Biological Sequence Analysis Spring 2009 Dr. Richard Friedman
(212)851-4765 friedman@cancercenter.columbia.edu 824 ICRC

Lesson 13: Functional Genomics II: Functional Genomics Databases and Overrepresentation Analysis and Clustering of Microarray Data

In this class we are going to learn the basics of using web-accessible function and pathway databases, and how to find the functions and pathways associated with a set of differentially expressed genes.

Theory

1. Pevsner on functional databases:
   A. OMIM p. 659-661.
   B. GO 243-246.
   C. KEGG 258-263.

2. Ontools. Read sections on Onto-Express and Pathway-Express.

   http://vortex.cs.wayne.edu/projects.htm

3. Pevsner on clustering or microarray data: p. 203-214.

Theory of Hierarchical Clustering (Unsupervised Learning)

d_{xy}= distance in expression patterns between genes x and y.

$$d_{xy} = 1 - r_{xy} = 1 - \frac{1}{N}\sum_{i=1,N}\left(\frac{x_i}{\sqrt{\sum_{i=1,N} x_i^2}}\frac{y_i}{\sqrt{\sum_{i=1,N} y_i^2}}\right)$$

r_{xy}= modified Pearson correlation between expression of genes x and y with zero mean.
N= Number of samples.

Then a tree is inferred based on distances.

Summary of Commands:

Note: In this document different fonts have different meanings:

Times is used to explain commands.

Courier is used to indicate commands and command options.

Courier italics are used to indicate command parameters, for example, filenames.

Courier bold is used to indicate commands that are not displayed.

Courier bold italics are used to indicate computer-generated output.

Helvetica is used to indicate menu items.

http://www.ncbi.nlm.nih.gov/entrez Accesses the NCBI ENTREZ web page

Gene Diverse information on genes including molecular interactions.

OMIM Descriptive genotype-phenotype relationships in humans.

OMIA Descriptive genotype-phenotype relationships in organisms other than humans.

http://www.geneontology.org/ Gene Ontology Gives Biological Process, Molecular function, Cellular Component for proteins/

Gene symbol/name, exact match To search for a gene/protein
This will give you all of the terms associated with a protein.

Terms To search for a term. Following the links associated with a term will give you a definition of the term and the names of all of the proteins associated with that term.

http://bond.unleashedinformatics.com/Action?
BOND: Biomolecular Object Network Database.
A database that gives interaction partners between biological molecules.

Search bind using a field specific query Enables user to specify the kind of search.

Field: *Interaction description* contains *at least one of the following words*
Generates a list of interactions in which a specified molecule plays a role.

On the list page clicking on an:
interaction identifier gives a description of the interaction in question.

On the interaction page, clicking on:

expand all	gives a more detailed picture of the interaction in question.

http://www.genome.jp/kegg/	KEGG: Kyoto Enyclopedia of Genes and Genomes.
Click on KEGG Table of Contents	
KEGG PATHWAY	Encyclopedia of Pathways.
KEGG GENES	Encyclopedia of Genes.
SEARCH	
KEGG for *subject*	Searches Gene database for name of gene.
GENES for *genename*	Searches Gene database for name of gene.
PATHWAYS for *pathway*	Searches Pathway database for name of pathway.
SEARCH *organismcode* for *genename*	Searches for name of gene in specific organism.
Organism	Searches for organism code.
GENES for *Genename*	Searches Gene database for name of gene.
PATHWAYS for *Pathway*	Searches Pathway database for name of pathway.

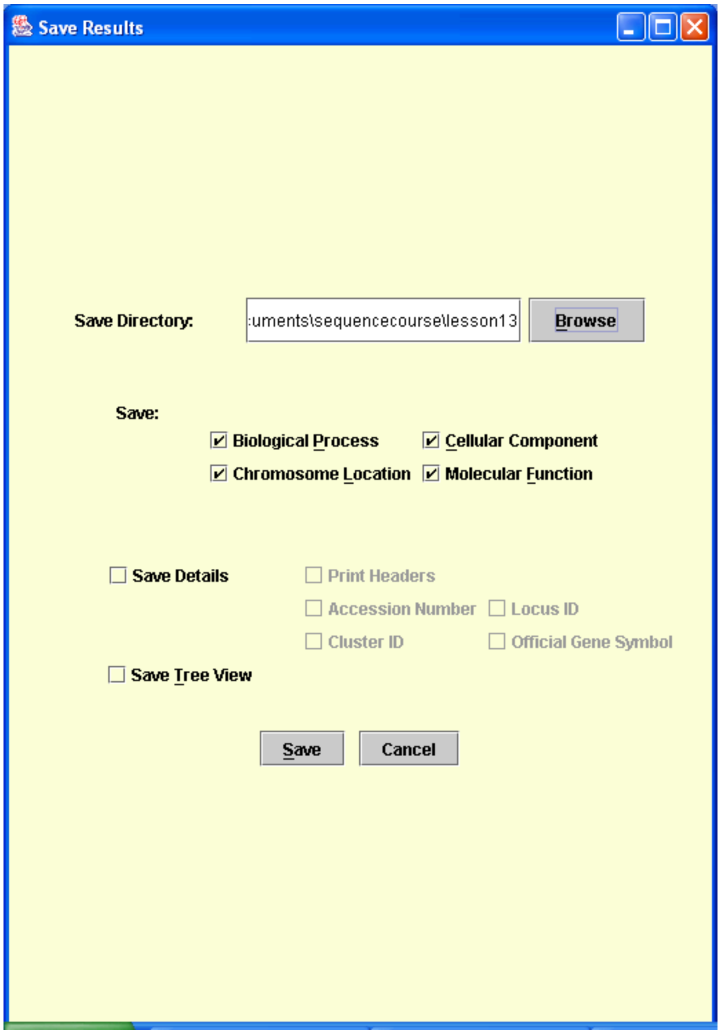
Analysis with Onto-tools

A. Preprocessing for ontools:
   1. Move output files to a directory that you can work in.
   2. Open output files with Microsoft Excel. Autoformat column length and save file as Excel Workbook with new name.
   3. Copy all genes with B >0 to new Excel Workbook and save as Workbook with new name.
   4. Delete all columns but the Affymetrix probeset identifier. Delete the header row. save as text file.

B. Analysis with Onto-Express and Pathway-Express.

http://vortex.cs.wayne.edu/ontoexpress/	Onto-Express web-site.
Login:	Login to web-site.
Onto-express	Select Onto-Express for finding overrepresented Gene Ontology Categories.

Fill out the ontoexpress input window as below (settings for experiment- the parameters will of course vary with other files, organisms, and chip).



Save Directory: :uments\sequencecourse\lesson13

Save:
☑ Biological Process ☑ Cellular Component
☑ Chromosome Location ☑ Molecular Function

□ Save Details
□ Print Headers
□ Accession Number □ Locus ID
□ Cluster ID □ Official Gene Symbol

□ Save Tree View

Onto-express output files:

`bioprocess`	Overrepresented biological process categories.
`molecular`	Overrepresented molecular function categories.
`cellular`	Overrepresented cellular location categories.
`chromosome`	Assignment of genes to chromosomes.

To process output files with Excel:

Open	Opens file from inside Excel.
Files as type *`textfiles`*	Lists text files in the menu.
Delimited	Reads file as entried separated by delimiters.
Tab	Recognizes tabs as delimiters.
Semi-colon	Recognizes semi-colons as delimiters.
Format->Column->Autofit	Fits selected columns.

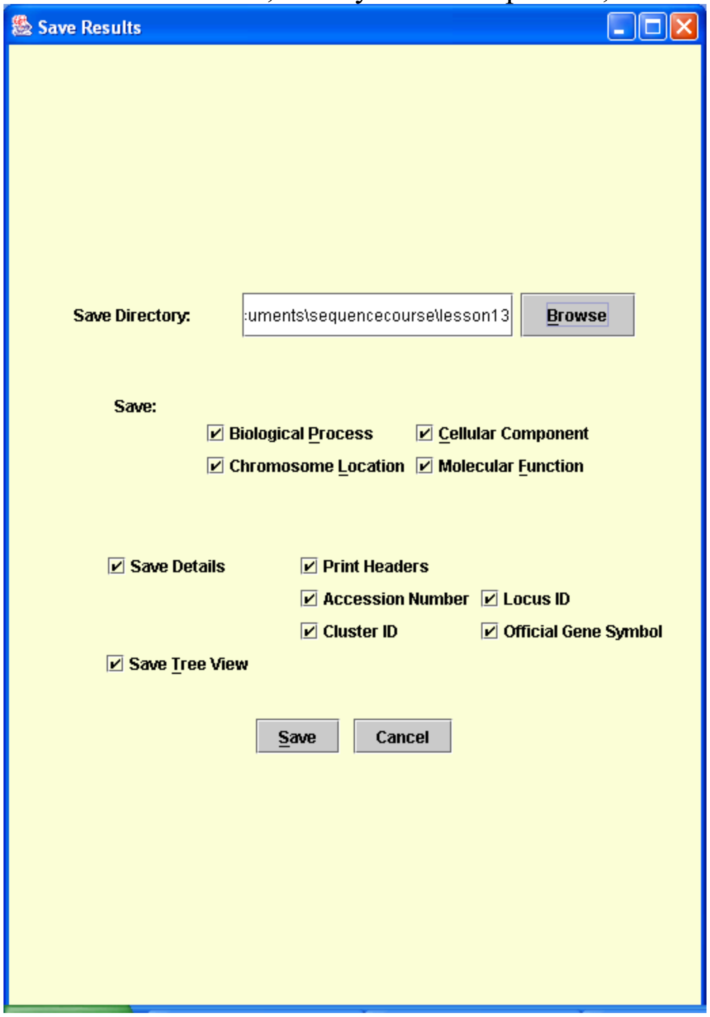
Data->Sort — Sorts by the selected column.

Expand Selection — Sorts rest of row according to the selected column.

File->Save As — Save as file with name typed in box.

Save As type *Microsoft Excel Workbook* — Save as file in Excel Format.

1. Open each of these files with Excel (check delimited and tab and semicolon).Autofit selection for column, sort by corrected p-value, and save as a Workbook.



Save Directory: :uments\sequencecourse\lesson13

Save:

☑ Biological Process ☑ Cellular Component

☑ Chromosome Location ☑ Molecular Function

☑ Save Details ☑ Print Headers

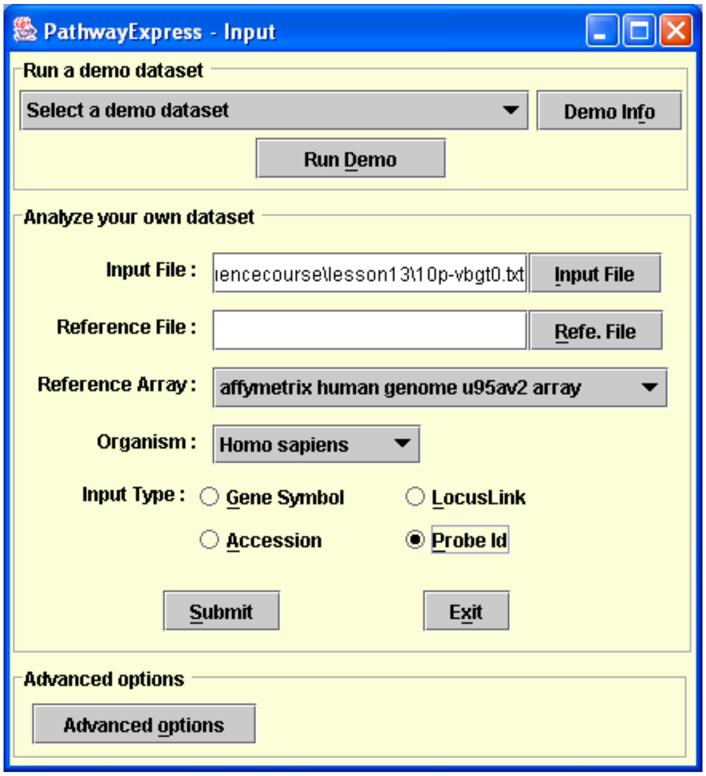
☑ Accession Number ☑ Locus ID

☑ Cluster ID ☑ Official Gene Symbol

☑ Save Tree View

The above box will output the same 4 files as before only with the genes included. In addition, it will output, “treeview”, a file with all of the gene ontology categories detected and “treeview_input”, a file with all of the gene-ontology categories detected along with input genes from the list. The above files cannot be sorted.by corrected p-values without getting mixed up, because there is more than one row per Gene Ontology category.

Pathway-Express — Select Pathway-Express for finding overrepresented KEGG pathwayCategories. Use list of pathways with log fold changes for input.



Select a demo dataset

Input File : iencecourse\lesson13\10p-vbgt0.txt

Reference File :

Reference Array : affymetrix human genome u95av2 array

Organism : Homo sapiens

Input Type : ○ Gene Symbol ○ LocusLink ○ Accession ● Probe Id

Advanced options

Corrections = fdr — Corrects pathways for false discoveries with false discovery rate.

Pathway express output-windows and options

Bar Graph — Display of overrepresented pathways..

Pathway Details — Shows list of pathways in tables.

(Right mouse button over pathway link) — Lists pathway options.

Show Pathway Genes — Changes pathway in Pathway Gene Details.

Show Pathway Details — Shows KEGG diagram of pathways with genes on list indicated.

Save Table — Saves Pathway Details file.

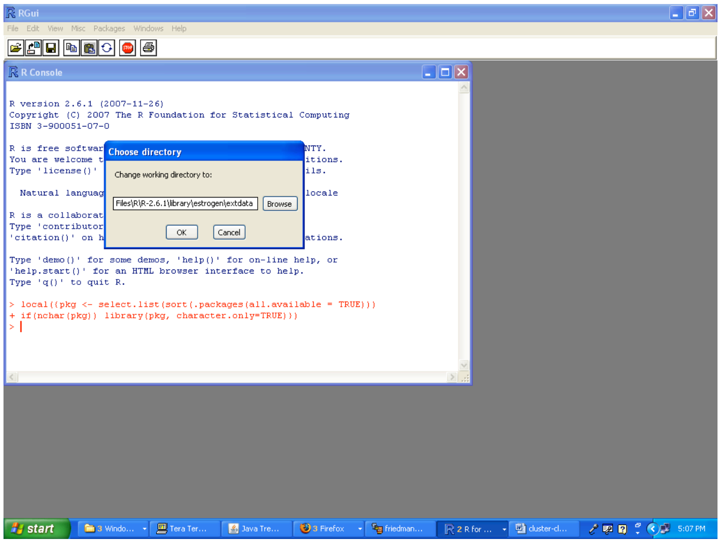
Pathway Gene Details — Shows list of genes in selected pathway and indicates whether or not they are part of the input list.

(Right mouse button over genelink) — Lists gene options options, which are analogous to Pathway Details options.

Input Details — Shows list of genes in input lists and indicates whether or not they are part of a pathway. Mouse options analogous to other windows.

Hierarchical clustering with Cluster 3.0 and JavaTreeview.

1. Start R and go to the estrogen working directory.

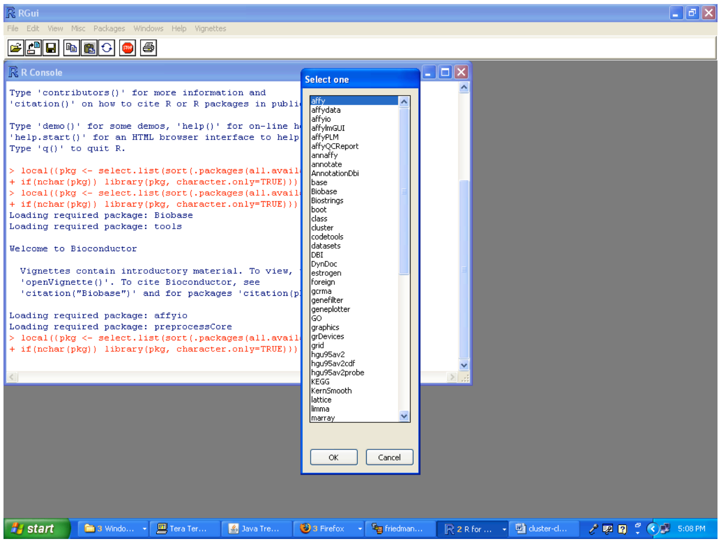


```
R version 2.6.1 (2007-11-26)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free softwar                                        NTY.
You are welcome t                                        itions.
Type 'license()'                                         ils.

  Natural languag                                        locale

R is a collaborat
Type 'contributor
'citation()' on h                                        ations.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> local({pkg <- select.list(sort(.packages(all.available = TRUE)))
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
> 
```

Change working directory to:

Files\R\R-2.6.1\library\estrogen\extdata

2. Load the Affy Program:



```
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publi

Type 'demo()' for some demos, 'help()' for on-line h
'help.start()' for an HTML browser interface to help
Type 'q()' to quit R.

> local({pkg <- select.list(sort(.packages(all.avail
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
> local({pkg <- select.list(sort(.packages(all.avail
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
Loading required package: Biobase
Loading required package: tools

Welcome to Bioconductor

  Vignettes contain introductory material. To view,
  'openVignette()'. To cite Bioconductor, see
  'citation("Biobase")' and for packages 'citation(p

Loading required package: affyio
Loading required package: preprocessCore
> local({pkg <- select.list(sort(.packages(all.avail
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
```

3. In R window:

```
estrogenEset<-ReadAffy()
```

Loads Estrogen cell files into Expression object.

```
estrogenmas5 <-mas5(estrogenEset)
```

Normalizes cel files by the MAS5 algorithm.

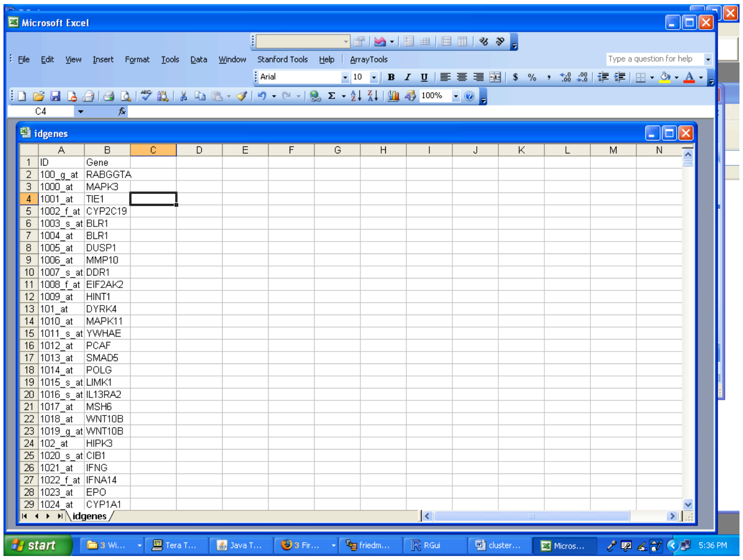
```
write.exprs(estrogenmas5,"estrogenmas5.txt")
```

Saves normalized arrays as text file.

4. Cut and paste file to directory which you are working in for lab.

5. Copy file with differential expression values from last week to file called id_gene.
   A, Open it with excel.
   B. Delete all columns other than Gene-and-symbol.
   C. Sort by ID.
   D. Rename Symbol Column "Gene".

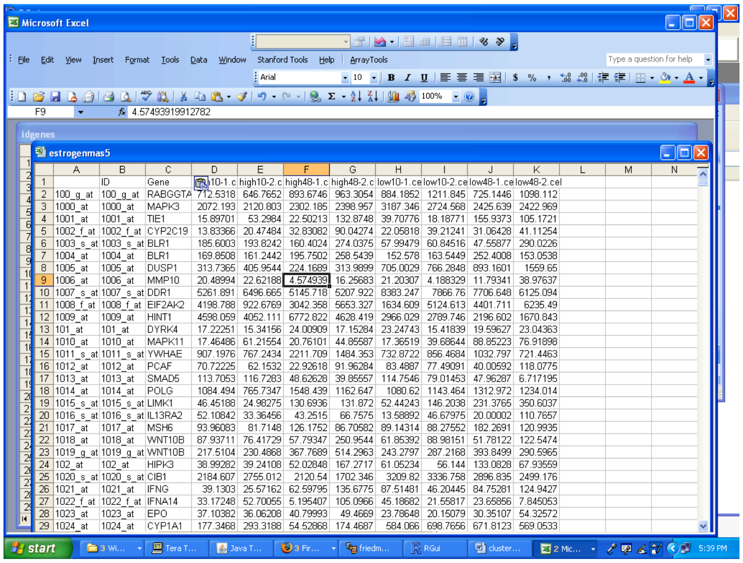
The spreadsheet should look like this:



ID	Gene
100_g_at	RABGGTA
1000_at	MAPK3
1001_at	TIE1
1002_f_at	CYP2C19
1003_s_at	BLR1
1004_at	BLR1
1005_at	DUSP1
1006_at	MMP10
1007_s_at	DDR1
1008_f_at	EIF2AK2
1009_at	HINT1
101_at	DYRK4
1010_at	MAPK11
1011_s_at	YWHAE
1012_at	PCAF
1013_at	SMAD5
1014_at	POLG
1015_s_at	LIMK1
1016_s_at	IL13RA2
1017_at	MSH6
1018_at	WNT10B
1019_g_at	WNT10B
102_at	HIPK3
1020_s_at	CIB1
1021_at	IFNG
1022_f_at	IFNA14
1023_at	EPO
1024_at	CYP1A1

6. Open Estrogenmas5 within Excel.
   A. Make 2 new columns.
   B. Paste in the contents of ID gene so that the id column should be aligned

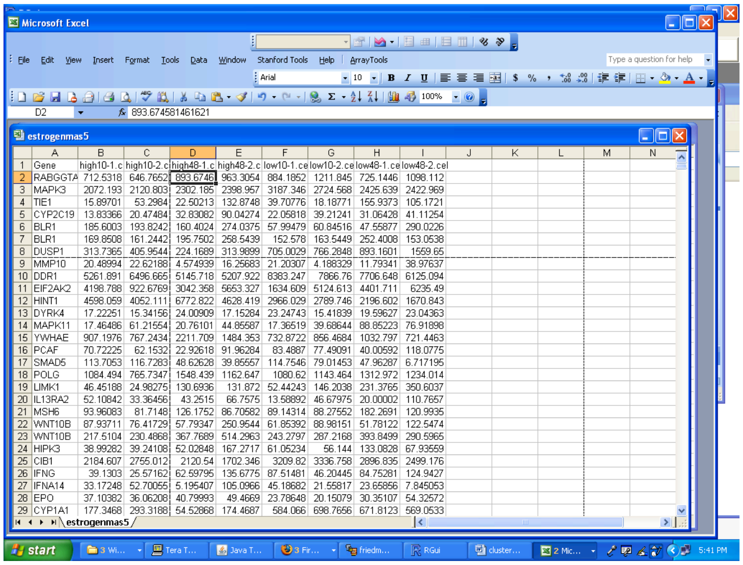
The file should look like this:


Microsoft Excel
File Edit View Insert Format Tools Data Window Stanford Tools Help ArrayTools
Type a question for help
Arial 10
100%
idgenes
estrogenmas5
start 3 Wi… Tera T… Java T… 3 Fir… friedm… RGui cluster… 2 Mic… 5:39 PM

F9 4.57493919912782

	A	B	C	D	E	F	G	H	I	J	K
1		ID	Gene	h10-1.c	high10-2.c	high48-1.c	high48-2.c	low10-1.ce	low10-2.ce	low48-1.ce	low48-2.cel
2	100_g_at	100_g_at	RABGGTA	712.5318	646.7652	893.6746	963.3054	884.1852	1211.845	725.1446	1098.112
3	1000_at	1000_at	MAPK3	2072.193	2120.803	2302.185	2398.957	3187.346	2724.568	2425.639	2422.969
4	1001_at	1001_at	TIE1	15.89701	53.2984	22.50213	132.8748	39.70776	18.18771	155.9373	105.1721
5	1002_f_at	1002_f_at	CYP2C19	13.83366	20.47484	32.83082	90.04274	22.05818	39.21241	31.06428	41.11254
6	1003_s_at	1003_s_at	BLR1	185.6003	193.8242	160.4024	274.0375	57.99479	60.84516	47.55877	290.0226
7	1004_at	1004_at	BLR1	169.8508	161.2442	195.7502	258.5439	152.578	163.5449	252.4008	153.0538
8	1005_at	1005_at	DUSP1	313.7365	405.9544	224.1689	313.9899	705.0029	766.2848	893.1601	1559.65
9	1006_at	1006_at	MMP10	20.48994	22.62188	4.574939	16.25683	21.20307	4.188329	11.79341	38.97637
10	1007_s_at	1007_s_at	DDR1	5261.891	6496.665	5145.718	5207.922	8383.247	7866.76	7706.648	6125.094
11	1008_f_at	1008_f_at	EIF2AK2	4198.788	922.6769	3042.358	5653.327	1634.609	5124.613	4401.711	6235.49
12	1009_at	1009_at	HINT1	4598.059	4052.111	6772.822	4628.419	2966.029	2789.746	2196.602	1670.843
13	101_at	101_at	DYRK4	17.22251	15.34156	24.00909	17.15284	23.24743	15.41839	19.59627	23.04363
14	1010_at	1010_at	MAPK11	17.46486	61.21554	20.76101	44.85587	17.36519	39.68644	88.85223	76.91898
15	1011_s_at	1011_s_at	YWHAE	907.1976	767.2434	2211.709	1484.353	732.8722	856.4684	1032.797	721.4463
16	1012_at	1012_at	PCAF	70.72225	62.1532	22.92618	91.96284	83.4887	77.49091	40.00592	118.0775
17	1013_at	1013_at	SMAD5	113.7053	116.7283	48.62628	39.85557	114.7546	79.01453	47.96287	6.717195
18	1014_at	1014_at	POLG	1084.494	765.7347	1548.439	1162.647	1080.62	1143.464	1312.972	1234.014
19	1015_s_at	1015_s_at	LIMK1	46.45188	24.98275	130.6936	131.872	52.44243	146.2038	231.3765	350.6037
20	1016_s_at	1016_s_at	IL13RA2	52.10842	33.36456	43.2515	66.7575	13.58892	46.67975	20.00002	110.7657
21	1017_at	1017_at	MSH6	93.96083	81.7148	126.1752	86.70582	89.14314	88.27552	182.2691	120.9935
22	1018_at	1018_at	WNT10B	87.93711	76.41729	57.79347	250.9544	61.85392	88.98151	51.78122	122.5474
23	1019_g_at	1019_g_at	WNT10B	217.5104	230.4868	367.7689	514.2963	243.2797	287.2168	393.8499	290.5965
24	102_at	102_at	HIPK3	38.99282	39.24108	52.02848	167.2717	61.05234	56.144	133.0828	67.93559
25	1020_s_at	1020_s_at	CIB1	2184.607	2755.012	2120.54	1702.346	3209.82	3336.758	2896.835	2499.176
26	1021_at	1021_at	IFNG	39.1303	25.57162	62.59795	135.6775	87.51481	46.20445	84.75281	124.9427
27	1022_f_at	1022_f_at	IFNA14	33.17248	52.70055	5.195407	105.0966	45.18682	21.55817	23.65856	7.845053
28	1023_at	1023_at	EPO	37.10382	36.06208	40.79993	49.4669	23.78648	20.15079	30.35107	54.32572
29	1024_at	1024_at	CYP1A1	177.3468	293.3188	54.52868	174.4687	584.066	698.7656	671.8123	569.0533

7. Delete the 2 id columns so that the spreadsheet looks like this:

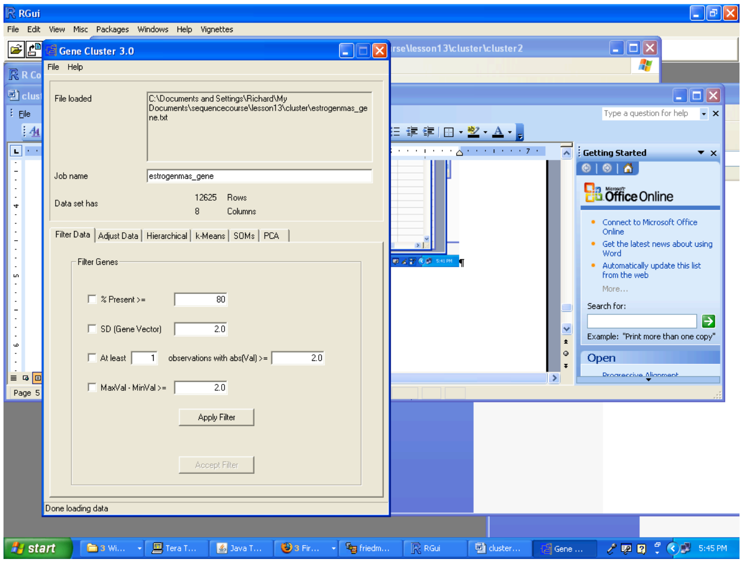

Microsoft Excel
File Edit View Insert Format Tools Data Window Stanford Tools Help ArrayTools
Type a question for help
Arial 10
100%
estrogenmas5
estrogenmas5
start 3 Wi… Tera T… Java T… 3 Fir… friedm… RGui cluster… 2 Mic… 5:41 PM

D2 893.674581461621

	A	B	C	D	E	F	G	H	I
1	Gene	high10-1.c	high10-2.c	high48-1.c	high48-2.c	low10-1.ce	low10-2.ce	low48-1.ce	low48-2.cel
2	RABGGTA	712.5318	646.7652	893.6746	963.3054	884.1852	1211.845	725.1446	1098.112
3	MAPK3	2072.193	2120.803	2302.185	2398.957	3187.346	2724.568	2425.639	2422.969
4	TIE1	15.89701	53.2984	22.50213	132.8748	39.70776	18.18771	155.9373	105.1721
5	CYP2C19	13.83366	20.47484	32.83082	90.04274	22.05818	39.21241	31.06428	41.11254
6	BLR1	185.6003	193.8242	160.4024	274.0375	57.99479	60.84516	47.55877	290.0226
7	BLR1	169.8508	161.2442	195.7502	258.5439	152.578	163.5449	252.4008	153.0538
8	DUSP1	313.7365	405.9544	224.1689	313.9899	705.0029	766.2848	893.1601	1559.65
9	MMP10	20.48994	22.62188	4.574939	16.25683	21.20307	4.188329	11.79341	38.97637
10	DDR1	5261.891	6496.665	5145.718	5207.922	8383.247	7866.76	7706.648	6125.094
11	EIF2AK2	4198.788	922.6769	3042.358	5653.327	1634.609	5124.613	4401.711	6235.49
12	HINT1	4598.059	4052.111	6772.822	4628.419	2966.029	2789.746	2196.602	1670.843
13	DYRK4	17.22251	15.34156	24.00909	17.15284	23.24743	15.41839	19.59627	23.04363
14	MAPK11	17.46486	61.21554	20.76101	44.85587	17.36519	39.68644	88.85223	76.91898
15	YWHAE	907.1976	767.2434	2211.709	1484.353	732.8722	856.4684	1032.797	721.4463
16	PCAF	70.72225	62.1532	22.92618	91.96284	83.4887	77.49091	40.00592	118.0775
17	SMAD5	113.7053	116.7283	48.62628	39.85557	114.7546	79.01453	47.96287	6.717195
18	POLG	1084.494	765.7347	1548.439	1162.647	1080.62	1143.464	1312.972	1234.014
19	LIMK1	46.45188	24.98275	130.6936	131.872	52.44243	146.2038	231.3765	350.6037
20	IL13RA2	52.10842	33.36456	43.2515	66.7575	13.58892	46.67975	20.00002	110.7657
21	MSH6	93.96083	81.7148	126.1752	86.70582	89.14314	88.27552	182.2691	120.9935
22	WNT10B	87.93711	76.41729	57.79347	250.9544	61.85392	88.98151	51.78122	122.5474
23	WNT10B	217.5104	230.4868	367.7689	514.2963	243.2797	287.2168	393.8499	290.5965
24	HIPK3	38.99282	39.24108	52.02848	167.2717	61.05234	56.144	133.0828	67.93559
25	CIB1	2184.607	2755.012	2120.54	1702.346	3209.82	3336.758	2896.835	2499.176
26	IFNG	39.1303	25.57162	62.59795	135.6775	87.51481	46.20445	84.75281	124.9427
27	IFNA14	33.17248	52.70055	5.195407	105.0966	45.18682	21.55817	23.65856	7.845053
28	EPO	37.10382	36.06208	40.79993	49.4669	23.78648	20.15079	30.35107	54.32572
29	CYP1A1	177.3468	293.3188	54.52868	174.4687	584.066	698.7656	671.8123	569.0533

8. Save file as text file.called “estrogenmas5gene”

9. Open Cluster 3.0 on the PC and load mas5gene



Gene Cluster 3.0

File loaded: C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\estrogenmas_gene.txt

Job name: estrogenmas_gene

Data set has: 12625 Rows, 8 Columns

Filter Genes

- □ % Present >= 80
- □ SD (Gene Vector) 2.0
- □ At least 1 observations with abs(Val) >= 2.0
- □ MaxVal - MinVal >= 2.0

Done loading data

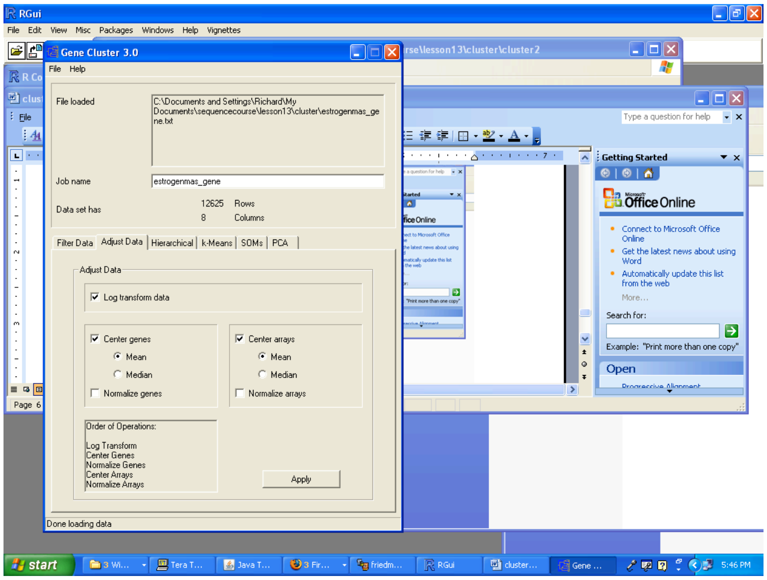
RGui
File Edit View Misc Packages Windows Help Vignettes
File Help
Filter Data
Adjust Data
Hierarchical
k-Means
SOMs
PCA
Apply Filter
Accept Filter
Getting Started
Office Online
Connect to Microsoft Office Online
Get the latest news about using Word
Automatically update this list from the web
More...
Search for:
Example: "Print more than one copy"
Open
start
5:45 PM

10. Click adjust data. Check

A. Log Transform

B. Center Genes

C. Center Arrays



Gene Cluster 3.0

File loaded: C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\estrogenmas_gene.txt

Job name: estrogenmas_gene

Data set has: 12625 Rows, 8 Columns

Adjust Data

- ☑ Log transform data
- ☑ Center genes: ● Mean ○ Median
- □ Normalize genes
- ☑ Center arrays: ● Mean ○ Median
- □ Normalize arrays

Order of Operations:
Log Transform
Center Genes
Normalize Genes
Center Arrays
Normalize Arrays

Done loading data

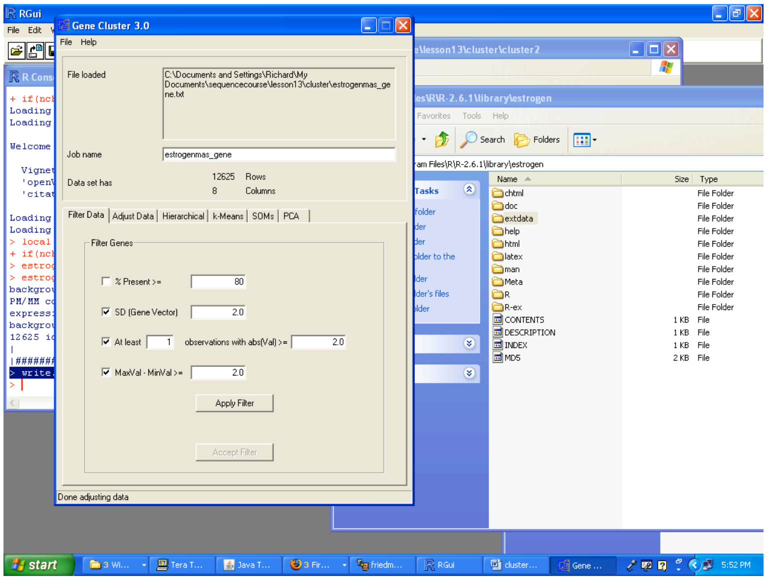
RGui
File Edit View Misc Packages Windows Help Vignettes
File Help
Filter Data
Adjust Data
Hierarchical
k-Means
SOMs
PCA
Apply
Getting Started
Office Online
Connect to Microsoft Office Online
Get the latest news about using Word
Automatically update this list from the web
More...
Search for:
Example: "Print more than one copy"
Open
start
5:46 PM

D. Click Apply

Click filter data

11. In filter data Check

    A. SD (Gene Vector) 2.0
    B. At least one observation with abs(val) >= 2.0
    C. MaxVal-MinVal>=2.0



Gene Cluster 3.0

File loaded: C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\estrogenmas_gene.txt

Job name: estrogenmas_gene

Data set has: 12625 Rows, 8 Columns

Filter Genes

☐ % Present >= 80

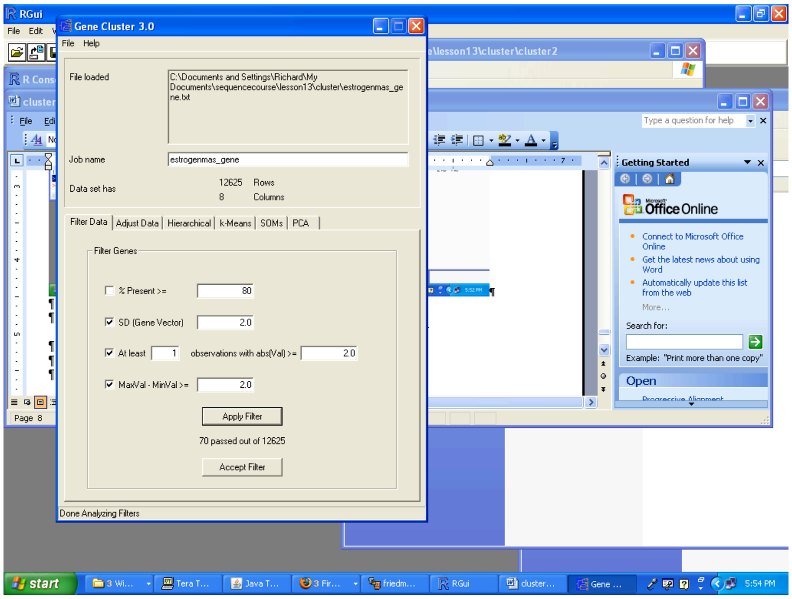
☑ SD (Gene Vector) 2.0

☑ At least 1 observations with abs(Val) >= 2.0

☑ MaxVal - MinVal >= 2.0

Done adjusting data

    D. Click Apply filter. The GUI should say “70 passed out of 12625”



Gene Cluster 3.0

File loaded: C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\estrogenmas_gene.txt

Job name: estrogenmas_gene

Data set has: 12625 Rows, 8 Columns

Filter Genes

☐ % Present >= 80

☑ SD (Gene Vector) 2.0

☑ At least 1 observations with abs(Val) >= 2.0

☑ MaxVal - MinVal >= 2.0

70 passed out of 12625

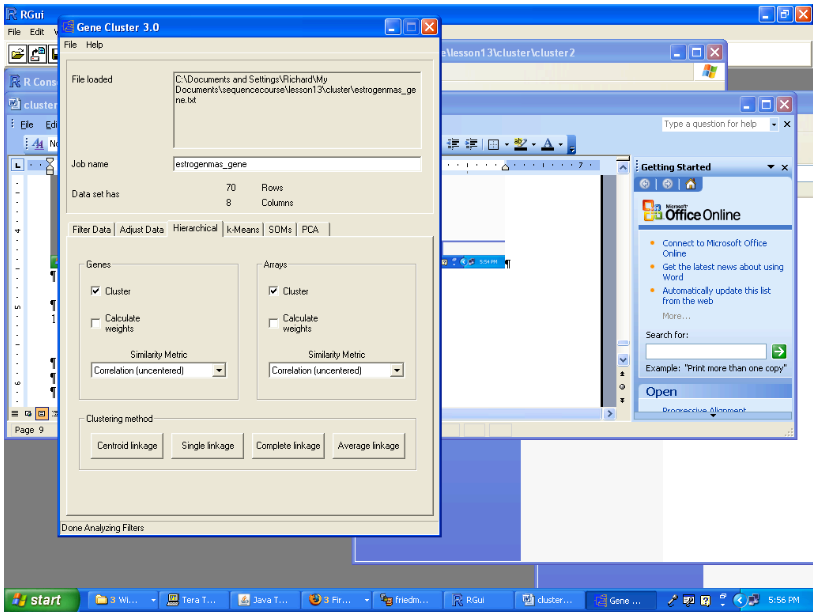
Done Analyzing Filters

    E. Click Accept filter

12, Click on Hierarchical tab. Check

A. Cluster Genes.

B. Cluster Arrays



Gene Cluster 3.0

File loaded: C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\estrogenmas_gene.txt

Job name: estrogenmas_gene

Data set has 70 Rows 8 Columns

Genes: ☑ Cluster □ Calculate weights Similarity Metric: Correlation (uncentered)

Arrays: ☑ Cluster □ Calculate weights Similarity Metric: Correlation (uncentered)

Done Analyzing Filters

File Help
Filter Data | Adjust Data | Hierarchical | k-Means | SOMs | PCA
Clustering method
Centroid linkage
Single linkage
Complete linkage
Average linkage
Getting Started
Office Online
Connect to Microsoft Office Online
Get the latest news about using Word
Automatically update this list from the web
More...
Search for:
Example: "Print more than one copy"
Open
Type a question for help
start
RGui
5:56 PM

Then click “Average linkage”

Files entitled:

A. estrogenmas_gene.atr (ATR)

B. estrogenmas_gene. (CDT)

C. estrogenmas_gene. (ATR)

Should appear in your directory. You should now be ready to display the cluster diagram and heatmap.

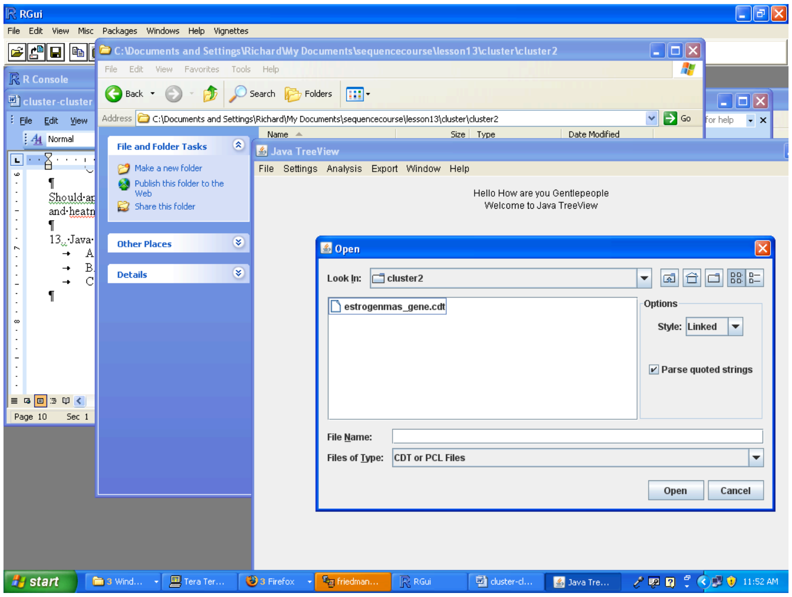
13. Java Treeview

A. Open Java Treeview.

B. Select Open-> File

C. Scroll to directory containing CDT file

D. Open CDT file



C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\cluster2

Java TreeView

Hello How are you Gentlepeople
Welcome to Java TreeView

Open

Look In: cluster2

estrogenmas_gene.cdt

Options

Style: Linked

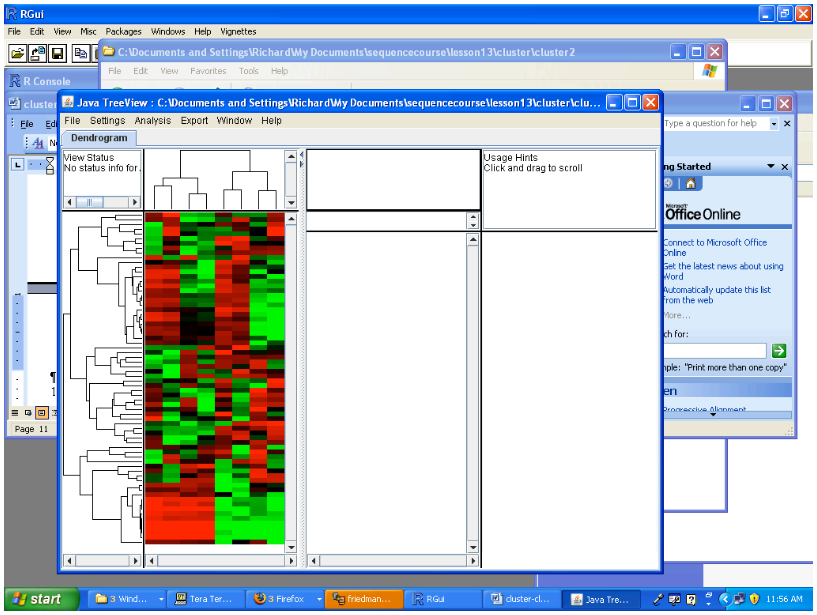
☑ Parse quoted strings

File Name:

Files of Type: CDT or PCL Files

Open Cancel

14. You should get a heat map that looks like this.



Java TreeView : C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\clu…

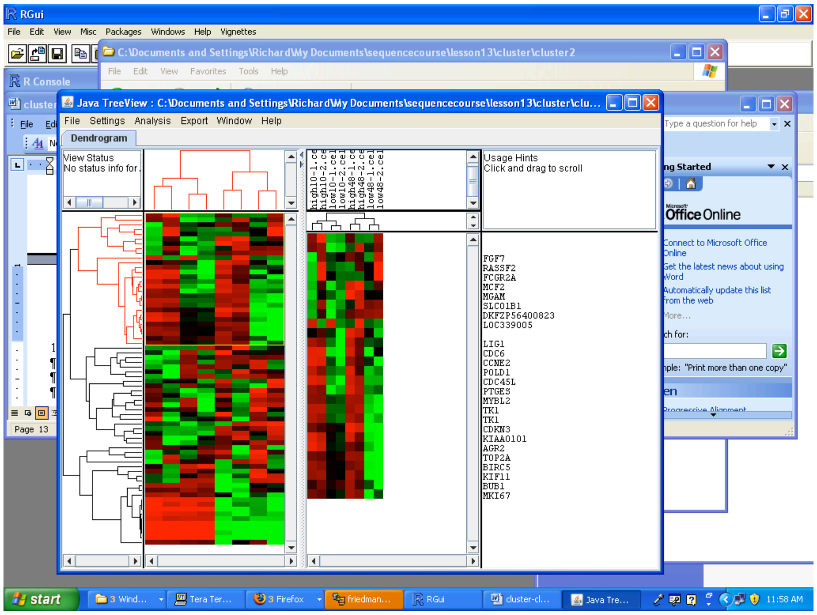
Dendrogram

View Status
No status info for

Usage Hints
Click and drag to scroll

RGui
File Edit View Misc Packages Windows Help Vignettes
R Console
File Settings Analysis Export Window Help
Type a question for help
Office Online
Connect to Microsoft Office Online
Get the latest news about using Word
Automatically update this list from the web
start
11:56 AM

15. You can select and magnify clusters by clicking on nodes:

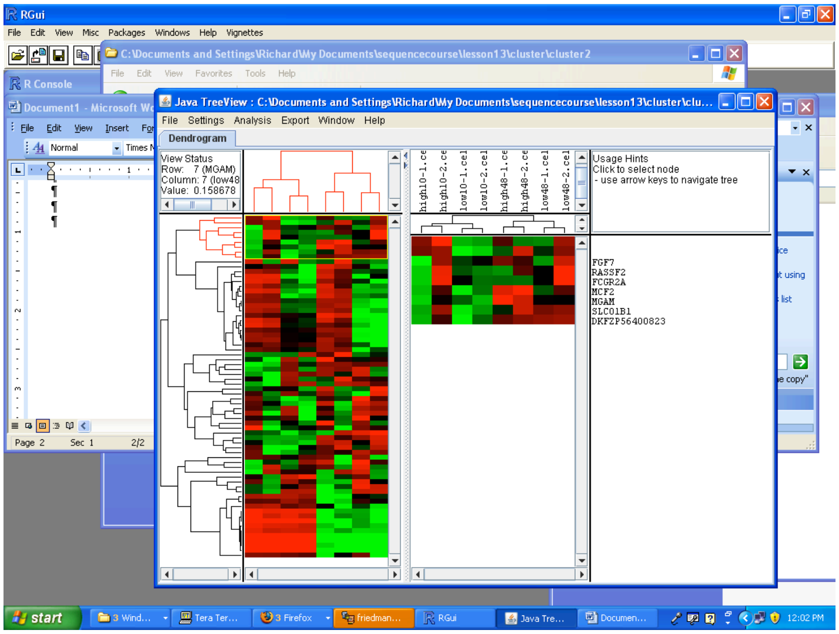

RGui
File Edit View Misc Packages Windows Help Vignettes
C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\cluster2
Java TreeView : C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\clu...
File Settings Analysis Export Window Help
Dendrogram
start
11:58 AM

View Status
No status info for

Usage Hints
Click and drag to scroll

```
FGF7
RASSF2
FCGR2A
MCF2
MGAM
SLCO1B1
DKFZP56400823
LOC339005

LIG1
CDC6
CCNE2
POLD1
CDC45L
PTGES
MYBL2
TK1
TK1
CDKN3
KIAA0101
AGR2
TOP2A
BIRC5
KIF11
BUB1
MKI67
```

16. The resulting heat map has good expansion for exploring clusters.


RGui
File Edit View Misc Packages Windows Help Vignettes
C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\cluster2
Java TreeView : C:\Documents and Settings\Richard\My Documents\sequencecourse\lesson13\cluster\clu...
File Settings Analysis Export Window Help
Dendrogram
high10-1.ce
high10-2.ce
low10-1.cel
low10-2.cel
high48-1.ce
high48-2.ce
low48-1.cel
low48-2.cel
start
12:02 PM

View Status
Row: 7 (MGAM)
Column: 7 (low48
Value: 0.158678

Usage Hints
Click to select node
- use arrow keys to navigate tree

```
FGF7
RASSF2
FCGR2A
MCF2
MGAM
SLCO1B1
DKFZP56400823
```

Lab

1. Characterize the function, interactions and pathways of the c-src (or the protein of your choice) using web-accessible databases. Save the necessary information.
2. What is the phenotype associated with the BRCA1 185DELAG mutation in humans?
3. From the list of differentially expressed genes obtained in the 10 hour estrogen experiment explained in part I, generate the following lists in Excel Workbook format:

   A. The Biological Function Gene-ontology values overrepresented in the Estrogen experiment sorted by corrected p-values.
   B. The Biological Function Gene-ontology values overrepresented in the Estrogen experiment with gene symbols and other identifiers (unsorted).
   C. The KEGG pathways overrepresented in the Estrogen experiment.

4. Produce a heatmap representing the gene expression patterns of the estrogen dataset from the last lab clustered by both gene and array.