

MetaOmGraph User Guide

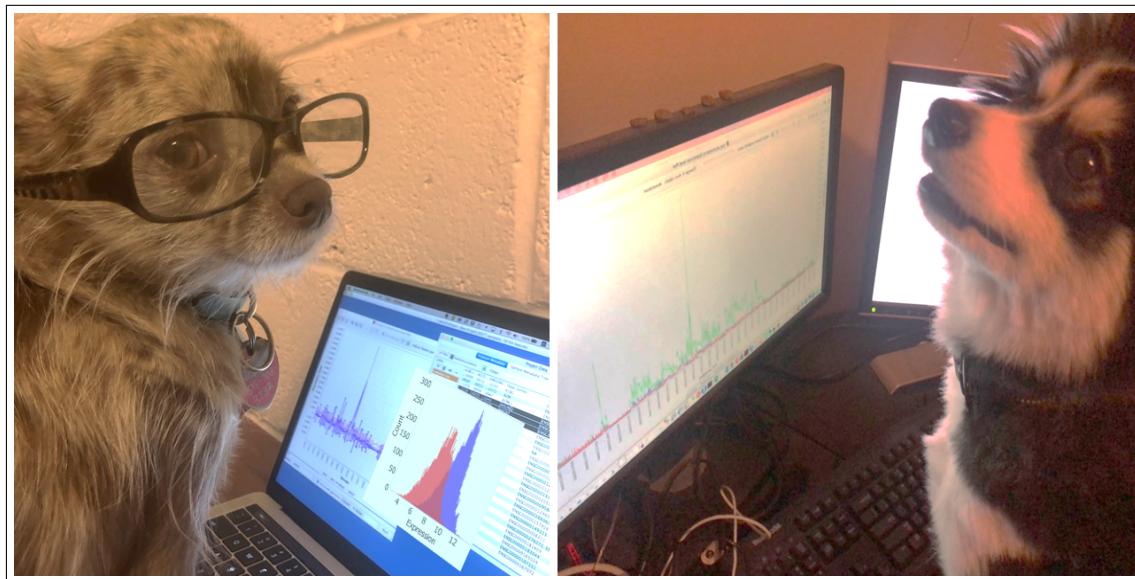
Urminder Singh and Eve Syrkin Wurtele
Bioinformatics and Computational Biology,
Center for Metabolic Biology,
and Genetics, Development, and Cell Biology
Iowa State University

October 28, 2019

Website: http://metnetweb.gdcb.iastate.edu/MetNet_MetaOmGraph.htm

GitHub (Source Code): <https://github.com/urmi-21/MetaOmGraph/>

Citation: Singh, Urminder, Manhoi Hur, Karin Dorman, and Eve Syrkin Wurtele. "MetaOmGraph: a workbench for interactive exploratory data analysis of large expression datasets." bioRxiv (2019): 698969.



Contents

1	Introduction	3
1.1	How to use this guide	3
1.2	Contributors	4
1.3	Citation	4
1.4	License	4
2	THE BASICS: Downloading and Using MOG	5
2.1	Download MOG	5
2.2	Download a pre-compiled project	5
2.3	Start MetaOmGraph	6
3	Open a Project	7
4	The Main MOG GUI: Window to Data Analysis	8
4.1	Save a current project	11
4.2	Open an existing project	12
5	Sort, subset, transform, analyze, and reorder the Data	14
5.1	Choosing the Replicate Column	14
5.2	Transform Data	14
5.3	Feature Lists	14
5.4	Search or Filter Sample Metadata	18
6	Coexpression Analysis	23
6.1	Correlation Between Features	23
6.2	Correlation Matrices	25
6.3	Correlation Between Samples	26
7	Visualizations	28
7.1	Line Charts	28
7.2	Line Chart with Averaged Replicates	31
7.3	Scatter Plot	32
7.4	Box Plot	33
7.5	Histogram	34
7.6	Volcano Plot	35
7.7	Correlation Histogram	36
8	Differential Expression Analysis	37
8.1	Differential Expression Analysis window	37
8.2	Differential Expression Results window	39
9	Create Your Own Projects	41
9.1	Input format	41
9.2	Start a new project	41
10	Interface to R	46
10.1	R script format	46
11	Change Project Properties and MOG Properties	47
11.1	Project properties	47
11.2	MOG properties	48

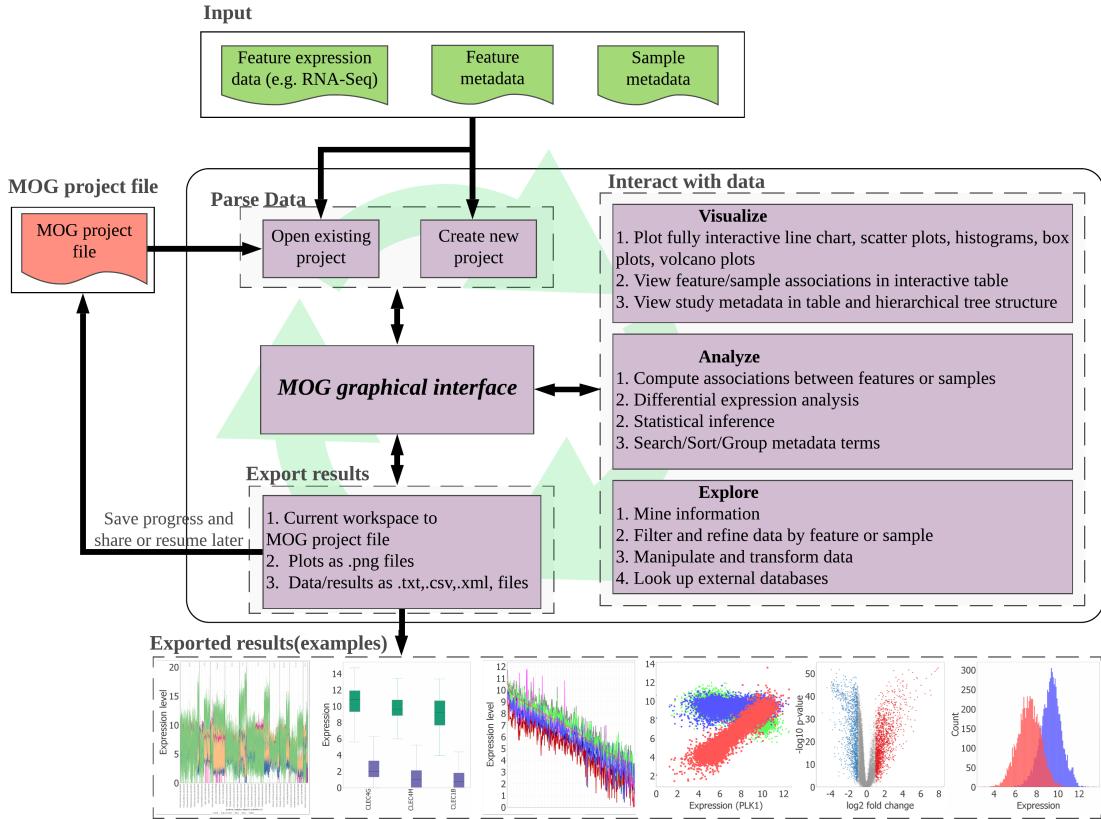


Figure 1: An overview of MOG’s modules.

1 Introduction

MetaOmGraph (**MOG**) is user-centered software written in Java to interactively explore and visualize large datasets. MOG can handle big datasets by efficient handling of data files. This is achieved via a combination of data indexing and buffering schemes.

MOG is specialized for biological expression datasets, and it is designed to be flexible to accommodate different types of data (e.g., taxes, finances, sports, revenues). MOG allows a user to analyse the data and its underlying metadata together; this adds another dimension to the analyses and provides flexibility in data exploration. It combines the ability

to handle very large heterogeneous data sets in real time with statistical analysis, list-making, and visualization capabilities. It also provides an interface to the R statistical platform, enabling use of the full range of R’s statistical and visualization capabilities for smaller-data analysis.

1.1 How to use this guide

MOG is an interactive software with lots of functionality and a fairly simple and intuitive GUI. First time users are encouraged read [THE BASICS](#) section, which describes how to get started with MOG. User seeking help with specific topics can directly proceed to that section. We have several large projects for particular organisms (*A thaliana*, human, yeast, maize) for you to use. If you’d like to create your own MOG projects, check out the [CREATING PROJECTS](#) section.

1.2 Contributors

1.3 Citation

Please cite MOG as: Singh, Urminder, Manhoi Hur, Karin Dorman, and Eve Syrkin Wurtele. "MetaOmGraph: a workbench for interactive exploratory data analysis of large expression datasets." bioRxiv (2019): 698969.

1.4 License

This work is licensed under the [MIT license](#).

Glossary

Feature-The item being examined. In this guidebook we use transcripts and genes as example features.

Sample- A representative part or a single item from a larger whole; a finite part of a statistical population whose properties are studied to gain information about the whole. [This definition can be confusing, since biologists often define a sample to be composed of multiple replicates, and many major databases do as well.]

Metadata-Additional information about a feature ("feature metadata") or a sample ("sample metadata").

GUI- Graphical User Interface. The windows the user sees and interacts with.

MOG Project- A MOG project has two main data components: the *feature metadata and data* and the *sample metadata*.

The Feature Metadata Table provides an interface to interact with the feature data. The Sample Metadata Tree and Sample Metadata Table provide interfaces to interact with the sample metadata. The Tree and Table are representations of the identical metadata *object*. MOG projects are saved to *.mog* files which can be read by MetaOmGraph.

2 THE BASICS: Downloading and Using MOG

2.0.1 System Requirements

Operating Systems

- Windows 10/8.x/7
- Mac
- Linux

Recommended Minimum Hardware

- 1.3 GHz or faster processor
- 2 GB RAM or greater
- 1024*768 screen resolution

Required Software

- Java Runtime Environment 8 (or higher)
- R 3.4 (or higher) [optional]

2.1 Download MOG

MOG is freely available to download from http://metnetweb.gdcb.iastate.edu/MetNet_MetaOmGraph.htm. Click the download button, and then download the .zip file. Unzip the downloaded file to get a .jar file, this is the MOG program.

MOG's source code is available at <https://github.com/urmi-21/MetaOmGraph/>.

2.2 Download a pre-compiled project

It's a lot easier to get started with an existing project. So select a project from On the same page that you download

MOG from (http://metnetweb.gdcb.iastate.edu/MetNet_MetaOmGraph.htm), there are many pre-compiled MOG projects to choose from. These are vetted projects, most containing public transcriptomics data of thousands of samples of RNA-Seq (from NCBI-SRA) or Affymetrix, and their metadata. Because we download existing metadata, it is only as accurate and comprehensive as what the researchers entered when they submitted it.

http://metnetweb.gdcb.iastate.edu/MetNet_MetaOmGraph.htm and download it. **Keep each of the three project files (the .mog file, the metadata file, and the data file) in the same directory** (It is OK to add other files to the folder).

2.3 Start MetaOmGraph

DOUBLE CLICK on the **.jar** file icon.

After starting MOG correctly, MOG's welcome dialog (Figure 2) will be displayed on the screen.

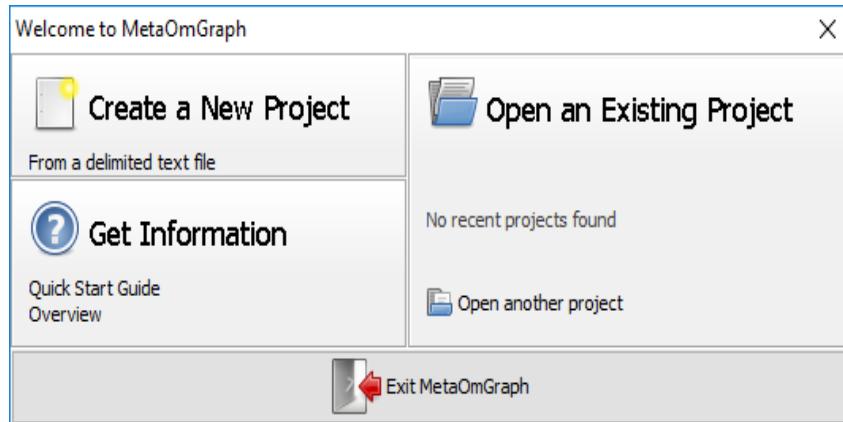


Figure 2: MOG welcome dialog.

3 Open a Project

The first time you work with MOG, MOG welcome dialog will not have an "existing project" (These will appear later, based on the projects that you open).

CLICK on the "**Open another project**" icon, and locate a project file in your file browser. (see [Section 2.2](#)). Follow the prompts and the MOG main window will open.

CURRENT PROJECTS. We have placed large, pre-made projects on the website for your use. These include: human, yeast, and Arabidopsis RNA-Seq data and metadata; Arabidopsis microarray data and metadata; and more (See [Create Your Own Projects](#)).

4 The Main MOG GUI: Window to Data Analysis

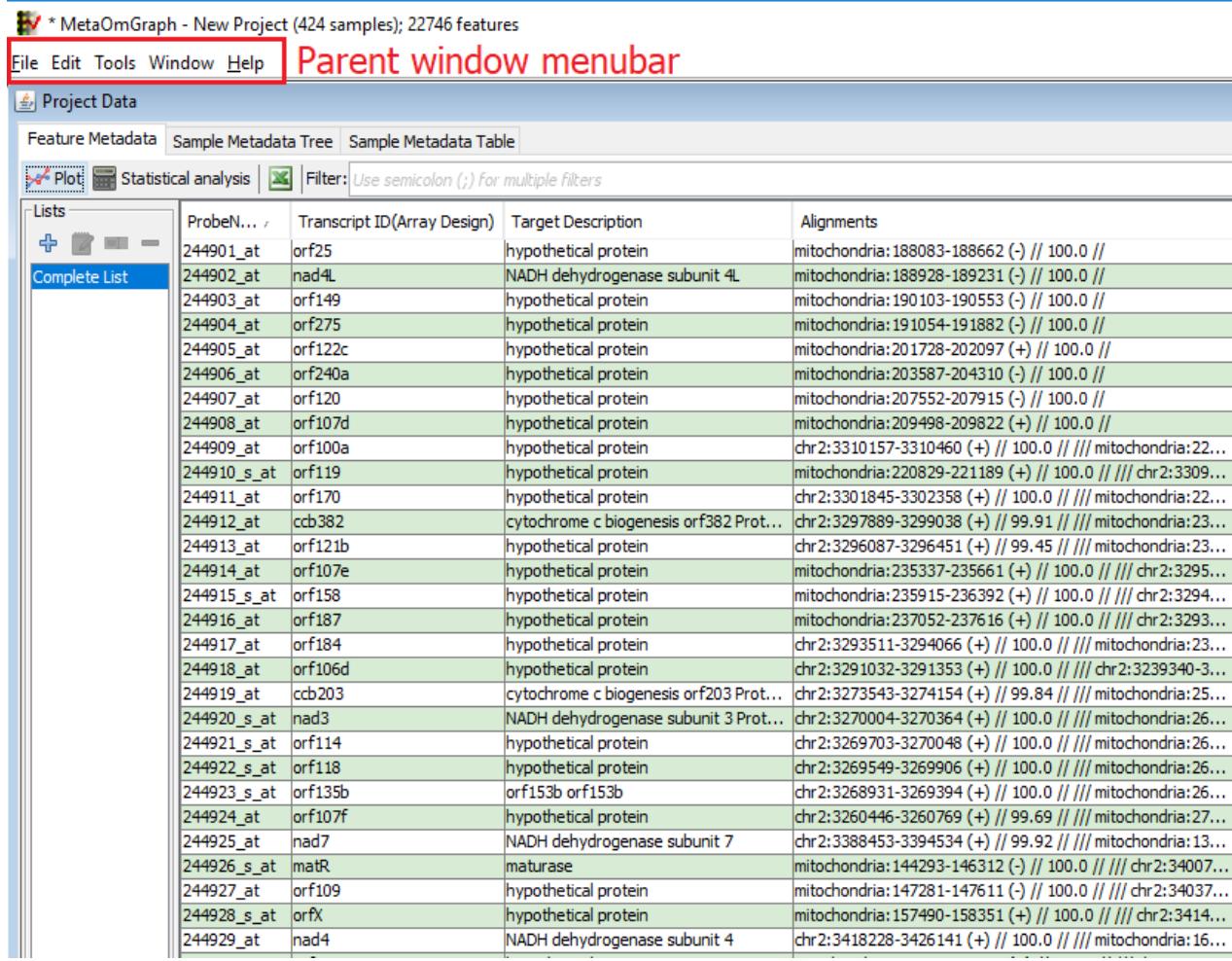


Figure 3: MOG main window.

Click on an existing .mog project; MOG will import the data and metadata and create a new .mog project. The "[Main MetaOmGraph](#)" window will be visible (Figure 3) that displays the data and feature metadata. Above the main window is a menu bar. (In MAC, this menu bar is at the top left, and not attached to the main window.) The "[Main MetaOmGraph](#)" window is the place to access all the data. From it, you can perform analyses, and create visualizations. It contains three panes:

Feature Metadata Displays each **feature metadata columns** from the data file. The user can select or search the features based on their metadata. (In Figure 3, the "features" are transcripts.)

Sample Metadata Tree Displays the metadata describing each sample in a tree format. It allows a user to search or filter samples based on sample metadata.

Sample Metadata Table Displays the sample metadata in a table format. It allows a user to search or filter the samples based on their metadata. It includes options to analyse or plot metadata. The user can export the metadata in a text format or XML format.

4.0.1 The layout of the "Feature Metadata" Table

The "[Feature Metadata](#)" pane in the "[Main MetaOmGraph](#)" window lets a user search or select features (genes in this example), and make feature lists (leaf DWN and leaf UP in this example).

Feature Metadata Table Displays the unique ID for each feature (required) and metadata about each feature (optional). Users can use select or search features of interest in the table based on its metadata; these features can be

The screenshot shows the Feature Metadata pane with a red border around the main data area. The menu bar at the top includes options for Plot, Statistical analysis, Filter, and Menubar. The List panel on the left contains a 'Complete List' section with items like 'leaf_DWN_0.001' and 'leaf_UP_0.001'. The main table has columns: ProbeName, Transcript ID(Array Design), and Target Description. The data rows are as follows:

ProbeName	Transcript ID(Array Design)	Target Description
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydrogenase subunit 4L
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein
244910_s_at	orf119	hypothetical protein
244911_at	orf170	hypothetical protein

Feature metadata table

Figure 4: Feature Metadata pane. The first column, ProbeName, contains the unique feature IDs.

analyzed or visualized.

Menubar The menubar directly above the "Feature Metadata" table provides options to screen features, analyze the data, visualize the data, save and search. Options are:

1. Plot: Reveals options to visualize the feature data across the samples.
2. Statistical analysis: Reveals options to determine coexpression or cluster the features based on their numerical values across the samples.
3. Export: Exports the metadata displayed in the feature metadata table to a .txt file. Note if a sublist has been selected (see List Panel), then only the metadata of the features in that sub-list will be exported.
4. Filter: Type into the *textfield* to filter the Feature Metadata Table, retaining only those rows that match user-input text .

List Panel The List Panel (far left) contains user-created lists for the project. A list is a collection of features; in this example, one list contains those genes that are downregulated in leaf compared to pollen (leafDWN). Lists can be added, edited, merged, named and saved.

1. Opens the "Create New List" dialog, through which a new list can be created (see [Feature Lists](#)).
2. Opens a user-selected list within the "Create New List" dialog and enables the user to edit the list.
3. Allows a user to rename a selected list (ends up far more useful than one might anticipate).
4. Deletes the list that the user has selected.

The lists can be manipulated in more complex ways (see [Feature Lists](#)).

4.0.2 The "Sample Metadata Tree" pane

Menubar The upper bar is a menubar containing the options:

1. File: Exports the sample metadata as an XML file
2. Search: Opens a search dialog, from which the user can search the metadata using multiple queries. The matching results are highlighted.
3. Edit: Shows the search dialog. This allows the user to filter metadata based on queries. The filtered metadata could be reversed. See [Section 5.3](#) for details.

Feature Metadata Sample Metadata Tree Sample Metadata Table

Menubar

The screenshot shows the 'Sample Metadata Tree' pane. At the top, there are three tabs: 'Feature Metadata', 'Sample Metadata Tree' (which is selected and highlighted in blue), and 'Sample Metadata Table'. Below the tabs is a menu bar with 'File', 'Search', 'Edit', and 'Help'. The main area is divided into two panes. The left pane is a hierarchical tree view with a yellow border around the selected node. The right pane is a table with a green border, showing metadata for the selected node.

Attribute	Value
Experiment	Targets of the mci genes.
lab_head	Dr Brendan Davies
experimenter_name	Dr Brendan Davies
experiment_type	genetic_modification_design
experiment_name	Targets of the mci genes.
experiment_factor	NA
experiment_description	In Antirrhinum, the equivalent mutant to the Arabidopsis factor homologous to CUC1 and CUC2. Yeast two-hybrid analysis in Arabidopsis, the closest homologues to TIC110 and CUC2. We have identified insertions in both TIC110 and CUC2. We expect TIC110 to be evolutionarily conserved interactors, we expect TIC110 to be involved in the boundaries of organ identity out of register with CUC2. The heterozygote is identical to the tcp13 homozygote. We have identified insertions in both TIC110 and CUC2. We expect TIC110 to be evolutionarily conserved interactors, we expect TIC110 to be involved in the boundaries of organ identity out of register with CUC2. The heterozygote is identical to the tcp13 homozygote. The aim of this microarray experiment is to generate these plants. The experimental material will serve as internal replicates. The fact that all of the mutant combinations observed show similar homeotic mutants, very similar sets of floral genes being expressed. Our tissue samples will be wild type and taking a large number of flowers and buds we hope to identify the genes involved in organ identity.

Metadata in tree format

Metadata in table for selected node

Figure 5: Sample Metadata Tree pane.

Tree panel The left panel is a hierarchical representation of the sample metadata. Any node selected in the tree updates the content in the table at the right.

Table panel When a node is selected in the left panel, the table shows the metadata for the selected node and all its child nodes.

4.0.3 The "Sample Metadata Table"

Menubar The menubar contains the options:

1. File: Export options for the metadata tables as .tsv file.
2. Analyze: Analyze the data (sample-wise) in the .mog project. See [Sample Correlations](#).
3. View: Plot the data (rows v columns). See [Visualizations](#).
4. Search: Opens the search dialog. The user can search the metadata using multiple queries; MOG highlights the matched results. See [Metadata search](#).
5. Edit: Contains options for filtering and removing metadata. This allows the user to filter metadata based on text queries. (The metadata filter can be reversed). See [Metadata search](#) for details.

Menubar

Experiment	alias_or_g...	chip	chip_name	developm...	diseased	experime...	experime...	experime...	experime...	experime...
AtGenExpress...	Col-0	ATGE_31_A2	ATGE_31_A2	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_32_A2	ATGE_32_A2	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_33_A	ATGE_33_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_34_A	ATGE_34_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_35_A	ATGE_35_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_36_A	ATGE_36_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_37_A	ATGE_37_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_39_A	ATGE_39_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_40_A	ATGE_40_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_41_A	ATGE_41_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_42_B	ATGE_42_B	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_43_A	ATGE_43_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_45_A	ATGE_45_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	dv3-7	ATGE_53_A	ATGE_53_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	fy1-12	ATGE_54_A	ATGE_54_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	ap1-15	ATGE_55_A	ATGE_55_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	ap2-6	ATGE_56_A	ATGE_56_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	ap3-6	ATGE_57_A	ATGE_57_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	ag-12	ATGE_58_A	ATGE_58_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	ufo-1	ATGE_59_A	ATGE_59_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_73_A	ATGE_73_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_76_A	ATGE_76_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_92_A	ATGE_92_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_5_A	ATGE_5_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_10_A	ATGE_10_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	glIT	ATGE_11_A	ATGE_11_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_12_A	ATGE_12_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_13_A	ATGE_13_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...
AtGenExpress...	Col-0	ATGE_14_A	ATGE_14_A	NA	NA	AtGenExpress...	NA	AtGenExpress...	genetic_modifi...	AtGenExpress...

Figure 6: Sample Metadata Table pane.

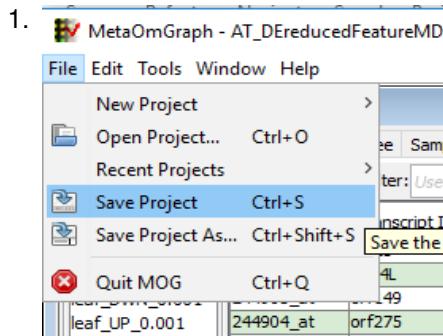
6. Help: Opens the help page.

Table panel The sample metadata is displayed in a table in the "Sample Metadata Table". This table shows the metadata for the samples currently included in the project. Individual rows from this table could be selected using the mouse. The table columns could be sorted by clicking the header of each column.

4.1 Save a current project

At any point, the current project can be saved. Once a project is saved, the user can directly open this project and can share it with others. Saving a project saves all the lists and analysis done with the project. MOG projects are saved with a *.mog* extension.

To save a project:



CLICK "File" in the "parent window"

- CLICK "Save" to save the project. (If saving a new project or renaming an existing project, CLICK "Save As" and name the project.)

Note: MOG asks a user to save the project if s/he tries to exit MOG with unsaved changes in the project. When a project is saved MOG remembers the location of the data and metadata files, and other configurations used when creating the project, and thus can re-open the files exactly as created.

4.2 Open an existing project

When starting MOG, any existing project can be opened from the "Welcome Dialog" (Figure 7). The right section in the "Welcome Dialog" shows a list of recently opened projects. Any project can be clicked and opened. If the project you want is not listed, CLICK on the "Open another project" option in the "Welcome Dialog", browse for the project, and open it.

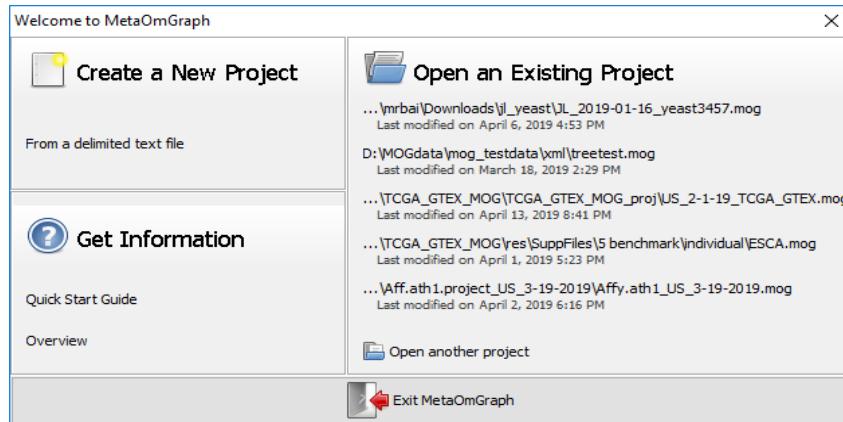


Figure 7: MOG welcome dialog with a list of recent projects.

When *already* using MOG, to close the current project and open another existing project:

1. CLICK "File" in the "parent window" (see Figure 8).
2. CLICK "Open Project". A file-chooser dialog will appear. Locate the required *.mog* and CLICK "Open". *OR*
3. For recently-used projects, CLICK "Recent Project" (see Figure 8). A file-list will appear. Select the *.mog* project you want.

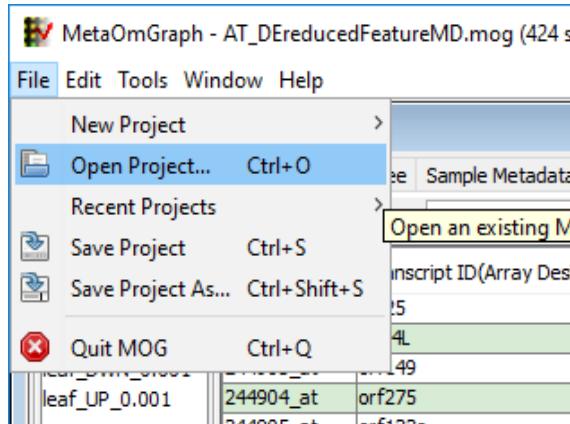


Figure 8: Click on Open Project or Recent Project to open a new project.

4.2.1 Troubleshooting opening projects

If MOG fails to open an existing *.mog* project, please check the following:

- Make sure that the **data** and **metadata** files used to create the project are the same folder (directory) as the *.mog* file. If the files have been moved, MOG will ask the user to locate the files.
- Make sure that the **data** and **metadata** files have not changed since the saving of the project. For example, changing the columns in the either of these files can cause parsing errors, and require making a new project.
- Make sure the *.mog* project file hasn't been modified by any program other than MOG.

- If sharing a MOG project with another user, make sure to share the **data**, **metadata** and **.mog** files in a single folder. To open a shared project, keep these three files in a common folder.
- Make sure the **.mog** file is compatible with the version of MOG being used to open it. All MOG projects created using version 1.5.5 or earlier are not compatible with MOG version 1.6.1 and later.

If this information doesn't resolve the issue then please contact the developers of MOG for assistance.

5 Sort, subset, transform, analyze, and reorder the Data

MOG's data-metadata linkage enables it to map each data point in the data file to a vector of metadata terms associated with the samples. Such mapping transforms a two dimensional data file into a virtually multidimensional dataset that can be explored from different perspectives.

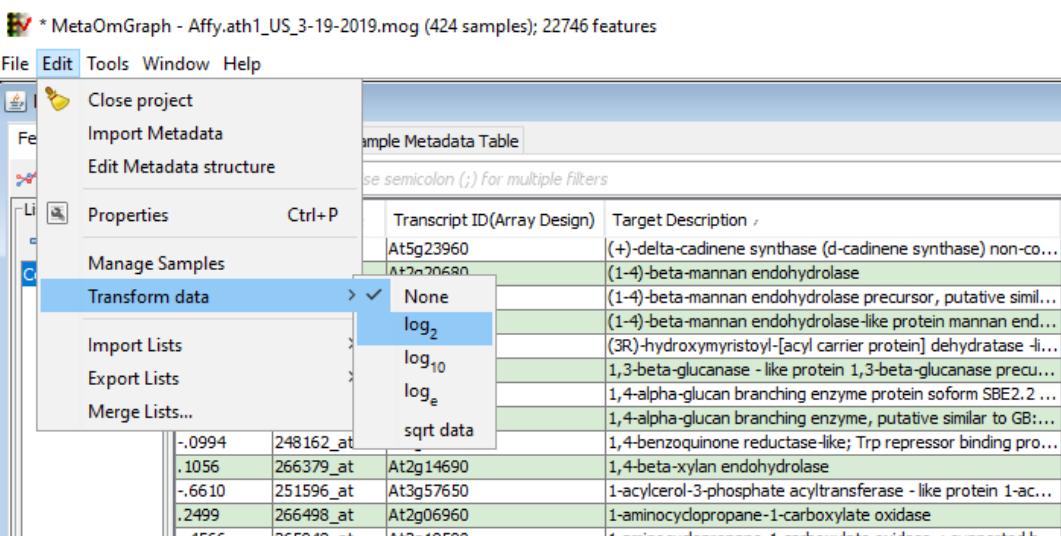
MOG links the two data components virtually upon parsing the data and metadata files. This means that manipulating the samples through the metadata table or metadata tree will also affect the feature data. E.g., if a user analyzes a subset of RNA-Seq samples, the Pearson correlation between two genes will change according to their values in that subset.

5.1 Choosing the Replicate Column

MOG keeps a variable called "Replicate column" which stores a chosen factor (or column), from the sample metadata, by the user. This column is then used in plotting average values (by averaging over the replicates) or for defining blocks for the permutation test. By default, the replicate column is chosen to be the one which is at the topmost level in the hierarchical structure defined by the user (see [Section "Metadata Table to Tree"](#)). A replicate column could be changed via the properties panel.

5.2 Transform Data

MOG has options to apply commonly-used transformations to the data. Apply a data transformation by:

1.  From the "Parent Window Menubar", CLICK "Edit – > Transform Data"
The screenshot shows the MOG software window. The menu bar at the top has 'File', 'Edit' (which is currently selected), 'Tools', 'Window', and 'Help'. Below the menu bar is a toolbar with icons for project management. The main area is titled 'Sample Metadata Table' and contains a table with columns for Transcript ID (Array Design), Target Description, and several rows of gene information. On the left, there is a sidebar with options like 'Properties', 'Manage Samples', 'Transform data' (which is highlighted in blue), 'Import Lists', 'Export Lists', and 'Merge Lists...'. A dropdown menu under 'Transform data' shows options: 'None', 'log₂', 'log₁₀', 'log_e', and 'sqrt data'. The 'log₂' option is currently selected.

Transcript ID (Array Design)	Target Description
At5g23960	(+)-delta-cadinene synthase (d-cadinene synthase) non-co...
At5g23960	(1-4)-beta-mannan endohydrolase
	(1-4)-beta-mannan endohydrolase precursor, putative simil...
	(1-4)-beta-mannan endohydrolase-like protein mannan end...
	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase -li...
	1,3-beta-glucanase - like protein 1,3-beta-glucanase precu...
	1,4-alpha-glucan branching enzyme protein isoform SBE2.2 ...
	1,4-alpha-glucan branching enzyme, putative similar to GB:...
	1,4-benzoquinone reductase-like; Trp repressor binding pro...
.1056	1,4-beta-xylan endohydrolase
-.6610	1-acylcerol-3-phosphate acyltransferase - like protein 1-ac...
.2499	1-aminocyclopropane-1-carboxylate oxidase
-4566	1-aminocyclopropane-1-carboxylate oxidase - supported h...
2. Choose a data transform (e.g., log₂). After the transformation is applied, all subsequent analysis/visualizations will use the transformed data.

5.3 Feature Lists

The features in the data file can be arranged into virtual subsets (lists). This allows for easy access to features of interest. For example, from a RNA-Seq dataset of human disease, the user can make a list of all the genes up-regulated in breast cancer.

5.3.1 Create a New List

1.

ProbeName	Transcript ID (Array Design)	Target Description
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydratase
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein

From the "Feature Metadata Table", select the rows to be included in the list.

(This is just one way to make a list. Lists can also be created from imported texts files, or after statistical analyses (e.g., see [Section 8.2](#).)

2.

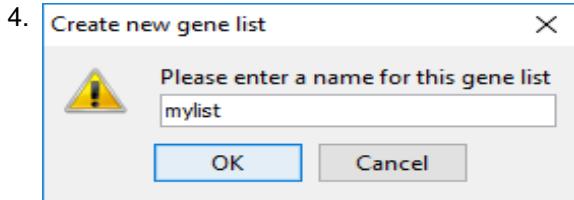
The 'Create a new list' button is highlighted with a red circle.

In the List Panel (to left of the "Feature Metadata Table"), CLICK the + icon ("Create New List").

3.

ProbeName	Transcript ...	Target Des...
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydratase
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein

A "Create New List" panel will be displayed. The table on the right ("In List") displays the features of the new list; the table on the left ("Not In List") displays left-out features. This panel allows a user to add or remove features from the list. Users can search features by CLICKing the "Search" button at top of each table. Once you're done, CLICK "Create".



Enter a name and CLICK "OK" to save the list.

5. **Feature Metadata** **Sample Metadata Tree** **Sample Metadata Table**

	265073_...	245106_...	ProbeN... /	Transcript ID(Array Design)
-.1123	-.2556	244901_at	orf25	
-.2249	-.1002	244902_at	nad4L	
-.1976	.2693	244903_at	orf149	
.0577	.1732	244904_at	orf275	
.1984	-.1933	244905_at	orf122c	
-.0988	-.1406	244906_at	orf240a	
-.0650	-.0939	244907_at	orf120	
-.0860	-.2380	244908_at	orf107d	
-.0748	-.1432	244909_at	orf100a	
.0126	-.2053	244910_s_at	orf119	

The new list will be added to the List Panel.

5.3.2 Edit a List

1. **Feature Metadata** **Sample Metadata Tree** **Sample Metadata Table**

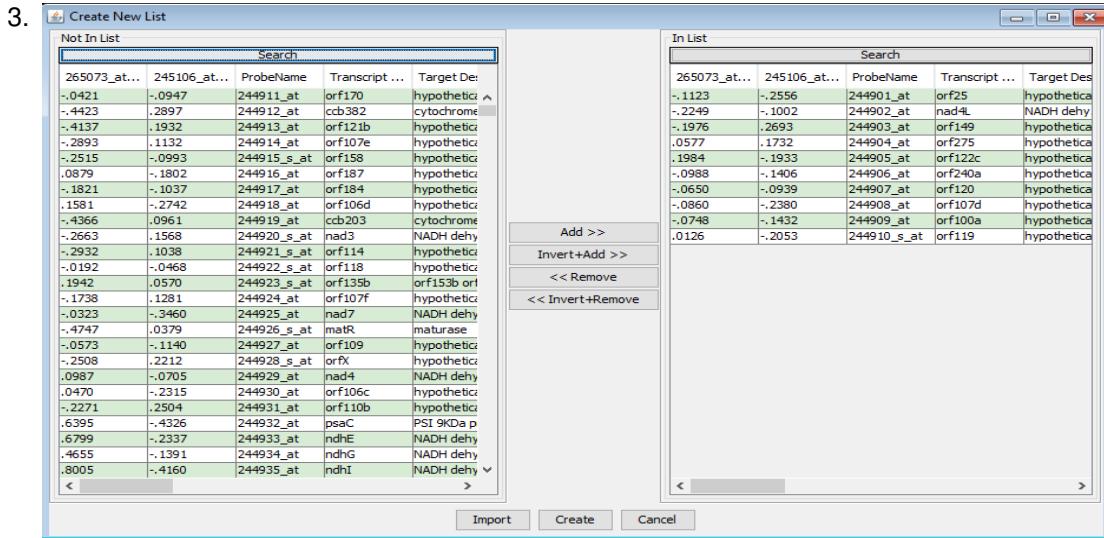
	265073_...	245106_...	ProbeN... /	Transcript ID(Array Design)
-.1123	-.2556	244901_at	orf25	
-.2249	-.1002	244902_at	nad4L	
-.1976	.2693	244903_at	orf149	
.0577	.1732	244904_at	orf275	
.1984	-.1933	244905_at	orf122c	
-.0988	-.1406	244906_at	orf240a	
-.0650	-.0939	244907_at	orf120	
-.0860	-.2380	244908_at	orf107d	
-.0748	-.1432	244909_at	orf100a	
.0126	-.2053	244910_s_at	orf119	

In the List Panel (to left of the "Feature Metadata Table"), select the list to be edited.

2. **Feature Metadata** **Sample Metadata Tree** **Sample Metadata Table**

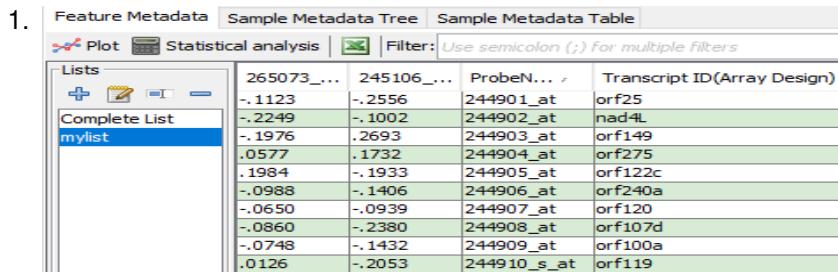
	265073_...	245106_...
-.0988	-.1406	
-.0650	-.0939	
-.0860	-.2380	

CLICK the "Edit List" Button.

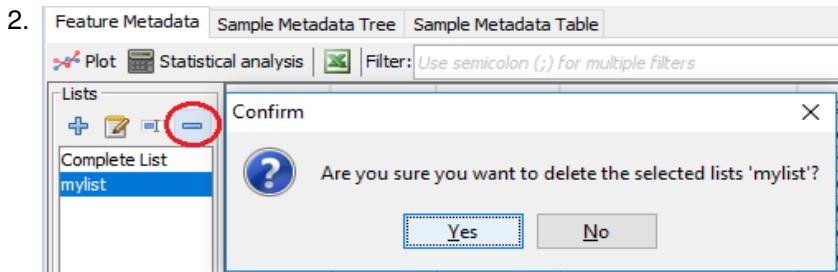


A "Create New List" panel will be displayed. Using this panel, the list can be edited as described in Step 3 of creating a new list. After editing the list CLICK OK to save the changes.

5.3.3 Delete a List



In the list panel (left of "Feature Metadata" pane), select the list to be deleted.



CLICK the - icon ("Delete List") and CLICK "OK" when asked for confirmation.

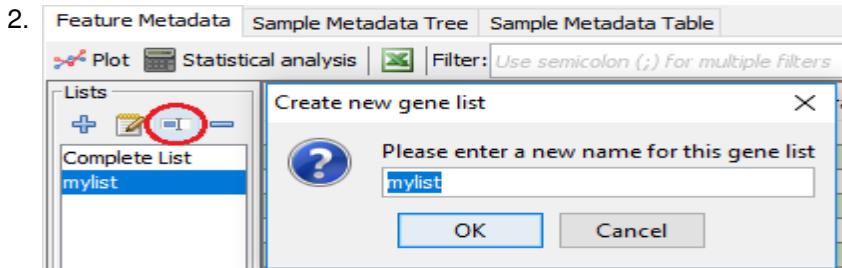
5.3.4 Renaming List

To rename an existing list perform the following steps:

1.

	265073_...	245106_...	ProbeN...	Transcript ID(Array Design)
-.1123	-.2556	244901_at	orf25	
-.2249	-.1002	244902_at	nad4L	
-.1976	.2693	244903_at	orf149	
.0577	.1732	244904_at	orf275	
.1984	-.1933	244905_at	orf122c	
-.0988	-.1406	244906_at	orf240a	
-.0650	-.0939	244907_at	orf120	
.0860	.2380	244908_at	orf107d	
-.0748	-.1432	244909_at	orf100a	
.0126	-.2053	244910_s_at	orf119	

In the list panel located in the left of "Feature Metadata" pane, select the list to be renamed (mylist in this example).



CLICK the "Rename List" icon and enter a new name.

5.4 Search or Filter Sample Metadata

The search and filter operations allow the user to choose or exclude samples for analysis or visualization. For example, for a human RNA-Seq project, the user may want to analyze only the samples from liver tissue. The filtered-out samples will be excluded from any subsequent analyses done with MOG; they are not forever gone, a user can include them back in.

5.4.1 Search Sample Metadata

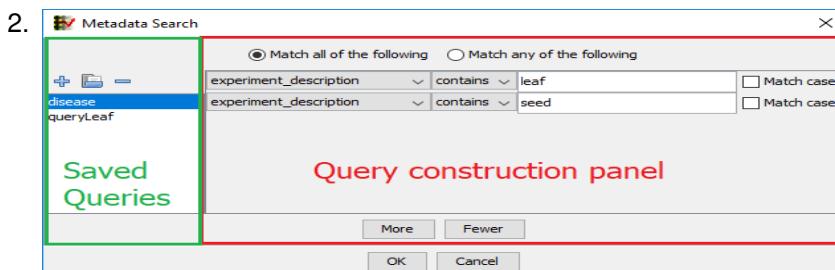
MOG allows a user to search the sample metadata using queries constructed via the "Metadata Search" panel [5.4.2](#). This panel allows a user to interactively create complex search queries by combining simple queries.

To search metadata:

1.

Experiment	alias_of	Search	chip_name
AtGenExpress:....	Col-0	Clear Last Search	GE_31_A2
AtGenExpress:....	Col-0	ATGE_32_A2	ATGE_32_A2
AtGenExpress:....	Col-0	ATGE_33_A	ATGE_33_A

From the "Sample Metadata Table pane" or "Sample Metadata Tree" pane select "Search" from the menubar.



A "Metadata Search" panel will be displayed. Create the search query in the "Metadata Search" panel (see Section [5.4.2](#)). and press OK.

3. Feature Metadata Sample Metadata Tree Sample Metadata Table

File Analyze View Edit Search Help

Experiment	alias_or_g...	chip	chip_name	developm...	diseased	experimen...
AtGenExpress:... Col-0		AtGen_A-1_17...	AtGen_A-1_17...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-5_21...	AtGen_A-5_21...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-9_22...	AtGen_A-9_22...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-13_2...	AtGen_A-13_2...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-18_2...	AtGen_A-18_2...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-21_2...	AtGen_A-21_2...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-25_2...	AtGen_A-25_2...	3.7	Normal	AtGenExpress:...
AtGenExpress:... Col-0		AtGen_A-29_2...	AtGen_A-29_2...	3.7	Normal	AtGenExpress:...

The search results will be highlighted in the Sample Metadata Table or the Sample Metadata Tree.

5.4.2 Metadata Search Panel

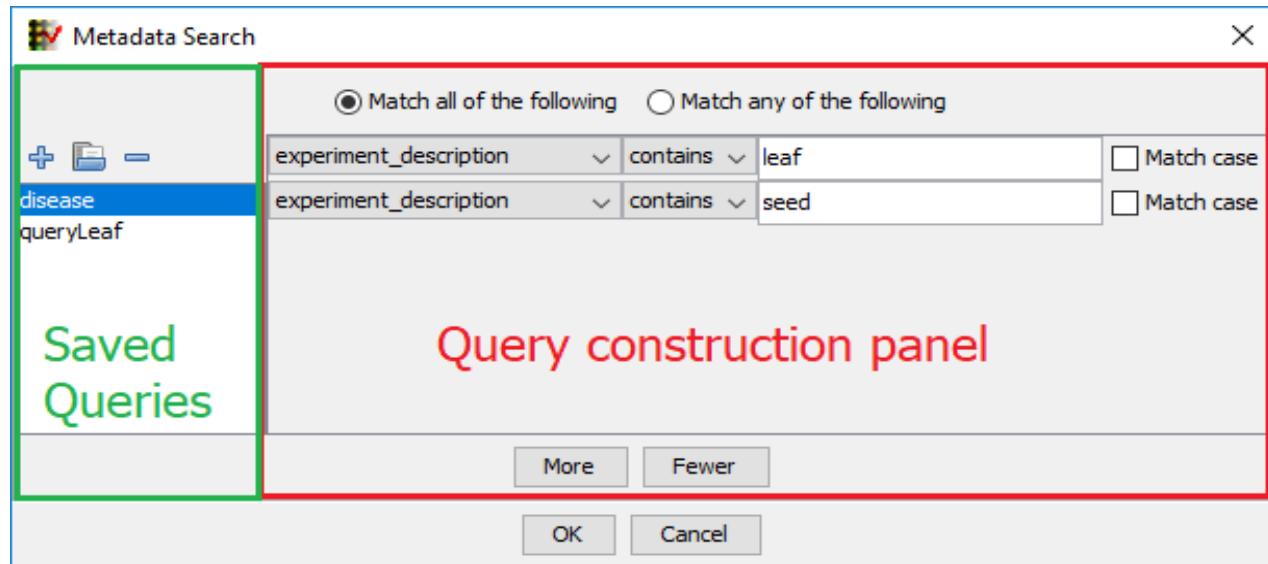
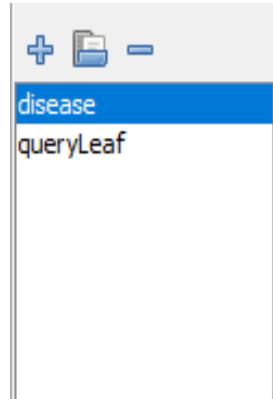


Figure 9: Metadata Search Panel.

The "Metadata Search" panel allows a user to enter search queries to search the metadata. It is divided into a query component (left) and a query construction component (right).

- "Query" panel



The "Query" panel shows the saved queries. A query can be saved or removed by CLICKing the icons at the top.

DOUBLE CLICKing an existing query loads that query to the "Query Construction" panel.

- "Query Construction" panel

Match all of the following Match any of the following

experiment_description	contains	leaf	<input type="checkbox"/> Match case
experiment_description	contains	seed	<input type="checkbox"/> Match case

More **Fewer**

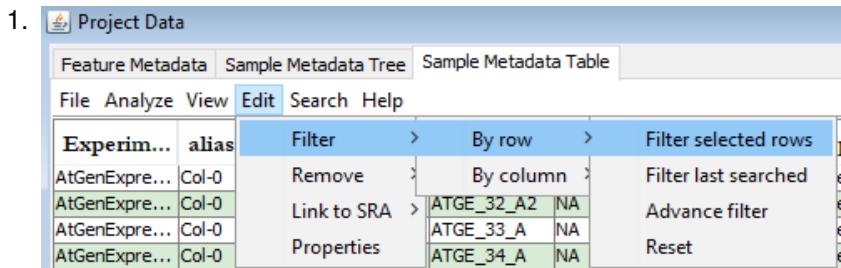
The "Query Construction" panel is where the query is constructed. Each row represents a single query. Multiple queries could be combined using the top radio buttons which can match all of the queries or match at least one of the query. CLICK the More button to add more search queries and Fewer to remove queries starting from the bottom.

TIPS: Pay attention to whether the "match all", or "match any" button is selected. Don't forget to save the search terms if you wish to use them again or keep a record of them.

5.4.3 Filter Samples

Samples in a project can be filtered-out *temporarily*. After filtering, the MOG Sample Metadata Tree and Sample Metadata Table will display only the included samples. All subsequent analysis will be limited to the included samples.

To filter samples:



In "[Sample Metadata Table pane](#)" select "Edit – > Filter – > By Row" from the Menubar. There are four filter options:

- Filter Selected Rows:** Filters selected rows (samples) in the metadata table. The user can specify to either keep or remove samples from the current filter.
- Filter Last Searched:** The most recently selected samples (highlighted due to the search operation) are shown. The user can specify to either keep or remove samples.
- Advance Filter:** This option will open the "[Advance Sample Filter](#)" window (described below) from which users can interactively filter samples. **Note** the same "[Advance Sample Filter](#)" window can be accessed from the "Parent window's menubar" by CLICKing "Edit – > Manage Samples".
- Reset:** This will reset the filter applied on the metadata and re-include all samples in the project.

- Once a filter is applied to sample metadata the sample metadata table and sample metadata tree will reflect those changes.

5.4.4 Advance Sample Filter Window

The "[Advance Sample Filter](#)" window allows a user to flexibly manage the samples in the project (Figure10).

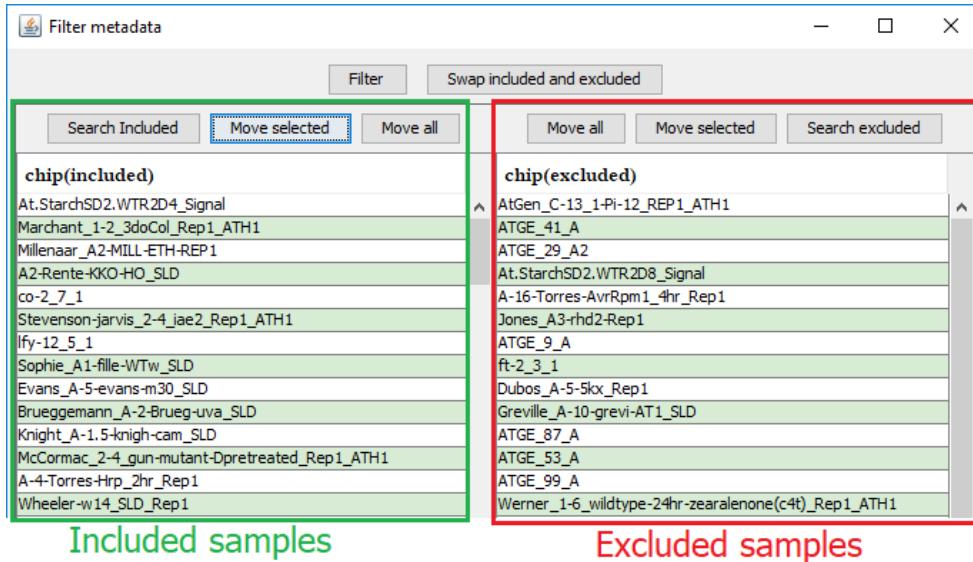


Figure 10: Advance Sample Filter Window.

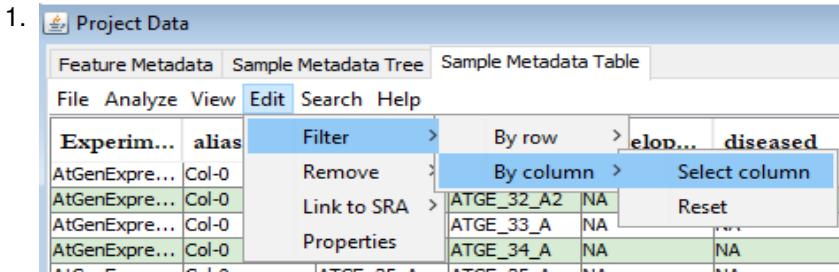
- This window contains a top button panel and two panels to display two tables.
- The **left panel** contains table and a button panel for the currently **included samples** in the project. The buttons "Search Included" will open up a "[Metadata Search](#)" panel to search in the currently included samples only. The samples could be searched, selected and moved to the excluded table using the "Move selected button" at the top.
- The **right panel** contains table and a button panel for the currently **excluded samples** in the project. The buttons "Search Excluded" will open up a "[Metadata Search](#)" panel to search in the currently excluded samples only. The samples could be searched, selected and moved to the excluded table using the "Move selected button" at the top.
- Once all user-determined samples are moved to the included or excluded sides of the tables, press the "Filter" button at the top button panel to set the included samples.

5.4.5 Filter Sample Metadata Columns

Entire Metadata Columns can be filtered out. This doesn't alter the analysis but does pare down the number of metadata columns in the .mog project.

Filter Columns		
Column number	Header	Keep/Remove
1	Experiment	Keep
2	alias_or_genotype	Keep
4	chip_name	Keep
5	development_stage_Paradigm	Remove
6	diseased	Keep
7	experiment_description	Keep
8	experiment_factor	Keep
9	experiment_name	Keep
10	experiment_type	Keep
11	experimenter_name	Keep
12	genetic_background	Keep
13	genetic_variation	Keep

Figure 11: Metadata Column Filter Window.



In "Sample Metadata Table pane" select "Edit – > – > By Column" from the menubar. There are two options:

- (a) **Select columns** will open a "Metadata Column Filter" window (Figure 11). In this window select the columns to remove and press "Done".
OR
- (b) Press **Reset** to remove any filter applied on the metadata, and include all metadata columns in the project.

2. Once columns are filtered, the Sample Metadata Table will reflect those changes.

6 Coexpression Analysis

Multiple methods are implemented in MOG to find correlations and associations among features or samples.

All methods statistical methods are accessed by CLICKing



Statistical analysis button in the menubar of the "Feature Metadata" pane.

The "Analyze" button in the menubar of the "Sample Metadata Table pane" displays the options available to calculate statistical associations between and among samples.

Coexpression methods implemented in MOG:

Pearson Correlation detects the linear dependency between two variables X and Y.

Spearman Correlation coefficient measures monotonic relationships between two variables X and Y.

Mutual Information (MI) quantifies the amount of information shared between two random variables. MI for two discrete random variables X and Y, having the joint probability $p(x, y)$ and marginal probabilities $p(x)$ and $p(y)$ respectively, is defined as:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

6.1 Correlation Between Features

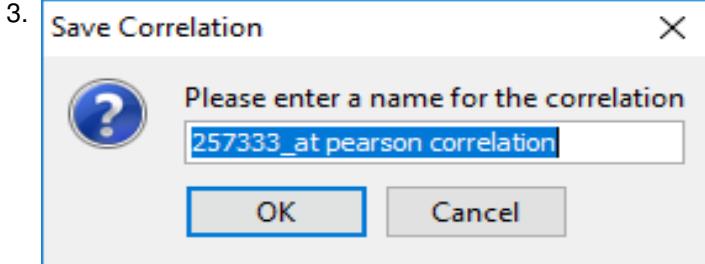
Correlations between features are helpful in finding features that are coexpressed across samples. To calculate pairwise correlations of a selected feature with all the other features in the list:

1.
The screenshot shows a table with columns: ProbeName, Transcript ID(Array D...), and Target Description. The rows list various probe names like 244950_at, 257333_at, etc., along with their corresponding target descriptions such as cytochrome c oxidase subunit 2 and 1.

In the "Feature Metadata" pane, select a single row to calculate the correlation against.

2.
The screenshot shows the "Statistical analysis" menu open. Under the "Correlation" section, "Mutual Information" is selected. A dropdown menu shows options like Mutual Information(No pval), Mutual Information(permute within groups), Mutual Information(permute all), Mutual Information matrix, and Relatedness matrix.

CLICK the "Statistical Analysis" button in the "Feature Metadata" pane menubar.



Select a coexpression method (Pearson, Spearman, or MI). When prompted, enter a name to save your results. This name will be saved in the "Feature Metadata" table.

4.

The screenshot shows the 'Feature Metadata' table. A red arrow points from the 'Filter' bar to the 'Created correlation column' label. The 'Created correlation column' label is in red text next to the table. The table lists various genes and their correlation values with the feature '257333_at pearson correlation'. The last row, which is circled in green, represents the correlation between 'cox1' and 'orf25' with a value of 0.7490.

	ProbeName	Transcript ID(Array Design)
1.0000	257333_at	cox1
.8457	263509_s_at	At2g07687
.8320	244943_at	nad9
.8144	265235_s_at	At2g07719
.8127	257321_at	orf106f
.8081	244950_at	cox2
.7954	244944_s_at	rpl16.mitochondria
.7833	265232_s_at	At2g07715
.7827	266044_s_at	At2g07725
.7766	244951_s_at	ccb452
.7669	263505_s_at	orf215b
.7573	257334_at	orf111d
.7490	244901_at	orf25

Correlation b/w cox1 and orf25 is 0.7490

The i^{th} row in the column represents the computed correlation value of selected feature with the i^{th} feature.

The column containing the coexpression results is saved with the MOG project. Correlation columns can be deleted by selecting "Remove Correlation" in step 2. The "View correlation details" option in step 2 shows the details (if available) about the computed correlation. For example, if a correlation is computed with a P-Value then the p-values could be seen with "View correlation details".

To remove a MOG-computed correlation:

1.

The screenshot shows the 'Statistical analysis' pane. A red arrow points from the 'Filter' bar to the 'Remove Correlation' button. The 'Remove Correlation' button is highlighted. Below it, a list of correlations is shown, including 'A1BG spearman correlation', 'ACCS mutual information', and 'SHROOM2 pearson correlation'. There is also a 'Remove all correlations' option.

CLICK the "Statistical Analysis" button in the "Feature Metadata" pane menubar. Select "Remove Correlation". This will display a list of saved correlations. Select the ones which you want removed.

6.1.1 P-value computation for correlations

MOG applies a permutation test to compute the p-values for correlations. MOG uses multithreading to speed-up this intensive computation, processing each permuted dataset in parallel (Algorithm 1). MOG provides Bonferroni corrections and Benjamini–Hochberg (BH) corrections to adjust the p-values for multiple comparisons.

For permutation (8) (Algorithm 1) the user can select blocks to permute the data, such that the data in the blocks are exchangeable under the null hypothesis. The blocks are determined by the experimental design. For, example if the metadata is structured in the tree of Figure 25, the user may choose to permute data only within samples. MOG keeps a variable called "Replicate column" which stores a chosen factor, from the sample metadata, by the user (see Section 5.1).

Algorithm 1 Calculate significance

```
1:  $P \leftarrow$  number of permutations
2:  $T \leftarrow$  number of threads
3:  $X \leftarrow$  expression values of first gene
4:  $Y \leftarrow$  expression values of second gene
5:  $\rho \leftarrow \text{association}(X, Y)$ 
6: extremes  $\leftarrow 0$ 
7: Execute in parallel P times using T threads :
8:      $X^* \leftarrow \text{permute}(X)$ 
9:      $\rho^* \leftarrow \text{association}(X^*, Y)$ 
10:    if  $|\rho^*| \geq |\rho|$  then
11:       extremes  $\leftarrow$  extremes + 1
12: end parallel
13:  $p\text{-value} \leftarrow \frac{\text{extremes}}{P}$ 
```

6.2 Correlation Matrices

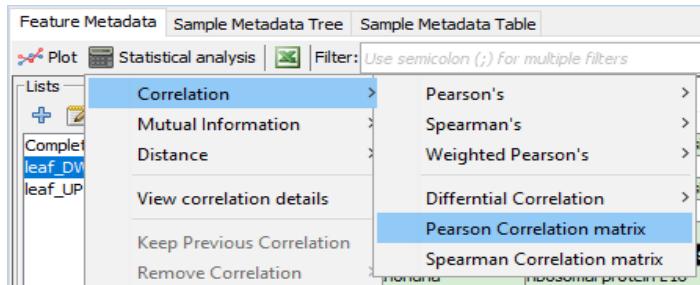


Figure 12: Selecting Pearson correlation matrix.

Pairwise matrices of correlations of each feature to each other feature serve as the basis for clustering. This has been implemented for Pearson and Spearman correlations.

1. To compute Pearson Correlation matrix: Select "Differential correlation" option through "Statistical Analysis – > Correlation – > Pearson correlation matrix (see Figure 12).
2. To compute Pearson Correlation matrix: Select "Differential correlation" option through "Statistical Analysis – > Correlation – > Spearman correlation matrix.
3. Enter file name to save the results.

Note: Computing all pairwise correlations for thousands of features can take significant time (minutes) depending on the number of samples. To compute pairwise correlations for only a given set of features, make a list of those features and select that list before the analysis (see [Feature Lists](#)).

6.2.1 Differential Correlation Analysis

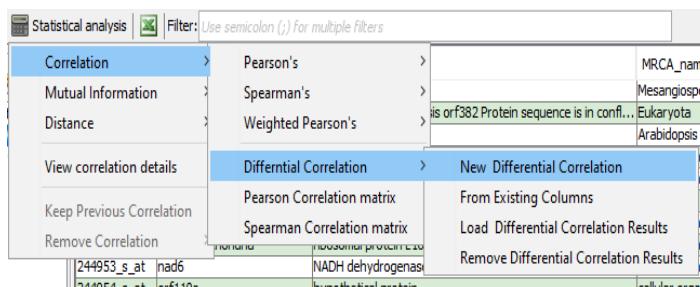


Figure 13: Selecting differential correlation.

MOG can test whether features are differentially correlated between two groups using Fisher transformation method.

Differential correlations means that the patterns of coexpression in a group of samples differs from that in a different group of samples. Features with differential correlation may reveal changes in biological interactions under different sets of conditions.

To compute differential correlation for a given feature under between two groups of samples:

1. Select "Differential correlation" option through "Statistical Analysis – > Correlation – > Differential correlation – > New Differential Correlation (see Figure 13).
2. A dialog identical to "[Differential Expression Analysis](#)" window will appear.
3. Add samples to the groups (see "[Differential Expression Analysis](#)" window).
4. CLICK OK. The results will be displayed in a new window.

To compute differential correlation from existing Pearson's correlation columns:

1. Select "Differential correlation" option through "Statistical Analysis – > Correlation – > Differential correlation – > From Existing Columns (see Figure 13).
2. Select the first correlation column and enter $N1$ which is the number of samples used in the calculation.
3. Select the second correlation column and enter $N2$ which is the number of samples used in the calculation.
4. The results will be displayed in a new window.

Note: Because the goal of finding differential correlation is to identify if the correlations of a feature with other features changes over different groups, the user must carefully choose the correlation columns (which contains the correlation of a feature with other features under different samples) and enter the correct group sizes. The interpretation of the results is up to the user.

6.3 Correlation Between Samples

Correlation between samples can reveal how similar various different samples are to one another other. This can be helpful in revealing similarity between samples from different biological groups and evaluating the importance of technical effects. It also can determine whether a set of "replicates" are similar or very diverse

To find the correlation between selected samples:

1.

Experiment	alias_or_genotype	chip	chip_name
AtGenExpre...	Col-0	ATGE_31_A2	ATGE_31_A2
AtGenExpre...	Col-0	ATGE_32_A2	ATGE_32_A2
AtGenExpre...	Col-0	ATGE_33_A	ATGE_33_A
AtGenExpre...	Col-0	ATGE_34_A	ATGE_34_A
AtGenExpre...	Col-0	ATGE_35_A	ATGE_35_A
AtGenExpre...	Col-0	ATGE_36_A	ATGE_36_A
AtGenExpre...	Col-0	ATGE_37_A	ATGE_37_A
AtGenExpre...	Col-0	ATGE_39_A	ATGE_39_A
AtGenExpre...	Col-0	ATGE_40_A	ATGE_40_A
AtGenExpre...	Col-0	ATGE_41_A	ATGE_41_A
AtGenExpre...	Col-0	ATGE_42_B	ATGE_42_B
AtGenExpre...	Col-0	ATGE_43_A	ATGE_43_A

In the "[Sample Metadata Table pane](#)", select the required rows (each row represents a sample).

2.

Experiment	Cosine similarity	Pearson Correlation
AtGenExpre...	Col-0	
AtGenExpre...	Col-0	
AtGenExpre...	Col-0	

In the "[Sample Metadata Table pane](#)" menubar, select the "Analyze" option and choose appropriate method.

3. File View Plot Edit

Var X	Var Y	Pearson's correlation
ATGE_31_A2	ATGE_39_A	0.789
ATGE_32_A2	ATGE_39_A	0.817
ATGE_32_A2	ATGE_36_A	0.508
ATGE_32_A2	ATGE_37_A	0.879
ATGE_32_A2	ATGE_34_A	0.828
ATGE_32_A2	ATGE_35_A	0.851
ATGE_32_A2	ATGE_33_A	0.945
ATGE_37_A	ATGE_39_A	0.811
ATGE_34_A	ATGE_40_A	0.887
ATGE_31_A2	ATGE_32_A2	0.96
ATGE_31_A2	ATGE_40_A	0.817
ATGE_33_A	ATGE_40_A	0.836
ATGE_36_A	ATGE_40_A	0.394
ATGE_35_A	ATGE_40_A	n/a

The results are displayed in a new window. The results could be saved to file by going to File – > Export.

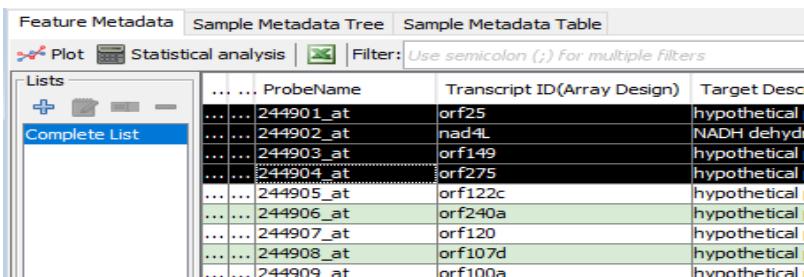
7 Visualizations

All visualizations generated in MOG are interactive and the charts and graphs can be manipulated and customized in many ways.

7.1 Line Charts

Line charts allow multiple features to be plotted together across all the samples on an X-Y plane. They visualize trends in data across the samples, and make it easy to compare one feature to another. (For example, what are the expression patterns of multiple members of a gene family. Which ones are up-regulated in particular samples. What are their relative levels of expression across the samples.)

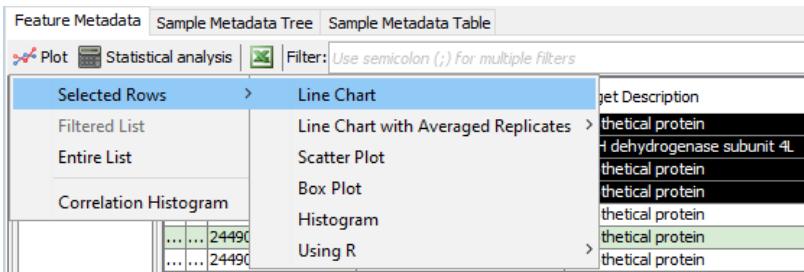
To create a line chart:

1. 

The Feature Metadata pane displays a table of genes. The columns are ProbeName, Transcript ID(Array Design), and Target Description. The genes listed are: 244901_at, orf25, hypothetical protein; 244902_at, nad4L, NADH dehydrogenase subunit 4L; 244903_at, orf149, hypothetical protein; 244904_at, orf275, hypothetical protein; 244905_at, orf122c, hypothetical protein; 244906_at, orf240a, hypothetical protein; 244907_at, orf120, hypothetical protein; 244908_at, orf107d, hypothetical protein; 244909_at, orf100a, hypothetical protein.

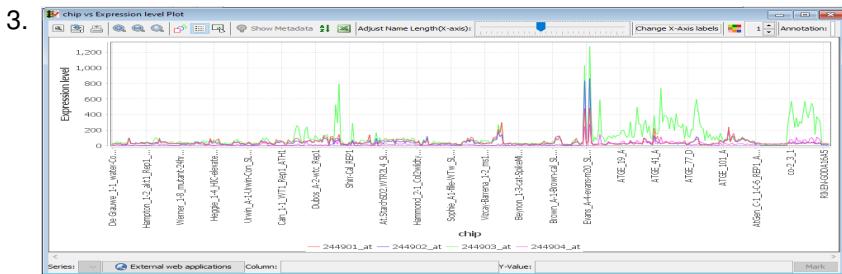
Select rows to plot from the "Feature Metadata" pane. E.g., to get a set of desired genes do a statistical analysis so genes highly correlated to your gene of interest will be at the top of the chart, make a custom list, or filter based on a characteristic of interest.

Note: Selecting too many rows will slow down the performance of MOG for big datasets. E.g., don't try to plot hundreds or thousands of genes at once.

2. 

The Feature Metadata pane shows the 'Selected Rows' menu. Under the 'Line Chart' section, 'Line Chart with Averaged Replicates' is selected. Other options include Scatter Plot, Box Plot, Histogram, and Using R.

CLICK Plot – > Selected Rows – > Line Chart from the menubar at the top of the "Feature Metadata" pane,



A new window containing the line chart appears with:

- A *display area* containing the line chart plot.
- A *toolbar* to manipulate the line chart (top of window).
- A *status bar* that displays status of selected points in the chart (bottom of window).

7.1.1 The Line Chart Display Area

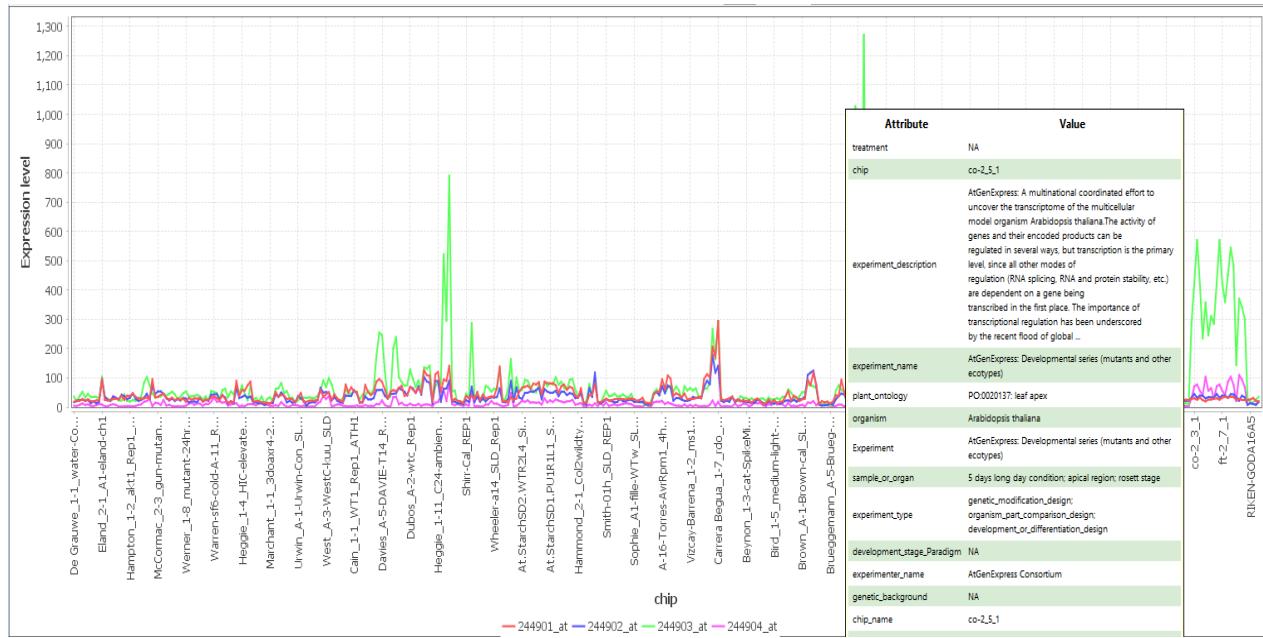


Figure 14: Line Chart Display Area.

The X-axis represents the samples and the Y-axis represents the value of each feature. Each line is a series of connected points that represents an individual feature and its changes in value across all samples. At the bottom is a legend showing the color of line that represent each feature.

SELECT AN AREA holding left-CLICK to zoom the line chart to that area.

HOVER a mouse over data points to reveal the metadata associated with that sample.

CLICK the mouse on a data point to show the metadata for that sample in the "Sample Metadata Tree" (Figure 14).

DOUBLE-CLICK the mouse on a line to open a window to change the color of that gene/feature.

7.1.2 Line Chart Toolbar

The line chart toolbar (Figure 15) lets a user customize, interact with, and export the chart.



Figure 15: Line Chart Toolbar.

Toolbar components are:

1. Set properties: Opens a new dialog box to change chart background color, fonts, and axis labels.
2. Export: Exports the chart as a .png file. **Note** The user can enter dimensions of image to determine resolution (e.g. for publication).
3. Print: Opens a dialog box to print chart.
4. +Zoom: Zooms in both axes.
5. -Zoom: Zooms out both axes.
6. Reset Zoom: Fits the entire chart in the display area.

7. Data points: Toggles visible data points.
8. Legend: Toggle visible legend.
9. Clears any markers in the chart. (See [7.1.4](#))
10. Metadata: Displays the metadata from the "Sample Metadata Tree" pane, for selected sample.
11. Sort/group. Pull-down menu contains many options for sorting and grouping the chart. (See [7.1.4](#))
12. Interactively select and save chart data to .png file.
13. Adjust slider to change lengths of x-axis labels.
14. Change X-Axis labels: Lists metadata headers and replaces X-Axis label with user-selected header.
15. Color: Choose a new color scheme for all series.
16. Line thickness: Adjust this spinner to alter line thickness.
17. Annotation: Write a custom annotation on the chart.

7.1.3 Line Chart Status Bar

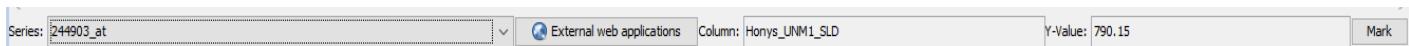


Figure 16: Line Chart Status Bar.

The line chart status bar (bottom of chart) displays information about the selected data point in the chart:

1. Series: Shows which feature is currently selected and its key metadata.
2. External web apps: CLICK to search external websites using the name of a selected series.
3. Column: Shows the sample name for the selected point.
4. Y-value: Shows the y value of the selected point.
5. Mark: CLICK to annotated a selected data point with X and Y axis information.

7.1.4 Sort and Group Line Chart

Lets the user explore the data by quickly reordering it and grouping it in different ways (see Figure [17](#)):

- Default: Sorts the x-axis by the order of the samples in the .mog file.
- Sample Name: Sorts the x-axis alphabetically by the unique identifiers of the samples.
- X-axis labels: Sorts the x-axis by the current x-axis labels. X-axis labels can be changed through the [line chart toolbar](#) of the line chart toolbar.
- Expression level: Sorts the x-axis by the decreasing order of the series selected from the pull-down menu.
- Group by Metadata: Groups the chart by categories in the metadata. CLICK the Sort/group button in the [Line Chart Toolbar](#) to open the menu and choose "Group by Metadata". Different categories may be combined to get multiple groupings by selecting the "More..." option (see Figure [17](#)).

Note: each time a grouping operation is performed, the groups in the chart are displayed using markers. The markers remain in the chart unless cleared by the user using the [line chart toolbar](#) or another grouping is chosen.

- Group by Query: Groups the samples by selected metadata. CLICK to open the search panel and create a metadata query for grouping.

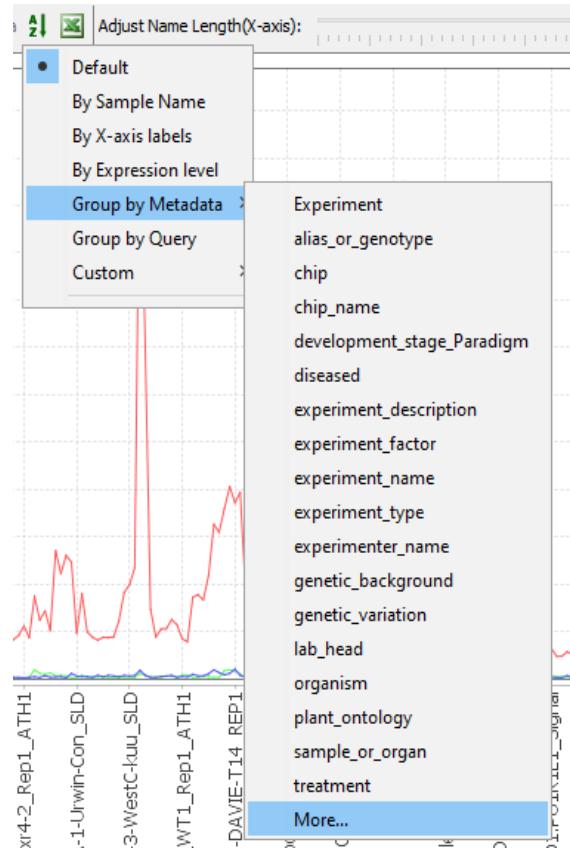


Figure 17: Group line chart by metadata.

- Custom: Groups samples according to user specifications. Custom sorts can be saved for future analyses of the dataset.

7.2 Line Chart with Averaged Replicates

A line chart with averaged replicates is a line chart in which the data of replicates is averaged before plotting. Section 5.1 describes how to specify a **factor** for replicates in the sample metadata.

To create a line chart with averaged replicates:

1.

ProbeName	Transcript ID(Array Design)	Target Description
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydrogenase (ubiquinone)
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein

In the "Feature Metadata" pane, select rows to plot. **Note:** Selecting too many rows may slow down the performance of MOG.

2.

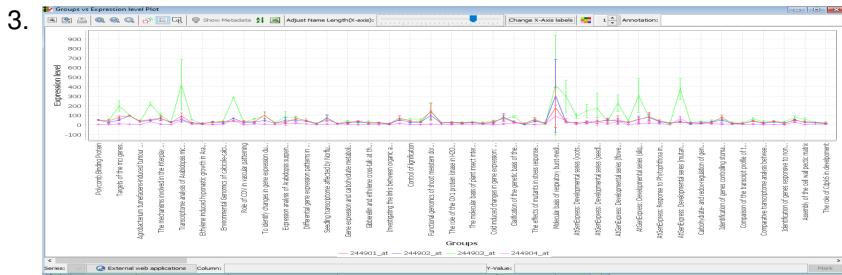
Selected Rows > Line Chart
 Line Chart with Averaged Replicates > Default grouping

Line Chart with Averaged Replicates > Choose grouping

Scatter Plot
 Box Plot
 Histogram
 Using R

Correlation Histogram

In the menubar located at the top in the "Feature Metadata" pane, CLICK Plot – >Selected Rows – >Line Chart with Averaged Replicates. Choose "default grouping" to average the samples under the default metadata header. To choose a different grouping, CLICK "Choose grouping".



A line chart with means +- standard deviations of each value for each feature is displayed.

7.3 Scatter Plot

A scatter plot visualizes data as the values for one feature in a each of a set of samples v.s. the value for another feature across the same samples. It can reveal relationships between the features. MOG can plot multiple features (pairwise) in a scatter plot.

To create a scatter plot with selected features, perform the following steps:

1.

ProbeName	Transcript ID(Array Design)	Target Description
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydrogenase (ubiquinone) 4L subunit
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein

In the "Feature Metadata" pane, select rows to plot. **Note:** Selecting hundreds of rows will slow down the performance of MOG.

2.

Selected Rows > Line Chart
 Line Chart with Averaged Replicates > Scatter Plot

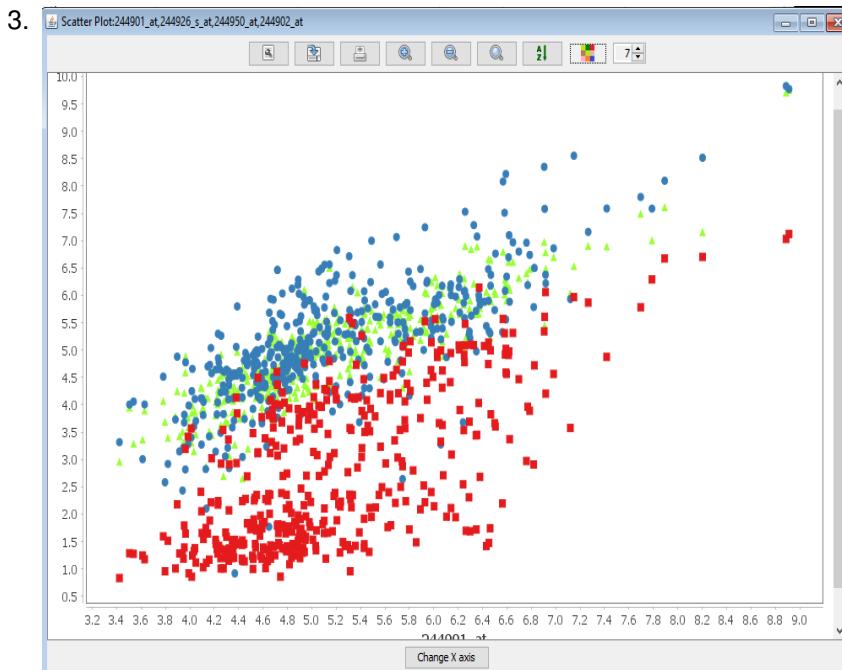
Line Chart with Averaged Replicates > Default grouping

Scatter Plot > Choose grouping

Box Plot
 Histogram
 Using R

Correlation Histogram

In the menubar at the top of the "Feature Metadata" pane, CLICK Plot – >Selected Rows – >Scatter Plot.



A new frame containing the scatter plot is displayed. Each color in the scatter plot compares the values for a different pair of features. In this example, the features are genes. The red dots compare the expression values of gene ATMG00640 (X-axis) and gene ATMG00520 (Y-axis) in each sample. The blue dots compare the values of gene ATMG00640 (X-axis) and gene ATMG00650 (Y-axis). The toolbar of the scatter plot is similar to that of the [line chart](#) thus a user can interact with the data and metadata in multiple ways.

7.4 Box Plot

A box plot helps understand the distribution of the data. MOG can plot multiple box plots side-by-side, for direct comparisons.

To create a box plots:

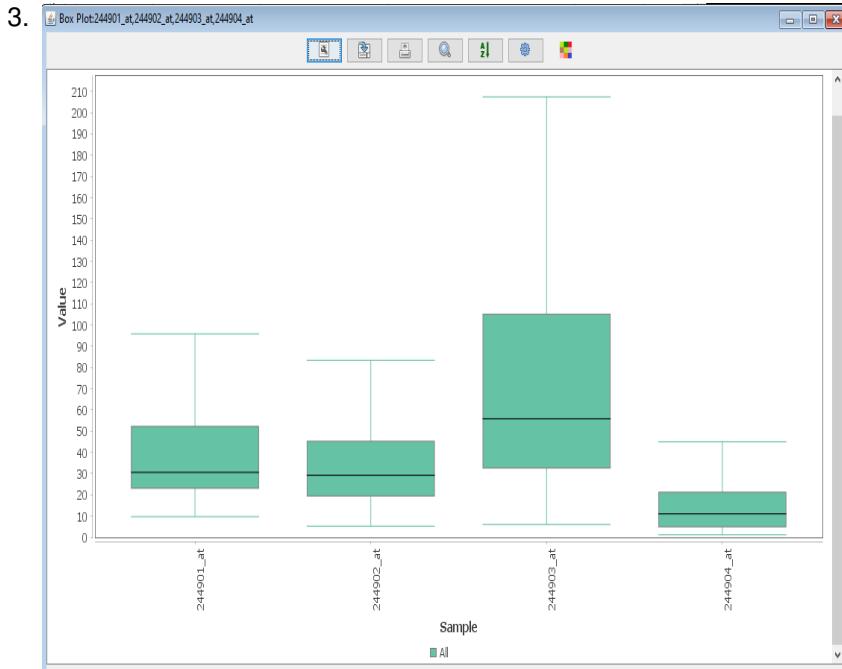
1.

ProbeName	Transcript ID(Array Design)	Target Description
244901_at	orf25	hypothetical protein
244902_at	nad4L	NADH dehydrogenase (ubiquinone) 4L subunit
244903_at	orf149	hypothetical protein
244904_at	orf275	hypothetical protein
244905_at	orf122c	hypothetical protein
244906_at	orf240a	hypothetical protein
244907_at	orf120	hypothetical protein
244908_at	orf107d	hypothetical protein
244909_at	orf100a	hypothetical protein

In the ["Feature Metadata"](#) pane, select rows to plot the in the chart. **Note** Selecting too many rows may slow down the performance of MOG.

2.

In the menubar located at the top in the ["Feature Metadata"](#) pane, CLICK Plot – >Selected Rows – >Box Plot.



A box plot is displayed in a new frame. The toolbar of box plots is similar to that of [line chart](#).

7.5 Histogram

The histograms visualize the distribution of values in a dataset for selected feature(s). MOG can plot histograms for multiple features side-by-side, which facilitates direct comparison.

To create a histogram:

1.

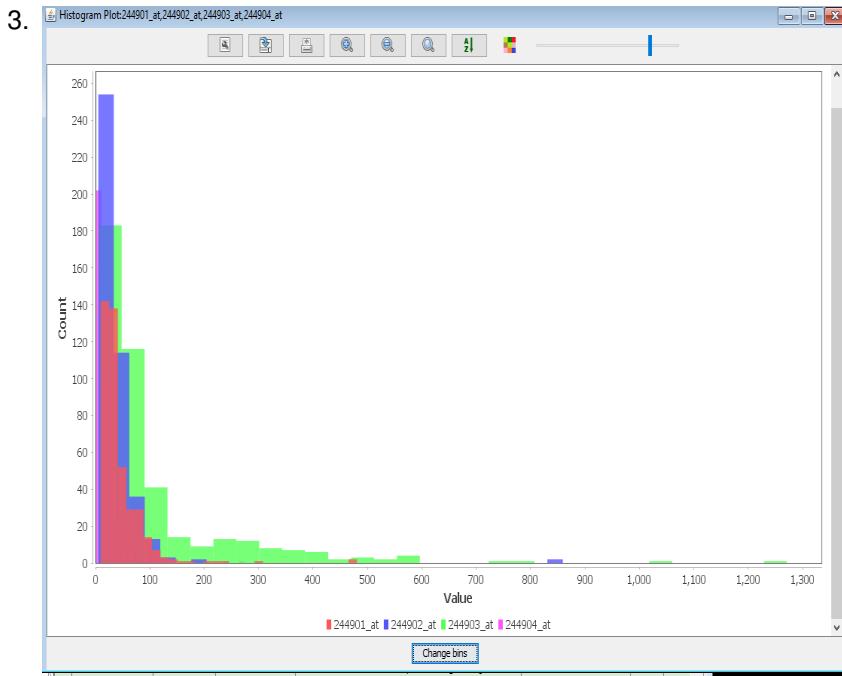
	ProbeName	Transcript ID(Array Design)	Target Description
...	244901_at	orf25	hypothetical protein
...	244902_at	nad4L	NADH dehydrogenase (ubiquinone) 4-like
...	244903_at	orf149	hypothetical protein
...	244904_at	orf275	hypothetical protein
...	244905_at	orf122c	hypothetical protein
...	244906_at	orf240a	hypothetical protein
...	244907_at	orf120	hypothetical protein
...	244908_at	orf107d	hypothetical protein
...	244909_at	orf100a	hypothetical protein

In the "Feature Metadata" pane, select the rows to plot. **Note** Selecting too many rows may slow down the performance of MOG.

2.

- Plot Statistical analysis
- Selected Rows > Line Chart
- Line Chart with Averaged Replicates >
- Scatter Plot
- Box Plot
- Histogram
- Using R

In the menubar at the top in the "Feature Metadata" pane, CLICK Plot – >Selected Rows – >Histogram.



A histogram is displayed in a new window. In the histogram shown, four genes were selected for plotting. Each color represents the distribution of a particular gene. Bin size was 100 samples. The green gene has a broader distribution and is more highly expressed. Although four genes are plotted, the values for one gene (pink) are obscured. CLICKing on that particular gene will bring its values to the forefront.

The toolbar of the histogram plot is similar to that of the [line chart](#). The number of bins the data points are grouped in can be changed by CLICKing the button at the bottom of the window.

7.6 Volcano Plot

The volcano plot is a scatterplot that visualizes the statistical significance (P-value) that two features are differentially expressed (on the Y-axis) versus the fold-change between the two features (on the X-axis) (see Figure 18). The user can select which sets of samples to compare (e.g., all *leaf* samples versus all *root* samples). Section [Differential Expression Analysis](#) describes how to calculate differential expression (DEA) in MOG .

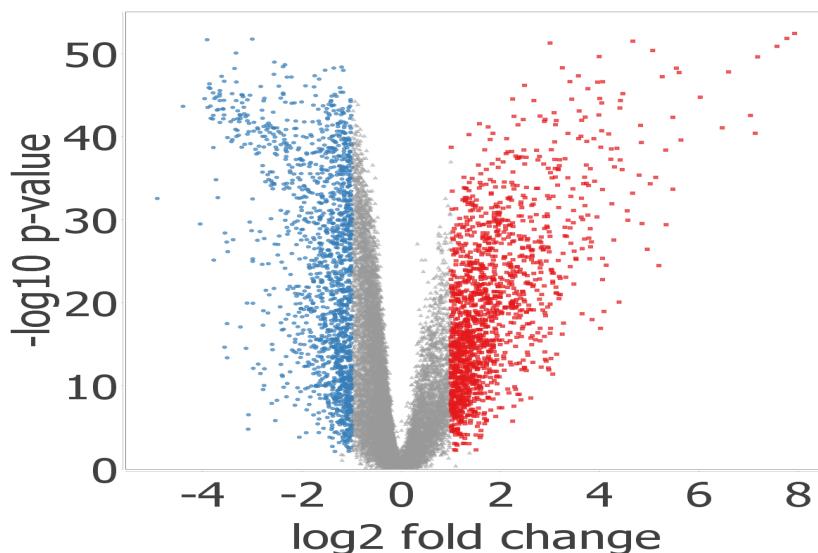
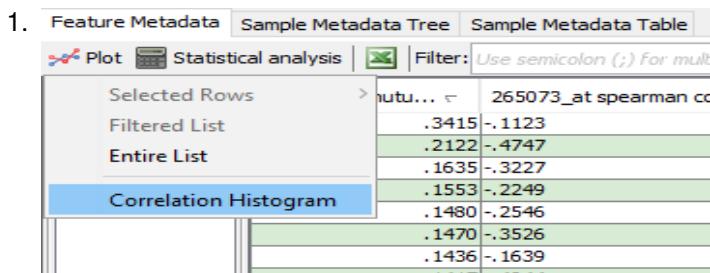


Figure 18: An example volcano plot generated using MOG.

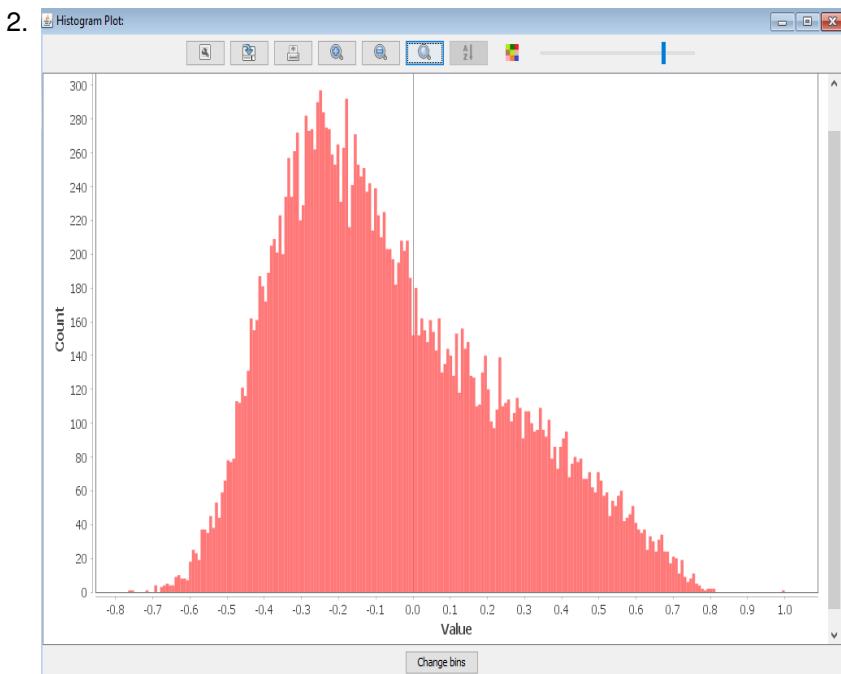
7.7 Correlation Histogram

A correlation histogram visualizes the distribution of correlation values across a given set of samples(see [Correlations Between Feature](#)). MOG shows the number of correlations of a given correlation value for pairwise comparisons of user-selected features.

To plot a correlation histogram:



In the menubar located at the top in the "[Feature Metadata](#)" pane, CLICK Plot – >Correlation Histogram and choose the column containing correlations to be plotted.



A correlation histogram plot is displayed in a new window. This correlation histogram shows the distribution of pairwise Spearman correlation values for a user-selected gene compared to all other genes. Bin size is 10.

8 Differential Expression Analysis

The goal of differential expression analysis (DEA) is to identify features having different values across two groups of samples. MOG has five different tests to determine the statistical significance of fold-change values, each based on different assumptions. These tests help the user to identify which features are differentially-expressed. Selecting the most appropriate method and interpreting the results is up to the user.

Perform a DEA:

1. MetaOmGraph - Affy.ath1_US_3-19-2019.mog (424 samples); 22746 features

The screenshot shows the software's main window with a menu bar (File, Edit, Tools, Window, Help). A 'Project' icon is visible. On the left, there are three tabs: 'Feature' (selected), 'Plot', and 'External web applications'. Under 'Feature', there are 'Lists' and 'Complete List' buttons. The 'Tools' menu is open, showing 'External web applications', 'Search by expression level', 'Differential expression analysis' (which is also selected), and 'Perform DEA'. The 'Perform DEA' option is highlighted with a blue border. Below this, a table displays differential expression results. The columns include ProbeName, TranscriptID(Array Design), Target description, and several other columns. The first row shows a significant hit for At5g23960.

ProbeName	TranscriptID(Array Design)	Target description
.1044	249760_at	At5g23960 (+)-delta-cadinene synthase (d-cadiner)
-.1269	265431_at	At2g20680 (1-4)-beta-mannan endohydrolase
-.1366	259442_at	At1g02310 (1-4)-beta-mannan endohydrolase precursor
.0363	251069_at	At5g01930 (1-4)-beta-mannan endohydrolase-like protein
.1290	250470_at	At5g10160 (3R)-hydroxymyristoyl-[acyl carrier protein]

From the parent window menubar, CLICK "Tools – > Differential Expression Analysis – > Perform DEA". This will open the "Differential Expression Analysis" window.

2. In the "Differential Expression Analysis" window select two sample groups and a feature list. CLICK OK to start the analysis.
3. The results will be displayed in a "Differential Expression Results" window.

8.1 Differential Expression Analysis window

The "Differential Expression Analysis" window provides the interface to select the groups and features for differential expression analysis, and to choose and run the DEA.

The "Differential Expression Analysis" window has the components (left to right, top to bottom):

1. Save results with MOG Checkbox: Saves the DEA results to the MOG project. **Note** the MOG project also must be saved in-order to retrieve the saved DEA results.
2. Select feature list **Complete List** Select feature list: Choose the features on which DEA analysis will be performed.
3. Select method **M-W U test** Method: Choose the statistical test to be applied.
4. **n=123** n= : Displays total number of samples currently included in first group (left table).
5. Group name: **Group Seed** Group name: User inputs a name for the first group in this text field (left table).
6. **>>** Button: Moves user-selected samples (rows) from the first group to the second group. (left table)

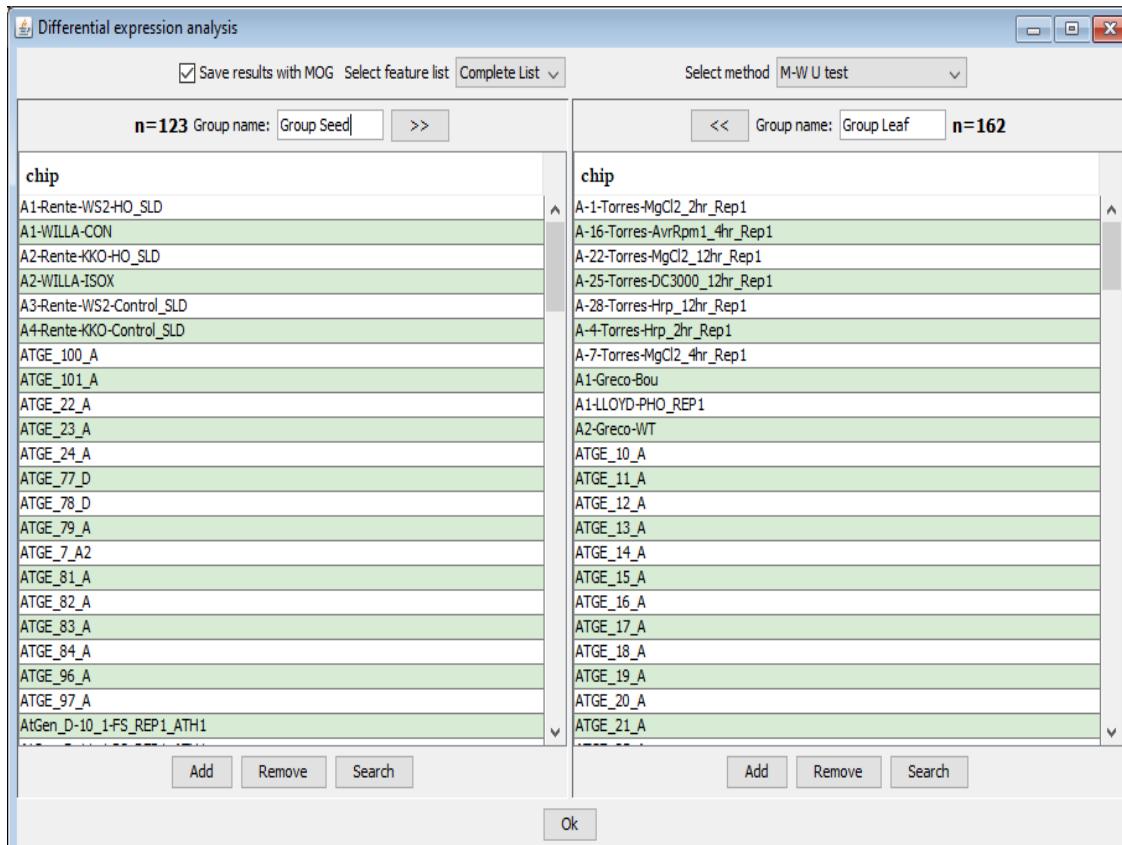


Figure 19: Differential Expression Analysis Window.

7. Button: Moves user-selected samples (rows) from the second group to the first group.(right table).
8. Group name: User inputs a name for the second group in this text field. (right table).
9. n=: Displays the total number of samples currently included in second group (right table).

10.

The table on the left of the window (enlarged) shows the samples ($n=123$) in the user-designated "Seed" group from a set of microarray studies. Buttons at the bottom add or remove samples. "Add" button opens up the "Metadata Search" panel where a user can search for samples by metadata and add these samples to the group. "Remove" button removes the selected rows from the table. "Search" opens up "Metadata Search" panel where a user can search the samples already included in the list by metadata. The search results are moved to top and selected.

11. Press this button to start the differential expression analysis.

8.2 Differential Expression Results window

The "Differential Expression Results" window displays the results as an interactive, sortable table. The columns are:

- **Name**: displays IDs of each feature.
- **Mean(log(Group1))** contains the geometric means of the corresponding features in first group.
- **Mean(log(Group2))** contains the geometric mean of the corresponding features in second group.
- **logFC** contains the fold-change of Group 1 vs Group 2 for each feature.
- **Test p-val** contains the p-values from the test used.
- **Adj. p-val** contains the corrected p-values.

The menubar in the "Differential Expression Results" window contains options to save, filter and visualize the results:

File menu: saves results as a text file.

Edit :

- **Export Selected to List**: Creates a new list (see Section 5.3) in MOG's list panel with features (rows) selected from the Result Table.
- **P-Value filter**: Filters results based on a p-value
- **P-Value correction**: Selects the p-value correction method (e.g., Benjamini-Hochberg or Bonferroni)

Fold change results

File Edit Plot

Name	Mean(log(Group Seed))	Mean(log(Group Leaf))	logFC	M-W U test pval	Adj pval
266353_at	6.6774	1.3671	5.3103	,0000	,0000
260668_at	7.4522	2.3390	5.1132	,0000	,0000
246855_at	7.5787	2.5214	5.0573	,0000	,0000
245555_at	7.0625	2.0092	5.0534	,0000	,0000
253667_at	6.9316	2.0037	4.9279	,0000	,0000
257947_at	6.5941	1.7618	4.8323	,0000	,0000
253024_at	6.5634	1.7607	4.8027	,0000	,0000
247333_at	6.5314	1.8283	4.7030	,0000	,0000
252882_at	6.7400	2.0498	4.6902	,0000	,0000
250090_at	8.0495	3.4345	4.6150	,0000	,0000
260130_s_at	7.2486	2.6903	4.5583	,0000	,0000
250500_at	7.0668	2.5624	4.5044	,0000	,0000
250059_at	6.2967	1.8014	4.4953	,0000	,0000
248252_at	6.9039	2.5042	4.3997	,0000	,0000
251226_at	5.7522	1.4767	4.2755	,0000	,0000
246825_at	6.0311	1.8019	4.2292	,0000	,0000
262260_at	8.3269	4.1428	4.1841	,0000	,0000
258080_at	6.1072	1.9371	4.1701	,0000	,0000
254726_at	6.9097	2.7545	4.1553	,0000	,0000
259009_at	10.2545	6.1185	4.1360	,0000	,0000
266330_at	7.6567	3.5732	4.0835	,0000	,0000
259478_at	6.0457	1.9940	4.0517	,0000	,0000
254234_at	8.0996	4.0803	4.0193	,0000	,0000
254718_at	5.8591	1.9011	3.9580	,0000	,0000
247297_at	7.9343	4.0122	3.9221	,0000	,0000
267121_at	5.6284	1.7147	3.9137	,0000	,0000

Figure 20: Differential Expression Analysis Window.

Plot : Options to visualize the results. (The options for visualizing specific selected features are discussed in Section 7.
To visualize the full results:

- **FC histogram**: Plots a histogram of the log fold change values.
- **Volcano Plot** Plots a volcano plot.

9 Create Your Own Projects

How to create, save, and open new MOG projects. This section describes the required format for MOG input data, and details the MOG GUI to create projects and interact with the project data.

9.1 Input format

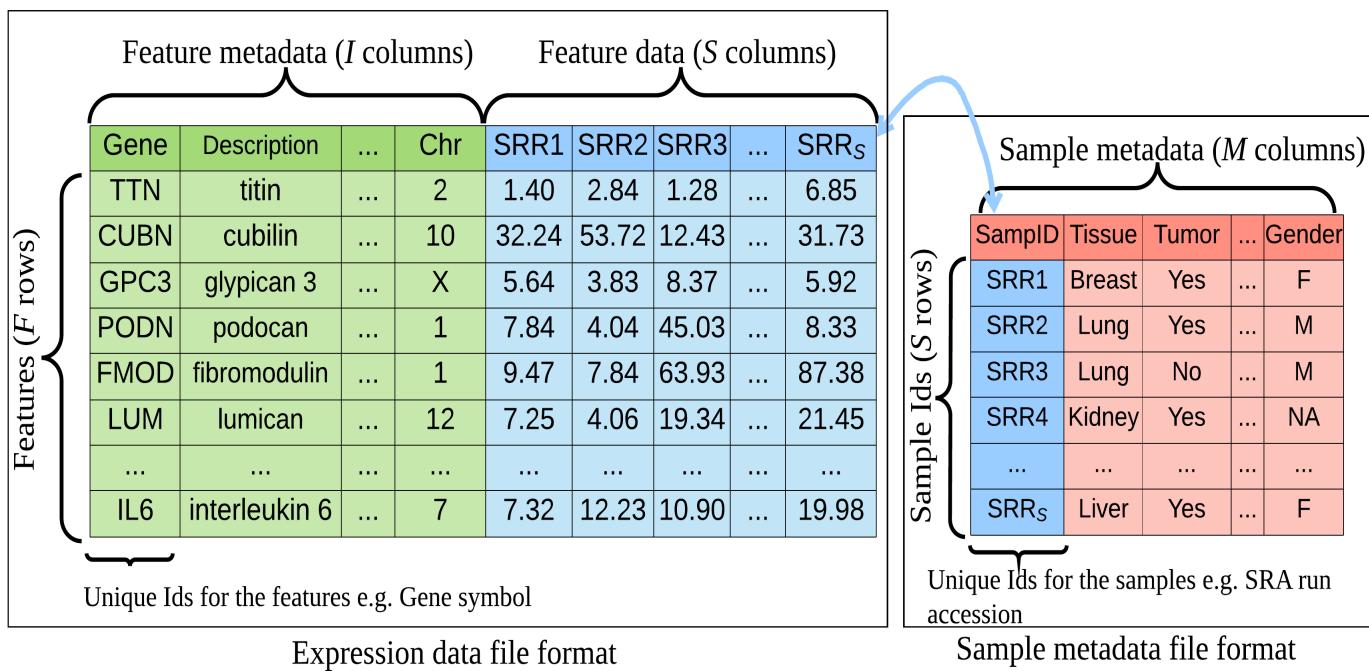


Figure 21: MOG input data format.

After starting MOG, the user can read their own data and metadata using MOG. The data and metadata should be present in two separate files which are in text-delimited format. The supported delimiter characters are tab, comma, semicolon and space.

Both data and metadata files should be in a specific delimited format (see Figure 21). The data file is a matrix containing the measured expression values of the features (rows) over the samples (columns). For example: expression of genes (each gene is a feature) over a number of microarray studies; transcript counts (each transcript is a feature) over multiple RNA-Seq runs. The file can have first I columns as feature metadata columns (Figure 21) which can have information about the features. For genes these might be: name, phylostratum, description, protein encoded, tertiary structure of encoded protein, gene type etc. The first I columns are **feature metadata columns**; the first column is a unique ID for each row (here, gene symbol). The latter (S) columns contain expression values of F features over S samples.

A metadata file should be a delimited text file with rows as samples and columns as metadata attributes. It is a matrix of S rows by M columns (Figure 21). Each row in the sample metadata file corresponds to a sample in the expression data file. The M columns are the metadata attributes of each experimental analysis (e.g., run for RNA-Seq data, chip for microarray data). A column, in the metadata file, contains the unique sample IDs that link to the columns in the data matrix. This column is referred to as the **sample id column** in MOG. If samples are missing from metadata file, MOG handles the missing information by producing an empty metadata row.

9.2 Start a new project

9.2.1 The "Create New Project" dialog

To start a new MOG project from the welcome dialog, follow the following steps:

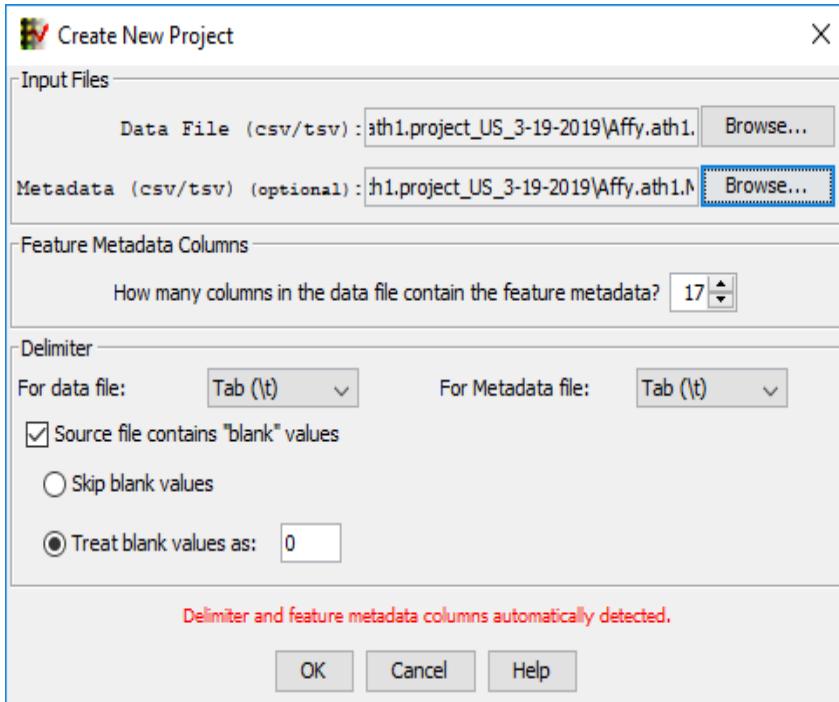


Figure 22: Create New Project dialog.

1. CLICK on the "From a delimited text file" option located in the upper-left quadrant of the "[Welcome Dialog](#)". CLICK- ing this option will create another dialog titled "[Create New Project](#)" (see Figure 22).
2. In the "[Create New Project](#)" dialog CLICK the "Browse" button next to "Data File" to locate the data file.
3. (Optional) In the "[Create New Project](#)" dialog CLICK the "Browse" button next to "Metadata" to locate the metadata file.
4. Under the "Feature Metadata Columns" section, enter the number of **feature metadata columns** present in the data file (see Section 9.1).
5. In the delimiter section select the correct delimiter for the data and the metadata files.
6. If the data file contain missing values choose an option to either skip the rows with missing values or treat the missing values as 0 or some other number.
7. CLICK OK.

9.2.2 The "[Import Metadata](#)" window

If metadata file was provided in step 3, a new window, "[Import Metadata](#)", will be displayed (Figure 23). This window displays a table which gives a preview of the metadata file selected. This window also shows basic summary of the metadata file by displaying:

Total Rows: This is the total number of rows in the metadata file (excluding the first row, which is treated as the header).

Total Columns: The total number of columns in the metadata file.

Extra Samples: The number of sample IDs in the "Sample ID column" of the *metadata* file that don't match the sample IDs in the *data* file. These rows are ignored by MOG.

Missing Samples: The number of sample IDs in the data that are missing from the "Sample ID column" in the metadata file. MOG generates empty metadata for such samples.

If the metadata file looks incorrect from the preview then a new metadata file could be loaded by performing the following steps:

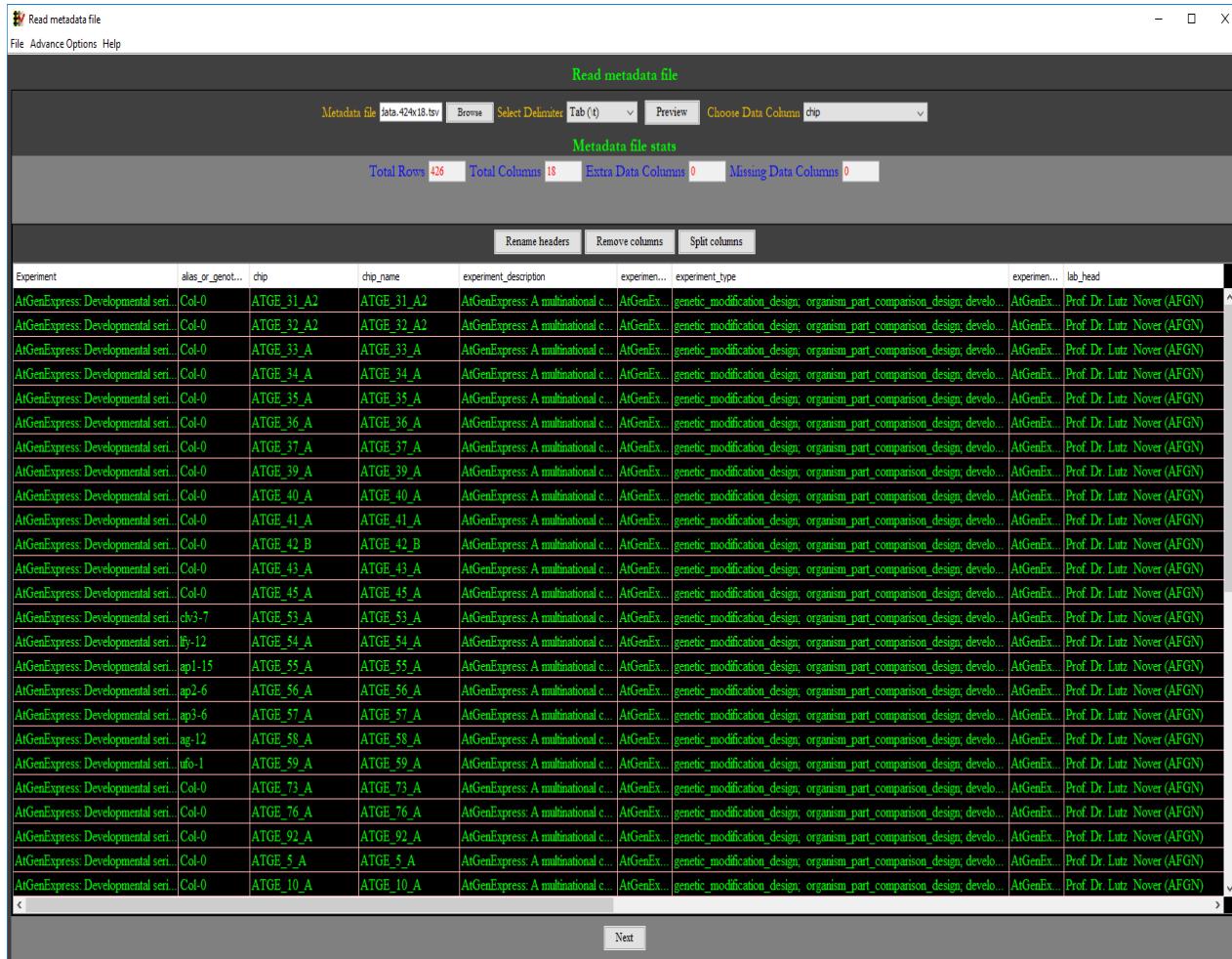


Figure 23: MOG read metadata window.

- In the "Import Metadata" window, CLICK the browse button next to the "Metadata file" label to select the metadata file.
- In the "Import Metadata" window, select the delimiter for the selected metadata file.
- In the "Import Metadata" window, CLICK preview to preview the metadata file.

If the data in the preview looks OK then proceed with the metadata import step. Perform the following steps in order to import the metadata into MOG:

- In the "Import Metadata" window, select the "Sample Id Column" located in the topmost panel. The "Sample Id Column" is the column in the metadata file which contains the sample id. Although, MOG will automatically detect this column based on the sample ids from the data file the user should check if the correct column is selected.
- CLICK "Next" located in the bottom most panel.

9.2.3 The "Metadata Table to Tree" window

After CLICKing "Next" in step 2 "Import Metadata" window will close and a new window "Metadata Table to Tree" will appear. This window provides an interface to interactively map the tabular metadata into a hierarchical tree structure.

To enable MOG to read the sample metadata, the user maps the sample metadata columns to a hierarchical structure which organizes the tabular data into a tree-like structure. A hierarchical view of the sample metadata can efficiently display the metadata at different levels of the hierarchy which makes the metadata more understandable for analysis. This hierarchical structure should be based on the organization of the metadata elements. Public repositories such as SRA, GEO and TCGA follow a hierarchical metadata schema with nested metadata elements are nested.

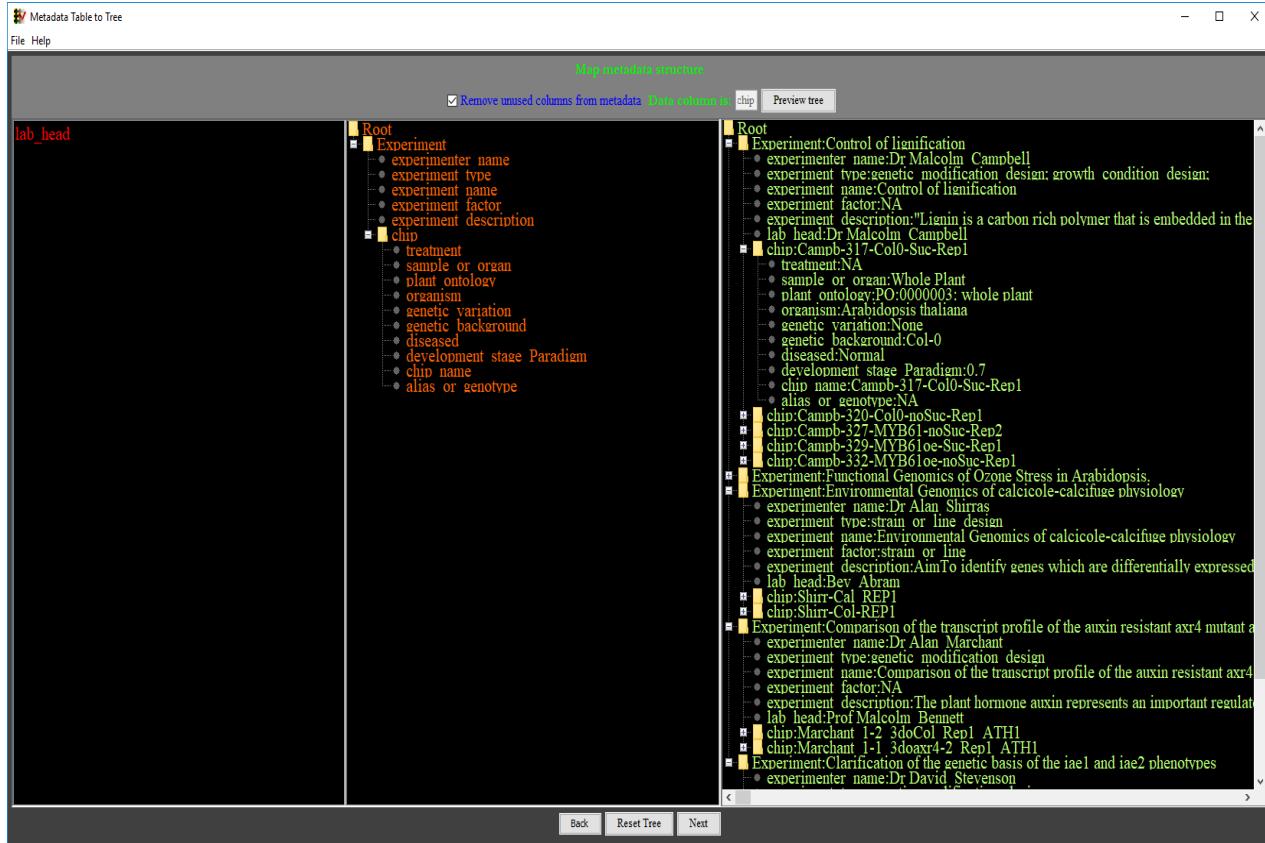


Figure 24: MOG metadata table to tree window.

For example, consider a workflow which is structured as follows:

- An experiment consists of multiple studies.
- Each study independently probes transcriptomic profiles of different samples.
- Each sample is independent biological material obtained from different sources.
- Each run is actual sequencing experiment performed on a given biological sample.

The above hypothetical structure would look like Figure 25.

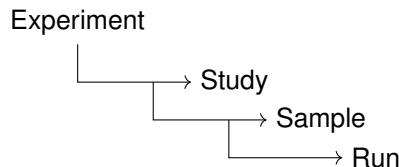


Figure 25: An example hierarchy of metadata columns.

The "Metadata Table to Tree" window is divided into three panes:left, center and right. Each panes displays different types of information about the metadata.

Left pane Shows all headers (column names) in the metadata file as a list.

Center pane Shows a user-built tree structure. Initially this tree is set to only one element, the "Root". The tree is built interactively by the user by dragging metadata headers from the left pane to the center pane.

Right pane Displays the tree structure in a hierarchical structure. This tree is generated once the "preview" button is CLICKed, based on the tree structure built by the user in the center pane.

To map the tabular metadata into a hierarchical tree structure, perform the following steps in the "["Metadata Table to Tree"](#)" window:

1. Drag the column headers from the left pane to the tree in the center pane.
2. Check the option to remove the unused column headers. The columns once removed are could not be included in the project later.
3. If required, CLICK "preview tree" to see preview of the mapped metadata.
4. CLICK "Next" if the tree structure has been created.

***If the data and metadata are properly formatted, MOG will then import the data and its metadata and a new project will be created. The "["Main MetaOmGraph" window](#)" will be visible (Figure 3)

10 Interface to R

Users can execute their own R scripts using MOG. This allows access to numerous statistical libraries which R has. The user need to write their R scripts to use data from MOG. MOG provides a convenient way for users to select relevant features and samples and forward this data to an R script. The R script makes use of this API and uses the input data to perform the computations or visualizations.

The R path can be set via MOG properties. See Section [11.2](#).

10.1 R script format

This section explains how to write an R script to be executed with MOG. To execute an R script from MOG, the R script should take following arguments:

1. The first argument is the path to the "data file" which is generated by MOG and stored in the project's directory. This file is a tab-delimited file. The file contains feature as rows and samples as columns.
2. The second argument is the path to the "metadata file" which is the metadata file used in the MOG project.
3. The third argument is the path to the output directory. When executing an R script, MOG asks an output directory name which is created under the project directory. Using this path all output from R could be saved under this directory.

Example scripts are available at: <https://github.com/urmi-21/MetaOmGraph/tree/master/rscripts>

11 Change Project Properties and MOG Properties

A user can change the behaviour and appearance of MOG and tweak parameters according to her/his requirements. Two types of properties can be altered: project properties and MOG properties. Each can be accessed from the parent window's upper menubar: "Edit – > Properties" (Figure 26).

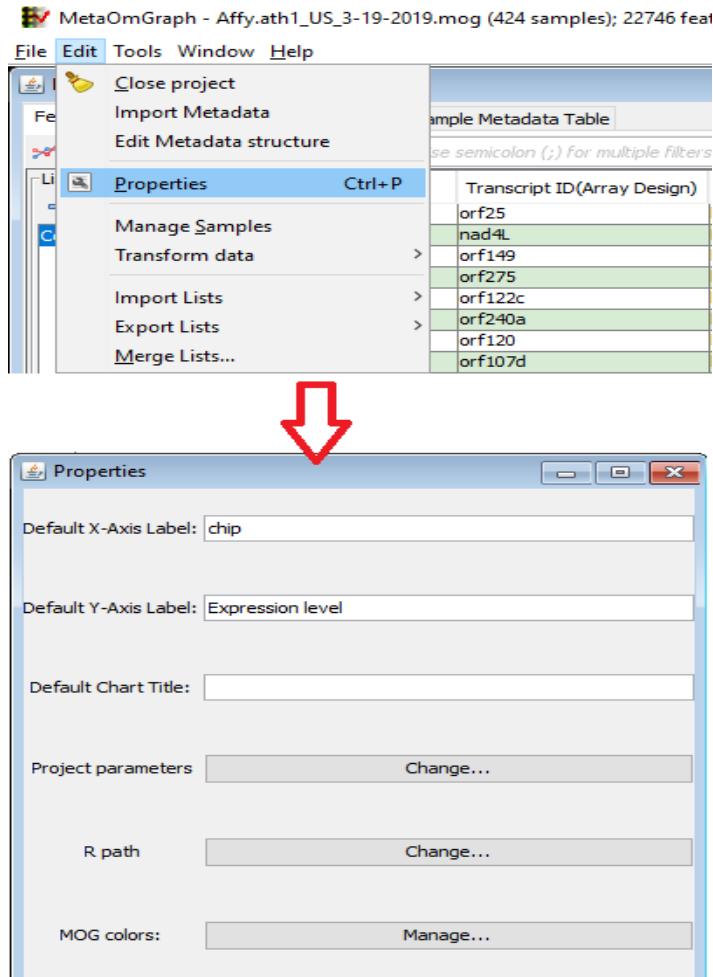


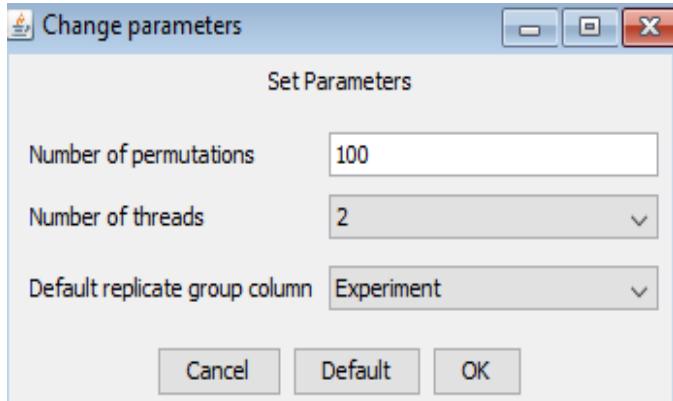
Figure 26: Properties Window.

11.1 Project properties

Project properties are specific for a given *project* and are saved with that project file.

From the Properties Window (Figure 26) a user can change text on visualizations, and calculation parameters.

1. "Default X-Axis label" is the default label of the X-Axis in visualizations (line chart, box plot, and histogram).
2. "Default Y-Axis label" is the default label of the X-Axis in visualizations (like line chart, box plot, and histogram.)
3. "Default Chart Title" is the default title displayed in line charts.
4. "Project parameters" lets a user set the numbers of permutations and threads used in the calculations, rather than use the default values. It opens a small dialog box.



Number of Permutations are the number of permutations used in computing the p-values (see 1). A higher setting will take a longer time to compute p values but give more accurate estimates. A lower setting will have a higher error in p value estimation but compute more quickly. The precision of the permutation tests also depend on the number of samples in each group. Setting the permutations to 100 allows a minimum p-value of 0.001.

Number of Threads: The number of threads used in multi-threading operations. To utilize maximum CPU, enter a higher value. The maximum value depends on the processor, we recommend a number equal to the number of cores in your processor, e.g., for a quad-core processor choose 4.

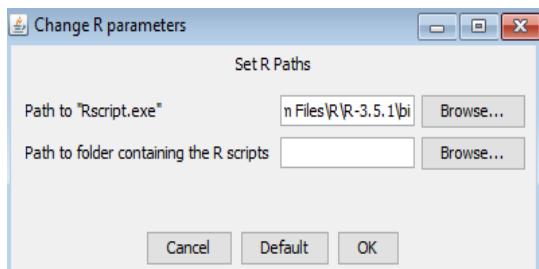
Default Replicate Group: Chooses the default replicate column using information in the sample describing the samples. The user can override the default and select the parameters to designate the replicate groups (see [Replicate Column](#)).

11.2 MOG properties

MOG properties are central to MOG and are saved in a binary file named "**metaomgraph.prefs**". This file is created when MOG is executed for the first time, and is updated when modified by the MOG user. If this file is deleted, a new file is created with default settings.

The following MOG properties can be changed from the Properties Window (Figure 26).

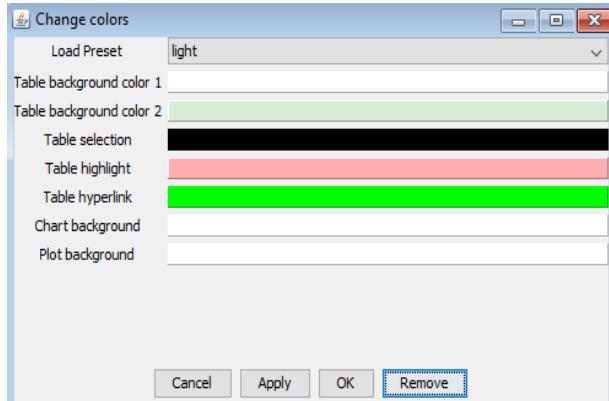
1. "R path" lets the user set a path to R which can be used with MOG (see Section 10). This opens a new dialog where user can enter the Path to R. If an R path is not defined by the user, MOG automatically detects the default path based on the operating system.



Path to Rscript.exe is the path to R executable which is used to execute R scripts.

Path to folder containing R Scripts is the path to a directory which contains scripts to be used with MOG. This parameter is optional; it is not required to execute R scripts.

2. MOG colors: Changes the appearance of MOG, MOG visualizations and MOG output. All table and background colors can be set via the "Change colors" dialog.



Load Preset loads saved color profiles. Light, sky and dark profiles are provided with MOG by default. A user can customize and create new color themes by choosing the colors. CLICKing "Apply" will apply the new theme by default to all subsequent MOG projects. If the user-customized theme isn't saved, the user will be prompted to save it. Any saved theme can be applied to a project from the "Load Preset" pull-down menu.