tree calculation setup" dialog, alongside to the "Fields" button, the information on number of selected fields will appear (fig. 3.5).



SELTAG: tree calculation setup

Select fields

Fields... (62 of 68 selected)

Correlation setup

Correlation type: Pearson r

□ Save tree

File name:

Tree options

Distance type: 1 -Rij

Clustering rule: UPGMA

Figure 3.5.

1. Information on number of selected fields. **2.** List of correlation types. **3.** List of distance types. **4.** List of clustering rules. **5.** Confirmation button.

3.6. Further actions are required:

3.6.1. To choose the appropriate variant from the “Correlation type” list (fig. 3.6.1.).

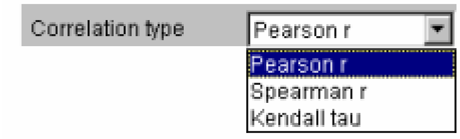


Figure 3.6.1.

3.6.2. To choose the type of distances (that are calculated in dependence on correlation coefficient Rij) from the “Distance type” list (fig. 3.6.2.).

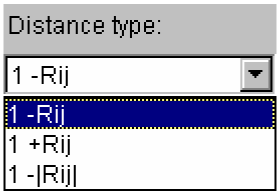


Figure 3.6.2.

3.6.3. To choose the clustering method from the "Amalgamation rule" list (fig. 3.6.3.).

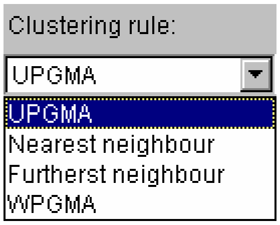


Figure 3.6.3.

3.6.4. To press the “OK” button.
Example of settings is shown on fig. 3.5.

3.7. It will cause the “Tree Diagram” dialog with obtained fields tree diagram to appear. On figures 3.7-3.10 the diagrams obtained with use of different knot binding ways are shown. To build the diagrams the following parameters shown on fig. 3.5 (Pearson’s correlation and 1-Rij type of distance) were used. The figure 3.7 represents the results of using the UPGMA knots binding way, the fig. 3.8 – the WPGMA one, the fig 3.9 – the Furthest neighbor type, and the fig 3.10 – the Nearest neighbor one.

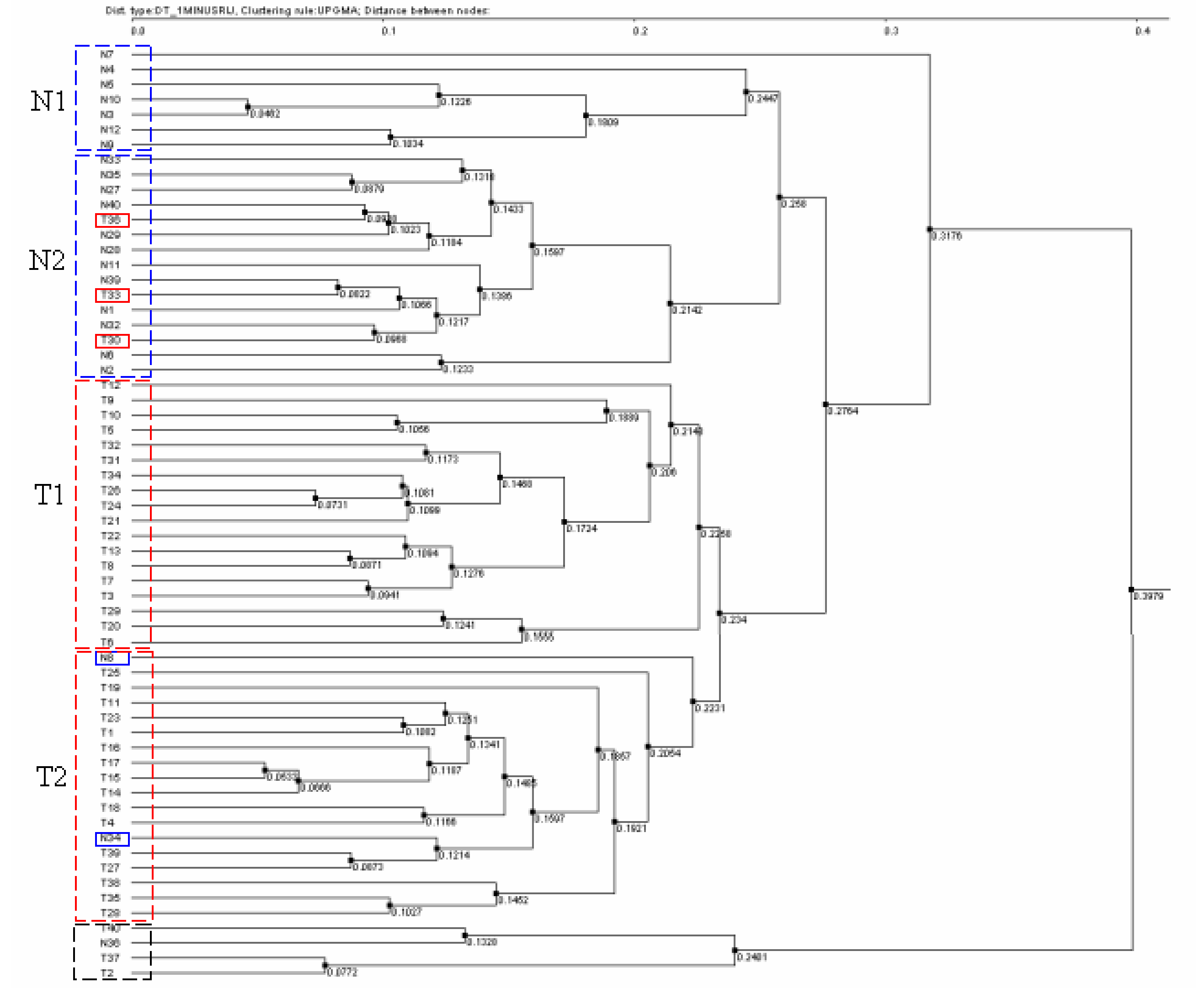

Dist. type:DT_1MINUSRLI, Clustering rule:UPGMA; Distance between nodes:
0.0
0.1
0.2
0.3
0.4
N1
N2
T1
T2
N7
N4
N5
N10
N3
N12
N9
N33
N35
N27
N40
T36
N29
N28
N11
N39
T33
N1
N32
T30
N6
N2
T12
T9
T10
T5
T32
T31
T34
T26
T24
T21
T22
T13
T8
T7
T3
T29
T20
T6
N8
T25
T19
T11
T23
T1
T16
T17
T15
T14
T18
T4
N34
T39
T27
T38
T35
T28
T40
N36
T37
T2
0.0462
0.1226
0.1809
0.1034
0.2447
0.0879
0.1318
0.0938
0.1023
0.1104
0.1433
0.1597
0.0823
0.1066
0.1217
0.1398
0.0968
0.1233
0.2142
0.258
0.3176
0.2764
0.1889
0.1056
0.2148
0.206
0.1173
0.1469
0.1081
0.0731
0.1099
0.1734
0.1094
0.0871
0.1276
0.0941
0.2258
0.1241
0.1555
0.234
0.1251
0.1093
0.1341
0.1107
0.0633
0.0666
0.1485
0.1166
0.1597
0.1214
0.0873
0.1867
0.2054
0.2231
0.1921
0.1452
0.1027
0.1320
0.0772
0.2481
0.3979

Figure 3.7.

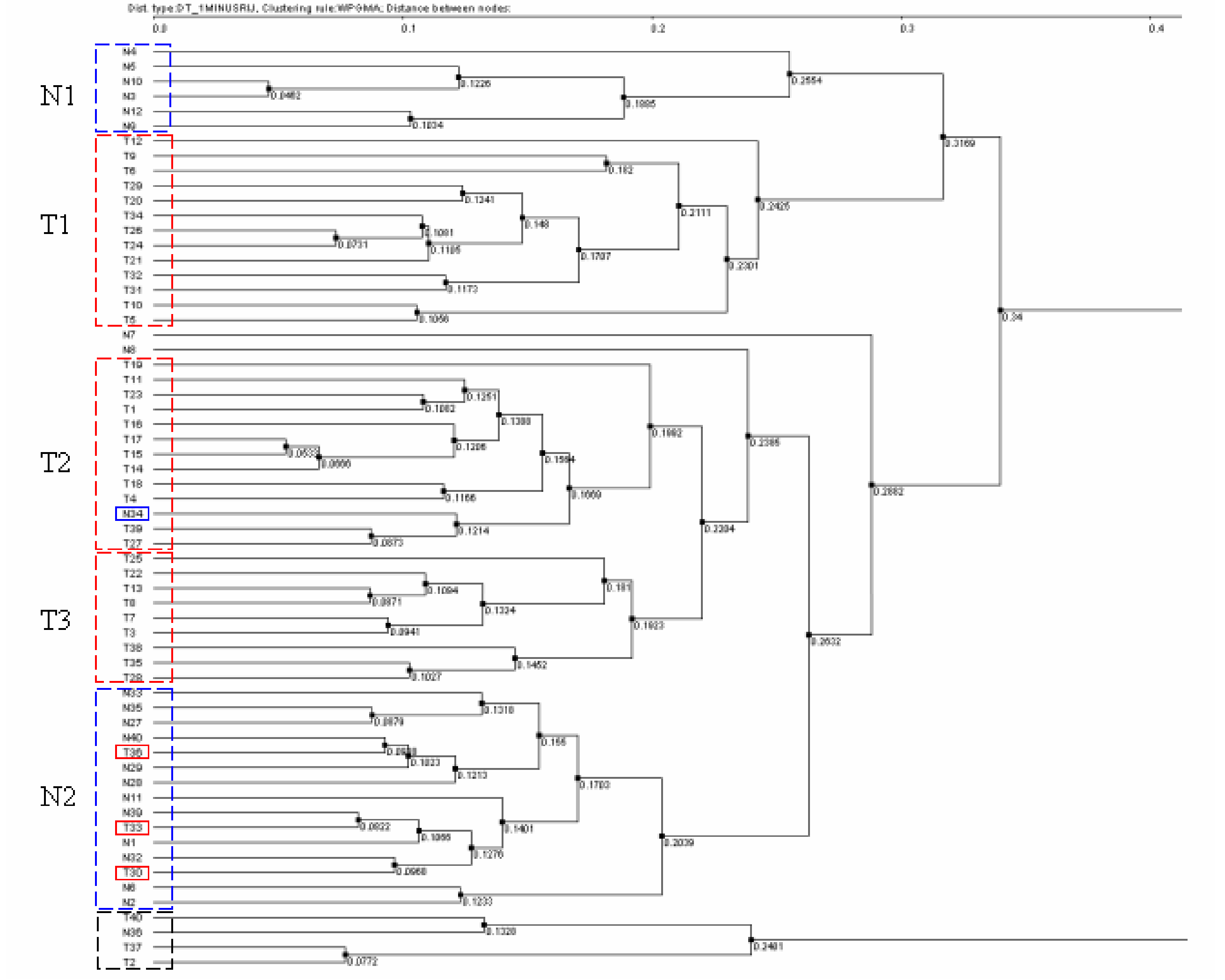

Dist. type:DT_1MINUSRIJ, Clustering rule:WPGMA, Distance between nodes:
0.0
0.1
0.2
0.3
0.4
N1
T1
T2
T3
N2
N4
N5
N10
N3
N12
N9
T12
T9
T6
T29
T20
T34
T26
T24
T21
T32
T31
T10
T5
N7
N8
T19
T11
T23
T1
T16
T17
T15
T14
T18
T4
N34
T39
T27
T25
T22
T13
T8
T7
T3
T38
T35
T28
N33
N35
N27
N40
T36
N29
N28
N11
N39
T33
N1
N32
T30
N6
N2
T40
N36
T37
T2
0.0482
0.1226
0.1034
0.1895
0.2554
0.182
0.1341
0.1081
0.0731
0.1105
0.148
0.1707
0.2111
0.1173
0.1056
0.2301
0.2426
0.3169
0.34
0.1251
0.1082
0.1388
0.1206
0.0633
0.0806
0.1594
0.1166
0.1669
0.1214
0.0873
0.1992
0.2385
0.2204
0.1094
0.0871
0.1324
0.181
0.0941
0.1923
0.1452
0.1027
0.2882
0.1318
0.0879
0.156
0.0848
0.1023
0.1213
0.1703
0.0922
0.1055
0.1401
0.1276
0.0968
0.2039
0.1233
0.2632
0.1328
0.2481
0.0772

Figure 3.8.

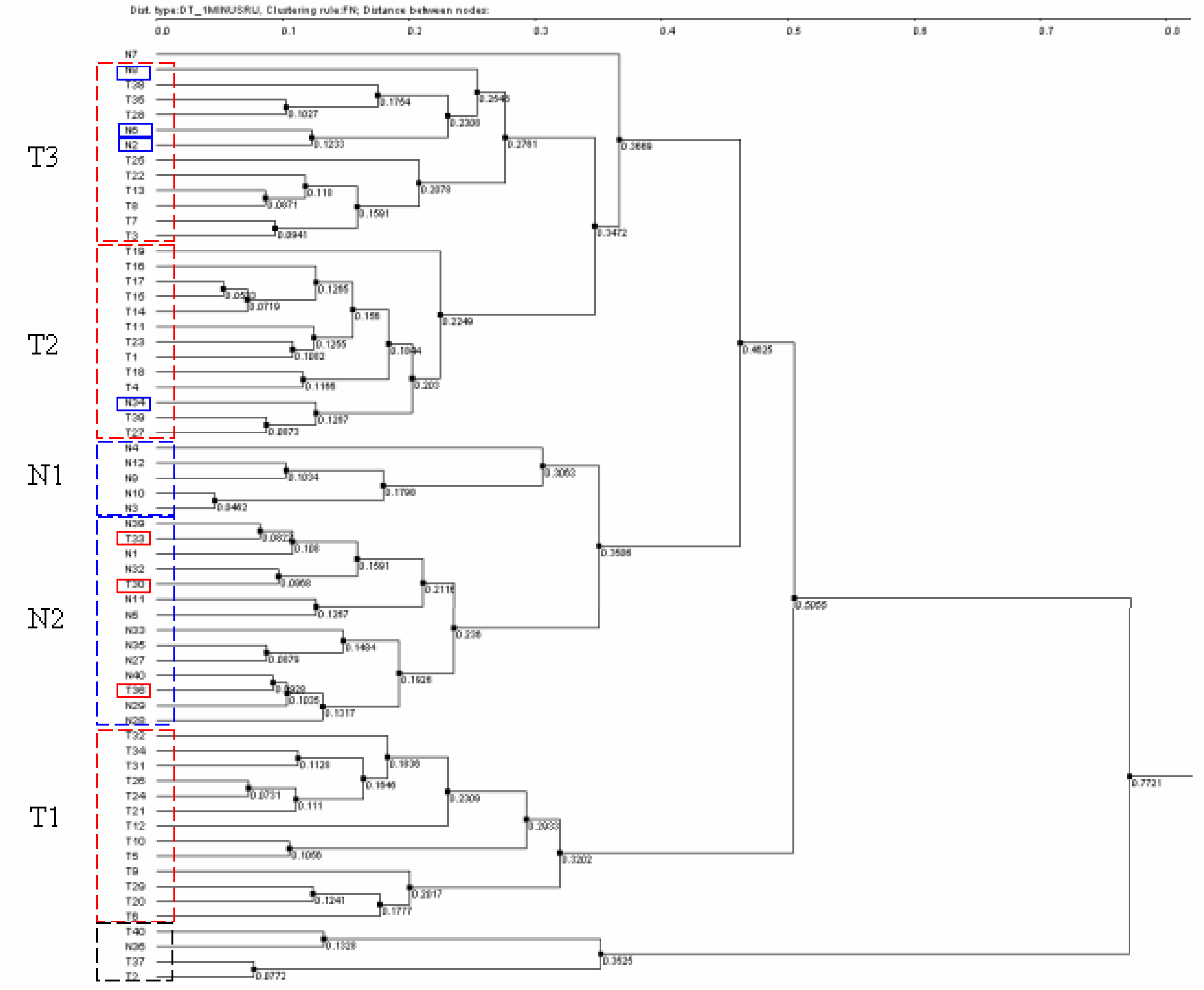

Dist. type:DT_1MINUSRU, Clustering rule:FN; Distance between nodes:
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
T3
T2
N1
N2
T1
N7
N8
T38
T35
T28
N5
N2
T25
T22
T13
T8
T7
T3
T19
T16
T17
T15
T14
T11
T23
T1
T18
T4
N34
T39
T27
N4
N12
N9
N10
N3
N39
T33
N1
N32
T30
N11
N6
N33
N35
N27
N40
T36
N29
N28
T32
T34
T31
T26
T24
T21
T12
T10
T5
T9
T29
T20
T6
T40
N36
T37
T2
0.1754
0.1027
0.2548
0.2308
0.1233
0.2761
0.118
0.0871
0.2878
0.1591
0.0941
0.3669
0.3472
0.1265
0.0533
0.0719
0.158
0.2249
0.1255
0.1082
0.1844
0.1188
0.203
0.1287
0.0873
0.3063
0.1034
0.1798
0.0462
0.4625
0.0822
0.108
0.1591
0.0968
0.2116
0.1267
0.236
0.3506
0.1484
0.0879
0.1926
0.0929
0.1035
0.1317
0.5955
0.1128
0.1838
0.1546
0.0731
0.111
0.2309
0.2933
0.1056
0.3202
0.2017
0.1241
0.1777
0.7721
0.1328
0.3525
0.0772

Figure 3.9.

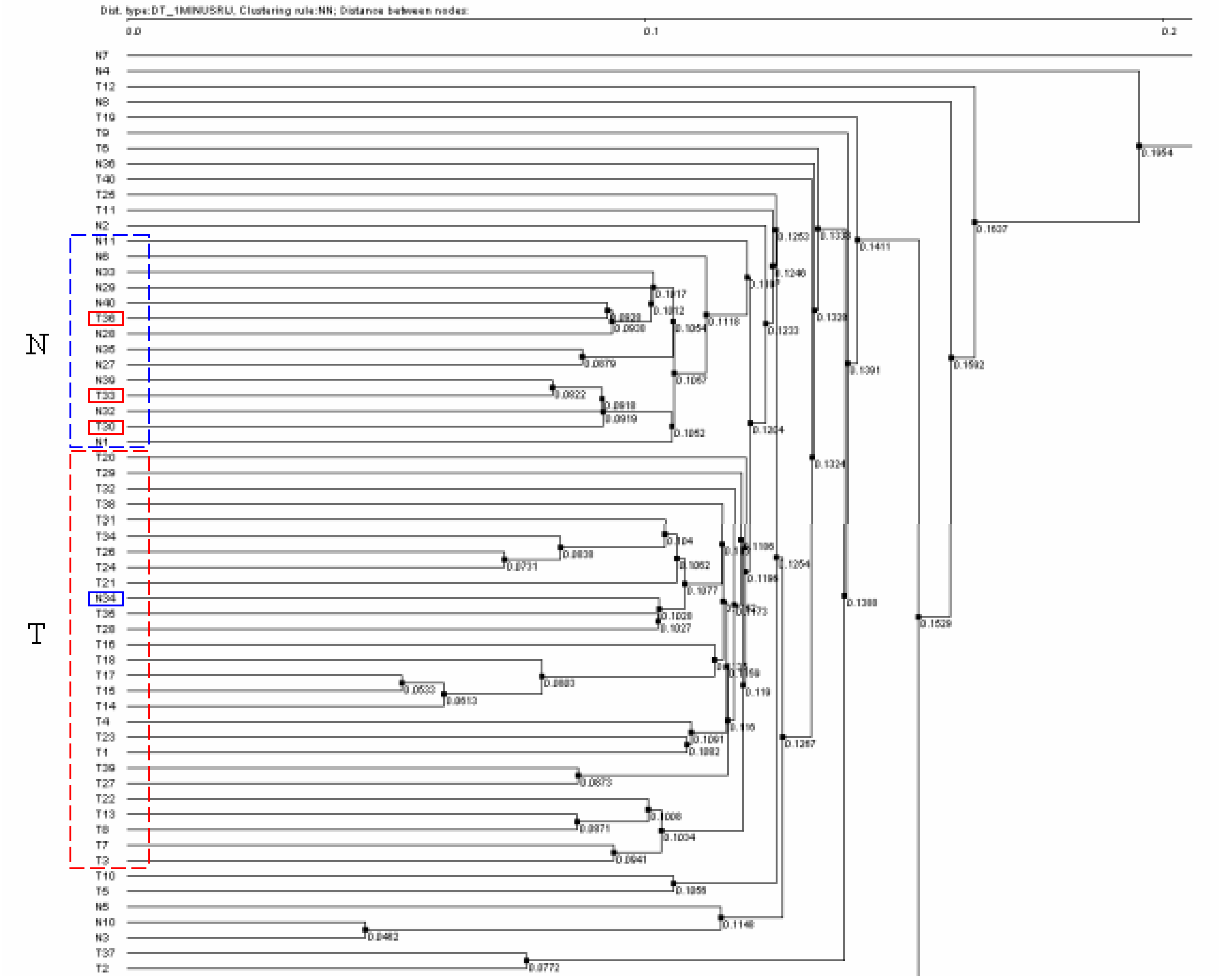


Figure 3.10.

On all diagrams shown the tissues are clearly divided into cancerous and normal ones, and for all diagrams the clusters of normal tissues contain the cancerous tissues T30, T33 and T36, while the clusters of cancerous tissues include the normal ones N34 (for all 4 diagrams) and N8 (for those built with use of Furthest neighbor and UPGMA types) that is in accordance to results obtained by Alon *et al.*, 1999. Three diagrams (UPGMA, WPGMA and Furthest neighbor) contain a small cluster of N36, T2, T37 and T40 tissues that is stably being excluded from common pull.

On comparison of contents for diagrams built with use of UPGMA, WPGMA and Furthest neighbor methods, it is clear that (table 1):

- The first cluster of normal tissues (N1) contains the common for all methods tissues N3, N4, N9, N10 and N12. In the article for this cluster described N3, N4, and N10 ones.
- The second cluster of normal tissues (N2) contains the common for all methods tissues N1, N11, N27, N28, N29, N32, N33, N35, N39, and N40 as well as previously mentioned T30, T33 and T36. In the article for this cluster described N1, N11, N28, N32, N35, N39 and T30, T33 and T36 ones.

- The cluster of cancerous tissues T1 contains the common for all three methods tissues T5, T6, T9, T10, T12, T20, T21, T24, T26, T29, T31, T32 and T34, and content of this cluster is identical for WPGMA and Furthest neighbor methods.
- The cluster of cancerous tissues T2 contains the common for all three methods tissues N34, T1, T4, T11, T14, T15, T16, T17, T18, T19, T23, T27 and T39, and content of this cluster is identical for WPGMA and Furthest neighbor methods.
- The cluster of cancerous tissues T3 contains the common for WPGMA and Furthest neighbor methods tissues T3, T7, T8, T13, T22, T25, T28, T35 and T38.

Thus, the analysis of clusters content by the methods described shows the stability of data clustering process.

Table 1: Content of clusters for methods UPGMA and Furthest neighbor (FN), and for results described in the article. In the “original” columns the order of cluster fields corresponding to tree is shown, the “sort” one contains the fields sorted by numbers. Clusters enumeration corresponds to that on figures 3.8-3.10. The numbers for normal tissues that are included in cancerous cluster, as well as that for cancerous ones that are included in normal cluster, are shown in red.

Cluster name	**paper**		**UPGMA**		**FN**		**WPGMA**	
	original	**sort**	**original**	**sort**	**original**	**sort**	**original**	**sort**
T1	T16	T1	T12	T3	T32	T5	T12	T5
	T28	T4	T9	T5	T34	T6	T9	T6
	T13	T5	T10	T6	T31	T9	T6	T9
	T9	T8	T5	T7	T26	T10	T29	T10
	T21	T9	T32	T8	T24	T12	T20	T12
	T35	T10	T31	T9	T21	T20	T34	T20
	T10	T13	T34	T10	T12	T21	T26	T21
	T27	T15	T26	T12	T10	T24	T24	T24
	T8	T16	T24	T13	T5	T26	T21	T26
	T5	T21	T21	T20	T9	T29	T32	T29
	T4	T26	T22	T21	T29	T31	T31	T31
	T1	T27	T13	T22	T20	T32	T10	T32
	T15	T28	T8	T24	T6	T34	T5	T34
	T26	T35	T7	T26				
			T3	T29				
			T29	T31				
			T20	T32				
			T6	T34				
T2	T17	N34	N8	N8	T19	N34	T19	N34
	T25	T14	T25	N34	T15	T1	T11	T1
	T18	T17	T19	T1	T17	T4	T23	T4
	T23	T18	T11	T4	T16	T11	T1	T11
	T31	T20	T23	T11	T14	T14	T16	T14
	T20	T23	T1	T14	T11	T15	T17	T15
	N34	T24	T16	T15	T23	T16	T15	T16
	T24	T25	T17	T16	T1	T17	T14	T17
	T29	T29	T15	T17	T18	T18	T18	T18
	T38	T31	T14	T18	T4	T19	T4	T19
	T14	T32	T18	T19	N34	T23	N34	T23
	T40	T38	T4	T23	T39	T27	T39	T27
	T32	T40	N34	T25	T27	T39	T27	T39
			T39	T27				
			T27	T28				
			T38	T35				
			T35	T38				
			T28	T39				

T3	T39	N8			N8	N2	T25	T3
	T11	N12			T38	N6	T22	T7
	T6	T3			T35	N8	T13	T8
	T19	T6			T28	T3	T8	T13
	T12	T7			N6	T7	T7	T22
	T22	T11			N2	T8	T3	T25
	T34	T12			T25	T13	T38	T28
	T7	T19			T22	T22	T35	T35
	N8	T22			T13	T25	T28	T38
	T3	T34			T8	T28		
	N12	T39			T7	T35		
					T3	T38		
N1	N4	N2	N7	N3	N4	N3	N4	N3
	N33	N3	N4	N4	N12	N4	N5	N4
	N7	N4	N5	N5	N9	N9	N10	N5
	T37	N5	N10	N7	N10	N10	N3	N9
	N5	N7	N3	N9	N3	N12	N12	N10
	N27	N10	N12	N10			N9	N12
	N3	N27	N9	N12				
	N2	N29						
	N40	N33						
	N36	N36						
	T2	N40						
	N29	T2						
	N10	T37						
N2	N9	N1	N33	N1	N39	N1	N33	N1
	T30	N6	N35	N2	T33	N5	N35	N2
	T36	N9	N27	N6	N1	N11	N27	N6
	N6	N11	N40	N11	N32	N27	N40	N11
	T33	N28	T36	N27	T30	N28	T36	N27
	N11	N32	N29	N28	N11	N29	N29	N28
	N1	N35	N28	N29	N5	N32	N28	N29
	N39	N39	N11	N32	N33	N33	N11	N32
	N28	T30	N39	N33	N35	N35	N39	N33
	N35	T33	T33	N35	N27	N39	T33	N35
	N32	T36	N1	N39	N40	N40	N1	N39
			N32	N40	T36	T30	N32	N40
			T30	T30	N29	T33	T30	T30
			N6	T33	N28	T36	N6	T33
			N2	T36			N2	T36

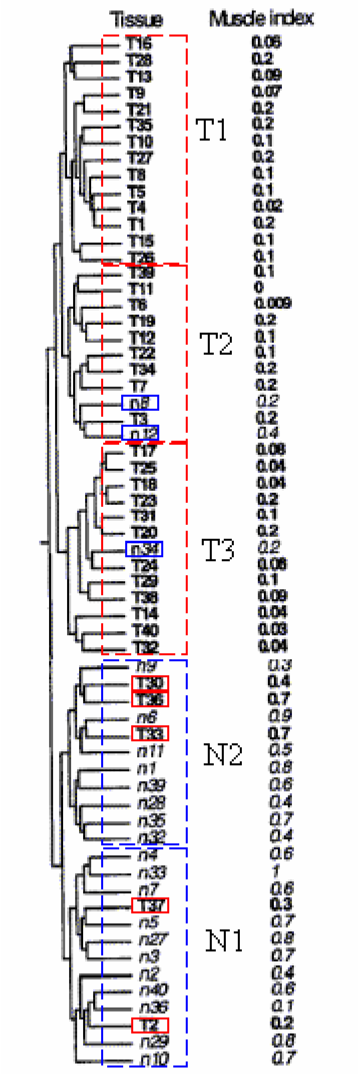

Tissue
Muscle index
T1
T2
T3
N2
N1
T16 0.06
T28 0.2
T13 0.09
T9 0.07
T21 0.2
T35 0.2
T10 0.1
T27 0.2
T8 0.1
T5 0.1
T4 0.02
T1 0.2
T15 0.1
T26 0.1
T39 0.1
T11 0
T6 0.009
T19 0.2
T12 0.1
T22 0.1
T34 0.2
T7 0.2
n8 0.2
T3 0.2
n12 0.4
T17 0.08
T25 0.04
T18 0.04
T23 0.2
T31 0.1
T20 0.2
n34 0.2
T24 0.08
T29 0.1
T38 0.09
T14 0.04
T40 0.03
T32 0.04
n9 0.3
T30 0.4
T36 0.7
n6 0.9
T33 0.7
n11 0.5
n1 0.8
n39 0.6
n28 0.4
n35 0.7
n32 0.4
n4 0.6
n33 1
n7 0.6
T37 0.3
n5 0.7
n27 0.8
n3 0.7
n2 0.4
n40 0.6
n36 0.1
T2 0.2
n29 0.8
n10 0.7

Figure 3.11.

4. References

1. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, and Levine AJ (1999) Broad patterns of gene expression revealed by clustering of tumor and normal colon tissues probed by oligonucleotide arrays, *Proc. Natl. Acad. Sci. US*A, **96**, 6745-6750.
2. Sneath P.H.A., Sokal R.R. (1973) Numerical Taxonomy. The principles and practice of numerical classification. San Francisco: W.H. Freeman and Co.