Expression Predictor version 4.0 (ExP)

ExP is a java-based software for microarray data analysis.
Credit to Dr. Jeff Chou, National Institute of Environmental Health and Science
Contact information: chou@niehs.nih.gov
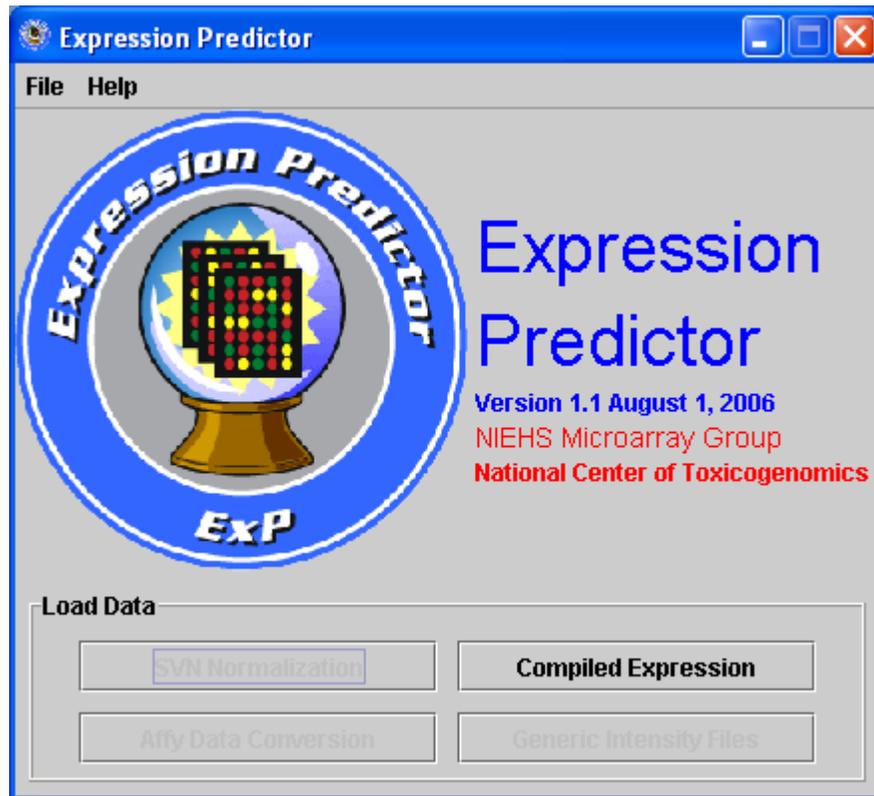
Compiled Expression –
Usage - Perform microarray gene expression data analysis
Load and open a tab delimited compiled (log2 ratio) text file.
Sample file - SampleData.txt;
Pop up the interface shown below:

SampleData.txt

File Analysis Data Help

Sample Name Row (size):	0	Data Start Column:	-1
Bio Replicate Row(n above threshold):	-1	Data Start Row:	-1
Cy3/Cy5 Label Row(r threshold):	-1	Cell Line Name Row:	-1
Tech Replicate Row (column number):	-1	Cell Line Profile Row:	-1

Click on Column and Select Column Name

column 0	column 1	column 2	column 3	column 4	column 5	column 6
Sample	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...
BioReplicate	UV_sham_6h	UV_sham_6h	UV_2h	UV_2h	UV_6h	UV_6h
TechReplic...	F1_UV_sha...	F1_UV_sha...	F1_UV_2h	F1_UV_2h	F1_UV_6h	F1_UV_6h
Label	C3	C5	C3	C5	C3	C5
GeneID	F1	F1	F1	F1	F1	F1
1	-0.046696782	0.046696812	0.030646205	0.13795961	0.006732792	0.08377281
2	0.054012865	-0.05401288	0.16290237	0.07522814	0.19827135	0.044337735
3	0.1443618	-0.14436185	0.08035207	-0.064905465	0.11411697	-0.07437456
4	0.022420406	-0.022420406	-0.2559638	0.02268219	0.12662931	0.28083417
5	0.18099713	-0.1809971	-0.07456559	-0.13778028	-0.02415818	-0.2082063
6	-0.46131432	0.46131432	-0.39849812	0.62430674	-0.35597903	0.73895085
7	-0.010273963	0.010273993	-0.12052919	0.041918784	-0.09796527	-0.063718796
8	-0.029606283	0.029606283	-0.003385067	0.063922346	0.111272454	0.029764593
9	-0.045791864	0.045791864	0.023365736	0.16324155	0.025440037	0.19879265
10	0.05489844	-0.05489844	0.03900063	-0.15642853	0.45993918	0.23945796
11	0.2763187	-0.27631867	-0.01924783	-0.5560247	0.4243609	-0.1313864
12	0.028856635	-0.028856635	0.6823792	0.5369602	0.3291458	0.19542366
13	0.0671255	-0.06712556	0.048854113	-0.14271098	-0.0238626	-0.13167113
14	0.13097304	-0.13097306	0.14296263	0.031564564	0.087766826	0.03199944
15	-0.13486221	0.13486221	-0.83240724	-0.5489509	-0.40199488	-0.3023231
16	0.029723316	-0.029723316	-0.033204883	0.064878374	-0.021936297	0.11216915
17	0.011399657	-0.011399627	-0.047936082	-0.03531015	0.06708273	0.08328408
18	-0.027516425	0.027516425	0.20526785	0.29785347	0.15431207	0.2547446
19	-0.6608701	0.66087	-0.22422123	-0.30219984	-1.5673342	0.2814893
20	0.6793933	-0.6793933	0.97240543	0.35104322	0.4290147	0.20181033
21	0.018202052	-0.018202066	-0.06574224	-0.076055765	0.04300806	0.010858715

(1) Highlight the cell and right click on mouse on "Select Data Start Column" to select the data start column. Again click on "Select Data Start Row" to the data start row.

SampleData.txt

File   Analysis   Data   Help

Sample Name Row (size): 0   Data Start Column: -1

Bio Replicate Row(n above threshold): -1   Data Start Row: -1

Cy3/Cy5 Label Row(r threshold): -1   Cell Line Name Row: -1

Tech Replicate Row (column number): -1   Cell Line Profile Row: -1

Click on Column and Select Column Name

	column 0	column 1	column 2	column 3	column 4	column 5	column 6
	Sample	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...	F1-hTERT_...
	BioReplicate	UV_sham_6h	UV_sham_6h	UV_2h	UV_2h	UV_6h	UV_6h
	TechReplic...	F1_UV_sha...	F1_UV_sha...	F1_UV_2h	F1_UV_2h	F1_UV_6h	F1_UV_6h
	Label	C3	C5	C3	C5	C3	C5
	GeneID	F1	F1	F1	F1	F1	F1
1		-0.046				05961	0.006732792	0.08377281
2		0.054				22814	0.19827135	0.044337735
3		0.144				905465	0.11411697	-0.07437456
4		0.022				68219	0.12662931	0.28083417
5		0.180				78028	-0.02415818	-0.2082063
6		-0.461				30674	-0.35597903	0.73895085
7		-0.010				918784	-0.09796527	-0.063718796
8		-0.029				322346	0.111272454	0.029764593
9		-0.045				24155	0.025440037	0.19879265
10		0.054				42853	0.45993918	0.23945796
11		0.276				0247	0.4243609	-0.1313864
12		0.028				3602	0.3291458	0.19542366
13		0.0671255	-0.06712556	0.048854113	-0.14271098	-0.0238626	-0.13167113
14		0.13097304	-0.13097306	0.14296263	0.031564564	0.087766826	0.03199944
15		-0.13486221	0.13486221	-0.83240724	-0.5489509	-0.40199488	-0.3023231
16		0.029723316	-0.029723316	-0.033204883	0.064878374	-0.021936297	0.11216915
17		0.011399657	-0.011399627	-0.047936082	-0.03531015	0.06708273	0.08328408
18		-0.027516425	0.027516425	0.20526785	0.29785347	0.15431207	0.2547446
19		-0.6608701	0.66087	-0.22422123	-0.30219984	-1.5673342	0.2814893
20		0.6793933	-0.6793933	0.97240543	0.35104322	0.4290147	0.20181033
21		0.018202052	-0.018202066	-0.06574224	-0.076055765	0.04300806	0.010858715

Pop-up menu (overlaying columns 2–4):
- **Select Data Start Column**
- **Select Data Start Row**
- **Replicate Row**
- **Sample Name Row**
- **Dye Swap Correction Row Selection** ▶
- **Cell Line Row Selection** ▶
- **Plot Selected Rows Data** ▶
- **Row Profile Analysis** ▶
- **Internet Search** ▶

(2) Highlight the cell where bio replicate assigned and right click on mouse on "Replicate Row" to select the biological replicate row

(3) Pull down "Analysis" and click left mouse on "EPIG" to pop up EPIG interface.

```
    (4)  Now you get EPIG interface.
Initially the pattern list includes all the genes.
Click left mouse on "Pattern Extraction (red)" – EPIG starts working now. Wait 5
minutes or so. No need change the parameters.
```
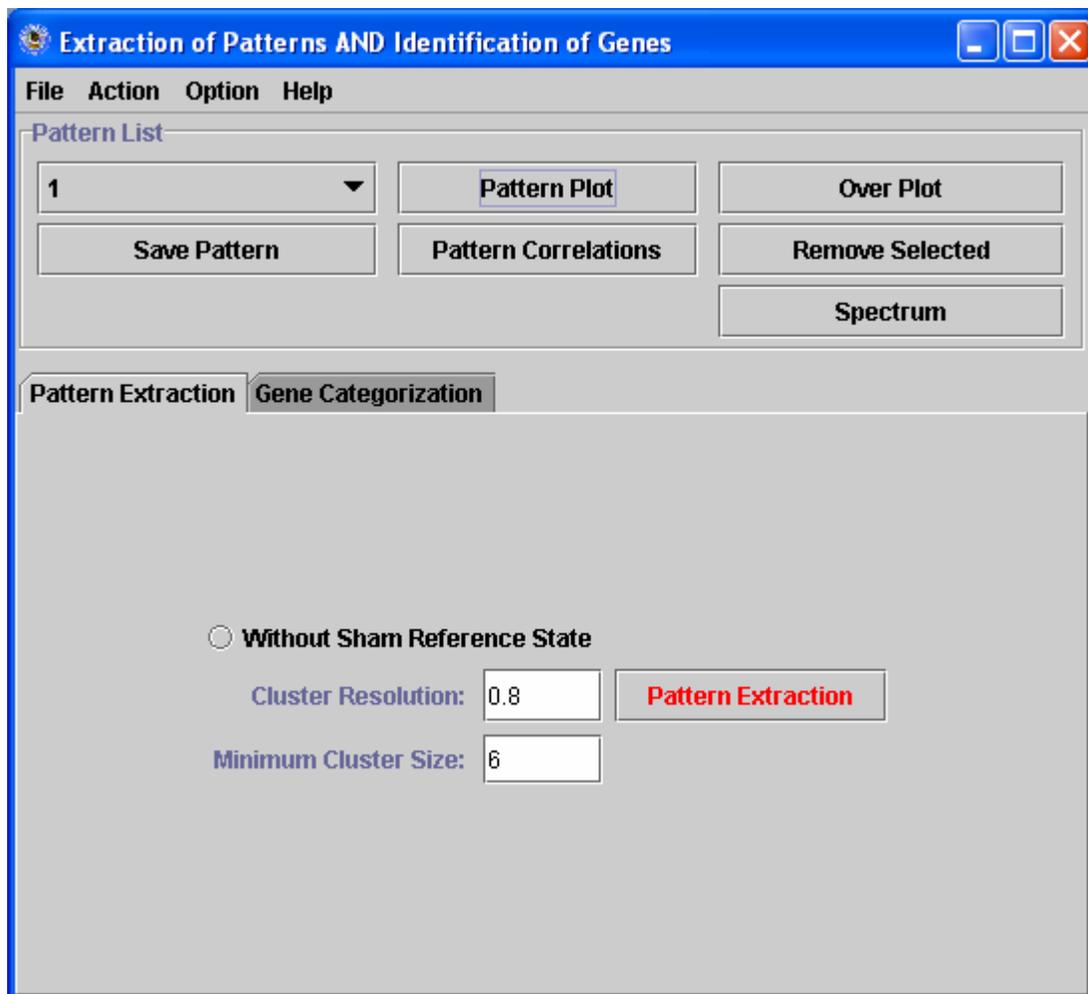
At the end of run, you get this. It extracts 18 patterns (left click on the pattern name to see the list).
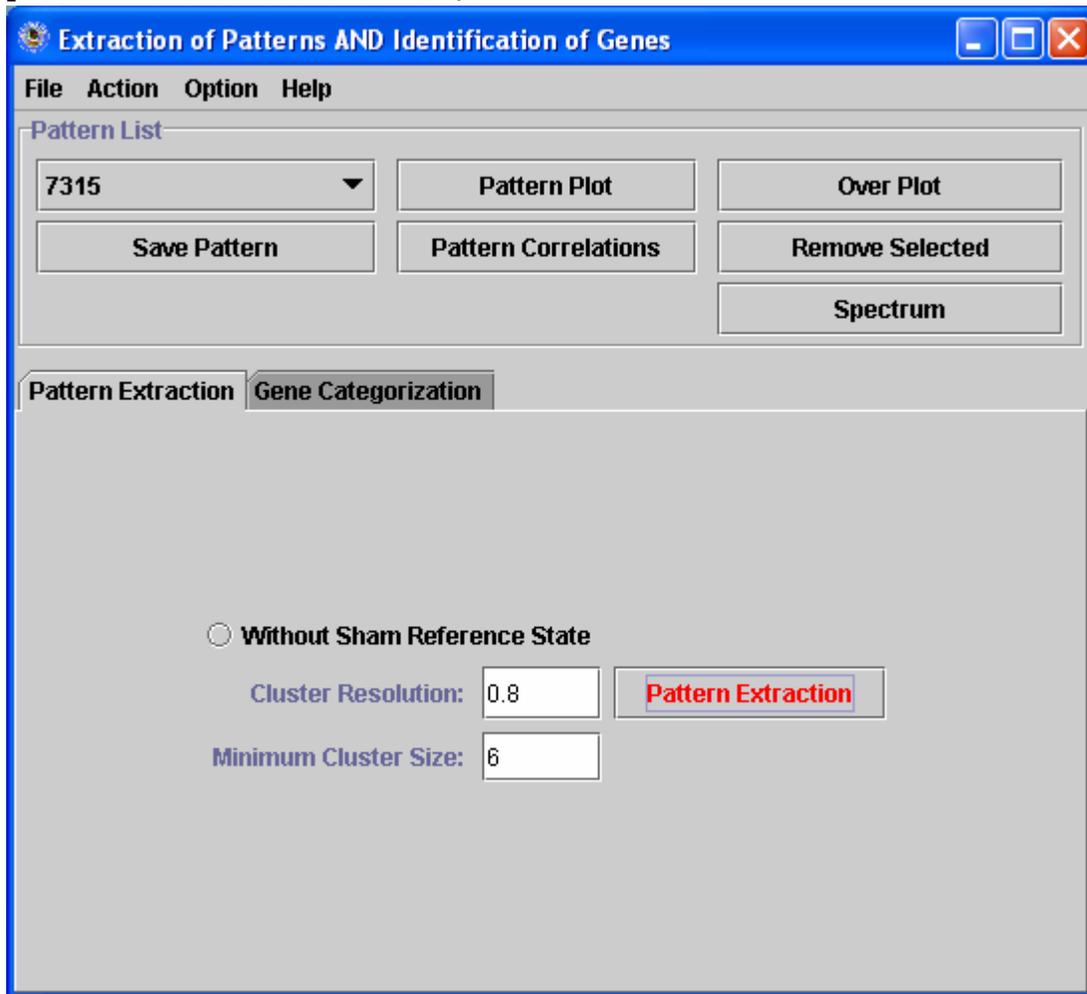
Also poped-up is the graphic display of the patterns below.
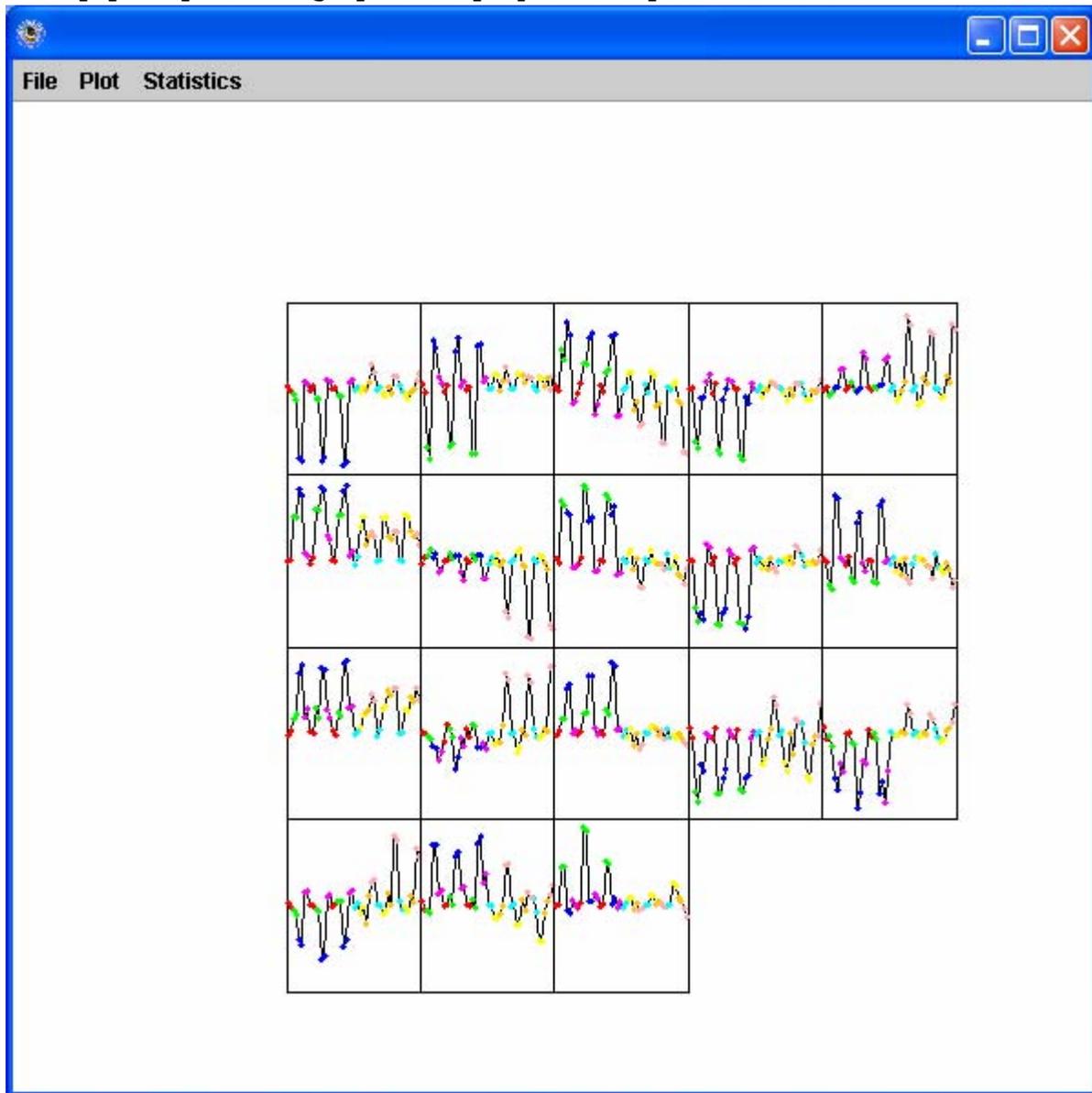
Option – before doing "Gene Categorization", you may check each of the extracted patterns. Patterns are listed according to their SNR values.
"Pattern Plot" is to plot the selected pattern
"Correlation" is to get pattern-wised correlation r-values displaying in message board.
"Pattern Save" is to save the extracted patterns with a given name.
"Remove Selected" is to remove the selected pattern from the list if you consider this pattern is too noisy or too similar to another pattern by checking their correlations, if so, you may remove the one which is lower (i.e. lower SNR value) in the list.

Click on "Gene Categorization" to get the interface below.

Then click on "Display Selection" to pup-up three data sets in three tables. The selected genes meet the criteria Magnitude > 0.5, SNR > 3 and r-value > 0.64 (i.e. 0.8*0.8). With the given criteria, EPIG will pop-up three tables.
One is selected and correlated genes (there are 2661 genes in the given sample data), which are meet all the three criteria.
Another is selected but low correlated genes (there are 65 genes), which meet Magnitude and SNR criteria, but not r-value.
The third is unselected gene list (there are 14360 genes), which do not meet all the criteria.

There are 2661 selected genes are correlated..

File   Analysis   Plot

Save Data
Save in Seperate Files
Close

		Correlation	PattenName2	Correlation2
		0.99222565	5013/878	0.8579949
		0.99092865	5013/878	0.889452
		0.989847	5013/878	0.84245855
5538	7315/5538	0.98833996	5013/878	0.8798103
7315	7315/12311	0.98761064	5013/878	0.8860135
8057	7315/8057	0.9830175	5013/878	0.833999
5275	7315/5275	0.98031986	10341/11613	0.8234075
9547	7315/9547	0.97886443	5013/878	0.8393851
12311	7315/12311	0.9759975	5013/878	0.84799325
12923	7315/12923	0.9757885	5013/878	0.8632866
8440	7315/8912	0.9755967	5013/878	0.8368473
3479	7315/3479	0.9740267	11291/13521	0.8612664
8543	7315/8543	0.9733468	11291/13521	0.8557763
5767	7315/5767	0.9729082	5013/878	0.8172825
8632	7315/8632	0.97153556	5013/878	0.84368
682	7315/682	0.9705987	5013/878	0.8256601
8384	7315/8384	0.97041476	5013/878	0.8254891
2682	7315/2682	0.970231	5013/878	0.7977683
13333	7315/13333	0.9702078	5013/11991	0.8903974
6372	7315/6372	0.96980506	5013/878	0.8222407
11805	7315/11805	0.969055	5013/878	0.88849086
12814	7315/12814	0.96834815	5013/11991	0.9318951
2668	7315/2668	0.9676526	11291/13521	0.84999985
681	7315/5767	0.9676502	10341/11613	0.8035436
10086	7315/3479	0.9675084	11291/13521	0.88005406
9170	7315/12923	0.9668191	5013/535	0.87292665
14423	7315/15528	0.96598506	5013/13787	0.89917904
5963	7315/5963	0.965807	5013/878	0.8247796
8884	7315/8884	0.96498	11291/13521	0.82018363

You can save the displayed data by clicking on "Save Data" with a given file name.
In the case of the selected and correlated genes, you can save each pattern associated data by clicking on "Save in Separate Files" with default names. The output files are named e.g. like this "selectedGenesInPattern_1_616.txt", which corresponds to pattern 1 and 616 genes in this pattern.

Columns in the Output files:
Column F – Magnitude is log2 ratio, you may convert it to ratio by taking 2-based exponential to it.
Column H – SNR as defined in EPIG paper.
Column I is a p value corresponding to SNR.
Column L is a p value corresponding to one-way ANOVA.
Column C is correlation r-value corresponding to associated pattern (Column B).
From Column F to the last column are the input data.