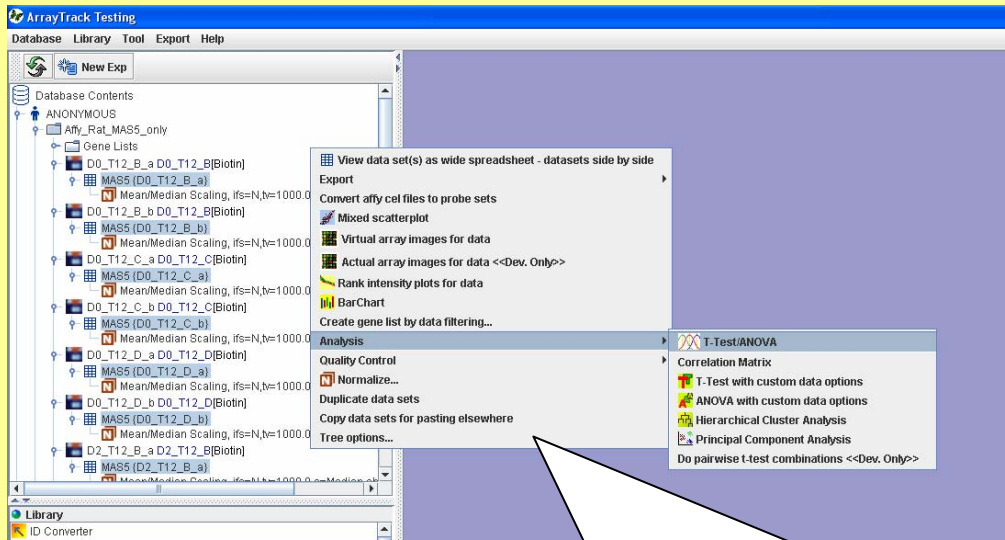


Tutorial 1:

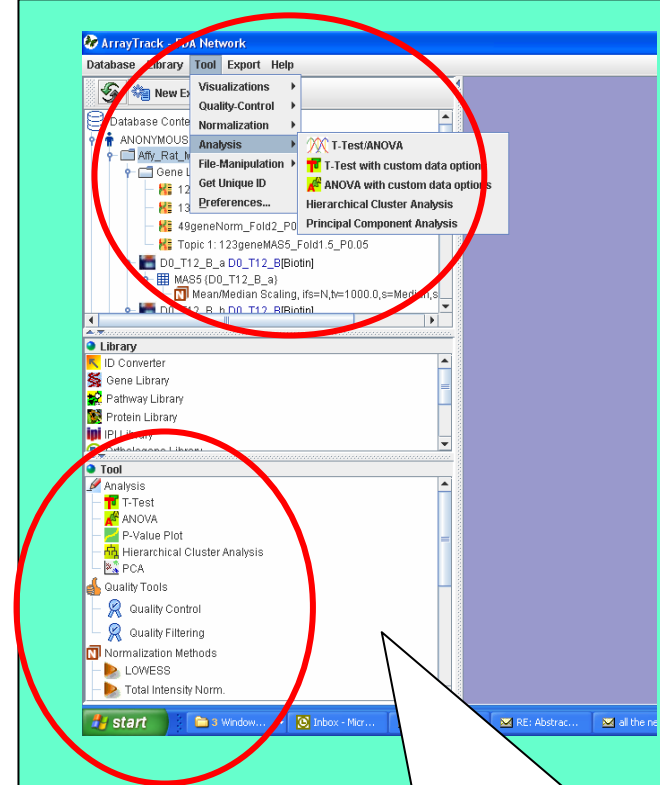
Comparing two groups (e.g.,
treated vs control groups)



Useful Tips



Right-click on a (set of) selected dataset(s) under the Database Contents tree to access the applicable TOOL functions. **This is one of the most frequently used functions.**



Many functions in ArrayTrack are accessible from multiple paths, for example, left-side window panels, pull down menus and right mouse-click options. **(Analyze the external datasets by accessing the tools from the left window panel)**

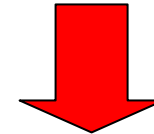
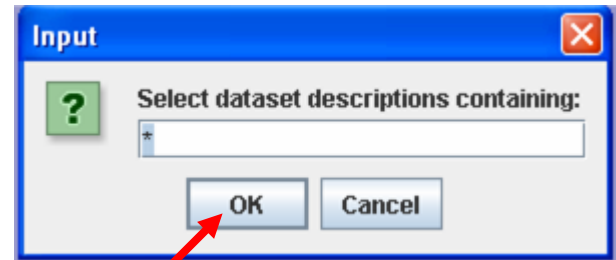
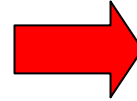
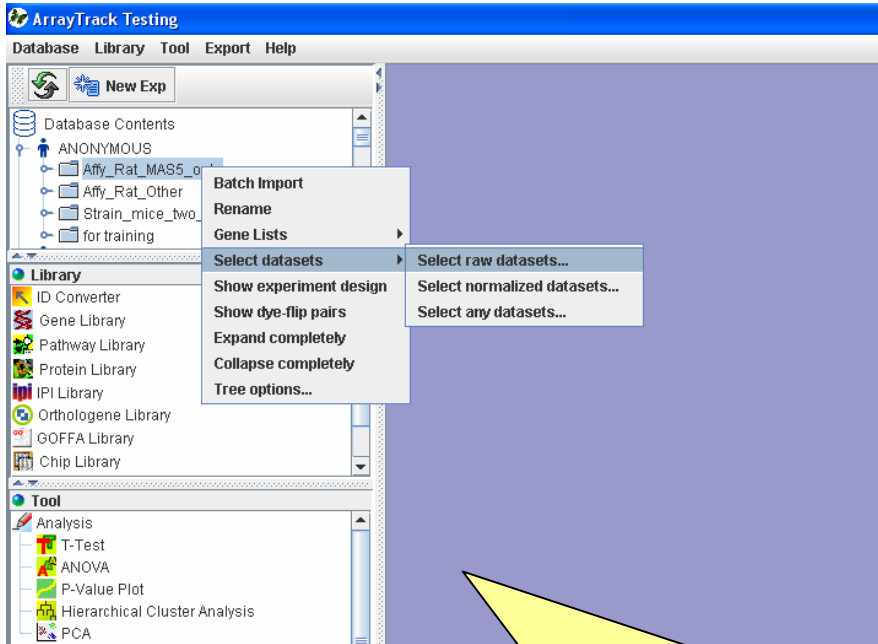
Bookmark the website for the convenience of accessing ArrayTrack **(Recommended)**



Other Tips

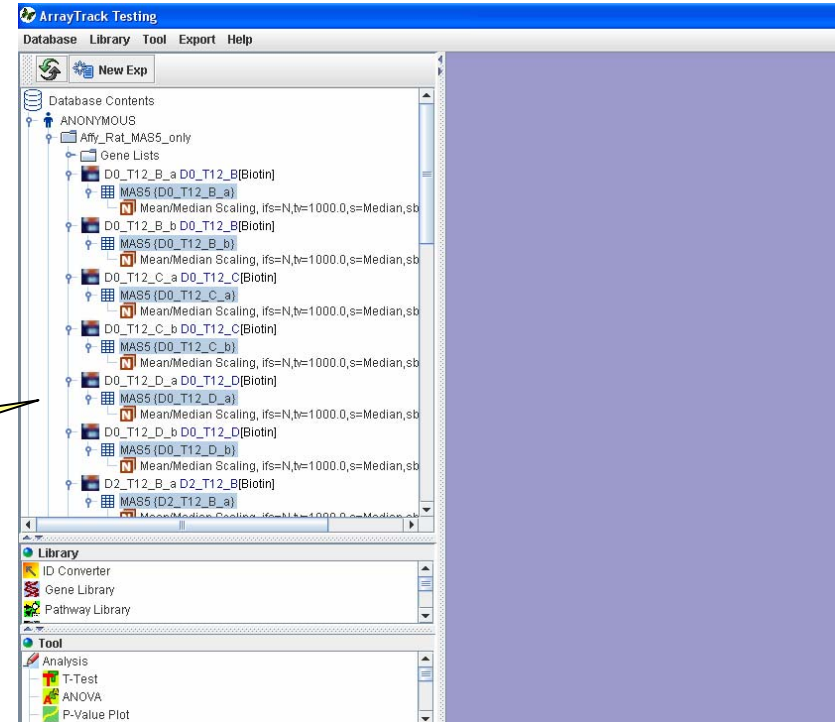
- Multiple sets of arrays can be selected by a combination of mouse-click and SHIFT-CTRL keys.
- Most functions come with default parameter settings. If you do not know a better setting, use the default.
- All Spreadsheet viewers share similar functions, e.g. Copy/Paste of selected table content.
- You can find ArrayTrack websites by searching the keyword “ArrayTrack” in any web searching engine (e.g., Google)

Step 1 - Select “Dataset” for Analysis

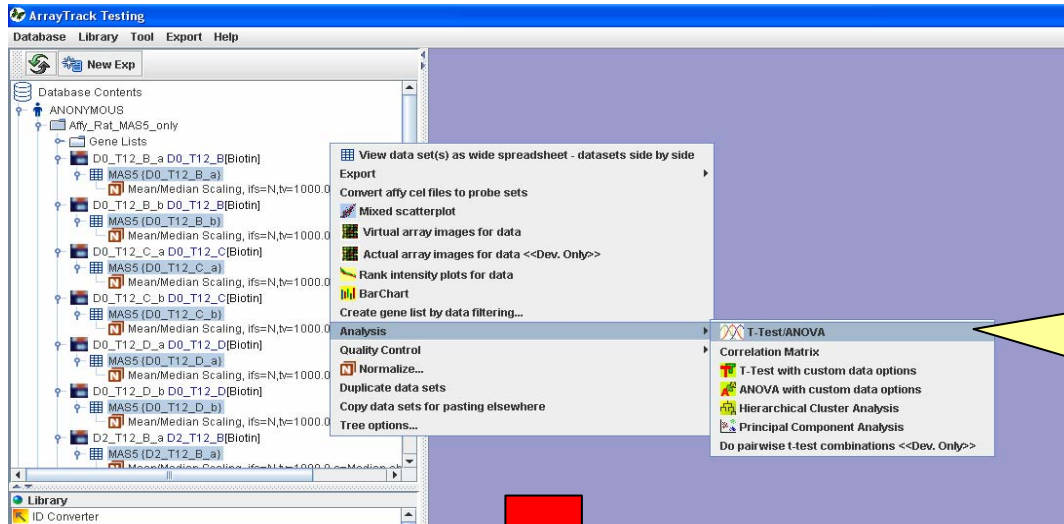


Right click the experiment “Affy_Rat_MAS5_Only” and select “select datasets” then “select raw dataset...”

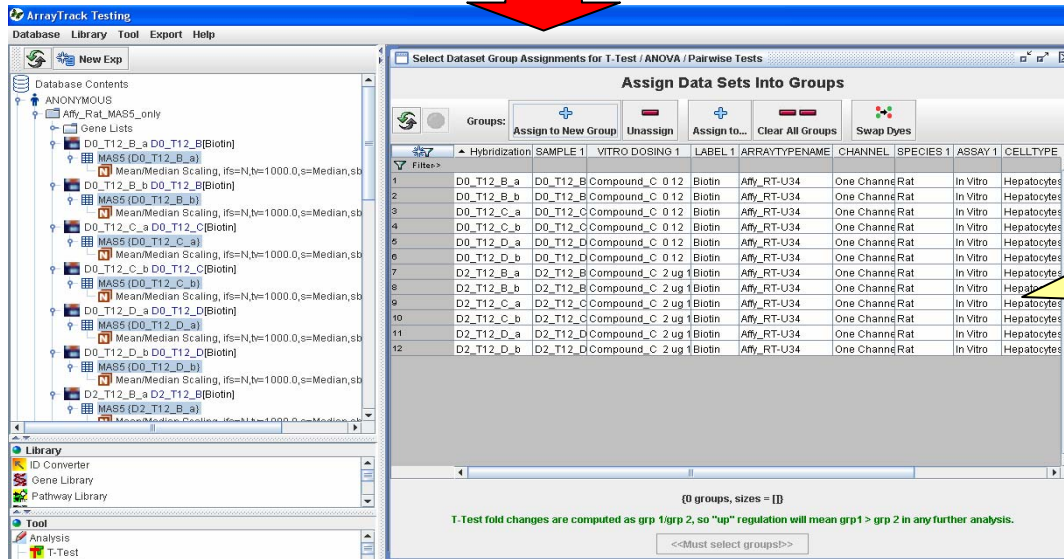
- 12 arrays are highlighted.
- Move the mouse to any place of this panel (or window) and right click, the analysis options are shown in the next slide



Step 2 - Select an analysis method



- Select “T-Test/ANOVA”
- Note: there are other analysis methods available in this menu.



A table is pop-up with the array on the row and the experimental description on the column.

Step 3 – Assign arrays into groups

Select Dataset Group Assignments for T-Test / ANOVA / Pairwise Tests

Assign Data Sets Into Groups

Groups **Assign to New Group** Unassign Assign to... Clear All Groups Swap Dyes

Filter->	Hybridization	SAMPLE ID	VITRO DOSING	LABEL	ARRAYTYPE	CHANNEL	SPECIES	ASSAY	CELLTYPE	
1	2	D0_T12_B_a	D0_T12_B	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
2	2	D0_T12_B_b	D0_T12_B	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
3	2	D0_T12_C_a	D0_T12_C	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
4	2	D0_T12_C_b	D0_T12_C	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
5	2	D0_T12_D_a	D0_T12_D	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
6	2	D0_T12_D_b	D0_T12_D	Compound_C 0 12	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
7	1	D2_T12_B_a	D2_T12_B	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
8	1	D2_T12_B_b	D2_T12_B	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
9	1	D2_T12_C_a	D2_T12_C	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
10	1	D2_T12_C_b	D2_T12_C	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
11	1	D2_T12_D_a	D2_T12_D	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes
12	1	D2_T12_D_b	D2_T12_D	Compound_C 2 ug 1	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	Hepatocytes

2 groups, sizes = [6, 6]

T-Test fold changes are computed as $grp\ 1/grp\ 2$, so "up" regulation will mean $grp\ 1 > grp\ 2$ in any further analysis.

Next >

There are total 12 arrays. The arrays whose name starting with D2 are the treated sample while those starting with D0 are the controls

- Select the arrays starting with D0 (treated samples, from 7 to 12) and click "Assign to New Group"
- Repeat the above process for the 1st 6 arrays

Pay attention on this statement and click "next"

Step 4 – Run “T-Test” with different options

Select Dataset Group Assignments for T-Test / ANOVA / Pairwise Tests

Test Type (Consistent with group selections)

T-Test

T-Test Options

P values from dist.: Welch t-test Simple t-test One class vs. mean:

P values from permutations: All Limit to:

Filtering with a gene list

Only include genes from gene list

Gene identifiers to include

Genbank Acc Gene Mfr ID LOCUSID UNIGENEID GENENAME

CLONEID GEN_DESCR_MFR REFSEQ SPOTID

Data options

Subtract backgrounds when present (raw datasets only)

Apply logarithm to expression values

Exclude spots flagged as bad

Run t-test on a predefined gene list (see Topic 6)

Select “Do Test” if you don’t know which option should be used.

Step 5 – Determine differentially expressed genes

File	Selected-Spot	All Spots	Advanced						
	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold (
1	AA108277	AA108277...	298444	Hsp105_pr...		516484	0.0006	4.1286	4.128
2	AA108308	AA108308...	314856			516485	0.8507	1.0548	1.054
3	AA108308	AA108308...				516486	0.0617	1.2367	0.808
4	AA684537	AA684537...	294964	Ndutf5_pr...		516487	0.0939	1.2459	0.802
5	AA684929	AA684929...				516488	0.2649	1.2663	0.789
6	AA684960	AA684960...				516489	0.3371	1.2892	0.775
7	AA684963	AA684963...	293702	LOC293702		516490	0.5633	1.0773	0.928
8	AA685112	AA685112...	293652	Ndufs8_pr...		516491	0.1383	1.2356	0.809
9	AA685152	AA685152...	25490	Nedd8		516492	0.1094	1.7581	0.568
10	AA685376	AA685376...				516493	0.4103	1.2636	0.791
11	AA685876	AA685876...	304805	Nqo3a2_pr...		516494	0.0184	1.4646	0.6
12	AA685903	AA685903...	362862	Tra1_predi...		516495	0.3889	1.1378	
13	AA686031	AA686031...	301458	Ndufs1		516496	0.14	1.2384	0.807
14	AA686579	AA686579...	301442	Sumo1_pr...		516497	0.9218	1.0433	1.043

1031 genes

Significance Filtering

P Values <

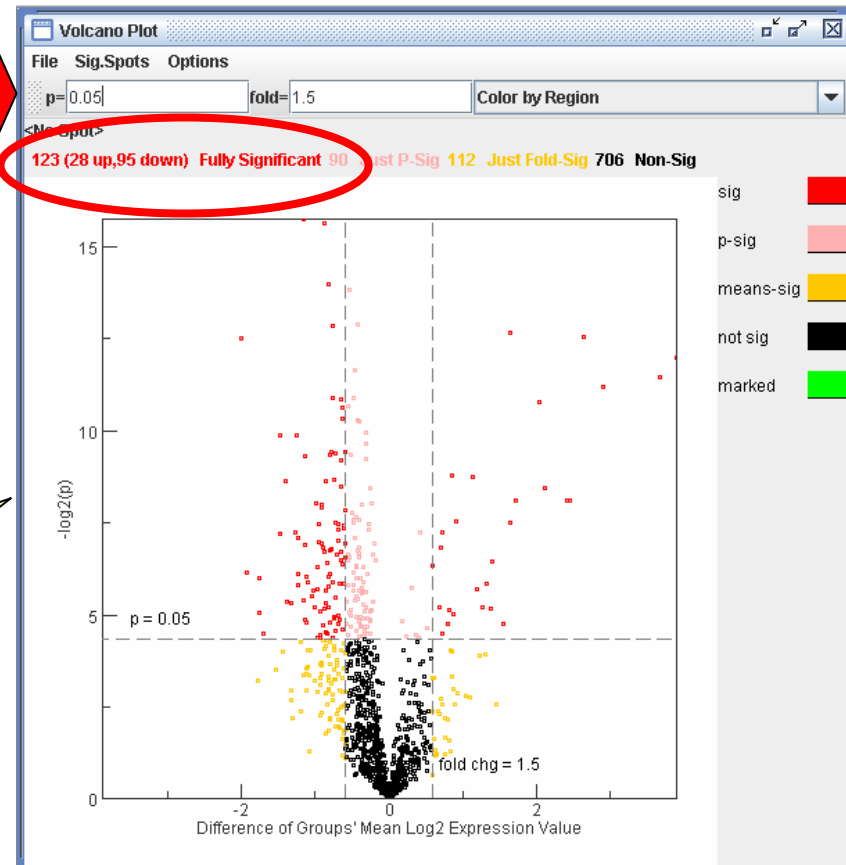
Target False Discovery Rate (FDR):

Select # genes by lowest p-values

Mean Channel Intensities > Bad Flags <=

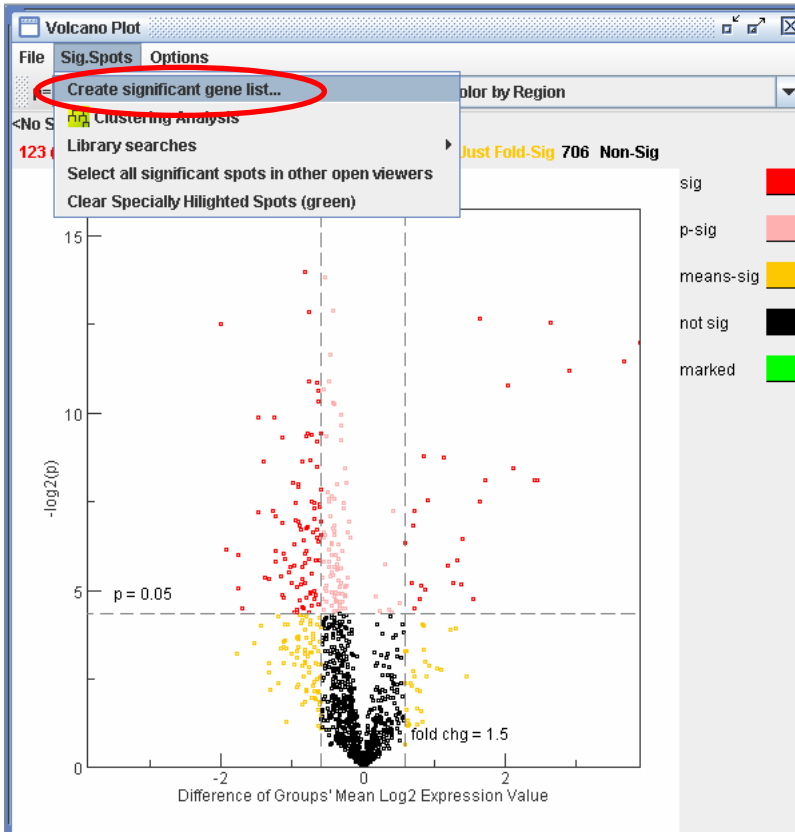
Abs Fold Change >

- T-test results with p-value is highlighted
- Many options are available under the table. We recommend to use “Volcano Plot”.



- Volcano plot results
- 123 genes have $p < 0.05$ and Fold Change > 1.5

Step 6 – Save the significant gene list (recommended)



ArrayTrack Testing

Database Library Tool Export Help

New Exp

Database Contents

ANONYMOUS

Affy_Rat_MAS5_only

Gene Lists

123geneMAS5_Fold1.5_P0.05

132geneNorm_Fold1.5P0.05

49geneNorm_Fold2_P0.05

Topic 1: 123geneMAS5_Fold1.5_P0.05

Library

ID Converter

Gene Library

Pathway Library

Tool

Analysis

T-Test

Create Gene List

Create gene list named: Topic 1: 123geneMAS5_Fold1.5_P0.05

Save into experiment: Affy_Rat_MAS5_only

Under gene list group (optional):

Description

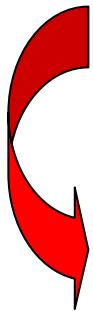
a. 6 treated samples vs 6 controls;

b. Welch t-test (default in ArrayTrack);

c. $p < 0.05$ and $FC > 1.5$ was used to select DEGs;

d. 123 significant genes are identified.

OK Cancel



Step 7 – Biological interpretation using ArrayTrack pathway, GO and other tools

Clicking the three buttons in red circle will direct you to step 8, 9, 10 correspondingly, see the next three slides

Individual gene annotation and analysis

Pathway analysis

GO-based functional analysis

The screenshot shows the ArrayTrack Testing interface. On the left, a tree view under 'Database Contents' shows a folder 'Afny_Rat_MAS5_only' containing 'Gene Lists'. One list, 'Topic 1: 123geneMAS5_Fold1.5_P0.05', is selected and circled in red with the text 'Double click' next to it. A red arrow points from this list to the 'Genes' button in the 'SIGNIFICANT_GENE_LIST' window. The 'SIGNIFICANT_GENE_LIST' window has a toolbar with buttons for 'Chrom', 'Lib', 'Genes', 'Proteins', 'Pathways', 'GOFA', and 'Orthologene'. The 'Genes', 'Pathways', and 'GOFA' buttons are circled in red. A red arrow points from the 'Pathways' button to the 'Pathway analysis' text box. Another red arrow points from the 'GOFA' button to the 'GO-based functional analysis' text box. The main window displays a table with columns: GENELIST_NAME *, EXPID *, REGG, NAME, LOCUSID, FOLD, and PVALUE. The table contains 15 rows of data, each representing a gene from the selected list.

	GENELIST_NAME *	EXPID *	REGG	NAME	LOCUSID	FOLD	PVALUE
1	Topic 1: 123geneMAS5_Fold1.5_P0.0	650		predicted	288444	4.1286	0.0006
2	Topic 1: 123geneMAS5_Fold1.5_P0.0	650			300721	5.3215	0.0038
3	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AA848503	Hspatam Hspa1b		14.9075	0.0003
4	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AB003400	Dao1	114027	0.6656	0.0107
5	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF021935	Cdc42bpa	114116	0.4238	0.0184
6	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF025670	Casp6	83584	0.2951	0.0161
7	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF025670	Casp6	83584	0.4238	0.0149
8	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF031657	Zfp94	499095	0.4941	0.0203
9	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF051943	Nme6	58964	0.5452	0.0282
10	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF051943	Nme6	58964	0.2628	0.0145
11	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AF084205	Taok1	286993	0.5826	0.0092
12	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AJ005424	Mapk7	114509	0.5151	0.0082
13	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AJ005425	Mer2d	81518	0.5585	0.0121
14	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	AJ223083	Rxrg	83574	2.5979	0.0285
15	Topic 1: 123geneMAS5_Fold1.5_P0.0	650	D14014	Ccnd1	58919	0.6225	0.0065

Step 8 – Individual gene annotation and analysis using GeneLib

Summary info for the highlighted gene

Link to other pub databases for the highlighted gene

The screenshot shows the GeneLib software interface. On the left, there is a sidebar with search options and a search bar. The main window displays a table of genes with columns for ID, Gene Name, and Description. The gene 'Mpg' (row 7) is highlighted. Two red circles highlight the 'More Info...' and 'Link To...' dropdown menus, which are open and show various database links and summary information for the selected gene.

Specify ID Type:

- GenBankAcc
- UnigenID
- LocusID
- SwissProtAcc
- IMAGEID
- GEN_ID_MFR
- GeneName

Hs Mm Rn

Enter Searching Data:

Message:
unique search ID number :91

Missing number : 0;
Missing list of search ids(LOCUSID):

Filter->	geneid as INPUT	LOCUSID	GENENAME	DESCR
1	24296		Cyp1a1	cytochrome P450, family 1, subfamily 1, member 1
2	24404		Gpx1	glutathione peroxidase 1
3	24451		Hmox1	hemoxylin 1
4	24471		Hspb1	heat shock protein 70 kDa class B member 1
5	24472		Hspa1a	heat shock protein 70 kDa class A member 1
6	24553		Met	methionine synthase
7	24561		Mpg	N-methylpurine-DNA glycosylase
8	24565		Abcc1	ATP-binding cassette, sub-family C (CFTR/MRP), member 1
9	24567		Mt1a	Metallothionein
10	24605		Nras	neuroblastoma RAS viral (v-ras) oncogene homolog
11	24646		Abcb1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
12	24811		Tap1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
13	24842		Tp53	tumor protein p53
14	24862		Ugt2b	UDP glycosyltransferase 2 family, polypeptide B
15	24891		Abcb4	ATP-binding cassette, sub-family B (MDR/TAP), member 4
16	25035		Dia1	diaphorase 1
17	25283		Gclc	glutamate-cysteine ligase, catalytic subunit
18	25315		Ephx1	epoxide hydrolase 1
19	25355		Ste	sulfotransferase, estrogen preferring
20	25406		Cd44	CD44 antigen
21	25420		Cryab	crystallin, alpha B
22	25493		Nfkbia	nuclear factor of kappa light chain gene enhancer in B-cells inhibitor, alpha 1
23	25591		Adprt	ADP-ribosyltransferase 1
24	25625		Tnfrsf1a	tumor necrosis factor receptor superfamily, member 1a
25	25675		Hmgcr	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
26	25717		Tgfb3	transforming growth factor, beta 3
27	26759		Bach	brain acyl-CoA hydrolase
28	26844		26844	zinc finger protein 26, C2H2-type like 1

Step 9 – Pathway analysis using KEGG

the up- and down-regulated genes that involves the same pathways

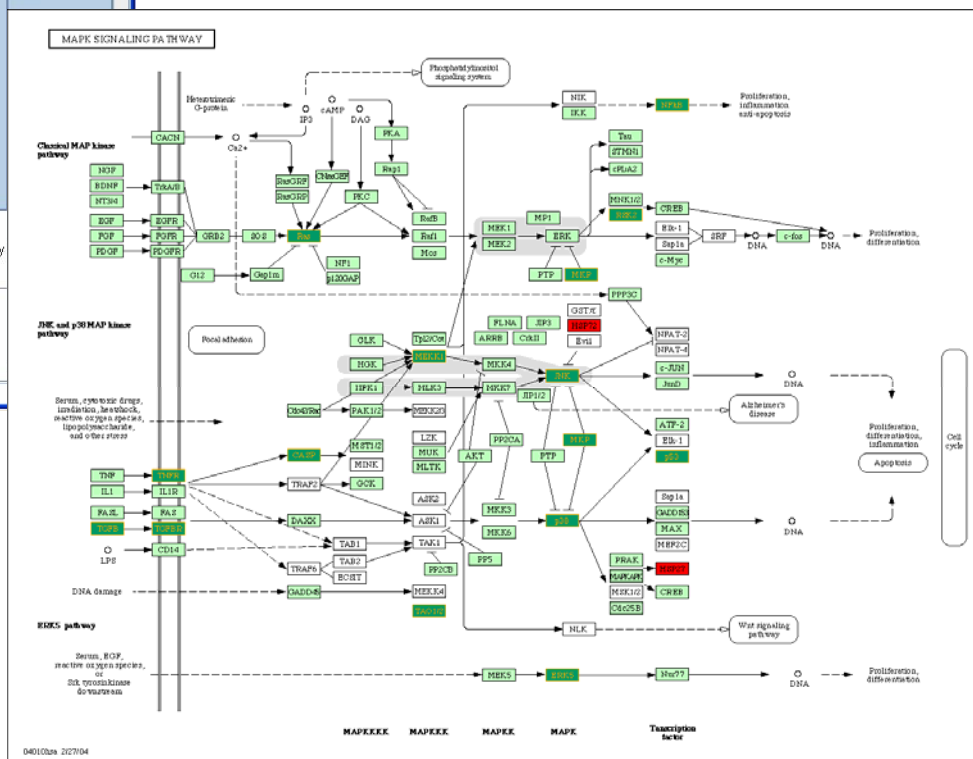
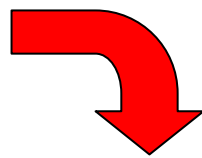
Statistical significance

KEGG Pathway(rno)

Gene name	Map	Category	Fisher P-value
↓ Casp6(83584)	MAPK signaling pathway(mo04010)	Regulatory pathway	0E-8
↑ Hspb1(24471)			
↑ Hspb1(24471)			
↓ Map3k1(116667)			
↓ Map3k1(116667)			
↓ Nras(24605)			
↓ Nfkb1(81736)			
↓ Mapk9(50658)			
↓ Mapk9(50658)			
↓ Mapk7(114509)			
↓ Mapk14(81649)			
↓ Map3k1(116667)			
↑ Hspb1(24471)			
↓ Tp53(24842)			
↓ Trnfrs1a(25625)	Tetrachloroethene degradation(mo00625)	Biosynthesis of Secondary Metabolites/Metabolic pathway	
↓ Tgfbr2(81810)			
↓ Tgfbr3(25717)			
↓ Taok1(286993)			
↓ Rps6ka1(81771)	Metabolism of xenobiotics by cytochrome P450(mo00980)	Regulatory pathway	
↑ Hspa1a(24472)			
↓ Dusp6(116663)			
↓ Dusp7(300980)			
↓ Casp6(83584)			

Input genes = 118, 54 genes found, 64 not found, Total 70 pathway maps.

Double click the pathway



MKP Down-regulation

HSP72 Up-regulation

Step 10 – GO-based analysis using GOFFA

The screenshot displays the GOFFA (GO-based analysis using GOFFA) software interface. The main window is titled "Go Term Cluster" and is divided into several panels:

- Select data type:** Includes radio buttons for GenBankAcc, UnigeneID, LocusID (selected), SwissProtAcc, Gene name (Official/Synonym), and IPI protein (Human/Mouse/Rat).
- Select array type:** A dropdown menu showing "ABI_Rat-Genome-Survey".
- Input Data:** A list of gene IDs including 84394, 9480, 221188, 84493, 84074, 81610, 8120, 243389, 79136, 19337, 64537, 192274, 72039, 293597, 71790, 79857, 76072, 18419, 100637, 84166, 29509, 20085, 116599, 394266, 320394, 89872, 20308, 316620, 77552, and 108670.
- Tree Window:** A hierarchical tree of GO terms. The root is "all(417) P=1.000000 E=1.00". The tree is expanded to show "biological process(377) P=0.267641 E=1.01", "physiological process(350) P=0.138708 E=1.03", and "response to stimulus(115) P=0.136045 E=1.09". The "response to stimulus" branch is highlighted in blue. A search box at the bottom of the tree window allows for filtering terms by GO term, gene name, p-value, or E-value.
- Gene List:** A table on the right side of the window listing genes. The columns are "No", "LocusID", and "Gene Name(official)". The list includes genes such as hihba, il11, Cd80, Spp1, Casp3, RGD:727815, Prlr, Fcgr2b, Cdkn1a, Cd5, Lcp2, Irf1, C3, Aps, Fn1, Sele, Tnfrsf1a, Mgl1, Cd14, Mefv, Ccr2, Alox5, C4a, S100a8, Cxcl10, Cxcl9, Hpse, Cystsr1, C2, C4bpa, Clu, Ctl, Cd53, Ubd, Cd24, Ocil, Umod, Fost1, Tnfrsf4, Tap1, Psmb8, RGD:727827, Gmp2, RGD:619831, Il24, Il1rn, Psmb9, Stat5a, Il6r, R11-M3, Lbp, Lyz, Cd44, Kcmm1, Tlr3, Kcmm4, and Cxcr3.

Total original submit =867, Found=417 with GO term

Tree Window – This is the default view of GOFFA, which enables the hierarchical display of the GO terms in a tree-like format; p- and E-values as well as the number of genes are also displayed for each GO term. E-values >1 are shown in green and those <1 in red, respectively denoting greater or lesser prevalence of the GO term in DEG's rather than in the overall experimental platform. The user can query the tree by GO term, gene name/symbol, p-value and E-value with functions below the view. The query-matched GO terms are highlighted as blue.

Step 10 – GO-based analysis using GOFFA (cont.)

The screenshot displays the GO Term Cluster software interface, which is used for GO-based analysis. It features two main windows: 'Terms' (A) and 'Genes' (B).

Window A: Terms

No	Term	GO ID	Level	Average	P value (A)	Gene Hits	E value
1	response to external stimulus	GO:0009605	4.00	0.000000	55.00	1.99	
2	amino acid and derivative metabolism	GO:0009611	5.00	0.000000	40.00	3.72	
3	organic acid metabolism	GO:0009682	5.00	0.000000	45.00	2.22	
4	defense response	GO:0009652	5.00	0.000000	57.00	7.32	

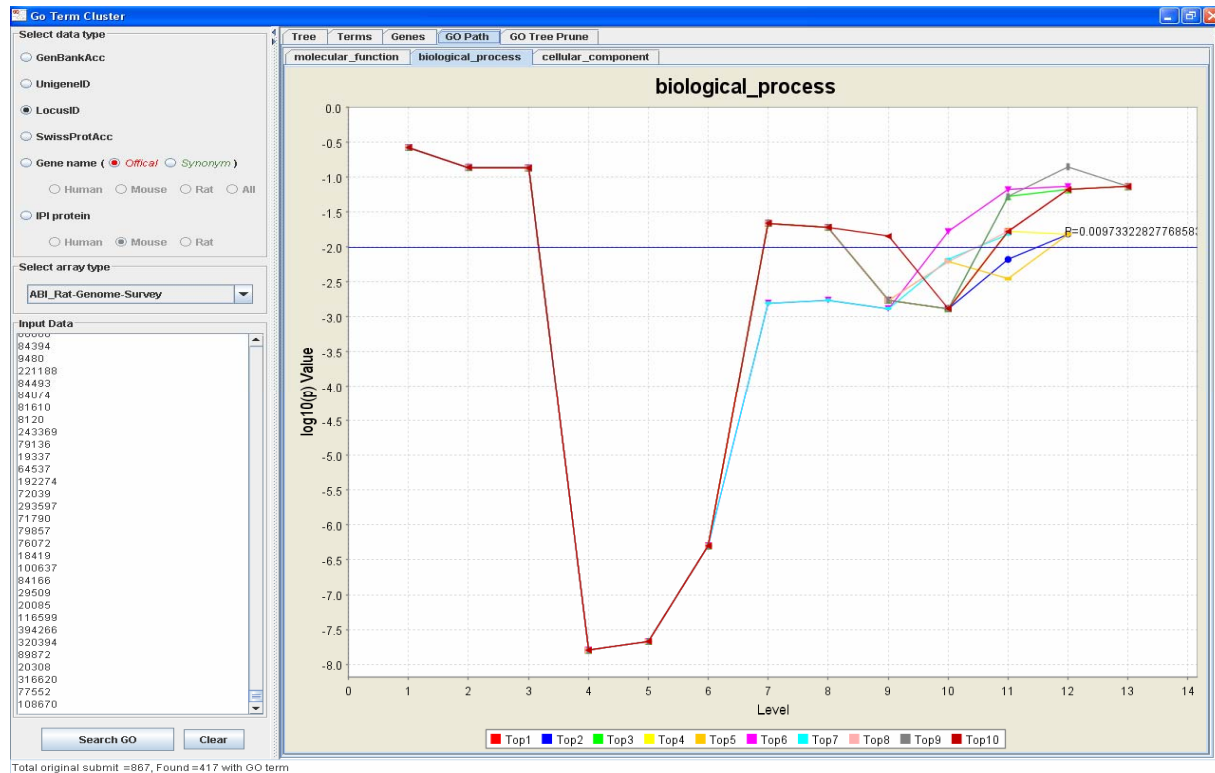
Window B: Genes

No	Gene	Term	GO ID	Level	Average	P value (Average)	Gene Hits	E value
3099	Emp3	death	GO:0016265	3.00	0.010881	33.00	1.51	
3099	Casp3	death	GO:0016265	3.00	0.010881	33.00	1.51	
3100	Bcl2	death	GO:0016265	3.00	0.010881	33.00	1.51	
3101	Tnfrsf4	defense response	GO:0009652	5.00	0.000000	57.00	2.12	

The interface includes various filters and options on the left side, such as 'Select data type' (GenBankAcc, UnigeneID, LocusID, SwissProtAcc), 'Gene name' (Official, Synonym), 'Input Data' (Human, Mouse, Rat, All), and 'Select array type' (ABI_Rat_Genome-Survey). At the bottom, it shows 'Total original submit = 867, Found = 417 with GO term'.

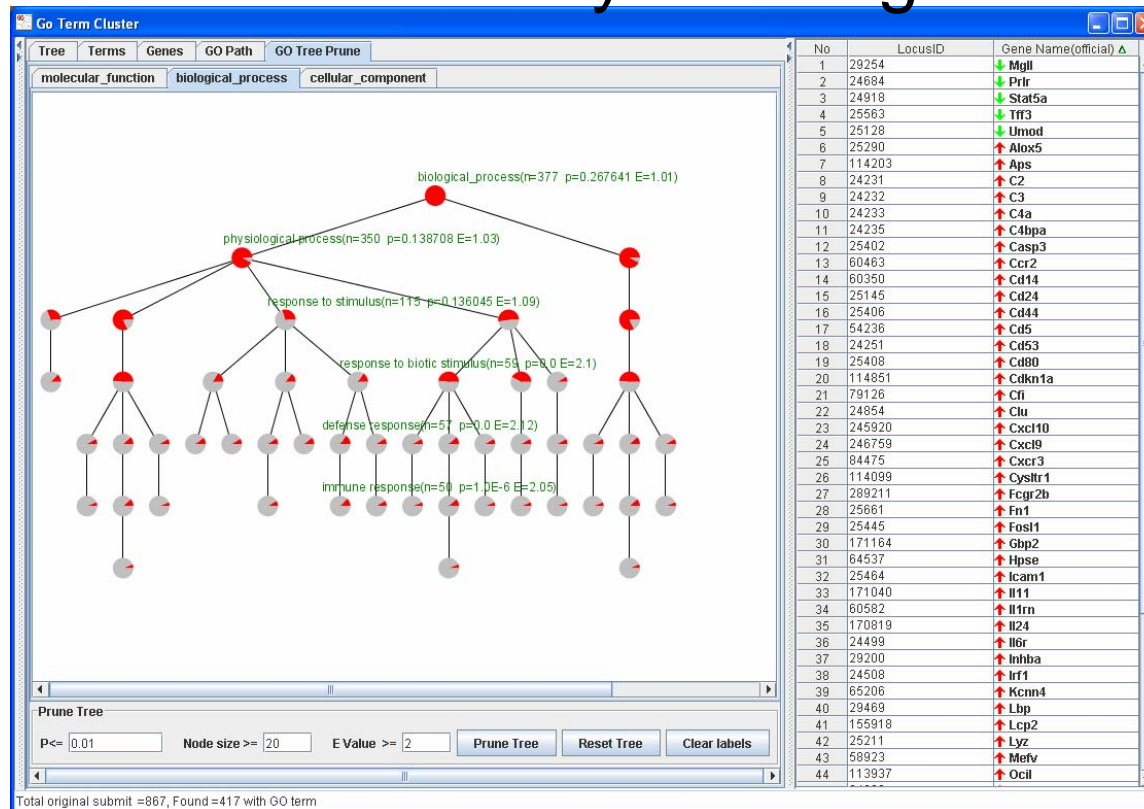
Terms and Genes Windows – The Terms Window (A) and Genes Window (B) summarize the findings associated with GO terms and genes respectively in a table format along with various statistical parameters (e.g., p- and E-values). Each View contains three tabs corresponding to three categories of GO (molecular functions, biological processes and cellular components). The table can be sorted in every column by clicking on the column header. Sorting on multiple columns is also supported (pressing Ctrl key while clicking on the second column header for sorting). Both copy/paste and export functions are available to transfer data to external software.

Step 10 – GO-based analysis using GOFFA (cont.)



GO Path – GO Path is sorted by descending statistical significance based on an inverse Chi-Squared test and the GOFFA Tree Paths (i.e., linked GO terms) that are displayed from high to low at each hierarchical level. GO Path plots the top ten paths with solid circles representing the GO terms on the path. The number in X-axis represents the hierarchical level to which the GO term belongs and the Y-axis ($\log p$) indicates the statistical significance of the term. A colored legend for the top 10 paths is located beneath the plot. Clicking any circle in a path in the plot or its corresponding color key launches a Tree View with the selected path highlighted in blue. Other features are also available from a popup menu by right clicking the plot, including zoom in/out, export figure, etc.

Step 10 – GO-based analysis using GOFFA (cont.)



GO TreePrune - This function allows the user to filter out nodes and thus reduce the complexity of a tree by specifying the p- and E-value as well as the user-defined number of genes at the end node. A GO term is represented by a sectored pie, where the red sector shows the percentage of DEGs associated with the term. The individual genes associated with each term are displayed in the right panel by single clicking the term. The annotation of a term can be turned on or off by double clicking the term. Each term (red circle) is movable by mouse dragging, which is convenient when working on a dense tree or with many annotations. The tree diagram can be zoomed or moved by dragging with the right or left mouse button held down.