



Table of Contents

1 Working with Experiments	2
2 MIAME	3
3 Filtration	3
4 Normalization	5
5 Analysis	7
6 Annotations	10
7 BioAnalysis	12
8 Auto Guide	15
9 Annotation Settings	16
10 Utilities	19
11 View	22
12 Pathways	23
13 Show Data	24
14 Others	27

1 Working with Experiments

Q. Can I export an experiment created in Genowiz and import it back in Genowiz installed on other machine?

A. Yes, you can perform import/export of experiment along with associated data in Genowiz. (**File > Import Experiment and File > Export Experiment**)

Q. What are the different microarray data formats supported by Genowiz?

A. Genowiz supports GenePix, ScanArray, QuantArray, GemTools (Incyte), ImaGene, Agilent, MAS 4.0, MAS 5.0, RMA and RES formats. These file formats are automatically recognized by Genowiz and can be uploaded in a single step. Files in any other data format can also be uploaded using “Customized Uploader”. The customized uploader appears when data uploaded from **File > Import > Data** is not recognized automatically by Genowiz. Affymetrix raw data binary files (.CHP and .CEL) can be imported from **File > Import > Affymetrix Raw Data**.

Q. Can excel (.xls) files be imported?

A. No, only tab delimited text files can be imported in Genowiz.

Q. How does Genowiz handles replicate genes?

A. While uploading data, Genowiz merges replicate genes on the basis of user selected method (i.e. mean/median).

Q. I have uploaded 6 Affymetrix .CEL files and the corresponding library files from the internet and stored them on my C: drive. In Genowiz, I have created an experiment, and I am trying to upload the Affymetrix data as per the manual, but the program does not upload my files.

A. To upload Affymetrix .CHP or .CEL files, use **Import > Affymetrix Raw Data** option. Select required options based on files to be uploaded. The uploaded library file should correspond to the .CEL or .CHP files. MAS 5, RMA and PLIER algorithms are available for uploading .CEL files.

Q. I am working with Affymetrix data pre-processed by MAS 5.0/RMA/PLIER, do I need to normalize data again, if yes then which Normalization method is suggested?

A. Pre-processing algorithm takes care of normalization of data, therefore you don't need to normalize it again in Genowiz. However, if sample distribution indicates requirement of Normalization, then you could use Sample Normalization method (**Normalization > Sample Normalization > Median/Mean Centering**).

2 MIAME

Q. How can I store my experiment details?

A. Experiment details can be stored using MIAME module of Genowiz which is an information management system for all microarray-related experiments and is designed in compliance with MIAME standards to facilitate easy data entry.

Q. Is Multiple Experiment Analysis Support provided by Genowiz?

A. Genowiz does offer support for multiple experiment analysis with single instance of the software.

3 Filtration

Q. My data has replicated samples. How is such data handled by Genowiz?

A. Genowiz has options to merge replicated samples. Replicated samples can be used to filter out genes that show non-significant expressions across samples.

Q. What are the non-parametric tests available in Genowiz to filter out the non-significant genes?

A. Non-parametric tests do not assume normally distributed data to filter out the non differentially expresses genes. Genowiz supports Mann Whitney U-test and J5 test for two groups of samples and Kruskal-Wallis test for more than two groups.

Q. What does p-Value indicate in non-significant tests?

A. p-Value indicates the probability that difference in expression could have been caused by chance. Lower the probability, lower the error and higher the confidence that the gene is differentially expressed.

Q. When should I use multiple testing correction?

A. In testing for differentially expressed genes, p-Value indicates the probability that the gene's expression is different between two or more groups by chance alone. If 10,000 genes are tested with a probability threshold of 0.05, 500 genes ($10,000 * 0.05$) could be found to be significant by chance alone. Multiple testing correction corrects the p-Value of each gene to control overall error rate to be less than or equal to user specified p-Value.

Q. Which statistical test should I use to filter out non-significant genes?

A. The choice of the test depends on the assumptions that are met by your data. When you assume that samples are from two populations with equal variances, go for standard two-sample t-test. Otherwise, use t-test with Welch correction. It provides valid t-test when the variances are not homogeneous. If the samples in two groups are dependent and observations are paired, go for paired t-test.

ANOVA (Analysis of Variance) is used for more than two groups of samples. When there are two factors that affect the gene expression, two-way ANOVA is used to find the differentially expressed genes between groups of factors independently and also between combinations of two factors.

Q. Which filtration technique should I use if my data is very large?

A. Use Fold change to filter data.

Q. What is difference between Fold change single channel and Fold change double channel?

A. Apply Fold change double channel for cDNA arrays and Fold change single channel for Affymetrix arrays or other arrays, where only one dye is used. In single channel, you have to define control sample.

Q. How does Genowiz treat missing values?

A. Missing values can be filtered out using **Filtration > Missing Values** (or) can be replaced with the average intensity of nearest neighbors using **Normalization > Impute Missing Values**.

4 Normalization

Q. Why should we perform Normalization of data?

A. Microarray experiment includes several steps which could act as source for experimental bias/non-biological variation. Normalization is used to nullify this non-biological variations from samples.

Normalization also makes samples in experiment comparable for downstream data analysis.

Q. How can I know whether my data needs to be normalized or not?

A. User can take help of various plots like box-plot/scatter plot (**View > Plots**) to view sample distribution. Box plot is a very useful tool to visualize and compare sample statistics like mean/median/interquartile range/max-min of data. Normalization is not required if mean/median of all samples in experiment is similar.

Q. I would like to have some information regarding the algorithm used by the program to perform the double channels normalization.

A. For implementing Double channel normalization methods in Genowiz, following papers have been referred:

- Yang, Y.H., Dudoit, S., Luu, P., and T.P. Speed. (2001) Normalization for cDNA Microarray Data. SPIE BiOS 2001, San Hose, California, January 2001.
- Bolstad, B.M., Irizarry R. A., Astrand M., and Speed, T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics* 19(2):185-193

Q. Does Genowiz support normalization for dye-swap experiments?

A. Yes, Genowiz supports normalization for dye-swap experiments through **Normalization > Dye**

Swap Experiment. You have to define the dye-swap groups in order to perform dye-swap normalization.

Q. When do we use loess normalization?

A. When there is intensity dependent dye bias, we go for loess normalization. The relationship between dye-bias and intensity can be seen in an MA-plot, which is a scatter plot of the log-ratios called M-values (minus) against the log-intensities called A-values (add) for an array. Imbalance of the hybridization intensities of the different dyes can be seen as a curve in the plot. (**View > Plots > M/A Scatter Plot**).

Q. Do you have any subgrid normalization procedures in Genowiz?

A. Genowiz considers all the spots printed by the same print-tip as subgrid and print-tip loess normalization can be performed on each subgrid.

Q. What is the objective of Transformation and which transformation method should be used for expression ratios?

A. The expression ratio is a relevant way of representing expression differences in a very intuitive manner. For example, genes that do not differ in their expression level will have an expression ratio of 1. However, this representation may be unhelpful when one has to represent up-regulation and down-regulation. For example, a gene that is up-regulated by a factor of 4 has an expression ratio of 4 ($R/G = 4G/G = 4$). However, for the case where a gene is down regulated by a factor of 4, the expression ratio becomes 0.25 ($R/G = R/4R = 1/4$). Thus up-regulation is blown up and mapped between 1 and infinity, whereas down-regulation is compressed and mapped between 0 and 1. To eliminate this inconsistency in the mapping interval, one can perform transformations of the expression ratio, namely, logarithmic transformation. A better transformation procedure is to take the logarithm base 2 value of the expression ratio {i.e. $\log_2(\text{expression ratio})$ }.

This has the major advantage that it treats differential up-regulation and down-regulation equally, and also has a continuous mapping space. For example, if the expression ratio is 1, then $\log_2(1)$ equals 0 represents no change in expression. If the expression ratio is 4, then $\log_2(4)$ equals +2 and for

expression ratio of $\log_2(1/4)$ equals -2. Thus, in this transformation the mapping space is continuous and up-regulation and down-regulation are comparable.

Q. How are the outliers treated in Genowiz?

A. Outliers can be removed by using **Normalization > Set Minimum Maximum Values** option. In this, all the values below the minimum and above the maximum are set to the same.

Q. Can I transform Signal Log Ratios (SLR) to their corresponding fold change values?

A. You could get fold change values from Signal Log Ratios from Affymetrix comparative experiment design. (**Normalization > Fold Change Signal Log Ratio.**)

5 Analysis

Q. How do I know the optimum number of clusters that my data fall in?

A. It requires pre-existing knowledge or one has to go for trial and error methods. So, in the case of partitional clustering methods (K-Means, Self Organizing Maps and Gene Shaving), one should do clustering many times, checking for homogeneity within clusters.

To a certain extent, hierarchical clustering methods (Single Linkage, Average Linkage and Complete Linkage) give an idea of the clusters present in the data.

Q. How do I remove heterogeneity of data that is present even after clustering?

A. You can re-cluster the clusters. Select cluster of your interest and again perform clustering on it. Repeat this until you get a homogeneous cluster.

Q. What is the purpose of 2D PCA?

A. 2D PCA gives a sense as to how many natural groupings exist in the data space. PCA can be done before the cluster analysis to capture the cluster structure (e.g., to know a good value of K in K-means clustering).

Q. What is the best distance measure to be used while clustering?

A. The choice of the distance measure depends on the area of application and the sort of similarities one would like to detect. For example, if the gene expression measurements for all samples in one gene are three times the expression measurements in the other gene, those two genes would be considered distant using Euclidean distance metric, but close using correlation coefficient (because correlation coefficient considers only change pattern). Manhattan distance is more robust against outliers. Euclidean distance is the preferred one to successfully group similar data items.

Q. I am interested in two or more clusters and want to merge them? Can I do that with Genowiz?

A. Selected clusters generated from either K-means, SOM or Gene Shaving can be merged. Right-click on selected clusters and select **Merge** option. Specify a name for the merged cluster. It is added to the analysis tree. Information about replicates in the merged cluster is displayed.

Q. Among the hierarchical clustering methods present in Genowiz, how do I choose the best one for my dataset?

A. In single linkage, distance between two clusters is the distance between the two closest objects in two clusters. As it ignores the distance between remote objects in two clusters, it tends to produce chained clusters. It is useful in clustering evolutionarily related entities, for example, taxonomically related species.

Complete linkage is the opposite of single linkage and considers distance between two furthest points in two clusters as the distance between them. This is useful when objects of same cluster are expected to be far from each other. This is not recommended if there is a lot of noise in the data.

Average linkage uses average of all pairwise distances between points in two clusters as the distance between two clusters. As it doesn't produce chained clusters and doesn't give weight to outliers, this is the preferred one in most of the situations.

Q. Is there any option in Genowiz to see cluster similarity?

A. The best option here is to use Self Organizing Maps (SOM) technique. It not only clusters the data

but also arranges the clusters in such a way that neighboring clusters are similar when compared to distant clusters.

Q. What is the main purpose of Principal Component Analysis (PCA)?

A. Principal Component Analysis is a technique to reduce the dimensionality of the data set. As a result of PCA, you get principal components that are linear combinations of original variables. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. By using only the first few principal components, it is possible to reduce the dimensions of the data, while maintaining the maximum possible variance.

Q. What is the significance of "Scree plot" in PCA?

A. It is the plot showing eigenvalues in PCA and can be used to decide the number of components to be retained. Look for an “elbow” in the plot and the components before the elbow can be taken for further analysis.

Q. How do I know the sample classes after performing classification using Discriminant PCA or SVM?

A. Classes can be distinguished by means of a separator in expression matrix. Samples falling in Class 1 are arranged before the separator and of Class 2 after the separator.

Q. I am interested in two or more clusters and want to merge them? Can I do that with Genowiz?

A. Selected clusters generated from either K-means, SOM or Gene Shaving can be merged. Right-click on selected clusters and select **Merge** option. Specify a name for the merged cluster It is added to the analysis tree. Information about replicates in the merged cluster is displayed.

6 Annotations

Q. Can I choose to view a group of annotations of interest, instead of defining them each time?

A. You can create annotation views from **Annotations > Create Annotation Views** for a selected database. Created views are reflected in “View Annotations” window. Selecting a view here will display information for annotation fields of that view alone. Editing of created views can be done either from the “View Annotations” window or from “Create Annotation Views” window. A view can be set as default also from here. The Bio Annotation Viewer and the **Annotations** tab below the expression matrix will display fields from the default view corresponding to the Annotation Settings specified. For User Annotation Settings if the default view includes some custom fields that are not present in the annotation file selected (to retrieve annotations) the Bio Annotation Viewer does not open. You can change the default view in this case to work with the Bio Annotation Viewer.

Q. Can I obtain a tabular view of annotations for all/group of genes?

A. Navigate to **Annotations > View Annotations**. The “Bio Annotation Viewer” window appears showing annotations for all genes/identifiers in the selected node. Select checkboxes/annotation views to view annotations of interest.

Q. I want to change the URL for a particular annotation field. How do I do that?

A. URLs for annotation fields in various annotation settings can be changed from **Utilities > Preferences > Annotations > URL Editor**. They can also be changed from “Bio Annotation Viewer” window in the **Connect** field. Click on the cell of an annotation field, for which you want to change the URL. The URL is displayed in the Connect field. Edit the URL. The Update URL icon is enabled. Click on the icon to update the URL for the selected annotation field for a particular annotation setting.

Q. Can I save/export genes from the annotations table or create a gene list?

A. Yes. Click **Save** button. In the “Save Data” dialog, select **All Genes** or **Select Genes** option. In case of the later, select genes from the annotations table. Click **OK** button. “Save Data” dialog appears. Enter name. Click **OK** button to create a gene list. The gene list gets added to the analysis tree with the

name specified and can be subjected to further analysis.

Export can be performed similarly by clicking on **Export** button. Exporting can be done in several formats namely .txt, .csv, .html, .xls and .xml.

Q. Can I perform search in the tabulated list of biological annotations?

A. Yes. You can perform search in the tabulated list of biological annotations. Select **Search** tab in the Bio-Annotation Viewer. Define the search term and click **Find** button. Ensure that at least one annotation field or annotation view is defined before executing search.

Q. Can I view/export expression data along with annotations for my genes/identifiers?

A. Yes. In the Bio-Annotation Viewer, click **Data** button. From the list of Available Samples, move samples of interest to the Selected Samples panel. Click **OK** button. Depending on the properties of the node on which the Bio-Annotation Viewer is opened, data for the selected sample is displayed. Raw, normalized, fold change values and p-Values are displayed in the Gene Data table depending on the sample and node properties. Selecting a row in the Annotations table highlights the corresponding row in the Gene Data table and vice versa.

The annotations can be exported along with the expression data for selected samples by clicking on **Export** button in desired format.

Q. Can I view and export annotation, expression value, fold change values, p-Value simultaneously?

A. Yes, you could view and export data with annotation with Bio-Annotation Viewer (**Annotations > View Annotations**).

Q. Where can I see the annotations for a selected gene?

A. Select a gene from the expression matrix, its annotation information will be displayed in the **Annotations** tab below the expression matrix. You can also select a gene from the expression matrix, navigate to **Annotations > Gene Information** or double click on a gene from the expression matrix. This opens the Gene Information Panel showing annotation details for the selected gene. Annotations

are retrieved from the current selected annotation settings.

Q. Can the list of positively or negatively correlated genes be saved as a gene list?

A. Yes. Right click on the list of positively or negatively correlated genes. Click on **Save** button. In the resulting “Save Data” dialog, specify a name and click **OK** button. A gene list is created with the name and added to the analysis tree.

7 BioAnalysis

Q. How do I build gene ontology or perform functional classification for my genes/identifiers?

A. Select a valid data node on which gene ontology is to be performed. Navigate to **Bio Analysis > Build Gene Ontology**. Gene ontology is displayed in the **Gene Ontology** tab. The ontology is built in **Experiment Name** folder. The ontology node is created with the node name on which gene ontology is built followed by the annotation settings used to build the ontology. A General Information tag added to the same node depicts the file or chip name in case of User/NetAffx™ annotation settings.

In the **Analysis** tab all ontology nodes are saved under the **Gene Ontology** folder.

For new organisms/genomes, ensure that Genowiz annotations are updated, for working with Genowiz annotation settings or an annotation file/NetAffx™ chip is uploaded for working with User/NetAffx™ annotation settings. Gene Ontology cannot be performed if annotations are absent in the selected annotation settings.

Q. Can I view gene ontologies built on different experiments simultaneously?

A. Yes. Open both the experiments. Double click on the **Gene Ontology** tree nodes that you want to visualize, in the **Gene Ontology** folder of respective experiments. The trees will be shown in their respective Experiment folders in the **Gene Ontology** tab.

Q. Can I save the genes in a functional cluster as a gene list?

A. Yes. Select the functional term that you are interested in from the respective tree in the **Gene Ontology** tab. Right click on the node. Select **Save Data** option. Specify a name to save the gene list

and click **OK** button. A gene list is created with the specified name and added to the **Analysis** tree.

Q. What do the numbers in parenthesis mean?

A. After building gene ontology, a number in parenthesis appears for all functional terms that have genes. Refer to the table for details.

Annotation Settings	Count in Parenthesis
Genowiz	<p>If gene ontology is performed on main node: Count is shown as : (a/b) where b= number of genes for the function in Gene database a= number of genes in the function from the main node</p> <p>If gene ontology is performed on any other node: Count is shown as: (a/b/c) where c= number of genes for the function in Gene database b= number of genes in the function from the main node a= number of genes for the function from the node on which gene ontology is performed</p>
User	<p>If gene ontology is performed on main node: Count is shown as : (a) where a= number of genes in the function from the main node</p> <p>If gene ontology is performed on any other node: Count is shown as: (a/b) where b= number of genes in the function from the main node a= number of genes for the function from the node on which gene ontology is performed.</p>
NetAffx	<p>If gene ontology is performed on main node: Count is shown as : (a) where a= number of genes in the function from the main node</p> <p>If gene ontology is performed on any other node: Count is shown as: (a/b) where b= number of genes in the function from the main node a= number of genes for the function from the node on which gene ontology</p>

	is performed

Q. What does Z-Score signify in GO?

A. The significance with which each gene has been assigned to a particular ontology is given by the z-score. A report showing this information is also provided. Higher the z-score more is the function/process/component affected.

Q. I want to view “apoptosis” in the Gene Ontology tree. Do I need to expand all the nodes and perform a manual search?

A. Manual search need not be done. Right click on any node in the **Gene Ontology** tree and select the **Search** option. Enter the text that you are looking for e.g. "apoptosis". Search will be performed in the tree and the corresponding node will be highlighted. The expression matrix will also be created if genes are present.

Q. How do I generate pathways?

A. Select the node on which pathway analysis is to be performed. This should be a valid data node. Else a message appears “Select a valid data node”. Navigate to **Bio Analysis > View Pathways**. If pathways are not present for the organism, a message appears “Data not available for viewing pathways. It would take a few minutes to download from KEGG. Do you want to continue?” Click Yes. If internet is up and running and there are no firewall problems in accessing KEGG database, pathways are downloaded from KEGG. If the download fails, a message appears "Unable to download from KEGG. Do you want to update pathways from alternate location?" Click **Yes** button. A folder selector appears. Select the folder that contains all your downloaded pathways. A message appears. "Download successful. Generating pathways". Click **OK** button. The pathway editor opens.

Non academic users need to have a license for downloading pathways from KEGG.

All pathways built are saved in the **Pathways** folder in the **Analysis** tree. To open a pathway from one experiment or multiple experiments, double click on a pathway node in the **Pathways** folder. The

Pathway Editor opens. The pathway node is created with the node name on which pathways are built followed by the annotation settings used to build it. A General Information tag added to the same node depicts the file or chip name in case of User/NetAffx annotation settings.

For new organisms/genomes added, ensure that Genowiz annotations are updated, for working with Genowiz annotation settings or an annotation file/NetAffx chip is uploaded for working with User/NetAffx annotation settings. Pathways cannot be performed if annotations are absent in the selected annotation settings.

8 Auto Guide

Q. How can I automate my analysis flow in Genowiz?

A. Analysis can be automated in Genowiz with the help of Auto Guide. Auto Guide provides user friendly interface to create and execute work flows.

Q. How can Auto Guide help me in analysis of my data?

A. Genowiz can help users to perform analysis with default auto guides, which are specific to data type (Single channel/Double channel) and also take cares of various scenarios like presence of replicate sample, Affymetrix quality calls, Dye-swap experimental design, etc. (**Auto Guide > Open > Default Workflows**).

Q. How can I create Auto Guide and use it for routine analysis?

A. Users can create customized analysis flow (**Auto Guide > New**). Analysis flow already created by users can be used for routine analysis.

Guidelines to work with Auto Guide:

1. Click on an option from any of the analysis tabs.
2. Click on Auto Guide Layout to paste an icon of the corresponding analysis..
3. To indicate the order in which analyses are to be performed, you can establish connections between these analyses.

- To connect two analyses, select the first analysis icon and right click.
 - In resulting popup menu, select **Connections** option.
 - Drag the mouse from the first icon to second one.
4. To move a selected icon, right click and select **Move** option from resulting popup menu.
 5. To delete a selected icon, right click and select **Delete** option from resulting popup menu.
 6. To clear the layout section, right click and select **Clear** option from resulting popup menu.
 7. Once you have defined all connections and flow of analysis, click **Run** button.
 8. Enter parameters for analysis, wherever necessary and click **Validate** button. To stop execution of the Auto Guide, click **Stop** button.
 9. Result is saved in Genowiz after successful execution of Auto Guide.

9 Annotation Settings

Q. What do the Annotation Settings mean?

A. Genowiz supports three types of annotation settings, namely Genowiz, User and NetAffx™. Genowiz annotations are the default settings.

- Genowiz annotations refer to annotations that are provided along with the application for default organisms and for new organisms, annotations can be obtained by connecting to NCBI Gene database.
- User annotations refer to annotations that can be imported by users. If researchers have some proprietary annotation files or annotation files from different databases; after importing this file into Genowiz, annotations can be retrieved from this file.
- NetAffx™ annotations refer to annotations retrieved from the Affymetrix chips. Once a chip is added to Genowiz, annotations can be obtained.

Q. I define Custom ID as the identifier during experiment creation. Only the User Annotations option is enabled. Why is this so?

A. The **Annotation Setting** options are enabled and disabled depending on the identifier selected at the time of experiment creation. Please refer to the table for details on which annotation settings are supported for which identifiers.

Identifier	Genowiz	User	NetAffx
ProbeSet ID	-	Yes	Yes
Custom ID	-	Yes	-
ORF ID	Yes	Yes	Yes
GenBank Accession	Yes	Yes	Yes
RefSeq Accession	Yes	Yes	Yes
Entrez Gene ID	Yes	Yes	Yes

Q. How do I add an annotation file and connect to User Annotation Settings?

A. Navigate to **Annotation Settings > Configure... > User Annotations** option. Click + button. Select the annotation file to be imported and click **Open** button in the “Open” dialog. “Save Data” dialog with the file name appears. Change the name if required. The Select Organism field is disabled if the file is added after experiment creation; else; the organism for which the file is to be added has to be selected. Click **Save** button. In the resulting “Annotation Uploader” window, select the **Gene Identifier** column (mandatory) and other columns as required. Click **Save** button. In the “Column Mappings” dialog, click **Continue** button. The file gets added to the database and is reflected in **User Annotations** drop down. Click **Apply** button to apply the settings to retrieve annotations from this file. In case there are more than one file for an organism, annotations are retrieved from the first file by default.

Q. The fields in my annotation file are not listed among the Available Fields in Annotations Uploader? How do I define them?

A. In “Annotations Uploader” window, enter the field name of interest in the New Field text field and click **Add** button. This field gets added to the list of Available Fields in “Annotations Uploader” window. The field can now be selected and assigned to the corresponding column in the annotation file.

Q. How many new fields can be added to the list of Available Fields in the Annotations Uploader?

A. 20 new fields can be defined and added to the list of Available Fields in the Annotations Uploader.

Q. How do I add an Affymetrix chip and connect to NetAffx™ Annotation Settings?

A. Navigate to **Annotation Settings > Configure... > NetAffx™** option. Click + button. “Add Chip” dialog appears. Select the organism for which the chip is to be added. In case, the chip is added after experiment creation, the organism name is displayed by default. Enter the chip name. The chip can either be added from the local folder or by connecting to NetAffx™ analysis center (you should have an account with Affymetrix). In case of the later, ensure that the chip name specified is as mentioned by Affymetrix or else a message appears stating "Chip not found". Click **Add** button to add the chip. The chip gets added to the database and is reflected in NetAffx™ drop down. Click **Apply** button to apply the settings to retrieve annotations from this chip.

In case there are more than one chip for an organism, annotations are retrieved from the first chip by default.

Q. What is the Map Gene Identifier option? When does it get enabled?

A. On account of Genowiz annotation update, if there is any change in the database column names for the identifiers defined at the time of experiment creation, the “Map Gene Identifier” dialog appears when a user tries to access any bioanalysis modules like Gene Ontology, Pathways, View Annotations, Gene Information Panel etc. The option is also enabled under **Annotation Settings**.

This primarily helps users to establish a connecting link between the identifier defined at the time of experiment creation with the corresponding column in the database. Once the link is established

correctly, annotations can be retrieved. If a wrong link is defined, annotations cannot be retrieved, and a message appears "No annotations found. Change Annotation Settings".

User can navigate to **Annotation Settings > Map Gene Identifier** to edit the settings.

Only the Gene Identifier drop down is enabled in case there is a change in the database column pertaining to the identifier defined at the time of experiment creation. The **Pathway Identifier** drop down is enabled when there is a change in the database column name pertaining to pathway identifiers namely ORF ID/Entrez Gene ID.

In case identifiers like ORF ID, RefSeq Accession, GenBank Accession or Entrez Gene ID are defined and NetAffx is the defined annotation setting; as soon as **Apply** button is clicked in the **Annotation Settings > Configure...** dialog, "Map Gene Identifiers" dialog appears. Select a field to map the defined gene identifier at the time of experiment creation to the corresponding column in the NetAffx chip and click Apply. Annotations will be retrieved. The mapping can be edited from **Annotation Settings > Map Gene Identifier**. In case of incorrect mapping, annotations cannot be retrieved and a message appears "No annotations found. Change Annotation Settings".

10 Utilities

Q. How do I search for genes specific to a pattern?

A. Select **Utilities > Pattern Simulation**. Define your pattern and Genowiz will search for the genes with similar expression pattern.

Q. Can I update annotations, gene ontology and pathway details?

A. The annotations present in Genowiz annotation settings can be updated from **Utilities > Preferences > Annotations > Updates > Genowiz**. Relevant check boxes can be selected to update Gene Ontology, Pathways, Organism Details or Gene Annotations. This would update the relevant information for all the organisms in the application. After updating, **Report** button is enabled. Report displays the update status.

This is a time taking process and also requires a lot of free disk space.



GENOWIZ FAQs

Q. The updates seems to fail. What could be the reason?

A. Please ensure that there is sufficient free disk space (~1 GB) and internet connectivity. There should be no firewall issues and you should have download permissions. If it still fails, try the **Local Folder** option in Update Report. Non academic users need a license agreement to download pathways from KEGG.

Q. What does the Local Folder option in “Update Report” dialog do? Can I update annotations, gene ontology or pathways from alternate location?

A. In case update fails from internet, one can update annotations, gene ontology, pathways and organism details from an alternate location or the **Local Folder** option from the “Update Report” dialog.

Please ensure that the following files are present in the folder you select from the local disk.

Option Selected for Update	Files Required for Update from Local Folder
Gene Ontology	URL: http://archive.geneontology.org/latest-full/ Required File: go_YYYYMM-termdb-tables.tar.gz; where YYYY stands for the year and MM stands for the month. e.g. (go_200802-termdb-tables.tar.gz)
Pathways	All pathway maps and the KGML Data.html file must be present in a folder. The folder should be named with the NCBI Taxon ID of the organism for which the pathways need to be updated. If pathways have to be updated for all organism's then the folders for each organism should be present in one consolidated folder. Non academic users need a license agreement to download pathways from KEGG.



GENOWIZ FAQs

Organism Details	URL: ftp.ncbi.nih.gov/pub/taxonomy/ Required File: taxdump.tar.gz
Gene Annotations	URL: ftp.ncbi.nlm.nih.gov/gene/DATA/ Required Files: gene2go.gz, gene2accession.gz, gene_info.gz, gene2unigene, gene2pubmed.gz, mim2gene

Q. Can I update annotations for NetAffx™ chips?

A. Annotations for existing NetAffx™ chips can be updated from **Utilities > Preferences > Annotations > Updates > NetAffx™**. The **Existing Chips** option is selected by default. Annotations can be updated for all chips of all organisms; for all chips of selected organism; for selected chip of selected organism. Annotations can be updated either by connecting to NetAffx™ analysis center or from local disk.

Q. Can I add an organism/genome of interest in Genowiz?

A. Genowiz supports 7 organisms by default. To add a new organism/genome, navigate to **Utilities > Preferences > Organism > Add Organism** or in “New Experiment” dialog, click **Add Organism** link. “Add Organism” window opens. Click **Add** button and enter scientific name of the organism in **Species Scientific Name** field. If the scientific name matches with that specified in NCBI Taxonomy database, then organism details are saved. If present, KEGG organism code for the new organism is retrieved.

Q. The organism name I have entered is the correct scientific name, but still a message appears "NCBI Taxon ID not found. Cannot add the organism".

A. The organism name has to be entered exactly as specified at NCBI Taxonomy database. For example, if the organism that needs to be added is Escherichia coli K-12 MG1655, it has to be specified as Escherichia coli K12.

Q. I entered an organism, a message appears "KEGG Organism Code not found".

A. If the organism name is not specified as defined in KEGG, this message appears. You can still add the organism. The pathways module however, will not work for such organisms.

In case you want to work with the pathways module, please contact [support@ocimum.com](#) for assistance.

Q. What does Pathway Identifier mean?

A. Pathway identifier is required for mapping genes in the uploaded data onto pathway maps. The identifier can be either ORF ID or Entrez Gene ID. After saving an organism, select the row and click **Edit** button. Now select the identifier from the drop down that appears on clicking the Pathway Identifier cell. If no identifier is selected, Entrez Gene ID is considered by default.

11 View

Q. What are the different types of plots or graphs generated by Genowiz?

A. Some plots generated by Genowiz are Box-Plot, Scatter-Plot, M/A Scatter plot, CH1/CH2 Scatter Plots, Line Graph, Bar Graph, Mean Graph, Venn-Diagram, Pie-Chart, Scree Plot and PCA graph.

Q. Where can I see mean graphs of individual clusters?

A. Mean graph is displayed by default for any selected cluster in **Mean Graph** tab. For a better view of any graph for the selected cluster, click on **View > Plots** menu, and select the required graph

Q. How can I interpret Box-plot?

A. Box-plot is a convenient way to visualize sample distribution. Box-plot implemented in Genowiz gives information about sample statistics like minimum, lower quartile, mean, median, upper quartile, inter-quartile range, maximum, and outliers in data. Box-plot can be used to compare sample distribution of samples in experiment.

Q. Can I view scatter plot with fold change line?

A. You can view fold change lines in scatter plot by right-clicking on the plot, and selecting **Graphs >**

Settings. In the resulting dialog, enter fold change value in Line at field. Genes identified at user specified fold change cutoff can be saved/exported for further analysis.

Q. Is there an option to change the color of the expression matrix?

A. Color of expression matrix can be changed from **View > Heat Map > Color Attributes**. Colors of missing values, borders, gene names, sample names and background can be changed from “Color Attributes” window.

Q. What is a heat map?

A. Heat map is a two dimensional representation of expression intensities of genes in various samples. X-axis represents samples and Y-axis represents genes. Heat map is also referred as expression matrix.

Q. Can I change the order of samples after uploading data into Genowiz?

A. Yes, you can reorder the samples using **View > Heat Map > Shuffle Samples**.

12 Pathways

Q. What do the numbers in pathway node parenthesis mean?

A. Against each pathway term, there is a number in parenthesis. For e.g. (a/b), where:
a = number of genes participating in the pathway from the node on which the pathways are built.
b = number of genes participating in the pathway as obtained from database.

Q. What are the red colored boxes on the pathway map?

A. The genes from the expression data (the node on which pathways have been built) are mapped onto pathways. The mapped genes are highlighted in red. Mouse pointer or clicking on them shows their information.

Q. Can I search for pathways of interest?

A. Right click on any node in the **Pathway** tree and select the **Search** option. Enter the text that you are

looking for e.g. "Pyruvate Metabolism". Search will be performed in the tree and the corresponding node will be highlighted; the pathway map and expression matrix are also created.

Q. Can I save the genes in a pathway as a gene list?

A. Yes. Select the pathway term that you are interested in from the **Pathway** tree in the Pathway Editor. Right click on the node. Select **Save Data** option. Specify a name to save the gene list and click **OK** button. A gene list is created with the specified name and added to the **Analysis** tree. Pathway nodes that do not have genes mapped to them cannot be saved as gene lists.

Q. Can I generate a pathway report?

A. Click **File > Pathway Report** in the Pathway Editor.

Q. What if I want to create my own pathways or edit an existing pathway?

A. **File > Create Pathway** option will enable creation of new pathways. To edit an existing pathway, open a pathway, make changes to it and click on **File > Save Pathway**. The pathway will be saved with the specified name and details under the **Custom Pathway** node.

13 Show Data

Q. Can I demarcate the genes of a particular cluster in any other analysis?

A. You can define a specific public cluster by selecting a cluster and selecting Set Public Cluster from pop-up menu of the heat map. All genes in a public cluster are marked with user-defined color on the right side of the heat map and are reflected in the course of analysis, if present.

Q. How do I export data to a physical file?

A. Right click on expression matrix, select **Show Data** option. This pops up a table with list of Genes and Expression values for all samples. You can export this data as a tab delimited file by clicking **Save** button.

Q. Can I view and export p-Value for differentially expressed genes identified by statistical filtration methods?

A. Yes, you could view and export p-Value for differentially expressed genes by right clicking on expression matrix and selecting **Show Data** option. (Check **p-Value** to view p-Values.)

Q. Can I view and export fold change values for differentially expressed genes identified by Fold change analysis modules?

A. Yes, you can view and export fold change for differentially expressed genes. Right click on expression matrix and select **Show Data** option. (Check Fold Change Values to view fold change values.)

Q. What are the memory requirements for Genowiz?

A. Although Genowiz will often run with 1GB RAM, we recommend 2 GB of memory for better performance.

Q. Unable to connect "localhost" at "3306". What is this error message?

A. This error message pops up when Genowiz is unable to connect to the database server. For trouble shooting click on **Open Connection** button, which pops-up a “Database Connection” dialog. The same dialog can also be opened from **File > Database** option. The error could occur if the database server is not started. The error could also be because the parameters specified in the “Database Connection” dialog are incorrect. Make sure that **Port** number specified is the same as the port on which the server is running and the rest of the parameters are as given below.

Type: MySQL (Database server vendor)

Server: localhost (The machine on which MySQL database server is running with Genowiz database.)

Database Name: genowiz

User Name: genowiz_admin

Password: microarray

Enter the above details and save the connection with a name, and also set this connection as default one. Try connecting now. On successful connection, a message - **Successfully connected** is displayed.



GENOWIZ FAQs

You can close “Database” dialog and start working with the software. If you face any other problems, check permission on the Genowiz installed folder to all users of your system.

Please make sure that the database server is up and running before opening Genowiz.

Q. What operating systems does Genowiz support?

A. Genowiz runs on Windows 98, Windows NT, Windows 2000, Windows XP, Macintosh OS, and Linux.

Q. My computer has enough memory. But Genowiz reports "Out of memory error". How do I fix this problem?

A. In case you encounter "Out of memory error" even though you have enough memory, then you need to change the 'Genowiz.lax' file in Genowiz installation folder. Open the Genowiz.lax file in any editor. The parameters 'lax.nl.java.option.java.heap.size.initial' and 'lax.nl.java.option.java.heap.size.max' specify the minimum and maximum heap size allocated to Genowiz respectively. You can edit this value to increase the memory as per our recommended values

Total Memory	<i>lax.nl.java.option.java.heap.size.min</i> (bytes)	<i>ax.nl.java.option.java.heap.size.max</i> (bytes)
1 GB	536870912	805306368
2 GB	805306368	1717986918

If the values exceed 60% of your total memory, it would lead to poor performance. The lax.nl.java.option.java.heap.size.max parameter should not exceed 1.6 GB for Windows, 2 GB for Linux and 2GB for MAC OS X. The above specifications are applicable only for 32 bit machines. Values can be further increased for 64 bit machines. Restart system after making changes to .lax file.

Q. How can I contact the technical team in case of any queries?

A. Use the **Technical Support** option from **Help** menu to e-mail the query details. You may also



GENOWIZ FAQs

contact via phone, e-mail or 24x5 live support on the web.

Q. I have created datasets and I want to find genes common to these datasets. How to I find?

A. Select the dataset nodes and then go to **Utilities > Gene Lists Comparison**. You will see a Venn diagram representation of nodes. You can view presence of genes by choosing among the combinations available.

14 Others

Q. What is the meaning of error message "Select a valid data node"?

A. Genowiz performs analysis on selected node. If selected node does not hold any data for further analysis, it pops up this error message. Ensure that the selected node has genes associated with it.

Q. How is Genowiz unique from other tools available?

A. Genowiz out performs other tools with its extensive biological analysis, automated work flows and robust statistical analysis. Moreover, it is very user friendly.