

ConvertToMAGETAB Documentation

Module Name: ConvertToMAGETAB

Description: A module to export data from GenePattern in MAGE-TAB

format

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

Summary

The ConvertToMAGETAB module exports microarray data from GenePattern in MAGE-TAB format. MAGE-TAB is a standard for the representation of microarray expression data. 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.

A MAGE-TAB data set includes 4 different types of files:

- Investigation Description Format (IDF)
- Array Design Format (ADF)
- Sample and Data Relationship Format (SDRF)
- raw and processed data files (TXT)

This module takes a GCT or RES file containing microarray data, and optionally takes a CLS file with them, as input. The GCT file is converted into the MAGE-TAB data matrix format. The CLS file is used to create a concise SDRF file with enough information to recreate the GCT/CLS file using the MAGETABImportViewer.

The GCT and CLS files do not contain enough information to create the IDF and ADF files. The IDF and ADF files are not necessary for using the expression data. However, if you would like to annotate your MAGE-TAB SDRF file and add an IDF file, you can use the <u>Annotare</u> tool to do so.

Usage/Example

The input file (in either GCT or RES format) is required, while the CLS file is optional. However, if you do not specify a CLS file, then no SDRF file is generated. You must also specify:

- sample reference label: This specifies the type of samples listed. It is a
 reference to one of the acceptable column names that can appear in an
 SDRF file. For example, the "Hybridization REF" label specifies that the
 samples listed are hybridizations referenced from the Hybridization
 column of the SDRF file for this MAGE-TAB data matrix file.
- feature reference label: This specifies the type of features listed, using one
 of the types acceptable for MAGE-TAB. It is a reference to one of the



acceptable column names that can appear in an ADF file. For example, the Composite REF label specifies that the features listed are composite elements and biologically relevant sequences such as genes; these features are referenced from the Composite Element column of the ADF file for this MAGE-TAB data matrix file.

- intensity type label: This is any label that describes the type of intensity values in the GCT or RES file. For example, if the values in the GCT file are expression ratios, a good label might be "Ratios" or "Expression Ratios."
- output prefix

Parameters

Name	Description
input.file	A dataset file - GCT, RES. Required.
cls.file	A class file - CLS. Optional.
sample.reference.label	Label that specifies the type of samples listed in the first row of the MAGE-TAB data matrix file. (default value: Hybridization REF)
feature.reference.label	Label that specifies the type of features listed in the second row of the MAGE-TAB data matrix file. (default value: CompositeElement REF)
intensity.type.label	A label that that describes the type of intensity values in the GCT or RES file.
output.prefix	<input.file_basename></input.file_basename>

Platform Dependencies

Module type: Preprocess & Utilities

CPU type: Any
OS: Any
Language: Java