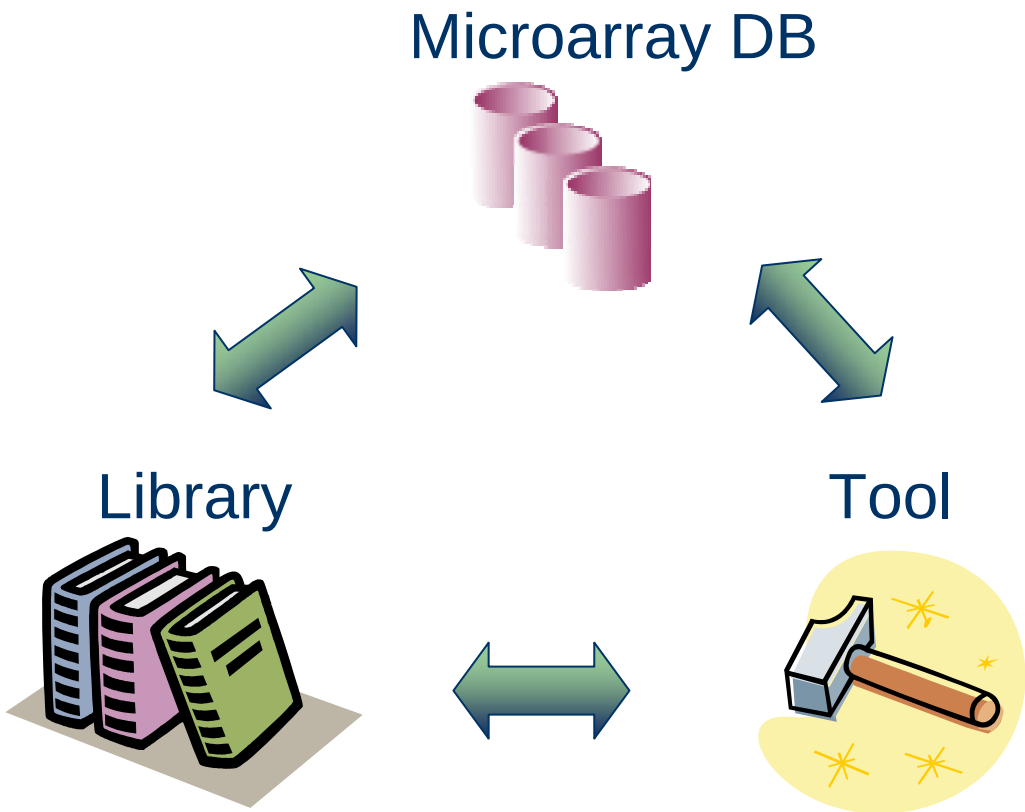
ArrayTrack$_{3.4.0}$ Demonstration

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

ArrayTrack Overview



An integrated environment for microarray data management, analysis and interpretation.

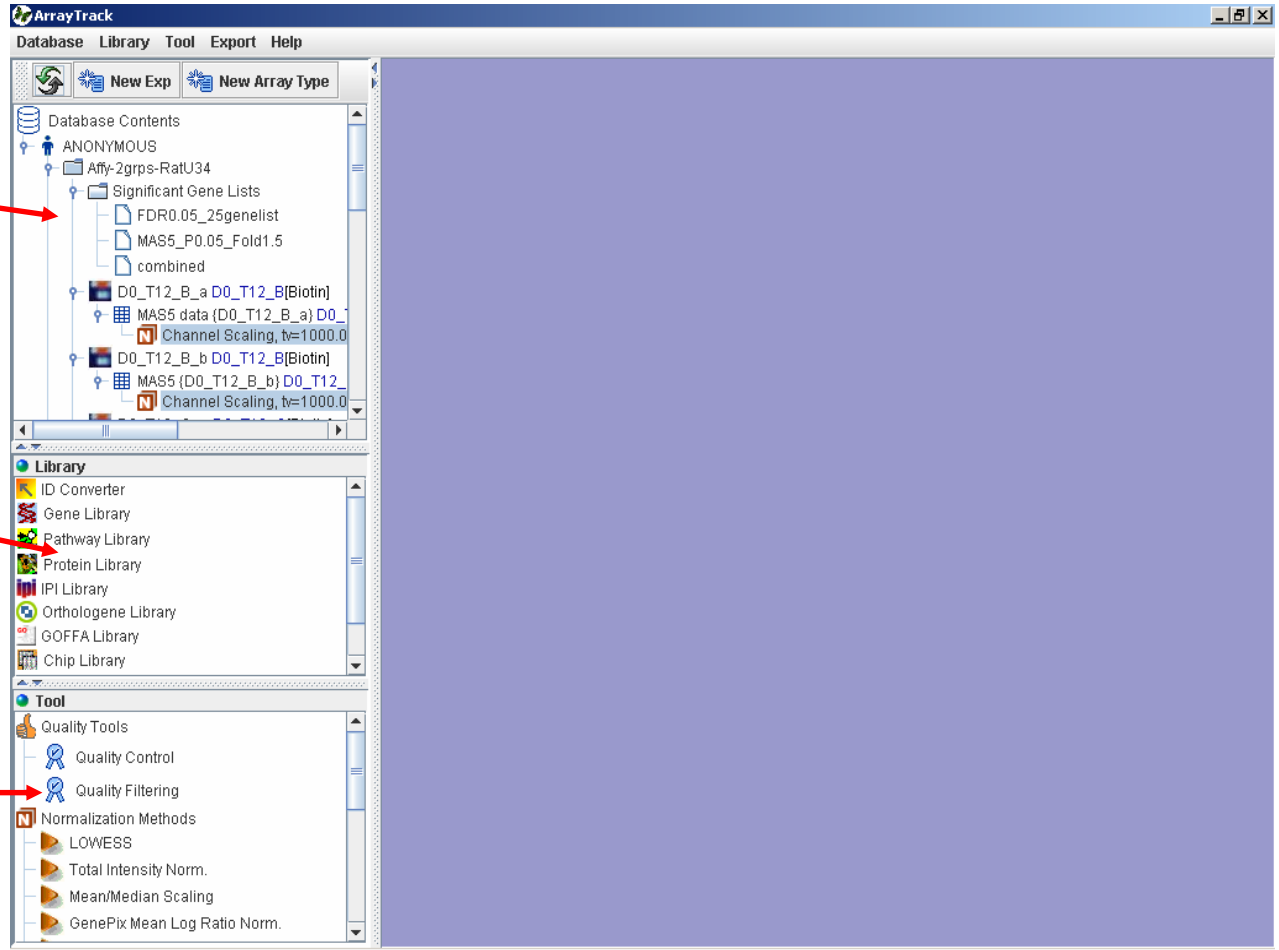
Using ArrayTrack, the user can apply analysis method from Tool to microarray DB and then get information from the linked LIB for biological interpretation

ArrayTrack Overview

Microarray DB

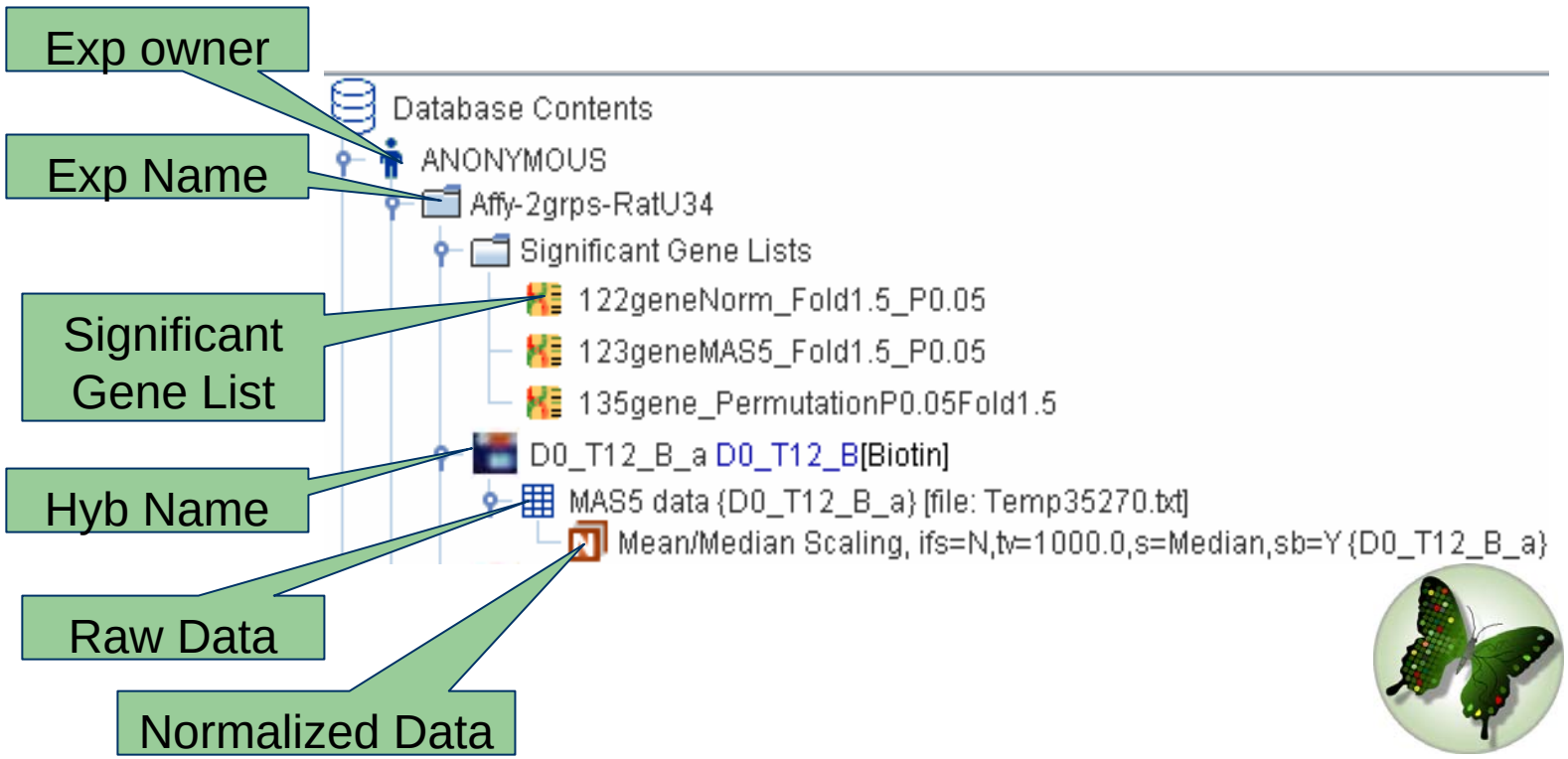
Libraries

Tools



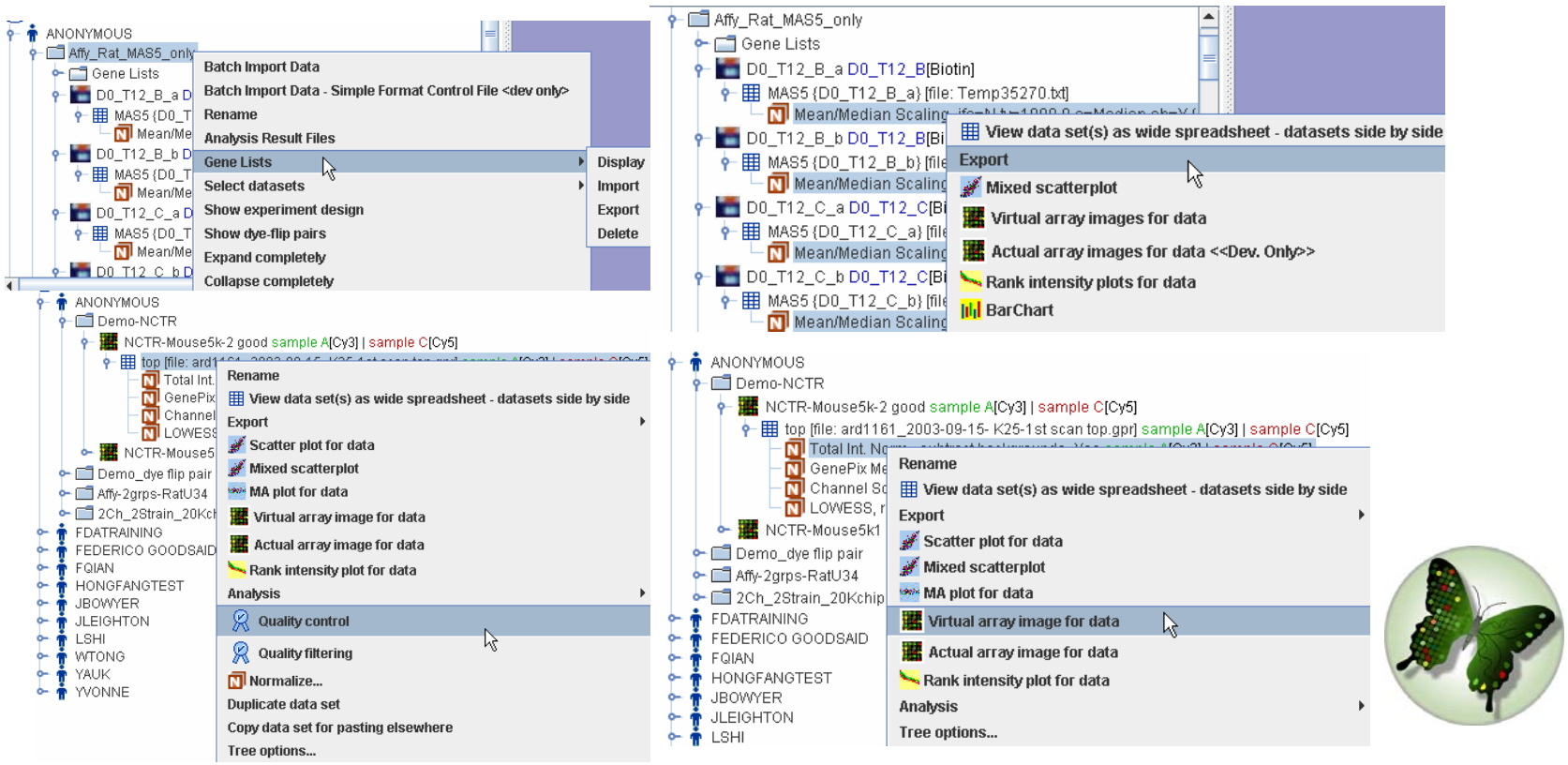
Microarray DB

Data is organized as a hierarchical tree structure:



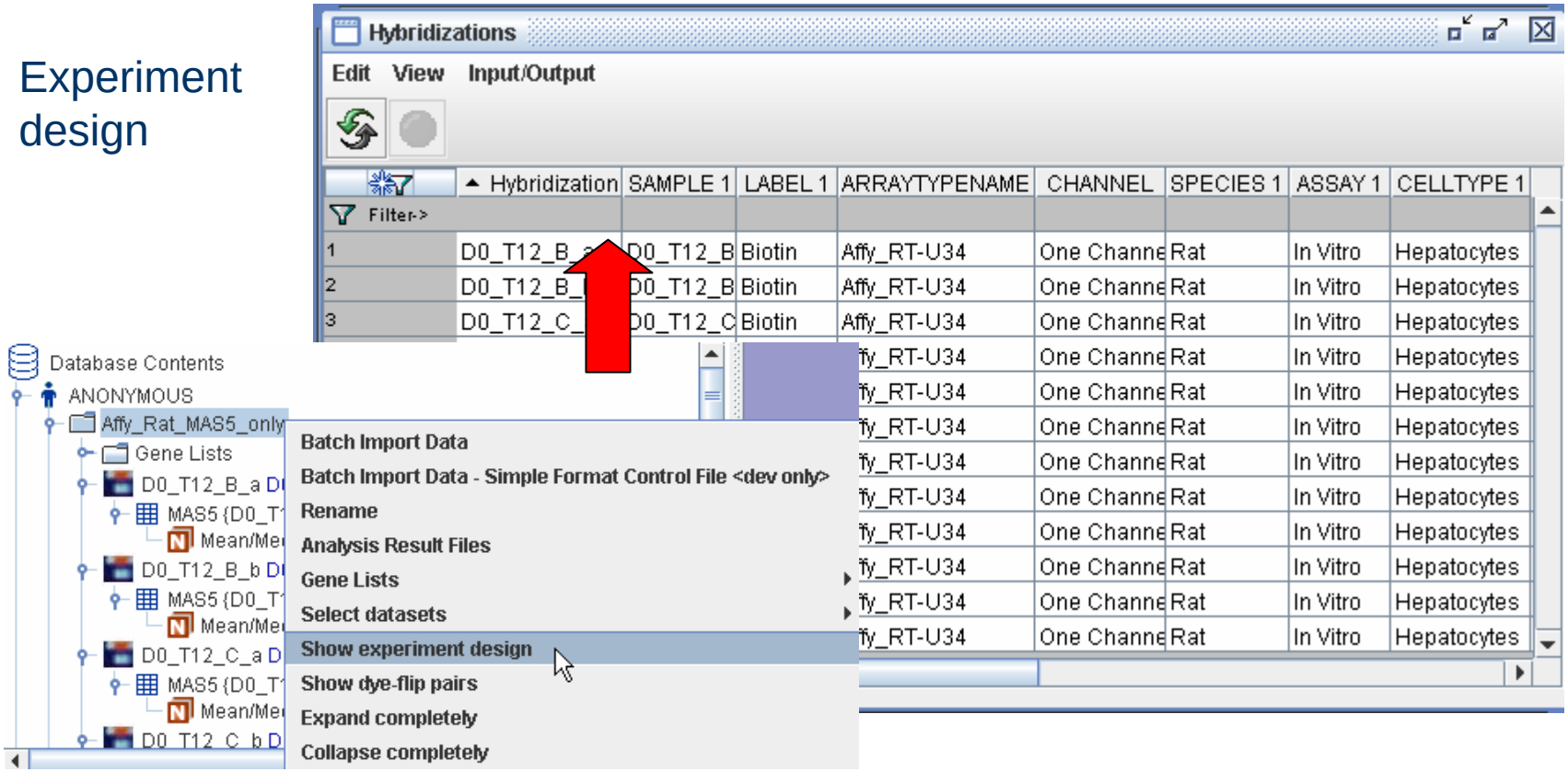
Exploring DB

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



Exploring DB

Experiment design



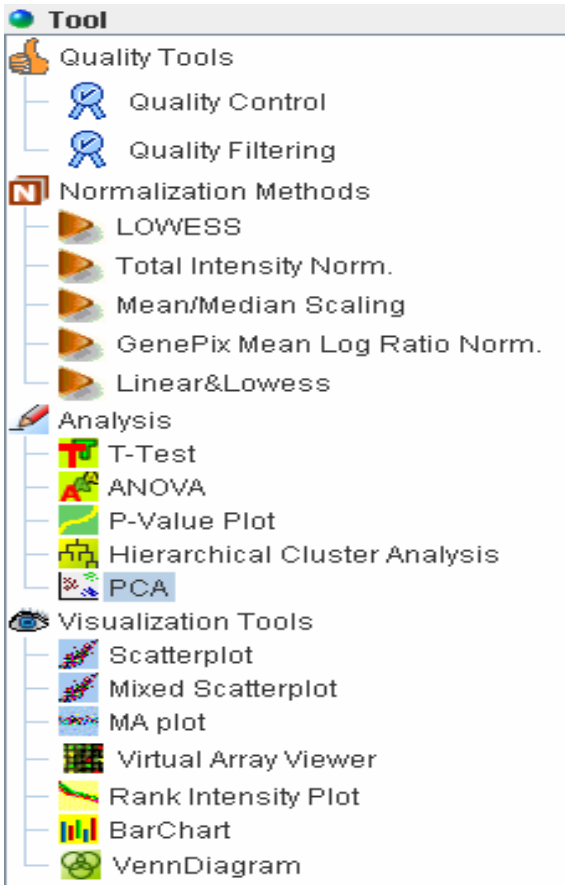
Searching Libraries

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

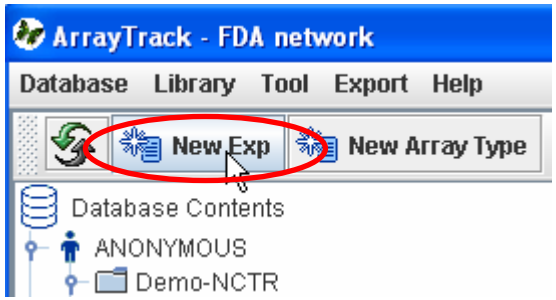
- Visualization Tools

Provides a direct view to identify abnormalities within data.

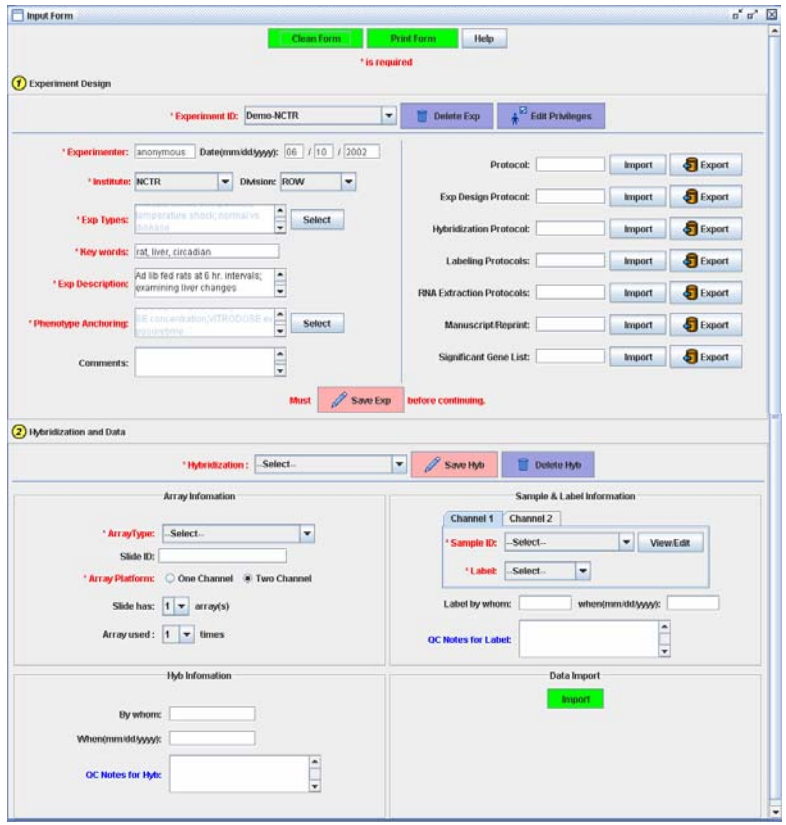
Microarray Data Processing

- 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



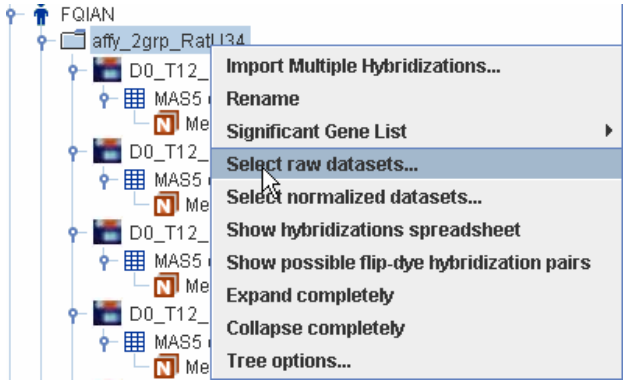
1. Create experiment
2. Create array type (if not existed in AT Chip LIB)
3. Use batch import wizard (see detail in Tutorial 9)

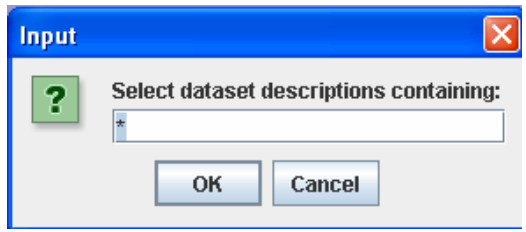


Normalization

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

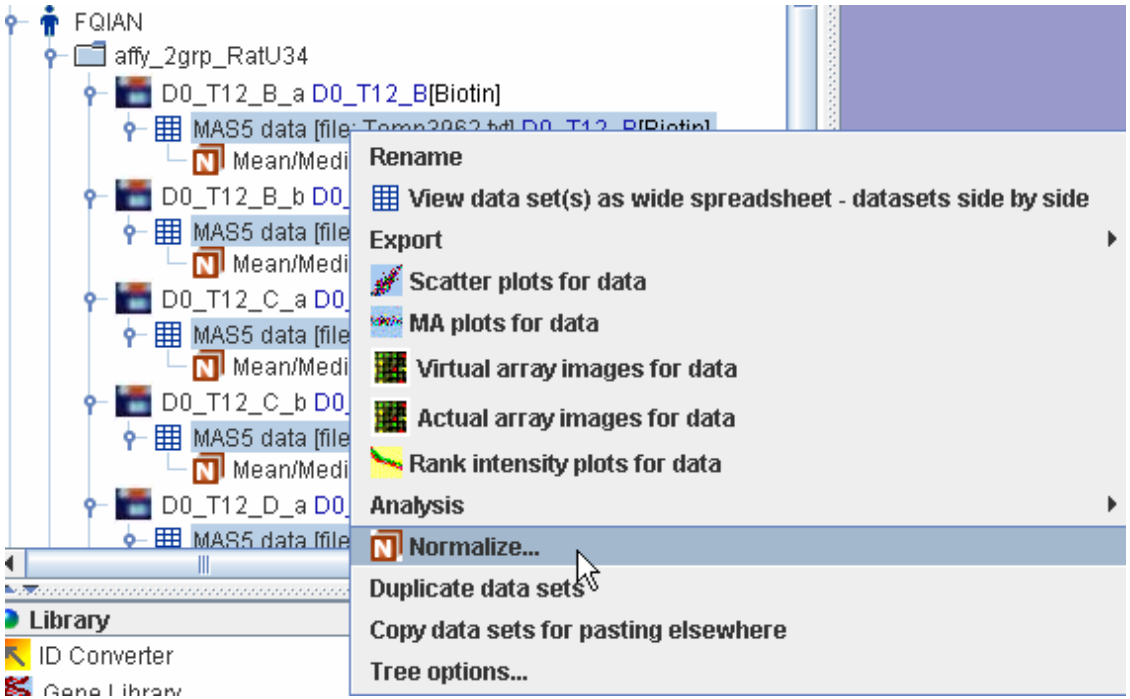
1. Right-click an experiment
2. Select “raw datasets…”





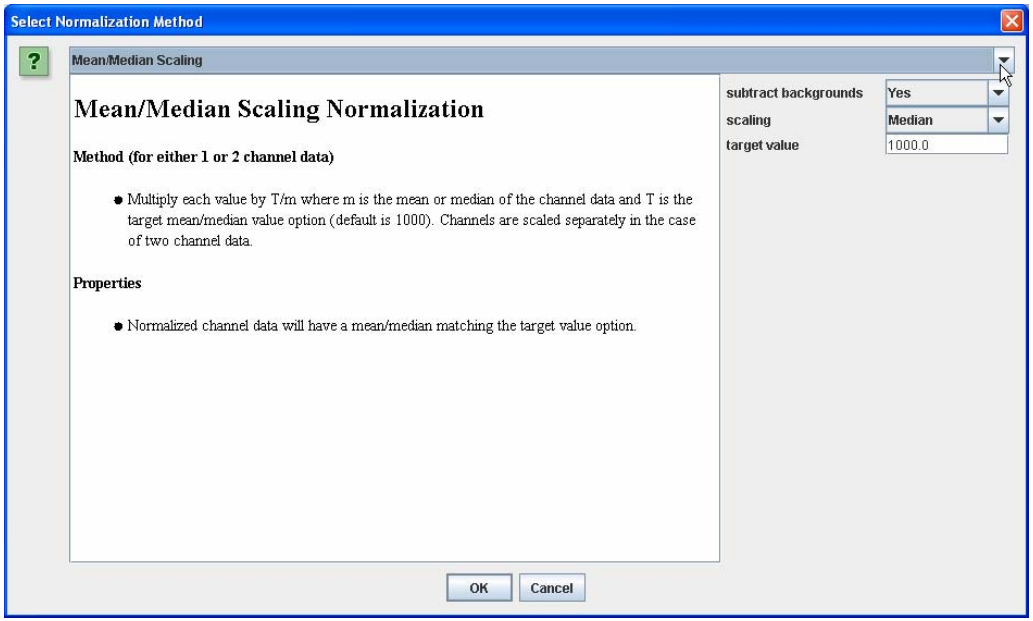
Normalization (-continued)

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: **Error rate for the exp**

Single testing: → 1 gene P<0.05 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***

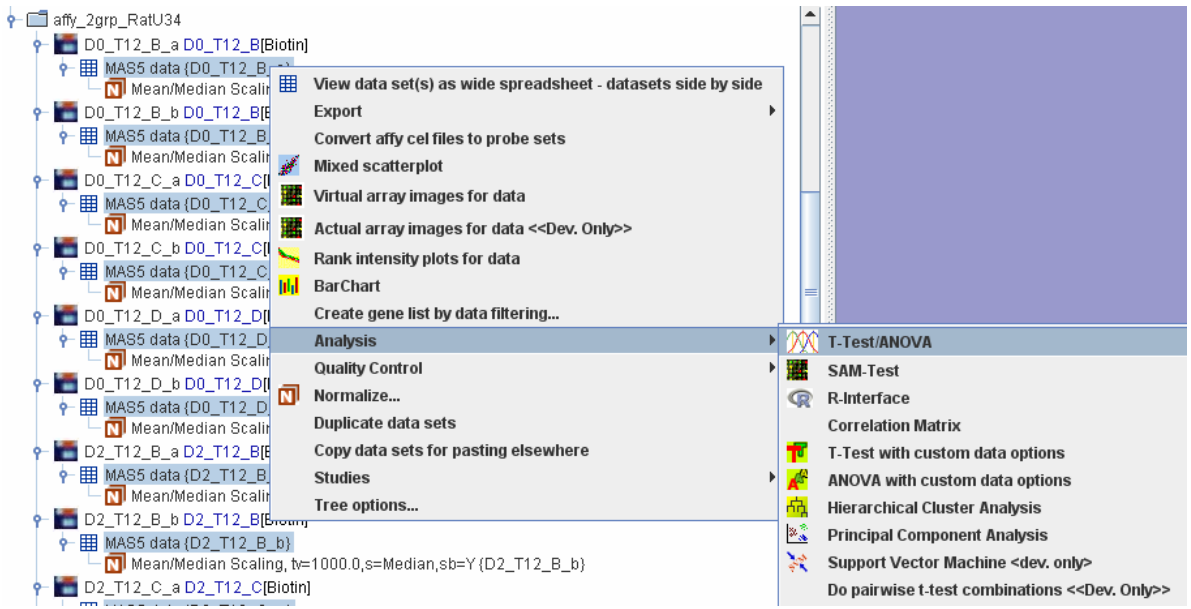
False discovery rate (e.g., Benjamini & Hochberg, p-value plot)

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

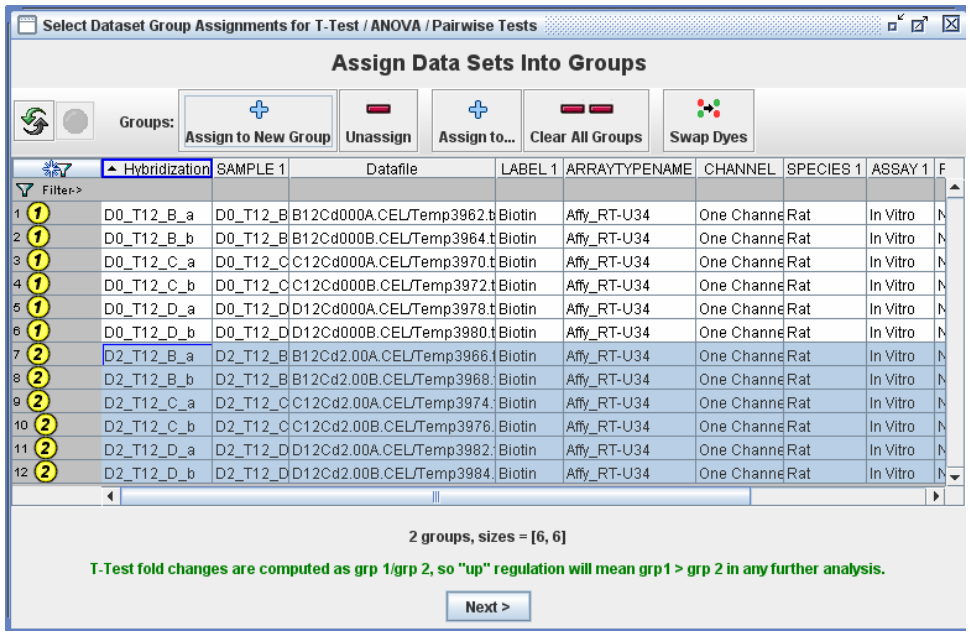
Volcano plot (combination of p and fold change)

Gene Selection

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



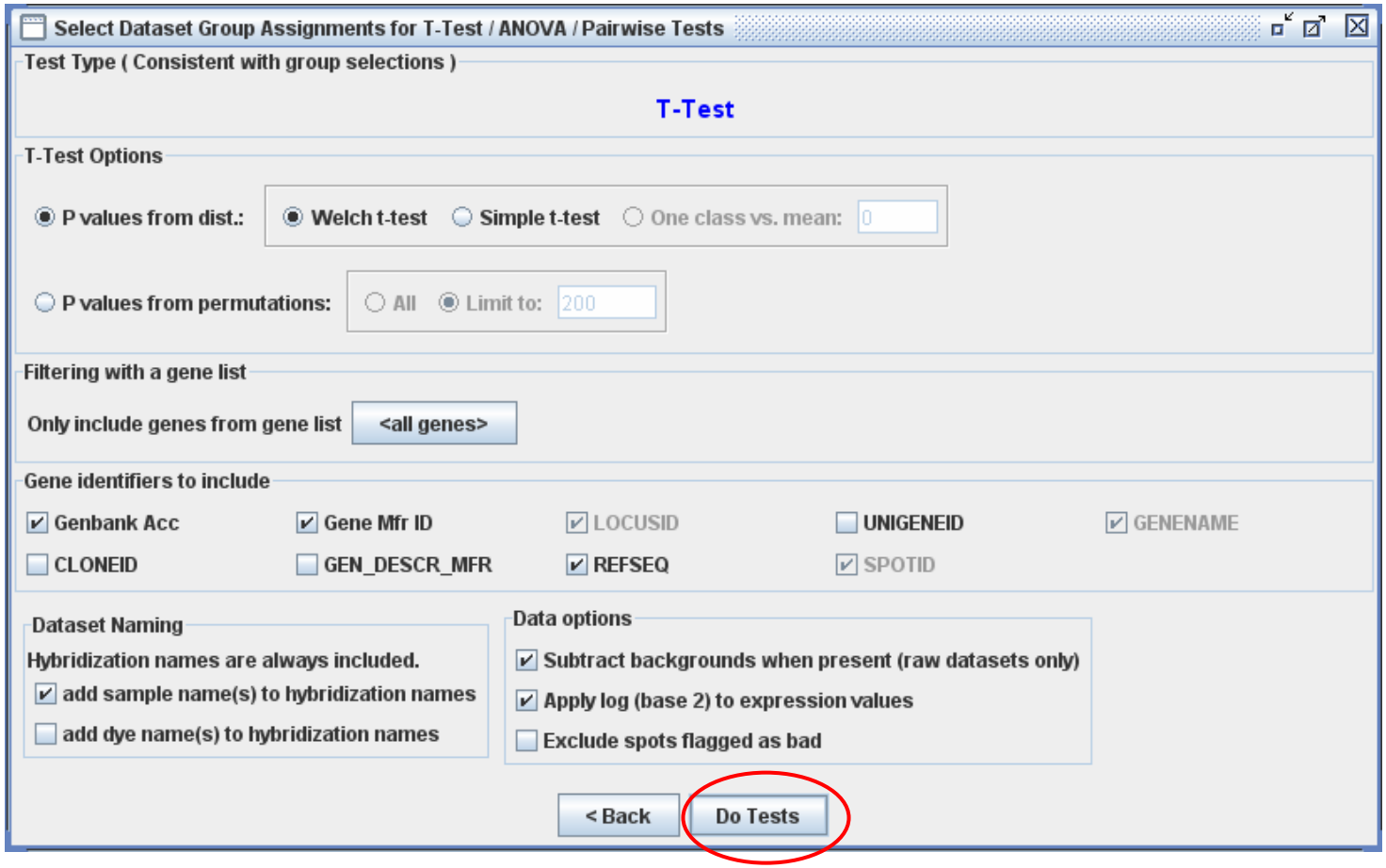
Gene Selection (-continued)



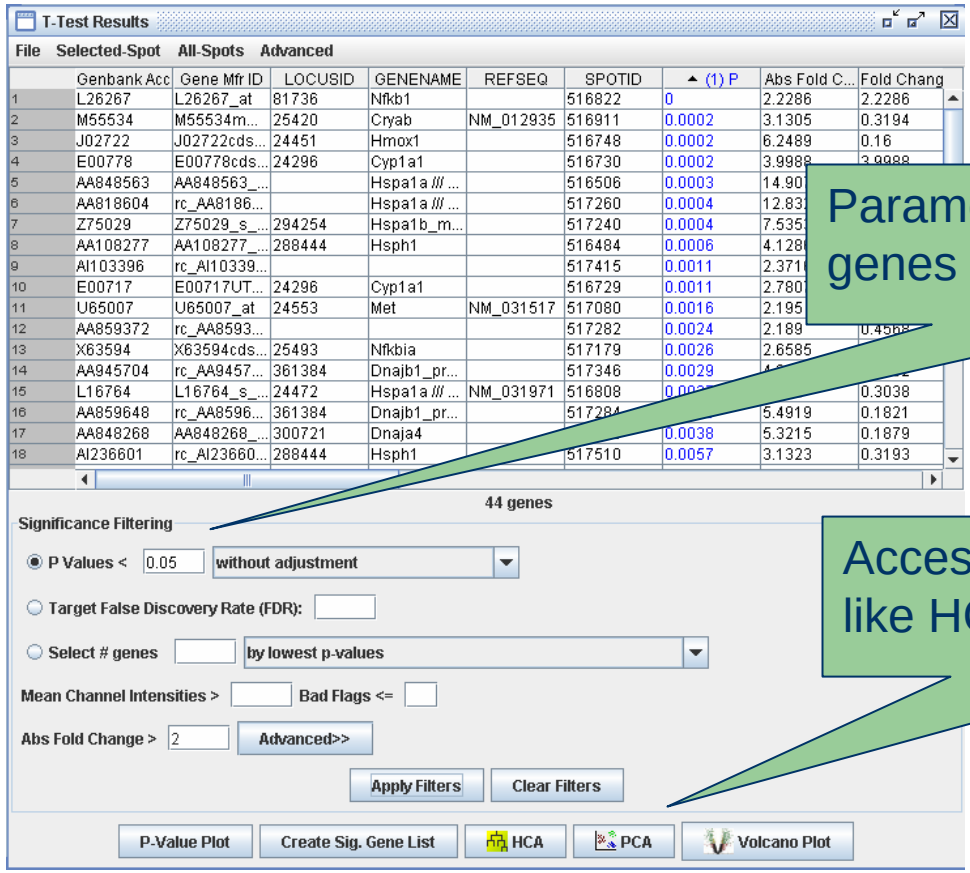
Assign the data into 2 groups
-different dose
-different time
-or different animal

Note: always put control in group 2

Gene Selection (-continued)



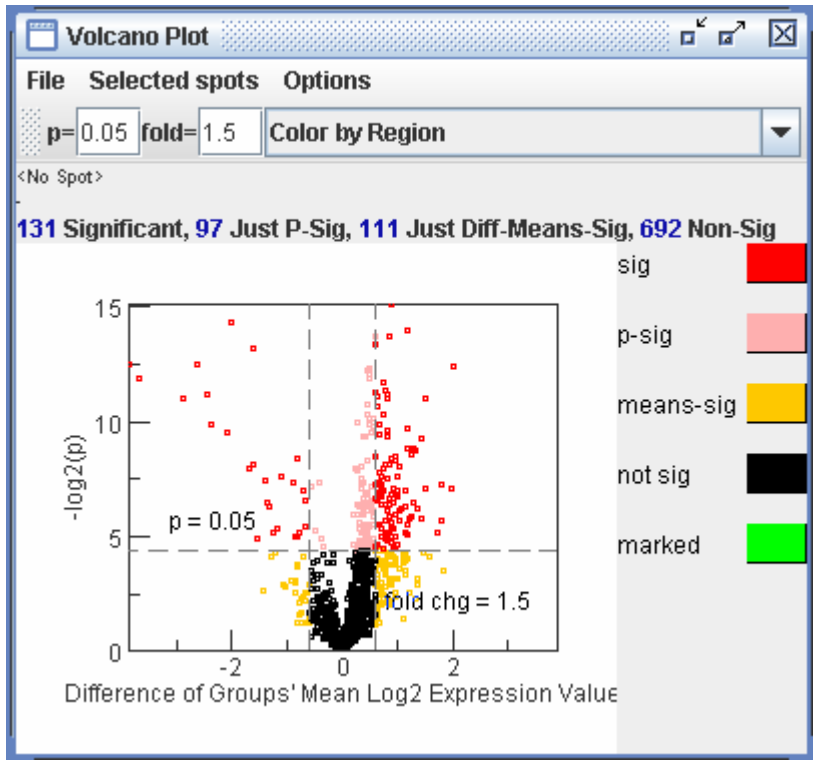
Gene Selection (-continued)



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)

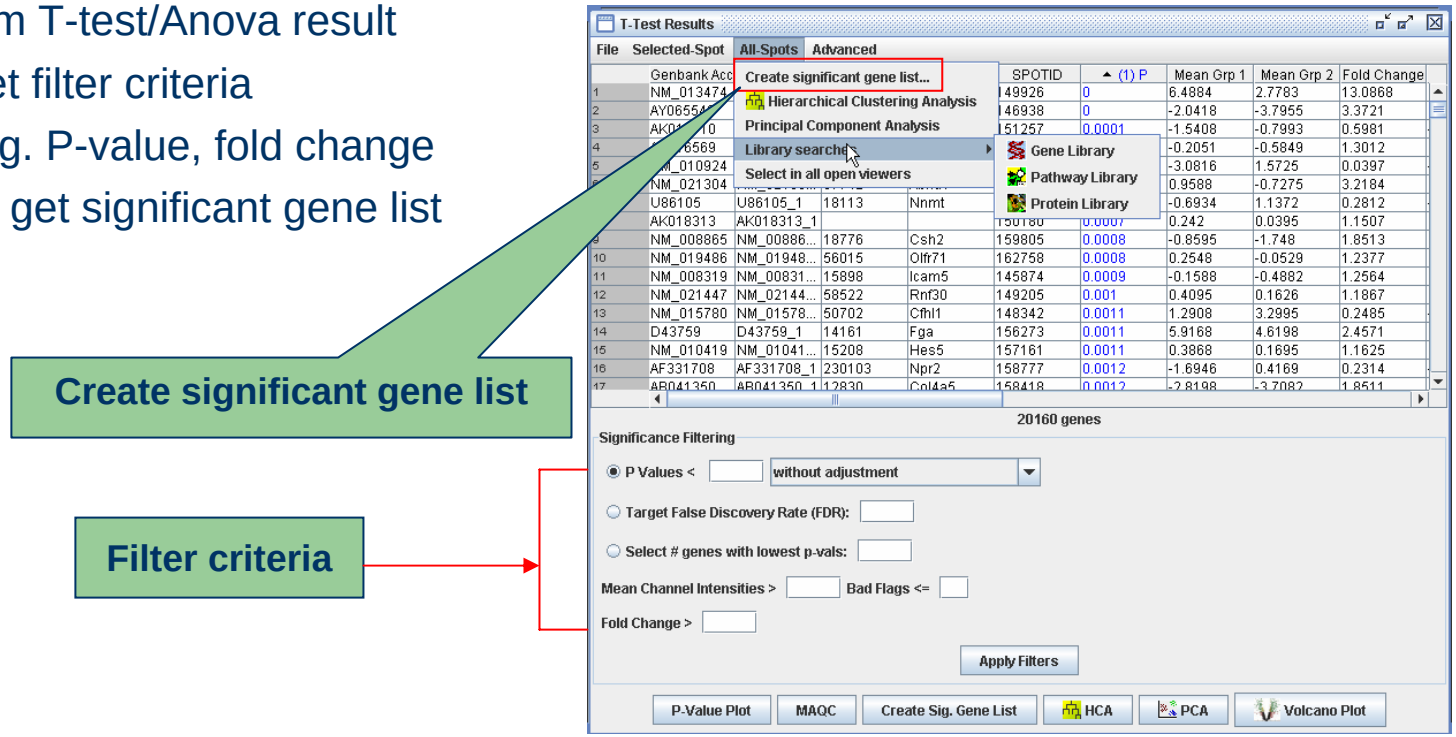
- Create significant gene list

  From T-test/Anova result

  set filter criteria

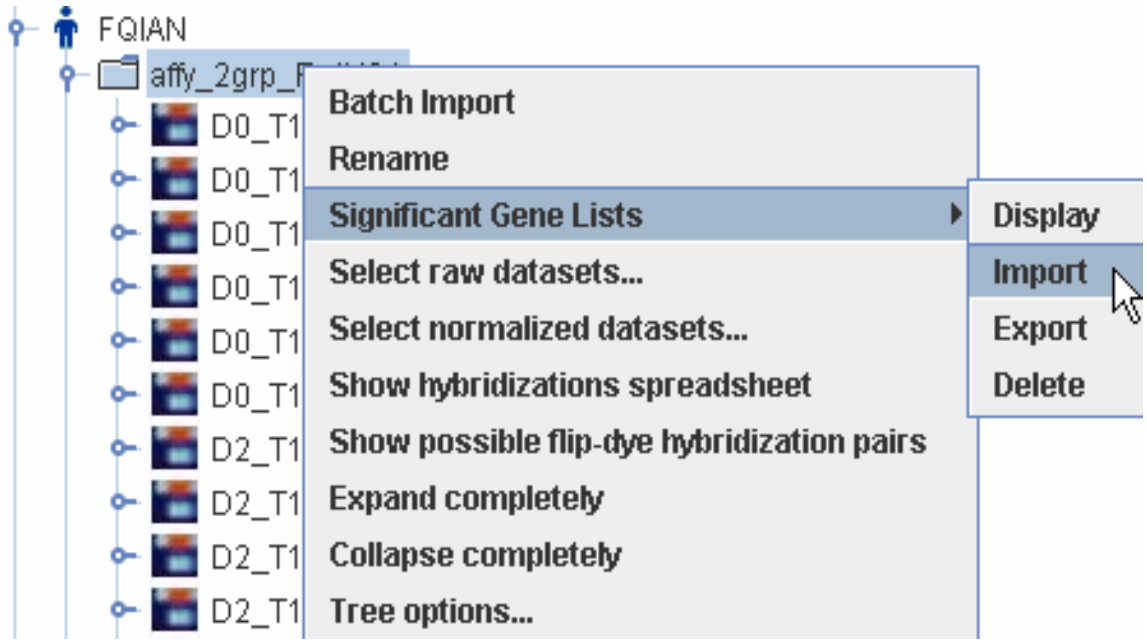
  e.g. P-value, fold change

  to get significant gene list



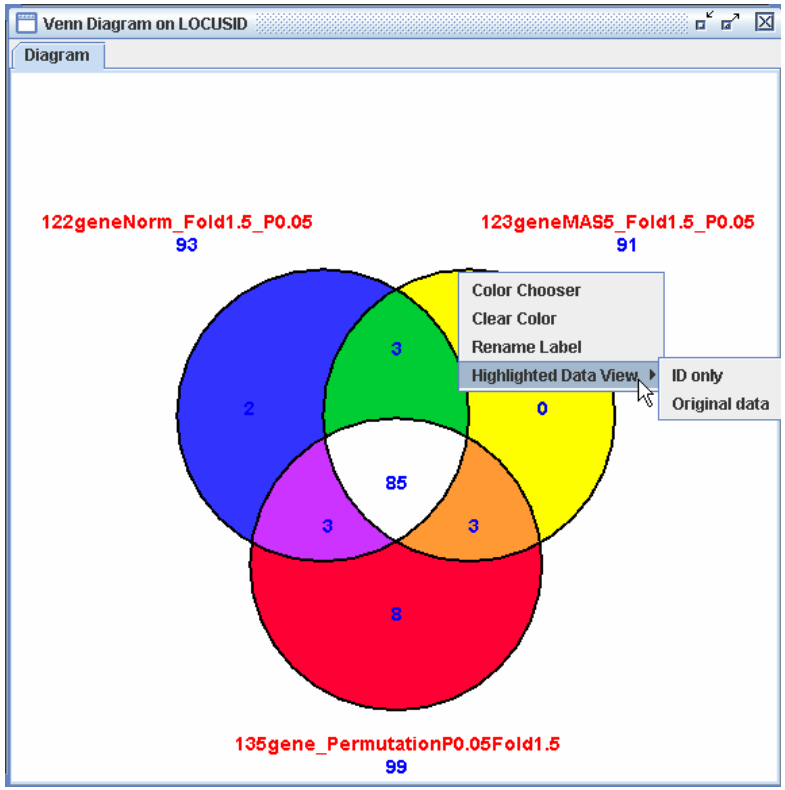
Significant Gene List(-cont)

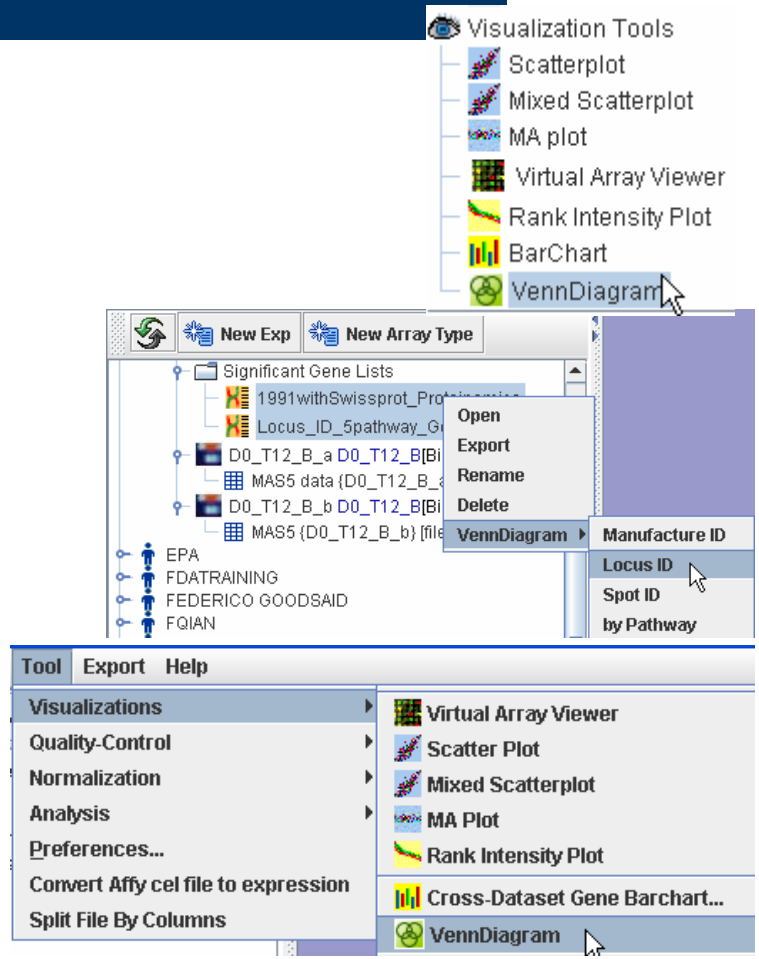
- **Display/Import/Export/Delete Significant Gene List**
  Right-click the experiment name
  select Significant Gene Lists
    select Display
      Import
      Export
      Delete



Significant Gene List(-cont)

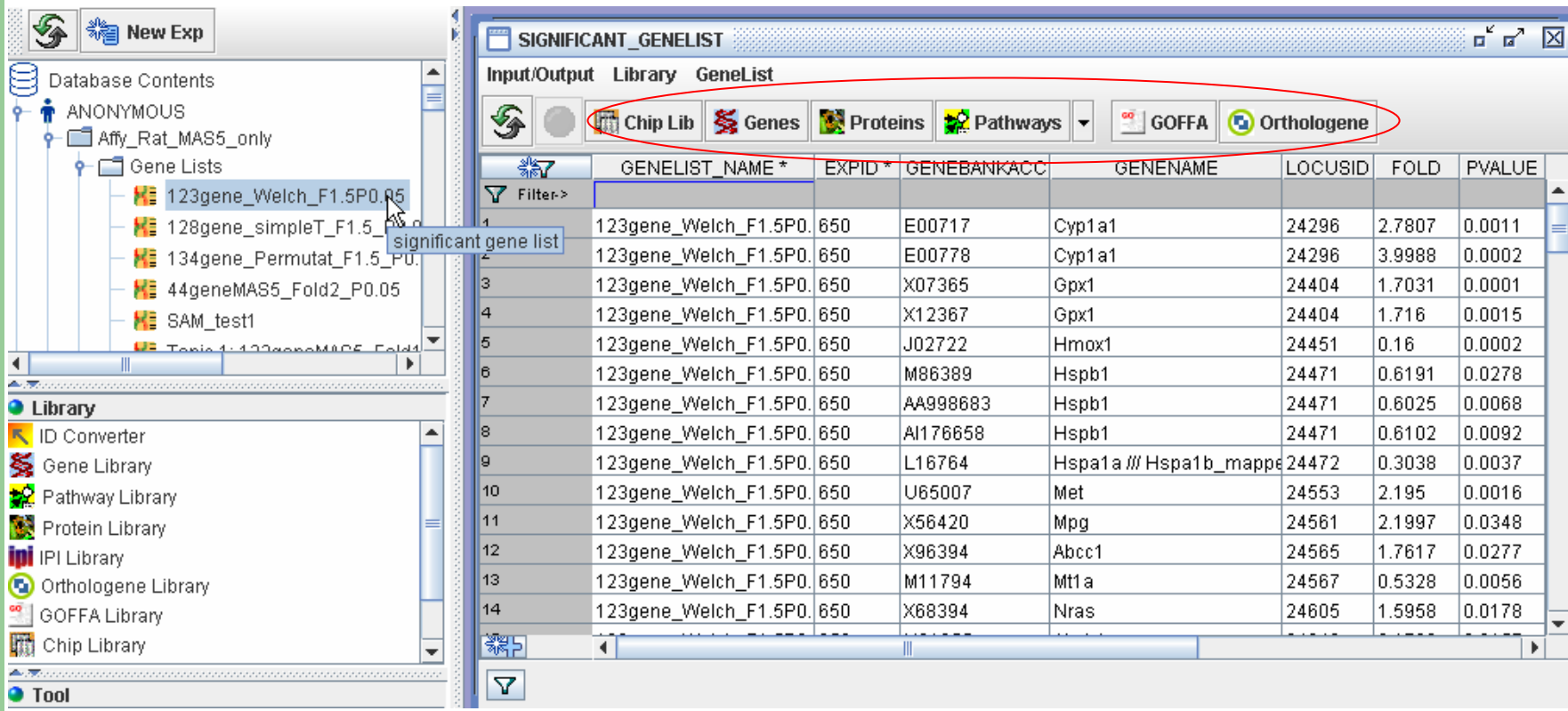
- **VennDiagram -** get common genes from 2~3 significant gene lists



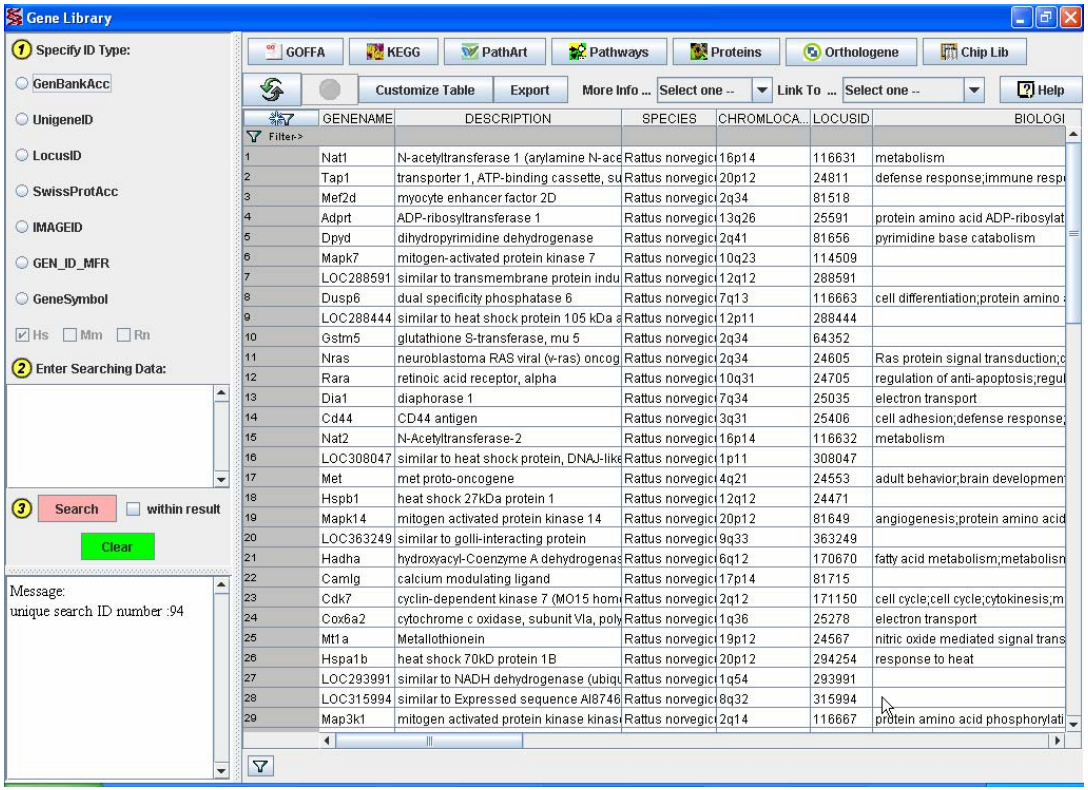


Interpretation

Link the significant genes to Gene Library for data interpretation



Interpretation (continued)



- 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

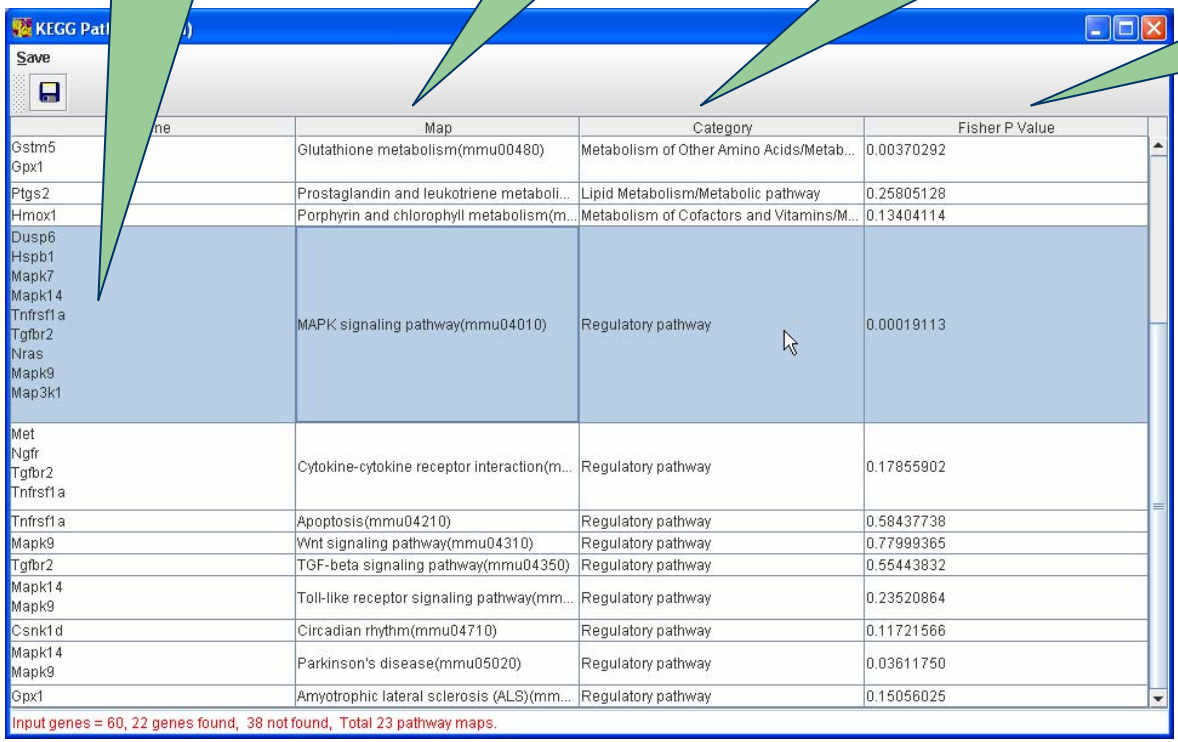
Interpretation (continued) Kegg

Genes involved in a pathway

Pathway name

Pathway category

Statistical significance of the pathway



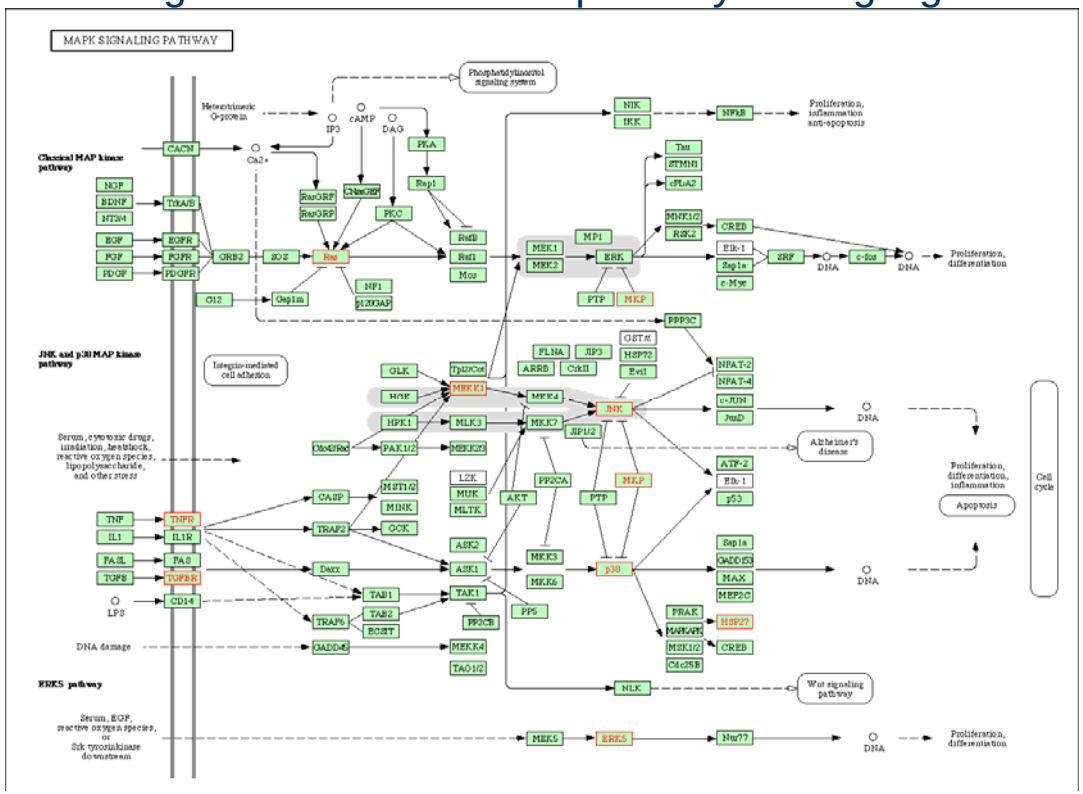
Gene	Map	Category	Fisher P Value
Gstm5 Gpx1	Glutathione metabolism(mmu00480)	Metabolism of Other Amino Acids/Metab...	0.00370292
Ptgs2	Prostaglandin and leukotriene metaboli...	Lipid Metabolism/Metabolic pathway	0.25805128
Hmox1	Porphyrin and chlorophyll metabolism(m...	Metabolism of Cofactors and Vitamins/M...	0.13404114
Dusp6 Hspb1 Mapk7 Mapk14 Tnfrsf1a Tgfbr2 Nras Mapk9 Map3k1	MAPK signaling pathway(mmu04010)	Regulatory pathway	0.00019113
Met Ngfr Tgfbr2 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
Tgfbr2	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

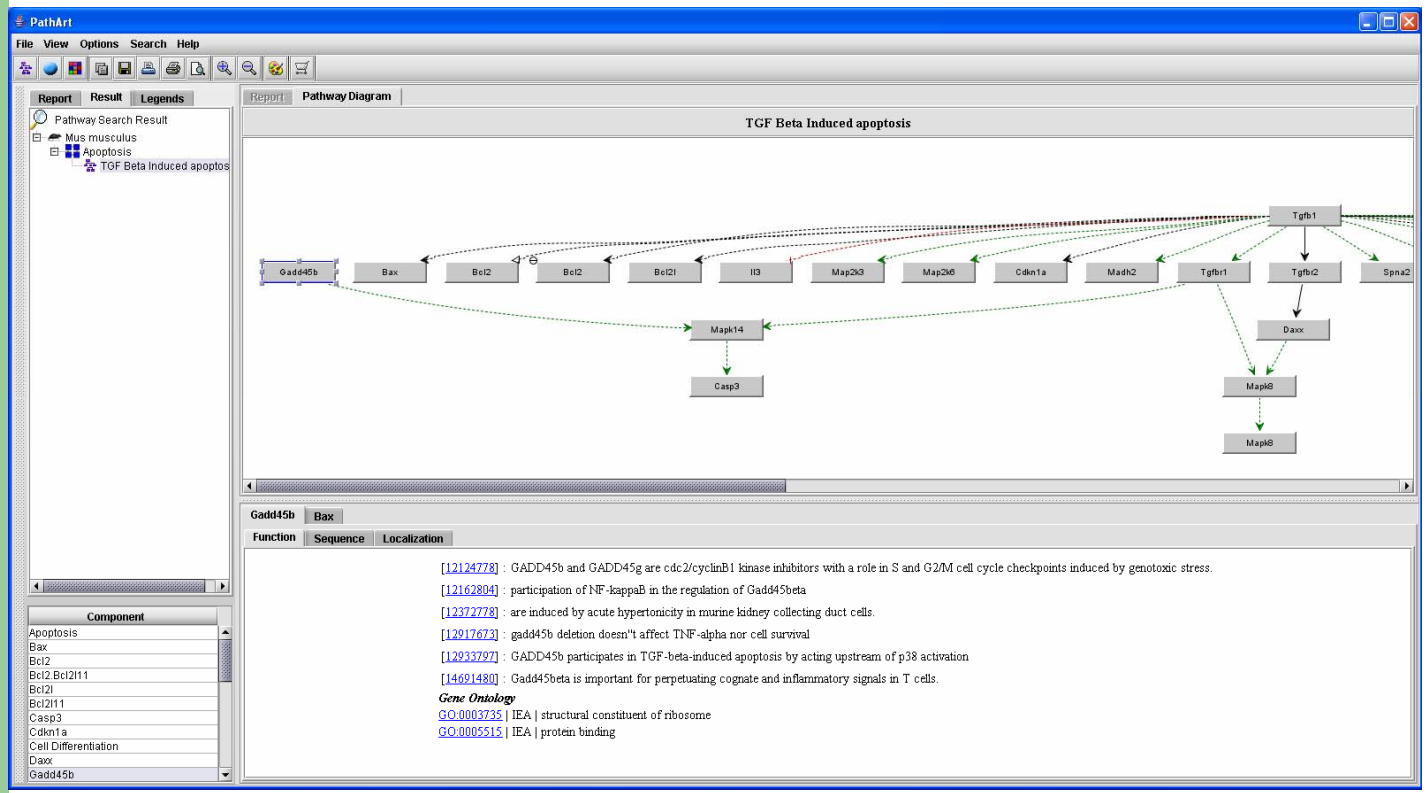
Pa t for Mus muscu

Save dit

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
Mef2d	T-cell apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.149045
Mapk14 Tgfbr2	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
Tgfbr2	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) PathArt



Interpretation (continued) GOFFA

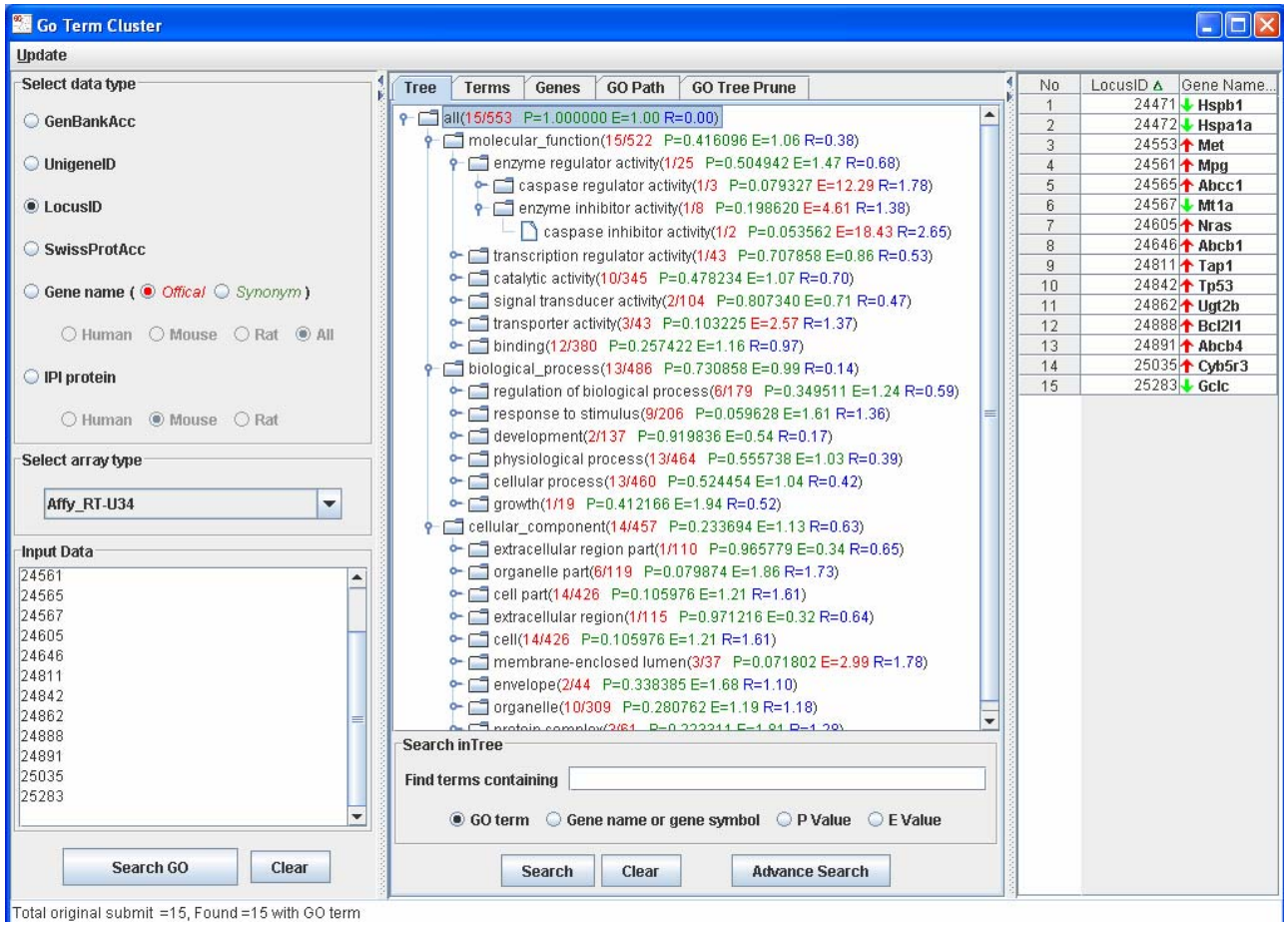
- GOFFA – Gene Ontology For Functional Analysis
- Developed based on Gene Ontology(GO) database
- Grouping the genes into functional classes
- GO- three ontologies

  -Molecular function: activities performed by individual gene products at the molecular level, such as catalytic activity, transporter activity, binding.

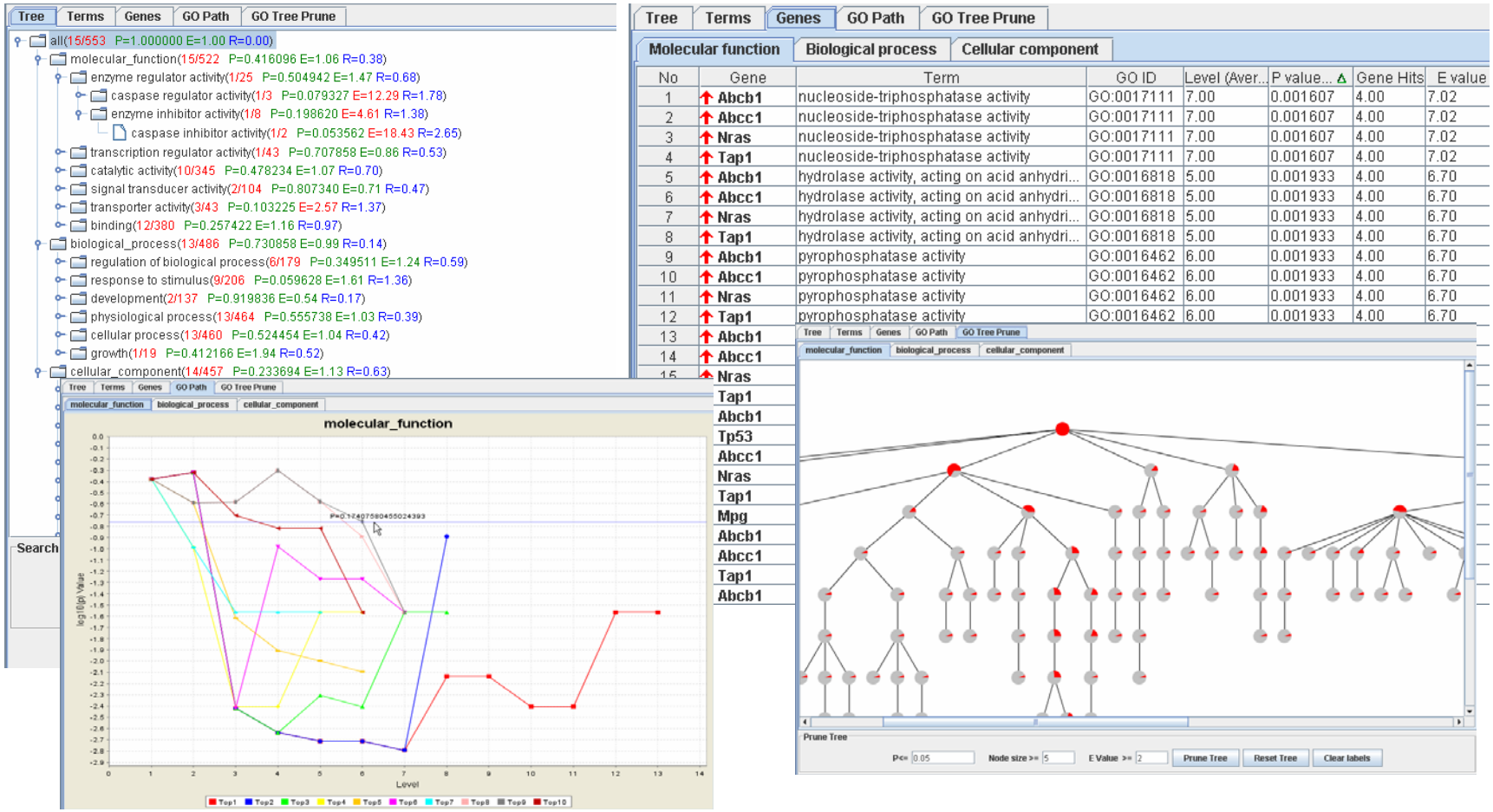
  -Biological process: broad biological goals accomplished by ordered assemblies of molecular functions, such as cell growth, signal transduction, metabolism.

  -Cellular component: the place in the cell where a gene product is found, such as nucleus, ribosome, proteasome.

Interpretation (continued) GOFFA



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)
Search
molecular_function
biological_process
cellular_component
molecular_function
log10(p) Value
Level
Top1 Top2 Top3 Top4 Top5 Top6 Top7 Top8 Top9 Top10
Molecular function
Biological process
Cellular component
No
Gene
Term
GO ID
Level (Aver...
P value...
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
Nras
Tap1
Abcb1
Tp53
Abcc1
Nras
Tap1
Mpg
Abcb1
Abcc1
Tap1
Abcb1
Prune Tree
P<= 0.05
Node size >= 5
E Value >= 2
Prune Tree
Reset Tree
Clear labels

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

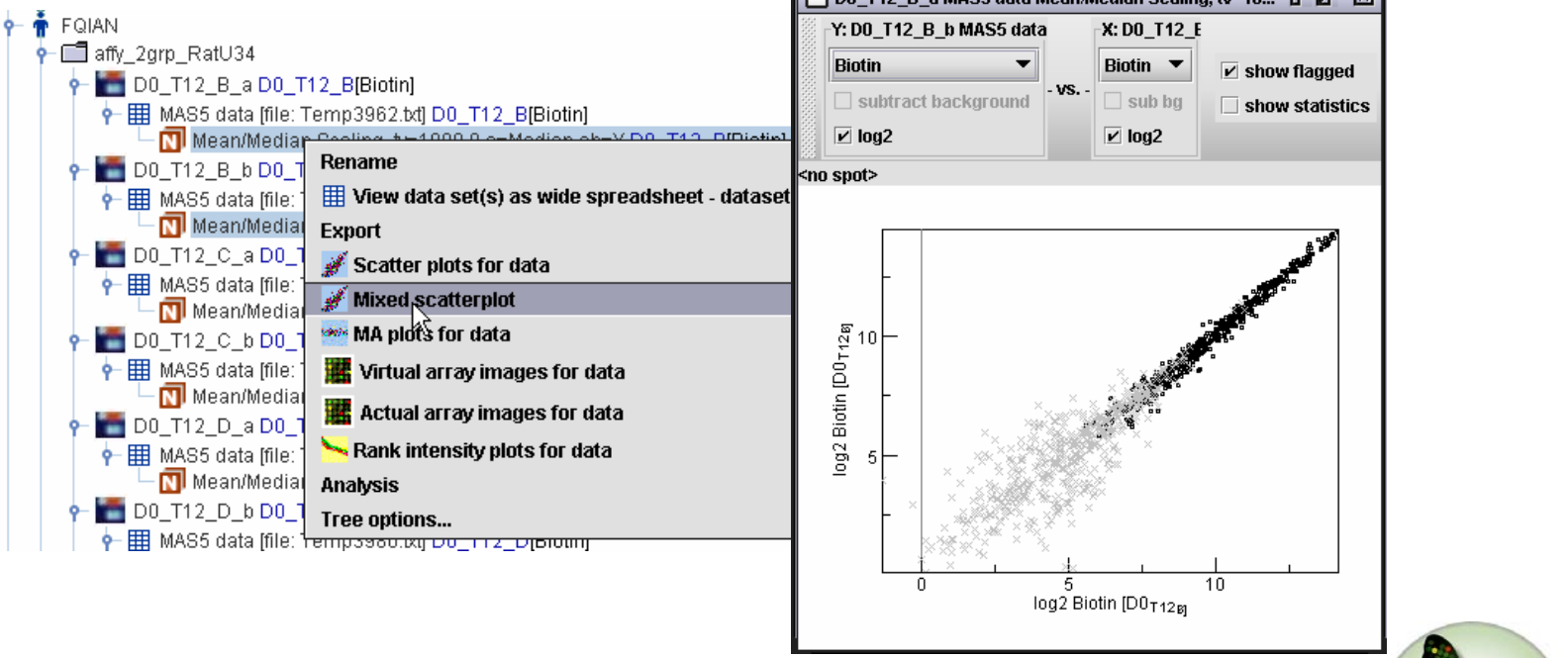
  displays expression data for a single gene across multiple arrays within the same experiment or across different experiment.
- Principal Component Analysis (PCA)

  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

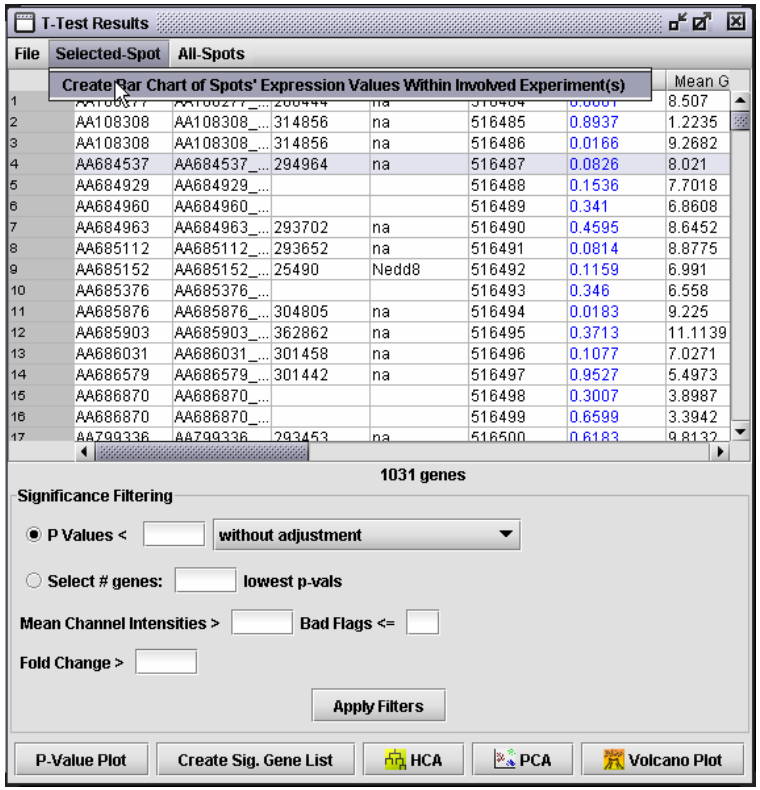
  is comprised of agglomerative methods and divisive methods that finds clusters of observati
  data set.

Data Exploring scatter plot



Data Exploring

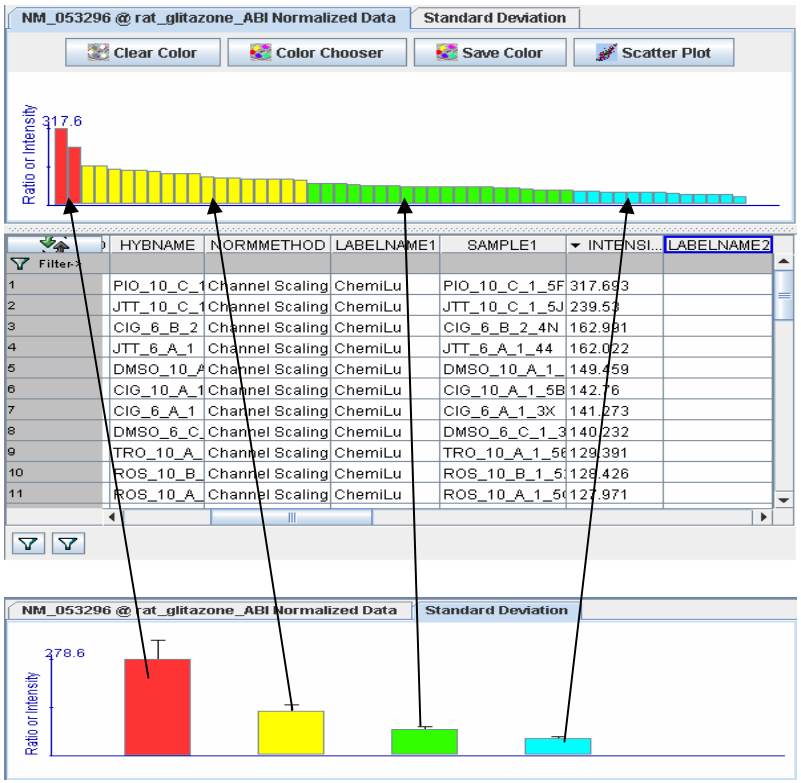
Bar chart



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



Grouping multiple arrays marked in different colors.

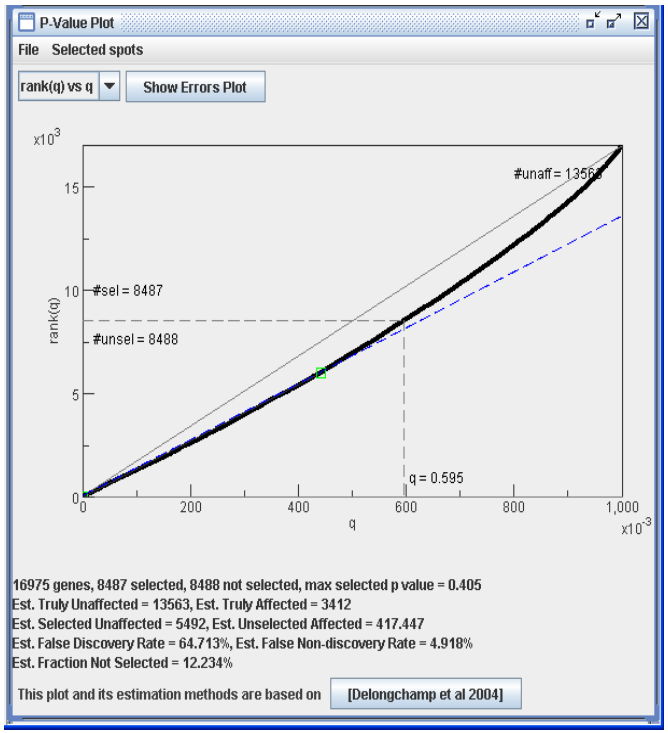
Group color could be saved.

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.

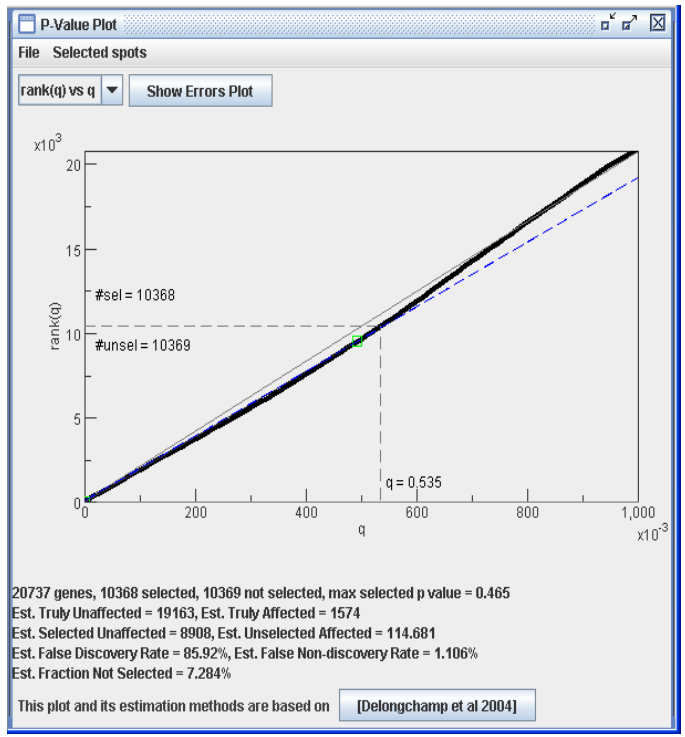
Data Exploring

P-value plot

Some effect



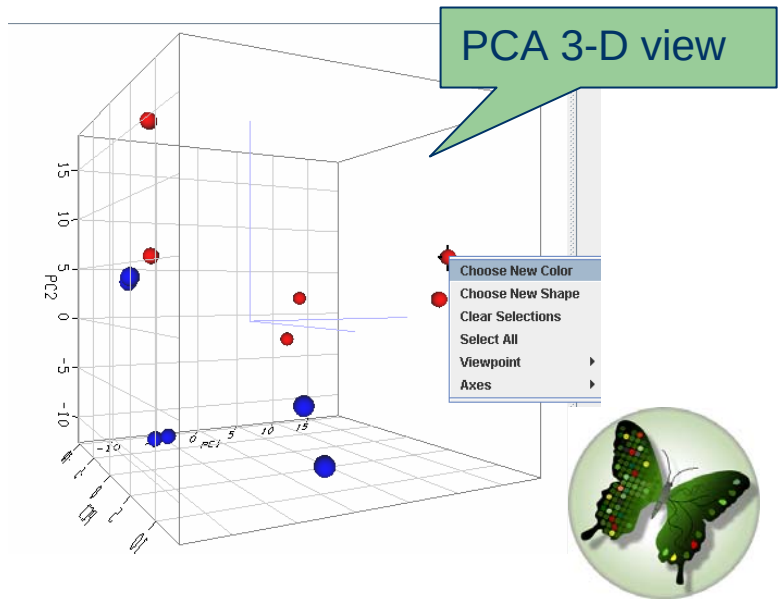
Slightly or no effect



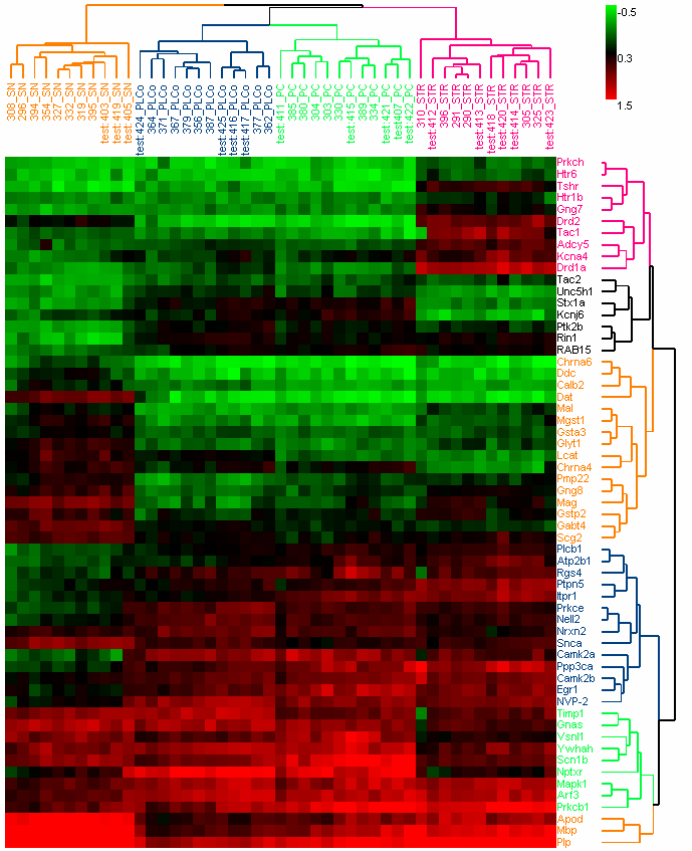
P plot curve is closer to diagonal line, the less treatment effect there is.

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

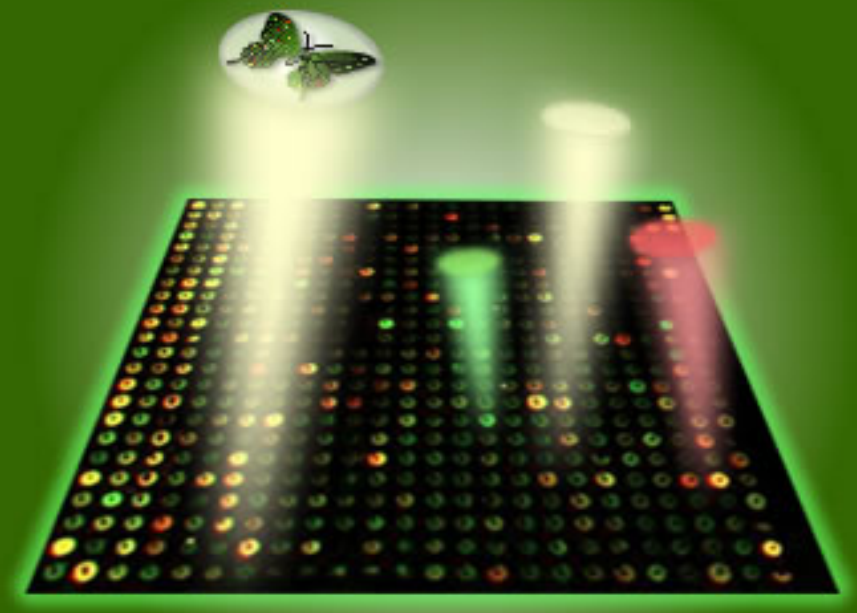
  http://weblaunch.nctr.fda.gov/jnlp/arraytrack/citrix/index.html

- FDA External:
  http://edkb.fda.gov/webstart/arraytrack/index.html

Technical Support

NCTRBioinformaticsSupport@nctr.fda.gov

Thank you!



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