Manual

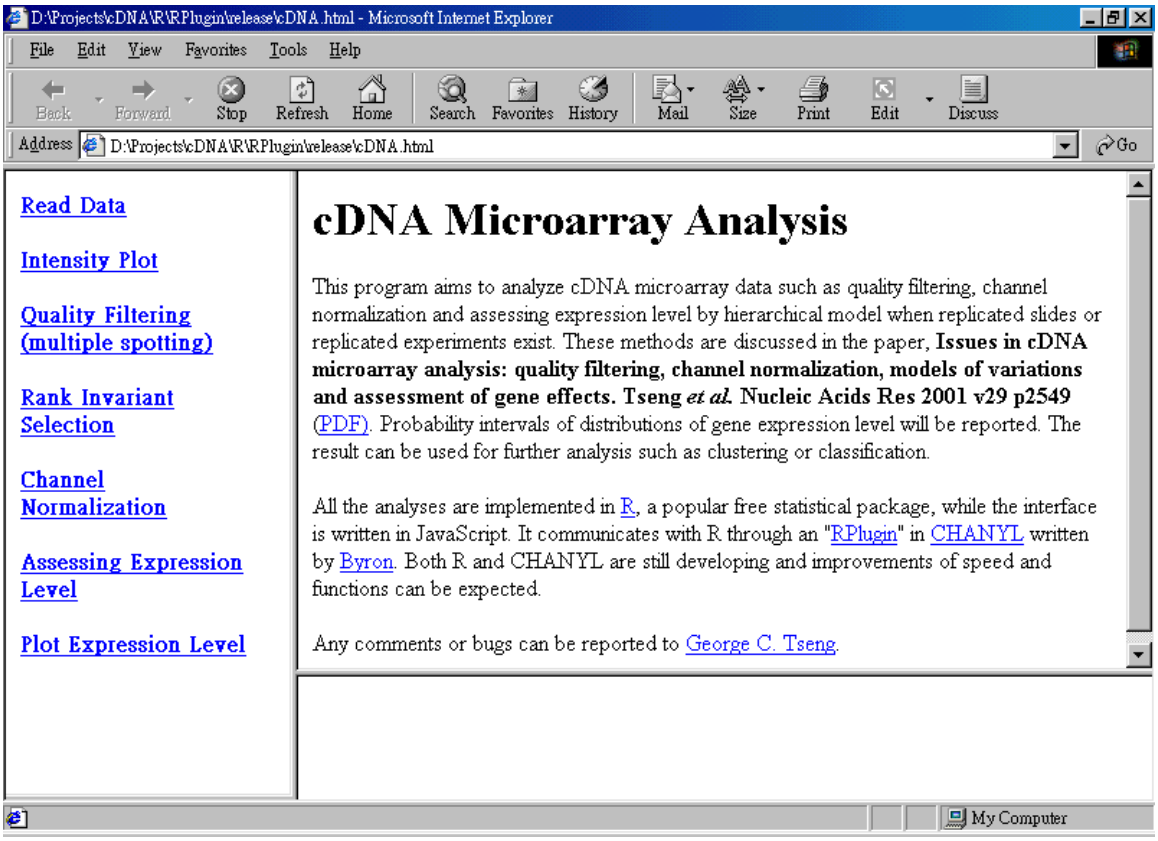
This program requires installation of two packages.
(1) R installation:
Download and install R from the homepage: http://cran.r-project.org/
(2) RPlugin installation:
Install Rplugin from http://www.pitt.edu/~ctseng/research/Rplugin/Rplugin.html

Paper

All the methods is this program are presented in the paper:
Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Tseng *et al.* Nucleic Acids Res 2001 v29 p2549 (PDF)

Installation of this program

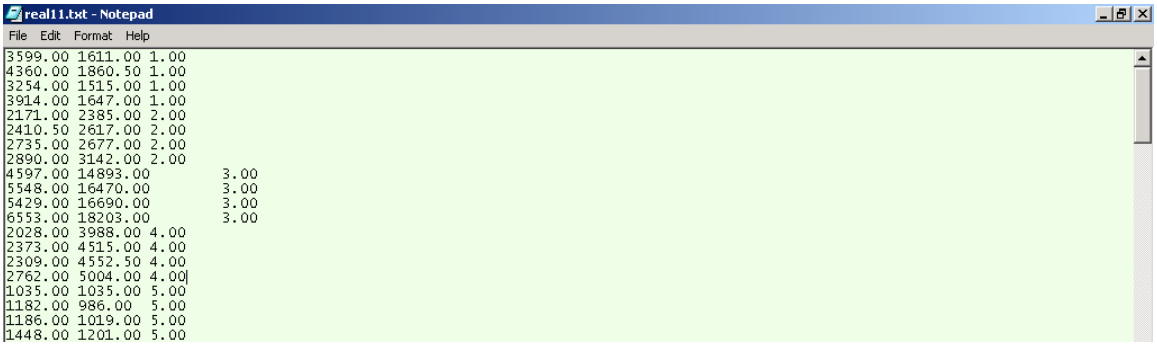
Unzip the file "cDNA.zip" into the same folder. It contains "cDNA.html", "menu.html" and "cDNAfunction.txt". Simply double click "cDNA.html" to begin the program.



Data Format

The input data format is shown in the following (Fig 0.1). Each row represents data of one spot (gene) in the microarray with Cy3 intensity in the first column, Cy5 in the

second and gene number in the third. In this example, each gene is spotted four times on a slide. The data is tab-delimited which can usually be exported from Excel, R or Splus. Note no header (description of column names) is allowed.


real11.txt - Notepad

File Edit Format Help

```
3599.00 1611.00 1.00
4360.00 1860.50 1.00
3254.00 1515.00 1.00
3914.00 1647.00 1.00
2171.00 2385.00 2.00
2410.50 2617.00 2.00
2735.00 2677.00 2.00
2890.00 3142.00 2.00
4597.00 14893.00        3.00
5548.00 16470.00        3.00
5429.00 16690.00        3.00
6553.00 18203.00        3.00
2028.00 3988.00 4.00
2373.00 4515.00 4.00
2309.00 4552.50 4.00
2762.00 5004.00 4.00
1035.00 1035.00 5.00
1182.00 986.00  5.00
1186.00 1019.00 5.00
1448.00 1201.00 5.00
```

Fig 0.1

Warnings:

1. Internet Explorer is required to run this program since “RPlugin” doesn’t support Netscape Navigator yet.
2. Currently “RPlugin” is vulnerable to any error message. It will break down the whole program. Thus correct variable names (case-sensitive) and data format are required.
3. Do not save the program in a directory with space, like “c:\my folder”. A space in IE is recognized as “%20” and will crash the program.
4. Remember to check “with header at the first line” in the “Read Data” menu if the data file being imported contains a header at the first line.
5. If you run R1.3.1 or up and find some problems, please install R1.3.0. The current RPlugin is designed for this version.
6. Due to the heavy computation when assessing expression level in hierarchical experiment structure, it is usually better to compute only several hundreds of genes at a time. For example, input “1:300” in the “genes to compute expression level” option to compute the first 300 genes, save the result at “result1” and then input “301:600” to compute the next 300 and save the result at “result2”.
7. When dealing with larger number of genes and replicates, it is recommended to directly run in R for more efficient input and computation. An example file, “simple example of R procedures.txt”, containing commands to implement these functions in R is located in the subdirectory, “example”. The subdirectory also contains a small dataset of microarray experiment used in the NAR paper. Note in R the data is manipulated from command lines and some familiarity of R programming is required to implement the functions.
8. Some recently added analysis functions are not described below but should easily be understood and implemented.
9. A known bug: There is usually an error message after closing the explorer.

Analysis:

The following example uses the 125-gene project data in our paper.

1. Read data:

Click on “Read Data” on the left frame. Array data can be read by selecting the data file of a microarray slide and give it a variable name in this program (Fig 1.1). In this manual we denote by “comp1S1” the first slide of the first comparative experiment (which means different samples used in Cy3 and Cy5) and “calib2S3” the third slide of the second calibration experiment (which means same control samples used in both Cy3 and Cy5). Here experiments mean replicate experiments that the whole experiment are redone and slides mean replicate slides. Note that the data file should conform the specific format described in previous section.

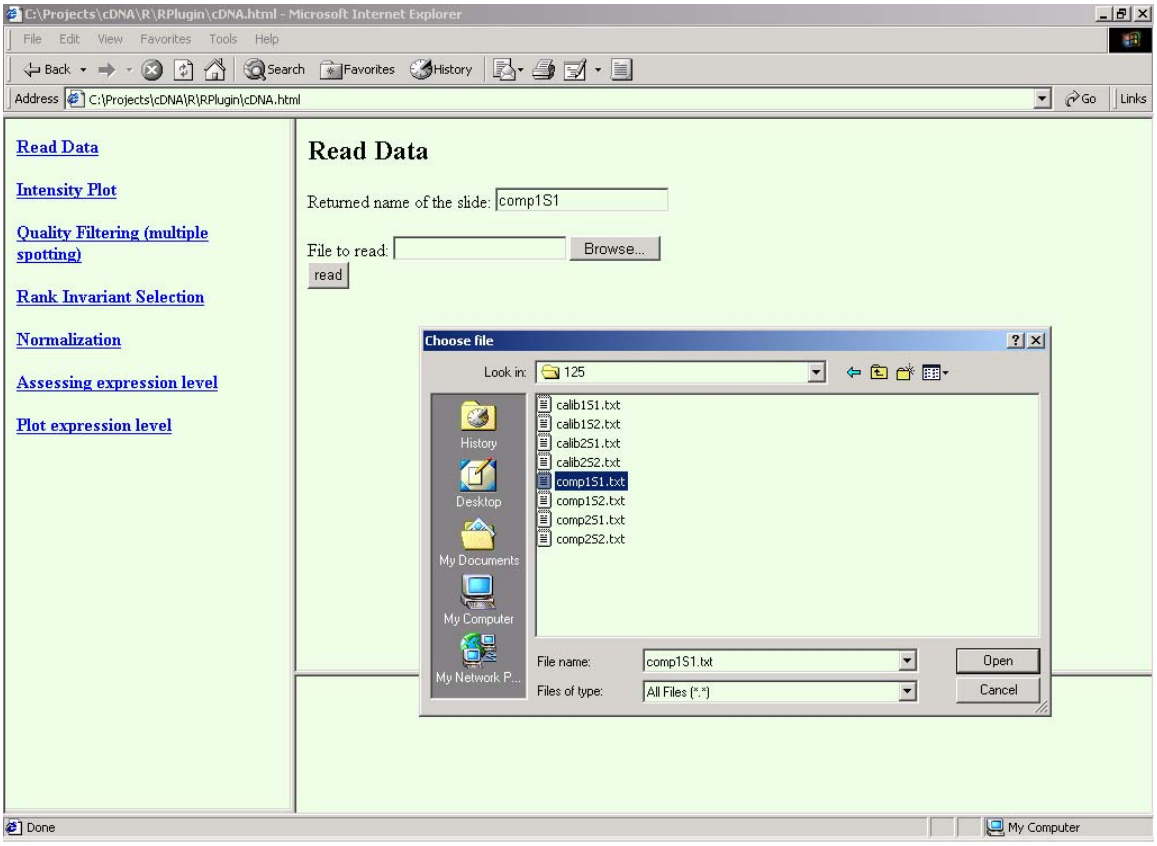


Fig 1.1

The data imported into the program will be shown on a pop-up window. The name of this data variable will be added to the variable list in the lower frame (Fig 1.2).

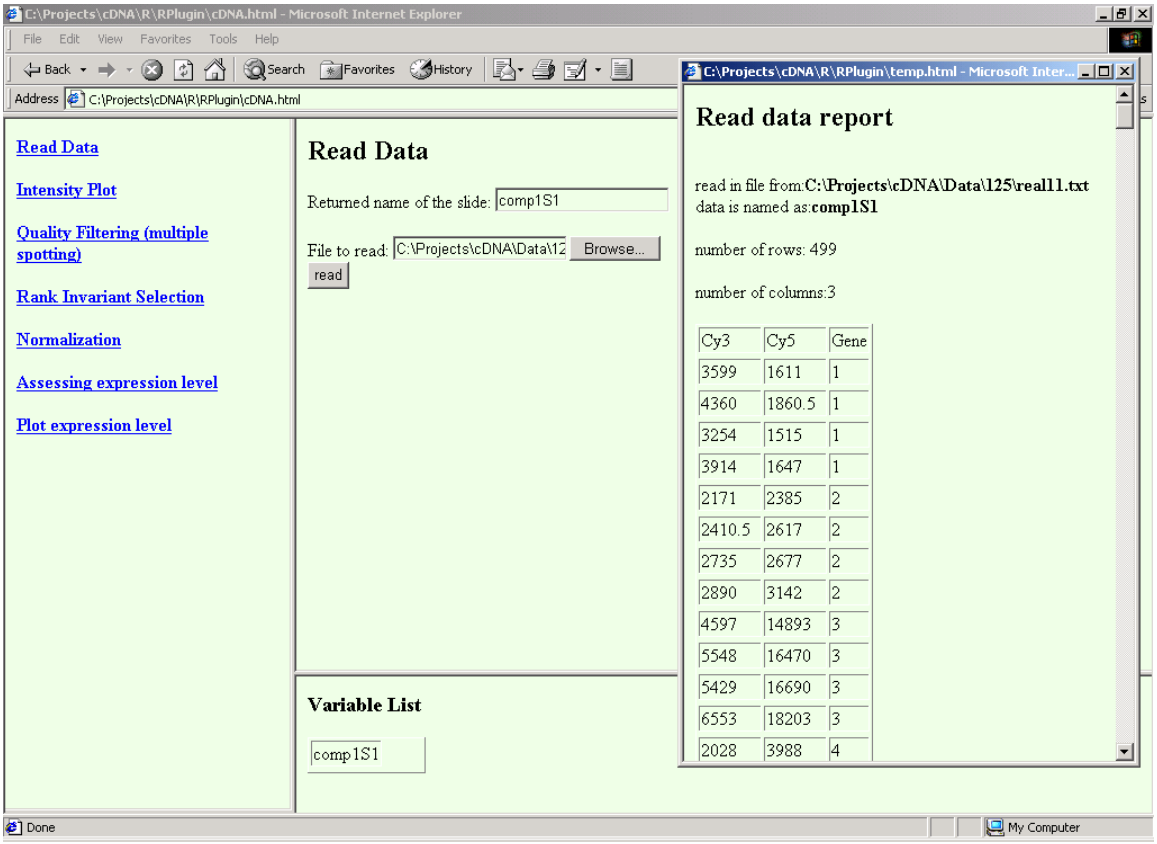


Fig 1.2

2. Intensity plot

Click on “Intensity Plot” on the left frame. After giving the variable name of the target slide in both “Cy5 versus Cy3 plot” and “M-A plot”, a window will pop up with the desired plot. You can save the plot by right-clicking the mouse and choose “Save as postscript”. The file will be saved as “*.ps”. (Fig 2.1, 2.2)

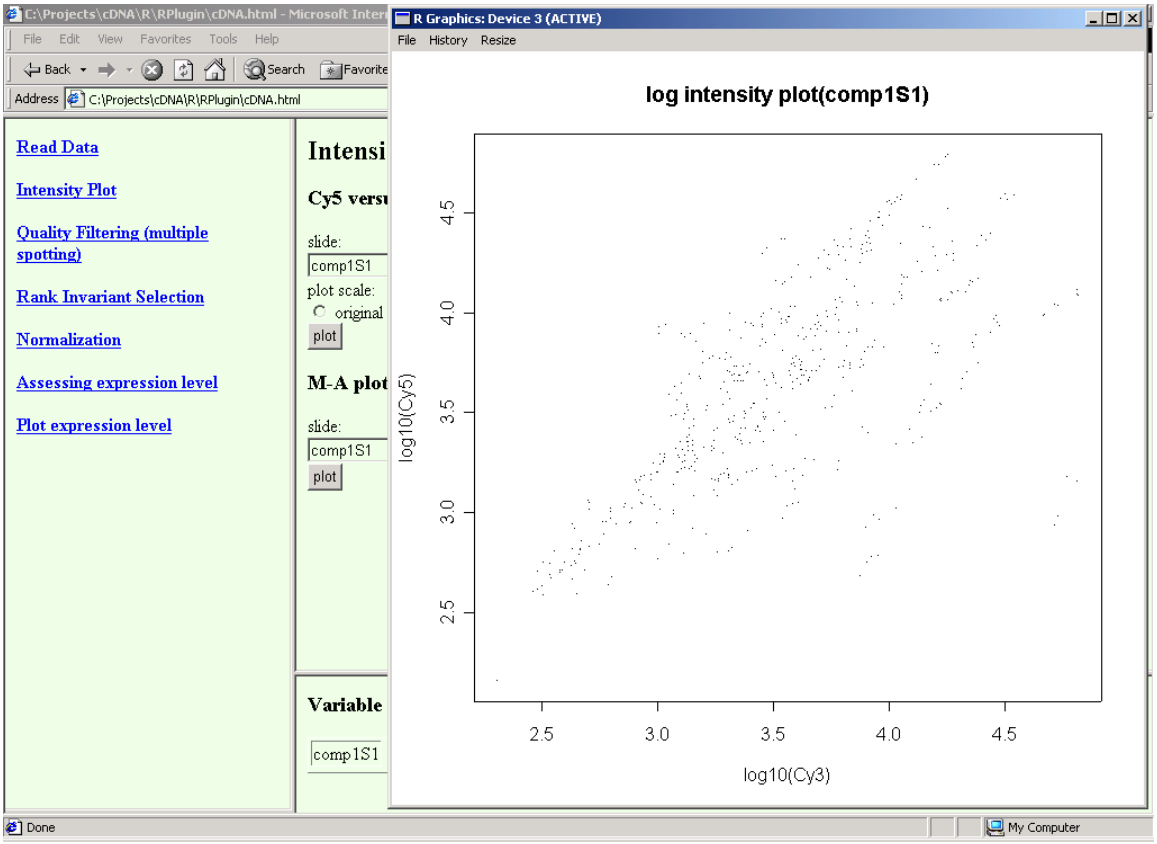

C:\Projects\cDNA\R\RPlugin\cDNA.html - Microsoft Inter
File Edit View Favorites Tools Help
Back Search Favorite
Address C:\Projects\cDNA\R\RPlugin\cDNA.html
Read Data
Intensity Plot
Quality Filtering (multiple spotting)
Rank Invariant Selection
Normalization
Assessing expression level
Plot expression level
Intensi
Cy5 versu
slide:
comp1S1
plot scale:
original
plot
M-A plot
slide:
comp1S1
plot
Variable
comp1S1
R Graphics: Device 3 (ACTIVE)
File History Resize
log intensity plot(comp1S1)
log10(Cy5)
log10(Cy3)
2.5
3.0
3.5
4.0
4.5
Done
My Computer

Fig 2.1

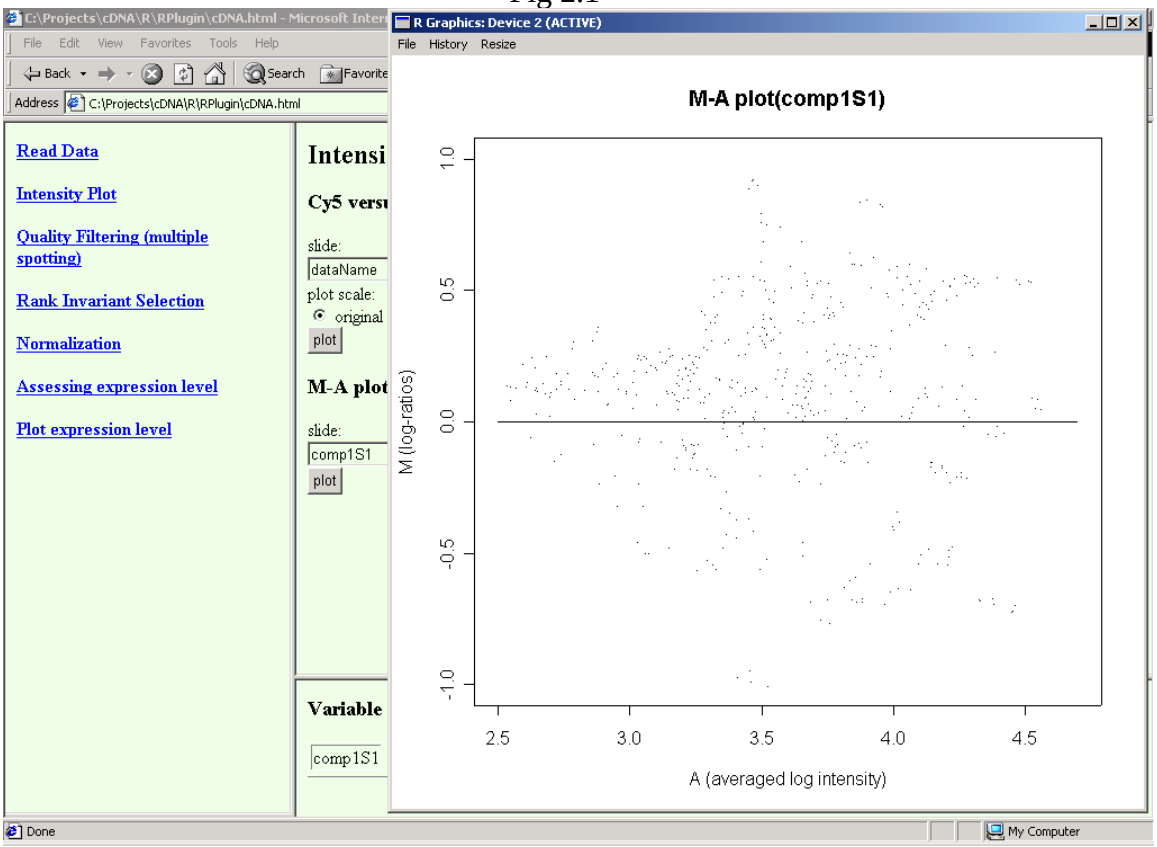

C:\Projects\cDNA\R\RPlugin\cDNA.html - Microsoft Inter
File Edit View Favorites Tools Help
Back Search Favorite
Address C:\Projects\cDNA\R\RPlugin\cDNA.html
Read Data
Intensity Plot
Quality Filtering (multiple spotting)
Rank Invariant Selection
Normalization
Assessing expression level
Plot expression level
Intensi
Cy5 versu
slide:
dataName
plot scale:
original
plot
M-A plot
slide:
comp1S1
plot
Variable
comp1S1
R Graphics: Device 2 (ACTIVE)
File History Resize
M-A plot(comp1S1)
M (log-ratios)
A (averaged log intensity)
1.0
0.5
0.0
-0.5
-1.0
2.5
3.0
3.5
4.0
4.5
Done
My Computer

Fig 2.2

3. Quality filtering

Click on “Quality Filtering” on the left frame. Give the variable name of the slide being filtered and the return name of the filtered slide (comp1S1 and comp1S1F respectively in the example). (Fig 3.1) Two windows will pop up, one showing the data after filtering and the other showing the plot of intensity versus quality index. There is also an option to save the data after filtering. Note that the quality filtering only applies to multiple spotting designs. Array data after filtering can be saved into a file. (Fig 3.2)

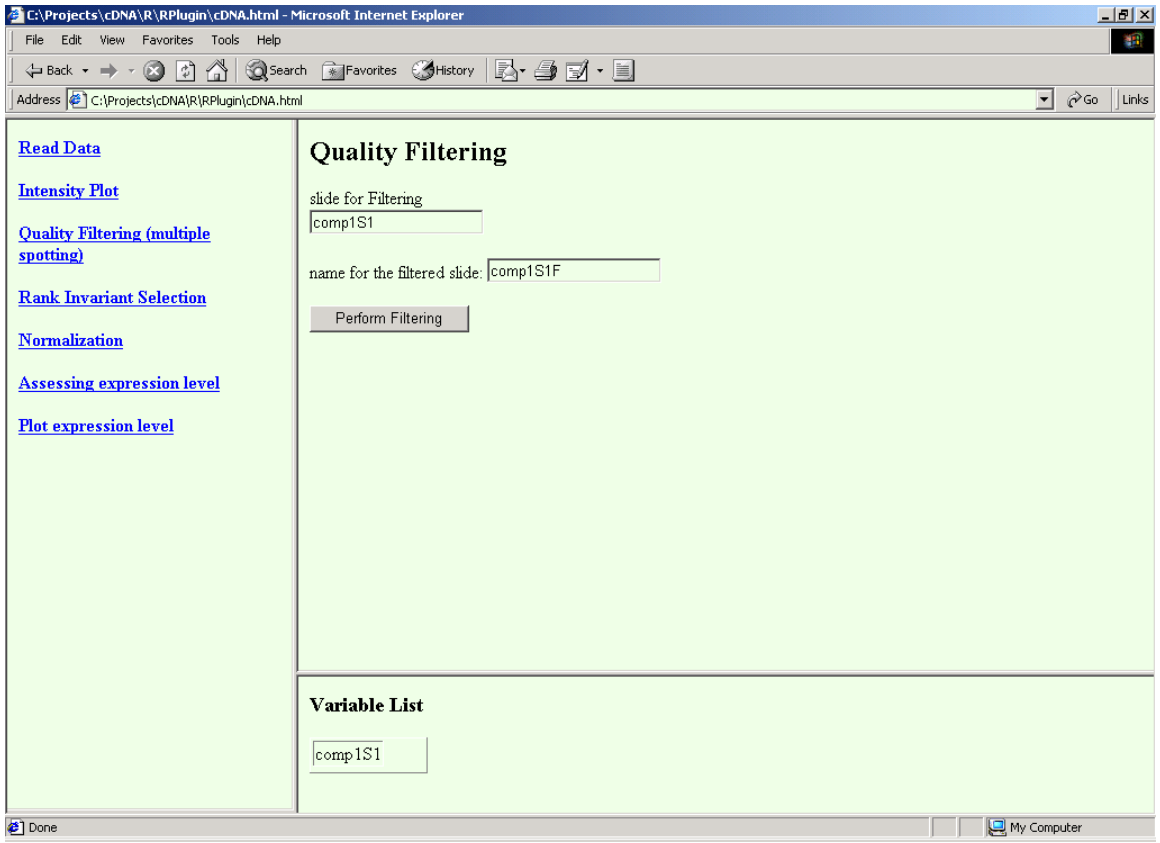


Fig 3.1

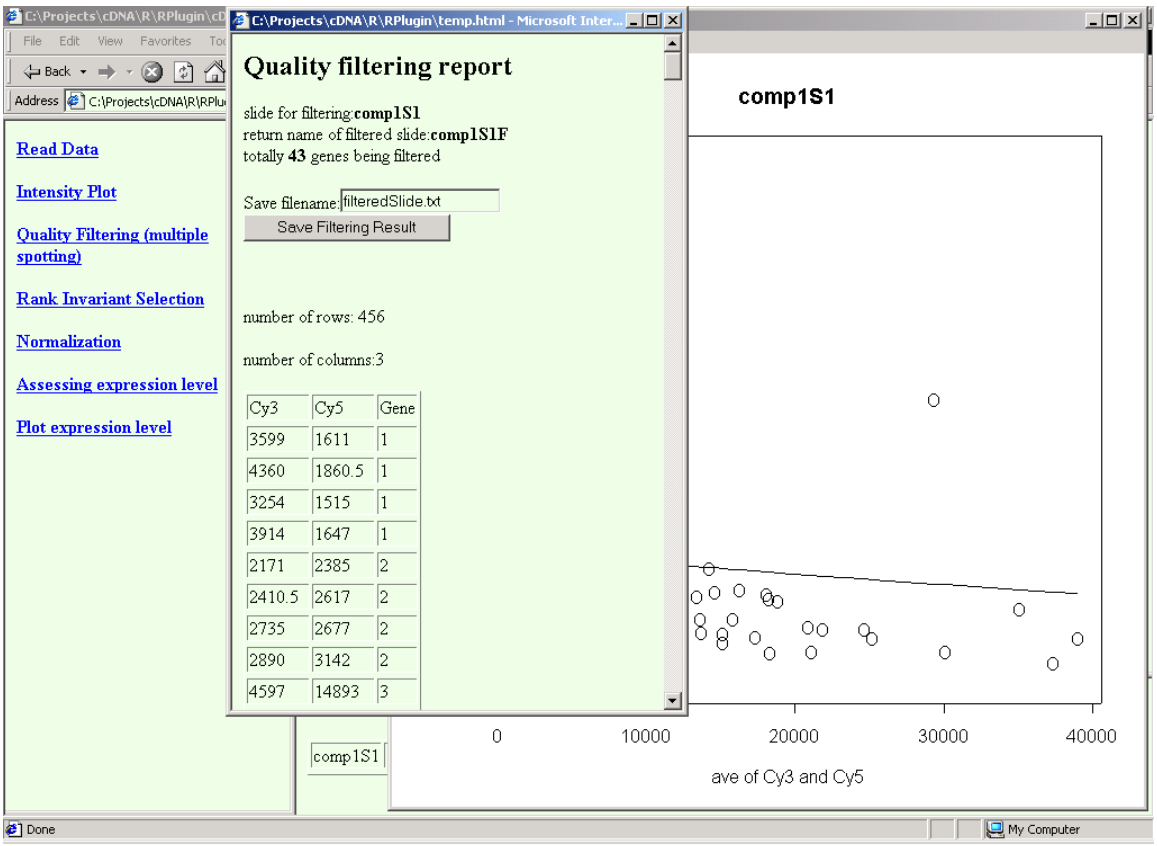


Fig 3.2

4. Rank invariant selection

Click on “Rank Invariant Selection” on the left frame. Input the variable name of the slide for selection and the return variable name of selected genes. When the number of genes is large (several thousands), iterative selection would help to select more conserved genes. In our 125-gene example, we can only apply non-iterative selection. For details of other parameters, please refer the original paper. After selection, a pop-up window will report the result. The selected genes can be saved into a file. (Fig 4.1)

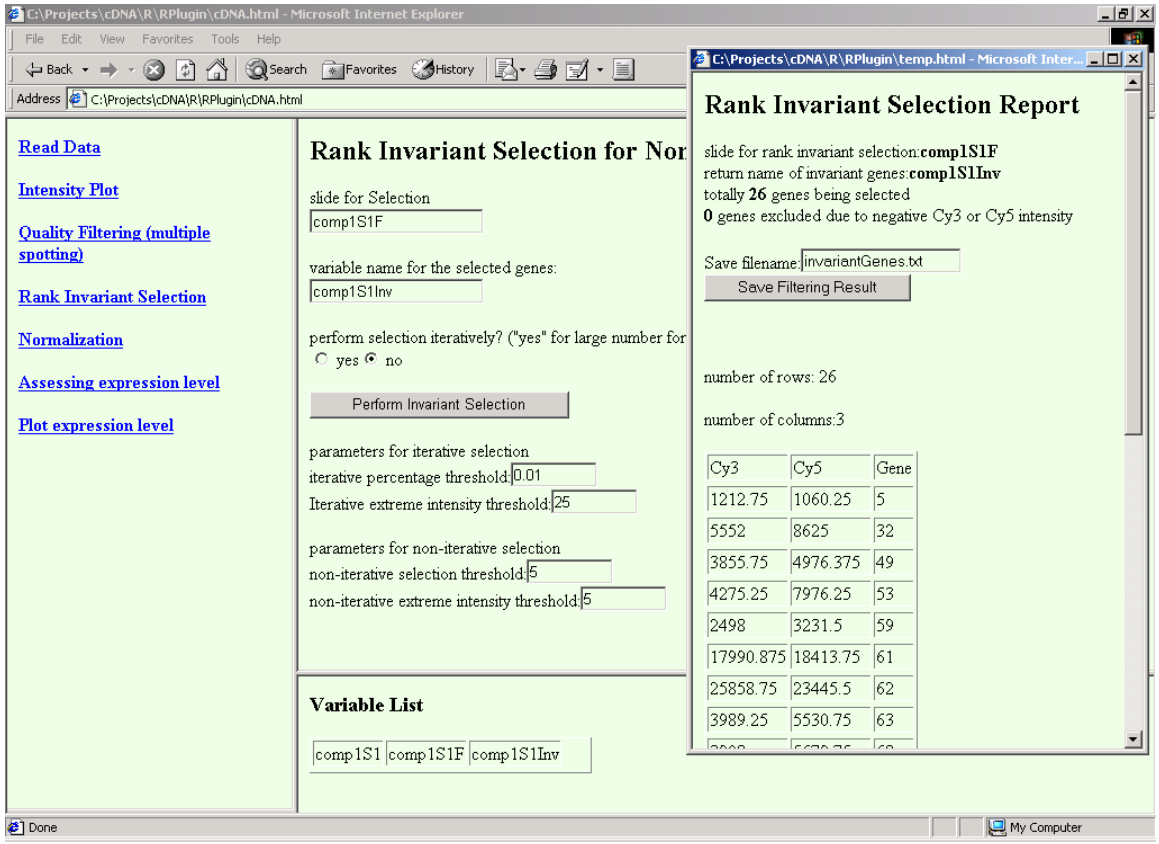


Fig 4.1

5. Channel normalization

Click on “Channel normalization” on the left frame. Array data of comparative or calibration experiment can be normalized here. In calibration experiment, it is appropriate to use all the genes for normalization. In comparative experiment, however, an extra set of invariant genes selected in step 4 should be given for normalization. (Fig 5.1)

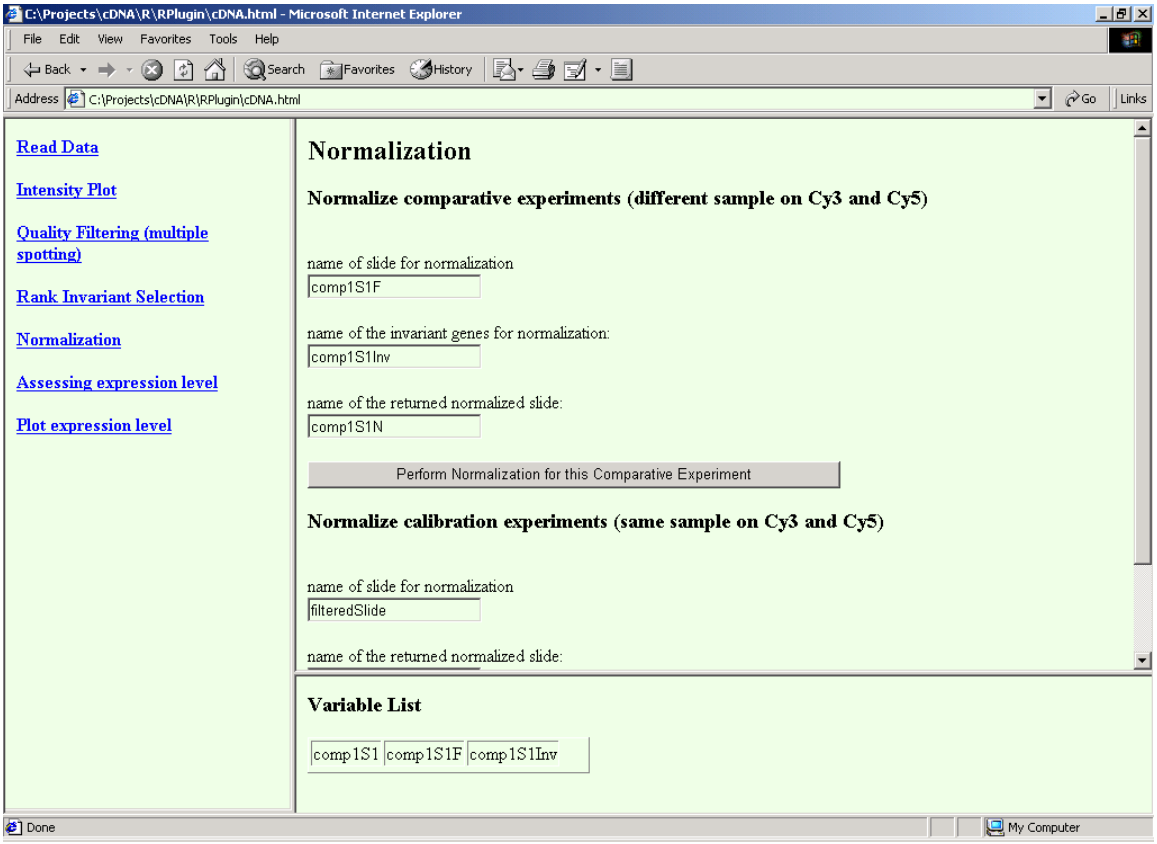


Fig 5.1

Two windows will pop up. One shows the data with normalized log-ratios. The other shows the normalization curve, data points (green) and selected invariant genes (red). The normalized result can be saved into file for further analysis. (Fig 5.2)

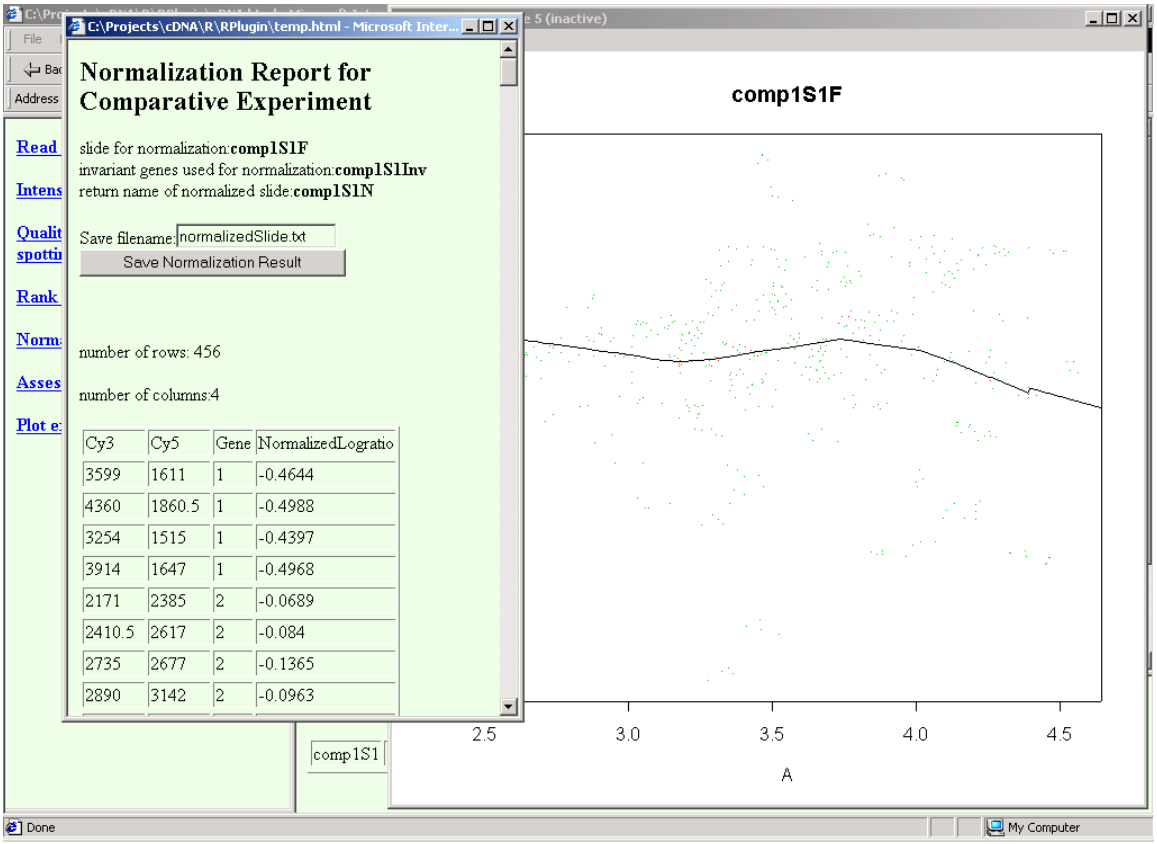


Fig 5.2

6. Assessing expression level

After we have normalized all the slides in our experiments, we are able to assess the expression level by the hierarchical structure, namely multiple experiments with multiple slides in each experiment. This program also works for multiple slides in only one experiment. When calibration experiments are available, they are used to obtain gene-specific prior information for reproducibility variation. Otherwise this prior is obtained from comparative experiment itself and is no longer gene-specific. For more details, refer to the original paper.

Click on “Assessing expression level” on the left frame. Here we show an example of two comparative experiments containing two slides without calibration.

(1) Input “2” comparative experiments in “Without calibration experiment”. (Fig 6.1) Then click “Next”.

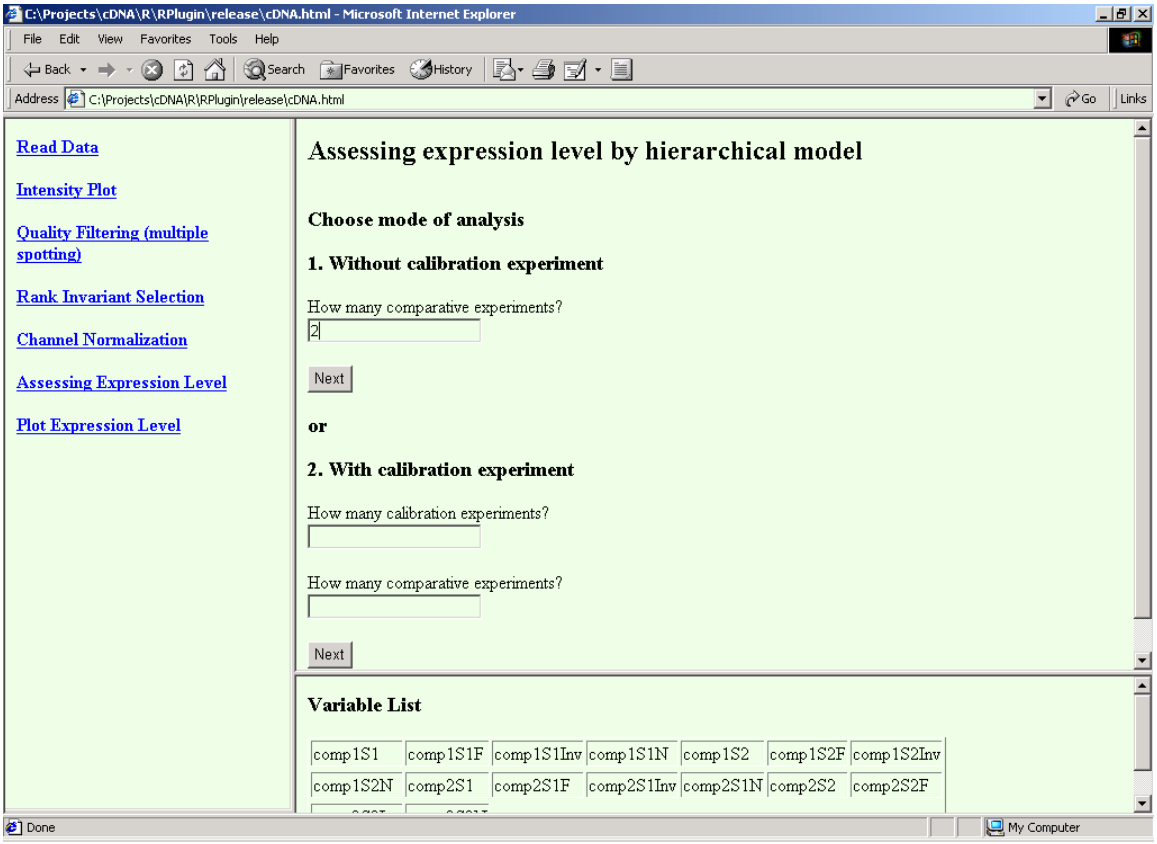


Fig 6.1

(2) The program will further ask for the number of slides in each experiment. Here we have “2” slides in experiment 1 and “2” slides in experiment 2. (Fig 6.2) Click “Next”.

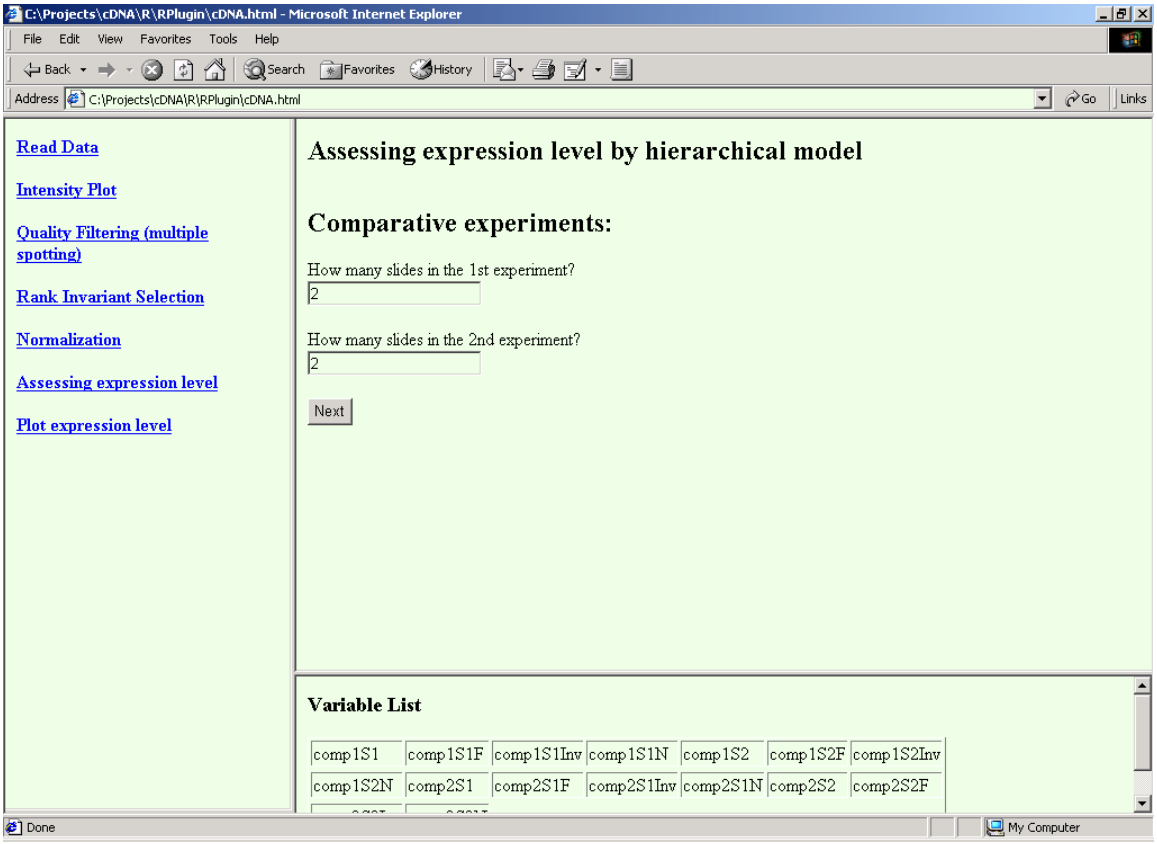


Fig 6.2

(3) The program then asks for the variable name of the data of each slide. (Fig 6.3) We also need to input the number of genes in the study. Note the maximum of the gene number should be given here. Besides we can specify the gene numbers we want to assess. Due to speed problem, it would be better to assess, say 500 genes, at a time and save them into different result names to avoid long waiting time that computer doesn't respond. Click "Run simulation to assess expression level".

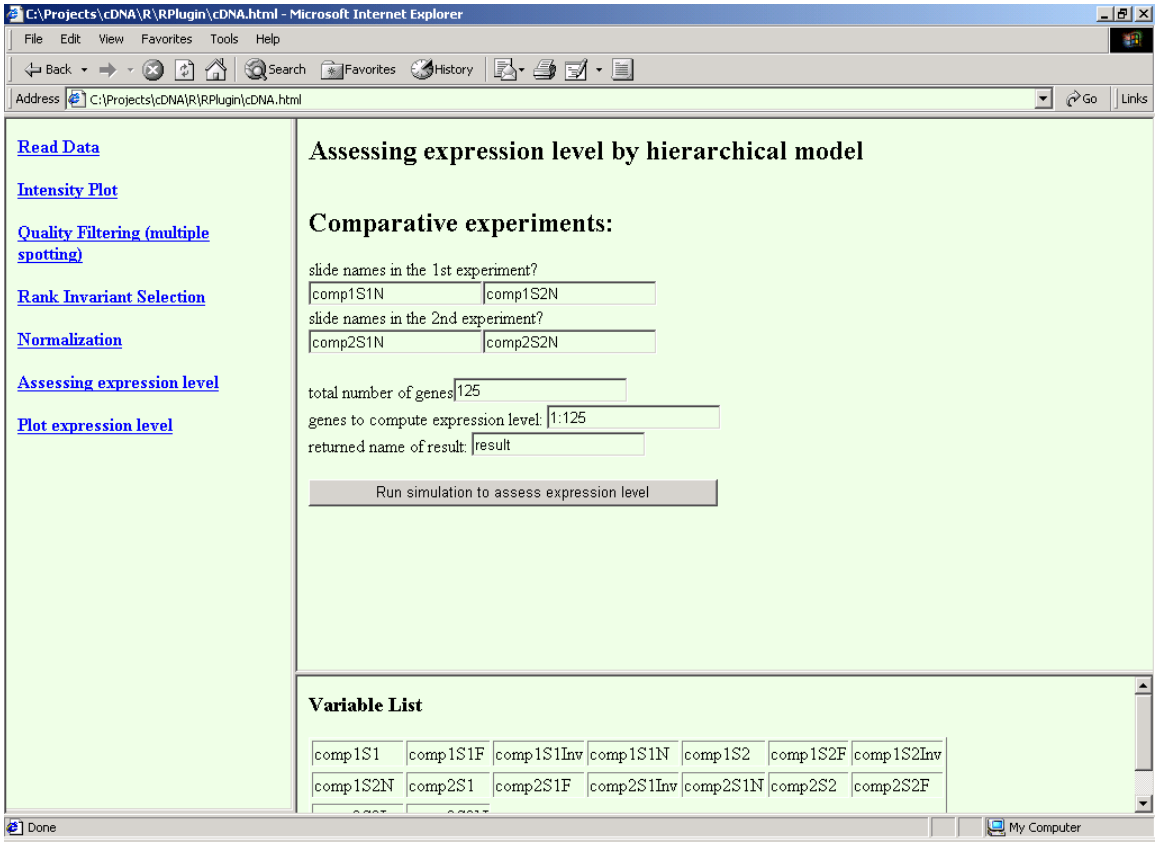


Fig 6.3

After some computation time depending on the complexity of the hierarchical structure and number of genes being computed, a pop-up window will show the result. The result contains the 95% probability coverage interval of the posterior distribution of gene expression level. The score shows the probability that the expression level (log-ratio) of a gene is less than 0. Thus a score very close to 1 (0) means down(up)-regulated. Note since the distribution has to be computed by simulation, the result will be slightly different if you rerun it. The result can be saved for further analysis. (Fig 6.4)

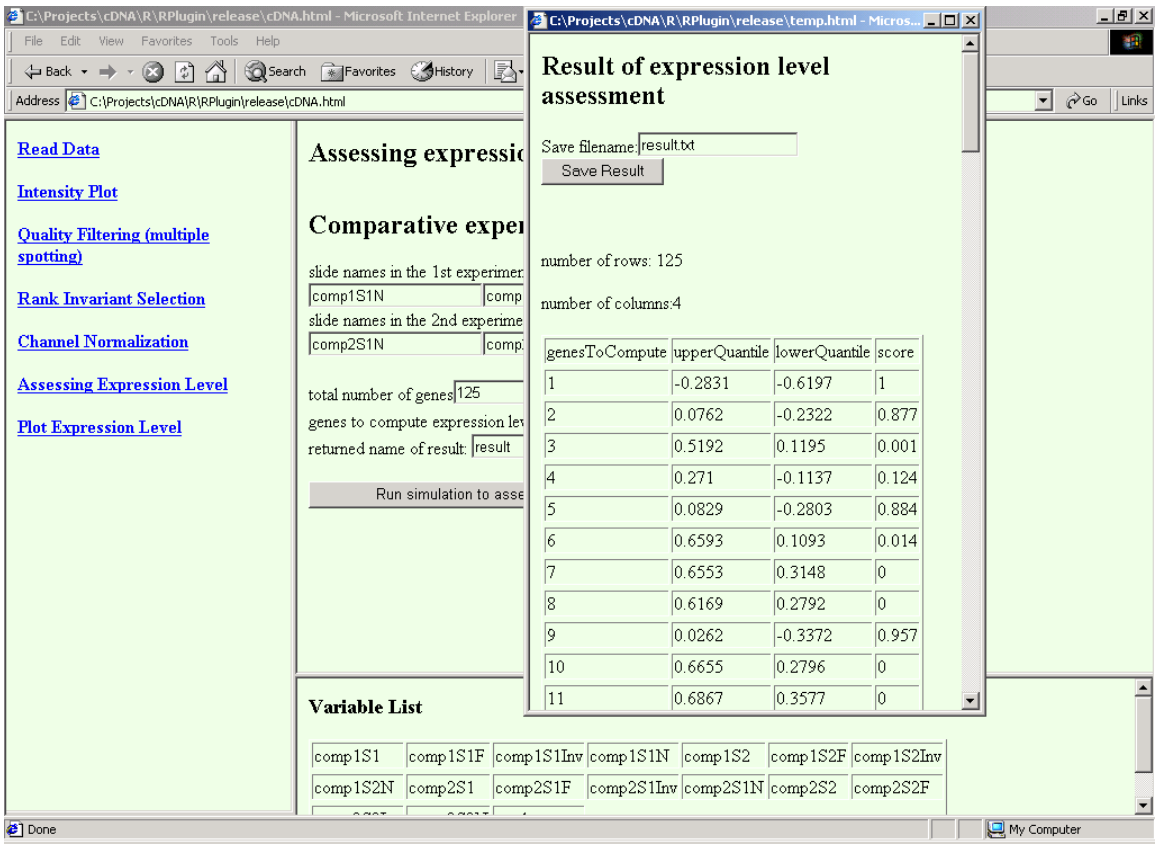


Fig 6.4

7. Plot expression level

Click on “Plot expression level” on the left frame. (Fig 7.1)

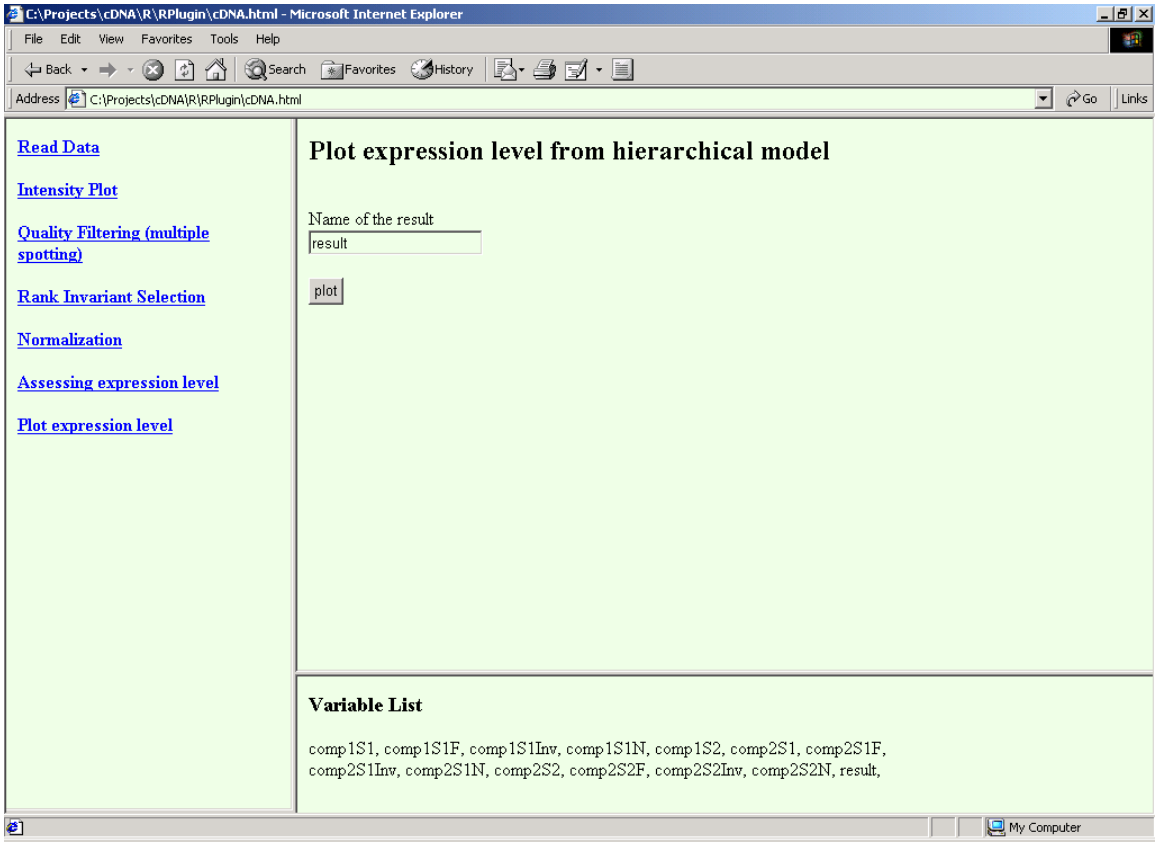


Fig 7.1

Click on “plot”. A pop-up window will show the expression level graphically. Use “PageUp” and “PageDown” buttons to view the graphs. The rectangles are the 95% probability coverage of the log-ratio expression level. If a rectangle lies under(above) 0, then it’s down(up)-regulated. (Fig 7.2)

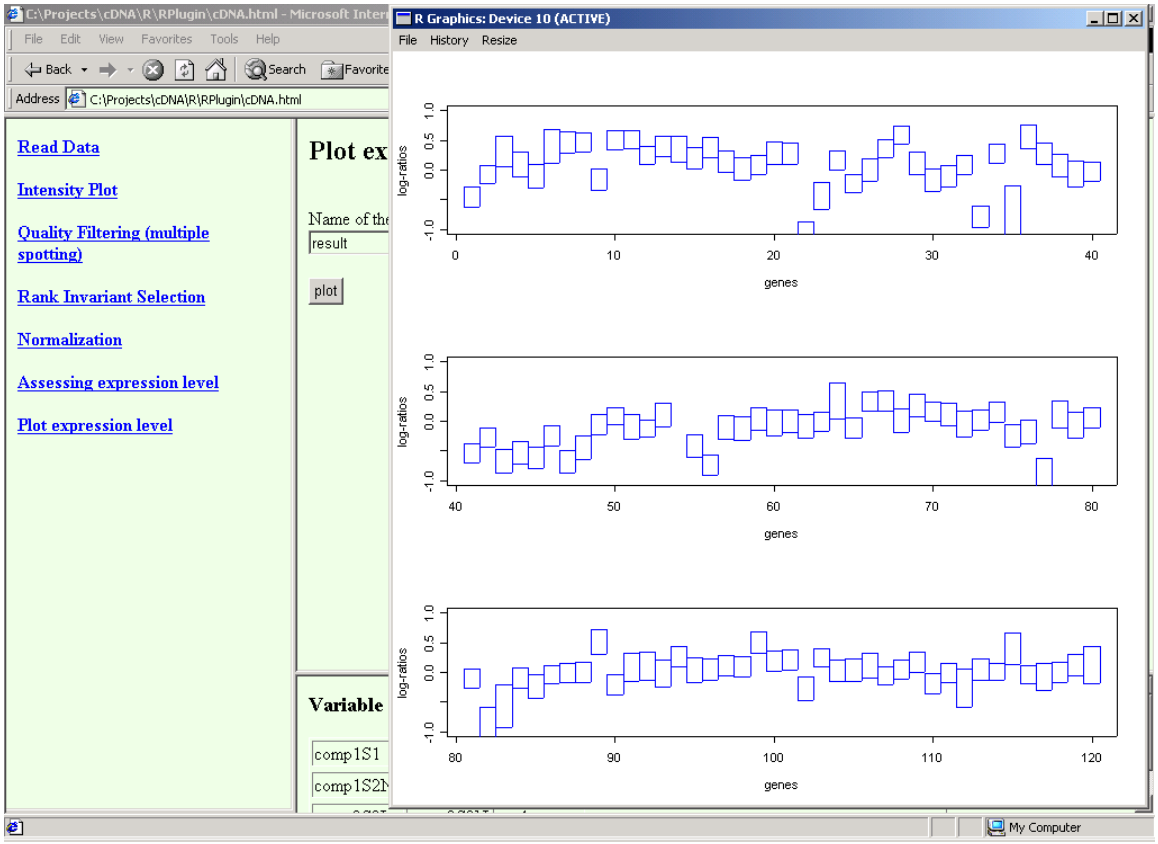

C:\Projects\cDNA\R\RPlugin\cDNA.html - Microsoft Inter
File Edit View Favorites Tools Help
Back Search Favorite
Address C:\Projects\cDNA\R\RPlugin\cDNA.html
Read Data
Intensity Plot
Quality Filtering (multiple spotting)
Rank Invariant Selection
Normalization
Assessing expression level
Plot expression level
Plot ex
Name of the
result
plot
Variable
comp1S1
comp1S2
R Graphics: Device 10 (ACTIVE)
File History Resize
log-ratios
genes
My Computer

Fig 7.2