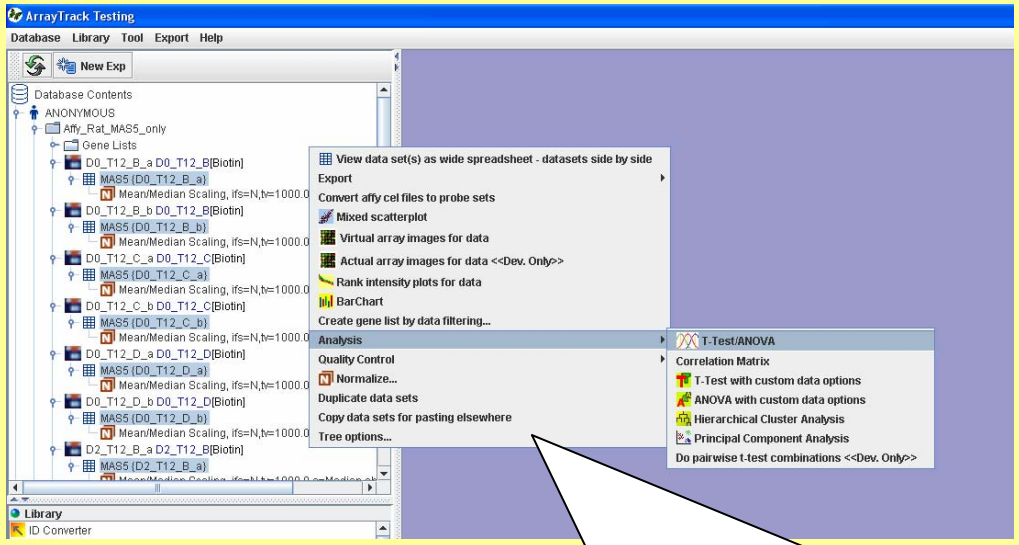
Tutorial Topic	Tutorial Content
1. Comparing two groups (e.g., treated vs control groups). *REQUIRED*	Statistical methods (t-test, permutation t-test, ANOVA) for group comparison. ▪Differentially Expressed Genes (DEGs) identification ▪Biological interpretation (individual gene analysis) using LIB ▪Pathway analysis (KEGG, PathArt, IPA, MetaCore, Key Molnet) ▪Gene Ontology analysis using GOFFA
2. Comparing multiple groups	e.g., multiple doses, time points
3. VennDiagram	Determine the common genes/pathways/functions shared by two or three gene lists (extended to cross-experiment and –platform comparison and systems biology) ▪Apply VennDiagram to the external files
4. Data exploring tools	Principal Component Analysis (PCA) ▪Hierarchical Cluster Analysis (HCA) ▪Apply HCA and PCA to the external files ▪Extensive features in HCA
5. Assessing gene expression profiles using BarChart	Access BarChart from the TOOL box ▪Access BarChart from the t-test result table ▪Access BarChart from ChipLib and other Libs ▪How to use BarChart for cross-experiment comparison • Assign group by color
6. GeneList - An important concept in ArrayTrack	Create a gene list through data filtering and statistical analysis ▪Import/export a gene list ▪Conduct normalization filtered by a gene lists ▪Conduct statistical analysis (t-test/ANOVA, PCA, HCA and others) based on a gene list ▪Export a dataset by specifying the gene list (extended for cross-platform and cross-experiment comparison)
7. Normalization methods	For Affymetrix platform: MAS5, RMA, DChip, Plier, Plier+16 ▪For other platforms: 7 methods (e.g., LOWESS)
8. How to create your own workspace	Copy/Paste/Duplicate an experiment
9A Import	Manual import and batch import
9B Export	Options of data exporting ▪Export a selected dataset with specifying a sub list of gene ▪Export multiple experiments and/or platforms using selected geneID types (e.g., RefSeq)
10. Other useful functions	Correlation matrix •IDConverter – converting one gene ID to another (e.g., from AffyID to AgilentID or GeneBank#, or LocusLinkID or vice verse) •ScatterPlot – pair-wise plot •JoinTable – Combine two tables •SplitTable – If a table contains multiple hybridization data in column with genes in row, the function split the table into individual tables with single hybridization data. •GetUniqueID – If a table contains duplicated IDs, the function pick out the unique IDs
11. Basic scripting for querying (raw and normalized) data and table	Query data from the database tree (How to use *) e.g., *EST, EST*, *EST*=EST ▪Query data in tables e.g., contain, like (%) and inlist
12. Using SAM through ArrayTrack	Provide SAM analysis tool for identifying statistical significant genes in a set of microarray experiments
13. SimpleTox Format	•For animal toxicology and toxicogenomics study, integrate microarray data with toxicological data. •Data are searchable for toxicological information and microarray information. •Batch import SimpleTox data including toxicological endpoints and microarray data. •Contains required/optional and customized fields, not limited to non-clinical study.

Array Track

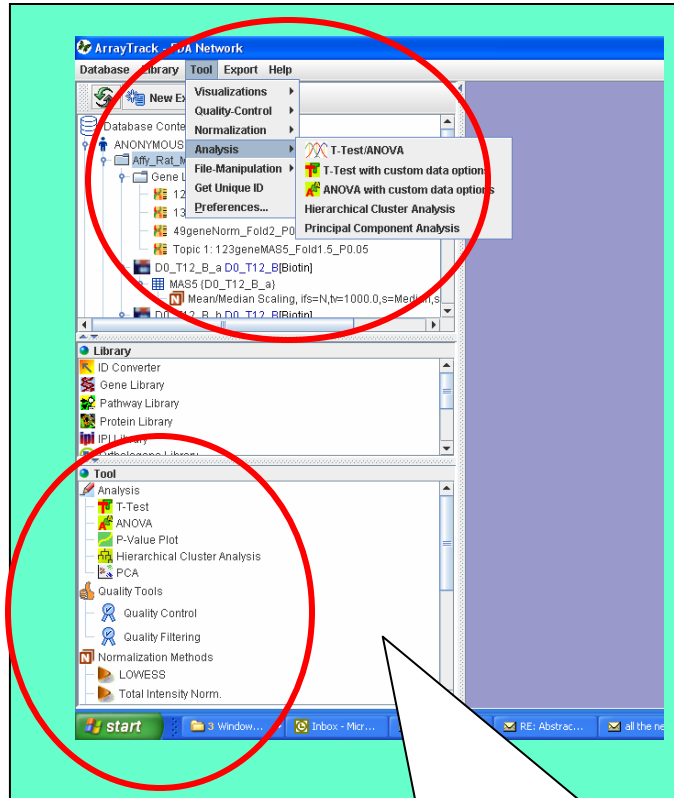
Tutorial 1:

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

Useful Tips



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



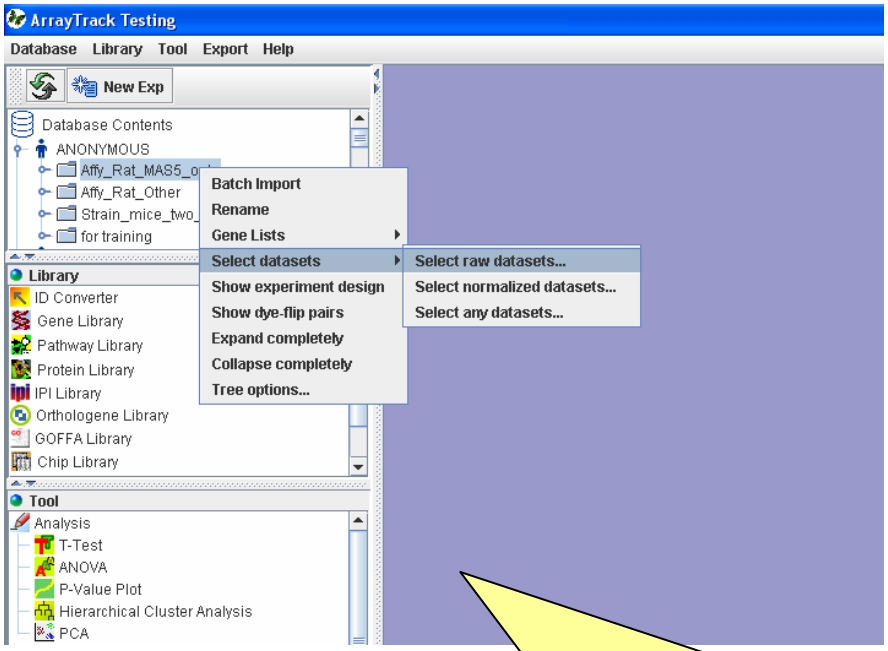
Many functions in ArrayTrack are accessible from multiple paths, for example, left-side window panels, pull down menus and right mouse-click options. (***Analyze the external datasets by accessing the tools from the left-window panel***)

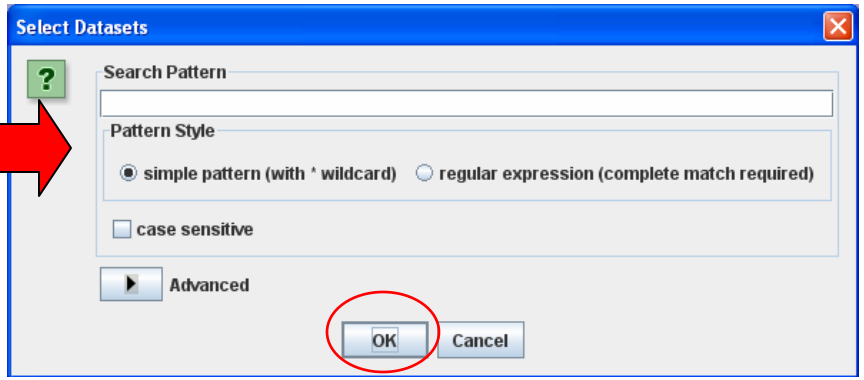
Bookmark the website for the convenience of accessing ArrayTrack (***Recommended***)

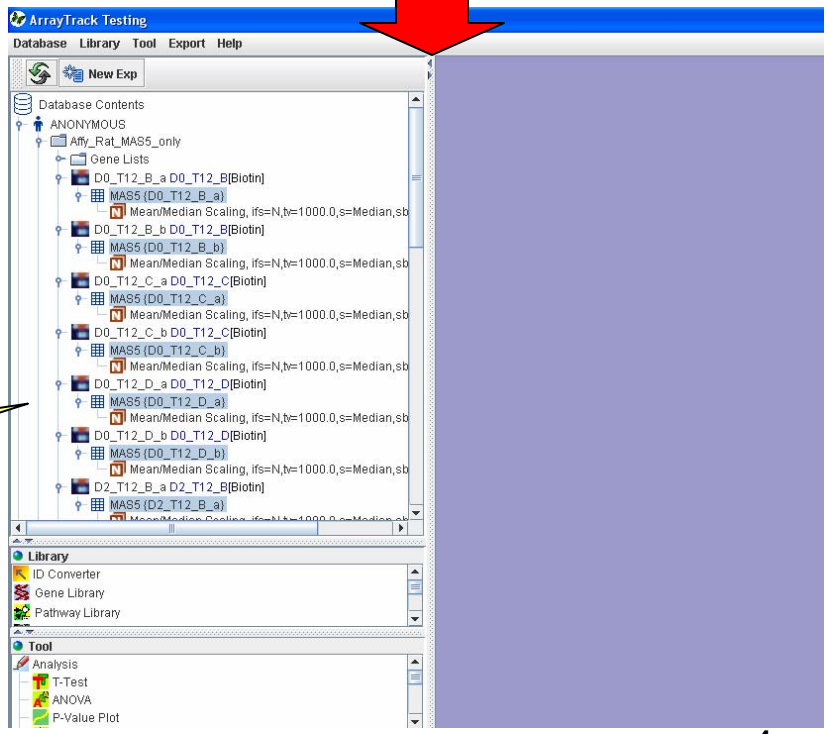
Other Tips

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

Step 1 - Select “Dataset” for Analysis



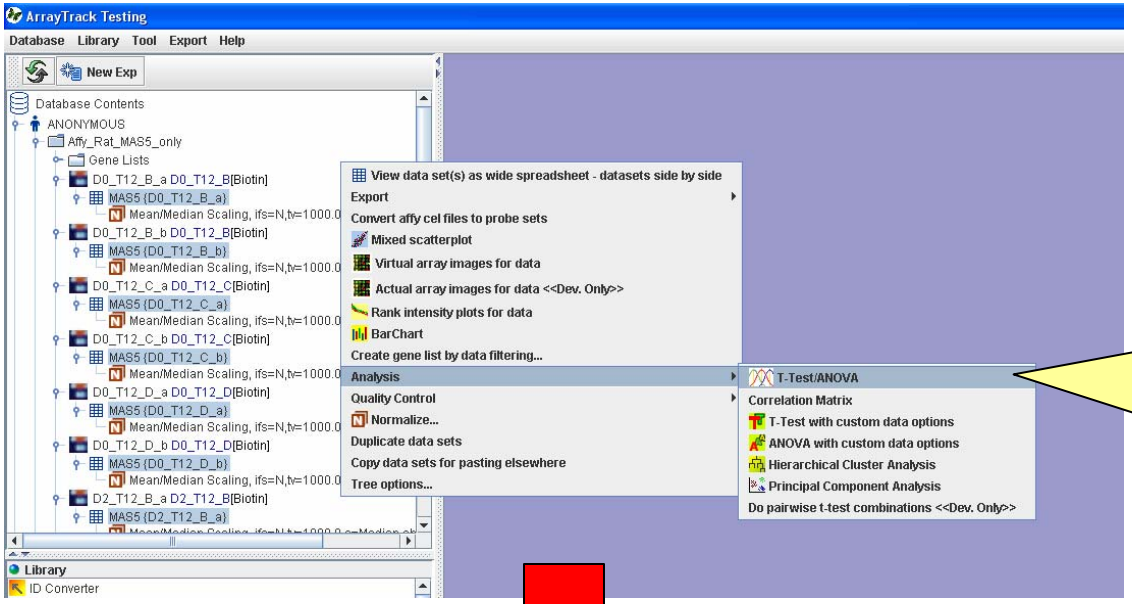




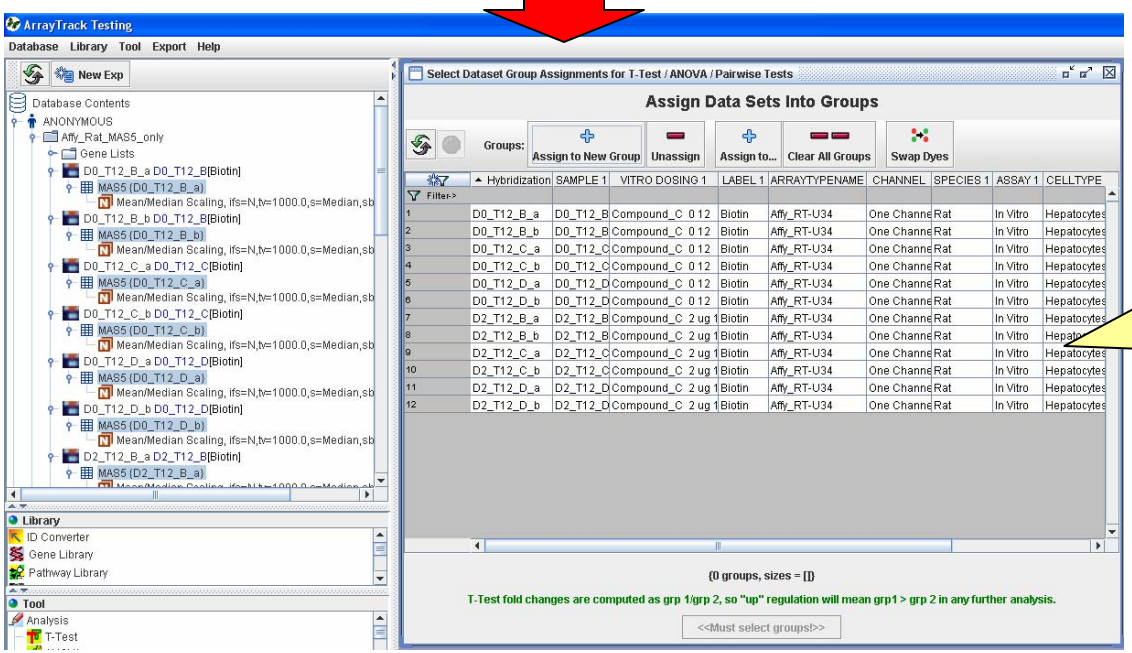
Right click the experiment “Affy_Rat_MAS5_Only” and select “select datasets” → “select raw dataset…”

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

Step 2 – Select an analysis method

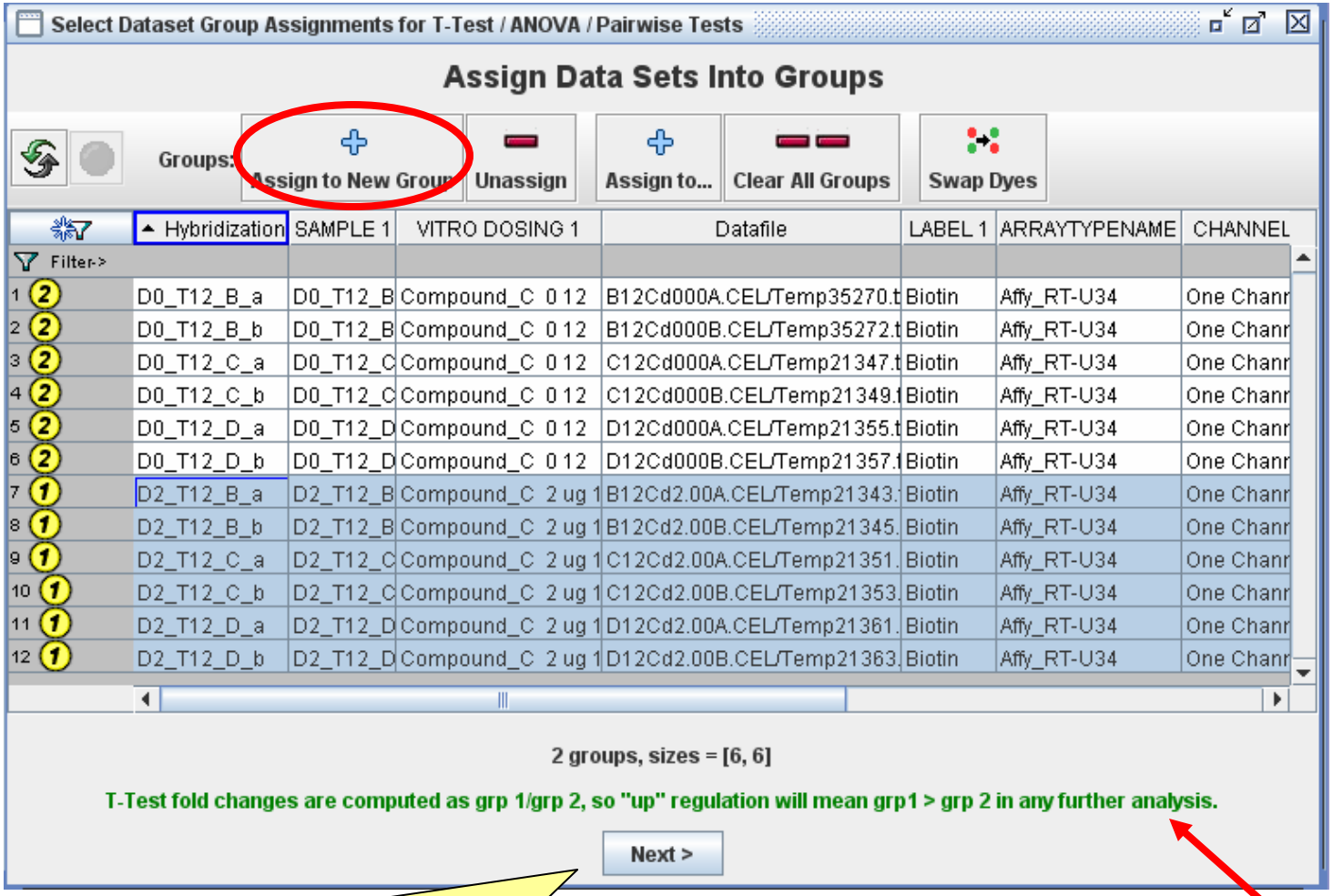


- Select “T-Test/ANOVA”
- Note: there are other analysis methods available in this menu.



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

Step 3 – Assign arrays into groups



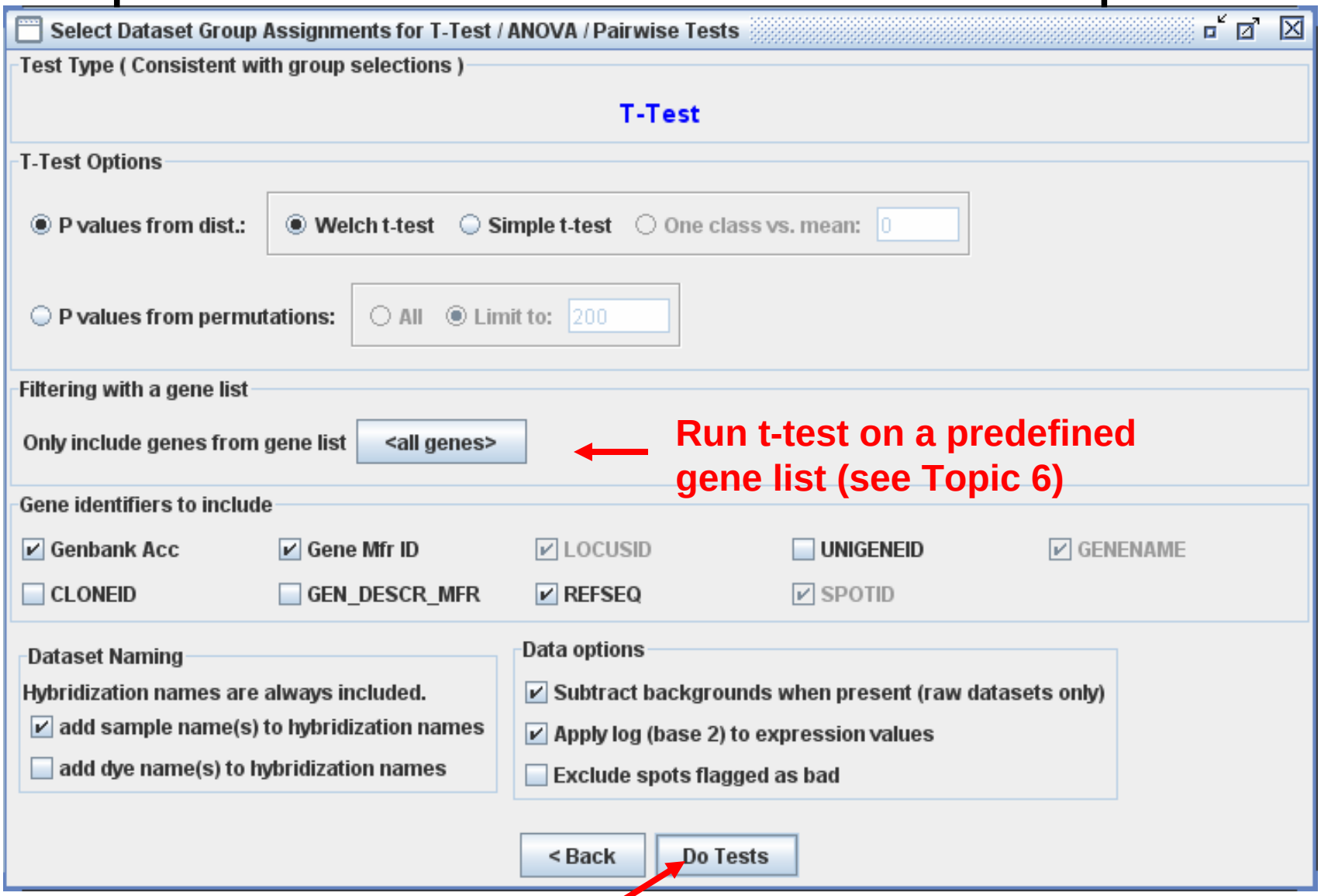
		Hybridization	SAMPLE 1	VITRO DOSING 1	Datafile	LABEL 1	ARRAYTYPENAME	CHANNEL
1	2	D0_T12_B_a	D0_T12_B	Compound_C 0 12	B12Cd000A.CEL/Temp35270.t	Biotin	Affy_RT-U34	One Chann
2	2	D0_T12_B_b	D0_T12_B	Compound_C 0 12	B12Cd000B.CEL/Temp35272.t	Biotin	Affy_RT-U34	One Chann
3	2	D0_T12_C_a	D0_T12_C	Compound_C 0 12	C12Cd000A.CEL/Temp21347.t	Biotin	Affy_RT-U34	One Chann
4	2	D0_T12_C_b	D0_T12_C	Compound_C 0 12	C12Cd000B.CEL/Temp21349.t	Biotin	Affy_RT-U34	One Chann
5	2	D0_T12_D_a	D0_T12_D	Compound_C 0 12	D12Cd000A.CEL/Temp21355.t	Biotin	Affy_RT-U34	One Chann
6	2	D0_T12_D_b	D0_T12_D	Compound_C 0 12	D12Cd000B.CEL/Temp21357.t	Biotin	Affy_RT-U34	One Chann
7	1	D2_T12_B_a	D2_T12_B	Compound_C 2 ug 1	B12Cd2.00A.CEL/Temp21343.	Biotin	Affy_RT-U34	One Chann
8	1	D2_T12_B_b	D2_T12_B	Compound_C 2 ug 1	B12Cd2.00B.CEL/Temp21345.	Biotin	Affy_RT-U34	One Chann
9	1	D2_T12_C_a	D2_T12_C	Compound_C 2 ug 1	C12Cd2.00A.CEL/Temp21351.	Biotin	Affy_RT-U34	One Chann
10	1	D2_T12_C_b	D2_T12_C	Compound_C 2 ug 1	C12Cd2.00B.CEL/Temp21353.	Biotin	Affy_RT-U34	One Chann
11	1	D2_T12_D_a	D2_T12_D	Compound_C 2 ug 1	D12Cd2.00A.CEL/Temp21361.	Biotin	Affy_RT-U34	One Chann
12	1	D2_T12_D_b	D2_T12_D	Compound_C 2 ug 1	D12Cd2.00B.CEL/Temp21363.	Biotin	Affy_RT-U34	One Chann

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

- Select the arrays starting with D2 (treated samples, from 7 to 12) and click “Assign to New Group” (group 1)
- Select the arrays starting with D0 (control samples, from 1 to 6) and click “Assign to New Group” (group 2)

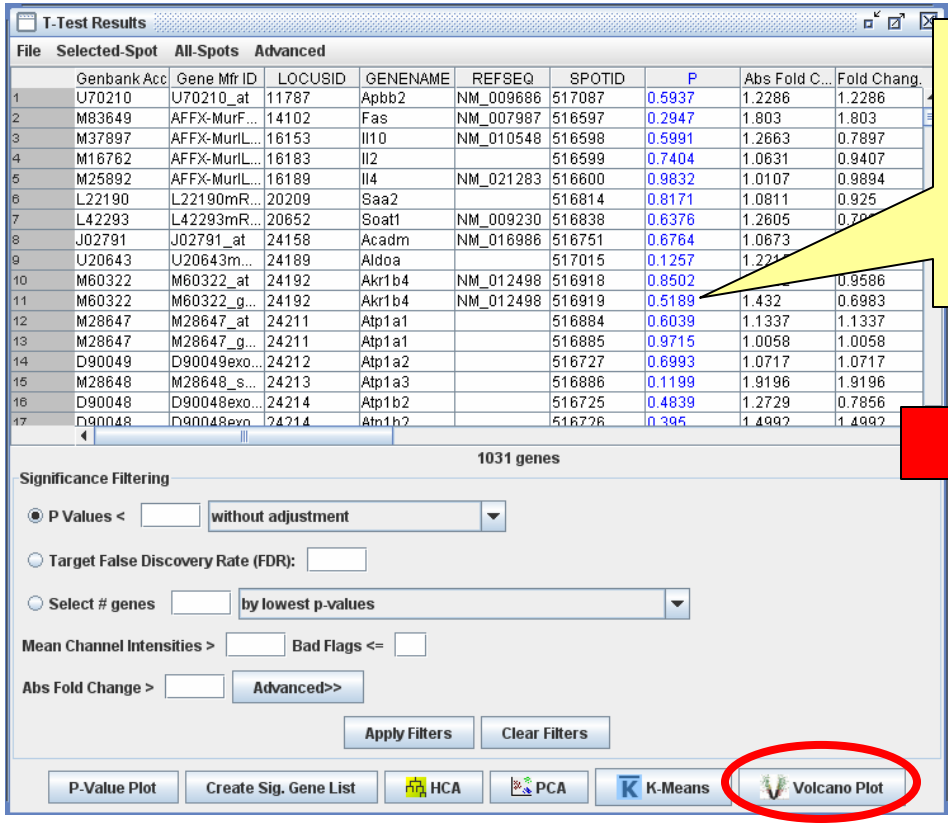
Pay attention on this statement and click “next”

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

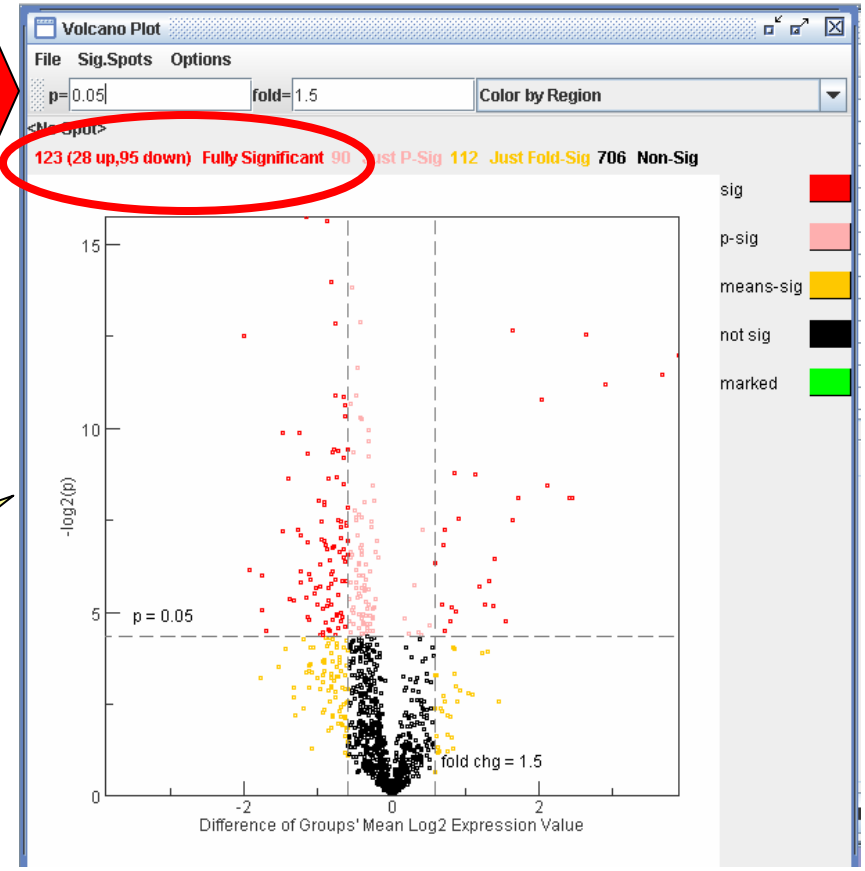


Select “Do Test” if you don’t know which option should be used.

Step 5 – Determine differentially expressed genes

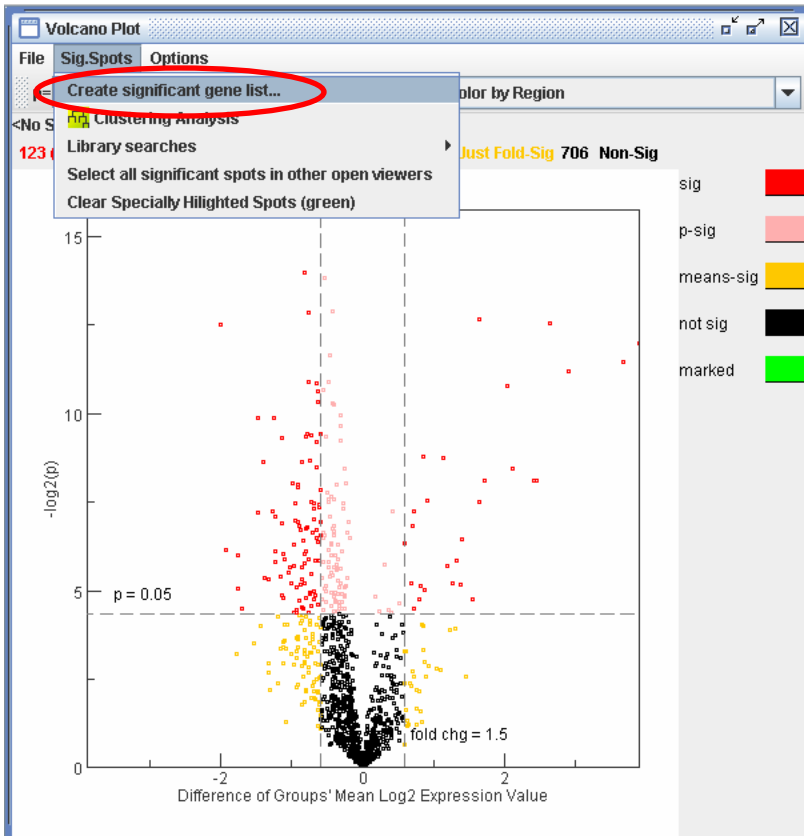


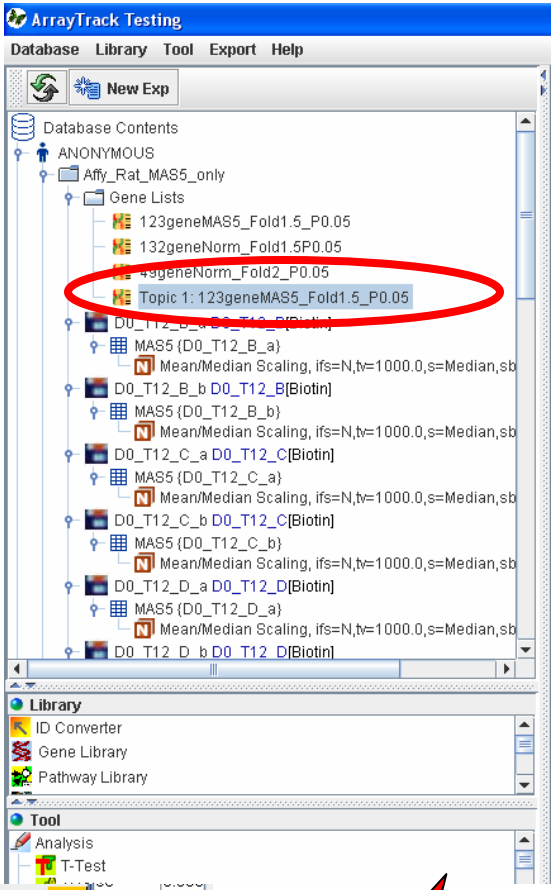
- T-test results with p-value is highlighted
- Many options are available under the table. We recommend to use “Volcano Plot”.

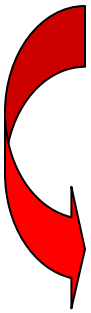


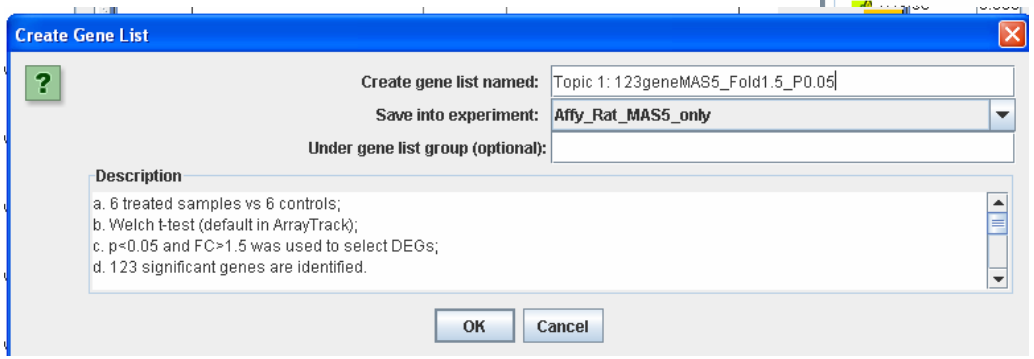
- Volcano plot results
- 123 genes have p<0.05 and Fold Change >1.5

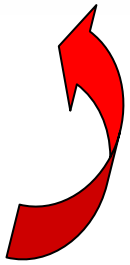
Step 6 – Save the significant gene list (recommended)











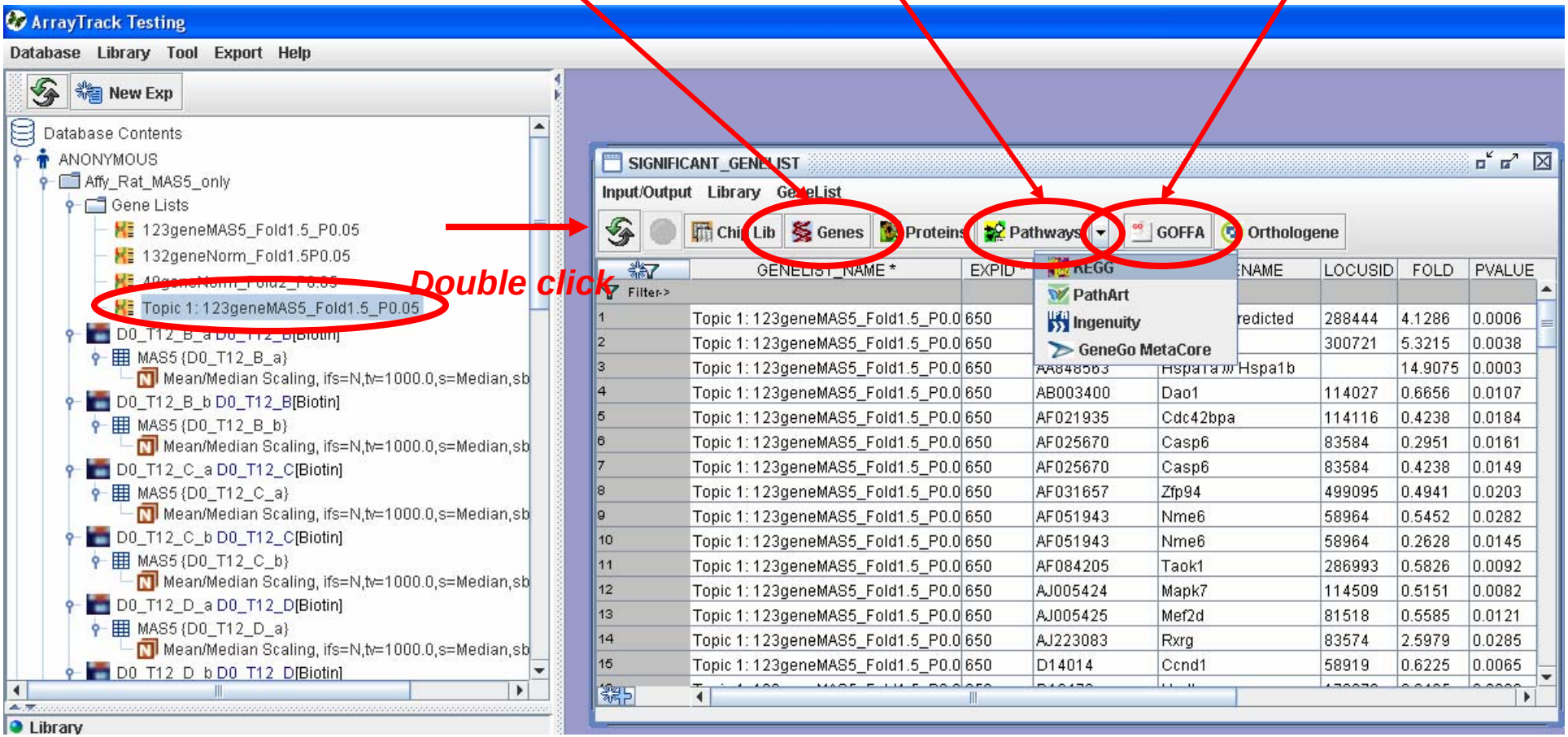
Step 7 – Biological interpretation using ArrayTrack pathway, GO and other tools

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

Individual gene annotation and analysis

Pathway analysis

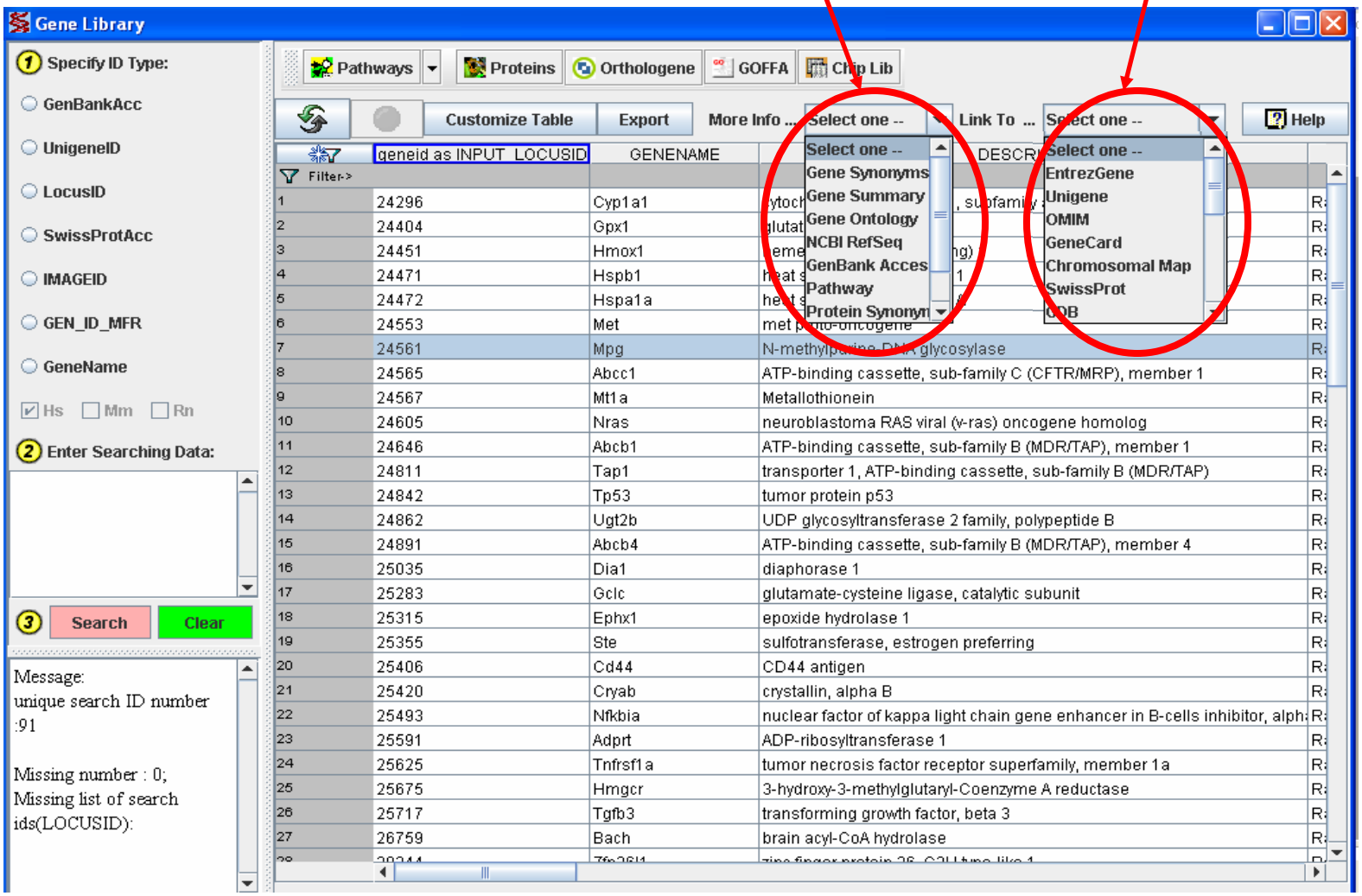
GO-based functional analysis



Step 8 – Individual gene annotation and analysis using GeneLib

Summary info for the highlighted gene

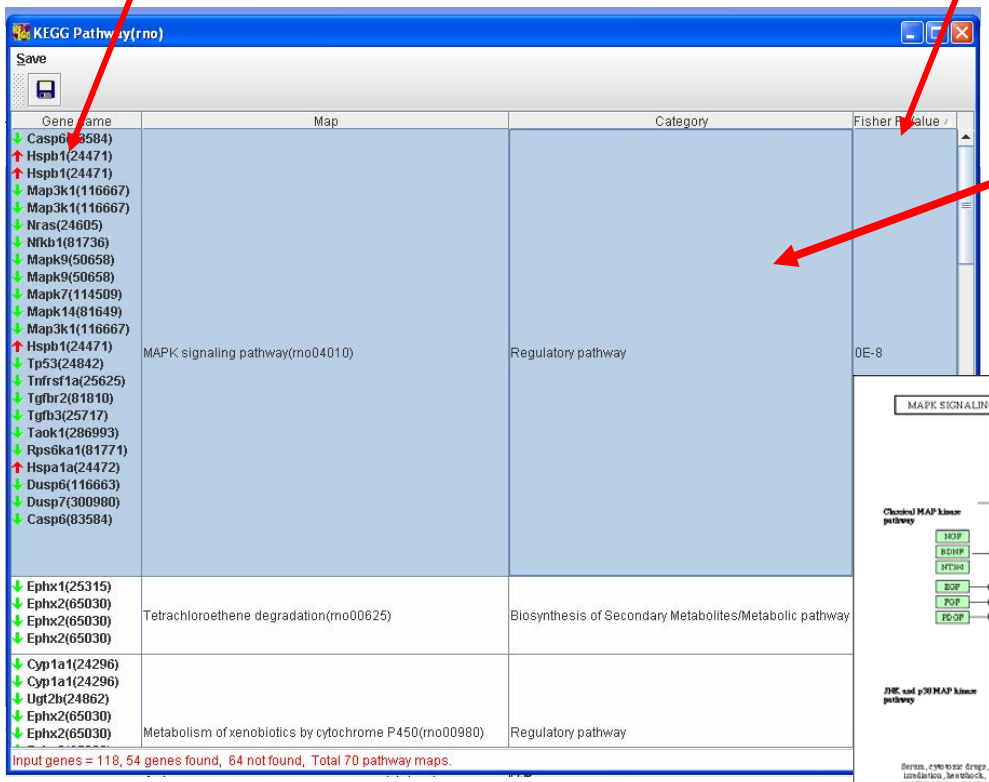
Link to other pub databases for the highlighted gene



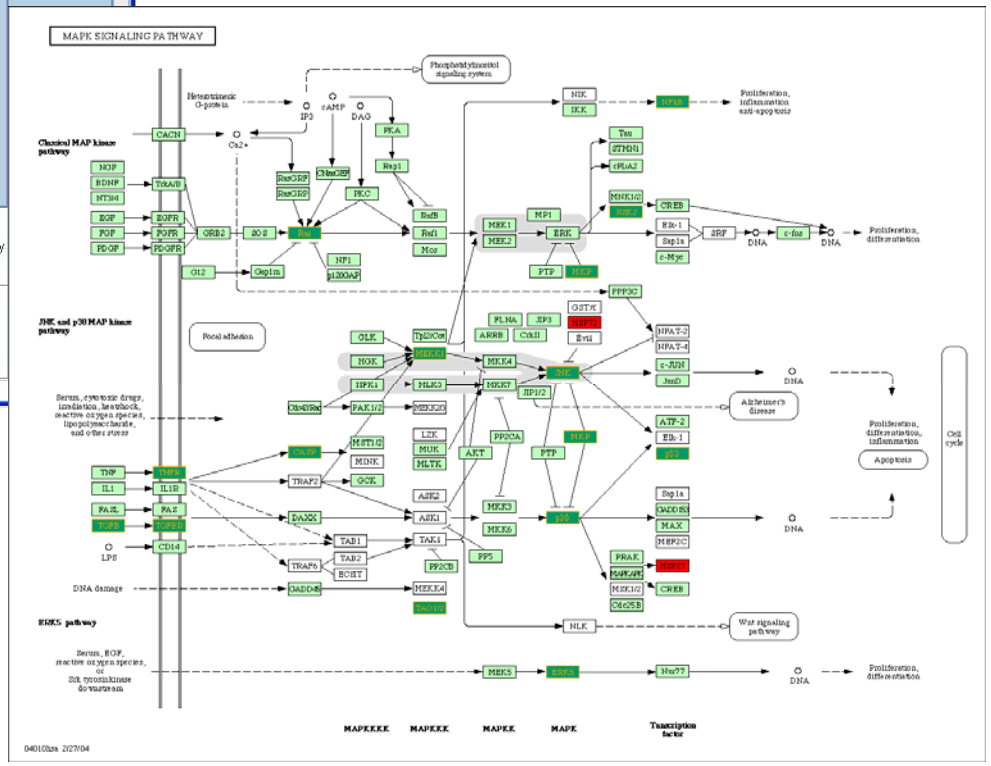
Step 9 – Pathway analysis using KEGG

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

Statistical significance

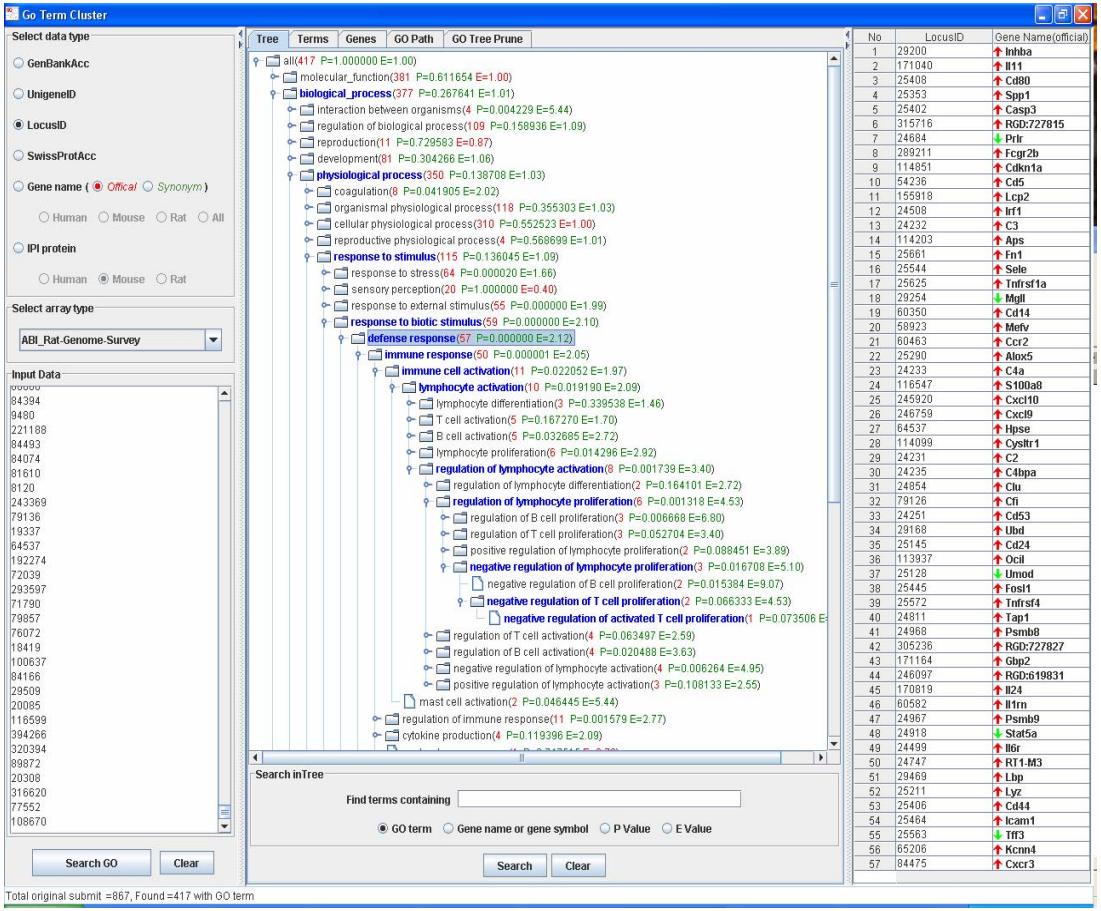


Double click the pathway



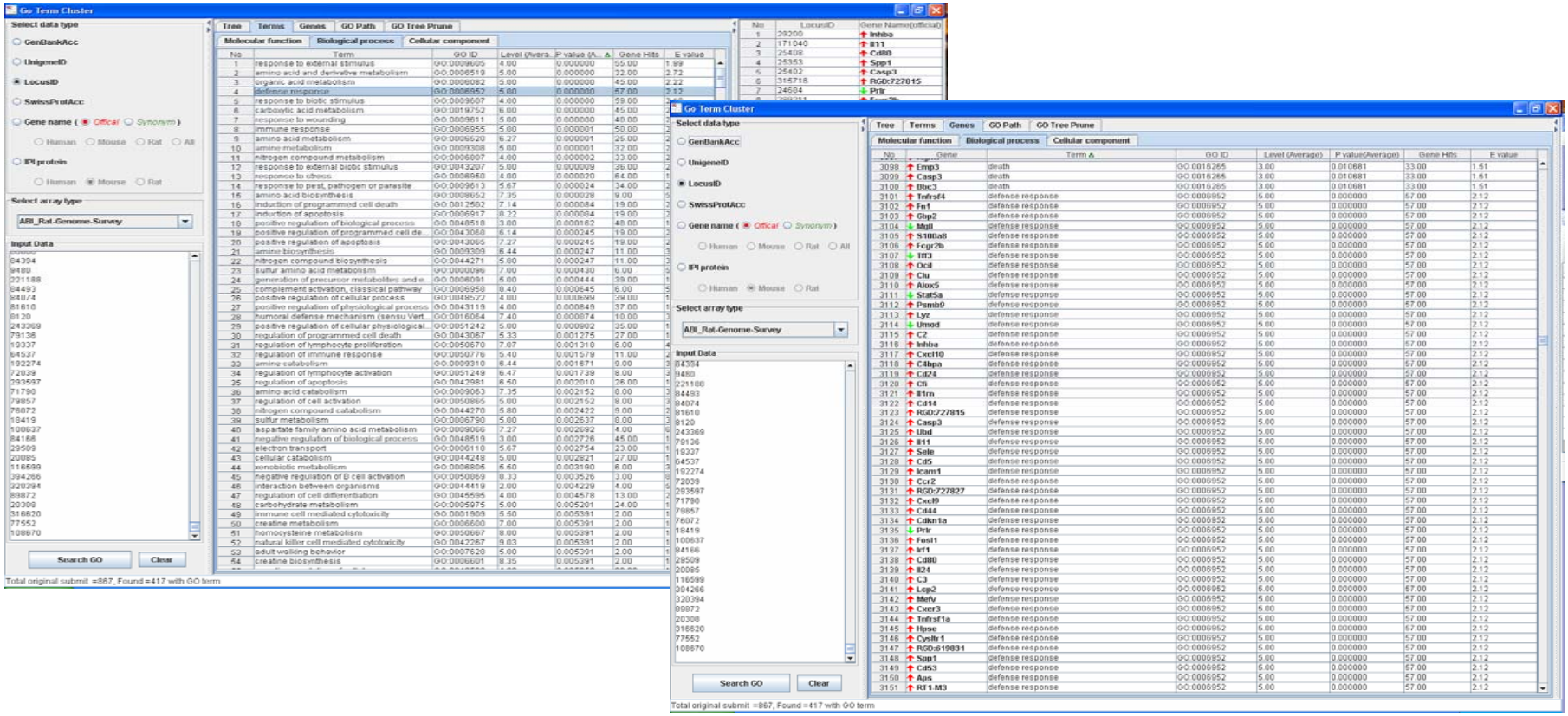
MKP	Down-regulation
HSP72	Up-regulation

Step 10 – GO-based analysis using GOFFA



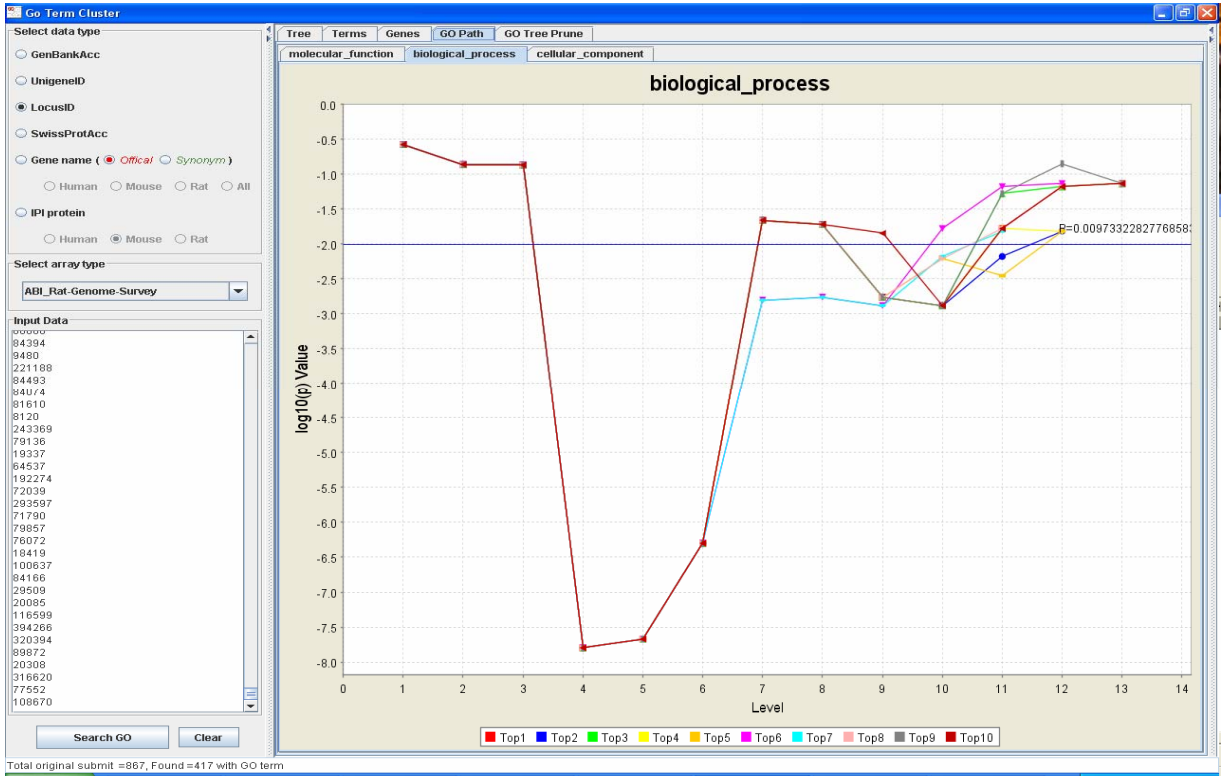
Tree Window – This is the default view of GOFFA, which enables the hierarchical display of the GO terms in a tree-like format; p- and E-values as well as the number of genes are also displayed for each GO term. E-values >1 are shown in green and those <1 in red, respectively denoting greater or lesser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

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



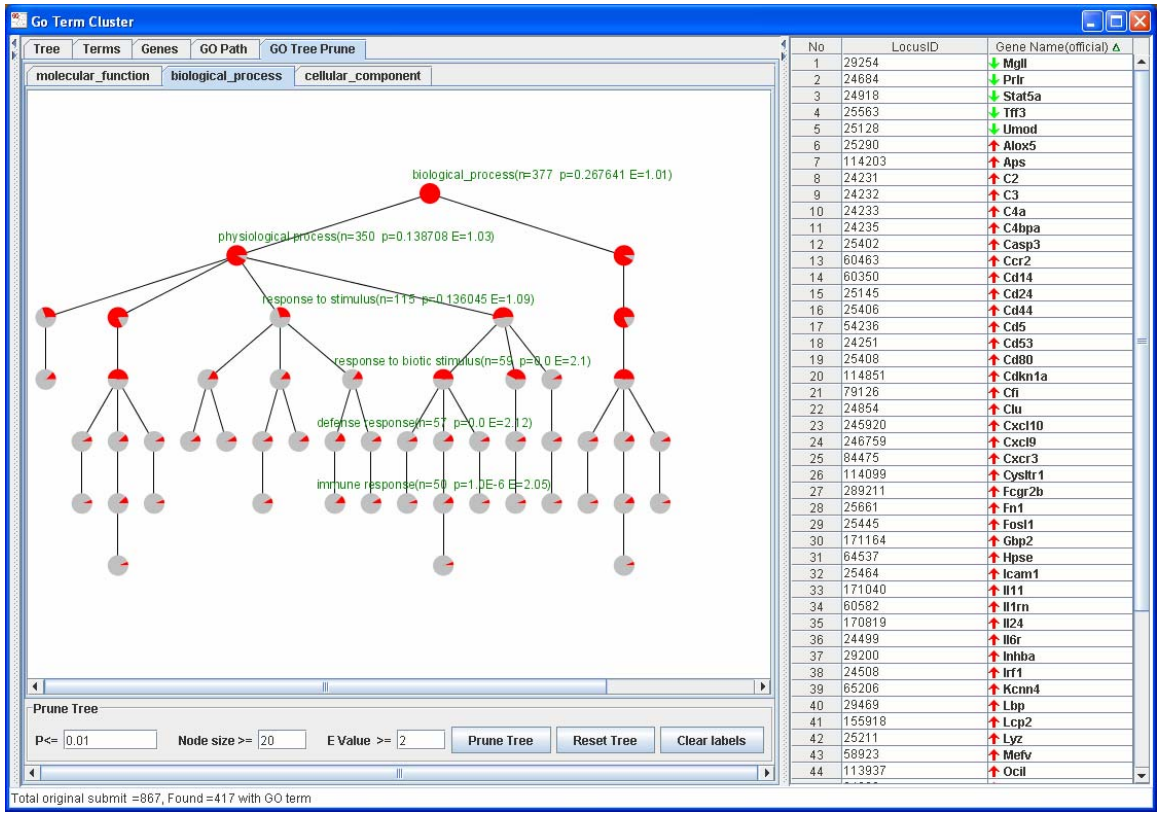
Terms and Genes Windows – The Terms Window (A) and Genes Window (B) summarize the findings associated with GO terms and genes respectively in a table format along with various statistical parameters (e.g., p- and E-values). Each View contains three tabs corresponding to three categories of GO (molecular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

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



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

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



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

Array Track

Tutorial 2: Comparing multiple groups

Comparing Multiple Groups

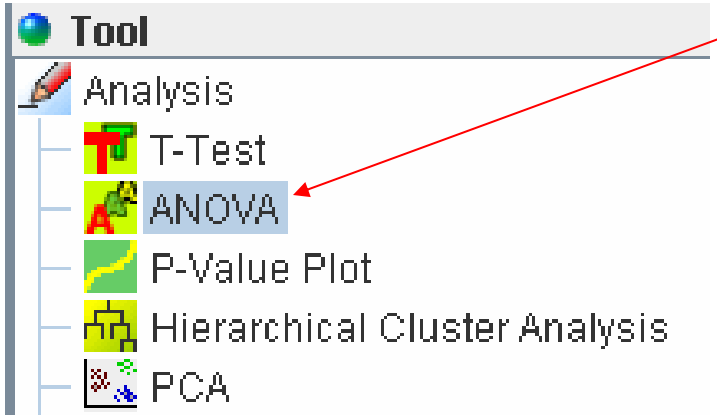
- User can choose ANOVA tool to compare multiple groups of data.
- There are three ways to activate ANOVA

  1. from TOOL panel.
  2. from pull-down menu.
  3. from selected dataset (database panel).

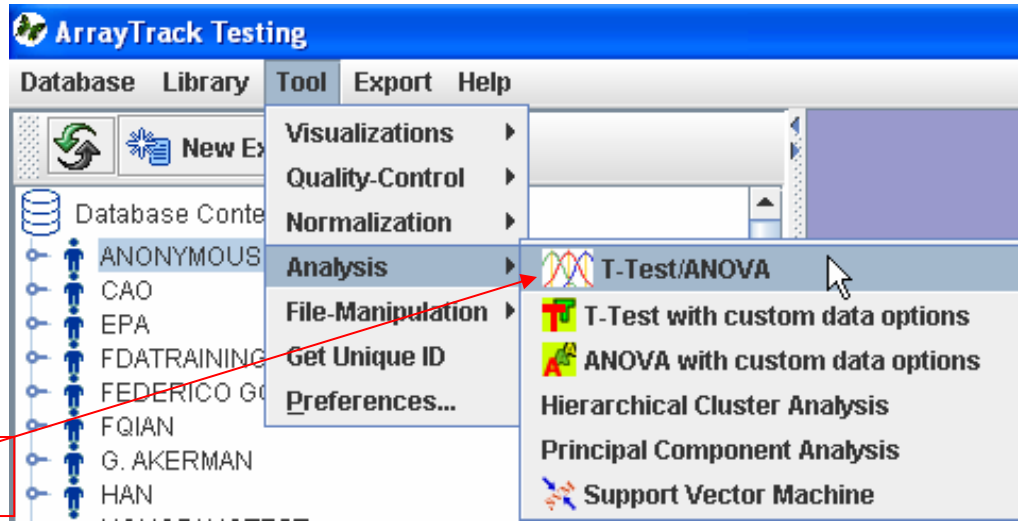
  We recommend the third way.

Comparing Multiple Groups

User can use ANOVA to compare multiple groups. Running ANOVA is similar to running T-test (see tutorial 1 for detail about T-test).



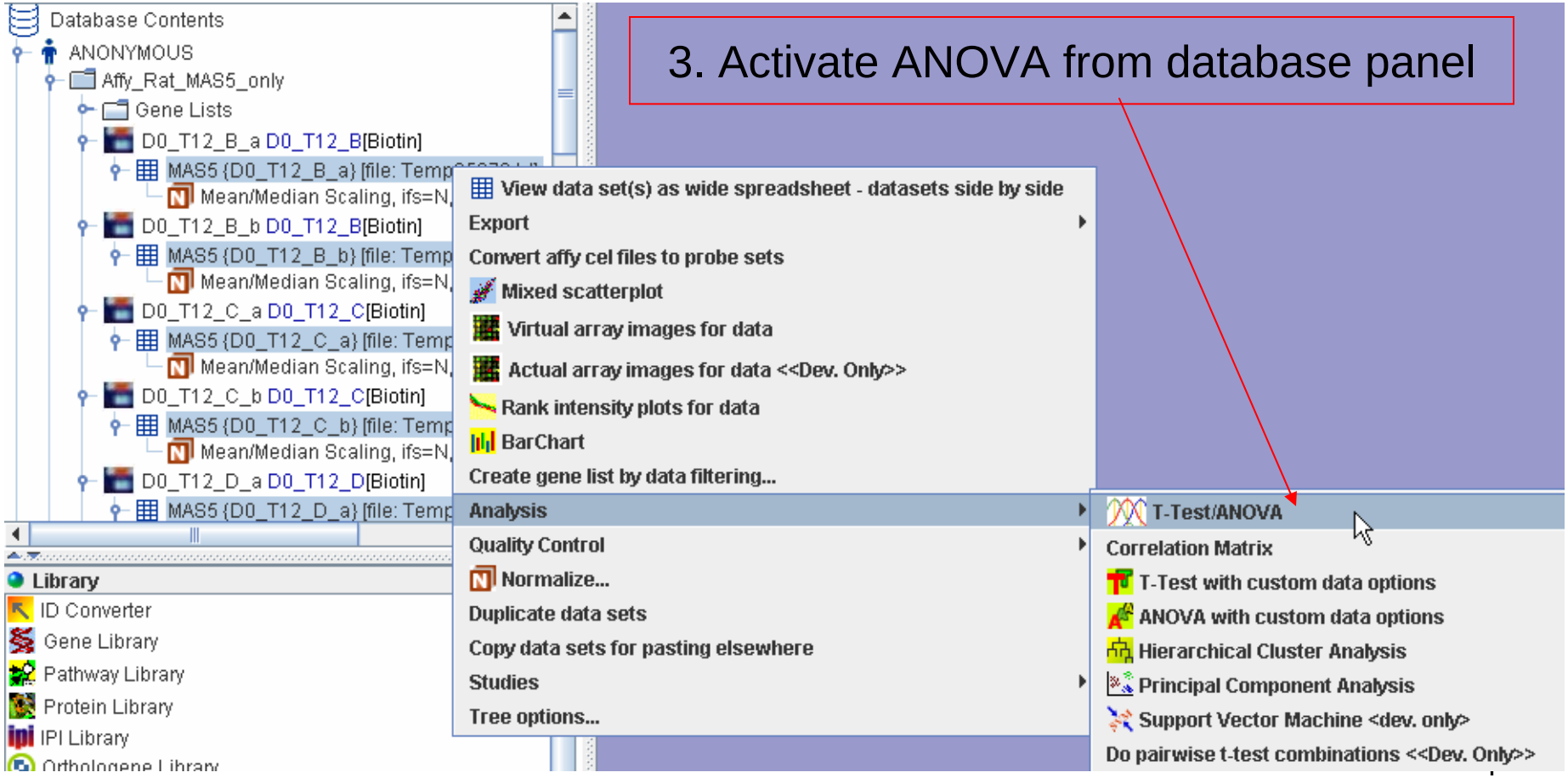
1. Activate ANOVA from Tool panel



2. Activate ANOVA from pull-down menu

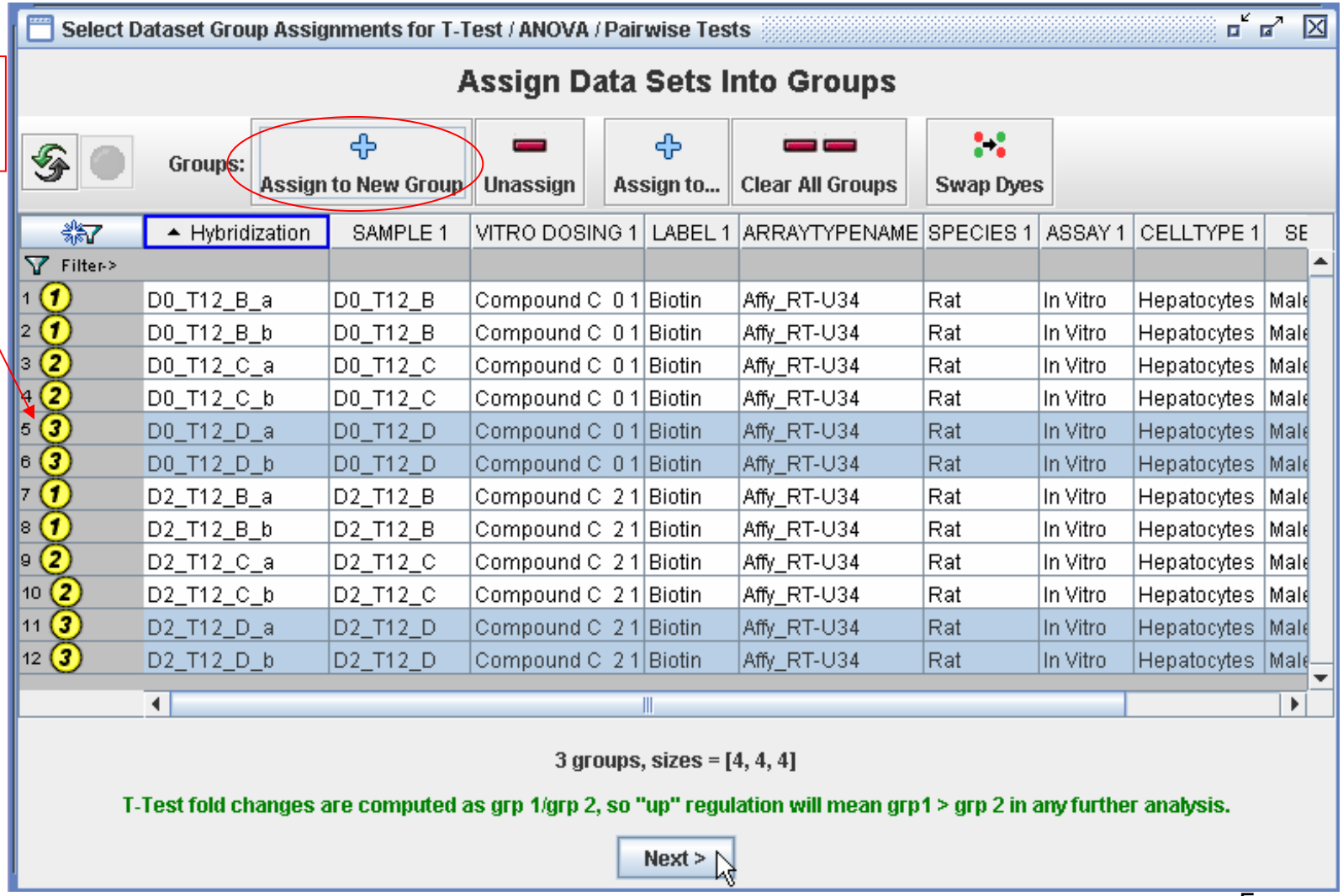
Comparing Multiple Groups – continued.

Right-click the selected datasets, choose “Analysis” -> T-Test/ANOVA.



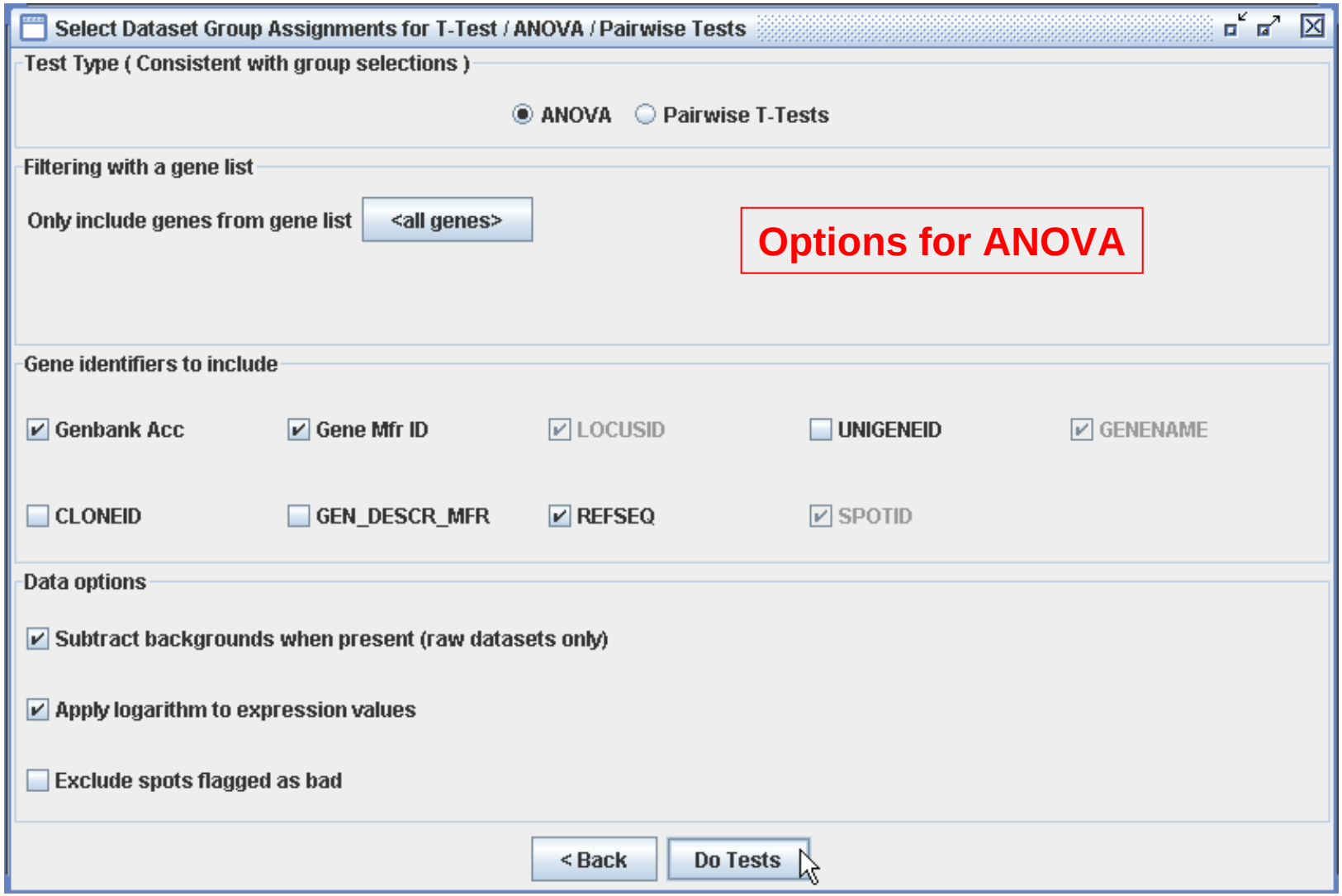
Comparing multiple groups - continued

Assign data into 3 groups



		Hybridization	SAMPLE 1	VITRO DOSING 1	LABEL 1	ARRAYTYPENAME	SPECIES 1	ASSAY 1	CELLTYPE 1	SE
1	1	D0_T12_B_a	D0_T12_B	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
2	1	D0_T12_B_b	D0_T12_B	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
3	2	D0_T12_C_a	D0_T12_C	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
4	2	D0_T12_C_b	D0_T12_C	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
5	3	D0_T12_D_a	D0_T12_D	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
6	3	D0_T12_D_b	D0_T12_D	Compound C 0 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
7	1	D2_T12_B_a	D2_T12_B	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
8	1	D2_T12_B_b	D2_T12_B	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
9	2	D2_T12_C_a	D2_T12_C	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
10	2	D2_T12_C_b	D2_T12_C	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
11	3	D2_T12_D_a	D2_T12_D	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male
12	3	D2_T12_D_b	D2_T12_D	Compound C 2 1	Biotin	Affy_RT-U34	Rat	In Vitro	Hepatocytes	Male

Comparing Multiple Groups - continued



Comparing Multiple Groups - continued

ANOVA result of 3 groups

ANOVA Results

File Selected-Spot All-Spots Advanced

	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	F Ratio	Num Gr...	Total Bad Fl...	Abs Fold C...	Fold Chang...	Diff Means ...	Grp 1 Size	Grp 1 Bad F...
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.6343	0.4791	3	12	1.491	1.491	0.5763	4	4
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.1063	2.9053	3	12	3.4635	3.4635	1.7922	4	4
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.1621	2.2421	3	12	2.5044	0.3993	-1.3244	4	4
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.3922	1.0405	3	12	1.3329	1.3329	0.4145	4	4
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.4113	0.9821	3	12	2.2239	2.2239	1.1531	4	4
6	L22190	L22190mR...	20209	Saa2		516814	0.1353	2.5191	3	9	2.0445	2.0445	1.0318	4	4
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.9597	0.0413	3	12	1.1901	1.1901	0.251	4	4
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.0011	15.9787	3	0	1.6787	0.5957	-0.7474	4	0
9	U20643	U20643m...	24189	Aldoa		517015	0.034	5.0392	3	0	1.4411	1.4411	0.5271	4	0
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.0036	11.1745	3	8	2.0299	0.4926	-1.0214	4	4
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.0539	4.1123	3	12	4.0823	0.245	-2.0294	4	4
12	M28647	M28647_at	24211	Atp1a1		516884	0.2325	1.7231	3	12	1.621	0.6169	-0.6969	4	4
13	M28647	M28647_g...	24211	Atp1a1		516885	0	72.5941	3	0	1.7933	0.5576	-0.8426	4	0
14	D90049	D90049exo...	24212	Atp1a2		516727	0.0842	3.2988	3	12	1.4824	0.6746	-0.568	4	4
15	M28648	M28648_s...	24213	Atp1a3		516886	0.2389	1.6855	3	12	2.3333	0.4286	-1.2224	4	4
16	D90048	D90048exo...	24214	Atp1b2		516725	0.062	3.8483	3	12	2.4469	0.4087	-1.291	4	4
17	D90048	D90048exo...	24214	Atp1b2		516726	0.0323	5.1517	3	12	3.1855	0.3139	-1.6715	4	4
18	L14680	L14680_at	24224	Bcl2	NM_016993	516803	0.4519	0.8686	3	12	1.6661	1.6661	0.7365	4	4
19	L14680	L14680_g_...	24224	Bcl2	NM_016993	516804	0.5709	0.5969	3	12	1.2519	0.7988	-0.3242	4	4
20	Z50051	Z50051_at	24235	C4bpa	NM_012516	517239	0.0019	13.5576	3	0	1.8727	0.534	-0.9052	4	0
21	X13933	X13933_s_...	24242	Calm1		517136	0.1177	2.7397	3	0	1.3836	0.7227	-0.4684	4	0
22	M11670	M11670_at	24248	Cat	NM_012520	516842	0.297	1.3936	3	0	1.3653	0.7324	-0.4492	4	0
23	AA926149	rc_AA9261...	24248	Cat		517330	0.73	0.326	3	12	1.6428	1.6428	0.7162	4	4
24	AA926149	rc_AA9261...	24248	Cat		517331	0.3518	1.1758	3	0	1.1945	0.8372	-0.2564	4	0
25	M60753	M60753_s...	24267	Comt		516920	0.0014	14.9522	3	0	2.2705	2.2705	1.183	4	0

1031 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 <=

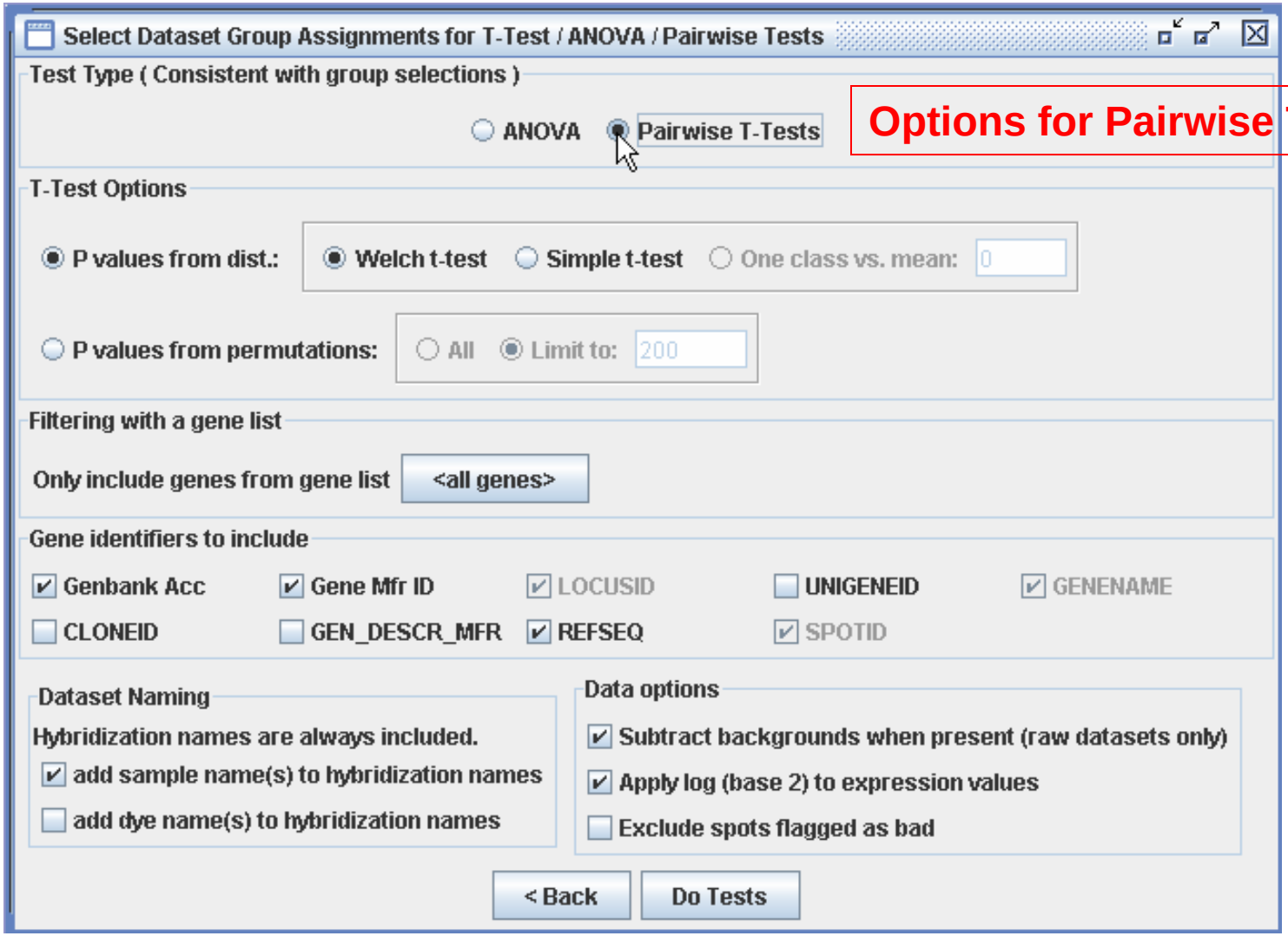
Abs Fold Change > 1.5 Advanced>>

Apply Filters Clear Filters

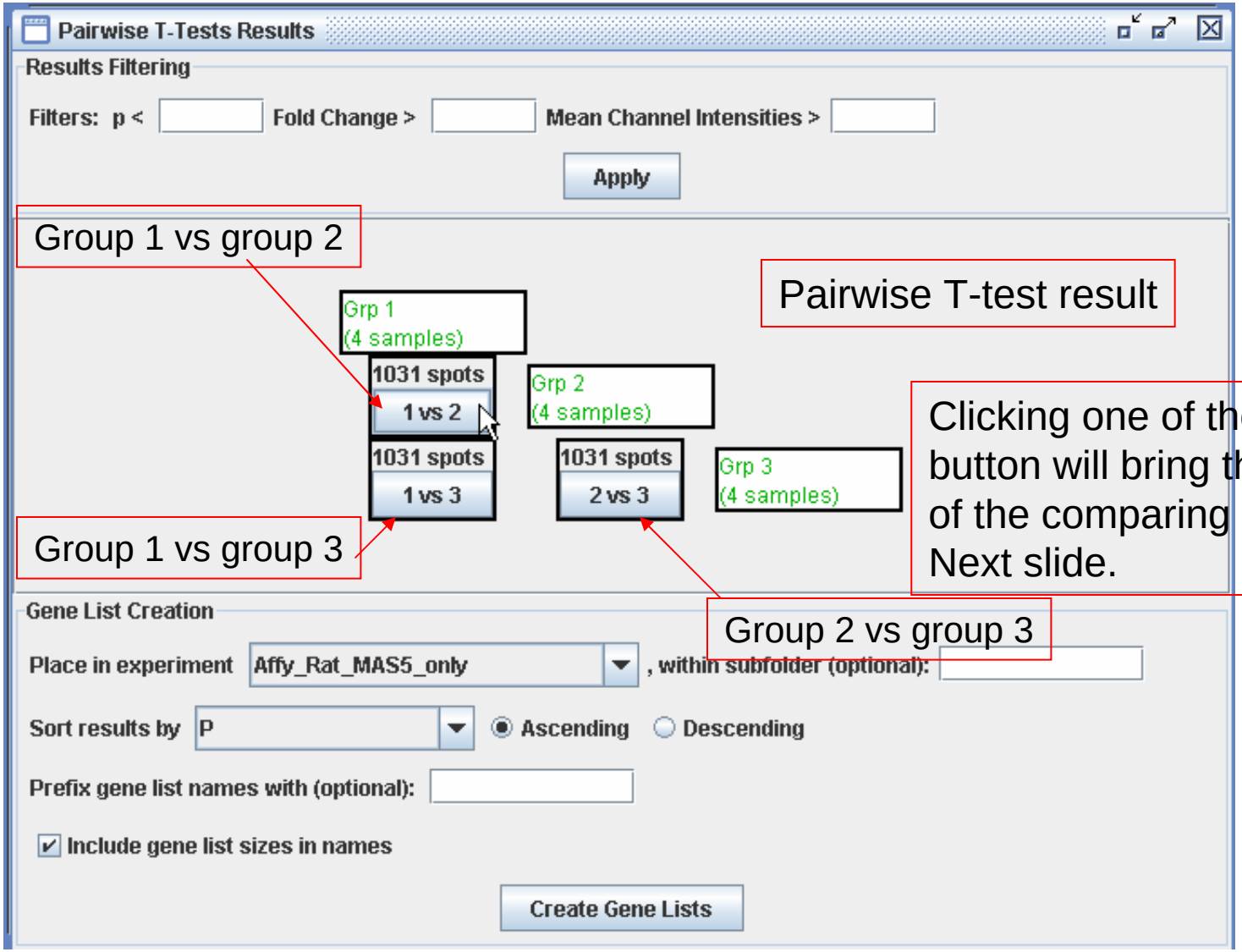
P-Value Plot Create Sig. Gene List HCA PCA K-Means Volcano Plot

User can further filter the results by setting criteria like P value, fold change, etc.

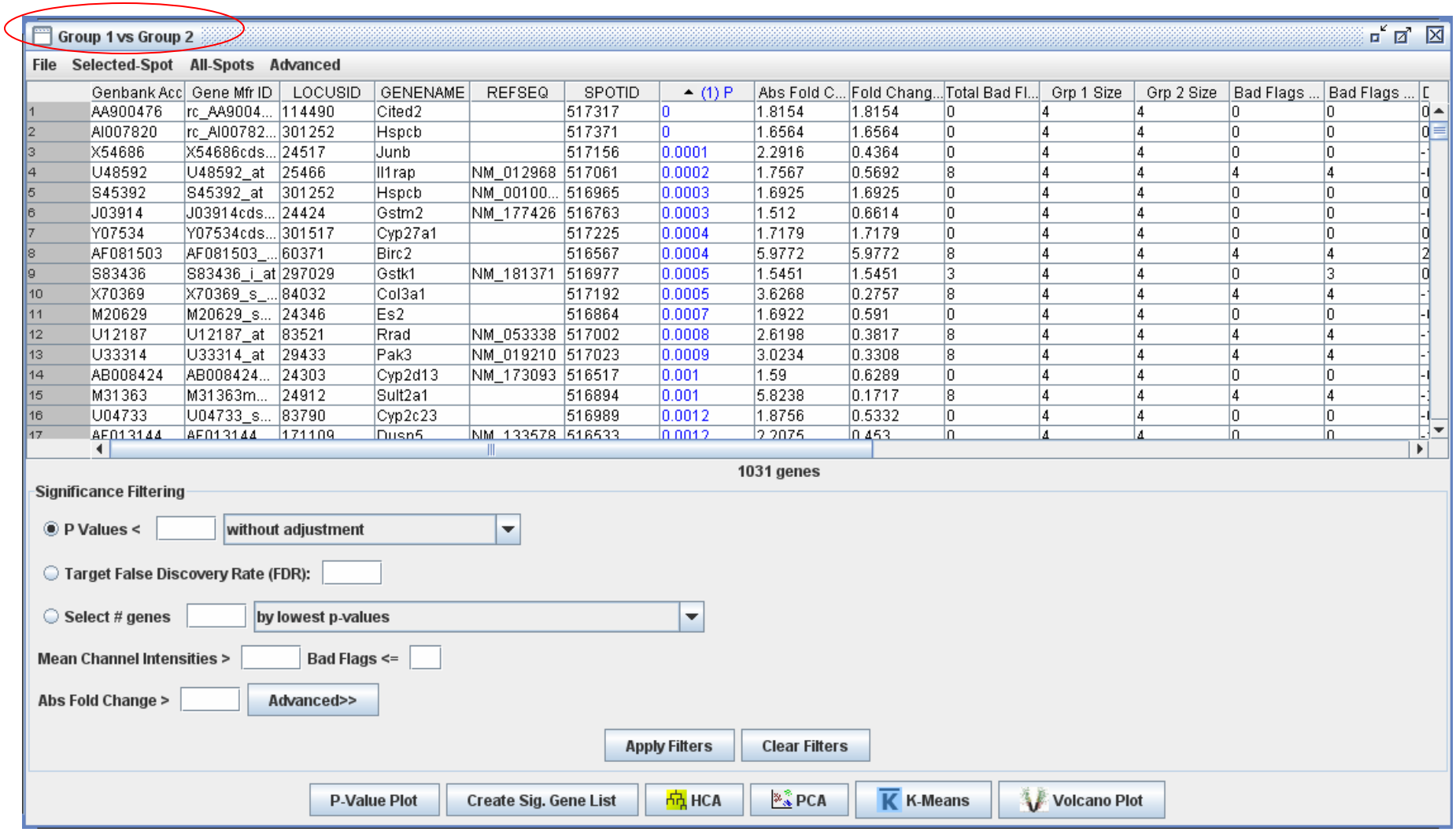
Comparing Multiple Groups - continued



Comparing Multiple Groups - continued



Comparing Multiple Groups - continued



	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	(1) P	Abs Fold C...	Fold Chang...	Total Bad Fl...	Grp 1 Size	Grp 2 Size	Bad Flags ...	Bad Flags ...	D
1	AA900476	rc_AA9004...	114490	Cited2		517317	0	1.8154	1.8154	0	4	4	0	0	0
2	AI007820	rc_AI00782...	301252	Hspcb		517371	0	1.6564	1.6564	0	4	4	0	0	0
3	X54686	X54686cds...	24517	Junb		517156	0.0001	2.2916	0.4364	0	4	4	0	0	-
4	U48592	U48592_at	25466	Il1rap	NM_012968	517061	0.0002	1.7567	0.5692	8	4	4	4	4	-
5	S45392	S45392_at	301252	Hspcb	NM_00100...	516965	0.0003	1.6925	1.6925	0	4	4	0	0	0
6	J03914	J03914cds...	24424	Gstm2	NM_177426	516763	0.0003	1.512	0.6614	0	4	4	0	0	-
7	Y07534	Y07534cds...	301517	Cyp27a1		517225	0.0004	1.7179	1.7179	0	4	4	0	0	0
8	AF081503	AF081503_...	60371	Birc2		516567	0.0004	5.9772	5.9772	8	4	4	4	4	2
9	S83436	S83436_i_at	297029	Gstk1	NM_181371	516977	0.0005	1.5451	1.5451	3	4	4	0	3	0
10	X70369	X70369_s_...	84032	Col3a1		517192	0.0005	3.6268	0.2757	8	4	4	4	4	-
11	M20629	M20629_s...	24346	Es2		516864	0.0007	1.6922	0.591	0	4	4	0	0	-
12	U12187	U12187_at	83521	Rrad	NM_053338	517002	0.0008	2.6198	0.3817	8	4	4	4	4	-
13	U33314	U33314_at	29433	Pak3	NM_019210	517023	0.0009	3.0234	0.3308	8	4	4	4	4	-
14	AB008424	AB008424...	24303	Cyp2d13	NM_173093	516517	0.001	1.59	0.6289	0	4	4	0	0	-
15	M31363	M31363m...	24912	Sult2a1		516894	0.001	5.8238	0.1717	8	4	4	4	4	-
16	U04733	U04733_s...	83790	Cyp2c23		516989	0.0012	1.8756	0.5332	0	4	4	0	0	-
17	AF013144	AF013144	171109	Dusp5	NM_133578	516533	0.0012	2.2075	0.453	0	4	4	0	0	-

Tutorial 3: VennDiagram

- **Why use VennDiagram:**

  VennDiagram is a tool that visually displays the number of common characteristics among 2~3 gene lists.

- **How to get Input to VennDiagram:**

  <1> Gene List saved in ArrayTrack (the most popular usage);

  <2> List saved on your local machine (can be gene/protein/metabolite list);

- **Functions associated with VennDiagram:**

  <1> **Common Genes:** based on different gene IDs (GenBank accession #, RefSeq, and etc)

  <2> **Common Pathways:** based on KEGG and PathArt

  Advanced: across different omics data, such as a gene list from microarray experiment, a protein list from proteomics study, and a metabolite list from metabolomics study

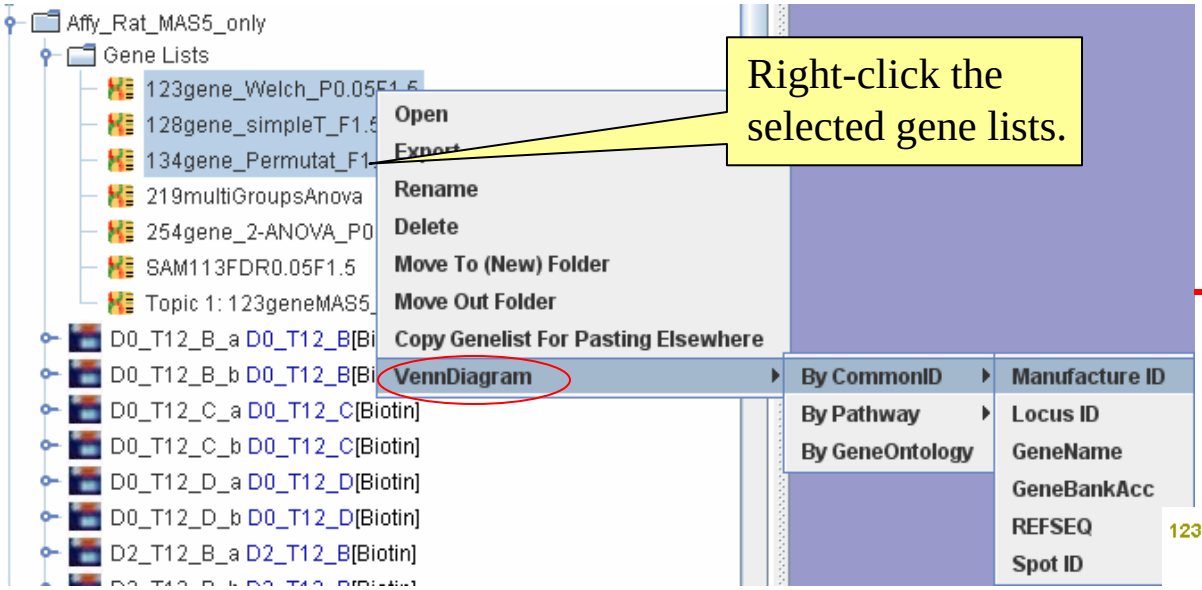
  <3> **Common Gene Ontology terms:**

  <4> **Compare two array types:**

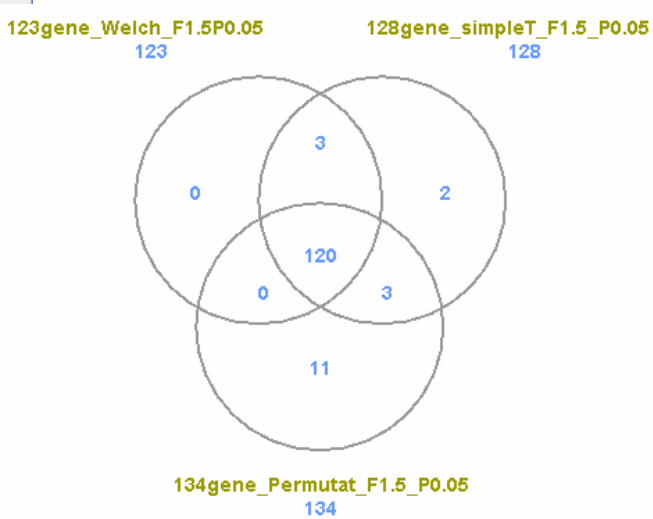
Common Genes

- Two ways to input to VennDiagram :

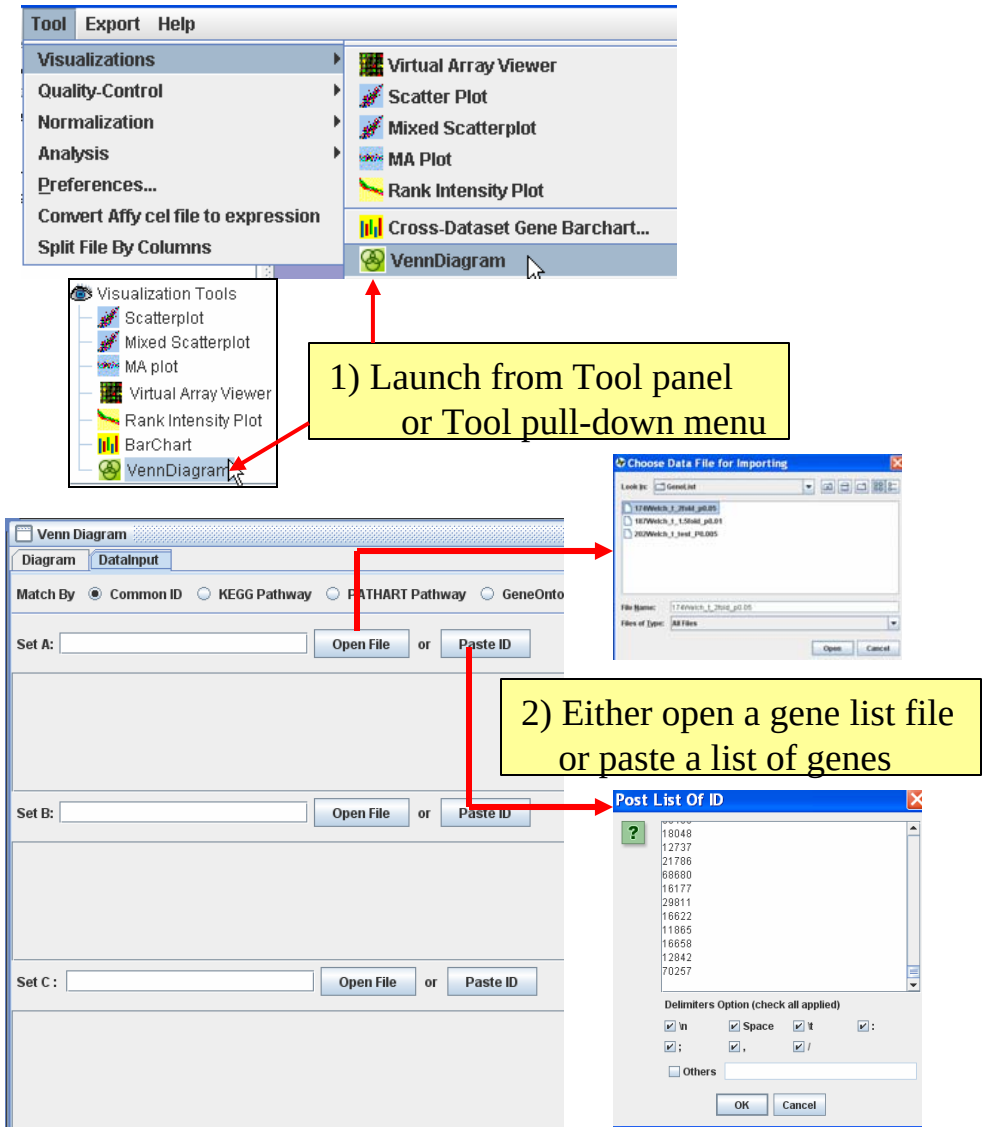
<1> when gene lists have been saved in ArrayTrack:

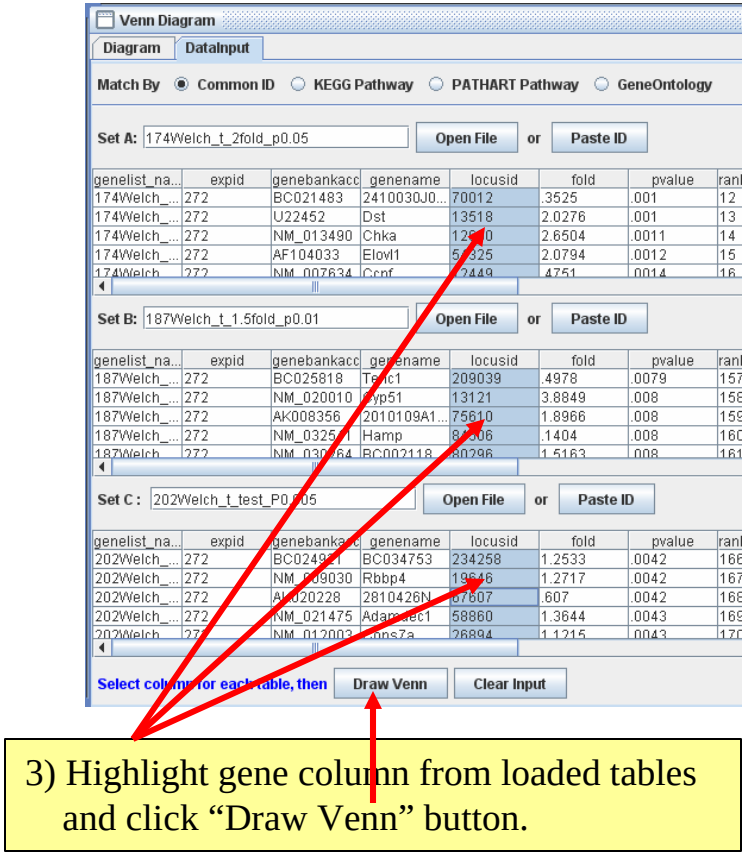


When genelists come from the same chip, Manufactory ID and Spot ID are preferred, otherwise use LocusID or other listed gene IDs

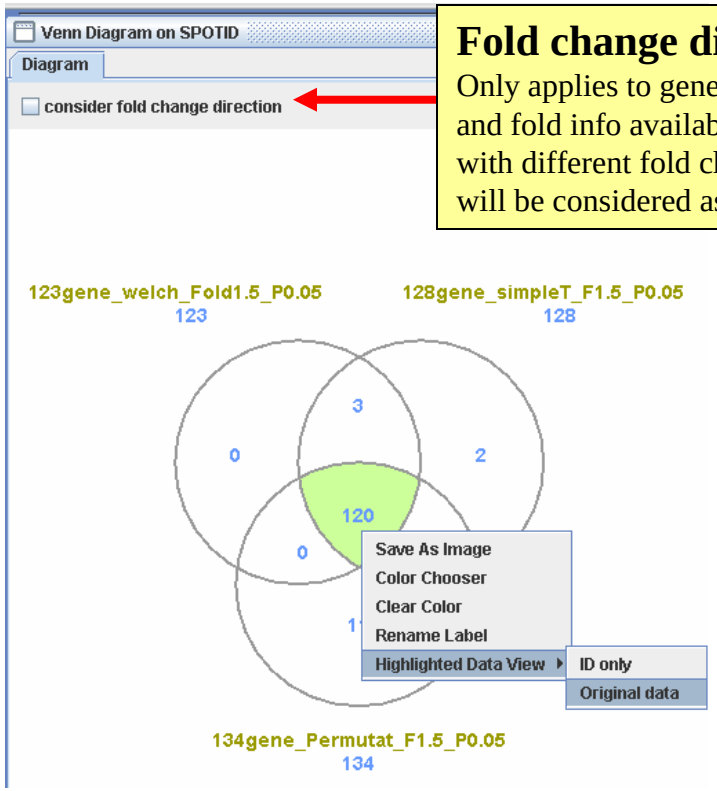


<2> when gene lists are saved in local machine:



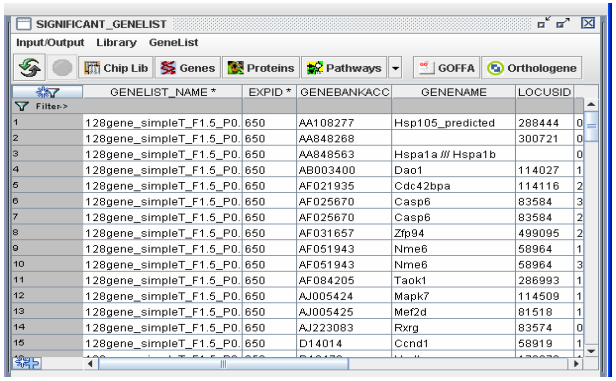


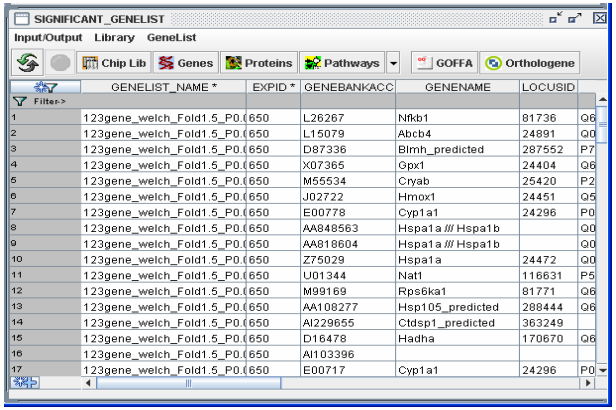
- Display original data for the selected sections

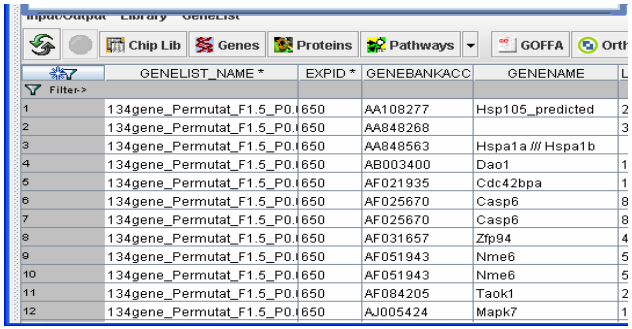


Fold change direction option

Only applies to genelists saved in ArrayTrack and fold info available. If gene id is same, but with different fold change direction, then it will be considered as different genes.







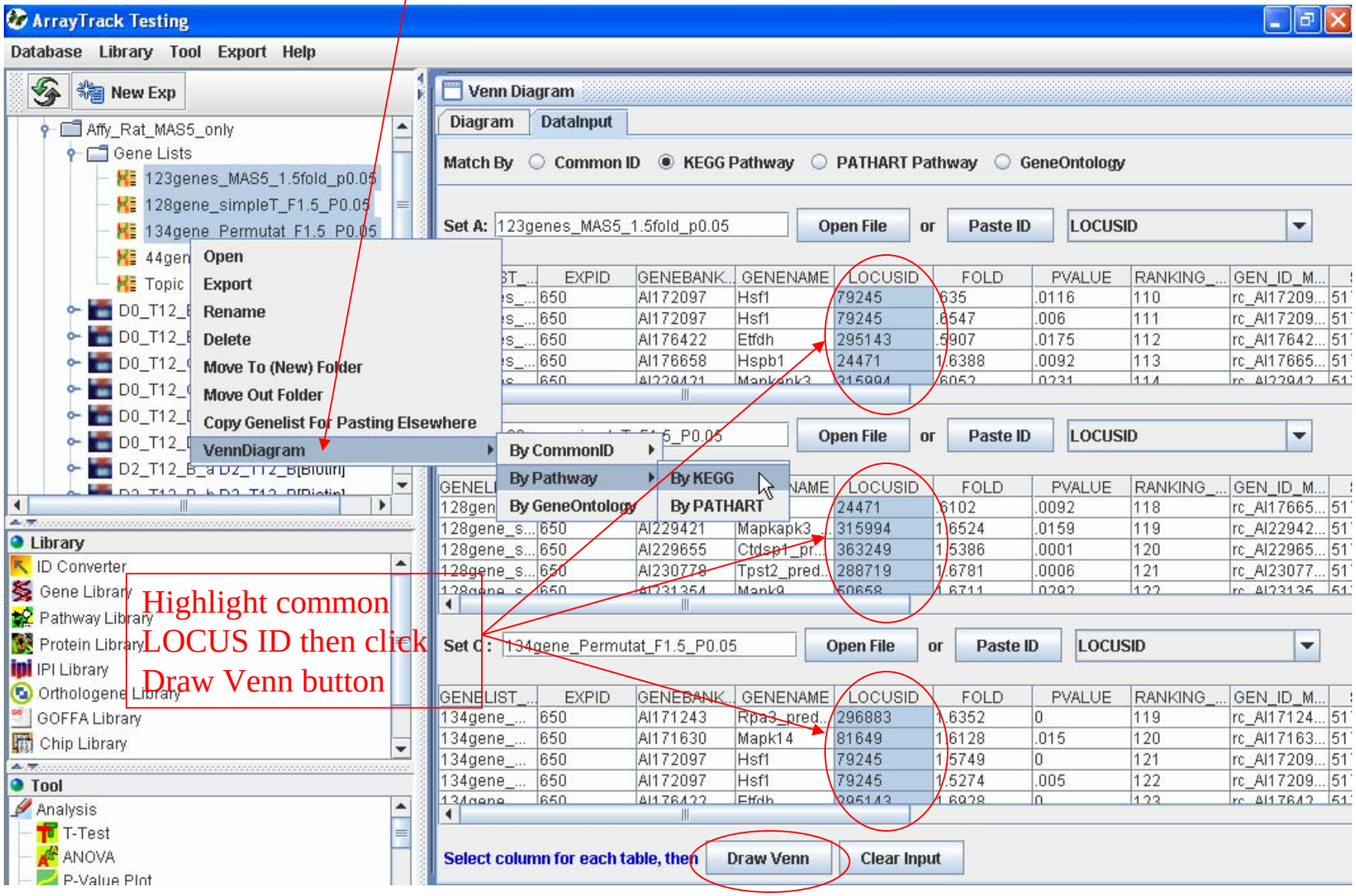
By clicking any sections from VennDiagram, the regions will be highlighted. Then right click to choose the options. One option allows display the selected genes in the original data sheet.

Common Pathways

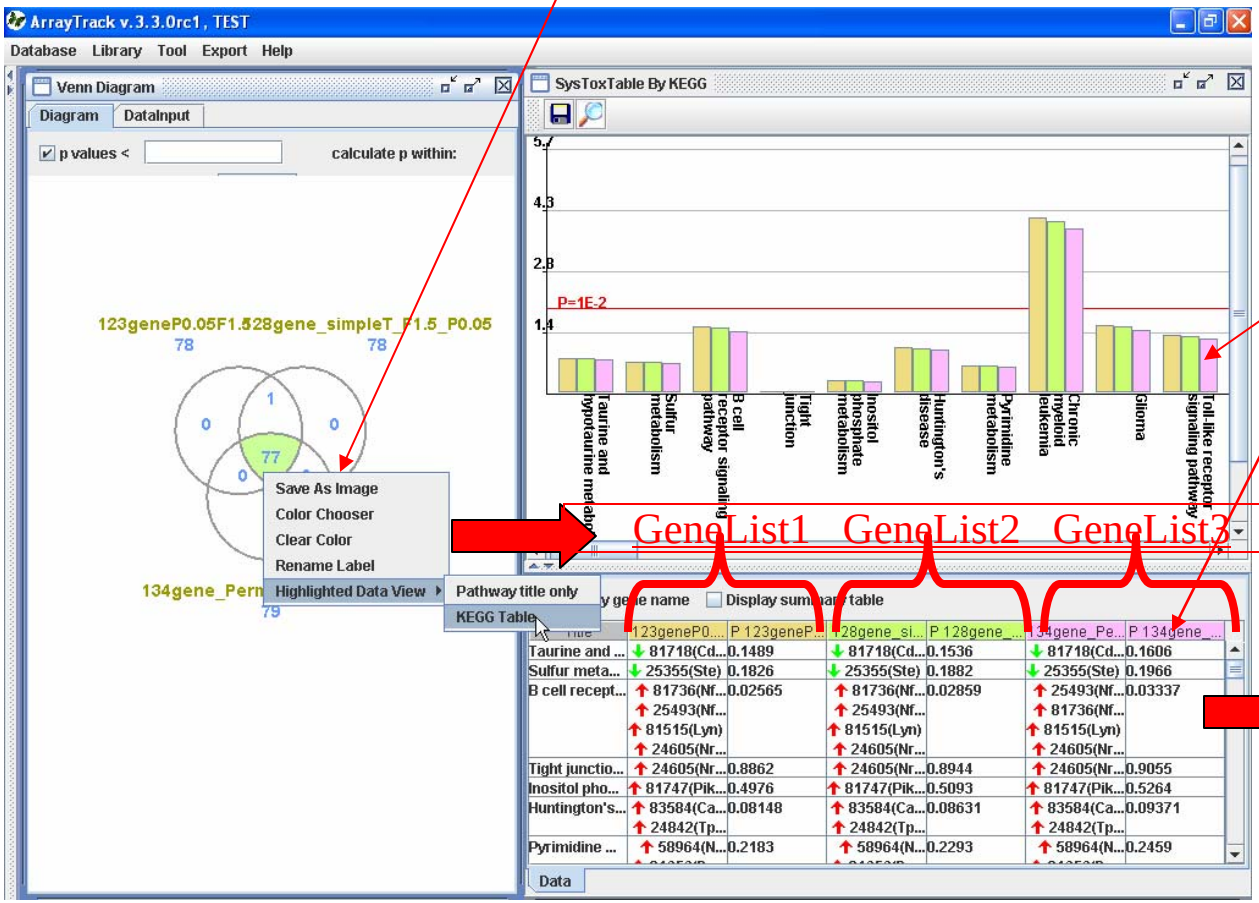
Using VennDiagram, the common pathways shared by two or three lists can be identified. The lists could be genes from microarray study, proteins from the proteomics study or metabolites from metabolomics. Please see Topic 6 for uploading the protein and metabolite list.

At this point, ArrayTrack provides KEGG and ParthArt pathways. Thus, the common pathways are separately identified for KEGG and PathArt.

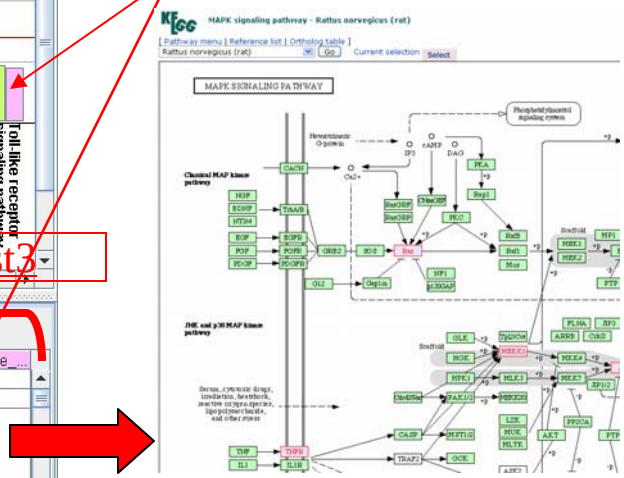
<1> Input: Right-click the selected gene lists, choose VennDiagram by Pathway/Kegg



<2> Display Pathway: Show common Pathway by right-click of the highlighted green area of the VennDiagram.

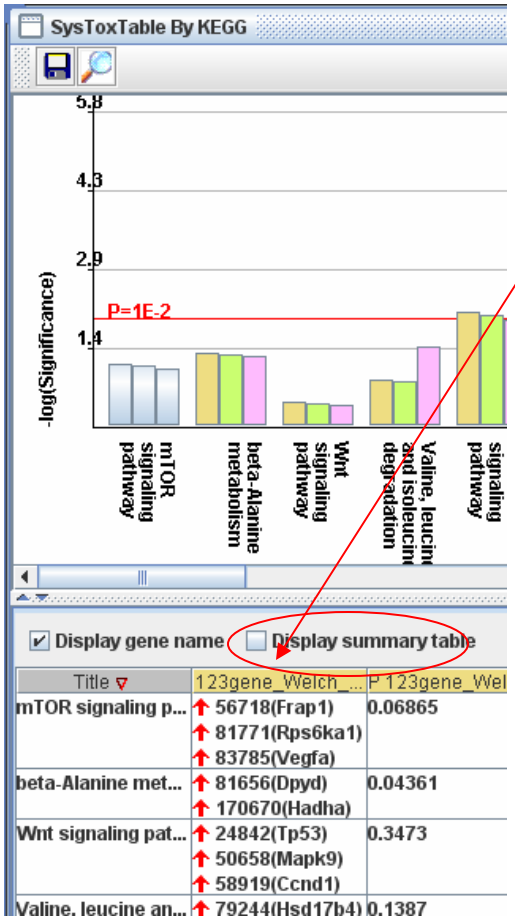


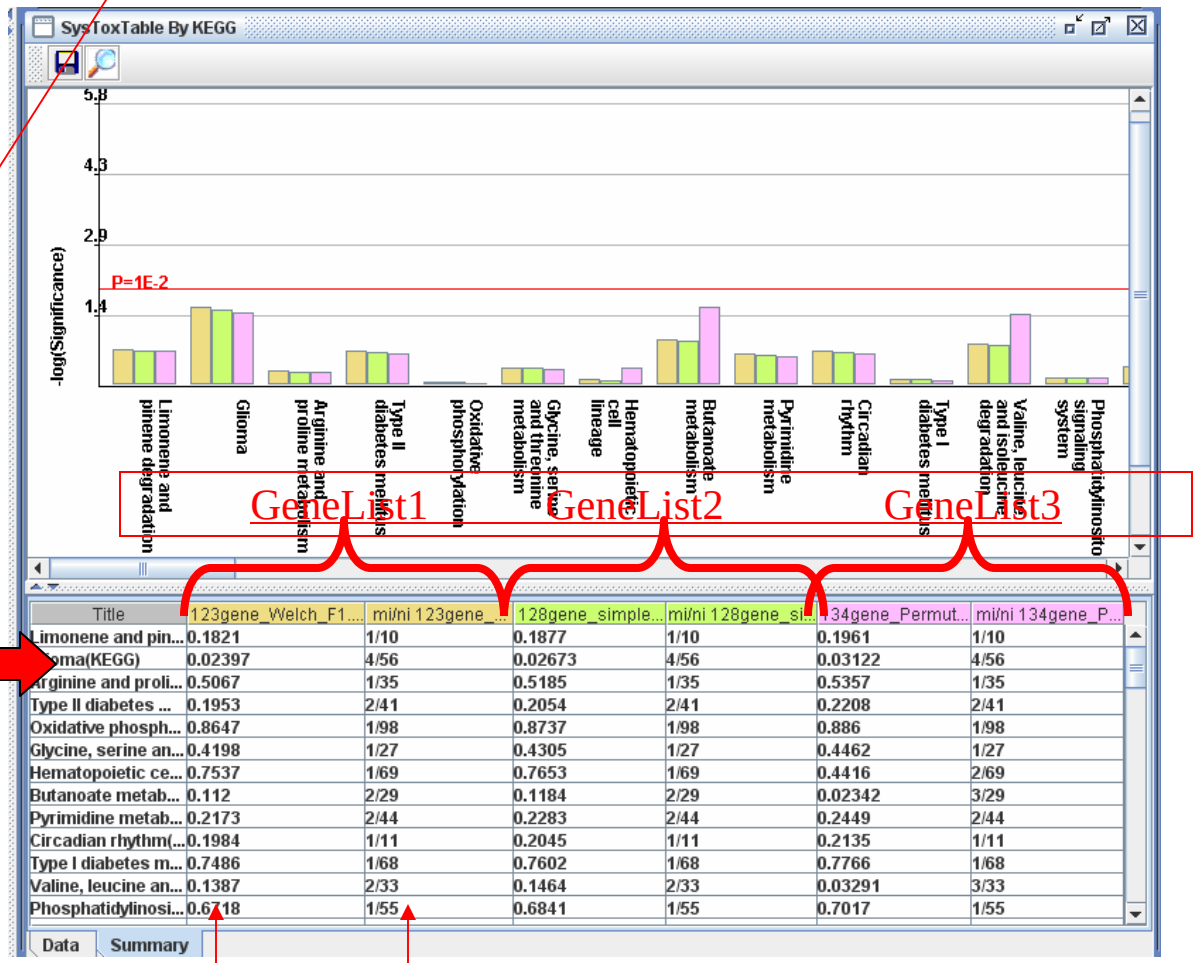
Different colors represent different gene lists.



Double click record on table to display pathway map (genes are highlighted)

<3> Display Summary: Select check box to show summary table for Common Pathway.





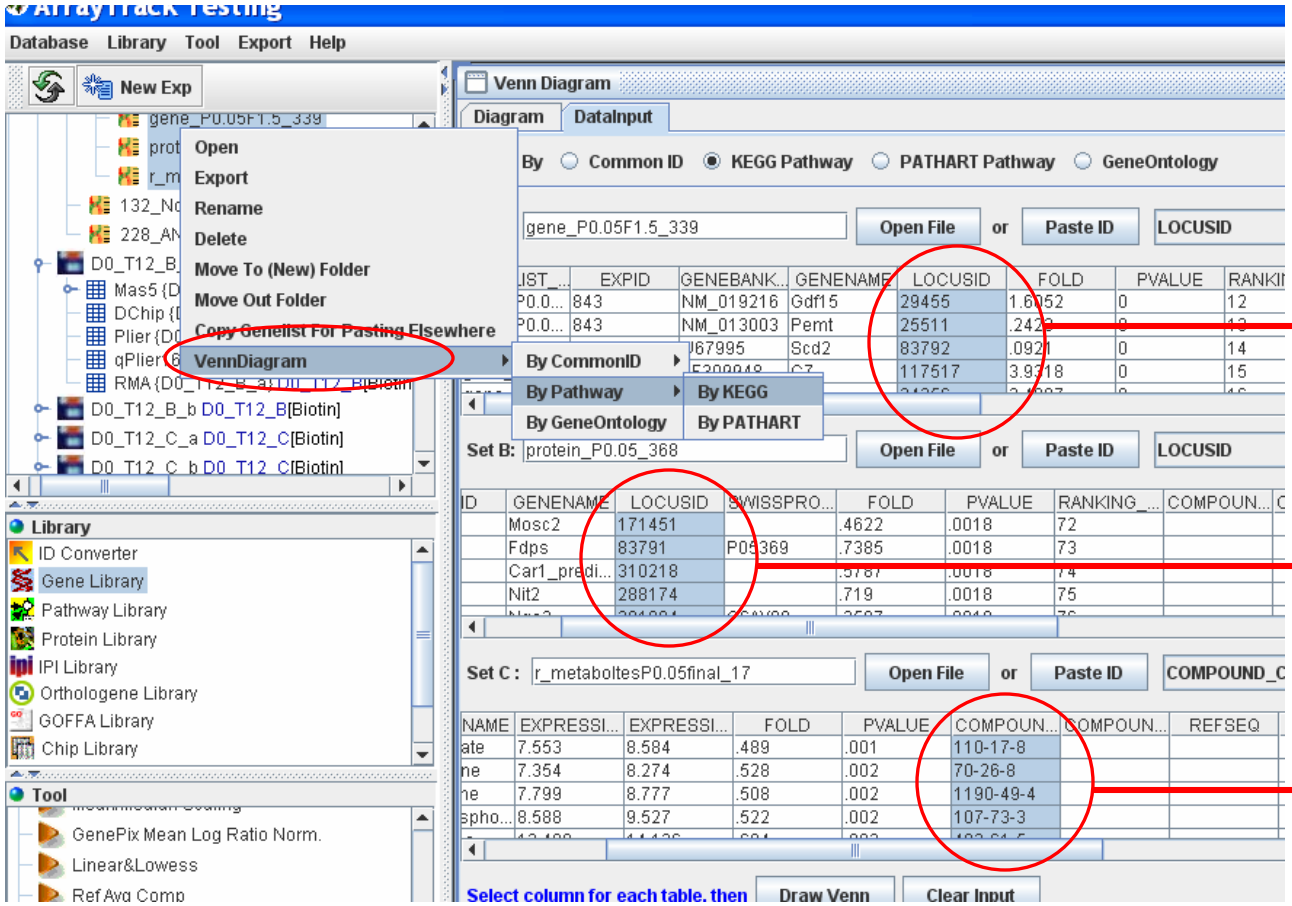
P value

1/55 : For pathway "Phosphatidylinositol signaling system",
55 genes are found in gene library for this pathway
1 gene is found in this genelist.

<4> **P value cut:** calculate p by Fisher Exact Test.

Common pathways:
Advanced: shared by the genes, proteins and metabolites

VennDiagram can also be used to find common pathway based on gene list, protein list and metabolite list. Rank of significant of P is displayed in summary table. Click individual KEGG pathway which links to KEGG website.

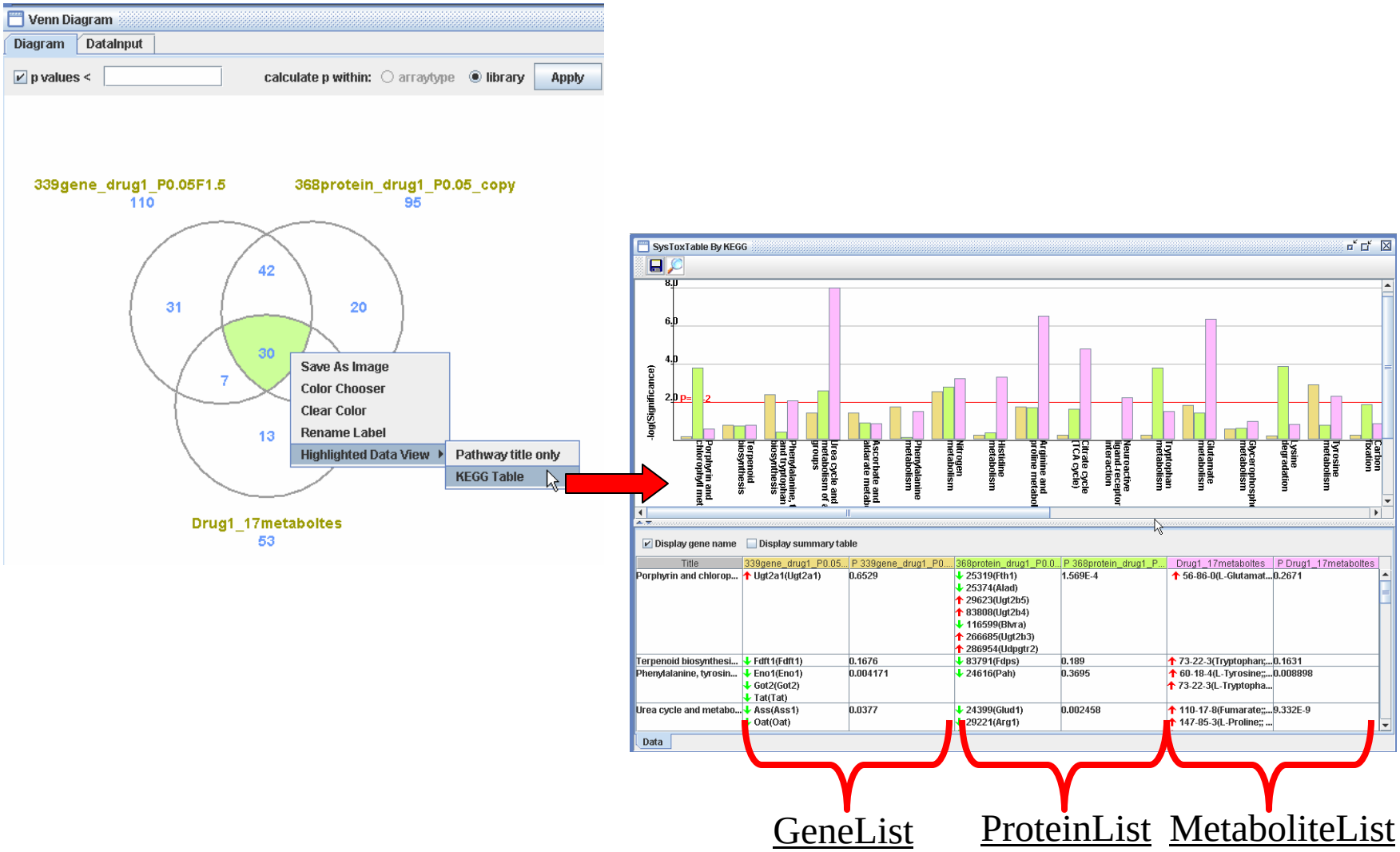


Genes
(i.e. locusid)

Proteins
(i.e. locusid or SwissProt ID)

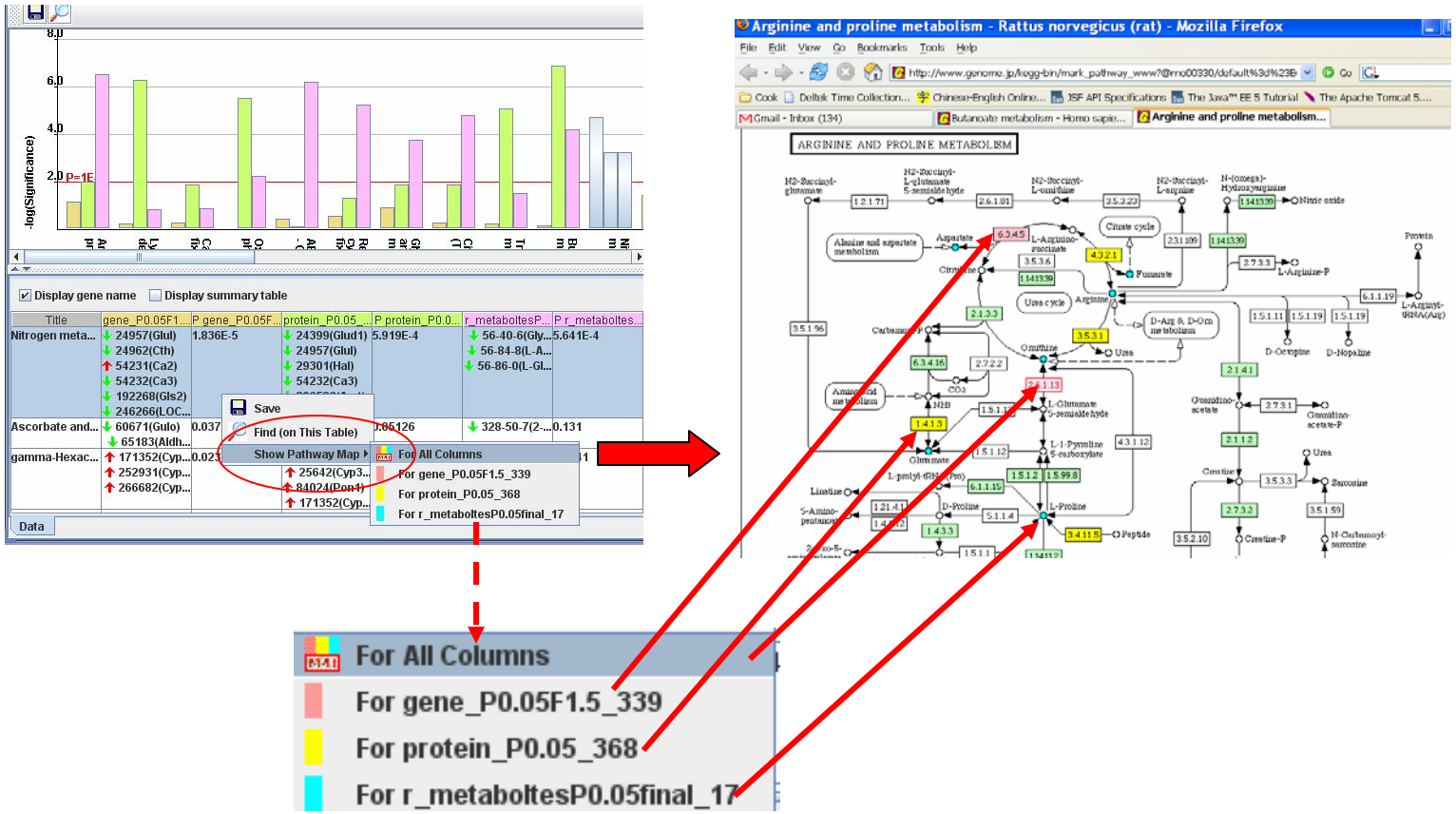
Metabolites
(i.e.Compound Cas #)

Common pathways:
Advanced: shared by the genes, proteins and metabolites



Common pathways
Advanced: shared by the genes, proteins and metabolites

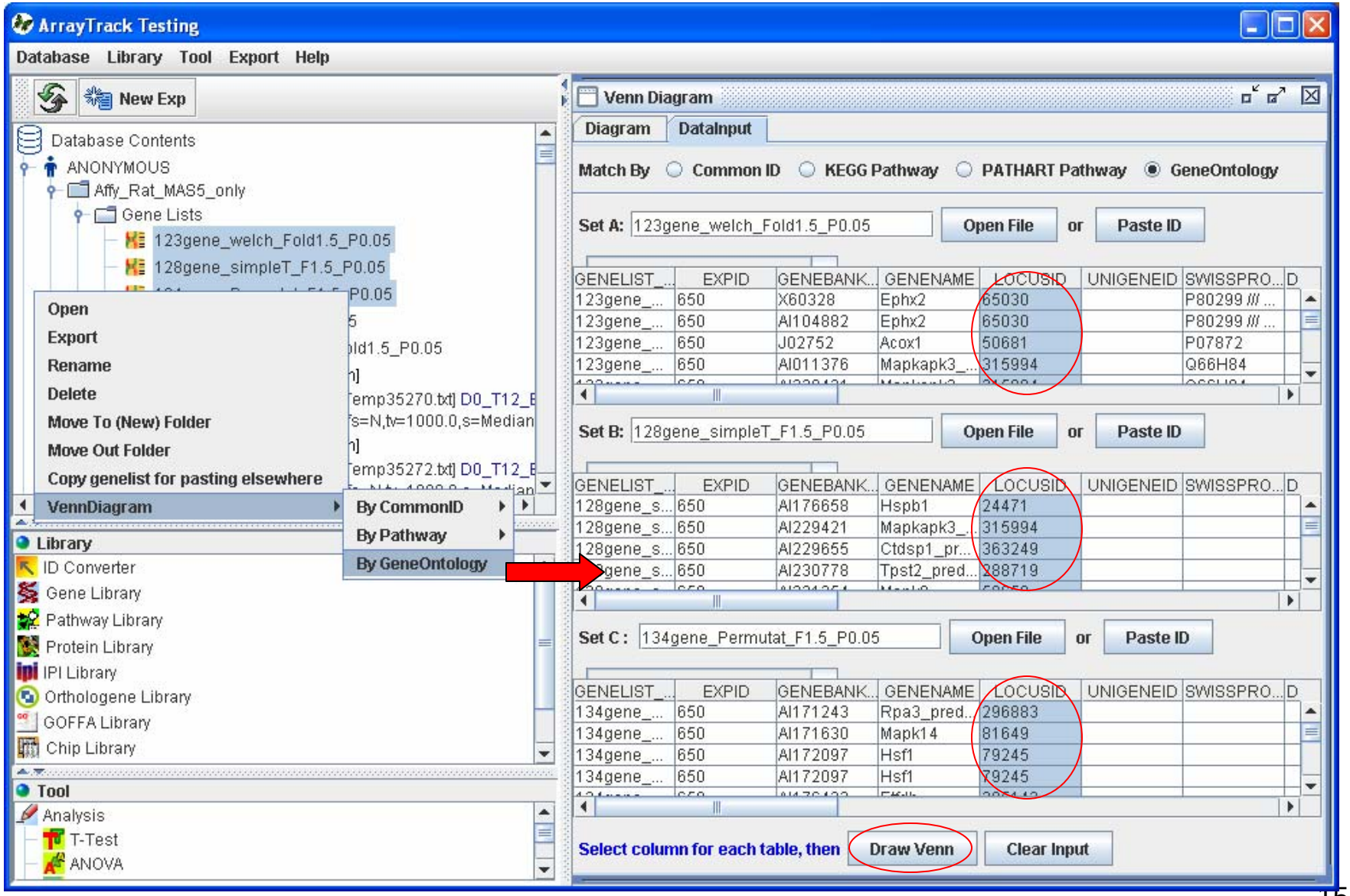
Double click a record from table, or highlight a record from table, right click and select “Show Pathway Map” - > “For All Columns”



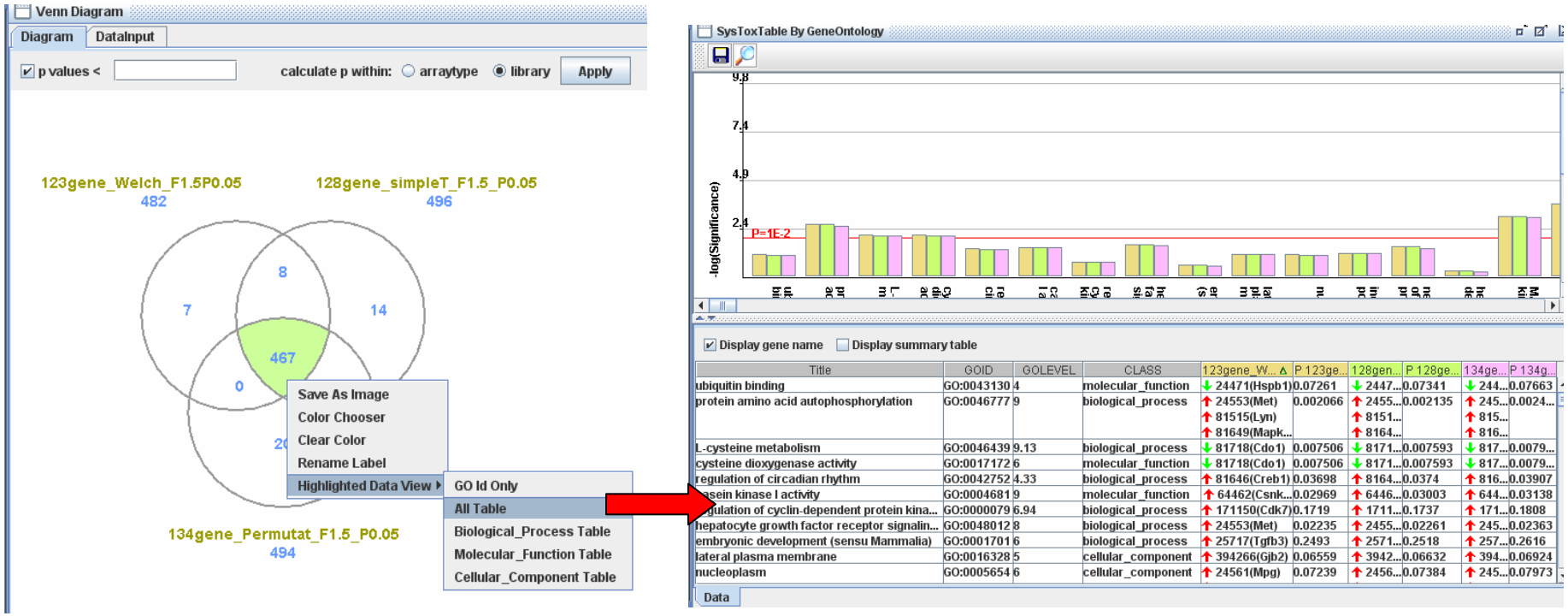
Common Gene Ontology terms

GOFFA (Gene Ontology For Functional Analysis) is a specific GO-based tool implemented in ArrayTrack. The common GeneOntology function identifies the common Gene Ontology terms shared by two or three gene lists. Identification of the common Gene Ontology terms is similar to the procedure used for determine the common pathways. The common biological process, molecular function and Cellular component terms are presented in results tables.

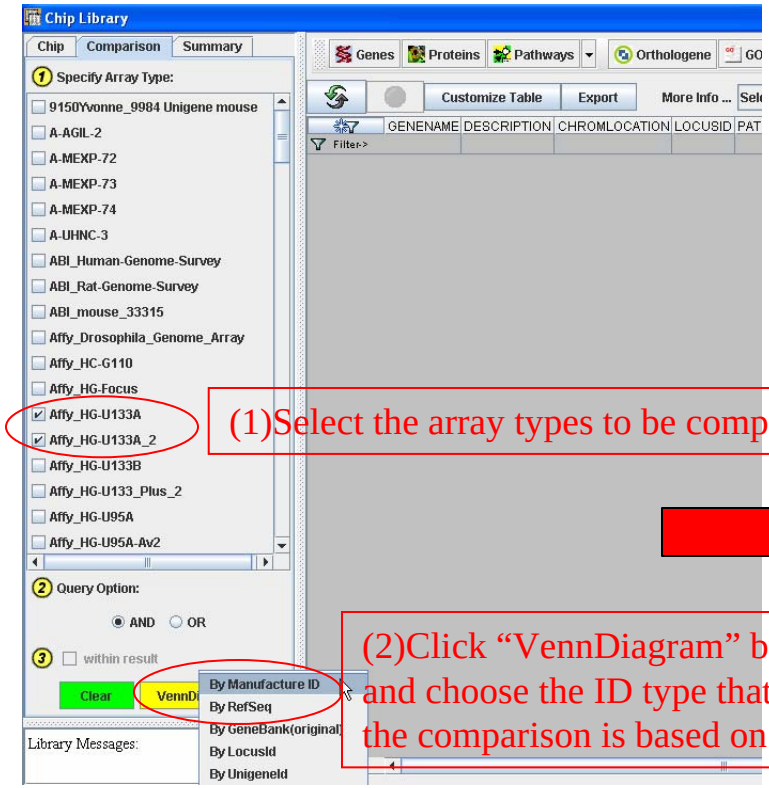
Highlight common LOCUS ID then click Draw Venn



Right click the highlighted green area, then choose “Highlighted Data View”, selecting Common GO tables (all table, Biological Process, Molecular Function and Cellular Component)



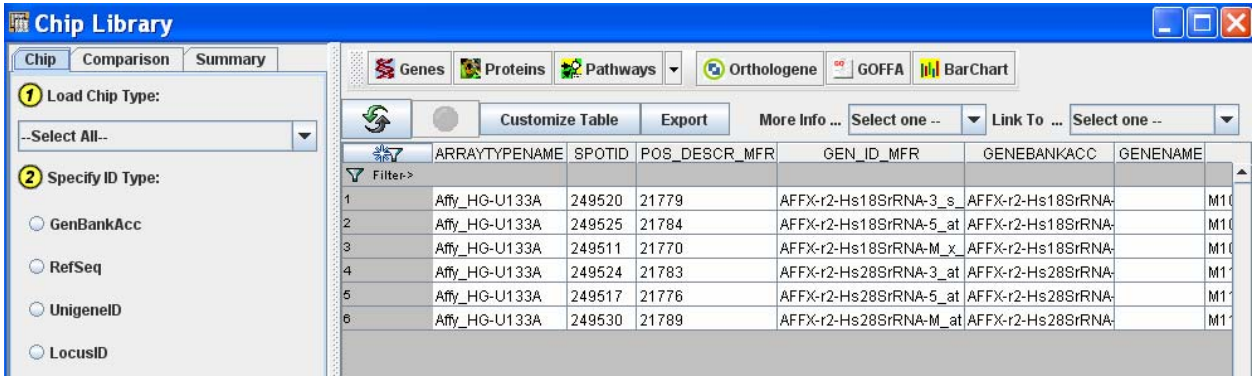
Compare two array types



(1)Select the array types to be compared

(2)Click "VennDiagram" button and choose the ID type that the comparison is based on

(3)Original data is displayed back into chip library table.



Array Track

Tutorial 4:
Data Exploring Tools

Data Exploring Tool

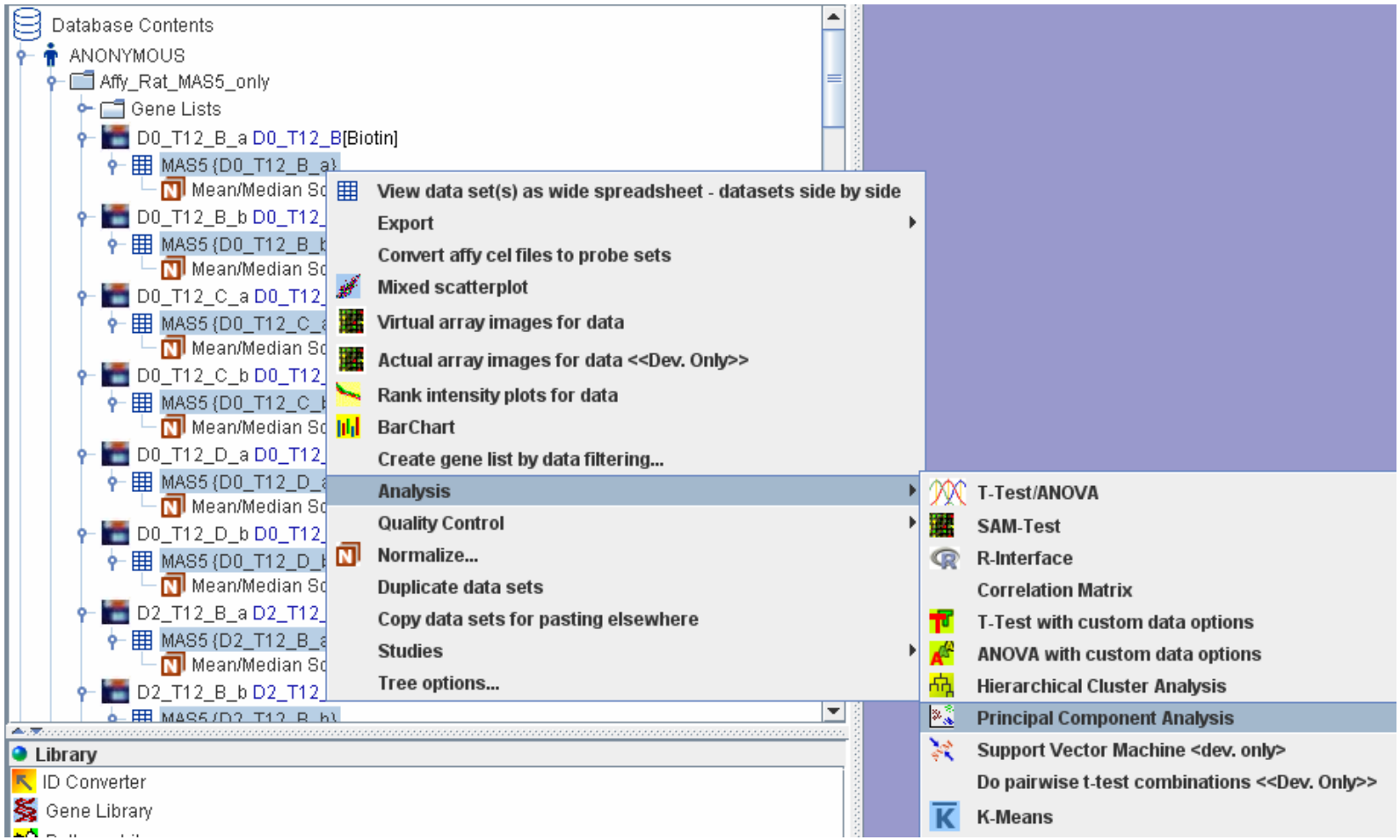
ArrayTrack provides some tools for data analysis. These tools can be accessed from the Tool panel, pull-down menu or database panel.

This tutorial will cover Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Correlation matrix, Scatter plot and K-Means.

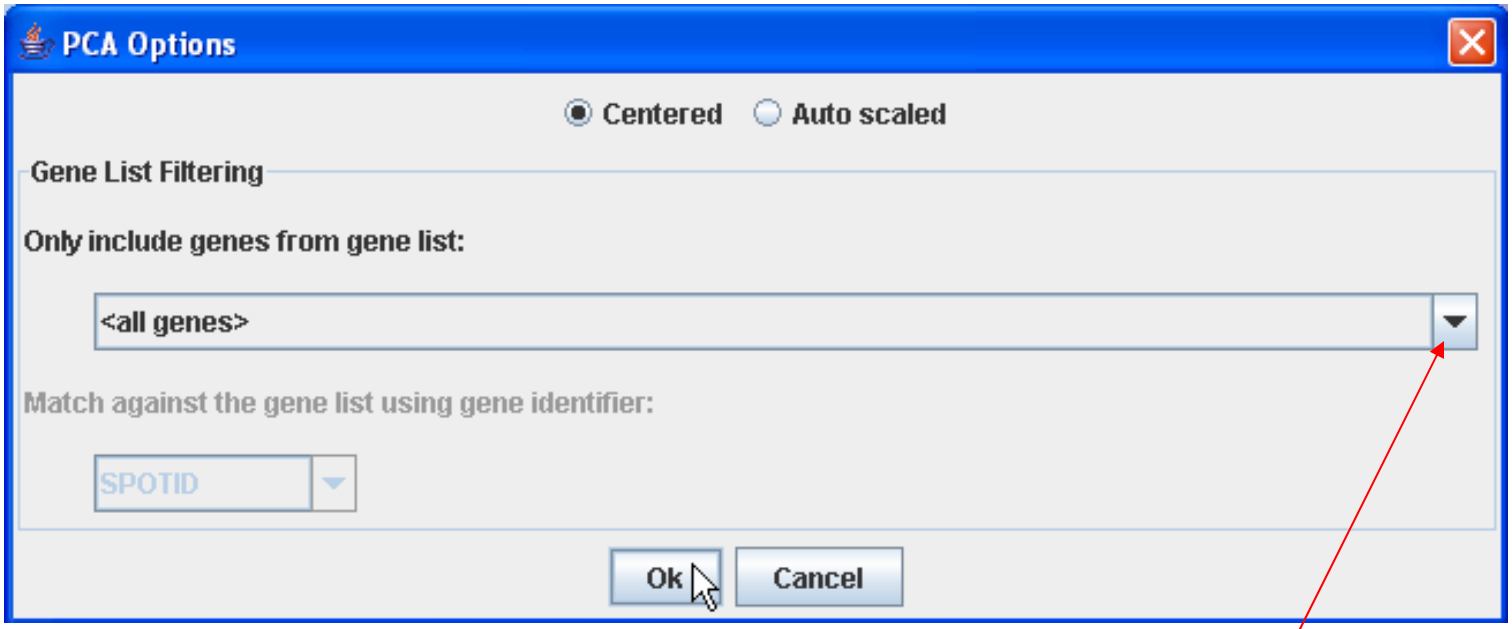
Principal Component Analysis (PCA)

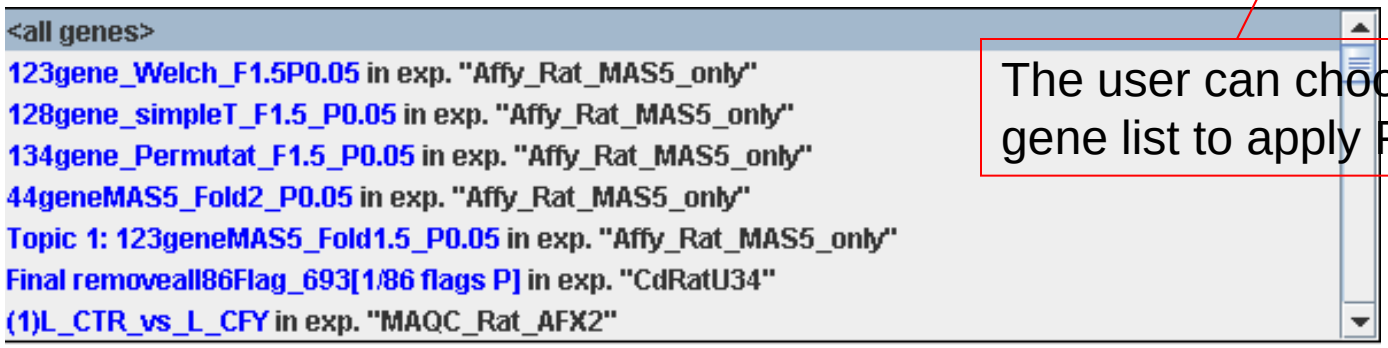
Select dataset, right-click, choose Analysis ->Principal Component Analysis

Right-click the selected data, choose “Analysis “ ->PCA



PCA - continued

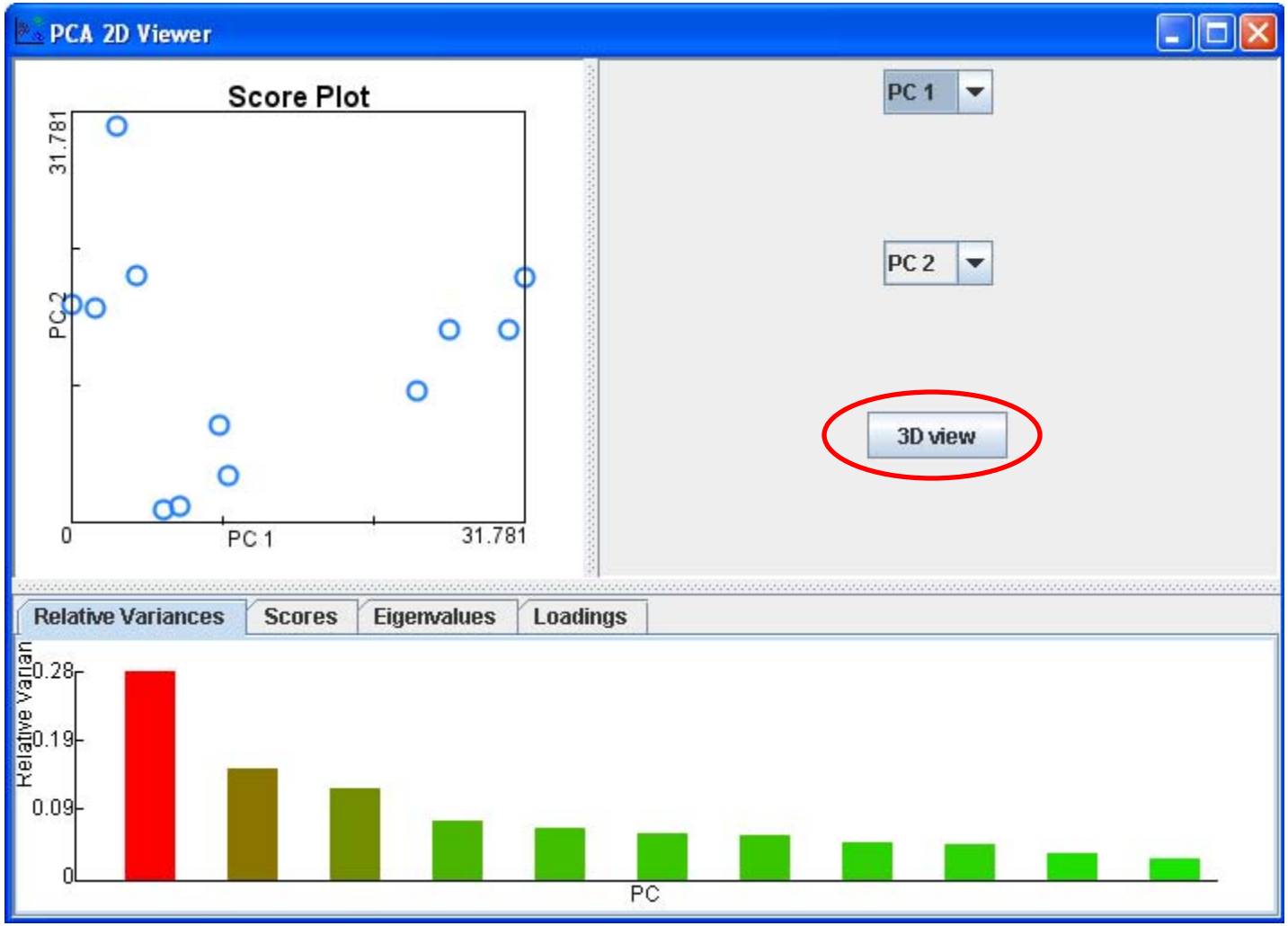




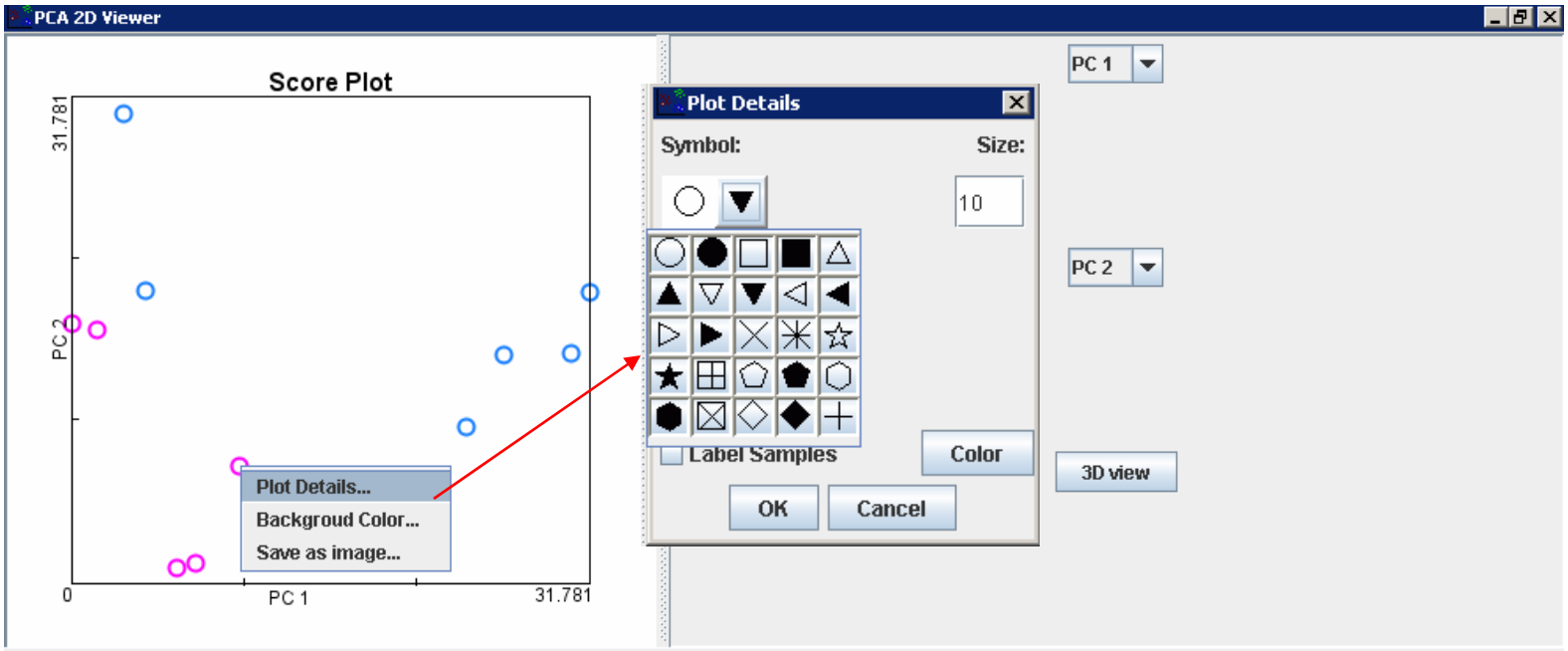
The user can choose a specific gene list to apply PCA

PCA - continued

This is the PCA 2D view. Click 3D view button will bring out the 3D image, see slide #7.



PCA - continued



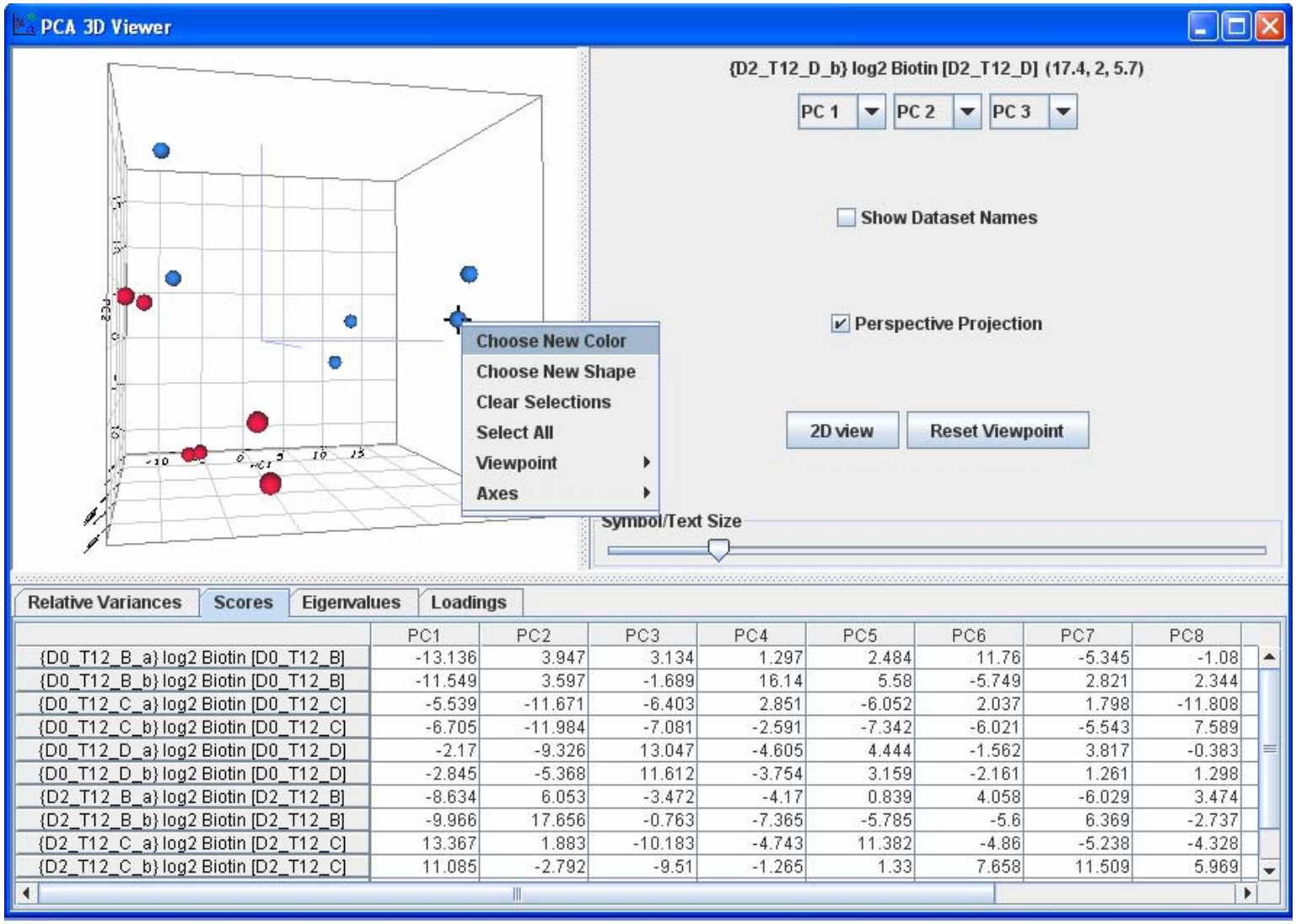
Relative Variances | Scores | Eigenvalues | Loadings

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
{D0_T12_B_a} log2 ...	-13.136	3.947	3.134	1.297	2.484	11.76	-5.345	-1.08	-8.292	0.949	-0.407
{D0_T12_B_b} log2 ...	-11.549	3.597	-1.689	16.14	5.58	-5.749	2.821	2.344	0.857	-0.197	0.441
{D0_T12_C_a} log2 ...	-5.539	-11.671	-6.403	2.851	-6.052	2.037	1.798	-11.808	3.318	0.724	-0.322
{D0_T12_C_b} log2 ...	-6.705	-11.984	-7.081	-2.591	-7.342	-6.021	-5.543	7.589	-4.763	-0.571	-0.048
{D0_T12_D_a} log2 ...	-2.17	-9.326	13.047	-4.605	4.444	-1.562	3.817	-0.383	-0.308	-2.452	7.788
{D0_T12_D_b} log2 ...	-2.845	-5.368	11.612	-3.754	3.159	-2.161	1.261	1.298	2.968	5.285	-8.243
{D2_T12_B_a} log2 ...	-8.634	6.053	-3.472	-4.17	0.839	4.058	-6.029	3.474	11.513	-3.398	1.007
{D2_T12_B_b} log2 ...	-9.966	17.656	-0.763	-7.365	-5.785	-5.6	6.369	-2.737	-2.731	-0.009	0.071
{D2_T12_C_a} log2 ...	13.367	1.883	-10.183	-4.743	11.382	-4.86	-5.238	-4.328	-2.11	2.197	0.584
{D2_T12_C_b} log2 ...	11.085	-2.792	-9.51	-1.265	1.33	7.658	11.509	5.969	-0.231	0.466	-0.541
{D2_T12_D_a} log2 ...	18.645	6.006	5.644	5.048	-7.659	0.922	-3.804	1.281	1.659	6.975	3.686
{D2_T12_D_b} log2 ...	17.447	2	5.664	3.157	-2.381	-0.482	-1.616	-1.619	-1.88	-9.969	-4.015

Under Score tab, there is the table displaying the scores for all the hybridizations. Select some Rows, the corresponding spots in the PCA plot will be highlighted. Right-click those spots Choose Plot Details, the user can define the shape of the spots and color of the spot

PCA - continued

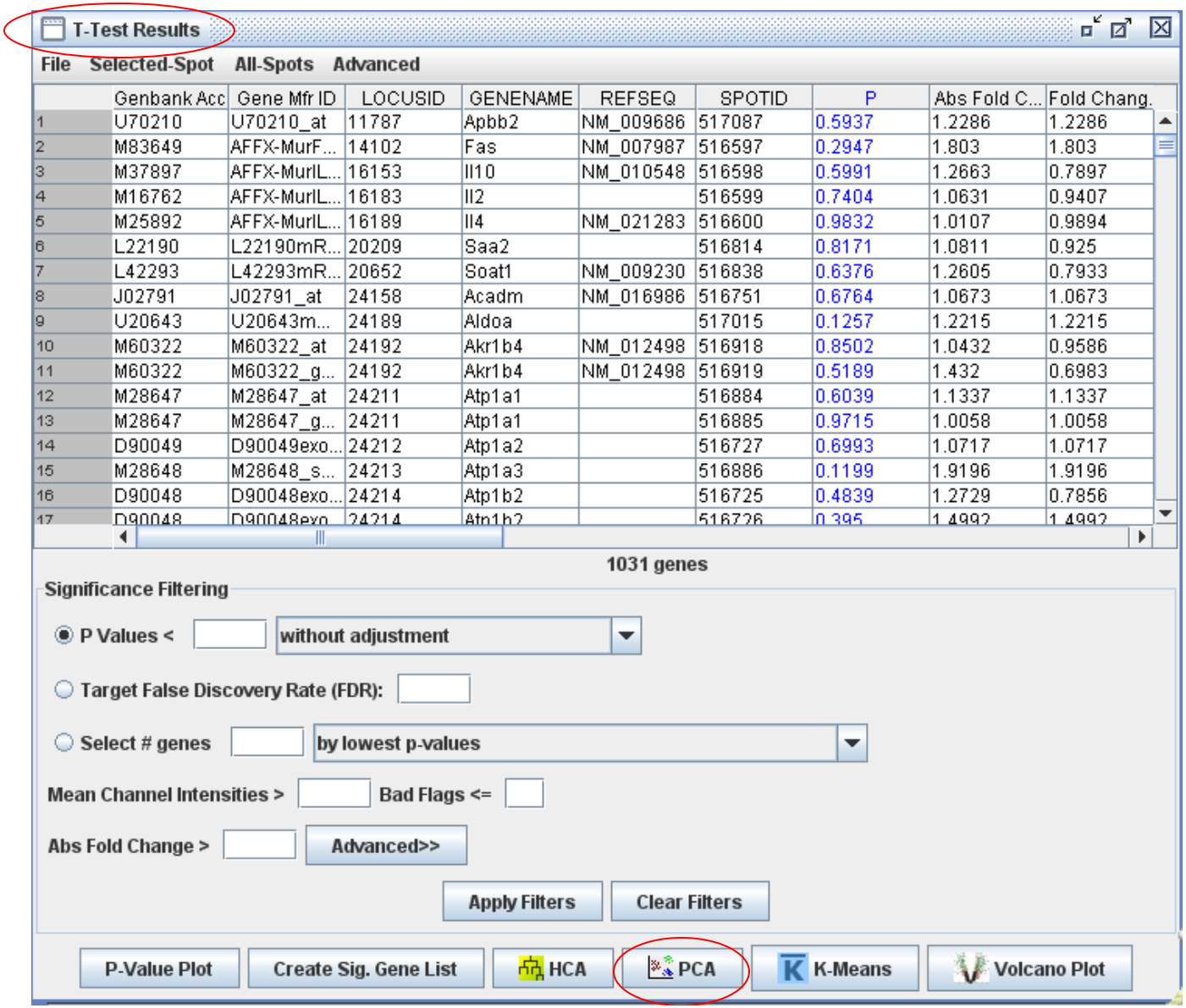
PCA 3D view



	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
{D0_T12_B_a} log2 Biotin [D0_T12_B]	-13.136	3.947	3.134	1.297	2.484	11.76	-5.345	-1.08
{D0_T12_B_b} log2 Biotin [D0_T12_B]	-11.549	3.597	-1.689	16.14	5.58	-5.749	2.821	2.344
{D0_T12_C_a} log2 Biotin [D0_T12_C]	-5.539	-11.671	-6.403	2.851	-6.052	2.037	1.798	-11.808
{D0_T12_C_b} log2 Biotin [D0_T12_C]	-6.705	-11.984	-7.081	-2.591	-7.342	-6.021	-5.543	7.589
{D0_T12_D_a} log2 Biotin [D0_T12_D]	-2.17	-9.326	13.047	-4.605	4.444	-1.562	3.817	-0.383
{D0_T12_D_b} log2 Biotin [D0_T12_D]	-2.845	-5.368	11.612	-3.754	3.159	-2.161	1.261	1.298
{D2_T12_B_a} log2 Biotin [D2_T12_B]	-8.634	6.053	-3.472	-4.17	0.839	4.058	-6.029	3.474
{D2_T12_B_b} log2 Biotin [D2_T12_B]	-9.966	17.656	-0.763	-7.365	-5.785	-5.6	6.369	-2.737
{D2_T12_C_a} log2 Biotin [D2_T12_C]	13.367	1.883	-10.183	-4.743	11.382	-4.86	-5.238	-4.328
{D2_T12_C_b} log2 Biotin [D2_T12_C]	11.085	-2.792	-9.51	-1.265	1.33	7.658	11.509	5.969

PCA - continued

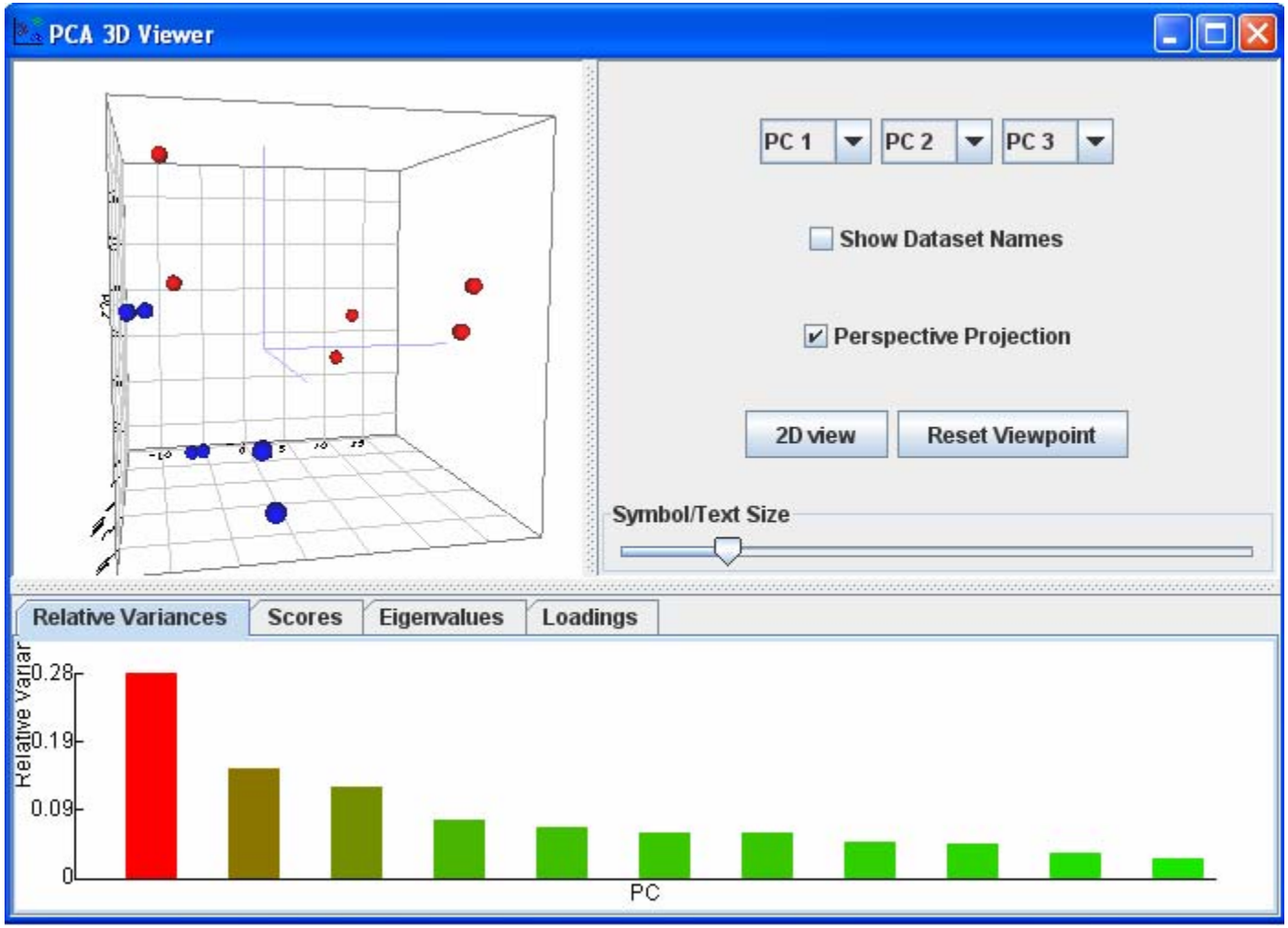
PCA can also be accessed from T-test result



	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang.
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.5937	1.2286	1.2286
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.2947	1.803	1.803
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.5991	1.2663	0.7897
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.7404	1.0631	0.9407
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.9832	1.0107	0.9894
6	L22190	L22190mR...	20209	Saa2		516814	0.8171	1.0811	0.925
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.6376	1.2605	0.7933
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.6764	1.0673	1.0673
9	U20643	U20643m...	24189	Aldoa		517015	0.1257	1.2215	1.2215
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.8502	1.0432	0.9586
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.5189	1.432	0.6983
12	M28647	M28647_at	24211	Atp1a1		516884	0.6039	1.1337	1.1337
13	M28647	M28647_g...	24211	Atp1a1		516885	0.9715	1.0058	1.0058
14	D90049	D90049exo...	24212	Atp1a2		516727	0.6993	1.0717	1.0717
15	M28648	M28648_s...	24213	Atp1a3		516886	0.1199	1.9196	1.9196
16	D90048	D90048exo...	24214	Atp1b2		516725	0.4839	1.2729	0.7856
17	D90048	D90048exo...	24214	Atp1b2		516726	0.395	1.4992	1.4992

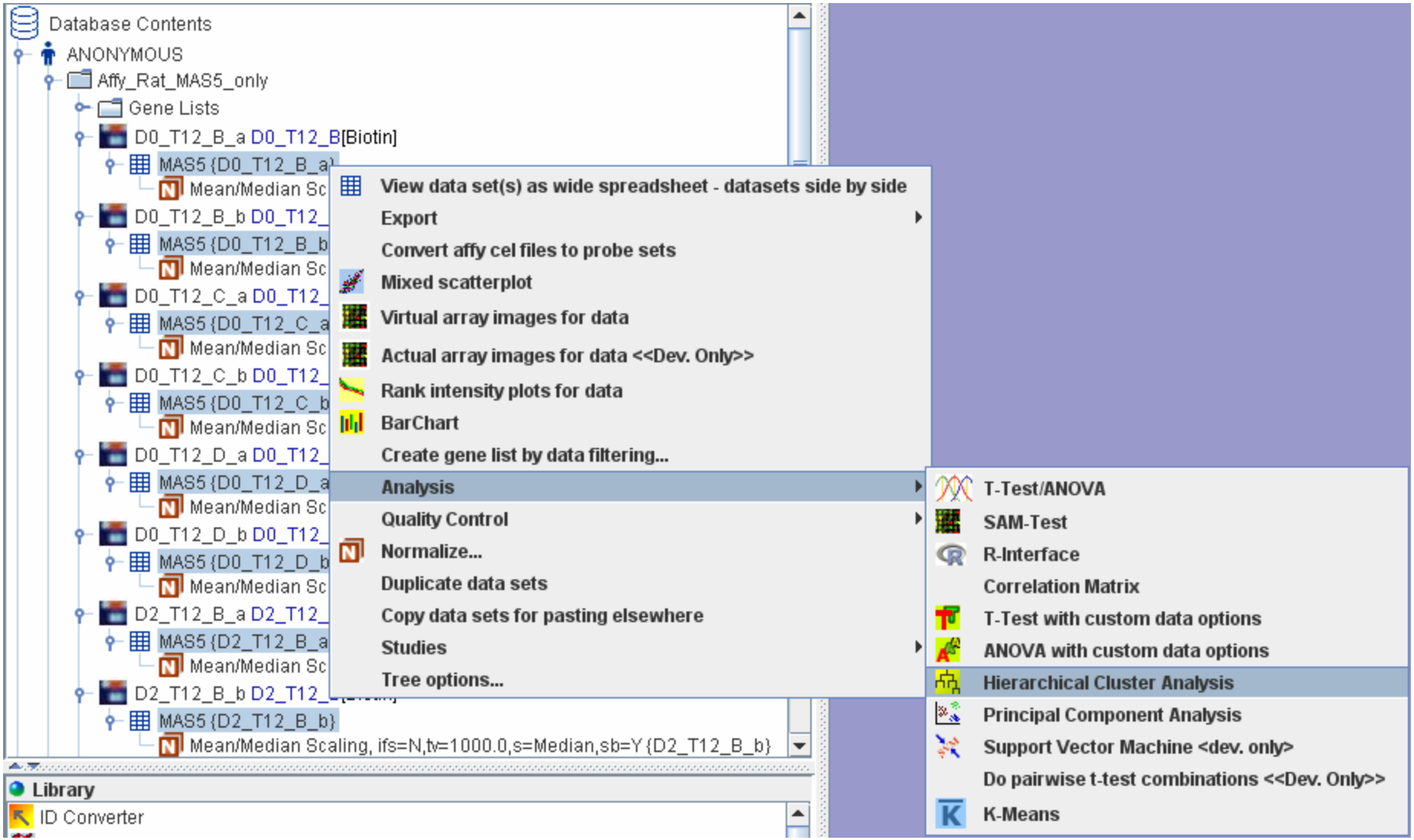
PCA - continued

PCA from T-test result. Since data are assigned in two groups in T-test, here The PCA 3D view automatically shows the two groups in two colors.

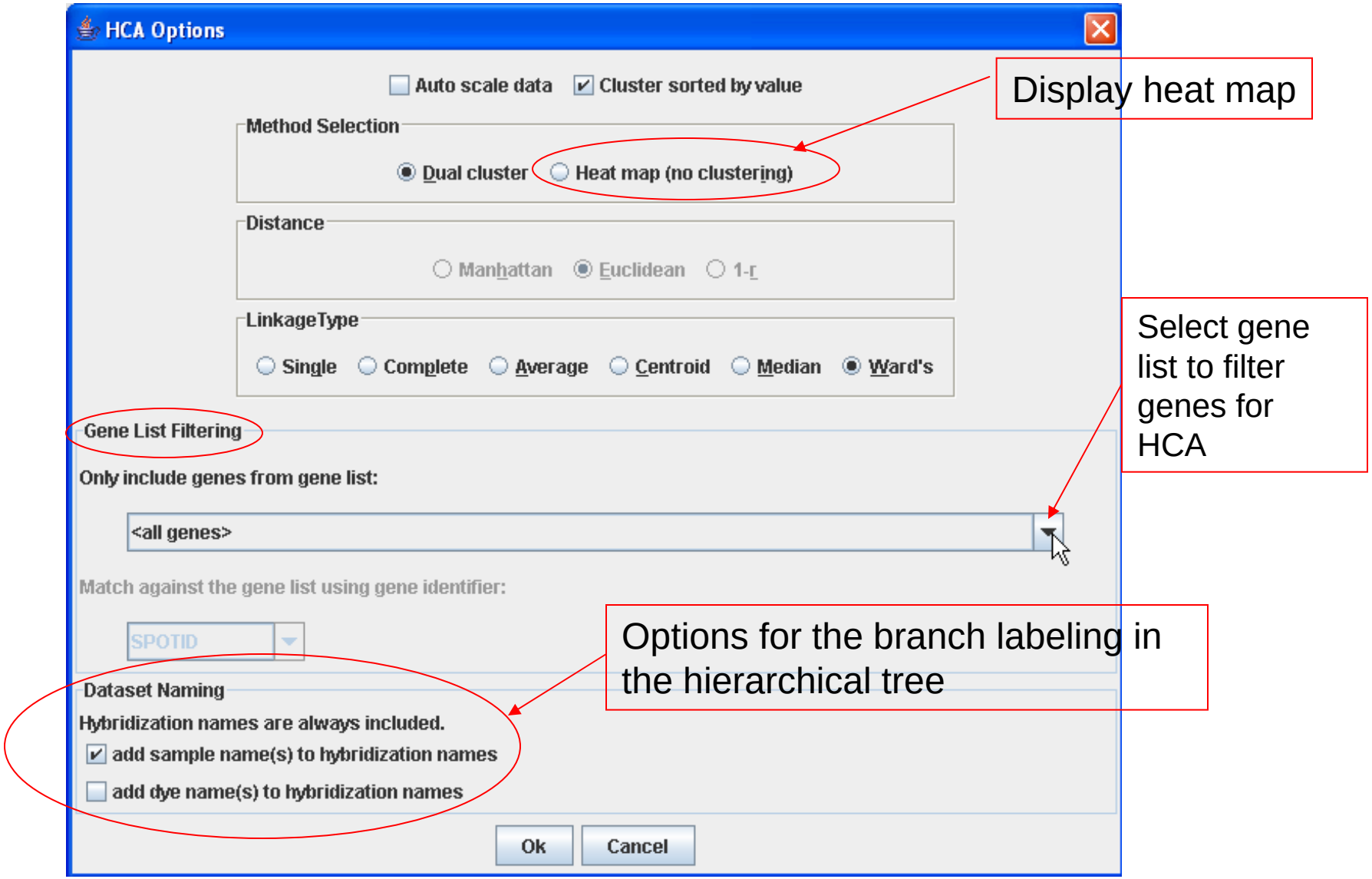


Hierarchical Cluster Analysis (HCA)

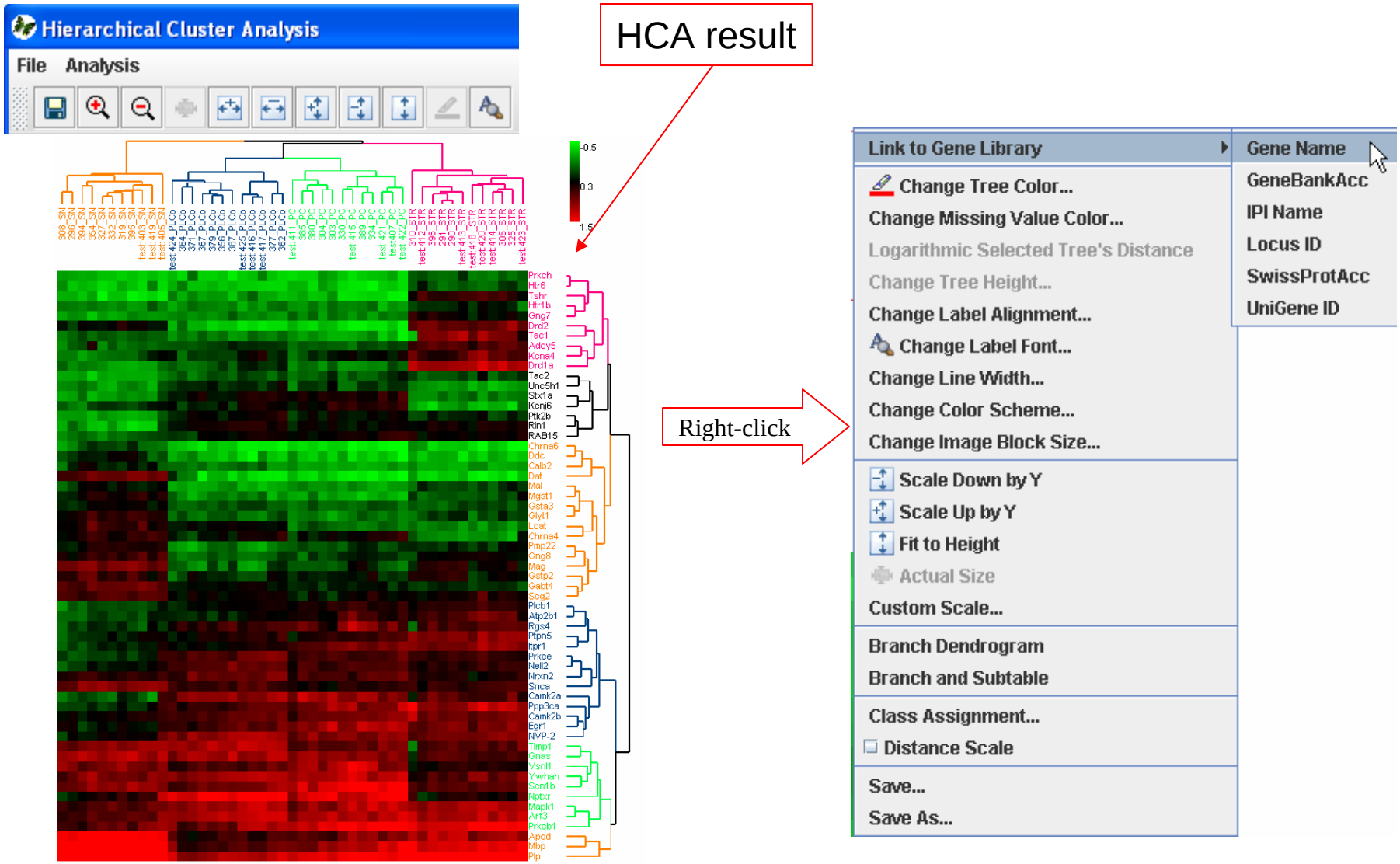
Select dataset, right-click, choose Analysis ->Hierarchical Cluster Analysis



HCA - continued

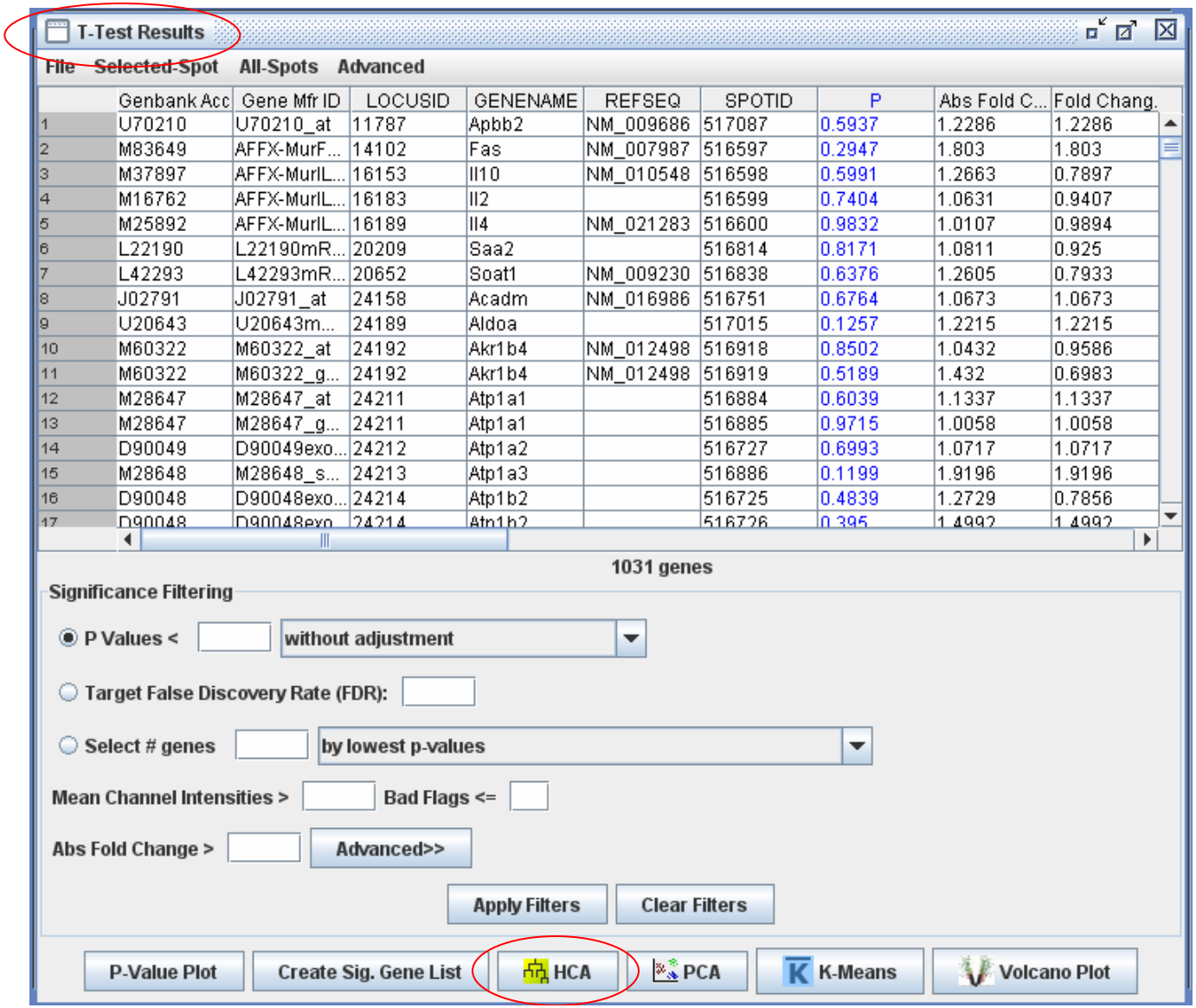


HCA - continued



HCA - continued

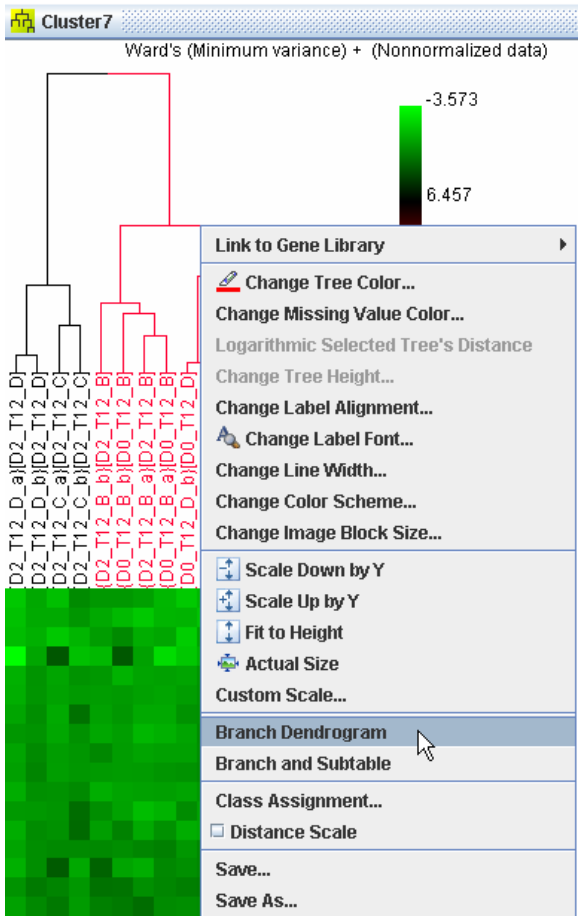
HCA can also be accessed from T-test result.

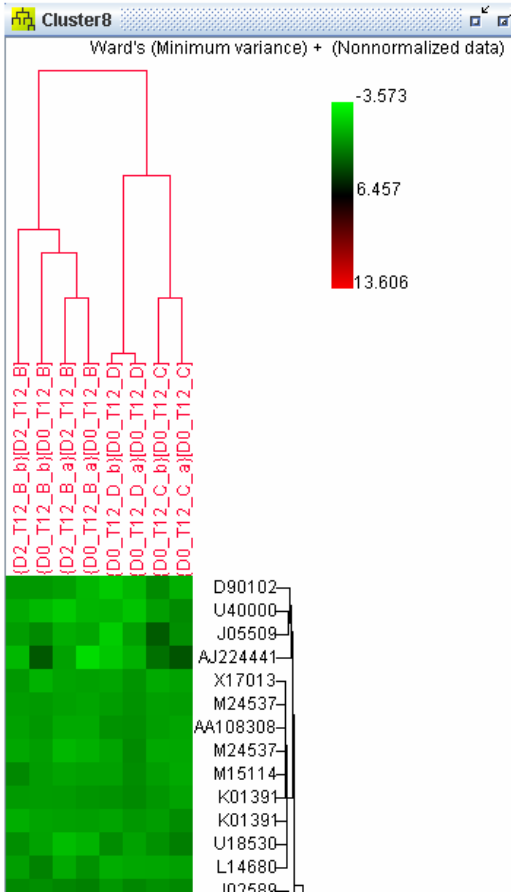


	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang.
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.5937	1.2286	1.2286
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.2947	1.803	1.803
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.5991	1.2663	0.7897
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.7404	1.0631	0.9407
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.9832	1.0107	0.9894
6	L22190	L22190mR...	20209	Saa2		516814	0.8171	1.0811	0.925
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.6376	1.2605	0.7933
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.6764	1.0673	1.0673
9	U20643	U20643m...	24189	Aldoa		517015	0.1257	1.2215	1.2215
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.8502	1.0432	0.9586
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.5189	1.432	0.6983
12	M28647	M28647_at	24211	Atp1a1		516884	0.6039	1.1337	1.1337
13	M28647	M28647_g...	24211	Atp1a1		516885	0.9715	1.0058	1.0058
14	D90049	D90049exo...	24212	Atp1a2		516727	0.6993	1.0717	1.0717
15	M28648	M28648_s...	24213	Atp1a3		516886	0.1199	1.9196	1.9196
16	D90048	D90048exo...	24214	Atp1b2		516725	0.4839	1.2729	0.7856
17	D90048	D90048exo...	24214	Atp1b2		516726	0.395	1.4992	1.4992

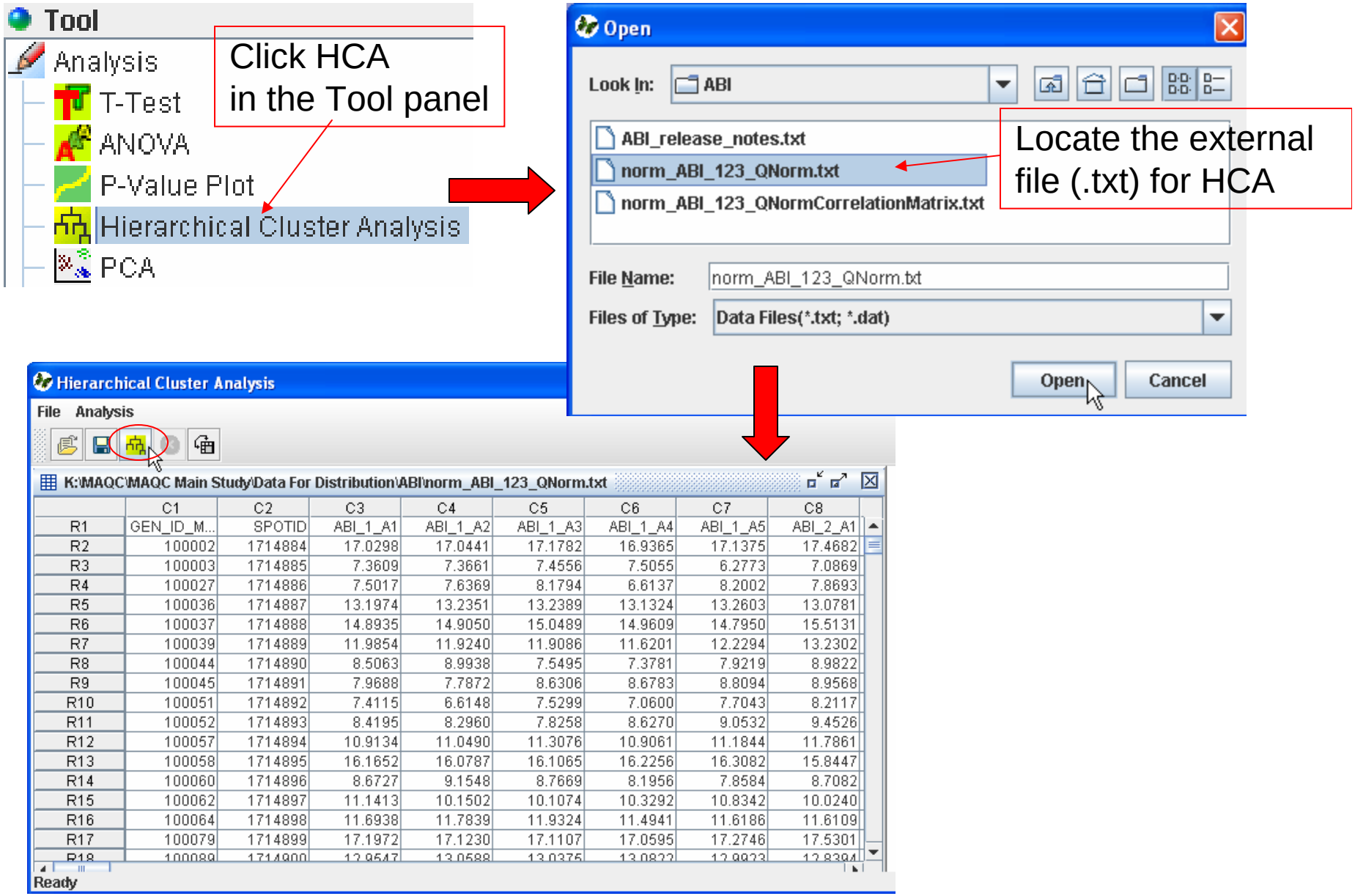
Extensive Features in HCA

- Zoom in and zoom out but clicking or
- Change the font and the color of the label for each branch of the tree.
- From the HCA plot there is a link to Gene Library according to the available IDs



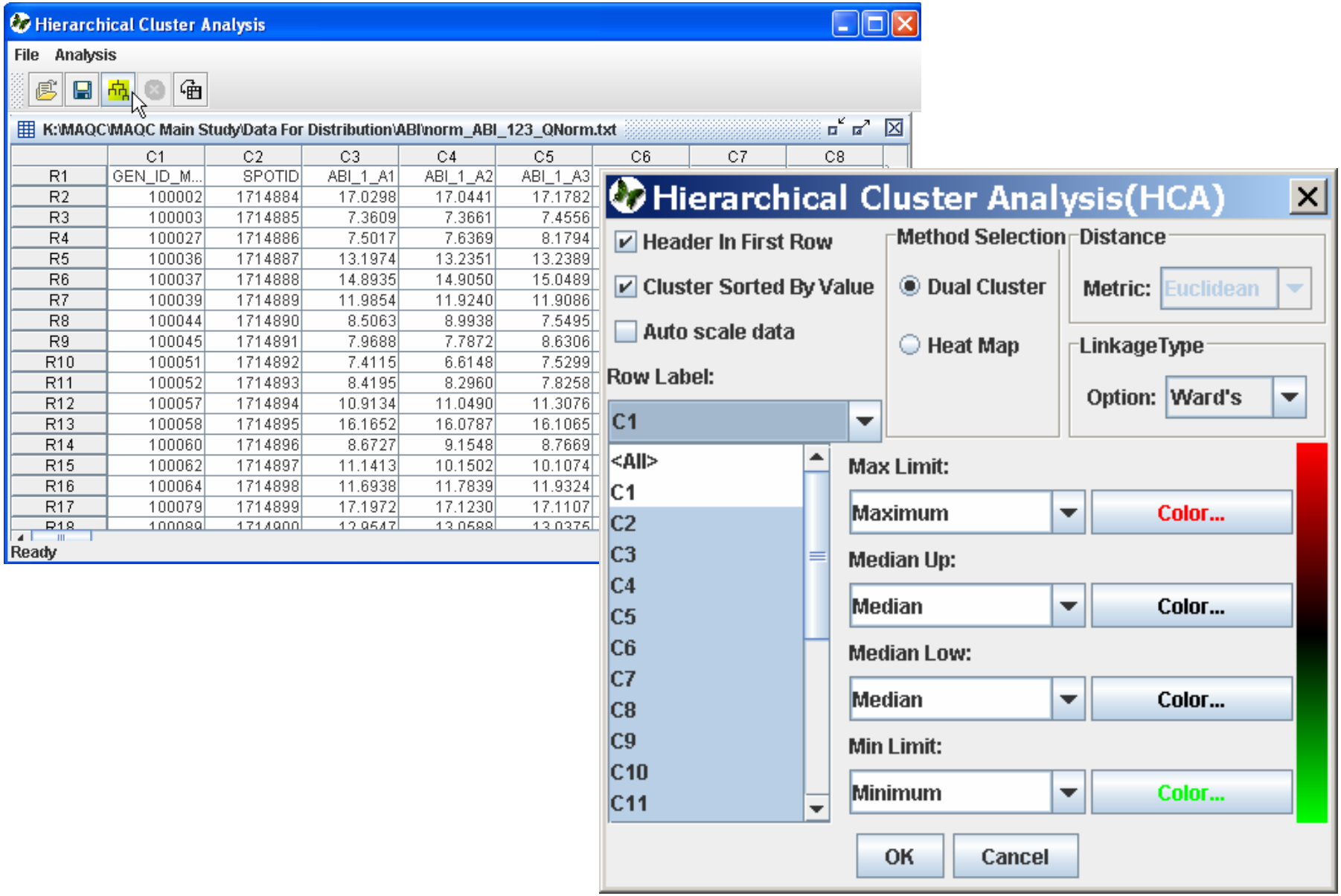


Apply HCA and PCA to the external files



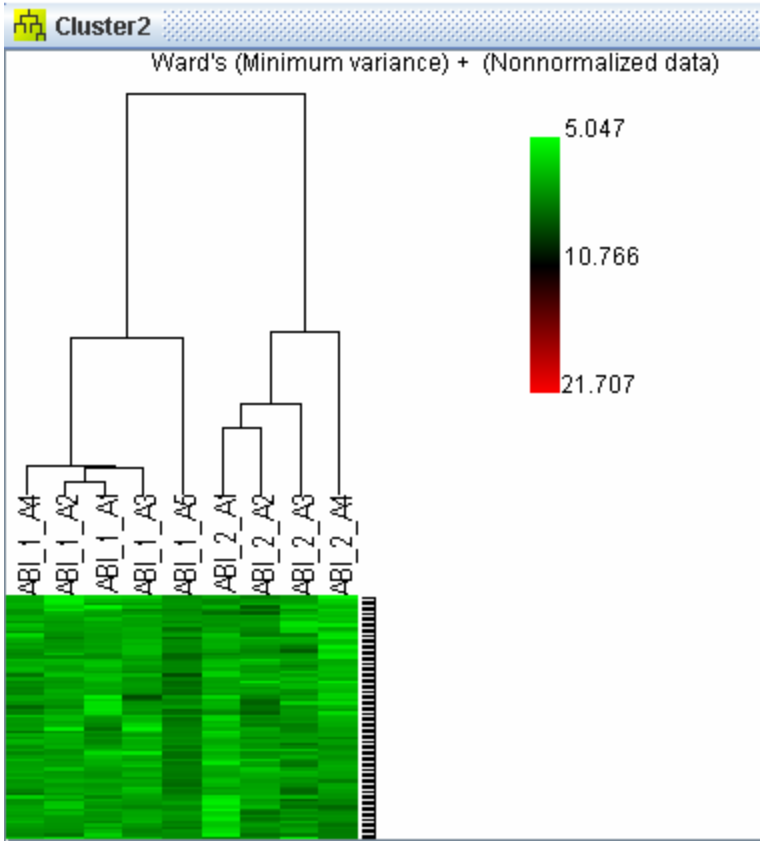
	C1	C2	C3	C4	C5	C6	C7	C8
R1	GEN_ID_M...	SPOTID	ABI_1_A1	ABI_1_A2	ABI_1_A3	ABI_1_A4	ABI_1_A5	ABI_2_A1
R2	100002	1714884	17.0298	17.0441	17.1782	16.9365	17.1375	17.4682
R3	100003	1714885	7.3609	7.3661	7.4556	7.5055	6.2773	7.0869
R4	100027	1714886	7.5017	7.6369	8.1794	6.6137	8.2002	7.8693
R5	100036	1714887	13.1974	13.2351	13.2389	13.1324	13.2603	13.0781
R6	100037	1714888	14.8935	14.9050	15.0489	14.9609	14.7950	15.5131
R7	100039	1714889	11.9854	11.9240	11.9086	11.6201	12.2294	13.2302
R8	100044	1714890	8.5063	8.9938	7.5495	7.3781	7.9219	8.9822
R9	100045	1714891	7.9688	7.7872	8.6306	8.6783	8.8094	8.9568
R10	100051	1714892	7.4115	6.6148	7.5299	7.0600	7.7043	8.2117
R11	100052	1714893	8.4195	8.2960	7.8258	8.6270	9.0532	9.4526
R12	100057	1714894	10.9134	11.0490	11.3076	10.9061	11.1844	11.7861
R13	100058	1714895	16.1652	16.0787	16.1065	16.2256	16.3082	15.8447
R14	100060	1714896	8.6727	9.1548	8.7669	8.1956	7.8584	8.7082
R15	100062	1714897	11.1413	10.1502	10.1074	10.3292	10.8342	10.0240
R16	100064	1714898	11.6938	11.7839	11.9324	11.4941	11.6186	11.6109
R17	100079	1714899	17.1972	17.1230	17.1107	17.0595	17.2746	17.5301
R18	100089	1714900	12.9547	13.0588	13.0375	13.0822	12.9923	12.8394

Apply HCA and PCA to the external files – continued



	C1	C2	C3	C4	C5	C6	C7	C8
R1	GEN_ID_M...	SPOTID	ABI_1_A1	ABI_1_A2	ABI_1_A3			
R2	100002	1714884	17.0298	17.0441	17.1782			
R3	100003	1714885	7.3609	7.3661	7.4556			
R4	100027	1714886	7.5017	7.6369	8.1794			
R5	100036	1714887	13.1974	13.2351	13.2389			
R6	100037	1714888	14.8935	14.9050	15.0489			
R7	100039	1714889	11.9854	11.9240	11.9086			
R8	100044	1714890	8.5063	8.9938	7.5495			
R9	100045	1714891	7.9688	7.7872	8.6306			
R10	100051	1714892	7.4115	6.6148	7.5299			
R11	100052	1714893	8.4195	8.2960	7.8258			
R12	100057	1714894	10.9134	11.0490	11.3076			
R13	100058	1714895	16.1652	16.0787	16.1065			
R14	100060	1714896	8.6727	9.1548	8.7669			
R15	100062	1714897	11.1413	10.1502	10.1074			
R16	100064	1714898	11.6938	11.7839	11.9324			
R17	100079	1714899	17.1972	17.1230	17.1107			
R18	100089	1714900	12.9547	13.0588	13.0375			

Apply HCA and PCA to the external files – continued

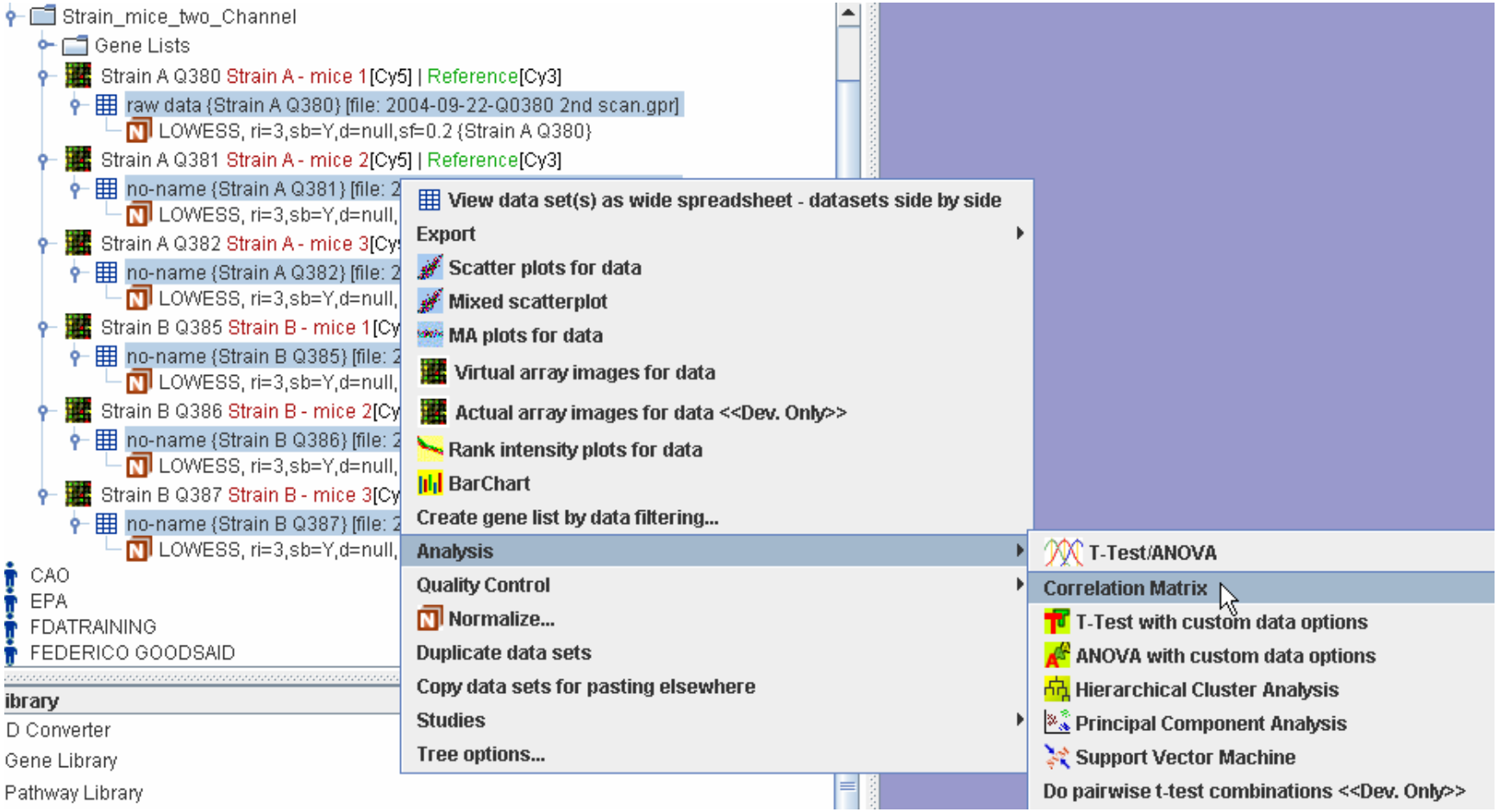


The procedure for applying PCA to the external files is same as applying HCA.

Correlation Matrix

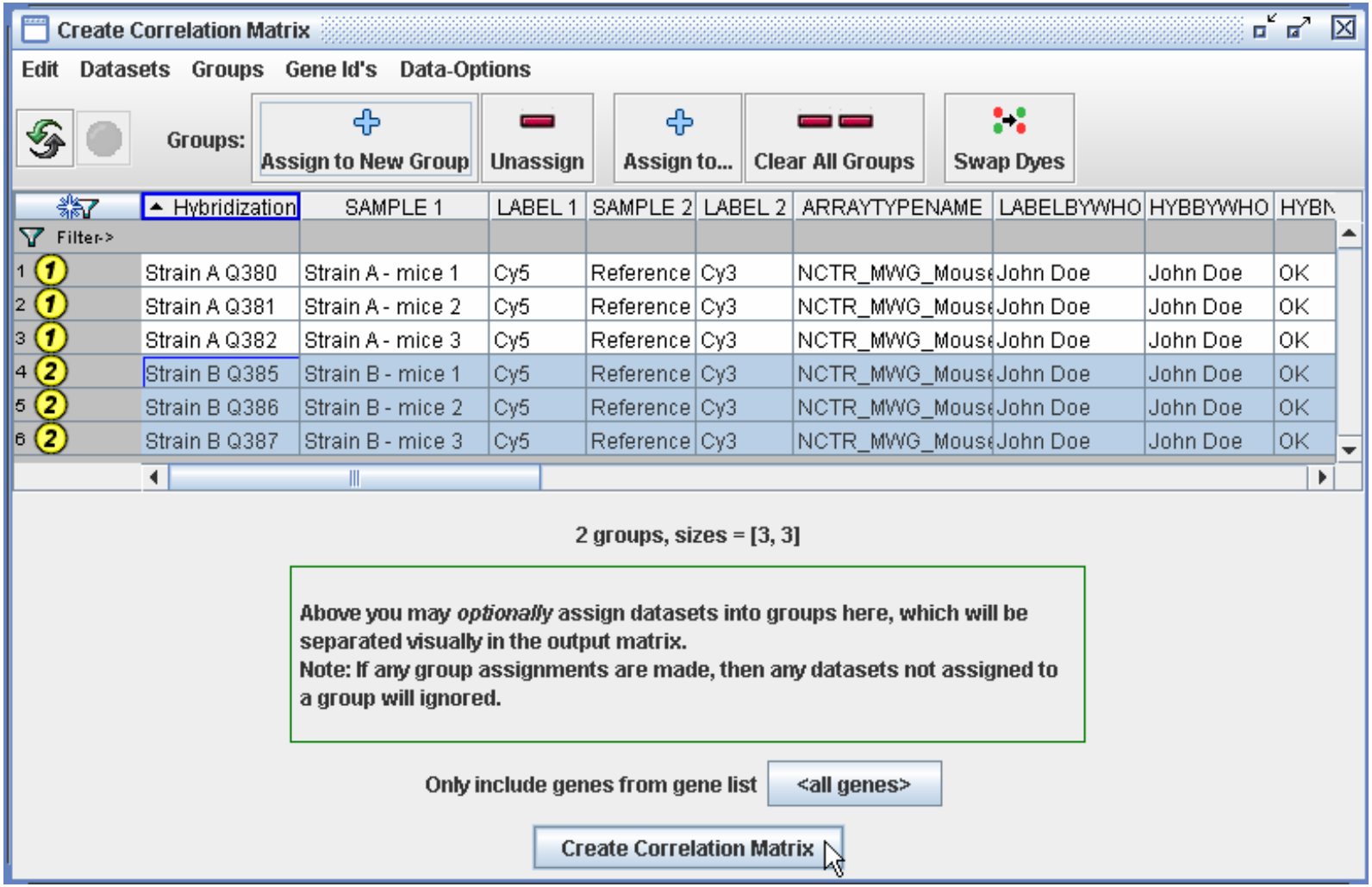
The correlation matrix function shows the correlation between column i and column j of the original matrix. It is used to find out the correlation between two groups of data in ArrayTrack.

Select data set, right-click, choose Analysis -> Correlation Matrix



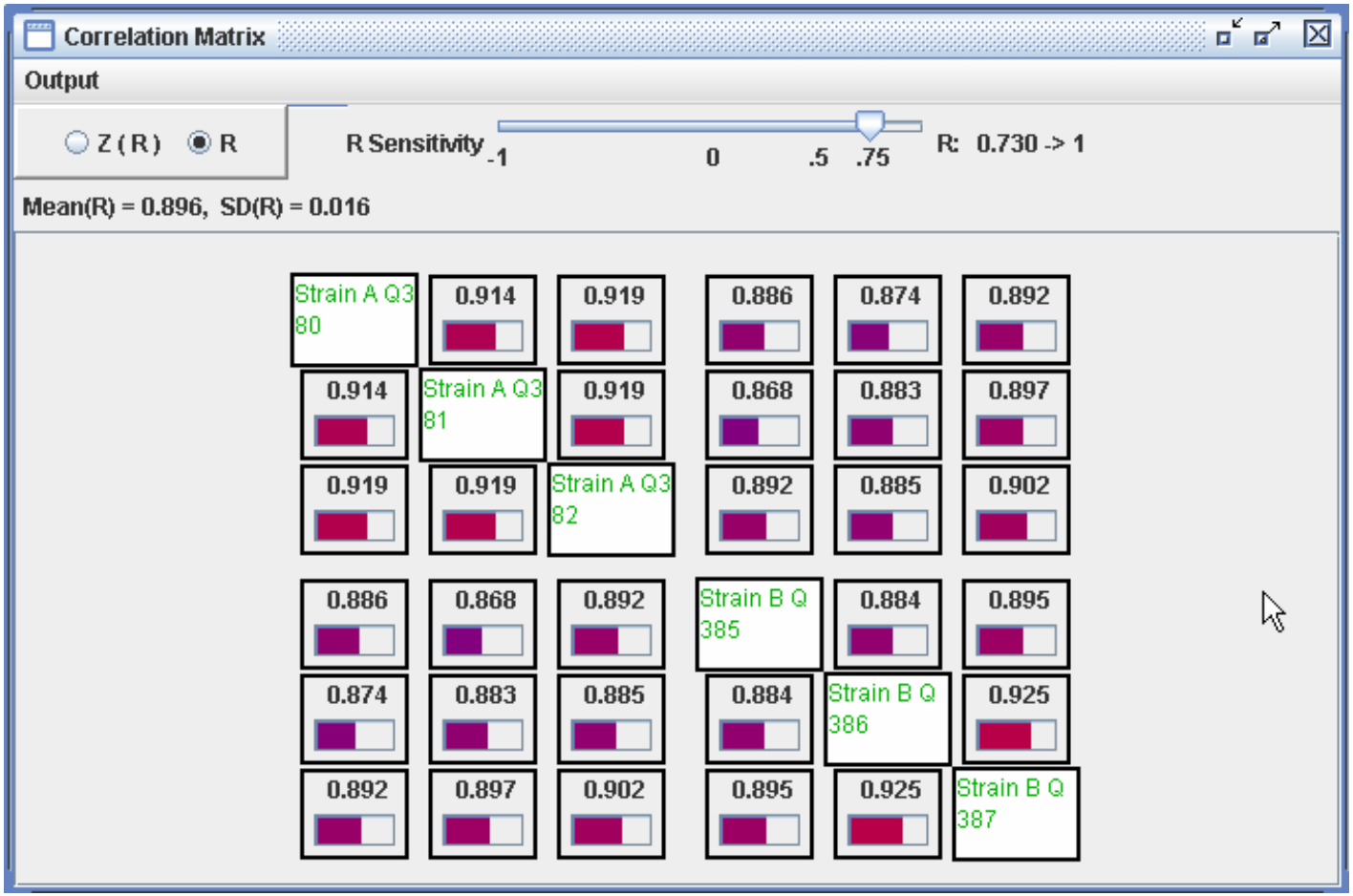
Correlation Matrix - continued

Assign dataset to groups, click “Create Correlation Matrix” button.



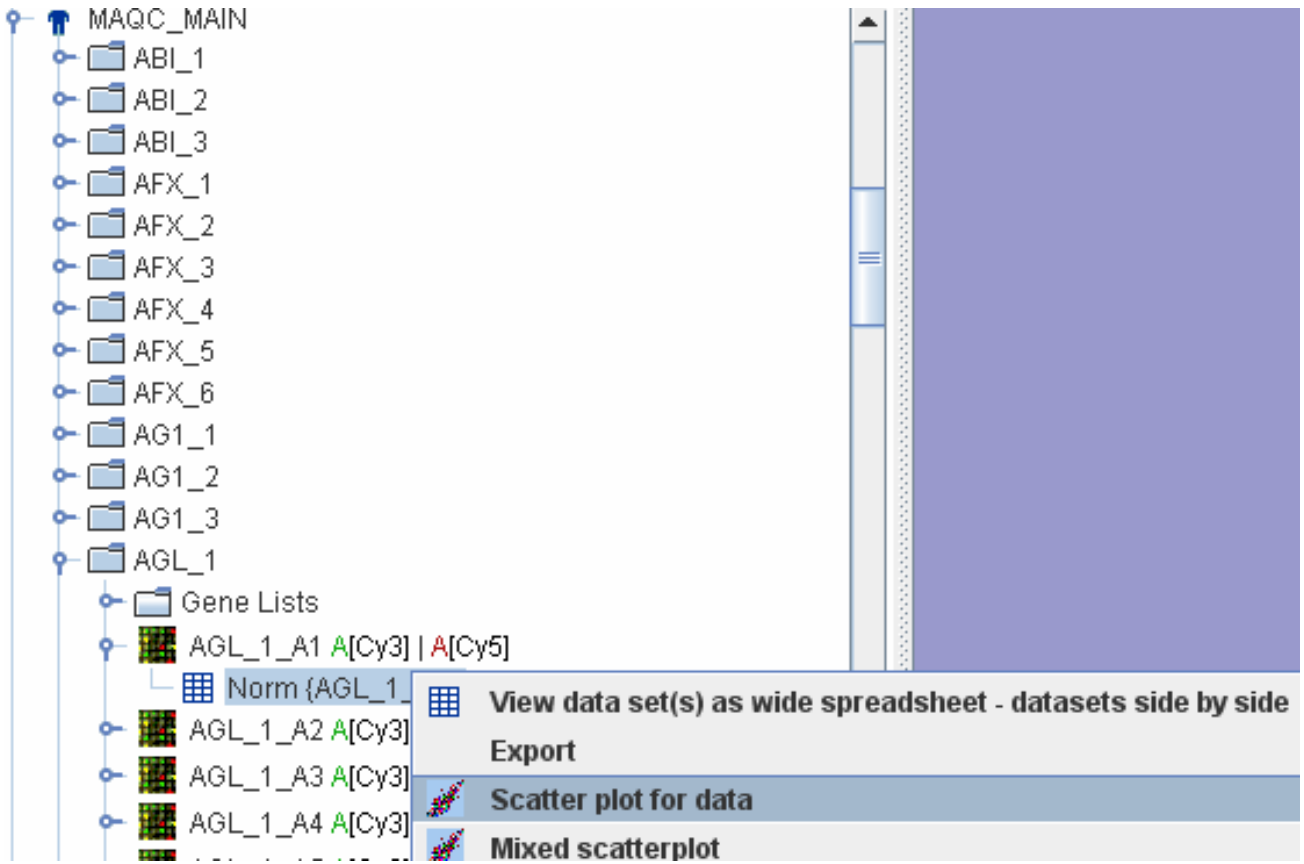
Correlation Matrix - continued

The diagonal elements of the correlation matrix will be 1 since they are the correlation of a column with itself. The correlation matrix is also symmetric since the correlation of column i with column j is the same as the correlation of column j with column i.

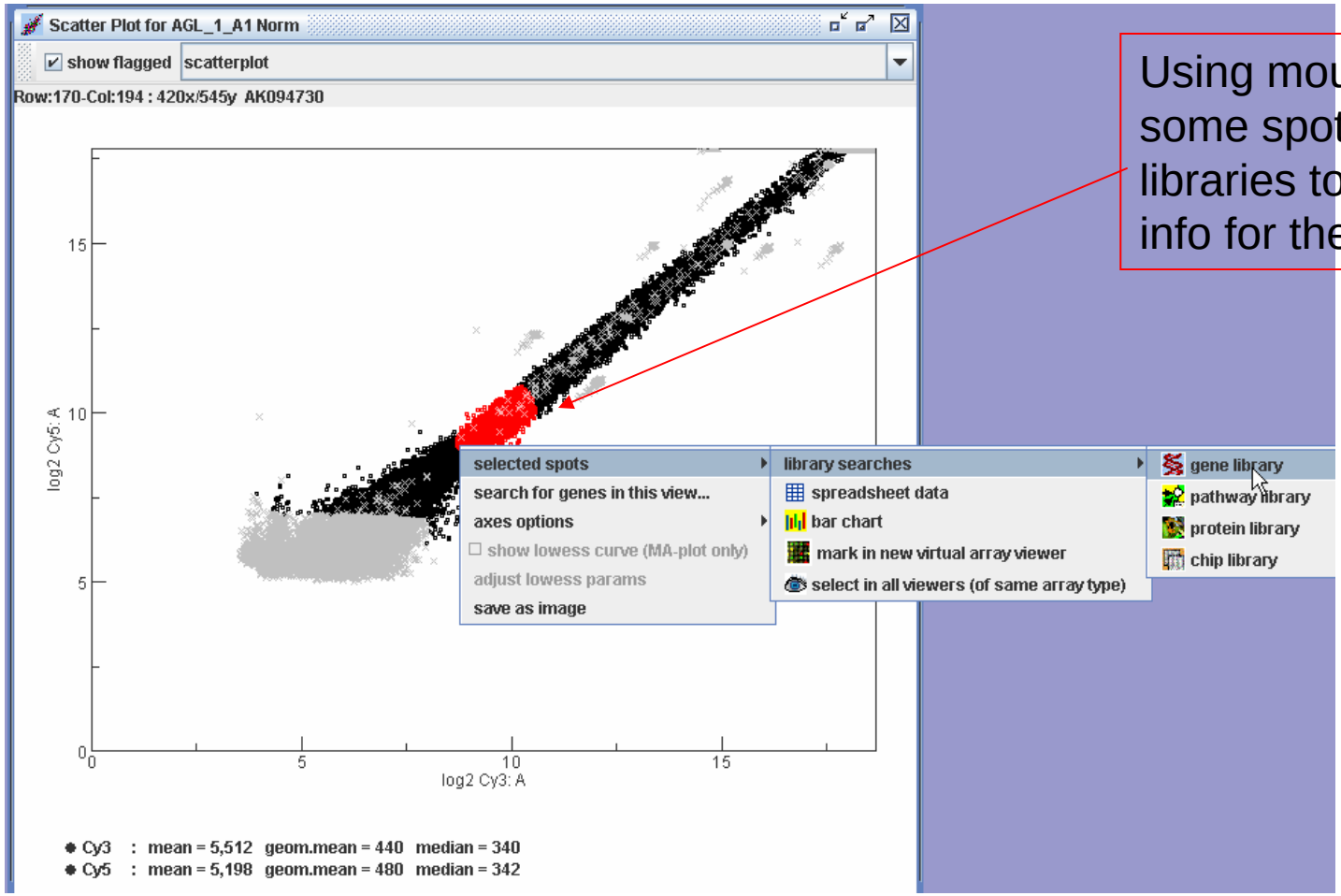


Scatter Plot

Scatter plot shows the fluorescence intensity data of the Cy3 channel versus those of the Cy5 channel for the same array.
Select data, right-click, choose “Scatter plot for data”.



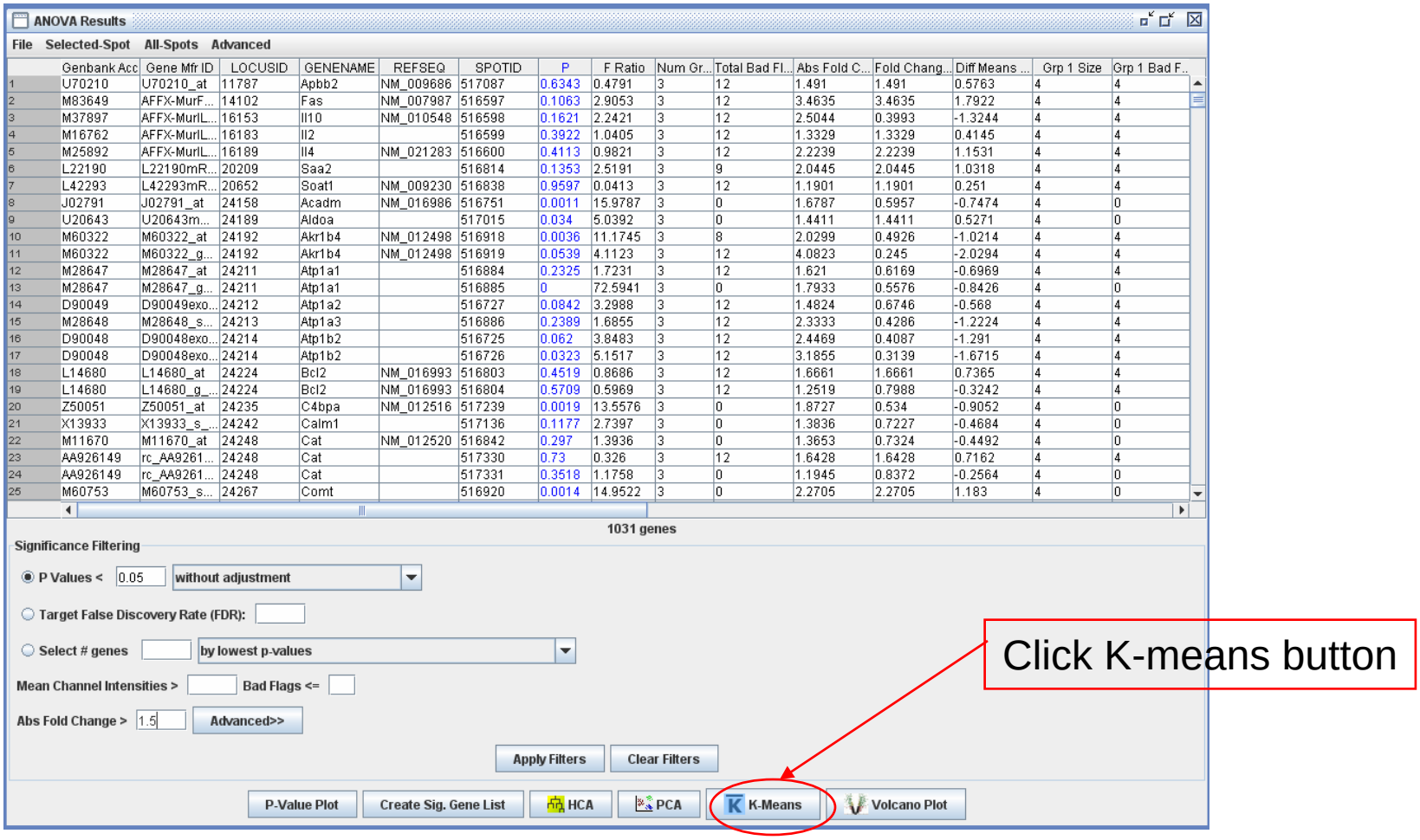
Scatter Plot - continued



Using mouse to circle some spots and approach libraries to get genome info for these spots

K-means

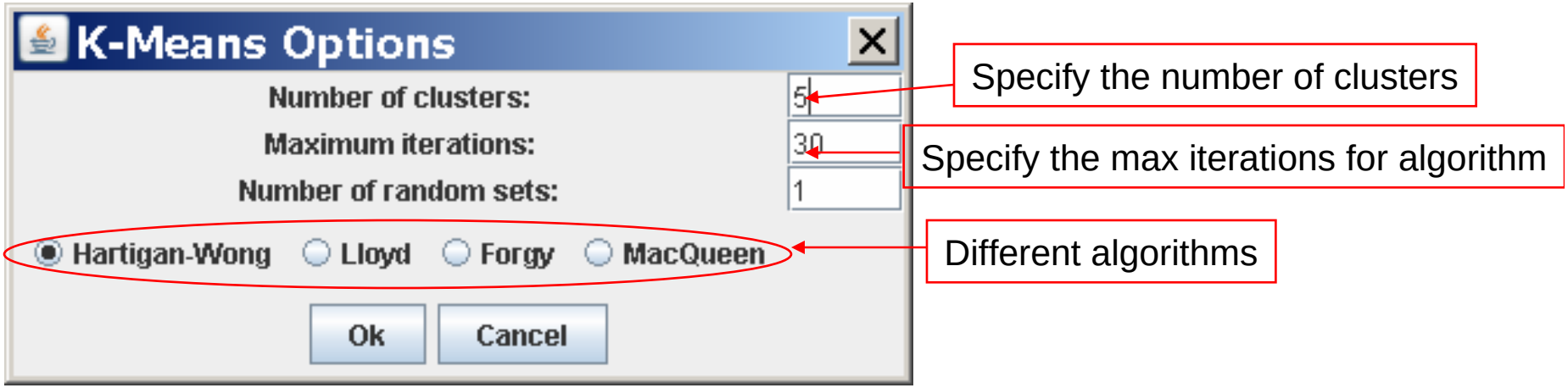
K-means is an algorithm to cluster objects into k clusters based on gene attributes, and calculate the centroid or mean point for each clusters. Users can access K-means from T-test/ANOVA result.



	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	F Ratio	Num Gr...	Total Bad Fl...	Abs Fold C...	Fold Chang...	Diff Means ...	Grp 1 Size	Grp 1 Bad F..
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.6343	0.4791	3	12	1.491	1.491	0.5763	4	4
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.1063	2.9053	3	12	3.4635	3.4635	1.7922	4	4
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.1621	2.2421	3	12	2.5044	0.3993	-1.3244	4	4
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.3922	1.0405	3	12	1.3329	1.3329	0.4145	4	4
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.4113	0.9821	3	12	2.2239	2.2239	1.1531	4	4
6	L22190	L22190mR...	20209	Saa2		516814	0.1353	2.5191	3	9	2.0445	2.0445	1.0318	4	4
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.9597	0.0413	3	12	1.1901	1.1901	0.251	4	4
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.0011	15.9787	3	0	1.6787	0.5957	-0.7474	4	0
9	U20643	U20643m...	24189	Aldoa		517015	0.034	5.0392	3	0	1.4411	1.4411	0.5271	4	0
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.0036	11.1745	3	8	2.0299	0.4926	-1.0214	4	4
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.0539	4.1123	3	12	4.0823	0.245	-2.0294	4	4
12	M28647	M28647_at	24211	Atp1a1		516884	0.2325	1.7231	3	12	1.621	0.6169	-0.6969	4	4
13	M28647	M28647_g...	24211	Atp1a1		516885	0	72.5941	3	0	1.7933	0.5576	-0.8426	4	0
14	D90049	D90049exo...	24212	Atp1a2		516727	0.0842	3.2988	3	12	1.4824	0.6746	-0.568	4	4
15	M28648	M28648_s...	24213	Atp1a3		516886	0.2389	1.6855	3	12	2.3333	0.4286	-1.2224	4	4
16	D90048	D90048exo...	24214	Atp1b2		516725	0.062	3.8483	3	12	2.4469	0.4087	-1.291	4	4
17	D90048	D90048exo...	24214	Atp1b2		516726	0.0323	5.1517	3	12	3.1855	0.3139	-1.6715	4	4
18	L14680	L14680_at	24224	Bcl2	NM_016993	516803	0.4519	0.8686	3	12	1.6661	1.6661	0.7365	4	4
19	L14680	L14680_g_...	24224	Bcl2	NM_016993	516804	0.5709	0.5969	3	12	1.2519	0.7988	-0.3242	4	4
20	Z50051	Z50051_at	24235	C4bpa	NM_012516	517239	0.0019	13.5576	3	0	1.8727	0.534	-0.9052	4	0
21	X13933	X13933_s_...	24242	Calm1		517136	0.1177	2.7397	3	0	1.3836	0.7227	-0.4684	4	0
22	M11670	M11670_at	24248	Cat	NM_012520	516842	0.297	1.3936	3	0	1.3653	0.7324	-0.4492	4	0
23	AA926149	rc_AA9261...	24248	Cat		517330	0.73	0.326	3	12	1.6428	1.6428	0.7162	4	4
24	AA926149	rc_AA9261...	24248	Cat		517331	0.3518	1.1758	3	0	1.1945	0.8372	-0.2564	4	0
25	M60753	M60753_s...	24267	Comt		516920	0.0014	14.9522	3	0	2.2705	2.2705	1.183	4	0

K-means (continued)

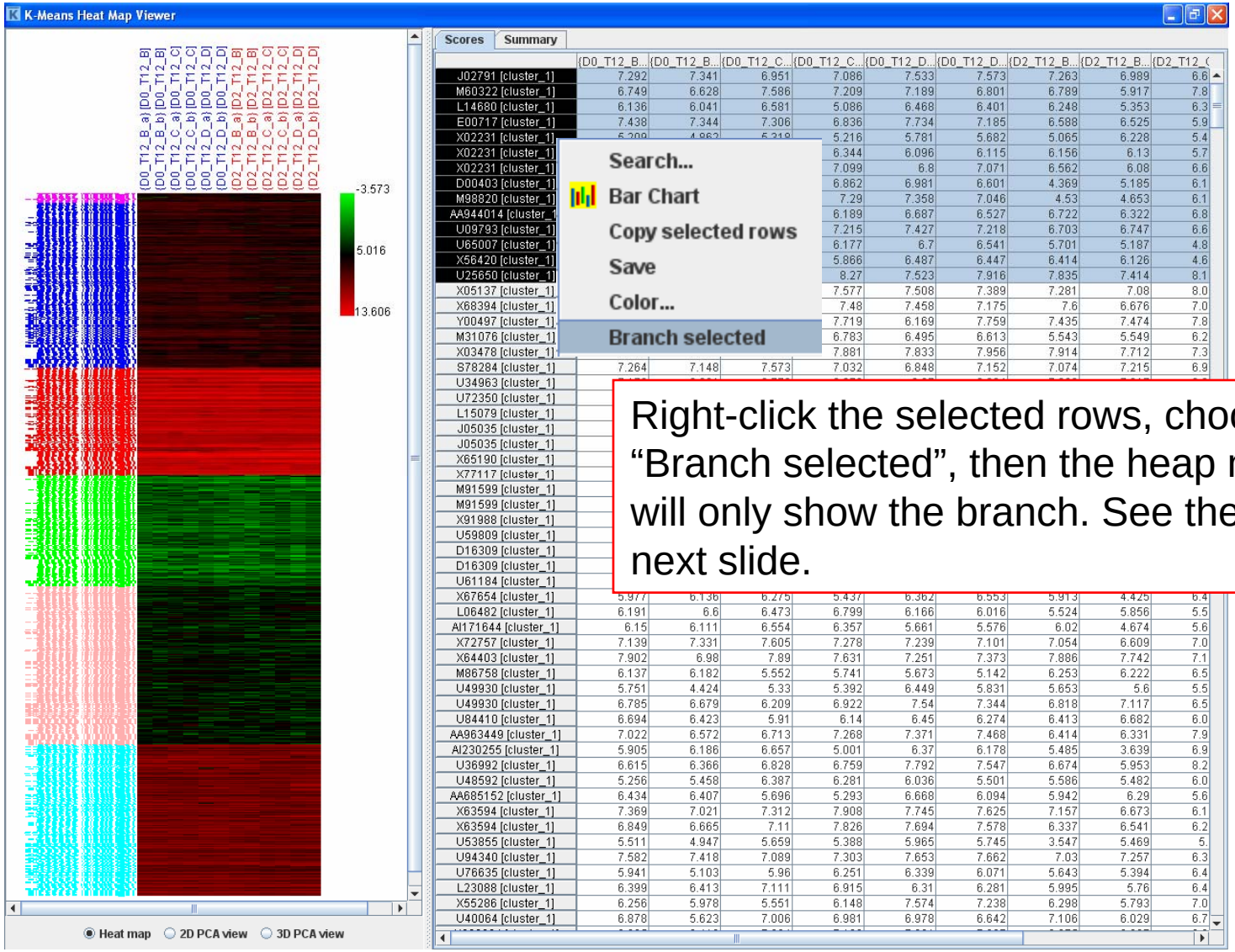
The following window will pop up with some default values



Click OK button

K-means (continued)

Clusters are highlighted in different colors

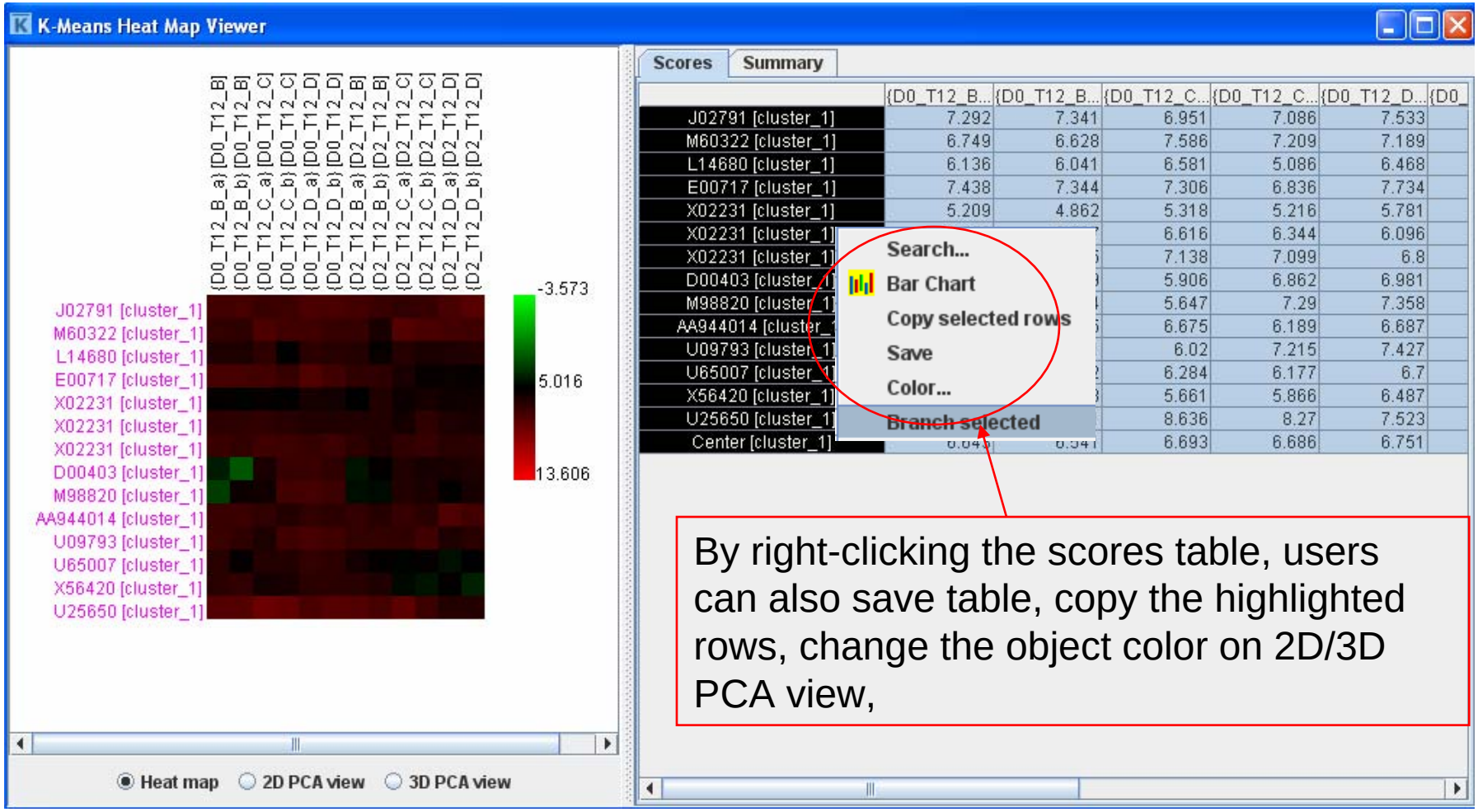


Right-click the selected rows, choose “Branch selected”, then the heap map will only show the branch. See the next slide.

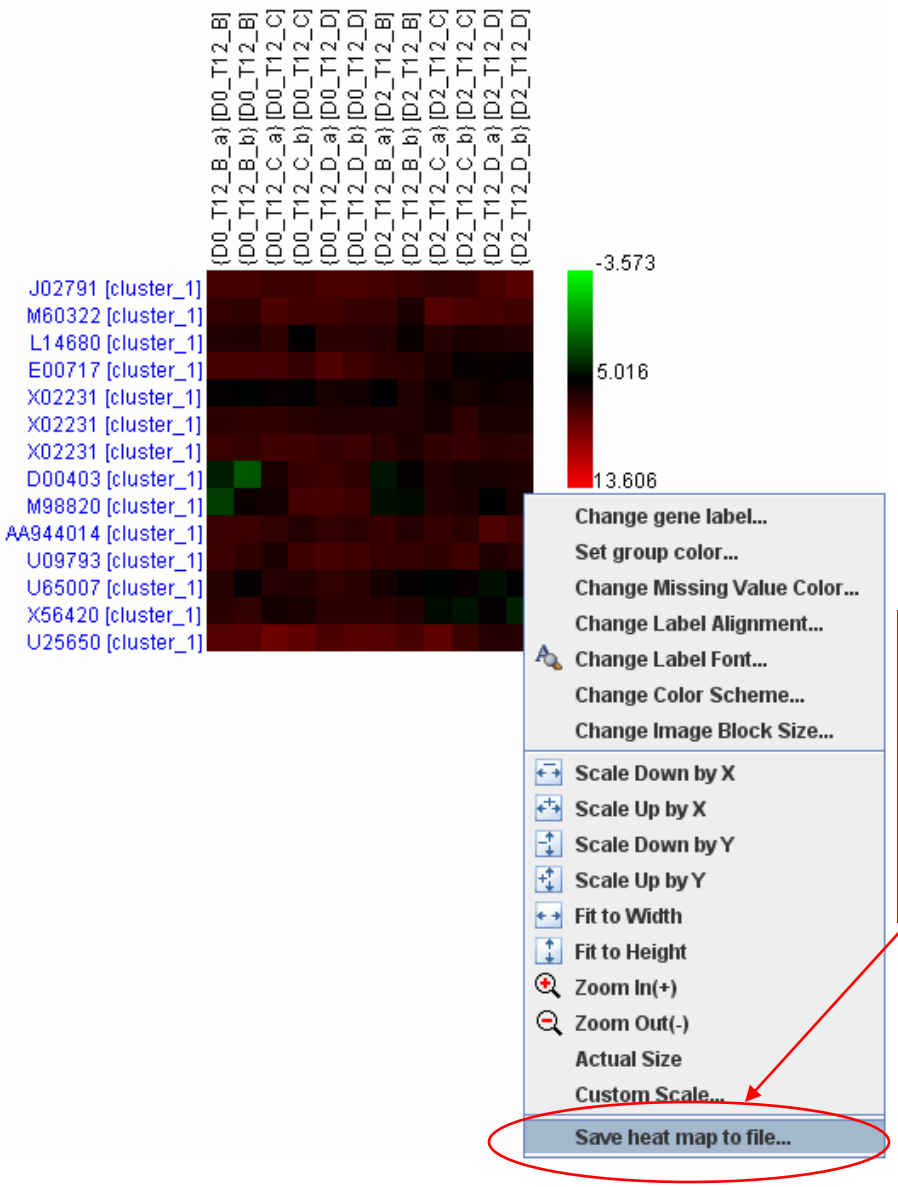
Options for heat map, 2D PCA view or 3D PCA view

K-means (continued)

A branch of the previous heat map

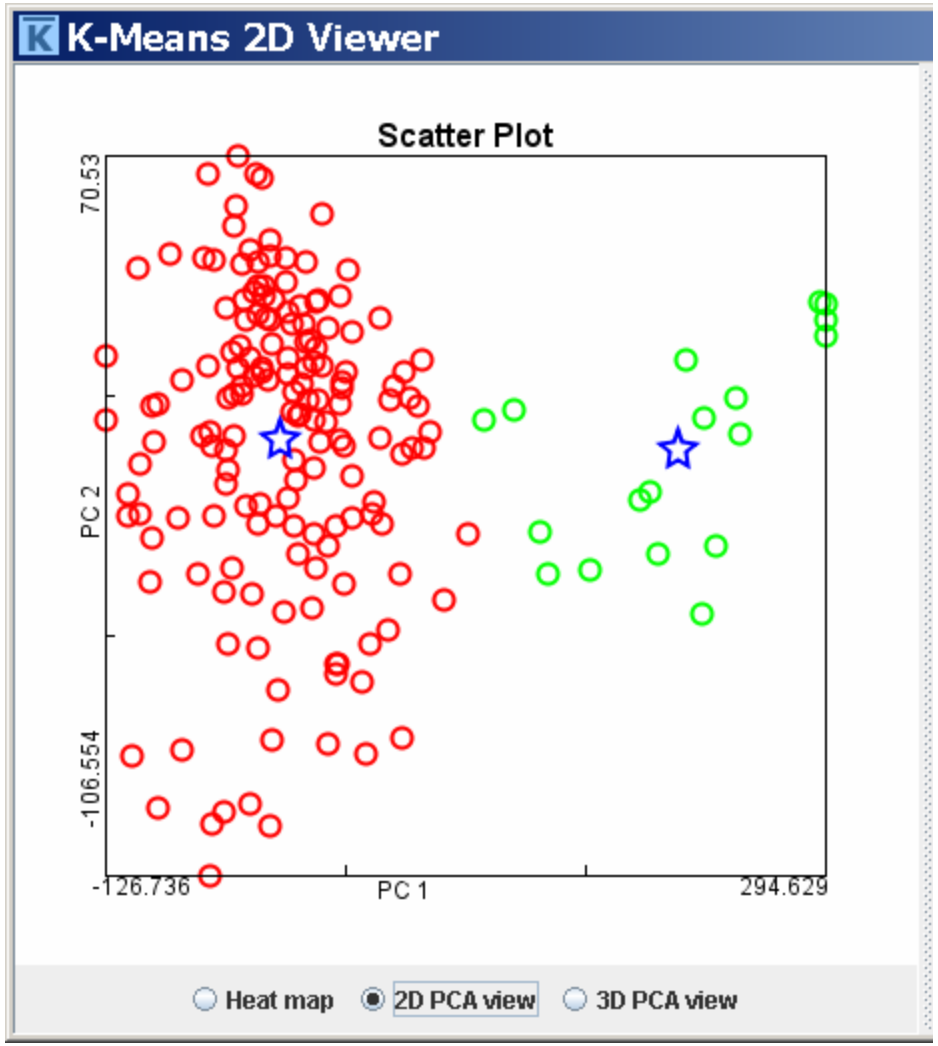


K-means (continued)



Right-click the heat map, user can choose “Save heat map to file…” to save the heat map. There are also many other options for users to change the looks of the heat map.

K-means (continued)

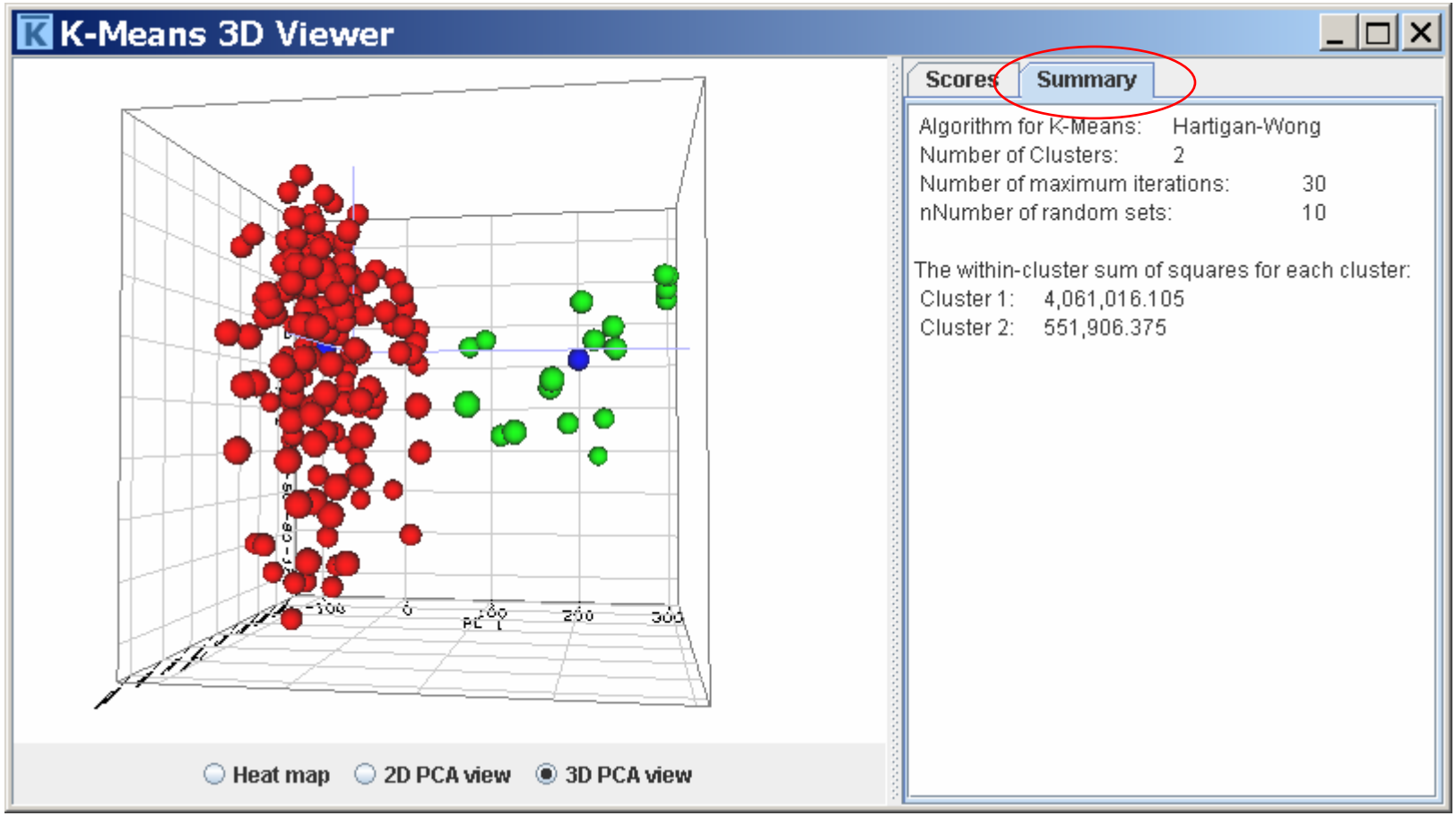


2D view of K-means. Each cluster is marked in different colors. The blue stars are centroids.

K-means (continued)

3D PCA view of K-means

Summary of K-means



Array Track

Tutorial 5:
Accessing gene expression profiles using BarChart

Why use Barchart:

displays expression data for a single gene across multiple arrays within the same experiment or across different experiments. It gives the user an overview of the differential expression levels of this gene across different samples.

How to get Barchart:

<1> from Gene List (the most popular usage);

<2> from Tool Panel;

<3> from T-test result;

<4> from Library windows, such as ChipLib or other Libs.

<5> Advanced: launch barchart from data content tree.

Functions associated with Barchart:

<1> query/sort barchart table;

<2> grouping bars with color;

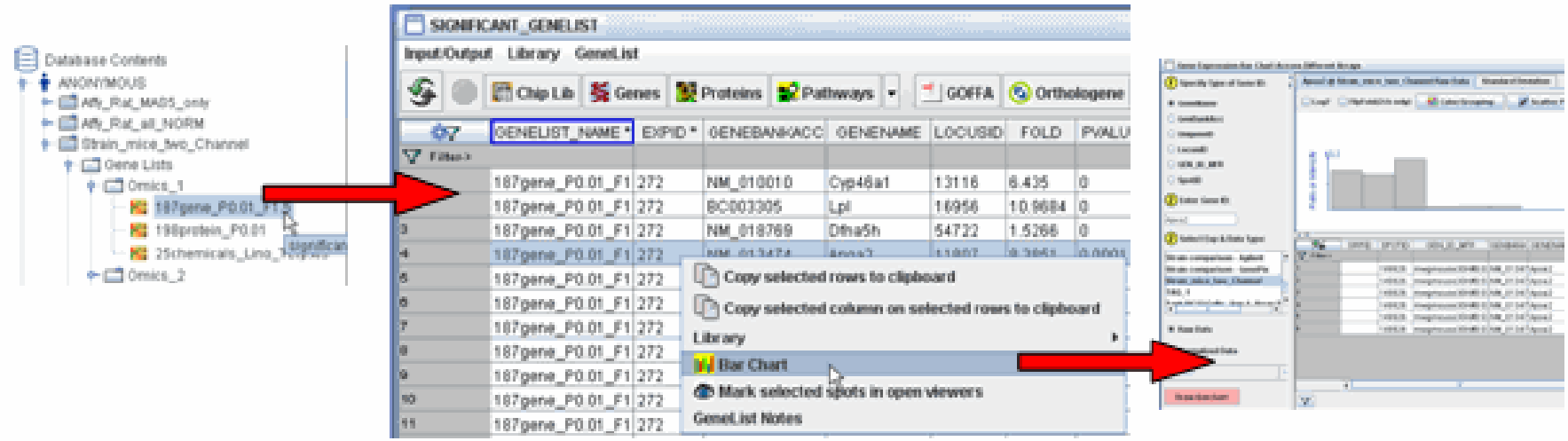
<3> view standard deviation;

<4> Others: apply log2, flip fold, link to Libraries, Scatter Plot etc;

<5> Advanced: cross multi-experiment comparison.

<1> From Gene List

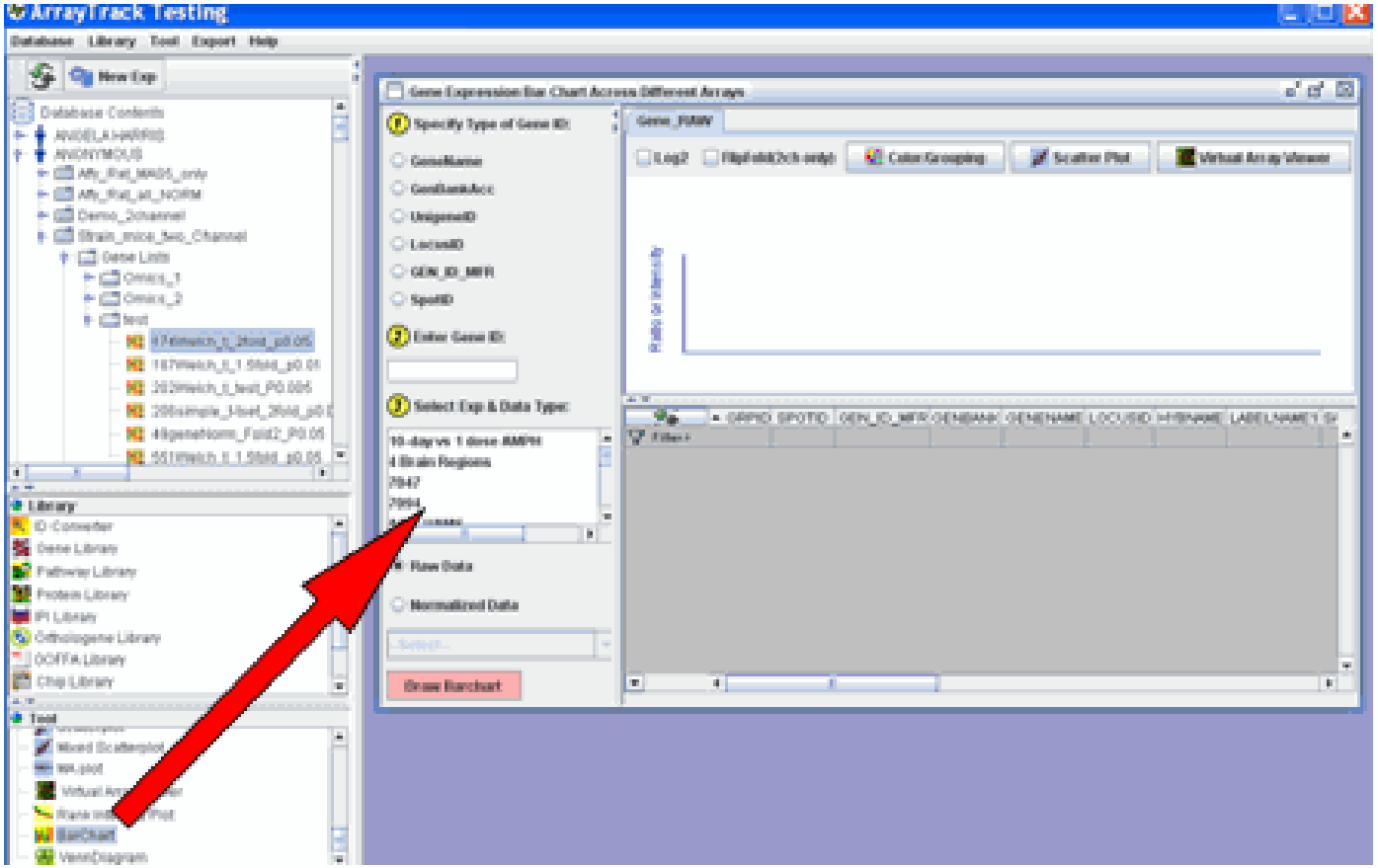
1) Double click a genelist node from database content tree in the top left window.
2) A genelist table will show up.
3) Highlight a record from genelist table, and right-clicking shows a popup menu.
4) Select ‘Bar Chart’ from the popup menu.
5) A Bar Chart window is displayed with the selected gene’s profile across all normalized data set within the specific experiment that this genelist belongs to.



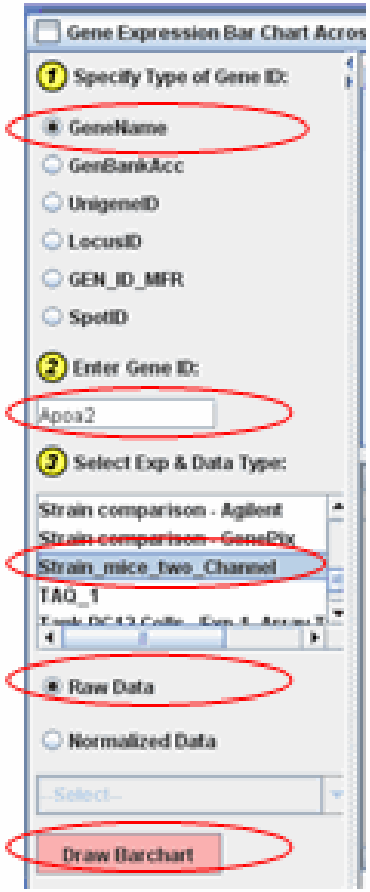
<2> From Tool panel

Use it when you know gene id and experiment name

1) Double click BarChart node from Tool Panel (the bottom left window)
2) A BarChart window is shown.



<2> From Tool panel (cont.)



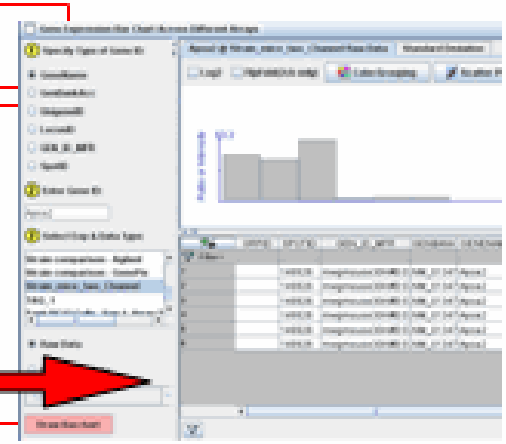
1) Specify gene type

2) Specify gene ID

3) Select experiment name

4) Select data type mod
   - Raw
   - Normalized(select normalization method from drop down list)

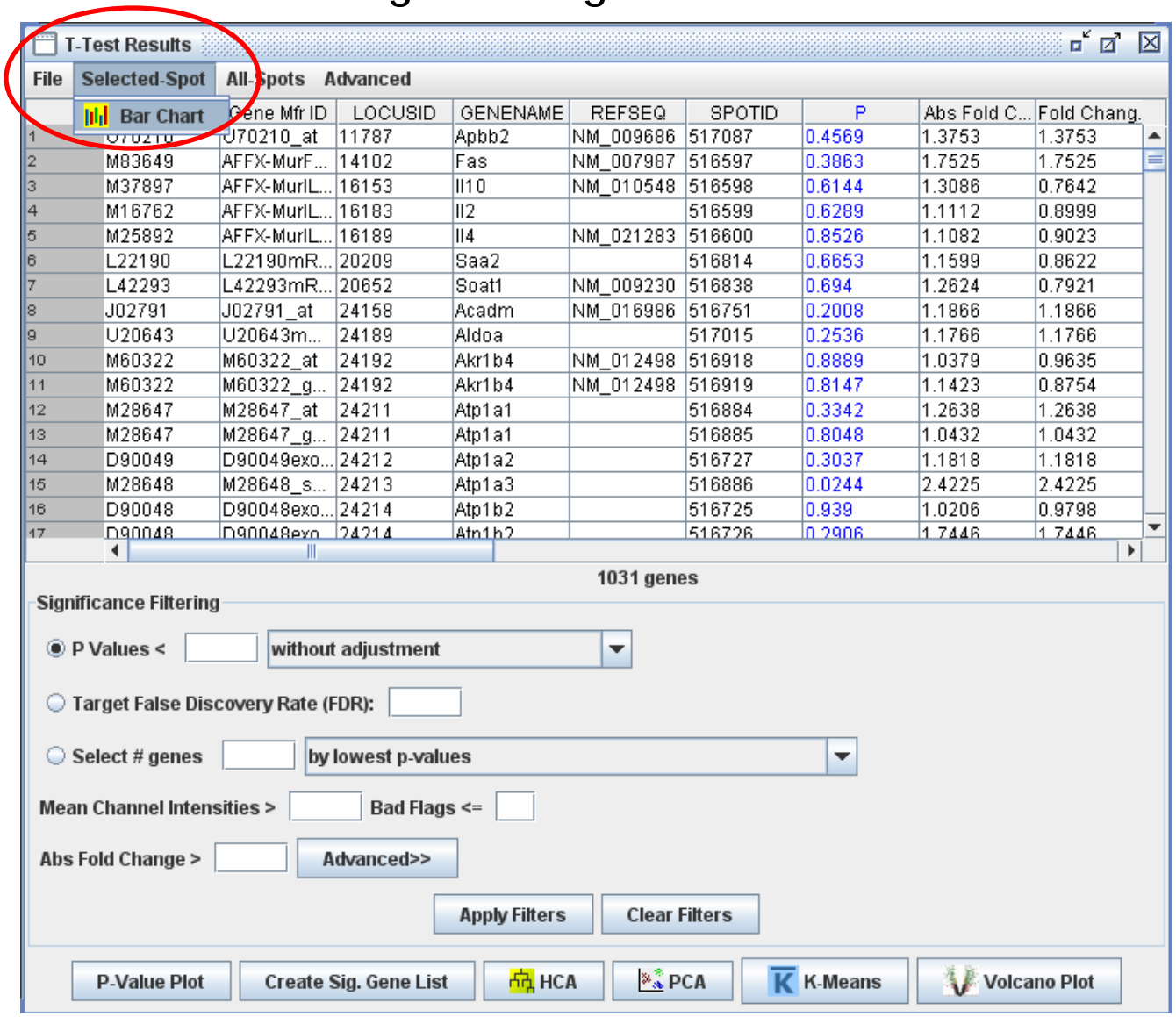
5) Click 'Draw Barchart' button



<3> From T-test Result

Please refer Tutorial 1 for generating T-test result.

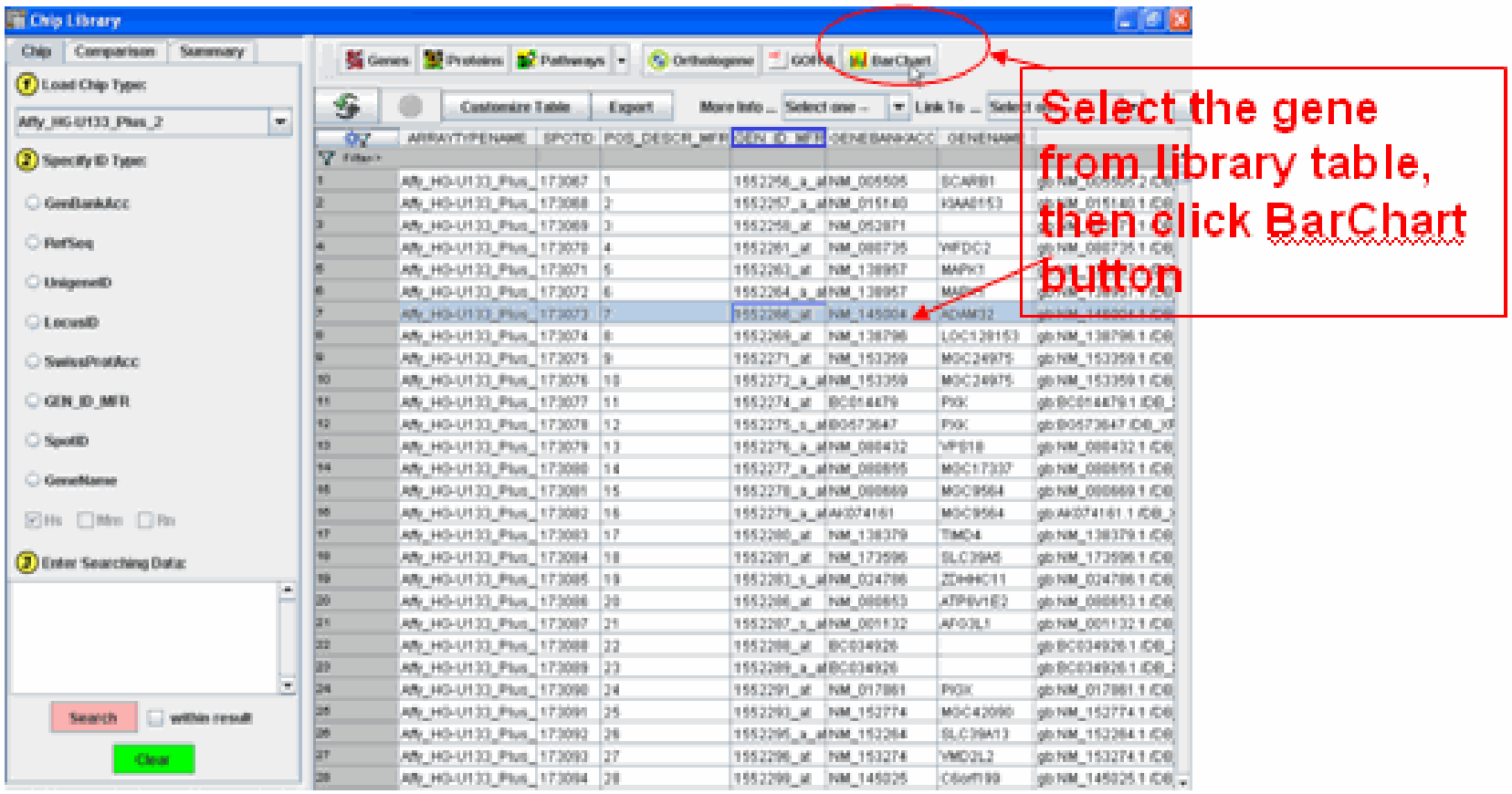
Select a gene in the T-test result and choose BarChart from “Selected-Spot” menu.



		Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang.
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.4569	1.3753	1.3753
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.3863	1.7525	1.7525
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.6144	1.3086	0.7642
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.6289	1.1112	0.8999
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.8526	1.1082	0.9023
6	L22190	L22190mR...	20209	Saa2		516814	0.6653	1.1599	0.8622
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.694	1.2624	0.7921
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.2008	1.1866	1.1866
9	U20643	U20643m...	24189	Aldoa		517015	0.2536	1.1766	1.1766
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.8889	1.0379	0.9635
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.8147	1.1423	0.8754
12	M28647	M28647_at	24211	Atp1a1		516884	0.3342	1.2638	1.2638
13	M28647	M28647_g...	24211	Atp1a1		516885	0.8048	1.0432	1.0432
14	D90049	D90049exo...	24212	Atp1a2		516727	0.3037	1.1818	1.1818
15	M28648	M28648_s...	24213	Atp1a3		516886	0.0244	2.4225	2.4225
16	D90048	D90048exo...	24214	Atp1b2		516725	0.939	1.0206	0.9798
17	D90048	D90048exo...	24214	Atp1b2		516726	0.2906	1.7446	1.7446

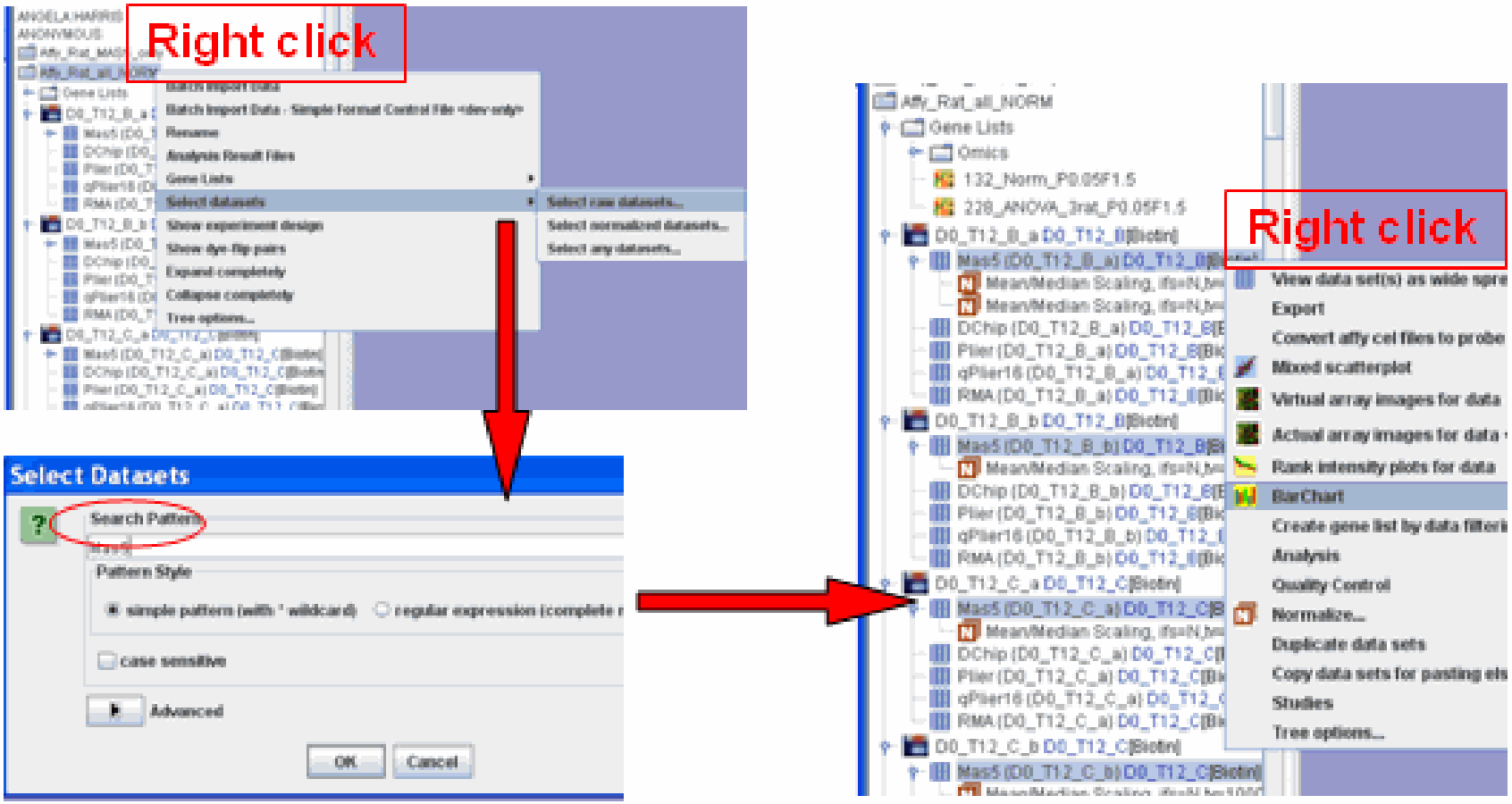
<4> From ChipLib and other Libs

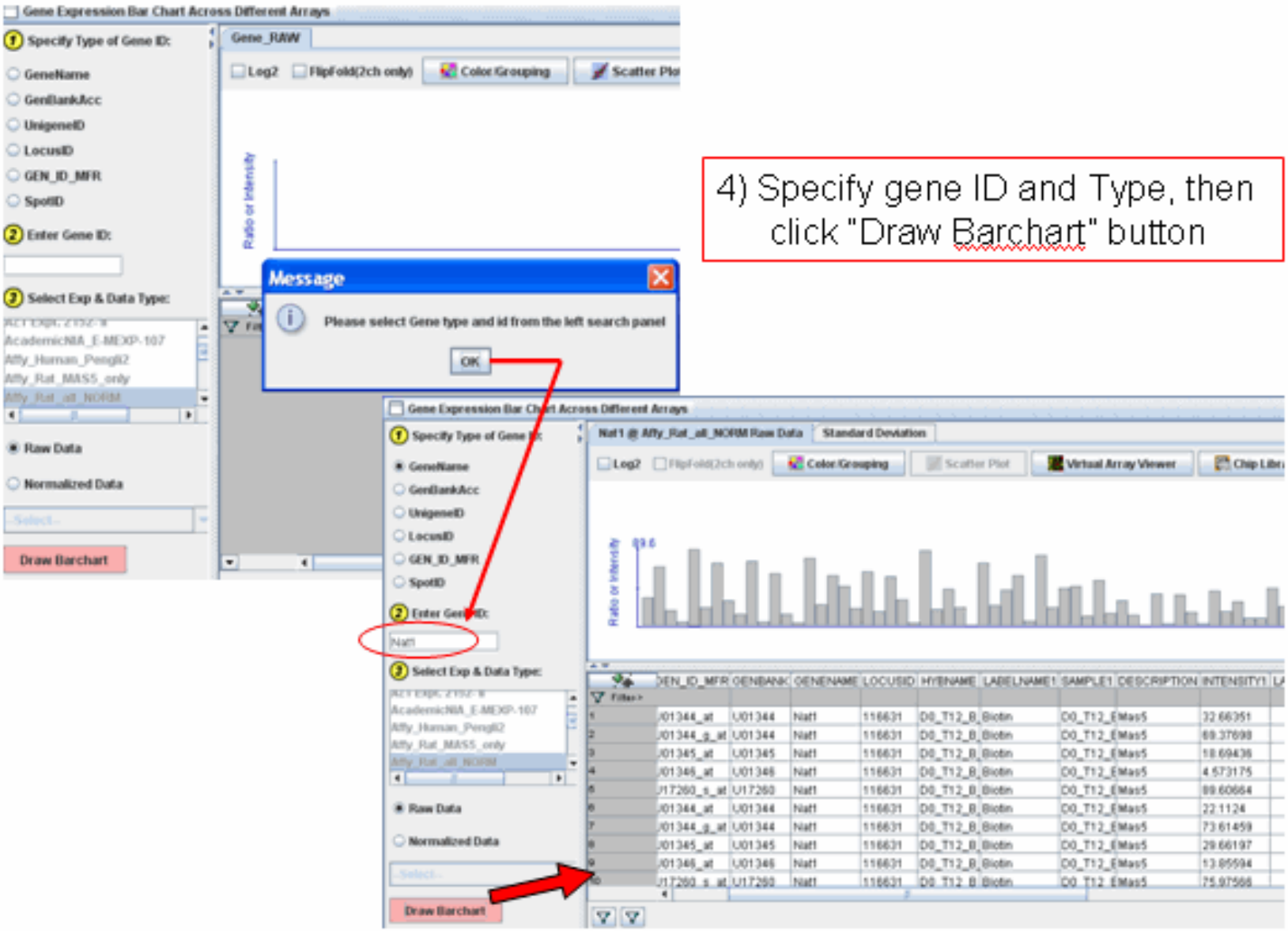
1) Select gene from library table
2) Click “BarChart” button on the top of tool bars



<5> Advanced: From Data Tree

1) Select experiment name, then right click to select datasets
2) Type in search pattern, matched data will be highlighted in the tree.
3) Right click and select “BarChart” to launch. (continue on next page..)

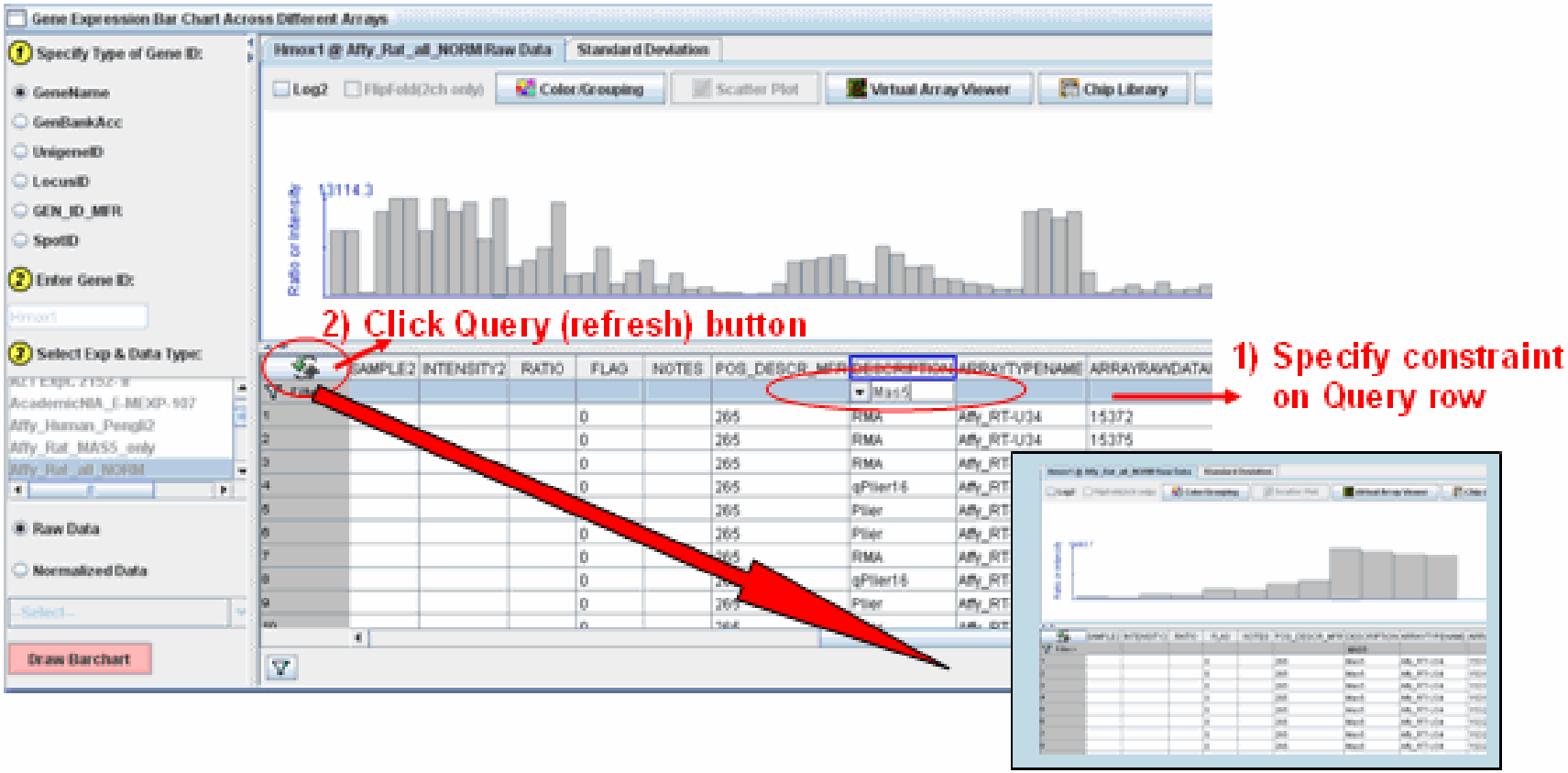



Gene Expression Bar Chart Across Different Arrays
Specify Type of Gene ID:
GeneName
GenBankAcc
UnigeneID
LocusID
GEN_ID_MFR
SpotID
Enter Gene ID:
Select Exp & Data Type:
Raw Data
Normalized Data
Draw Barchart
Gene_RAW
Log2
FlipFold(2ch only)
Color Grouping
Scatter Plot
Ratio or Intensity
Message
Please select Gene type and id from the left search panel
OK
Virtual Array Viewer
Standard Deviation
Natl
4) Specify gene ID and Type, then click "Draw Barchart" button

Functions: Query/Sort Barchart Table

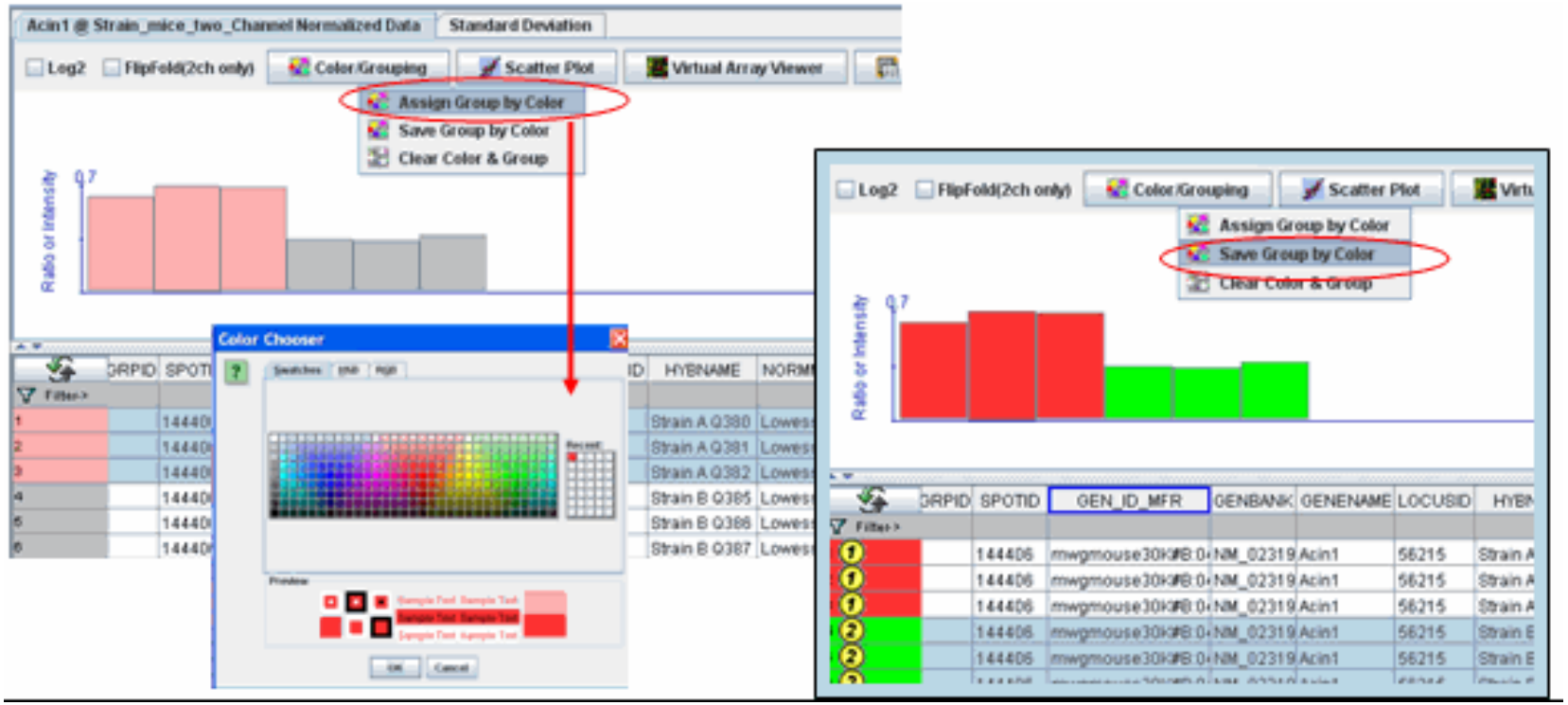
Query:
1) Type constraint in a particular field on the query row.
2) Click query button.
3) Barchart will be re-drawn based on the search result.

Sort: Click header to sort table by columns.



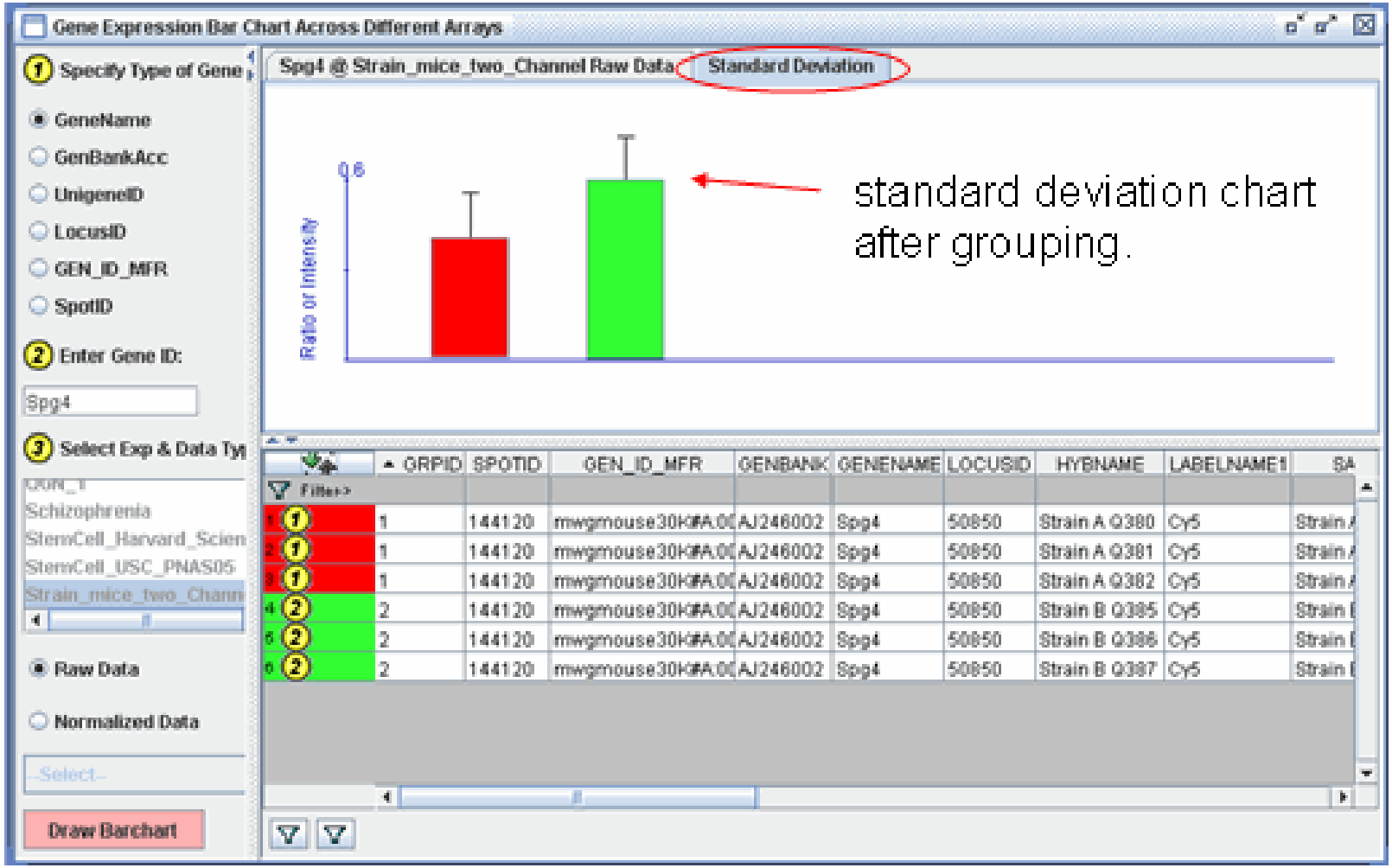
Functions: Grouping Bars

1) Select group of records from table or highlight bars from bar chart panel
2) Click “Color/Grouping” button and select “Assign Group by Color”
3) Once grouping is done, Click “Color/Grouping” button again and select “Save Group by Color”. (the grouping info will be saved for that particular experiment, so when you view other genes within the same experiment, the color of grouping will automatically be applied.)

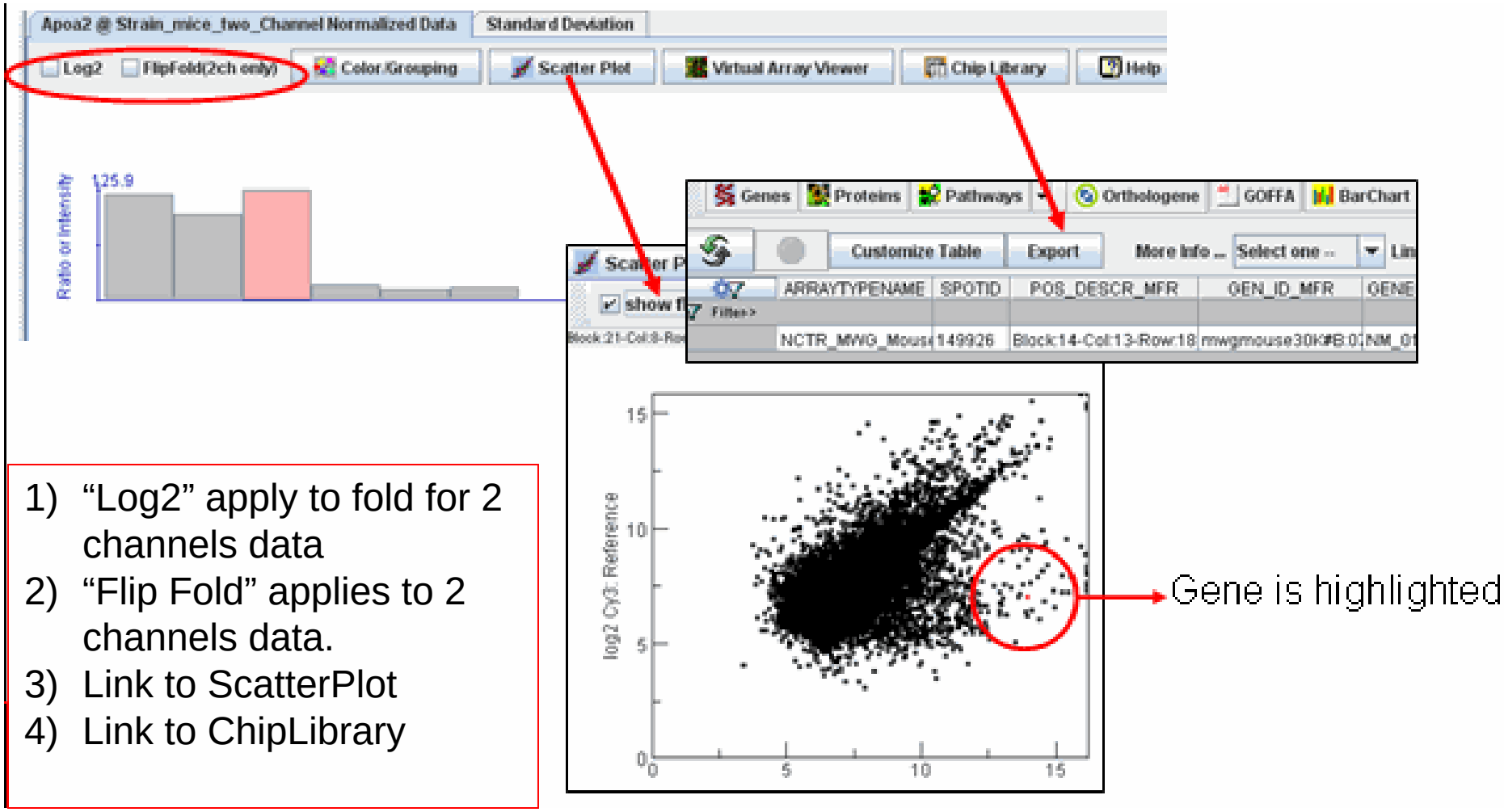


Functions: Standard Deviation

1) Click “Standard Deviation” tab on the top of barchart panel
(the standard deviation chart makes more sense after grouping.)



Functions: Others



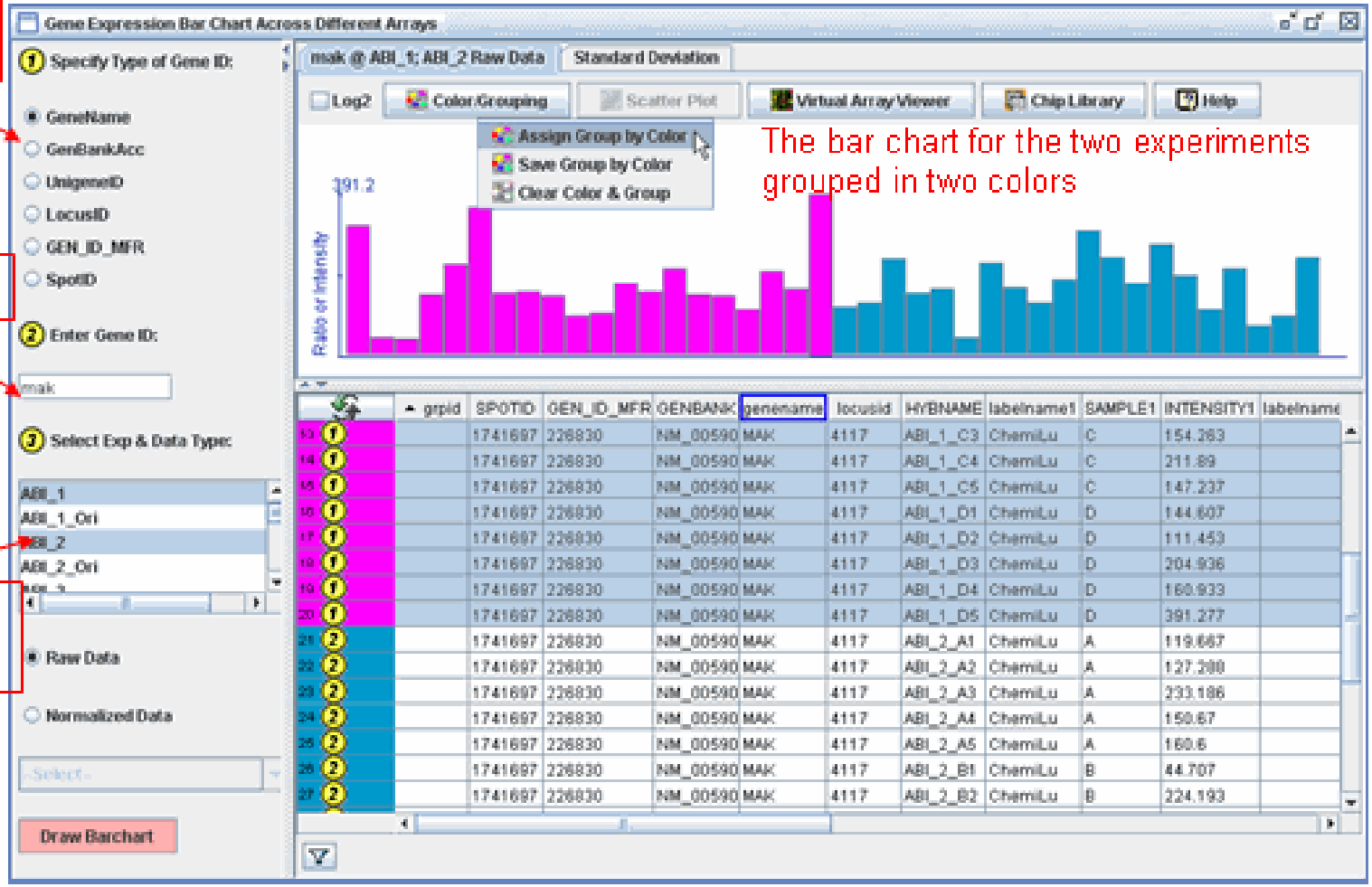
1) "Log2" apply to fold for 2 channels data
2) "Flip Fold" applies to 2 channels data.
3) Link to ScatterPlot
4) Link to ChipLibrary

Advanced Function: How to use BarChart for cross-experiment comparison

Specify the type of gene ID

Enter the gene ID

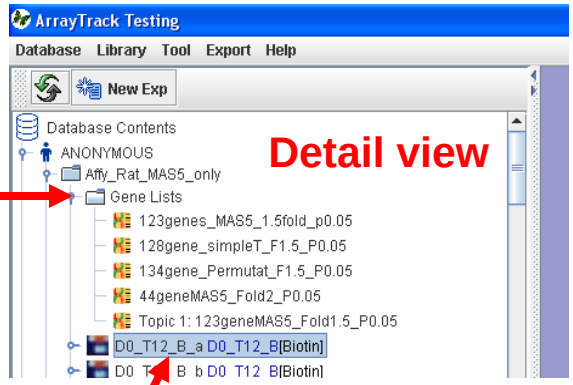
Select more than one exp



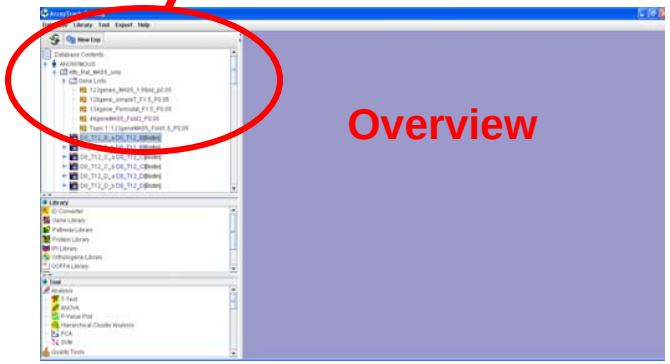
Tutorial 6:
GeneList – An important concept in ArrayTrack

Introduction

The Gene List folder contains gene lists that are usually derived from applying certain filtering criteria (e.g. P-value, fold change, etc) on the microarray data. As a common practice, a list of gene is identified by researchers and then carried forward for biological interpretation and/or further analysis.



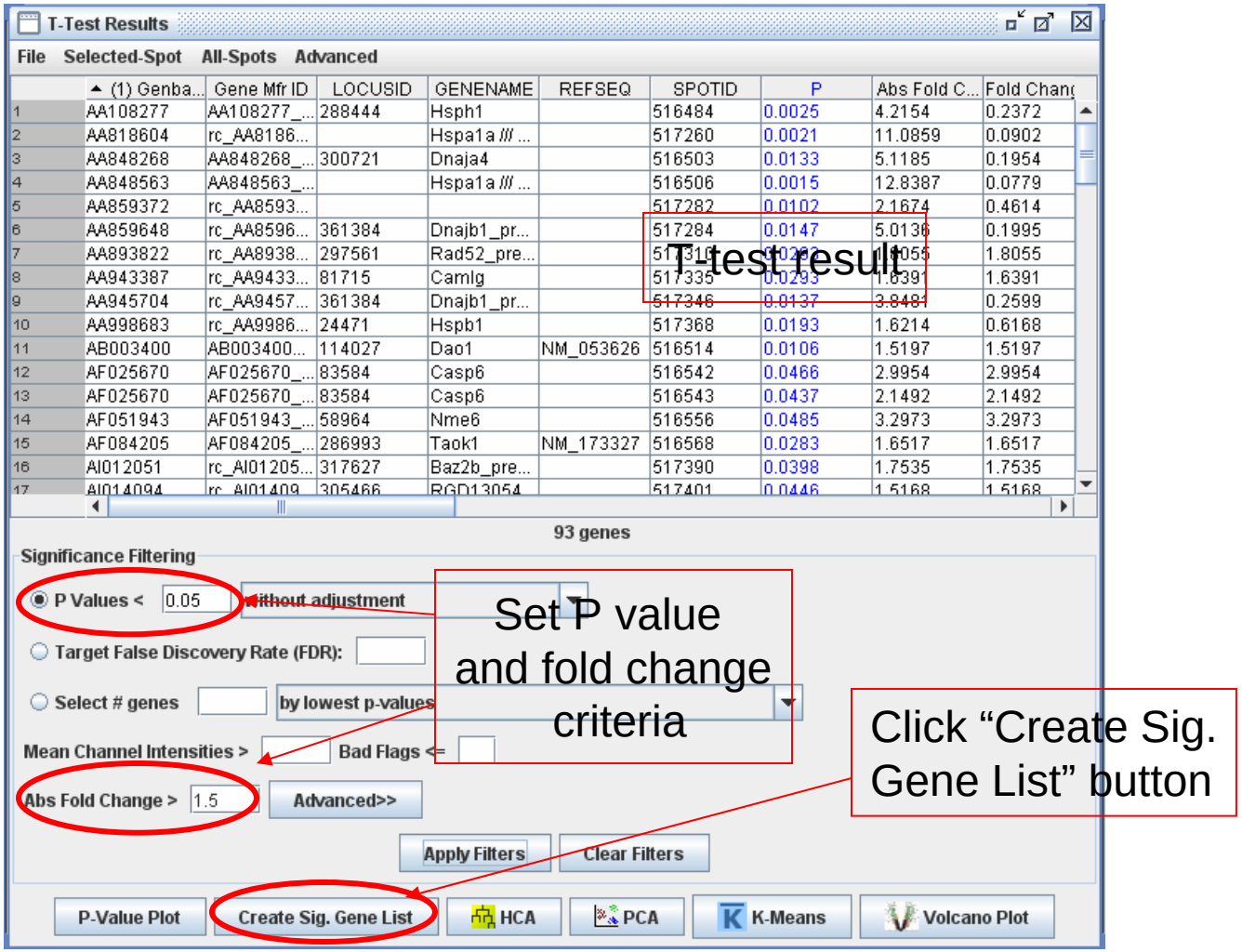
There are several ways to create a gene list in ArrayTrack:
- Create a gene list through statistical analysis (page 3)
- Create a gene list through volcano plot (page 5)
- Create a gene list through data filtering (page 6)
- Create a gene list through manual selection (page 7)
- Create a gene list through GOFFA (page 8)
- Create a gene list through pathway (page 10)
- Alternatively, the user can import a list of genes that is created outside of ArrayTrack. (page 12)



ArrayTrack allows the following operations solely based on gene lists:
- Normalization based on a gene list (page 21)
- Statistical analysis (t-test, PCA, HCA) applied on a gene list (page 22)
- Venn diagram drawn on two or three gene lists (page 19)
- Functional analysis on a gene list (page 26)
- Export a set of microarray with genes that are specified by a gene list (page 28)
- Miscellaneous: delete a gene list (page 17); create a Gene List subfolder (page 15)

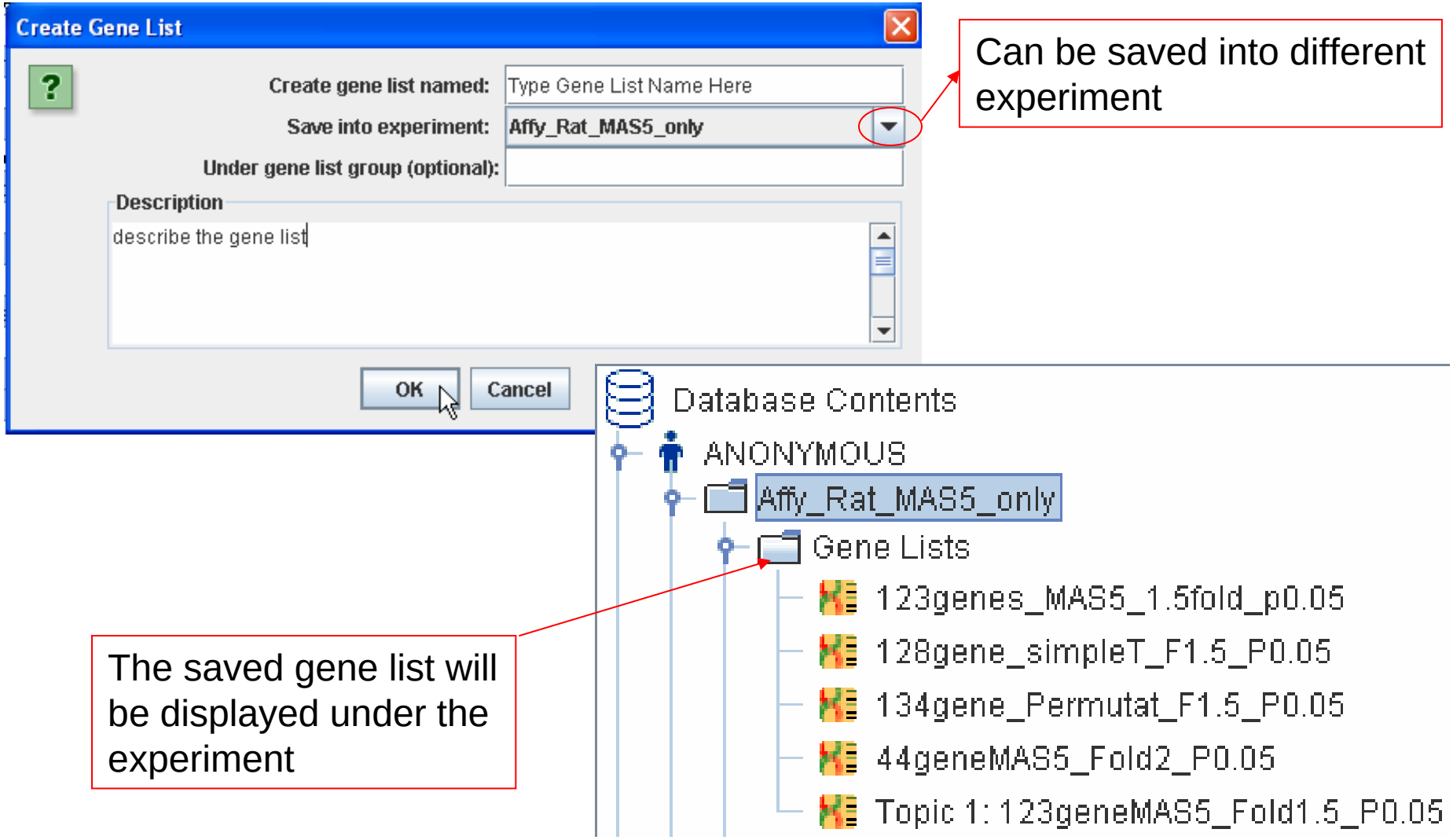
Create a gene list through statistical analysis

From data analysis results (e.g. T-test result. See Tutorial 1 for doing T-test).



	(1) Genba...	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang
1	AA108277	AA108277_...	288444	Hsph1		516484	0.0025	4.2154	0.2372
2	AA818604	rc_AA8186...		Hspa1a /// ...		517260	0.0021	11.0859	0.0902
3	AA848268	AA848268_...	300721	Dnaja4		516503	0.0133	5.1185	0.1954
4	AA848563	AA848563_...		Hspa1a /// ...		516506	0.0015	12.8387	0.0779
5	AA859372	rc_AA8593...				517282	0.0102	2.1674	0.4614
6	AA859648	rc_AA8596...	361384	Dnajb1_pr...		517284	0.0147	5.0136	0.1995
7	AA893822	rc_AA8938...	297561	Rad52_pre...		517310	[illegible]	[illegible]	1.8055
8	AA943387	rc_AA9433...	81715	Camlg		517335	0.0293	1.6391	1.6391
9	AA945704	rc_AA9457...	361384	Dnajb1_pr...		517346	0.0137	3.8481	0.2599
10	AA998683	rc_AA9986...	24471	Hspb1		517368	0.0193	1.6214	0.6168
11	AB003400	AB003400...	114027	Dao1	NM_053626	516514	0.0106	1.5197	1.5197
12	AF025670	AF025670_...	83584	Casp6		516542	0.0466	2.9954	2.9954
13	AF025670	AF025670_...	83584	Casp6		516543	0.0437	2.1492	2.1492
14	AF051943	AF051943_...	58964	Nme6		516556	0.0485	3.2973	3.2973
15	AF084205	AF084205_...	286993	Taok1	NM_173327	516568	0.0283	1.6517	1.6517
16	AI012051	rc_AI01205...	317627	Baz2b_pre...		517390	0.0398	1.7535	1.7535
17	AI014094	rc_AI01409...	305466	RGD13054		517401	0.0446	1.5168	1.5168

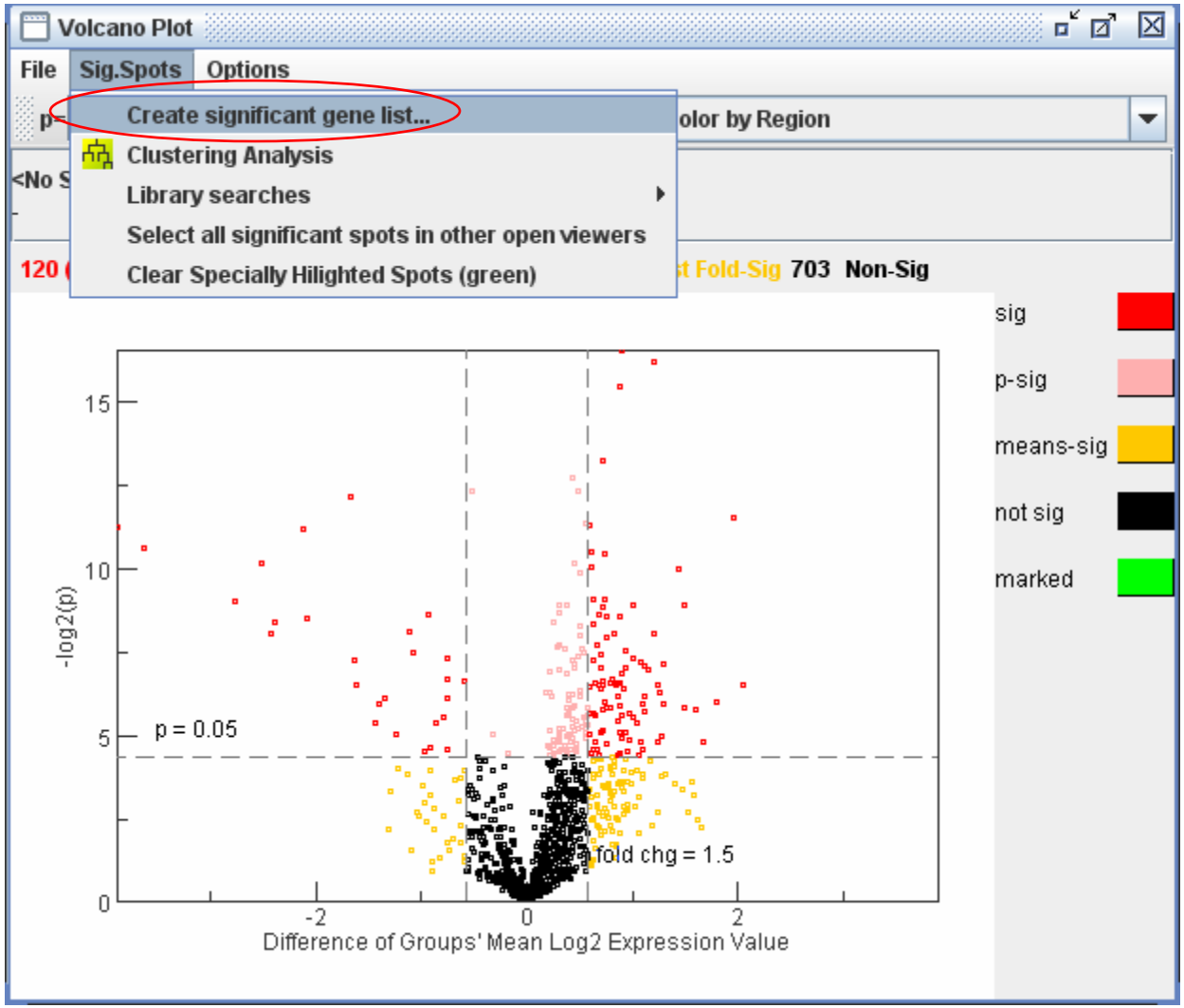
Create a gene list through statistical analysis - continued



Can be saved into different experiment

The saved gene list will be displayed under the experiment

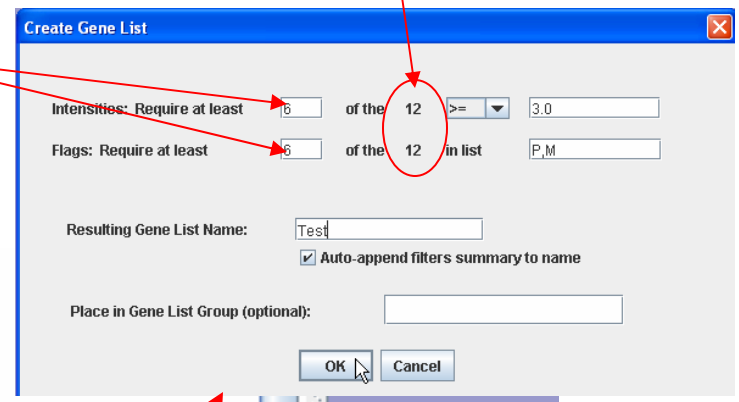
Create gene list through volcano plot



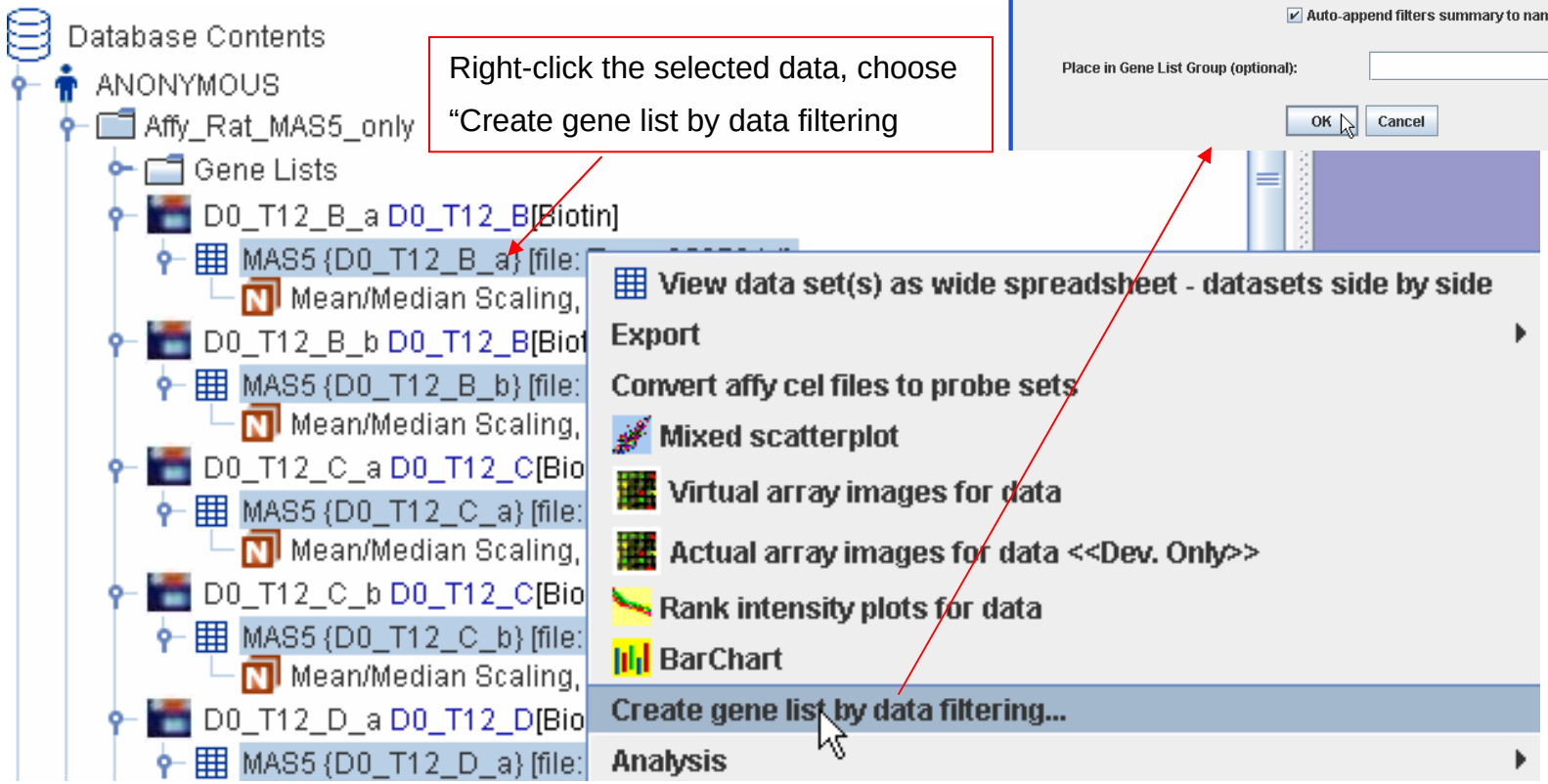
Create a gene list through data filtering

The total number of selected hybridizations

User can set the criteria (like flag, intensity value) to create a gene list, e.g find out all the genes which intensity is greater than 3.0 in at least 6 out of 12 hybridizations.



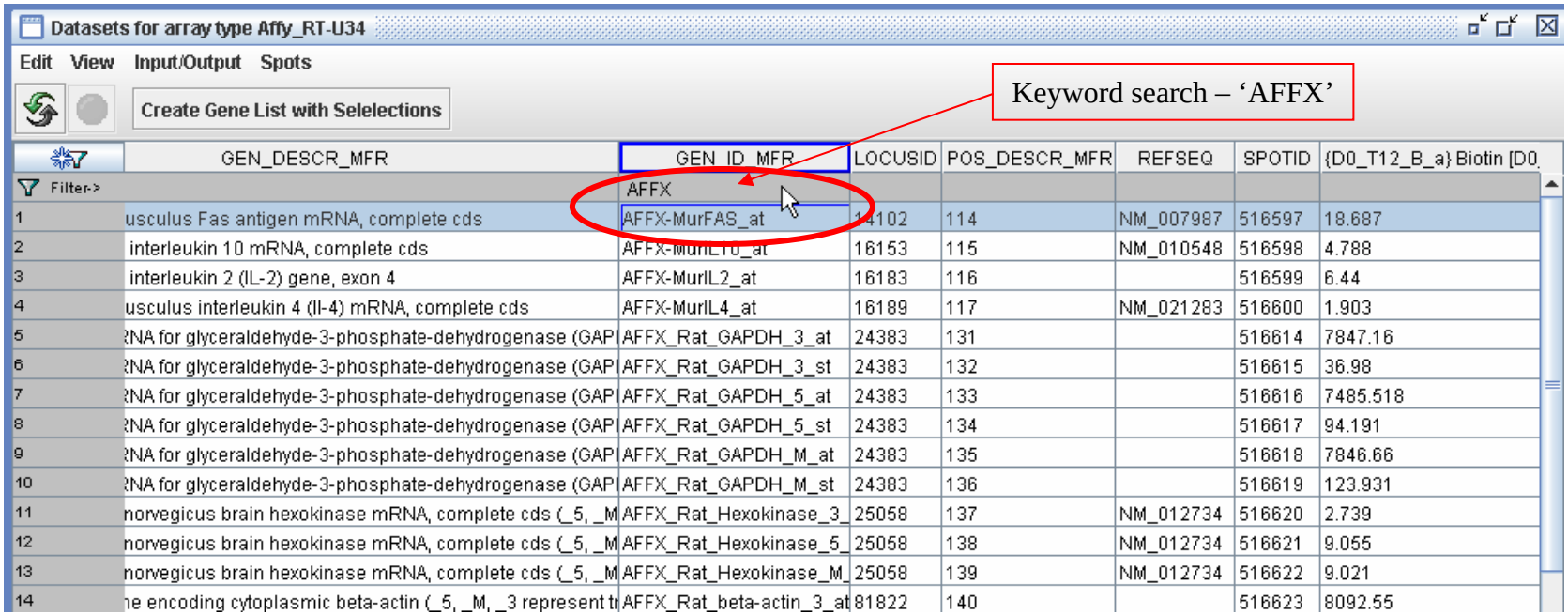
Right-click the selected data, choose "Create gene list by data filtering



Create a gene list through querying chip information

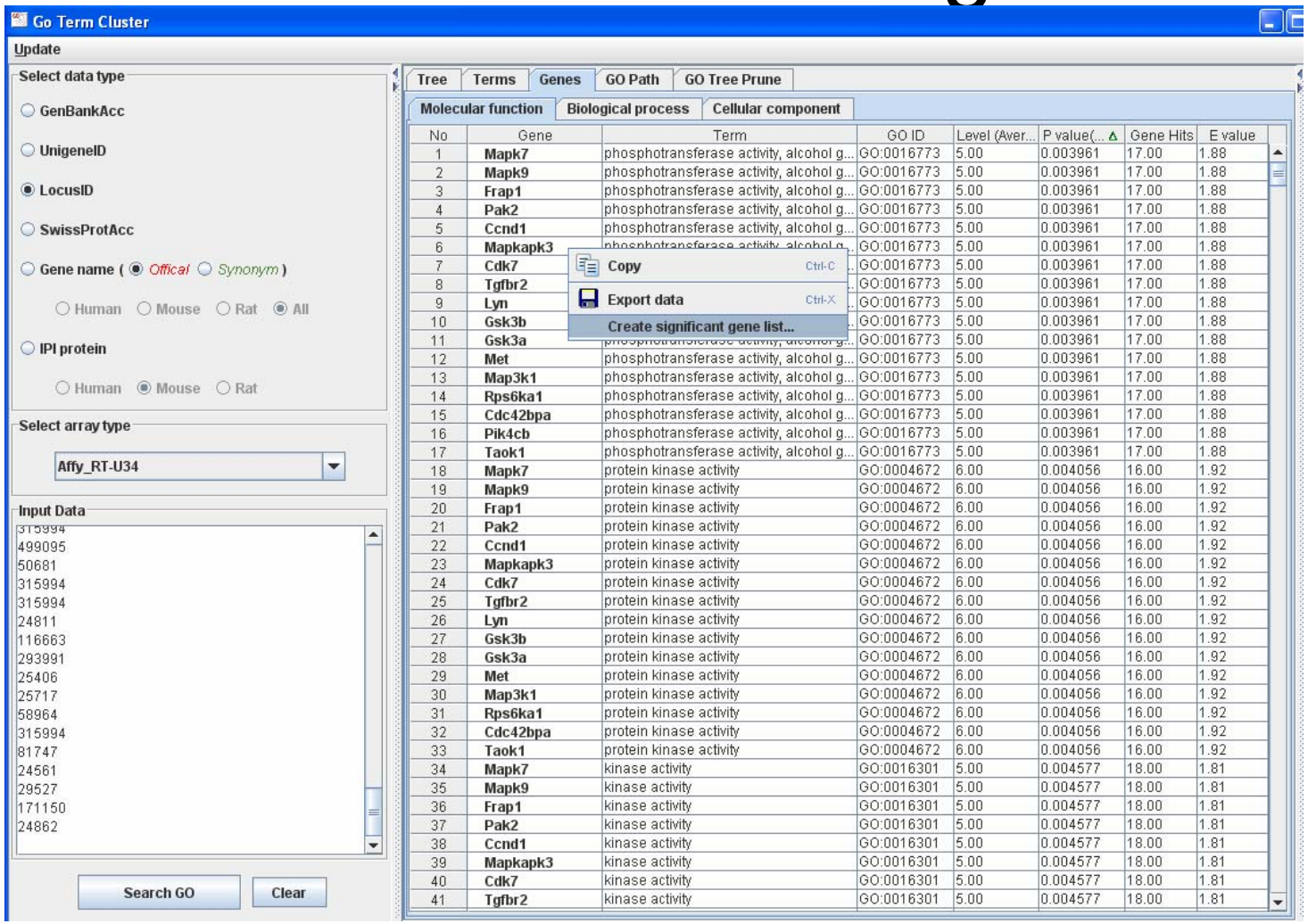
The user can filter out some genes (e.g. spike-in genes) to get a gene list for further normalization.

This image shows a raw dataset using filtering criteria to only display the genes the user inquired. User can open any raw data to get the following spreadsheet.



	GEN_DESCR_MFR	GEN_ID_MFR	LOCUSID	POS_DESCR_MFR	REFSEQ	SPOTID	{D0_T12_B_a} Biotin [D0
Filter->		AFFX					
1	usculus Fas antigen mRNA, complete cds	AFFX-MurFAS_at	14102	114	NM_007987	516597	18.687
2	interleukin 10 mRNA, complete cds	AFFX-MurIL10_at	16153	115	NM_010548	516598	4.788
3	interleukin 2 (IL-2) gene, exon 4	AFFX-MurIL2_at	16183	116		516599	6.44
4	usculus interleukin 4 (Il-4) mRNA, complete cds	AFFX-MurIL4_at	16189	117	NM_021283	516600	1.903
5	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_3_at	24383	131		516614	7847.16
6	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_3_st	24383	132		516615	36.98
7	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_5_at	24383	133		516616	7485.518
8	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_5_st	24383	134		516617	94.191
9	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_M_at	24383	135		516618	7846.66
10	RNA for glyceraldehyde-3-phosphate-dehydrogenase (GAP	AFFX_Rat_GAPDH_M_st	24383	136		516619	123.931
11	norvegicus brain hexokinase mRNA, complete cds (_5, _M	AFFX_Rat_Hexokinase_3_	25058	137	NM_012734	516620	2.739
12	norvegicus brain hexokinase mRNA, complete cds (_5, _M	AFFX_Rat_Hexokinase_5_	25058	138	NM_012734	516621	9.055
13	norvegicus brain hexokinase mRNA, complete cds (_5, _M	AFFX_Rat_Hexokinase_M_	25058	139	NM_012734	516622	9.021
14	ne encoding cytoplasmic beta-actin (_5, _M, _3 represent t	AFFX_Rat_beta-actin_3_at	81822	140		516623	8092.55

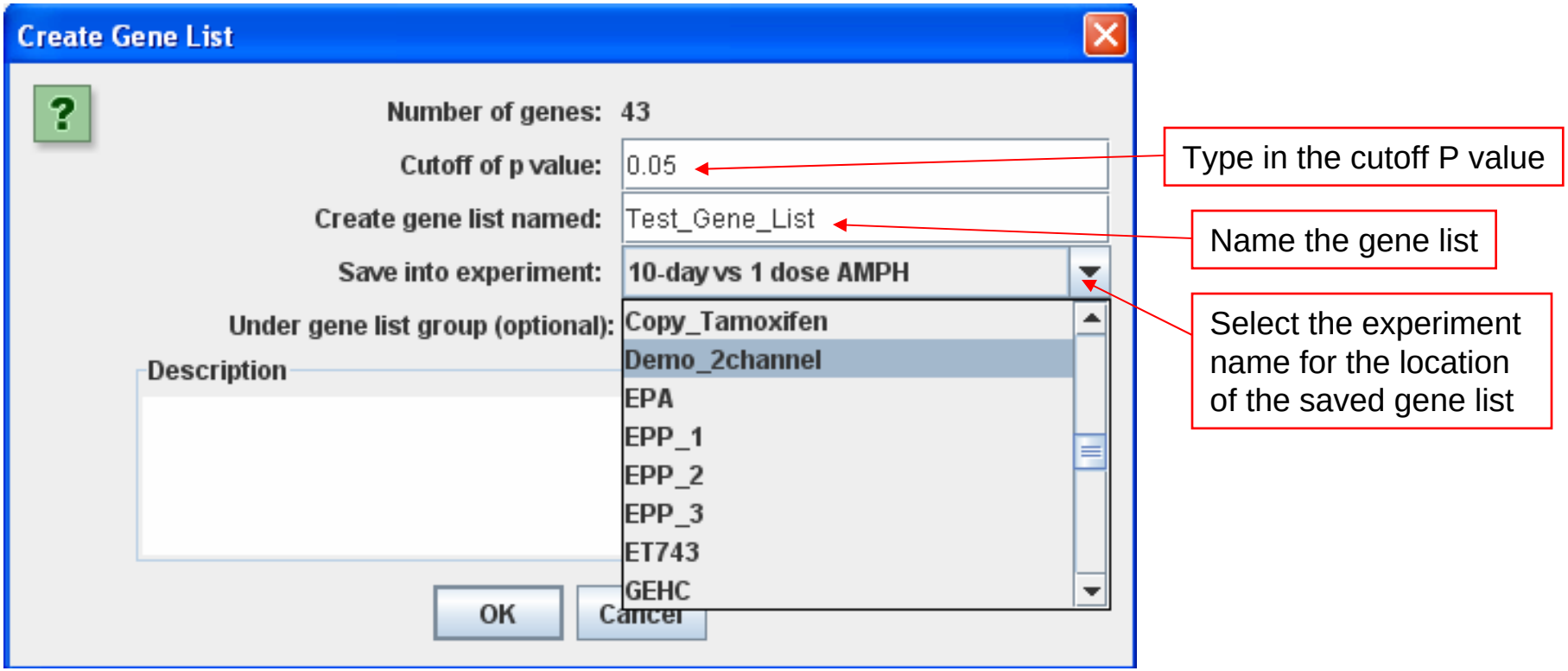
Create Gene List through GOFFA



No	Gene	Term	GO ID	Level (Aver...	P value(...	Gene Hits	E value
1	Mapk7	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
2	Mapk9	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
3	Frap1	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
4	Pak2	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
5	Ccnd1	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
6	Mapkapk3	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
7	Cdk7	[illegible]	GO:0016773	5.00	0.003961	17.00	1.88
8	Tgfbr2	[illegible]	GO:0016773	5.00	0.003961	17.00	1.88
9	Lyn	[illegible]	GO:0016773	5.00	0.003961	17.00	1.88
10	Gsk3b	[illegible]	GO:0016773	5.00	0.003961	17.00	1.88
11	Gsk3a	[illegible]	GO:0016773	5.00	0.003961	17.00	1.88
12	Met	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
13	Map3k1	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
14	Rps6ka1	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
15	Cdc42bpa	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
16	Pik4cb	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
17	Taok1	phosphotransferase activity, alcohol g...	GO:0016773	5.00	0.003961	17.00	1.88
18	Mapk7	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
19	Mapk9	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
20	Frap1	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
21	Pak2	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
22	Ccnd1	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
23	Mapkapk3	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
24	Cdk7	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
25	Tgfbr2	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
26	Lyn	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
27	Gsk3b	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
28	Gsk3a	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
29	Met	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
30	Map3k1	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
31	Rps6ka1	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
32	Cdc42bpa	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
33	Taok1	protein kinase activity	GO:0004672	6.00	0.004056	16.00	1.92
34	Mapk7	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
35	Mapk9	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
36	Frap1	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
37	Pak2	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
38	Ccnd1	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
39	Mapkapk3	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
40	Cdk7	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81
41	Tgfbr2	kinase activity	GO:0016301	5.00	0.004577	18.00	1.81

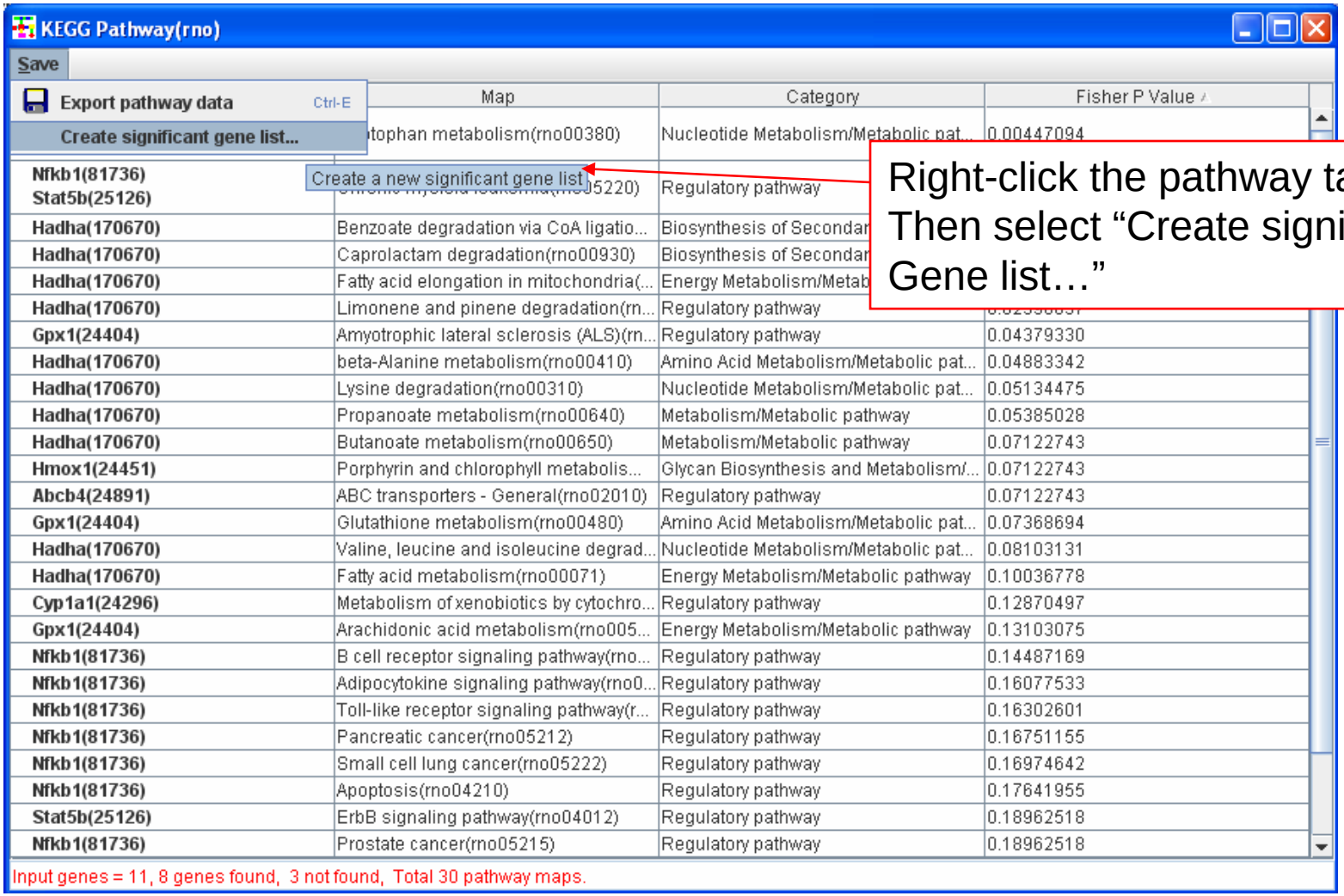
This is the GOFFA table. Click Gene tab, then right-click the gene table, select "Create significant gene list…"

Create Gene List through GOFFA



When the user type in the cutoff p value, the number of genes will change according to the value of P.

Create Significant Gene List through Pathway

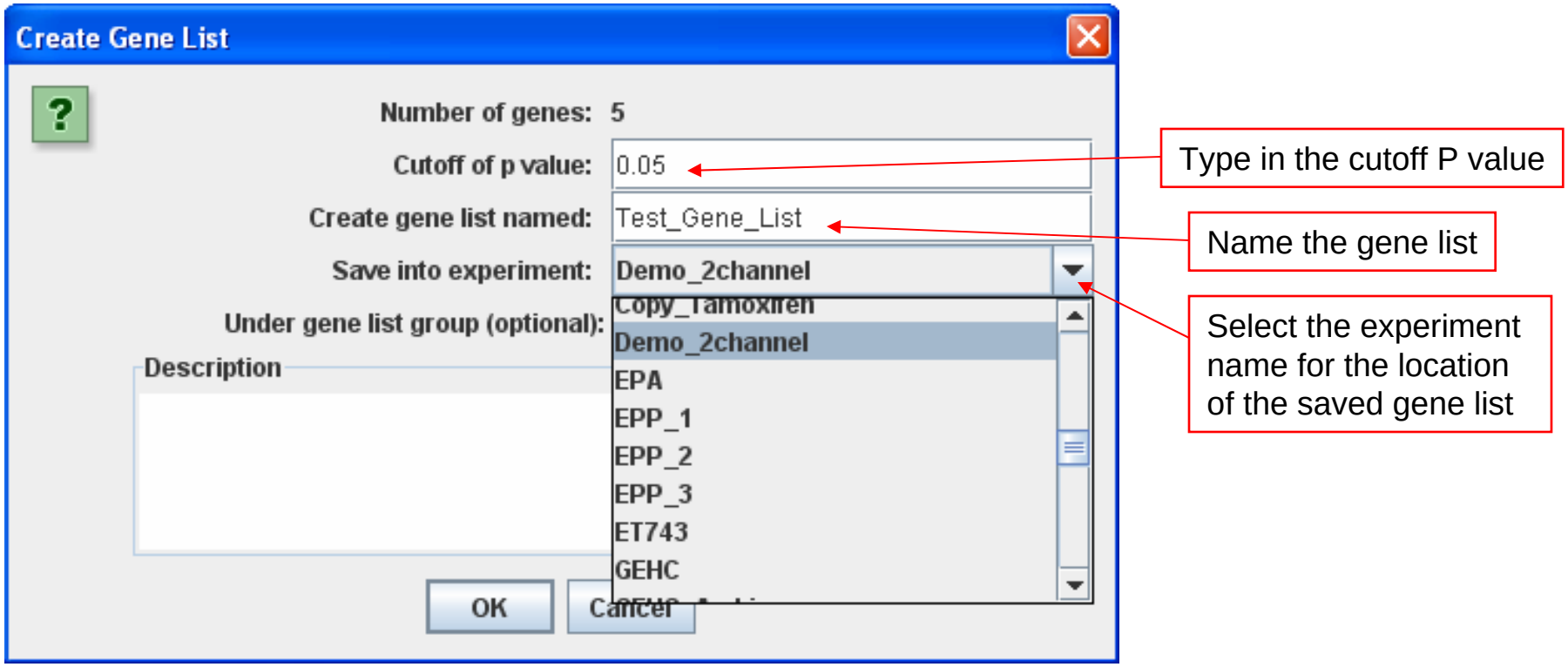


	Map	Category	Fisher P Value
	tophan metabolism(rno00380)	Nucleotide Metabolism/Metabolic pat...	0.00447094
Nfkb1(81736) Stat5b(25126)	5220)	Regulatory pathway	
Hadha(170670)	Benzoate degradation via CoA ligatio...	Biosynthesis of Secondar	
Hadha(170670)	Caprolactam degradation(rno00930)	Biosynthesis of Secondar	
Hadha(170670)	Fatty acid elongation in mitochondria(...	Energy Metabolism/Metab	
Hadha(170670)	Limonene and pinene degradation(rn...	Regulatory pathway	
Gpx1(24404)	Amyotrophic lateral sclerosis (ALS)(rn...	Regulatory pathway	0.04379330
Hadha(170670)	beta-Alanine metabolism(rno00410)	Amino Acid Metabolism/Metabolic pat...	0.04883342
Hadha(170670)	Lysine degradation(rno00310)	Nucleotide Metabolism/Metabolic pat...	0.05134475
Hadha(170670)	Propanoate metabolism(rno00640)	Metabolism/Metabolic pathway	0.05385028
Hadha(170670)	Butanoate metabolism(rno00650)	Metabolism/Metabolic pathway	0.07122743
Hmox1(24451)	Porphyrin and chlorophyll metabolis...	Glycan Biosynthesis and Metabolism/...	0.07122743
Abcb4(24891)	ABC transporters - General(rno02010)	Regulatory pathway	0.07122743
Gpx1(24404)	Glutathione metabolism(rno00480)	Amino Acid Metabolism/Metabolic pat...	0.07368694
Hadha(170670)	Valine, leucine and isoleucine degrad...	Nucleotide Metabolism/Metabolic pat...	0.08103131
Hadha(170670)	Fatty acid metabolism(rno00071)	Energy Metabolism/Metabolic pathway	0.10036778
Cyp1a1(24296)	Metabolism of xenobiotics by cytochro...	Regulatory pathway	0.12870497
Gpx1(24404)	Arachidonic acid metabolism(rno005...	Energy Metabolism/Metabolic pathway	0.13103075
Nfkb1(81736)	B cell receptor signaling pathway(rno...	Regulatory pathway	0.14487169
Nfkb1(81736)	Adipocytokine signaling pathway(rno0...	Regulatory pathway	0.16077533
Nfkb1(81736)	Toll-like receptor signaling pathway(r...	Regulatory pathway	0.16302601
Nfkb1(81736)	Pancreatic cancer(rno05212)	Regulatory pathway	0.16751155
Nfkb1(81736)	Small cell lung cancer(rno05222)	Regulatory pathway	0.16974642
Nfkb1(81736)	Apoptosis(rno04210)	Regulatory pathway	0.17641955
Stat5b(25126)	ErbB signaling pathway(rno04012)	Regulatory pathway	0.18962518
Nfkb1(81736)	Prostate cancer(rno05215)	Regulatory pathway	0.18962518

Input genes = 11, 8 genes found, 3 not found, Total 30 pathway maps.

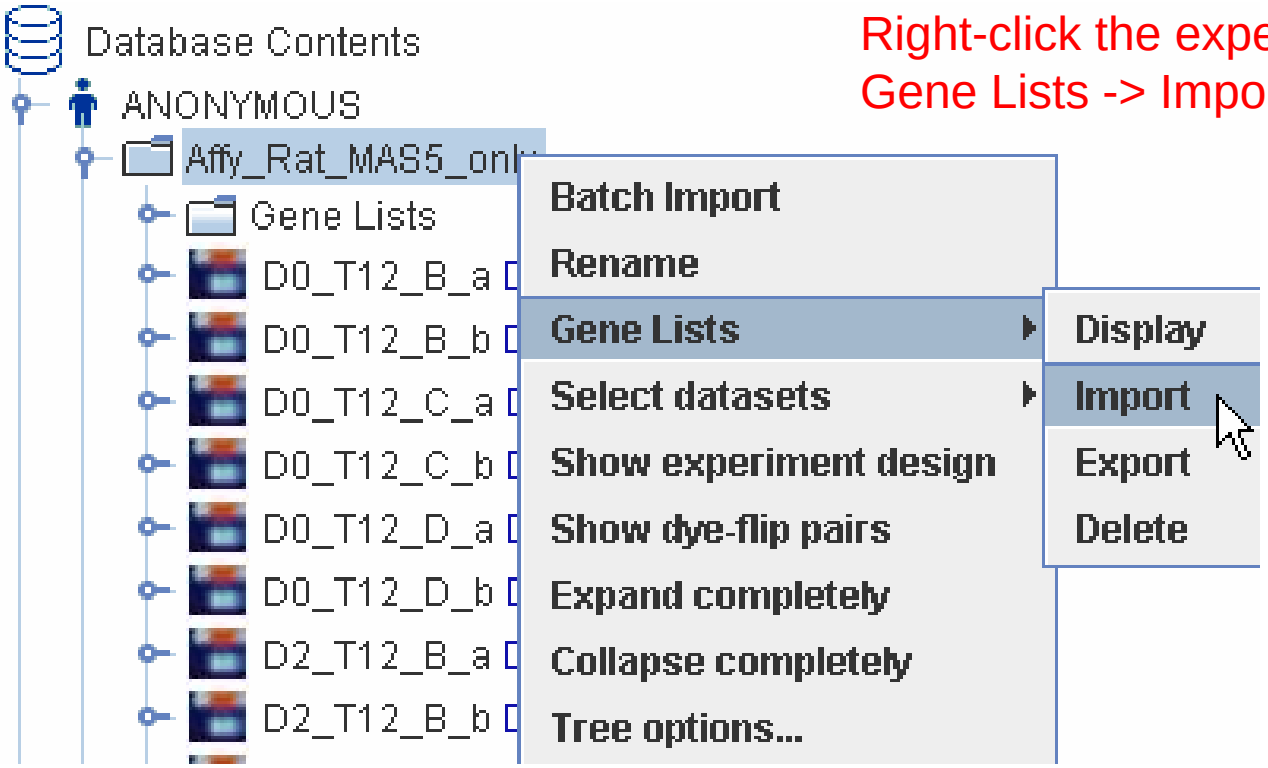
Right-click the pathway table, Then select “Create significant Gene list…”

Create Significant Gene List through Pathway

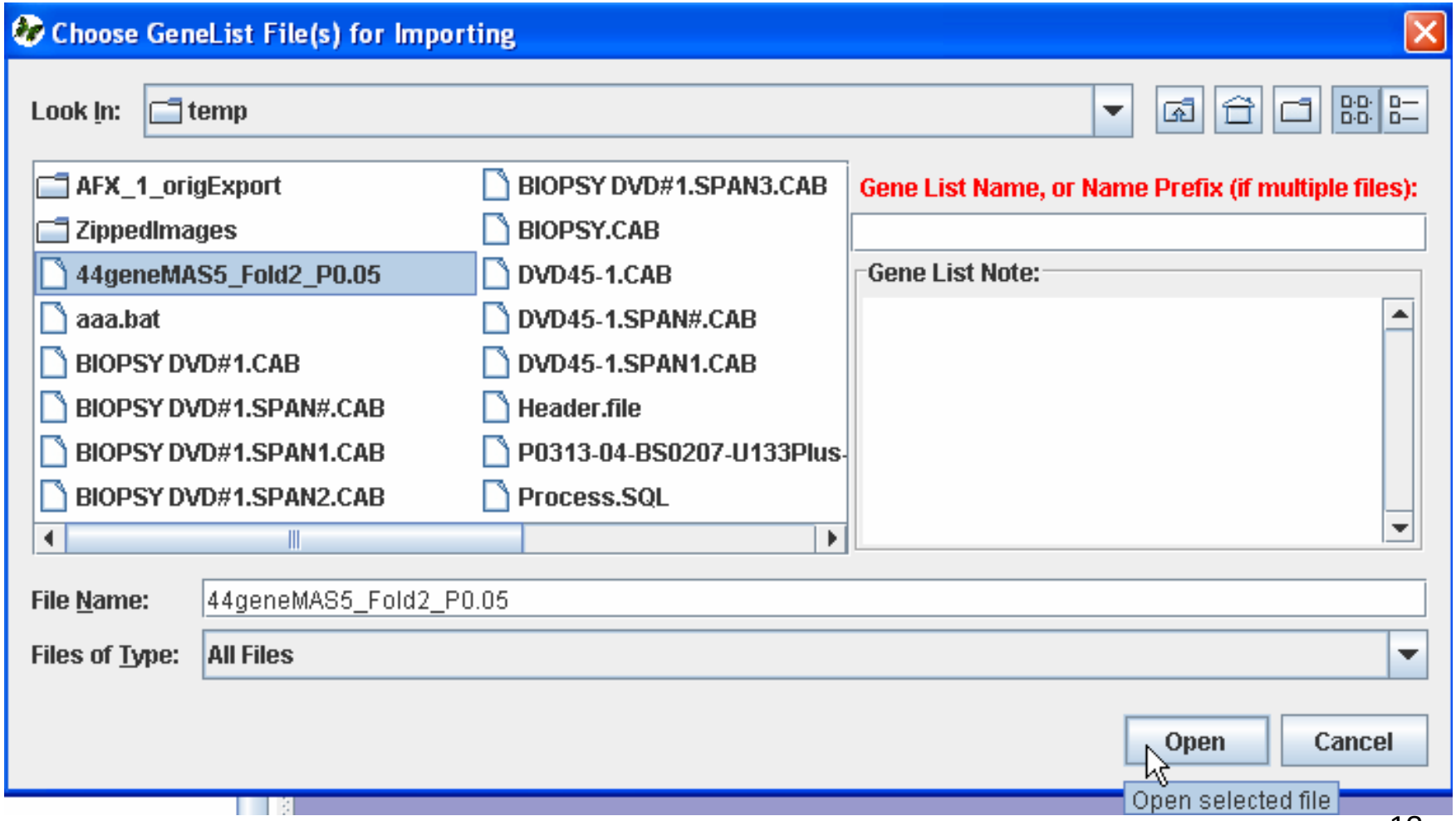


Import/Export a gene list

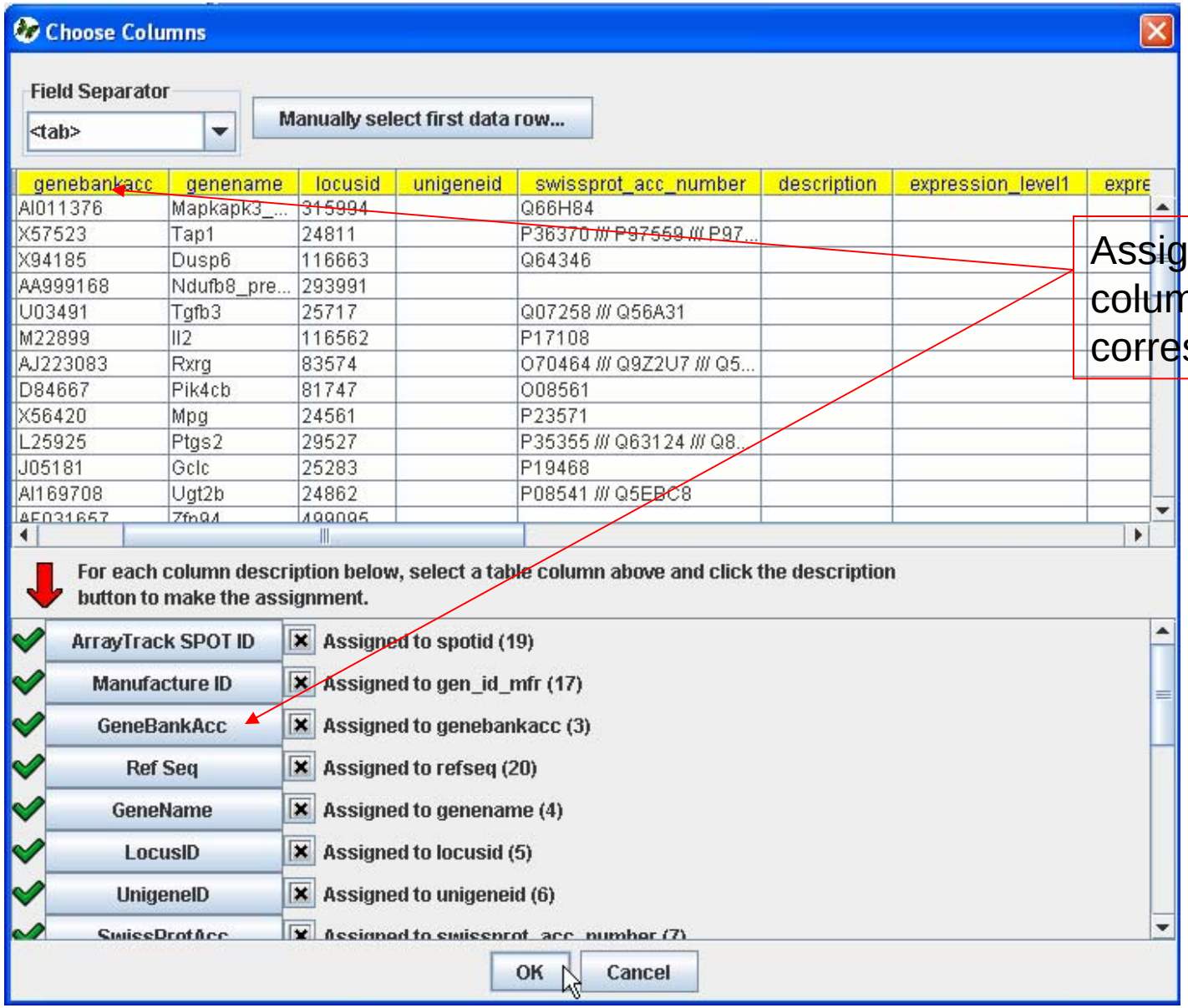
Right-click the experiment, then choose Gene Lists -> Import



Import/Export a gene list – cont.

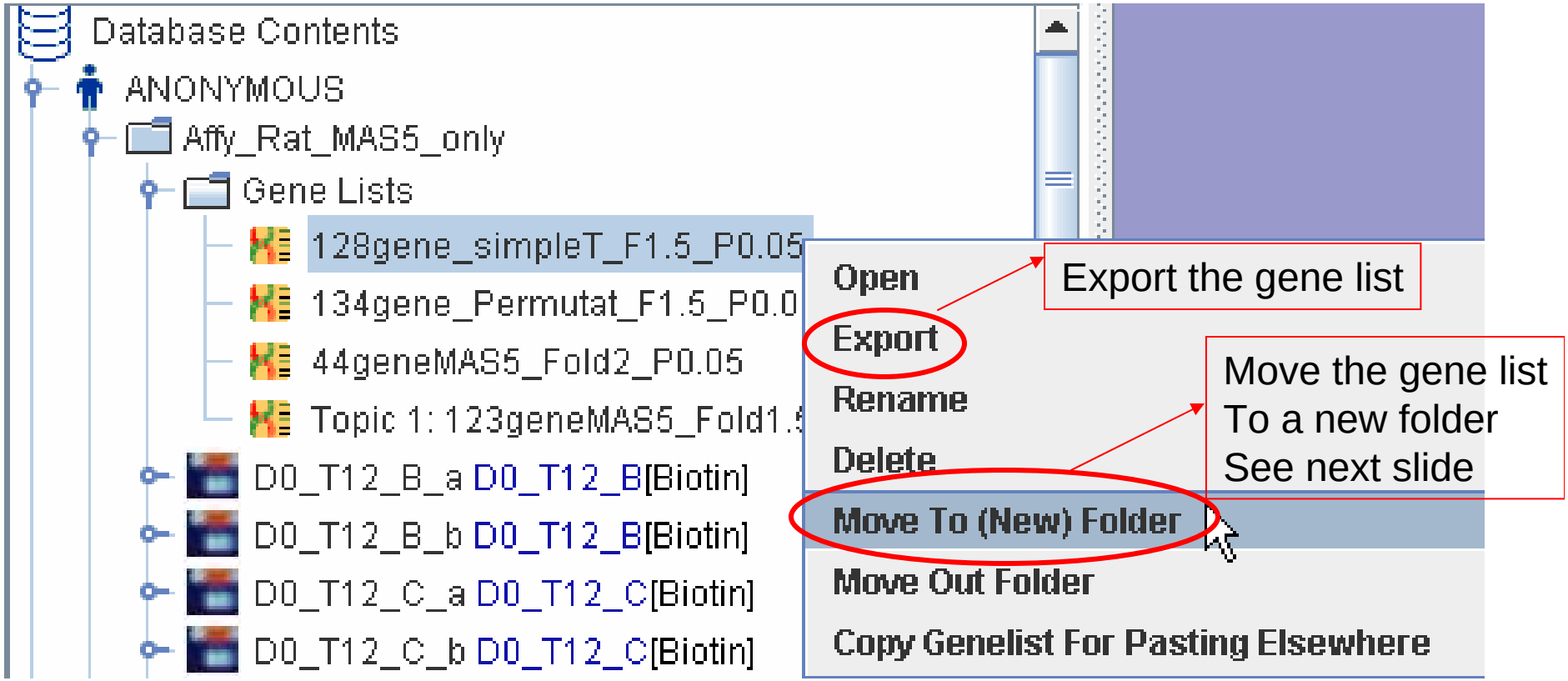


Import/Export a gene list – cont.



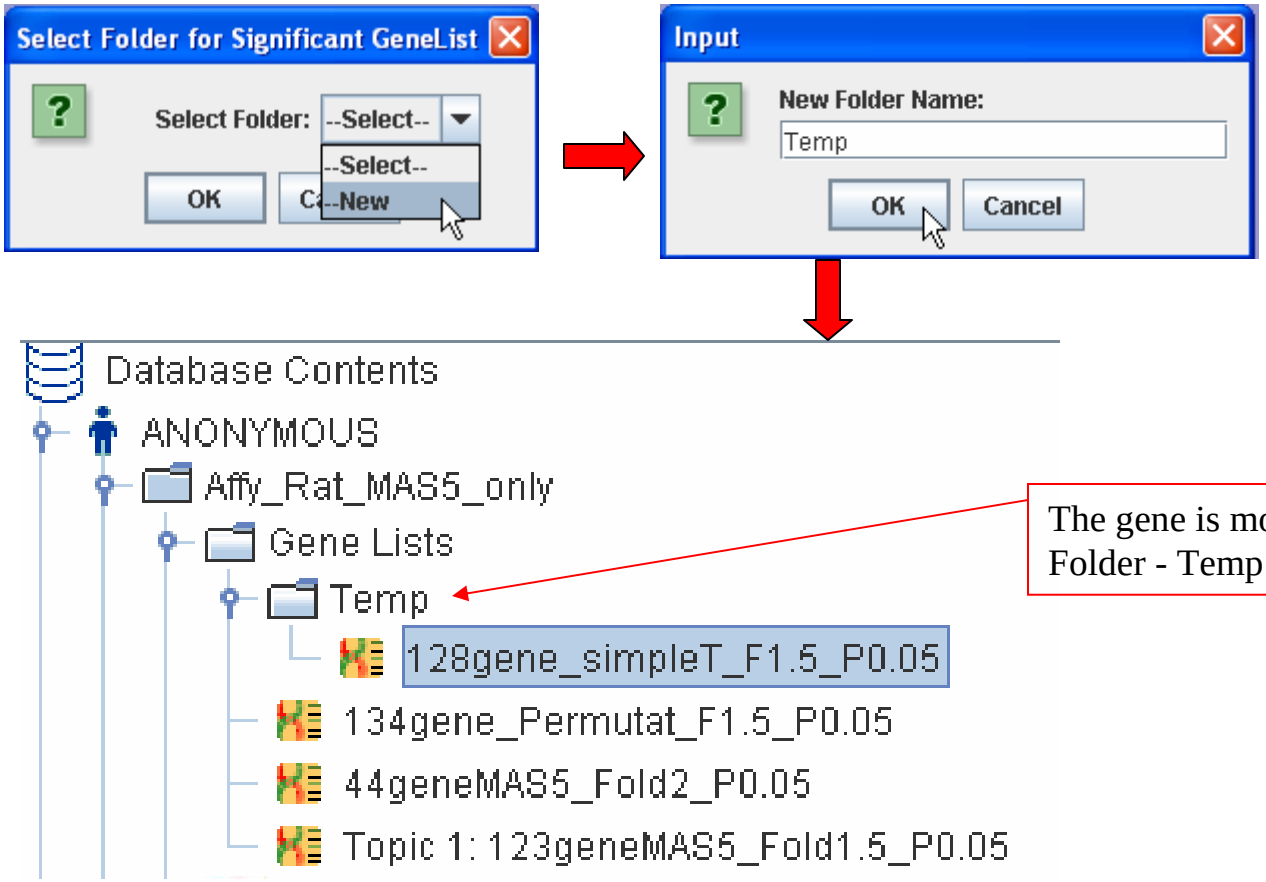
Assign table columns to the corresponding fields

Import/Export a gene list - continued

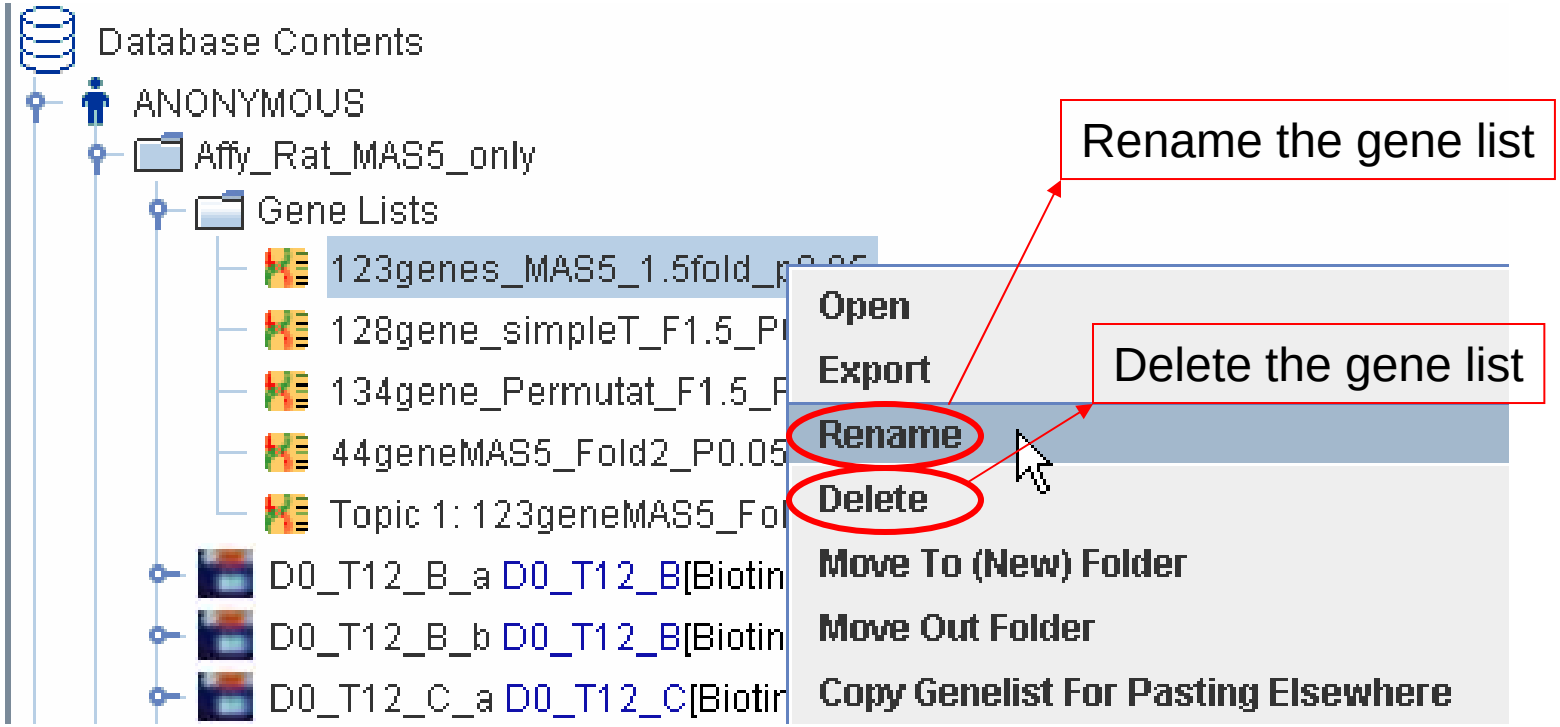


Import/Export a gene list - continued

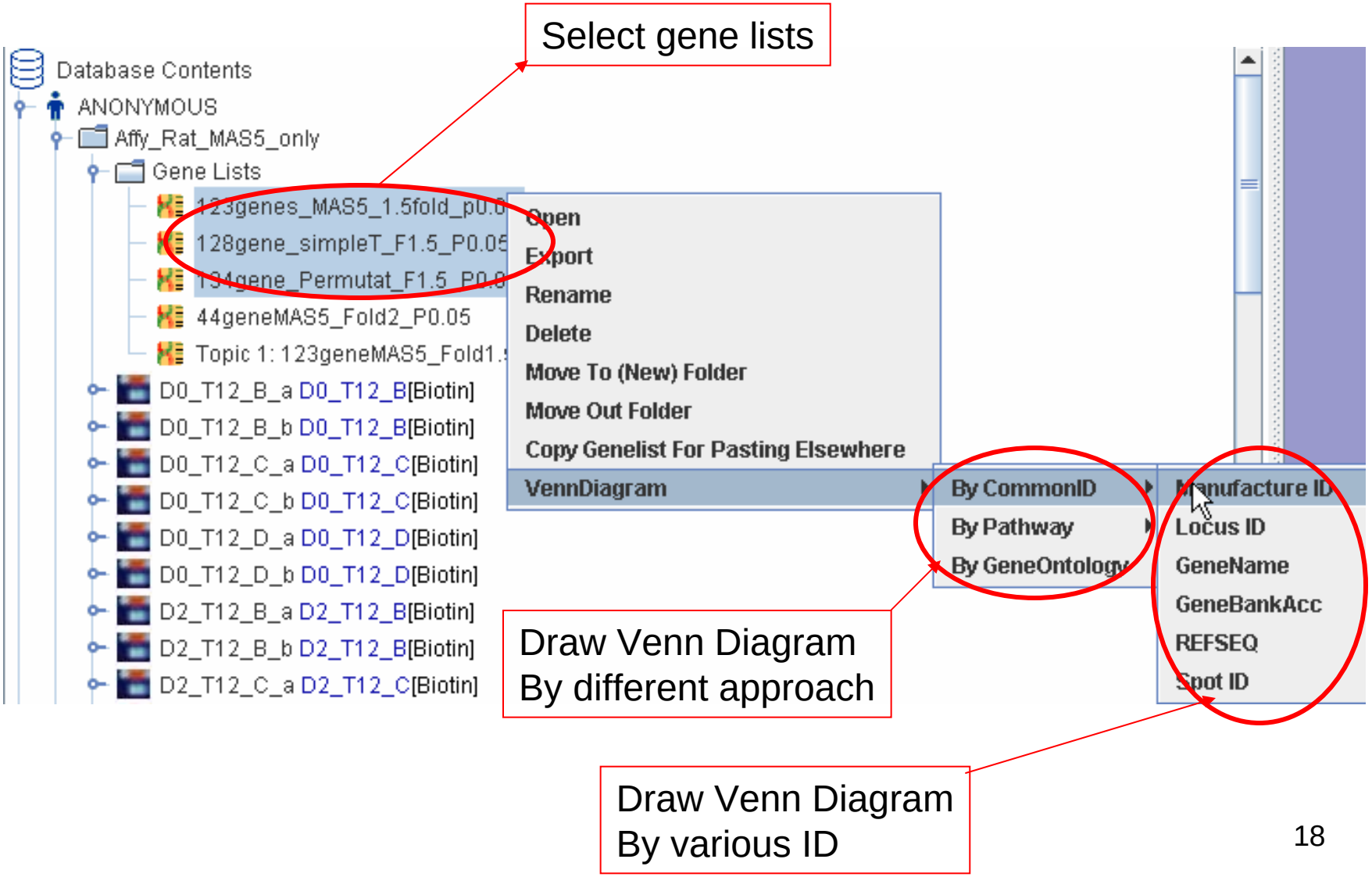
Move gene list to a new folder



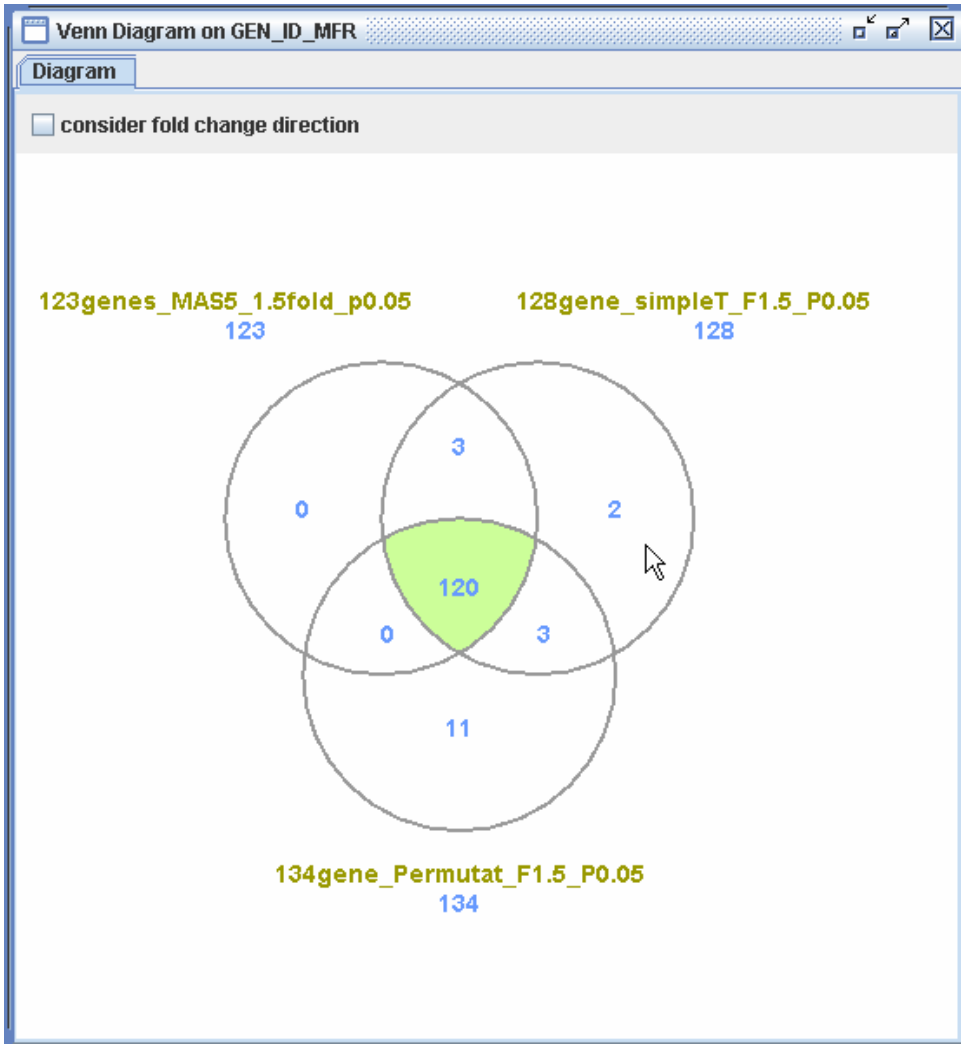
Import/Export a gene list - continued



Draw Venn Diagram from gene lists

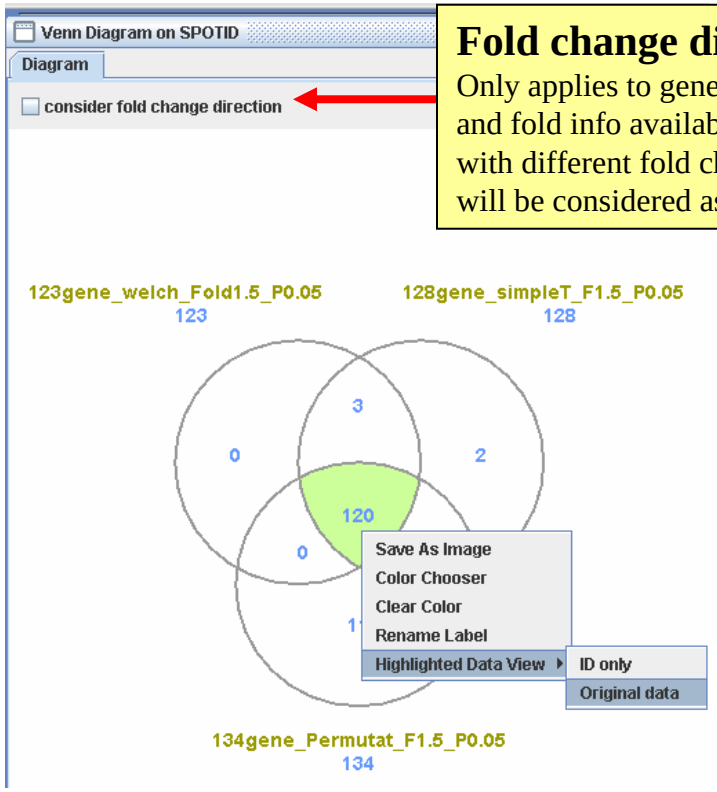


Draw Venn Diagram from gene lists -continued



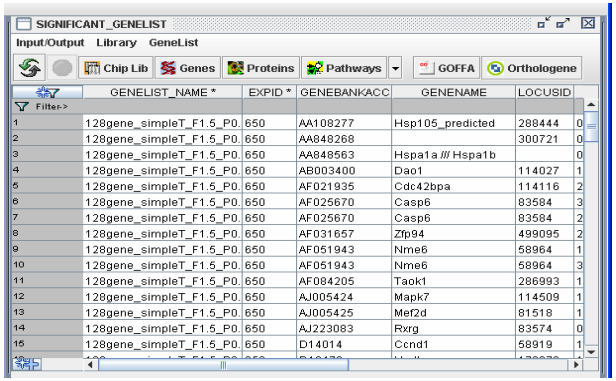
More detail about Venn Diagram is explained in Tutorial 3.

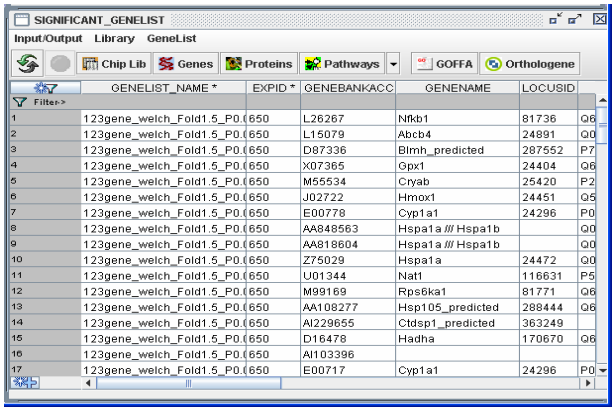
Draw Venn Diagram from gene lists -continued

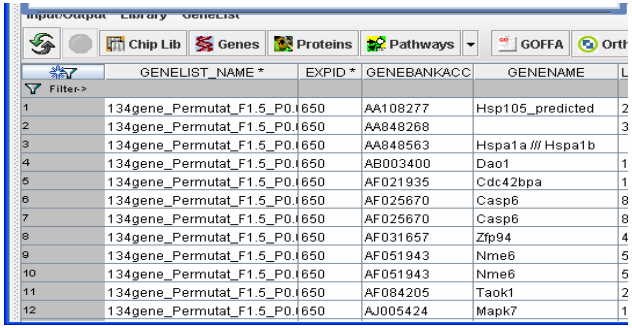


Fold change direction option

Only applies to genelists saved in ArrayTrack and fold info available. If gene id is same, but with different fold change direction, then it will be considered as different genes.



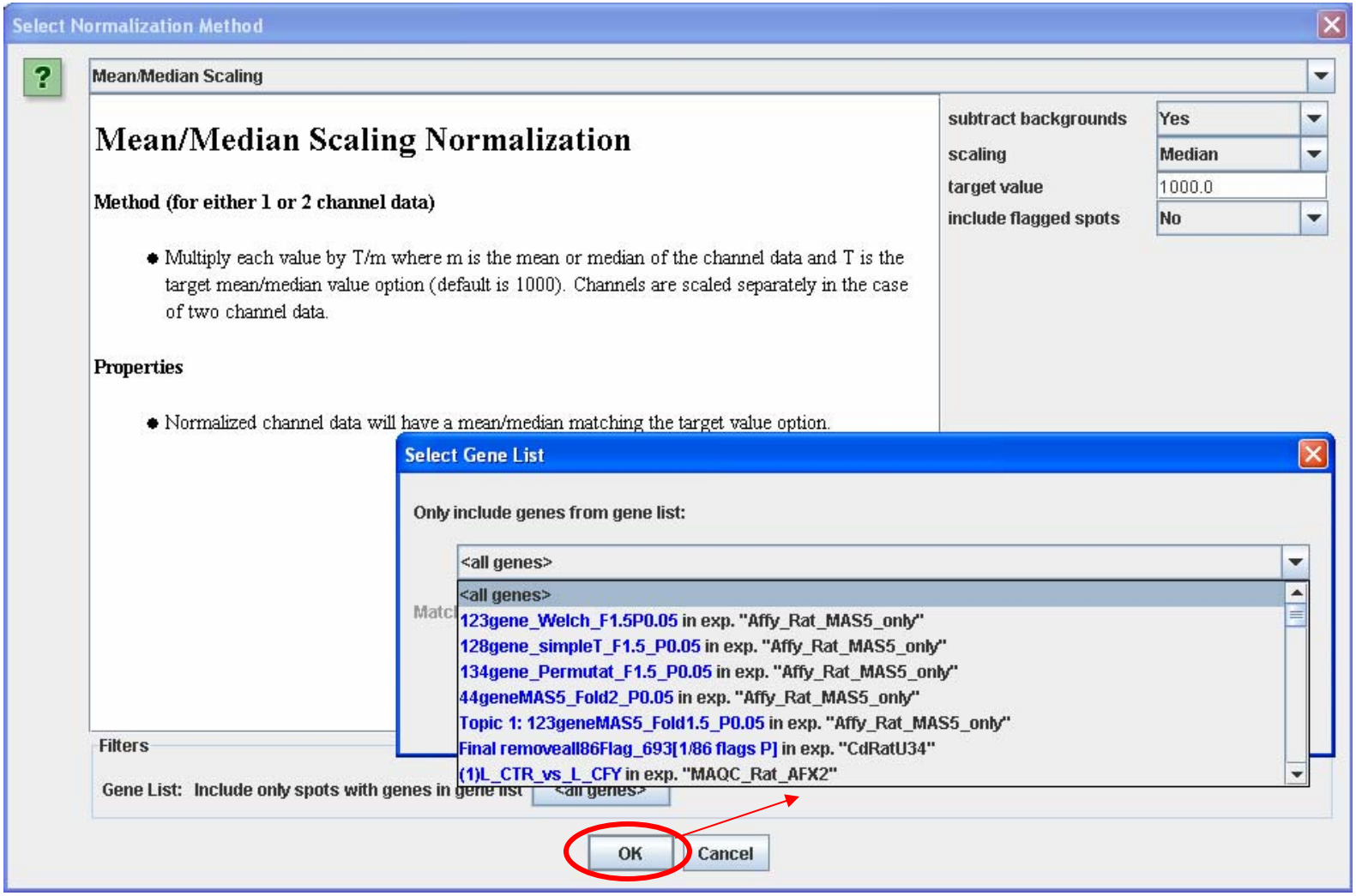




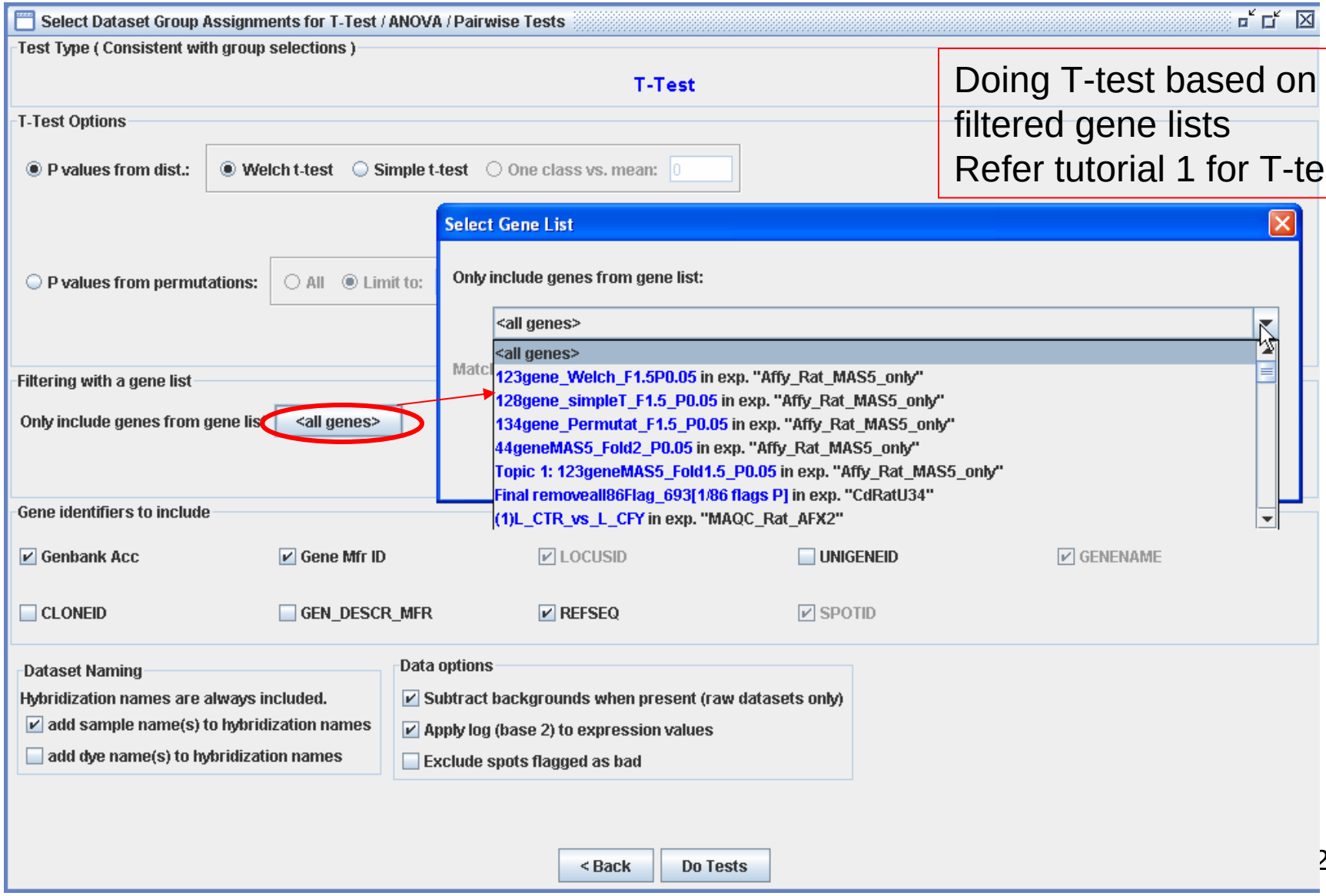
By clicking any sections from VennDiagram, the regions will be highlighted. Then right click to choose the options. One option allows display the selected genes in the original data sheet.

Conduct normalization filtered by a gene list

User can do data normalization based on a filtered gene list. Refer tutorial 7 for Normalization methods.



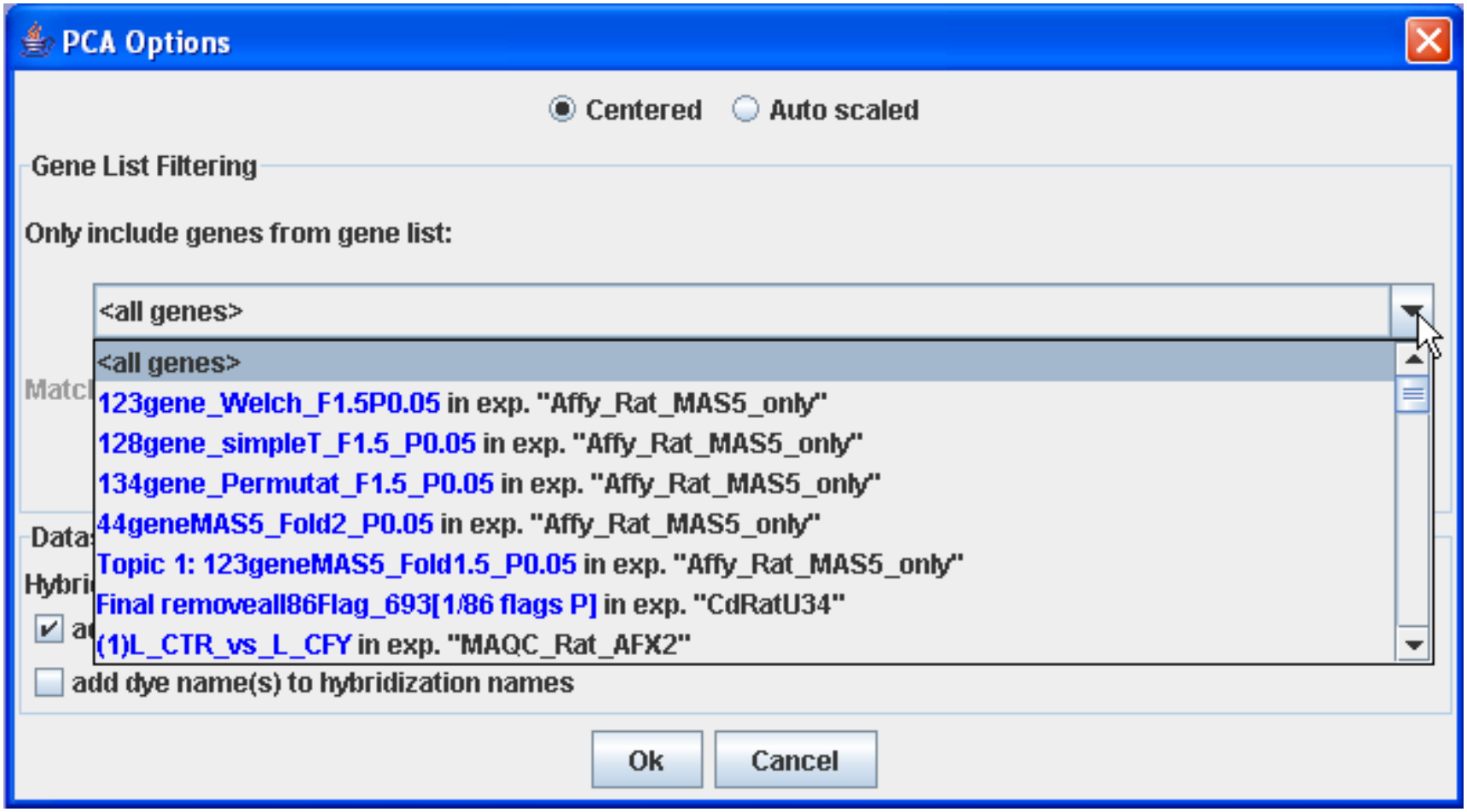
Conduct statistical analysis based on a gene list



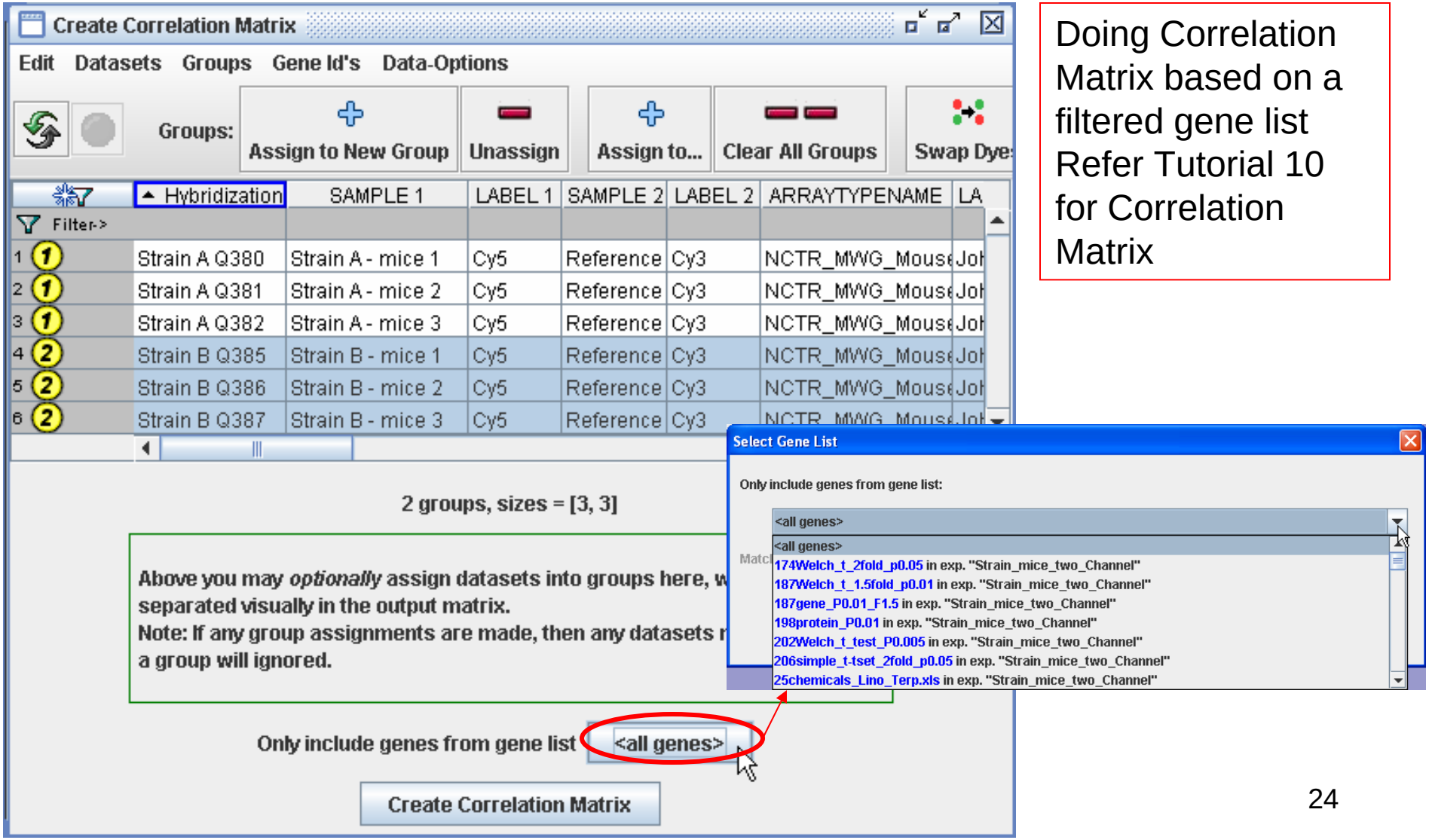
Doing T-test based on a filtered gene lists
Refer tutorial 1 for T-test

Conduct statistical analysis based on a gene list - continued

Doing PCA based on a filtered gene lists. Refer Tutorial 4 for PCA



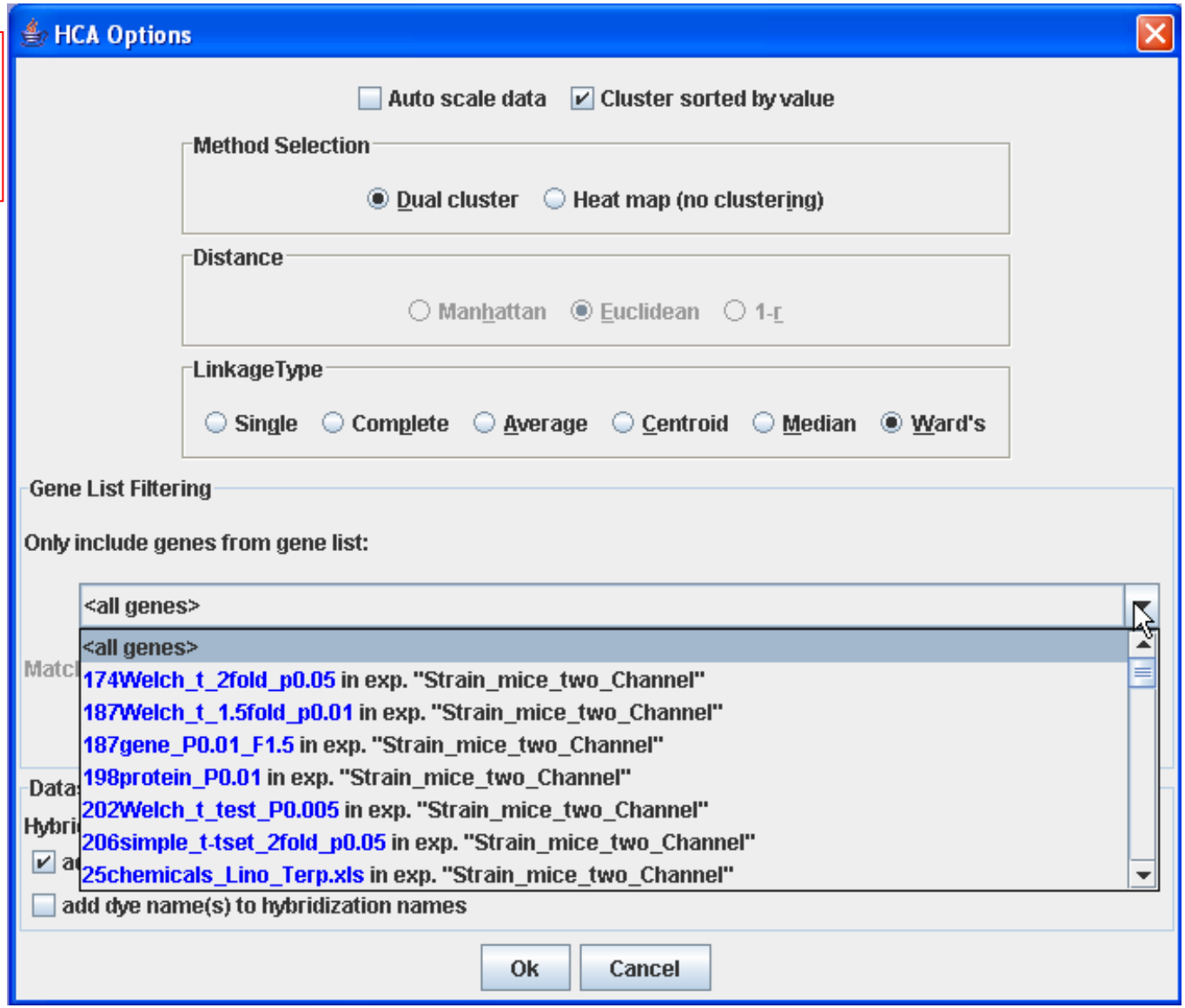
Conduct statistical analysis based on a gene list - continued



Doing Correlation Matrix based on a filtered gene list Refer Tutorial 10 for Correlation Matrix

Conduct statistical analysis based on a gene list - continued

Doing HCA based on
a filtered gene list
Refer Tutorial 4 for HCA

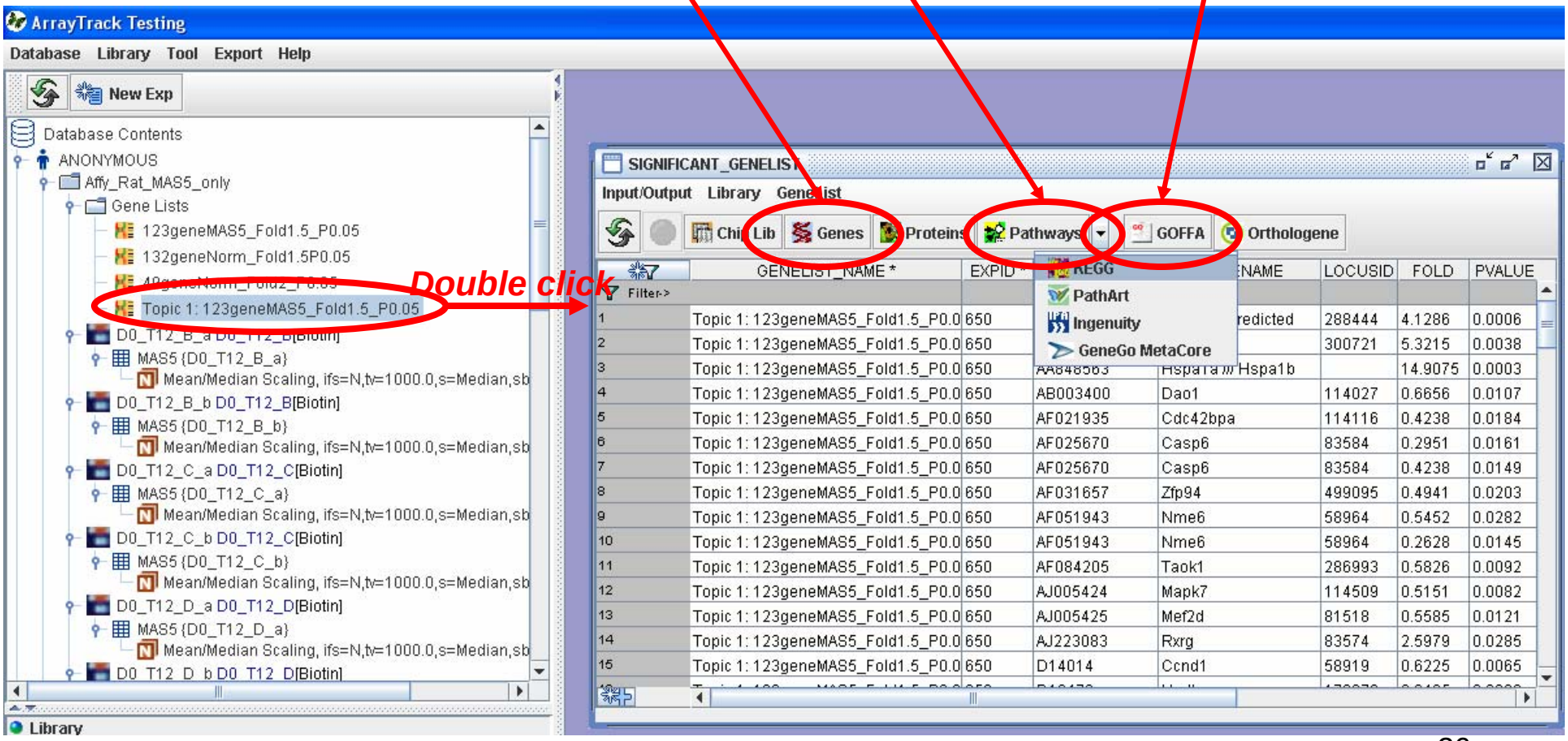


Biological interpretation using ArrayTrack Pathway, GO and other tools

Individual gene annotation and analysis

Pathway analysis

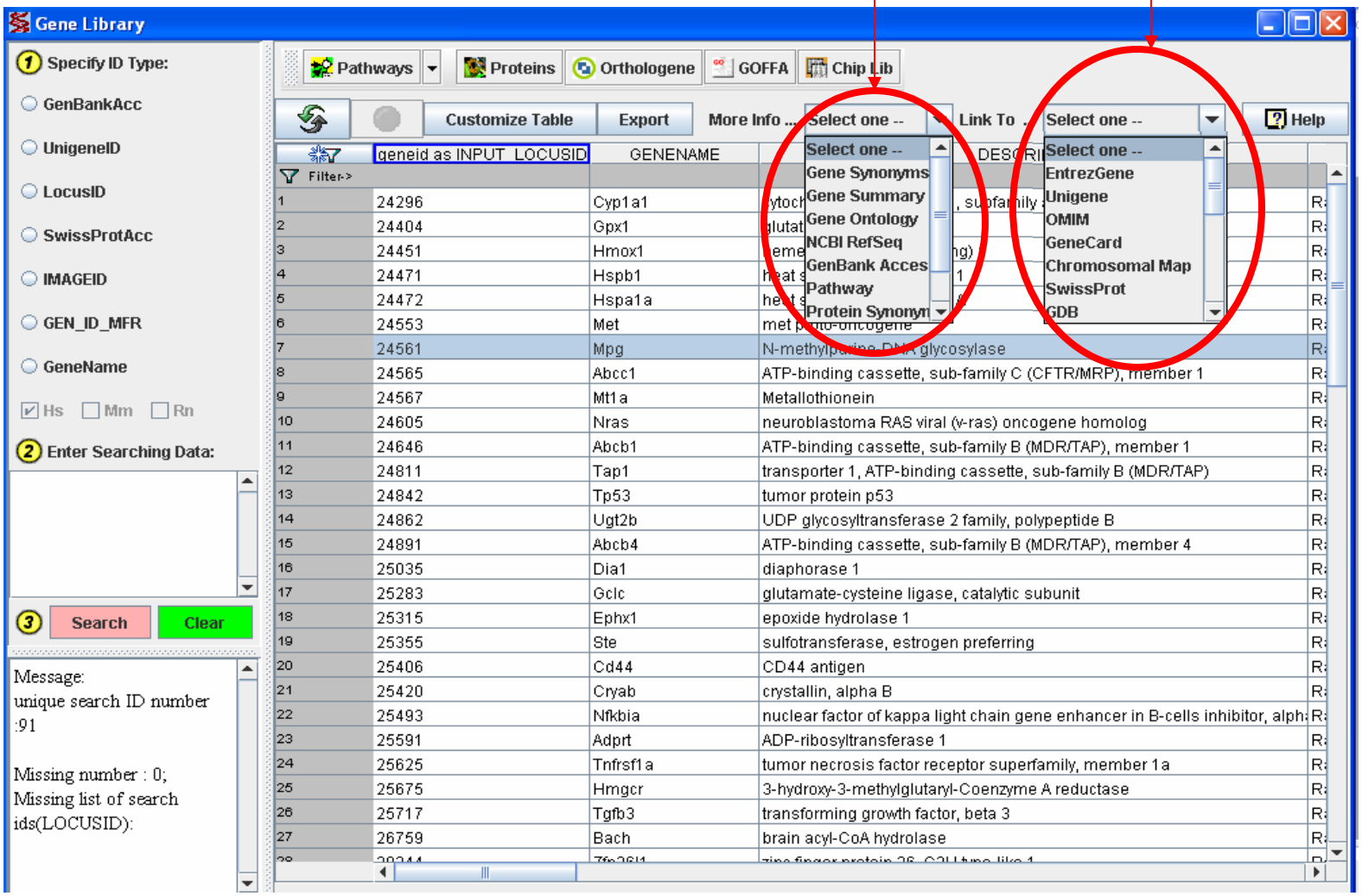
GO-based functional analysis



Individual gene annotation and analysis using GeneLib

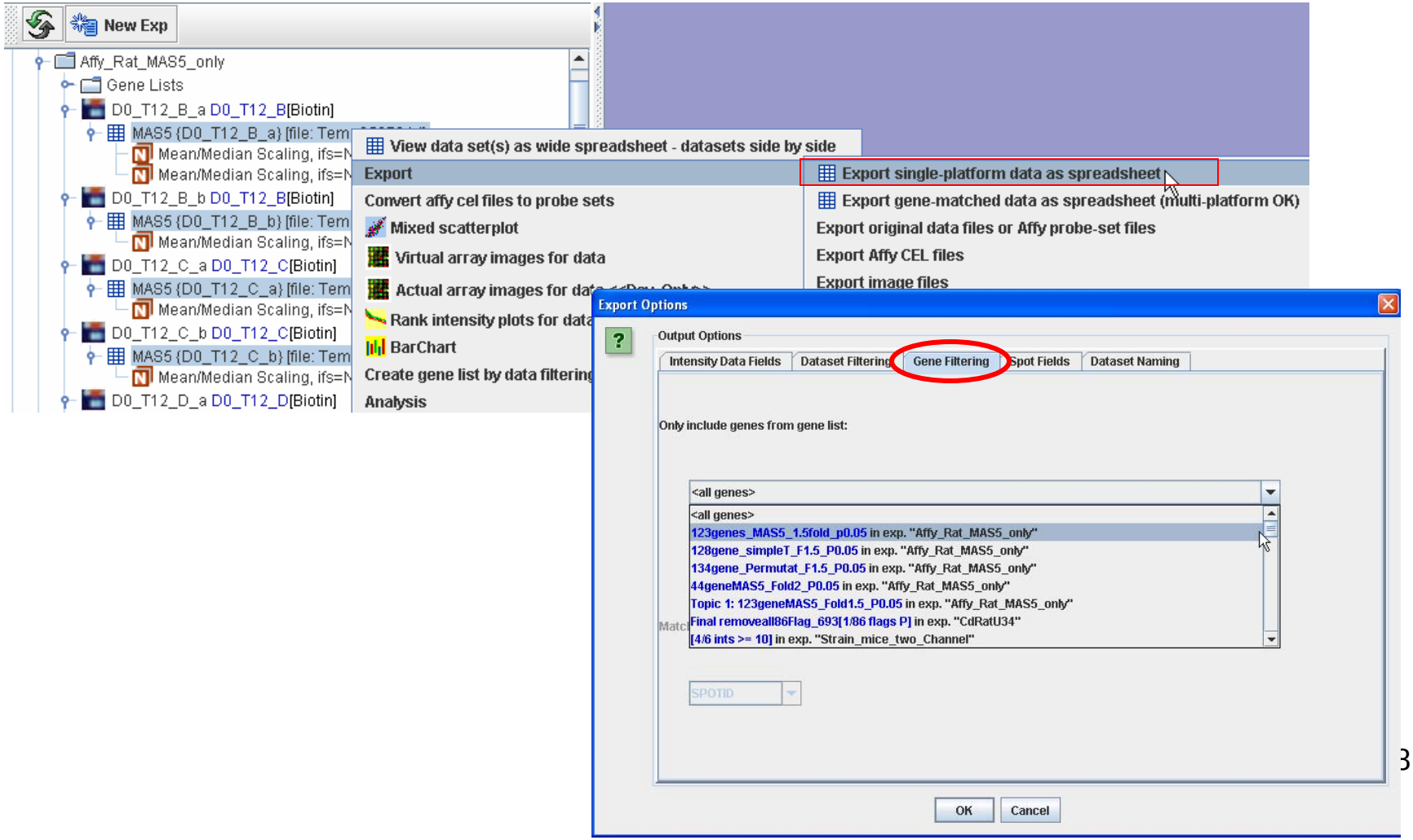
Summary info for the highlighted gene

Link to other pub databases for the highlighted gene



Export a dataset by specifying the gene list

The user can export data with only selected genes. Refer Tutorial 9 for data export.



Array Track

Tutorial 7: Normalization Methods

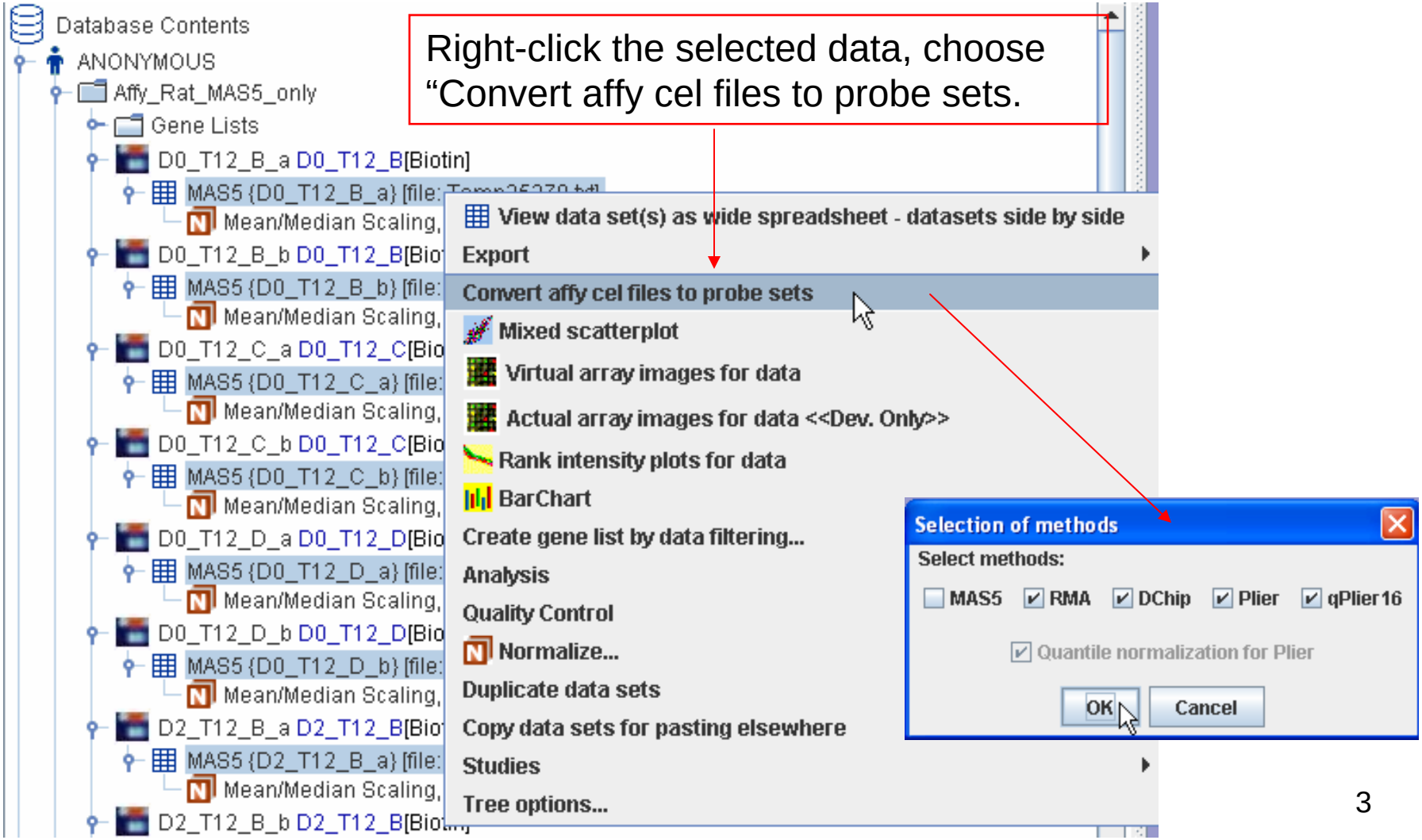
Normalization methods for Affymetrix data

- MAS 5
- RMA
- DChip
- Plier
- Plier +16

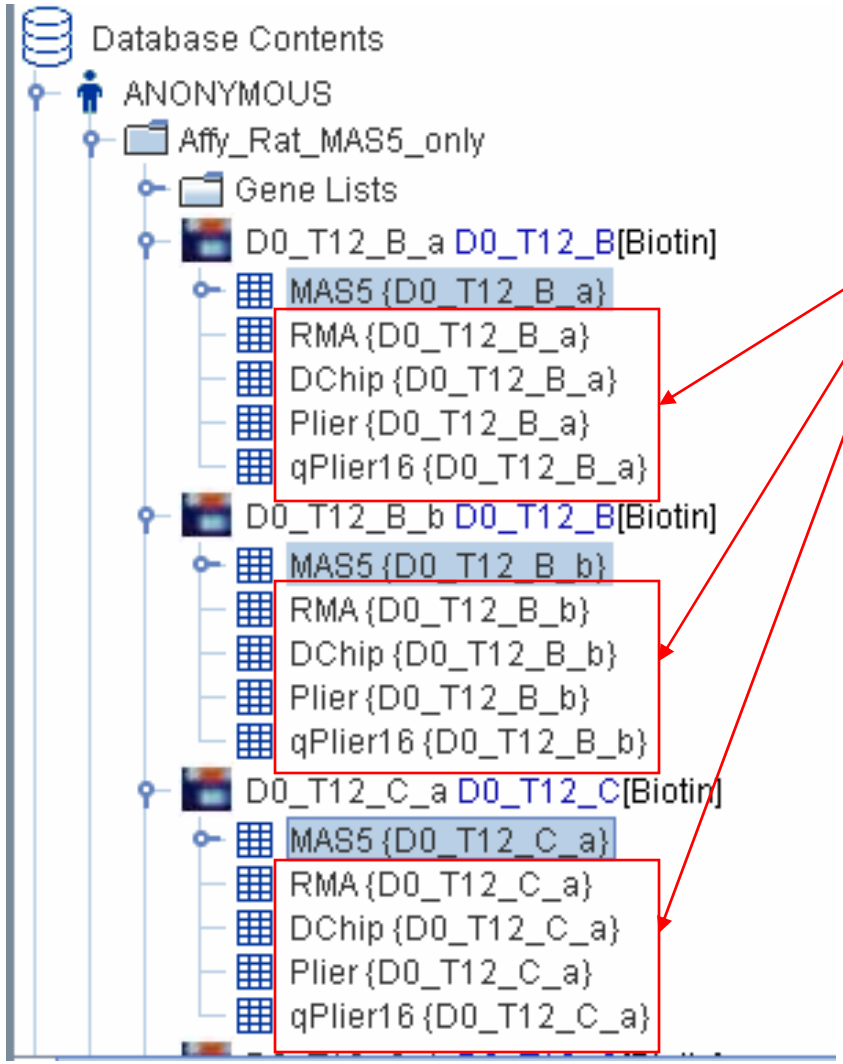
Can be applied on single array

Can be applied on multiple arrays

Normalization methods for Affymetrix data - continued



Normalization methods for Affymetrix data - continued



The converted probe sets files will be shown under the experiment.

Other normalization methods

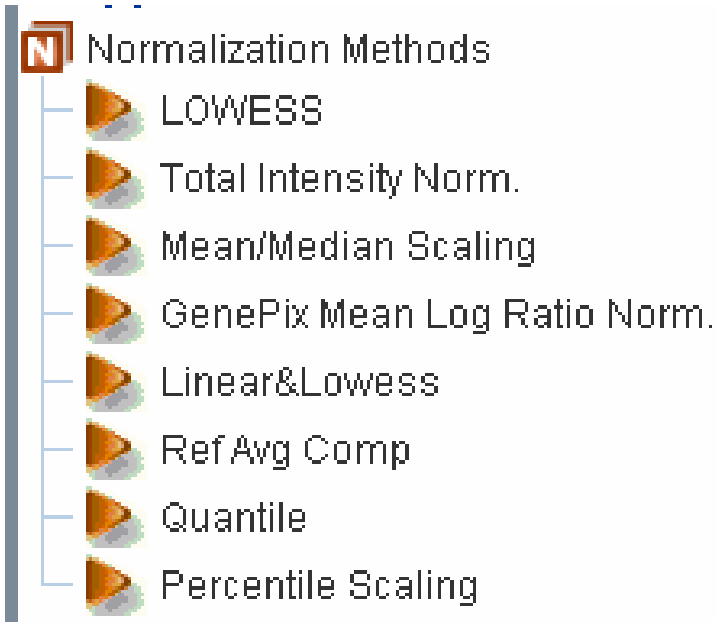
The following methods are for one and/or two channels

- Lowess
- Total intensity norm
- Linear Lowess
- GenePix Mean Log ratio
- Mean/Median scaling
- Percentile Scaling
- Ref Avg Comp
- Quantile

For 2 channel only

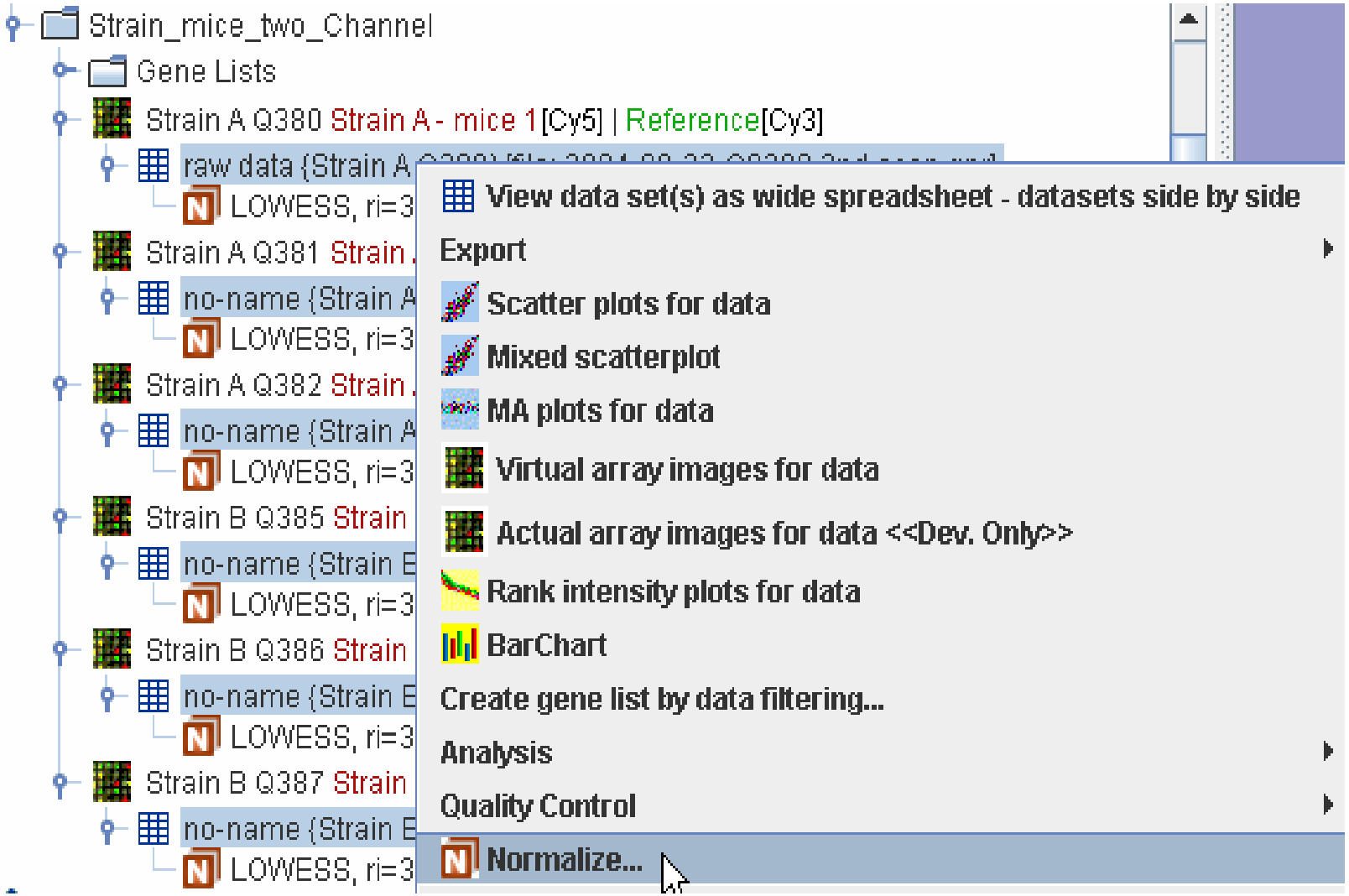
For either 1 or 2 channel

For 1 channel only



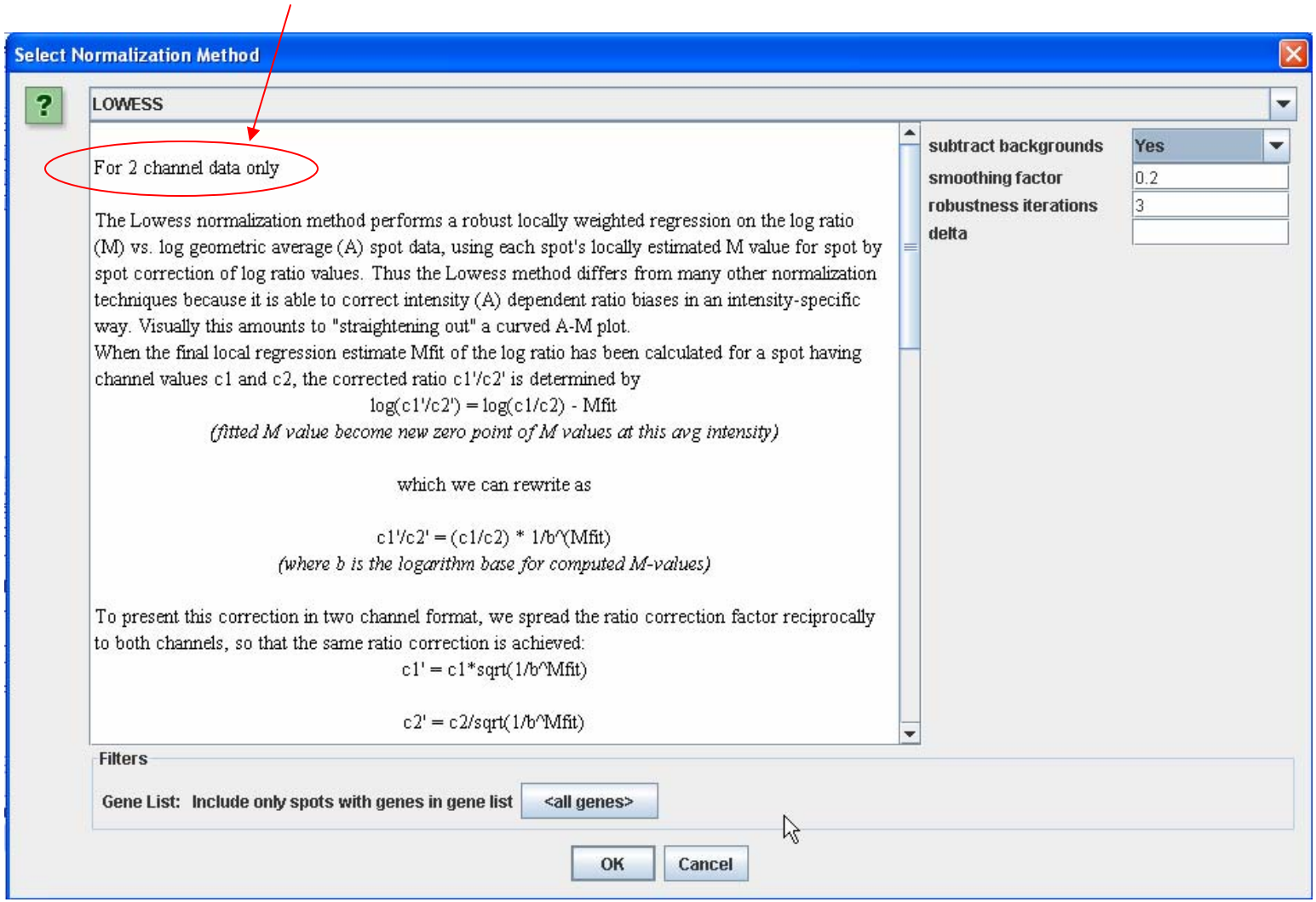
Normalize Data

Select the data, right-click, choose “Normalize…”

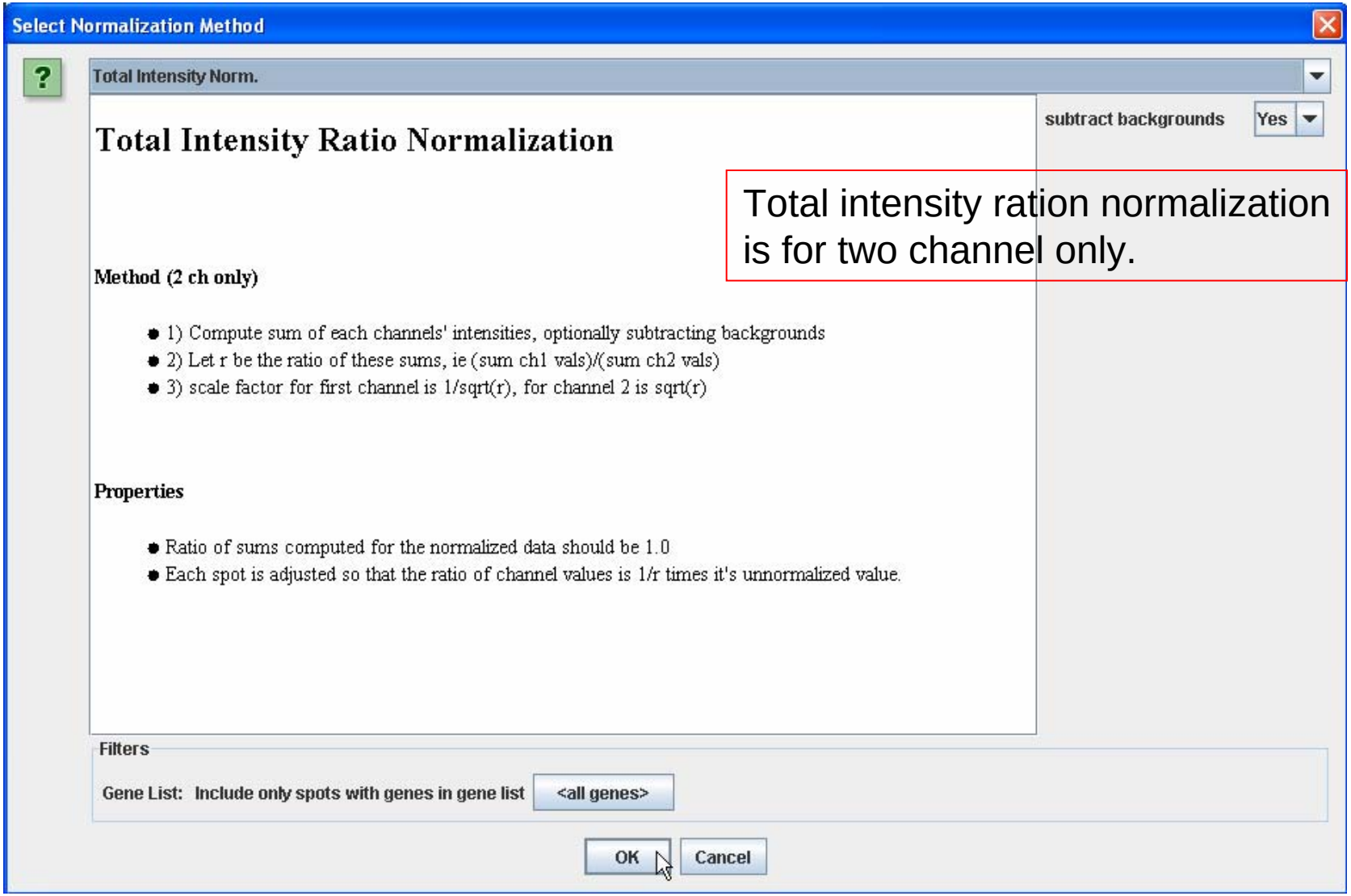


Lowess

Lowess is for two channel data only.

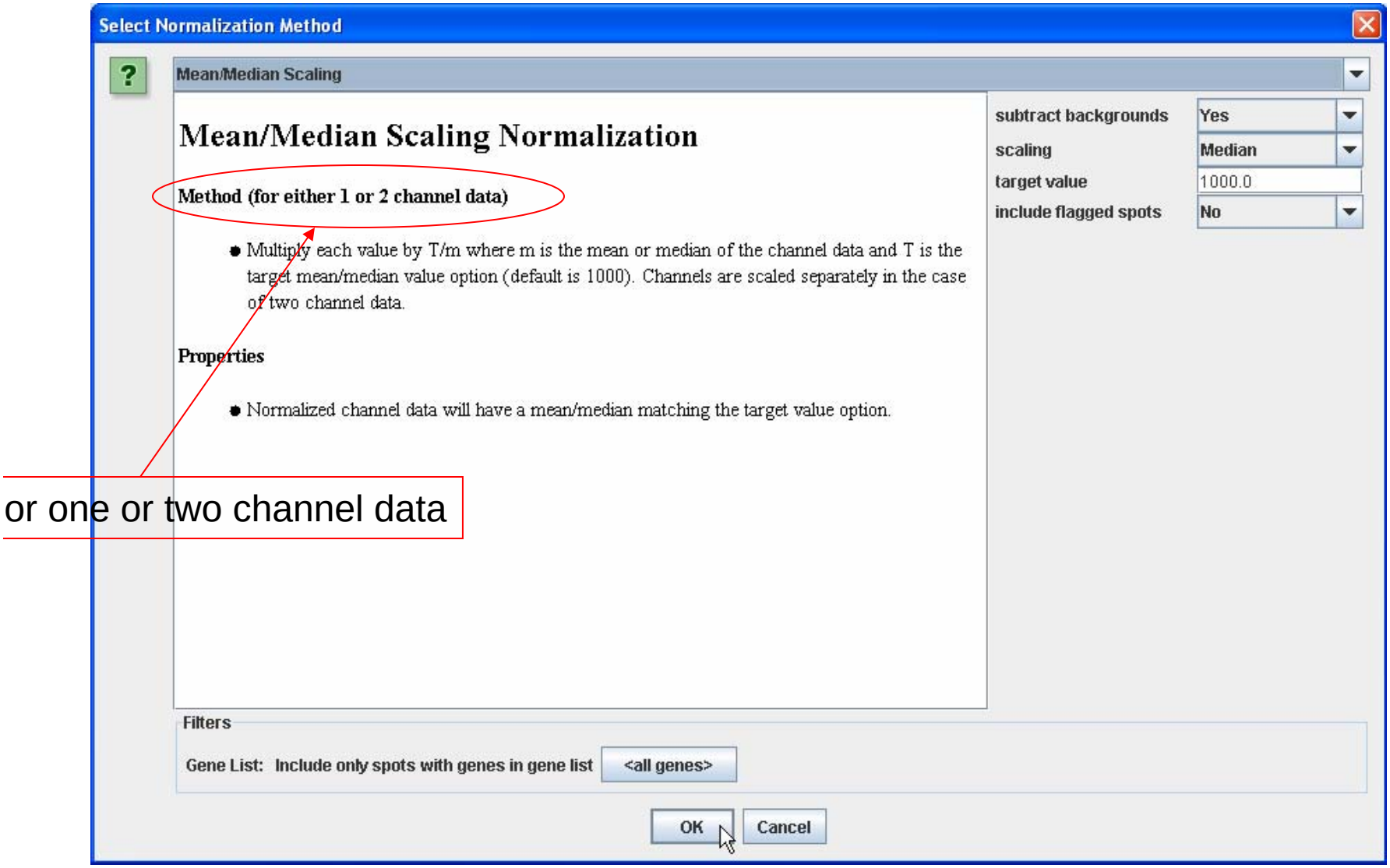


Total Intensity Ratio Normalization



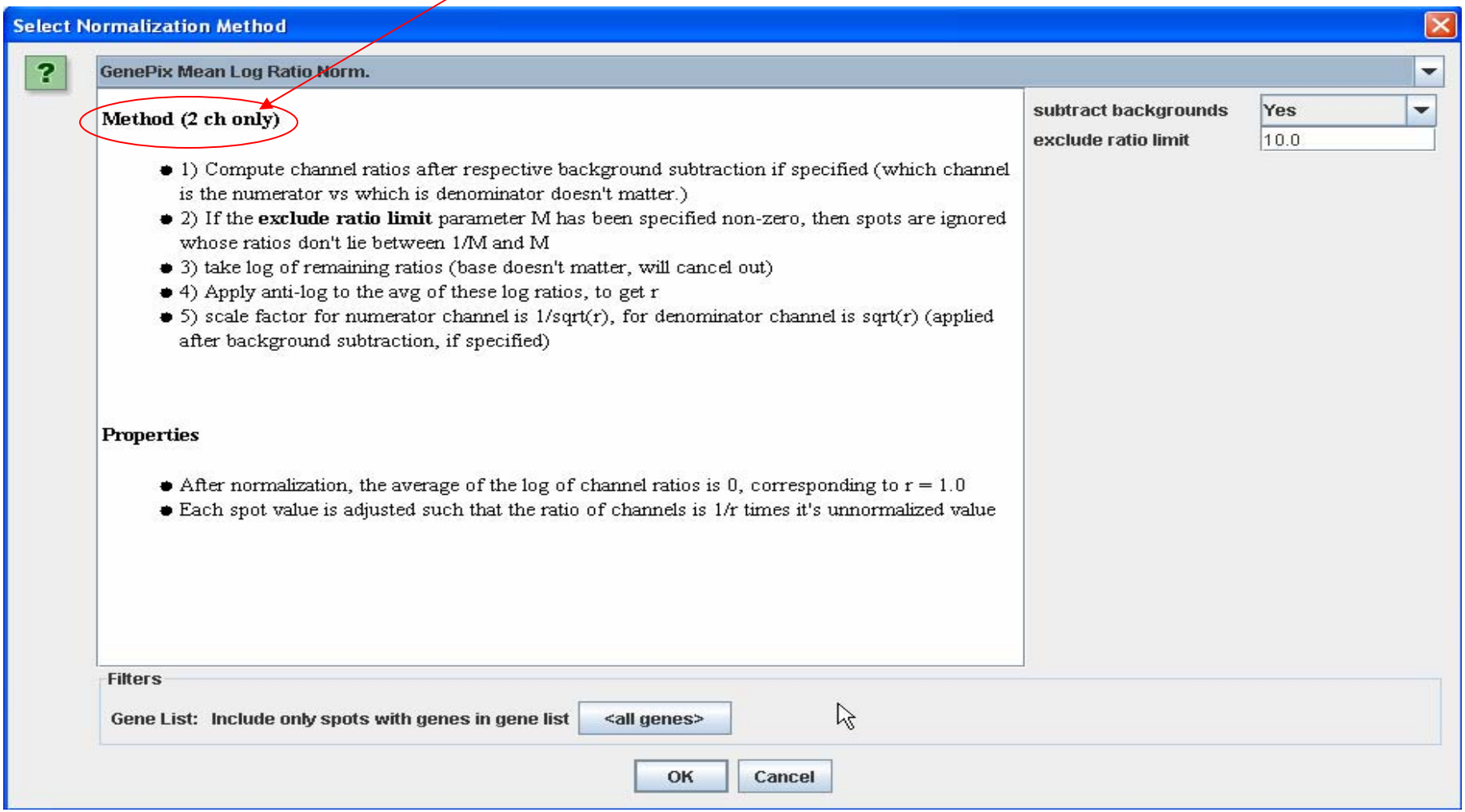
Total intensity ration normalization is for two channel only.

Mean/Median Scaling Normalization

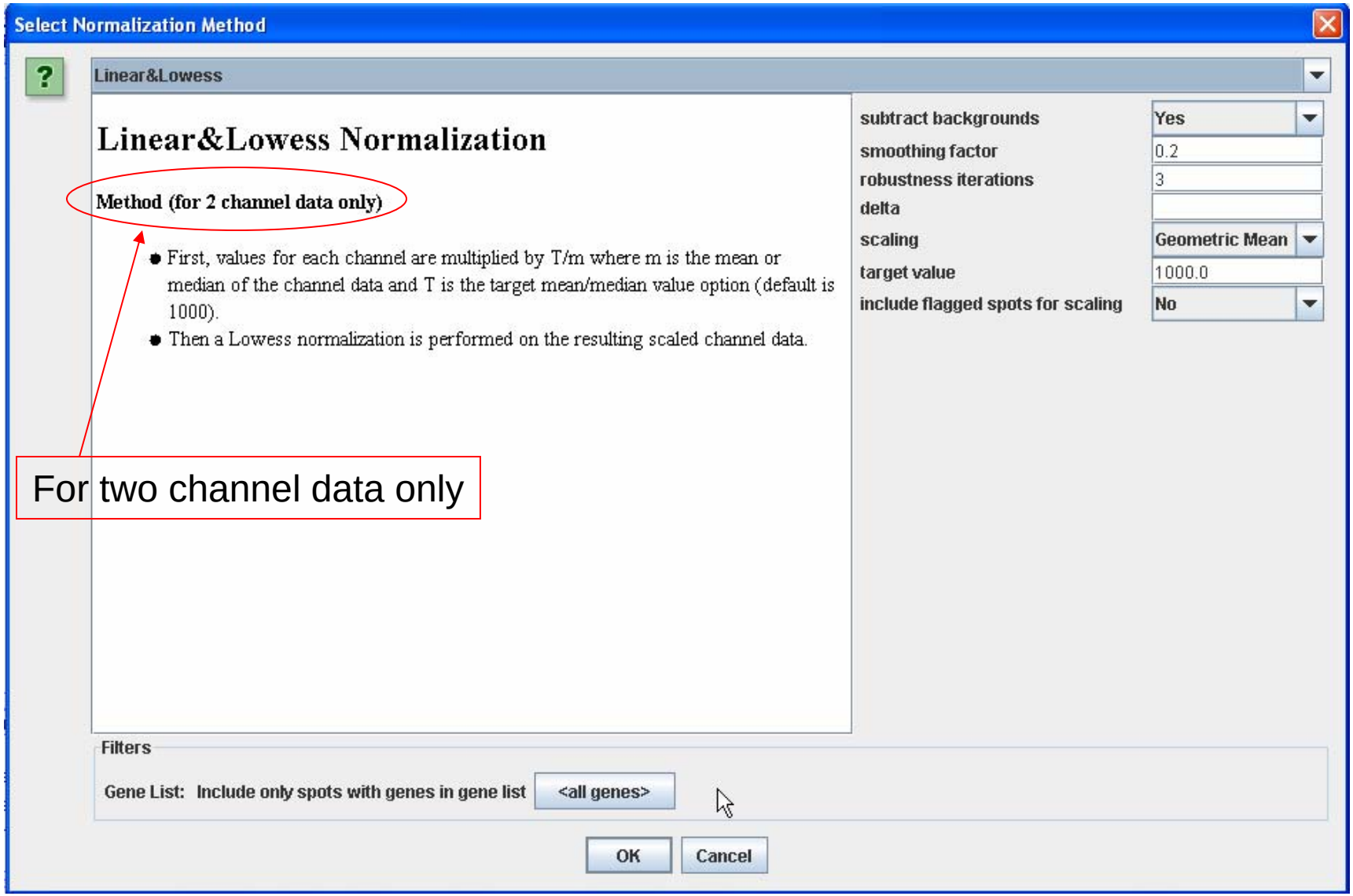


GenePix Mean Log Ration Normalization

This method is for two channel data only

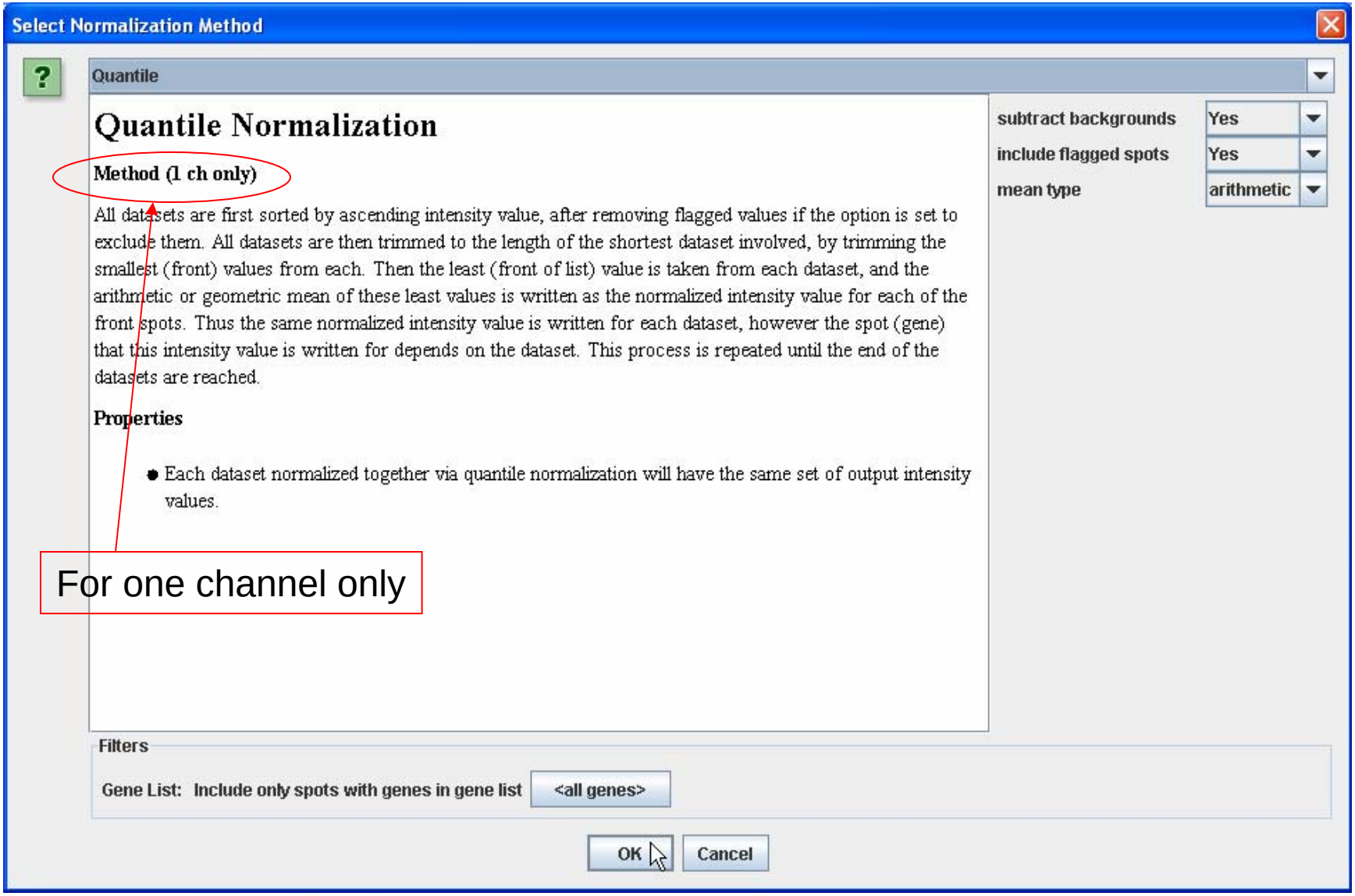


Linear & Lowess Normalization



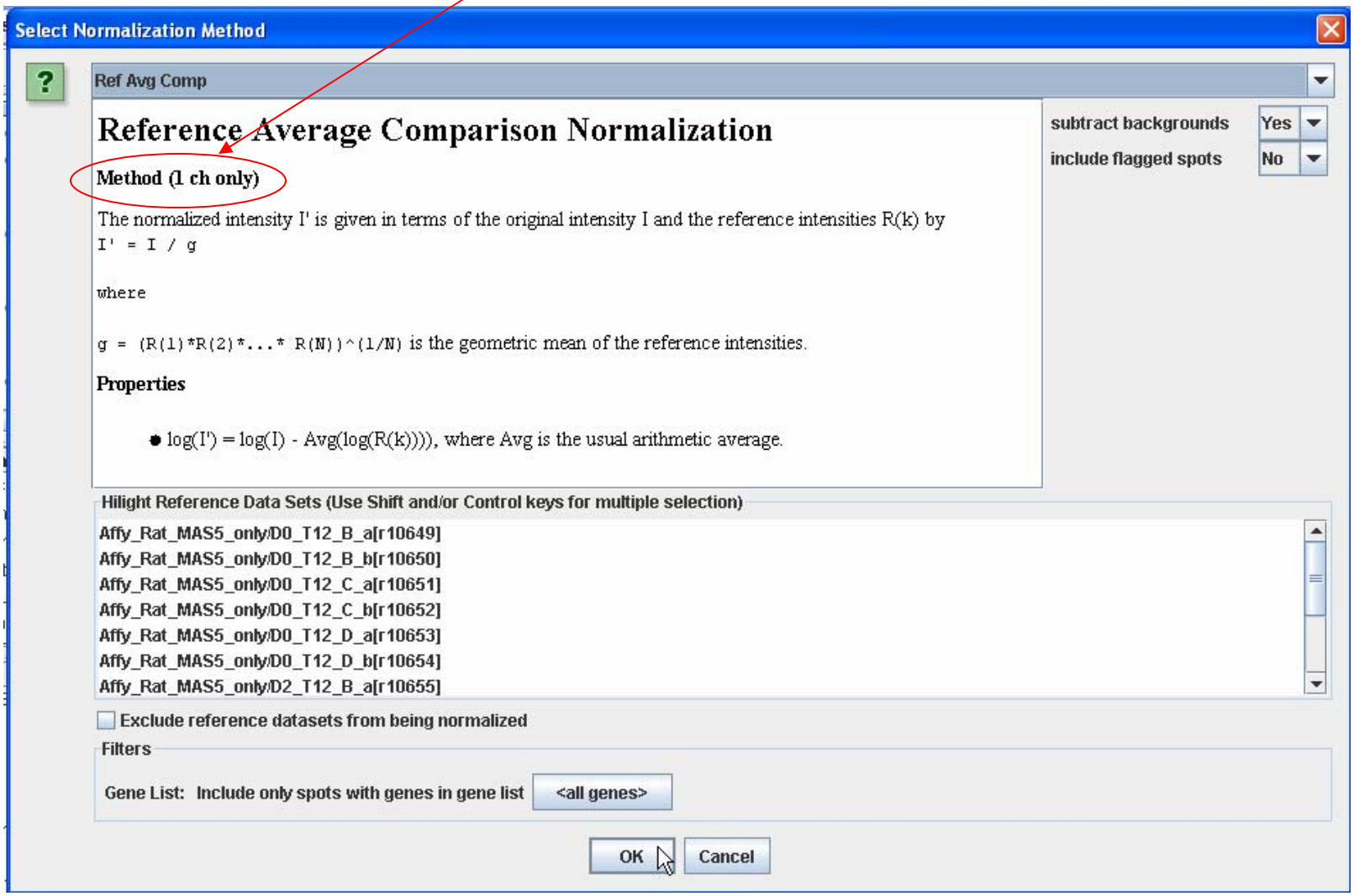
For two channel data only

Quantile Normaliztion



Reference Average Comparison Normaliztion

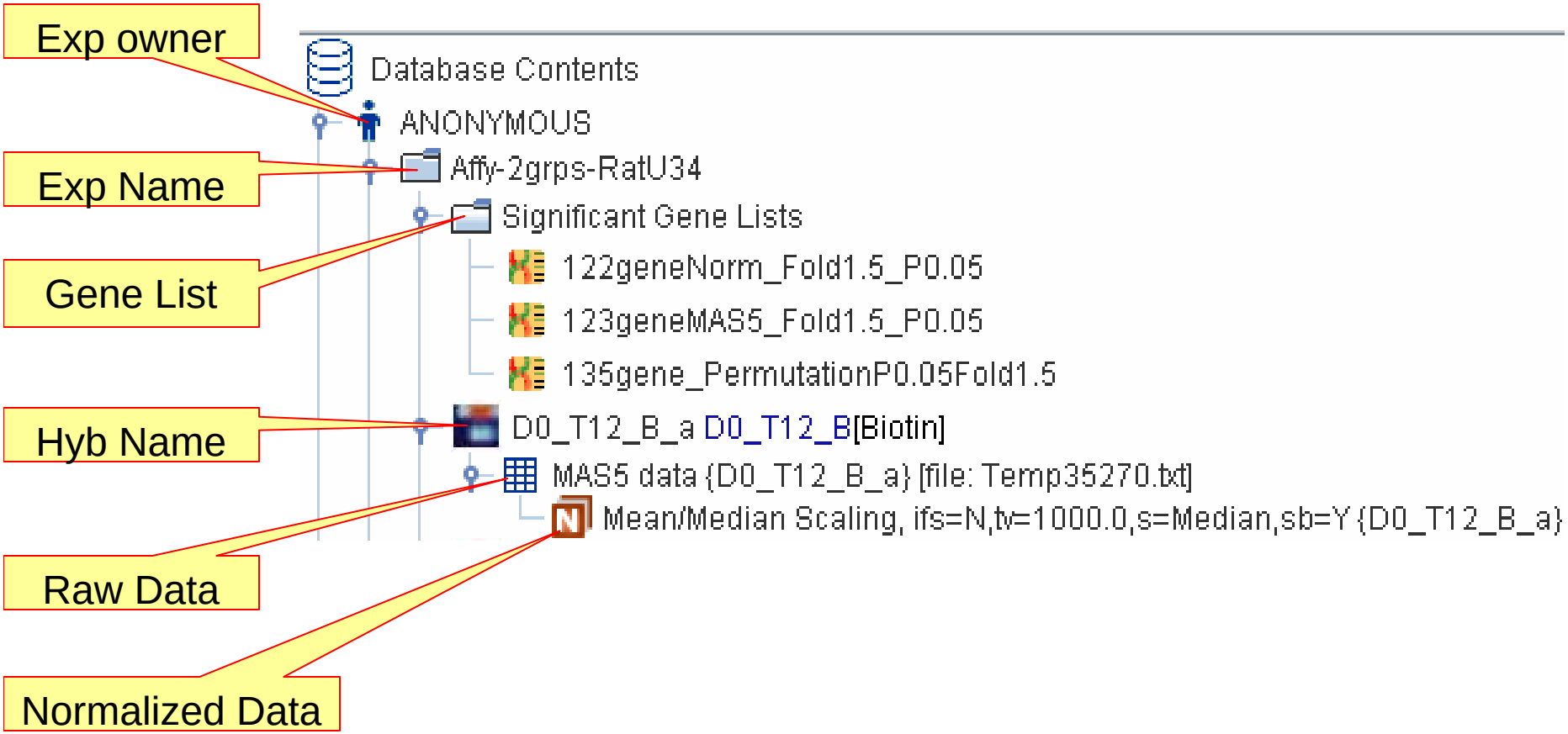
For one channel only



Tutorial 8:
How to create your own workspace

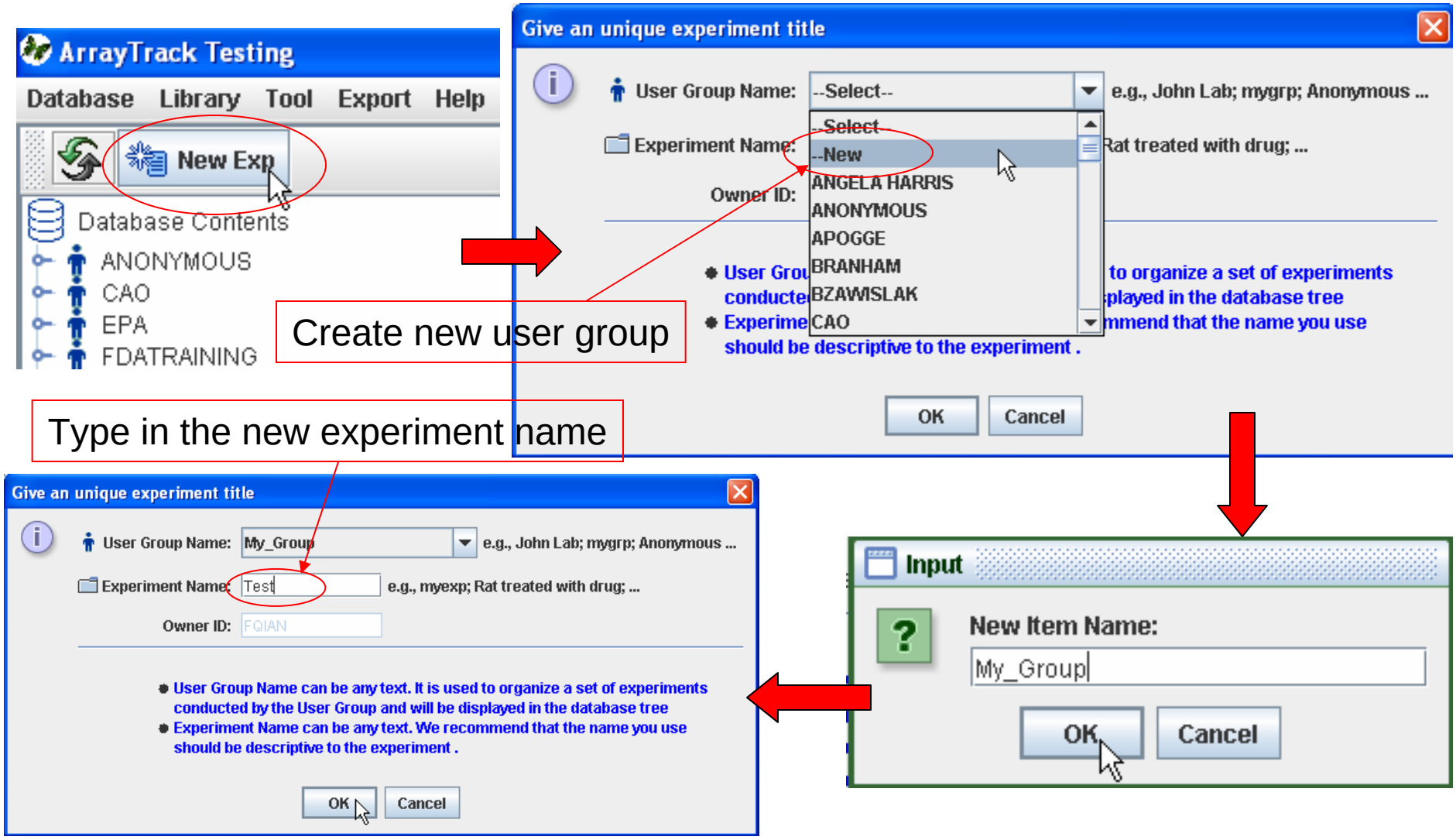
The structure of the data in ArrayTrack

Data is organized as a hierarchical tree structure:

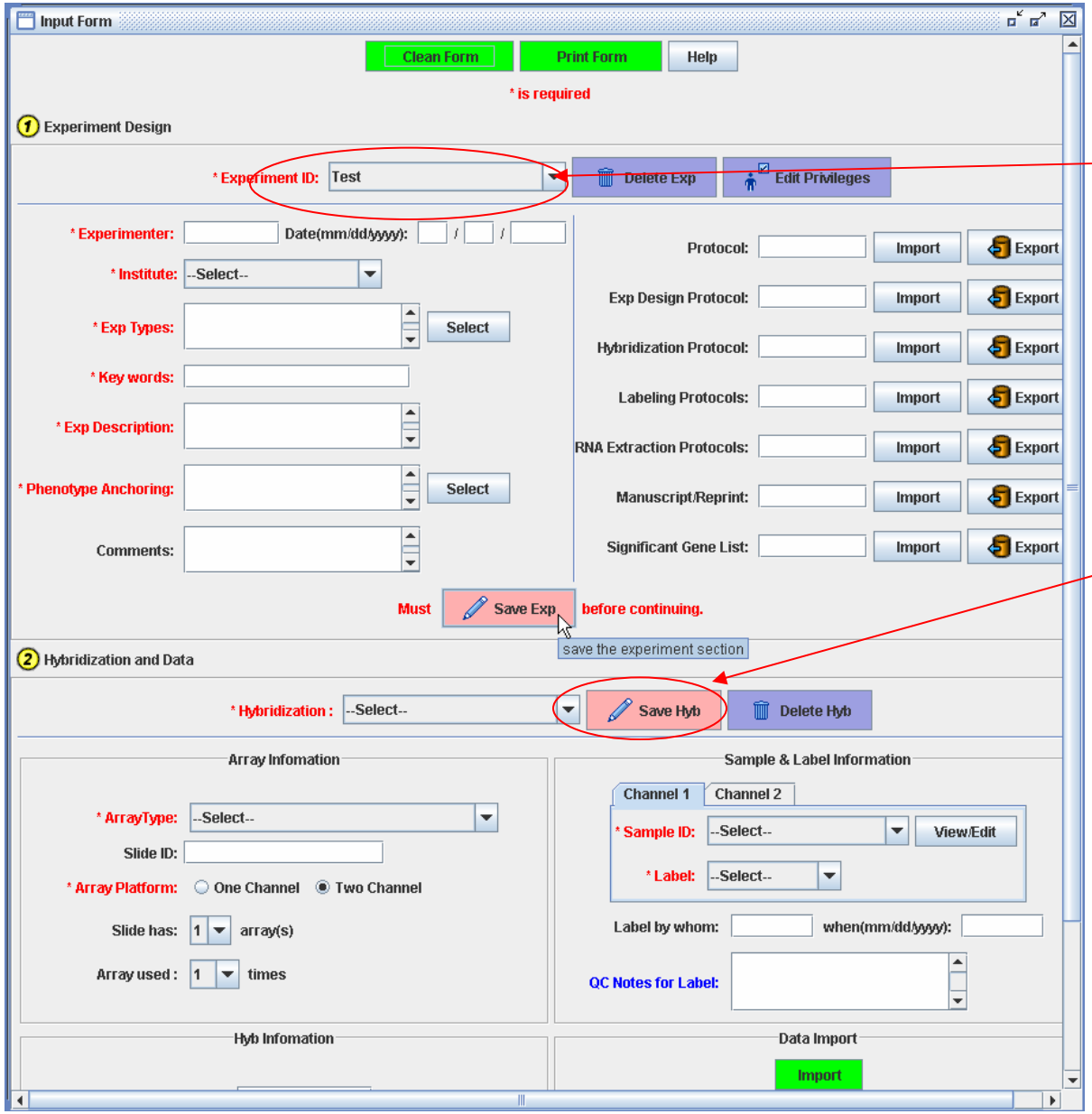


Create new user group and experiment

User can create his own group to congregate all the experiments under his own space



Create Experiment

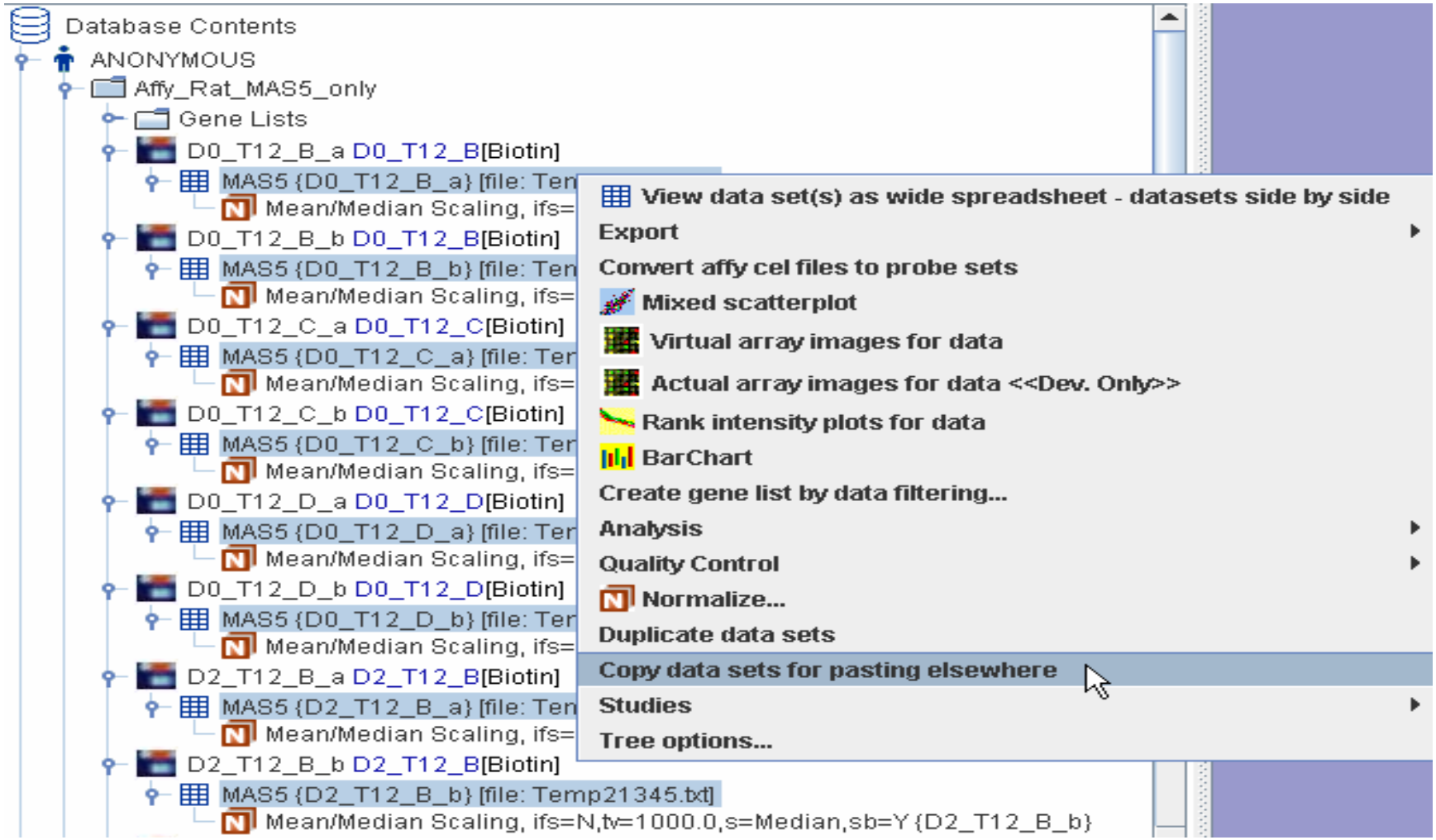


Experiment name

Must save experiment

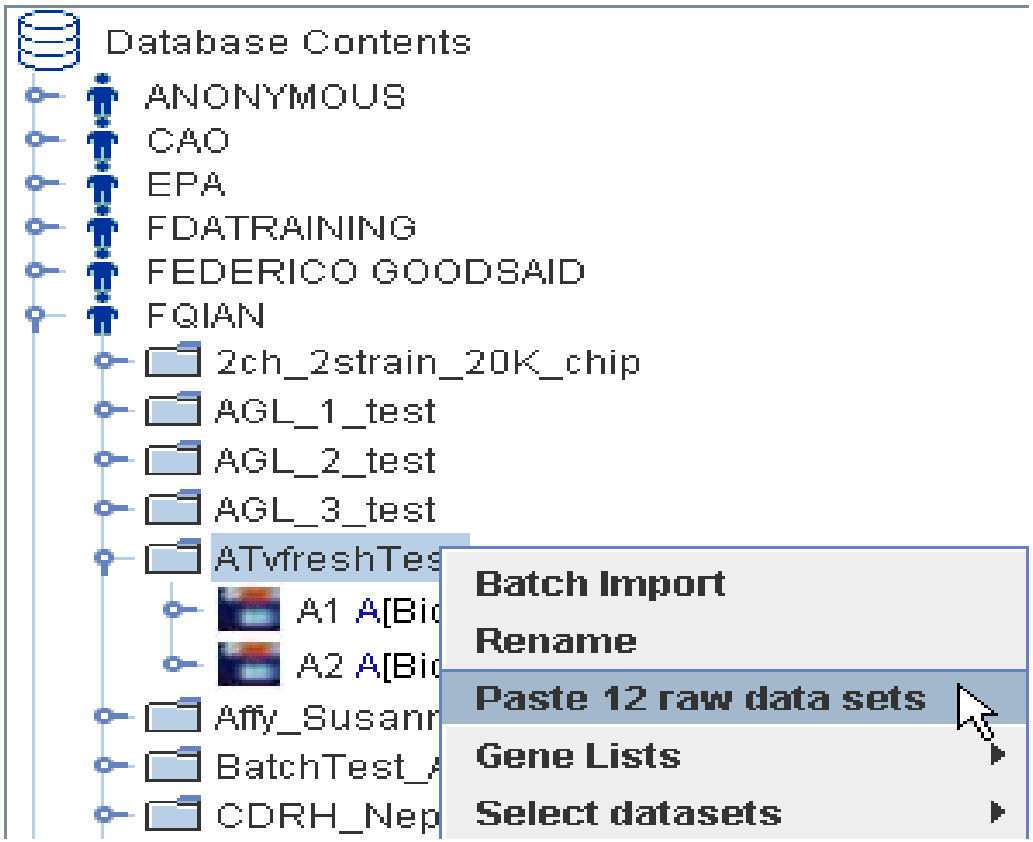
Copy Experiment

User can also copy experiment from other group and paste it under his own group



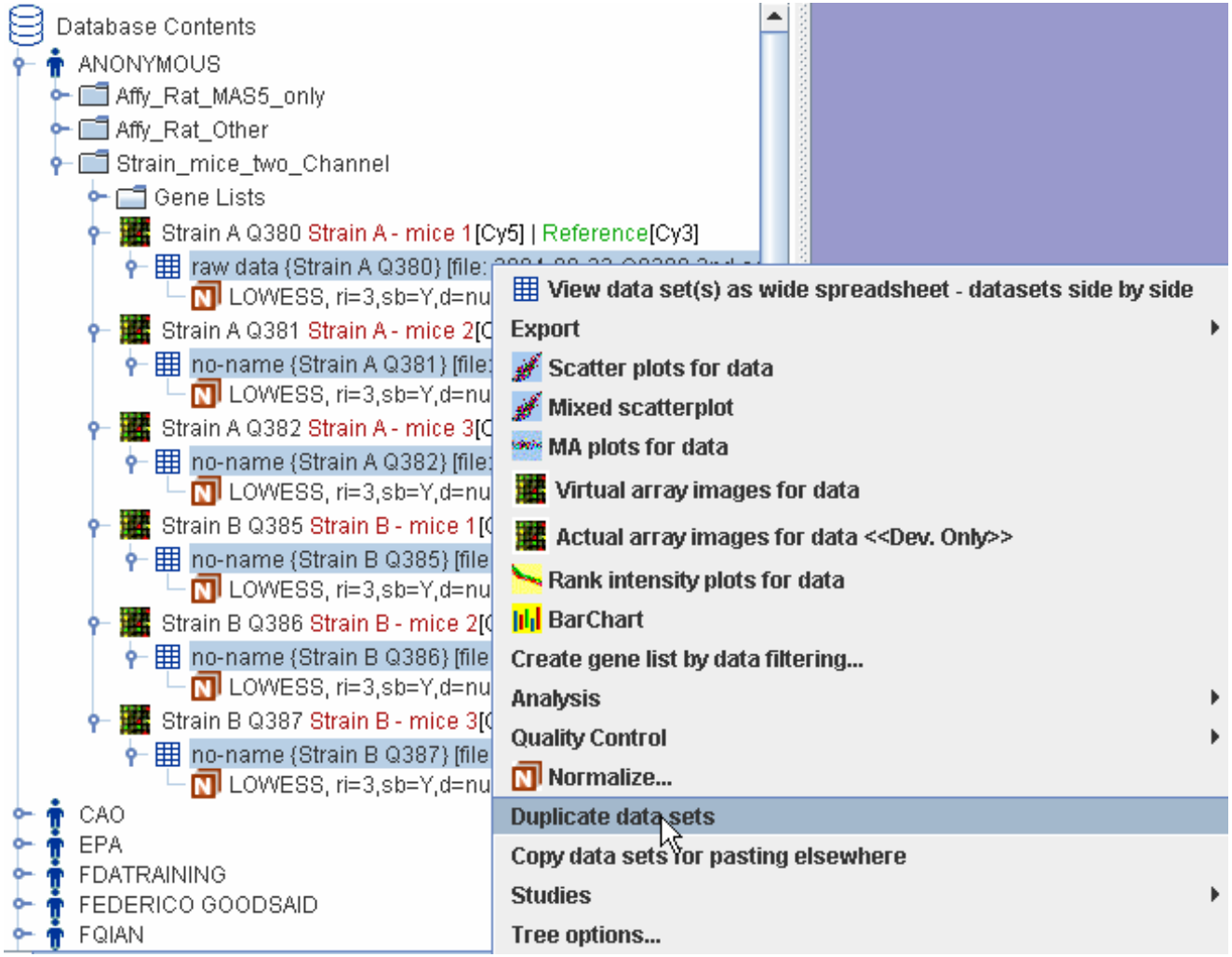
Select the hybridizations to be copied, right-click -> choose “Copy data sets for Pasting elsewhere

Copy experiment - continued



Right-click the other experiment (for storing the copied data), then choose “Paste # raw data sets”.

Duplicate an experiment



Select the data to be duplicated, right-click -> choose “Duplicate data sets”. The data sets will be duplicated under the same experiment folder. This is different from copying data that the copied data can be copied to a different experiment.

Array Track

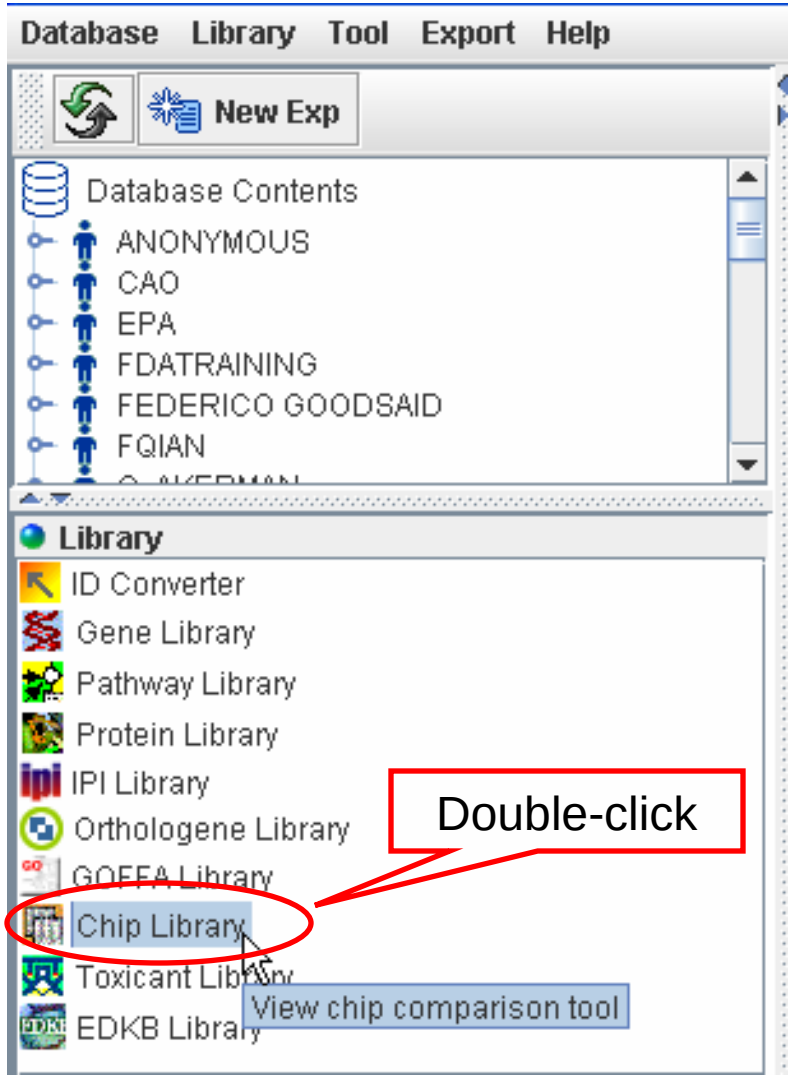
Tutorial 9A: Data Import

Data Import

Before you import any data, you need to create array type unless the array type is already exist in ArrayTrack (you can check all the existing array type from Chip Library). For Affymetrix array type, you can take it as it is. But for other array type (e.g. Agilent) you need to check the coordinates and gene ID to make sure they are exactly match with your data file.

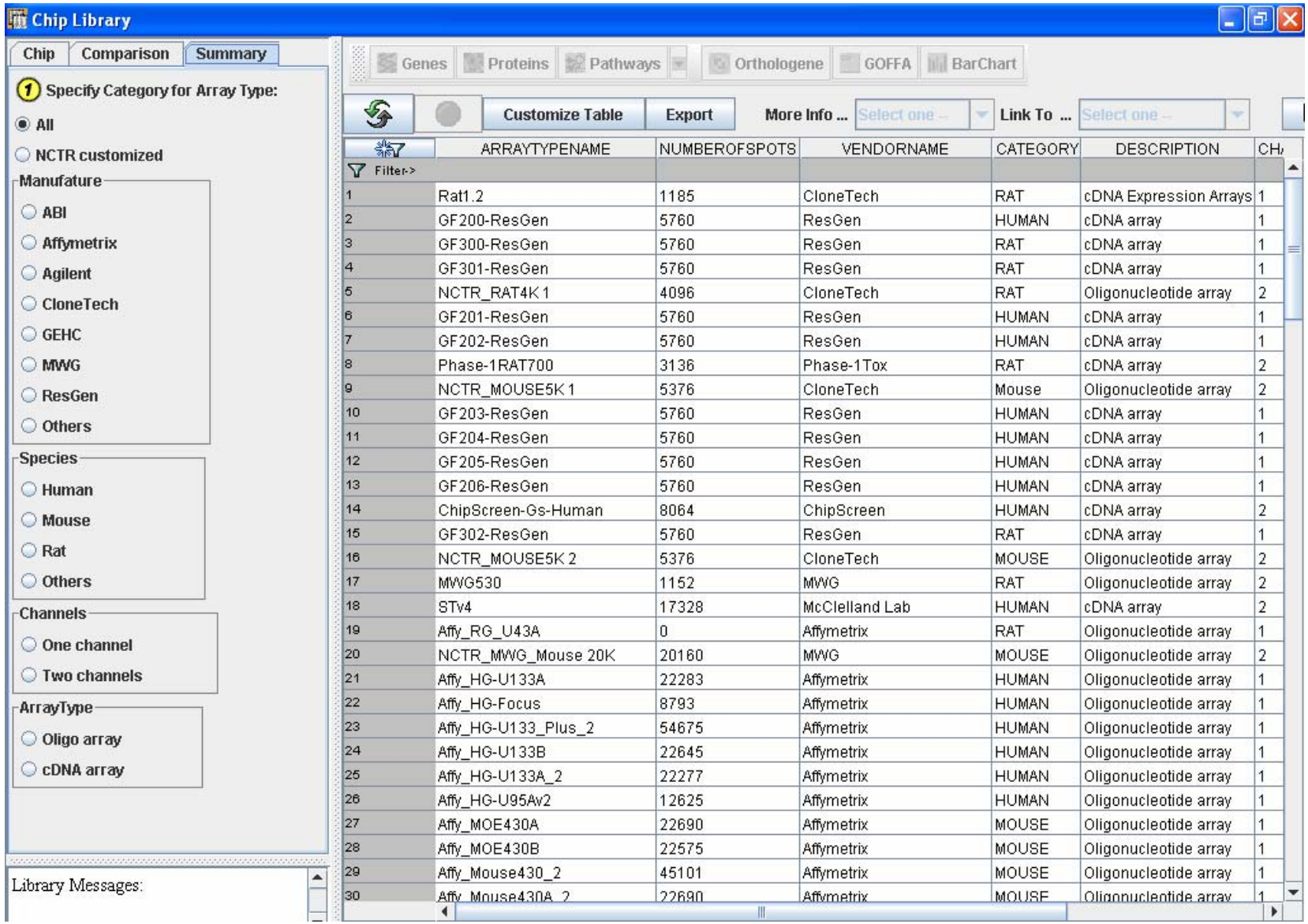
You also need to create hybridization file which will be used for batch import. The following slides will lead you through all the steps.

Check Existing Array Type



Chip Library stores all the array types that are available in ArrayTrack.

Check Existing Array Type



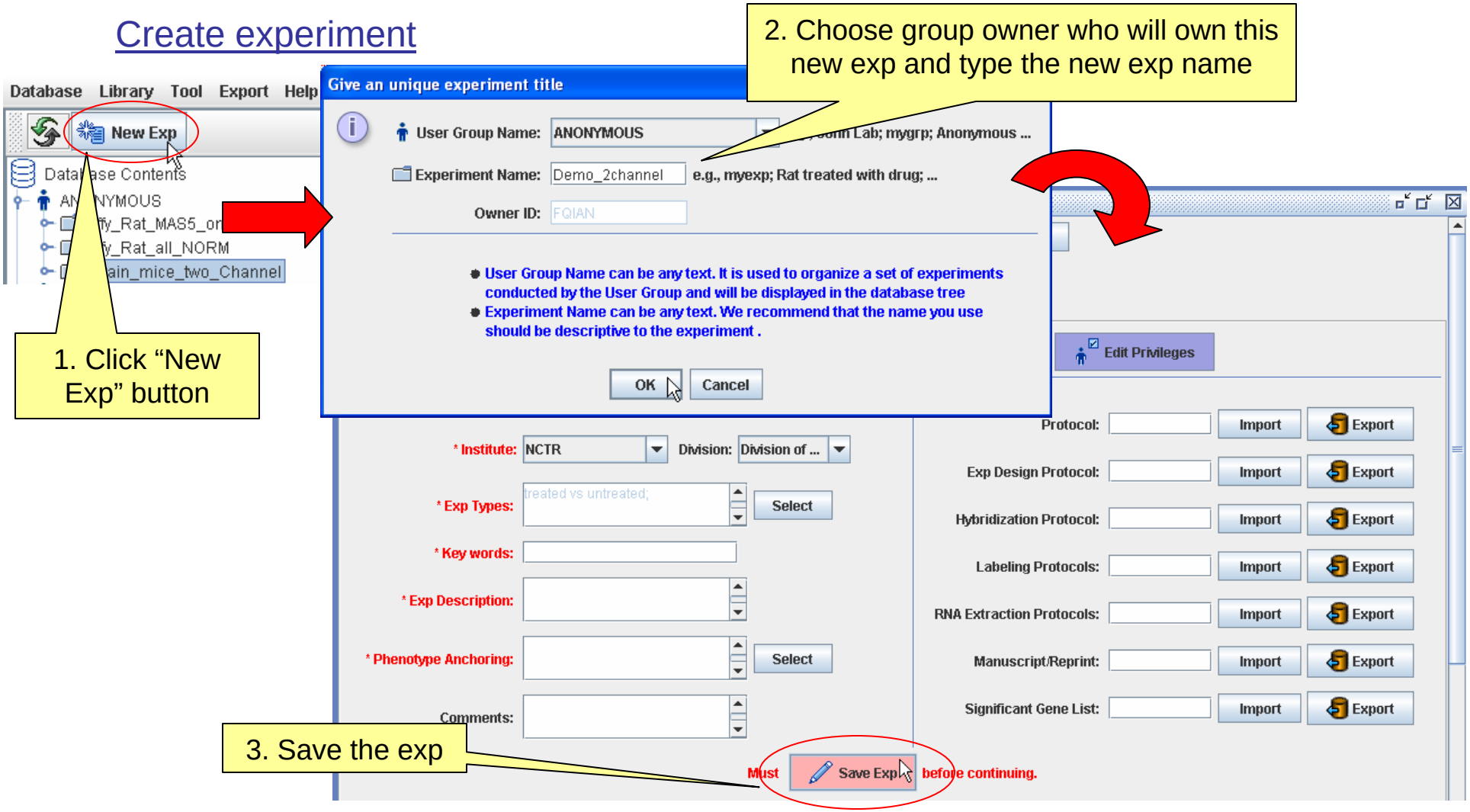
Check Existing Array Type

- If the array type for your data (e.g. Affy_HG U133_Plus_2) does not exist in ArrayTrack, you need to go to slide #34 (create array type). After you created array type, then come back to continue the next slides (create experiment).
- If the array type for your data does exist in ArrayTrack, you can continue to the next slide (create experiment.

Create experiment

Before importing data, you need to create an experiment to hold the data.

Create experiment



Create experiment

Input Form

Clean Form | Print Form | Help

* is required

1 Experiment Design

* Experiment ID: Demo_2channel | Delete Exp | Edit Privileges

* Experimenter: | Date(mm/dd/yyyy): / /

* Institute: --Select--

* Exp Types: | Select

* Key words:

* Exp Description:

* Phenotype Anchoring: | Select

Comments:

Protocol: | Import | Export

Exp Design Protocol: | Import | Export

Hybridization Protocol: | Import | Export

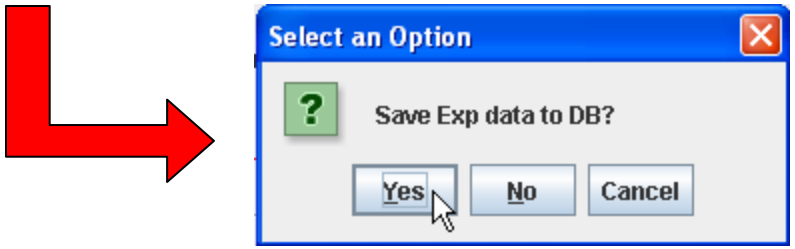
Labeling Protocols: | Import | Export

RNA Extraction Protocols: | Import | Export

Manuscript/Reprint: | Import | Export

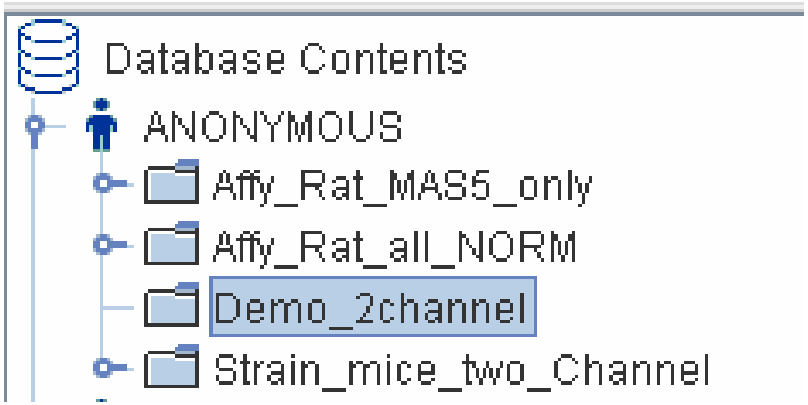
Significant Gene List: | Import | Export

Must Save Exp before continuing.



Create experiment

After clicking the save exp button, you will see the new experiment displaying under the designated owner.

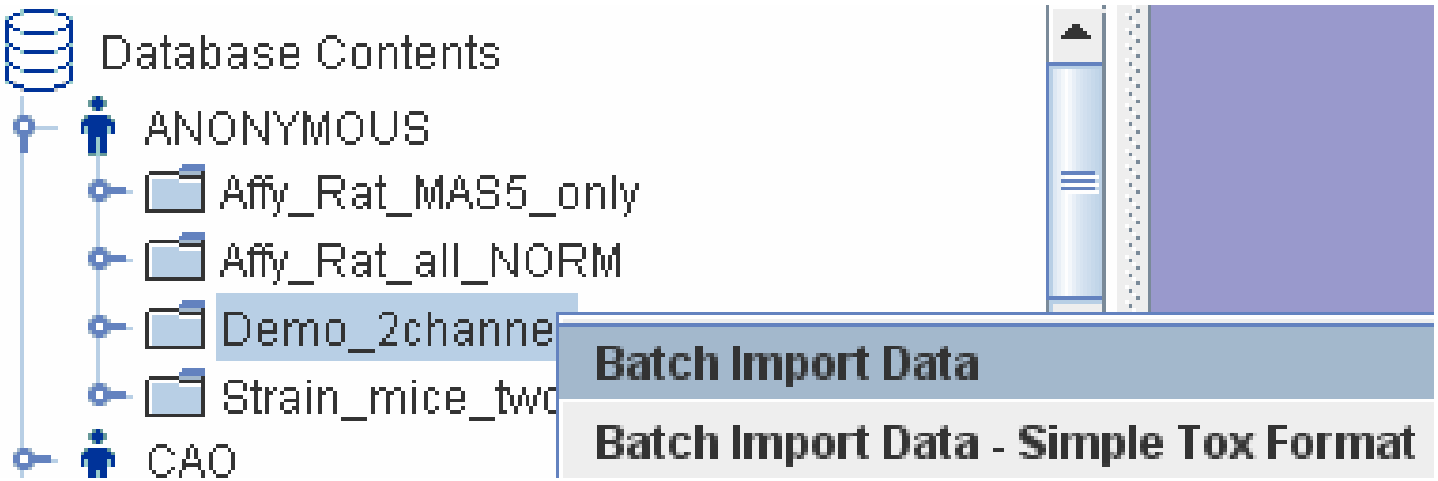


Batch Import

There are two formats for batch import:

- Regular batch import
- Simple Tox format – for animal toxicology study data submission, with combined and extracted main fields from SEND (Standard for Exchange of Nonclinical Data)

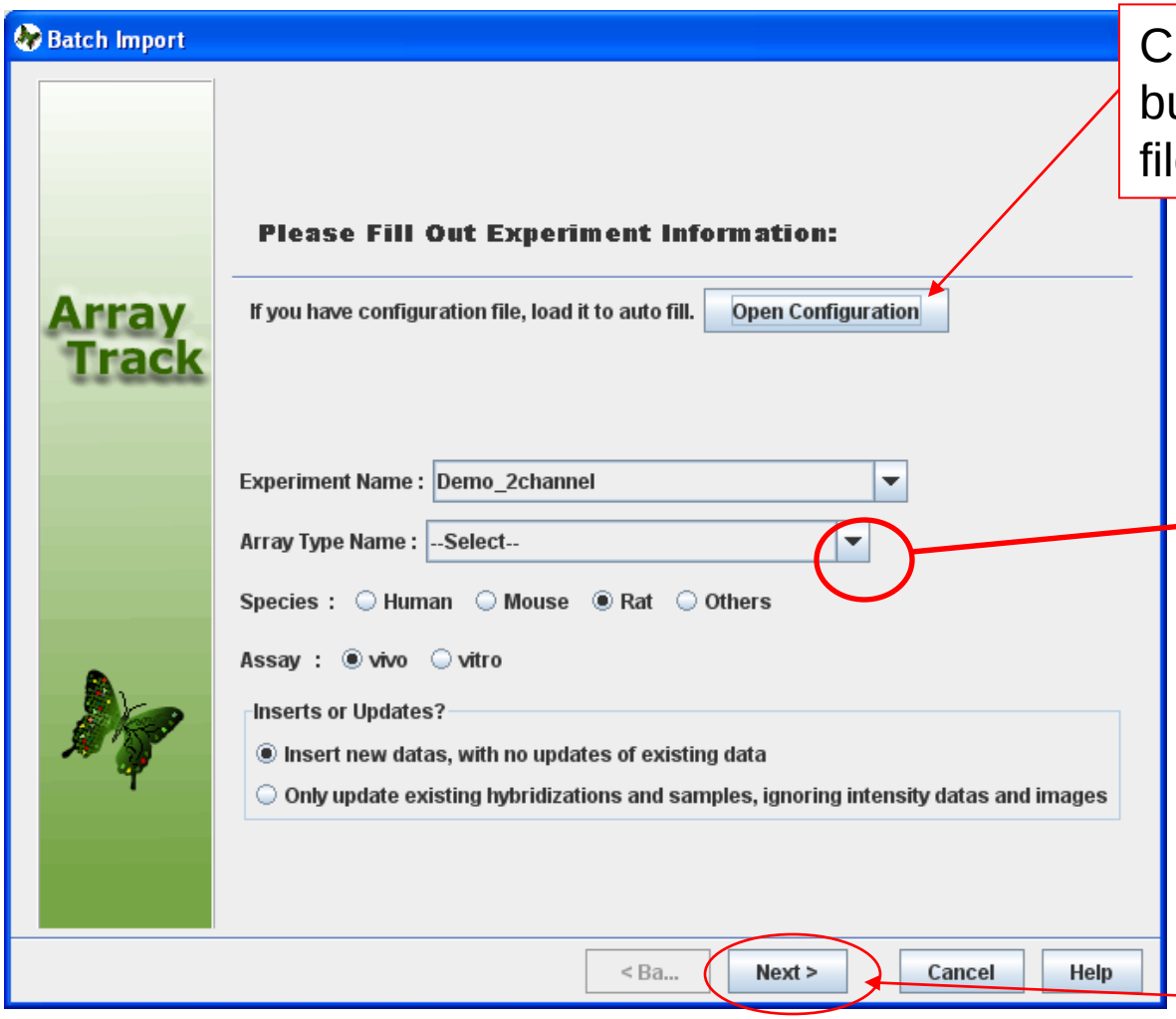
Batch Import - regular



Right-click the experiment, choose “Batch Import Data”

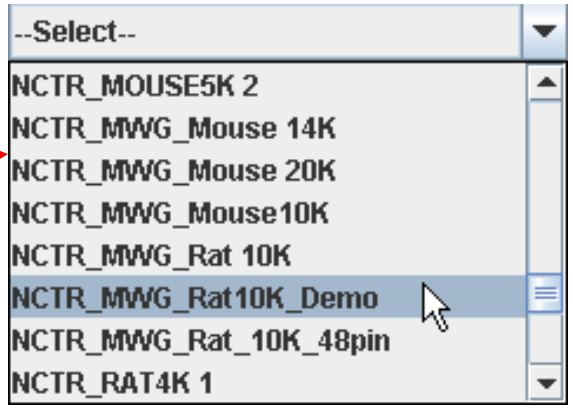
Batch Import Wizard

This is an example of importing 2-channel data.



Click Open Configuration button if you have configure file, if not just skip it.

Choose the array type



Click Next button

Batch Import

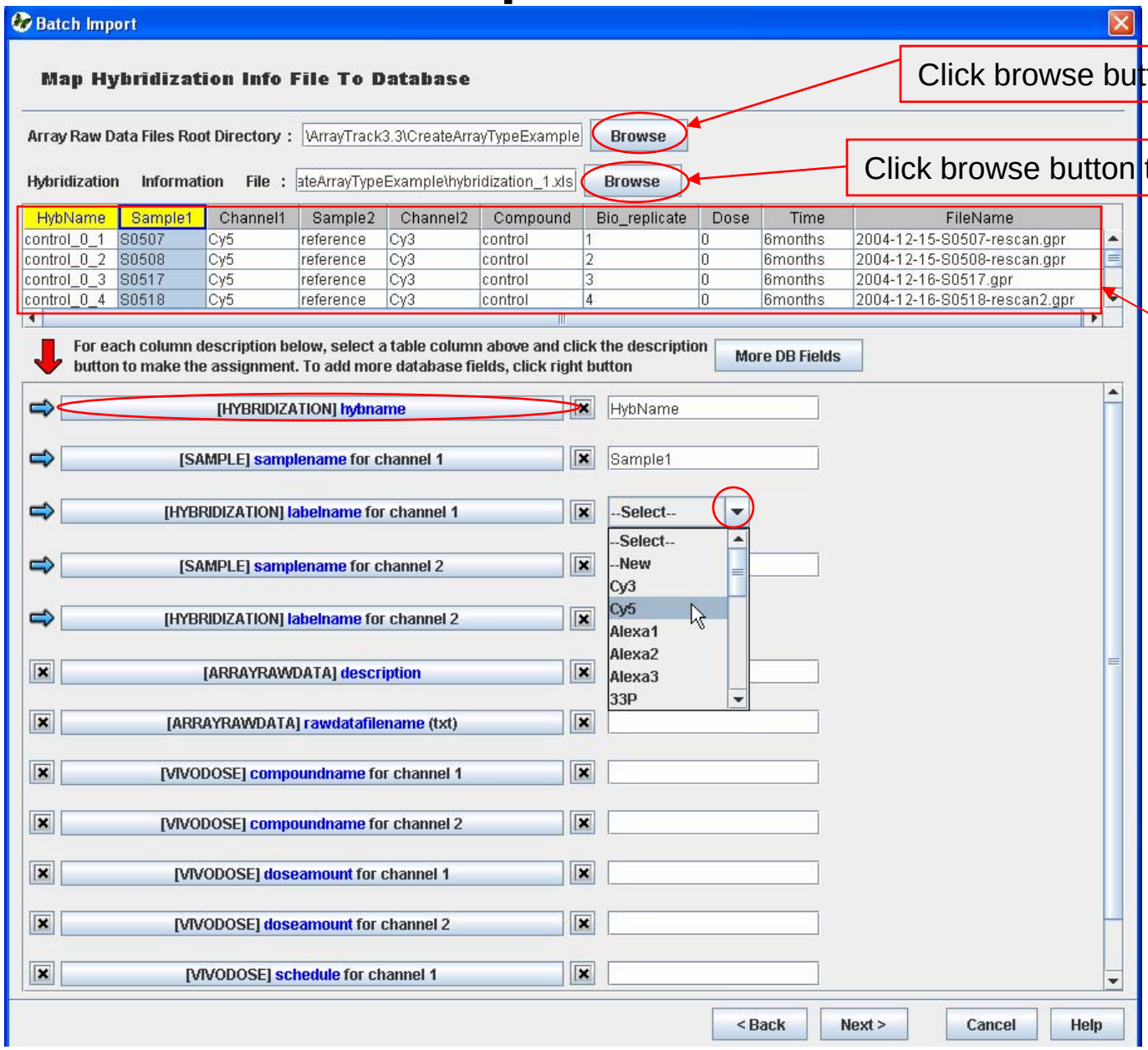
Use Excel to create a hybridization file like following. You will need this file for batch import.

hybridization_1

	A *	B *	C	D	E	F	G	H	I	J *
1	HybName	Sample1	Channel1	Sample2	Channel2	Compound	Bio_replicate	Dose	Time	FileName
2	control_0_1	S0507	Cy5	reference	Cy3	control	1	0	6months	2004-12-15-S0507-rescan.gpr
3	control_0_2	S0508	Cy5	reference	Cy3	control	2	0	6months	2004-12-15-S0508-rescan.gpr
4	control_0_3	S0517	Cy5	reference	Cy3	control	3	0	6months	2004-12-16-S0517.gpr
5	control_0_4	S0518	Cy5	reference	Cy3	control	4	0	6months	2004-12-16-S0518-rescan2.gpr
6	control_0_5	S0527	Cy5	reference	Cy3	control	5	0	6months	2004-12-17-S0527-rescan.gpr
7	Drug1_250_1	S0513	Cy5	reference	Cy3	Drug1	1	250	6months	2004-12-15-S0513.gpr
8	Drug1_250_2	S0514	Cy5	reference	Cy3	Drug1	2	250	6months	2004-12-15-S0514.gpr
9	Drug1_250_3	S0523	Cy5	reference	Cy3	Drug1	3	250	6months	2004-12-16-S0523.gpr
10	Drug1_250_4	S0524	Cy5	reference	Cy3	Drug1	4	250	6months	2004-12-16-S0524.gpr
11	Drug1_250_5	S0550	Cy5	reference	Cy3	Drug1	5	250	6months	2004-12-17-S0550.gpr
12	Drug1_500_1	S0515	Cy5	reference	Cy3	Drug1	1	500	6months	2004-12-15-S0515.gpr
13	Drug1_500_2	S0516	Cy5	reference	Cy3	Drug1	2	500	6months	2004-12-15-S0516.gpr
14	Drug1_500_3	S0525	Cy5	reference	Cy3	Drug1	3	500	6months	2004-12-16-S0525.gpr
15	Drug1_500_4	S0526	Cy5	reference	Cy3	Drug1	4	500	6months	2004-12-16-S0526.gpr
16	Drug1_500_5	S0549	Cy5	reference	Cy3	Drug1	5	500	6months	2004-12-17-S0549.gpr

* is required column. The others are optional columns.

Batch Import Wizard – continued

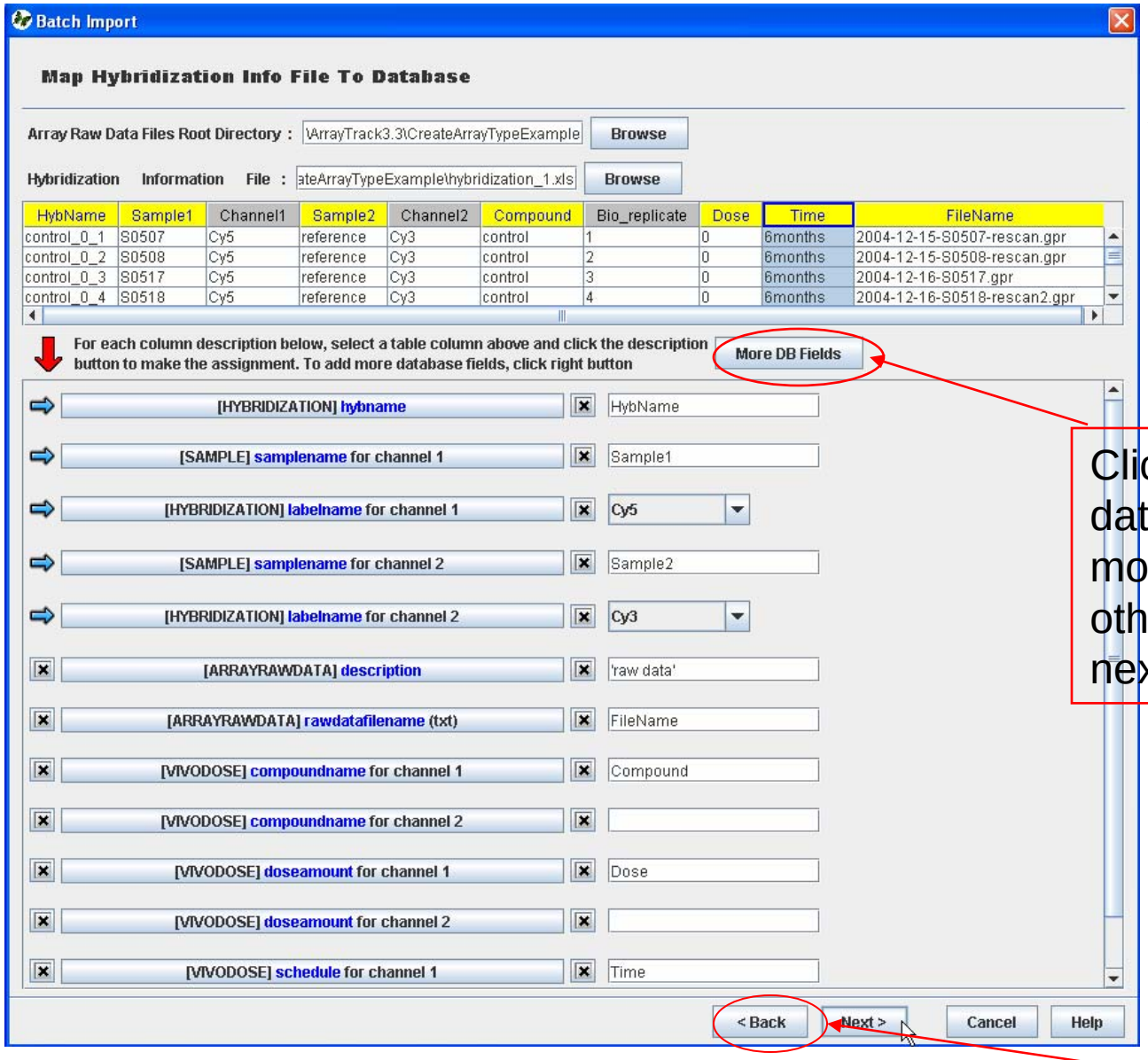


Click browse button to locate the data files

Click browse button to locate the hybridization files

Hybridization file

Batch Import Wizard - – continued



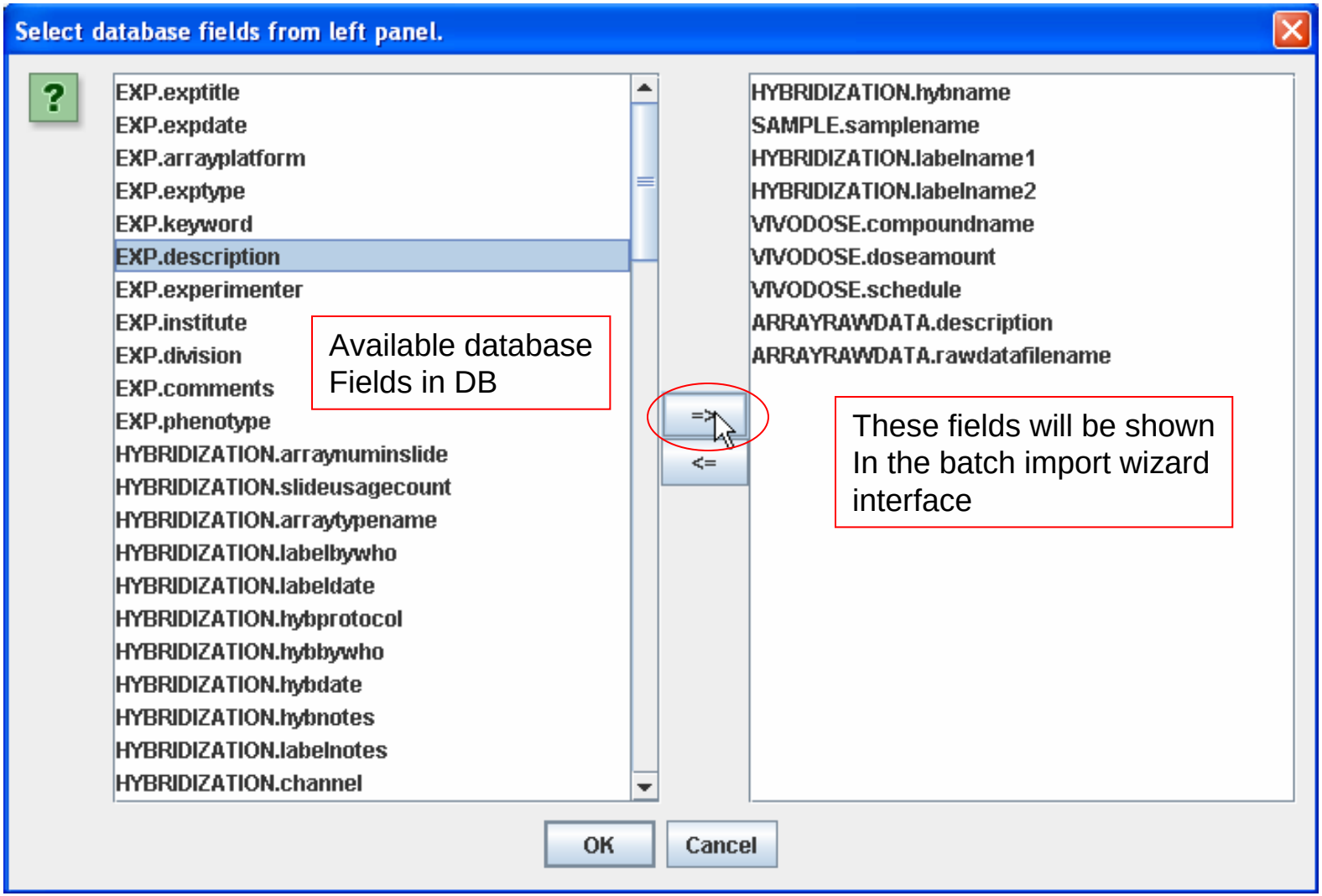
Assign the other columns to the corresponding fields

Click this button to get more database fields if you have more information to put in, otherwise skip this step. See next slide.

Click Next button

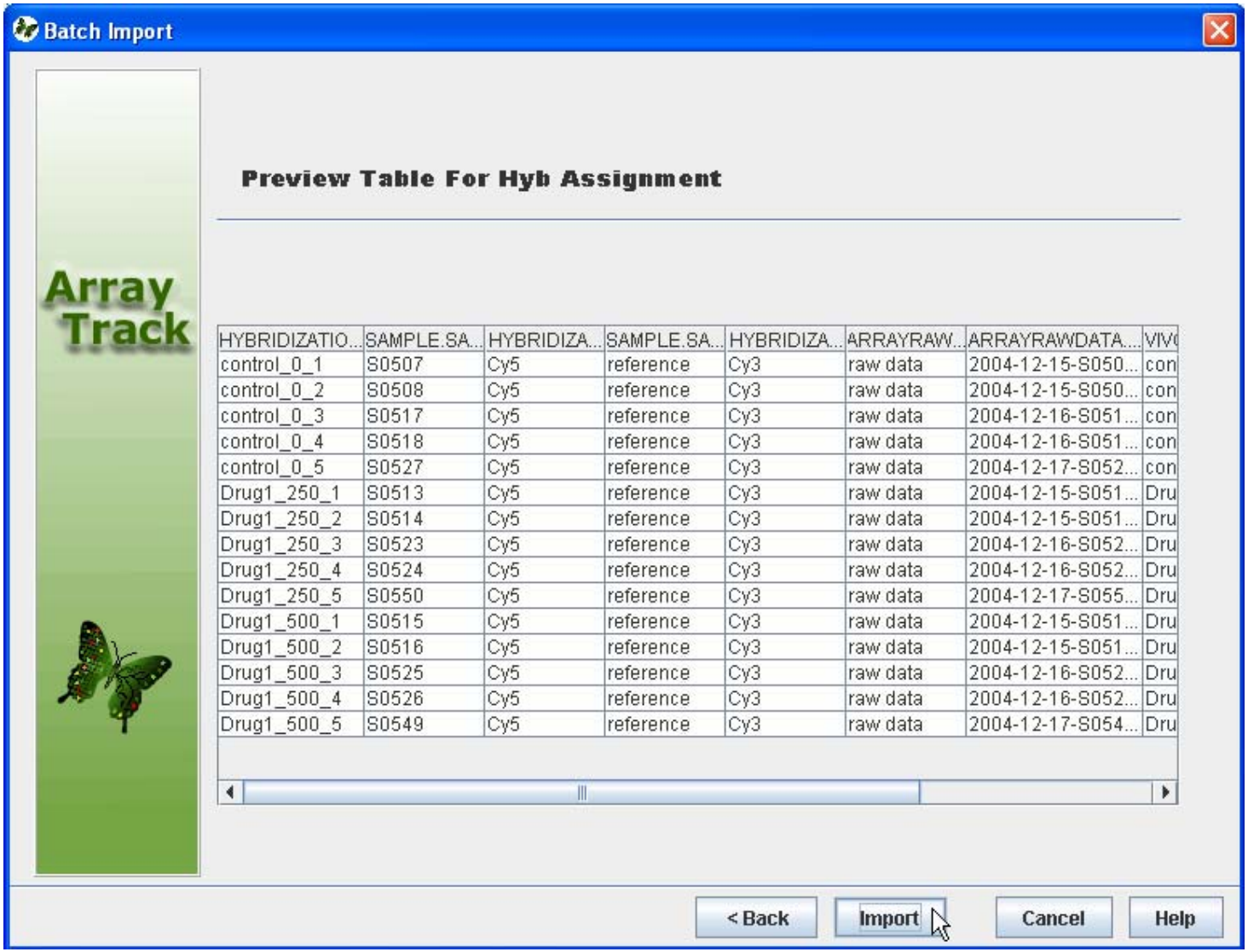
Batch Import Wizard - continued

Select the database fields in the left panel, click => to add in the right panel and this added field will be shown in the batch import interface.



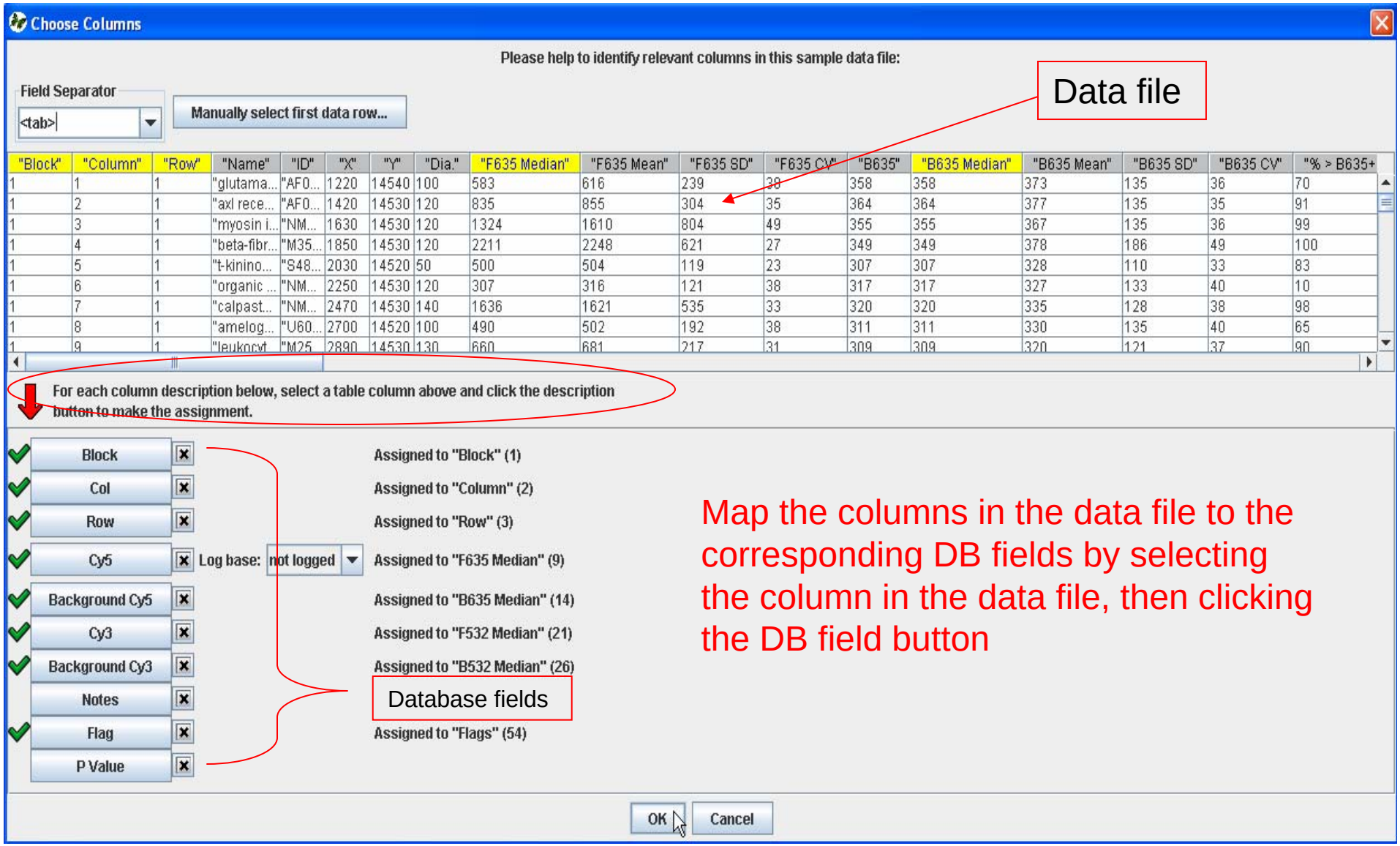
Batch Import Wizard - continued

Preview the table for hybridization assignment before import



HYBRIDIZATIO...	SAMPLE.SA...	HYBRIDIZA...	SAMPLE.SA...	HYBRIDIZA...	ARRAYRAW...	ARRAYRAWDATA....	VIV
control_0_1	S0507	Cy5	reference	Cy3	raw data	2004-12-15-S050...	con
control_0_2	S0508	Cy5	reference	Cy3	raw data	2004-12-15-S050...	con
control_0_3	S0517	Cy5	reference	Cy3	raw data	2004-12-16-S051...	con
control_0_4	S0518	Cy5	reference	Cy3	raw data	2004-12-16-S051...	con
control_0_5	S0527	Cy5	reference	Cy3	raw data	2004-12-17-S052...	con
Drug1_250_1	S0513	Cy5	reference	Cy3	raw data	2004-12-15-S051...	Dru
Drug1_250_2	S0514	Cy5	reference	Cy3	raw data	2004-12-15-S051...	Dru
Drug1_250_3	S0523	Cy5	reference	Cy3	raw data	2004-12-16-S052...	Dru
Drug1_250_4	S0524	Cy5	reference	Cy3	raw data	2004-12-16-S052...	Dru
Drug1_250_5	S0550	Cy5	reference	Cy3	raw data	2004-12-17-S055...	Dru
Drug1_500_1	S0515	Cy5	reference	Cy3	raw data	2004-12-15-S051...	Dru
Drug1_500_2	S0516	Cy5	reference	Cy3	raw data	2004-12-15-S051...	Dru
Drug1_500_3	S0525	Cy5	reference	Cy3	raw data	2004-12-16-S052...	Dru
Drug1_500_4	S0526	Cy5	reference	Cy3	raw data	2004-12-16-S052...	Dru
Drug1_500_5	S0549	Cy5	reference	Cy3	raw data	2004-12-17-S054...	Dru

Batch Import Wizard - continued

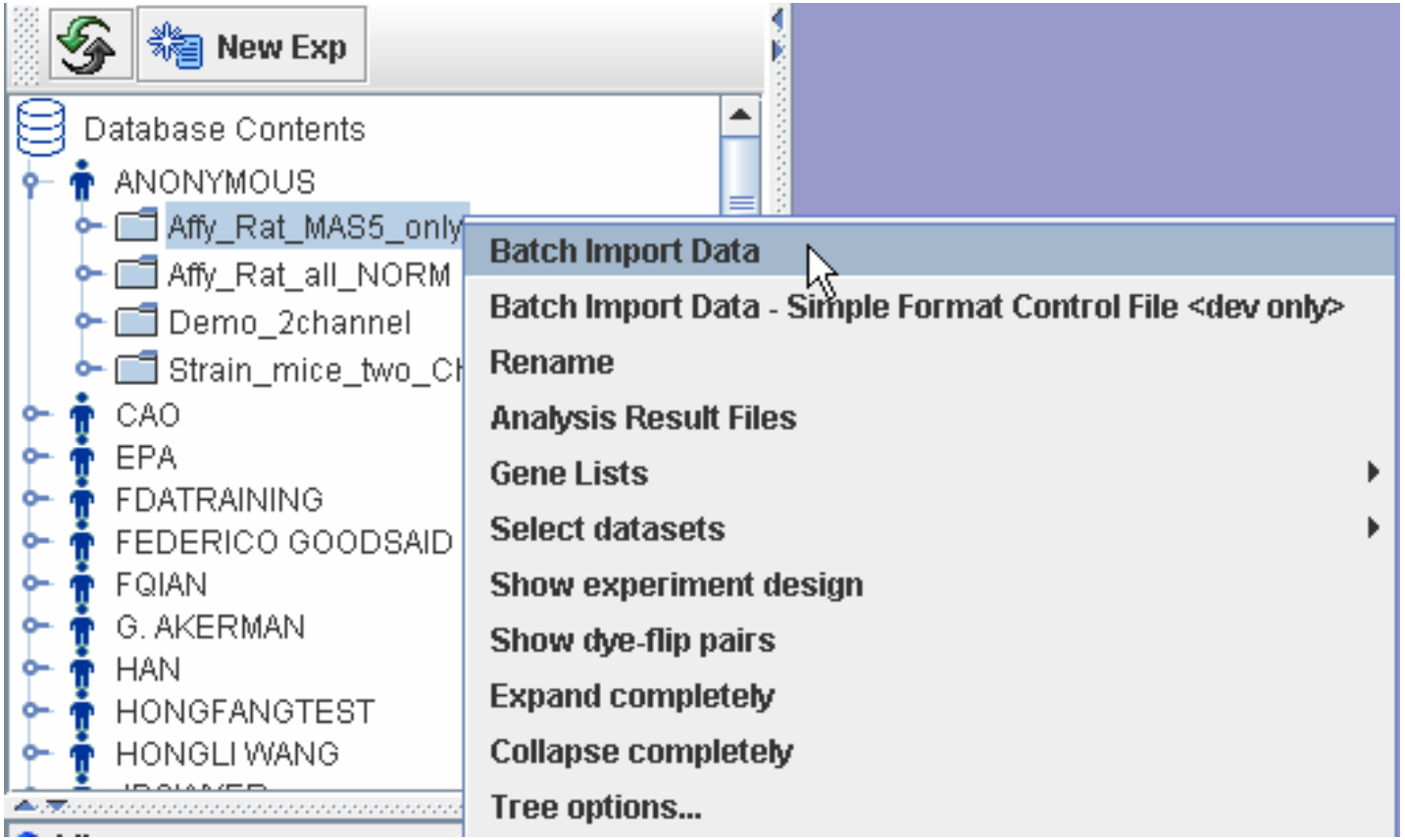


Map the columns in the data file to the corresponding DB fields by selecting the column in the data file, then clicking the DB field button

Batch Import Wizard - continued

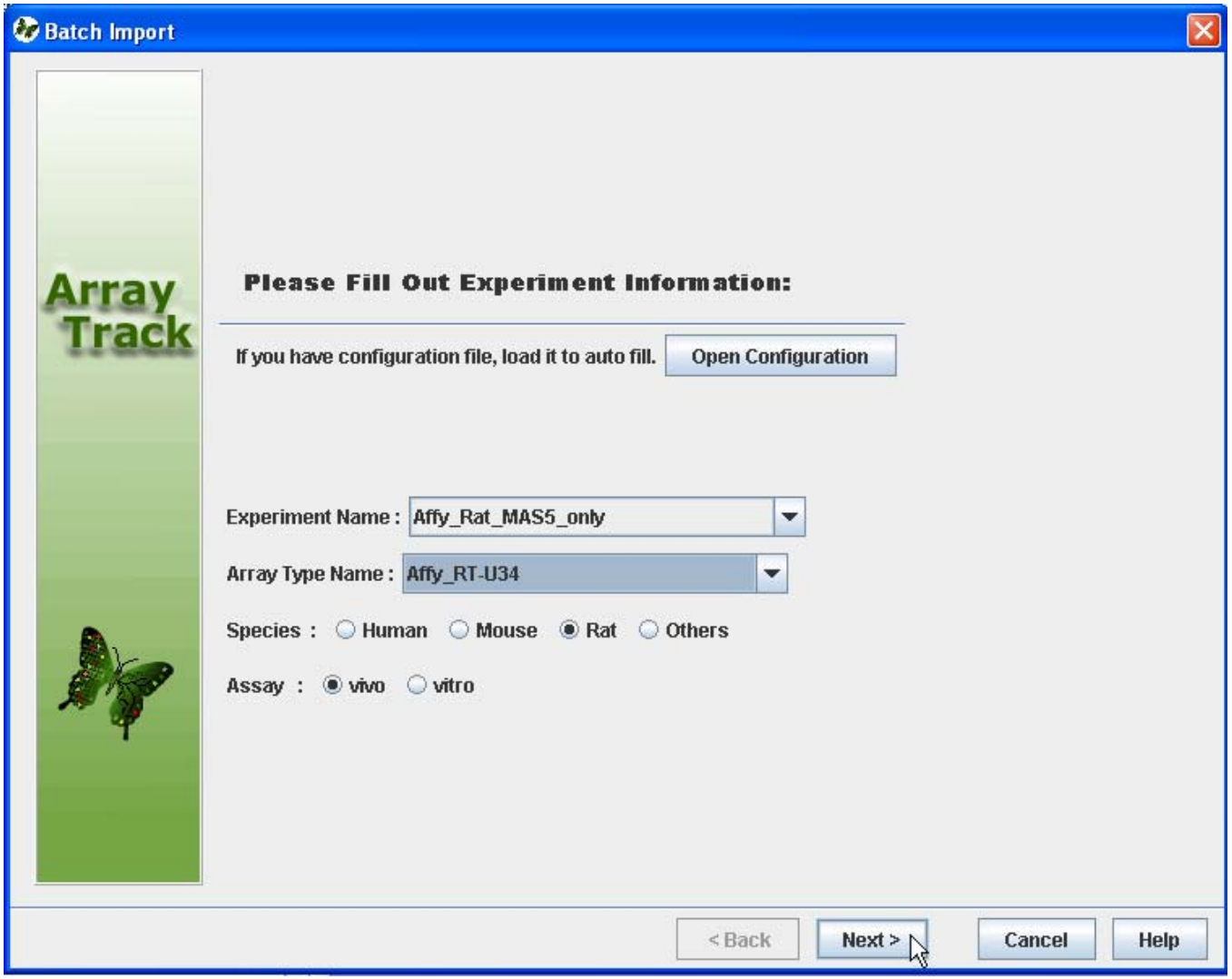
Data import is finished!

Batch Import Wizard
Import CEL File Data

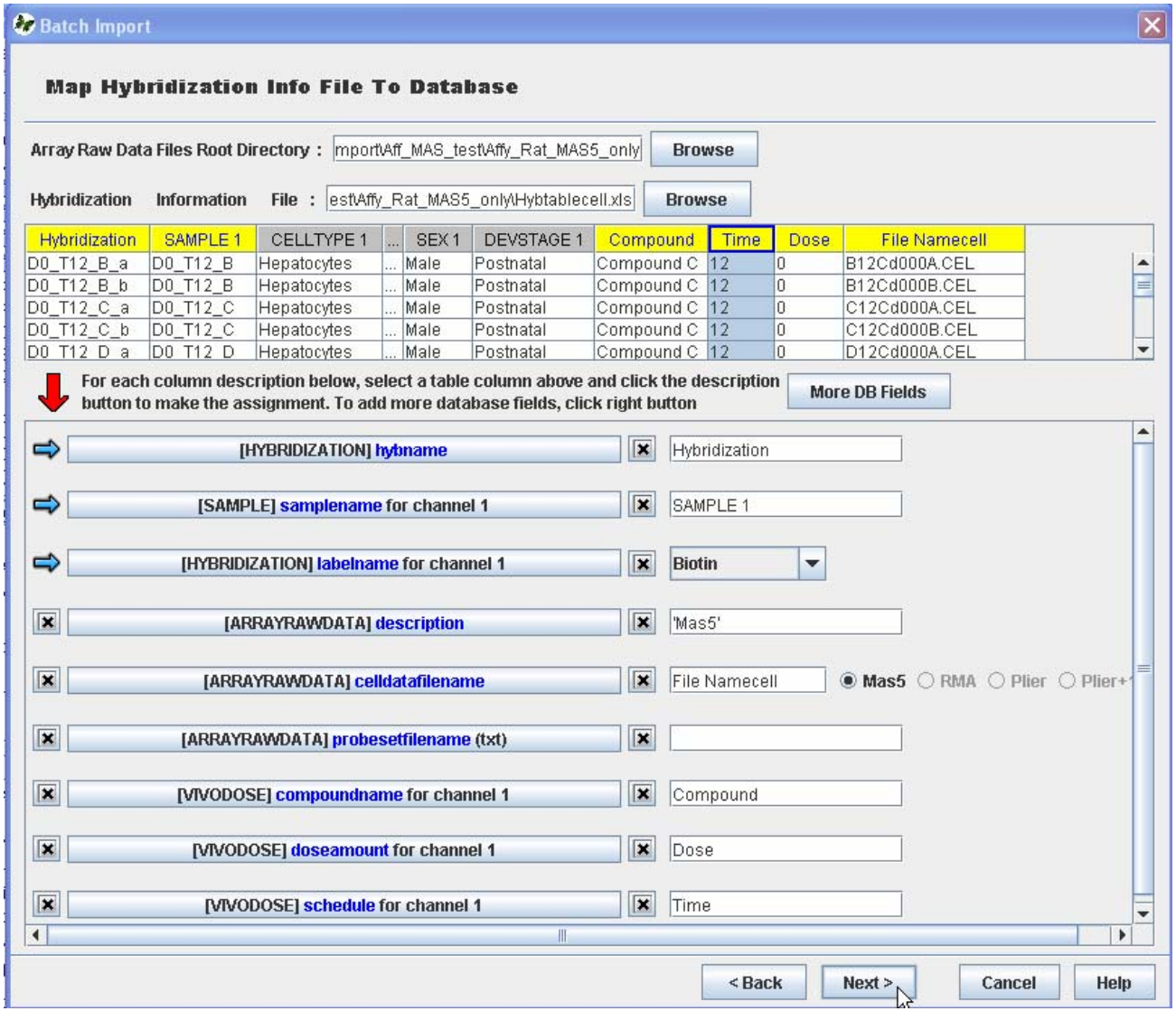


Right-click the experiment and choose “Batch Import Data”.

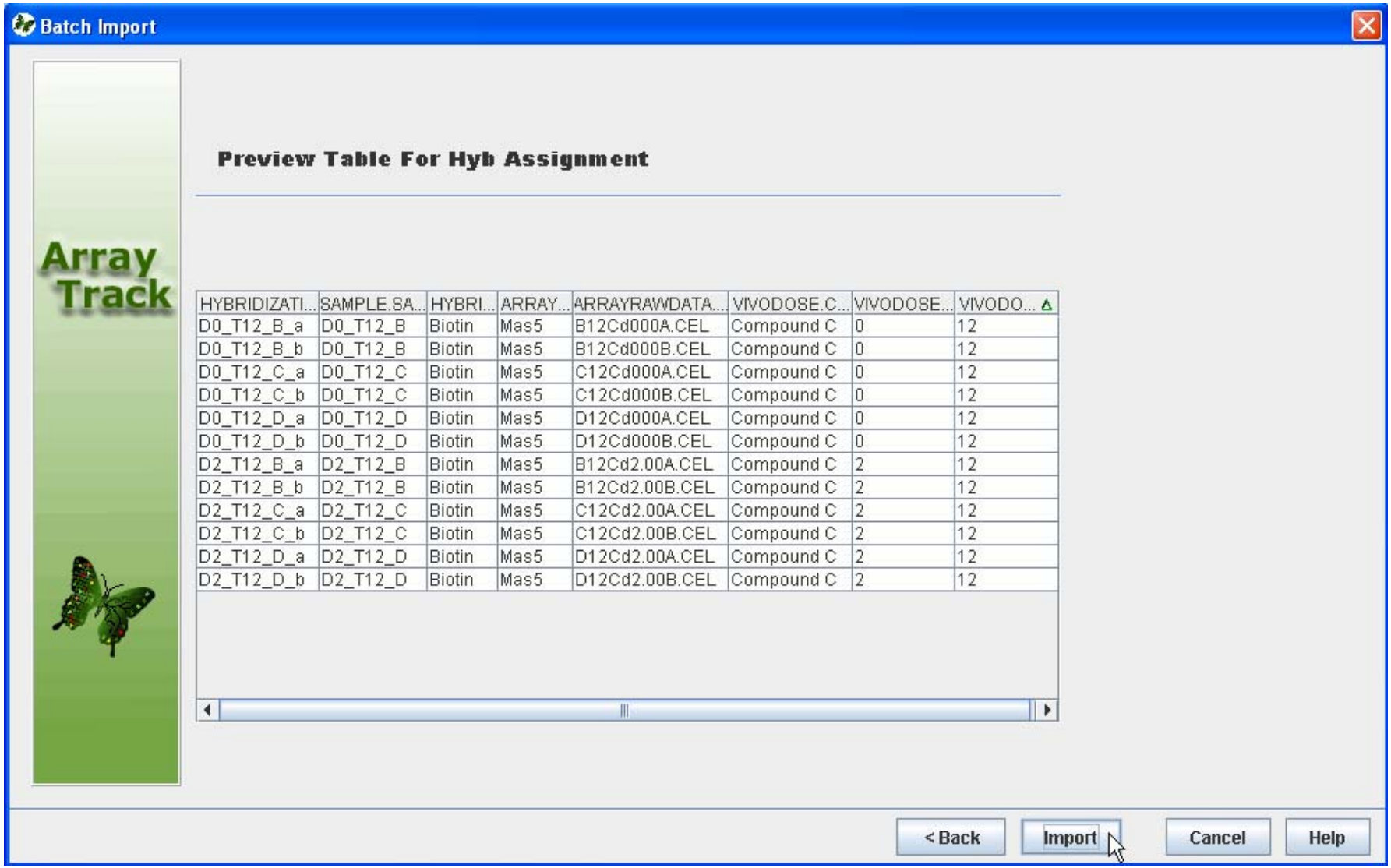
Batch Import Wizard
Import CEL File Data



Batch Import Wizard
Import CEL File Data



Batch Import Wizard
Import CEL File Data

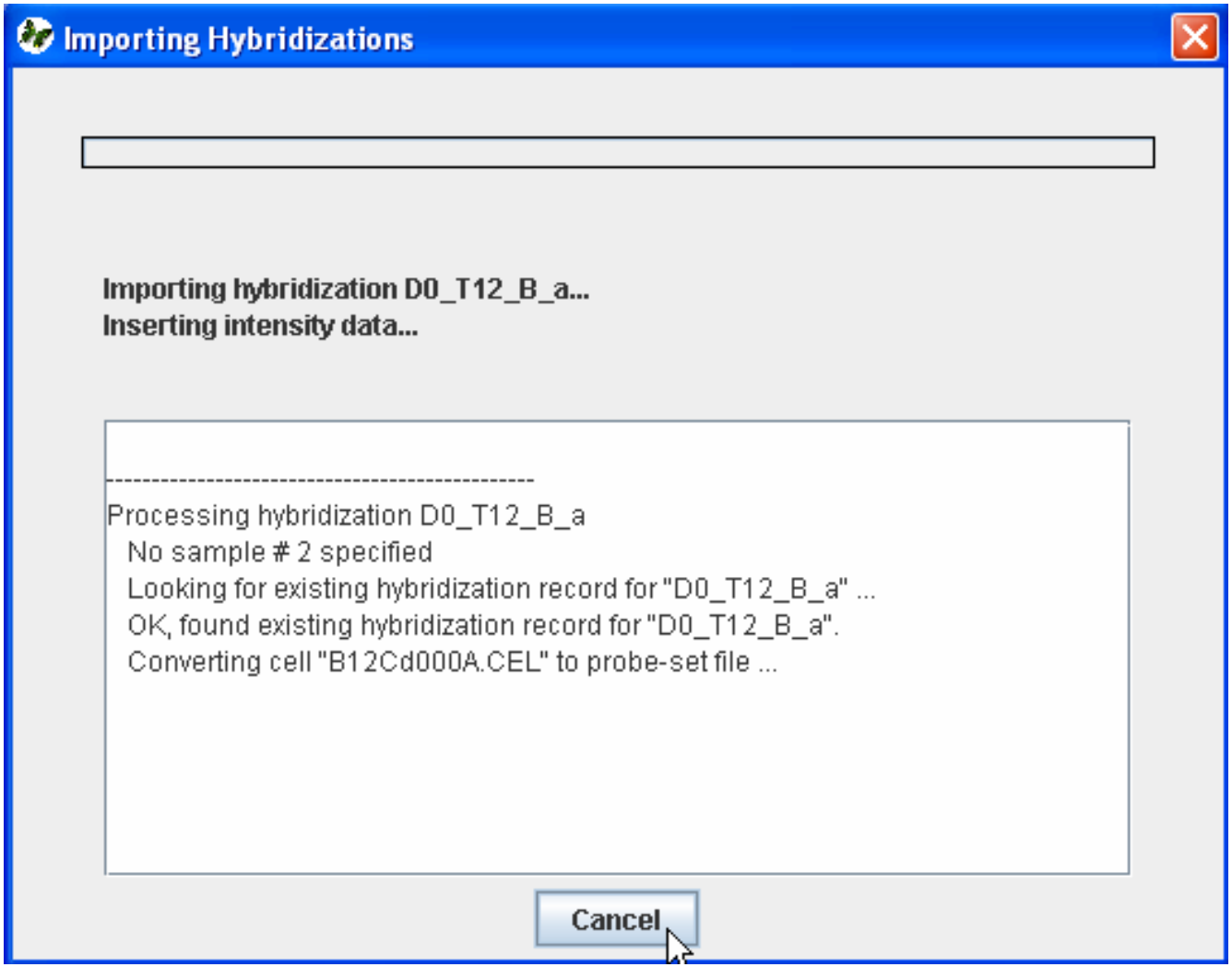


HYBRIDIZATI...	SAMPLE.SA...	HYBRI...	ARRAY...	ARRAYRAWDATA...	VIVODOSE.C...	VIVODOSE...	VIVODO...
D0_T12_B_a	D0_T12_B	Biotin	Mas5	B12Cd000A.CEL	Compound C	0	12
D0_T12_B_b	D0_T12_B	Biotin	Mas5	B12Cd000B.CEL	Compound C	0	12
D0_T12_C_a	D0_T12_C	Biotin	Mas5	C12Cd000A.CEL	Compound C	0	12
D0_T12_C_b	D0_T12_C	Biotin	Mas5	C12Cd000B.CEL	Compound C	0	12
D0_T12_D_a	D0_T12_D	Biotin	Mas5	D12Cd000A.CEL	Compound C	0	12
D0_T12_D_b	D0_T12_D	Biotin	Mas5	D12Cd000B.CEL	Compound C	0	12
D2_T12_B_a	D2_T12_B	Biotin	Mas5	B12Cd2.00A.CEL	Compound C	2	12
D2_T12_B_b	D2_T12_B	Biotin	Mas5	B12Cd2.00B.CEL	Compound C	2	12
D2_T12_C_a	D2_T12_C	Biotin	Mas5	C12Cd2.00A.CEL	Compound C	2	12
D2_T12_C_b	D2_T12_C	Biotin	Mas5	C12Cd2.00B.CEL	Compound C	2	12
D2_T12_D_a	D2_T12_D	Biotin	Mas5	D12Cd2.00A.CEL	Compound C	2	12
D2_T12_D_b	D2_T12_D	Biotin	Mas5	D12Cd2.00B.CEL	Compound C	2	12

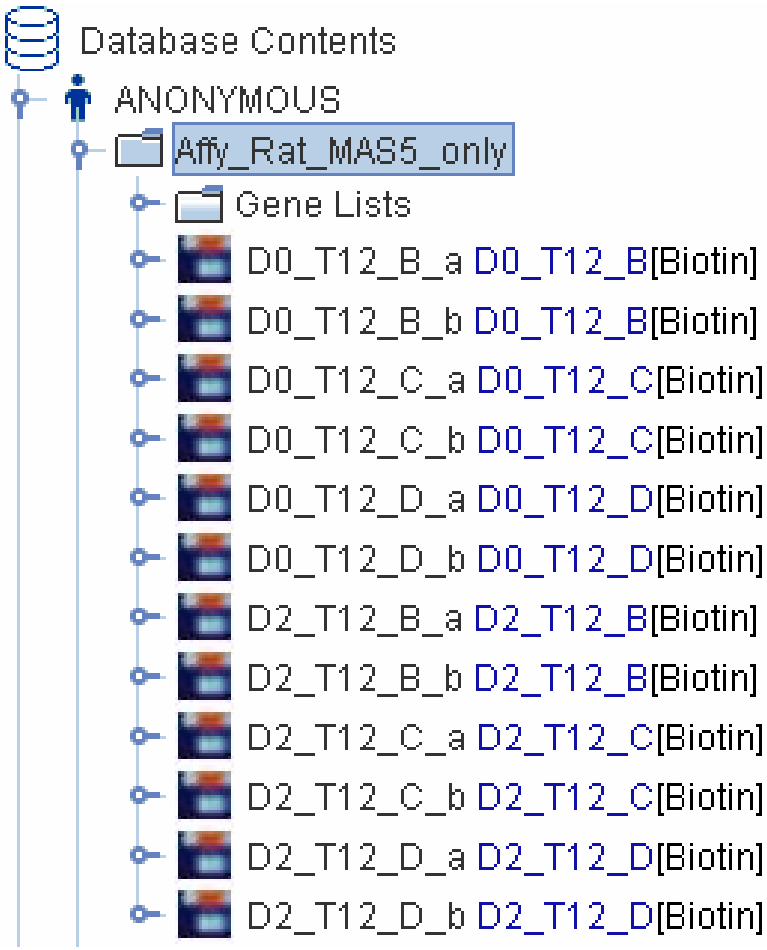
Batch Import Wizard
Import CEL File Data

Click Yes to save the configuration file which stores all the assignments, making it easier for the second time import in case of the first import failure.

Batch Import Wizard
Import CEL File Data

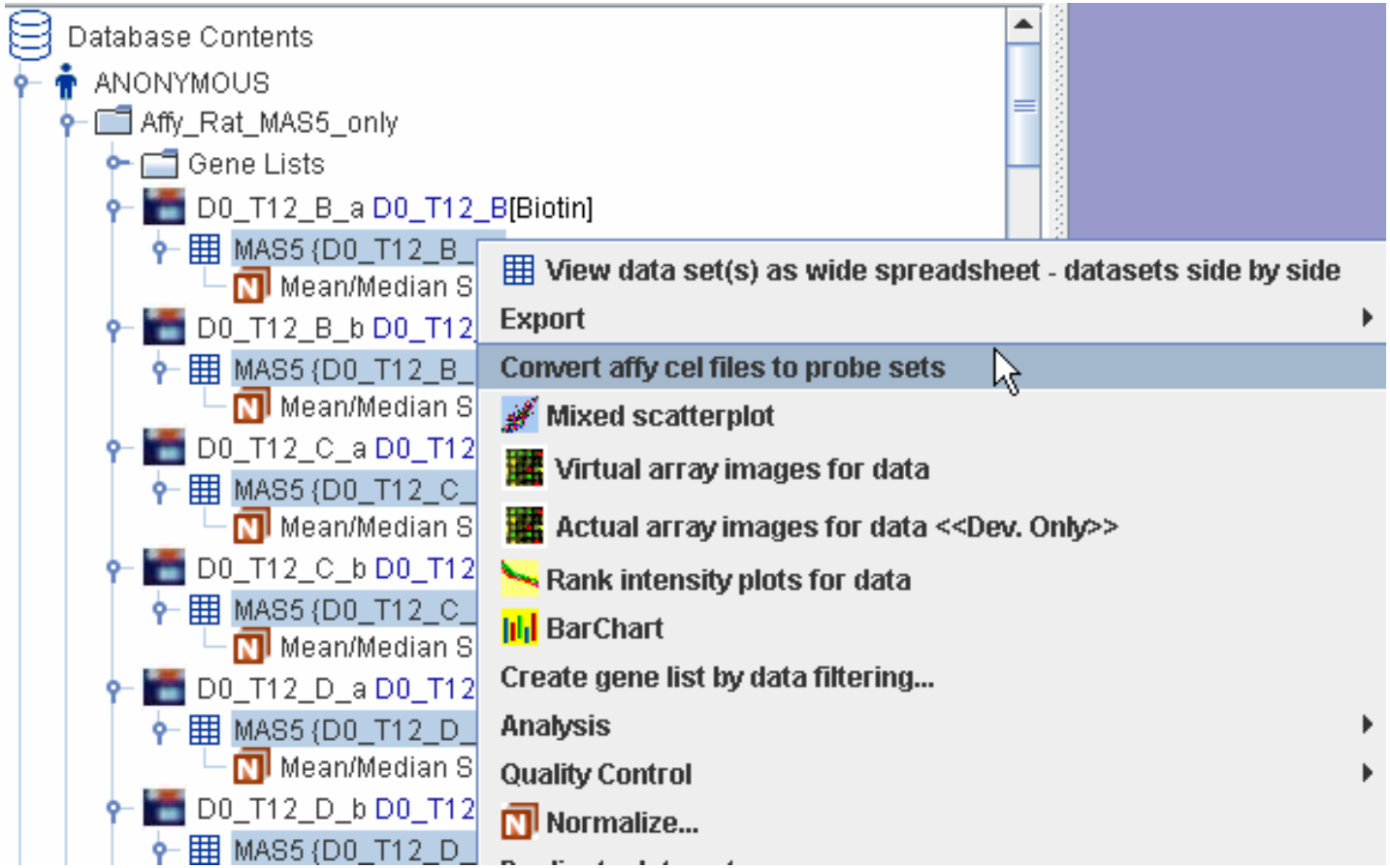


Batch Import Wizard
Import CEL File Data



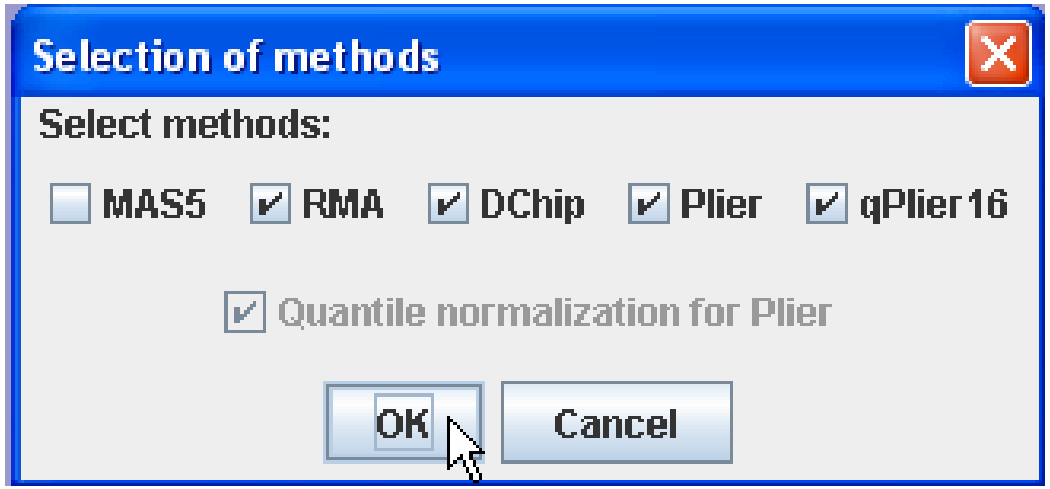
Import is finished.

Convert CEL file to probe set file

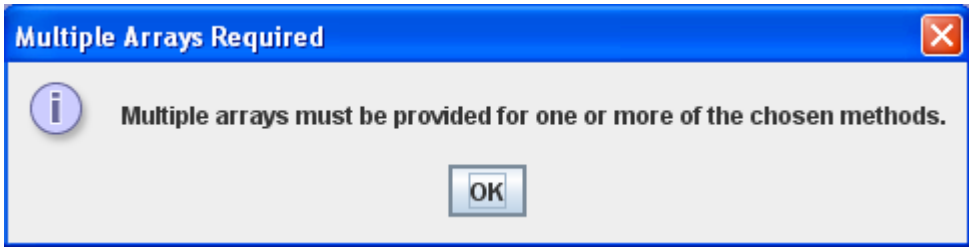


Convert CEL file to probe set file

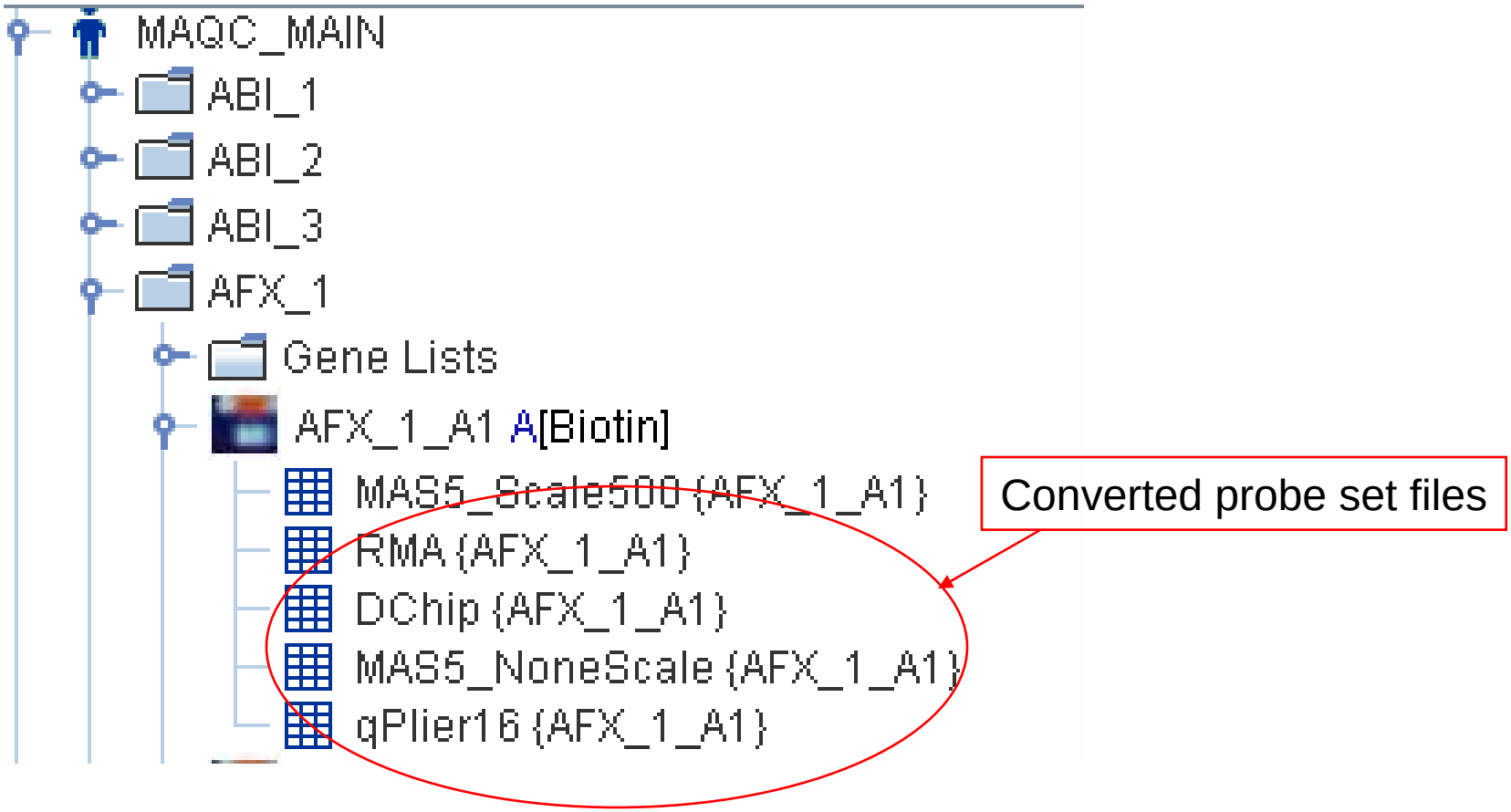
Select the data, right-click, choose “Convert affy cel files to probe sets. ArrayTrack can convert cel to RMA, DChip, qPlier and qPlier16



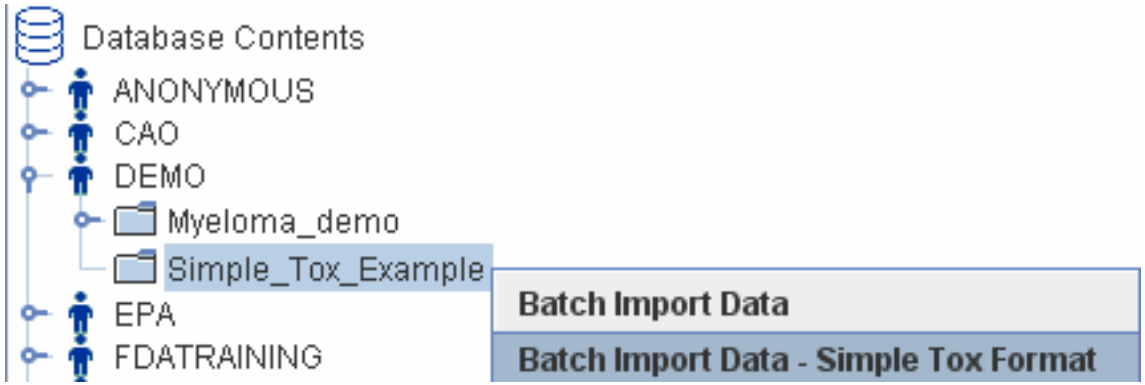
Keep in mind that you have to choose multiple data to do the converting of all the methods except MAS5. For example, you can not select just one data to do RMA converting, otherwise you will get the following message:



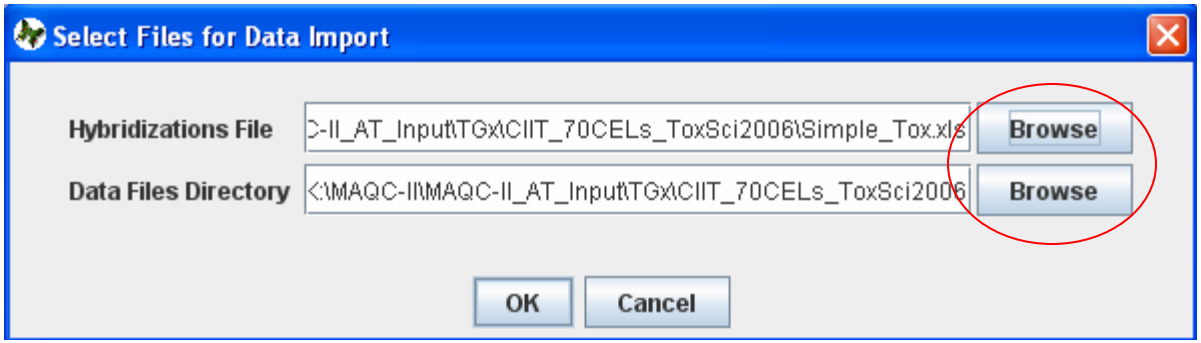
Convert CEL file to probe set file



Batch Import – Simple Tox Format



Right-click the experiment, choose “Batch Import Data – Simple Tox Format”



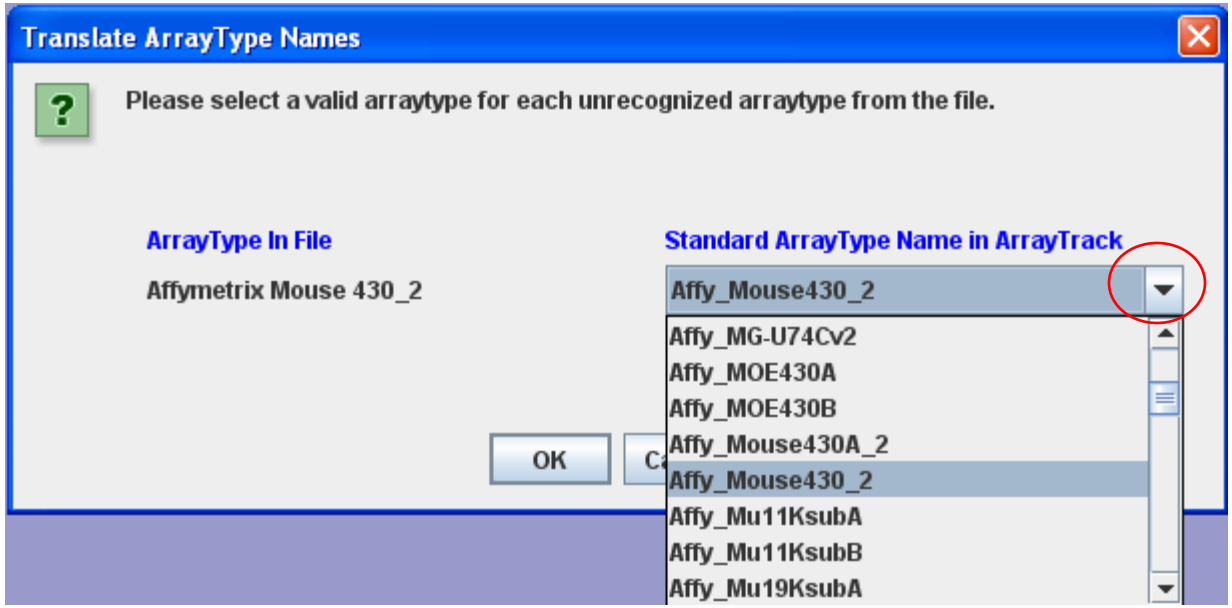
Click Browse button to locate the hybridization file and data files directory. The next slide shows an example of hybridization table file.

Batch Import – Simple Tox Format

This is an example of hybridization table file (which is attached). User can use this as a template to make his own hybridization file. Just make sure that your column titles are exactly same as this template.

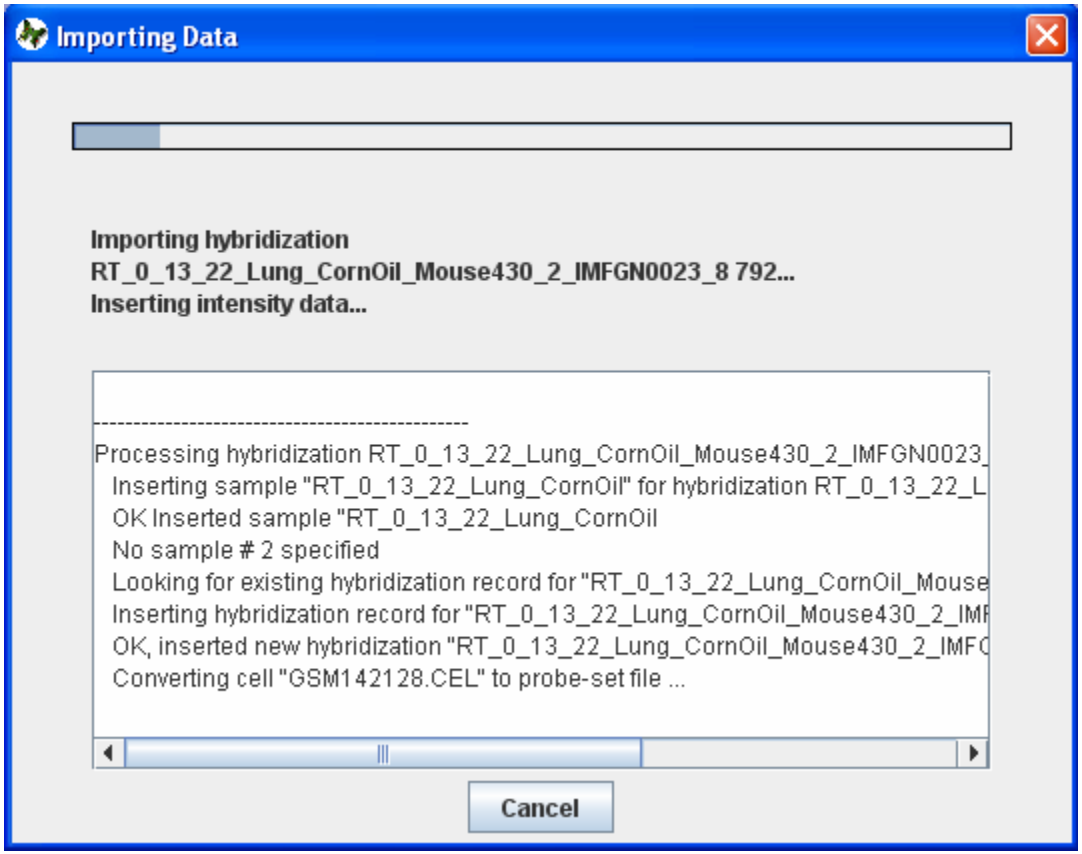
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Array_ID	Animal_ID	Institutio	StudyTitle	StudyTy	Class	ClassCo	ClassRe	Compou	CAS	Control	Treatme	Dose	DoseUn	HybName	Sample
2	792	1-22	Hamner	Hamner Mice Lung	Repeat Dose				Corn Oil		Y	C	0	mg/kg	RT_0_13_22_Lu	RT_0_13_
3	793	1-24	Hamner	Hamner Mice Lung	Repeat Dose				Corn Oil		Y	C	0	mg/kg	RT_0_13_24_Lu	RT_0_13_
4	794	1-25	Hamner	Hamner Mice Lung	Repeat Dose				Corn Oil		Y	C	0	mg/kg	RT_0_13_25_Lu	RT_0_13_
5	795	1-27	Hamner	Hamner Mice Lung	Repeat Dose				Rodent Chow		Y	C	0	ppm	RT_0_13_27_Lu	RT_0_13_
6	796	1-28	Hamner	Hamner Mice Lung	Repeat Dose				Rodent Chow		Y	C	0	ppm	RT_0_13_28_Lu	RT_0_13_
7	797	1-29	Hamner	Hamner Mice Lung	Repeat Dose				Rodent Chow		Y	C	0	ppm	RT_0_13_29_Lu	RT_0_13_
8	798	1-12	Hamner	Hamner Mice Lung	Repeat Do	Non-lungtu	NLT	NTP_No16	N-(1-napht	1465-25-4	N	NLT	2000	ppm	RT_2000_13_12	RT_2000_
9	799	1-14	Hamner	Hamner Mice Lung	Repeat Do	Non-lungtu	NLT	NTP_No16	N-(1-napht	1465-25-4	N	NLT	2000	ppm	RT_2000_13_14	RT_2000_
10	800	1-15	Hamner	Hamner Mice Lung	Repeat Do	Non-lungtu	NLT	NTP_No16	N-(1-napht	1465-25-4	N	NLT	2000	ppm	RT_2000_13_15	RT_2000_
11	801	1-2	Hamner	Hamner Mice Lung	Repeat Do	lung tumor	LT	NTP_No14	1,5-Naphth	2243-62-1	N	LT	2000	ppm	RT_2000_13_2_	RT_2000_

Batch Import – Simple Tox Format

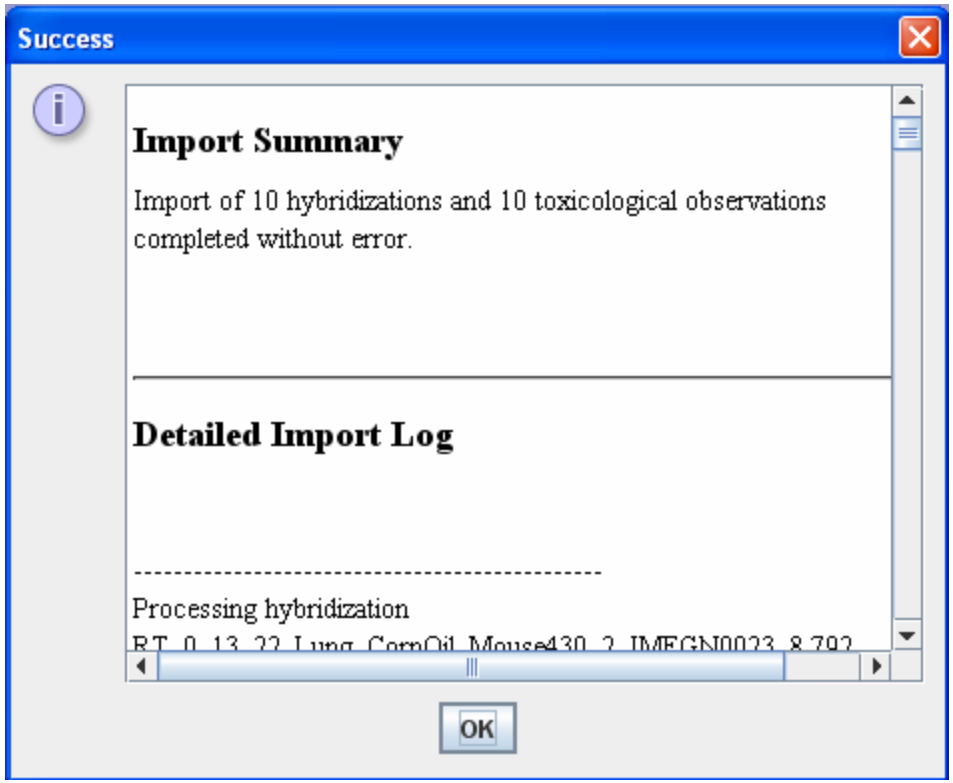


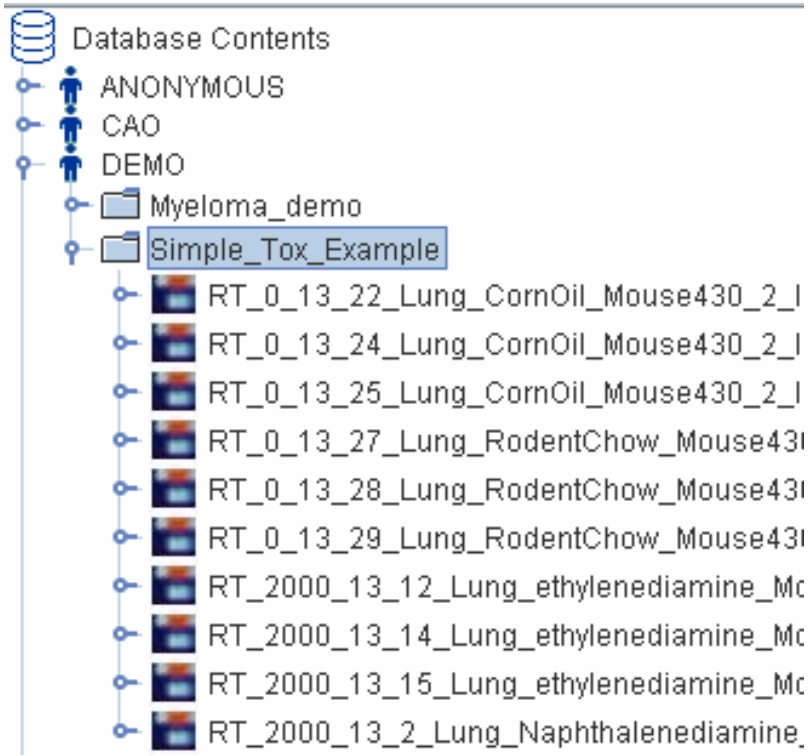
Select the right array type.

Batch Import – Simple Tox Format



Batch Import – Simple Tox Format

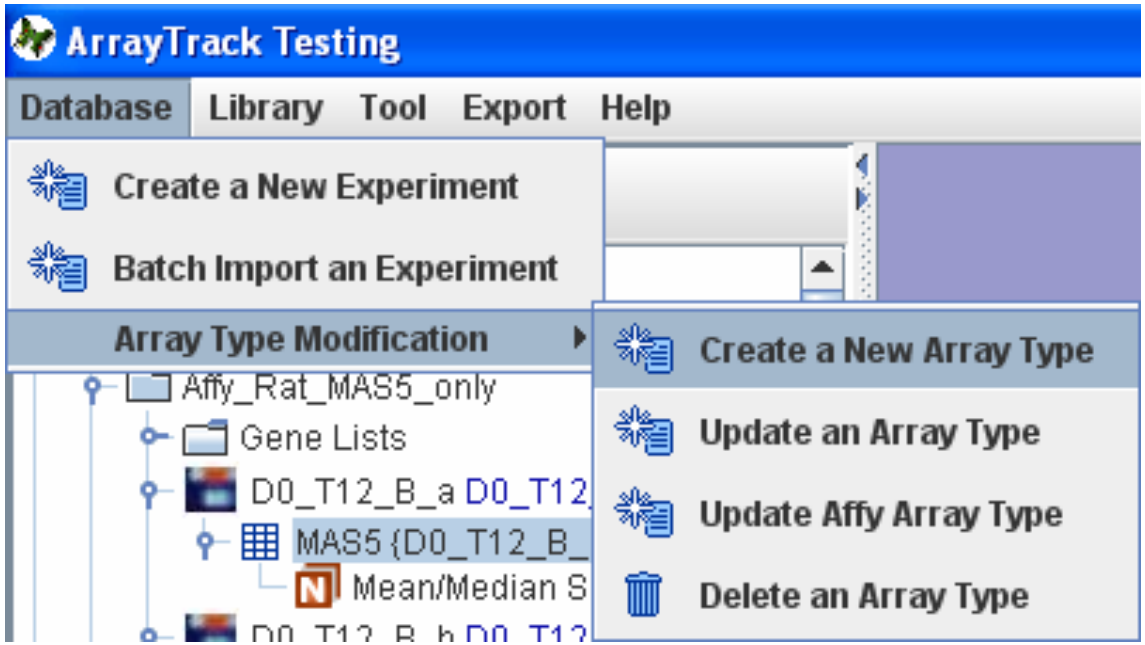




Batch import is successfully finished.

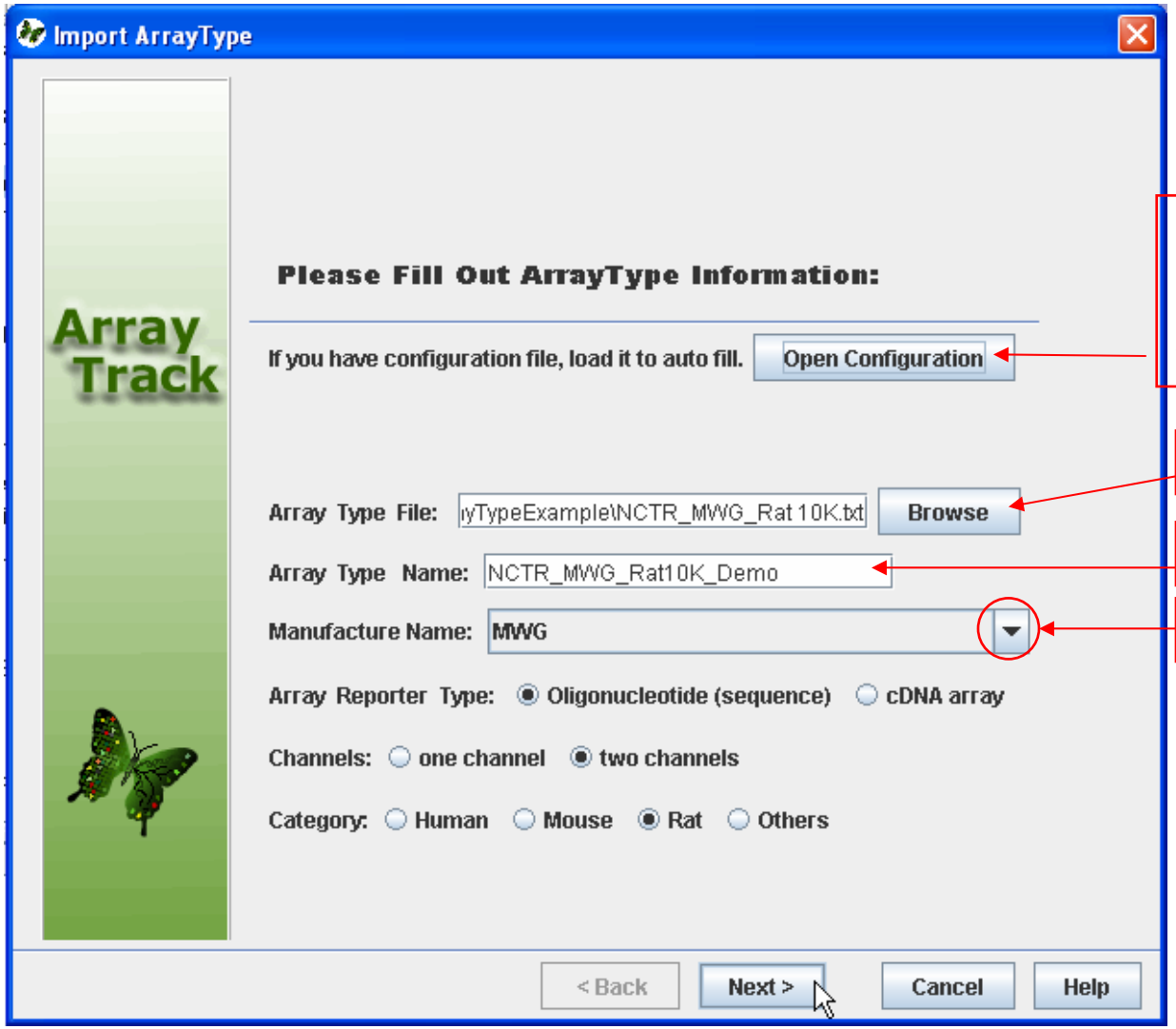
Create Array Type

Click “Database” menu, choose “Array Type Modification” ->”Create a new Array Type”.



Create Array Type - continued

The following is the example of creating array type for 2 channel data.



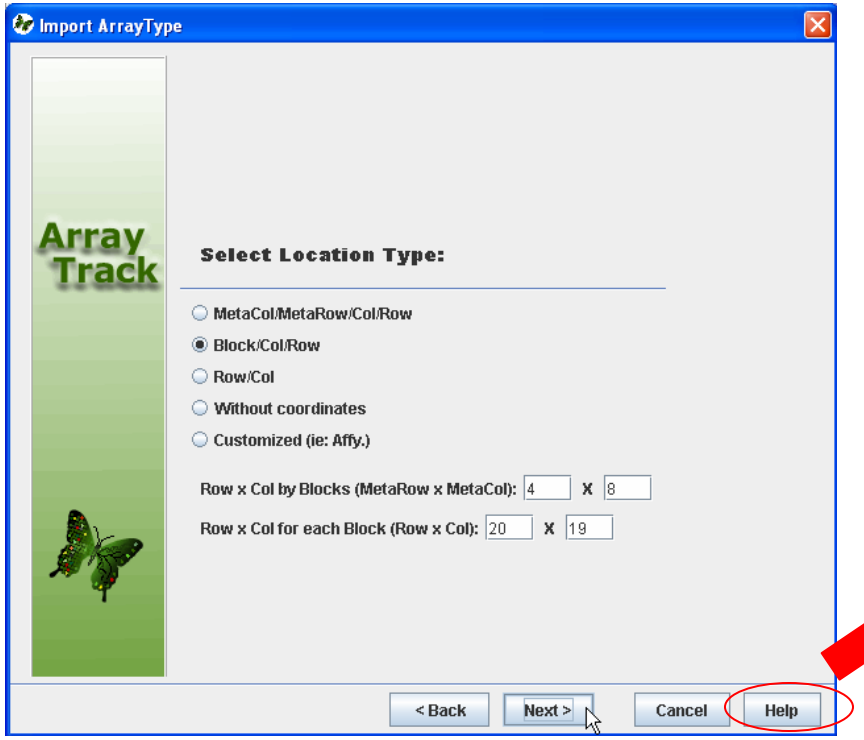
Ignore this button If you don't have configuration file. You will be asked to save the configuration file later, see slide #28

Locate the array type file

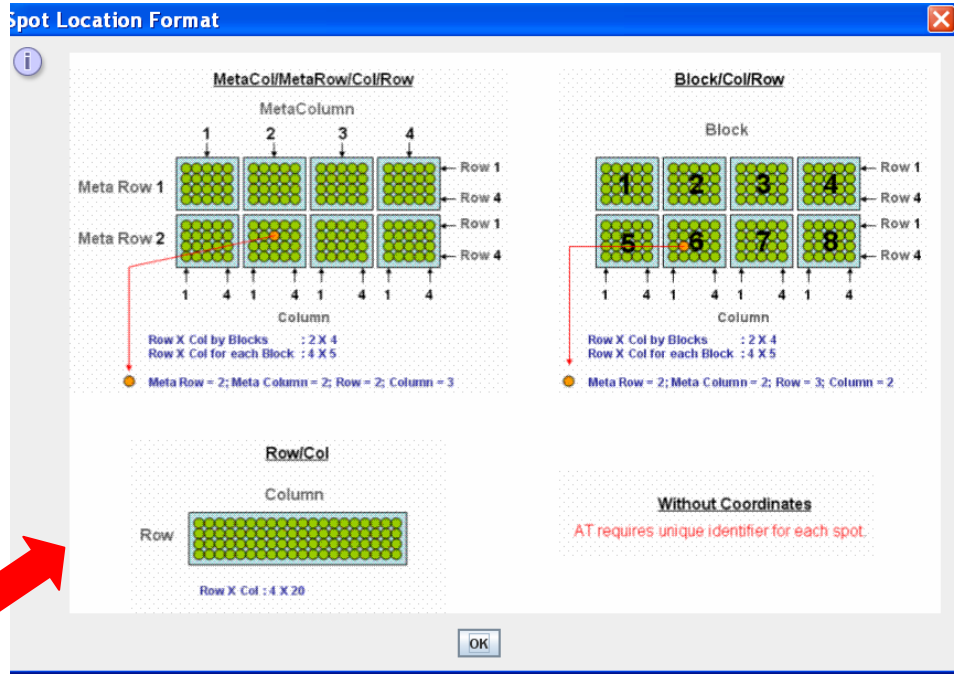
Type in the array type name

Choose the manufacture name

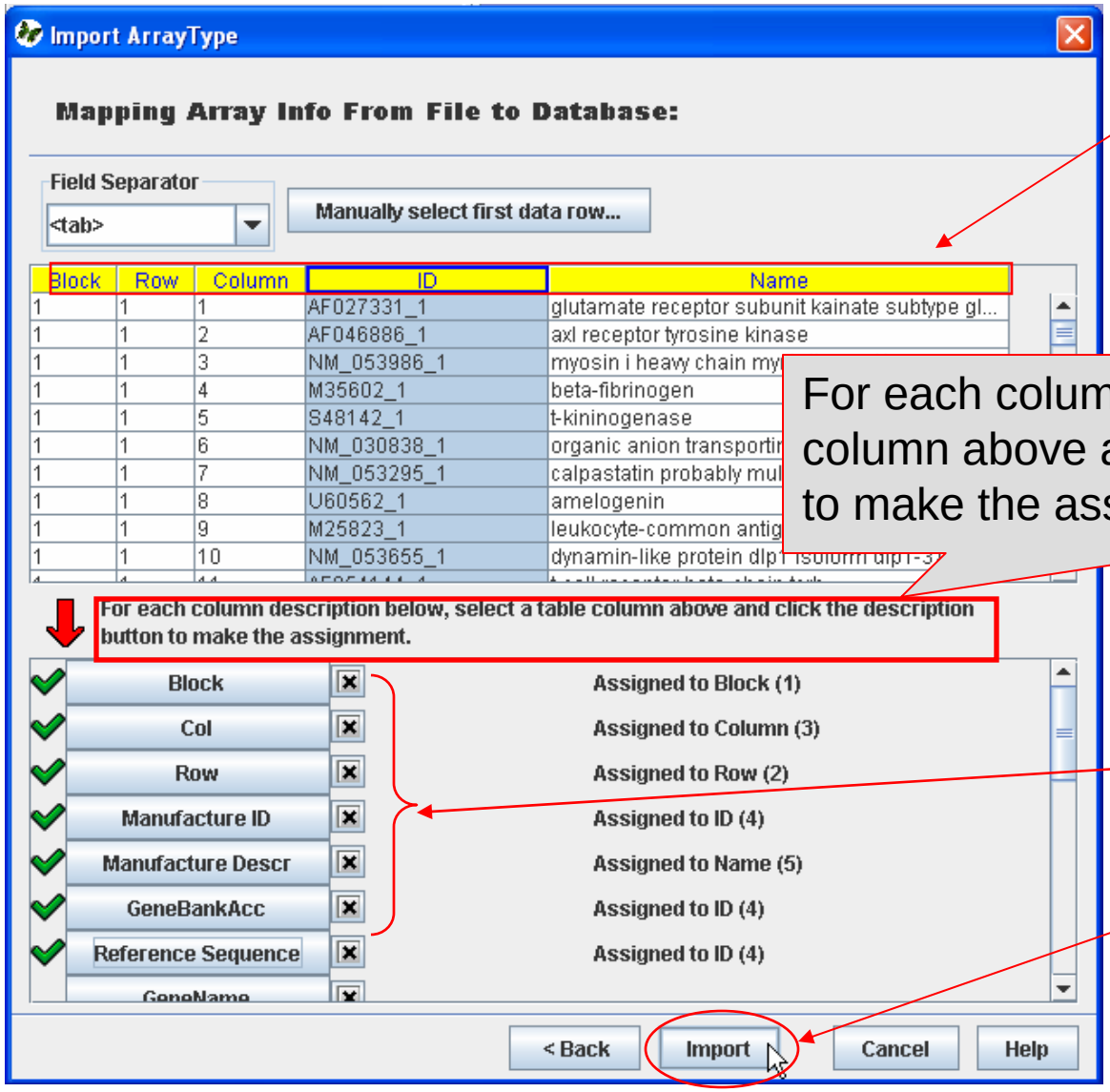
Create Array Type - continued



Click Help button to view the location definition



Create Array Type - continued



Column header

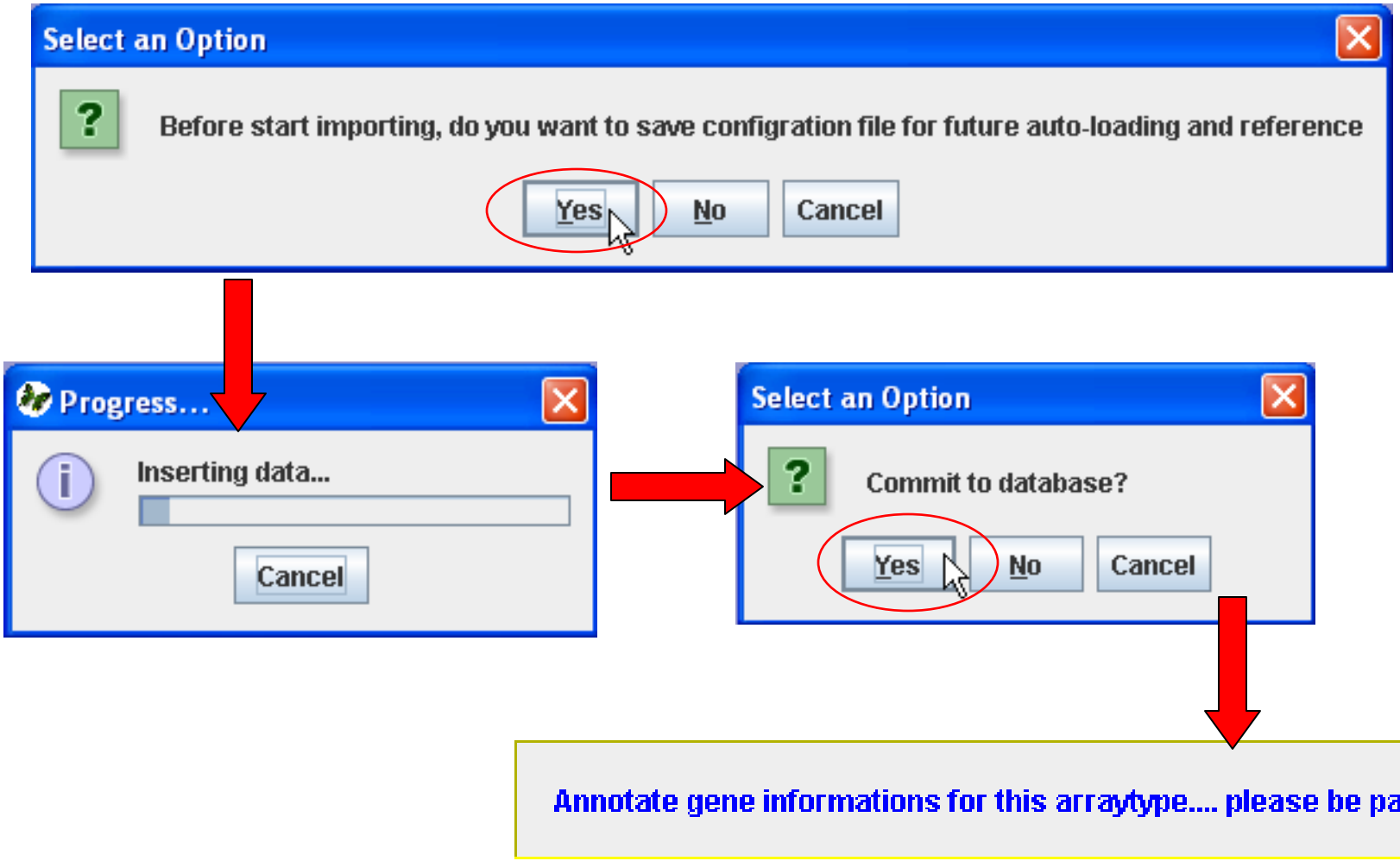
For each column description, select a table column above and click the description button to make the assignment

Description buttons

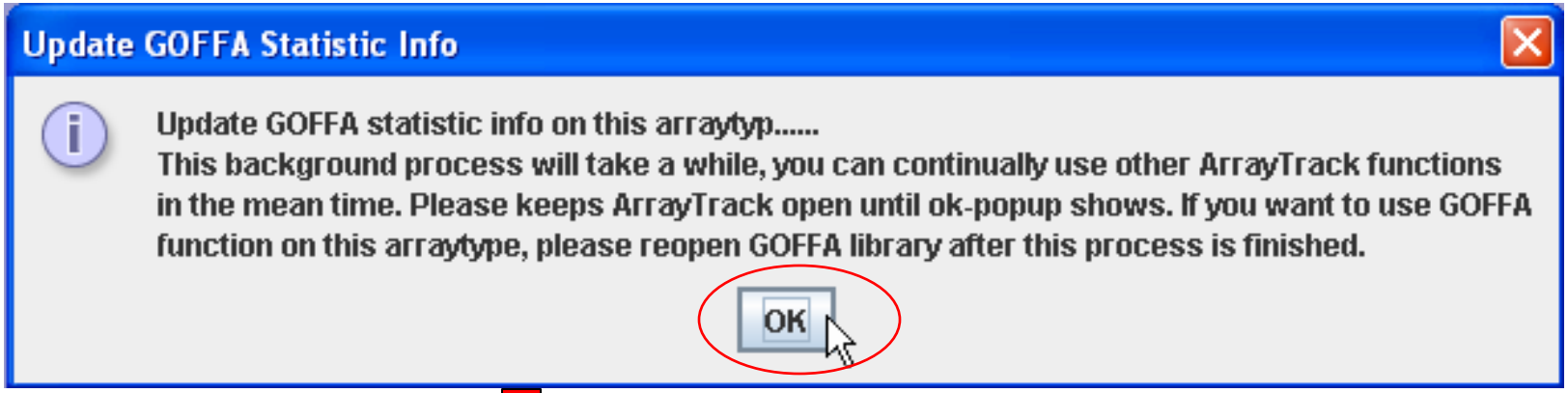
Click Import button

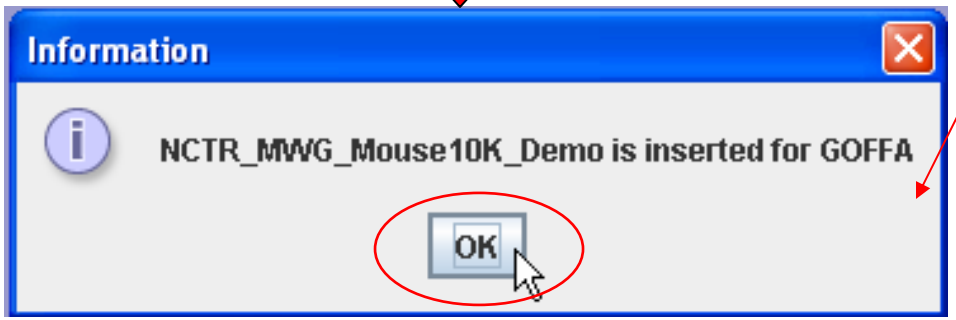
Create Array Type - continued

The user can save configuration file for future auto-loading and reference. Click Yes button to save to the local drive.



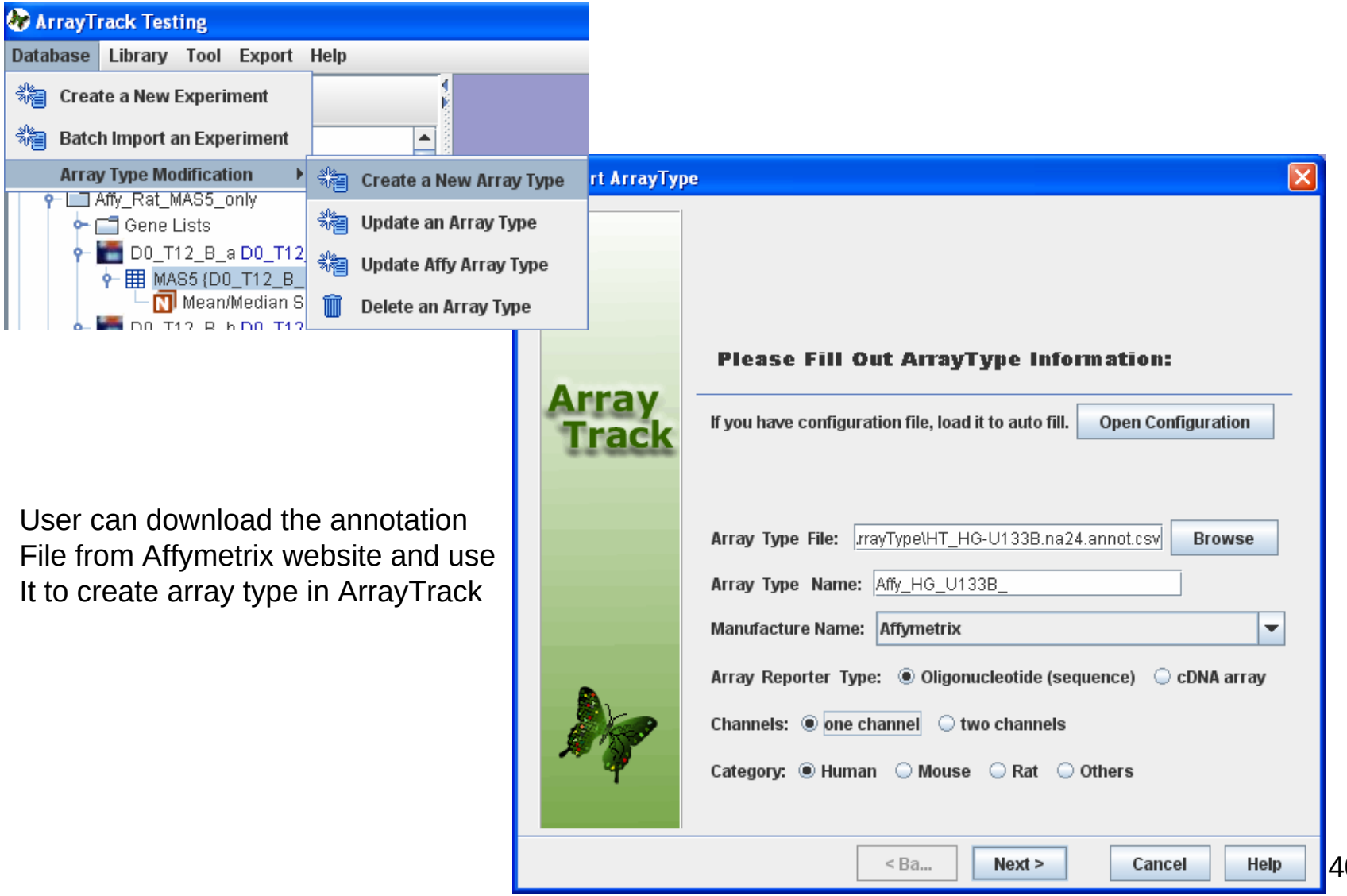
Create Array Type - continued





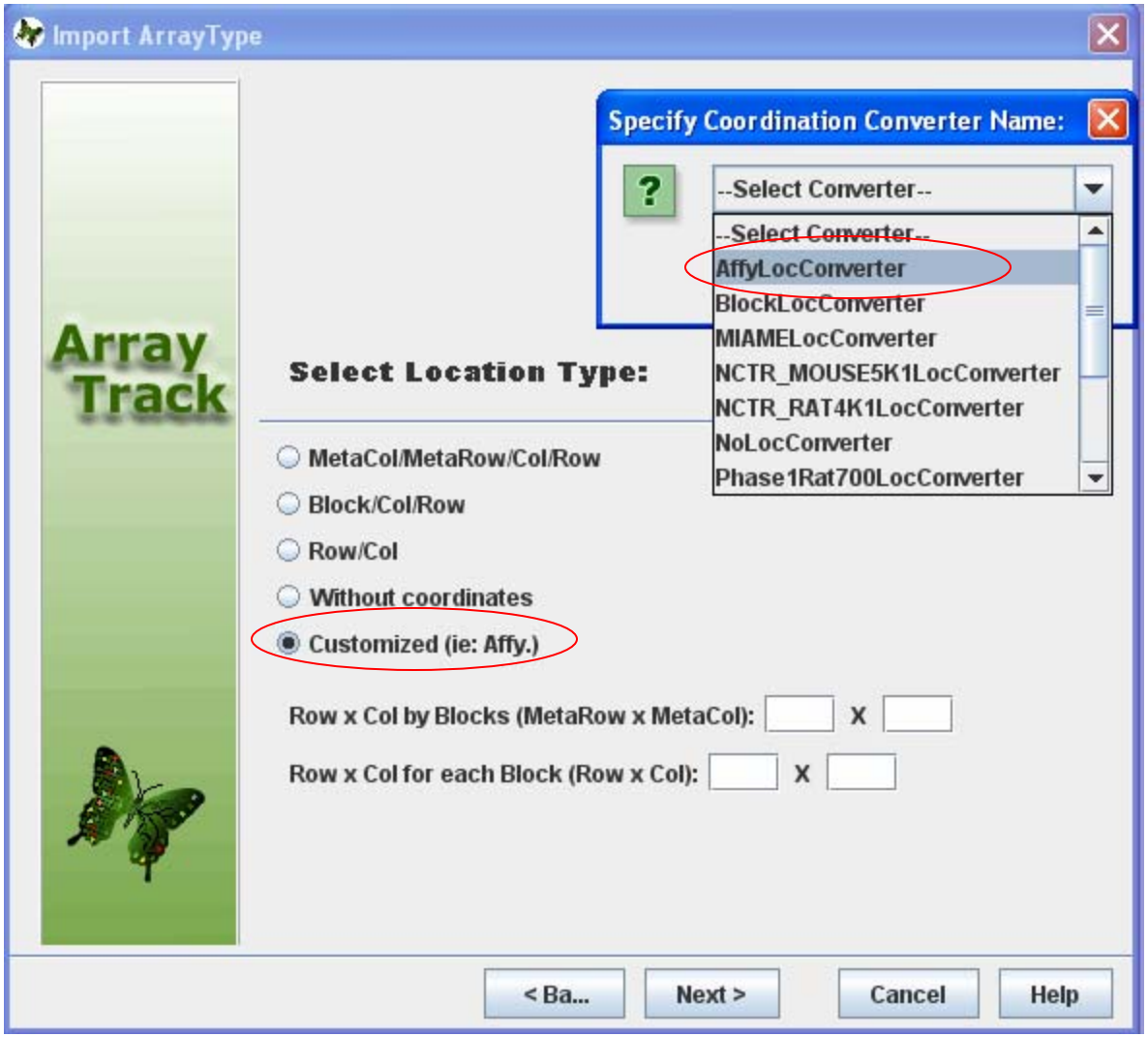
The array type is created successfully.

Create Array Type - Affymetrix

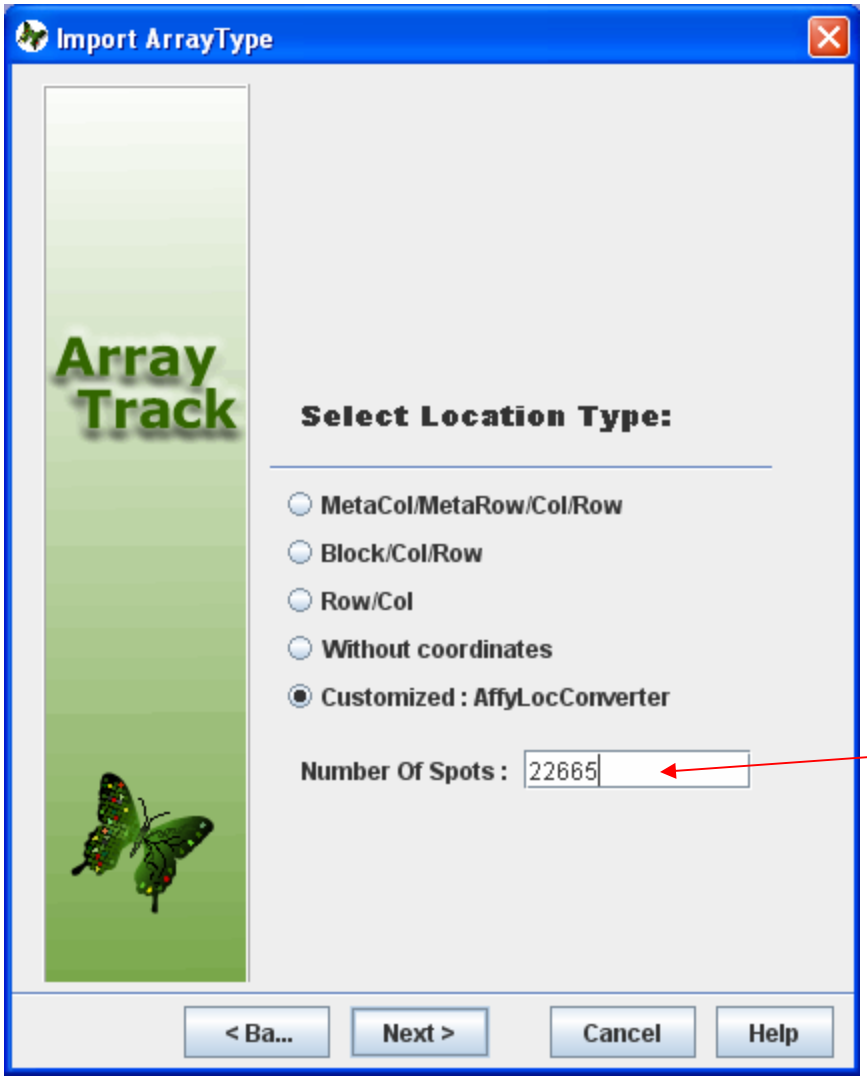


User can download the annotation File from Affymetrix website and use It to create array type in ArrayTrack

Create Array Type - Affymetrix



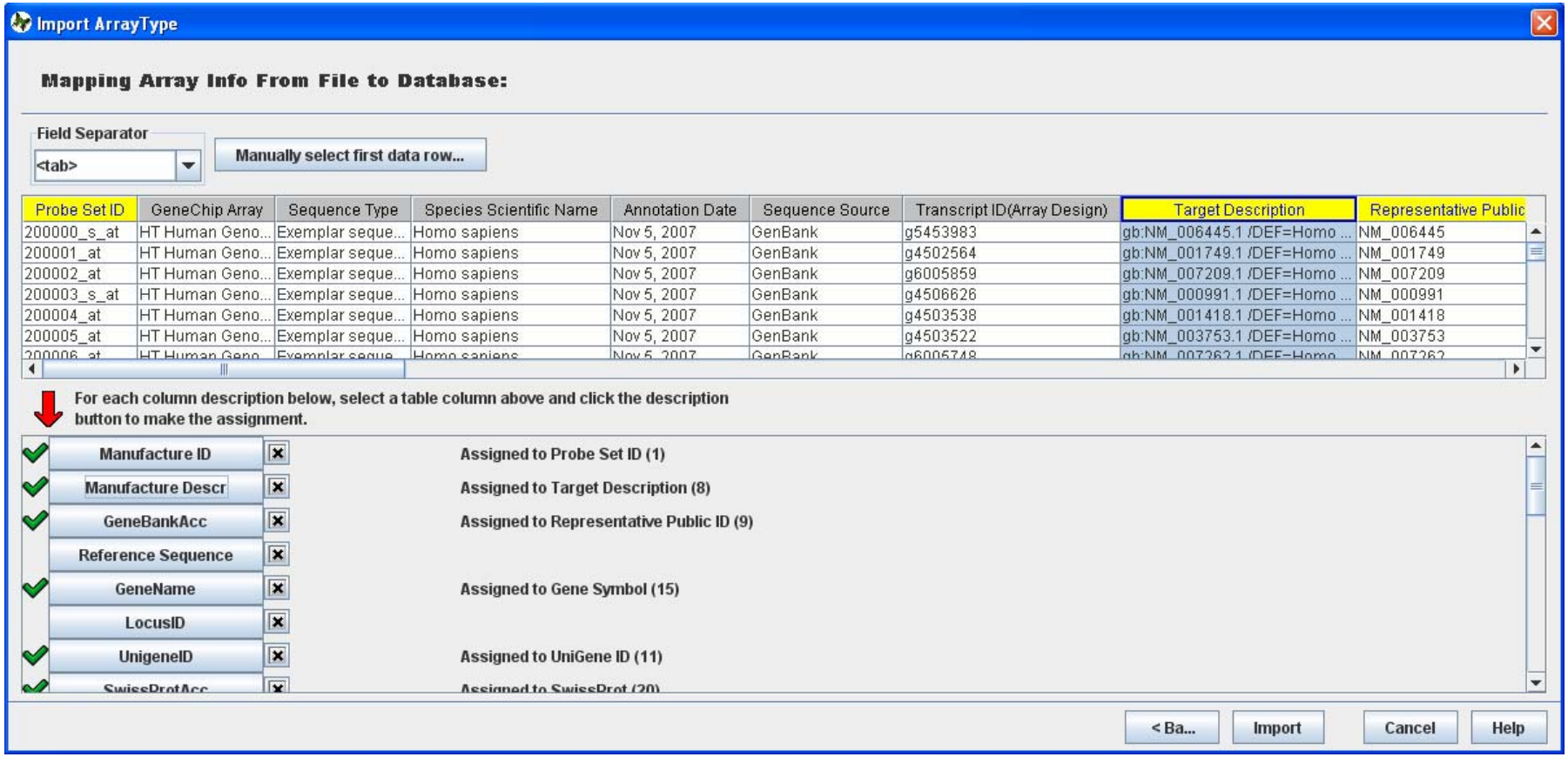
Create Array Type - Affymetrix



Specify the number of spots

Create Array Type - Affymetrix

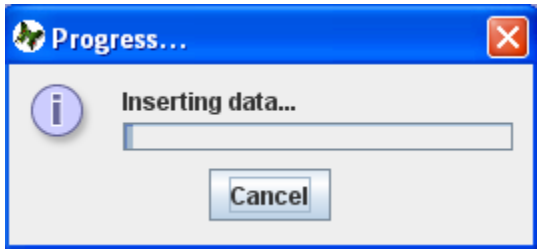
Map the columns to database fields then click “Import” button.
See the next slide for the suggested mapping fields.

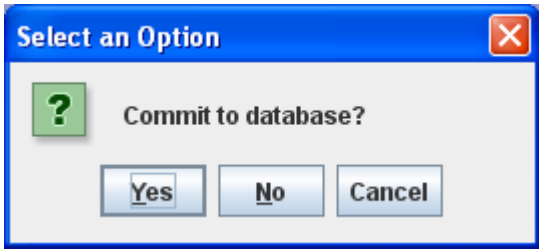


We suggest mapping the columns of Affy annotation file to the ArrayTrack fields according the following table.

Affymetrix annotation CSV file	ArrayTrack chip fields
Probe Set ID	GEN_ID_MFR
Strand	**Strand**
Sart	**BP_Start**
End	**BP_End**
Target Description	GEN_DESCR_MFR
Representative Public ID	GENEBANKACC
Archival UniGene Cluster	
UniGene ID	UNIGENEID
Genome Version	
Alignments	
Gene Title	DESCRIPTION
Gene Symbol	GENENAME
Chromosomal Location	**CHROMLOCATION**
Unigene Cluster Type	
Ensembl	**Ensembl**
Entrez Gene	LOCUSID
SwissProt	SWISSPROT_ACC_NUMBER
EC	
OMIM	OMIM
RefSeq Protein ID	**PROTEIN_REFSEQ**
RefSeq Transcript ID	REFSEQ
Gene Ontology Biological Process	BIOLOGICAL_PROCESS
Gene Ontology Cellular Component	CELULAR_COMPONENT
Gene Ontology Molecular Function	MOLECULAR_FUNCTION
Pathway	**PATHWAYS**

Create Array Type - Affymetrix

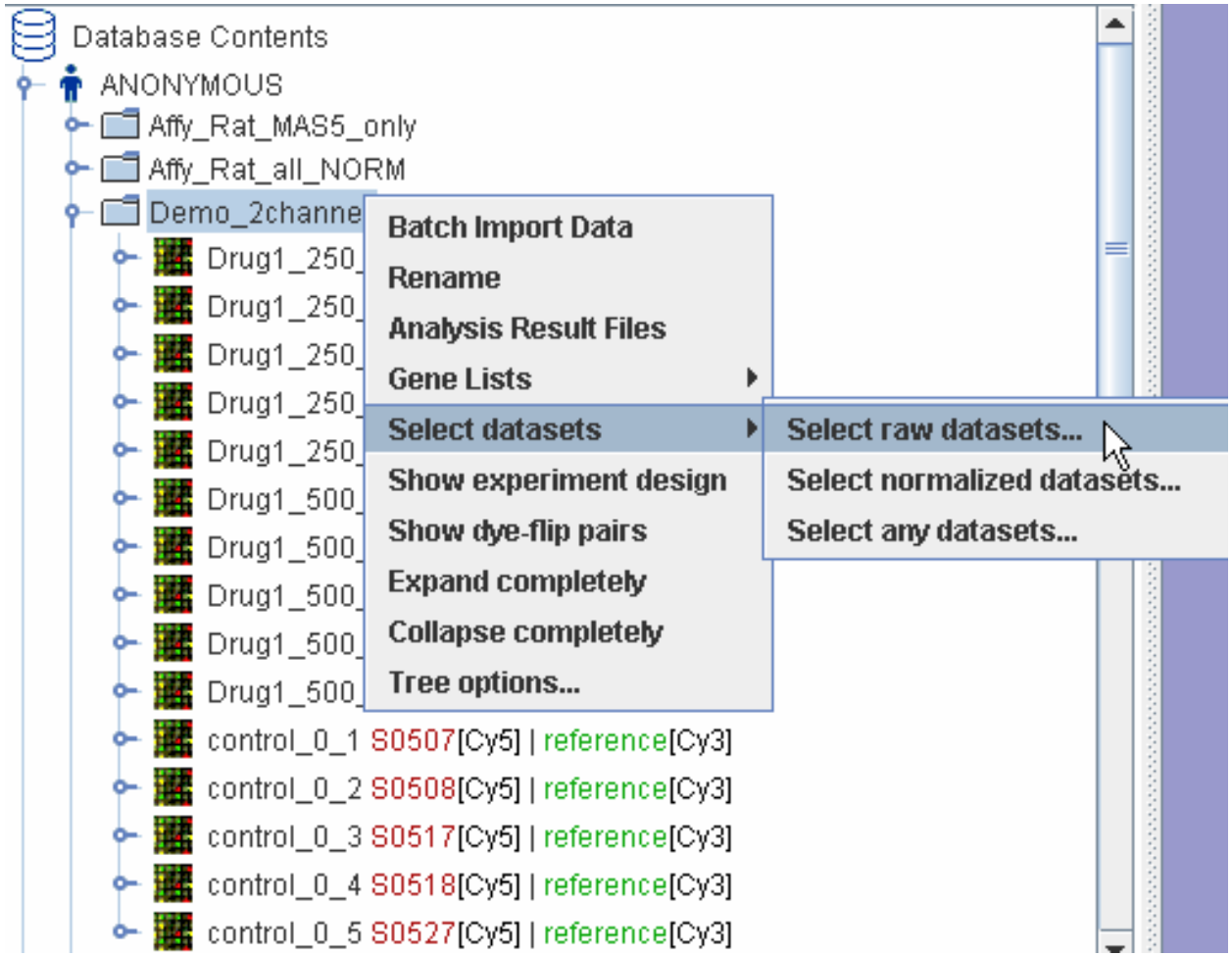




Array Track

Tutorial 9B: Data Export

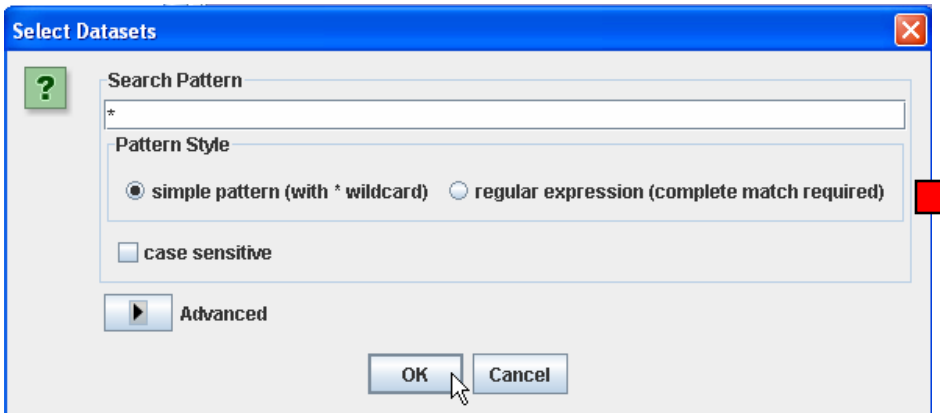
Data Export



Right-click the experiment, ->select datasets -> select raw datasets…if you want To export raw data.

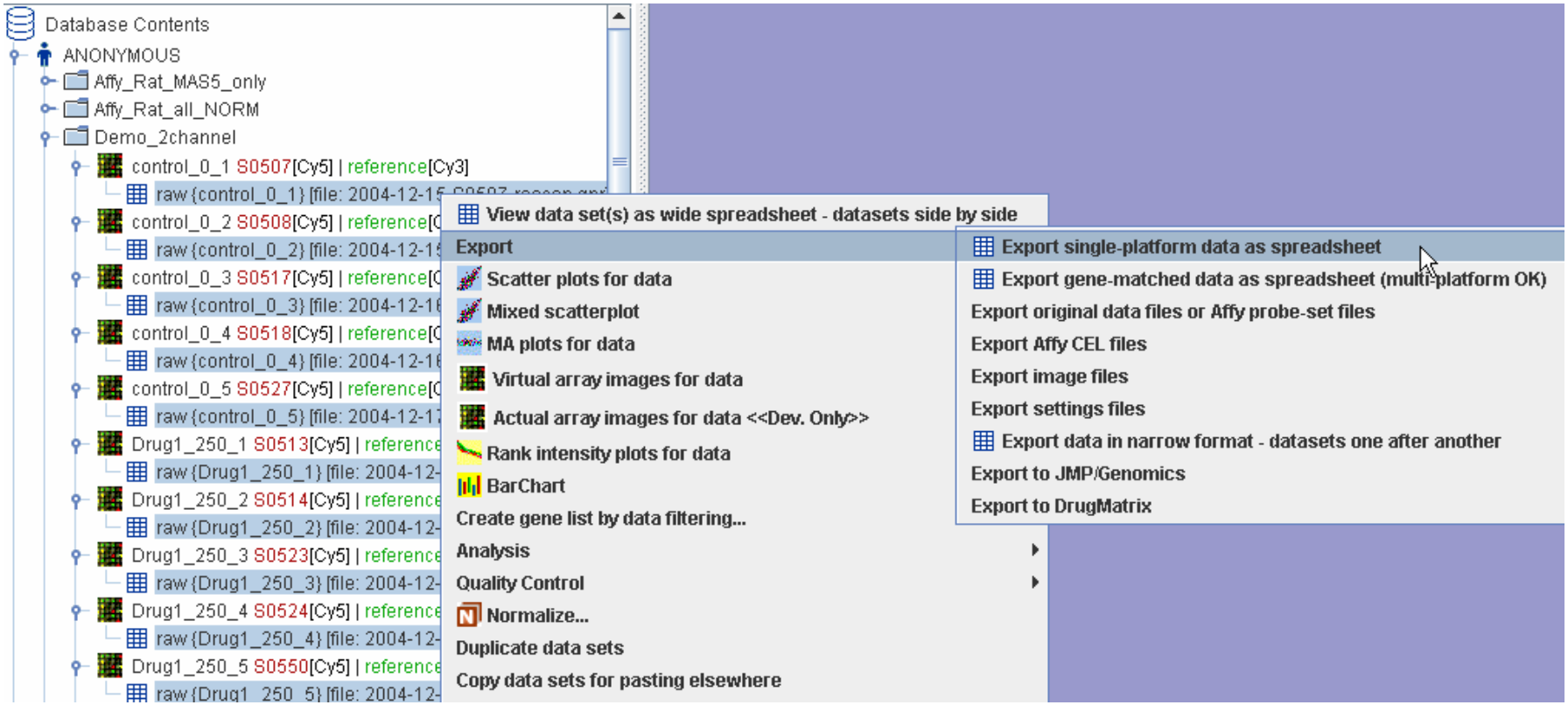
Note: user can also select normalized data to export.

Data Export



All the selected data will be highlighted.

Data Export – single-platform data spreadsheet

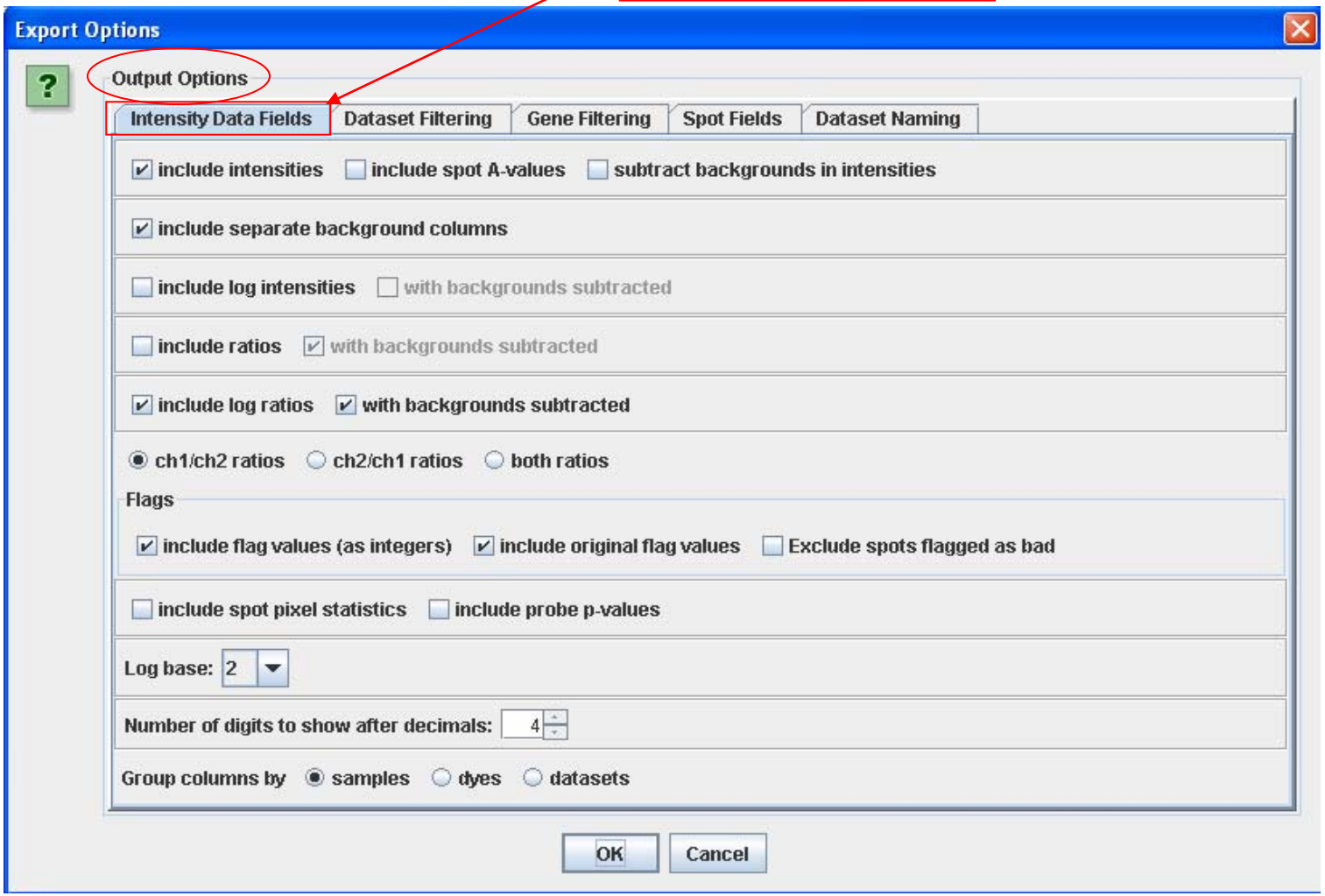


Right-click the highlighted data, select Export -> Export single-platform data as spreadsheet.

Data Export – single-platform data spreadsheet

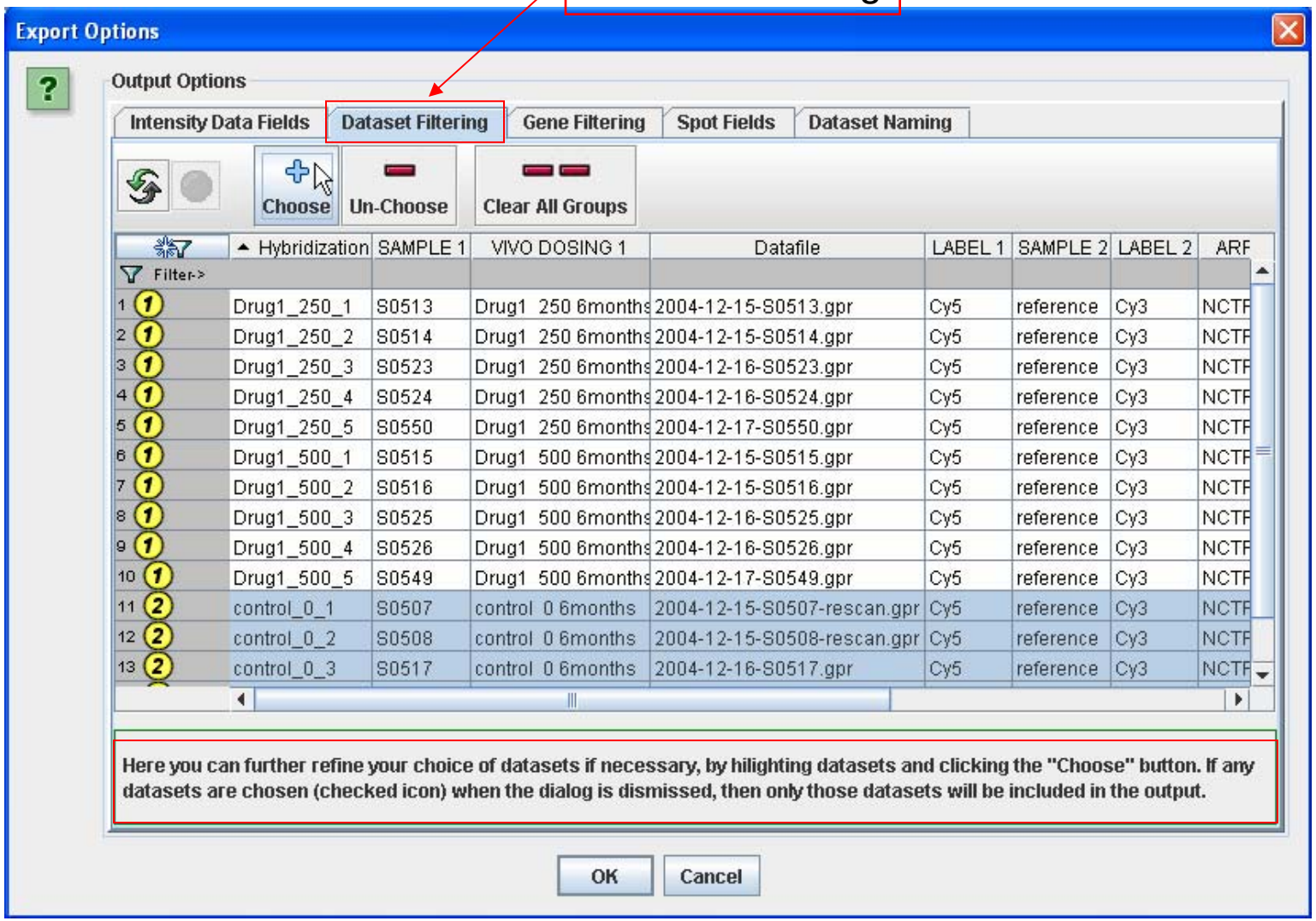
Choose the output options.

Intensity Data Fields

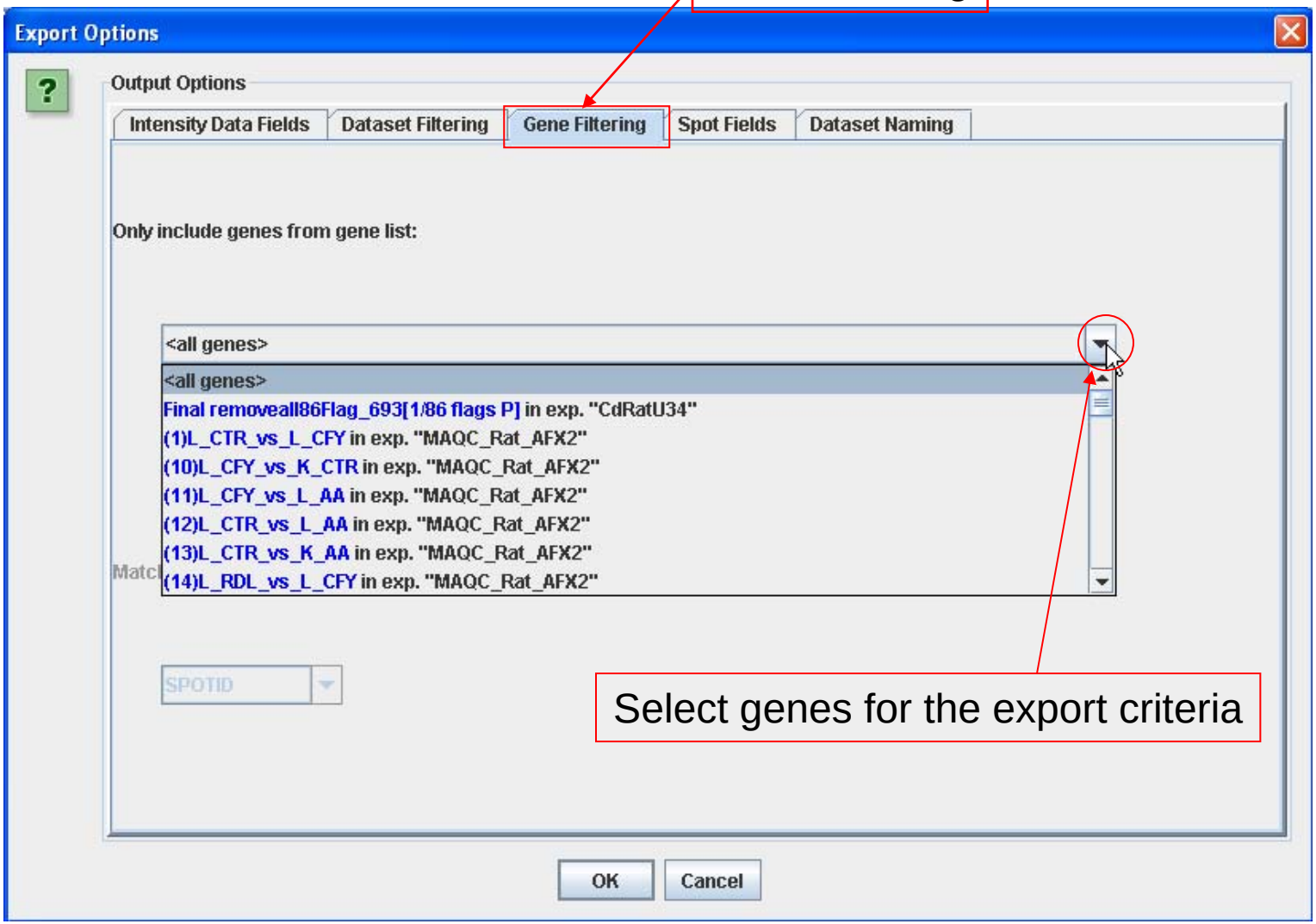


Data Export – single-platform data spreadsheet

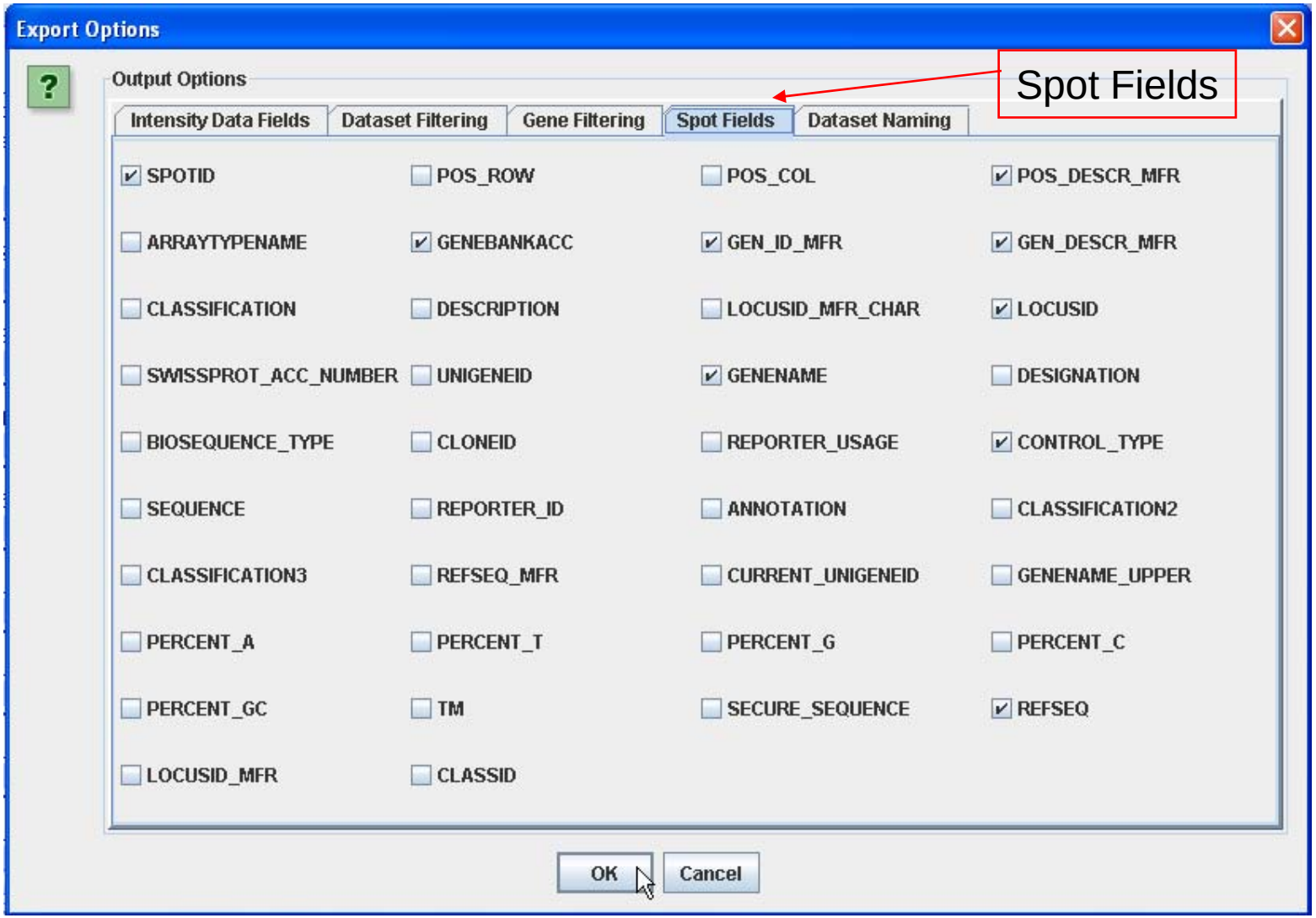
Dataset Filtering



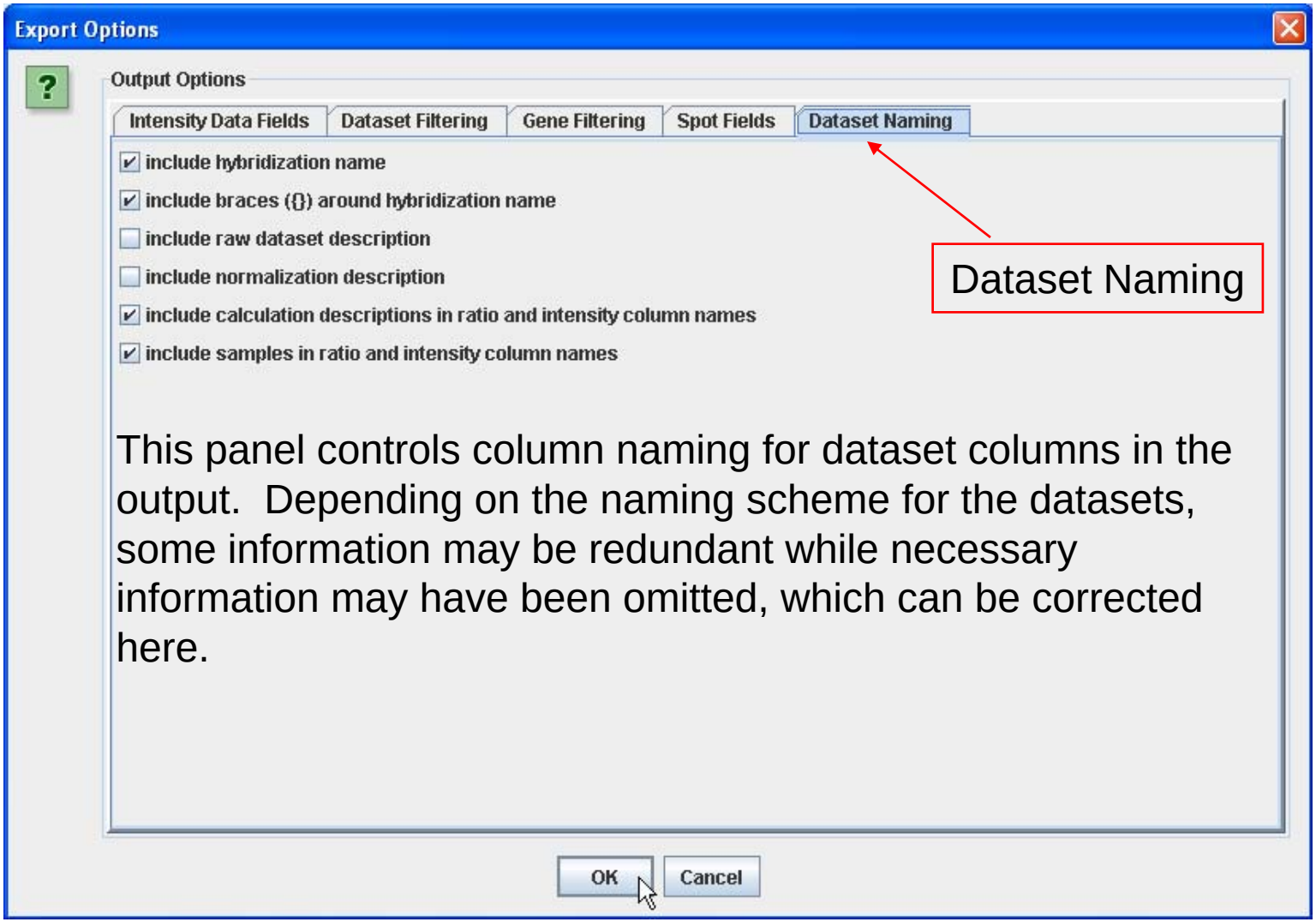
Data Export – single-platform data spreadsheet



Data Export – single-platform data spreadsheet

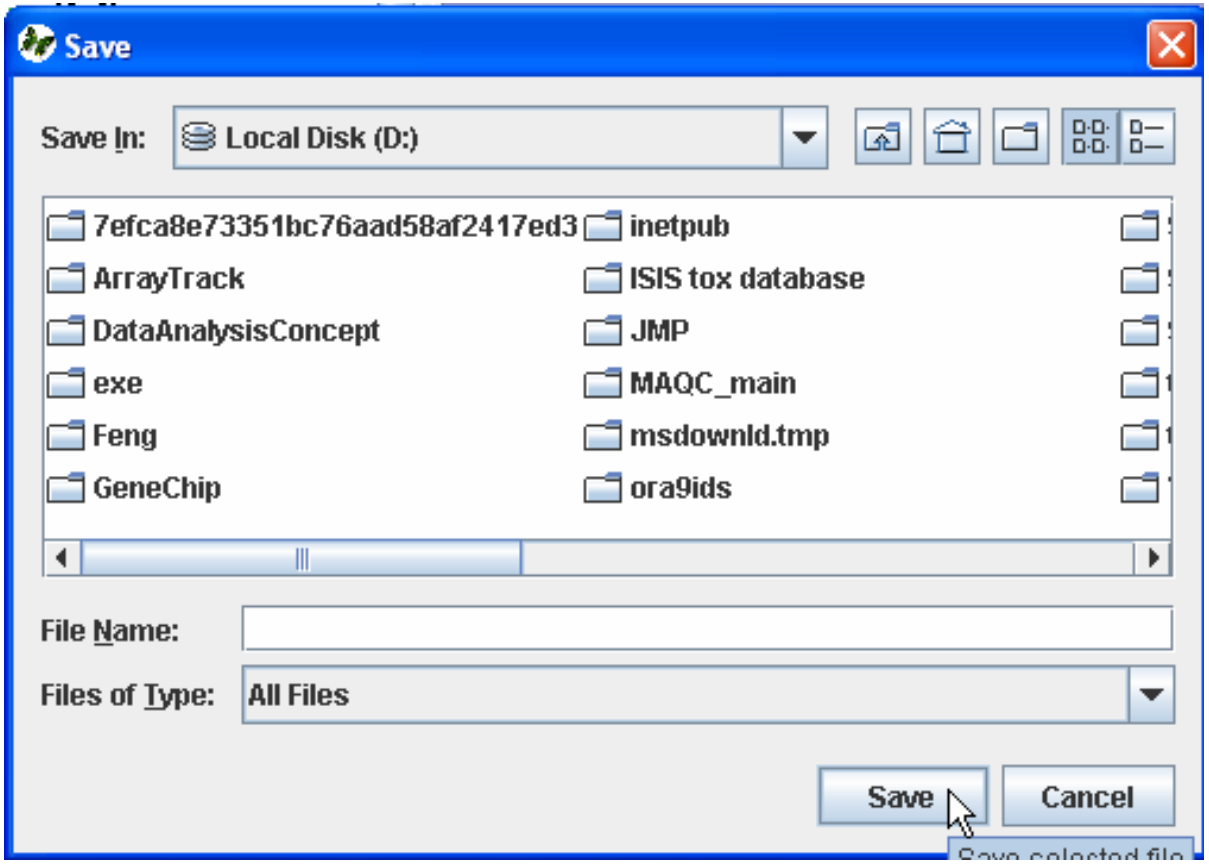


Data Export – single-platform data spreadsheet



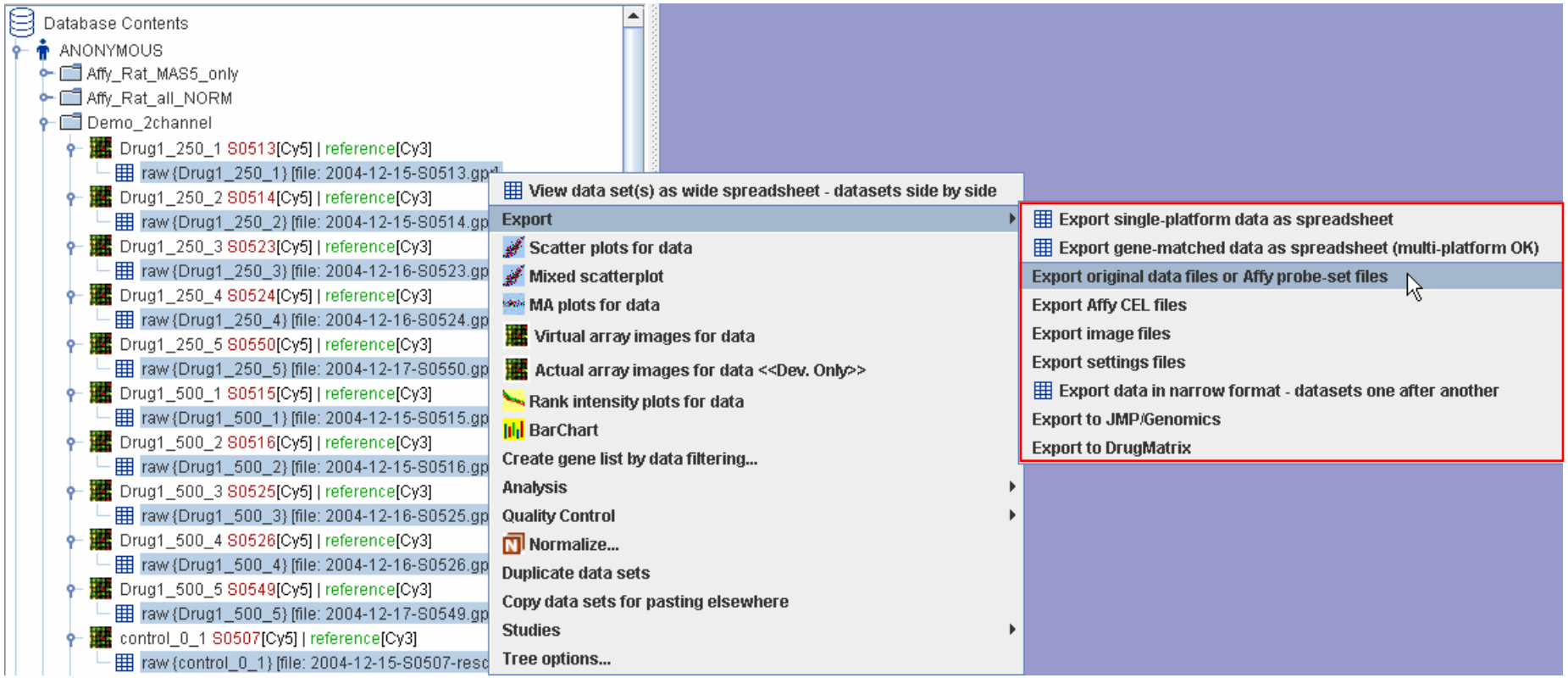
This panel controls column naming for dataset columns in the output. Depending on the naming scheme for the datasets, some information may be redundant while necessary information may have been omitted, which can be corrected here.

Data Export – single-platform data spreadsheet

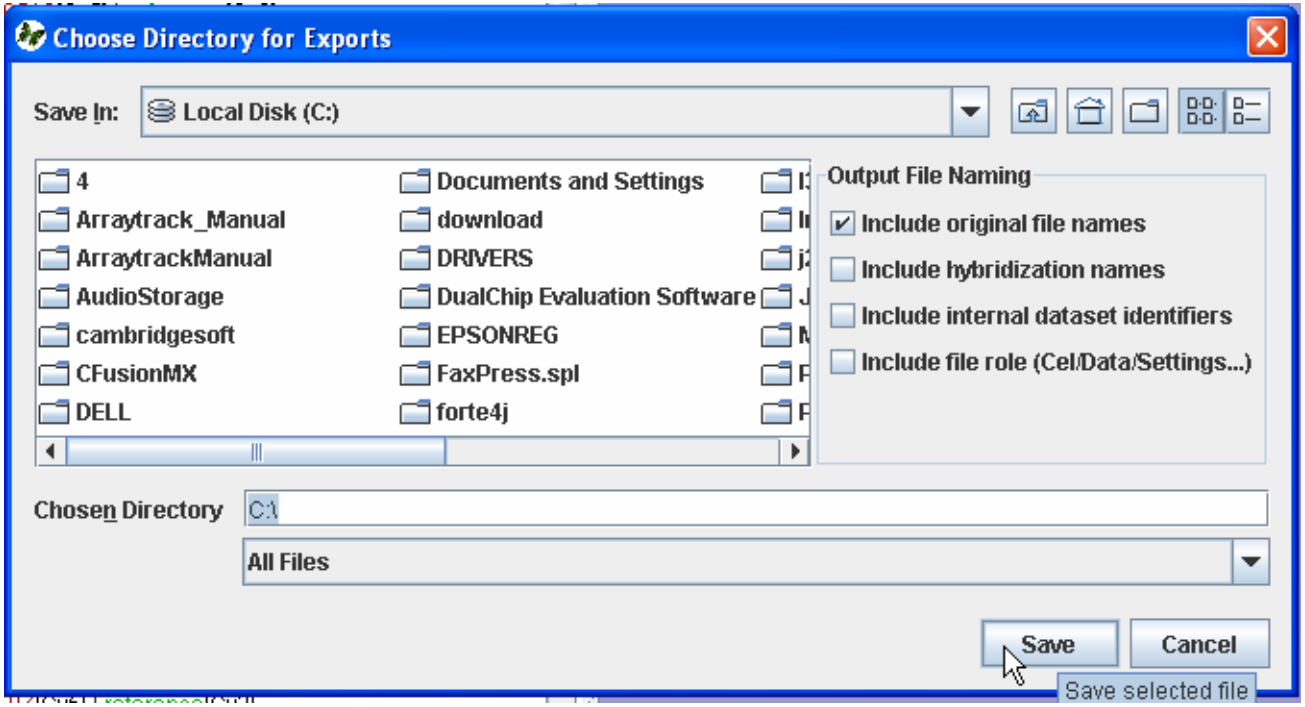


Save to your folder

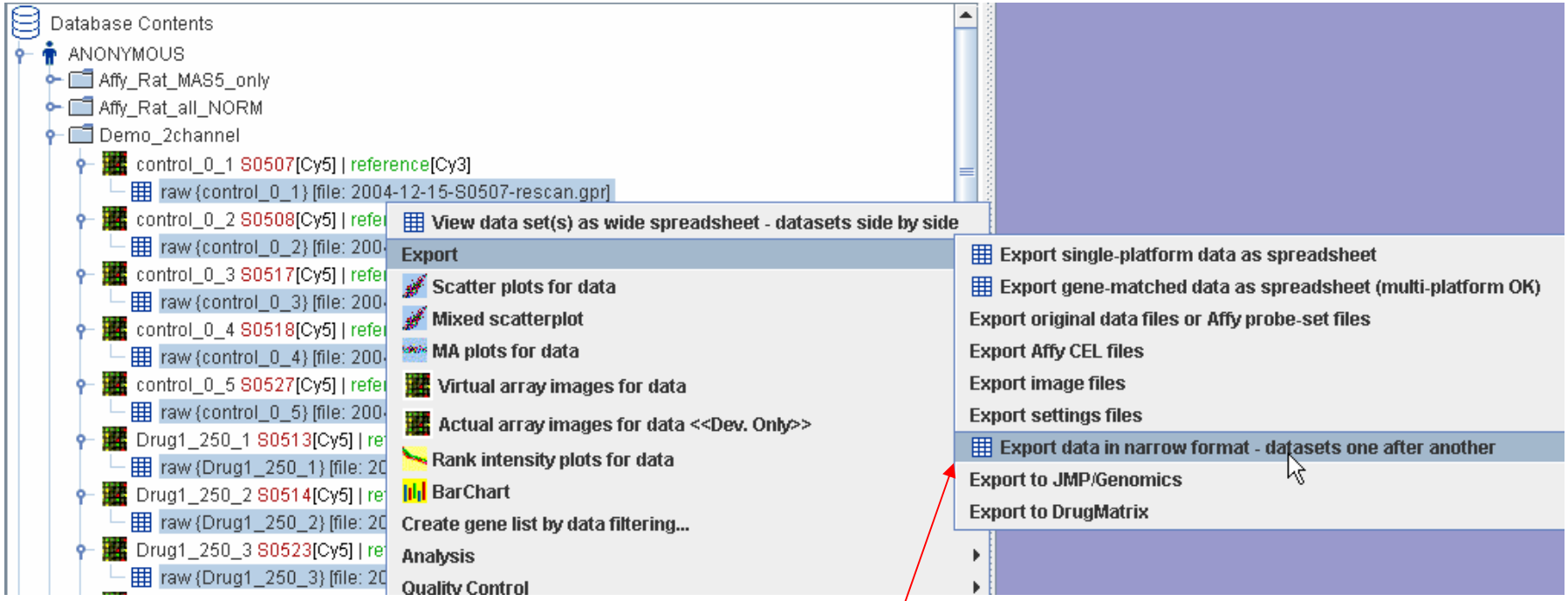
Data Export – Export original files



Data Export – original data files

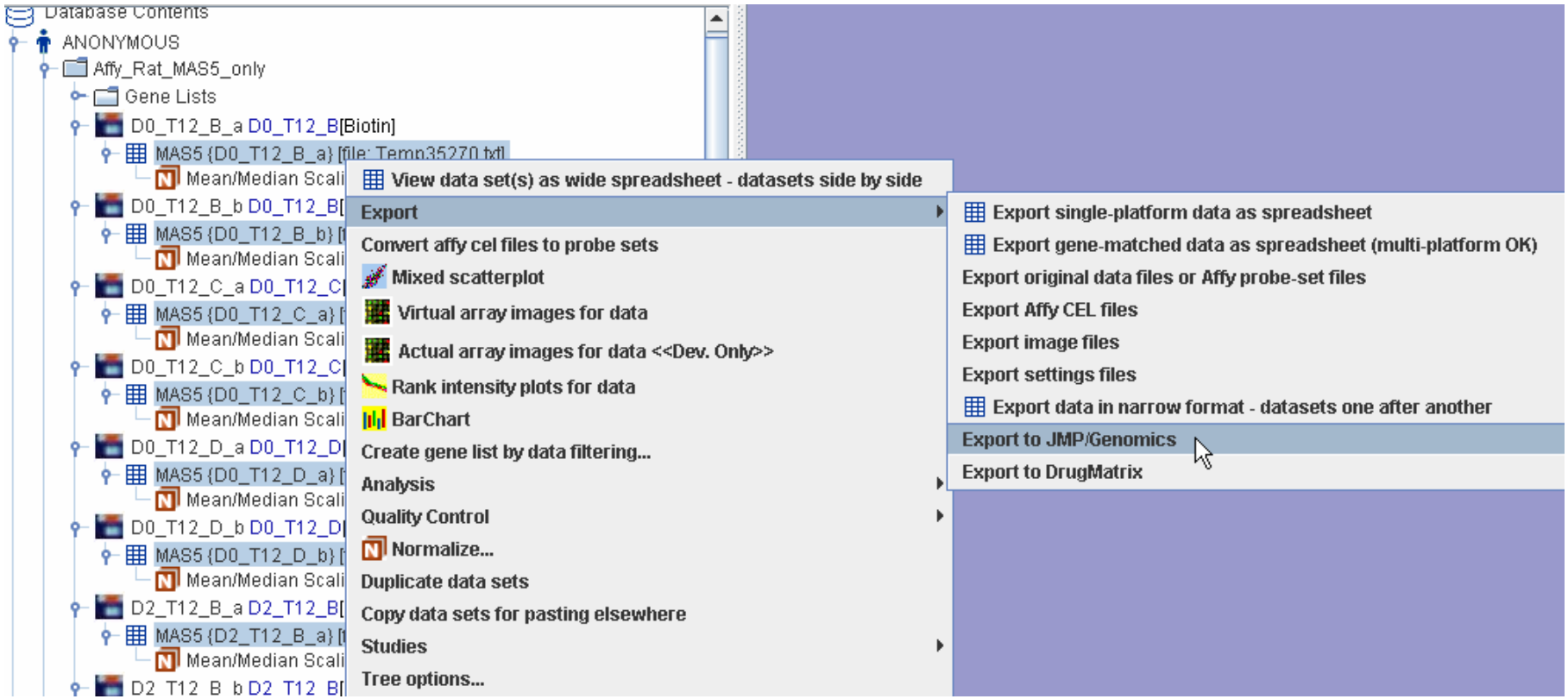


Data Export- in narrow format



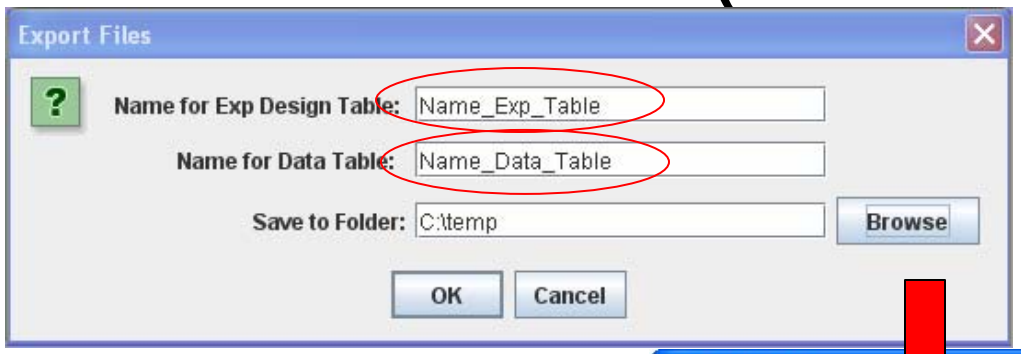
Choosing this option, you will save all the exported data in one file.

Export Data – to JMP/Genomics

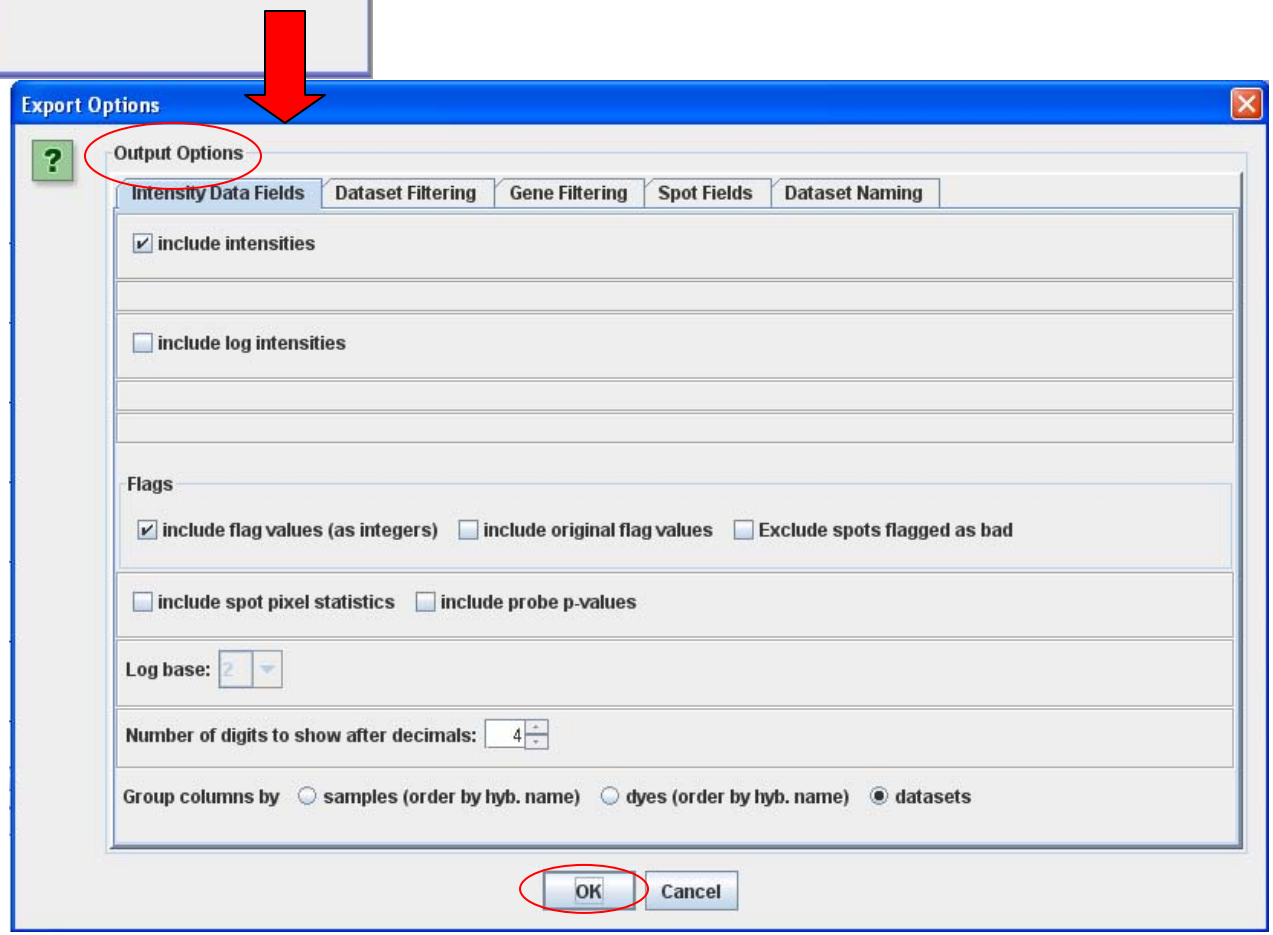


The user need JMP v6.0 or higher to open the exported files.

Export Data – to JMP/Genomics (continued)

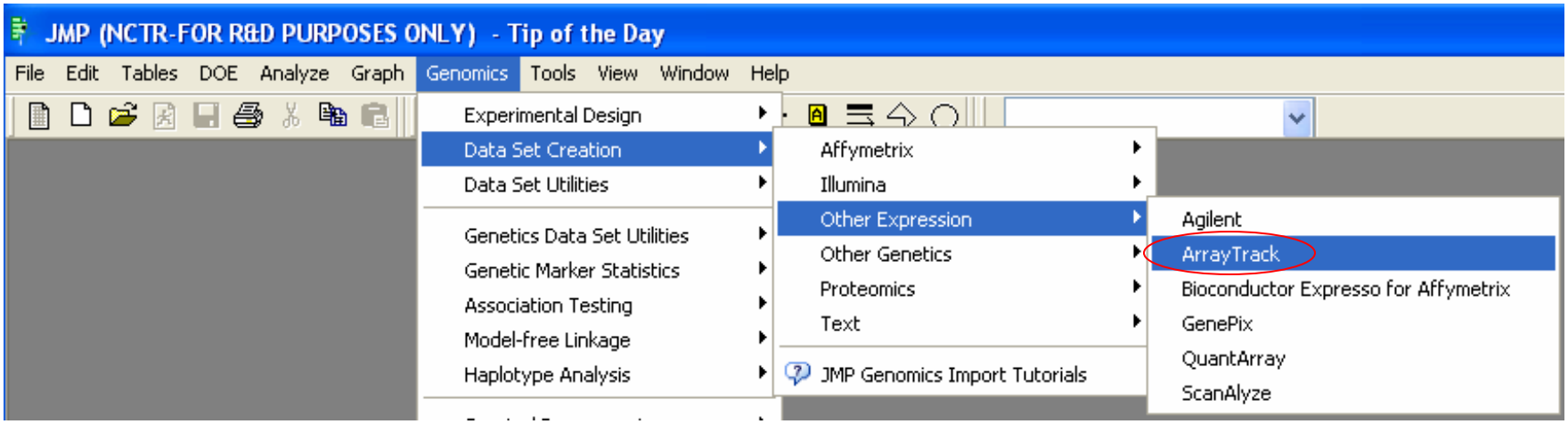


Type in the name for exp design table and data table. Then choose the location for the exported data.



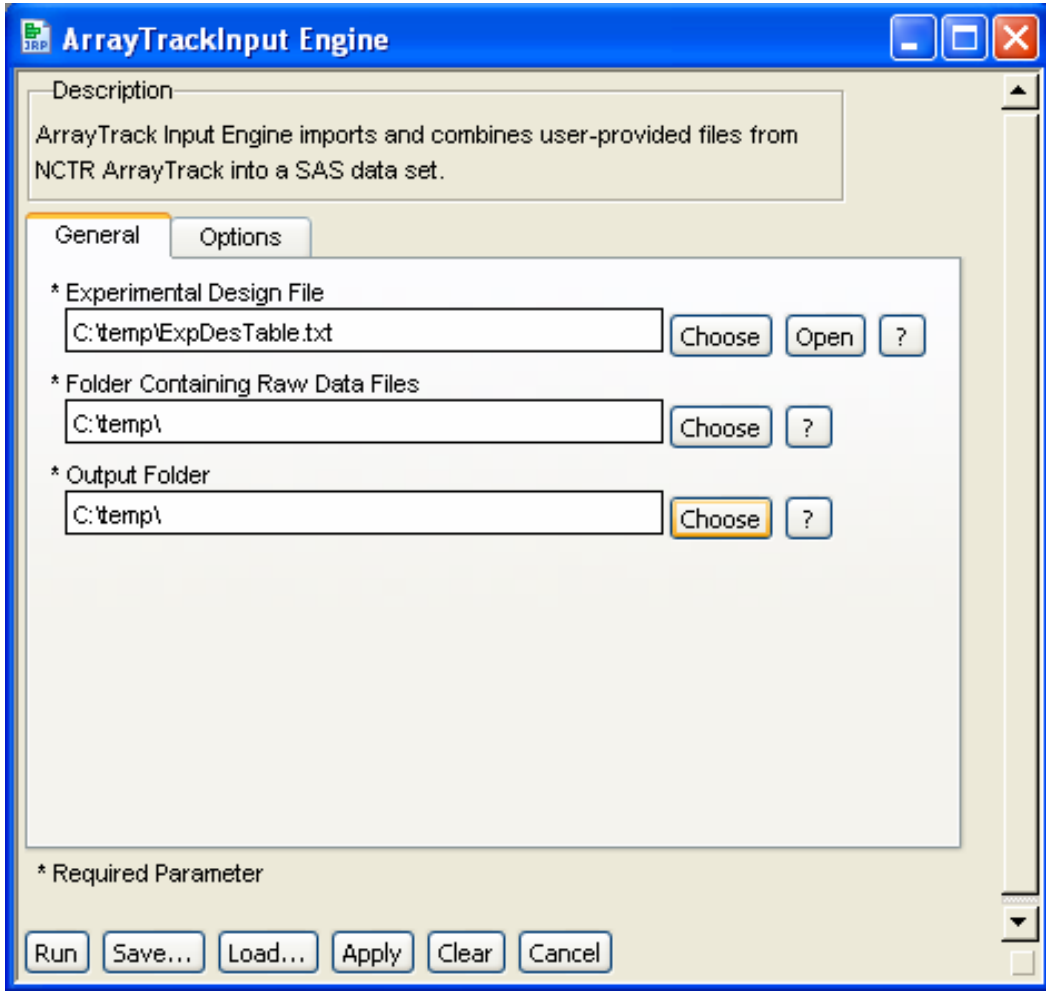
Export Data – to JMP/Genomics (continued)

Once the exported data are saved, open JMP, click Genomics pull-down menu, choose “Data set Creation -> Other Expression ->ArrayTrack”

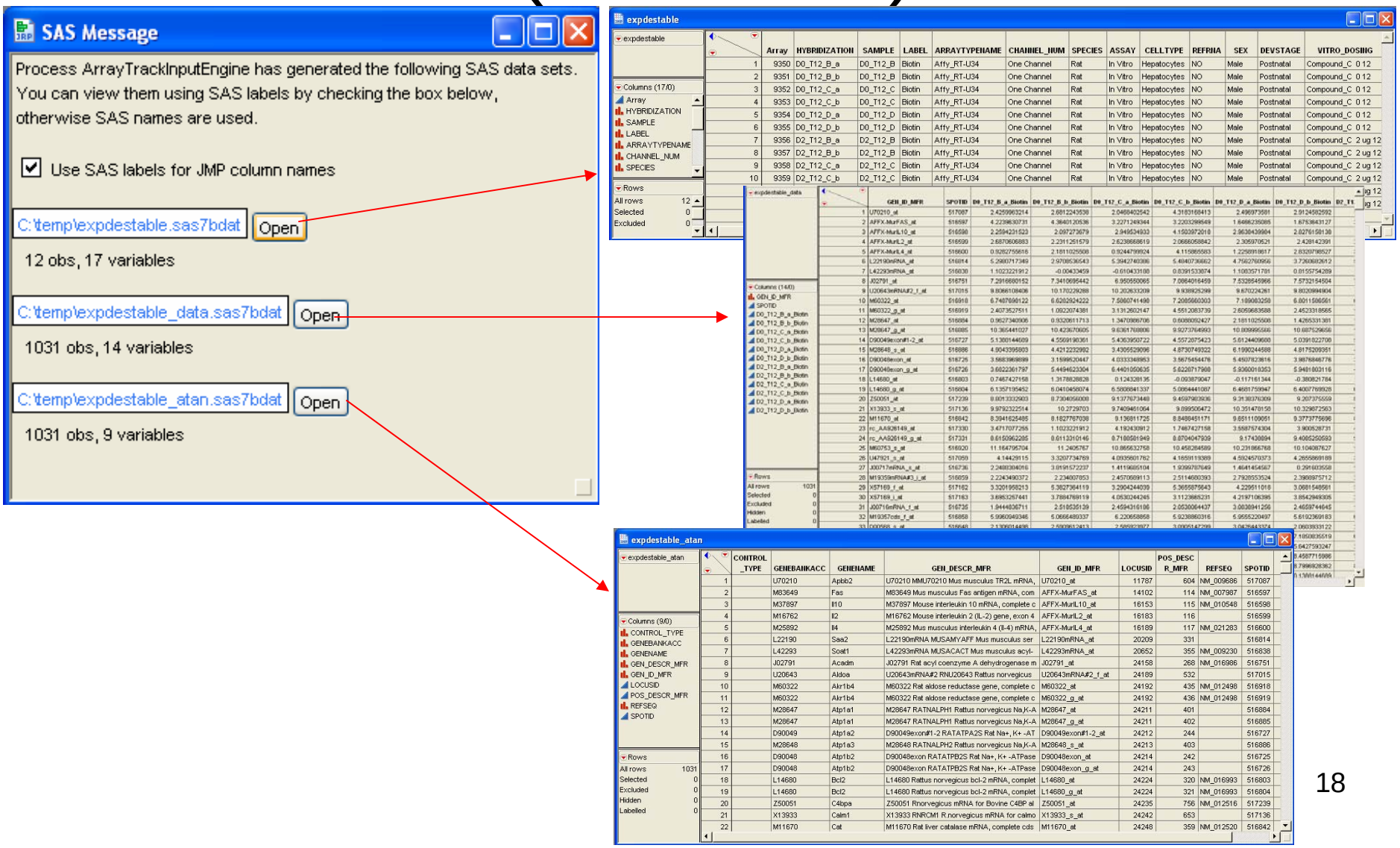


Export Data – to JMP/Genomics (continued)

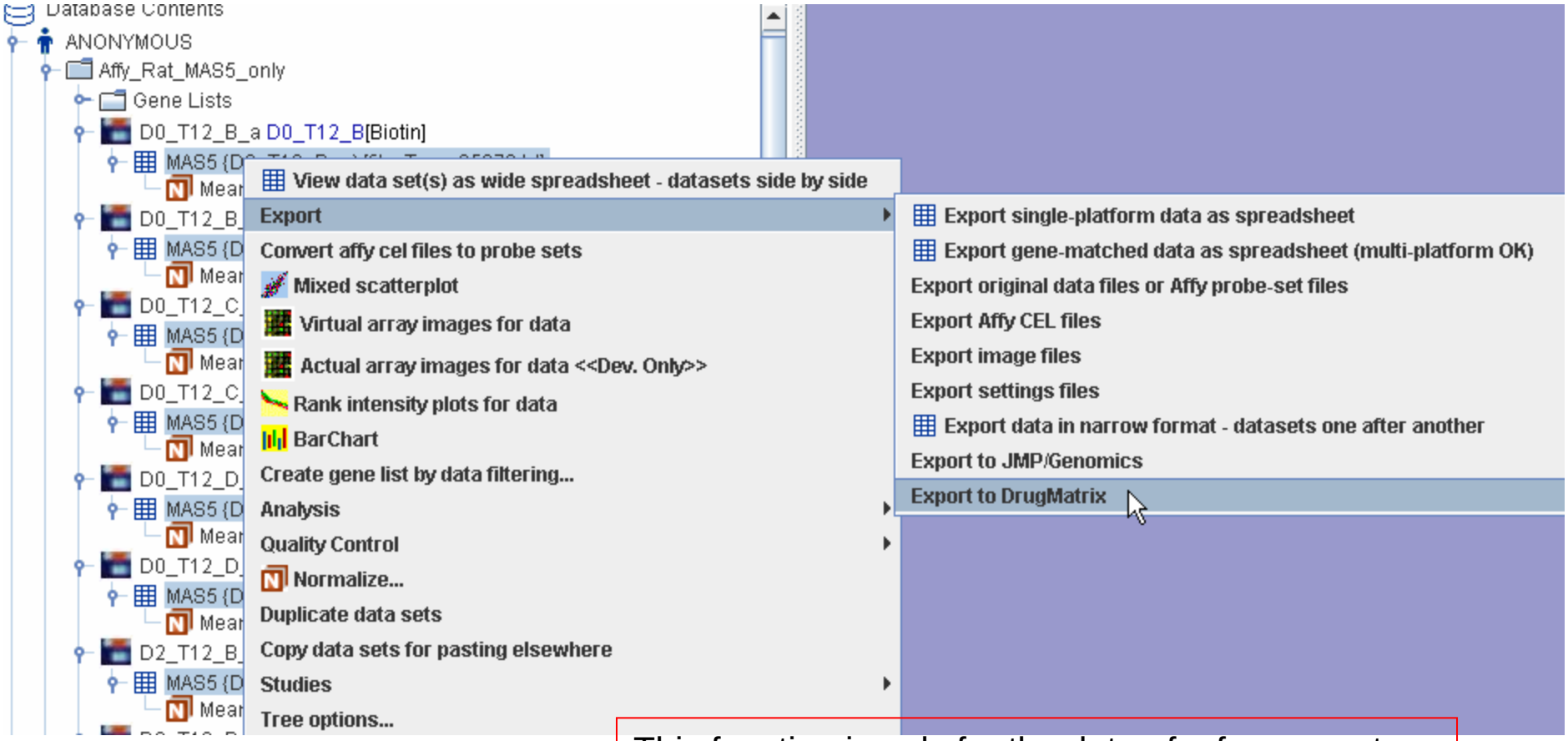
Choose the exp design file and the folder containing the data file.



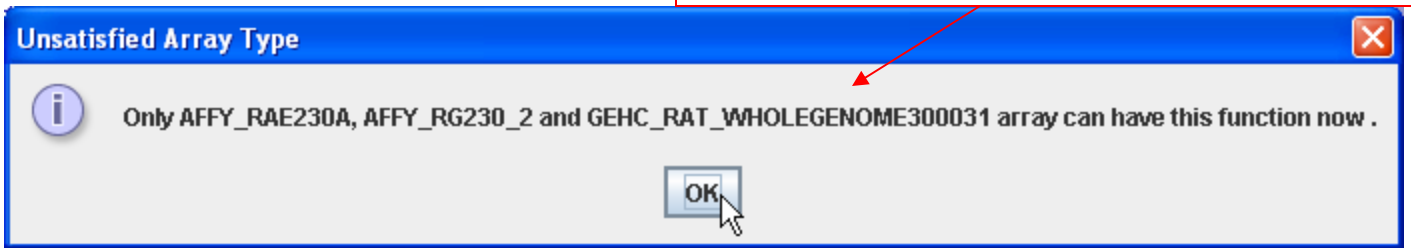
Export Data – to JMP/Genomics (continued)



Export data – to DrugMatrix



This function is only for the data of a few array type

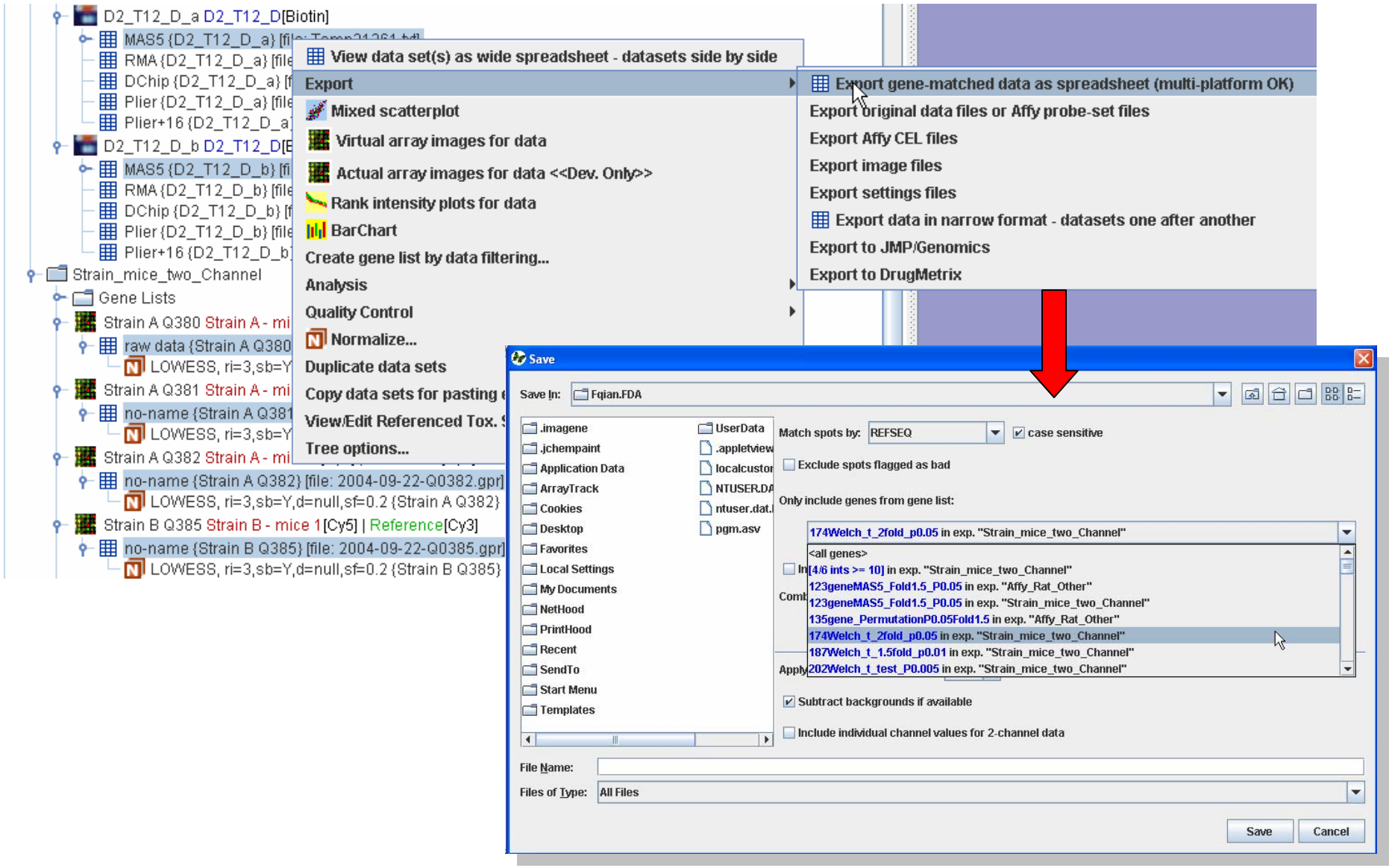


Mixed Platform Exporting

When datasets from multiple platforms are exported together, a gene identifier type must be chosen to serve as a “match field” (e.g. “REFSEQ”). Filtering based on a gene identifier may also be done.

For single-platform datasets the single-platform export should be used, as it supports more output options.

Mixed Platform Exporting



Array Track

Tutorial 10:
Other Useful Functions

Other Functions in ArrayTrack

1. ID converter
2. Join table
3. Split table
4. Get unique ID

ID Converter

ID converter is a very useful tool that converts one kind of ID to another when user searching libraries with only a few available ID types

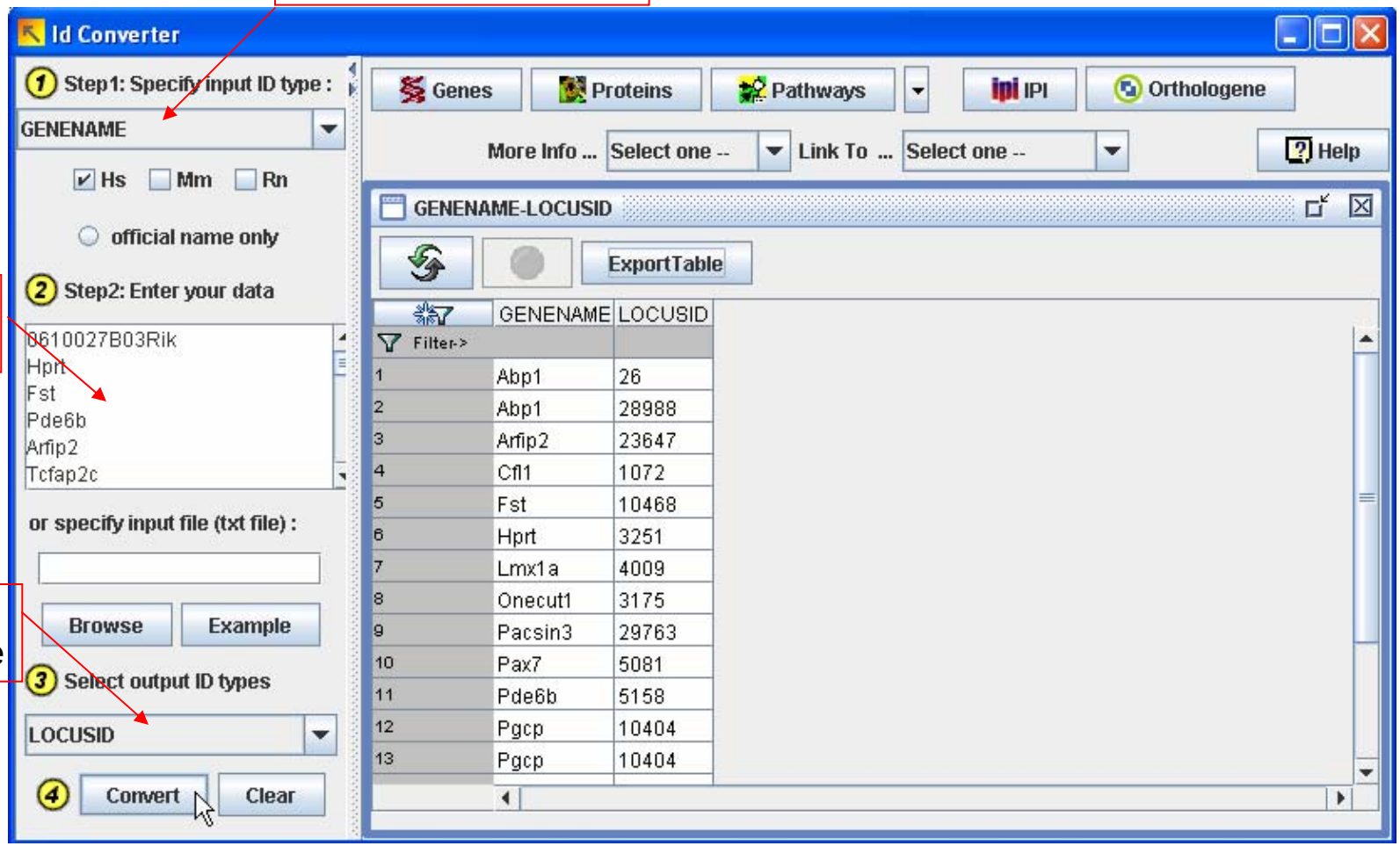
Double-click ID Converter
In the Library panel

ID Converter - continued

Select the input ID type

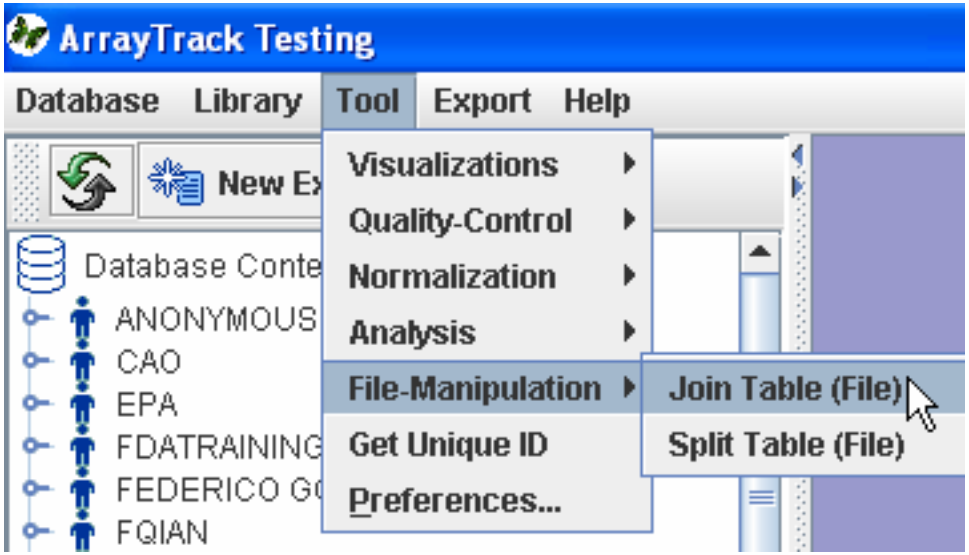
Paste the IDs in this box

Select the output ID type

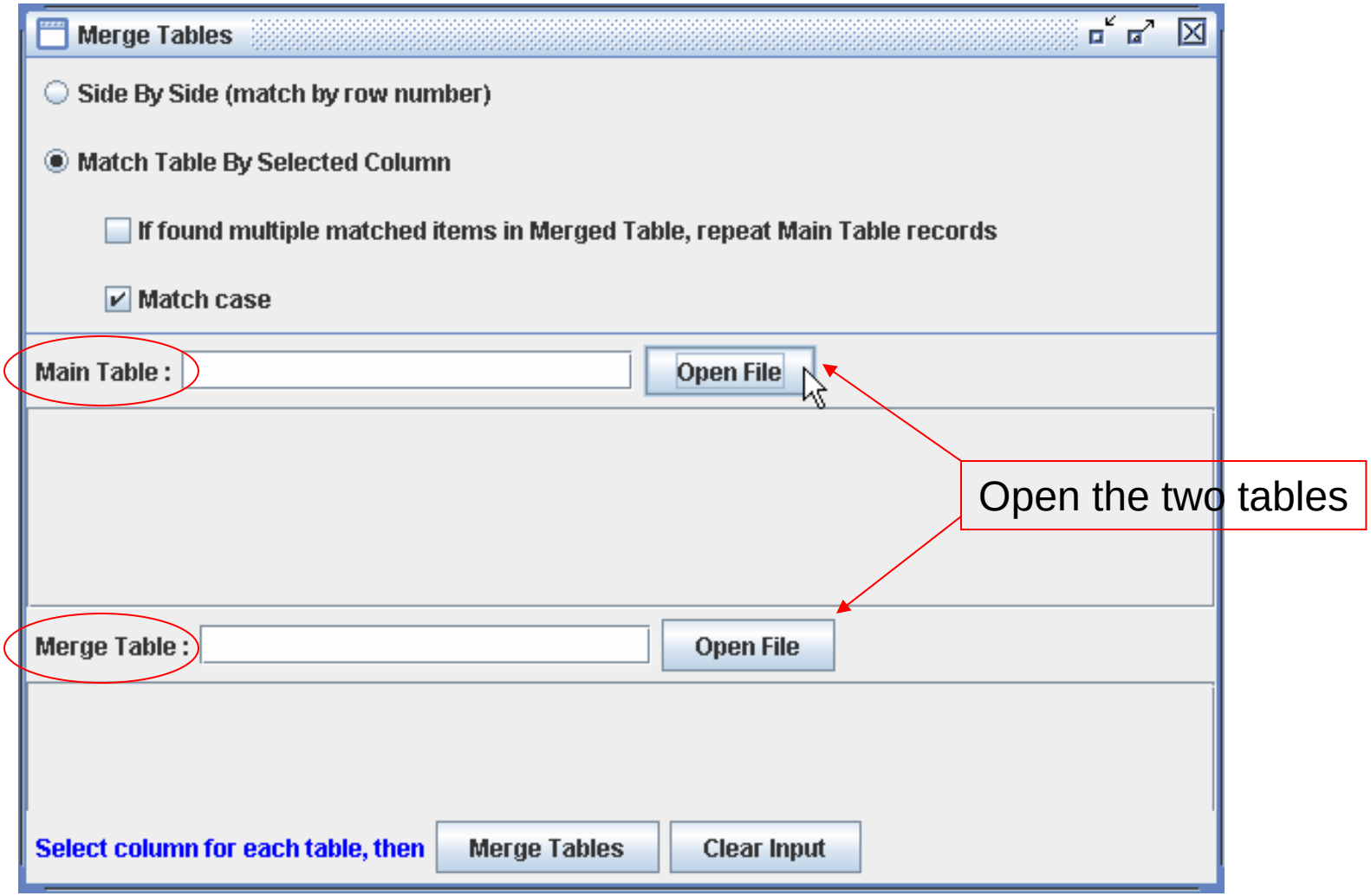


Join Table

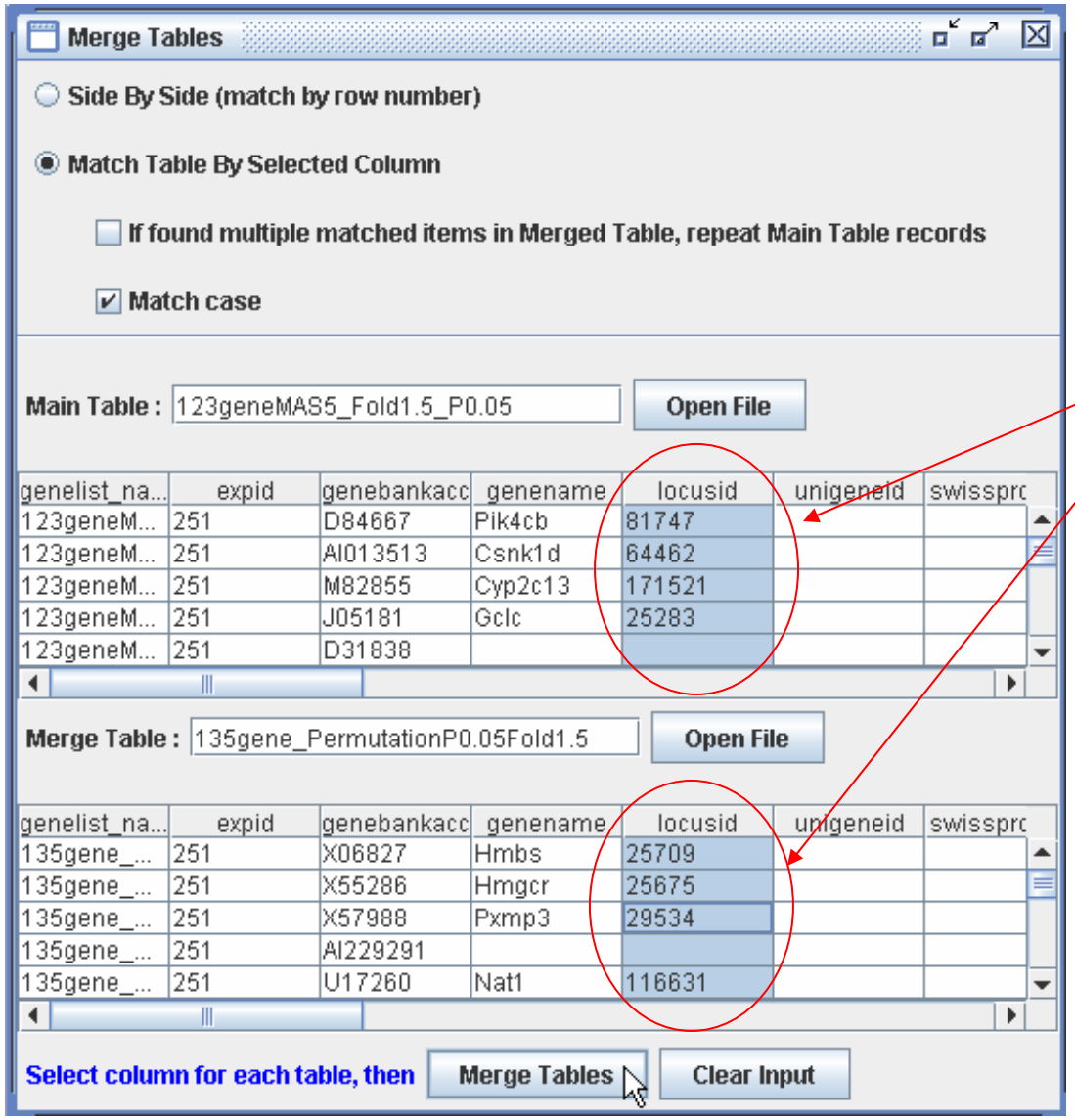
User can join two data tables according to a common ID



Join Table - continued



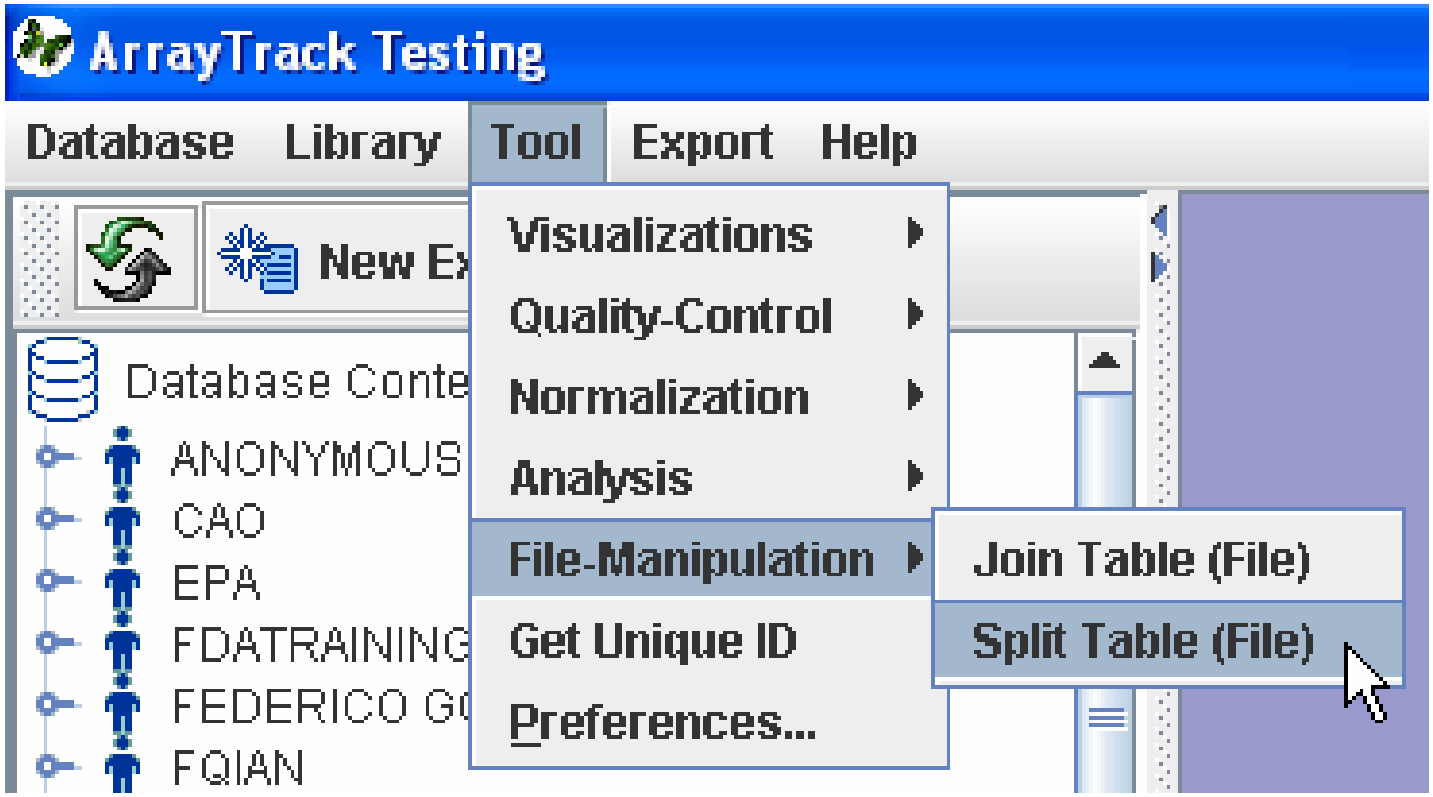
Join Table - continued



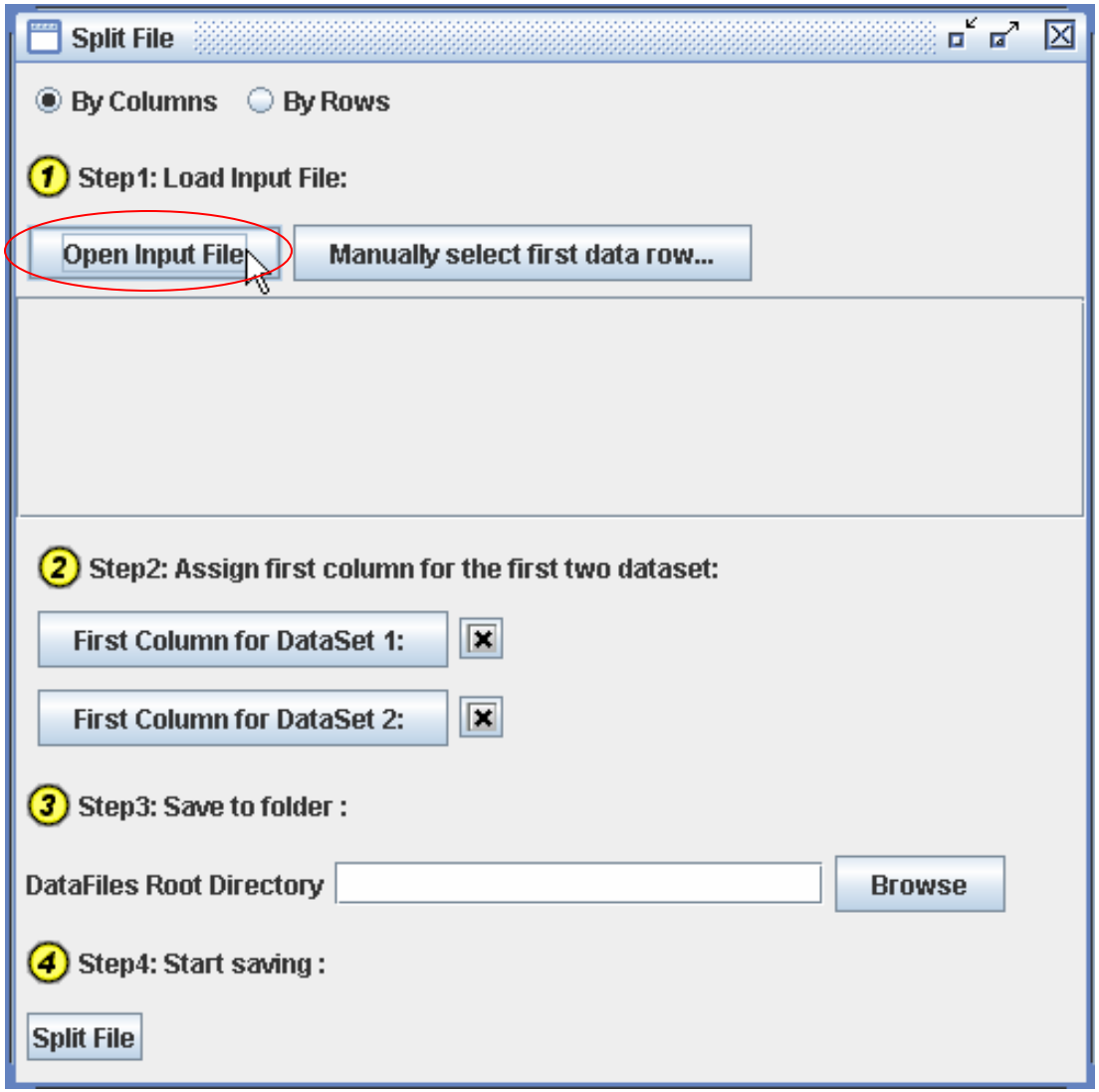
Highlight the columns for merging

Split Table

ArrayTrack provides a tool for splitting a big spreadsheet to separated files.

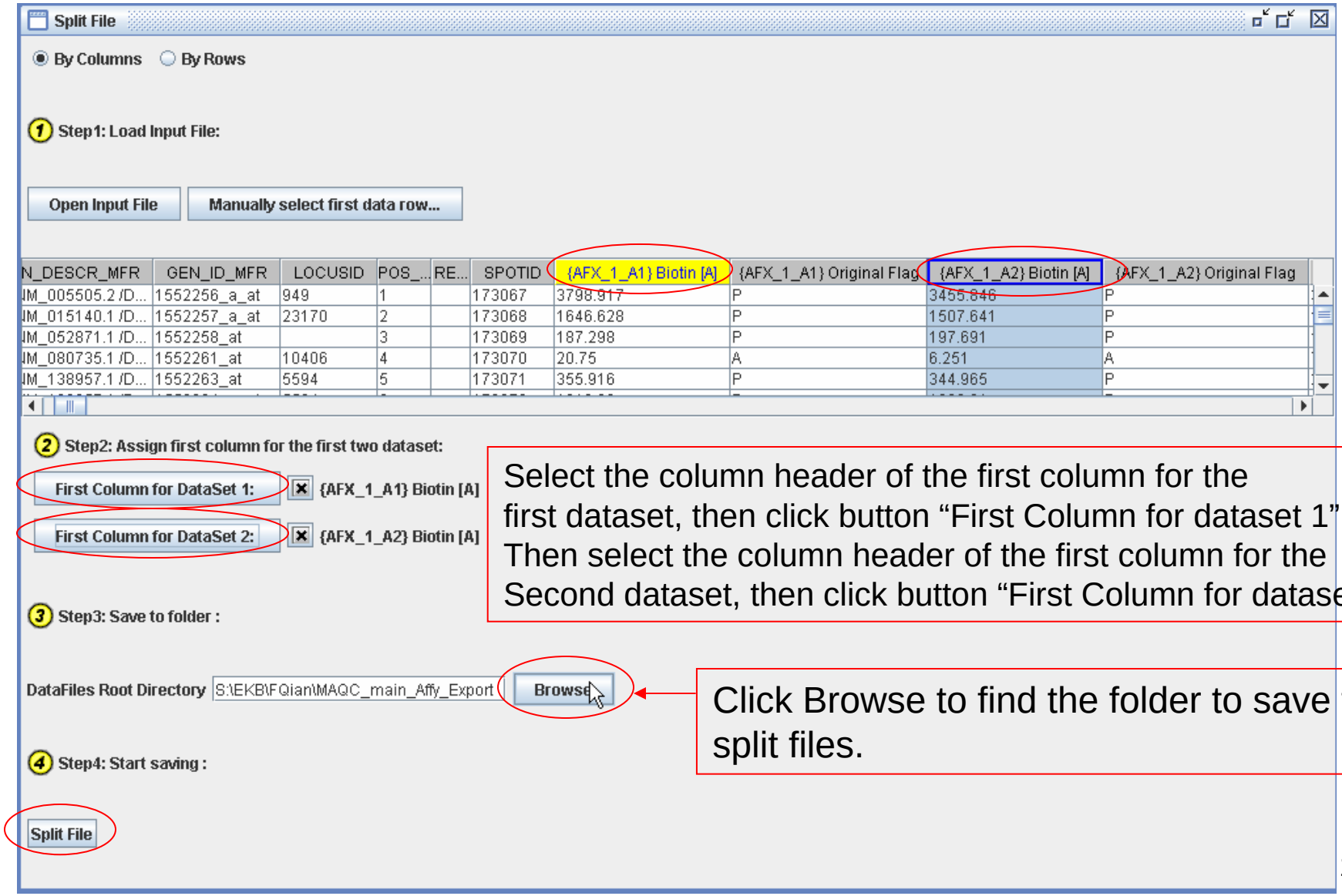


Split Table - continued



Click “Open Input File” button
To open the table file

Split Table - continued

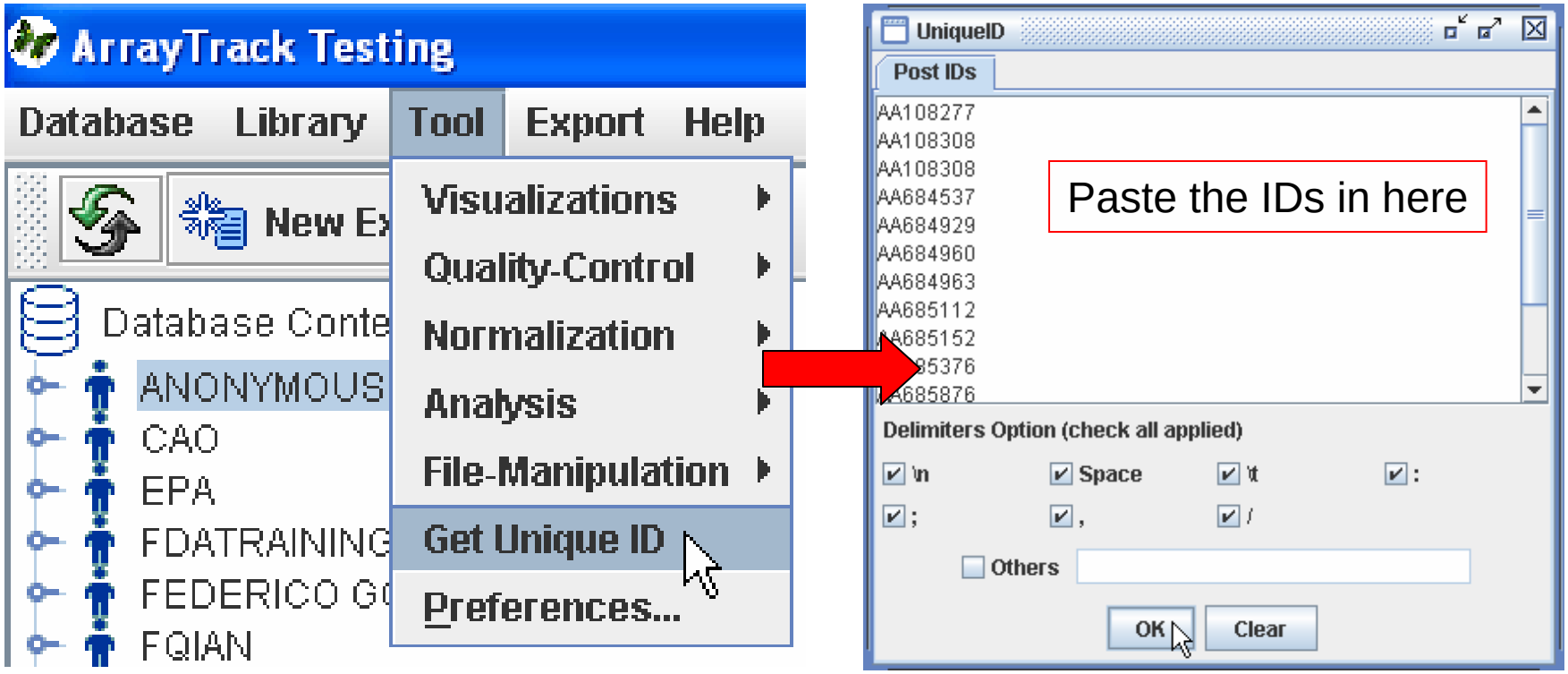


Select the column header of the first column for the first dataset, then click button “First Column for dataset 1”. Then select the column header of the first column for the Second dataset, then click button “First Column for dataset 2”.

Click Browse to find the folder to save the split files.

Get Unique ID

Using this tool, the user can find out how many unique IDs there are in a data file.



Get Unique ID - continued

The value in Column N represents the frequency of the ID in the data table

UniqueID

Post IDs | ID table

	ID	N
1	AA685112	1
2	AA685376	1
3	AA685903	1
4	AA108308	2
5	AA684963	1
6	AA684929	1
7	AA686031	1
8	AA686579	1
9	AA684960	1
10	AA685876	1
11	AA685152	1
12	AA108277	1
13	AA684537	1

Input Number:14; Returned Unique Number=13

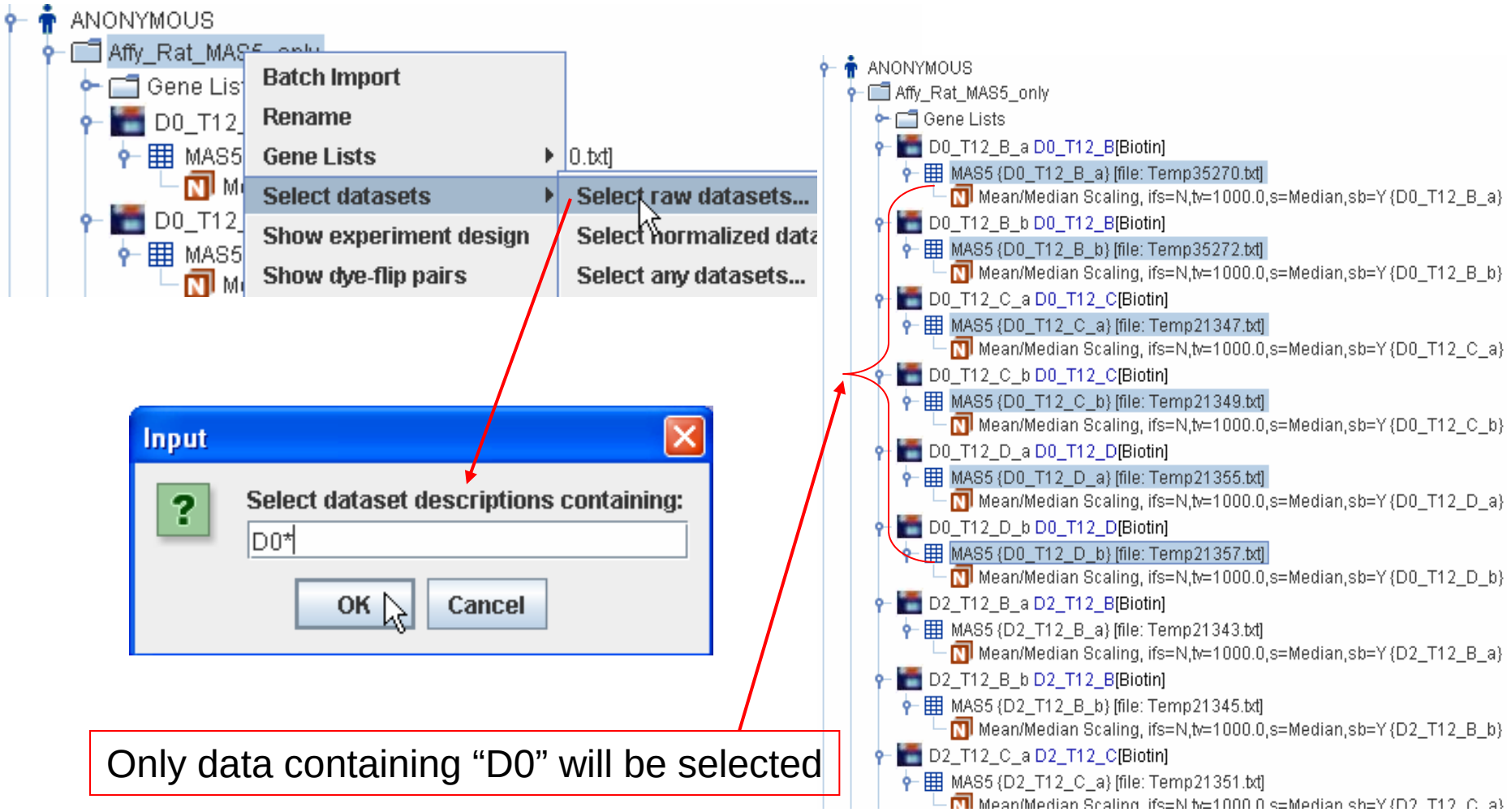
This is not an unique ID

Tutorial 11:
Basic scripting for querying (raw and normalized) data and table

Querying Data

ArrayTrack provides a way to select a group of data easily. Users can take advantage of this convenience to select a bunch of data for further data analysis.

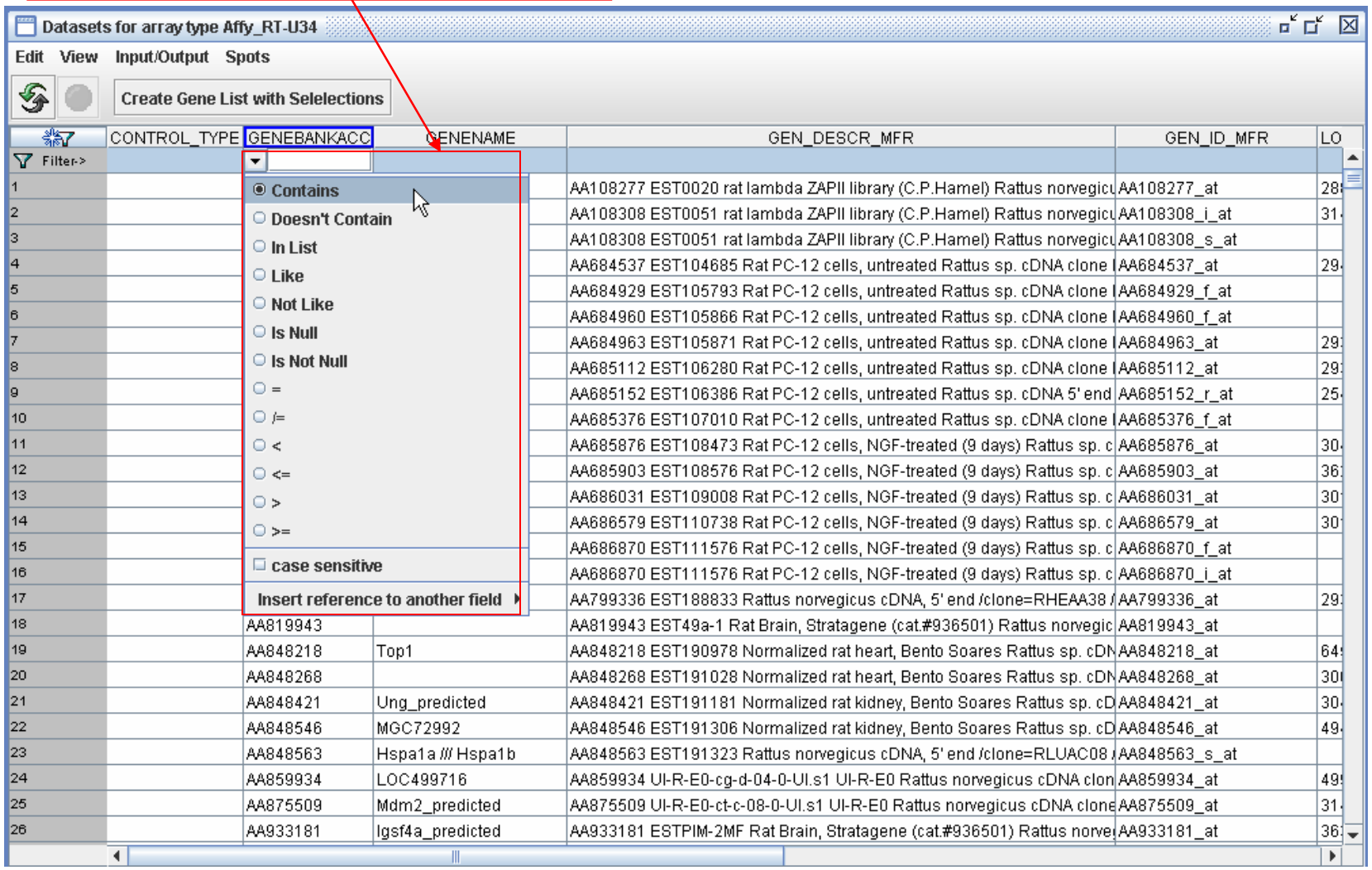
Query data from the database tree



Only data containing “D0” will be selected

Query data in tables

Various criteria for filtering data



Sort the table

Click fresh button

Click the column title to sort ascending ▲ or descending ▼

Datasets for array type Affy_RT-U34

Edit View Input/Output Spots

Create Gene List with Selelections

	CONTROL_TYPE	GENEBANKACC	GENENAME	GEN_DESCR_MFR	GEN_ID_MFR	LOCUSID
Filter->						
1		AA108277	Hsph1	AA108277 EST0020 rat lambda ZAPII library (C.F	AA108277_at	288444
2		AA108308	Mdm2_predicted	AA108308 EST0051 rat lambda ZAPII library (C.F	AA108308_i_at	314856
3		AA108308		AA108308 EST0051 rat lambda ZAPII library (C.F	AA108308_s_at	
4		AA684537	Ndufb5_predicted	AA684537 EST104685 Rat PC-12 cells, untreate	AA684537_at	294964
5		AA684929		AA684929 EST105793 Rat PC-12 cells, untreate	AA684929_f_at	
6		AA684960		AA684960 EST105866 Rat PC-12 cells, untreate	AA684960_f_at	
7		AA684963	Fkbp2_predicted	AA684963 EST105871 Rat PC-12 cells, untreate	AA684963_at	293702
8		AA685112	Ndufs8_predicted	AA685112 EST106280 Rat PC-12 cells, untreate	AA685112_at	293652
9		AA685152	Nedd8	AA685152 EST106386 Rat PC-12 cells, untreate	AA685152_r_at	25490
10		AA685376		AA685376 EST107010 Rat PC-12 cells, untreate	AA685376_f_at	
11		AA685876	Cyb5r1	AA685876 EST108473 Rat PC-12 cells, NGF-tre	AA685876_at	304805
12		AA685903	Tra1_predicted	AA685903 EST108576 Rat PC-12 cells, NGF-tre	AA685903_at	362862
13		AA686031	Ndufs1	AA686031 EST109008 Rat PC-12 cells, NGF-tre	AA686031_at	301458
14		AA686579	Sumo1	AA686579 EST110738 Rat PC-12 cells, NGF-tre	AA686579_at	301442
15		AA686870		AA686870 EST111576 Rat PC-12 cells, NGF-tre	AA686870_f_at	
16		AA686870		AA686870 EST111576 Rat PC-12 cells, NGF-tre	AA686870_i_at	
17		AA799336	Ndufab1_predicted	AA799336 EST188833 Rattus norvegicus cDNA,	AA799336_at	293453
18		AA819943		AA819943 EST49a-1 Rat Brain, Stratagene (cat.#	AA819943_at	
19		AA848218	Top1	AA848218 EST190978 Normalized rat heart, Ber	AA848218_at	64550
20		AA848268	Dnaja4	AA848268 EST191028 Normalized rat heart, Ber	AA848268_at	300721
21		AA848421	Ung	AA848421 EST191181 Normalized rat kidney, Be	AA848421_at	304577
22		AA848546	MGC72992	AA848546 EST191306 Normalized rat kidney, Be	AA848546_at	494345
23		AA848563	Hspa1a /// Hspa1b_mapped	AA848563 EST191323 Rattus norvegicus cDNA,	AA848563_s_at	
24		AA859934	LOC499716	AA859934 UI-R-E0-cg-d-04-0-UI.s1 UI-R-E0 Ratt	AA859934_at	499716
25		AA875509	Mdm2_predicted	AA875509 UI-R-E0-ct-c-08-0-UI.s1 UI-R-E0 Rattu	AA875509_at	314856
26		AA933181	Igsf4a	AA933181 ESTPIM-2MF Rat Brain, Stratagene (c	AA933181_at	363058
27		AA956437	RGD1559939_predicted	AA956437 UI-R-E1-fj-h-02-0-UI.s1 UI-R-E1 Rattu	AA956437_at	500006

Tutorial 12: Using SAM through ArrayTrack

SAM-Test

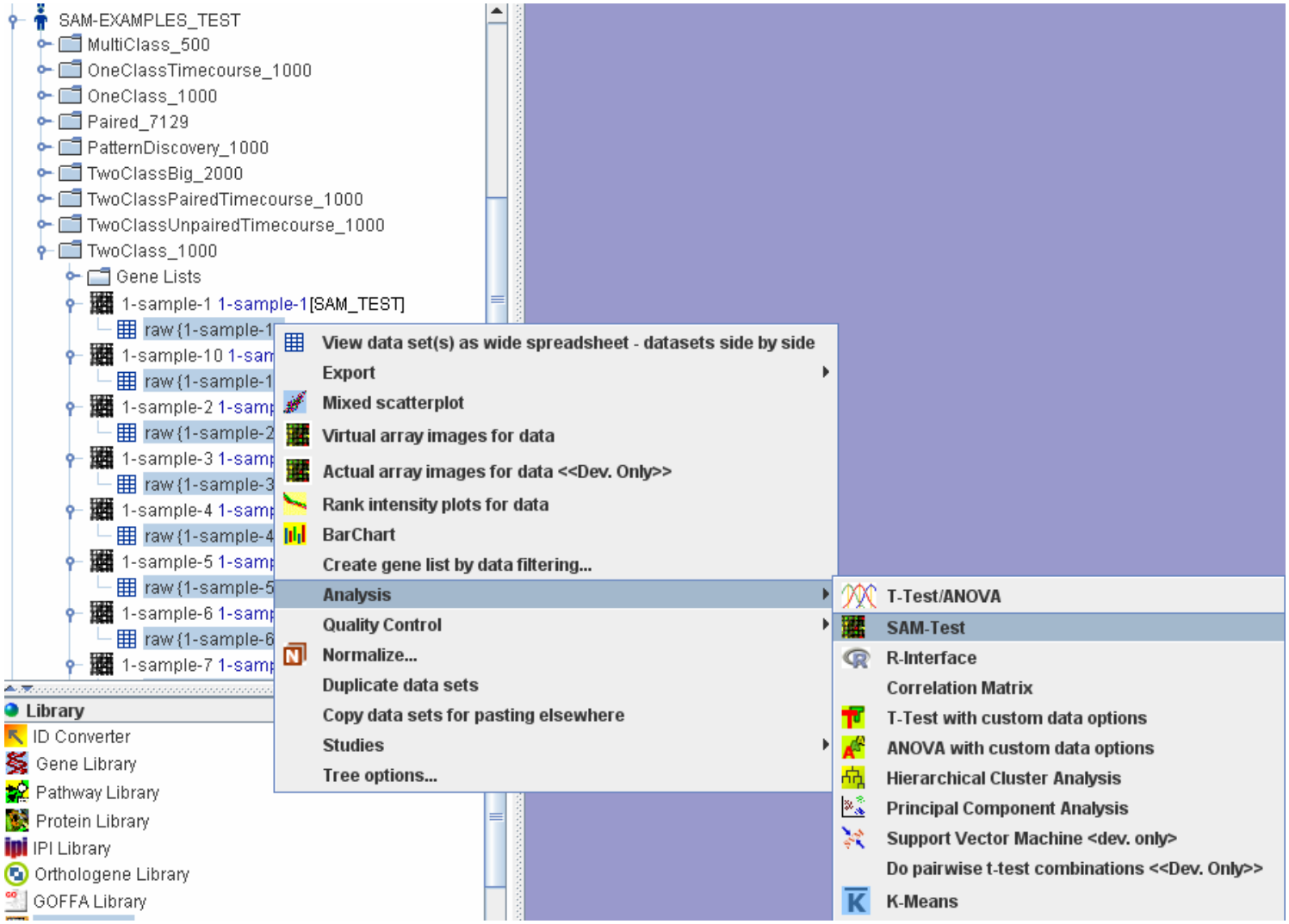
SAM (Significance Analysis of Microarrays) is an analysis tool for Identifying statistical significant genes in a set of microarray experiments.
http://www-stat.stanford.edu/~tibs/SAM/

Before using SAM-test tool in ArrayTrack, users should read SAM manual
http://www-stat.stanford.edu/~tibs/SAM/sam.pdf

The SAM tool in ArrayTrack includes following analysis types:

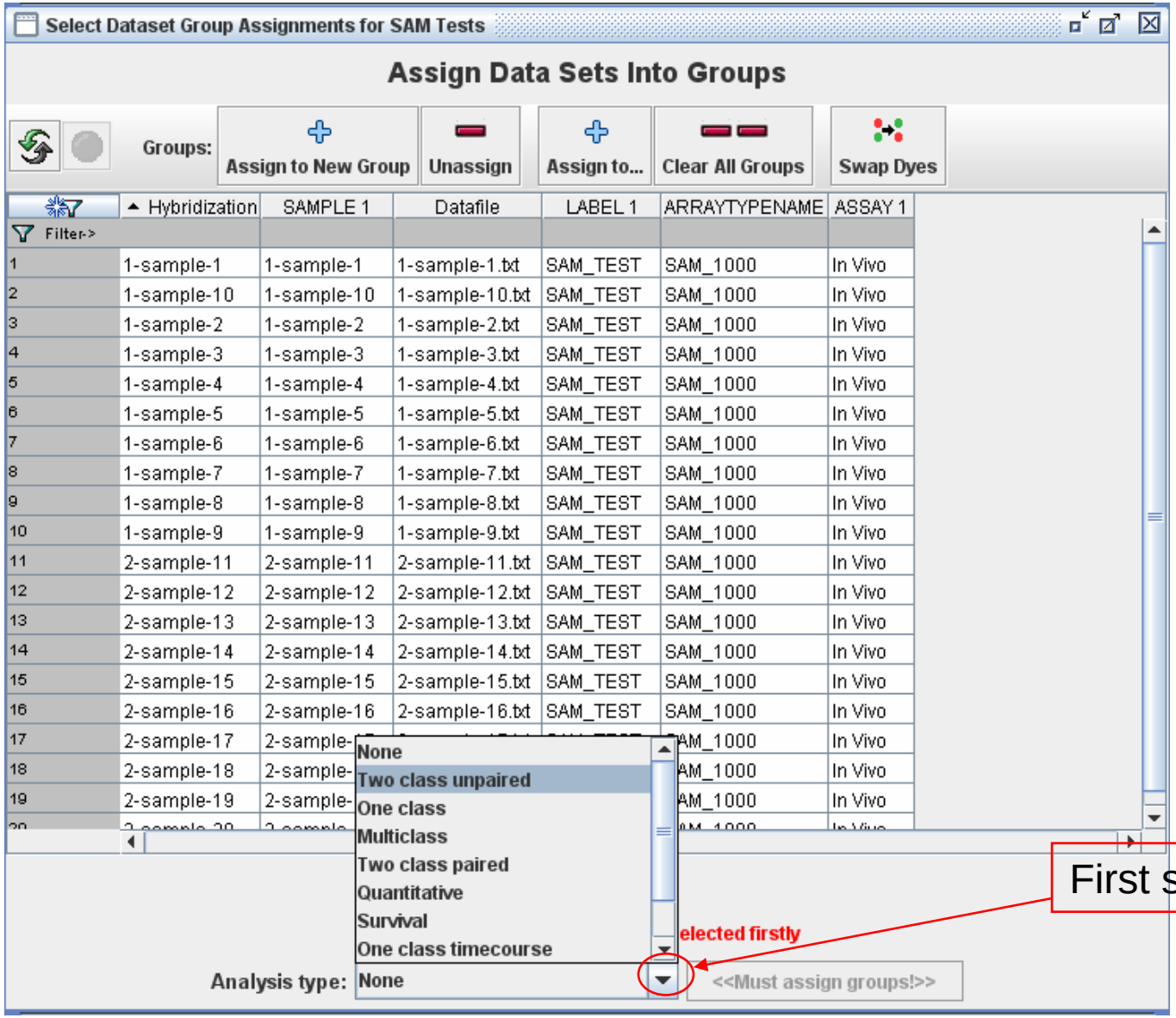
- Two class paired
- Two class unpaired
- One class
- Multi class
- Survival
- One class timecourse
- Two class unpaired timecourse
- Two class paired timecourse

SAM-Test: two class unpaired



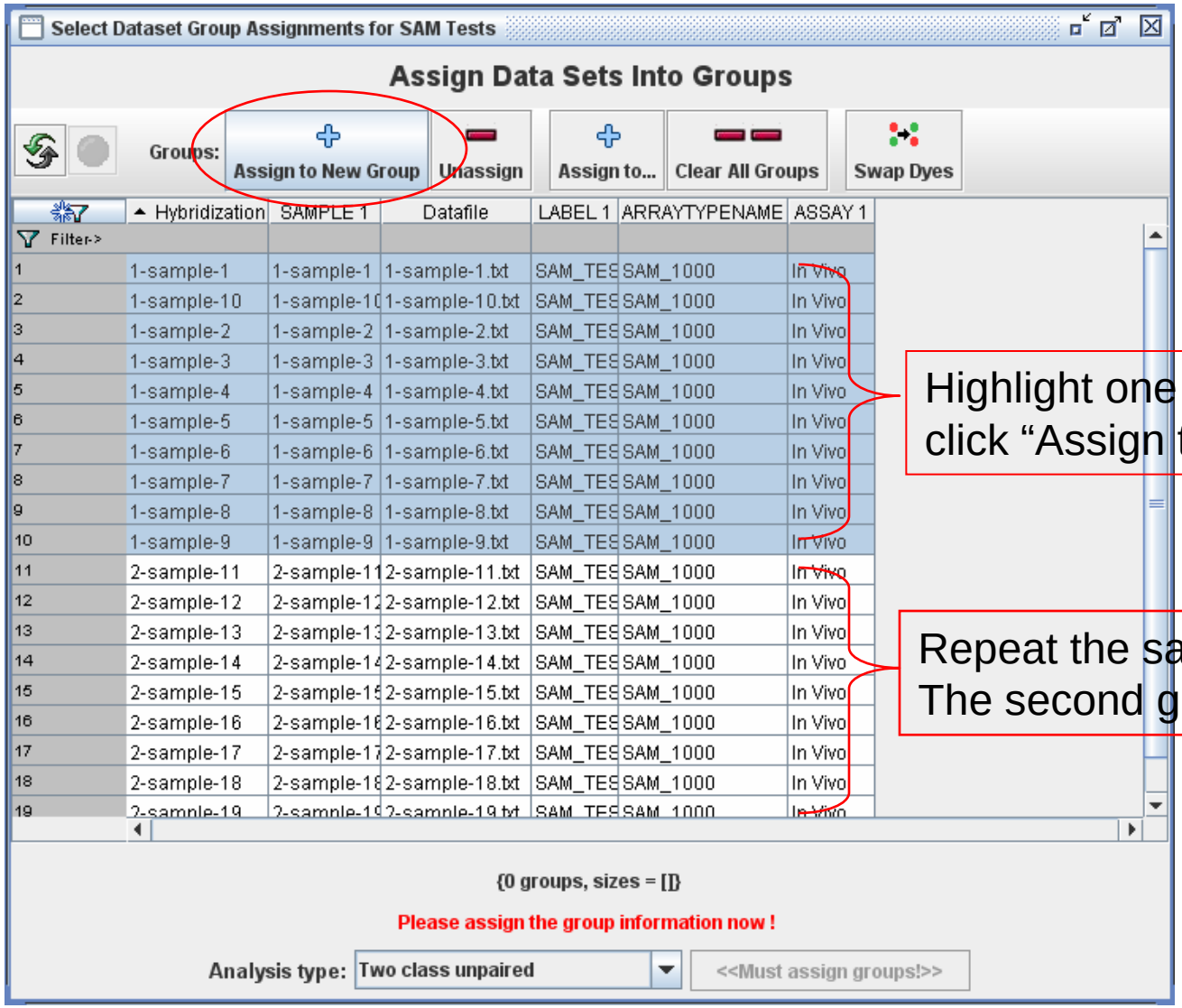
Select all the data, right-click, choose “Analysis -> SAM-Test”

SAM-Test: two class unpaired



First select analysis type

SAM-Test: two class unpaired

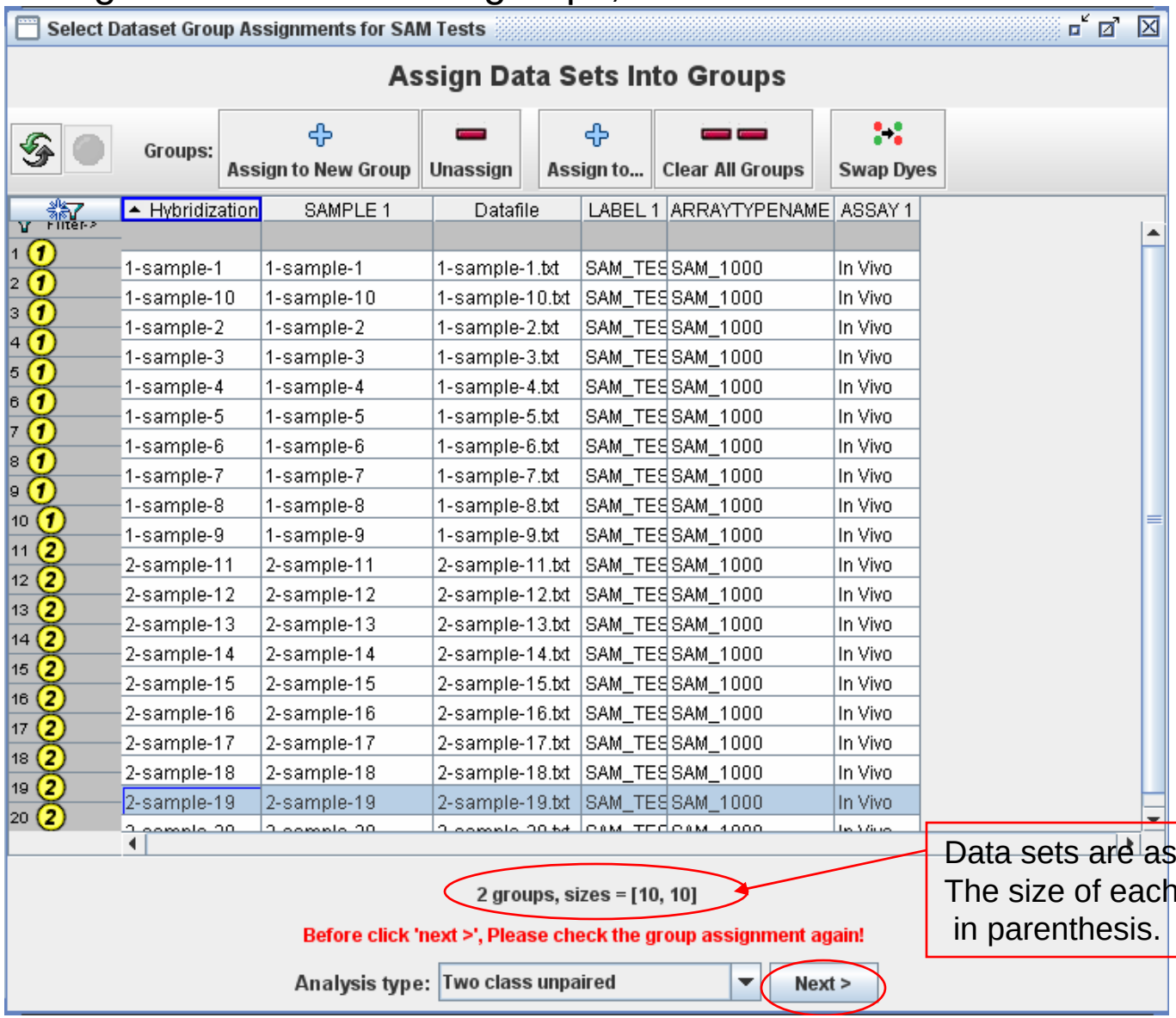


Highlight one group of data, then click “Assign to New Group” button.

Repeat the same step for assigning The second group.

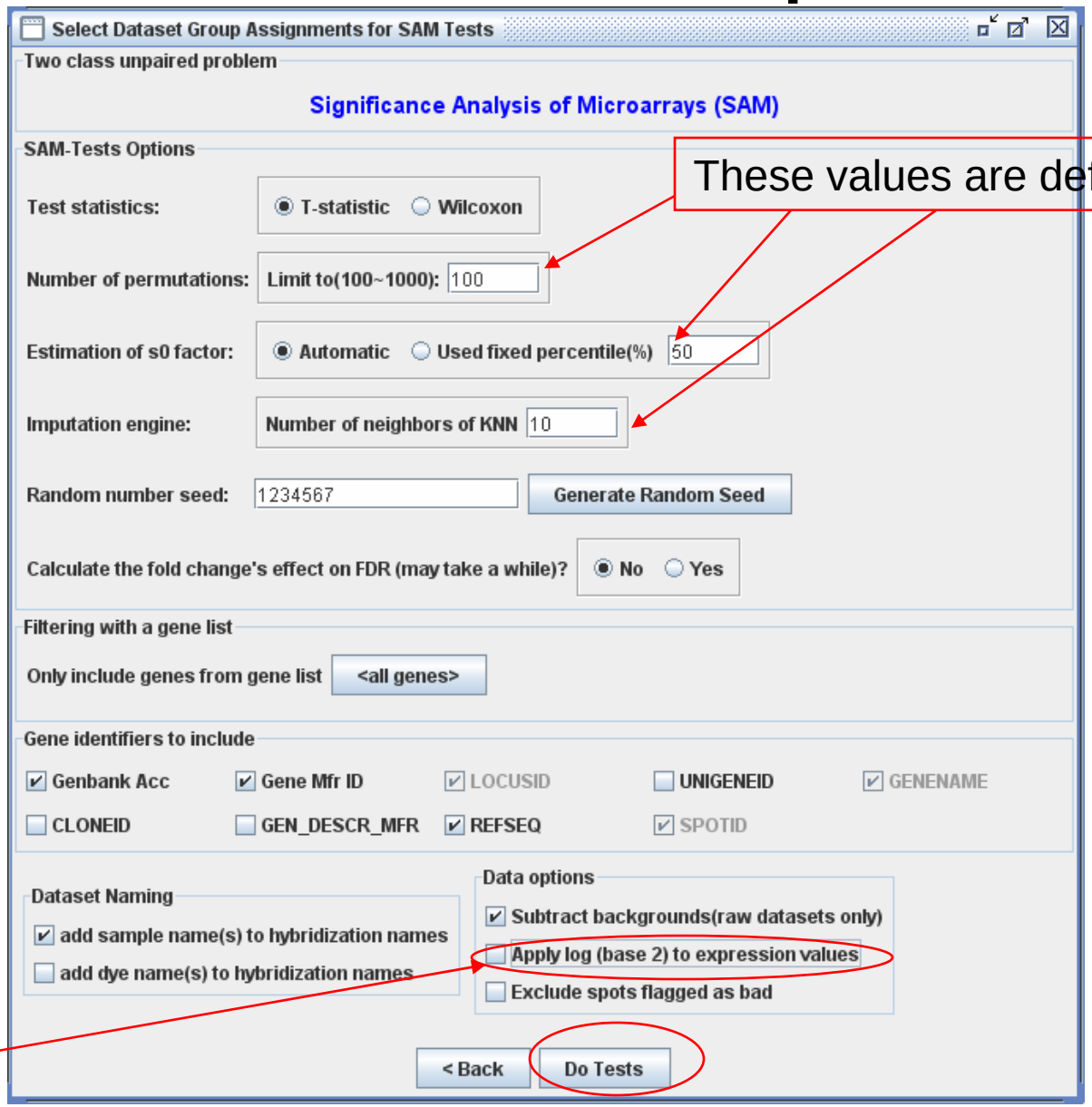
SAM-Test: two class unpaired

After assign data sets into two groups, click “Next” button.



Data sets are assigned to 2 groups. The size of each group is presented in parenthesis.

SAM-Test: two class unpaired



These values are default

If you use the sample data to compare with the SAM results in Excel, make sure “apply log…” is Unchecked.

SAM-Test: two class unpaired

SAM-Tests Results

File Selected-Spot All-Spots

	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	Observed s...	P value	Numerator(r)	Denominat...	Fold Change	q
1		GENE703				4661482	-0.0454	0.9194	-0.0565	1.2451	0.8345	89
2		GENE704				4661483	-0.1482	0.6589	-0.1534	1.0353	0.9822	84
3		GENE705				4661484	0.761	0.0832	0.9029	1.1864	2.2999	45
4		GENE706				4661485	-0.3265	0.4395	-0.3862	1.1827	0.7524	75
5		GENE707				4661486	0.0143	0.9672	0.0151	1.0555	0.9252	92
6		GENE708				4661487	0.5624	0.1804	0.651	1.1576	1.1782	86
7		GENE709				4661488	0.3802	0.3629	0.4447	1.1698	1.4819	91
8		GENE710				4661489	0.1115	0.7359	0.1147	1.029	1.0886	92
9		GENE711				4661490	-0.3565	0.3319	-0.3835	1.0757	0.8565	75
10		GENE712				4661491	0.1968	0.6172	0.2226	1.1314	1.1666	91
11		GENE713				4661492	0.1229	0.7774	0.1493	1.2145	1.0299	92
12		GENE714				4661493	-2.927	0	-3.4858	1.1909	0.0855	0
13		GENE715				4661494	-0.5412	0.1842	-0.6168	1.1398	0.5658	63
14		GENE716				4661495	-0.6519	0.0836	-0.7026	1.0778	0.6399	58
15		GENE717				4661496	-0.0329	0.9406	-0.0405	1.2304	1.0624	90
16		GENE718				4661497	-0.0177	0.9649	-0.0204	1.1509	1.1047	90
17		GENE719				4661498	-0.159	0.7136	-0.1913	1.2032	0.6056	84
18		GENE720				4661499	-0.5584	0.1313	-0.5993	1.0732	0.6625	63
19		GENE721				4661500	-0.5336	0.2539	-0.6746	1.2642	0.4535	63
20		GENE722				4661501	0.3128	0.4216	0.3501	1.1193	1.3825	91
21		GENE723				4661502	-0.6569	0.1064	-0.7384	1.124	0.5101	58
22		GENE724				4661503	0.2127	0.638	0.2648	1.2449	1.5199	91
23		GENE725				4661504	0.0909	0.8039	0.0987	1.0859	1.0696	92
24		GENE726				4661505	-1.6399	0.0006	-1.7439	1.0634	0.375	0
25		GENE727				4661506	0.3239	0.353	0.3405	1.0513	1.2123	91
26		GENE728				4661507	-0.2093	0.6108	-0.2434	1.1631	0.9654	80
27		GENE729				4661508	0.329	0.4524	0.3995	1.2143	1.4442	91
28		GENE730				4661509	0.0478	0.9102	0.0569	1.1914	1.1592	92

1000 genes

Significance Filtering

Target False Discovery Rate (FDR): Delta =

Select # genes by largest SAM absolute score

Mean Channel Intensities > Bad Flags <=

Abs Fold Change >

Apply Filters Clear Filters

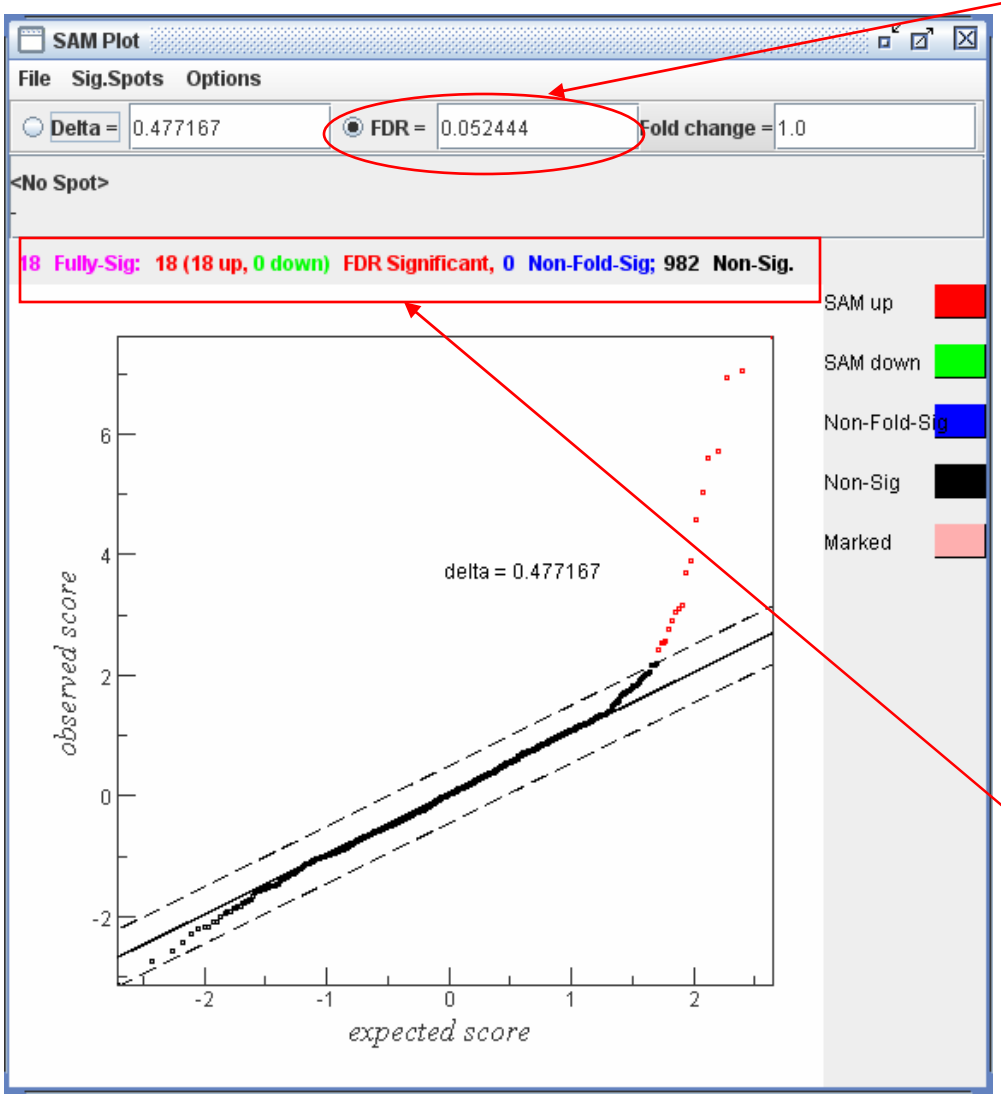
Create Sig. Gene List HCA PCA K-Means SAM Plot

Click SAM Plot button

SAM-Test: two class unpaired

This is the SAM plot

The default FDR <0.05.

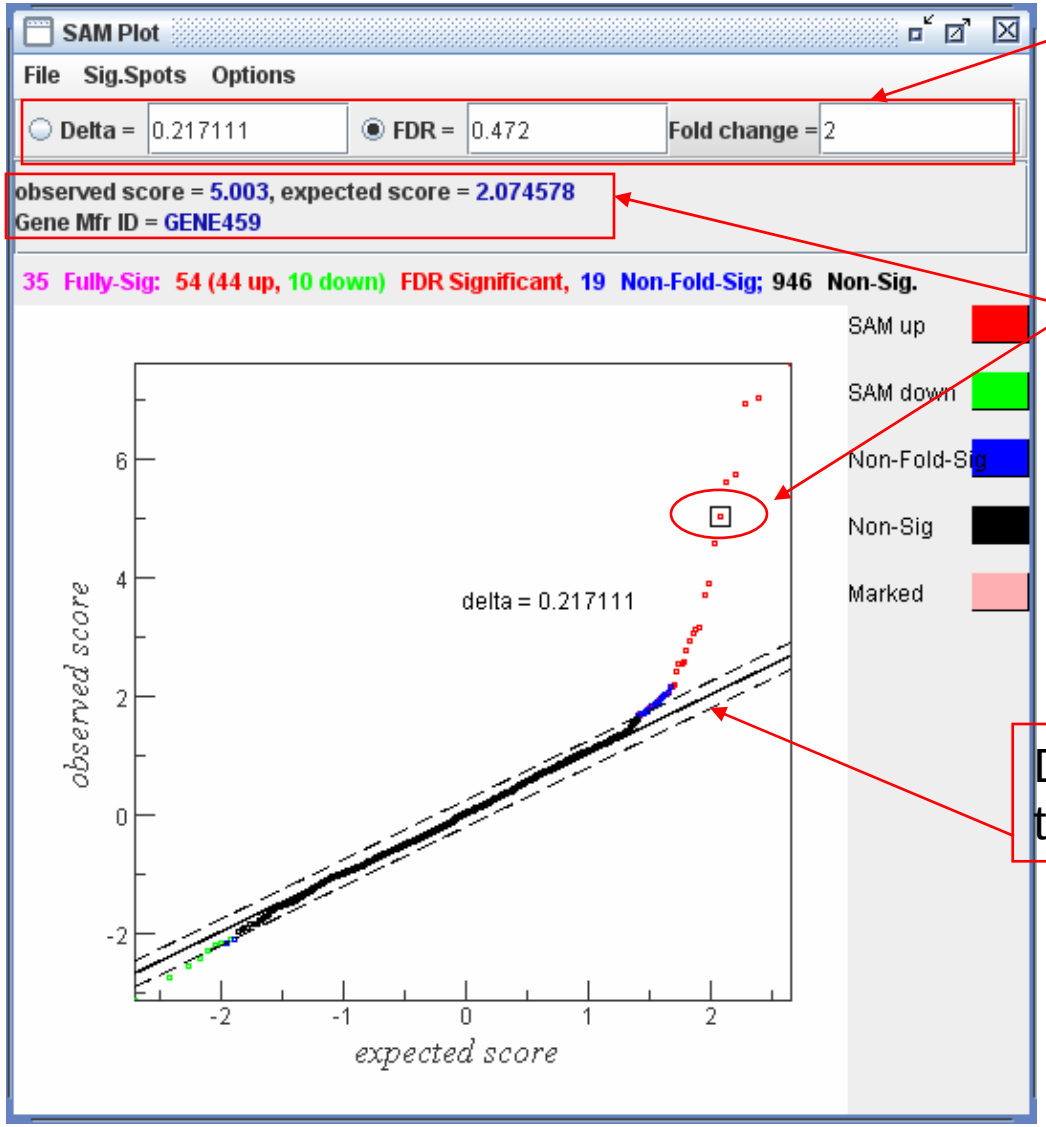


Color legend:
Red – SAM up
Green – SAM down
Blue – non fold significant
Black – non significant
Pink - marked

Display the numbers for significant genes according to the current delta, FDR and fold change values.

SAM-Test: two class unpaired

SAM Plot

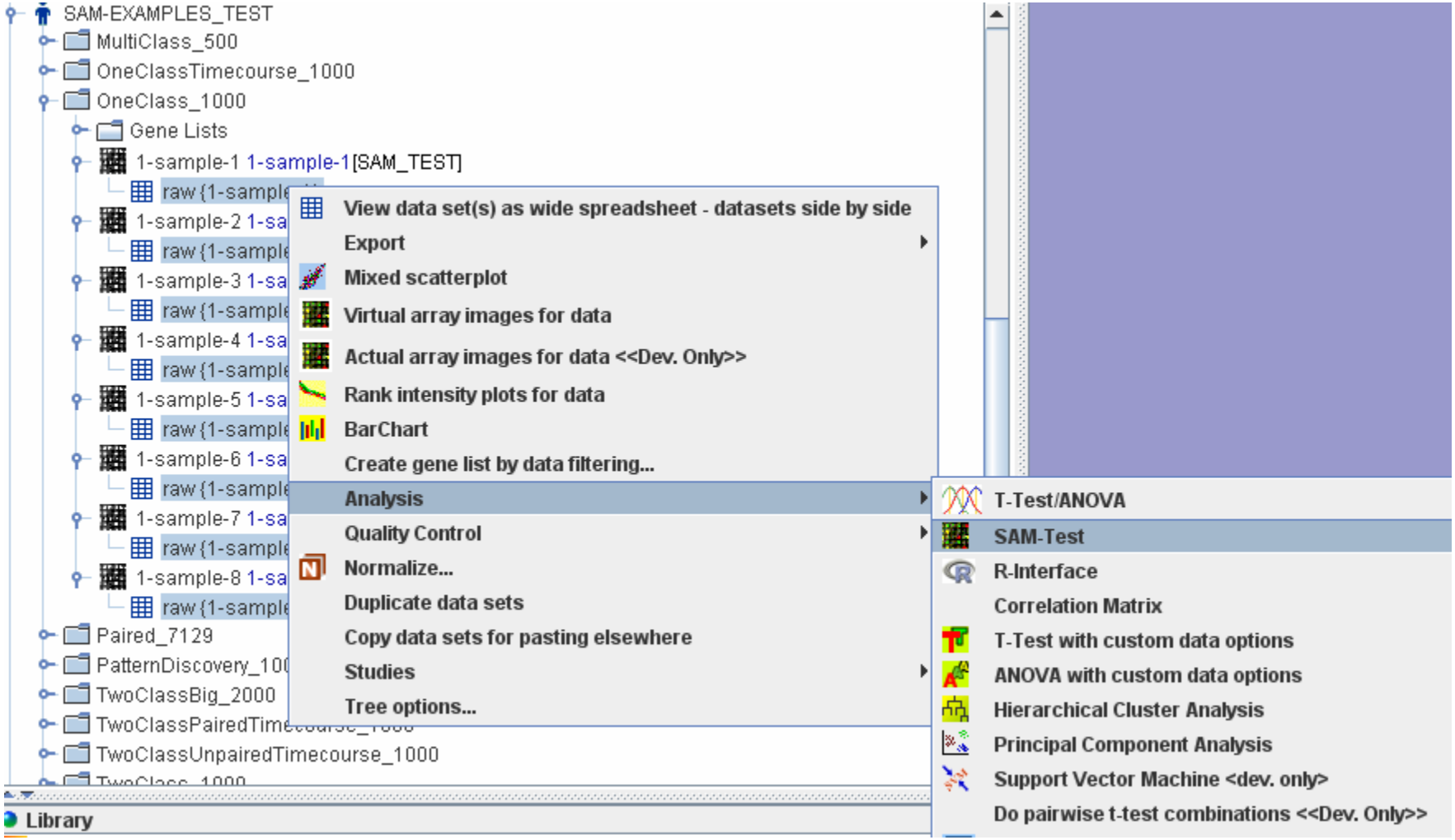


Users can type in numbers for Delta, FDR and fold change

Move mouse over a spot, the score value and gene name will be shown.

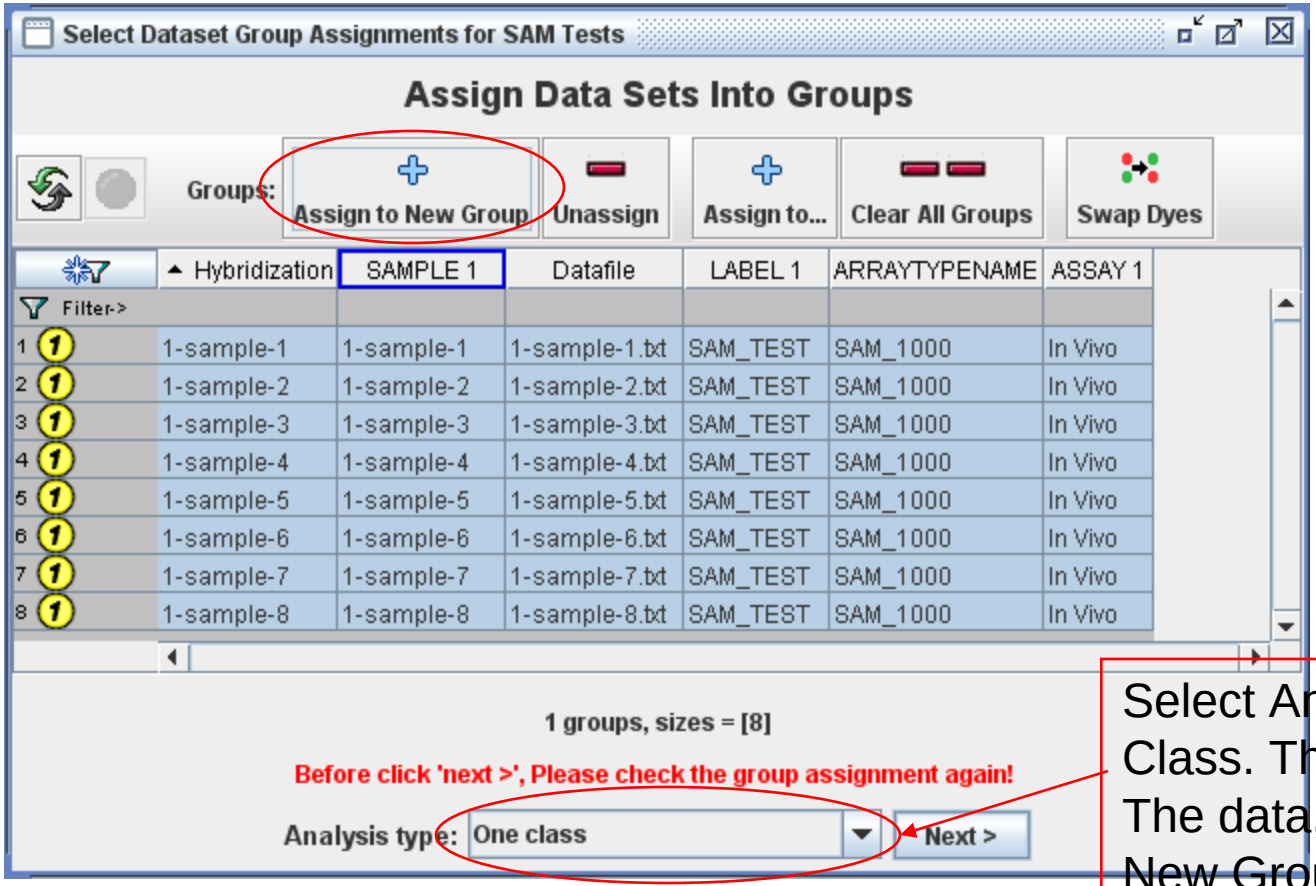
Dragging the dash line will change the delta, FDR value.

SAM-test: One class



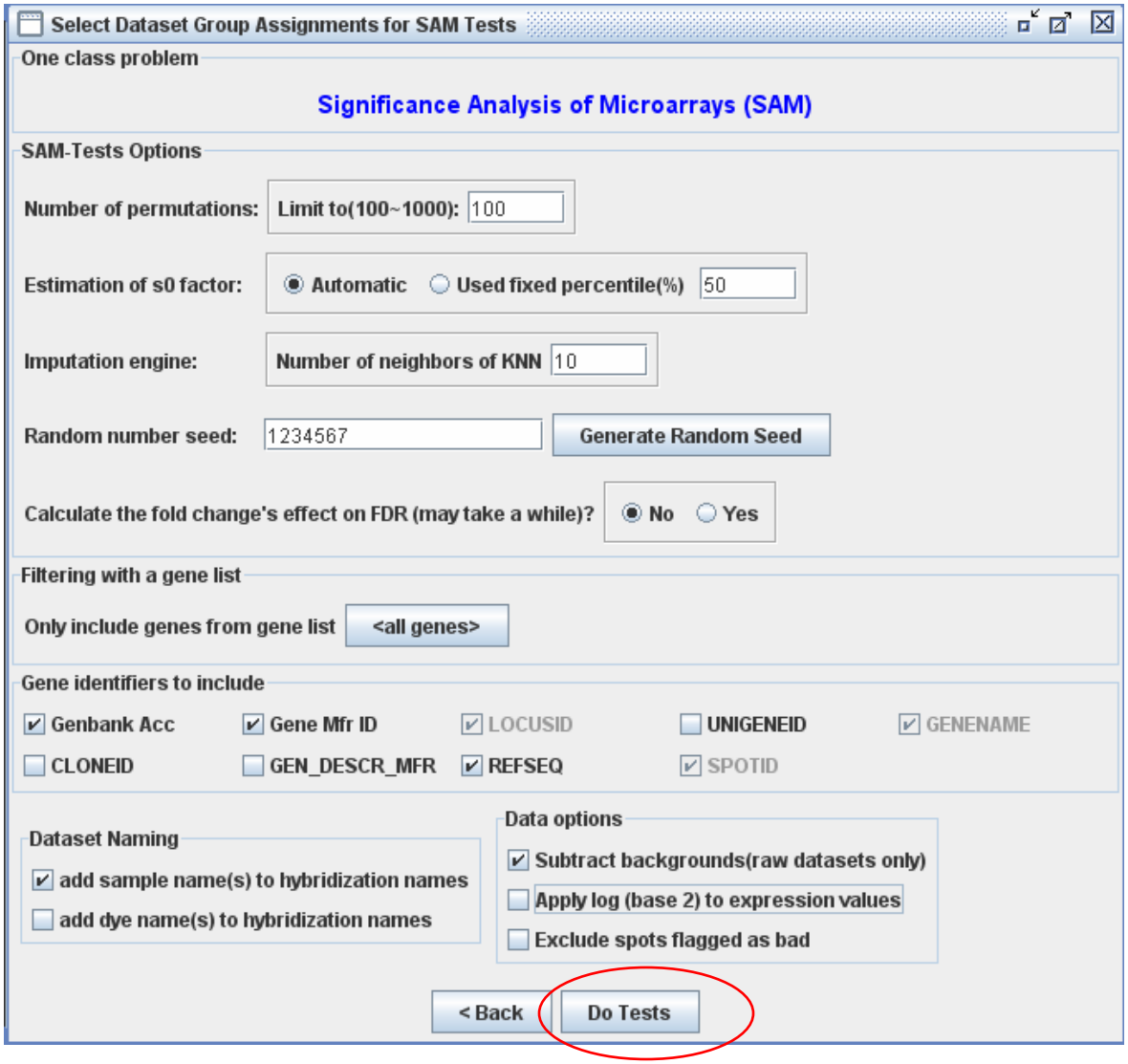
Right-click the selected data set, choose “Analysis” -> SAM-Test.

SAM-test: One class



Select Analysis type – One Class. Then highlight all The data, click “Assign to New Group” button.

SAM-test: One class



The value in the text boxes are default. Click “Do Tests” button.

SAM-test: One class

SAM-Tests Results

File Selected-Spot All-Spots

	Gene Mfr ID	Observed score (d)	P value	Numerator(r)	Denominat...	GEN...	q-value(%)	REF...	SPOTID
1	GENE703	-0.3922		-0.5909	1.5068		5.9055		4661482
2	GENE704	-0.4424	0.1608	-0.6147	1.3895		5.9055		4661483
3	GENE705	-0.2094	0.5751	-0.2568	1.2264		10.9112		4661484
4	GENE706	-0.468	0.3963	-0.741	1.5833		5.9055		4661485
5	GENE707	-0.1579		-0.1847	1.17		10.9112		4661486
6	GENE708	-0.7139	0.1181	-1.2662	1.7738		1.4605		4661487
7	GENE709	-0.5778	0.0789	-0.8629	1.4934		2.496		4661488
8	GENE710	-0.2387	0.6411	-0.3496	1.4643		10.334		4661489
9	GENE711	-0.4595	0.2318	-0.7098	1.5445		5.9055		4661490
10	GENE712	0.283	0.0852	0.3731	1.3186		14.253		4661491
11	GENE713	-0.1933		-0.2218	1.1474		10.9112		4661492
12	GENE714	-0.4287	0.3627	-0.6128	1.4294		5.9055		4661493
13	GENE715	-0.5146	0.2931	-0.9513	1.8486		3.1854		4661494
14	GENE716	-0.4829	0.1603	-0.7646	1.5834		3.1854		4661495
15	GENE717	0.4599	0.0169	0.6108	1.3282		12.4867		4661496
16	GENE718	-0.5346	0.3272	-1.0246	1.9166		3.1854		4661497
17	GENE719	0.0668	0.5303	0.0971	1.454		14.253		4661498
18	GENE720	-0.8768	0.1351	-1.6637	1.8975		0		4661499
19	GENE721	-0.3551		-0.494	1.3915		6.6517		4661500

1000 genes

Significance Filtering

Target False Discovery Rate (FDR): Delta =

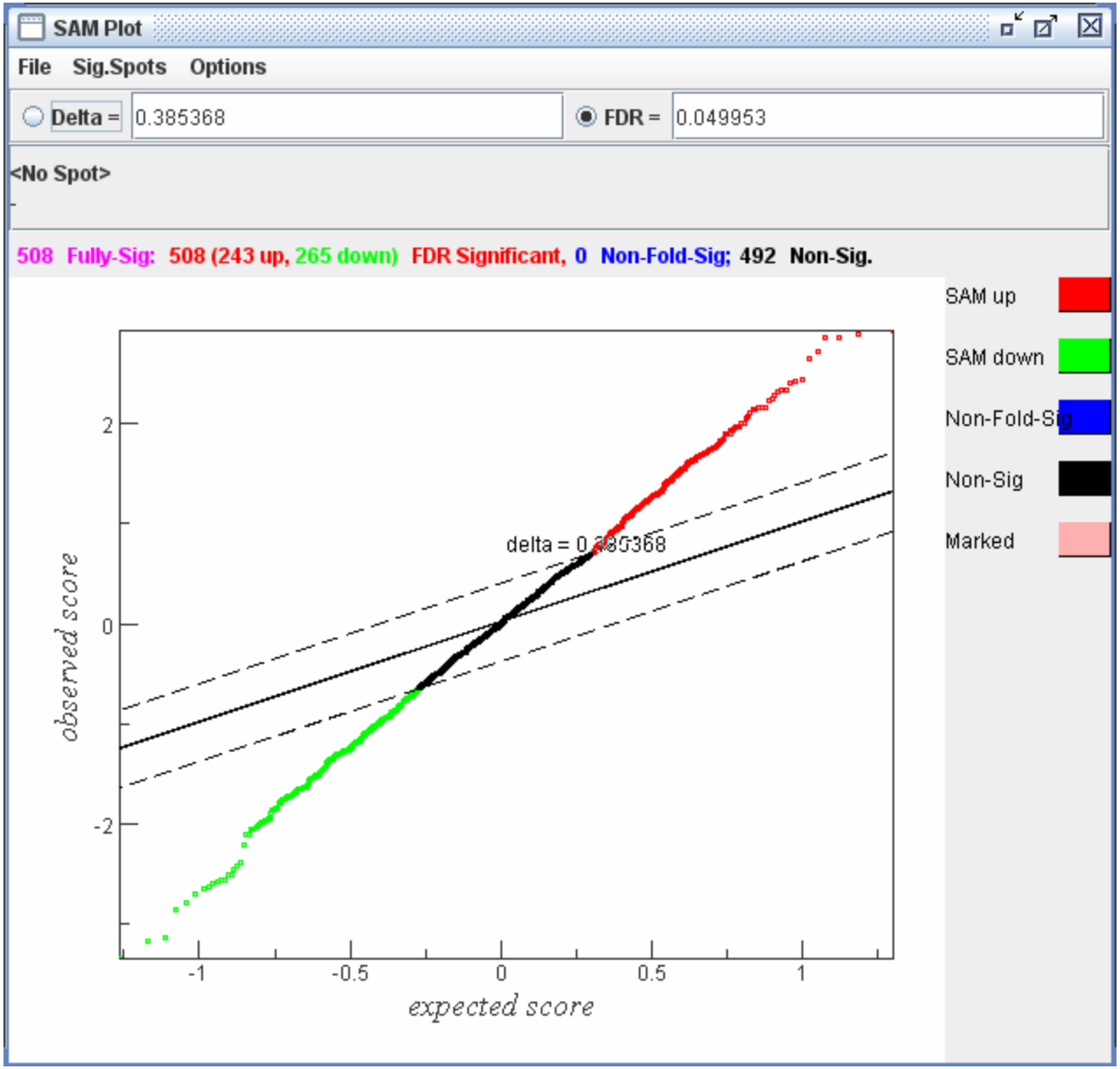
Select # genes by largest SAM absolute score

Mean Channel Intensities > Bad Flags <=

Apply Filters Clear Filters

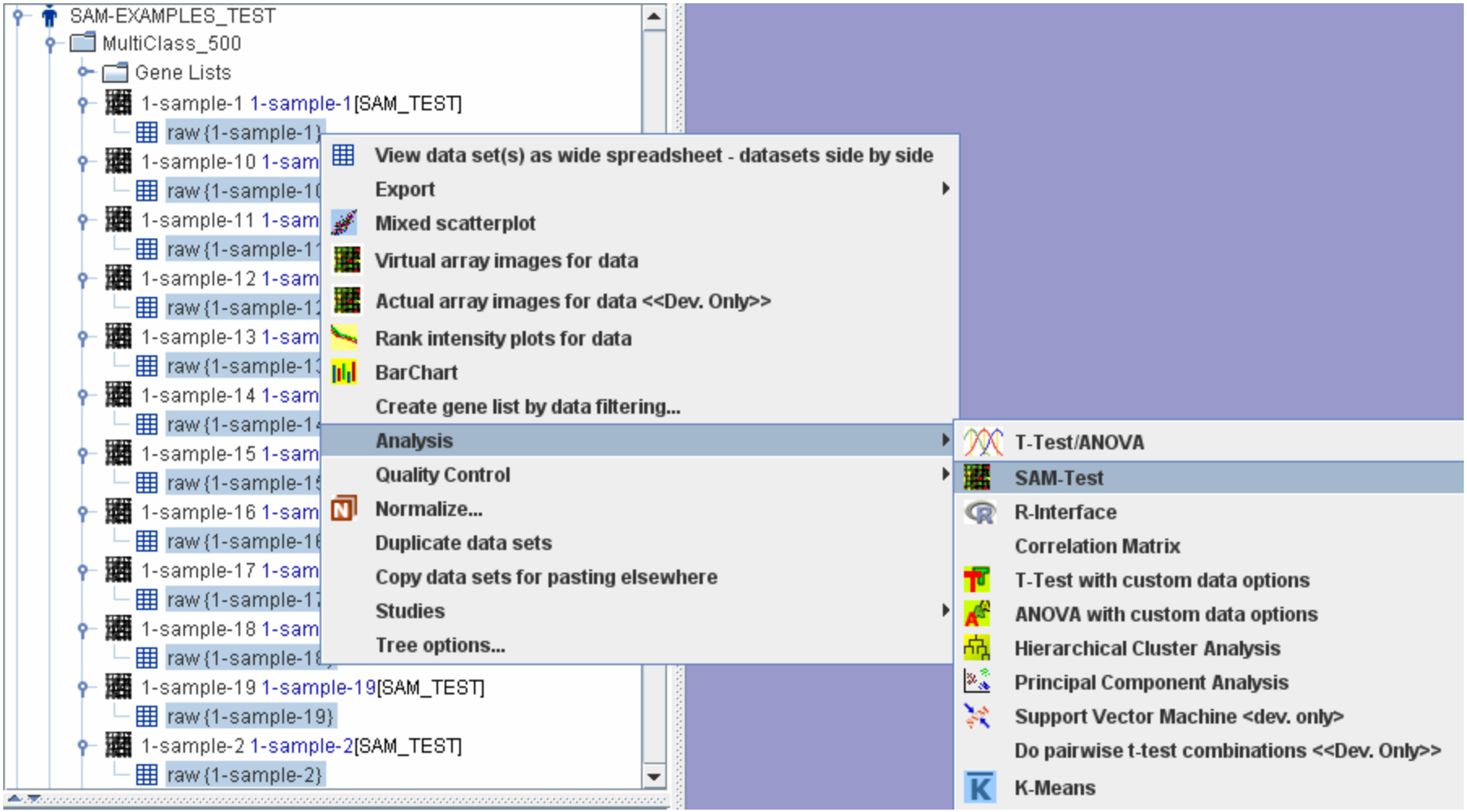
Create Sig. Gene List HCA PCA K-Means SAM Plot

SAM-test: One class

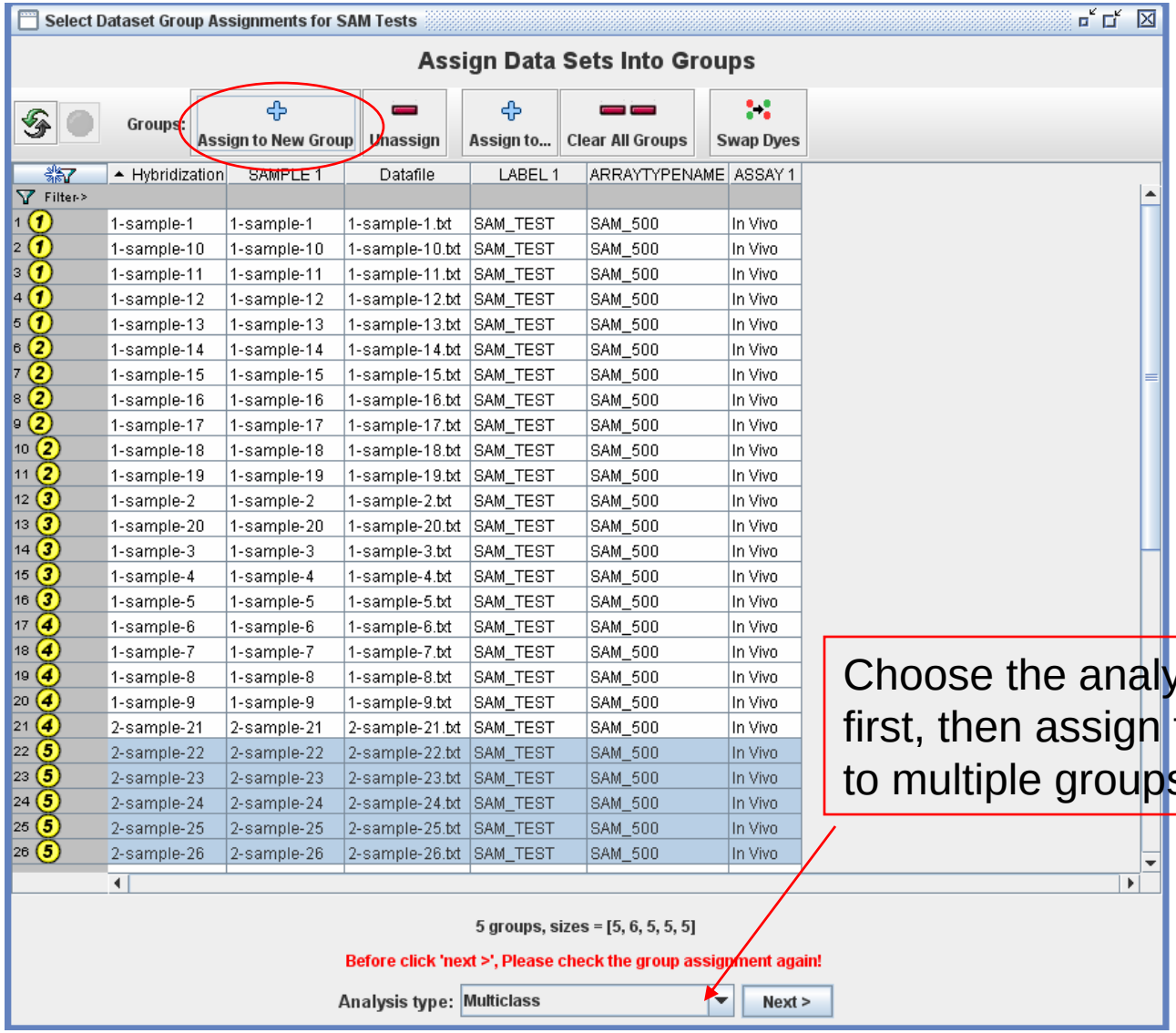


SAM-test: MultiClass

Right-click the selected data set, choose “Analysis” ->SAM-Test



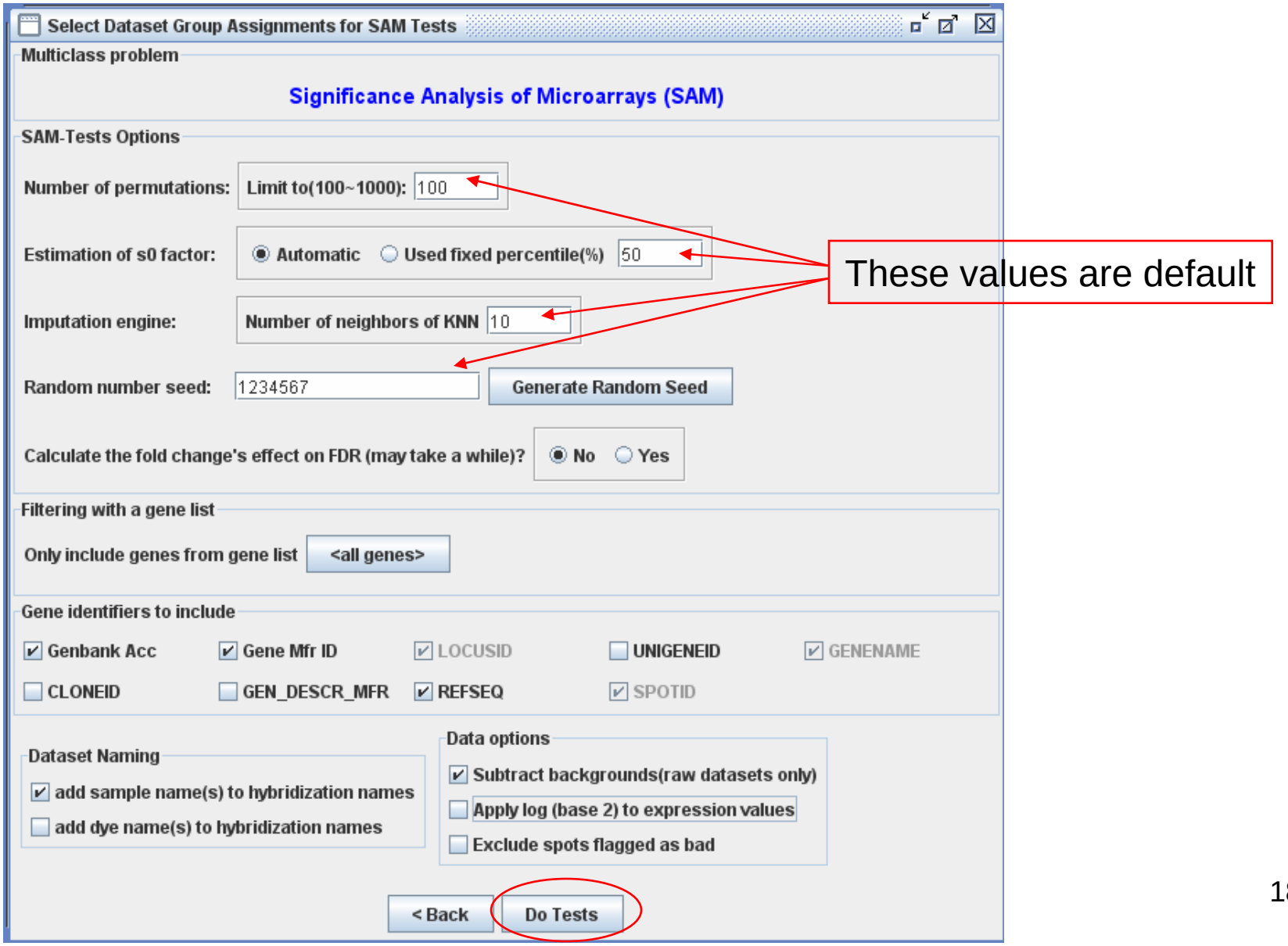
SAM-test: MultiClass



		Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPENAME	ASSAY 1
1	1	1-sample-1	1-sample-1	1-sample-1.txt	SAM_TEST	SAM_500	In Vivo
2	1	1-sample-10	1-sample-10	1-sample-10.txt	SAM_TEST	SAM_500	In Vivo
3	1	1-sample-11	1-sample-11	1-sample-11.txt	SAM_TEST	SAM_500	In Vivo
4	1	1-sample-12	1-sample-12	1-sample-12.txt	SAM_TEST	SAM_500	In Vivo
5	1	1-sample-13	1-sample-13	1-sample-13.txt	SAM_TEST	SAM_500	In Vivo
6	2	1-sample-14	1-sample-14	1-sample-14.txt	SAM_TEST	SAM_500	In Vivo
7	2	1-sample-15	1-sample-15	1-sample-15.txt	SAM_TEST	SAM_500	In Vivo
8	2	1-sample-16	1-sample-16	1-sample-16.txt	SAM_TEST	SAM_500	In Vivo
9	2	1-sample-17	1-sample-17	1-sample-17.txt	SAM_TEST	SAM_500	In Vivo
10	2	1-sample-18	1-sample-18	1-sample-18.txt	SAM_TEST	SAM_500	In Vivo
11	2	1-sample-19	1-sample-19	1-sample-19.txt	SAM_TEST	SAM_500	In Vivo
12	3	1-sample-2	1-sample-2	1-sample-2.txt	SAM_TEST	SAM_500	In Vivo
13	3	1-sample-20	1-sample-20	1-sample-20.txt	SAM_TEST	SAM_500	In Vivo
14	3	1-sample-3	1-sample-3	1-sample-3.txt	SAM_TEST	SAM_500	In Vivo
15	3	1-sample-4	1-sample-4	1-sample-4.txt	SAM_TEST	SAM_500	In Vivo
16	3	1-sample-5	1-sample-5	1-sample-5.txt	SAM_TEST	SAM_500	In Vivo
17	4	1-sample-6	1-sample-6	1-sample-6.txt	SAM_TEST	SAM_500	In Vivo
18	4	1-sample-7	1-sample-7	1-sample-7.txt	SAM_TEST	SAM_500	In Vivo
19	4	1-sample-8	1-sample-8	1-sample-8.txt	SAM_TEST	SAM_500	In Vivo
20	4	1-sample-9	1-sample-9	1-sample-9.txt	SAM_TEST	SAM_500	In Vivo
21	4	2-sample-21	2-sample-21	2-sample-21.txt	SAM_TEST	SAM_500	In Vivo
22	5	2-sample-22	2-sample-22	2-sample-22.txt	SAM_TEST	SAM_500	In Vivo
23	5	2-sample-23	2-sample-23	2-sample-23.txt	SAM_TEST	SAM_500	In Vivo
24	5	2-sample-24	2-sample-24	2-sample-24.txt	SAM_TEST	SAM_500	In Vivo
25	5	2-sample-25	2-sample-25	2-sample-25.txt	SAM_TEST	SAM_500	In Vivo
26	5	2-sample-26	2-sample-26	2-sample-26.txt	SAM_TEST	SAM_500	In Vivo

Choose the analysis type first, then assign the data set to multiple groups.

SAM-test: MultiClass



SAM-test: MultiClass

SAM-Tests Results

File Selected-Spot All-Spots

	Gene Mfr ID	SPOTID	Observed score (d)	P value	Numerator(r)	Denominat...	contrast-1	contrast
1	GENE199	4660478	0.0684	0.9532	0.0826	1.2088	0.6396	-0.1507
2	GENE200	4660479	0.1082	0.6209	0.1897	1.7535	0.2063	0.5882
3	GENE201	4660480	0.0632	0.9577	0.1119	1.7695	0.0768	-0.3344
4	GENE202	4660481	0.0548		0.0973	1.7752	0.2543	0.1381
5	GENE203	4660482	0.1619	0.1716	0.2372	1.4648	1.0482	0.0642
6	GENE204	4660483	0.071	0.8664	0.1442	2.0316	-0.0939	0.1011
7	GENE205	4660484	0.1747	0.1109	0.2407	1.3779	0.4285	0.4524
8	GENE206	4660485	0.087	0.6966	0.1474	1.6948	-0.0446	0.045
9	GENE207	4660486	0.1188	0.1882	0.1727	1.4537	0.0569	-0.0524
10	GENE208	4660487	0.1523	0.1006	0.2093	1.3743	0.3434	-0.6046
11	GENE209	4660488	0.105	0.4709	0.2001	1.9061	-0.4153	0.2514
12	GENE210	4660489	0.0902	0.7682	0.1677	1.8595	-0.3386	0.0254
13	GENE211	4660490	0.1079	0.6225	0.1816	1.6826	0.1645	0.428
14	GENE212	4660491	0.1174	0.3474	0.2096	1.7856	0.2178	-0.7163
15	GENE213	4660492	0.1563	0.1057	0.2266	1.4495	-0.2053	0.786
16	GENE214	4660493	0.0683		0.1305	1.9108	0.0806	0.1939
17	GENE215	4660494	0.2033	0.019	0.3867	1.9027	0.1648	0.3296
18	GENE216	4660495	0.1884		0.2603	1.3816	-1.5802	0.5408
19	GENE217	4660496	0.1153	0.664	0.1646	1.4273	0.2295	0.4566

500 genes

Significance Filtering

Target False Discovery Rate (FDR): Delta =

Select # genes by largest SAM absolute score

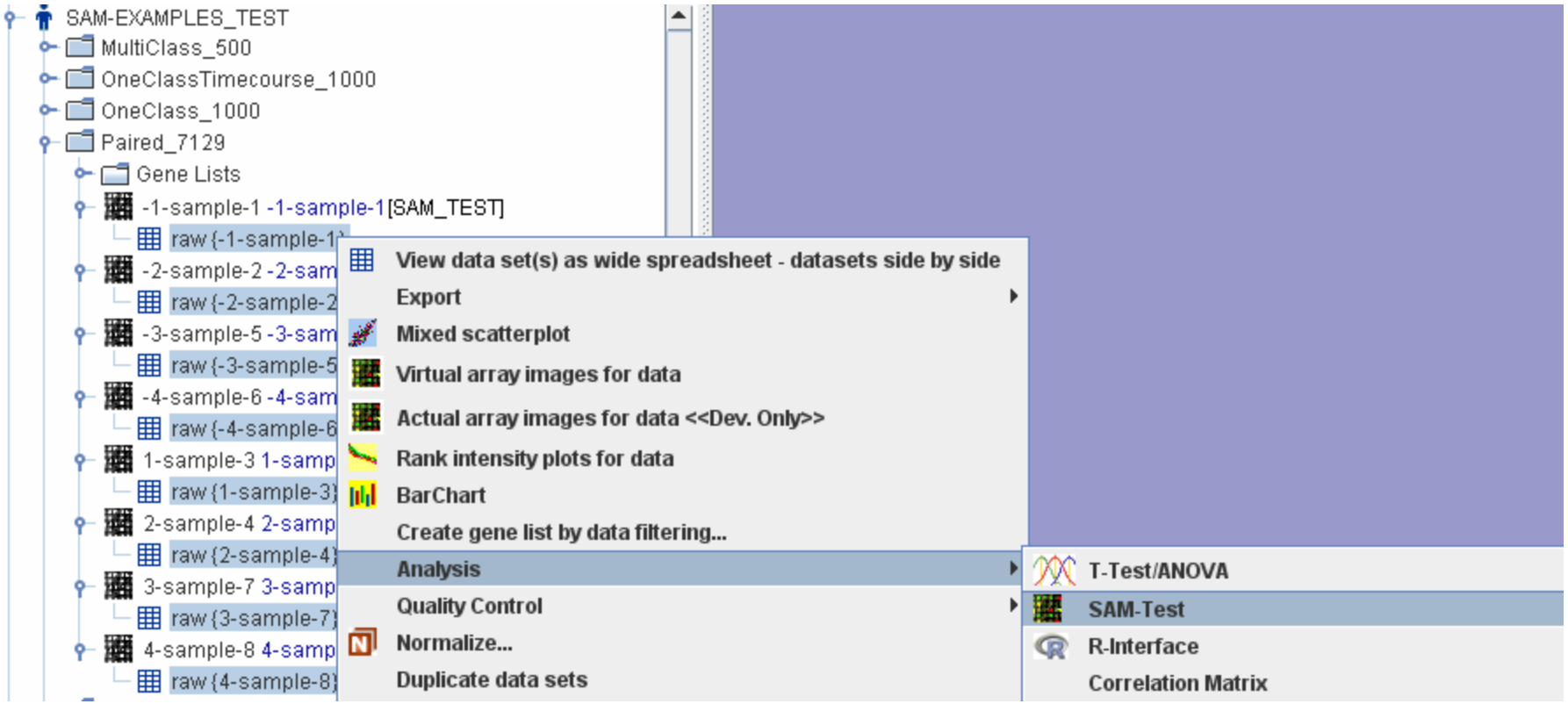
Mean Channel Intensities > Bad Flags <=

Apply Filters Clear Filters

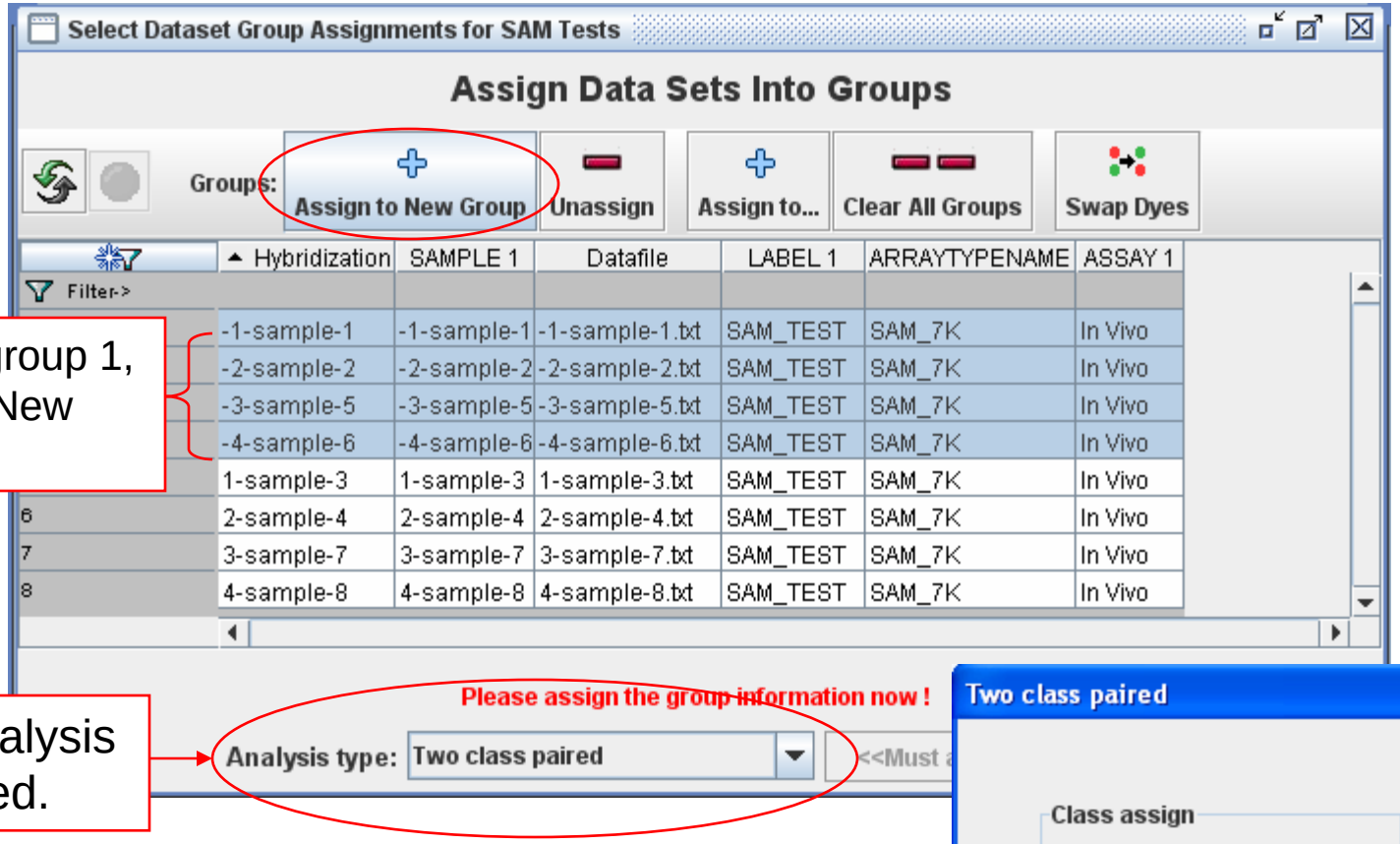
Create Sig. Gene List HCA PCA K-Means SAM Plot

SAM-Test: two class paired

Right-click the selected data set, choose “Analysis” ->SAM-Test.

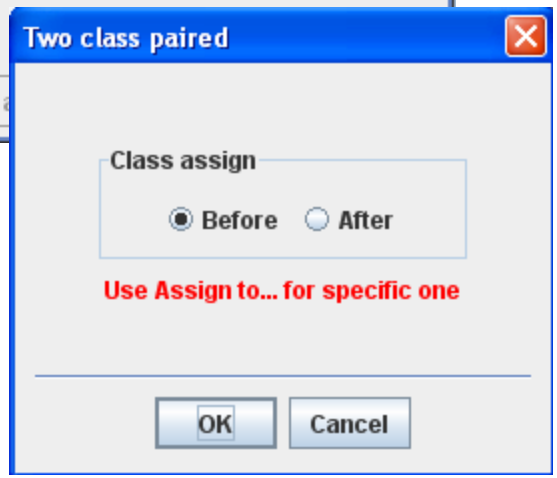


SAM-Test: two class paired



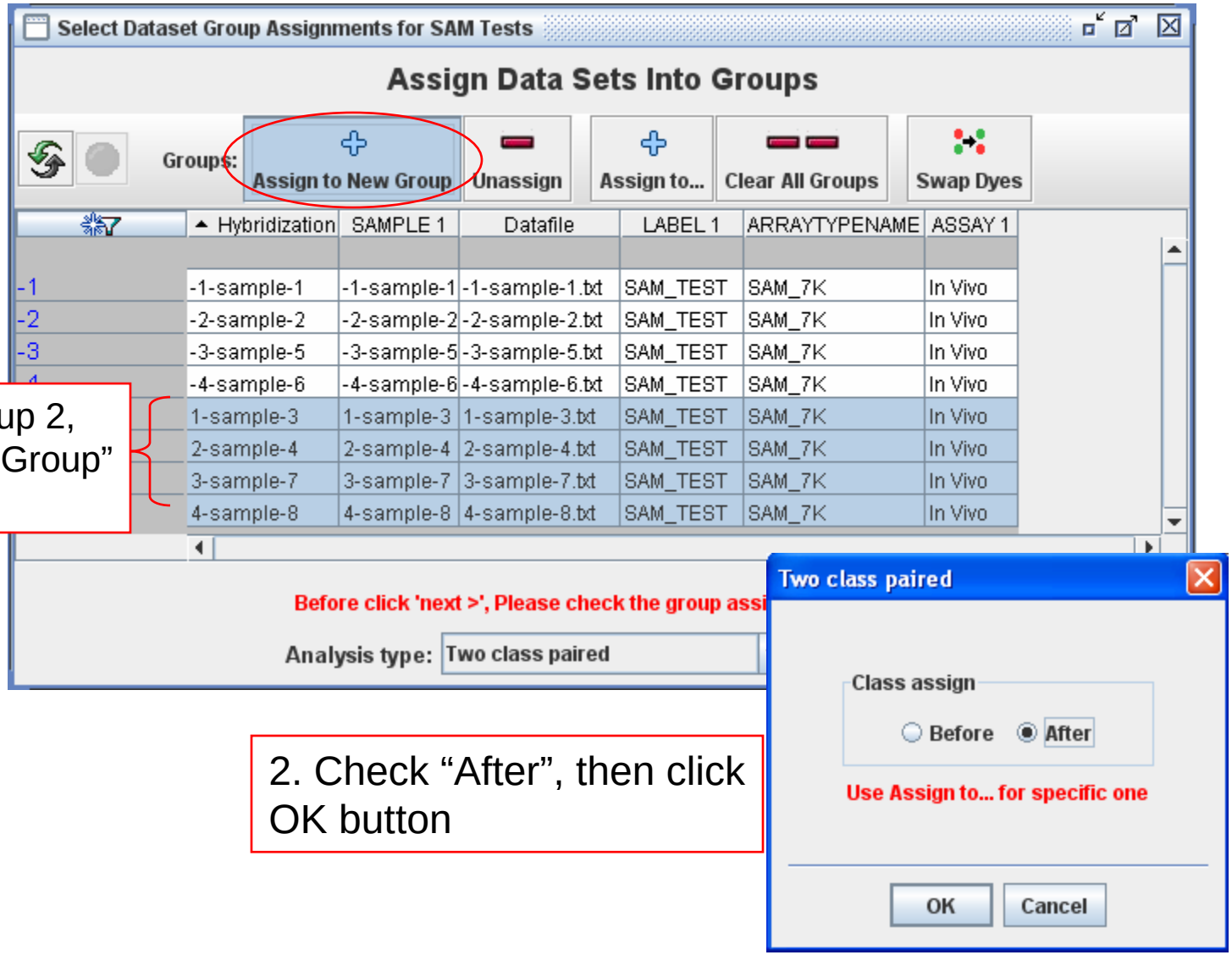
2 Select data for group 1, click “Assign to New Group” button.

1 Make sure analysis type is selected.



3 Check “Before”, then click OK button

SAM-Test: two class paired

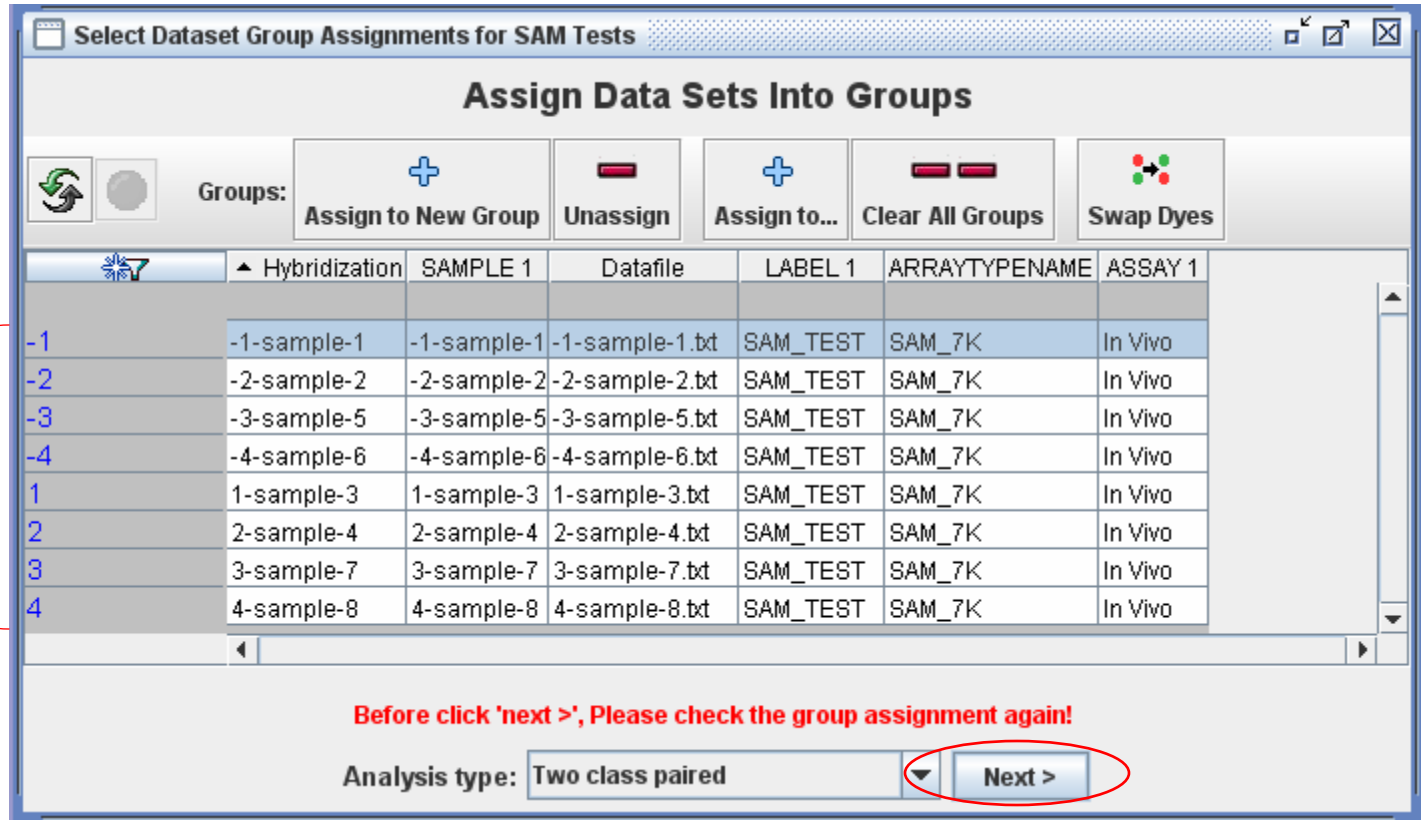


1. Select data for group 2, Click “Assign to New Group” Button.

2. Check “After”, then click OK button

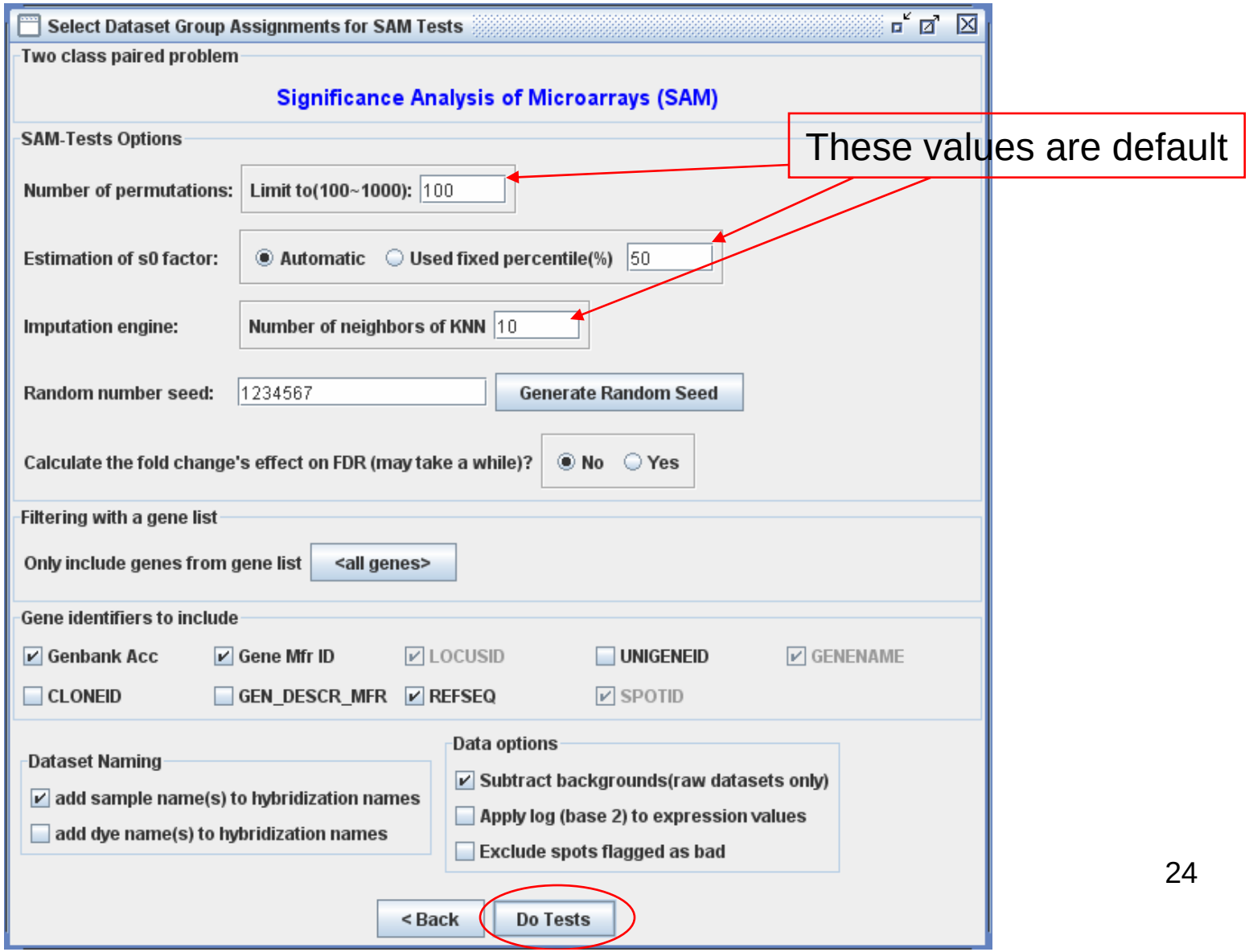
SAM-Test: two class paired

Data are assigned
In pairs



Click “Next” button.

SAM-Test: two class paired



SAM-Test: two class paired

SAM-Tests Results

File Selected-Spot All-Spots

	Gene Mfr ID	SPOTID	Observed score (d)	Numerator(r)	Denominat...	Fold Change	q-value(%)	localfdr
1	GENE870	4654020	0.4245	0.7516	1.7706	4.9263	80.8766	94.0522
2	GENE871	4654021	0.0677	0.1264	1.8682	3.0871	94.3878	100
3	GENE872	4654022	-0.2398	-0.4027	1.6793	0.8102	86.4446	94.1904
4	GENE873	4654023	-0.0792	-0.1556	1.9641	1.3397	91.078	99.4605
5	GENE874	4654024	0.0535	0.1033	1.9306	1.7484	94.3878	100
6	GENE875	4654025	-0.2488	-0.5128	2.0609	1.1078	86.4446	93.8744
7	GENE876	4654026	-0.2298	-0.3941	1.715	0.823	86.4446	94.5365
8	GENE877	4654027	-0.6001	-1.0456	1.7423	0.4787	77.8565	80.77
9	GENE878	4654028	-0.3534	-0.605	1.712	0.6823	86.4446	89.5737
10	GENE826	4653976	-0.0168	-0.0245	1.4572	1.0481	92.688	100
11	GENE827	4653977	-0.2985	-0.4591	1.5381	0.7289	86.4446	92.0149
12	GENE828	4653978	-0.1136	-0.2142	1.8858	1.3896	89.6273	98.399
13	GENE829	4653979	0.3222	0.5884	1.826	3.588	86.3568	95.7861
14	GENE830	4653980	0.0323	0.0533	1.6476	1.2171	94.3878	100
15	GENE831	4653981	0.0076	0.0122	1.5909	1.1026	94.3878	100
16	GENE832	4653982	-0.0855	-0.1446	1.6914	1.2943	91.078	99.2674
17	GENE833	4653983	0.3023	0.4876	1.6128	2.1278	86.4309	96.1364

7129 genes

Significance Filtering

Target False Discovery Rate (FDR): Delta =

Select # genes by largest SAM absolute score

Mean Channel Intensities > Bad Flags <=

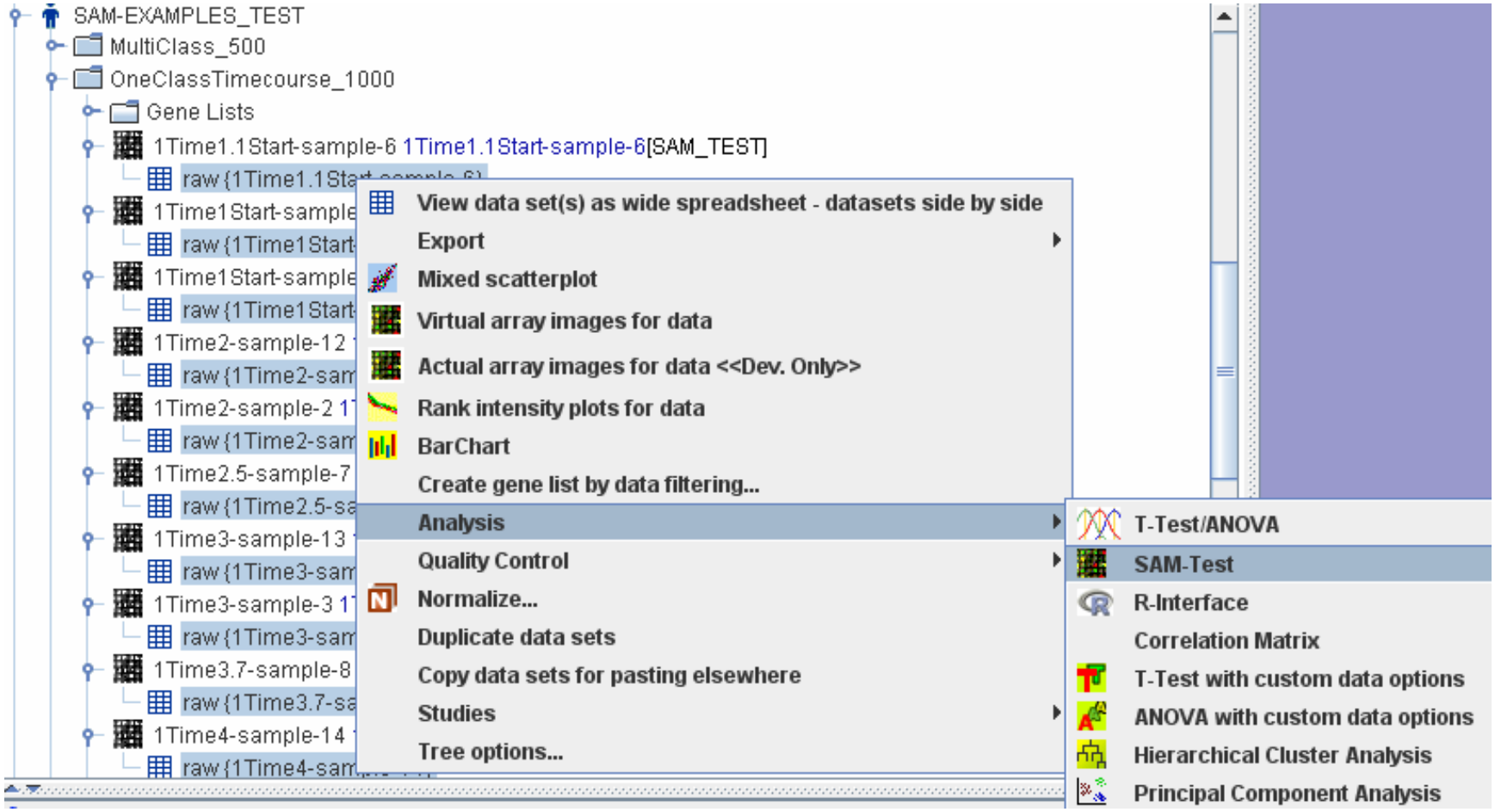
Abs Fold Change >

Apply Filters Clear Filters

Create Sig. Gene List HCA PCA K-Means SAM Plot

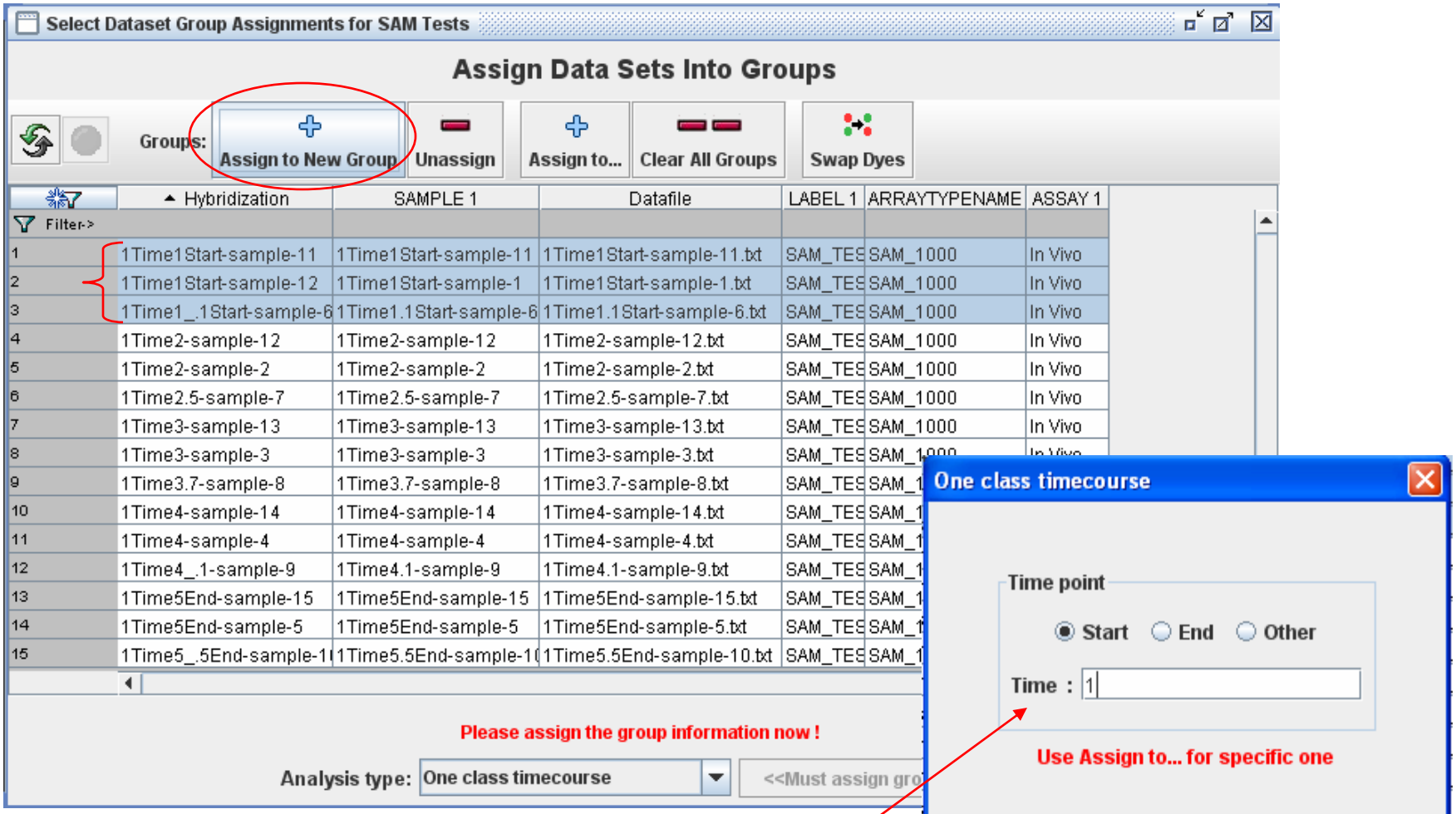
SAM-Test: Timecourse

One class timecourse



SAM-Test: one class timecourse

Highlight the first group, click “Assign to New Group” button.

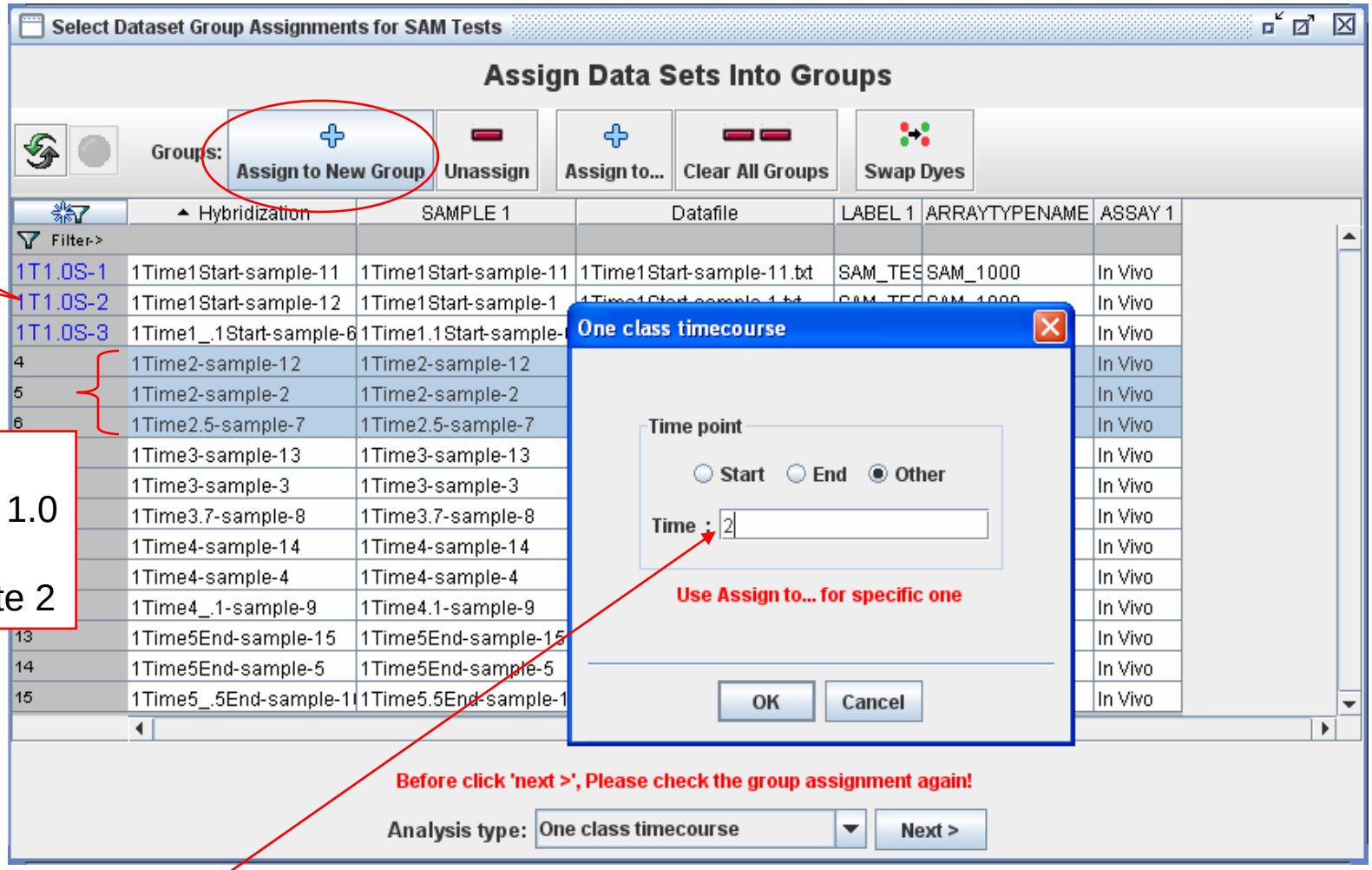


Check “Start” radio button, type in number “1” in the time text box. Click OK button.

SAM-Test: One Class Timecourse

1T1.0S-2

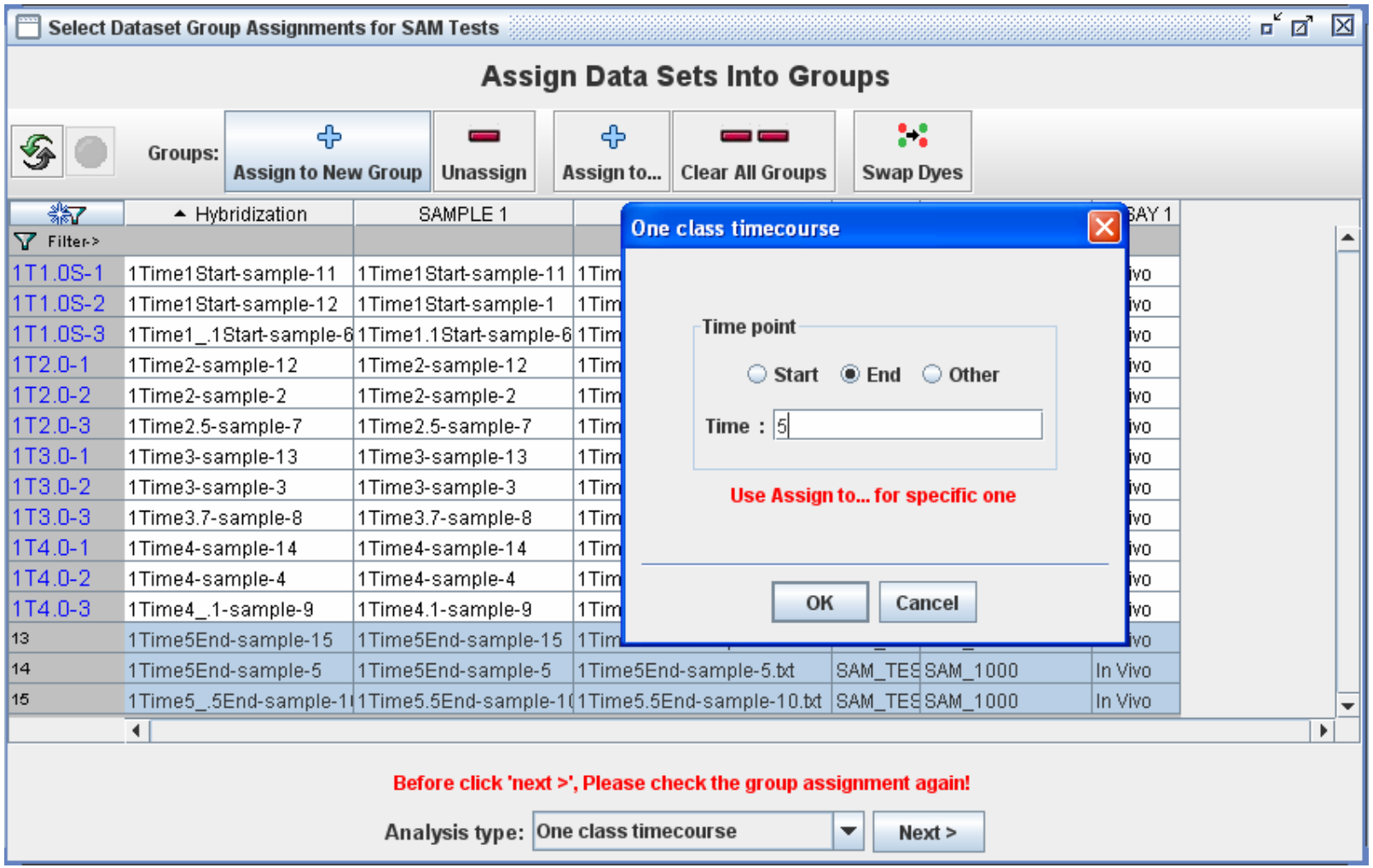
1 means class 1.
T1.0 means time 1.0
S means start
-2 means replicate 2



For second group, check “Other” radio button and type in number “2” for time course.

SAM-Test: One Class Timecourse

Repeat the previous steps for assigning the other groups. For the last group, check "End" radio button and type in the time. Click OK button.



SAM-Test: One Class Timecourse

This is what looks like after assigning all the groups.

Select Dataset Group Assignments for SAM Tests

Assign Data Sets Into Groups

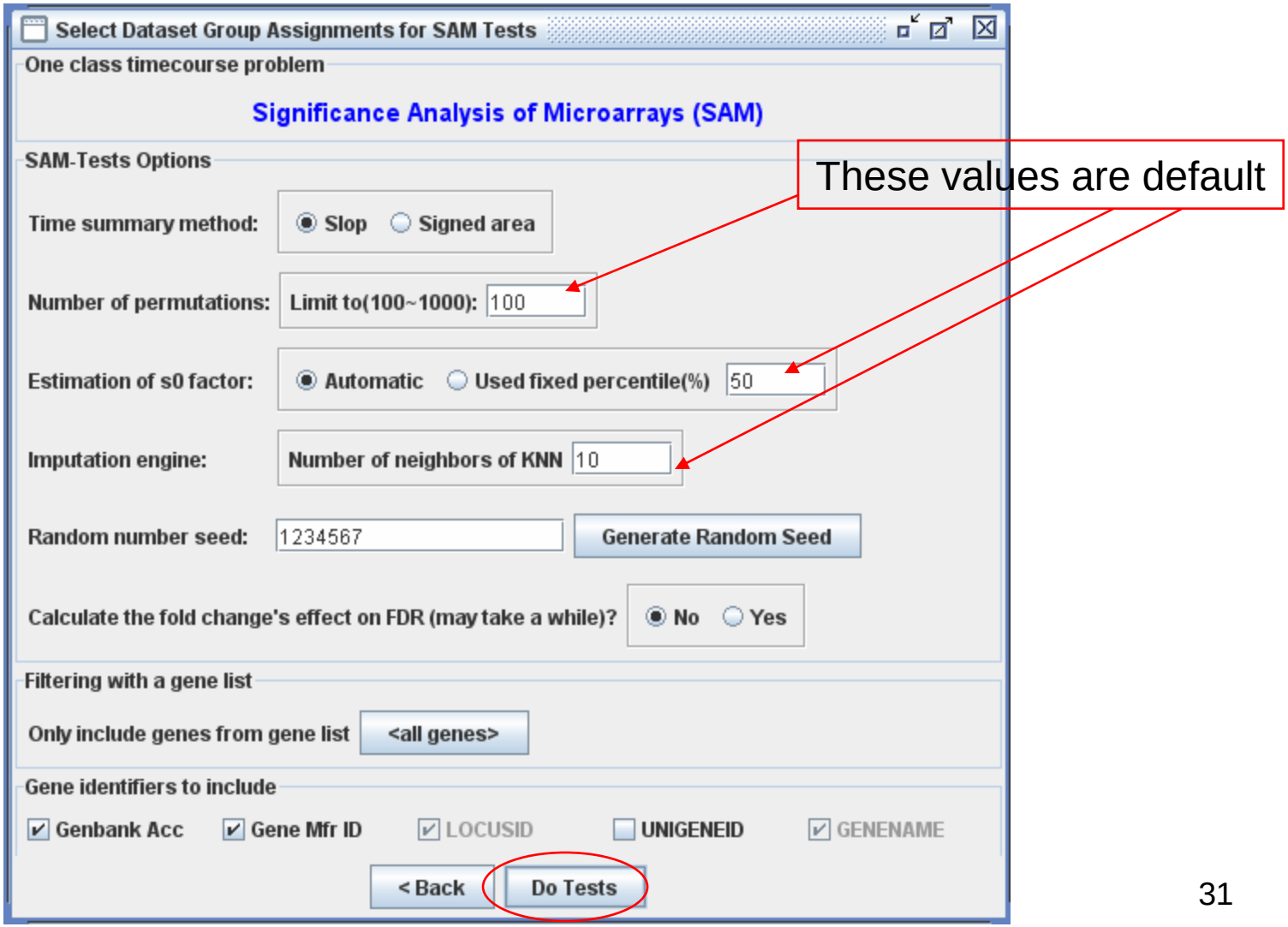
Groups: Assign to New Group | Unassign | Assign to... | Clear All Groups | Swap Dyes

	Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPENAME	ASSAY 1
Filter->						
Time 1 Start: 1T1.0S-1	1Time1Start-sample-11	1Time1Start-sample-11	1Time1Start-sample-11.txt	SAM_TEST	SAM_1000	In Vivo
1T1.0S-2	1Time1Start-sample-12	1Time1Start-sample-1	1Time1Start-sample-1.txt	SAM_TEST	SAM_1000	In Vivo
1T1.0S-3	1Time1_.1Start-sample-6	1Time1.1Start-sample-6	1Time1.1Start-sample-6.txt	SAM_TEST	SAM_1000	In Vivo
Time 2: 1T2.0-1	1Time2-sample-12	1Time2-sample-12	1Time2-sample-12.txt	SAM_TEST	SAM_1000	In Vivo
1T2.0-2	1Time2-sample-2	1Time2-sample-2	1Time2-sample-2.txt	SAM_TEST	SAM_1000	In Vivo
1T2.0-3	1Time2.5-sample-7	1Time2.5-sample-7	1Time2.5-sample-7.txt	SAM_TEST	SAM_1000	In Vivo
Time 3: 1T3.0-1	1Time3-sample-13	1Time3-sample-13	1Time3-sample-13.txt	SAM_TEST	SAM_1000	In Vivo
1T3.0-2	1Time3-sample-3	1Time3-sample-3	1Time3-sample-3.txt	SAM_TEST	SAM_1000	In Vivo
1T3.0-3	1Time3.7-sample-8	1Time3.7-sample-8	1Time3.7-sample-8.txt	SAM_TEST	SAM_1000	In Vivo
Time 4: 1T4.0-1	1Time4-sample-14	1Time4-sample-14	1Time4-sample-14.txt	SAM_TEST	SAM_1000	In Vivo
1T4.0-2	1Time4-sample-4	1Time4-sample-4	1Time4-sample-4.txt	SAM_TEST	SAM_1000	In Vivo
1T4.0-3	1Time4_.1-sample-9	1Time4.1-sample-9	1Time4.1-sample-9.txt	SAM_TEST	SAM_1000	In Vivo
Time 5 End: 1T5.0E-1	1Time5End-sample-15	1Time5End-sample-15	1Time5End-sample-15.txt	SAM_TEST	SAM_1000	In Vivo
1T5.0E-2	1Time5End-sample-5	1Time5End-sample-5	1Time5End-sample-5.txt	SAM_TEST	SAM_1000	In Vivo
1T5.0E-3	1Time5_.5End-sample-1	1Time5.5End-sample-1	1Time5.5End-sample-10.txt	SAM_TEST	SAM_1000	In Vivo

Before click 'next >', Please check the group assignment again!

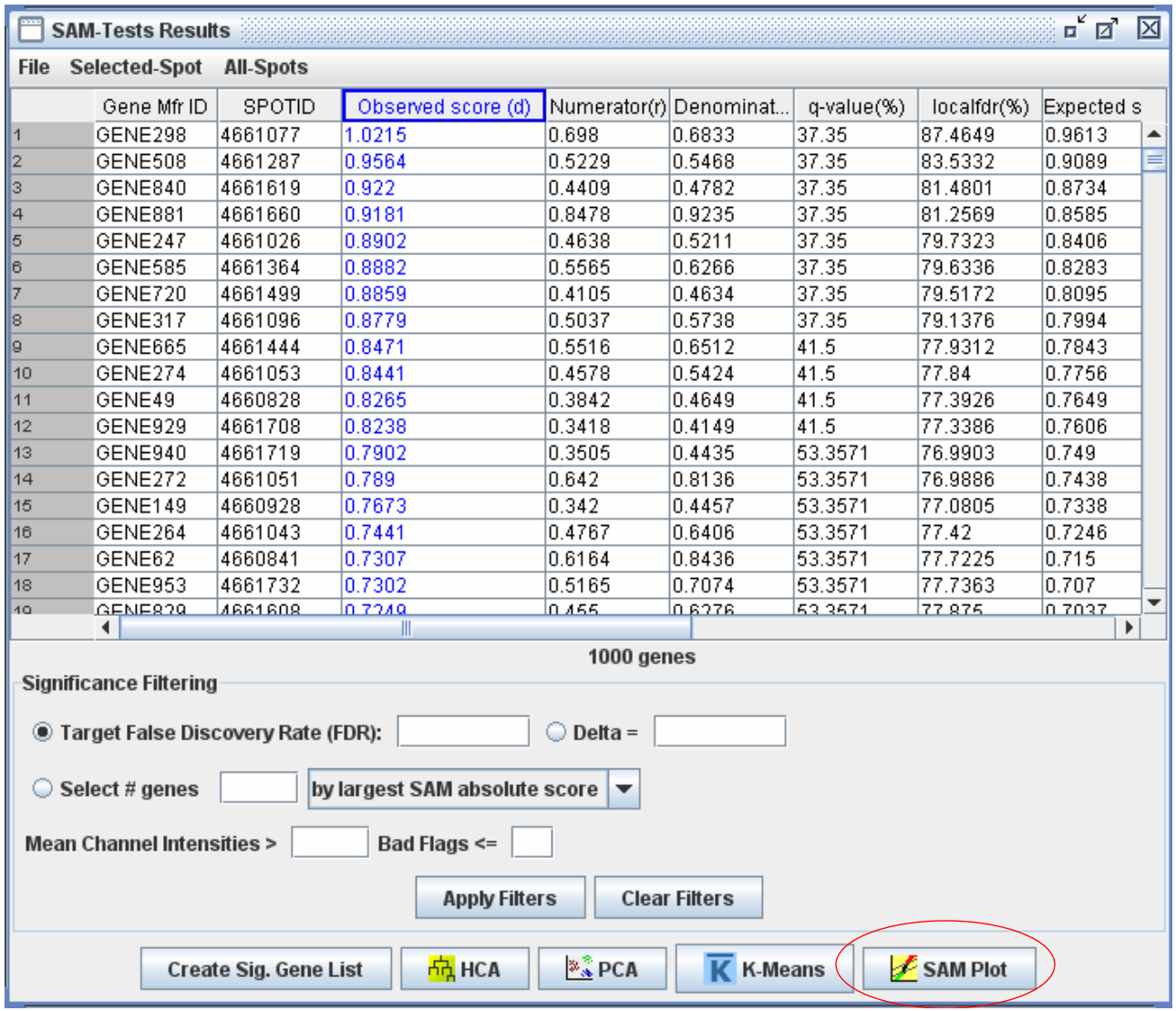
Analysis type: One class timecourse | Next >

SAM-Test: One Class Timecourse



SAM-Test: One Class Timecourse

One class time course result:



SAM-Tests Results

File Selected-Spot All-Spots

	Gene Mfr ID	SPOTID	Observed score (d)	Numerator(r)	Denominat...	q-value(%)	localfdr(%)	Expected s
1	GENE298	4661077	1.0215	0.698	0.6833	37.35	87.4649	0.9613
2	GENE508	4661287	0.9564	0.5229	0.5468	37.35	83.5332	0.9089
3	GENE840	4661619	0.922	0.4409	0.4782	37.35	81.4801	0.8734
4	GENE881	4661660	0.9181	0.8478	0.9235	37.35	81.2569	0.8585
5	GENE247	4661026	0.8902	0.4638	0.5211	37.35	79.7323	0.8406
6	GENE585	4661364	0.8882	0.5565	0.6266	37.35	79.6336	0.8283
7	GENE720	4661499	0.8859	0.4105	0.4634	37.35	79.5172	0.8095
8	GENE317	4661096	0.8779	0.5037	0.5738	37.35	79.1376	0.7994
9	GENE665	4661444	0.8471	0.5516	0.6512	41.5	77.9312	0.7843
10	GENE274	4661053	0.8441	0.4578	0.5424	41.5	77.84	0.7756
11	GENE49	4660828	0.8265	0.3842	0.4649	41.5	77.3926	0.7649
12	GENE929	4661708	0.8238	0.3418	0.4149	41.5	77.3386	0.7606
13	GENE940	4661719	0.7902	0.3505	0.4435	53.3571	76.9903	0.749
14	GENE272	4661051	0.789	0.642	0.8136	53.3571	76.9886	0.7438
15	GENE149	4660928	0.7673	0.342	0.4457	53.3571	77.0805	0.7338
16	GENE264	4661043	0.7441	0.4767	0.6406	53.3571	77.42	0.7246
17	GENE62	4660841	0.7307	0.6164	0.8436	53.3571	77.7225	0.715
18	GENE953	4661732	0.7302	0.5165	0.7074	53.3571	77.7363	0.707
19	GENE829	4661608	0.7249	0.455	0.6276	53.3571	77.875	0.7037

1000 genes

Significance Filtering

Target False Discovery Rate (FDR): Delta =

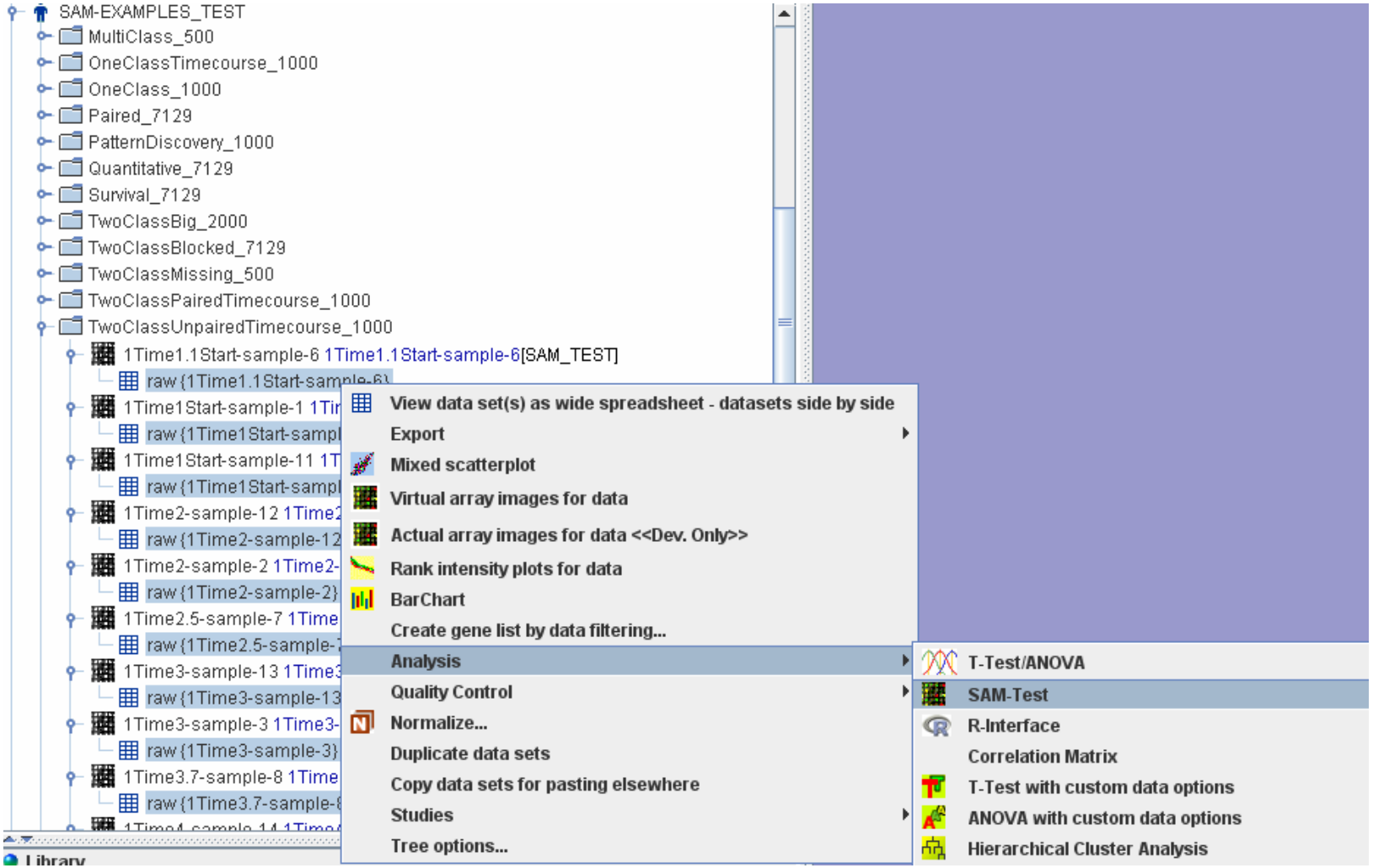
Select # genes by largest SAM absolute score

Mean Channel Intensities > Bad Flags <=

Apply Filters Clear Filters

Create Sig. Gene List HCA PCA K-Means SAM Plot

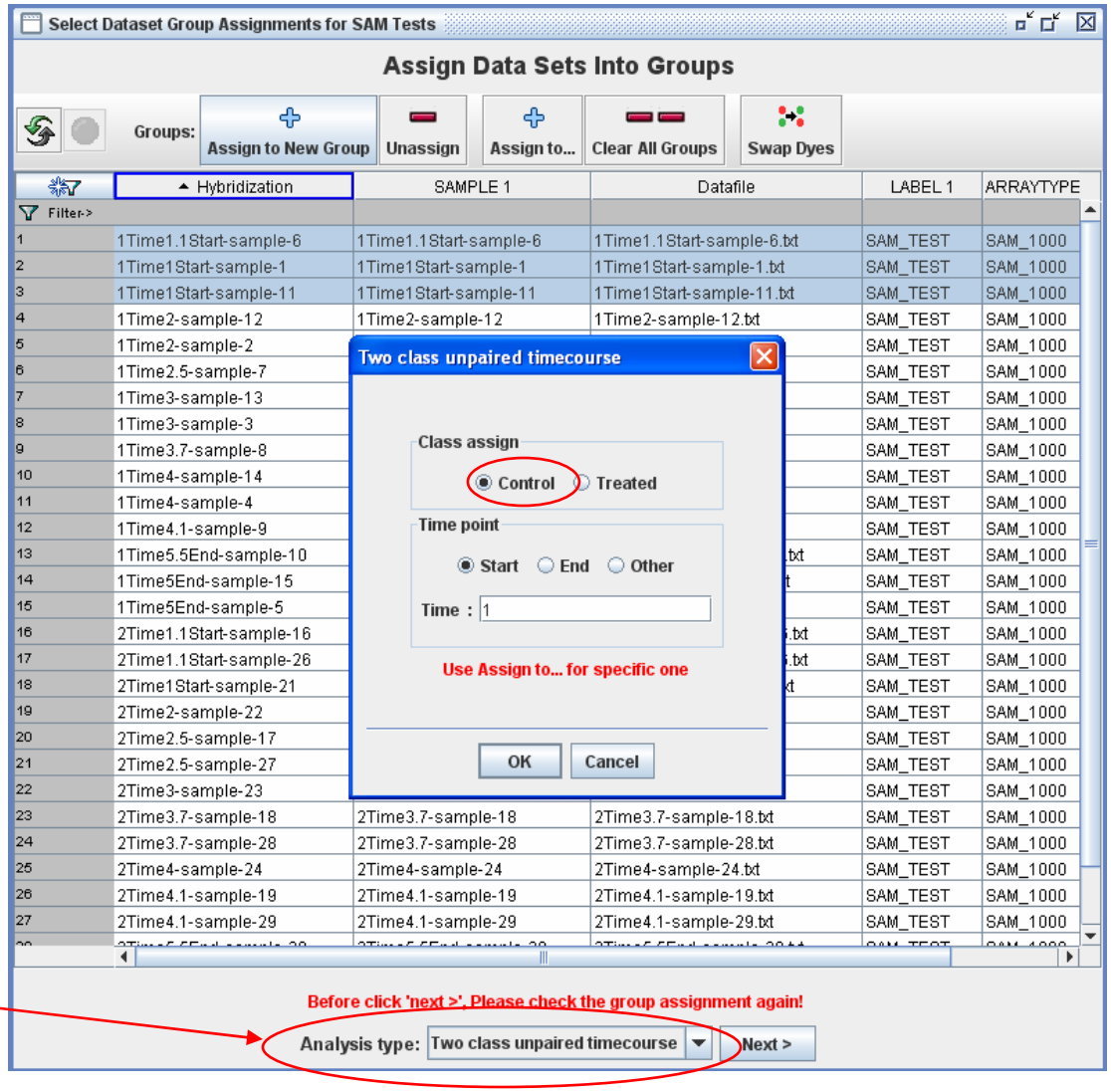
SAM-Test: Two Class Unpaired Timecourse



Right-click the selected data set, choose “Analysis” ->SAM-Test.

SAM-Test: Two Class Unpaired Timecourse

Here the steps are similar as one class time course. User assign start time point, middle point and end time point for class one. And then repeat the steps for class two. Just remember if class one is control group, then class two will be treated group.

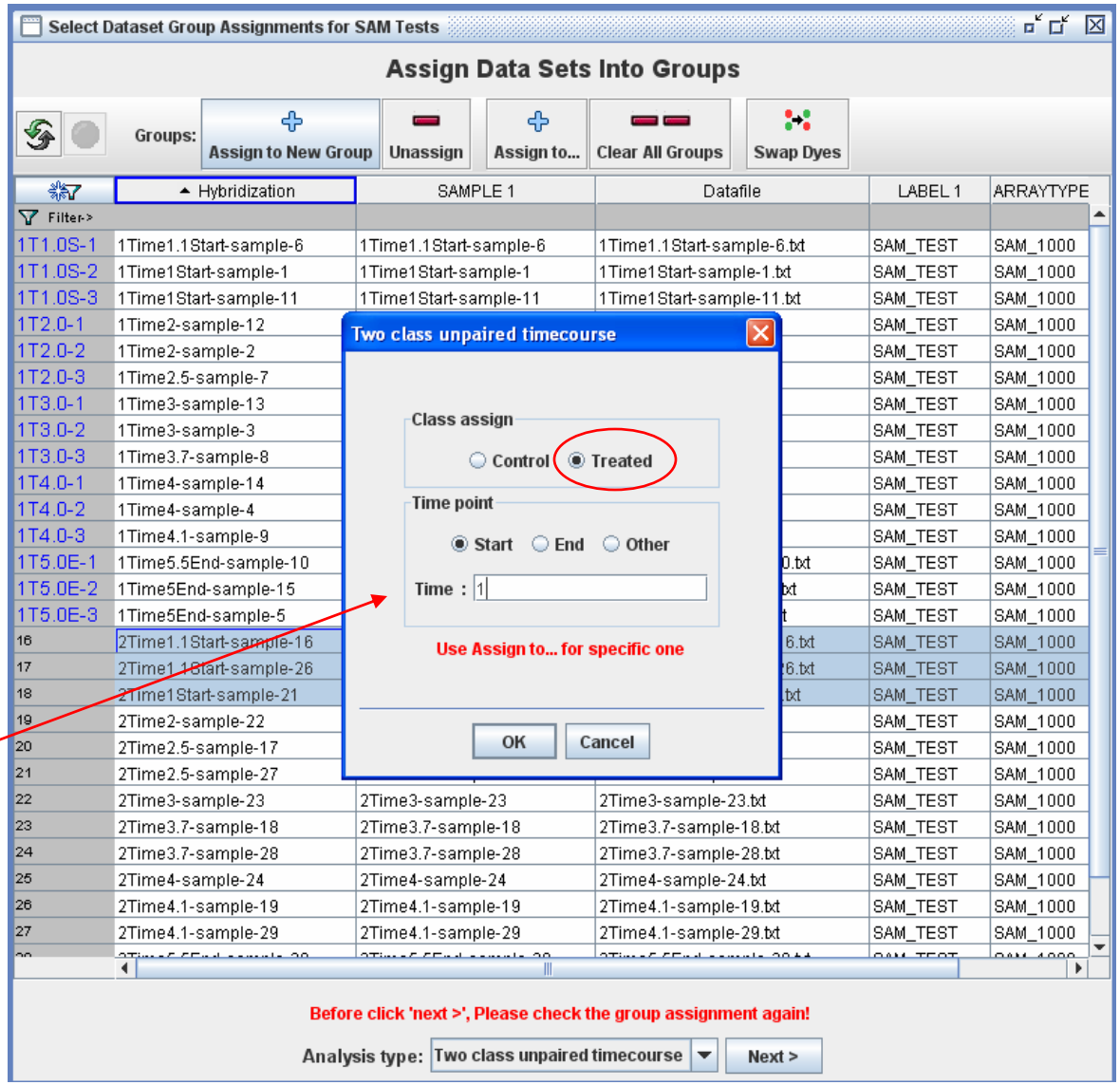


Make sure the analysis type is Selected before assigning class.

SAM-Test: Two Class Unpaired Timecourse

Class one from start to end time point.

Repeat the steps to assign class two. Class two is treated.



SAM-Test: Two Class Unpaired Timecourse

Here is what looks like after assigning.

Class one

Class two

Select Dataset Group Assignments for SAM Tests

Assign Data Sets Into Groups

Groups: Assign to New Group | Unassign | Assign to... | Clear All Groups | Swap Dyes

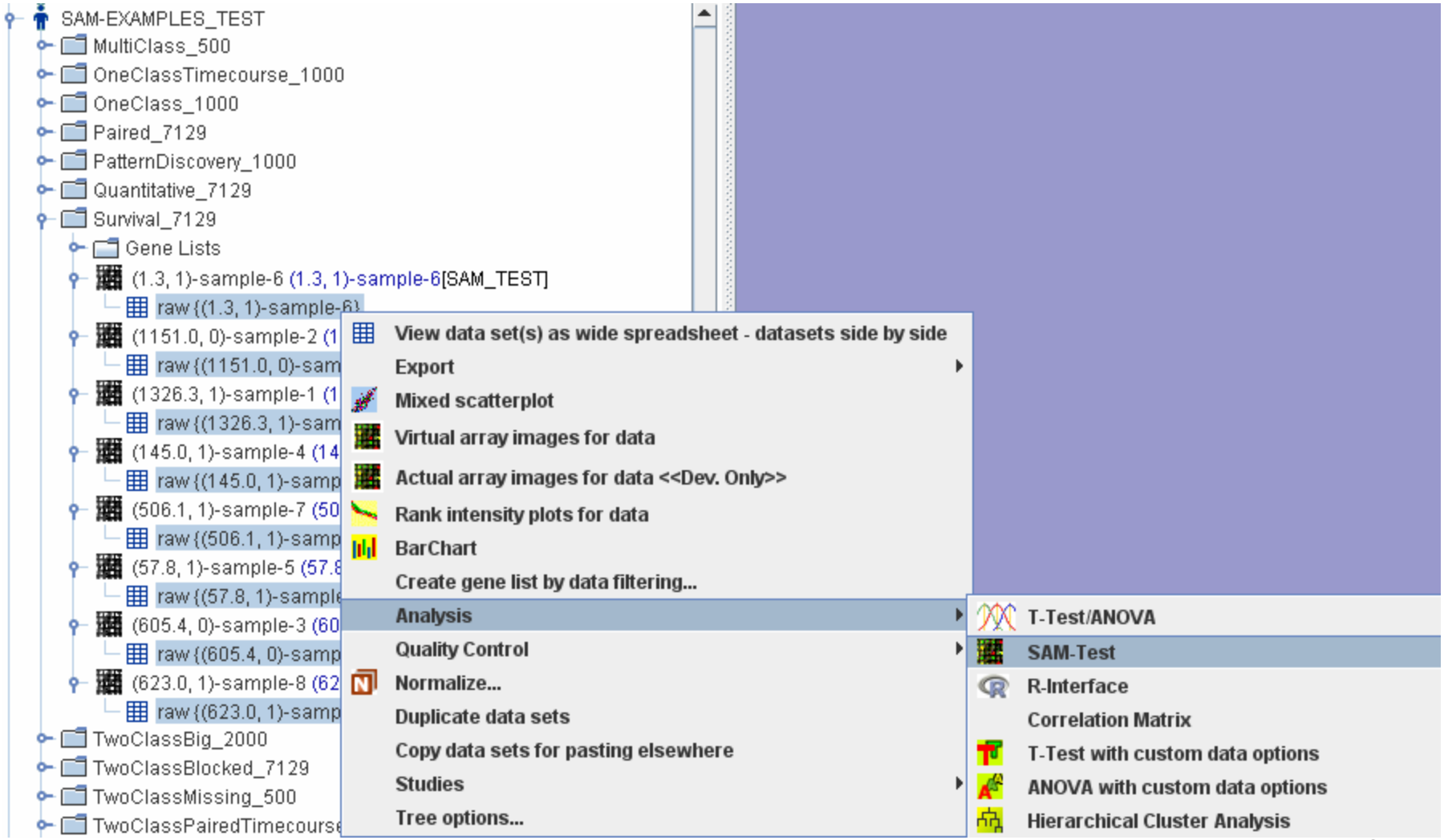
	Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPENAME	ASSAY 1
Filter->						
1T1.0S-1	1Time1.1Start-sample-6	1Time1.1Start-sample-6	1Time1.1Start-sample-6.txt	SAM_TEST	SAM_1000	In Vivo
1T1.0S-2	1Time1Start-sample-1	1Time1Start-sample-1	1Time1Start-sample-1.txt	SAM_TEST	SAM_1000	In Vivo
1T1.0S-3	1Time1Start-sample-11	1Time1Start-sample-11	1Time1Start-sample-11.txt	SAM_TEST	SAM_1000	In Vivo
1T2.0-1	1Time2-sample-12	1Time2-sample-12	1Time2-sample-12.txt	SAM_TEST	SAM_1000	In Vivo
1T2.0-2	1Time2-sample-2	1Time2-sample-2	1Time2-sample-2.txt	SAM_TEST	SAM_1000	In Vivo
1T2.0-3	1Time2.5-sample-7	1Time2.5-sample-7	1Time2.5-sample-7.txt	SAM_TEST	SAM_1000	In Vivo
1T3.0-1	1Time3-sample-13	1Time3-sample-13	1Time3-sample-13.txt	SAM_TEST	SAM_1000	In Vivo
1T3.0-2	1Time3-sample-3	1Time3-sample-3	1Time3-sample-3.txt	SAM_TEST	SAM_1000	In Vivo
1T3.0-3	1Time3.7-sample-8	1Time3.7-sample-8	1Time3.7-sample-8.txt	SAM_TEST	SAM_1000	In Vivo
1T4.0-1	1Time4-sample-14	1Time4-sample-14	1Time4-sample-14.txt	SAM_TEST	SAM_1000	In Vivo
1T4.0-2	1Time4-sample-4	1Time4-sample-4	1Time4-sample-4.txt	SAM_TEST	SAM_1000	In Vivo
1T4.0-3	1Time4.1-sample-9	1Time4.1-sample-9	1Time4.1-sample-9.txt	SAM_TEST	SAM_1000	In Vivo
1T5.0E-1	1Time5.5End-sample-10	1Time5.5End-sample-10	1Time5.5End-sample-10.txt	SAM_TEST	SAM_1000	In Vivo
1T5.0E-2	1Time5End-sample-15	1Time5End-sample-15	1Time5End-sample-15.txt	SAM_TEST	SAM_1000	In Vivo
1T5.0E-3	1Time5End-sample-5	1Time5End-sample-5	1Time5End-sample-5.txt	SAM_TEST	SAM_1000	In Vivo
2T1.0S-1	2Time1.1Start-sample-16	2Time1.1Start-sample-16	2Time1.1Start-sample-16.txt	SAM_TEST	SAM_1000	In Vivo
2T1.0S-2	2Time1.1Start-sample-26	2Time1.1Start-sample-26	2Time1.1Start-sample-26.txt	SAM_TEST	SAM_1000	In Vivo
2T1.0S-3	2Time1Start-sample-21	2Time1Start-sample-21	2Time1Start-sample-21.txt	SAM_TEST	SAM_1000	In Vivo
2T2.0-1	2Time2-sample-22	2Time2-sample-22	2Time2-sample-22.txt	SAM_TEST	SAM_1000	In Vivo
2T2.0-2	2Time2.5-sample-17	2Time2.5-sample-17	2Time2.5-sample-17.txt	SAM_TEST	SAM_1000	In Vivo
2T2.0-3	2Time2.5-sample-27	2Time2.5-sample-27	2Time2.5-sample-27.txt	SAM_TEST	SAM_1000	In Vivo
2T3.0-1	2Time3-sample-23	2Time3-sample-23	2Time3-sample-23.txt	SAM_TEST	SAM_1000	In Vivo
2T3.0-2	2Time3.7-sample-18	2Time3.7-sample-18	2Time3.7-sample-18.txt	SAM_TEST	SAM_1000	In Vivo
2T3.0-3	2Time3.7-sample-28	2Time3.7-sample-28	2Time3.7-sample-28.txt	SAM_TEST	SAM_1000	In Vivo
2T4.0-1	2Time4-sample-24	2Time4-sample-24	2Time4-sample-24.txt	SAM_TEST	SAM_1000	In Vivo
2T4.0-2	2Time4.1-sample-19	2Time4.1-sample-19	2Time4.1-sample-19.txt	SAM_TEST	SAM_1000	In Vivo
2T4.0-3	2Time4.1-sample-29	2Time4.1-sample-29	2Time4.1-sample-29.txt	SAM_TEST	SAM_1000	In Vivo
2T5.0E-1	2Time5.5End-sample-20	2Time5.5End-sample-20	2Time5.5End-sample-20.txt	SAM_TEST	SAM_1000	In Vivo
2T5.0E-2	2Time5.5End-sample-30	2Time5.5End-sample-30	2Time5.5End-sample-30.txt	SAM_TEST	SAM_1000	In Vivo
2T5.0E-3	2Time5End-sample-25	2Time5End-sample-25	2Time5End-sample-25.txt	SAM_TEST	SAM_1000	In Vivo

Before click 'next >', Please check the group assignment again!

Analysis type: Two class unpaired timecourse | Next >

Click Next button to get SAM-test

SAM-Test: Survival

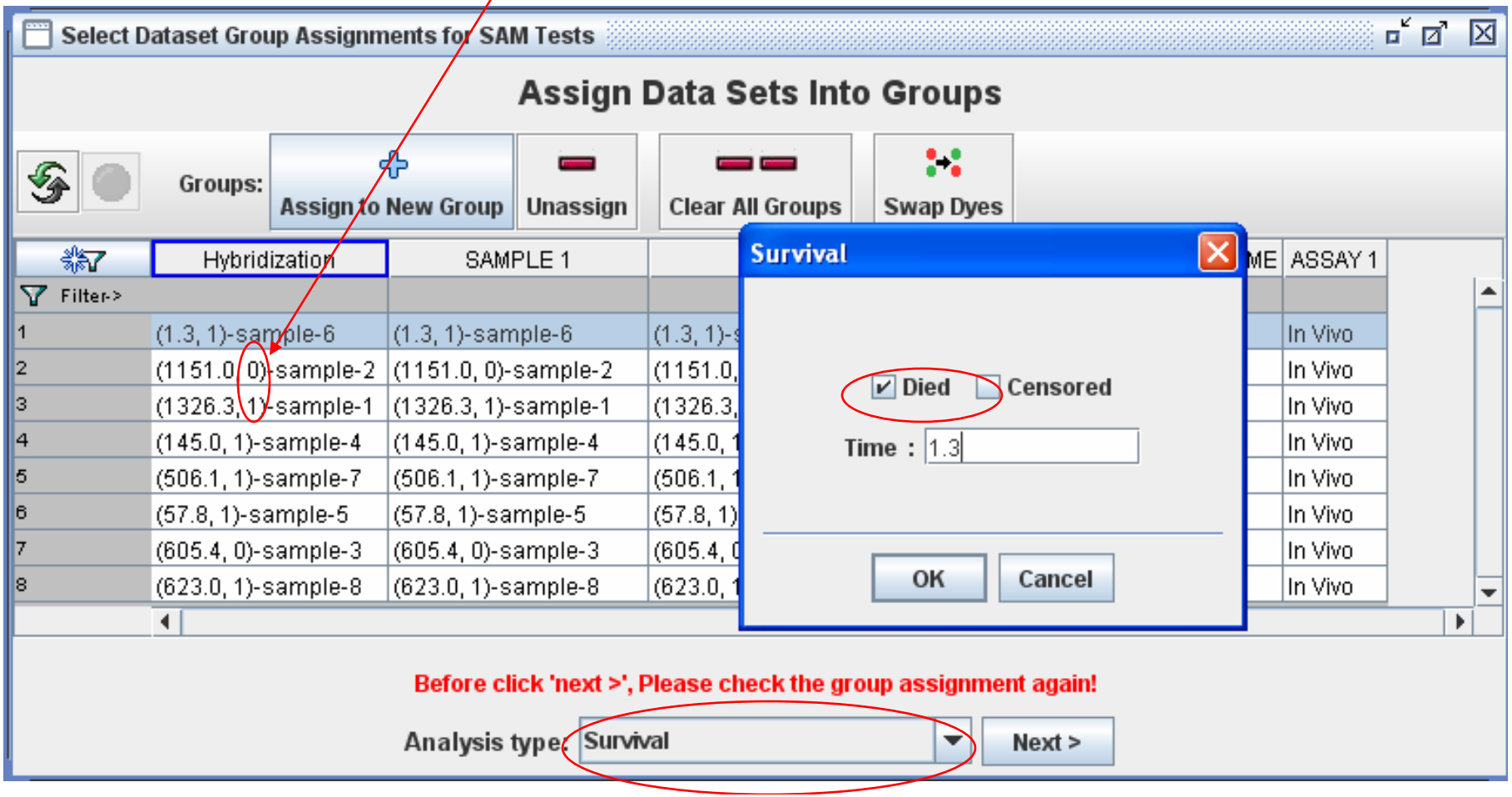


Right-click the selected data set, choose “Analysis” ->SAM-Test

SAM-Test: Survival

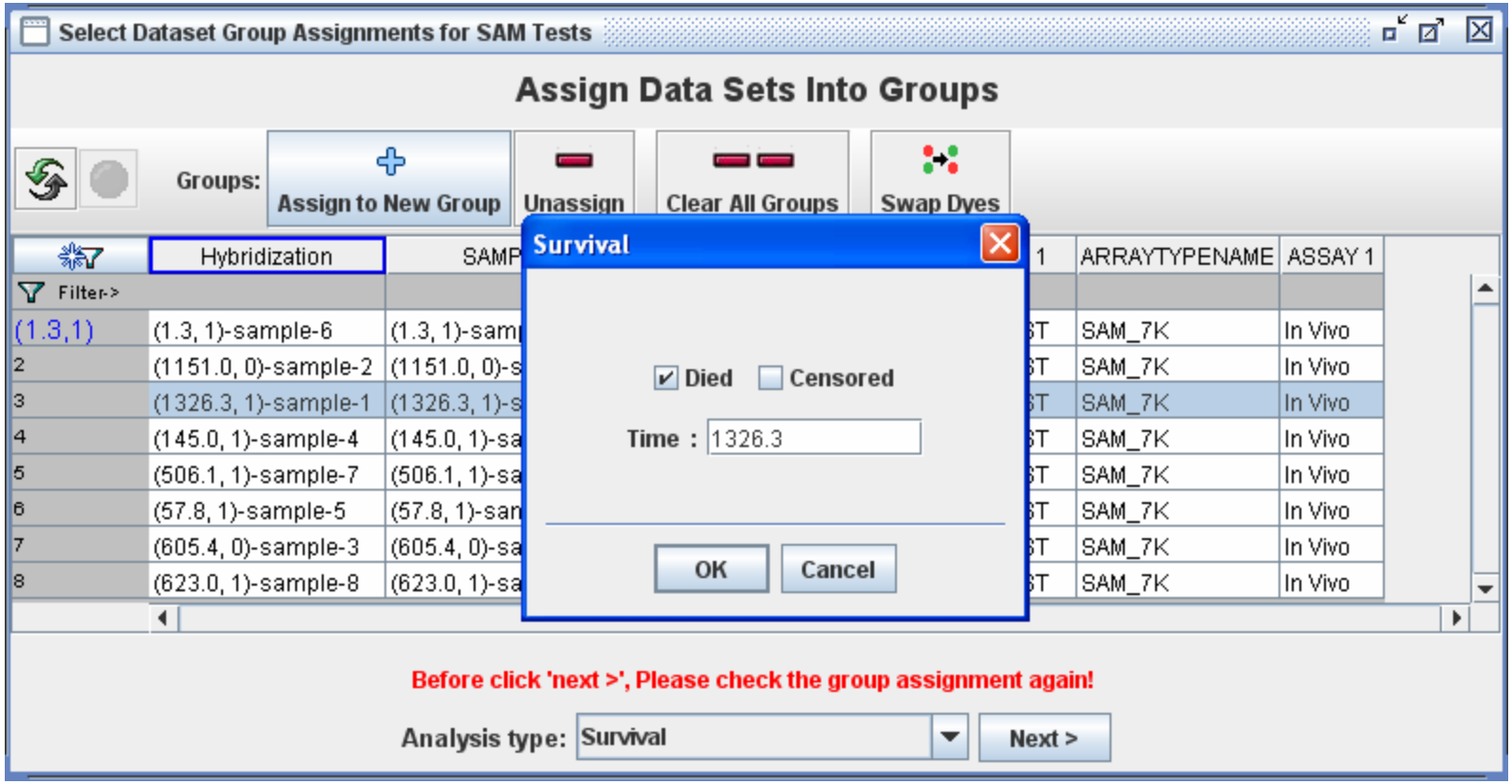
This data set has two groups. The first number in parenthesis represent time and the second number in parenthesis (1 or 0) represents “died” (1) or “censored”(0) group.

Select the first data – (1.3, 1)-sample-6, click “Assign to New Group” button. Type 1.3 in the time area.



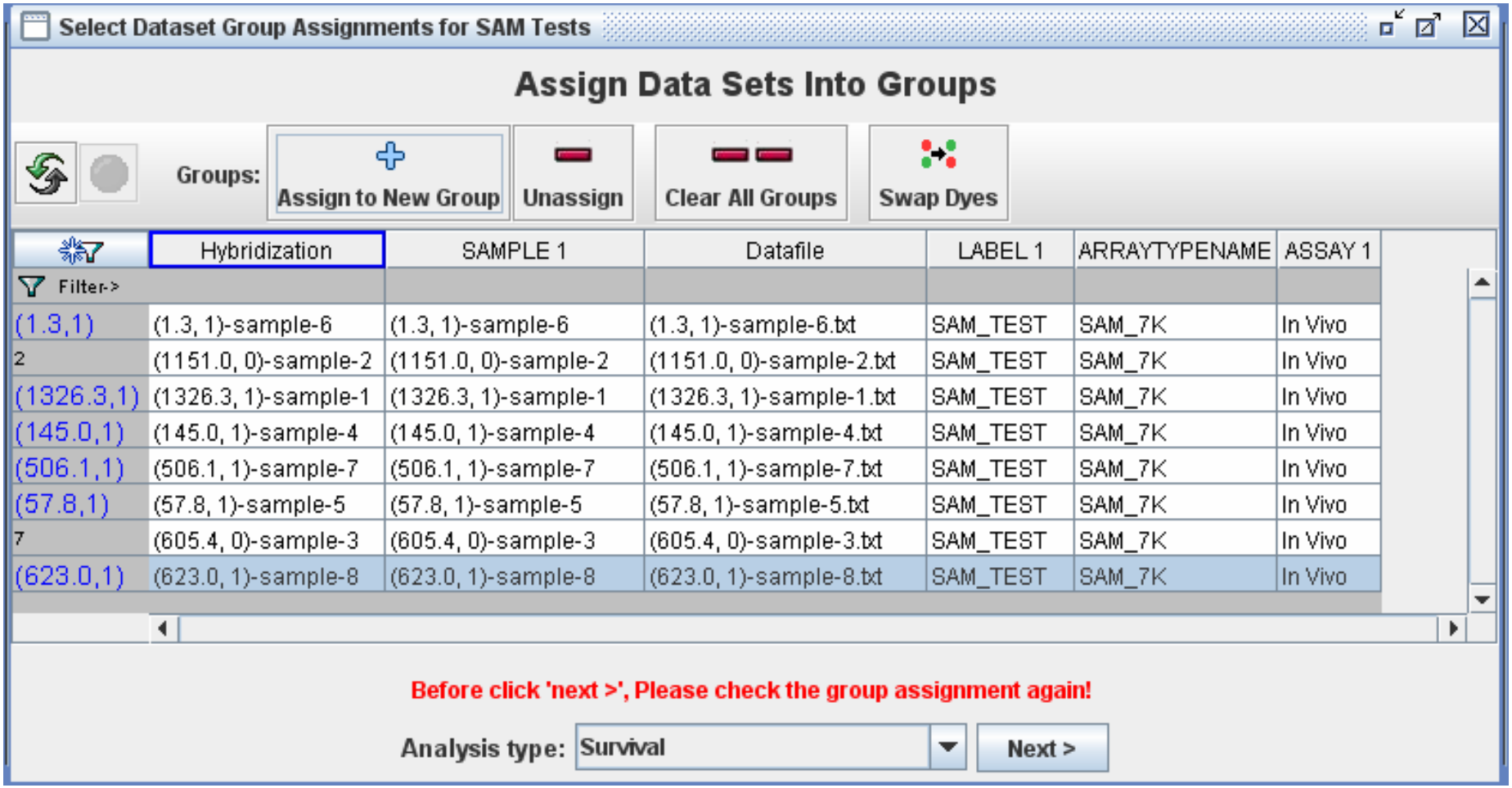
SAM-Test: Survival

Repeat the previous steps to assign the other data of group 1, type in the time.



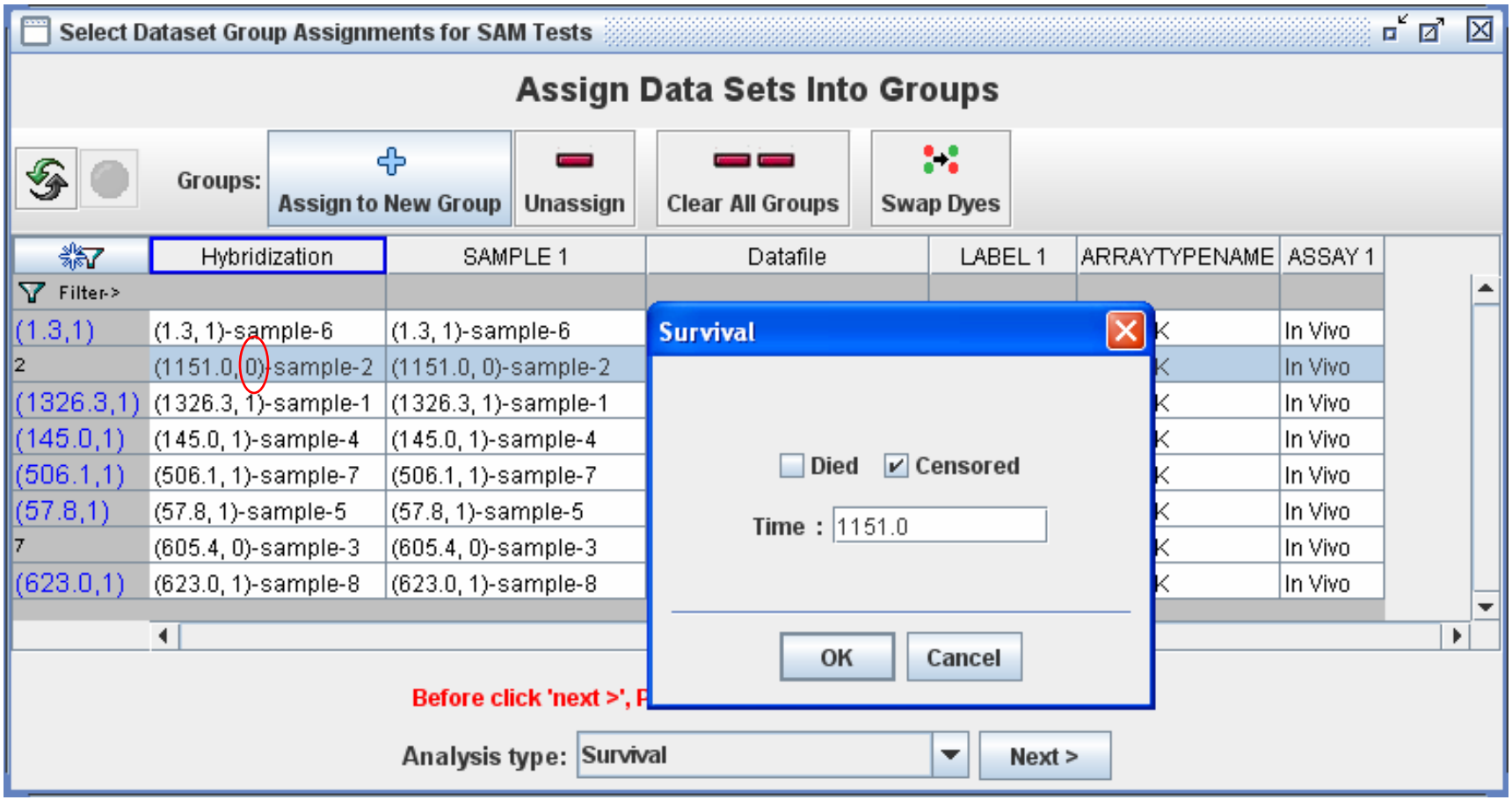
SAM-Test: Survival

This image shows all the data of group 1 has been assigned.



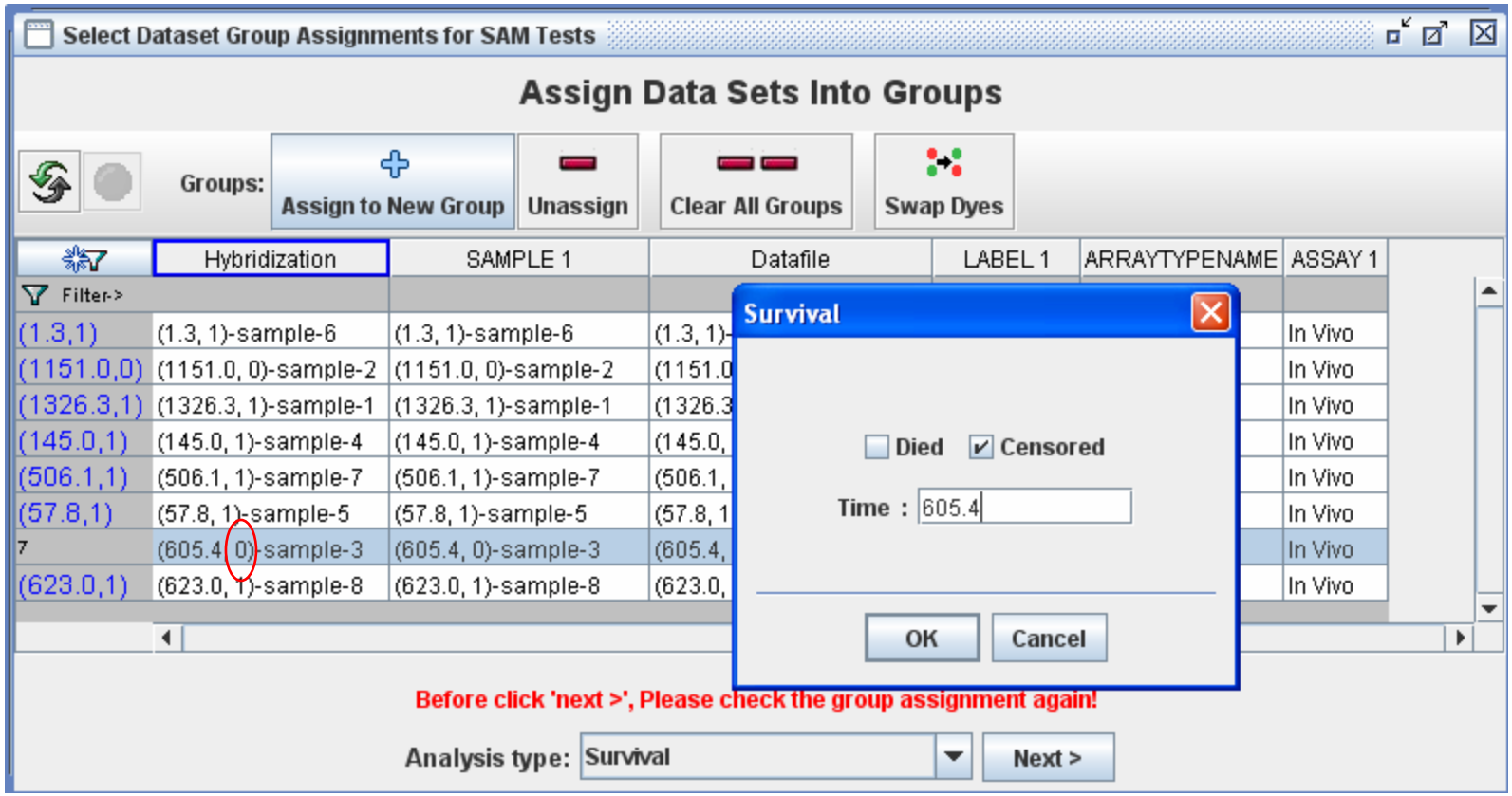
SAM-Test: Survival

Select the 2nd data which is in group 0, click “Assign to New Group” button. Check “Censored” check box and type in the time.



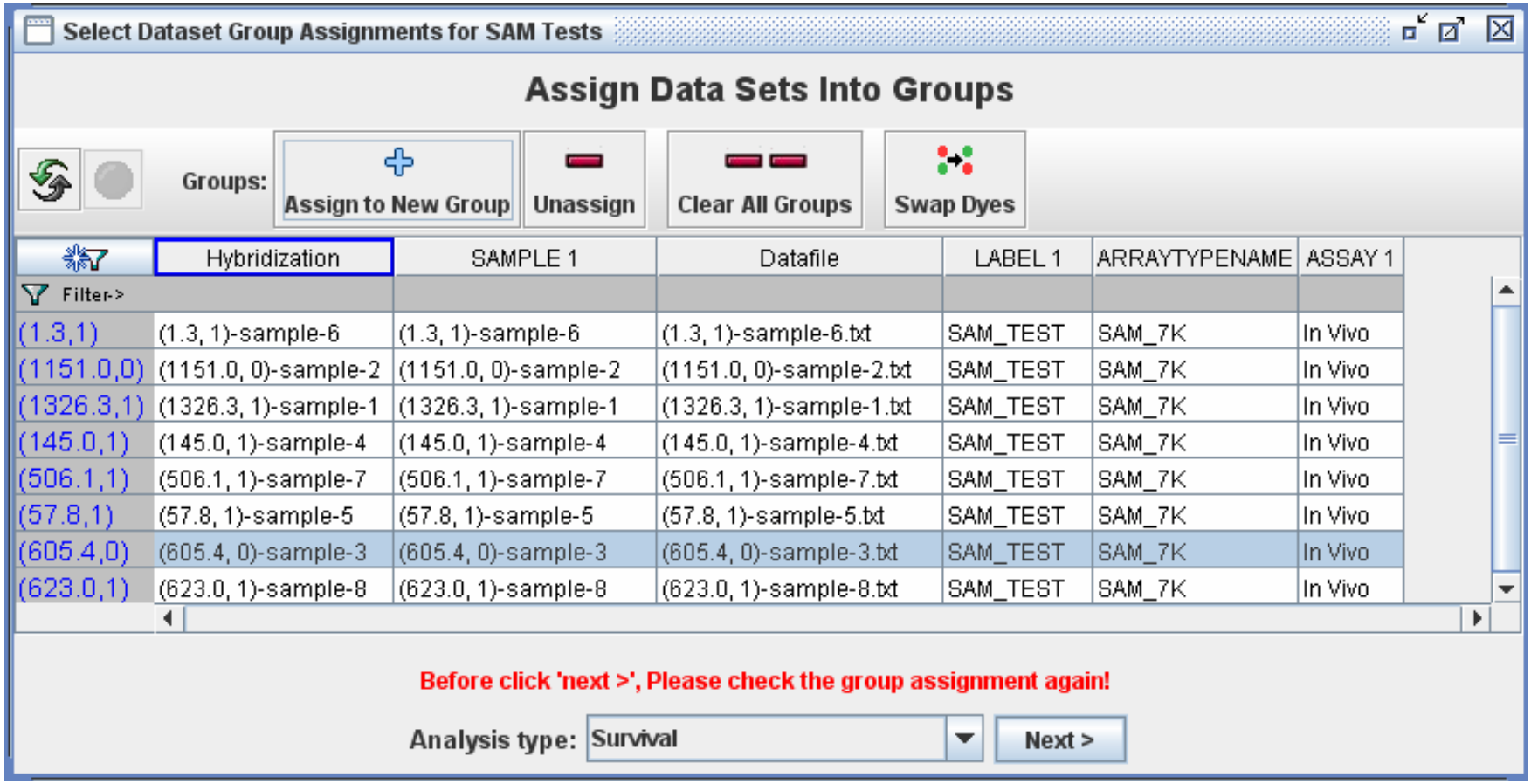
SAM-Test: Survival

Do the same steps for the other data of group 0.



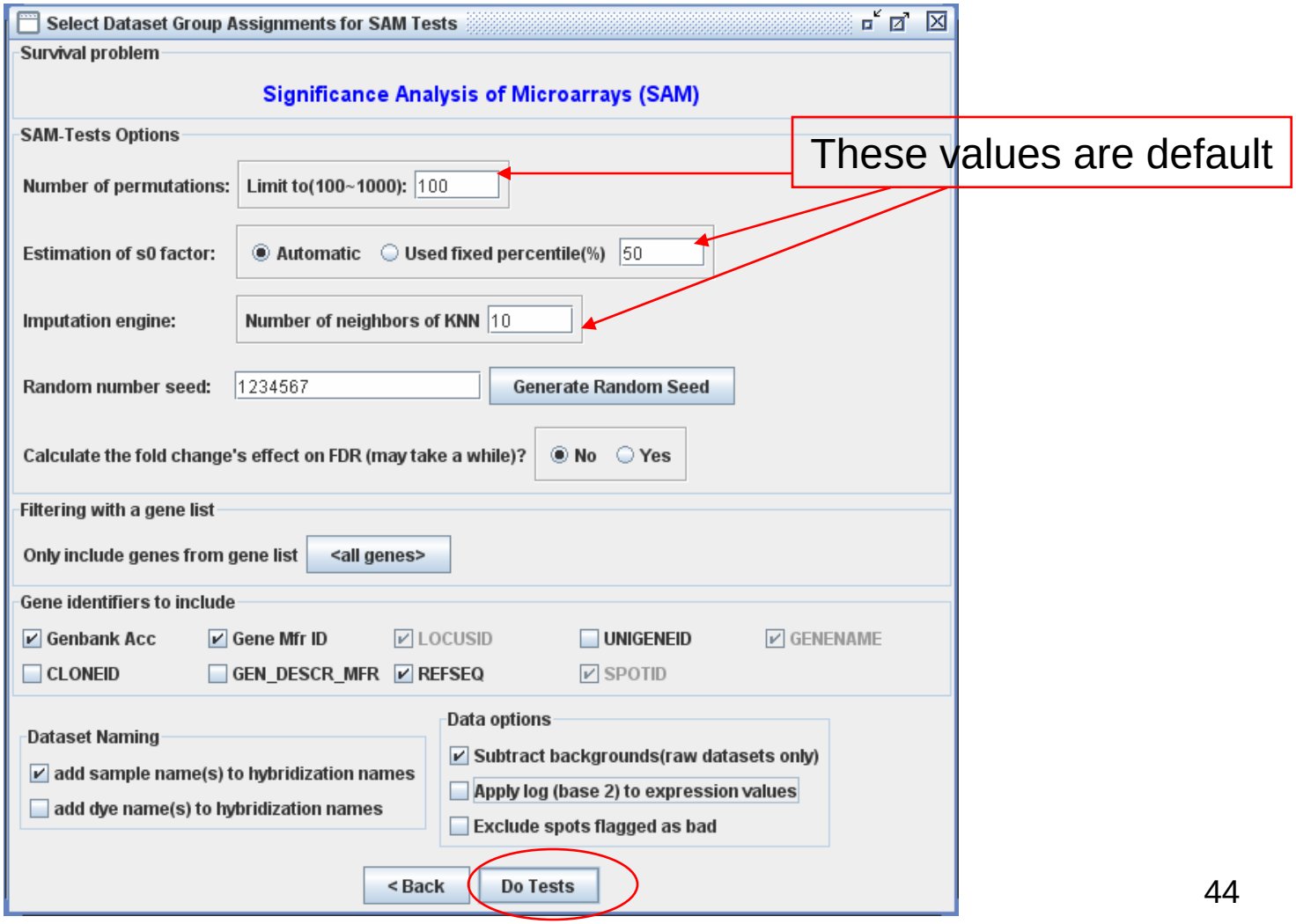
SAM-Test: Survival

Here is what looks like after assigning all the data. Click “Next” button.



	Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPENAME	ASSAY 1
(1.3,1)	(1.3, 1)-sample-6	(1.3, 1)-sample-6	(1.3, 1)-sample-6.txt	SAM_TEST	SAM_7K	In Vivo
(1151.0,0)	(1151.0, 0)-sample-2	(1151.0, 0)-sample-2	(1151.0, 0)-sample-2.txt	SAM_TEST	SAM_7K	In Vivo
(1326.3,1)	(1326.3, 1)-sample-1	(1326.3, 1)-sample-1	(1326.3, 1)-sample-1.txt	SAM_TEST	SAM_7K	In Vivo
(145.0,1)	(145.0, 1)-sample-4	(145.0, 1)-sample-4	(145.0, 1)-sample-4.txt	SAM_TEST	SAM_7K	In Vivo
(506.1,1)	(506.1, 1)-sample-7	(506.1, 1)-sample-7	(506.1, 1)-sample-7.txt	SAM_TEST	SAM_7K	In Vivo
(57.8,1)	(57.8, 1)-sample-5	(57.8, 1)-sample-5	(57.8, 1)-sample-5.txt	SAM_TEST	SAM_7K	In Vivo
(605.4,0)	(605.4, 0)-sample-3	(605.4, 0)-sample-3	(605.4, 0)-sample-3.txt	SAM_TEST	SAM_7K	In Vivo
(623.0,1)	(623.0, 1)-sample-8	(623.0, 1)-sample-8	(623.0, 1)-sample-8.txt	SAM_TEST	SAM_7K	In Vivo

SAM-Test: Survival



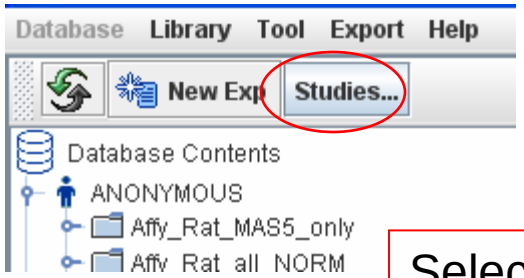
Array Track

Tutorial 13:
Simple Tox Format

SimpleTox for Toxicogenomics Study

- Used directly for data uploading for toxicogenomics experiment (non-clinical data and clinical data)

  http://www.cdisc.org/models/send/v1.4/index.html
- Rationale behind SimpleTox
  - To capture essential information related to microarray and toxicity study to attain sufficient information for cross-study analysis
  - To mimic the way of biologists managing/organizing data
  - To be compliant with MIAME and SEND
- Format
  - Excel spreadsheet format
  - Contains required and optional fields
  - Customization: Users be able to add new fields
  - Searchable

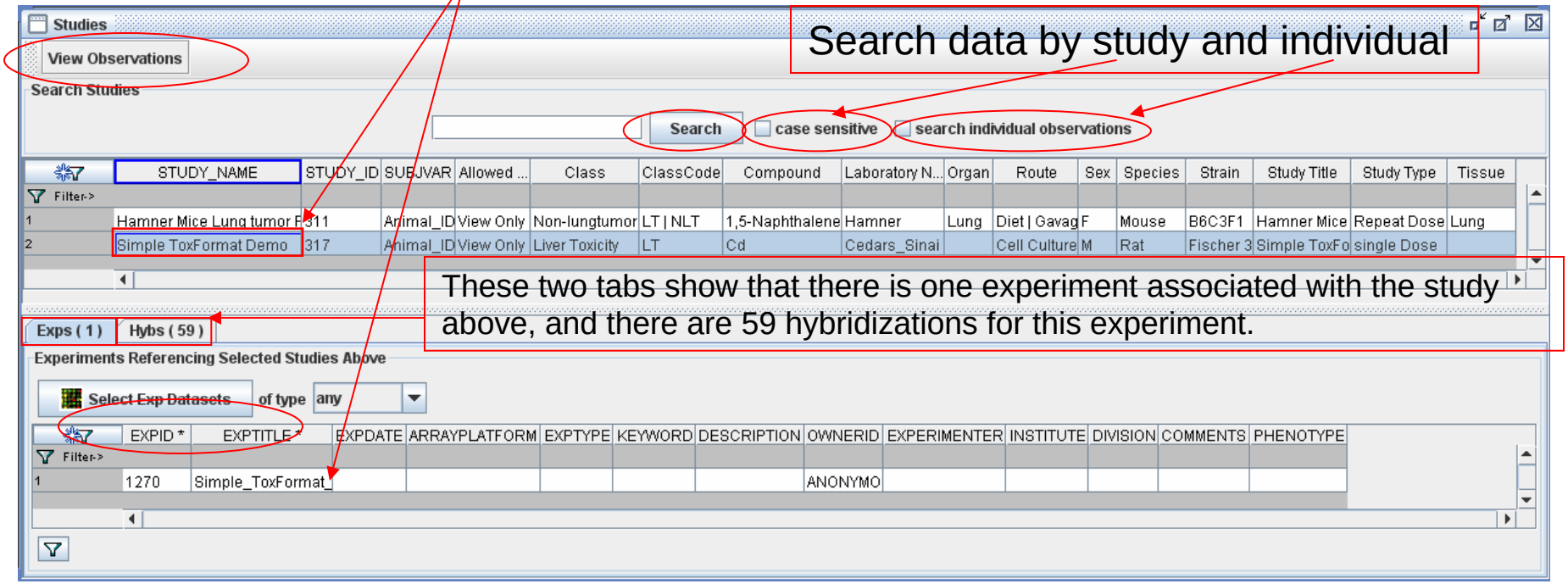


Click “Studies…” button to bring out the window below.

Select one study, the experiment that references the selected Study will be shown below

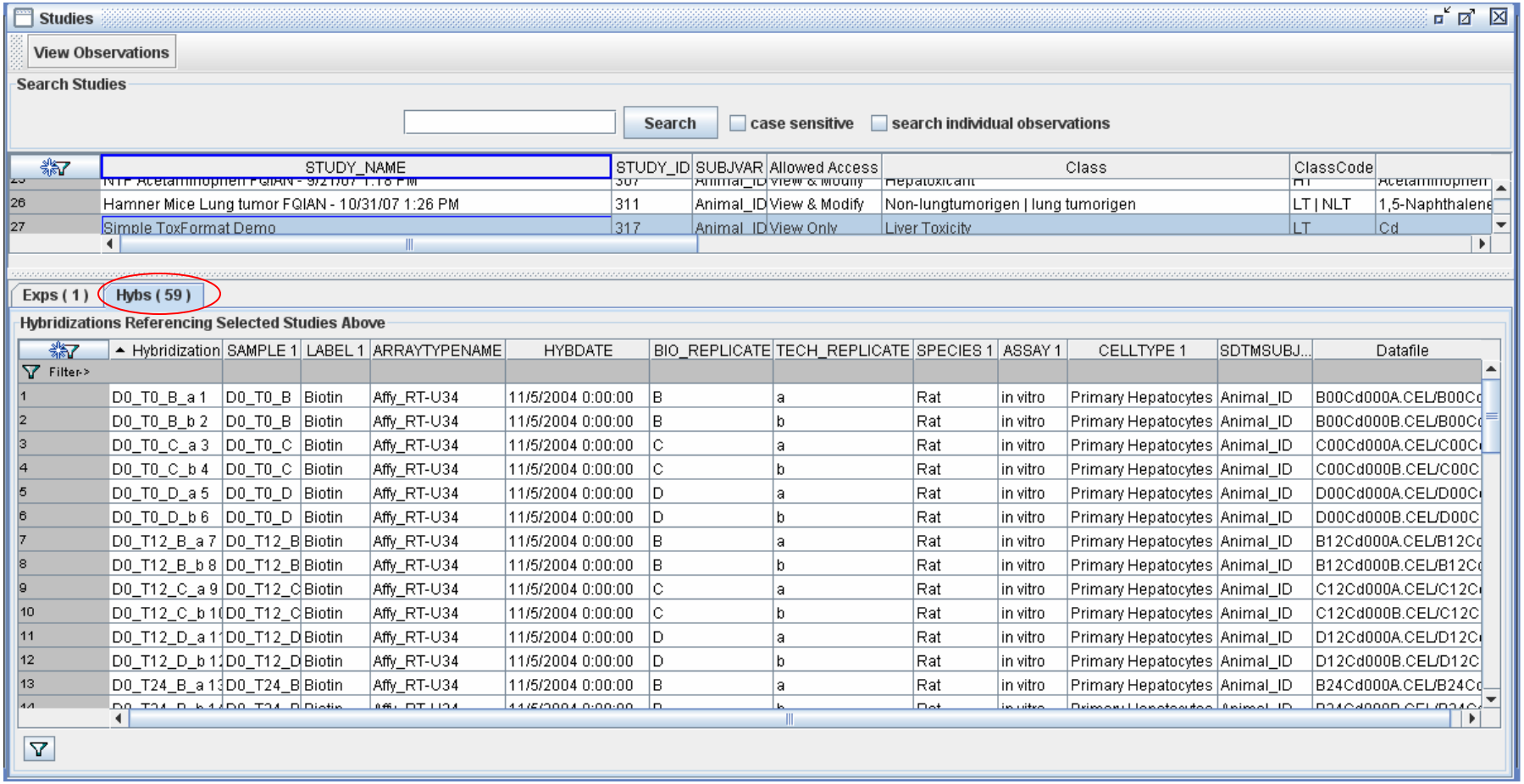
Search data by study and individual

These two tabs show that there is one experiment associated with the study above, and there are 59 hybridizations for this experiment.

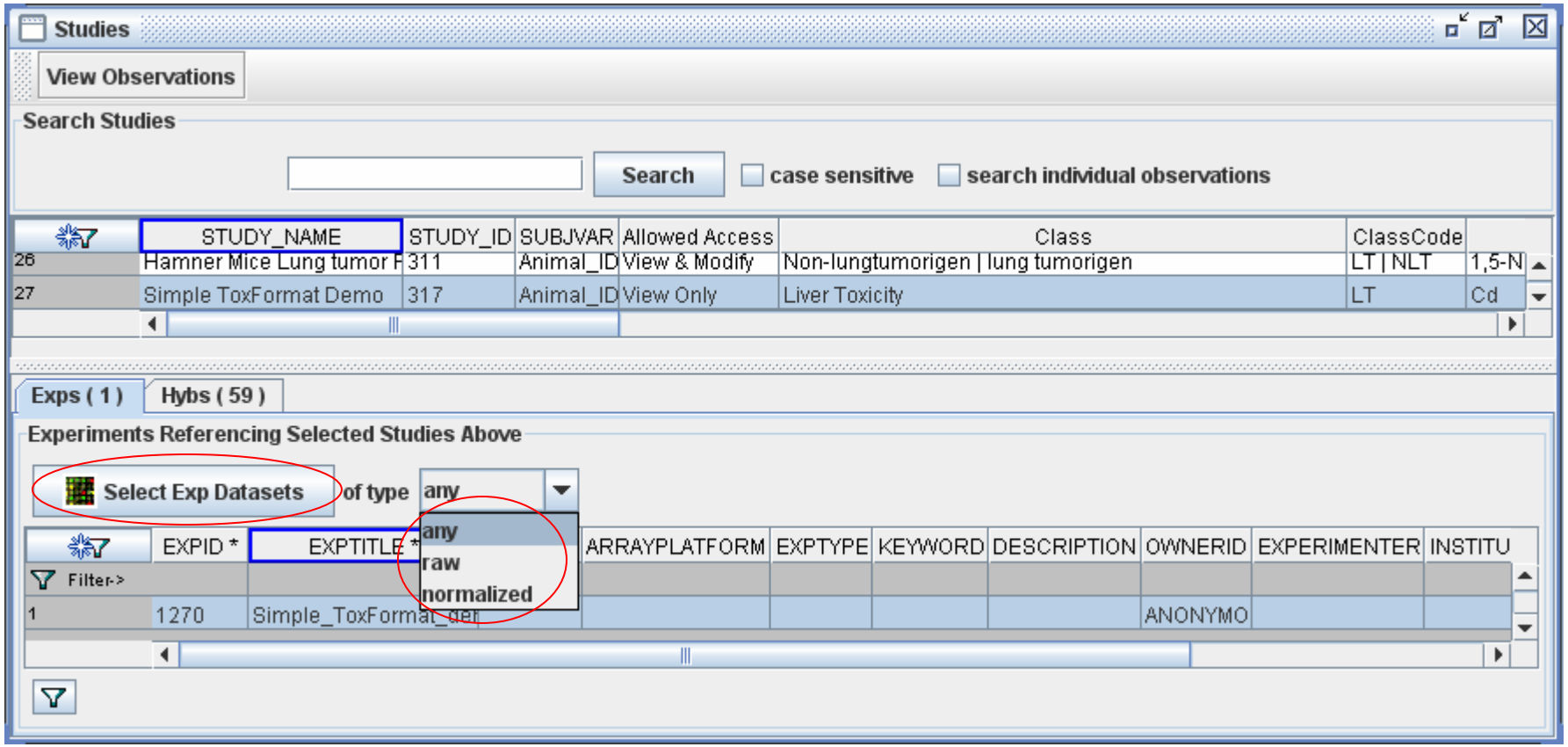


Select the study named “Simple_ToxFormat”, then click “View Observation”.

Click Hybs tab, the table below will display all the information about the hybridizations.



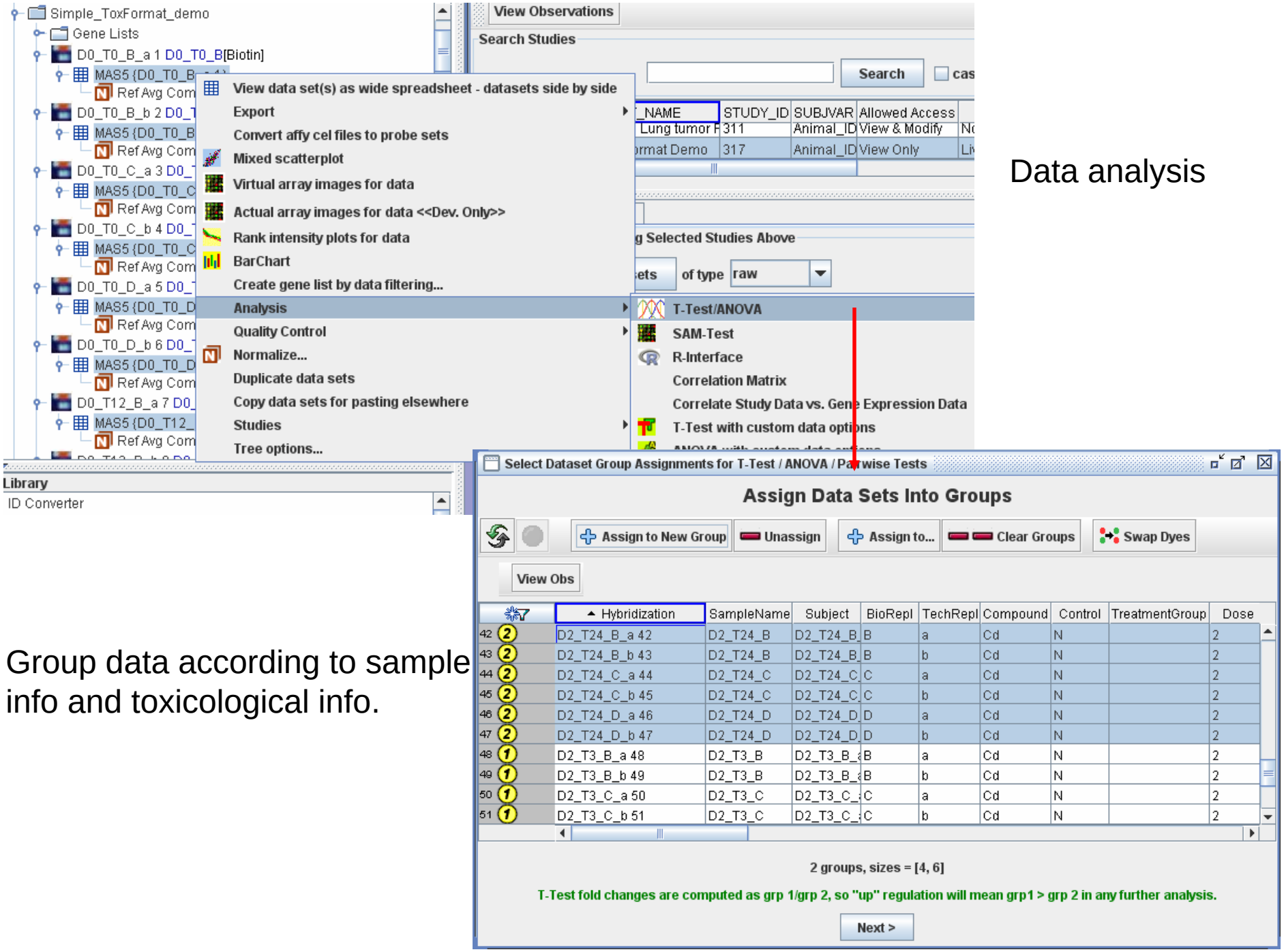
Highlight the exp and click “Select Exp Dataset” button will select all the data (or just raw/normalized data) in the data panel for data analysis purpose (see the screen shot in next slide)

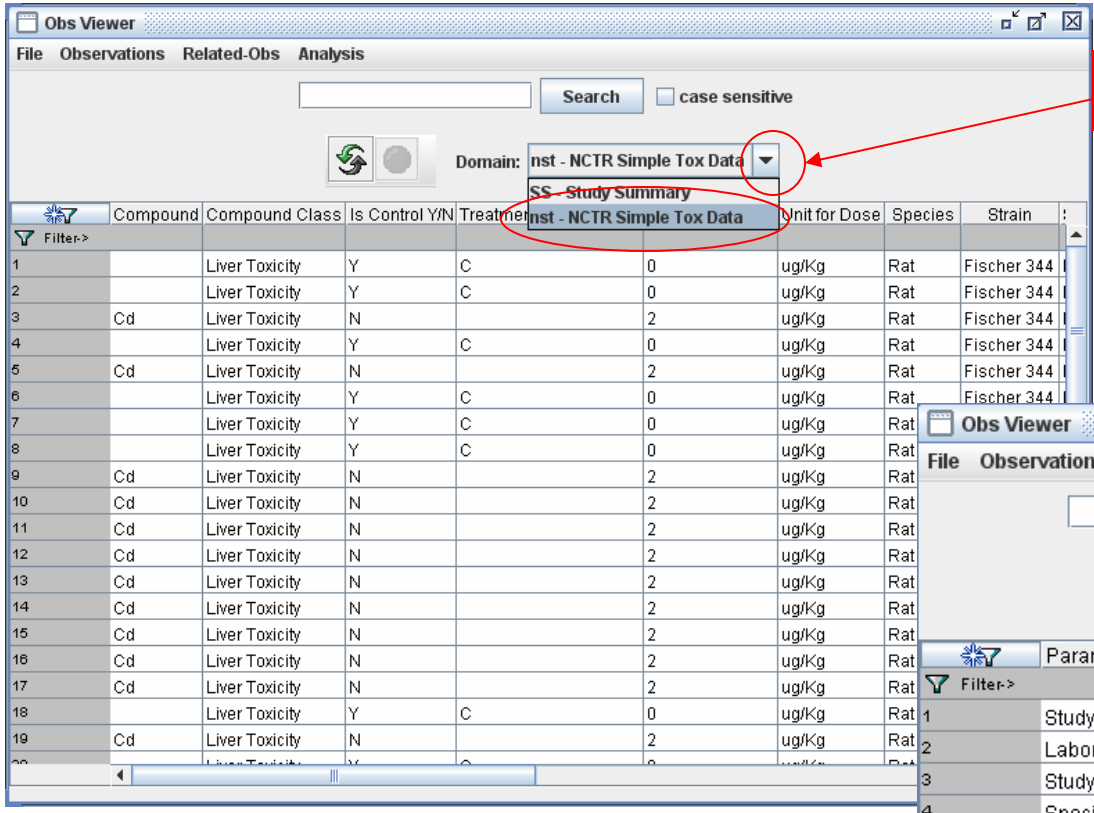


You can select the study named “Simple_ToxFormat”, then click “View Observation” to view the detail of the observation including sample info and toxicological info. See the observation viewer in slide 6.

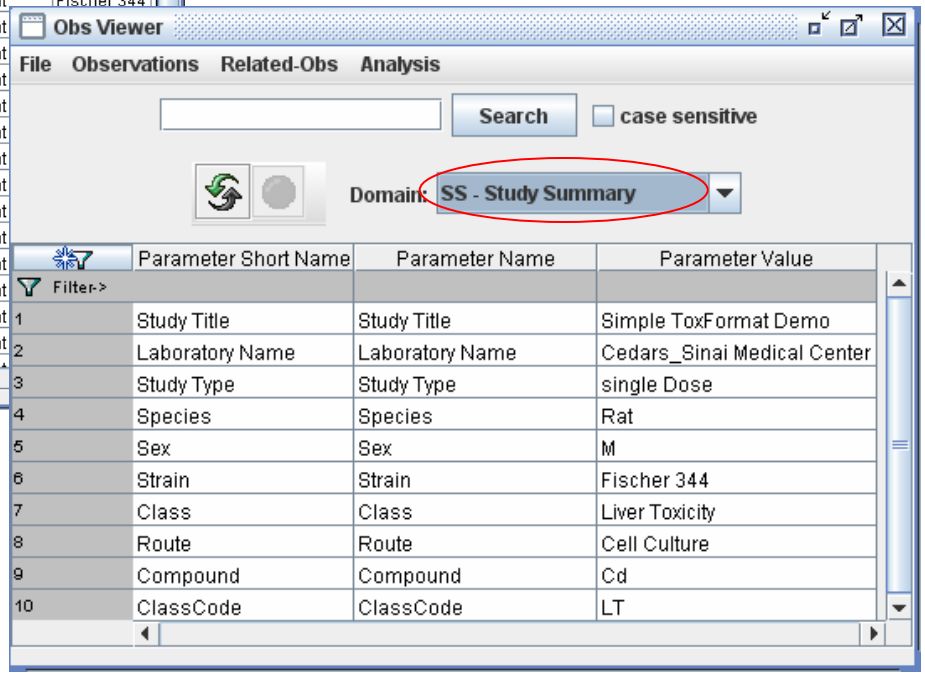
Data analysis

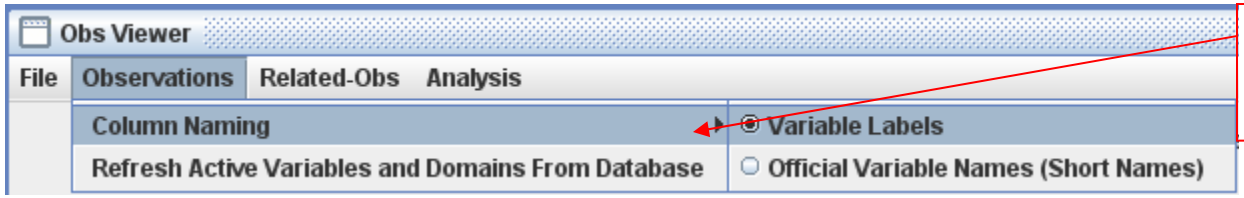
Group data according to sample info and toxicological info.





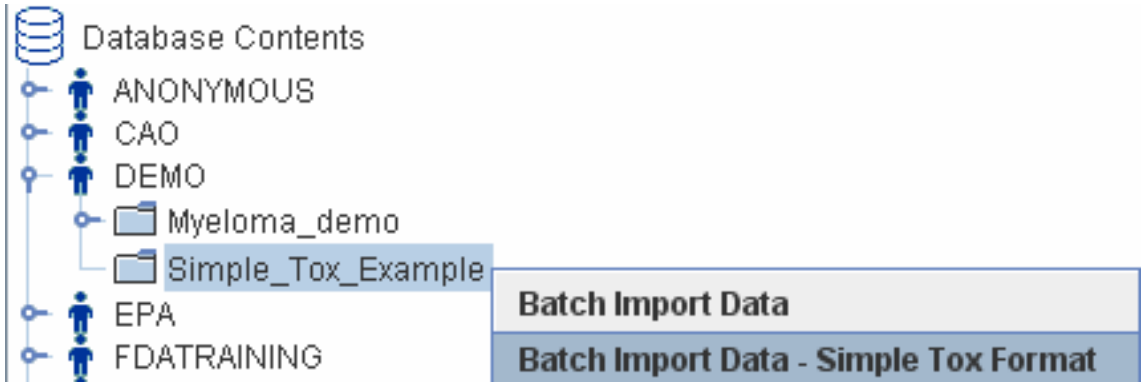
Switch to study summary domain



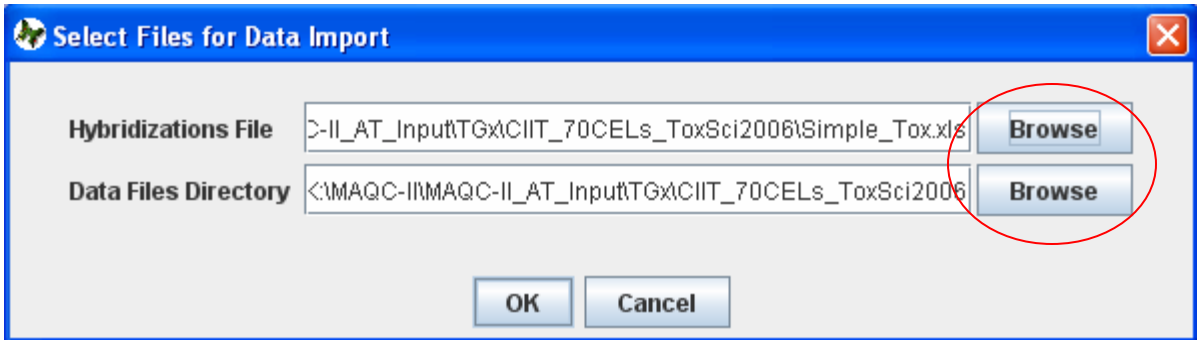


User can also change the display of column naming

Batch Import – Simple Tox Format



Right-click the experiment, choose “Batch Import Data – Simple Tox Format”



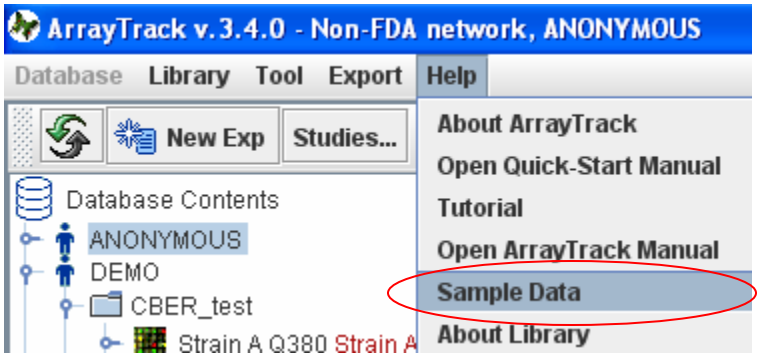
Click Browse button to locate the hybridization file and data files directory.
The next slide shows an example of hybridization table file.

Batch Import – Simple Tox Format

This is an example of hybridization table file (which can be downloaded from ArrayTrack help pull-down menu.). User can use this as a template to make his own hybridization file. Just make sure that your column titles are exactly same as this template. You are allowed to add new columns, however formatted columns Designed for toxicogenomics study will benefit for future across study analysis.

SimpleTox_Hamner_NoArraydata

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Array_ID	SubjectID	Institution	StudyTitle	StudyType	CompoundClass	ClassReference	Compound	CAS	Control	Treatmen	Dose	DoseUnit	HybName	SampleName
2	792	1-22	Hamner	Hamner Mice Lu	Repeat Dose			Corn Oil		Y	C	0	mg/kg	RT_0_13_22	RT_0_13_22_Lung_CornOil
3	793	1-24	Hamner	Hamner Mice Lu	Repeat Dose			Corn Oil		Y	C	0	mg/kg	RT_0_13_24	RT_0_13_24_Lung_CornOil
4	794	1-25	Hamner	Hamner Mice Lu	Repeat Dose			Corn Oil		Y	C	0	mg/kg	RT_0_13_25	RT_0_13_25_Lung_CornOil
5	795	1-27	Hamner	Hamner Mice Lu	Repeat Dose			Rodent Chow		Y	C	0	ppm	RT_0_13_27	RT_0_13_27_Lung_RodentCho
6	796	1-28	Hamner	Hamner Mice Lu	Repeat Dose			Rodent Chow		Y	C	0	ppm	RT_0_13_28	RT_0_13_28_Lung_RodentCho
7	797	1-29	Hamner	Hamner Mice Lu	Repeat Dose			Rodent Chow		Y	C	0	ppm	RT_0_13_29	RT_0_13_29_Lung_RodentCho
8	798	1-12	Hamner	Hamner Mice Lu	Repeat Dose	Non-lungtumorigen	NTP_No168	N-(1-naphthyl)et	1465-25-4	N	NLT	2000	ppm	RT_2000_13	RT_2000_13_12_Lung_ethylen
9	799	1-14	Hamner	Hamner Mice Lu	Repeat Dose	Non-lungtumorigen	NTP_No168	N-(1-naphthyl)et	1465-25-4	N	NLT	2000	ppm	RT_2000_13	RT_2000_13_14_Lung_ethylen
10	800	1-15	Hamner	Hamner Mice Lu	Repeat Dose	Non-lungtumorigen	NTP_No168	N-(1-naphthyl)et	1465-25-4	N	NLT	2000	ppm	RT_2000_13	RT_2000_13_15_Lung_ethylen
11	801	1-2	Hamner	Hamner Mice Lu	Repeat Dose	lung tumorigen	NTP_No143	1,5-Naphthalene	2243-62-1	N	LT	2000	ppm	RT_2000_13	RT_2000_13_2_Lung_Naphthal



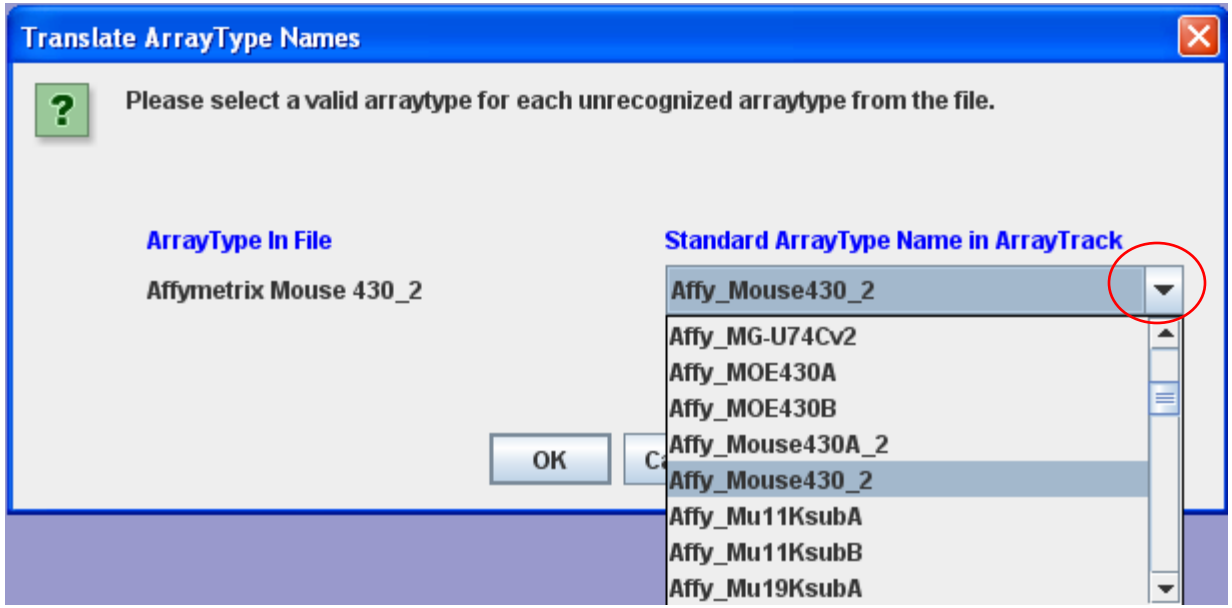
The next slide shows a table that explains the meaning of the column heads. The order for each column doesn't matter.

SimpleTox

SimpleTox Column Head	Description	Example
Institution	Laboratory or institution name	NCTR/FDA, EPA
DataFile	Microarray data file	GSM142129.CEL
HybName	User specified identifier for a hybridization name	APAP_D100_T6_Jun04; APAP_D0_T6_Jun04
SampleName	Sample name	APAP_Dose100_Time6; APAP_Dose0_Time6
Array_ID	User-specified identifier for a hybridization	1,2,3 … or p1002356
Label	RNA label reagent	Biotin; Cy3
ArrayType	Array type	Affymetrix Mouse 430_2
Subjiect_ID	Subject identifier	
StudyTitle	Study title	6 days repeating toxicity study
Tech_Rep	Technical replicates; microarray specific	A, B or C; 1,2, or 3
Bio_Rep	Biological replicates	A, B or C; 1,2, or 3
HybDate	Hybridization date	2/25/2007
StudyType	Study type	Single Dose Toxicity or Repeat Dose Toxicity
Compound	Compond name	Carbon Tetrachloride; Acetaminophen
……..	**……..**	**……..**

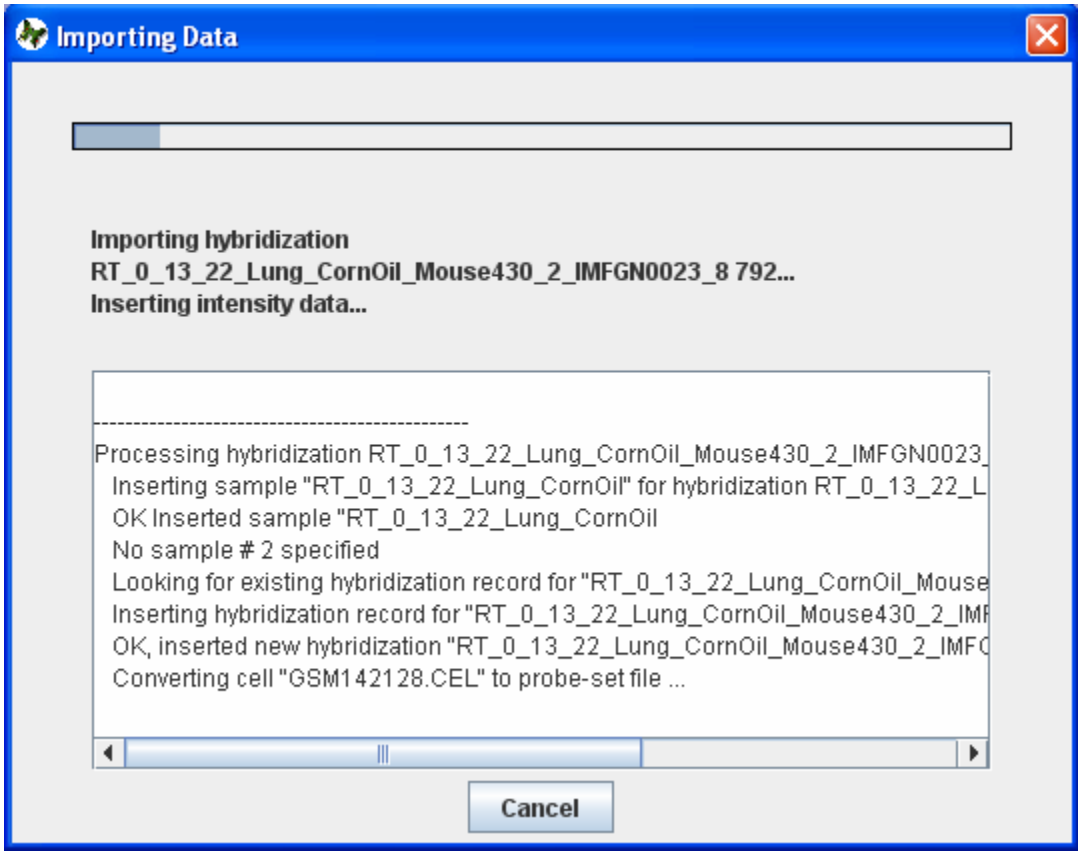
The red-colored part are required fields, while black-colored are optional fields. If you don't have Array_ID, you can use subject_ID as Array_ID.

Batch Import – Simple Tox Format

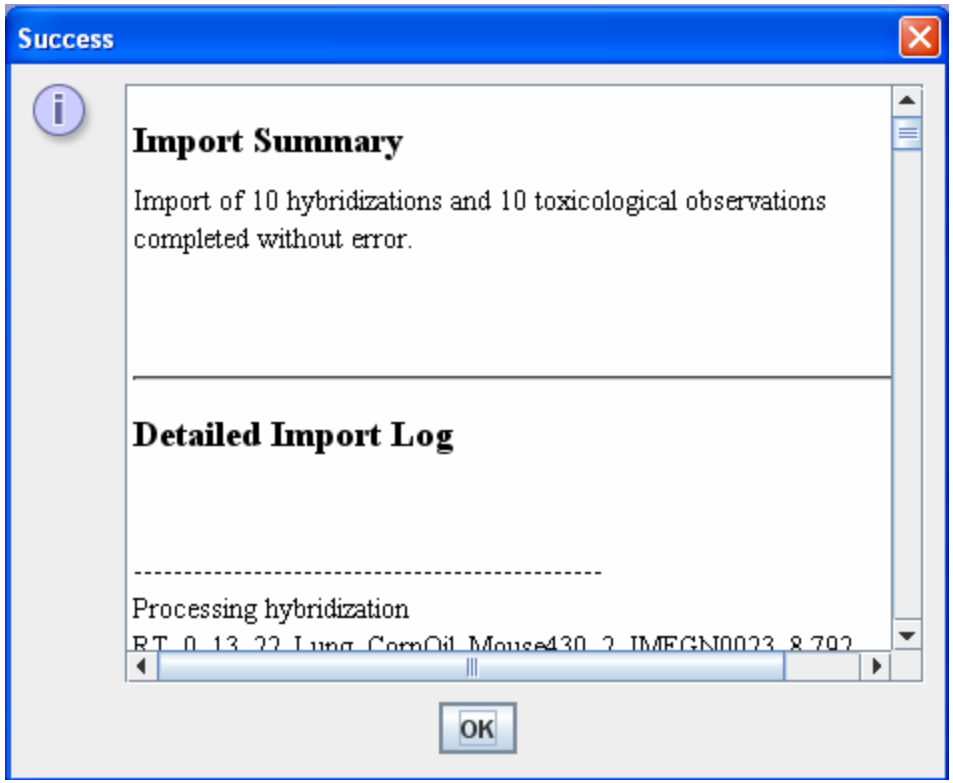


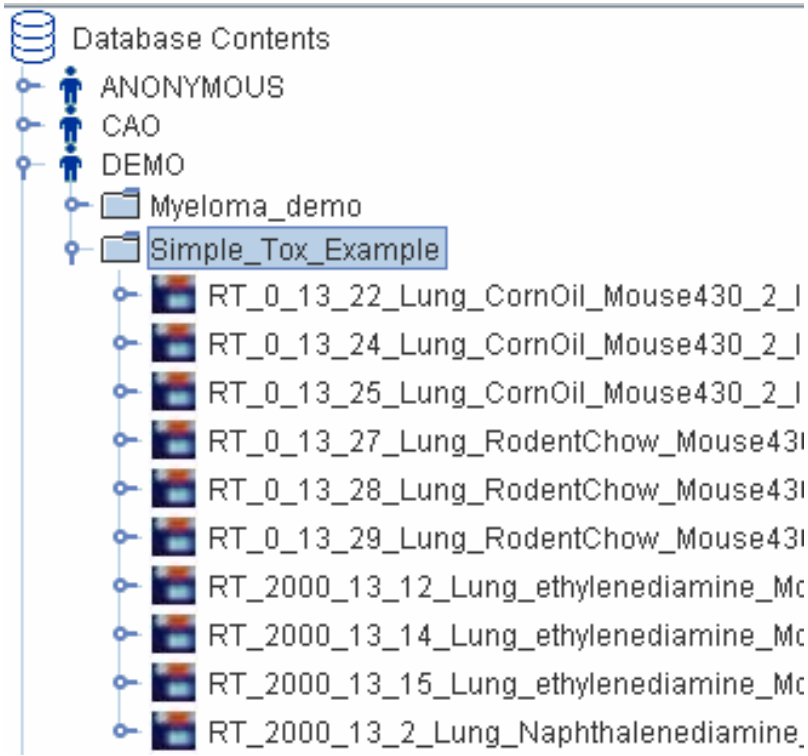
Select the right array type.

Batch Import – Simple Tox Format



Batch Import – Simple Tox Format

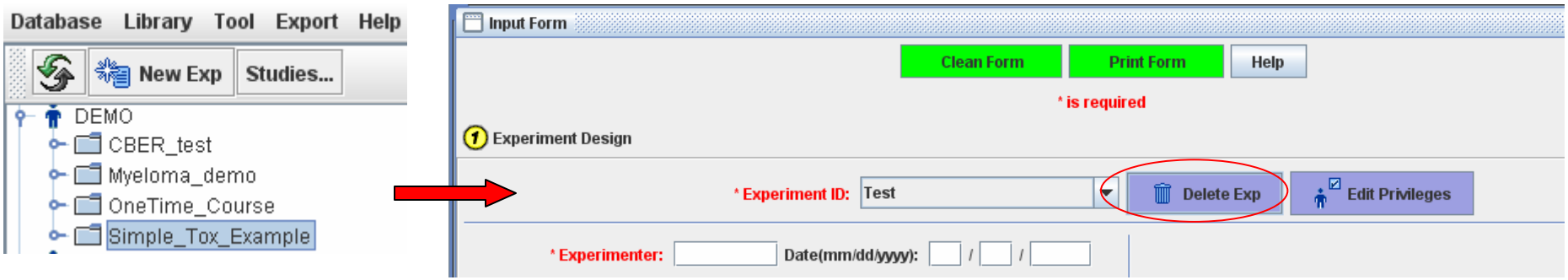




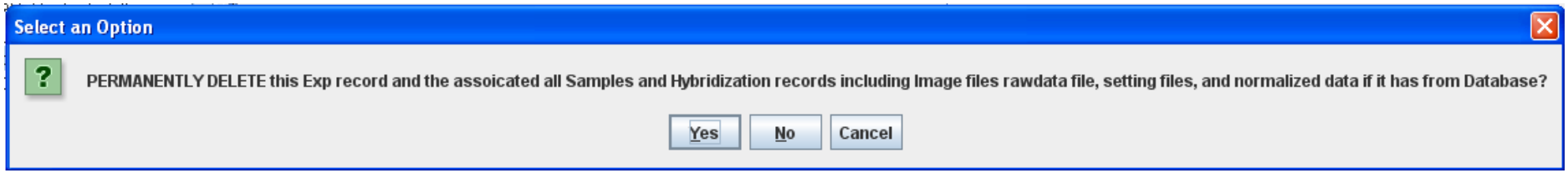
Batch import is successfully finished.

Delete Data

To delete SimpleTox data you have imported, you need to delete the experiment first and then delete the study.
Double-click the experiment name in the data panel to bring out the Input Form, then click “Delete Exp” button.

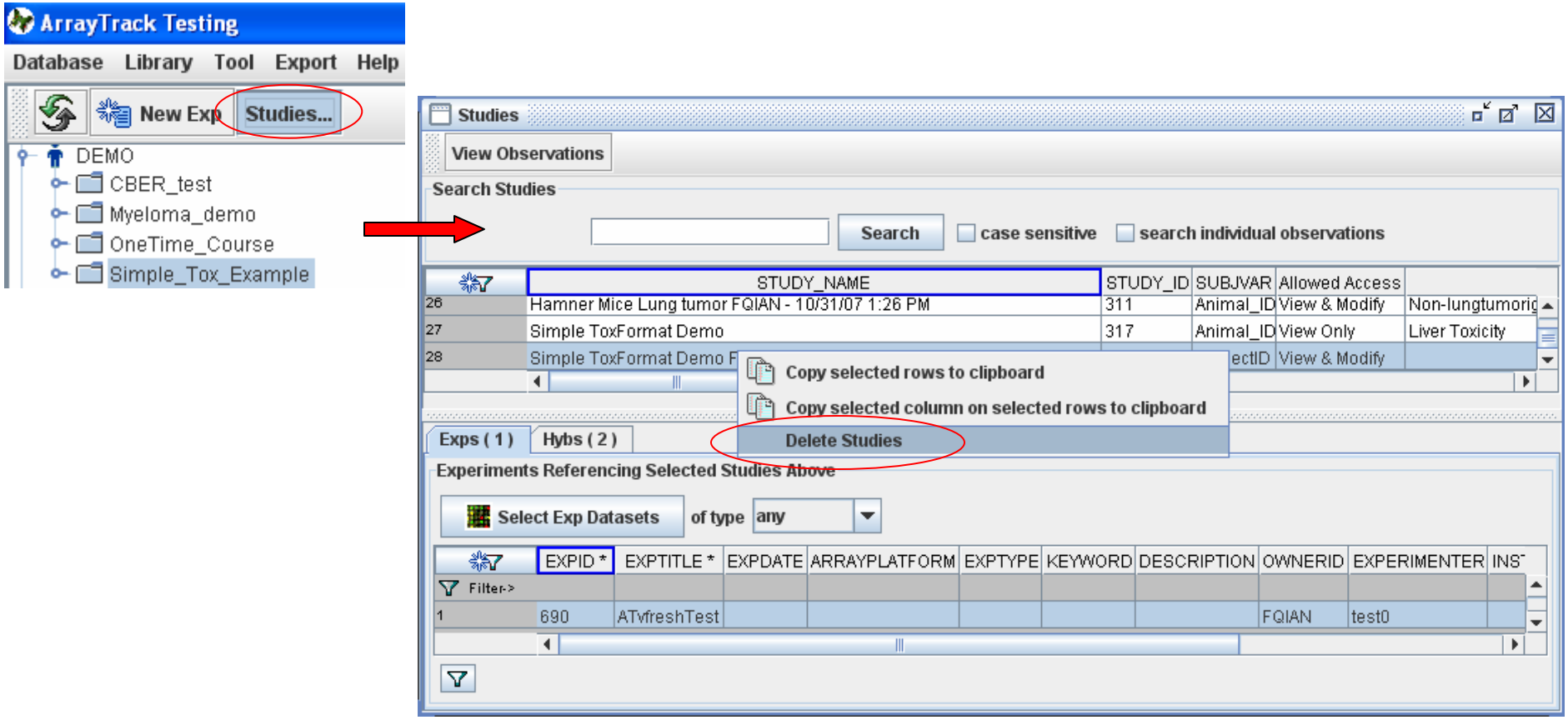
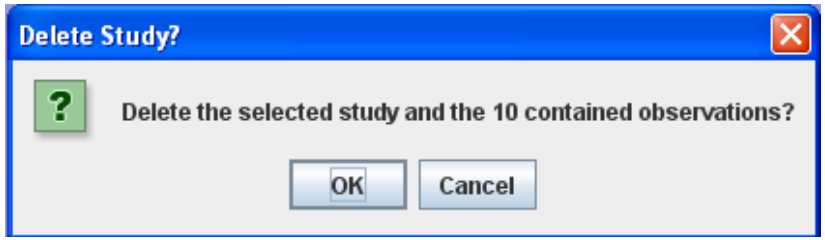


You will be asked the following question, click “Yes” button to permanently delete the experiment.



Delete Data

After deleted experiment, you need to delete the study. Select the study, right-click, Choose “Delete Studies”.



Click “OK” button.