

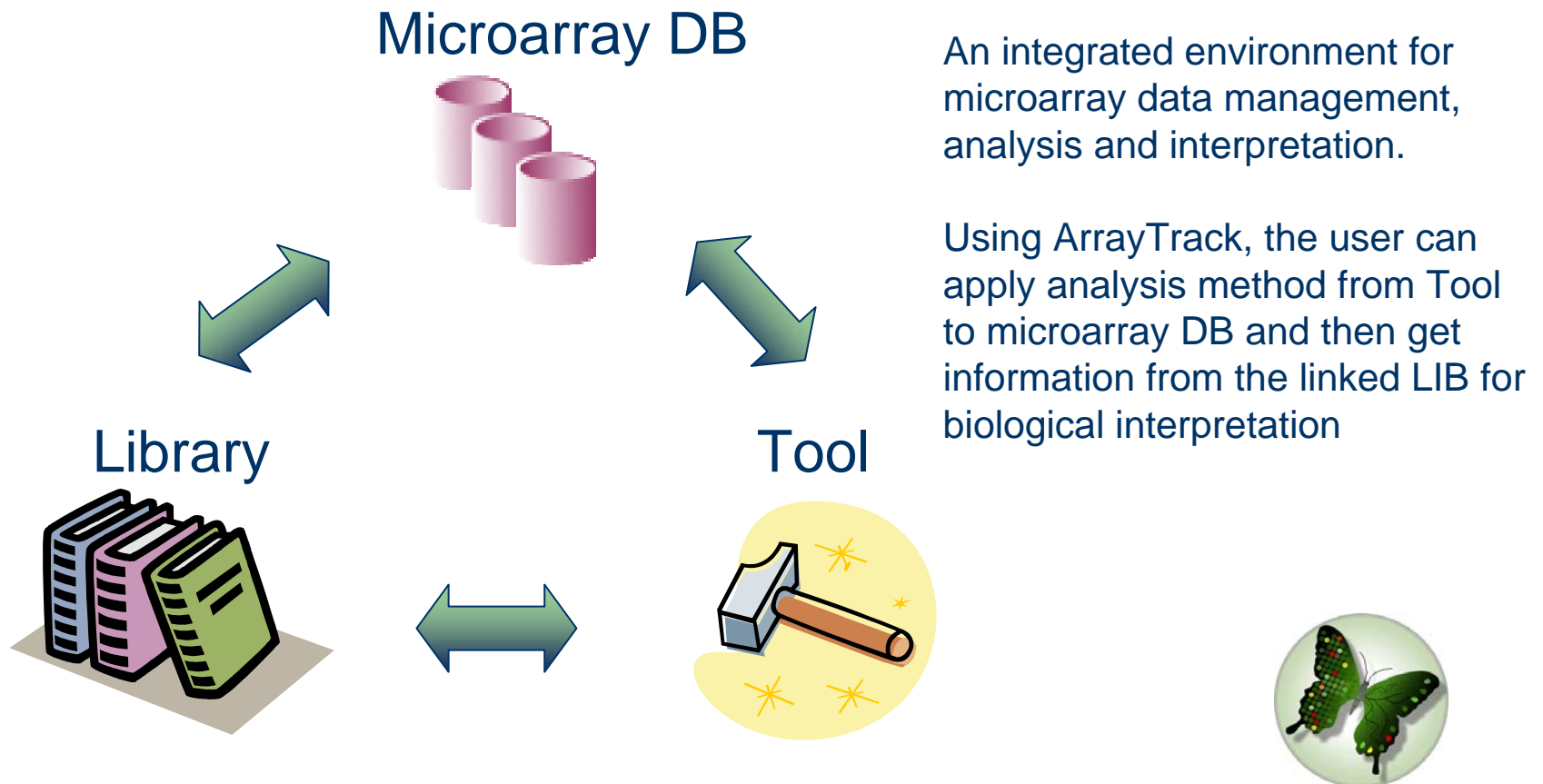


# *ArrayTrack<sub>3.4.0</sub> Demonstration*

National Center for Toxicological Research  
U.S. Food and Drug Administration  
3900 NCTR Road,  
Jefferson, AR 72079



# ArrayTrack Overview

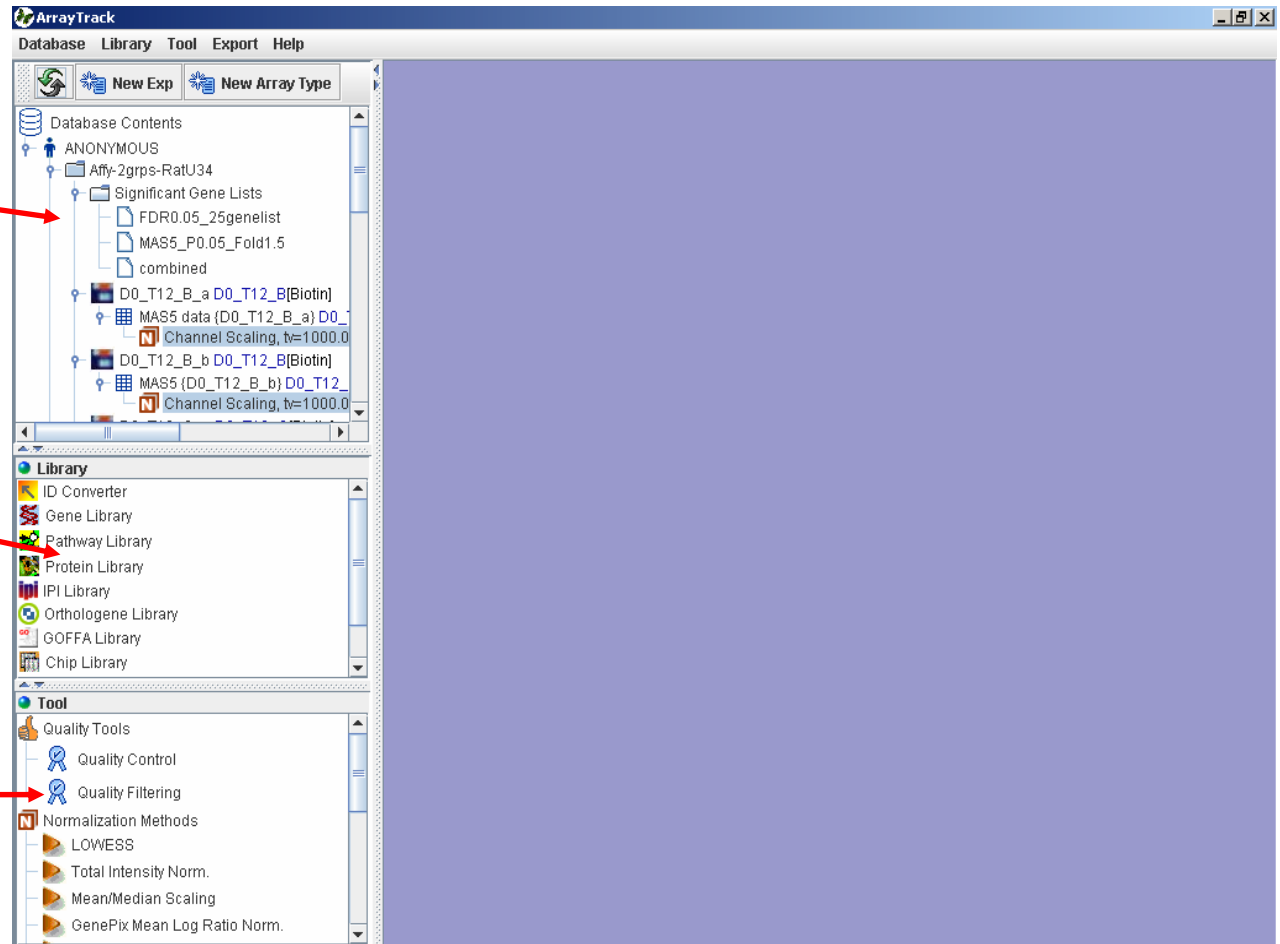


# ArrayTrack Overview

Microarray DB

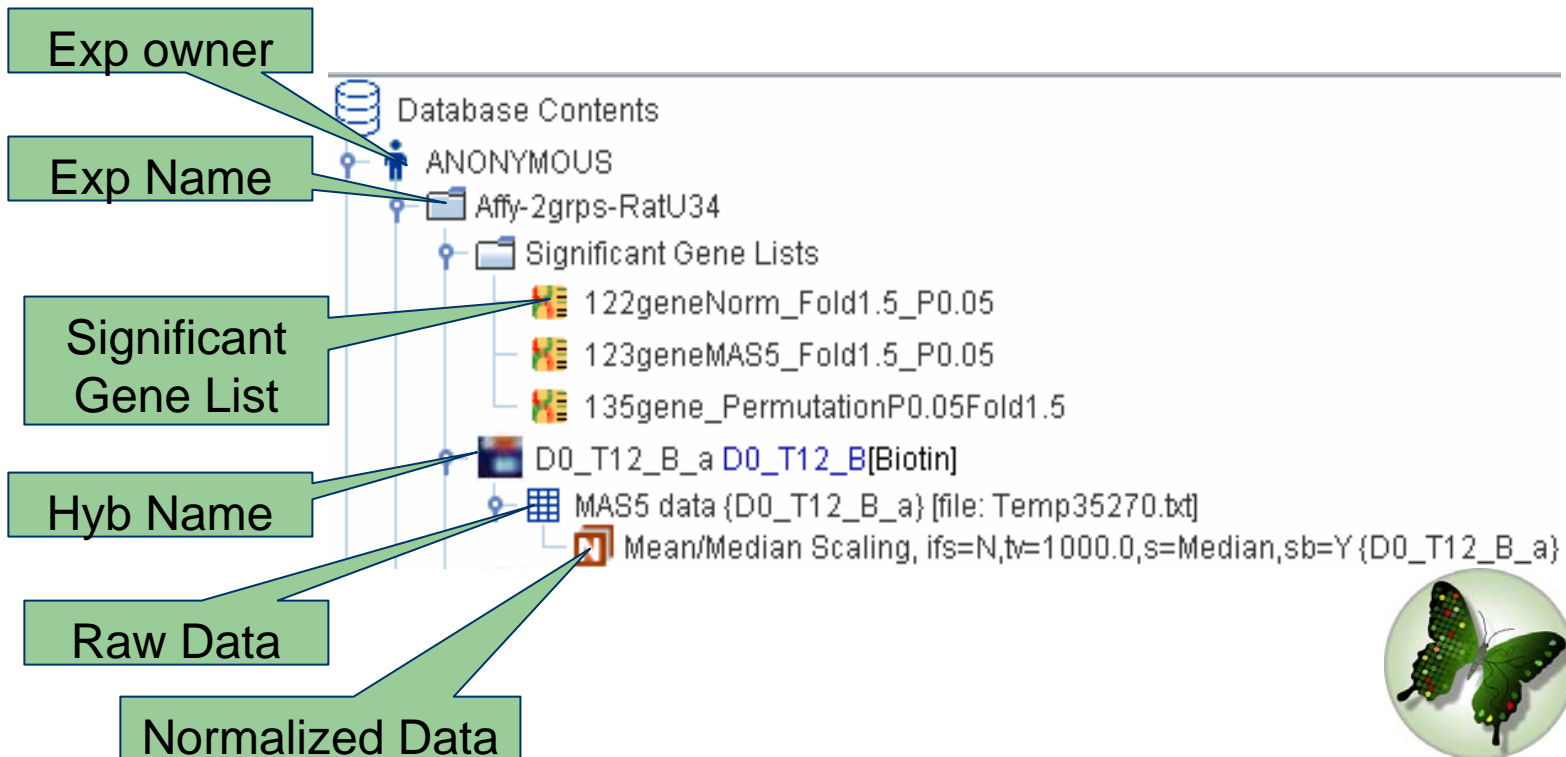
Libraries

Tools



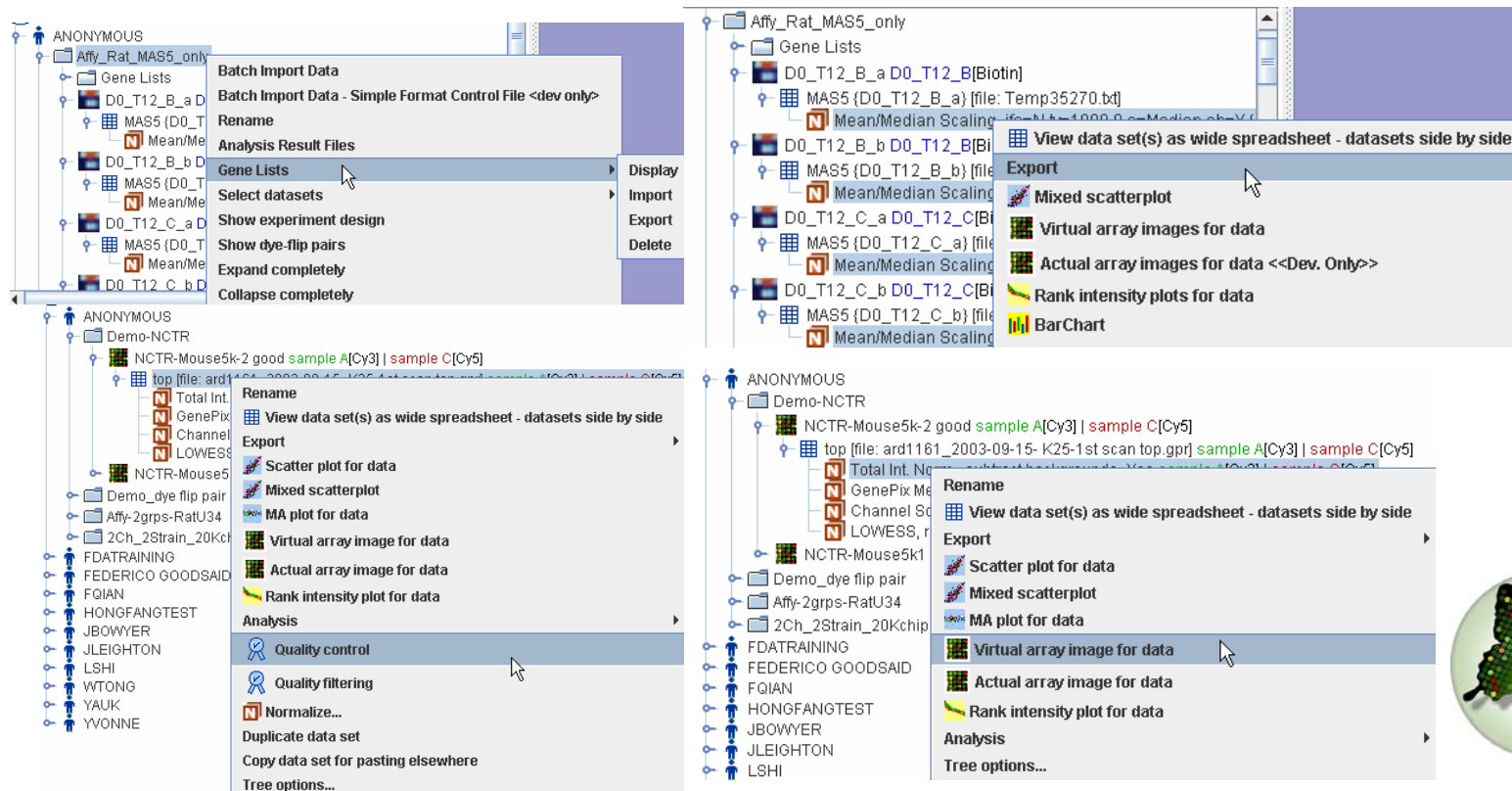
# Microarray DB

Data is organized as a hierarchical tree structure:



# Exploring DB

- Right-click exp, hybridization, raw data and normalized data to access various functions



The image displays three screenshots of the ArrayTrack software interface, illustrating the right-click context menus for different data types:

- Top Left Screenshot:** Shows a right-click context menu for a data set (e.g., MAS5 (D0\_T12\_B\_a)). The menu includes options such as "Batch Import Data", "Rename", "Analysis Result Files", "Gene Lists", "Select datasets", "Show experiment design", "Show dye-flip pairs", "Expand completely", and "Collapse completely".
- Top Right Screenshot:** Shows a right-click context menu for a data set (e.g., MAS5 (D0\_T12\_B\_a)). The menu includes options such as "View data set(s) as wide spreadsheet - datasets side by side", "Export", "Mixed scatterplot", "Virtual array images for data", "Actual array images for data <<Dev. Only>>", "Rank intensity plots for data", and "BarChart".
- Bottom Left Screenshot:** Shows a right-click context menu for a data set (e.g., NCTR-Mouse5k-2 good sample A[Cy3] | sample C[Cy5]). The menu includes options such as "Rename", "View data set(s) as wide spreadsheet - datasets side by side", "Export", "Scatter plot for data", "Mixed scatterplot", "MA plot for data", "Virtual array image for data", "Actual array image for data", "Rank intensity plot for data", "Analysis", "Quality control", "Quality filtering", "Normalize...", "Duplicate data set", "Copy data set for pasting elsewhere", and "Tree options...".
- Bottom Right Screenshot:** Shows a right-click context menu for a data set (e.g., NCTR-Mouse5k1). The menu includes options such as "Rename", "View data set(s) as wide spreadsheet - datasets side by side", "Export", "Scatter plot for data", "Mixed scatterplot", "MA plot for data", "Virtual array image for data", "Actual array image for data", "Rank intensity plot for data", "Analysis", and "Tree options...".





# Searching Libraries



There are nine libraries in ArrayTrack.  
All the libraries are interlinked.

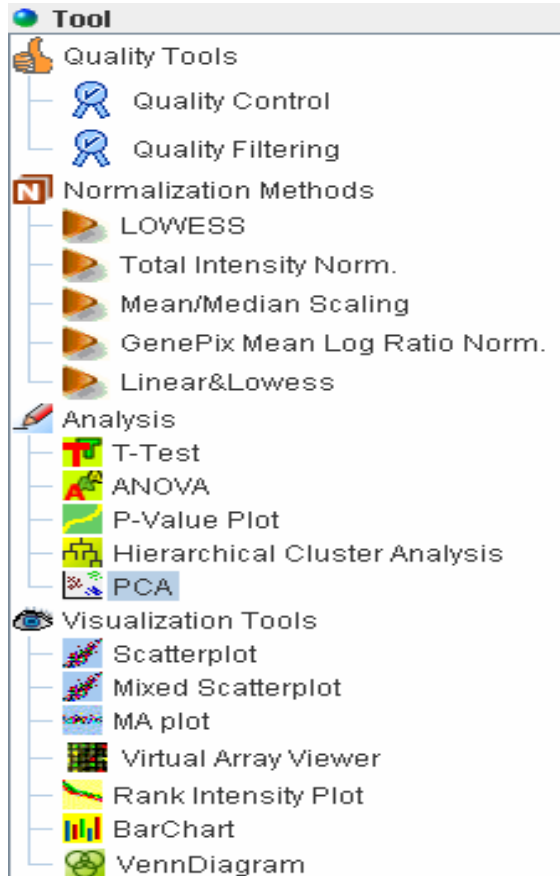
The libraries integrate gene, protein, pathway and other data allowing data interrogation and mining of data across data types.

Several ways to activate these libraries

- From library panel
- From the Library pull-down menu
- From the results of analysis (e.g. T-test)



# Analysis Tools



## Tools:

### •Quality Tools

Provides various visual plots and numerical parameters for measuring the quality of a hyb, and filtering the unwanted spots.

### •Normalization Methods

Correct systematic variations in microarray data introduced by experimental factors

### •Analysis

Provides different methods to evaluate the microarray data

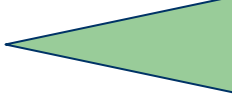
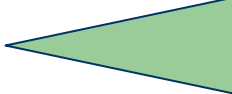
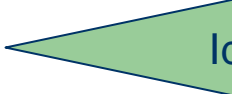

### •Visualization Tools

Provides a direct view to identify abnormalities within data.



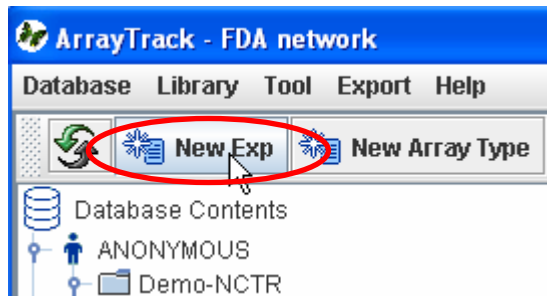


# Microarray Data Processing

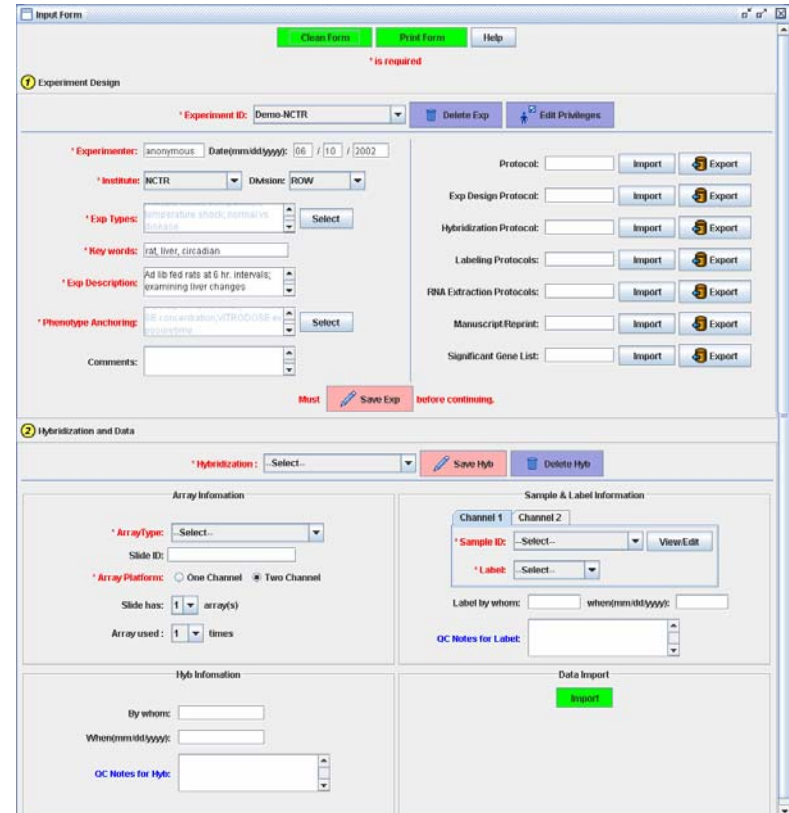
- Importing data  Loading array data into AT
- Normalization  Ensure cross-chip comparison
- Gene Selection  Identify a list of significant genes
- Interpretation  Interpret data using pathways and GO



# Importing Data



1. Create experiment
2. Create array type (if not existed in AT Chip LIB)
3. Use batch import wizard (see detail in Tutorial 9)



**Input Form**

Buttons: Clean Form, Print Form, Help

\* is required

**1) Experiment Design**

Experiment ID: Demo-NCTR [Delete Exp] [Edit Privileges]

\* Experimenter: anonymous Date(mm/dd/yyyy): 06 / 10 / 2002

\* Institute: NCTR Division: ROW

\* Exp Types: [Select]

\* Key words: rat, liver, circadian

\* Exp Descriptions: Ad lib fed rats at 6 hr intervals; examining liver changes

\* Phenotype Anchoring: [Select]

Comments: [Text Area]

Buttons: Import, Export (multiple)

Must Save Exp before continuing.

**2) Hybridization and Data**

Hybridization: [Select] [Save Hyb] [Delete Hyb]

**Array Information**

\* Array type: [Select]

Slide ID: [Text Field]

\* Array Platform:  One Channel  Two Channel

Slide has: [1] array(s)

Array used: [1] times

**Sample & Label Information**

Channel 1 Channel 2

\* Sample ID: [Select] [View/Edit]

\* Label: [Select]

Label by whom: [Text Field] when(mm/dd/yyyy): [Text Field]

OC Notes for Label: [Text Area]

**Hyb Information**

By whom: [Text Field]

When(mm/dd/yyyy): [Text Field]

OC Notes for Hyb: [Text Area]

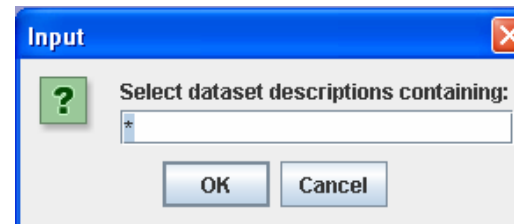
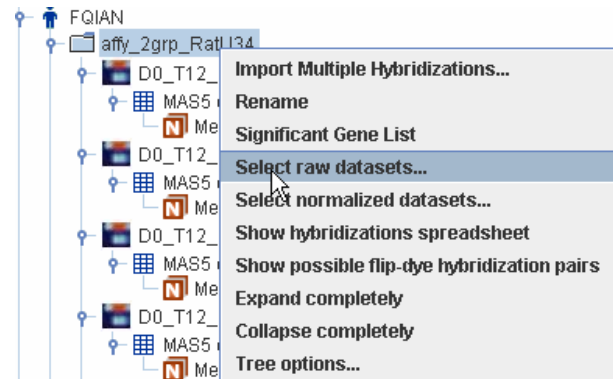
**Data Import**

Import

# Normalization

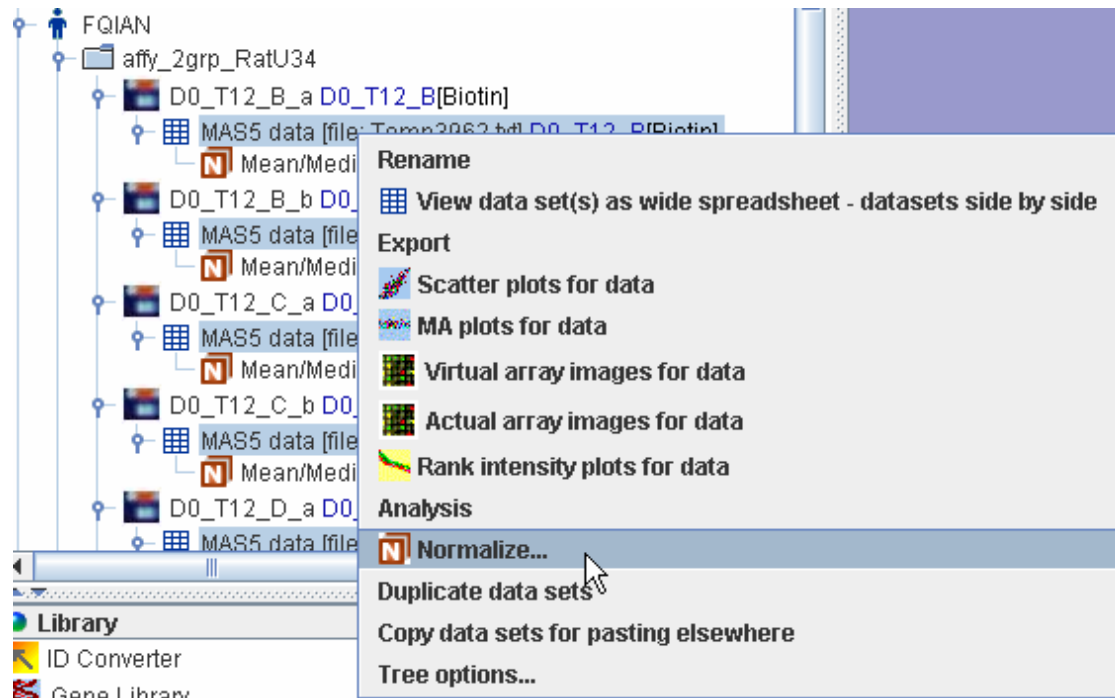
- To remove systematic variations across chips and ensure a valid cross-chip comparison

1. Right-click an experiment
2. Select “raw datasets...”



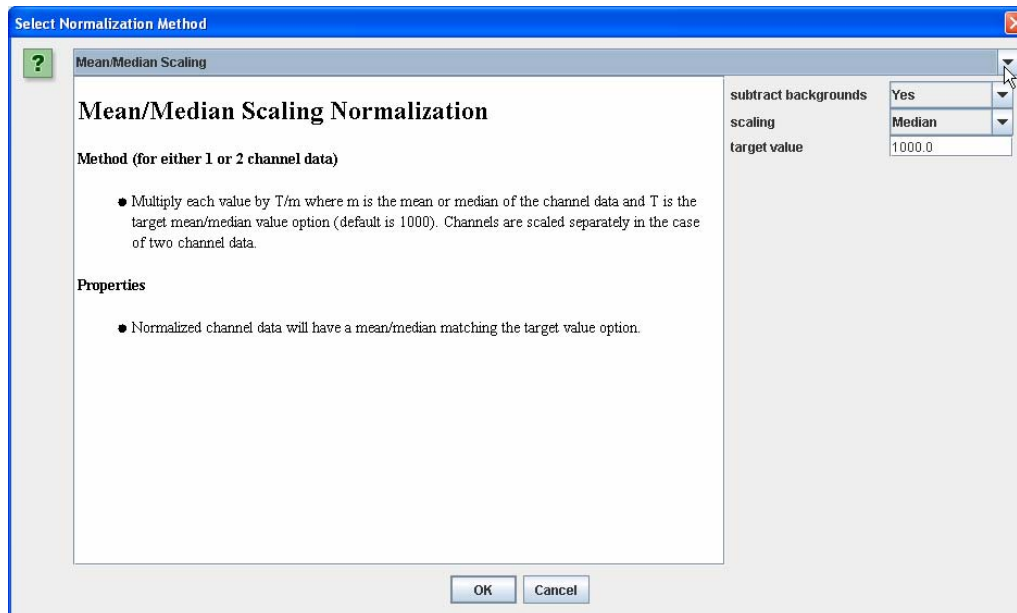
# Normalization (-continued)

3. Right-click any highlighted raw data
4. Select “Normalize...”



# Normalization (-continued)

- Choose normalization method



- For Affy data, choose “Mean/Median Scaling”
- For 2-channel data, the default method is “Lowess”



# Gene Selections

Determining a list of genes that are differentially expressed between Strain A and Strain B

## Two types of experiment:

Single testing:



1 gene  $P < 0.05$

## Error rate for the exp

low error rate

Multiple testing:



$n$  genes  $P = nP_i$

If  $P_i = 0.05$ , high error rate

e.g., If  $n = 10$  and  $P_i = 0.05$ ,  $P = 0.5$  for family-wise error

## Select a gene list based on:

P value



Bonferroni correction

$P_i/n$

*Low sensitivity*

*Low power*



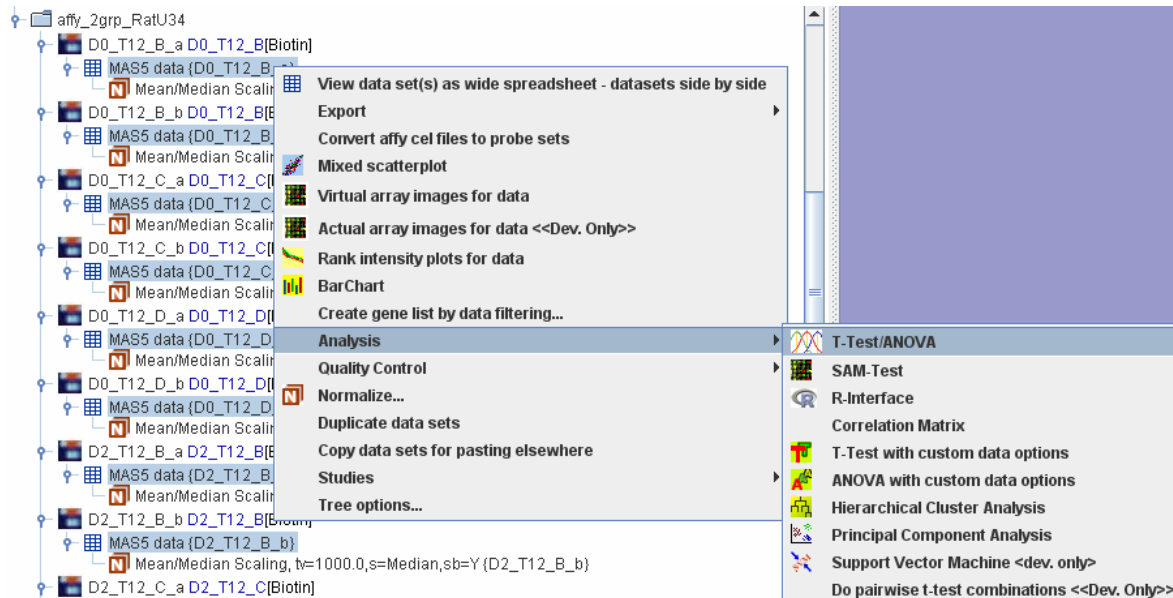
False discovery rate (e.g., Benjamini & Hochberg, p-value plot)

Permutation t-test (e.g., SAM)

*Volcano plot (combination of p and fold change)*

# Gene Selection

- Highlight and Right-click the experiment.
- Select “Normalized datasets...” and click “OK”
- Right-click the highlighted normalized data
- Choose “Analysis->T-test”



The screenshot displays the ArrayTrack software interface. On the left, a file tree shows a project named 'affy\_2grp\_RatU34' containing several data sets, including 'D0\_T12\_B\_a D0\_T12\_B[Biotin]', 'D0\_T12\_B\_b D0\_T12\_B[Biotin]', 'D0\_T12\_C\_a D0\_T12\_C[Biotin]', 'D0\_T12\_C\_b D0\_T12\_C[Biotin]', 'D0\_T12\_D\_a D0\_T12\_D[Biotin]', 'D0\_T12\_D\_b D0\_T12\_D[Biotin]', 'D2\_T12\_B\_a D2\_T12\_B[Biotin]', 'D2\_T12\_B\_b D2\_T12\_B[Biotin]', and 'D2\_T12\_C\_a D2\_T12\_C[Biotin]'. A context menu is open over the 'MAS5 data (D0\_T12\_B\_b)' item. The menu options include: 'View data set(s) as wide spreadsheet - datasets side by side', 'Export', 'Convert affy cel files to probe sets', 'Mixed scatterplot', 'Virtual array images for data', 'Actual array images for data <<Dev. Only>>', 'Rank intensity plots for data', 'BarChart', 'Create gene list by data filtering...', 'Analysis' (highlighted), 'Quality Control', 'Normalize...', 'Duplicate data sets', 'Copy data sets for pasting elsewhere', 'Studies', and 'Tree options...'. The 'Analysis' sub-menu is open, showing options: 'T-Test/ANOVA' (highlighted), 'SAM-Test', 'R-Interface', 'Correlation Matrix', 'T-Test with custom data options', 'ANOVA with custom data options', 'Hierarchical Cluster Analysis', 'Principal Component Analysis', 'Support Vector Machine <dev. only>', and 'Do pairwise t-test combinations <<Dev. Only>>'. A butterfly icon is visible in the bottom right corner of the interface.

# Gene Selection (-continued)

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

### Assign Data Sets Into Groups

Groups:

Filter->	Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPENAME	CHANNEL	SPECIES 1	ASSAY 1	F	
1	1	D0_T12_B_a	D0_T12_B	B12Cd000A.CELTemp3962.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
2	1	D0_T12_B_b	D0_T12_B	B12Cd000B.CELTemp3964.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
3	1	D0_T12_C_a	D0_T12_C	C12Cd000A.CELTemp3970.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
4	1	D0_T12_C_b	D0_T12_C	C12Cd000B.CELTemp3972.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
5	1	D0_T12_D_a	D0_T12_D	D12Cd000A.CELTemp3978.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
6	1	D0_T12_D_b	D0_T12_D	D12Cd000B.CELTemp3980.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
7	2	D2_T12_B_a	D2_T12_B	B12Cd2.00A.CELTemp3966.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
8	2	D2_T12_B_b	D2_T12_B	B12Cd2.00B.CELTemp3968.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
9	2	D2_T12_C_a	D2_T12_C	C12Cd2.00A.CELTemp3974.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
10	2	D2_T12_C_b	D2_T12_C	C12Cd2.00B.CELTemp3976.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
11	2	D2_T12_D_a	D2_T12_D	D12Cd2.00A.CELTemp3982.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N
12	2	D2_T12_D_b	D2_T12_D	D12Cd2.00B.CELTemp3984.t	Biotin	Affy_RT-U34	One Channe	Rat	In Vitro	N

2 groups, sizes = [6, 6]

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

Assign the data into 2 groups  
 -different dose  
 -different time  
 -or different animal

Note: always put control in group 2





# Gene Selection (-continued)

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**

<input checked="" type="checkbox"/> Genbank Acc	<input checked="" type="checkbox"/> Gene Mfr ID	<input checked="" type="checkbox"/> LOCUSID	<input type="checkbox"/> UNIGENEID	<input checked="" type="checkbox"/> GENENAME
<input type="checkbox"/> CLONEID	<input type="checkbox"/> GEN_DESCR_MFR	<input checked="" type="checkbox"/> REFSEQ	<input checked="" type="checkbox"/> SPOTID	

**Dataset Naming**

Hybridization names are always included.

add sample name(s) to hybridization names

add dye name(s) to hybridization names

**Data options**

Subtract backgrounds when present (raw datasets only)

Apply log (base 2) to expression values

Exclude spots flagged as bad

# Gene Selection (-continued)

**T-Test Results**

File	Selected-Spot	All-Spots	Advanced	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	(1) P	Abs Fold C...	Fold Chang
1	L26267	L26267_at	81736	Nfkb1					516822	0	2.2286	2.2286
2	M55534	M55534m...	25420	Cryab	NM_012935				516911	0.0002	3.1305	0.3194
3	J02722	J02722cds...	24451	Hmox1					516748	0.0002	6.2489	0.16
4	E00778	E00778cds...	24296	Cyp1a1					516730	0.0002	3.9988	3.9988
5	AA848563	AA848563...		Hspa1a // ...					516506	0.0003	14.90	
6	AA818604	rc_AA8186...		Hspa1a // ...					517260	0.0004	12.83	
7	Z75029	Z75029_s...	294254	Hspa1b_m...					517240	0.0004	7.535	
8	AA108277	AA108277...	288444	Hsph1					516484	0.0006	4.128	
9	A1103396	rc_A110339...							517415	0.0011	2.371	
10	E00717	E00717UT...	24296	Cyp1a1					516729	0.0011	2.780	
11	U65007	U65007_at	24553	Met	NM_031517				517080	0.0016	2.195	
12	AA859372	rc_AA8593...							517282	0.0024	2.189	0.4568
13	X63594	X63594cds...	25493	Nfkb1a					517179	0.0026	2.6585	
14	AA945704	rc_AA9457...	361384	Dnajb1_pr...					517346	0.0029	4.8	
15	L16764	L16764_s...	24472	Hspa1a // ...	NM_031971				516808	0.003		0.3038
16	AA859648	rc_AA8596...	361384	Dnajb1_pr...					517284		5.4919	0.1821
17	AA848268	AA848268...	300721	Dnaja4						0.0038	5.3215	0.1879
18	AI236601	rc_AI23660...	288444	Hsph1					517510	0.0057	3.1323	0.3193

44 genes

**Significance Filtering**

- P Values < 0.05 without adjustment
- Target False Discovery Rate (FDR):
- Select # genes by lowest p-values

Mean Channel Intensities > Bad Flags <=

Abs Fold Change > 2 **Advanced>>**

Apply Filters Clear Filters

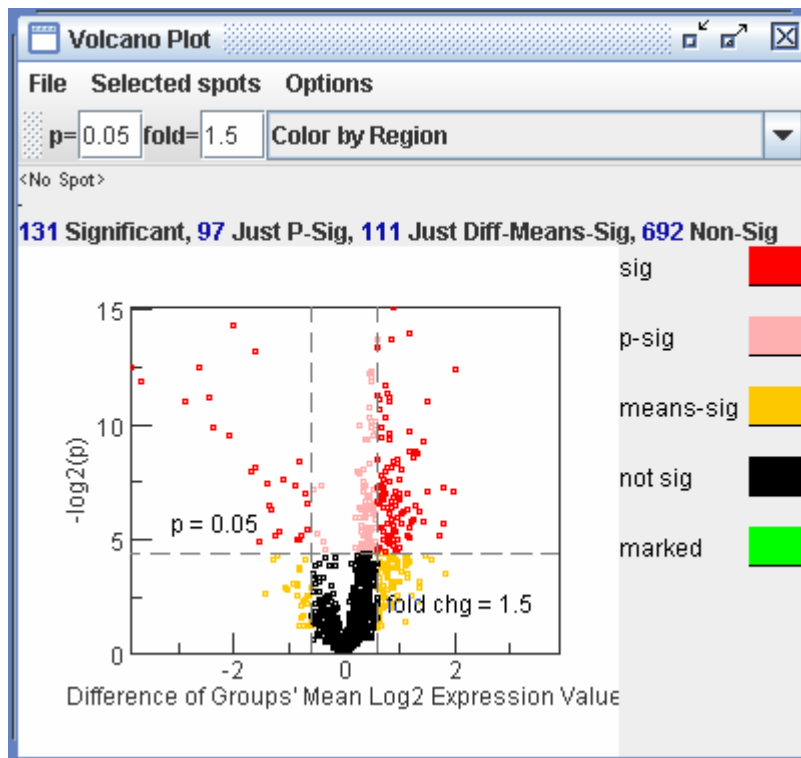
P-Value Plot Create Sig. Gene List HCA PCA Volcano Plot

Parameters used to filter the genes to get significant gene list

Access other analysis methods like HCA, PCA, Volcano plot, etc.



# Gene Selection (-continued)



Volcano Plot



# Significant Gene List

---

- Create
- Display
- Import
- Export
- Delete

# Significant Gene List(-cont)

- Create significant gene list

From T-test/Anova result

set filter criteria

e.g. P-value, fold change

to get significant gene list

Create significant gene list

Filter criteria

The screenshot shows the 'T-Test Results' window with a table of genes and a 'Significance Filtering' panel. A red box highlights the 'Create significant gene list...' option in the 'All Spots' menu. A green arrow points from this menu item to the 'Create significant gene list' text box. Another red arrow points from the 'Filter criteria' text box to the 'Significance Filtering' panel.

Genbank Acc	SPOTID	(1) P	Mean Grp 1	Mean Grp 2	Fold Change			
NM_013474	149926	0	6.4884	2.7783	13.0868			
AY065511	146938	0	-2.0418	-3.7955	3.3721			
AK011110	151257	0.0001	-1.5408	-0.7993	0.5981			
U8669			-0.2051	-0.5849	1.3012			
NM_010924			-3.0816	1.5725	0.0397			
NM_021304			0.9588	-0.7275	3.2184			
U86105	U86105_1	18113	Nnmt	-0.6934	1.1372	0.2812		
AK018313	AK018313_1			0.242	0.0395	1.1507		
NM_008865	NM_008865...	18776	Csh2	159805	0.0008	-0.8595	-1.748	1.8513
NM_019486	NM_01948...	56015	Olf71	162758	0.0008	0.2548	-0.0529	1.2377
NM_008319	NM_00831...	15898	Icam5	145874	0.0009	-0.1568	-0.4882	1.2564
NM_021447	NM_02144...	58522	Rrf30	149205	0.001	0.4095	0.1626	1.1867
NM_015780	NM_01578...	50702	Cfhl1	148342	0.0011	1.2908	3.2995	0.2485
D43759	D43759_1	14161	Fga	156273	0.0011	5.9168	4.6198	2.4571
NM_010419	NM_01041...	15208	Hes5	157161	0.0011	0.3868	0.1695	1.1625
AF331708	AF331708_1	230103	Npr2	158777	0.0012	-1.6846	0.4169	0.2314
AB041350	AB041350_1	12830	Col4a5	158418	0.0012	-2.8198	-3.7082	1.8511

20160 genes

Significance Filtering

P Values <  without adjustment

Target False Discovery Rate (FDR):

Select # genes with lowest p-vals:

Mean Channel Intensities >  Bad Flags <=

Fold Change >

Apply Filters

P-Value Plot MAQC Create Sig. Gene List HCA PCA Volcano Plot

# Significant Gene List(-cont)

- **Display/Import/Export/Delete Significant Gene List**

Right-click the experiment name

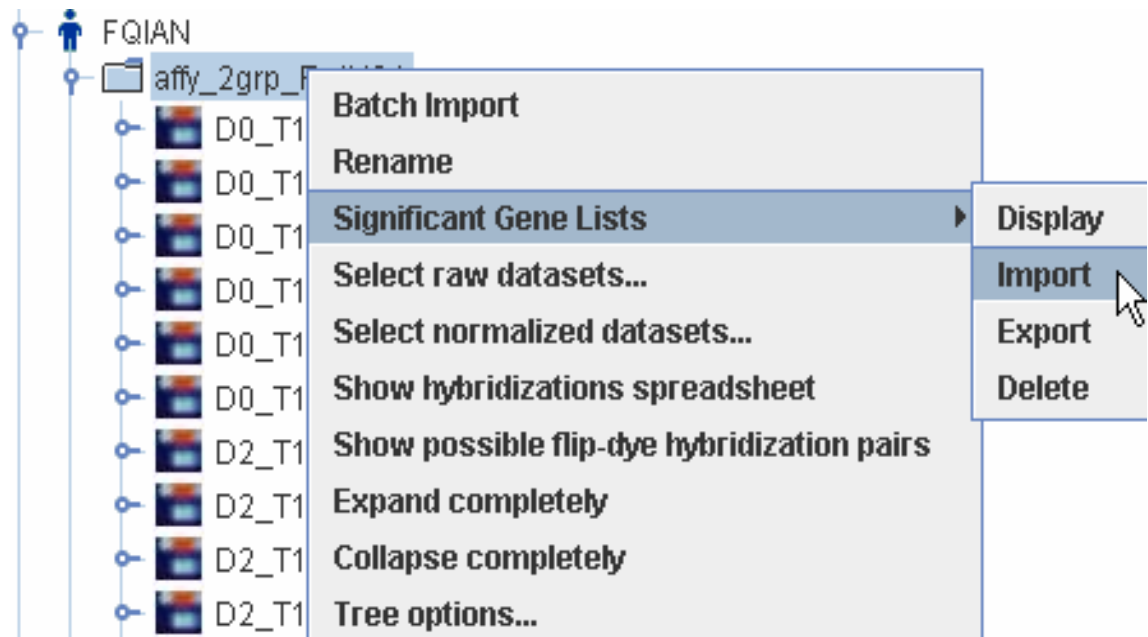
select Significant Gene Lists

select Display

Import

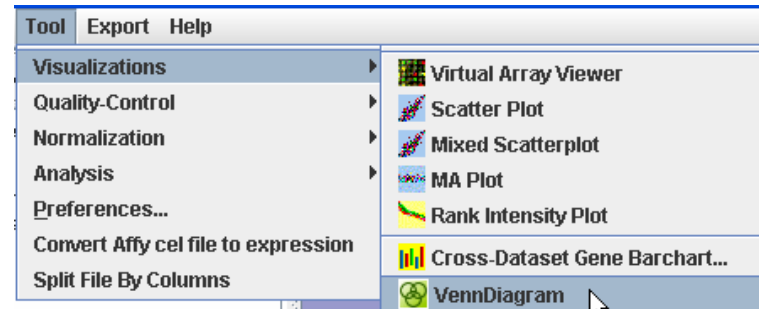
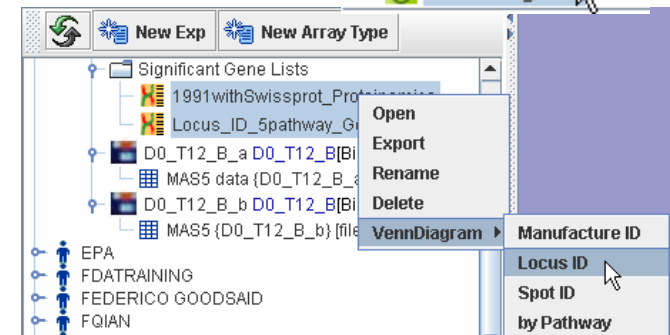
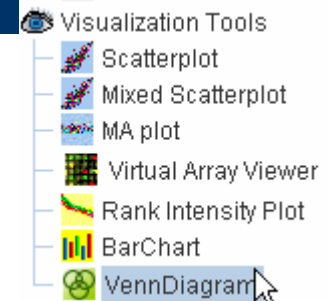
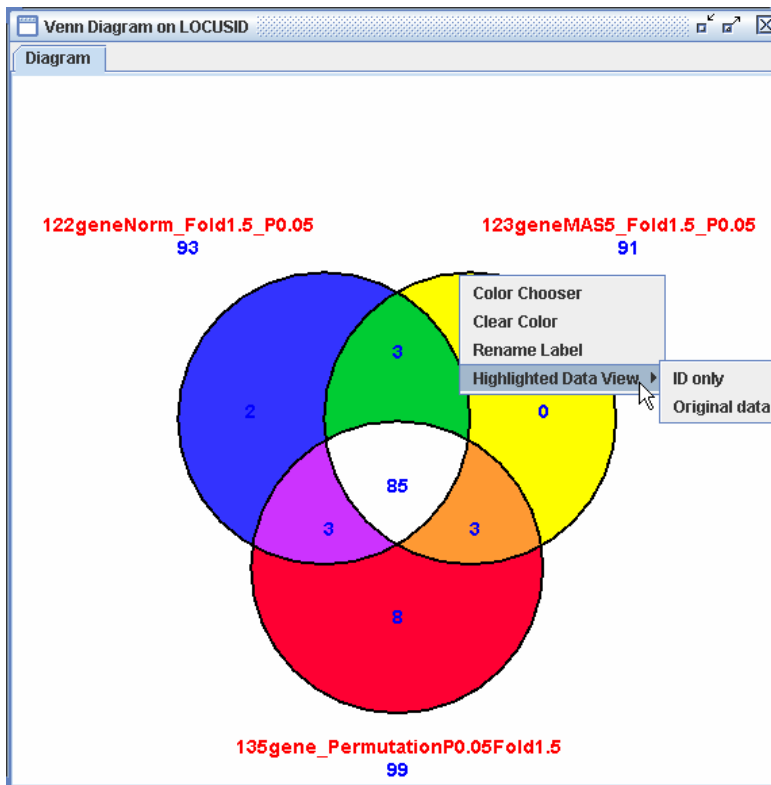
Export

Delete



# Significant Gene List(-cont)

- VennDiagram** - get common genes from 2~3 significant gene lists



# Interpretation

Link the significant genes to Gene Library for data interpretation

The screenshot shows the ArrayTrack software interface. On the left, a tree view under 'Database Contents' shows a folder 'Gene Lists' with a file '123gene\_Welch\_F1.5P0\_R5' selected. A tooltip 'significant gene list' points to this file. Below the tree is a 'Library' section with icons for 'ID Converter', 'Gene Library', 'Pathway Library', 'Protein Library', 'IPI Library', 'Orthologene Library', 'GOFFA Library', and 'Chip Library'. The main window is titled 'SIGNIFICANT\_GENELIST' and has tabs for 'Input/Output', 'Library', and 'GeneList'. The 'GeneList' tab is active, showing a table of gene data. A red oval highlights the 'Chip Lib', 'Genes', 'Proteins', 'Pathways', 'GOFFA', and 'Orthologene' buttons in the toolbar. The table below shows the following data:

	GENELIST_NAME *	EXPID *	GENEBANKACC	GENENAME	LOCUSID	FOLD	PVALUE
1	123gene_Welch_F1.5P0.	650	E00717	Cyp1a1	24296	2.7807	0.0011
2	123gene_Welch_F1.5P0.	650	E00778	Cyp1a1	24296	3.9988	0.0002
3	123gene_Welch_F1.5P0.	650	X07365	Gpx1	24404	1.7031	0.0001
4	123gene_Welch_F1.5P0.	650	X12367	Gpx1	24404	1.716	0.0015
5	123gene_Welch_F1.5P0.	650	J02722	Hmox1	24451	0.16	0.0002
6	123gene_Welch_F1.5P0.	650	M86389	Hspb1	24471	0.6191	0.0278
7	123gene_Welch_F1.5P0.	650	AA998683	Hspb1	24471	0.6025	0.0068
8	123gene_Welch_F1.5P0.	650	A1176658	Hspb1	24471	0.6102	0.0092
9	123gene_Welch_F1.5P0.	650	L16764	Hspa1a // Hspa1b_mappe	24472	0.3038	0.0037
10	123gene_Welch_F1.5P0.	650	U65007	Met	24553	2.195	0.0016
11	123gene_Welch_F1.5P0.	650	X56420	Mpg	24561	2.1997	0.0348
12	123gene_Welch_F1.5P0.	650	X96394	Abcc1	24565	1.7617	0.0277
13	123gene_Welch_F1.5P0.	650	M11794	Mt1a	24567	0.5328	0.0056
14	123gene_Welch_F1.5P0.	650	X68394	Nras	24605	1.5958	0.0178



# Interpretation (continued)

Gene Library

Specify ID Type:

- GenBankAcc
- UnigeneID
- LocusID
- SwissProtAcc
- IMAGEID
- GEN\_ID\_MFR
- GeneSymbol

Enter Searching Data:

Search  within result

Clear

Message:  
unique search ID number :94

GOFFA KEGG PathArt Pathways Proteins Orthologene Chip Lib

Customize Table Export More Info ... Select one ... Link To ... Select one ... Help

Filter>	GENENAME	DESCRIPTION	SPECIES	CHROMLOCA.	LOCUSID	BIOLOGI
1	Nat1	N-acetyltransferase 1 (arylamine N-ace	Rattus norvegicus	16p14	116631	metabolism
2	Tap1	transporter 1, ATP-binding cassette, su	Rattus norvegicus	20p12	24811	defense response;immune resp
3	Mezf2d	myocyte enhancer factor 2D	Rattus norvegicus	2q34	81518	
4	Adprt	ADP-ribosyltransferase 1	Rattus norvegicus	13q26	25591	protein amino acid ADP-ribosylat
5	Dpyd	dihydropyrimidine dehydrogenase	Rattus norvegicus	2q41	81656	pyrimidine base catabolism
6	Mapk7	mitogen-activated protein kinase 7	Rattus norvegicus	10q23	114509	
7	LOC288591	similar to transmembrane protein indu	Rattus norvegicus	12q12	288591	
8	Dusp6	dual specificity phosphatase 6	Rattus norvegicus	7q13	116663	cell differentiation;protein amino
9	LOC288444	similar to heat shock protein 105 kDa s	Rattus norvegicus	12p11	288444	
10	Gstm5	glutathione S-transferase, mu 5	Rattus norvegicus	2q34	64352	
11	Nras	neuroblastoma RAS viral (v-ras) oncog	Rattus norvegicus	2q34	24605	Ras protein signal transduction,c
12	Rara	retinoic acid receptor, alpha	Rattus norvegicus	10q31	24705	regulation of anti-apoptosis;regul
13	Dia1	diaphorase 1	Rattus norvegicus	7q34	25035	electron transport
14	Cd44	CD44 antigen	Rattus norvegicus	3q31	25406	cell adhesion;defense response;
15	Nat2	N-Acetyltransferase-2	Rattus norvegicus	16p14	116632	metabolism
16	LOC308047	similar to heat shock protein, DNAJ-like	Rattus norvegicus	1p11	308047	
17	Met	met proto-oncogene	Rattus norvegicus	4q21	24553	adult behavior;brain developmen
18	Hspb1	heat shock 27kDa protein 1	Rattus norvegicus	12q12	24471	
19	Mapk14	mitogen activated protein kinase 14	Rattus norvegicus	20p12	81649	angiogenesis;protein amino acid
20	LOC363249	similar to golli-interacting protein	Rattus norvegicus	9q33	363249	
21	Hadha	hydroxyacyl-Coenzyme A dehydrogenas	Rattus norvegicus	6q12	170670	fatty acid metabolism;metabolism
22	Camlg	calcium modulating ligand	Rattus norvegicus	17p14	81715	
23	Cdk7	cyclin-dependent kinase 7 (MO15 hom	Rattus norvegicus	2q12	171150	cell cycle;cell cycle;cytokinesis;m
24	Cox6a2	cytochrome c oxidase, subunit VIa, poly	Rattus norvegicus	1q36	25278	electron transport
25	Mt1a	Metallothionein	Rattus norvegicus	19p12	24567	nitric oxide mediated signal trans
26	Hspa1b	heat shock 70kD protein 1B	Rattus norvegicus	20p12	294254	response to heat
27	LOC293991	similar to NADH dehydrogenase (ubiqui	Rattus norvegicus	1q54	293991	
28	LOC315994	similar to Expressed sequence A18748	Rattus norvegicus	8q32	315994	
29	Map3k1	mitogen activated protein kinase kinas	Rattus norvegicus	2q14	116667	protein amino acid phosphorylati

- The significant genes are listed here in Gene Library.

- Can search and sort the Gene library

- There are links to other Libraries(Kegg, Pathart)



# Interpretation (continued)

- KEGG – Kyoto Encyclopedia of Genes and Genomes  
<http://www.genome.jp/kegg/>
- KEGG is a suite of databases and associated software.
- KEGG Pathway database provides the information of metabolic, regulatory and disease pathways; Most of them are metabolic pathways.



# Interpretation (continued)

PathArt (Jubilant) – a pathway database

- The Pathways (over 600 mammalian disease and signaling)
- The Pathways is a collection of manually curated information from literature and public domain databases.

## In ArrayTrack

	Human	Rat	Mouse
Kegg	200	187	193
PathArt	587	151	297



# Interpretation (continued) Kegg

Genes involved in a pathway

Pathway name

Pathway category

Statistical significance of the pathway

Gene	Map	Category	Fisher P Value
Gstm5 Gpx1	Glutathione metabolism(mmu00480)	Metabolism of Other Amino Acids/Metab...	0.00370292
Ptgs2	Prostaglandin and leukotriene metaboli...	Lipid Metabolism/Metabolic pathway	0.25805128
Hmox1	Porphyrin and chlorophyll metabolism(m...	Metabolism of Cofactors and Vitamins/M...	0.13404114
Dusp6 Hspb1 Mapk7 Mapk14 Tnfrsf1a Tgfr2 Nras Mapk9 Map3k1	MAPK signaling pathway(mmu04010)	Regulatory pathway	0.00019113
Met Ngfr Tgfr2 Tnfrsf1a	Cytokine-cytokine receptor interaction(m...	Regulatory pathway	0.17855902
Tnfrsf1a	Apoptosis(mmu04210)	Regulatory pathway	0.58437738
Mapk9	Wnt signaling pathway(mmu04310)	Regulatory pathway	0.77999365
Tgfr2	TGF-beta signaling pathway(mmu04350)	Regulatory pathway	0.55443832
Mapk14 Mapk9	Toll-like receptor signaling pathway(mm...	Regulatory pathway	0.23520864
Csnk1d	Circadian rhythm(mmu04710)	Regulatory pathway	0.11721566
Mapk14 Mapk9	Parkinson's disease(mmu05020)	Regulatory pathway	0.03611750
Gpx1	Amyotrophic lateral sclerosis (ALS)(mm...	Regulatory pathway	0.15056025

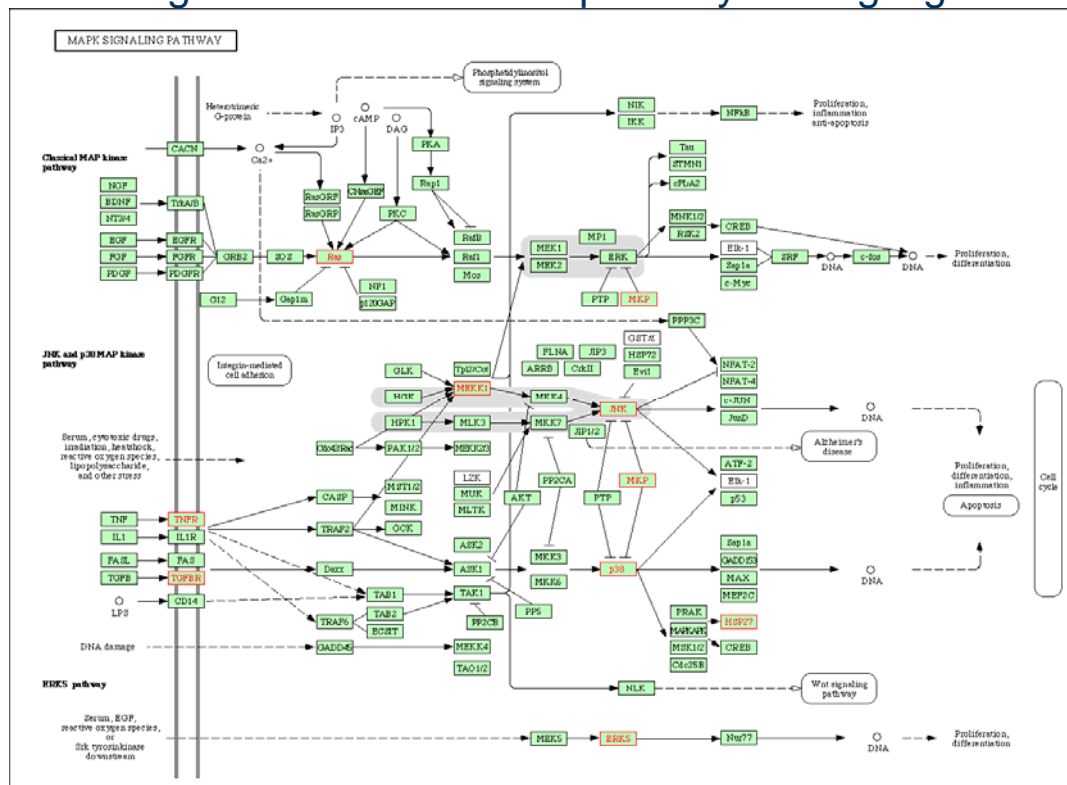
Input genes = 60, 22 genes found, 38 not found, Total 23 pathway maps.

Kegg Pathway



# Interpretation (continued) Kegg

Double-click a specific pathway, the pathway map will be displayed and the genes involved in the pathway are highlighted.



# Interpretation (continued) PathArt

Genes

Pathways

Physiology/disease

Statistical significance of the pathway

Component n...	Pathway Name	Organism	Physiology...	IS_PHYSIO...	Pathway Ty...	Fisher P V...
Mapk14	CCR5 Mediated Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.089682
Gclc	Tat Signaling Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.236507
Mapk14	AGEs Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.149045
Mapk14 Tnfrsf1a	Amyloidbeta-peptide Signal...	Mus muscu...	Alzheimers	Disease	Signaling	0.118019
Tnfrsf1a	IFN Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.634464
Mapk14	LPS Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.386674
Map3k1	DR4/5 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.064863
Mapk14	IGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.717471
Ngfr	NGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.204750
Mer2d	T-cell apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.149045
Mapk14 Tgfb2	TGF Beta Induced apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.051402
Tnfrsf1a	TNF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.776034
Mapk14	TNFR1 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.246824
Adprt	TRAIL Mediated Apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.378170
Mapk14	UV induced Antiapoptotic P...	Mus muscu...	Apoptosis	Physiology	Signaling	0.113880
Cd44	WNT Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.644709
Mapk14	Interleukin Signaling Pathw...	Mus muscu...	Atheroscle...	Disease	Signaling	0.926753
Met	HGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.064863
Tgfb2	TGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.267067
Cd44	CD44 Signaling Pathway	Mus muscu...	Cell Adhe...	Physiology	Signaling	0.052216
Tp53	ATR-ATM Signaling Pathway	Mus muscu...	Cell Cycle	Physiology	Signaling	0.204750
Map3k1 Vegf	Bcr-Abl Signaling Pathway	Mus muscu...	Chronic M...	Disease	Signaling	0.217718

Total genes =60, Found genes= 17, Found pathways= 58







# Interpretation (continued)

## GOFFA

- GOFFA – Gene Ontology For Functional Analysis
- Developed based on Gene Ontology(GO) database
- Grouping the genes into functional classes
- GO- three ontologies
  - Molecular function**: activities performed by individual gene products at the molecular level, such as catalytic activity, transporter activity, binding.
  - Biological process**: broad biological goals accomplished by ordered assemblies of molecular functions, such as cell growth, signal transduction, metabolism.
  - Cellular component**: the place in the cell where a gene product is found, such as nucleus, ribosome, proteasome.





# Interpretation (continued) GOFFA

**Go Term Cluster**

**Update**

Select data type

GenBankAcc

UnigenelID

LocusID

SwissProtAcc

Gene name (  Official  Synonym )

Human  Mouse  Rat  All

IPI protein

Human  Mouse  Rat

Select array type

Affy\_RT-U34

**Input Data**

24561  
24565  
24567  
24605  
24646  
24811  
24842  
24862  
24888  
24891  
25035  
25283

**Tree**   **Terms**   **Genes**   **GO Path**   **GO Tree Prune**

- all(15/553 P=1.000000 E=1.00 R=0.00)
  - molecular\_function(15/522 P=0.416096 E=1.06 R=0.38)
    - enzyme regulator activity(1/25 P=0.504942 E=1.47 R=0.68)
      - caspase regulator activity(1/3 P=0.079327 E=12.29 R=1.78)
      - enzyme inhibitor activity(1/8 P=0.198620 E=4.61 R=1.38)
        - caspase inhibitor activity(1/2 P=0.053562 E=18.43 R=2.65)
    - transcription regulator activity(1/43 P=0.707858 E=0.86 R=0.53)
    - catalytic activity(10/345 P=0.478234 E=1.07 R=0.70)
    - signal transducer activity(2/104 P=0.807340 E=0.71 R=0.47)
    - transporter activity(3/43 P=0.103225 E=2.57 R=1.37)
    - binding(12/380 P=0.257422 E=1.16 R=0.97)
  - biological\_process(13/486 P=0.730858 E=0.99 R=0.14)
    - regulation of biological process(6/179 P=0.349511 E=1.24 R=0.59)
    - response to stimulus(9/206 P=0.059628 E=1.61 R=1.36)
    - development(2/137 P=0.919836 E=0.54 R=0.17)
    - physiological process(13/464 P=0.555738 E=1.03 R=0.39)
    - cellular process(13/460 P=0.524454 E=1.04 R=0.42)
    - growth(1/19 P=0.412166 E=1.94 R=0.52)
  - cellular\_component(14/457 P=0.233694 E=1.13 R=0.63)
    - extracellular region part(1/110 P=0.965779 E=0.34 R=0.65)
    - organelle part(6/119 P=0.079874 E=1.86 R=1.73)
    - cell part(14/426 P=0.105976 E=1.21 R=1.61)
    - extracellular region(1/115 P=0.971216 E=0.32 R=0.64)
    - cell(14/426 P=0.105976 E=1.21 R=1.61)
    - membrane-enclosed lumen(3/37 P=0.071802 E=2.99 R=1.78)
    - envelope(2/44 P=0.338385 E=1.68 R=1.10)
    - organelle(10/309 P=0.280762 E=1.19 R=1.18)
    - protein complex(2/61 P=0.222211 E=1.81 R=1.29)

**Search in Tree**

Find terms containing

GO term    Gene name or gene symbol    P Value    E Value

No	LocusID	Gene Name...
1	24471	Hspb1
2	24472	Hspa1a
3	24553	Met
4	24561	Mpg
5	24565	Abcc1
6	24567	Mt1a
7	24605	Nras
8	24646	Abcb1
9	24811	Tap1
10	24842	Up53
11	24862	Ugt2b
12	24888	Bcl2l1
13	24891	Abcb4
14	25035	Cy5b3
15	25283	Gclc

Total original submit =15, Found =15 with GO term



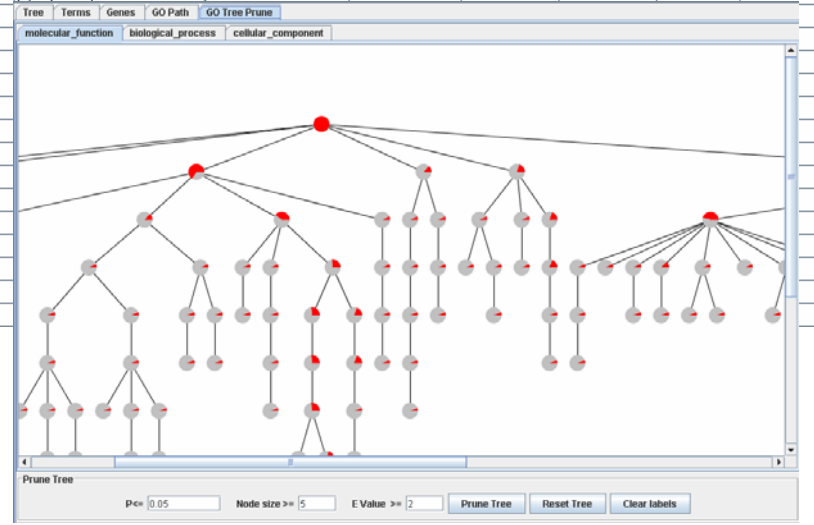
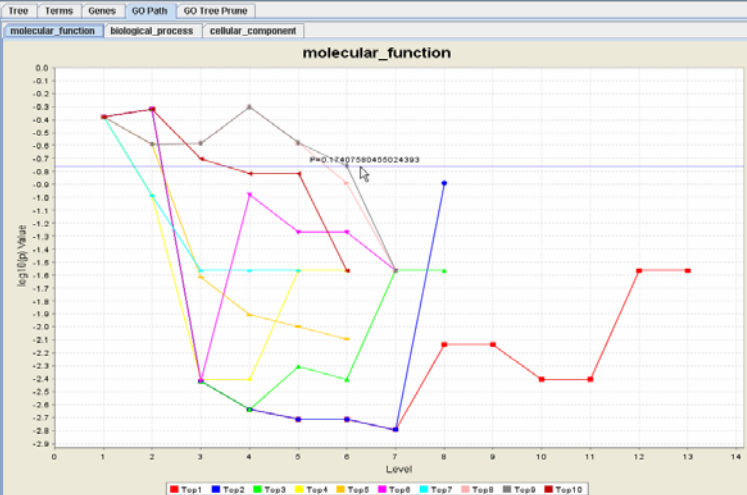
# GOFFA

Tree Terms Genes GO Path GO Tree Prune

- all(15/553 P=1.000000 E=1.00 R=0.00)
  - molecular\_function(15/522 P=0.416096 E=1.06 R=0.38)
    - enzyme regulator activity(1/25 P=0.504942 E=1.47 R=0.68)
      - caspase regulator activity(1/3 P=0.079327 E=12.29 R=1.78)
        - enzyme inhibitor activity(1/8 P=0.198620 E=4.61 R=1.38)
          - caspase inhibitor activity(1/2 P=0.053562 E=18.43 R=2.65)
      - transcription regulator activity(1/43 P=0.707858 E=0.86 R=0.53)
      - catalytic activity(10/345 P=0.478234 E=1.07 R=0.70)
      - signal transducer activity(2/104 P=0.807340 E=0.71 R=0.47)
      - transporter activity(3/43 P=0.103225 E=2.57 R=1.37)
      - binding(12/380 P=0.257422 E=1.16 R=0.97)
    - biological\_process(13/486 P=0.730858 E=0.99 R=0.14)
      - regulation of biological process(6/179 P=0.349511 E=1.24 R=0.59)
      - response to stimulus(9/206 P=0.059628 E=1.61 R=1.36)
      - development(2/137 P=0.919836 E=0.54 R=0.17)
      - physiological process(13/464 P=0.555738 E=1.03 R=0.39)
      - cellular process(13/460 P=0.524454 E=1.04 R=0.42)
      - growth(1/19 P=0.412166 E=1.94 R=0.52)
      - cellular\_component(14/457 P=0.233694 E=1.13 R=0.63)

Tree Terms Genes GO Path GO Tree Prune

Molecular function		Biological process		Cellular component			
No	Gene	Term	GO ID	Level (Aver...	P value... Δ	Gene Hits	E value
1	↑ Abcb1	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
2	↑ Abcc1	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
3	↑ Nras	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
4	↑ Tap1	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
5	↑ Abcb1	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
6	↑ Abcc1	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
7	↑ Nras	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
8	↑ Tap1	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
9	↑ Abcb1	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
10	↑ Abcc1	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
11	↑ Nras	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
12	↑ Tap1	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
13	↑ Abcb1						
14	↑ Abcc1						
15	↑ Nras						
16	↑ Tap1						



# Data Exploring

- **Scatter Plot**

plot the fluorescence intensity data of Cy3 vs Cy5 for the same array

- **Mixed Scatter Plot**

compare two arrays in one plot, applies to both 2-channel and 1-channel data.

- **Correlation Matrix**

Correlation Matrix shows the correlation between column  $i$  and column  $j$  of the original matrix. It visually shows the correlation between two groups of data.

- **Bar Chart**

displays expression data for a single gene across multiple arrays within the same experiment or across different experiment.

- **Principal Component Analysis (PCA)**

PCA is a way of identifying the data patterns and highlighting the data's similarity and difference

- **P-value Plot**

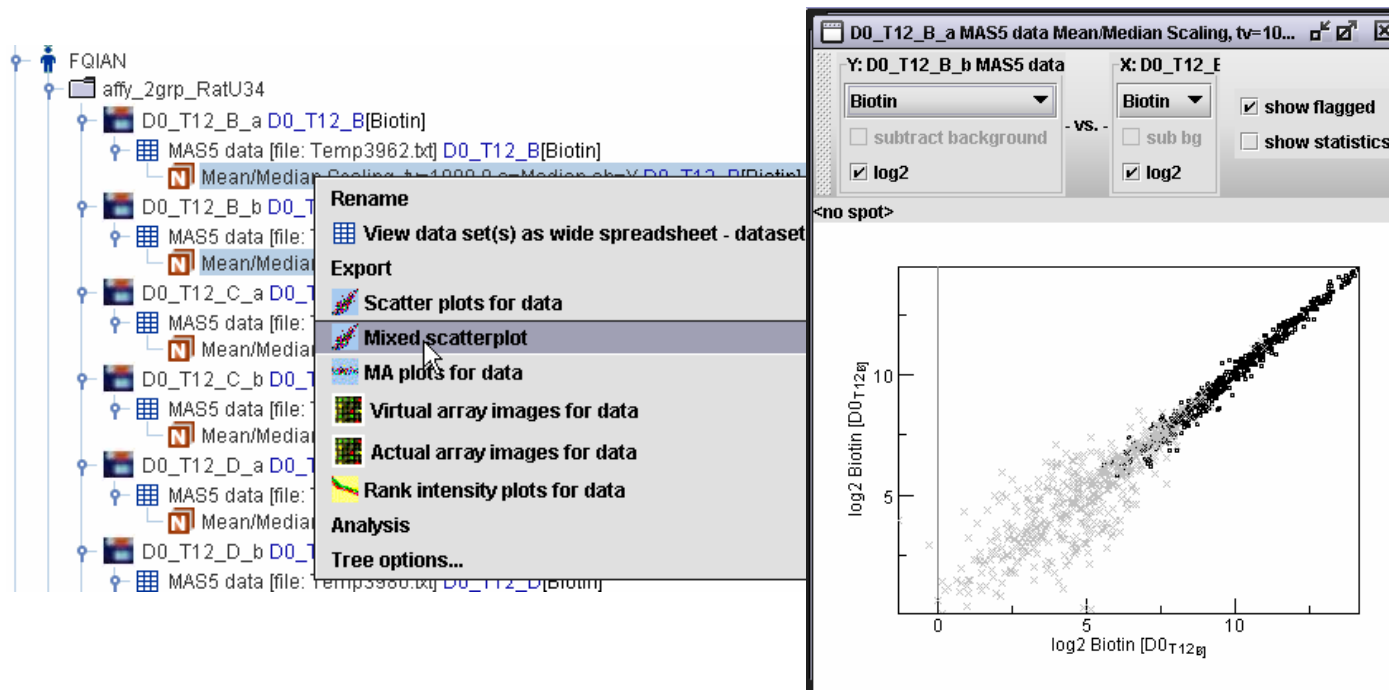
visual interpretation of P-value.

- **HCA**

is comprised of agglomerative methods and divisive methods that finds clusters of observational data set.



# Data Exploring scatter plot



# Data Exploring Bar chart

T-Test Results

File Selected-Spot All-Spots

Create Bar Chart of Spots' Expression Values Within Involved Experiment(s) Mean G

Spot	Gene	Exp1	Exp2	Exp3	P-Value	Mean G	
1	AA108308	AA108308	314856	na	516485	0.8937	8.507
2	AA108308	AA108308	314856	na	516486	0.0166	1.2235
3	AA108308	AA108308	314856	na	516486	0.0166	9.2682
4	AA684537	AA684537	294964	na	516487	0.0826	8.021
5	AA684929	AA684929			516488	0.1536	7.7018
6	AA684960	AA684960			516489	0.341	6.8608
7	AA684963	AA684963	293702	na	516490	0.4595	8.6452
8	AA685112	AA685112	293652	na	516491	0.0814	8.8775
9	AA685152	AA685152	25490	Nedd8	516492	0.1159	6.991
10	AA685376	AA685376			516493	0.346	6.558
11	AA685876	AA685876	304805	na	516494	0.0183	9.225
12	AA685903	AA685903	362862	na	516495	0.3713	11.1139
13	AA686031	AA686031	301458	na	516496	0.1077	7.0271
14	AA686579	AA686579	301442	na	516497	0.9527	5.4973
15	AA686870	AA686870			516498	0.3007	3.8987
16	AA686870	AA686870			516499	0.6599	3.3942
17	AA799336	AA799336	293453	na	516500	0.6183	9.8132

1031 genes

Significance Filtering

P Values <  without adjustment

Select # genes:  lowest p-vals

Mean Channel Intensities >  Bad Flags <=

Fold Change >

Apply Filters

P-Value Plot Create Sig. Gene List HCA PCA Volcano Plot

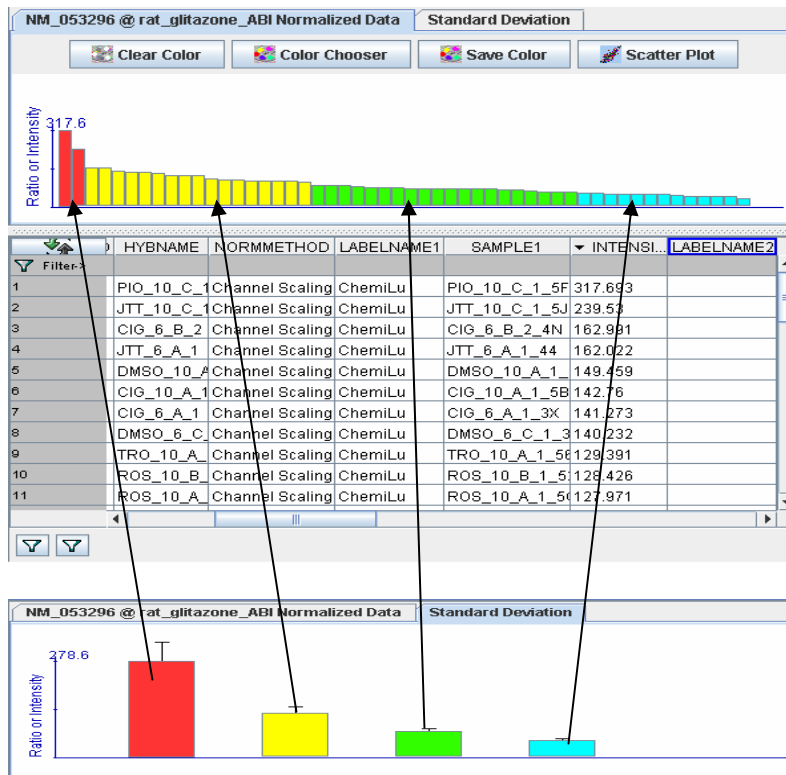
Access Bar chart from T-Test results.

Bar chart – display expression data for a single gene across multiple arrays in the same experiment or across different experiments.



# Data Exploring

## Bar chart continued



Grouping multiple arrays marked in different colors.

Group color could be saved.

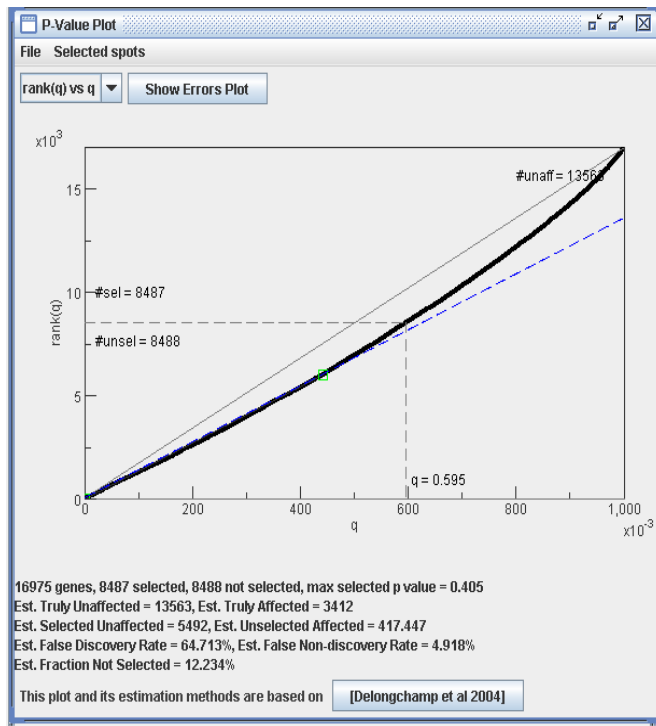
Standard deviation bar chart for the above groups. The bar height represents the mean intensity, while the T-line above the bar stands for the value of SD. The color for each bar echoes the color of the bar chart at the top, in the same order.



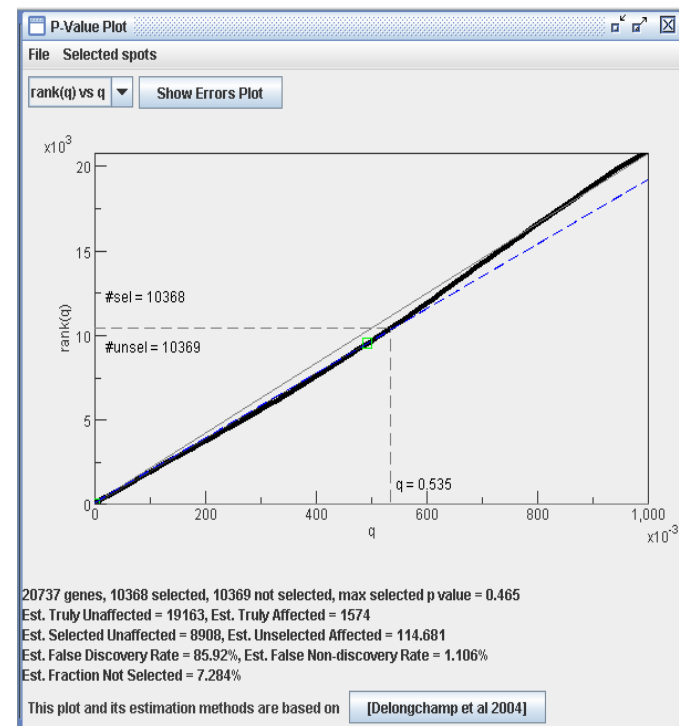
# Data Exploring

## P-value plot

### Some effect

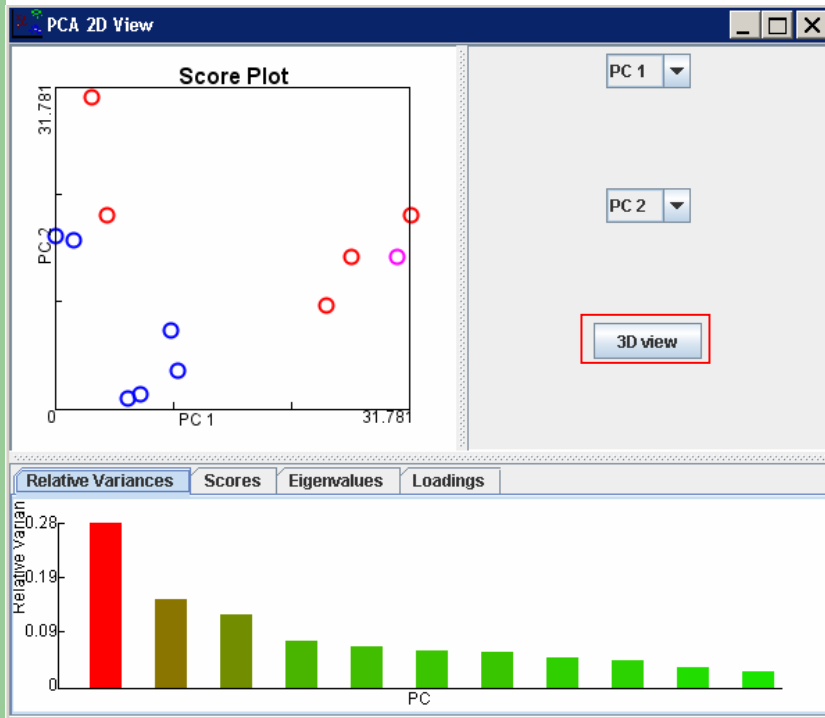


### Slightly or no effect

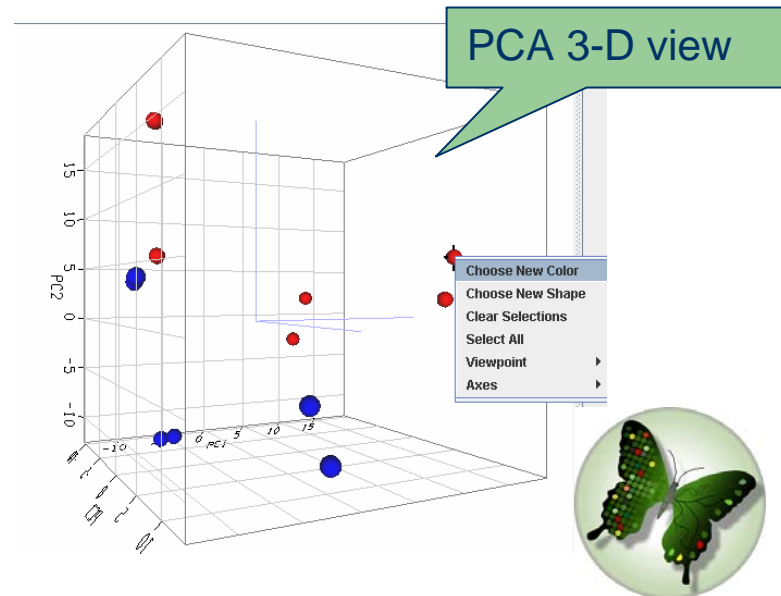


P plot curve is closer to diagonal line, the less treatment effect there is.

# Data Exploring PCA

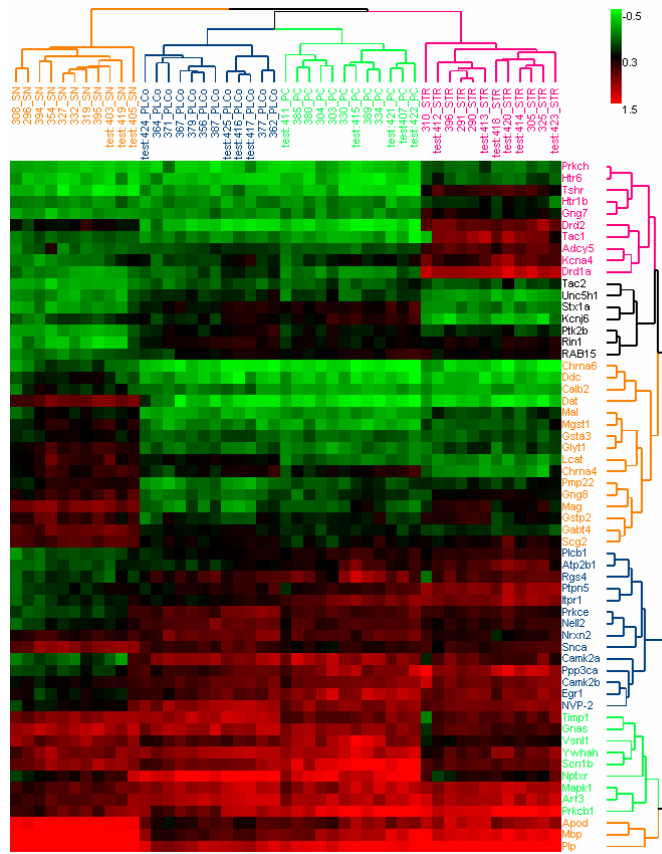


- Commonly used before gene selection
- To investigate the inter-sample relationship based on the gene expression profile
- Identify the outliers in biological/technical replicates
- View the variance of a multidimensional data





# Data Exploring HCA





# Accessing ArrayTrack

- FDA Internal:  
<http://weblaunch.nctr.fda.gov/jnlp/arraytrack/index.html>  
<http://weblaunch.nctr.fda.gov/jnlp/arraytrack/citrix/index.html>
- FDA External:  
<http://edkb.fda.gov/webstart/arraytrack/index.html>





# Technical Support

[NCTRBioinformaticsSupport@nctr.fda.gov](mailto:NCTRBioinformaticsSupport@nctr.fda.gov)

ArrayTrack is developed by the U.S. Food and Drug Administration, National Center for Toxicological Research (FDA/NCTR).

FDA/NCTR reserves all rights for the software .



# Thank you!



National Center for Toxicological Research  
U.S. Food and Drug Administration