GRACOMICS Instruction Manual

<Contents>

5. CONTACT

1. INTRODUCTION

1.1 What is the GRACOMICS?

GRACOMICS is a visualization program to view a set of multiple graphical results at a glance. Recent omics data have extremely a large number of parameters while having a small set of samples, this is also known as NP-problem and running an analysis with such data will return numerous amounts of results in p-dimensions. Additionally, comparing these results may not be simple as many different analysis specific results are produced. Consequently, the use of computer languages can′ t be avoided when performing such analyses. For example, R (reference) can be used to import data, analyze, and visualize certain data, but comparing those visualized results still remains a problem. In case the user is computer language friendly, such problem can be solved using R. TM4 (reference) is another program for omics data analysis that uses Graphic User interface (GUI) that ease the burden of programming skills yet, still do not provide a good comparison of multiple results.

The main hardship in comparing multiple results is that the results are solely dependent on the user-defined values such as the significant-level. However, GRACOMICS solves such problem, provides a full comparison of multiple results, and is a GUI based program even a novice in computer programmer can use. GRACOMICS is composed of 3 modules: Pairwise Correlation Plot to compare two results, Pairwise Interactive Plot to summarize the differences between two results more precisely with a click of a mouse, and Multiple Result Heat Map to compare two or more results comprehensively. Each module is expressed differently, yet can be used to provide mutual supplementation for each other. Especially, as a ′standalone′ java based program, GRACOMICS is a powerful tool that guides novice programmers to effectively compare several analysis results with ease.

1.2 What can GRACOMICS do?

GRACOMICS is useful when comparing multiple results from a single dataset at a glance. Following are some examples of such comparisons:

- In microarray data analysis, GRACOMICS can compare p-values of several statistical methods such as t-test, significance of microarray.
- In multiple regression analysis of statistics, GRACOMICS can find simultaneously significant result of p-values from several independent variables. For Example, in the model, Y<- SEX+AGE, we can find significant result for both test result of SEX and result of AGE.
- Comparisons of different models in Genome wide SNP dataset analysis.
- Comparing results from same models yet in different settings.

Since recent omics data analysis forces comparisons of repetitive results from different analytical approaches, GRACOMICS can be a useful addition to novice programmers and biologists when performing such comparisons analysis.

1.3 System Requirements

GRACOMICS will run on any machine that has a recent version of Java. This version of GRACOMICS requires Java 1.7.0_45 at least. Most of the development has been done using JDK v1.6 and some of the more advanced features require Java 1.7.0_45 or higher. See the section called Getting and Installing GRACOMICS for details on how to get Java.

2. INSTALLATION

2.1 Install and download

GRACOMICS is a java based program, and is distributed in *.jar form. The software is provided on the GRACOMICS homepage, and as a side note, JAVA launcher called *JRE* is an essential program that is needed to run GRACOMICS.

The most up to date version of GRACOMICS is always available from the GRACOMICS web pages.

http://bibs.snu.ac.kr/software/GRACOMICS/gracomics.php

Also, you can directly download the executable java file(.jar) from the following link: http://bibs.snu.ac.kr/software/GRACOMICS/gracomics.zip

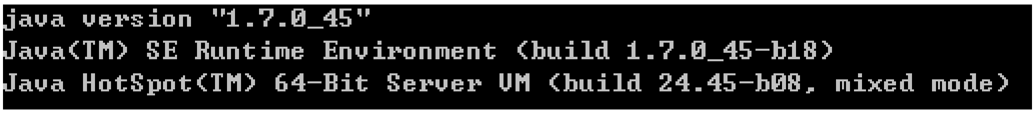
And following is the link for the *JRE* (launcher for .jar files):

http://www.oracle.com/technetwork/java/javase/downloads/index.html

To execute the GRACOMICS, you must register the following path: *jre-setup-folder/bin*.

If Java is successfully installed on the computer, following message will be on the screen.

- Java -version



```
java version "1.7.0_45"
Java(TM) SE Runtime Environment (build 1.7.0_45-b18)
Java HotSpot(TM) 64-Bit Server VM (build 24.45-b08, mixed mode)
```

2.2 Simple example

If all of the above contents are installed, GRACOMICS will run in any computer. With the simple command below, GRACOMICS can be executed from any folder, without additional install:

```
Java - jar gracomics.jar
```

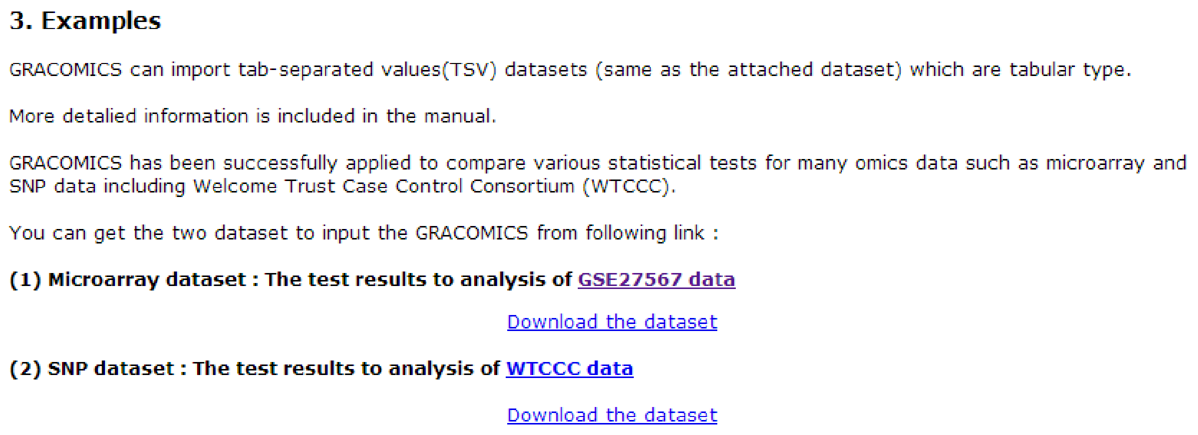
When the program is executed, the user can explore the functions of GRACOMICS using the dataset given on the zip file or web site.

http://bibs.snu.ac.kr/software/GRACOMICS/example_microarray.zip

<Microarray data>

http://bibs.snu.ac.kr/software/GRACOMICS/example_snp.zip

<SNP data>



3. Examples

GRACOMICS can import tab-separated values(TSV) datasets (same as the attached dataset) which are tabular type.

More detalied information is included in the manual.

GRACOMICS has been successfully applied to compare various statistical tests for many omics data such as microarray and SNP data including Welcome Trust Case Control Consortium (WTCCC).

You can get the two dataset to input the GRACOMICS from following link :

(1) Microarray dataset : The test results to analysis of GSE27567 data

Download the dataset

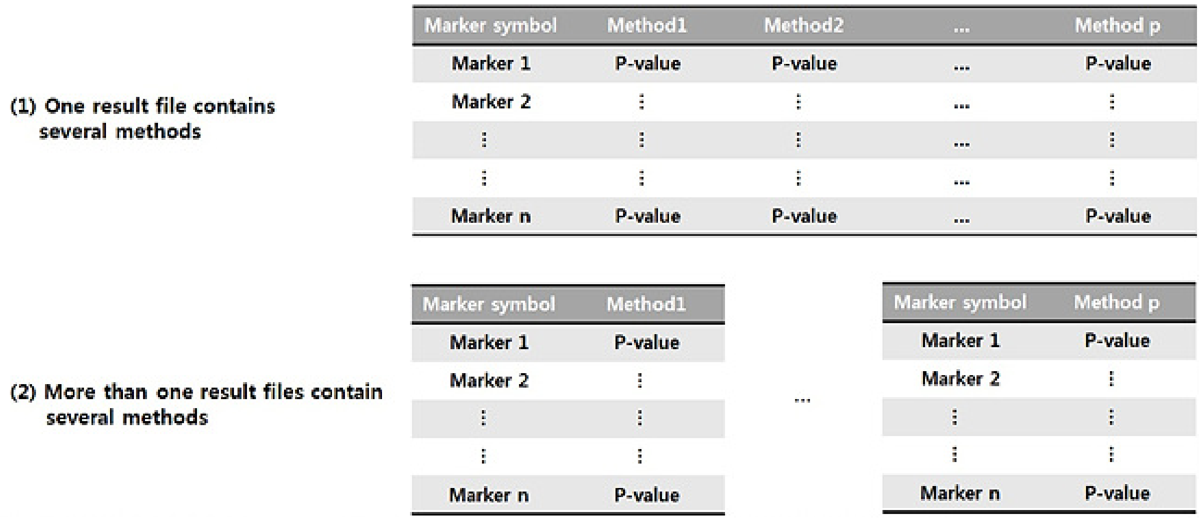
(2) SNP dataset : The test results to analysis of WTCCC data

Download the dataset

3. USAGE OF PROGRAM

3.1 INPUT DATA FORMAT

GRACOMICS can import tab-separated values (TSV) datasets (same as the attached dataset) which are tabular type and have the following structure:



(1) One result file contains several methods

Marker symbol	Method1	Method2	...	Method p
Marker 1	P-value	P-value	...	P-value
Marker 2	⋮	⋮	...	⋮
⋮	⋮	⋮	...	⋮
⋮	⋮	⋮	...	⋮
Marker n	P-value	P-value	...	P-value

(2) More than one result files contain several methods

Marker symbol	Method1
Marker 1	P-value
Marker 2	⋮
⋮	⋮
⋮	⋮
Marker n	P-value

...

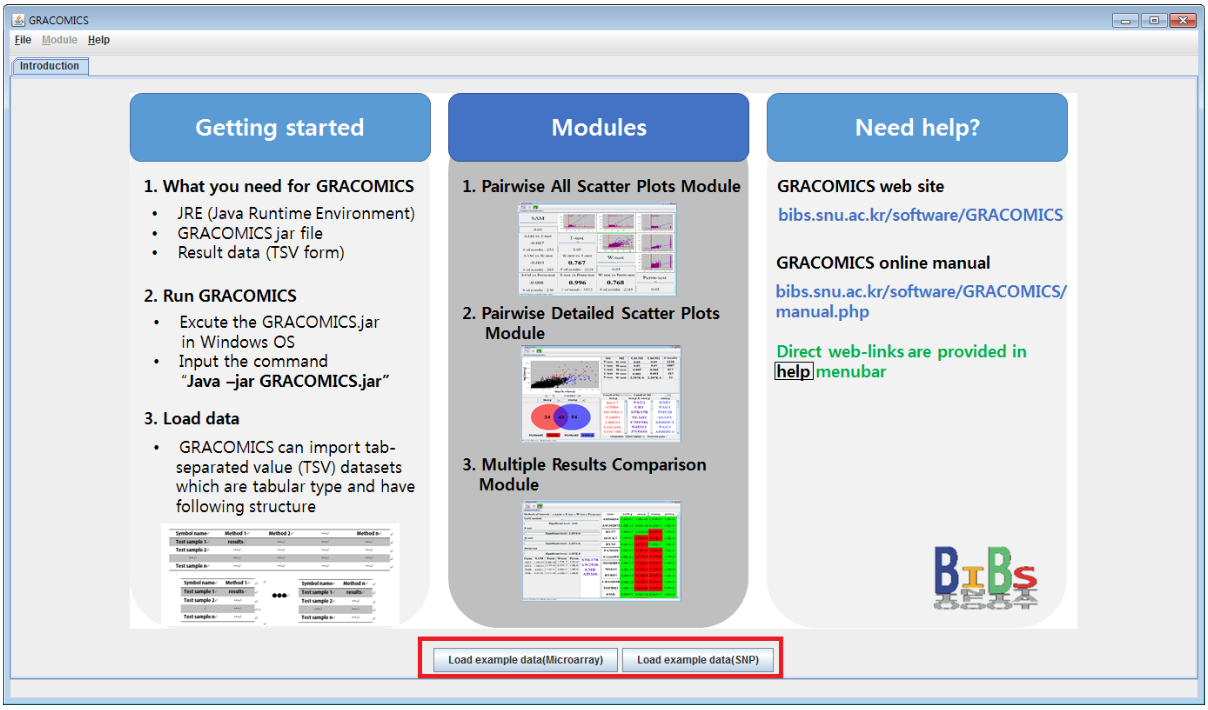
Marker symbol	Method p
Marker 1	P-value
Marker 2	⋮
⋮	⋮
⋮	⋮
Marker n	P-value

Also, GRACOMICS provides a module to open multiple files with similar format.

A detailed description of the file format can be observed from the example file on the website.

3.2 Loading dataset & prescreening or selecting methods

Following screen pops up when GRACOMICS is executed.



Getting started

1. What you need for GRACOMICS
   - JRE (Java Runtime Environment)
   - GRACOMICS jar file
   - Result data (TSV form)
2. Run GRACOMICS
   - Excute the GRACOMICS.jar in Windows OS
   - Input the command "Java –jar GRACOMICS.jar"
3. Load data
   - GRACOMICS can import tab-separated value (TSV) datasets which are tabular type and have following structure

Modules

1. Pairwise All Scatter Plots Module
2. Pairwise Detailed Scatter Plots Module
3. Multiple Results Comparison Module

Need help?

GRACOMICS web site

bibs.snu.ac.kr/software/GRACOMICS

GRACOMICS online manual

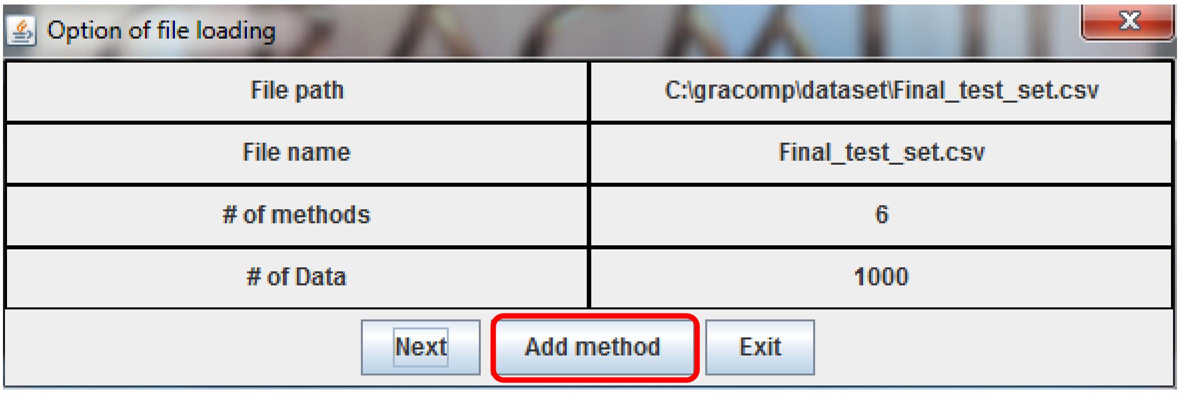
bibs.snu.ac.kr/software/GRACOMICS/manual.php

Direct web-links are provided in help menubar

<Figure 1> Main frame of the GRACOMICS program

This is the main page of the program which provides the total scheme of the program. GRACOMICS provides links to its given example datasets, by default, for tutorial purposes. General applicants can open files using the [File] -> [open] option on the menu bar (up top).

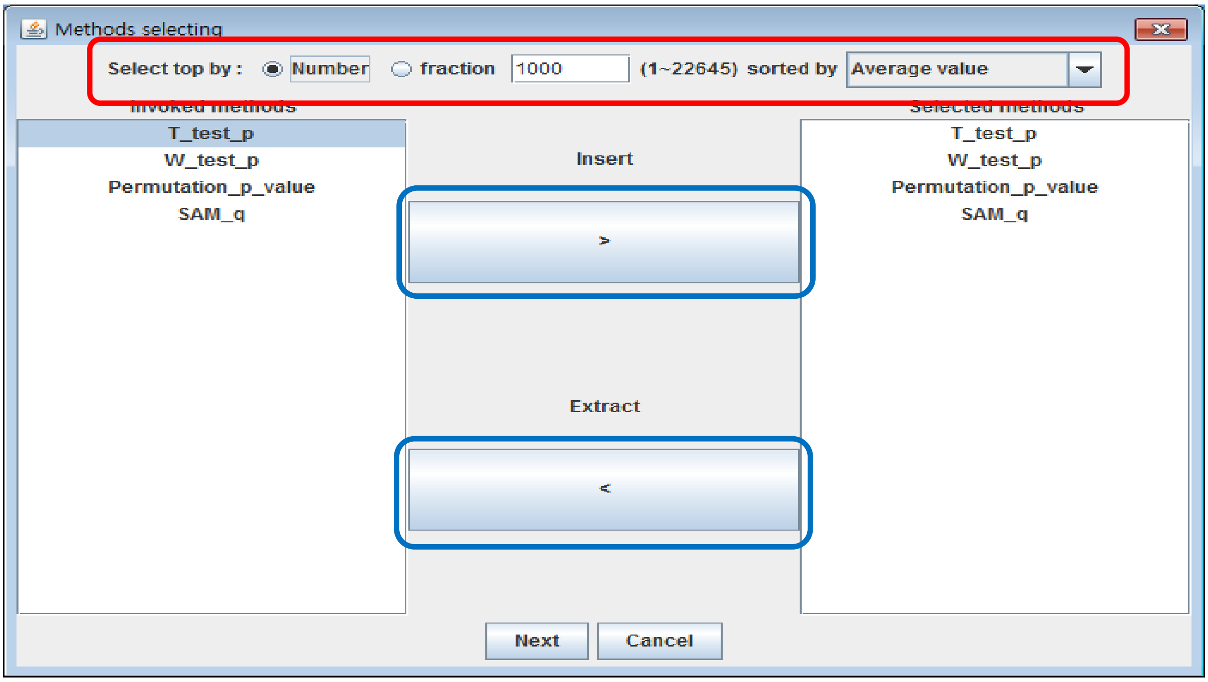
The files can be opened through the dialog box; the TSV files can have multiple methods or several files with single methods as shown in 3.1.



File path	C:\gracomp\dataset\Final_test_set.csv
File name	Final_test_set.csv
# of methods	6
# of Data	1000

<Figure 2> Dialog of file loading

When the file is opened, another dialog with a detailed description of the files will appear. [Add method] button, highlighted in red on the above picture, can be used to import more files. Press next if all methods are imported.



Select top by : ● Number ○ fraction 1000 (1~22645) sorted by Average value

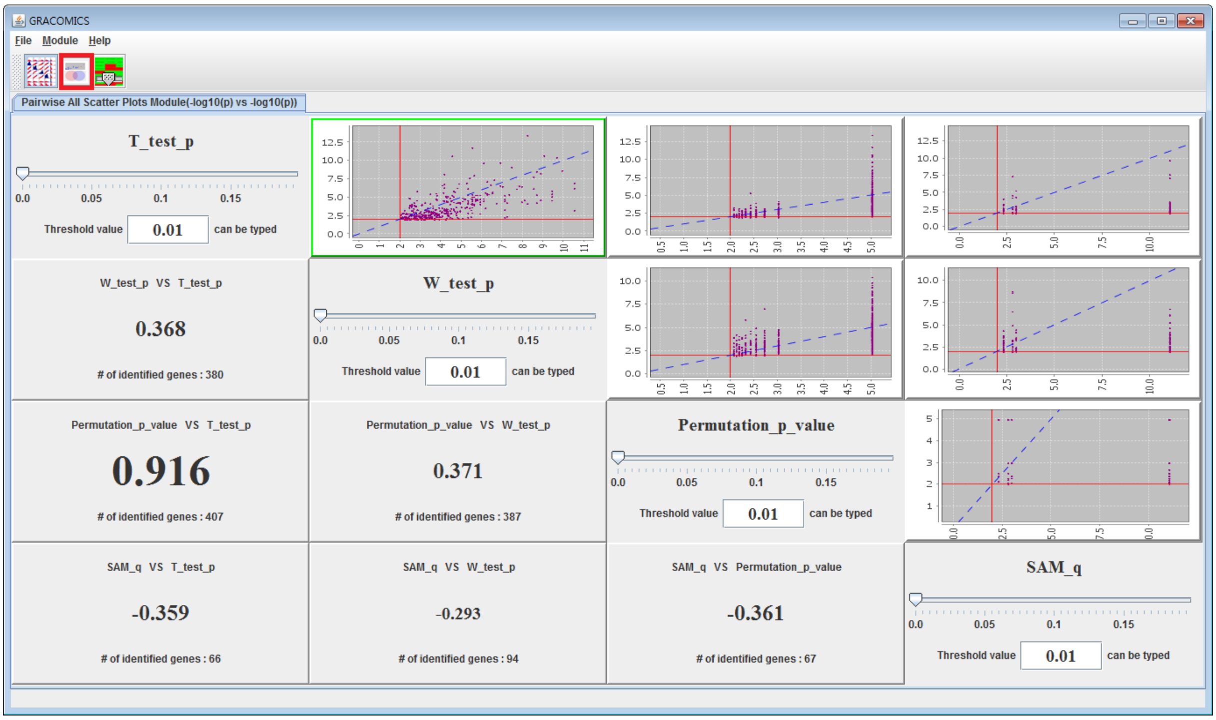
Invoked methods	Selected methods
T_test_p	T_test_p
W_test_p	W_test_p
Permutation_p_value	Permutation_p_value
SAM_q	SAM_q

<Figure 3> Method selection

In the method selection dialog, the user can select from the invoked pool of methods by shading the methods and pressing the insert button in blue. Also, the user may change the number of input data (preferably fewer than 5000 top

selection for smooth interactive nature of GRACOMICS). Usually in microarray data analysis, only a minority of the SNP data is selected; justifying 5000 is enough. Such process is called pre-screening; GRACOMICS provides top selection by average of all inputs and top selection via selected methods. As shown in the red box, top markers can be chosen by number or fraction; furthermore, these input-values can be sorted in various ways via comb-box, for a less constrained pre-screening process.

3.3 Pairwise All Scatter Plots Module (Pair-ASP)



T_test_p

Threshold value 0.01 can be typed

W_test_p VS T_test_p

0.368

# of identified genes : 380

W_test_p

Threshold value 0.01 can be typed

Permutation_p_value VS T_test_p

0.916

# of identified genes : 407

Permutation_p_value VS W_test_p

0.371

# of identified genes : 387

Permutation_p_value

Threshold value 0.01 can be typed

SAM_q VS T_test_p

-0.359

# of identified genes : 66

SAM_q VS W_test_p

-0.293

# of identified genes : 94

SAM_q VS Permutation_p_value

-0.361

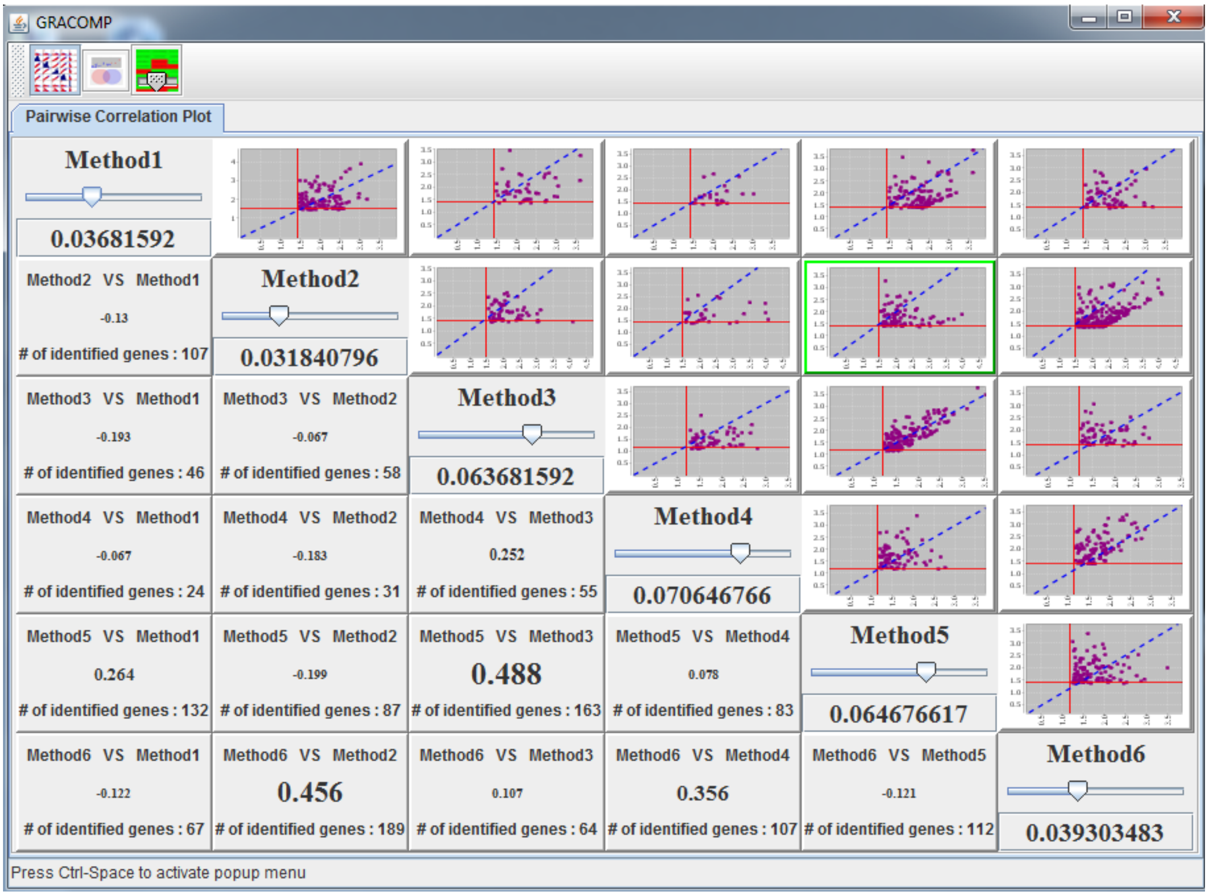
# of identified genes : 67

SAM_q

Threshold value 0.01 can be typed

<Figure 4> Pairwise All Scatter Plots Module (Pair-ASP)

If all the files and methods are selected, we can observe is the Pairwise Correlation Plot in the first module. This table provides correlation information of each method versus another in $-\log_{10}$ scale. Orthogonal panels contain their name of the method and a slide bar to control user defined value. Pair-ASP can provide interactive change in the scatter plots with simple dragging of the slide bar to desired user-defined value. On the panels south-western to the orthogonal panels, names of compared methods and the correlation-coefficient of the two methods are provided. Pair-ASP provides the correlation information of the input-methods and the significant results at a given user-defined value, the best part is, that it is interactive and global in a sense. To get more detailed view of the analyzed results, the user can use the Pairwise Detailed Scatter Plot Module (Pair-DSP) module either by clicking the button indicated in red in figure 4 or by clicking the plot itself as shown in figure 5.

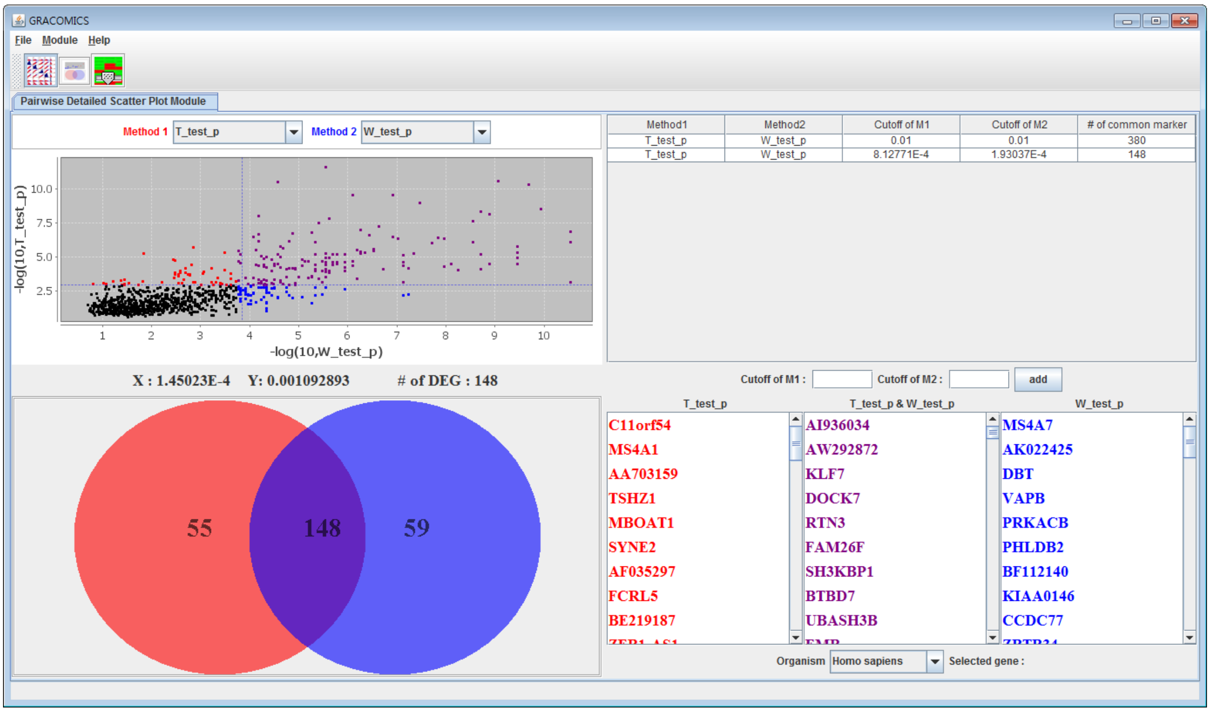


Method1 0.03681592					
Method2 VS Method1 -0.13 # of identified genes : 107	Method2 0.031840796				
Method3 VS Method1 -0.193 # of identified genes : 46	Method3 VS Method2 -0.067 # of identified genes : 58	Method3 0.063681592			
Method4 VS Method1 -0.067 # of identified genes : 24	Method4 VS Method2 -0.183 # of identified genes : 31	Method4 VS Method3 0.252 # of identified genes : 55	Method4 0.070646766		
Method5 VS Method1 0.264 # of identified genes : 132	Method5 VS Method2 -0.199 # of identified genes : 87	Method5 VS Method3 0.488 # of identified genes : 163	Method5 VS Method4 0.078 # of identified genes : 83	Method5 0.064676617	
Method6 VS Method1 -0.122 # of identified genes : 67	Method6 VS Method2 0.456 # of identified genes : 189	Method6 VS Method3 0.107 # of identified genes : 64	Method6 VS Method4 0.356 # of identified genes : 107	Method6 VS Method5 -0.121 # of identified genes : 112	Method6 0.039303483

Press Ctrl-Space to activate popup menu

<Figure 5>Interactive connection between Pair-ASP and Pair-DSP

3.4 Pairwise Detailed Scatter Plot Module (Pair-DSP)



Method 1 T_test_p Method 2 W_test_p

Method1	Method2	Cutoff of M1	Cutoff of M2	# of common marker
T_test_p	W_test_p	0.01	0.01	380
T_test_p	W_test_p	8.12771E-4	1.93037E-4	148

X : 1.45023E-4 Y: 0.001092893 # of DEG : 148

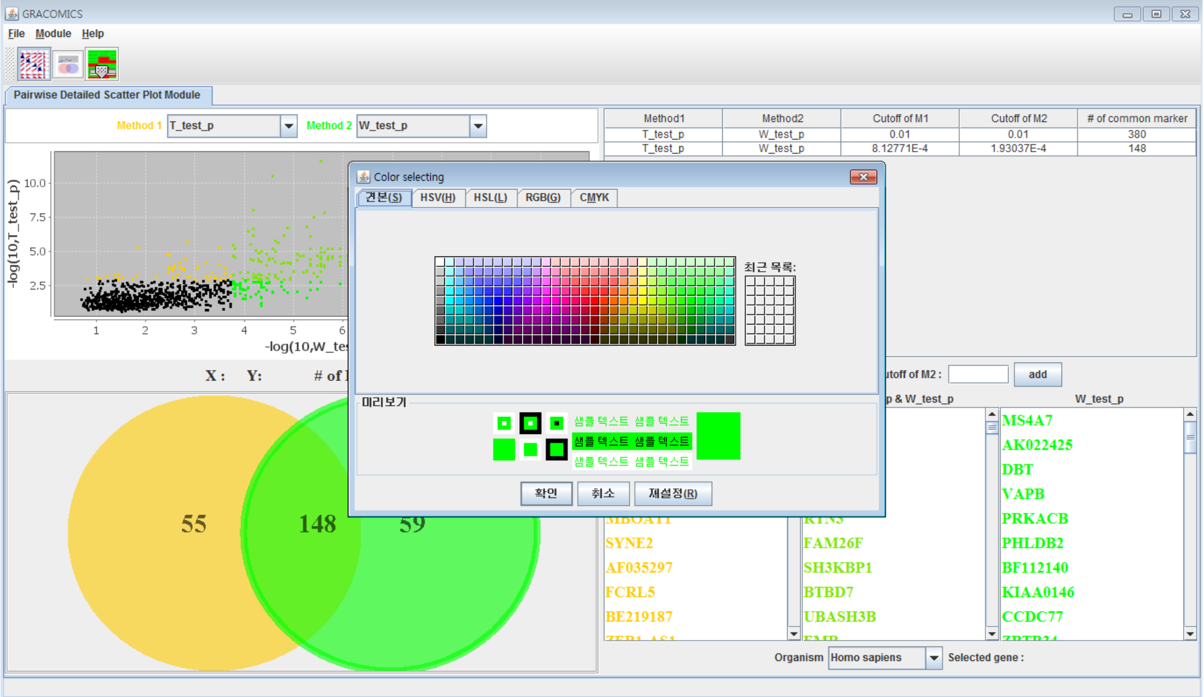
Cutoff of M1 :
Cutoff of M2 :

T_test_p	T_test_p & W_test_p	W_test_p
C11orf54	AI936034	MS4A7
MS4A1	AW292872	AK022425
AA703159	KLF7	DBT
TSHZ1	DOCK7	VAPB
MBOAT1	RTN3	PRKACB
SYNE2	FAM26F	PHLDB2
AF035297	SH3KBP1	BF112140
FCRL5	BTBD7	KIAA0146
BE219187	UBASH3B	CCDC77

Organism Homo sapiens Selected gene :

<Figure 6> Pairwise Detailed Scatter Plot Module (Pair-DSP)

By entering Pair-DSP using any of the two ways stated above, screen similar to figure 6 will pop up. Pair-DSP provides much more zoomed in scatter plot and select out the significant results much more easily. Using simple mouse events on the scatter plot, the user can manipulate the cutoffs which changes the resulting table and plots. Venn diagram on the left and the table on the right summarize the results and present to the user in an organized manner. For microarray data analysis, the list would contain the symbols of the significant genes on the table, and show the corresponding Venn diagram on the left. As seen in figure 6 and 7 the methods are color defined, the user can change the color that soothes users taste.



Method 1 T_test_p
Method 2 W_test_p

Method1	Method2	Cutoff of M1	Cutoff of M2	# of common marker
T_test_p	W_test_p	0.01	0.01	380
T_test_p	W_test_p	8.12771E-4	1.93037E-4	148

utoff of M2 :

SYNE2
AF035297
FCRL5
BE219187

p & W_test_p

FAM26F
SH3KBP1
BTBD7
UBASH3B

W_test_p

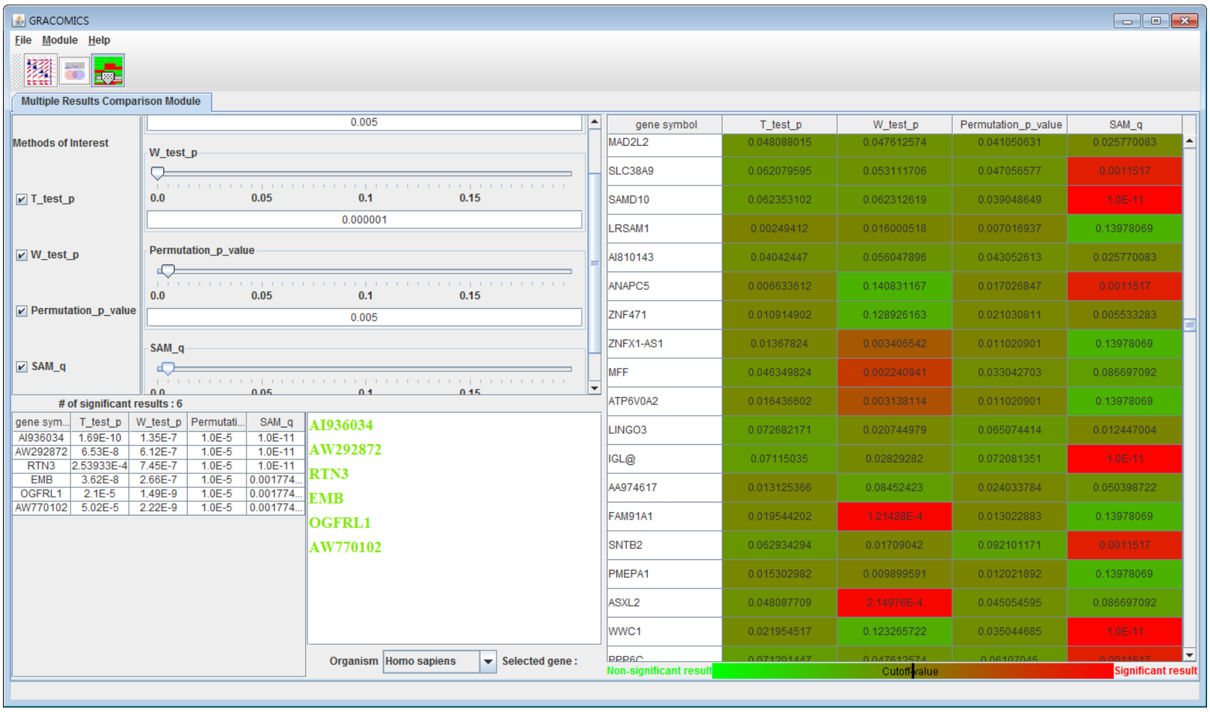
MS4A7
AK022425
DBT
VAPB
PRKACB
PHLDB2
BF112140
KIAA0146
CCDC77

Organism Homo sapiens

<Figure 7>Dialog to select expressing the methods color

3.5 Multiple Results Comparison Module (Multi-RC)

So far, Pair-ASP, and Pair-DSP were used to compare two statistical methods at a time; however, Multi-RC is a module to compare several methods at once.



☑ T_test_p
☑ W_test_p
☑ Permutation_p_value
☑ SAM_q

0.005

W_test_p: 0.000001

Permutation_p_value: 0.005

SAM_q

# of significant results : 6

gene sym...	T_test_p	W_test_p	Permutati...	SAM_q
AI936034	1.69E-10	1.35E-7	1.0E-5	1.0E-11
AW292872	6.53E-8	6.12E-7	1.0E-5	1.0E-11
RTN3	2.53933E-4	7.45E-7	1.0E-5	1.0E-11
EMB	3.62E-8	2.66E-7	1.0E-5	0.001774...
OGFRL1	2.1E-5	1.49E-9	1.0E-5	0.001774...
AW770102	5.02E-5	2.22E-9	1.0E-5	0.001774...

AI936034
AW292872
RTN3
EMB
OGFRL1
AW770102

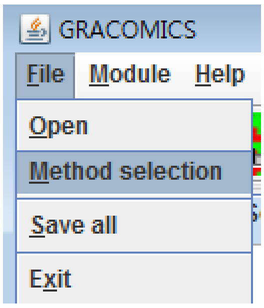
Organism: Homo sapiens
Selected gene :

gene symbol	T_test_p	W_test_p	Permutation_p_value	SAM_q
MAD2L2	0.048088015	0.047612574	0.041050631	0.025770083
SLC38A9	0.062079595	0.053111706	0.047056577	0.0011517
SAMD10	0.062353102	0.062312619	0.039048649	1.0E-11
LRSAM1	0.00249412	0.016000518	0.007016937	0.13978069
AI810143	0.04042447	0.056047896	0.043052613	0.025770083
ANAPC5	0.006633612	0.140831167	0.017026847	0.0011517
ZNF471	0.010914902	0.128926163	0.021030811	0.005533283
ZNFX1-AS1	0.01367824	0.003406542	0.011020901	0.13978069
MFF	0.046349824	0.002240941	0.033042703	0.086697092
ATP6V0A2	0.016436602	0.003138114	0.011020901	0.13978069
LINGO3	0.072682171	0.020744979	0.065074414	0.012447004
IGL@	0.07115035	0.02829282	0.072081351	1.0E-11
AA974617	0.013125366	0.08452423	0.024033784	0.050398722
FAM91A1	0.019544202	1.21428E-4	0.013022883	0.13978069
SNTB2	0.062934294	0.01709042	0.092101171	0.0011517
PMEPA1	0.015302982	0.009899591	0.012021892	0.13978069
ASXL2	0.048087709	2.14976E-4	0.045054595	0.086697092
WWC1	0.021954517	0.123265722	0.035044685	1.0E-11

<Figure 8> Multiple Results Comparison Module (Multi-RC)

As shown on figure 8, Multi-RC provides a Heat Map Table comparison of methods at a glance. By using the slide bar to manipulate the cutoffs for each method, the user can change the significant outcomes from independent cutoff values. The figure is an output from a microarray data; the genes are coded in gradation of colors between red and green, which indicates the marker′ s distance from the cut-off value using color intensities. The results of the signal table are summarized in a table located on the bottom left. This module interactively compares n-number of methods as a whole.

3.6 Options on the menu bar

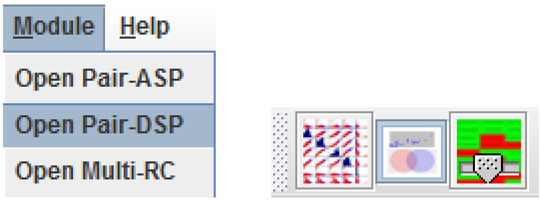


<Figure 9> [File] -> [Method selection] function

This option provides adding and subtracting of the methods in the current analysis. It is usually used for re-plotting the Pair-ASP to see global trend, or to exclude wrongfully included methods.

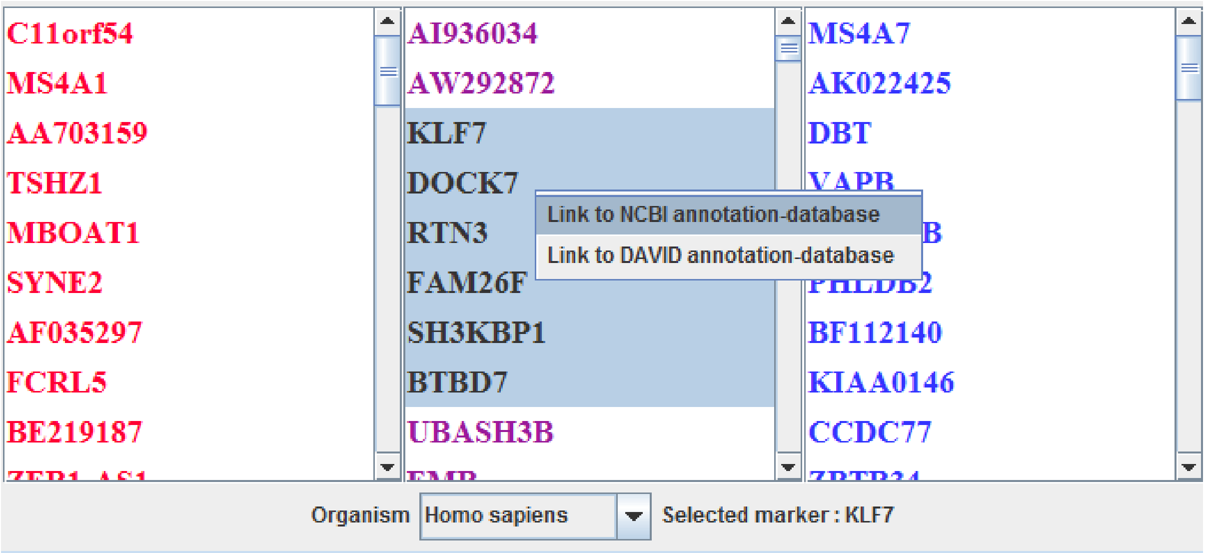
<Figure 10> [File] -> [Save all] function

This option saves all outputs generated in the current session; yes, all outputs can be individually saved, but this option is for convenience purposes.



<Figure 11> [Module] menu

3.6 Options on the menu



<Figure 12> Simple annotation function

GRACOMICS provides a direct link to the web-based open database, such as NCBI and DAVID, for the selected significant results. New window with results will appear if NCBI database is chosen, while for the DAVID query web-page, a new window will pop up with the marker-list copied on the clipboard.

5. Contact

This program can be downloaded on the following website:

http://bibs.snu.ac.kr/software/GRACOMICS/gracomics.php

The author will provide a reply to all questions, comments, and requests on the program, made to the following e-mail address.

nijorral@snu.ac.kr