

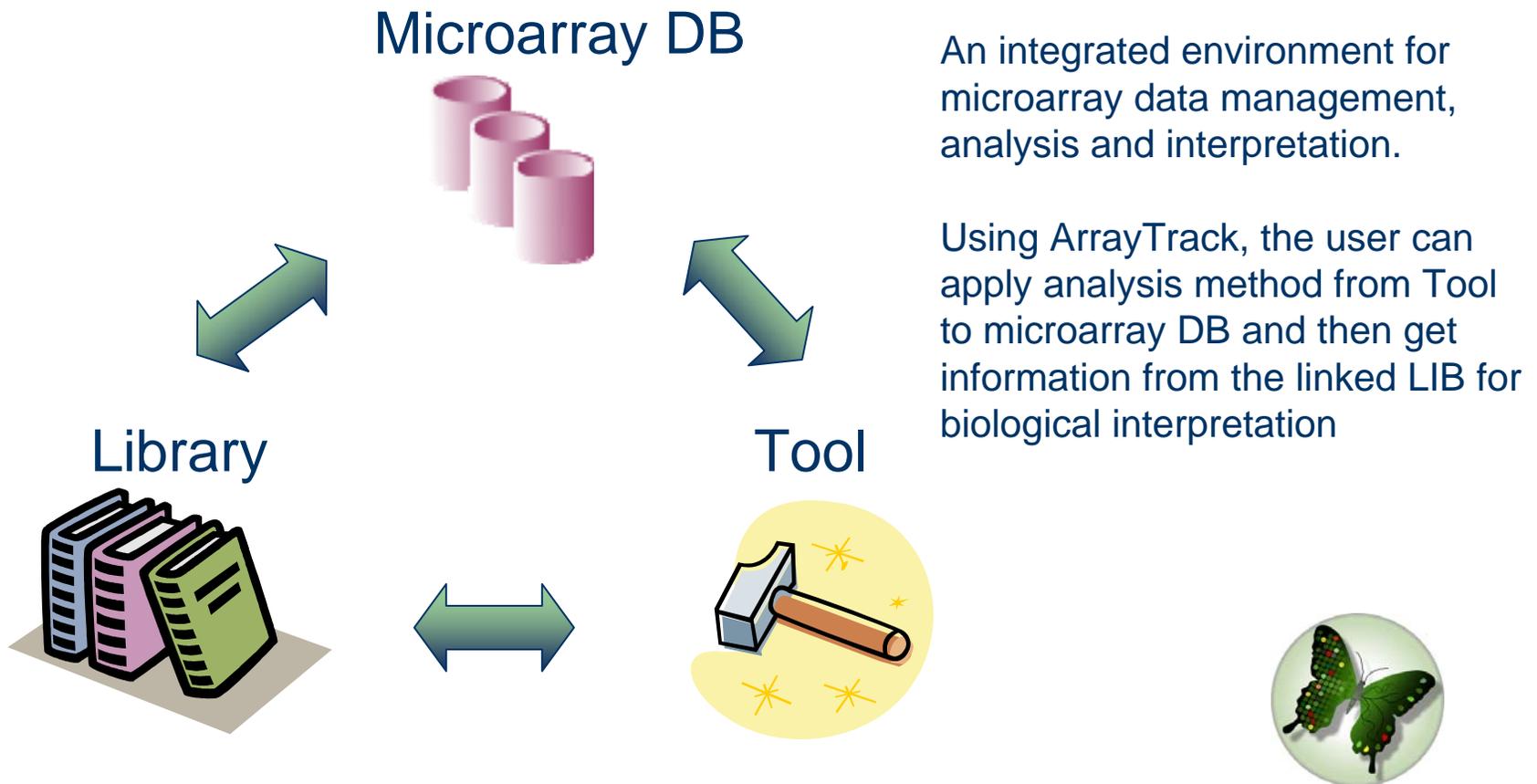


ArrayTrack_{3.5.0} Quick Start Manual

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



ArrayTrack Overview

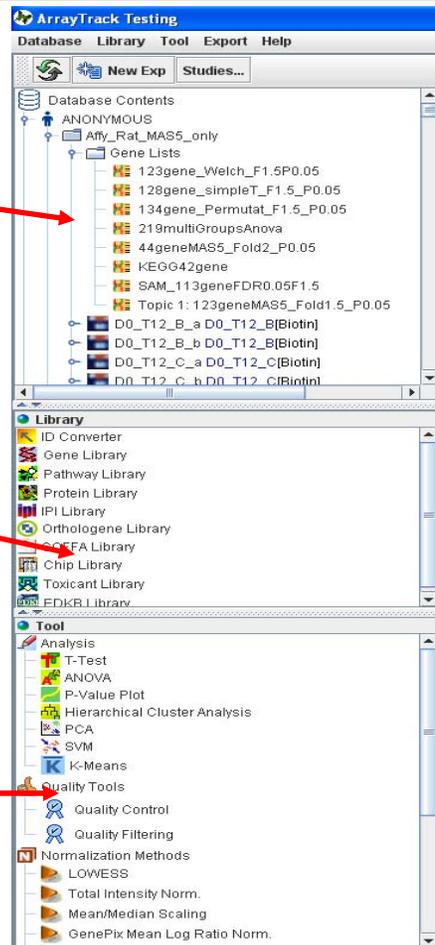


ArrayTrack Overview

Microarray DB

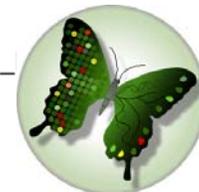
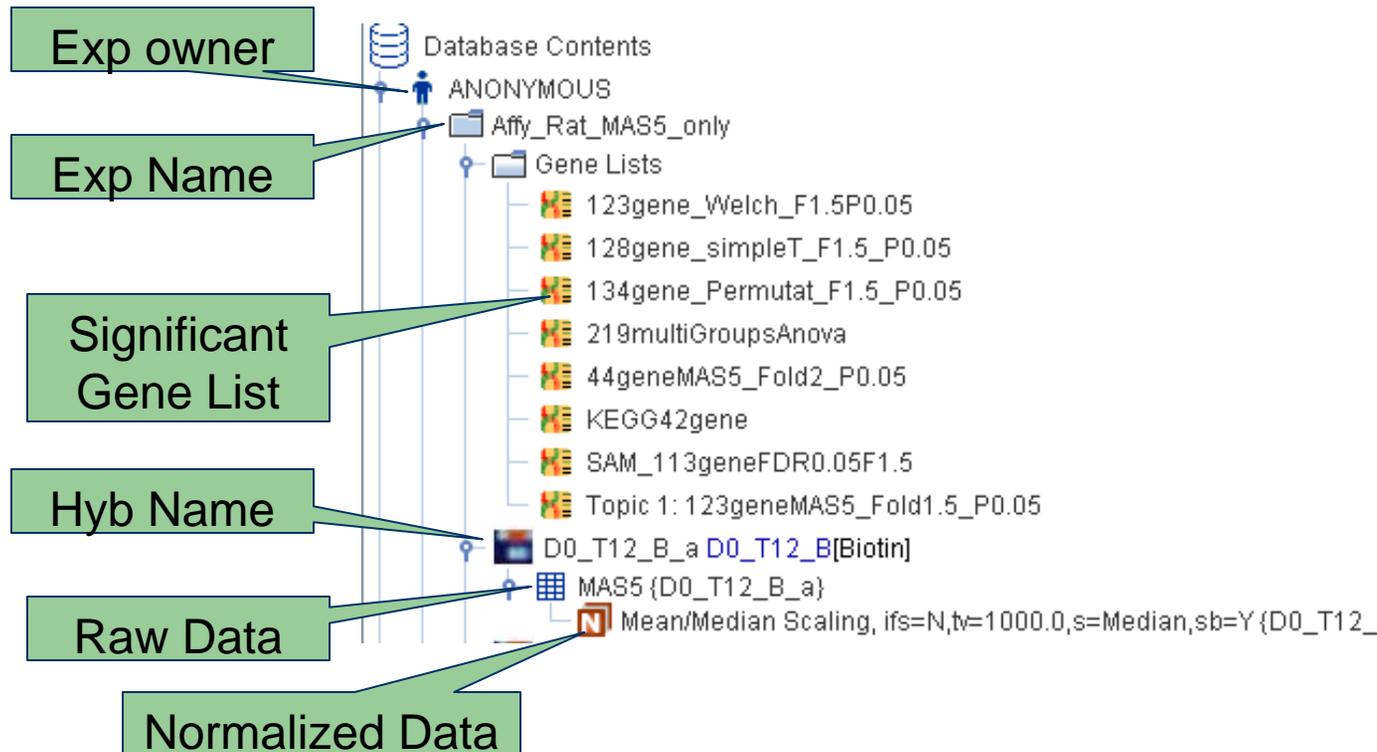
Libraries

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 screenshots illustrate the following context menu options:

- Batch Import Data**
- Batch Import Data - SimpleTox**
- Rename**
- Analysis Result Files**
- Gene Lists**
 - Display
- Select datasets**
- Import**
- Export**
- Delete**
- Show experiment design**
- Show dye-flip pairs**
- Expand completely**
- Collapse completely**
- Tree options...**

- Export**
 - Single-platform data as spreadsheet
 - Gene-matched data as spreadsheet (multi-platform OK)
 - Original data files or Affy probe-set files
 - Data in narrow format (datasets one after another)
 - Affy CEL files
 - Image files
 - Settings files
 - JMP/Genomics
 - DrugMatrix
 - R-Interface
- Analysis**
- Visualizations**
- Correlations**
- Convert Affy CEL files to probe sets**
- Create gene list by data filtering**
- Normalize**
- Quality control**
- Create duplicate data sets**
- Copy data sets for pasting elsewhere**
- Studies**
- Tree options**

- Statistics Tests**
 - T-Test/ANOVA
 - SAM-Test
 - Two-way ANOVA
 - T-Test with custom data options
 - ANOVA with custom data options
- Pattern Discovery**
 - Hierarchical Cluster Analysis
 - Principal Component Analysis
 - K-Means
- Model Prediction**
 - K-Nearest Neighbors
 - Linear Discriminant Analysis

Exploring DB

Experiment design

The screenshot shows a database application window titled "Hybridizations". The window has a menu bar with "Edit", "View", and "Input/Output". Below the menu bar are two icons: a refresh icon and a grey circle. The main area contains a table with the following columns: "Hybridization", "SAMPLE 1", "LABEL 1", "ARRAYTYPENAME", "CHANNEL", "SPECIES 1", "ASSAY 1", and "CELLTYPE 1". The table contains three rows of data. A red arrow points to the first row. A context menu is open over the table, listing various actions 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". The "Show experiment design" option is highlighted by the mouse cursor.

Hybridization	SAMPLE 1	LABEL 1	ARRAYTYPENAME	CHANNEL	SPECIES 1	ASSAY 1	CELLTYPE 1	
1	D0_T12_B_a	D0_T12_B	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
2	D0_T12_B_b	D0_T12_B	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
3	D0_T12_C_a	D0_T12_C	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes

Database Contents

- ANONYMOUS
 - Affy_Rat_MAS5_only
 - Gene Lists
 - D0_T12_B_a D
 - MAS5 (D0_T
 - Mean/Me
 - D0_T12_B_b D
 - MAS5 (D0_T
 - Mean/Me
 - D0_T12_C_a D
 - MAS5 (D0_T
 - Mean/Me
 - D0_T12_C_b D

Searching Libraries



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

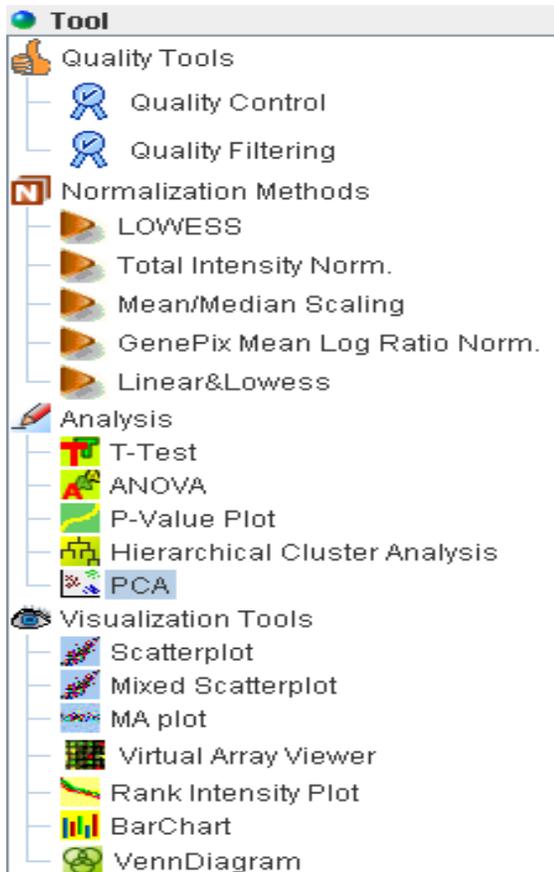
The libraries integrate gene, protein, pathway and other data allowing data interrogation and mining of data across data types.

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

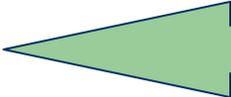
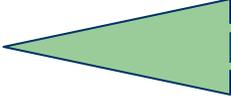
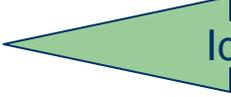
Provides different methods to evaluate the microarray data

•Visualization Tools

Provides a direct view to identify abnormalities within data.

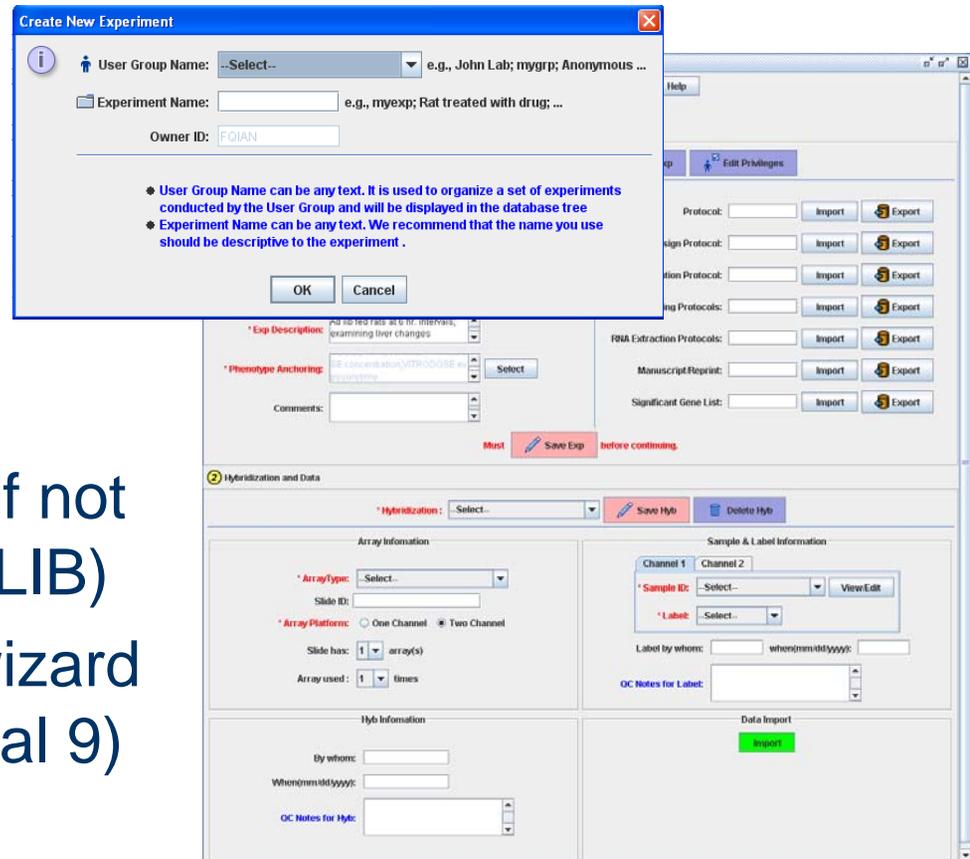
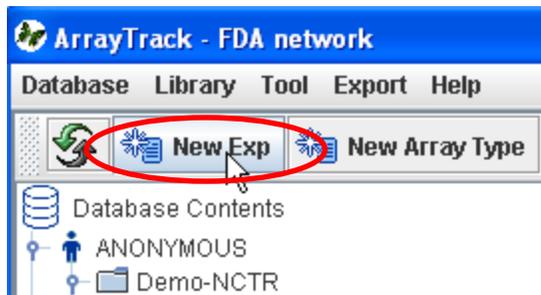


Microarray Data Processing

- 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



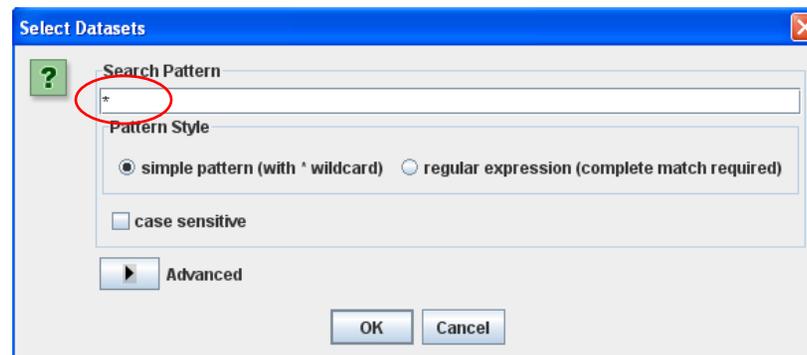
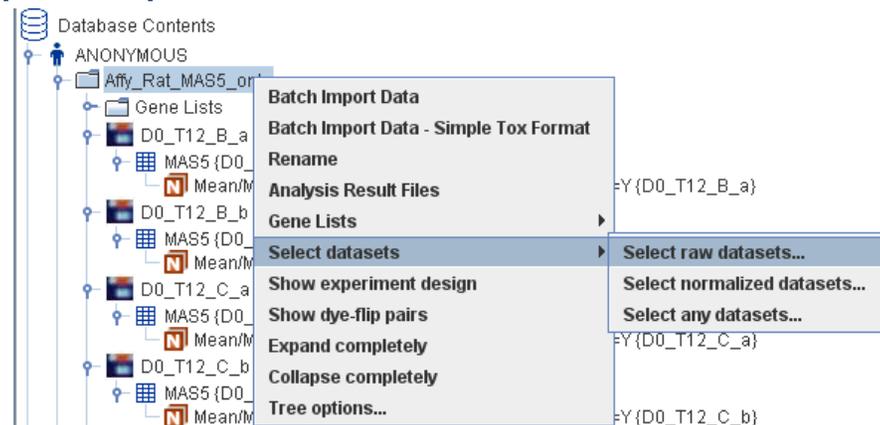
1. Create experiment
2. Create array type (if not existed in AT Chip LIB)
3. Use batch import wizard (see detail in Tutorial 9)

Normalization

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

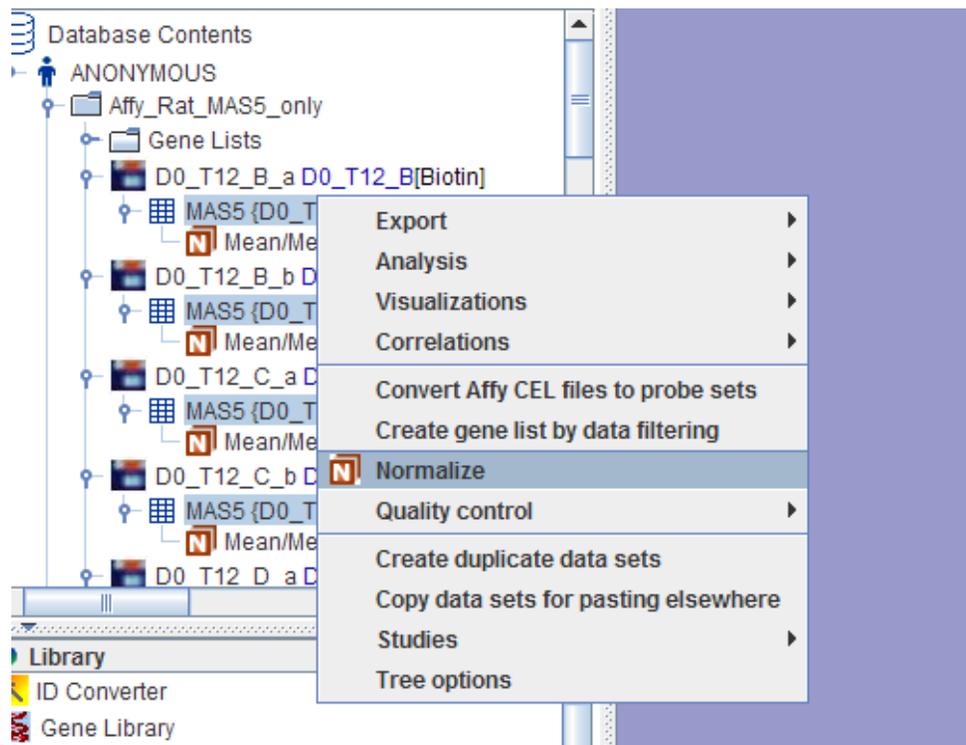
1. Right-click an experiment
2. Choose “Select datasets”
->”Select raw dataset...”

3. Type wildcard * if you want to select all the data.
Click OK button.



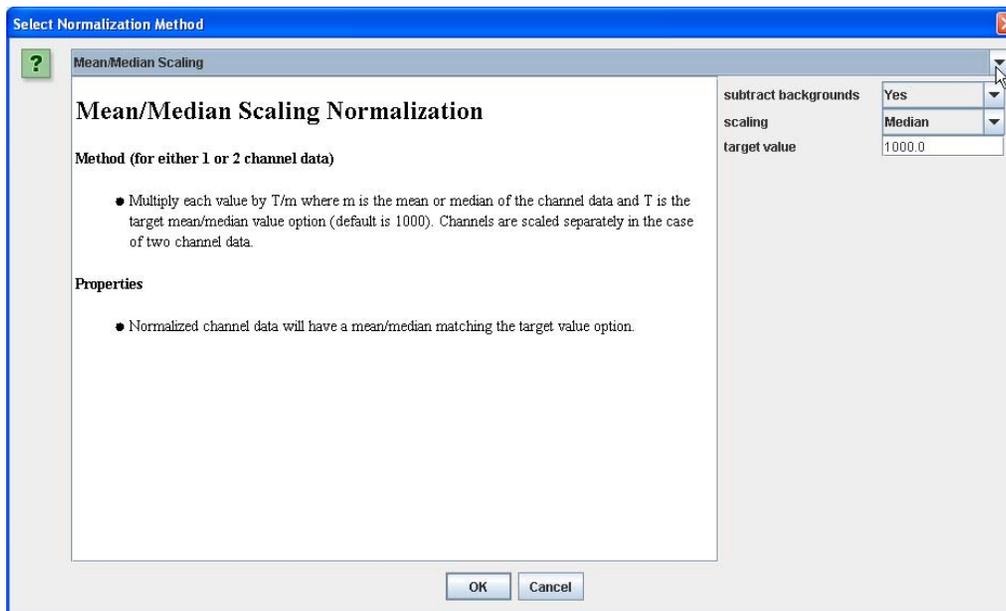
Normalization (-continued)

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 control and treatment.

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

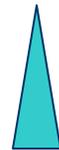


Bonferroni correction

P_i/n

Low sensitivity

Low power



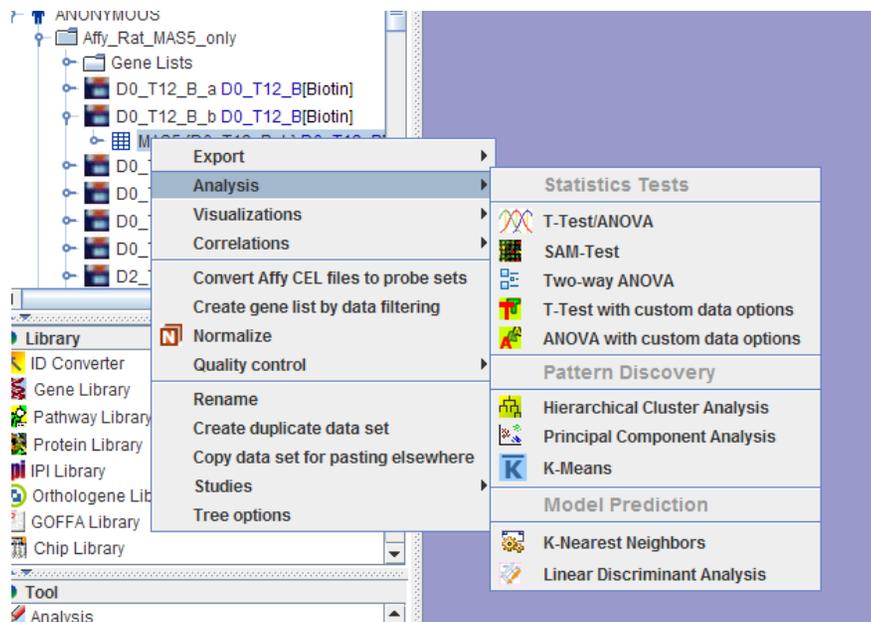
False discovery rate (e.g., Benjamini & Hochberg, p-value plot)

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

Volcano plot (combination of p and fold change)

Gene Selection

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



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	ARRAYTYPE NAME	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 grp 1/grp 2, so "up" regulation will mean grp1 > 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	L15079	L15079mR...	24891	Abcb4	NM_012690				516807	0	1.8323	1.8323
3	D87336	D87336_g...	287552	Blmh					516718	0.0001	1.7665	1.7665
4	X07365	X07365_s...	24404	Gpx1					517129	0.00		
5	M55534	M55534m...	25420	Cryab	NM_012935				516911	0.00		
6	J02722	J02722cds...	24451	Hmx1					516748	0.00		
7	E00778	E00778cds...	24296	Cyp1a1					516730	0.00		
8	AA848563	AA848563...		Hspa1a III ...					516506	0.00		
9	AA818604	rc_AA8186...		Hspa1a III ...					517260	0.00		
10	Z75029	Z75029_s...	294254	Hspa1b					517240	0.0064		
11	U01344	U01344_g...	116631	Nat1	NM_053853				516979	0.0005		
12	M99169	M99169_at	81771	Rps6ka1	NM_031107				516964	0.008	1.56	1.56
13	AA108277	AA108277...	288444	Hsph1					516484	0.0007	4.1286	0.2422
14	AI229655	rc_AI22965...	363249	Ctdsp1					517278	0.0007	1.5386	1.5386
15	D16478	D16478_g...	170670	Hadha	NM_130...				517415	0.0008	1.5541	1.5541
16	AI103396	rc_AI10339...							517415	0.0011	2.3716	2.3716
17	E00717	E00717LIT	24296	Cyp1...					516729	0.0011	2.7807	2.7807

123 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 > 1.5 **Advanced>>**

Apply Filters Clear Filters

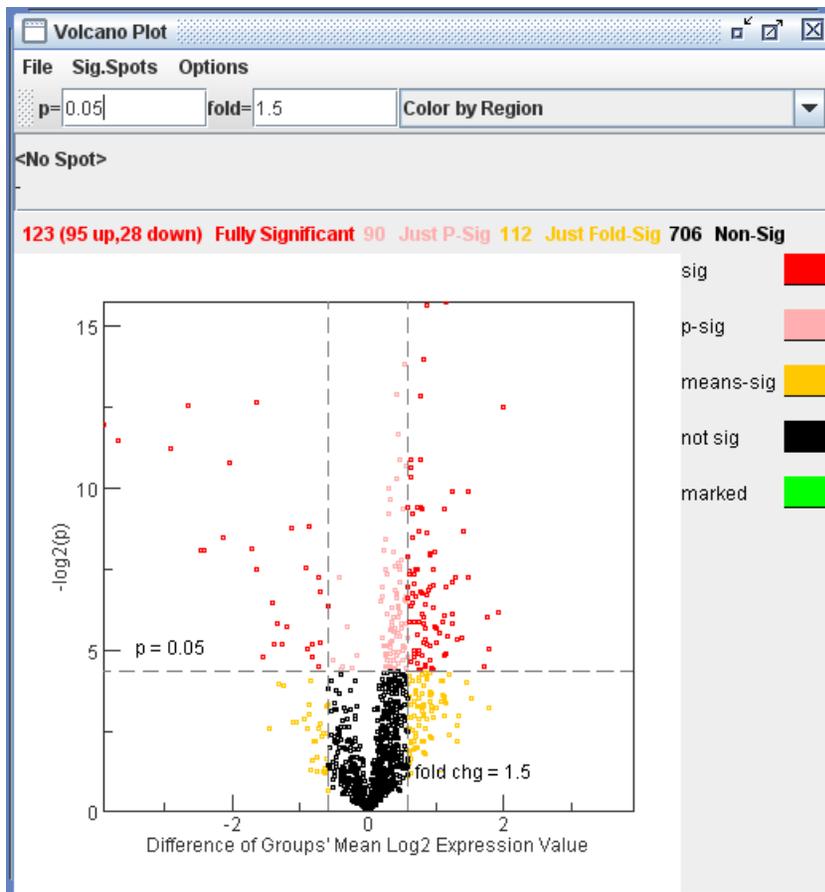
P-Value Plot Create Sig. Gene List HCA PCA K-Means Volcano Plot

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' software interface. A menu is open over the 'All Spots' column, with the option 'Create significant gene list...' highlighted. Other menu options include 'Hierarchical Clustering Analysis', 'Principal Component Analysis', 'Library search', and 'Select in all open viewers'. A secondary menu is open over the 'Library search' option, showing 'Gene Library', 'Pathway Library', and 'Protein Library'. Below the menu, a table of results is visible with columns for Genbank Acc, SPOTID, P, Mean Grp 1, Mean Grp 2, and Fold Change. At the bottom, a 'Significance Filtering' section contains radio buttons for 'P Values < [] without adjustment', 'Target False Discovery Rate (FDR): []', and 'Select # genes with lowest p-vals: []'. There are also input fields for 'Mean Channel Intensities > [] Bad Flags <= []' and 'Fold Change > []'. An 'Apply Filters' button is located below these fields. At the very bottom, there are buttons for 'P-Value Plot', 'MAQC', 'Create Sig. Gene List', 'HCA', 'PCA', and 'Volcano Plot'.

Genbank Acc	SPOTID	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			
U86105			-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_1	18776	Csh2	159805	0.0008	-0.8595	-1.748	1.8513
NM_019486	NM_019486_1	56015	Olfir71	162758	0.0008	0.2548	-0.0529	1.2377
NM_008319	NM_008319_1	15898	Icam5	145874	0.0009	-0.1598	-0.4882	1.2564
NM_021447	NM_021447_1	58522	Rnf30	149205	0.001	0.4095	0.1626	1.1867
NM_015780	NM_015780_1	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_010419_1	15208	Hes5	157161	0.0011	0.3868	0.1695	1.1625
AF331708	AF331708_1	230103	Npr2	158777	0.0012	-1.6946	0.4169	0.2314
AB041350	AB041350_1	112830	Col4a5	158418	0.0012	-2.8198	-3.7082	1.8511

Significant Gene List(-cont)

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

Right-click the experiment name

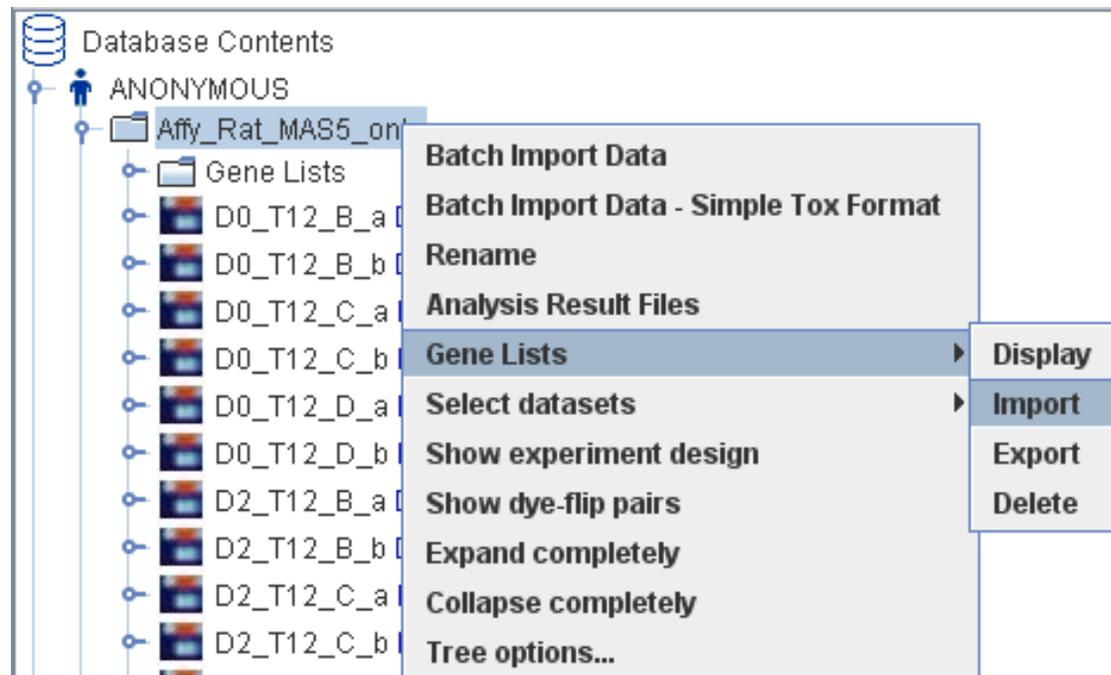
select Significant Gene Lists

select Display

Import

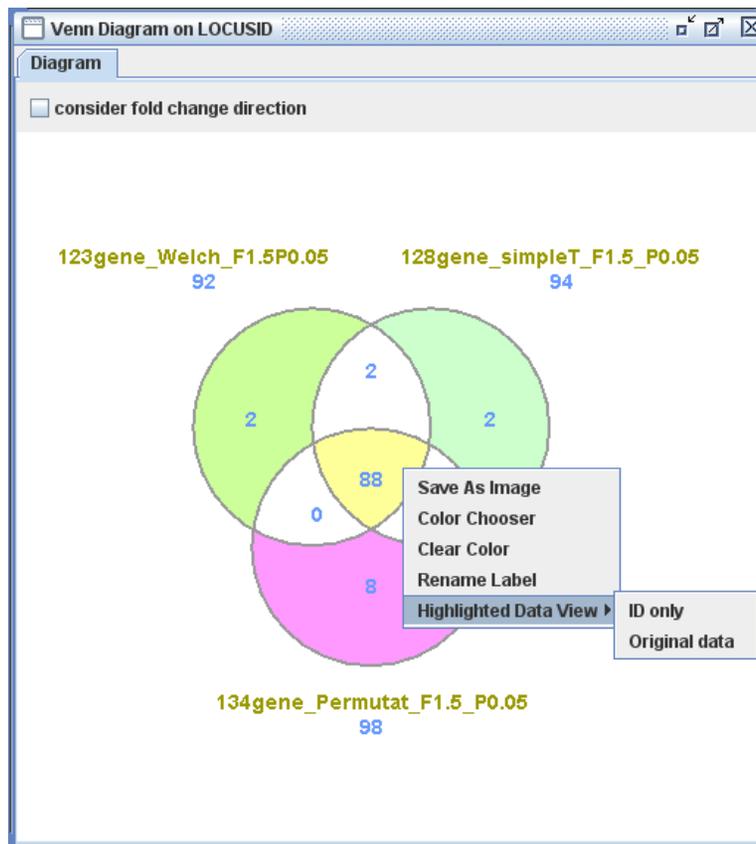
Export

Delete



Significant Gene List(-cont)

- **VennDiagram** - get common genes from 2~3 significant gene, protein, or metab lists.



Affy_Rat_MAS5_only

- Gene Lists
 - 123gene_Welch_F1.5P0.05
 - 128gene_simpleT_F1.5_P0.05
 - 134gene_Permutat_F1.5_P0.05
 - 219multiGroupsAnova
 - 44geneMAS5_Fold2_P0.05
 - KEGG42gene
 - SAM_113geneFDR0.05F1.5
 - Topic 1: 123geneMAS5_Fol
- 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)

Open
Export
Rename
Delete
Move To (New) Folder
Move Out Folder
Copy Genelist For Pasting Elsewhere
VennDiagram

Visualization Tools

- Scatterplot
- Mixed Scatterplot
- MA plot
- Virtual Array Viewer
- Rank Intensity Plot
- BarChart
- VennDiagram

By CommonID
By Pathway
By GeneOntology

Manufacture ID
Locus ID
GeneName
GeneBankAcc
REFSEQ
Spot ID

Tool Export Help

Visualizations

- Virtual Array Viewer
- Scatter Plot
- Mixed Scatterplot
- MA Plot
- Rank Intensity Plot
- Cross-Dataset Gene Barchart...
- VennDiagram

Interpretation

Link the significant genes to various libraries 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_F1.5P0.05' 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. On the right, a window titled 'SIGNIFICANT_GENELIST' displays a table with columns: GENELIST_NAME *, EXPID *, GENE BANKACC, GENENAME, LOCUSID, FOLD, and PVALUE. A red oval highlights the 'Input/Output' and 'Library' tabs, and the 'GeneList' tab. The table contains 14 rows of gene data.

	GENELIST_NAME *	EXPID *	GENE BANKACC	GENENAME	LOCUSID	FOLD	PVALUE
1	123gene_Welch_F1.5P0.05	650	E00717	Cyp1a1	24296	2.7807	0.0011
2	123gene_Welch_F1.5P0.05	650	E00778	Cyp1a1	24296	3.9988	0.0002
3	123gene_Welch_F1.5P0.05	650	X07365	Gpx1	24404	1.7031	0.0001
4	123gene_Welch_F1.5P0.05	650	X12367	Gpx1	24404	1.716	0.0015
5	123gene_Welch_F1.5P0.05	650	J02722	Hmox1	24451	0.16	0.0002
6	123gene_Welch_F1.5P0.05	650	M86389	Hspb1	24471	0.6191	0.0278
7	123gene_Welch_F1.5P0.05	650	AA998683	Hspb1	24471	0.6025	0.0068
8	123gene_Welch_F1.5P0.05	650	A1176658	Hspb1	24471	0.6102	0.0092
9	123gene_Welch_F1.5P0.05	650	L16764	Hspa1a // Hspa1b_mappe	24472	0.3038	0.0037
10	123gene_Welch_F1.5P0.05	650	U65007	Met	24553	2.195	0.0016
11	123gene_Welch_F1.5P0.05	650	X56420	Mpg	24561	2.1997	0.0348
12	123gene_Welch_F1.5P0.05	650	X96394	Abcc1	24565	1.7617	0.0277
13	123gene_Welch_F1.5P0.05	650	M11794	Mt1a	24567	0.5328	0.0056
14	123gene_Welch_F1.5P0.05	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:
 Hs Mm Rn

Search within result
Clear

Message:
unique search ID number :94

	GENENAME	DESCRIPTION	SPECIES	CHROMLOCA.	LOCUSID	BIOLOGI
1	Nat1	N-acetyltransferase 1 (arylamine N-ac	Rattus norvegicus	16p14	116631	metabolism
2	Tap1	transporter 1, ATP-binding cassette, su	Rattus norvegicus	20p12	24811	defense response;immune resp
3	Mef2d	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 (ubiqu	Rattus norvegicus	1q54	293991	
28	LOC315994	similar to Expressed sequence A18746	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	203	195	197
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 Tgfb2 Nras Mapk9 Map3k1	MAPK signaling pathway(mmu04010)	Regulatory pathway	0.00019113
Met Ngfr Tgfb2 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
Tgfb2	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

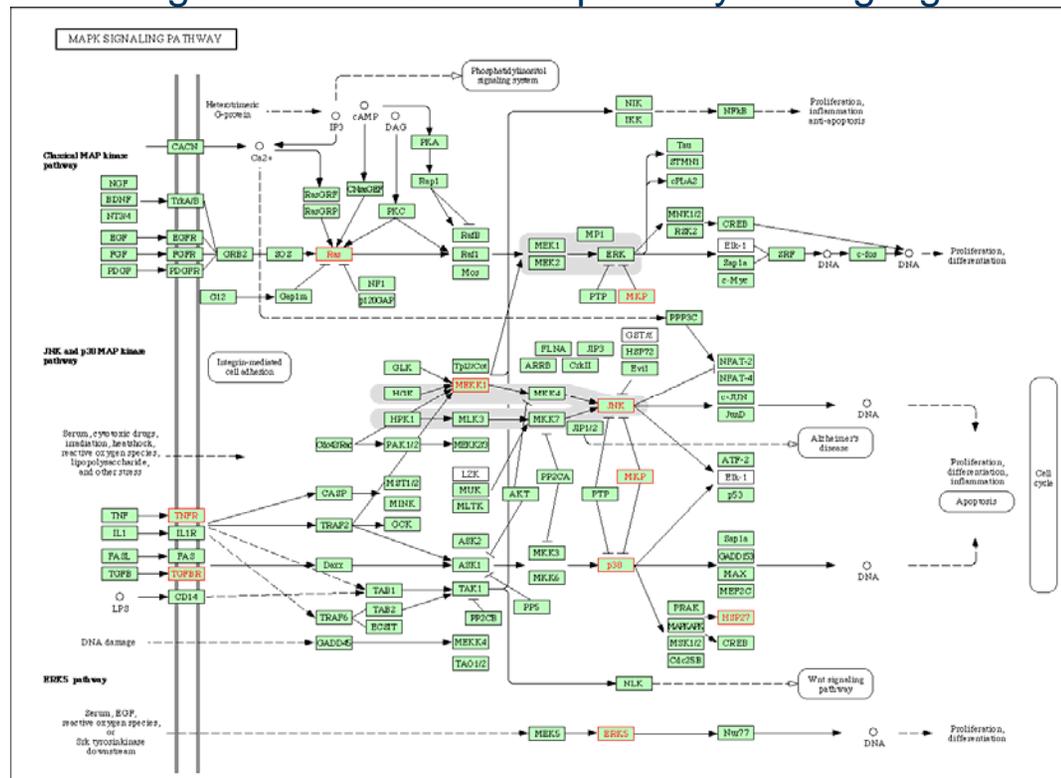
Input genes = 60, 22 genes found, 38 not found, Total 23 pathway maps.

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	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
Mef2d	T-cell apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.149045
Mapk14	Tgfb2	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	Bcr-Abl Signaling Pathway	Mus muscu...	Chronic M...	Disease	Signaling	0.217718
Vegf						

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.



GOFFA

The screenshot displays the GOFFA web application interface, which is used for analyzing Gene Ontology (GO) terms. The interface is divided into several main sections:

- Tree View (Left):** A hierarchical tree showing the structure of GO terms. The root is 'all(15/553 P=1.000000 E=1.00 R=0.00)'. Major categories include 'molecular_function', 'biological_process', and 'cellular_component'. Each category is further subdivided into more specific terms, with associated counts, P-values, E-values, and R-values.
- Table View (Top Right):** A table listing gene hits for selected GO terms. The columns are: No, Gene, Term, GO ID, Level (Aver...), P value... Δ, Gene Hits, and E value. The table shows hits for terms like 'nucleoside-triphosphatase activity' and 'pyrophosphatase activity'.
- GO path Graph (Bottom Left):** A line graph titled 'GO path' showing the relationship between GO terms. The x-axis is 'Level' (0-14) and the y-axis is 'log10(P) Value' (-2.9 to 0.0). Multiple colored lines represent different top 10 genes (Top1 to Top10). A specific path is highlighted with a P-value of 0.72407580455024393.
- GO tree prune (Bottom Right):** A large tree diagram titled 'GO tree prune' showing a pruned version of the GO hierarchy. The tree is rooted at the top and branches out into many smaller nodes. A legend on the left lists genes: Nras, Tap1, Abcb1, Abcc1, Mpg, Tap1, Abcb1, Abcc1, Tap1, Abcb1. Below the tree, there are controls for 'Prune Tree', 'Reset Tree', and 'Clear labels', along with parameters for P <= 0.05, Node size >= 5, and E Value >= 2.

Data Exploring

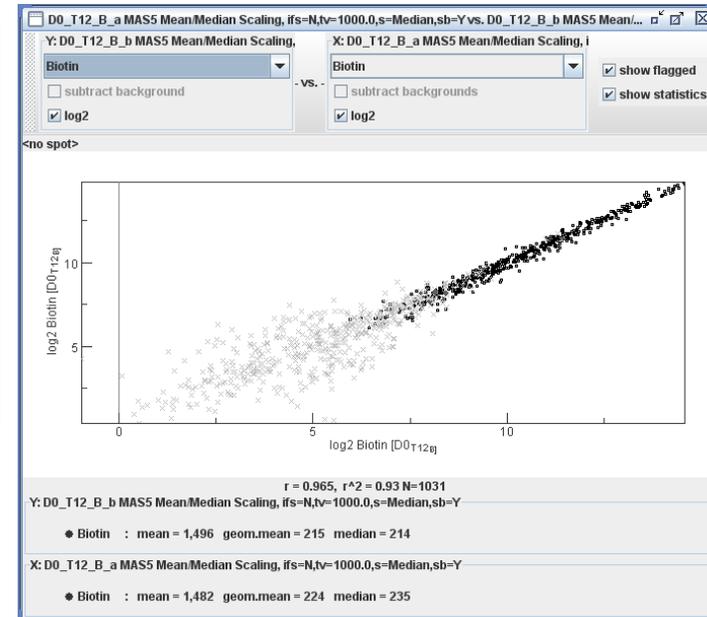
- **Scatter Plot**
plot the fluorescence intensity data of Cy3 vs Cy5 for the same array
- **Mixed Scatter Plot**
compare two arrays in one plot, applies to both 2-channel and 1-channel data.
- **Correlation Matrix**
Correlation Matrix shows the correlation between column i and column j of the original 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 maps multidimensional data into a low dimensional graph (usually two or three dimensions) to visually inspect the sample relationships.
- **P-value Plot**
visual interpretation of P-value distribution.
- **Hierarchical Clustering Analysis (HCA)**
A two-way HCA is applied to investigate the grouping of samples in terms of their similarities in gene expression profiles, as well as the grouping of genes in terms of their similarities of samples.



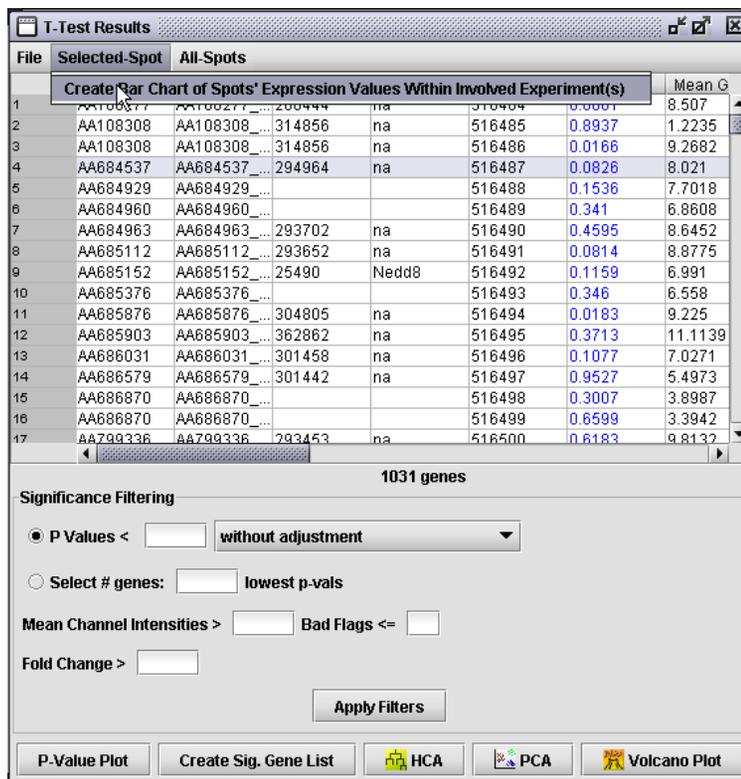
Data Exploring scatter plot

The screenshot shows the ArrayTrack interface with a menu open over the 'Correlations' section. The menu items are:

- Export
- Analysis
- Visualizations
- Correlations
 - Correlation matrix
 - Correlate study data vs. gene expression data
 - Mixed scatterplot
- Create gene list by data filtering
- Quality control
- Studies
- Tree options



Data Exploring Bar chart



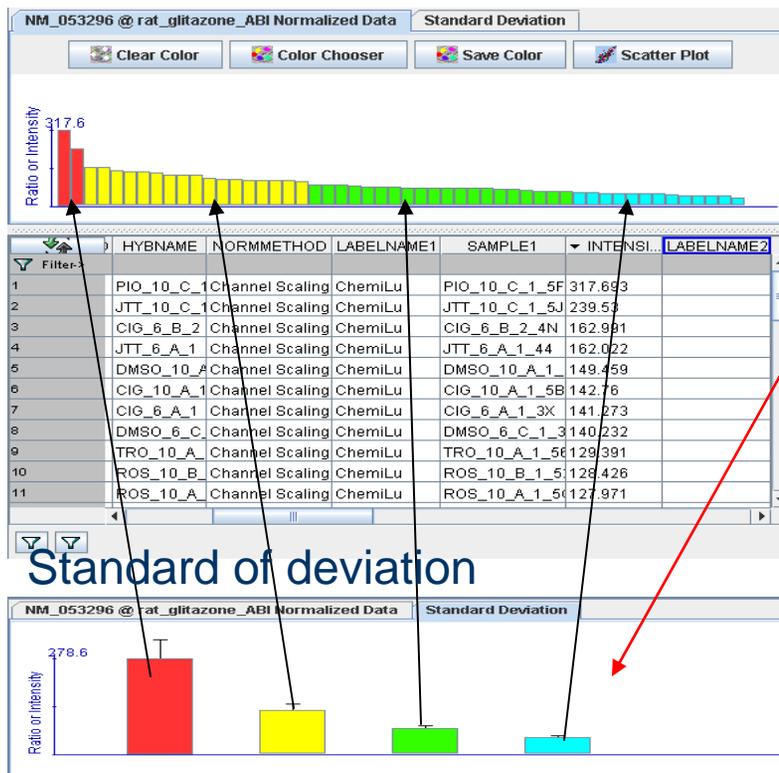
Access Bar chart from T-Test results.

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

Bar chart



Grouping multiple arrays across the same gene in different colors.

Group color could be saved.

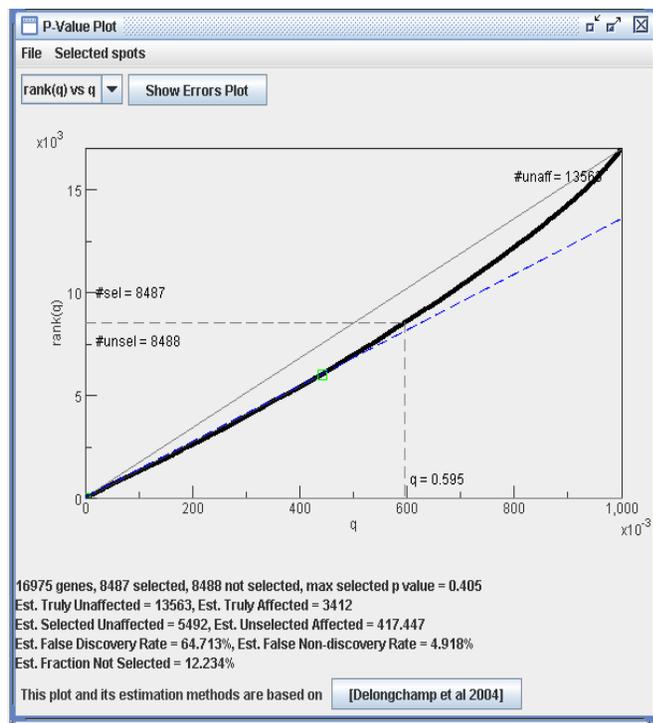
Standard deviation bar chart for the above groups. The bar height represents the mean intensity, while the T-line above the bar stands for the value of SD. The color for each bar echoes the color of the bar chart at the top, in the same order.



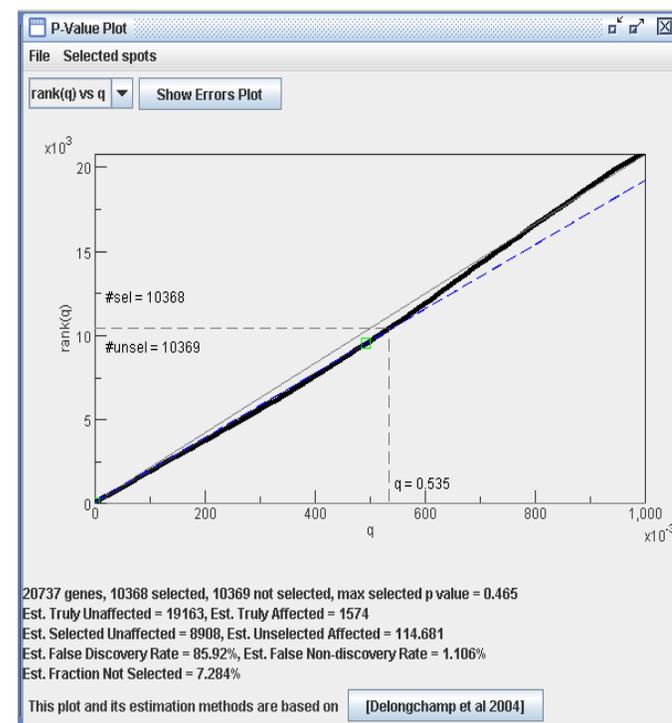
Data Exploring

P-value plot

Some effect

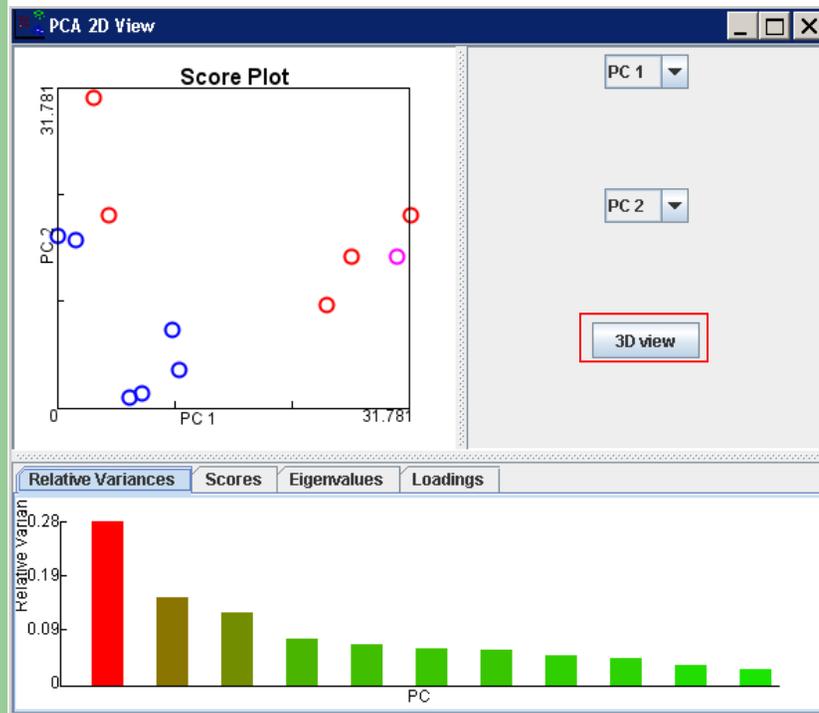


Slightly or no effect

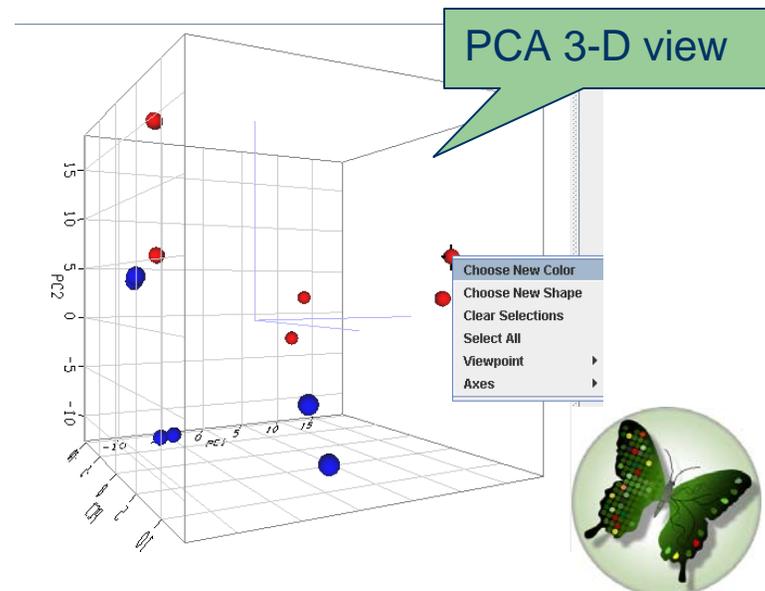


P plot curve is closer to diagonal line, the less treatment effect there is.

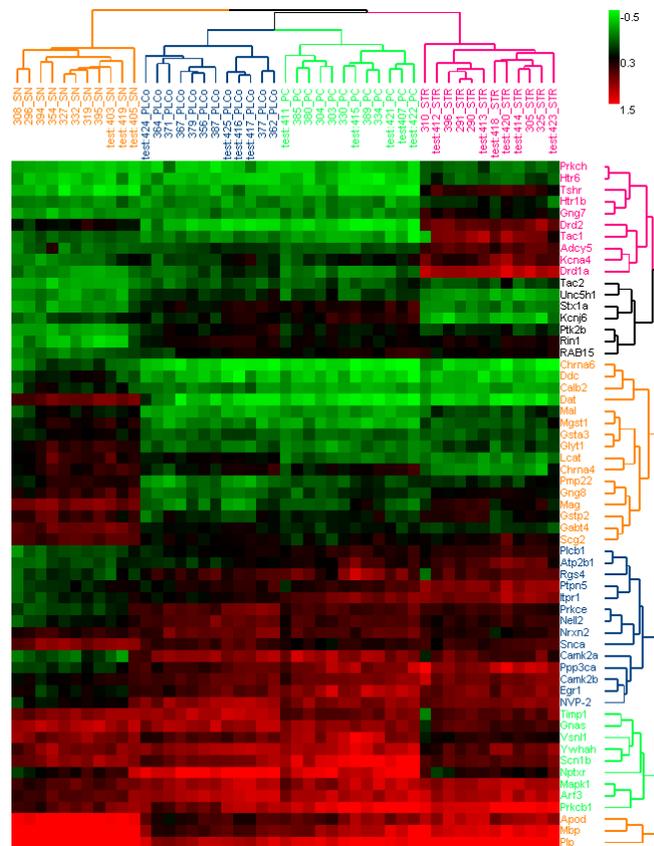
Data Exploring PCA



- 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



HCA is commonly used after gene selection for viewing the gene expression of profile.





Accessing ArrayTrack and documents

- 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>
- ArrayTrack documents (tutorials and manual)
<http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/tutorials.htm>



Important tips for using ArrayTrack

Many functions in ArrayTrack are accessible from multiple paths, for example, left-side window panels, pull-down menus and right mouse-click options.

- Right-click on a (set of) selected object(s) under the Database Contents tree to access the applicable TOOL functions.
- Multiple sets of arrays can be selected by a combinations 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. Each column can be sorted by clicking the column title.
- Tutorials for various functions.

<http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/tutorials.htm>





Technical Support

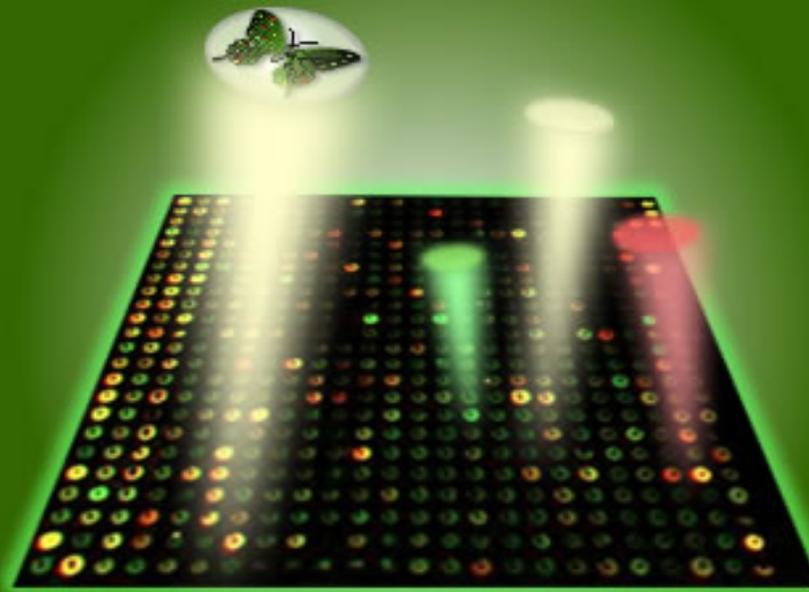
NCTRBioinformaticsSupport@nctr.fda.gov

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