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

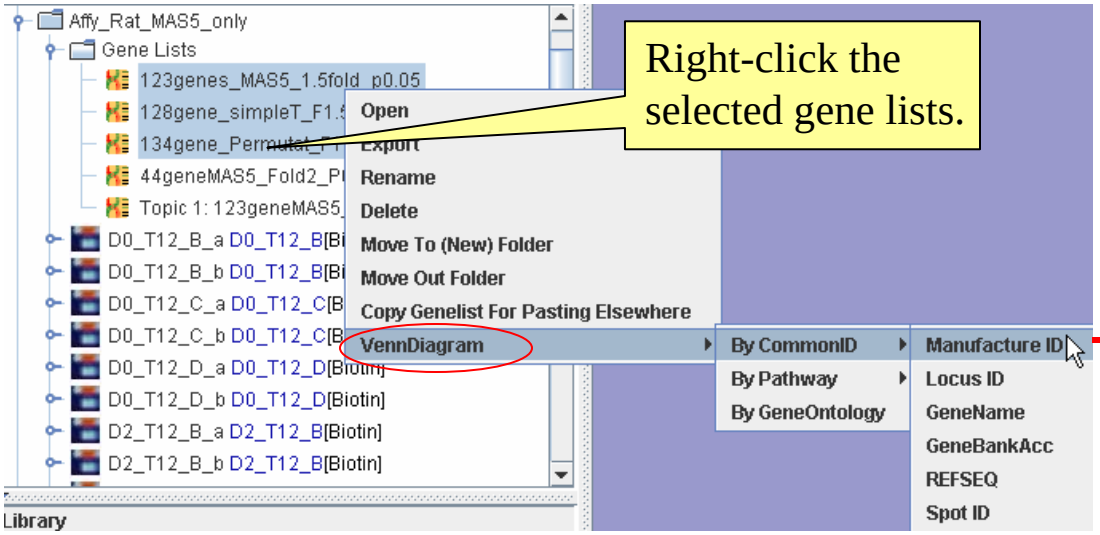
  <3> **Common Gene Ontology terms:**

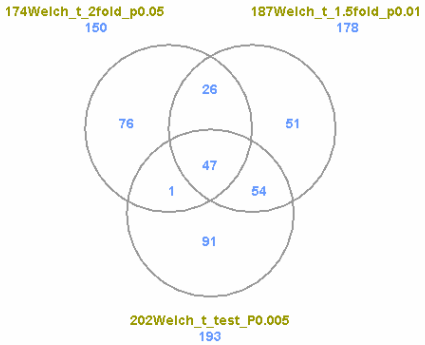
  <4> **Compare two array types:**

Common Genes

- Two ways to input to VennDiagram :

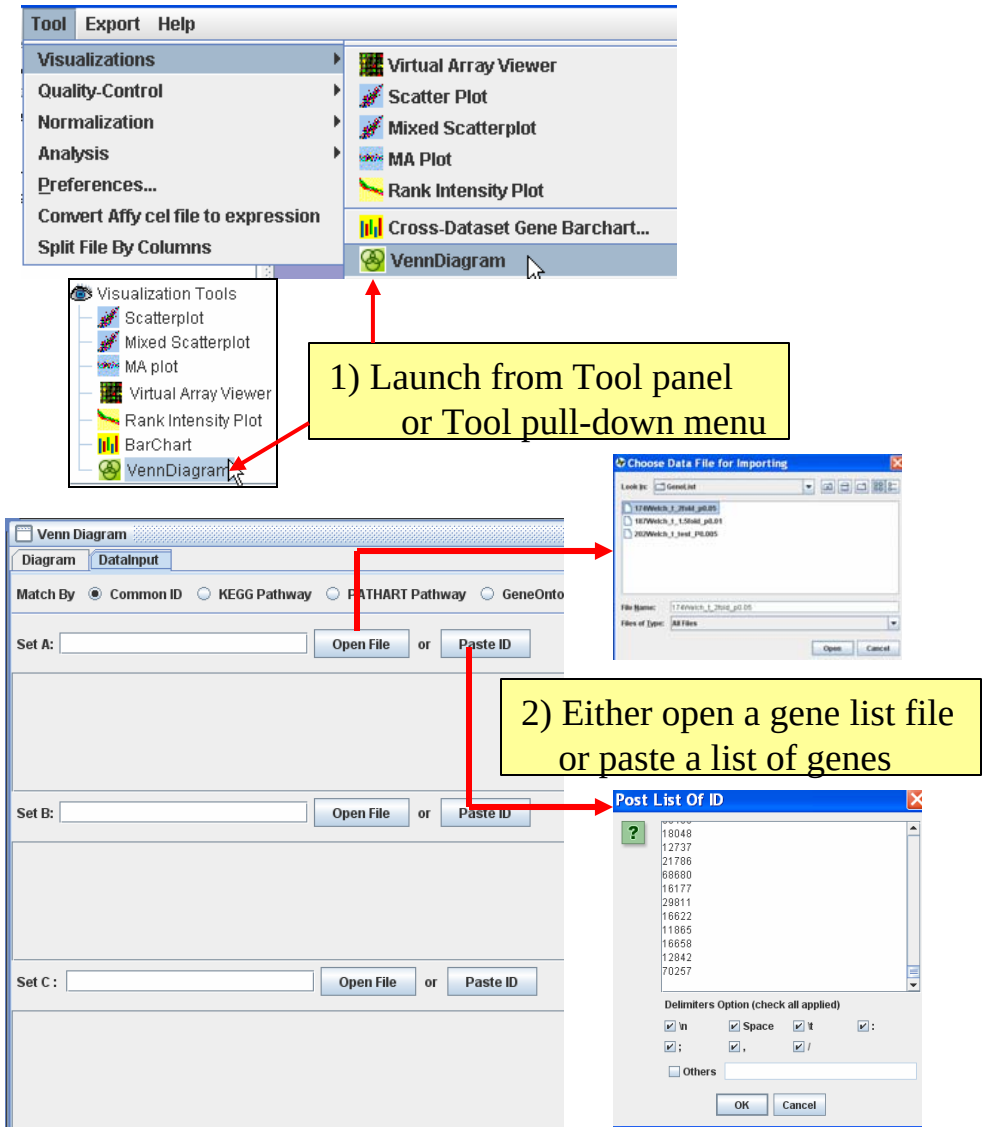
<1> when gene lists have been saved in ArrayTrack:

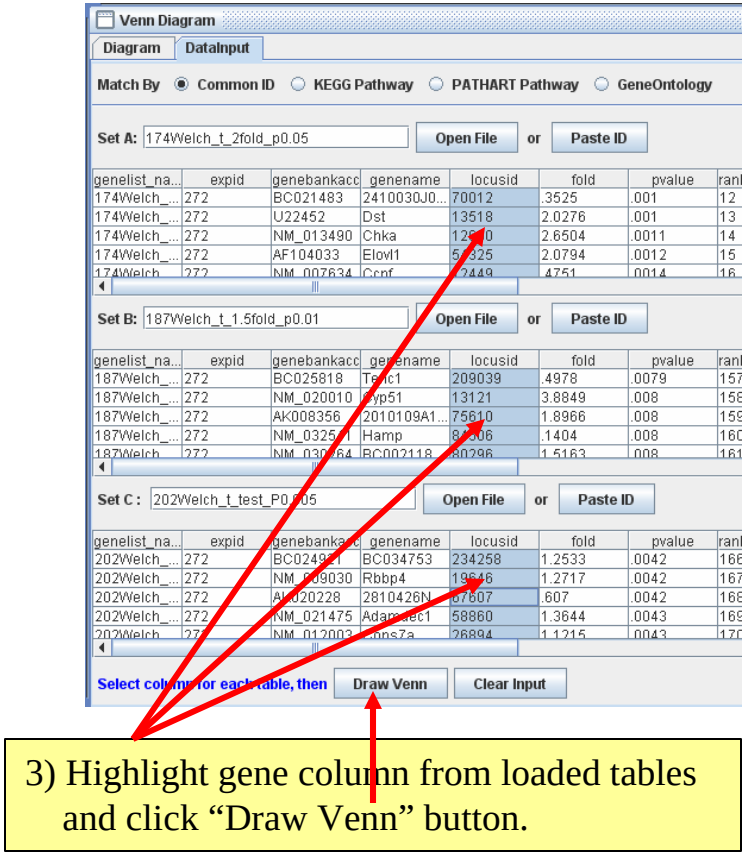




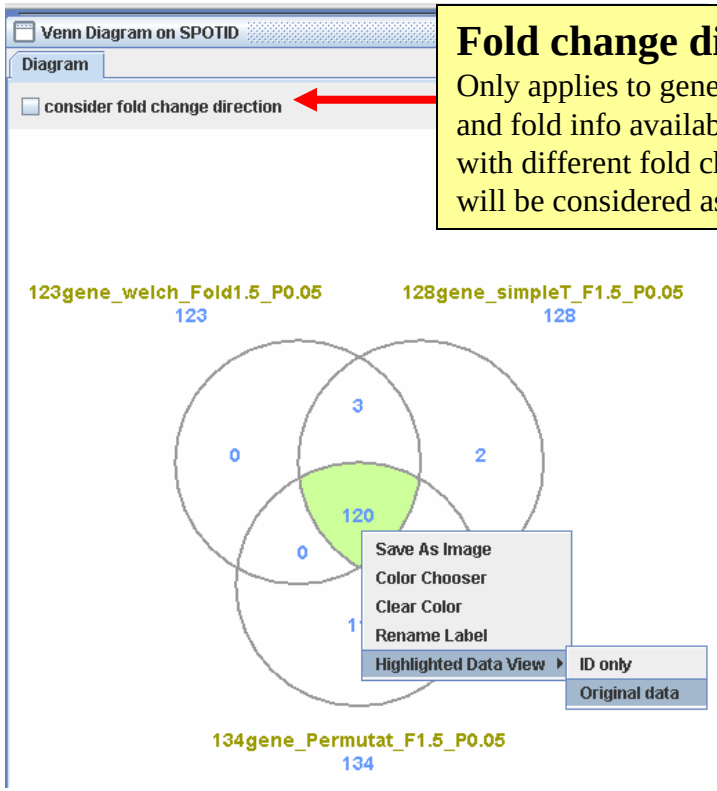
When genelists come from the same chip, Manufactory ID and Spot ID are preferred, otherwise use LocusID or other listed gene IDs

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



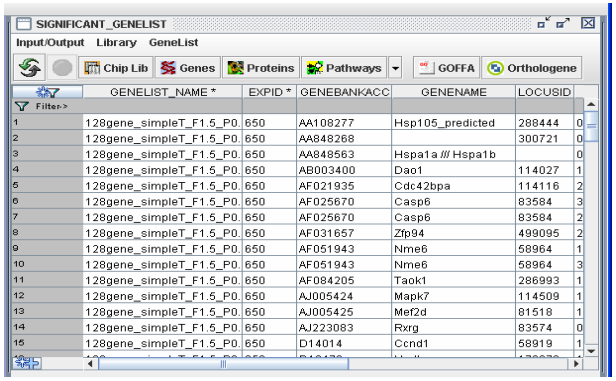


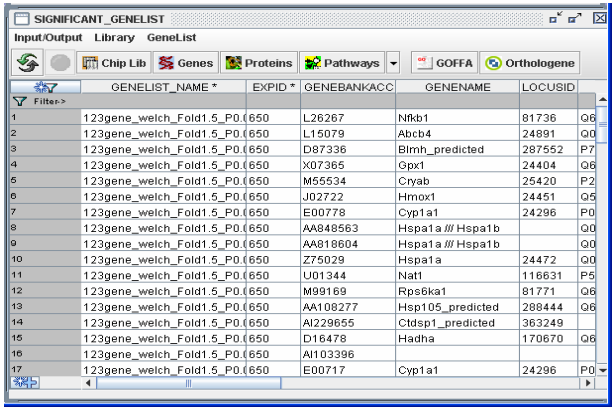
- Display original data for the selected sections

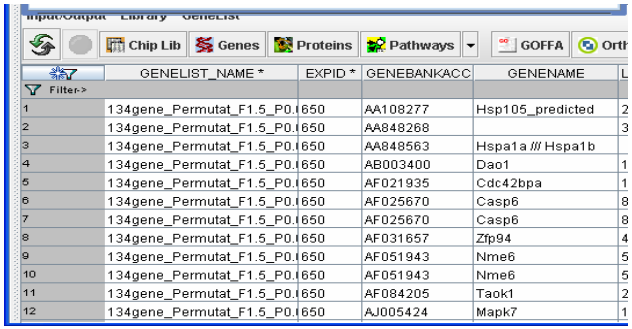


Fold change direction option

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.







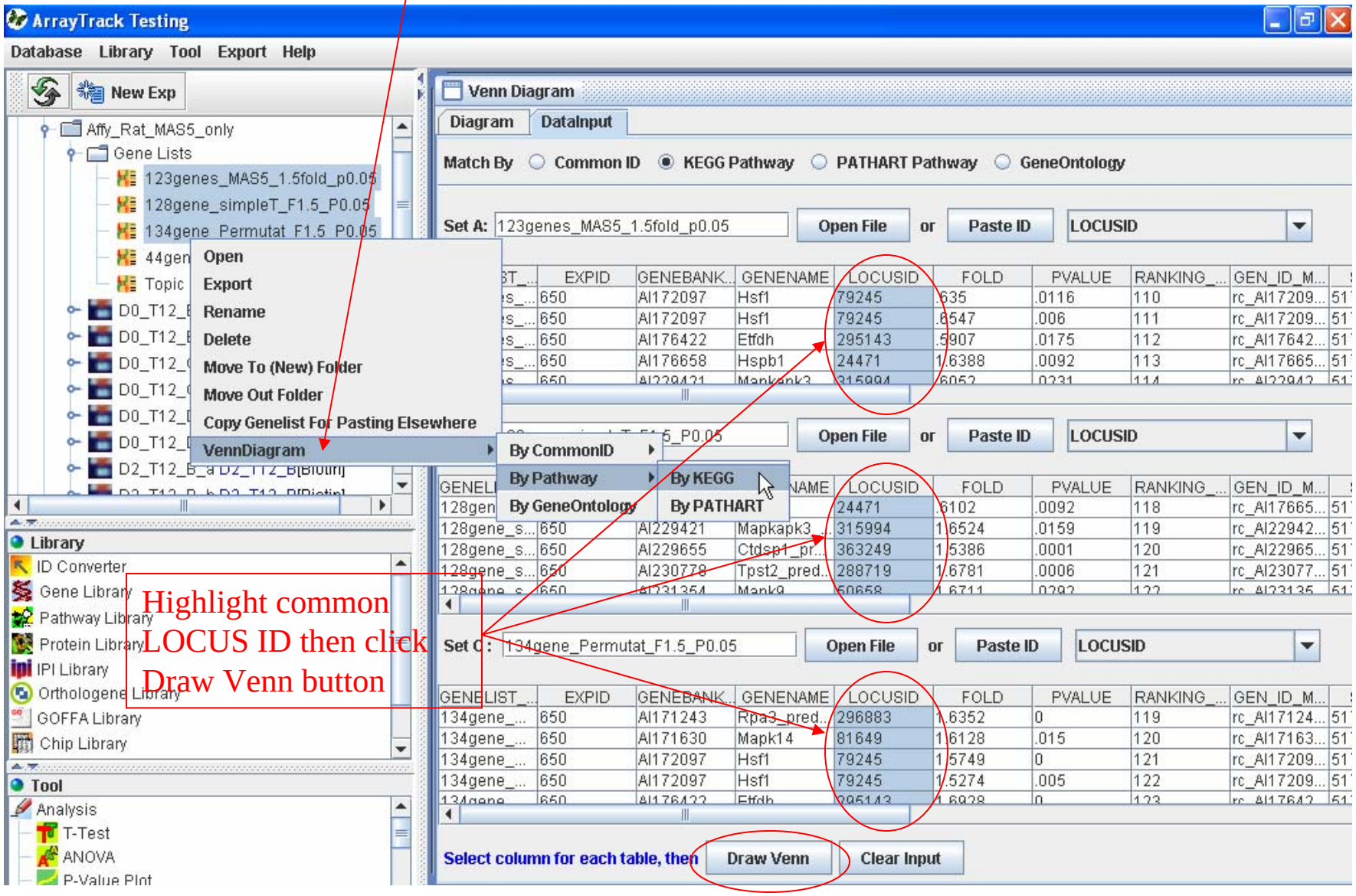
By clicking any sections from VennDiagram, the regions will be highlighted. Then right click to choose the options. One option allows display the selected genes in the original data sheet.

Common Pathways

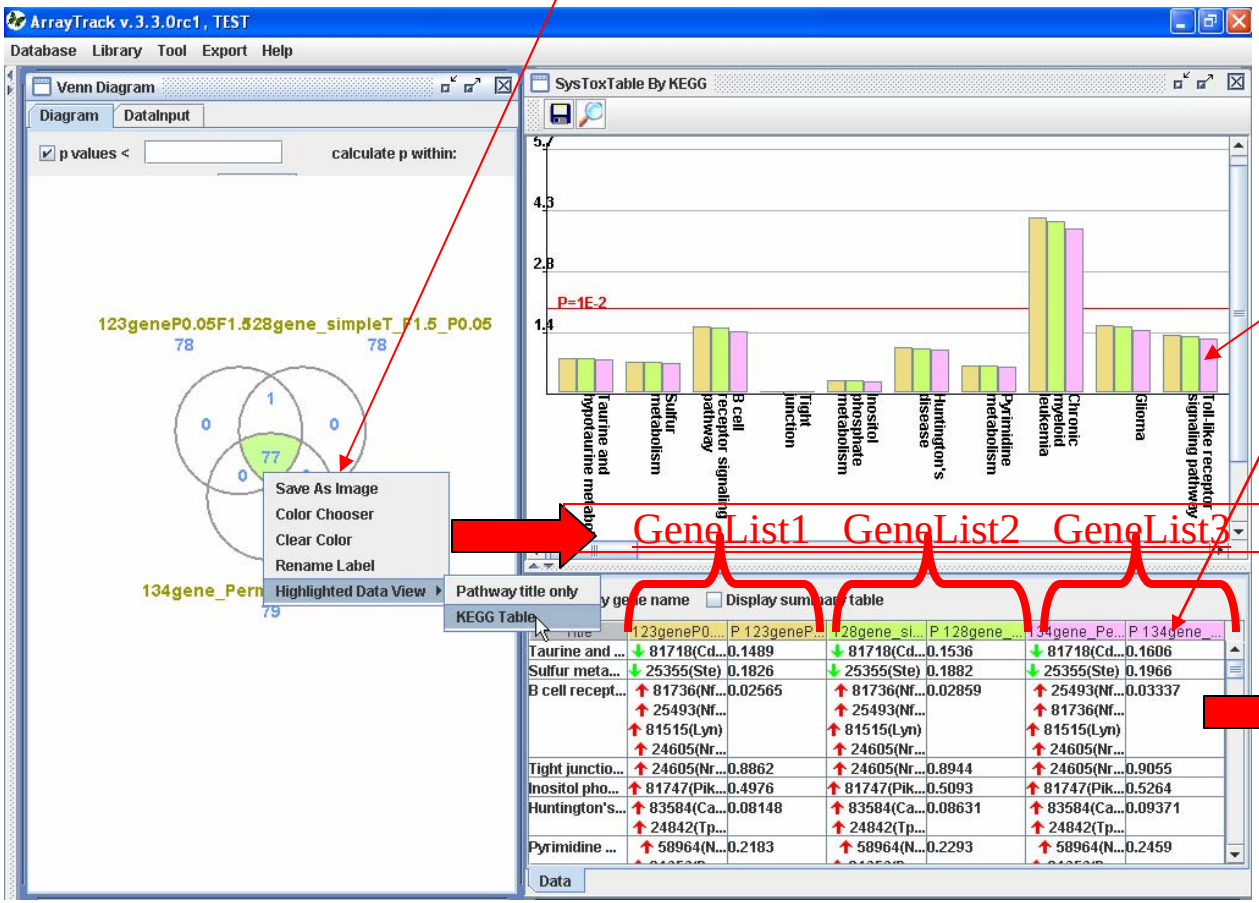
Using VennDiagram, the common pathways shared by two or three lists can be identified. The lists could be genes from microarray study, proteins from the proteomics study or metabolites from metabolomics. Please see Topic 6 for uploading the protein and metabolite list.

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

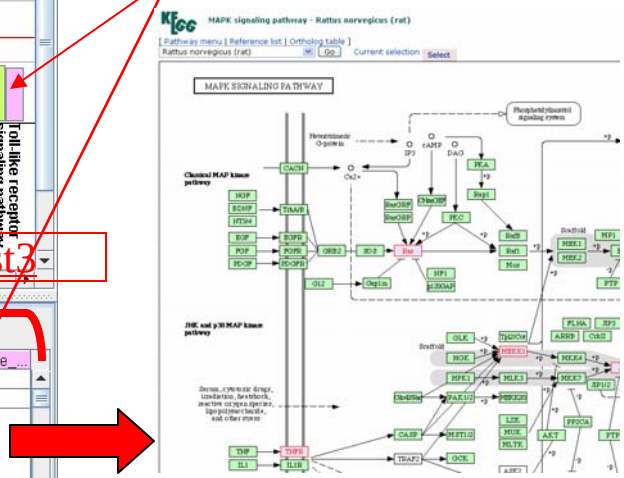
<1> Input: Right-click the selected gene lists, choose VennDiagram by Pathway/Kegg



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

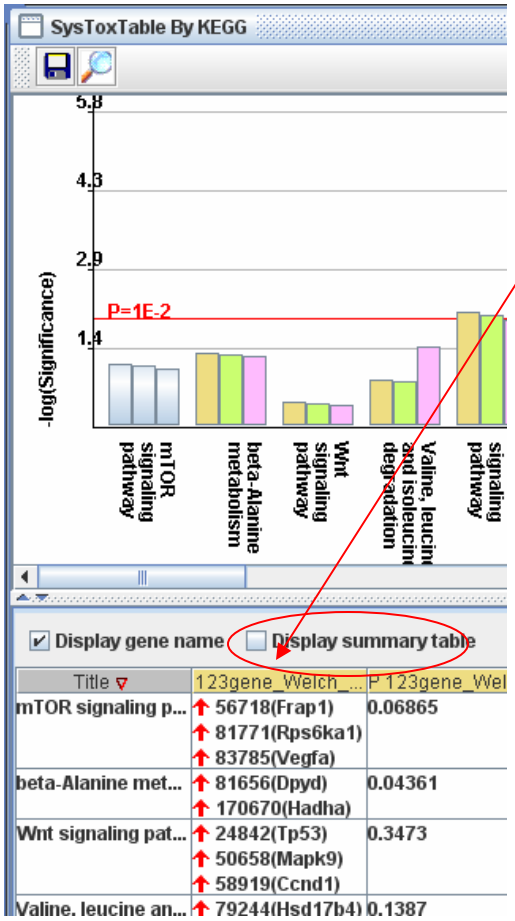


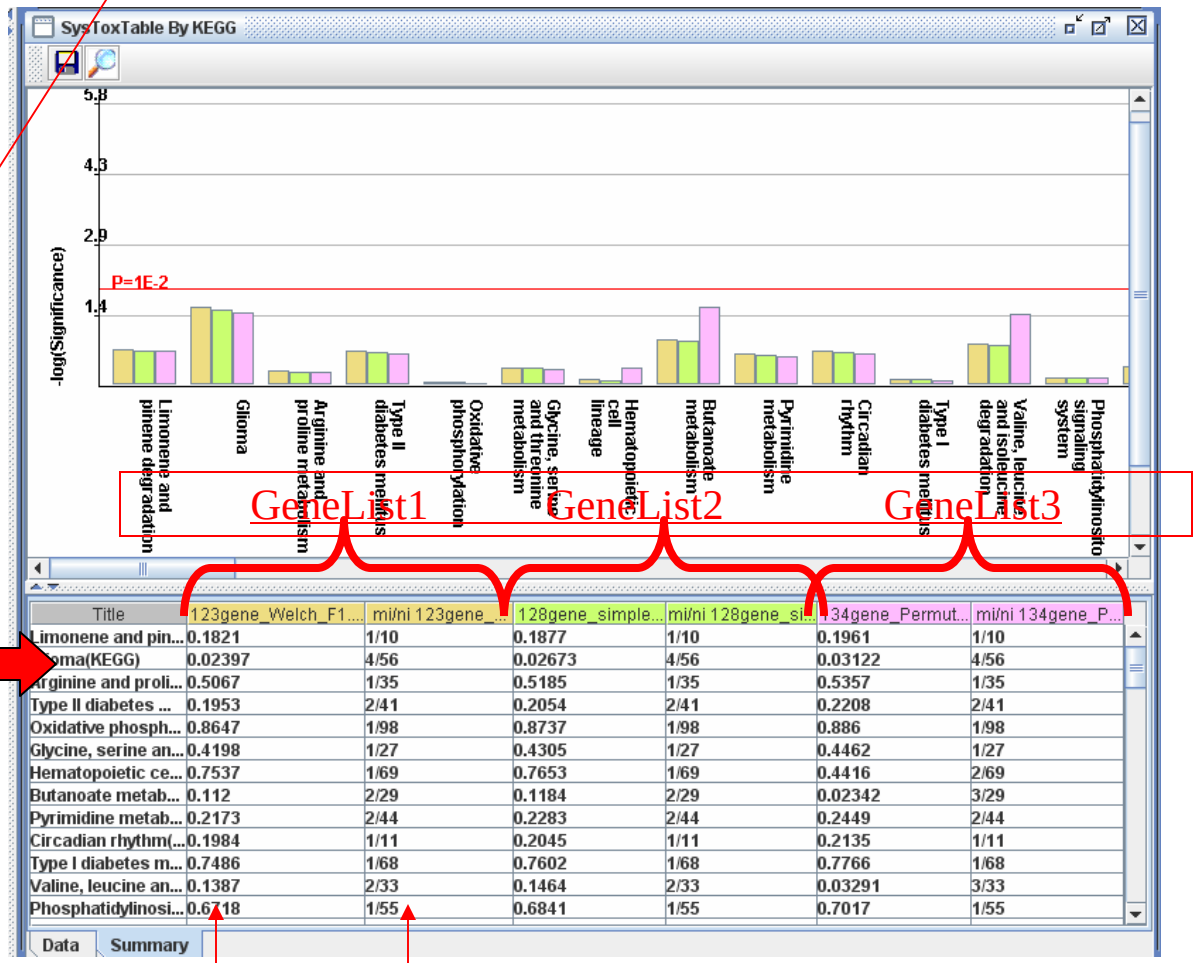
Different colors represent different gene lists.



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.





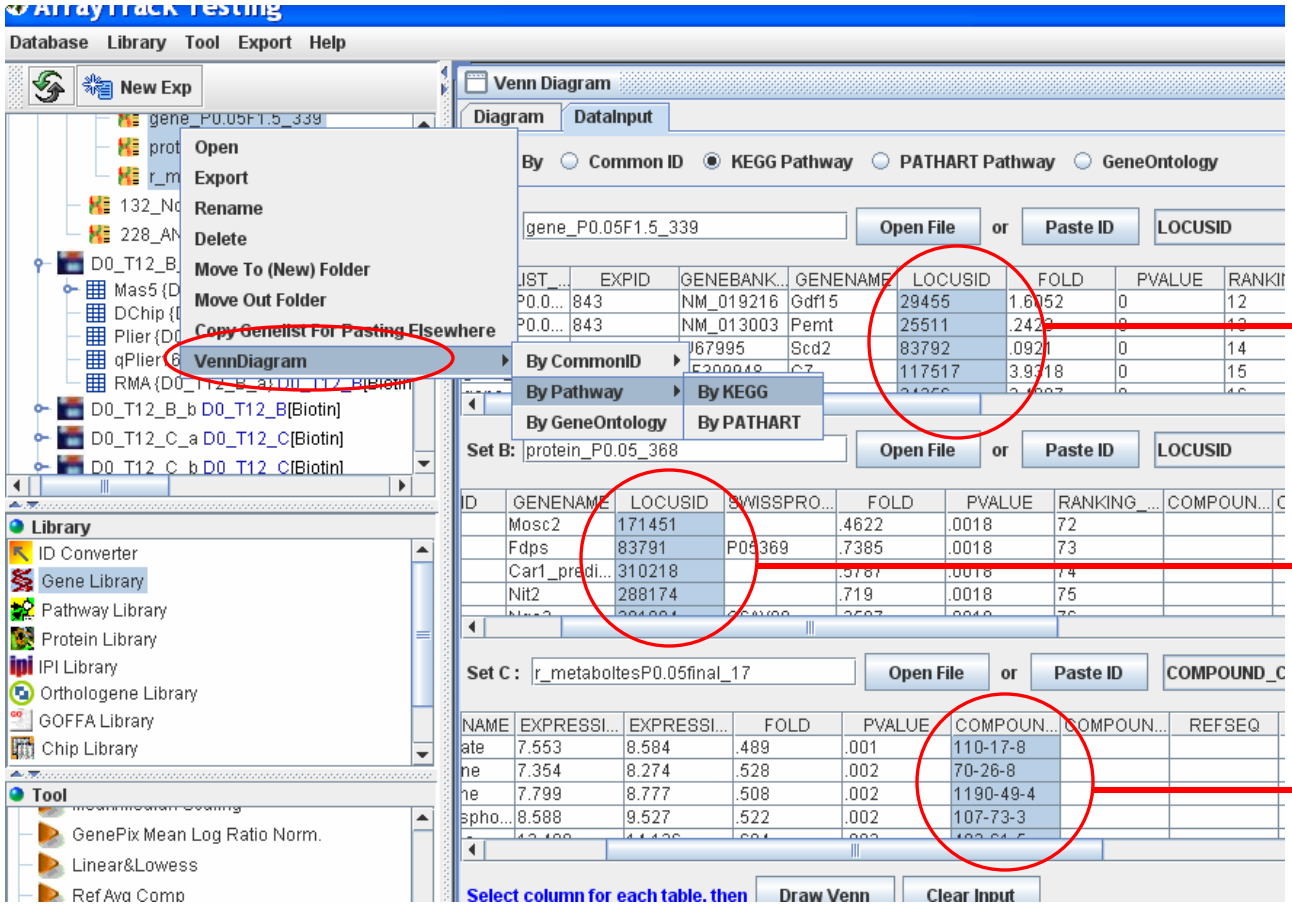
P value

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.

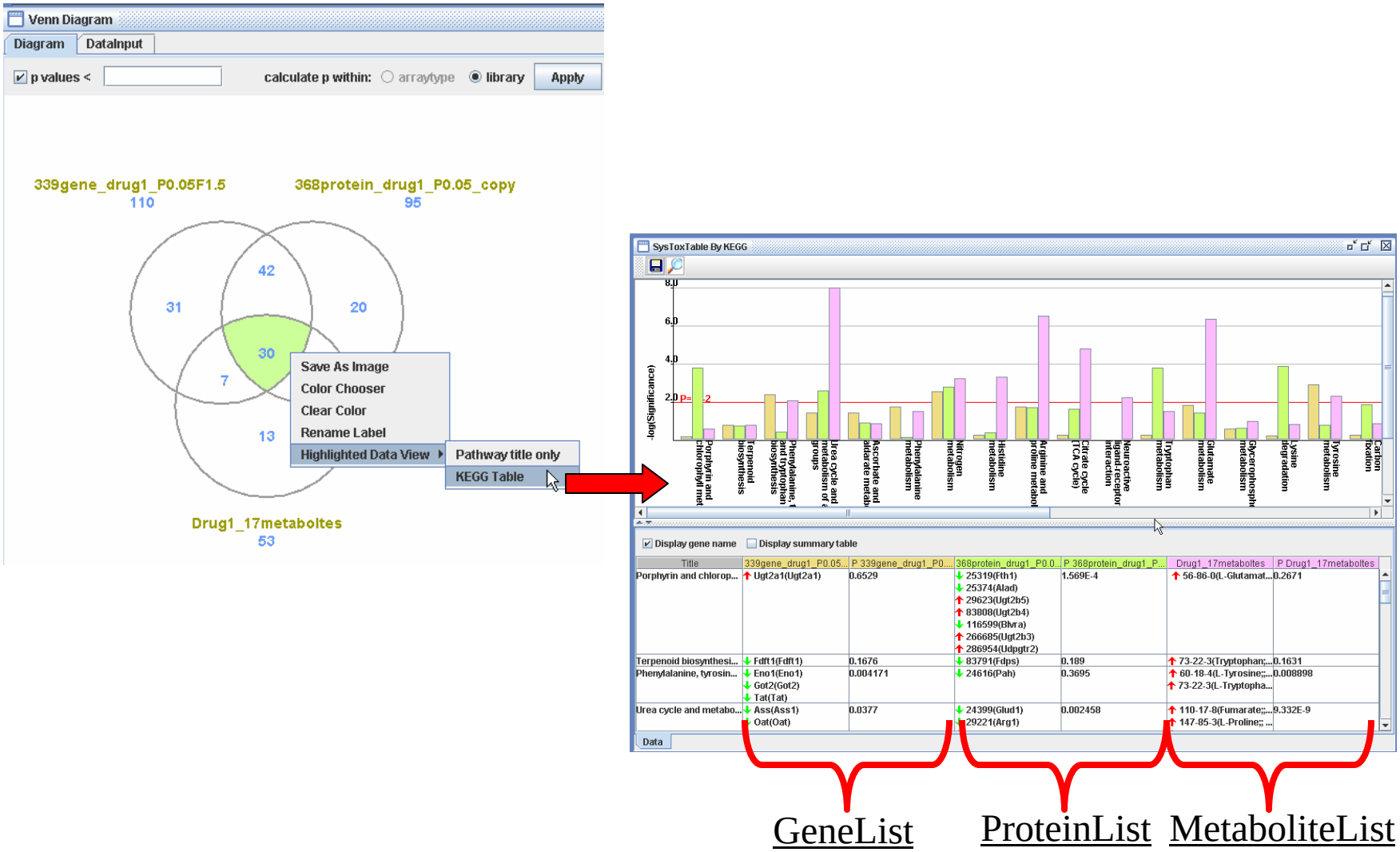


Genes
(i.e. locusid)

Proteins
(i.e. locusid or SwissProt ID)

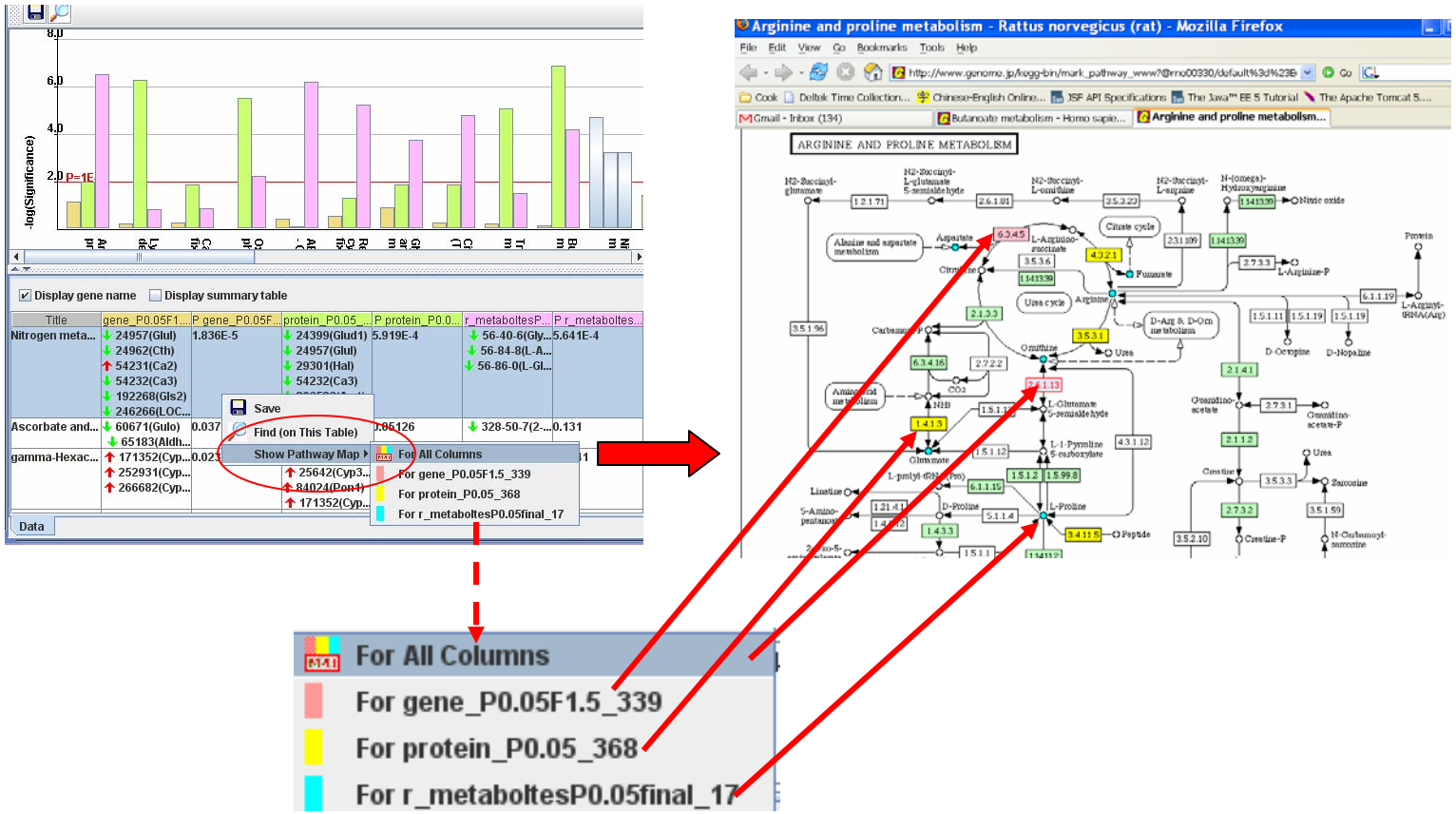
Metabolites
(i.e.Compound Cas #)

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



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

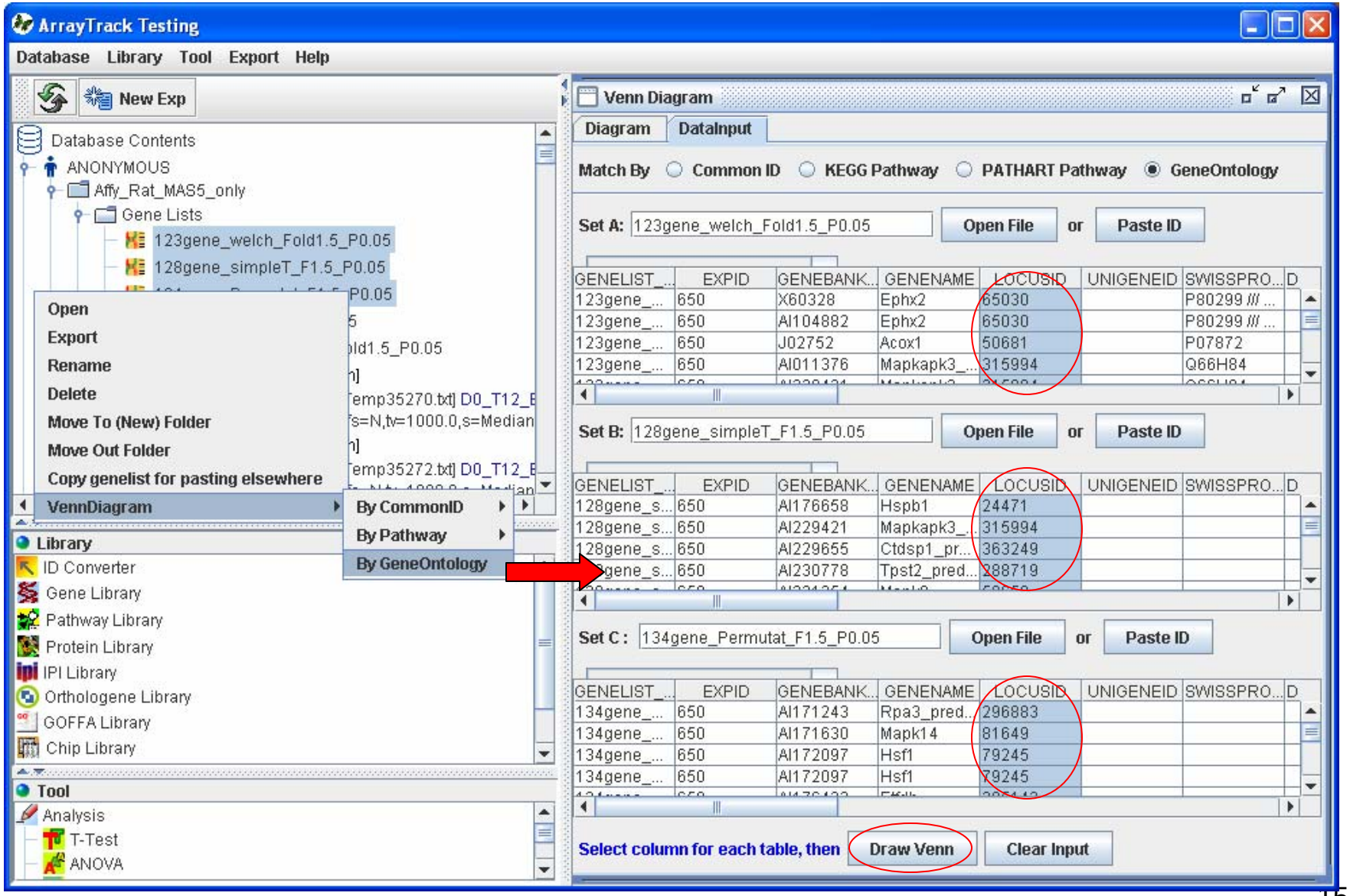
Double click a record from table, or highlight a record from table, right click and select “Show Pathway Map” - > “For All Columns”



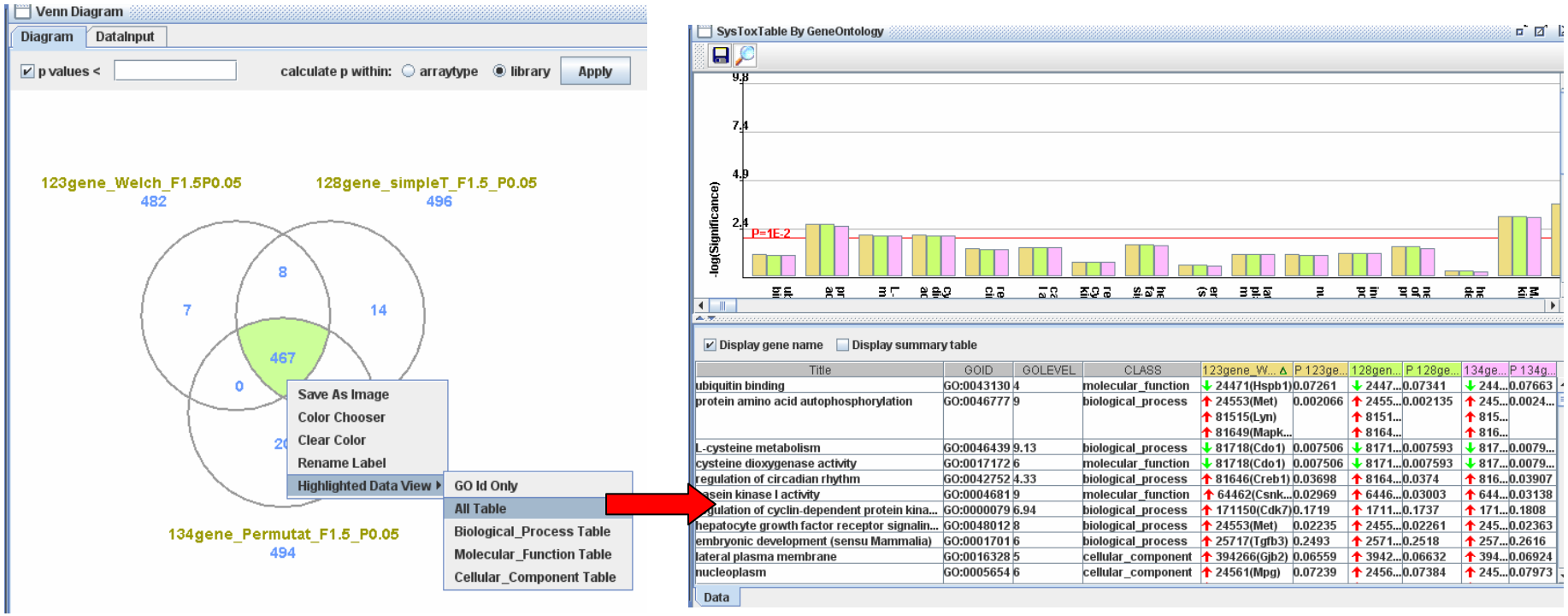
Common Gene Ontology terms

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.

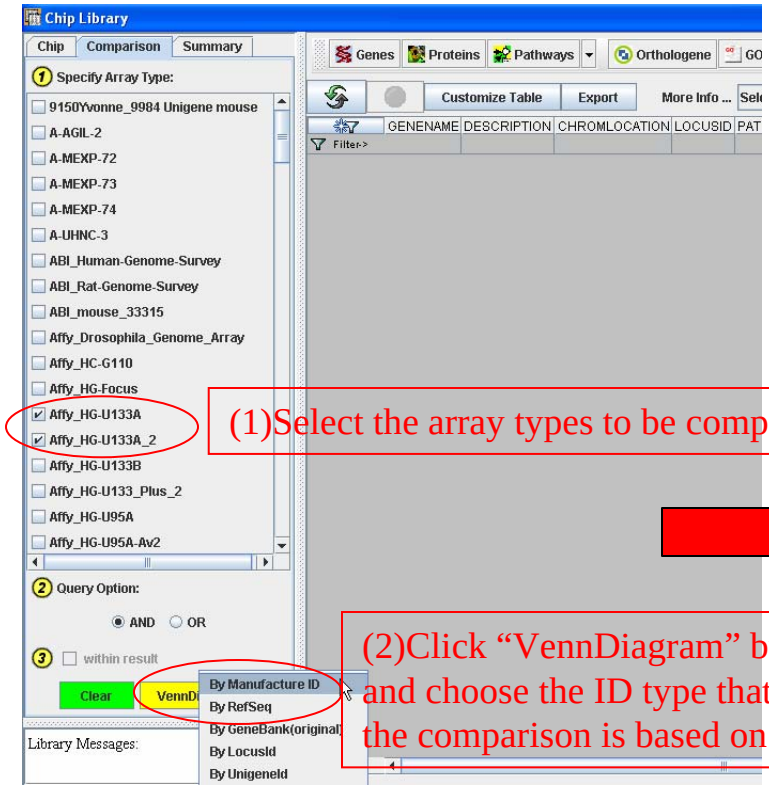
Highlight common LOCUS ID then click Draw Venn



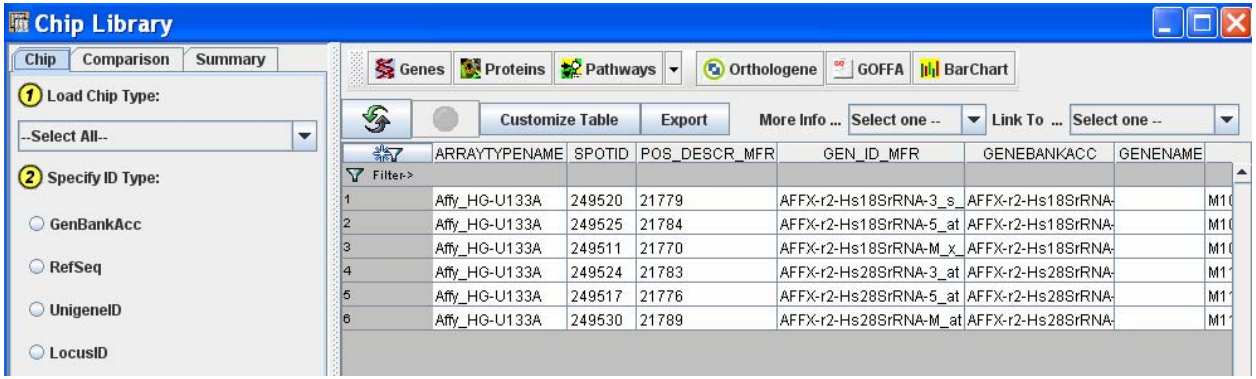
Right click the highlighted green area, then choose “Highlighted Data View”, selecting Common GO tables (all table, Biological Process, Molecular Function and Cellular Component)



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.