

MAGETABImportViewer Documentation

Module Name: MAGETABImportViewer

Description: A visualizer to import data in MAGE-TAB format into GenePattern

Author: Marc-Danie Nazaire, gp-help@broadinstitute.org

Summary

The MAGETABImportViewer imports data in MAGE-TAB format into the GenePattern .gct or .cls file formats. See http://www.mged.org/Workgroups/MAGE/mage.html for more information about MAGE, and http://www.mged.org/mage-tab/ for more information about the MAGE-TAB format.

This module takes a .zip file containing the MAGE-TAB files as input. Files containing MAGE-TAB data can be downloaded from ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) and Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/). The MAGETABImportViewer uses the MAGE-TAB parser library from http://limpopo.sourceforge.net/.

The MAGE-TAB format includes 4 file types:

- Investigation Description Format (.idf)
- Array Design Format (.adf)
- Sample and Data Relationship Format (.sdrf)
- Derived data matrix file (.txt)

MAGETABImportViewer requires only the .sdrf file and the derived data matrix file(s); if .idf or .adf files are included, they are not used by the module. The derived data matrix files are those listed in the .sdrf file under the Derived Data Matrix File column. All required files should be combined into a single .zip file for input into the module. In the case of multiple data matrix files being provided, the row names in each of the matrix files should match in order to merge samples from different files into one GenePattern gct formatted file.

Usage/Example

When you launch the MAGETABImportViewer module from GenePattern, you will see a series of screens that allow you to choose how your MAGE-TAB file will be handled. The MAGETABImportViewer uses the wizard format, so if at any time you want to change the choices you have made, you can use the *Back* button. The details of these screens are outlined below.



Step 1 of 5: Samples

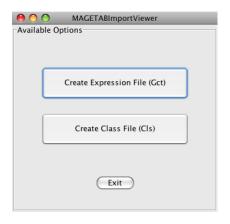
elect the samples that will appear in the gct file

Available Options:

Choose which type of file to generate:

- .GCT file to create an expression dataset
- .CLS file to create a class label file

To read more about these file formats, go <u>here</u>. Click the appropriate button to enter the sequence of steps to generate your file.



0

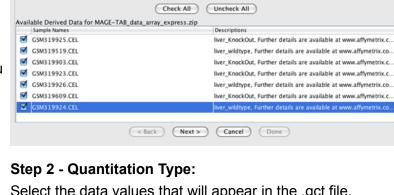
Show samples in: All files

Generating a .GCT File

Step 1 - Samples:

This screen allows you to select the samples you want to include in the output file. You can view all the samples from all the data files in the MAGE-TAB file (the default setting of the *Show samples in:* drop-down, as shown here), or you can select a single data file from which you can select samples.

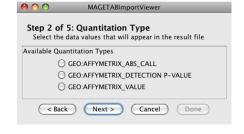
Check the desired samples and click *Next*.



Select the data values that will appear in the .gct file. Typically, a user selects a **quantitation type** that contains intensity values, signal values, or ratio values. For single-channel data, select the quantitation type that contains the signal data for the probe sets. In this example using Affymetrix data, the AFFYMETRIX_VALUE quantitation type was selected because it contains the signal values for the probe sets. In other Affymetrix-based MAGE-TAB files, this value is referred to as CHPSignal. For two-channel data, select the quantitation type that corresponds to the ratio values.

To determine the data values that correspond to the available quantitation types, view the second row of the data matrix (.txt) file, where the quantitation types appear as column headers.

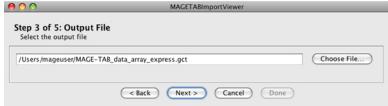
Note: If you select a non-numerical quantitation type, the resulting dataset will not be usable by most GenePattern modules.

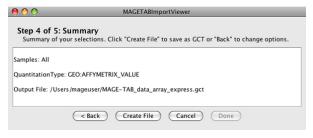




Step 3 - Output File:

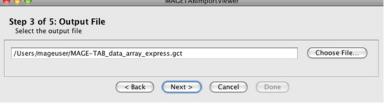
Choose the path and filename of the output file. Use the default filename or click Choose File... to specify an alternate file and click Next.





Step 5 - File Saved:

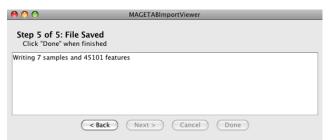
This window indicates that you have successfully imported your MAGE-TAB file. Click Done to close the wizard. The Available Options window (see above) will still be visible, allowing you to create additional expression and/or class files.



Step 4 - Summary:

Shows a summary of your selections. Click Back if you would like to change any

of the values or click Create File to continue.

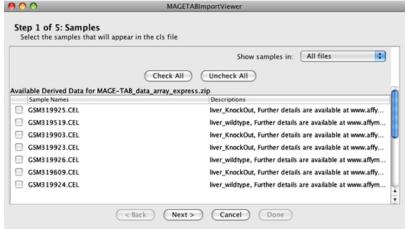


Generating a .CLS File

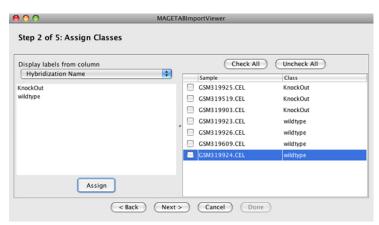
Step 1 - Samples:

This screen allows you to select the samples to use in the output file. Check the desired samples and click Next.

If this .cls file is being created for use with a .gct file, you should choose the same samples used for the .gct file.



GenePattern



Step 2 - Assign Classes:

This window lets you assign classes to a set of samples. This means that you are selecting phenotype or clinical labels for your samples. In the case illustrated here, the class "wild type" or "knockout" labels have been assigned to each of the samples.

First, select a **label** that represents the set of phenotypes/clinical attributes you want to use to label your samples. You will use the drop-down menu under

Display labels from column to do this. Each of these labels comes from the .sdrf file.

You will see a list of possible phenotypes or clinical attributes corresponding to the label type you have selected. These are the **classes** you will use to categorize the samples in your dataset.

For each class you are assigning:

- Select the class from the list that appears in the left-hand window.
- Select the checkboxes for the samples to be assigned this class in the right-hand window.
- Click Assign. This assigns the class to your selected samples and clears your selections in the left and right windows so you can assign another class to another group of samples.

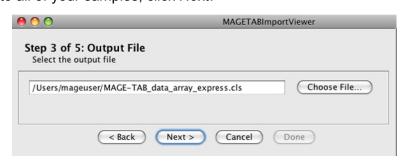
(Note that classes can also be edited manually in the Class column.)

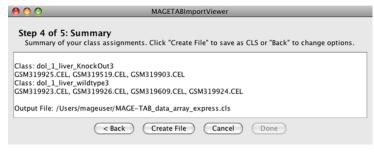
NOTE: Each sample can only be assigned to one class.

When you have assigned classes to all of your samples, click *Next*.

Step 3 - Output File:

Choose the path and filename of the output file. Use the default filename or click *Choose File...* to specify an alternate file and click *Next*.





Step 4 - Summary:

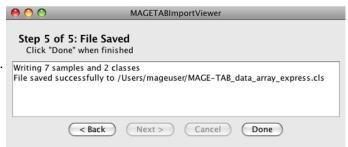
Shows a summary of your selections.

Click *Back* if you would like to change any of the values or click *Create File* to continue.



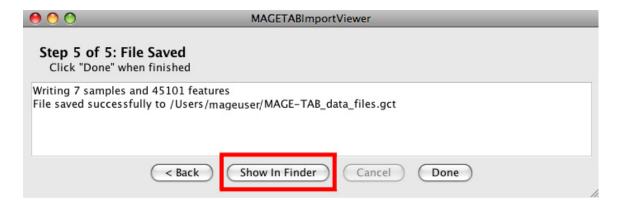
Step 5 - File Saved:

This window indicates that you have successfully imported your MAGE-TAB file. Click *Done* to close the wizard. The Available Options window (see above) will still be visible, allowing you to create additional expression and/or class files.



Other Capabilities

On a Macintosh that is running Java 6, the visualizer also allows you to see the newly-created .gct or .cls file in the Macintosh Finder application.



Parameters

Name	Description
sdrf file	The sdrf filetxt
data files	A zip file containing MAGE-TAB data matrix fileszip

Platform Dependencies

Module type: Visualizer

CPU type: any

OS: any

Language: Java

GenePattern

GenePattern Version Notes:

Date	Version	Description
11/20/12	2	Fixed bug with merging multiple data matrix files and also modified input parameters.