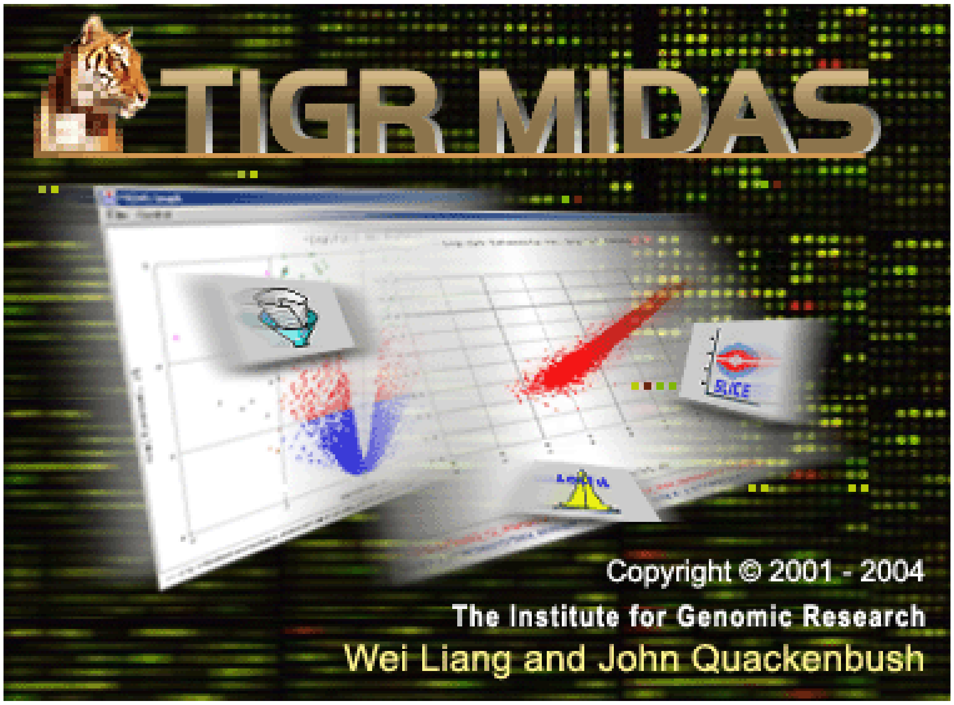

TIGR MIDAS
Copyright © 2001 - 2004
The Institute for Genomic Research
Wei Liang and John Quackenbush

TIGR MIDAS

Microarray Data Analysis System

Version 2.19
November 2004

Table of Contents

1 General Information

1.1 Obtaining MIDAS

www.tigr.org/software/tm4/

Maintainer / Contact Information

TIGR MIDAS Team – midas@tigr.org

Platform / System Requirements

Java Runtime Environment (JRE) 1.4.1 or later

1.2 Referencing MIDAS

Users of this program should cite:

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003 Feb;34(2):374-8.
http://www.tigr.org/software/tm4/menu/TM4_Biotechniques_2003.pdf

1.3 A note on non-Windows operating systems

The majority of our MIDAS development and testing was performed on Windows operating systems. Although MIDAS will run under other operating systems, there may be some incompatibilities or bugs revealed in this manner. Please report any such issues to midas@tigr.org.

MacOSX users can simulate the ‘right-click’ by using <option> click.

2 TIGR Software Overview

TIGR Microarray Data Analysis System is one member of a suite of microarray data management and analysis applications developed at The Institute for Genomic Research (TIGR). Within the suite, known as TM4, there are four programs: MADAM, Spotfinder, MIDAS and MeV. Together, they provide functions for managing microarray experimental conditions and data, converting scanned slide images into numerical data, normalizing the data and finally analyzing that normalized data. These tools are all OSI certified (see section 12) open-source and are freely available through the TIGR website, **www.tigr.org/software/tm4**.

The Microarray Data Manager (MADAM) is a data management tool used to upload, download, and display a plethora of microarray data to and in a database management system (MySQL). An interface to MySQL, Madam allows scientists and researchers to electronically record, capture, and administrate annotated gene expression and experiment data to be shared with and ultimately used by others within the scientific community.

TIGR Spotfinder is image-processing software created for analysis of the image files generated in microarray expression studies. TIGR Spotfinder uses a fast and reproducible algorithm to identify the spots in the array and provide quantification of expression levels.

TIGR Microarray Data Analysis System (MIDAS) is an application that allows user to perform normalization and other statistical data analysis, trim the raw experimental data, and create output for MeV.

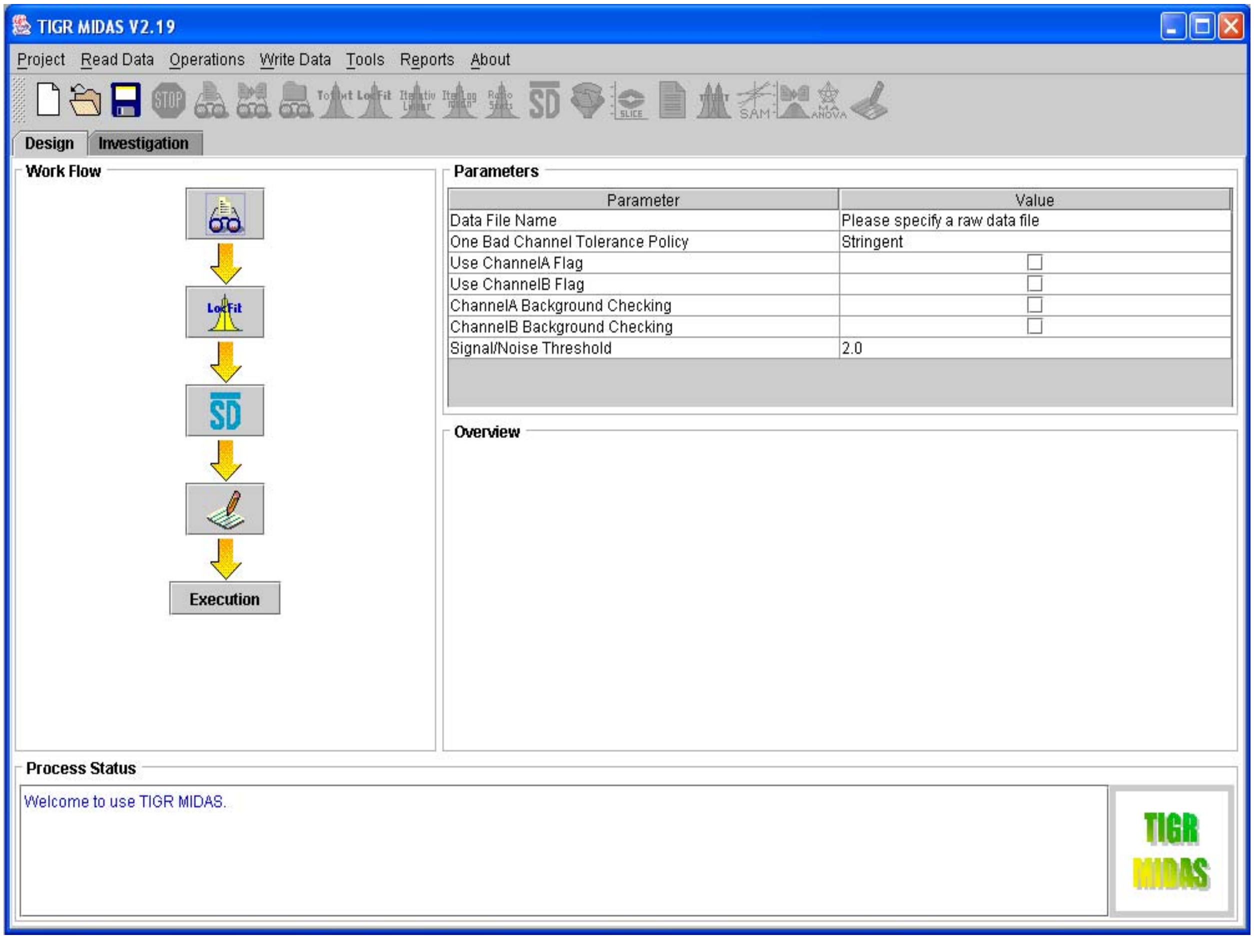
TIGR MultiExperiment Viewer (MeV) is an application that allows the viewing of processed microarray slide representations and the identification of genes and expression patterns of interest. Slides can be viewed one at a time in detail or in groups for comparison purposes. A variety of normalization algorithms and clustering analyses allow the user flexibility in creating meaningful views of the expression data.

3 Starting TIGR MIDAS

If using Windows, run the MIDAS batch file **midas.bat** to start the program. This batch file invokes the Java interpreter and stores input parameters. Similarly, if using Linux or Unix, run the **midas.sh** file. Macintosh users should double-click on the application file named **MidasMac**.

MIDAS user interface is composed of a menu bar and a work space tabbed with two panels: *Design panel* and *Investigation panel*.

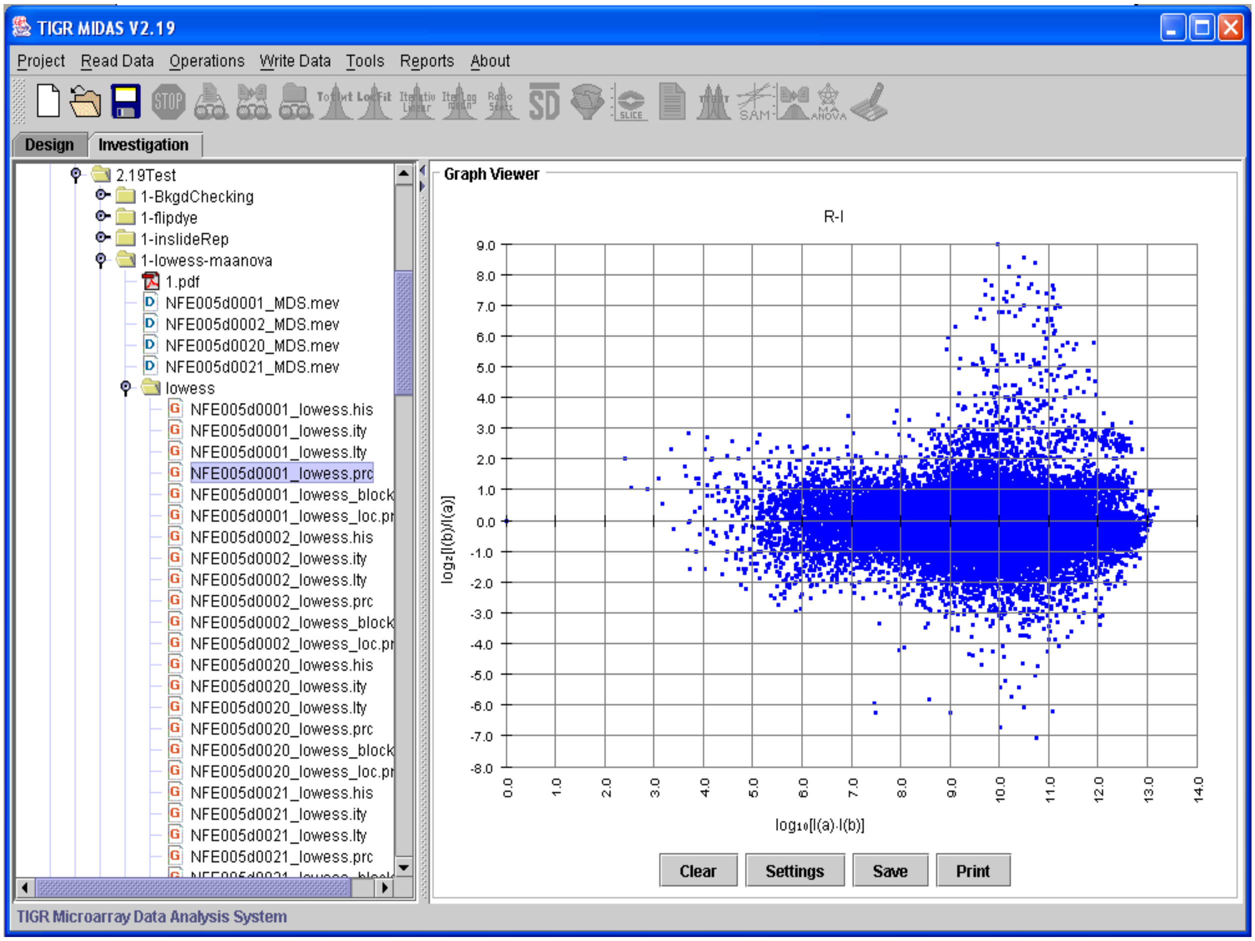
Design panel is the work place for user to define and execute analysis. It consists of a Work Flow window, a Parameters window, an Overview window and a Process Status window.



Parameter	Value
Data File Name	Please specify a raw data file
One Bad Channel Tolerance Policy	Stringent
Use ChannelA Flag	☐
Use ChannelB Flag	☐
ChannelA Background Checking	☐
ChannelB Background Checking	☐
Signal/Noise Threshold	2.0

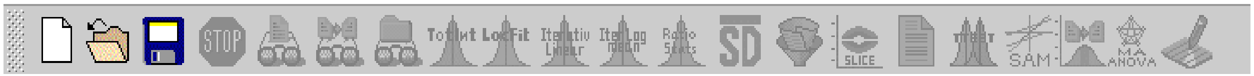
Welcome to use TIGR MIDAS.

Investigation panel is the work place for user to view and investigate the data and analysis plots.



The menu bar includes a text menu bar

and a graphical toolbar



You can perform operations from either the text menu bar or the graphical toolbar.

3.1 Project

Project in MIDAS is a working unit that contains all information about a data analysis operation, such as the process flow, parameter settings for each step in the flow, final outputs, analysis reports, error logs, plot files, etc. A MIDAS project is always connected to a directory in the file system of user's computer. This directory is usually referred as *project folder*.

The project folder can be opened by Pressing **Project/Open** if it already exists or prompt to be created by the user when the **Project/Save** button or **Execution** button is pressed.

Within the project folder, information about the data analysis process is stored in files. These files include *a project file, post-process output files, report files, an error log file (if any) and plot files.*

Project file records the analysis process flow and the parameters related to each process step. Project file is stored with the extension *.prj*.

Post-process output files are the analysis results of the input data files. They follow the same data format as the input files, which can be in either *.tav* format or *.mev.*

An error log files is a plain text file that records errors or exceptions that occurred during an analysis process.

Report files include a text report file (*.rpt*) and a PDF report file (*.pdf*). By default, only the text report is prepared for each MIDAS project. User can demand the PDF report also be prepared for each project by checking the **Report/Create PDF report** checkbox before executing the analysis.

Plot files are the data files that MIDAS uses to make the analysis graph.

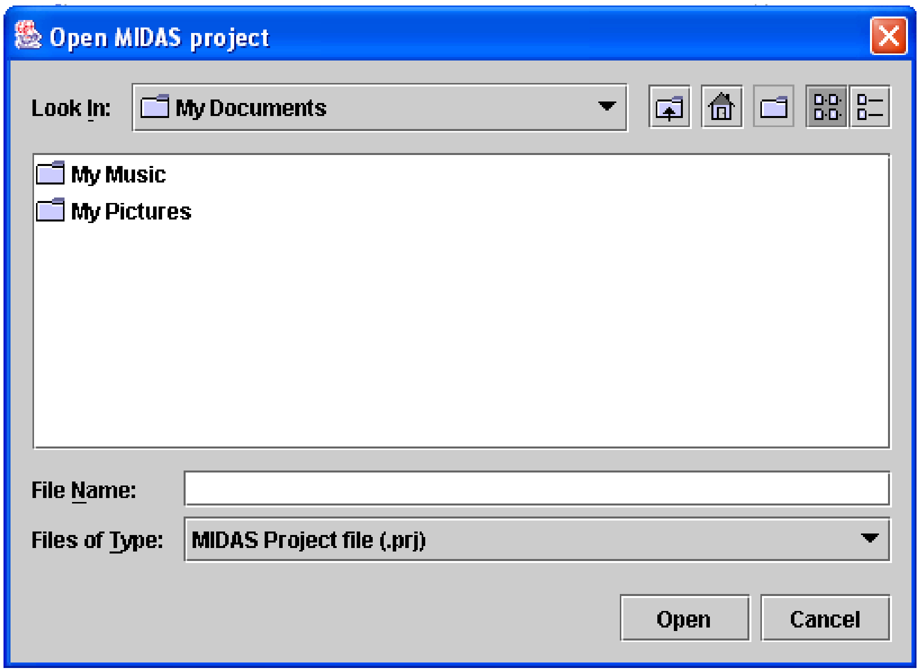
There are five commands under Project: **New**, **Open**, **Save**, **Stop and Exit**.

- **Project/New**

  This command can also be called by clicking 🗋 icon button in the graphical toolbar. It flushes the current process flow in the Work Flow window (Section 4) and reset all parameter settings to default values.

- **Project/Open**

  This command can also be called by clicking 📂 icon button in the graphical toolbar. It pops up an “Open MIDAS project” window for user’s selection of an existing MIDAS project file, opens the selected project file, displays the process flow in the Work Flow window and loads the parameters related to each process steps that are saved when the project file was created.



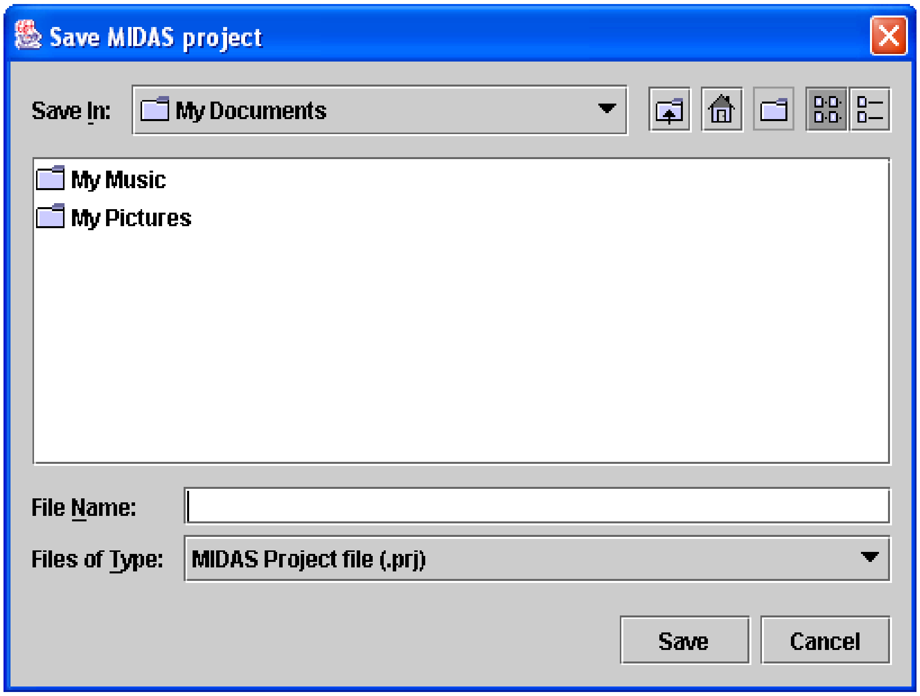
Look In: My Documents

File Name:

Files of Type: MIDAS Project file (.prj)

- **Project/Save**

This command can also be called by clicking 💾 icon button in the graphical toolbar. It pops up a “Save MIDAS project” window and saves the current working process flow displayed in the Work Flow window and all related parameter settings for each step into a *.prj* file.



Save In: My Documents

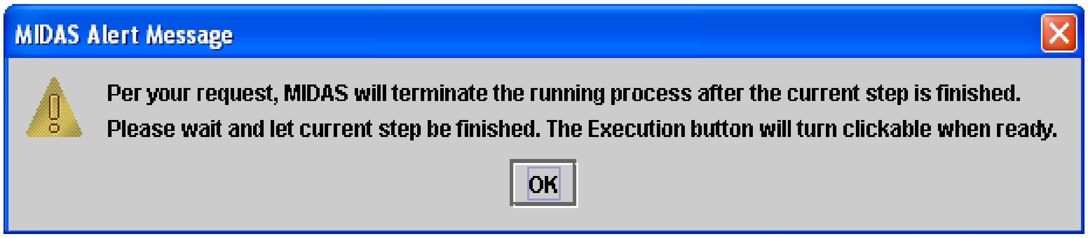
File Name:

Files of Type: MIDAS Project file (.prj)

- **Project/Stop**

This command can also be called by clicking STOP icon button in the graphical toolbar. It terminates the execution of a MIDAS project.

Note: Once the "Stop" request is issued by user, MIDAS needs to finish the current running process module before completely aborting the whole process flow and the "Execution" button becoming clickable again. A MIDAS alert message will be displayed to notify user about this delay.



Per your request, MIDAS will terminate the running process after the current step is finished.
Please wait and let current step be finished. The Execution button will turn clickable when ready.

- **Project/Exit**
  This command allows exiting the entire application.

3.2 Read Data

MIDAS accepts input data file(s) in TIGR TAV format (Appendix) and TIGR MEV format (Appendix).

Note: TIGR Microarray software group also developed data format conversion software called ExpressConverter, which can convert a variety of commonly used microrray data format such as GenePix format to TIGR TAV or TIGR MEV format. ExpressConverter is also OSI certified (see section 4) open-source and are freely available through the TIGR website, **www.tigr.org/software/tm4**.

Three data input modes are provided: *Read single data file* mode, *Read flip dye data file pair* mode and *Read data directory* mode.

3.2.1 Read single data file mode

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.1).

3.2.2 Read flip dye data file pair mode

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.2).

3.2.3 Read data directory mode

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be for setting the related parameters (Section 5.3).

3.3 Operations

MIDAS provides a variety of statistical methods to analyze the data. The methods include: Total intensity normalization, Locfit (LOWESS) normalization, Iterative linear regression normalization, Iterative log mean centering normalization, Ratio statistics normalization and confidence interval checking, Standard deviation regularization, Low intensity filtering, Slice analysis (z-score filtering), In-slide replicates analysis and Flip dye consistency checking.

3.3.1 Total intensity normalization

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.4).

3.3.2 Locfit (LOWESS) normalization

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.5).

3.3.3 Iterative linear regression normalization

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.6).

3.3.4 Iterative log mean centering normalization

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.7).

3.3.5 Ratio statistics normalization and confidence interval checking

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.8).

3.3.6 Standard deviation regularization

This command can also be called by clicking SD icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.9).

3.3.7 Low intensity filter

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.10).

3.3.8 Slice analysis

This command can also be called by clicking SLICE icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.11).

3.3.9 In-slide replicates analysis

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.12).

3.3.10 Cross-slide replicates Ttest

This command can also be called by clicking TTEST icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.13).

3.3.11 Cross-slide replicates SAM

This command can also be called by clicking SAM icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.14).

3.3.12 Flip dye consistency checking

This command can also be called by clicking icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.15).

3.3.13 MA-ANOVA

This command can also be called by clicking MA ANOVA icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.16).

3.4 Write Data

- **Write**

  This command can also be called by clicking ✏ icon button in the graphical toolbar. It displays a corresponding process icon button in the Work Flow window, which can be clicked for setting the related parameters (Section 5.17).

3.5 Tools

- **Set all parameters**
  This command allows all parameters settings for each process icon in the current working process flow be displayed and set altogether.

3.6 Report

- **Create TXT Report**
  This checkbox is checked to indicate a text-format report about the data analysis will be generated at the end of the data process. The report records general information about the data analysis (MIDAS version, time elapsed for the data process, project file name), parameters setting, as well as mid-calculation results for each individual data analysis step. This report is created by default.

- **Create PDF Report**
  This checkbox is checked to indicate a PDF-format report about the data analysis will be generated at the end of the data process. The report records general information about the data analysis (MIDAS version, time elapsed for the data process, project file name), parameters setting, as well as mid-calculation results for each individual data analysis step. Besides, useful graphs for each analysis step are also recorded.

  Note: This option requires more CPU time, RAM and other computer resources. For volume data processing, this option is strongly NOT recommended.

3.7 About

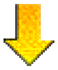
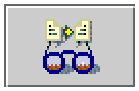
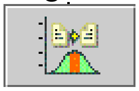
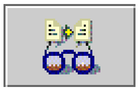
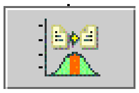
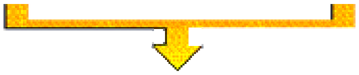
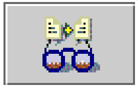
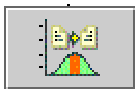
- **About**
  This command displays MIDAS branding and the copyright information.

- **System Information**
  This command displays system information about user’s computer.

- **Acknowledgement**
  This command displays contribution acknowledgement.

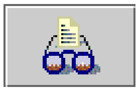
4 Work Flow window

Work Flow window is the working space that user define the analysis process sequence. Corresponding process icon buttons are displayed in this window to show the flow. Each neighboring process icon buttons are linked vertically with each other by the linking arrows. There are three types of linking arrows:

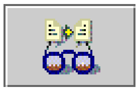
- Single linking arrow  links two process icon buttons except  and 
- Pair split linking arrow links  and other process icon buttons except 
- Pair merge linking arrow  links process icon buttons except  and 

Each process icon button represents a process step defined in the workflow.

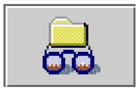
- **Read single data file process icon button**

   Clicking Read single data file process icon button invokes the Read single data file parameter sheet (Section 5.1) in the Parameters window.

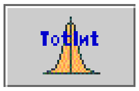
- **Read flip dye data file pair process icon button**

   Clicking Read flip dye data file pair process icon button invokes the Read flip dye data file pair parameter sheet (Section 5.2) in the Parameters window.

- **Read data directory process icon button**

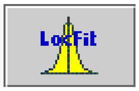
   Clicking Read data directory process icon button invokes the Read data directory parameter sheet (Section 5.3) in the Parameters window.

- **Total intensity normalization process icon button**

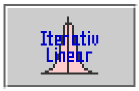
  Clicking Total intensity normalization process icon button invokes the Total intensity normalization parameter sheet (Section 5.4) in the Parameters window.

- **Locfit (LOWESS) normalization process icon button**

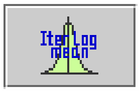
  Clicking Locfit (LOWESS) Normalization process icon button invokes the Locfit (LOWESS) Normalization parameter sheet (Section 5.5) in the Parameters window.

- **Iterative linear regression normalization process icon button**

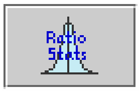
  Clicking Iterative linear regression normalization process icon button invokes the Iterative linear regression normalization parameter sheet (Section 5.6) in the Parameters window.

- **Iterative log mean centering normalization process icon button**

  Clicking Iterative log mean centering normalization process icon button invokes the Iterative log mean normalization parameter sheet (Section 5.7) in the Parameters window.

- **Ratio statistics normalization and CI checking process icon button**

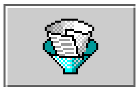
  

  Clicking Ratio statistics normalization and Confidence Interval checking process icon button invokes the Ratio statistics normalization and Confidence Interval checking parameter sheet (Section 5.8) in the Parameters window.

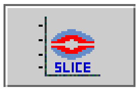
- **Standard deviation regularization process icon button**

  Clicking Standard deviation regularization process icon button invokes the Standard deviation regularization parameter sheet (Section 5.9) in the Parameters window.

- **Intensity filter process icon button**

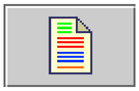

Clicking Intensity filter process icon button invokes the Intensity filter parameter sheet (Section 5.10) in the Parameters window.

- **Slice analysis process icon button**

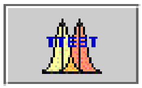


Clicking Slice analysis process icon button invokes the Intensity filter parameter sheet (Section 5.11) in the Parameters window.

- **In-slide replicates analysis process icon button**

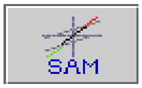

Clicking In-slide replicates analysis process icon button invokes the In-slide replicates analysis parameter sheet (Section 5.12) in the Parameters window.

- **Cross-slide replicates t-test process icon button**



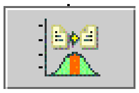
Clicking Cross-slide replicates t-test process icon button invokes the Cross-slide replicates t-test parameter sheet (Section 5.13) in the Parameters window.

- **Cross-slide replicates SAM process icon button**

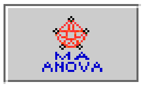


Clicking Cross-slide replicates SAM process icon button invokes the Cross-slide replicates one-class SAM parameter sheet (Section 5.14) in the Parameters window.

- **Flip dye consistency checking process icon button**

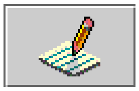

Clicking Flip dye consistency checking process icon button invokes the Flip dye consistency checking parameter sheet (Section 5.15) in the Parameters window.

- **MAANOVA process icon button**



MAANOVA checking process icon button invokes the MAANOVA parameter sheet (Section 5.16) in the Parameters window.

- **Write process icon button**

 Clicking Write process icon button invokes the Write parameter sheet (Section 5.17) in the Parameters window.

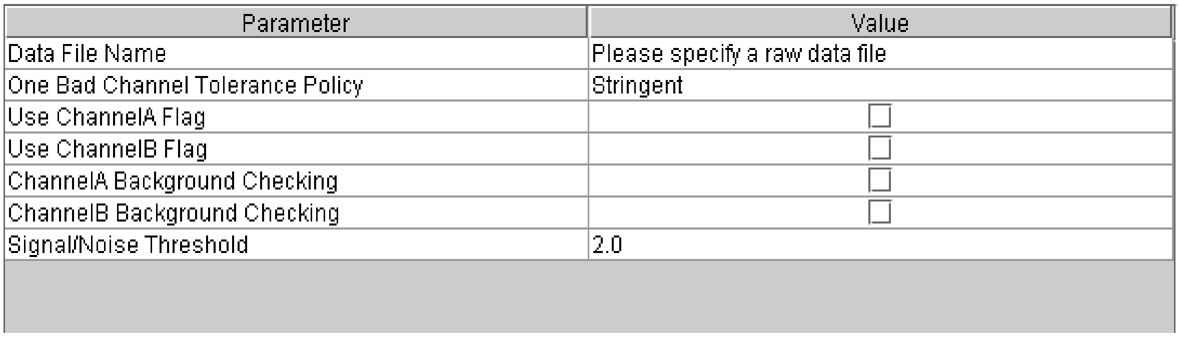
- **Execution process icon button**

Clicking Execution process icon button starts the execution of the process flow as user defined. During the execution, the Process Status window shows the running status of the flow.

5 Parameters window

Parameters windows are bound to process icon buttons in the Work Flow window (Section 4). The parameters for each process icon button are described below:

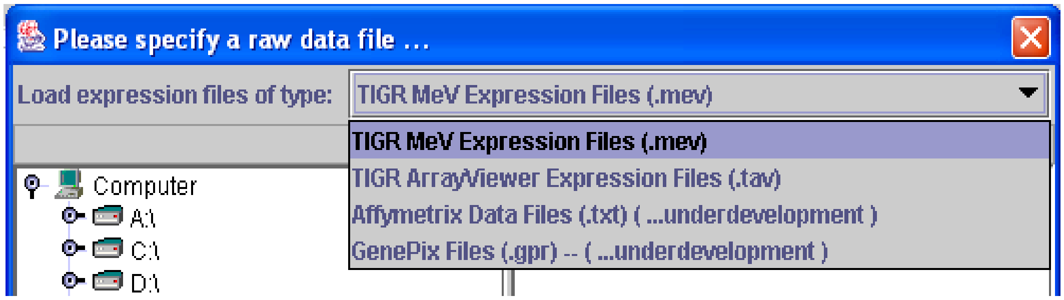
5.1 Read single data file parameters



Parameter	Value
Data File Name	Please specify a raw data file
One Bad Channel Tolerance Policy	Stringent
Use ChannelA Flag	□
Use ChannelB Flag	□
ChannelA Background Checking	□
ChannelB Background Checking	□
Signal/Noise Threshold	2.0

- **Data File Name**
  Input data file name can be selected from a pop-up file system browser window when the corresponding value cell is clicked.

  When the file system browser is displayed, first specify the file format of the input data files from the "Load expression files of type" drop-down list:



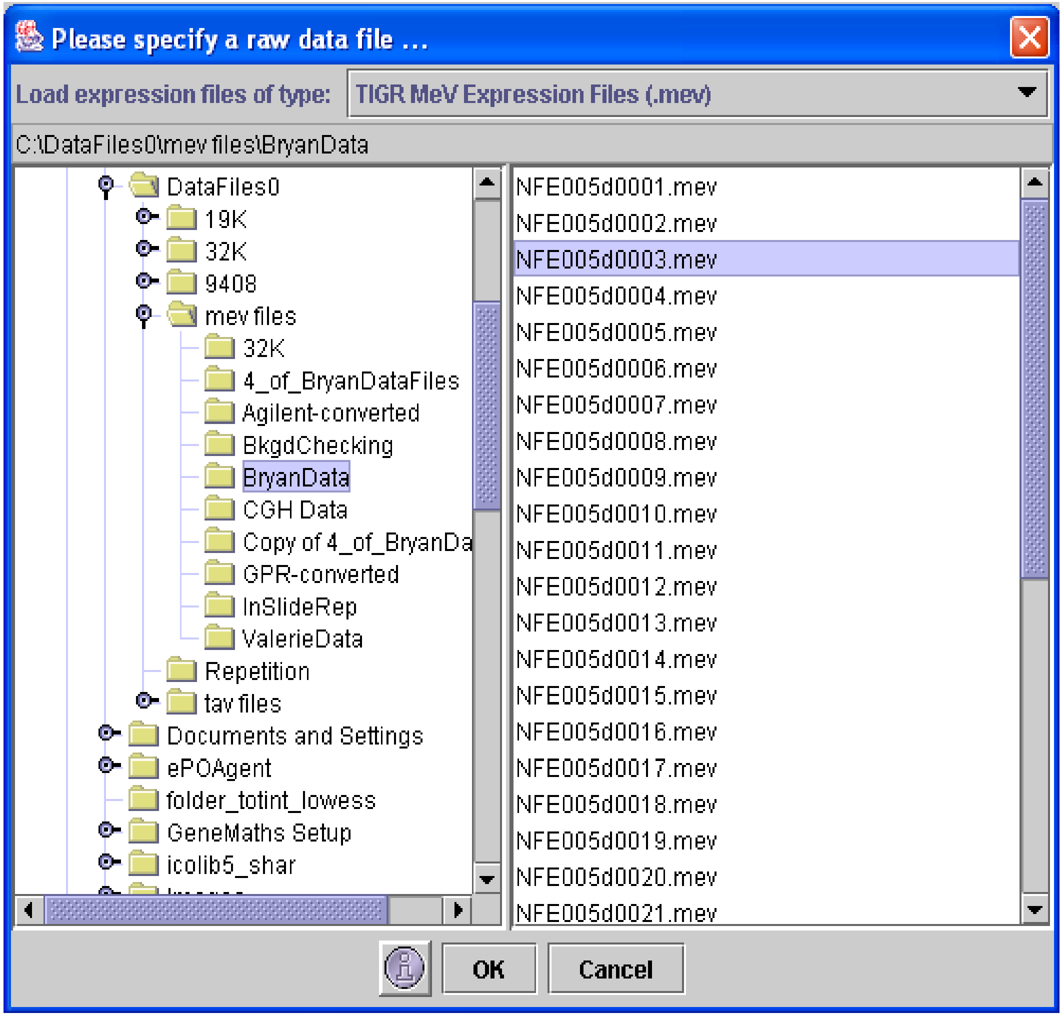
Please specify a raw data file ...

Load expression files of type: TIGR MeV Expression Files (.mev)

TIGR MeV Expression Files (.mev)
TIGR ArrayViewer Expression Files (.tav)
Affymetrix Data Files (.txt) (...underdevelopment)
GenePix Files (.gpr) -- (...underdevelopment)
Computer
A:\
C:\
D:\

In this release of the software, MIDAS only reads TIGR-MEV data format (Appendix) and TIGR-TAV data format (Appendix).

Expand the desired directory nodes in the left panel, click the directory where data files are stored, all data files with the specified format type in this directory will be displayed in the right panel. Select a file, by highlighting the file name in the file list panel.



Please specify a raw data file ...

Load expression files of type: TIGR MeV Expression Files (.mev)

C:\DataFiles0\mev files\BryanData

button

Display a brief overview of supported file types.

OK button

Confirm the selected file name.

Cancel button

Cancel the selection and close the file system browser.

- **One Bad Channel Tolerance Policy**
  When reading a data file, MIDAS reads channel A and channel B intensities for each spot, *preliminary filtering* is then applied by marking those spots with invalid intensities (intensity value less than 1) in both channels as "bad", these spots will be excluded in the down stream analysis defined in the process sequence.

  If a spot has one channel with valid intensity (intensity value not less than 1), but invalid intensity in the other channel, this spot can be either discarded or reserved by replacing the bad channel intensity value by some "fake" number, so that it can still be included in the downstream analysis.

The One Bad Channel Tolerance Policy parameter allows a user to indicate his/her preference on how to deal with one-bad-channel spots. There are two options provided:

Stringent

The one-bad-channel spot will be discarded, both channel A and channel B intensities will be reset to 0 and the spot will be excluded in the downstream analysis.

Generous

The one-bad-channel spot will be reserved; the bad channel intensity will be replaced by a "fake intensity" which is computed as described below:

1. Sort intensities for this channel for all spots;
2. Find the spot counts *N* for those having valid intensity values(greater than 1);
3. Find the lowest *N/10* spots that have valid intensity values;
4. The "fake intensity" is calculated as the arithmetic mean of those spot intensities found in step 3.

- **Use ChannelA Flag**
  When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

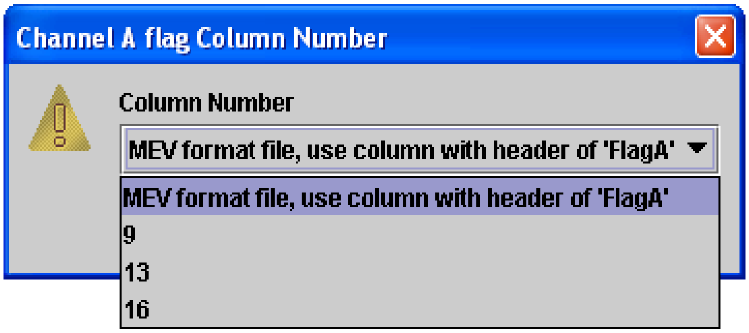
  If *Use ChannelA Flag* is checked, MIDAS will apply channel A flag filtering on each spot that has valid intensities by the steps described below:

  1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
  2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

  When channel A flag filtering is desired, checking the checkbox will bring up a Channel A flag Column Number window.

  If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagA'". MIDAS will check the channel A flags by locating the column with header name of "FlagA". (See TIGR-MEV data file format in Appendix).

  If the data file is in TIGR-TAV format, ChannelA flag column number can only be 9, 13 or 16. The column number can be selected from a drop-down list.


Channel A flag Column Number
Column Number
MEV format file, use column with header of 'FlagA'
MEV format file, use column with header of 'FlagA'
9
13
16

OK button
Confirm the selected channelA flag column number.
Cancel button
Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel A flag checking for each spot".

- **Use ChannelB Flag**
  When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

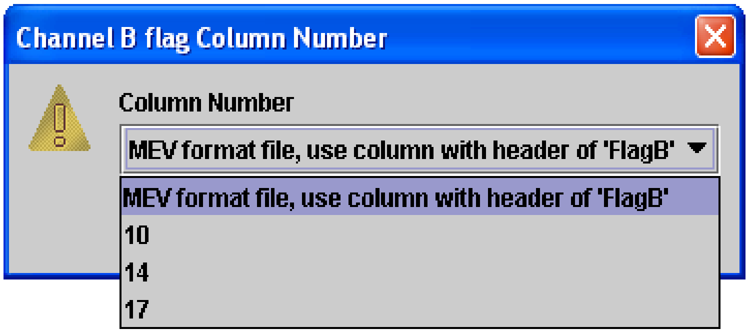
  If *Use ChannelB Flag* is checked, MIDAS will apply channel B flag filtering on each spot that has valid intensities by the steps described below:

  1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
  2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

  When channel B flag filtering is desired, checking the checkbox will bring up a Channel B flag Column Number window.

  If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagB'". MIDAS will check the channel B flags by locating the column with header name of "FlagB". (See TIGR-MEV data file format in Appendix).

  If the data file is in TIGR-TAV format, ChannelA flag column number can only be 10, 14 or 17. The column number can be selected from a drop-down list.


Channel B flag Column Number
Column Number
MEV format file, use column with header of 'FlagB'
MEV format file, use column with header of 'FlagB'
10
14
17

OK button

Confirm the selected channelB flag column number.

Cancel button

Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel B flag checking for each spot".

- **ChannelA Background Checking**

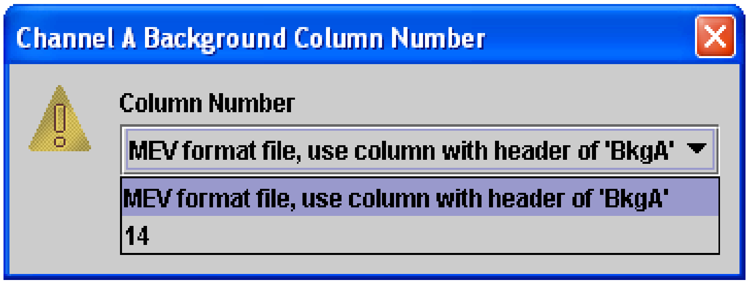
  When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

  If background checking is desired, checking the checkbox will bring up a Channel A Background Column Number window.

  If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'BkgA'". MIDAS will check the channel A background by locating the column with header name of "BkgA" (See TIGR-MEV data file format in Appendix).

  If the data file is in TIGR-TAV format, ChannelA background column number must be 14. The column number can be selected from a drop-down list.

  The algorithm for background checking is explained under caption "Signal/Noise Threshold" below.


Channel A Background Column Number
Column Number
MEV format file, use column with header of 'BkgA'
MEV format file, use column with header of 'BkgA'
14

OK button

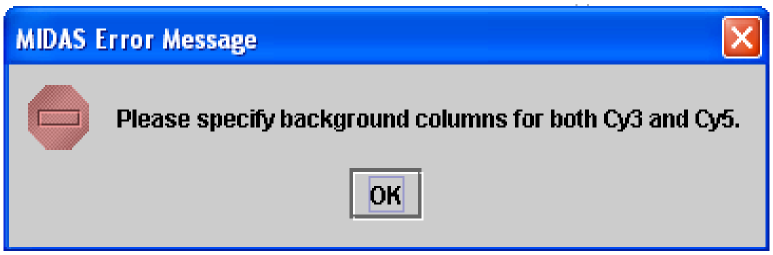
Confirm the selected channelA background column number.

Cancel button

Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel A background checking for each spot".

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



MIDAS Error Message

Please specify background columns for both Cy3 and Cy5.

- **ChannelB Background Checking**

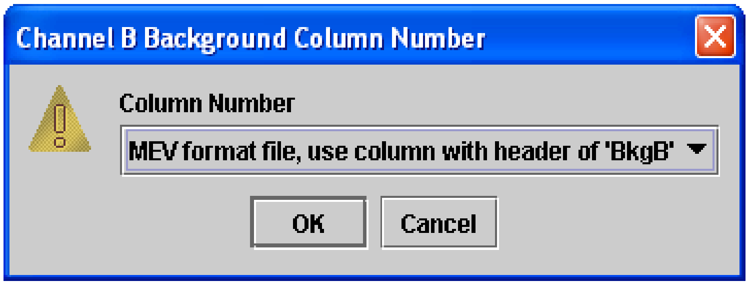
  When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

  If background checking is desired, checking the checkbox will bring up a Channel B Background Column Number window.

  If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'BkgB'". MIDAS will check the channel B background by locating the column with header name of "BkgB" (See TIGR-MEV data file format in Appendix).

  If the data file is in TIGR-TAV format, ChannelB background column number must be 15. The column number can be selected from a drop-down list.

  The algorithm for background checking is explained under caption "Signal/Noise Threshold" below.



Channel B Background Column Number

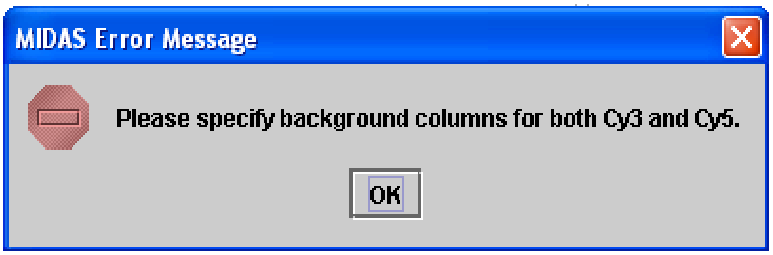
Column Number

MEV format file, use column with header of 'BkgB'

OK button
    Confirm the selected channelB background column.
Cancel button
    Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel B background checking for each spot".

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



Please specify background columns for both Cy3 and Cy5.

- **Signal/Noise Threshold**
  Only when both ChannelA Background Checking and ChannelB Background Checking are applied, spot's Signal/Noise ratios for each channel will be calculated and compared with the threshold set by this parameter. User can specify a Signal/Noise Threshold by clicking the corresponding value cell. MIDAS will apply the algorithm described below for each spot's background filtering.

  Assume:
      The user specified a Signal/Noise is *SN*;
      Background-corrected Intensity for channel A is denoted as *I(A)*;
      Background-corrected Intensity for channel B is denoted as *I(B)*;
      Background intensity for channel A is denoted as *Bkg(A)*;
      Background intensity for channel B is denoted as *Bkg(B)*;

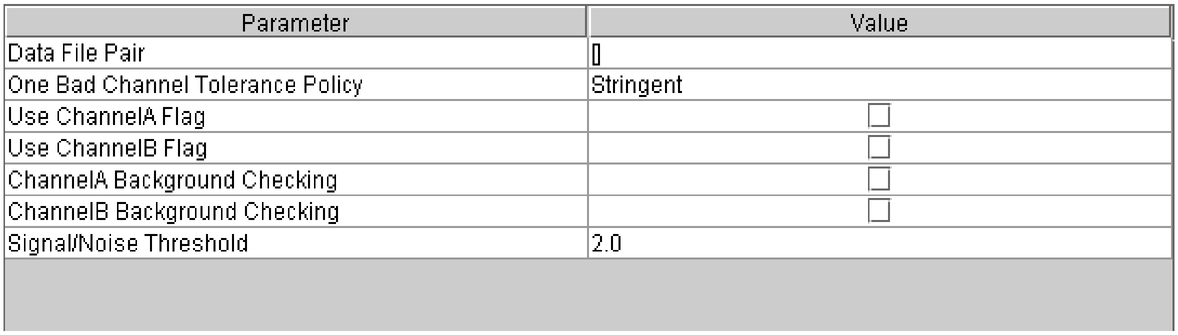
  For each spot, if

$$\frac{I(A)+Bkg(A)}{Bkg(A)} \le SN \text{ or } \frac{I(B)+Bkg(B)}{Bkg(B)} \le SN$$

  this spot will be marked as "bad", its *I(A)* and *I(B)* will be reset to 0s, so that the spot will be excluded in any down stream analysis.

  This parameter is set 2.0 by default, which means "Only if the background-corrected signal intensity is greater than or equal to the background intensity for both channel A and B of a spot, will this spot be considered as a 'good spot' and included in the downstream analysis".

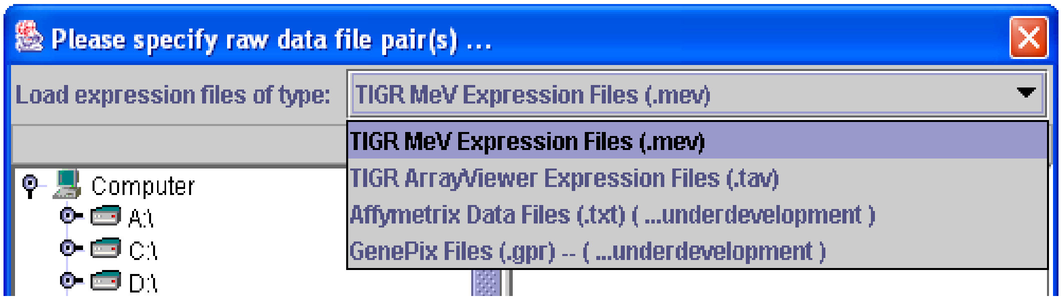
5.2 Read flip dye data file pair parameters



Parameter	Value
Data File Pair	[]
One Bad Channel Tolerance Policy	Stringent
Use ChannelA Flag	☐
Use ChannelB Flag	☐
ChannelA Background Checking	☐
ChannelB Background Checking	☐
Signal/Noise Threshold	2.0

- **Data File Pair**
  Input data file pair can be selected from a pop-up file system browser window when the corresponding value cell is clicked.

  When the file system browser is displayed, first specify the file format of the input data files from the "Load expression files of type" drop-down list:



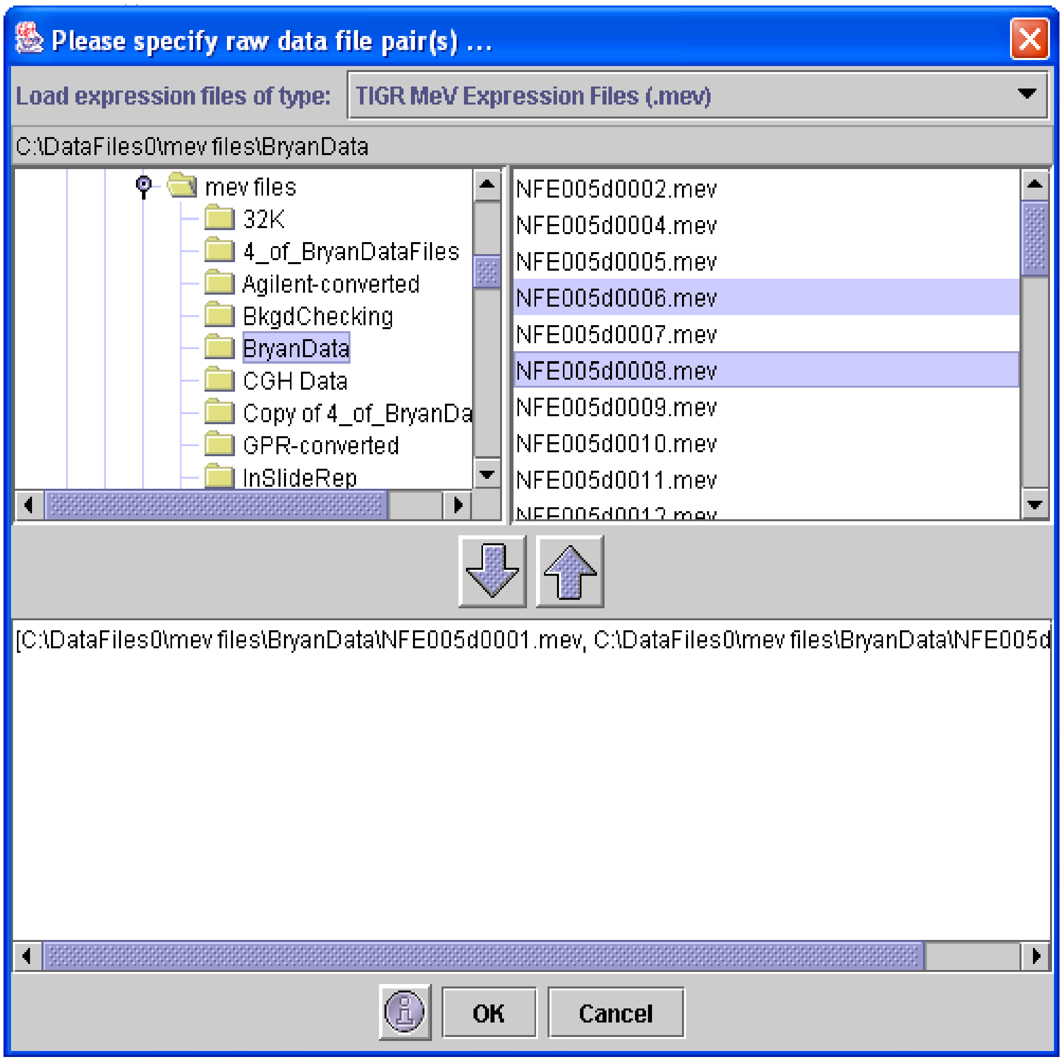
Please specify raw data file pair(s) ...

Load expression files of type: TIGR MeV Expression Files (.mev)

TIGR MeV Expression Files (.mev)
TIGR ArrayViewer Expression Files (.tav)
Affymetrix Data Files (.txt) (...underdevelopment)
GenePix Files (.gpr) -- (...underdevelopment)
Computer
A:\
C:\
D:\

In this release of the software, MIDAS only reads TIGR-MEV data format (Appendix) and TIGR-TAV data format (Appendix).

Expand the desired directory nodes in the upper-left panel, click the directory where data files are stored, all data files with the specified format type in this directory will be displayed in the upper-right panel. From the file list, highlight the first file of a flip dye pair, while it is highlighted, hold "Ctl" key and highlight the second file of the flip dye pair. When both file names are highlighted, press ⇩ button to pair them up. The file pair will be displayed in the lower panel (pair list panel). Multiple pairs can be selected by repeating the above steps. If two files are paired up by mistake, the pair can be de-selected from the pair list panel by highlighting the pair and clicking ⇧ button. The two files will return to the upper-right file list panel.



Please specify raw data file pair(s) ...

Load expression files of type: TIGR MeV Expression Files (.mev)

C:\DataFiles0\mev files\BryanData

[C:\DataFiles0\mev files\BryanData\NFE005d0001.mev, C:\DataFiles0\mev files\BryanData\NFE005d

ⓘ button
    Display a brief overview of supported file types.

OK button
    Confirm the selected file pair(s).

Cancel button
    Cancel the selection and close the file system browser.

- **One Bad Channel Tolerance Policy**

  When reading a data file, MIDAS reads channel A and channel B intensities for each spot, preliminary filtering is then applied by marking those spots with invalid intensities (intensity value less than 1) in both channels as “bad”, these spots will be excluded in the down stream analysis defined in the process sequence.

  If a spot has one channel with valid intensity (intensity value not less than 1), but invalid intensity in the other channel, this spot can be either discarded or reserved by replacing the bad channel intensity value by some "fake" number, so that it can still be included in the downstream analysis.

The One Bad Channel Tolerance Policy parameter allows a user to indicate his/her preference on how to deal with one-bad-channel spots. There are two options provided:

Stringent

> The one-bad-channel spot will be discarded, both channel A and channel B intensities will be reset to 0 and the spot will be excluded in the downstream analysis.

Generous

> The one-bad-channel spot will be reserved; the bad channel intensity will be replaced by a "fake intensity" which is computed as described below:
>
> 1. Sort intensities for this channel for all spots;
> 2. Find the spot counts *N* for those having valid intensity values(greater than 1);
> 3. Find the lowest *N/10* spots that have valid intensity values;
> 4. The "fake intensity" is calculated as the arithmetic mean of those spot intensities found in step 3.

- **Use ChannelA Flag**

Note: This parameter will be applied to both files in the flip dye pair.

When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

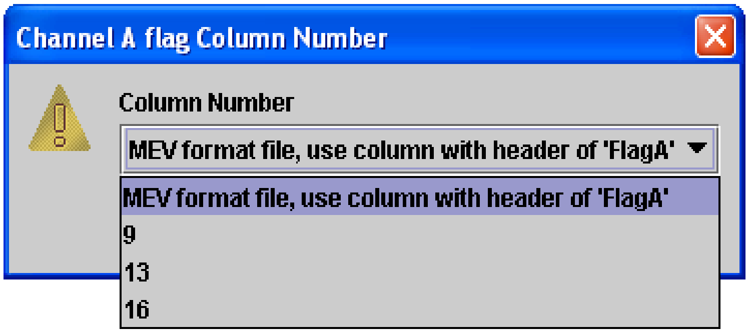
If *Use ChannelA Flag* is checked, MIDAS will apply channel A flag filtering on each spot that has valid intensities by the steps described below:

1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

When channel A flag filtering is desired, checking the checkbox will bring up a Channel A flag Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagA'". MIDAS will check the channel A flags by locating the column with header name of "FlagA". (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA flag column number can only be 9, 13 or 16. The column number can be selected from a drop-down list.


Channel A flag Column Number
Column Number
MEV format file, use column with header of 'FlagA'
MEV format file, use column with header of 'FlagA'
9
13
16

OK button
Confirm the selected channelA flag column number.
Cancel button
Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel A flag checking for each spot".

- **Use ChannelB Flag**

Note: This parameter will be applied to both files in the flip dye pair.

When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

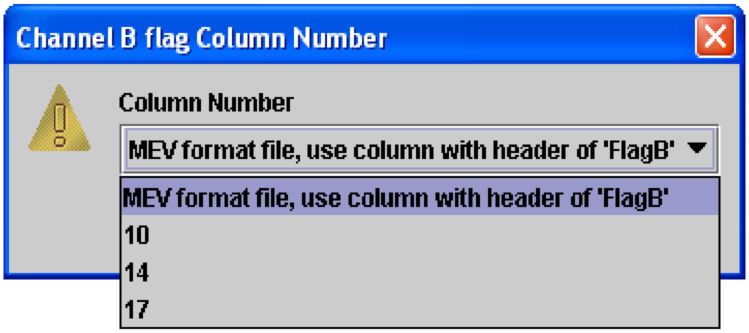
If *Use ChannelB Flag* is checked, MIDAS will apply channel B flag filtering on each spot that has valid intensities by the steps described below:

1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

When channel B flag filtering is desired, checking the checkbox will bring up a Channel B flag Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagB'". MIDAS will check the channel B flags by locating the column with header name of "FlagB". (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA flag column number can only be 10, 14 or 17. The column number can be selected from a drop-down list.



Channel B flag Column Number

Column Number: MEV format file, use column with header of 'FlagB'

OK button
Confirm the selected channelB flag column number.
Cancel button
Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel B flag checking for each spot".

- **ChannelA Background Checking**

Note: This parameter will be applied to both files in the flip dye pair.

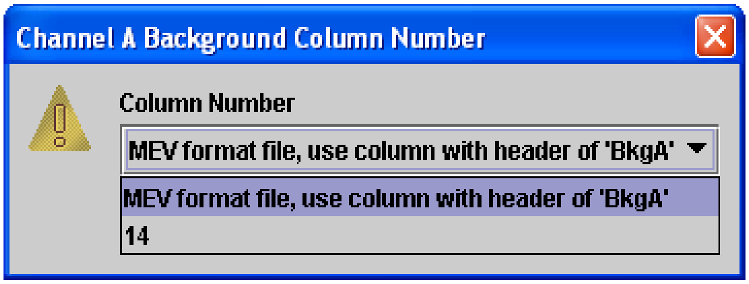
When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

If background checking is desired, checking the checkbox will bring up a Channel A Background Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'BkgA'". MIDAS will check the channel A background by locating the column with header name of "BkgA" (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA background column number must be 14. The column number can be selected from a drop-down list.

The algorithm for background checking is explained under caption "Signal/Noise Threshold" below.



Column Number

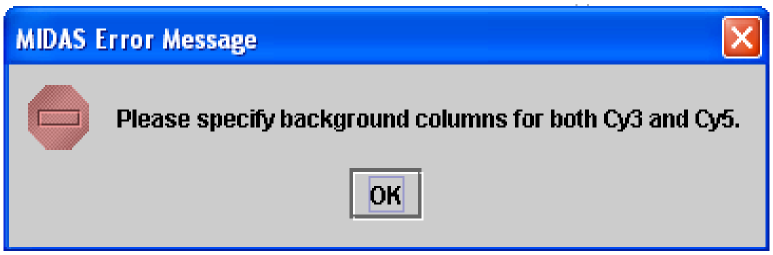
MEV format file, use column with header of 'BkgA'

MEV format file, use column with header of 'BkgA'
14

OK button
Confirm the selected channelA background column number.
Cancel button
Cancel the selection.

This parameter is set to false by default, which means “Do not apply channel A background checking for each spot”.

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



Please specify background columns for both Cy3 and Cy5.

- **ChannelB Background Checking**

Note: This parameter will be applied to both files in the flip dye pair.

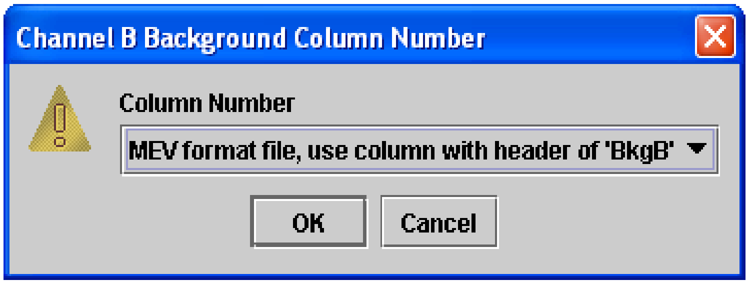
When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

If background checking is desired, checking the checkbox will bring up a Channel B Background Column Number window.

If the data file is in TIGR-MEV format, choose “MEV format file, use column with header of ‘BkgB’”. MIDAS will check the channel B background by locating the column with header name of “BkgB” (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelB background column number must be 15. The column number can be selected from a drop-down list.

The algorithm for background checking is explained under caption “Signal/Noise Threshold” below.

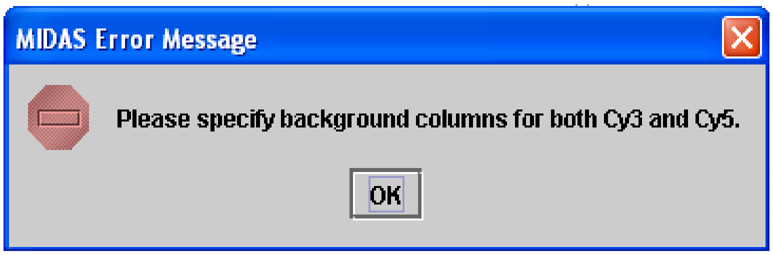


Channel B Background Column Number
Column Number
MEV format file, use column with header of 'BkgB'
OK
Cancel

OK button
    Confirm the selected channelB background column.
Cancel button
    Cancel the selection.

This parameter is set to false by default, which means “Do not apply channel B background checking for each spot”.

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



MIDAS Error Message

Please specify background columns for both Cy3 and Cy5.

- **Signal/Noise Threshold**
  Only when both ChannelA Background Checking and ChannelB Background Checking are applied, spot’s Signal/Noise ratios for each channel will be calculated and compared with the threshold set by this parameter. User can specify a Signal/Noise Threshold by clicking the corresponding value cell. MIDAS will apply the algorithm described below for each spot’s background filtering.

  Assume:
      The user specified a Signal/Noise is *SN*;
      Background-corrected Intensity for channel A is denoted as *I(A)*;
      Background-corrected Intensity for channel B is denoted as *I(B)*;
      Background intensity for channel A is denoted as *Bkg(A)*;
      Background intensity for channel B is denoted as *Bkg(B)*;

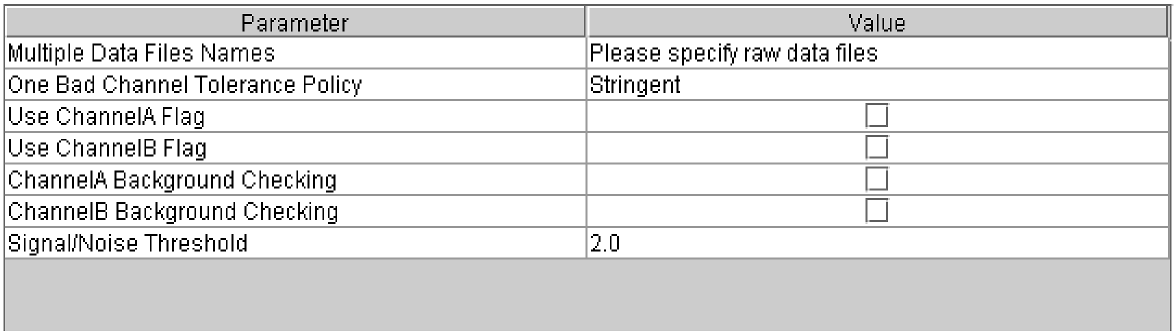
For each spot, if

$$\frac{I(A)+Bkg(A)}{Bkg(A)} \le SN \text{ or } \frac{I(B)+Bkg(B)}{Bkg(B)} \le SN$$

this spot will be marked as "bad", its *I(A)* and *I(B)* will be reset to 0s, so that the spot will be excluded in any down stream analysis.

This parameter is set 2.0 by default, which means "Only if the background-corrected signal intensity is greater than or equal to the background intensity for both channel A and B of a spot, will this spot be considered as a 'good spot' and included in the downstream analysis".

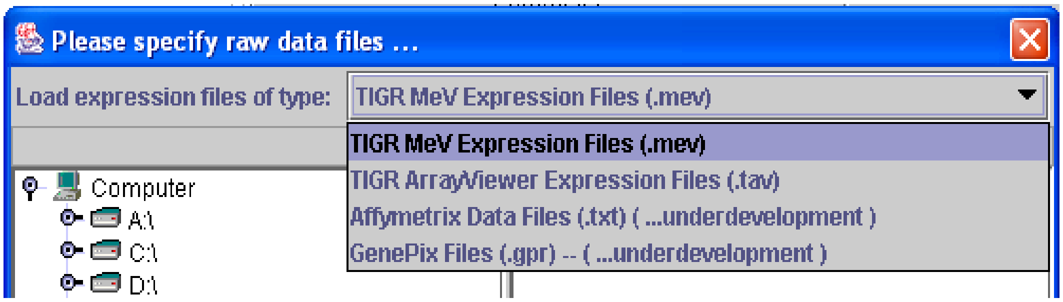
5.3 Read data directory parameters



Parameter	Value
Multiple Data Files Names	Please specify raw data files
One Bad Channel Tolerance Policy	Stringent
Use ChannelA Flag	□
Use ChannelB Flag	□
ChannelA Background Checking	□
ChannelB Background Checking	□
Signal/Noise Threshold	2.0

- **Multiple Data Files Names**
  Multiple input data files names can be selected from a pop-up file system browser window when the corresponding value cell is clicked.

  When the file system browser is displayed, first specify the file format of the input data files from the "Load expression files of type" drop-down list:



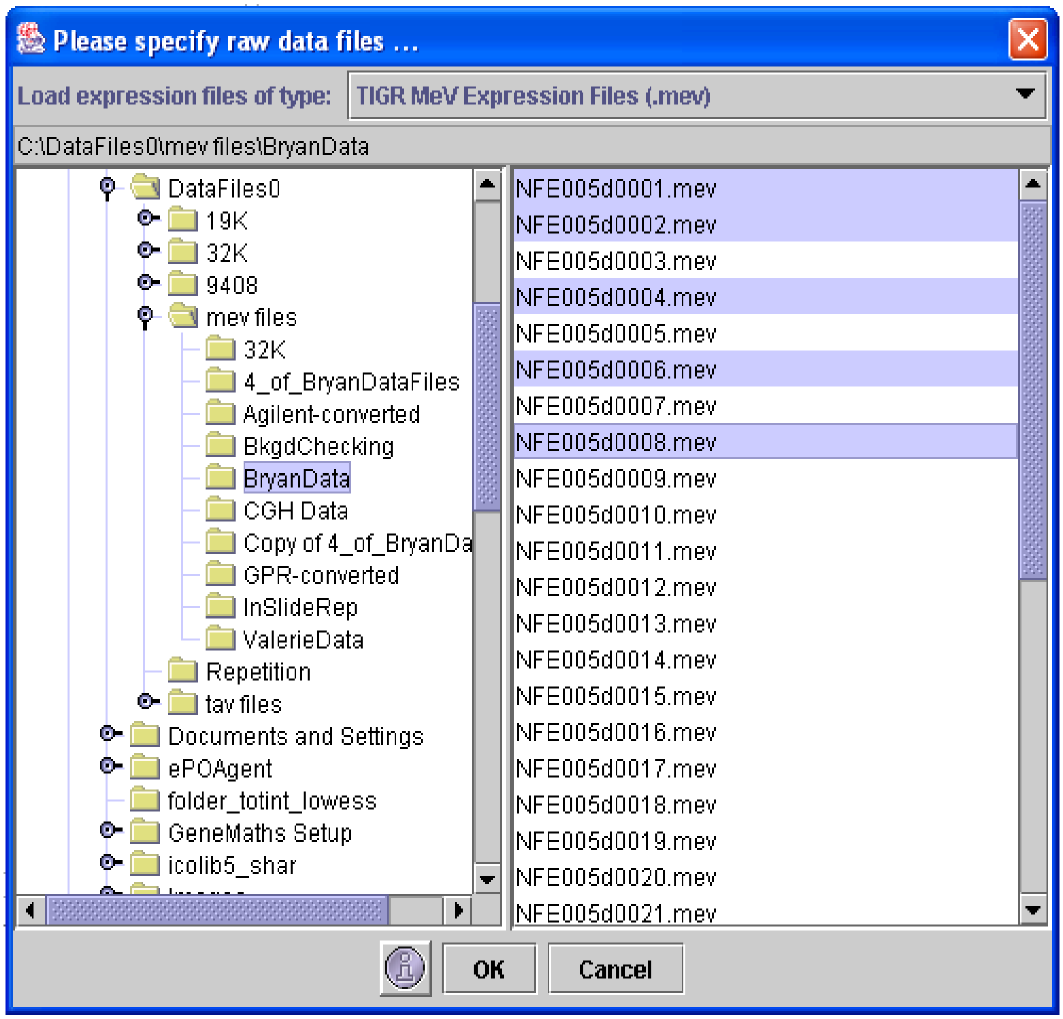
Please specify raw data files ...

Load expression files of type: TIGR MeV Expression Files (.mev)

TIGR MeV Expression Files (.mev)
TIGR ArrayViewer Expression Files (.tav)
Affymetrix Data Files (.txt) (...underdevelopment)
GenePix Files (.gpr) -- (...underdevelopment)
Computer
A:\
C:\
D:\

In this release of the software, MIDAS only reads TIGR-MEV data format (Appendix) and TIGR-TAV data format (Appendix).

Expand the desired directory nodes in the left panel, click the directory where data files are stored, all data files with the specified format type in this directory will be displayed in the right panel. Select the data files by holding “Ctl” key and highlighting the file names in the right panel.



Please specify raw data files …

Load expression files of type: TIGR MeV Expression Files (.mev)

C:\DataFiles0\mev files\BryanData

ⓘ button
Display a brief overview of supported file types.
OK button
Confirm the selected directory name.
Cancel button
Cancel the selection and close the file system browser.

- **One Bad Channel Tolerance Policy**
When reading a data file, MIDAS reads channel A and channel B intensities for each spot, preliminary filtering is then applied by marking those spots with invalid intensities (intensity value less than 1) in both channels as “bad”, these spots will be excluded in the down stream analysis defined in the process sequence.

If a spot has one channel with valid intensity (intensity value not less than 1), but invalid intensity in the other channel, this spot can be either discarded or reserved by replacing the bad channel intensity value by some "fake" number, so that it can still be included in the downstream analysis.

The One Bad Channel Tolerance Policy parameter allows a user to indicate his/her preference on how to deal with one-bad-channel spots. There are two options provided:

Stringent

The one-bad-channel spot will be discarded, both channel A and channel B intensities will be reset to 0 and the spot will be excluded in the downstream analysis.

Generous

The one-bad-channel spot will be reserved; the bad channel intensity will be replaced by a "fake intensity" which is computed as described below:

1. Sort intensities for this channel for all spots;
2. Find the spot counts *N* for those having valid intensity values(greater than 1);
3. Find the lowest *N/10* spots that have valid intensity values;
4. The "fake intensity" is calculated as the arithmetic mean of those spot intensities found in step 3.

- **Use ChannelA Flag**

Note: This parameter will be applied to all selected file.

When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

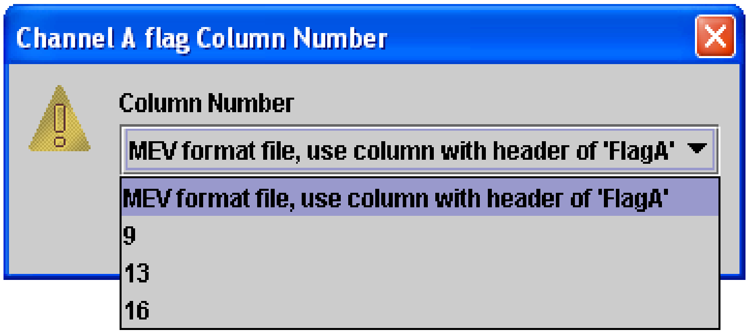
If *Use ChannelA Flag* is checked, MIDAS will apply channel A flag filtering on each spot that has valid intensities by the steps described below:

1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

When channel A flag filtering is desired, checking the checkbox will bring up a Channel A flag Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagA'". MIDAS will check the channel A flags by locating the column with header name of "FlagA". (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA flag column number can only be 9, 13 or 16. The column number can be selected from a drop-down list.



Channel A flag Column Number
Column Number
MEV format file, use column with header of 'FlagA'
MEV format file, use column with header of 'FlagA'
9
13
16

OK button

Confirm the selected channelA flag column number.

Cancel button

Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel A flag checking for each spot".

- **Use ChannelB Flag**

Note: This parameter will be applied to all selected files.

When quality control flags for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply flag filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above).

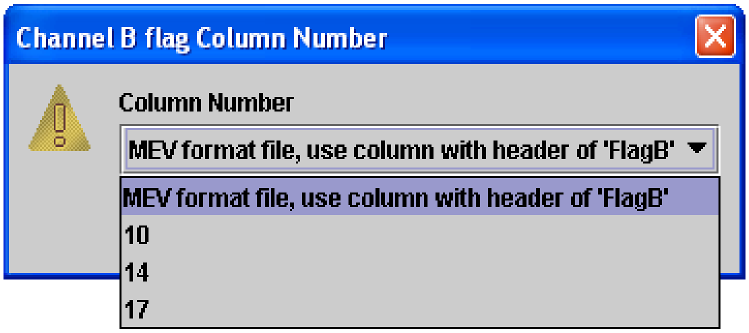
If *Use ChannelB Flag* is checked, MIDAS will apply channel B flag filtering on each spot that has valid intensities by the steps described below:

1. If the flag value of this spot specified in the flag column is "B" or "C" (Appendix), the spot will be marked as "good" and be included in the downstream analysis.
2. If the flag value of this spot specified in the flag column is other than "B" or "C", the spots will be marked as "bad" and be excluded in the downstream analysis.

When channel B flag filtering is desired, checking the checkbox will bring up a Channel B flag Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'FlagB'". MIDAS will check the channel B flags by locating the column with header name of "FlagB". (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA flag column number can only be 10, 14 or 17. The column number can be selected from a drop-down list.


Channel B flag Column Number
Column Number
MEV format file, use column with header of 'FlagB'
MEV format file, use column with header of 'FlagB'
10
14
17

OK button
Confirm the selected channelB flag column number.
Cancel button
Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel B flag checking for each spot".

- **ChannelA Background Checking**

Note: This parameter will be applied to all selected files.

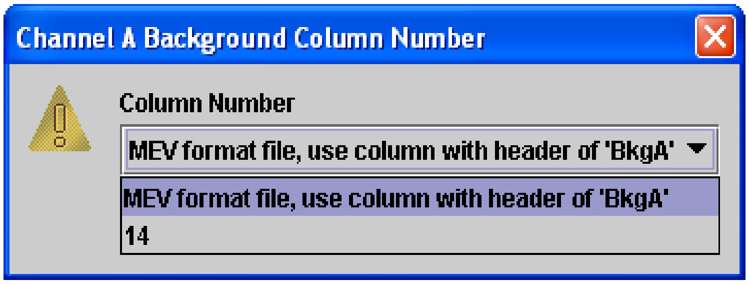
When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelA intensities, I(A), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

If background checking is desired, checking the checkbox will bring up a Channel A Background Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'BkgA'". MIDAS will check the channel A background by locating the column with header name of "BkgA" (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelA background column number must be 14. The column number can be selected from a drop-down list.

The algorithm for background checking is explained under caption "Signal/Noise Threshold" below.



Channel A Background Column Number

Column Number

MEV format file, use column with header of 'BkgA'

OK button

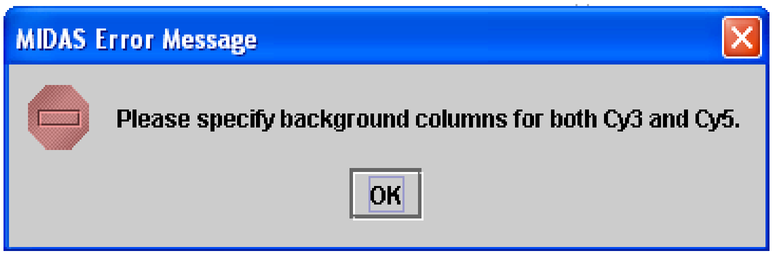
Confirm the selected channelA background column number.

Cancel button

Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel A background checking for each spot".

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



MIDAS Error Message

Please specify background columns for both Cy3 and Cy5.

- **ChannelB Background Checking**

Note: This parameter will be applied to all selected files.

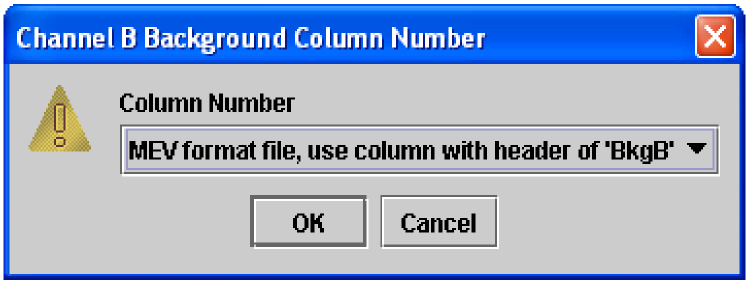
When channel background intensities for spots are available in the input data file, user can use this parameter to instruct MIDAS whether to apply a background filtering on channelB intensities, I(B), for those spots passed the preliminary filtering (see One Bad Channel Tolerance Policy above) and flag filtering (if applied).

If background checking is desired, checking the checkbox will bring up a Channel B Background Column Number window.

If the data file is in TIGR-MEV format, choose "MEV format file, use column with header of 'BkgB'". MIDAS will check the channel B background by locating the column with header name of "BkgB" (See TIGR-MEV data file format in Appendix).

If the data file is in TIGR-TAV format, ChannelB background column number must be 15. The column number can be selected from a drop-down list.

The algorithm for background checking is explained under caption "Signal/Noise Threshold" below.



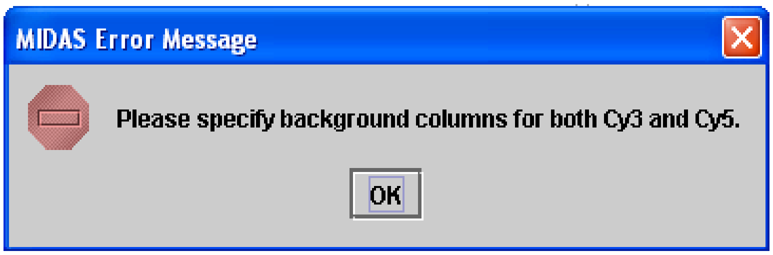
Column Number

MEV format file, use column with header of 'BkgB'

OK button
    Confirm the selected channelB background column.
Cancel button
    Cancel the selection.

This parameter is set to false by default, which means "Do not apply channel B background checking for each spot".

Note: If background checking is desired, both ChannelA Background Column Number and ChannelB Background Column Number must be selected. Otherwise, MIDAS will display an error message window as below:



Please specify background columns for both Cy3 and Cy5.

- **Signal/Noise Threshold**
  Only when both ChannelA Background Checking and ChannelB Background Checking are applied, spot's Signal/Noise ratios for each channel will be calculated and compared with the threshold set by this parameter. User can specify a Signal/Noise Threshold by clicking the corresponding value cell. MIDAS will apply the algorithm described below for each spot's background filtering.

  Assume:
      The user specified a Signal/Noise is *SN*;
      Background-corrected Intensity for channel A is denoted as *I(A)*;
      Background-corrected Intensity for channel B is denoted as *I(B)*;
      Background intensity for channel A is denoted as *Bkg(A)*;
      Background intensity for channel B is denoted as *Bkg(B)*;

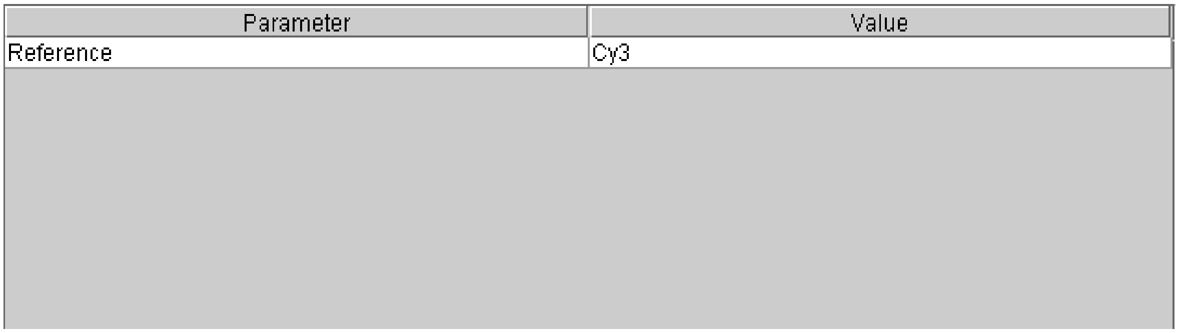
For each spot, if

$$\frac{I(A)+Bkg(A)}{Bkg(A)} \leq SN \text{ or } \frac{I(B)+Bkg(B)}{Bkg(B)} \leq SN$$

this spot will be marked as “bad”, its *I(A)* and *I(B)* will be reset to 0s, so that the spot will be excluded in any down stream analysis.

This parameter is set 2.0 by default, which means “Only if the background-corrected signal intensity is greater than or equal to the background intensity for both channel A and B of a spot, will this spot be considered as a ‘good spot’ and included in the downstream analysis”.

5.4 Total intensity normalization parameters [1]



Parameter	Value
Reference	Cy3

Total intensity normalization function normalizes channelA intensities *I(A)* or channelB intensities *I(B)* all of spots in the data file by scaling either channelA intensity or channelB intensity of each spot in the slide by a normalization factor *f*.

f is computed by the following formula:

$$f = \frac{\sum I(B)}{\sum I(A)}$$

for all those spots passed the upstream analysis steps in the defined process sequence

- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

  If Cy3 is selected as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be scaled by

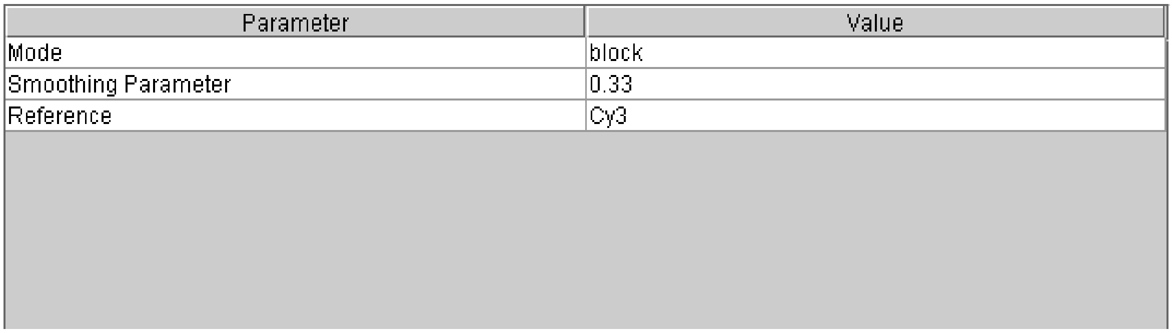
  $$New\ I(B) = Old\ I(B) / f$$

  If Cy5 is selected, as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be scaled by

$$New\ I(A) = Old\ I(A) * f$$

This parameter is set to Cy3 by default.

5.5 Locfit (LOWESS) normalization parameters [1], [2], [7]



Parameter	Value
Mode	block
Smoothing Parameter	0.33
Reference	Cy3

Locfit (LOWESS) normalization function normalizes channelA intensities *I(A)* or channelB intensities *I(B)* all of spots in the data file by applying LOWESS algorithm and adjusting either channelA intensity or channelB intensity of each spot by the computed LOWESS factors.

MIDAS provides two ways, or *modes*, to apply LOWESS algorithm on an input data file: either computing the LOWESS factor for each spot by assuming that all spots within a slide contribute to the bias of this spot's intensity (*Global mode*), or computing the LOWESS factor for each spot by assuming that only those spots within the same block as this spot contribute to the bias of this spot's intensity (*Block mode*). *Block* is defined by *meta-row* and *meta-column* data of each spot in the input data file. Spots within the same block should have the same meta-row and meta-column data. In TIGR-TAV format file, meta-row data is stored in column3; meta-column data is stored in column4. In TIGR-MEV format file, meta-row and meta-column are stored under the column with header MR and MC.

- **Mode**
  Either "Global" or "Block", depending on user's preference for applying LOWESS algorithm on the whole data set or each block data sets individually. This parameter is set to "Block" by default.

- **Smooth Parameter**
  Smooth parameter is a percentage number used by MIDAS to compute LOWESS factor for each spot. The higher the smooth parameter is set, the more severe the channelA or channelB intensity of raw input data will be adjusted. This parameter is set to 33% by default.

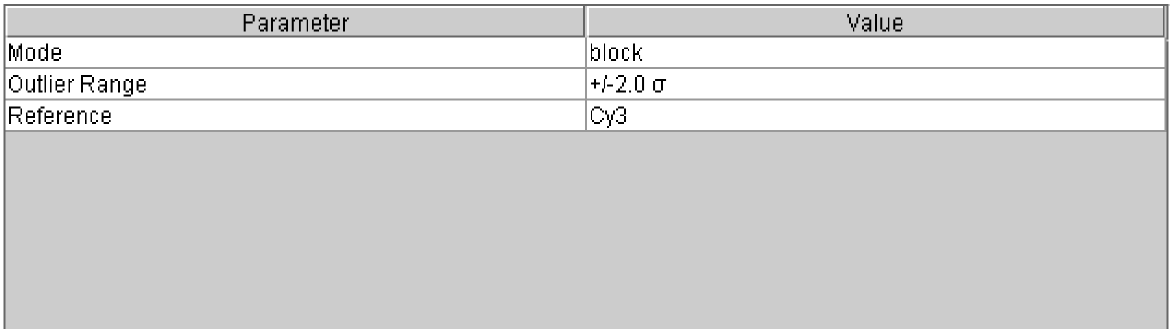
- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

If Cy3 is selected as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be adjusted by the calculated LOWESS factor for the spot.

If Cy5 is selected as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be adjusted by the calculated LOWESS factor for the spot..

This parameter is set to Cy3 by default.

5.6 Iterative linear regression normalization parameters [3]



Parameter	Value
Mode	block
Outlier Range	+/-2.0 σ
Reference	Cy3

Iterative linear regression normalization function normalizes channelA intensities *I(A)* or channelB intensities *I(B)* all of spots in the data file by applying the "Iterative Linear Regression" algorithm.

MIDAS provides two ways, or modes, to apply Iterative Linear Regression algorithm on an input data file: either computing the scaling factor for each spot by assuming that all spots within a slide contribute to the bias of this spot's intensity (*Global mode*), or computing the scaling factor for each spot by assuming that only those spots within the same block as this spot contribute to the bias of this spot's intensity (*Block mode*). *Block* is defined by *meta-row* and *meta-column* data of each spot in the input data file. Spots within the same block should have the same meta-row and meta-column data. In TIGR-TAV format file, meta-row data is stored in column3; meta-column data is stored in column4. In TIGR-MEV format file, meta-row and meta-column are stored under the column with header MR and MC.

- **Mode**
  Either "Global" or "Block", depending on user's preference for applying Iterative linear regression algorithm on the whole data set or each block data sets individually. It is set to "Block" by default.

- **Outlier Range**
  When Linear Regression is iteratively performed on $(\log_2 I(A), \log_2 I(B))$ data set within a slide (Global mode) or within each block of a slide (Block mode), outliers are

excluded from the data set from iteration to iteration to reduce the size of the data set. The iteration will stop when the correlation coefficients converge. Outlier Range specifies beyond how many folds of standard deviation of the data set a spot will be defined as outlier spot. It is set to "±2 SD" by default.

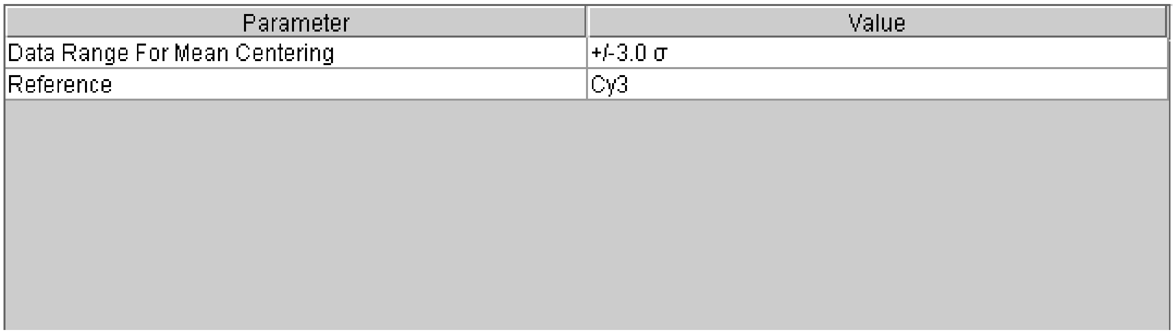
- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

  If Cy3 is selected, as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be adjusted by the calculated Iterative Linear Regression factor for the spot.

  If Cy5 is selected, as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be adjusted by the calculated Iterative Linear Regression factor for the spot.

  This parameter is set to Cy3 by default.

5.7 Iterative log mean centering normalization parameters [6]



Parameter	Value
Data Range For Mean Centering	+/-3.0 σ
Reference	Cy3

Iterative log-mean centering normalization function normalizes channelA intensities *I(A)* or channelB intensities *I(B)* all of spots in the data file by adjusting the mean of $\log_2 \frac{I(B)}{I(A)}$ to zero iteratively.

1st iteration: Calculate the mean of $\log_2 \frac{I(B)}{I(A)}$ from all spots within the slide

Adjust each spot's I(A) or I(B) within the slide so that for all spots within the slide, mean of $\log_2 \frac{I(B)}{I(A)} = 0$;

2nd iteration: Calculate the mean of $\log_2 \frac{I(B)}{I(A)}$ from those spots within a user-specified “Data Range For Mean Centering”;

Adjust each spot’s I(A) or I(B) within the slide so that for spots Within the “Data Range For Mean Centering”, mean of $\log_2 \frac{I(B)}{I(A)} = 0$;

Iteration continues until the mean of $\log_2 \frac{I(B)}{I(A)}$ stabilize.

- **Data Range For Mean Centering**
  After an iteration of Log-Mean Centering is performed, some spots within the slide are selected to calculate the mean of $\log_2 \frac{I(B)}{I(A)}$. The “Data Range For Mean Centering” defines which spots should be selected.

  This parameter is set to “±3 SD” by default.

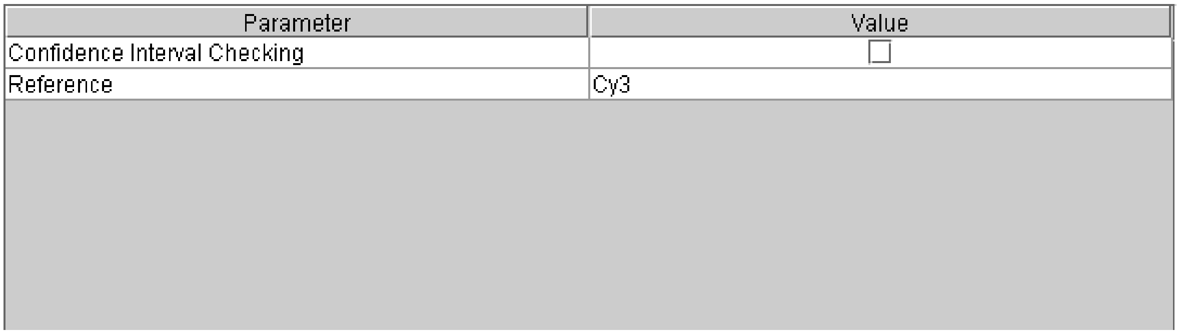
- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

  If Cy3 is selected as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be adjusted by the calculated Iterative Log-Mean Centering factor for the spot.

  If Cy5 is selected as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be adjusted by the calculated Iterative Log-Mean Centering factor for the spot.

  This parameter is set to Cy3 by default.

5.8 Ratio statistics normalization and CI checking parameters [4]

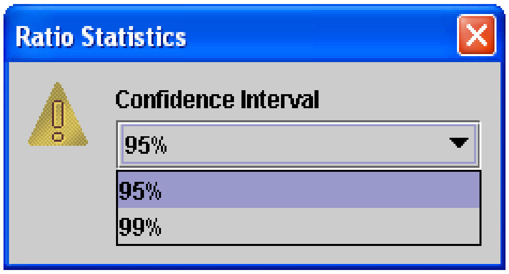


Parameter	Value
Confidence Interval Checking	□
Reference	Cy3

Ratios statistics normalization function normalizes channelA intensities *I(A)* or channelB intensities *I(B)* all of spots in the data file by adjusting the $\frac{I(B)}{I(A)}$ ratios, so that the $\frac{I(B)}{I(A)}$ ratio distribution follow the ratio statistics studied in the reference [4]. Please refer to the referenced paper for detailed explanation of the algorithm.

This normalization method also serves as a filtering method; user can also use it to perform confidence interval checking for spot quality control purpose if desired.

- **Confidence Interval**
  Checking the corresponding value cell indicates a confidence interval checking is applied. When checked, the pop-up window will be displayed for Confidence Interval range selection. The range can be either 95% or 99%.



If confidence interval checking is performed, those spots with $\frac{I(B)}{I(A)}$ ratio falling beyond the confidence interval range will be marked as “bad” in MIDAS memory and will be excluded in the downstream analysis.

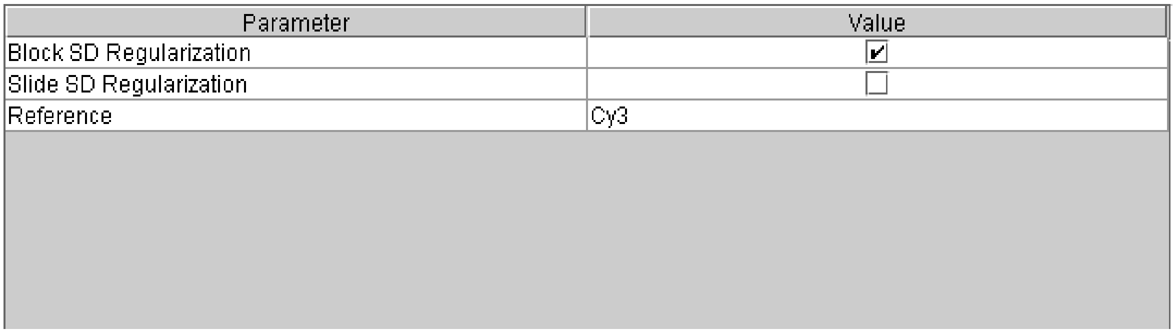
- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

  If Cy3 is selected as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be scaled by the calculated $\frac{I(B)}{I(A)}$ ratio for the spot.

  If Cy5 is selected as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be scaled by the calculated $\frac{I(B)}{I(A)}$ ratio for the spot.

  This parameter is set to Cy3 by default.

5.9 Standard deviation regularization parameters [1], [5]



Parameter	Value
Block SD Regularization	☑
Slide SD Regularization	□
Reference	Cy3

Based on the assumption that all spots within each block of a slide or all spots within each slide of multiple slides (when inputs are multiple data files) should have the same spread for $\log_2 \frac{I(B)}{I(A)}$, Standard deviation regularization function scales the channelA or channelB intensities of each spot so that the spots within each block of a slide or spots within each slide of multiple slides will have the same standard deviation for $\log_2 \frac{I(B)}{I(A)}$ distribution.

When input is single data file, standard deviation regularization can be applied only on Block basis. When inputs are multiple data files, both Block and Slide modes are available.

When Block SD Regularization is applied, *Block* is defined by *meta-row* and *meta-column* data of each spot in the input data file. Spots within the same block should have the same meta-row and meta-column data. In TIGR-TAV format file, meta-row data is stored in column3; meta-column data is stored in column4. In TIGR-MEV format file, meta-row and meta-column are stored under the column with header MR and MC.

- **Block SD Regularization**
  Checking this checkbox indicates Block standard deviation regularization is applied. This parameter is checked by default.

- **Slide SD Regularization**
  Checking this checkbox indicates Slide standard deviation regularization is desired. This parameter can only be checked when inputs are multiple data files.

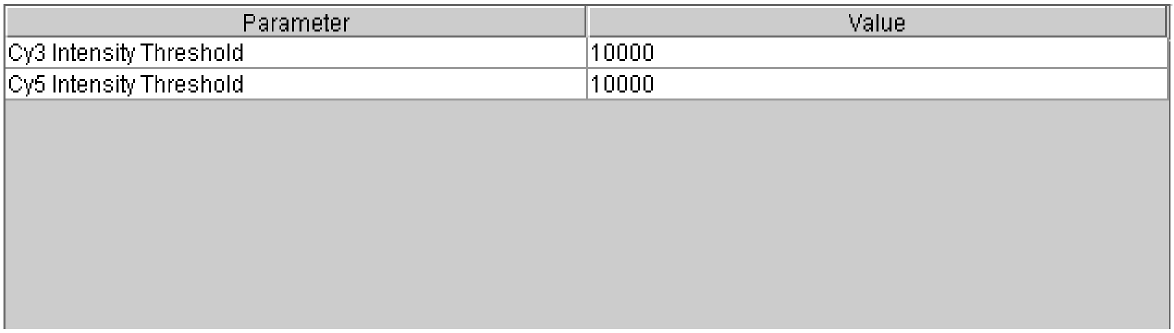
- **Reference**
  Either Cy3 or Cy5 can be selected as a reference by clicking the corresponding value cell.

  If Cy3 is selected as the reference, then *I(A)* of each spot will not be changed in the output file; however, *I(B)* of each spot will be adjusted by the calculated SD regularization factor for the spot.

If Cy5 is selected, as the reference, then *I(B)* of each spot will not be changed in the output file; however, the *I(A)* of each spot will be adjusted by the calculated SD regularization factor for the spot.

This parameter is set to Cy3 by default.

5.10 Low intensity filter parameters



Parameter	Value
Cy3 Intensity Threshold	10000
Cy5 Intensity Threshold	10000

Low intensity filter function allows filtering low intensity spots in a data file. The spots with channelA *I(A)* or channelB *I(B)* intensities lower than user-specified thresholds will be marked as "bad" and will be excluded in the downstream analysis.

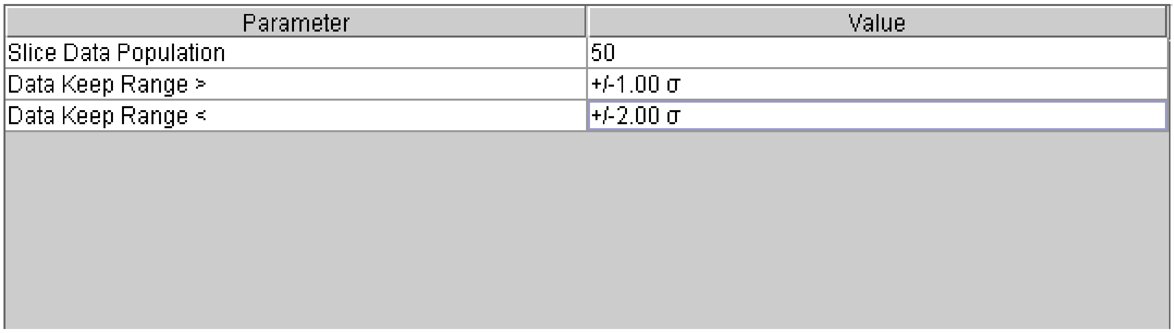
- **Cy3 Intensity Threshold**
  ChannelA intensity threshold can be set by clicking the corresponding value cell. MIDAS will compare the channelA intensity of each spot with the threshold. This parameter is set to 10000 by default.

  If *I(A)* ≤ Cy3 intensity threshold, *I(A)* and *I(B)* will be reset to 0s, so that the spot will not be included in the downstream analysis.

- **Cy5 Intensity Threshold**
  ChannelB intensity threshold can be set by clicking the corresponding value cell. MIDAS will compare the channelB intensity of each spot with the threshold. This parameter is set to 10000 by default.

  If *I(B)* ≤ Cy5 intensity threshold, *I(A)* and *I(B)* will be reset to 0s, so that the spot will not be included in the downstream analysis.

5.11 Slice analysis parameters [1], [7]



Parameter	Value
Slice Data Population	50
Data Keep Range >	+/-1.00 σ
Data Keep Range <	+/-2.00 σ

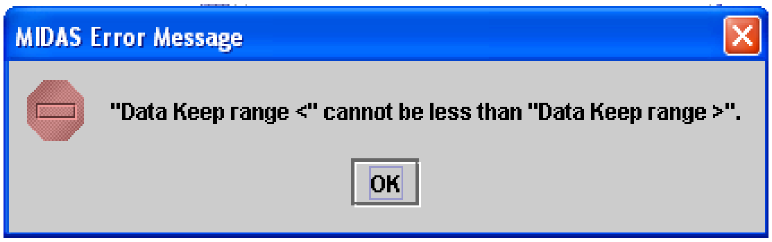
Slice analysis function allows user to classify and filter spots based on their expression levels.
Slice analysis converts spots' channelA intensities *I(A)* and channelB intensities *I(B)* to $\{\log_{10}(I(A)\cdot I(B))$ vs. $\log_2\frac{I(A)}{I(B)}\}$ data sets. It generates a series of dynamic "slide windows" along the $\log_{10}(I(A)\cdot I(B))$ axis. Each window contains a user-specified number (*Slice Data Population*) of data points. For each window, the mean and standard deviation of the $\log_2\frac{I(A)}{I(B)}$ values for the contained data points are calculated. Slice analysis then performs filtering based on the *Data Keep Range >* and *Data Keep Range <* parameters that user specified. The data points that fall beyond the range are marked as "bad" and excluded in the downstream analysis. Their channelA and channelB intensities are set to 0s.

- **Slice Data Population**
  The data point number within each dynamic "slide window" along the $\log_{10}(I(A)\cdot I(B))$ axis can be specified by clicking the corresponding value cell. This parameter is set to 50 by default.

- **Data Keep Range >**
  The lower limit for a desired data keep range can be set by clicking the corresponding value cell. It is in the form of folds of standard deviation and can be selected from the drop-down list. If "------" is selected, it means 0SD is selected.

- **Data Keep Range <**
  The upper limit for a desired data keep range can be set by clicking the corresponding value cell. It is in the form of folds of standard deviation and can be selected from the drop-down list. If "------" is selected, it means open upper limit.

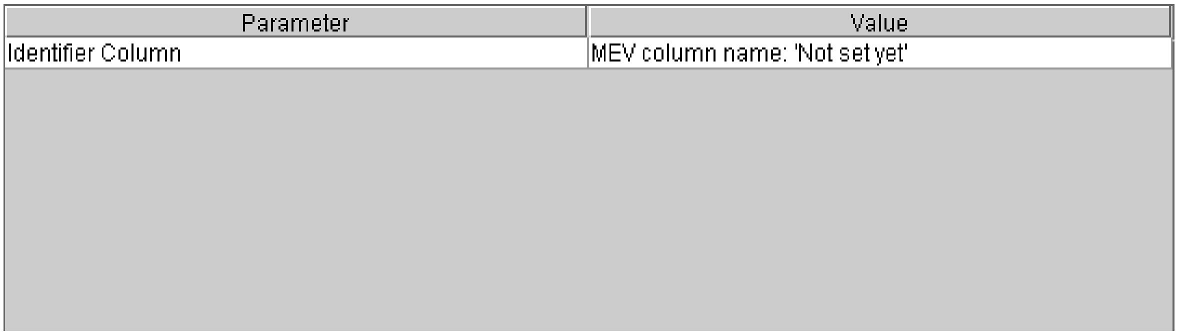
  Note: If "Data Keep Rang <" (upper limit) is less than "Data Keep Range >" (lower limit), a MIDAS error message will be displayed as below.



MIDAS Error Message

"Data Keep range <" cannot be less than "Data Keep range >".

5.12 In-slide replicates analysis parameters



Parameter	Value
Identifier Column	MEV column name: 'Not set yet'

Note: In-slide replicates analysis is not available when inputs are file pair(s).

In-slide replicates analysis is a replica averaging function that works on data file(s) with replicated spots. It groups and merges the replicated spots by presenting the channelA intensities and channelB intensities as the geometric means of the replica.

Let $I(A)_i$ denotes the channelA intensity of the ith replicated spots in a replica group with n spots;

$I(B)_i$ denotes the channeB intensity of the ith replicated spots in a replica group with n spots;

$I(A)'$ denotes the averaged channelA intensity of replica group with n spots;

$I(B)'$ denotes the averaged channelB intensity of replica group with n spots;

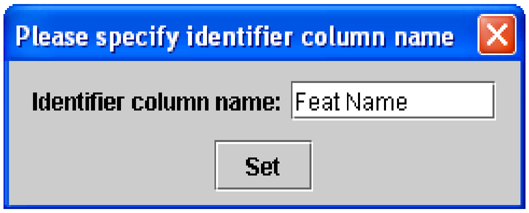
The following equations are used to average the replicated spots' intensities:

$$I(A)' = \sqrt[n]{\prod_{i=1}^{n} I(A)_i} \qquad I(B)' = \sqrt[n]{\prod_{i=1}^{n} I(B)_i}$$

- **Identifier Column**
  In order to perform In-slide replicates analysis, an identifier for each spot is required to instruct MIDAS which spots belong to the same replica group. Thus an identifier

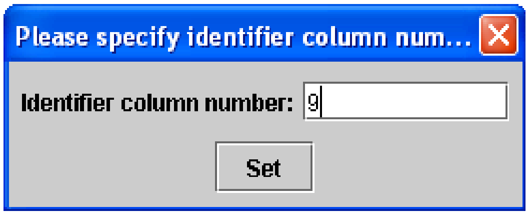
column that registers an identifier value for each spot in the data file is required for this operation. The replicated spots should have the same identifier value. The identifier column is specified by user.

If the input file is in TIGR-MEV format, the identifier column should be specified by keying in the column header name (Appendix).



Identifier column name: Feat Name

If the input file is in TIGR-TAV format, the identifier column should be specified by keying in the column number (Appendix).

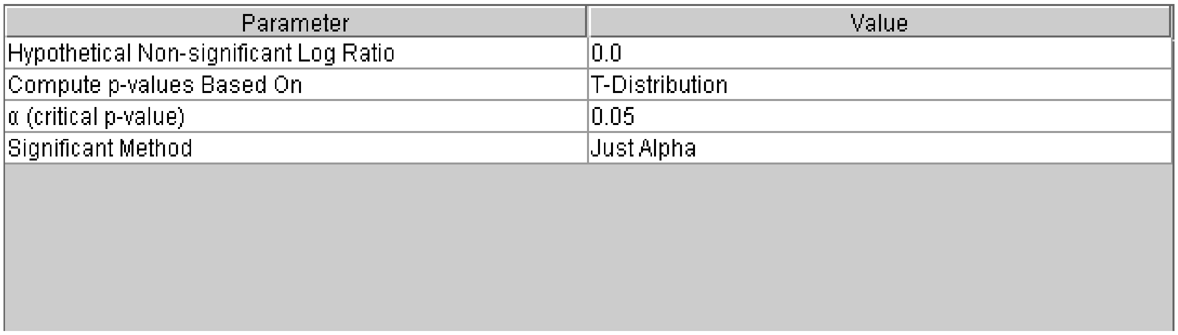


Identifier column number: 9

5.13 Cross-slide replicates t-test parameters [8]

One-class t-test can be applied to a group of data files representing replicated experiments to identify significant genes. In the one-class t-test design, user specifies a hypothetical log-ratio value as the non-significant criteria. Those genes whose mean log-ratios over all included experiments are significantly different from this hypothetical non-significant log-ratio are selected as significant genes. The "significantly" in the above sentence is statistically evaluated by comparing the calculated p-values of each gene with a user specified critical p-value.

Note: Cross-slide replicates t-test module is only available when the input is a data file directory, which contains a group of data files representing replicated experiments.



Parameter	Value
Hypothetical Non-significant Log Ratio	0.0
Compute p-values Based On	T-Distribution
α (critical p-value)	0.05
Significant Method	Just Alpha

- **Hypo Non-significant Log Ratio**
  Hypothetical non-significant log-ratio is the criteria that user sets to determine if a gene should be believed to be expressed significantly. This parameter is set to 0 by default.

- **Compute p-value Based On**
  This parameter is used to indicate the method by which p-values are determined for each gene.

  T-distribution

  Using this option a gene's p-value is taken directly from the theoretical t-distribution based on the gene's calculated t-value.

  All possible permutation

  Using this option a gene's p-value is determined by forming a distribution based on permutations of the data for that gene. Permutation means randomly picking some elements of the gene expression vector and changing their values, such that the new value of the changed element is:

  New value = original value – 2 ×(original value – hypothetical non-significant log ratio)

  All possible permutation allows every possible combination of elements in the gene expression vector be sign-flipped.

  Specify permutation number

  Using this option a gene's p-value is determined by forming a distribution based on a specified number of permutations of the data for that gene. Permutation means randomly picking some elements of the gene expression vector and changing their values, such that the new value of the changed element is:

  New value = original value – 2 ×(original value – hypothetical non-significant log ratio)

- **α (critical p-value)**
  This parameter allows a user to enter the alpha or critical p-value. The calculated p-value of each gene is compared with this critical p-value for significant gene identification.

- **Significant Method**
  The p-values for each gene can be adjusted to correct for the large number of observations (genes) and the increased possibility of considering a gene without a real significant change to be considered significant. Three options are given for this adjustment.

  Just Alpha

  By using this option, the alpha is not altered.

Standard Bonferroni

In the standard Bonferroni correction, the user-specified alpha is divided by the number of genes to give the critical p-value. This is much more stringent than using an uncorrected alpha.

Adjusted Bonferroni

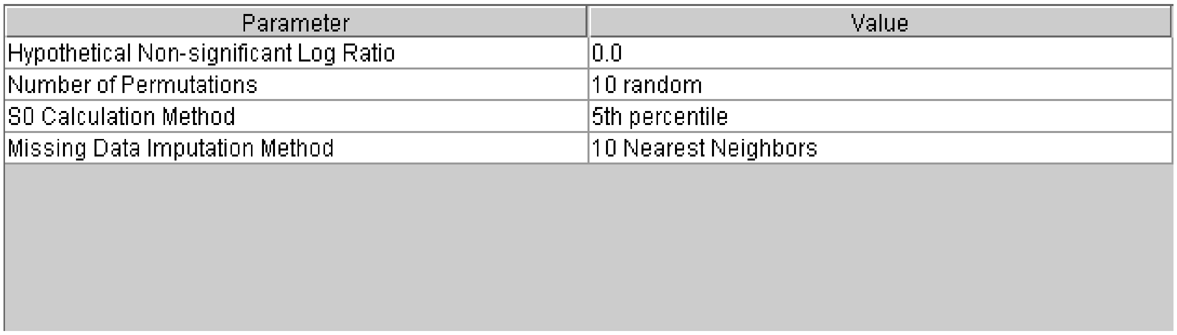
In the adjusted Bonferroni correction, the t-values for all the genes are ranked in descending order. For the gene with the highest t-value, the critical p-value becomes α/n, where n is the total number of genes; for the gene with the second-highest t-value, the critical p-value becomes α/(n-1), and so on. The stringency of this correction falls somewhere between no correction and the standard Bonferroni.

5.14 Cross-slide replicates SAM parameters [9]

SAM or Significance Analysis of Microarrays is a significant genes identification method. A valuable feature of SAM is that it gives estimates of the False Discovery Rate (FDR), which is the proportion of genes likely to have been identified by chance as being significant. Furthermore, SAM is an interactive algorithm. It allows users to eyeball the distribution of the test statistic, and then set thresholds for significance (through the tuning parameter Δ) after looking at the distribution.

In this release of the software, only one-class SAM is implemented, which applies to a group of data files representing replicated experiments. In one-class SAM, user specifies a non-significant log-ratio value against which the mean expression of each gene across the replica is tested. A gene is considered significant if its mean log-ratio over all replicates is significantly different from the user-specified non-significant log-ratio. The "significantly" in the above sentence is statistically evaluated in three steps:

1. Compare the observed d-value of each gene with the expected d-value of it calculated from a permutation analysis.
2. Estimate the falsely discovered gene number or False Discovery Rate (FDR) of each gene.
3. User should use the interactive SAM plot to adjust the Δ value and reach a set of differentially expressed genes with satisfied low falsely discovered gene number of FDR.

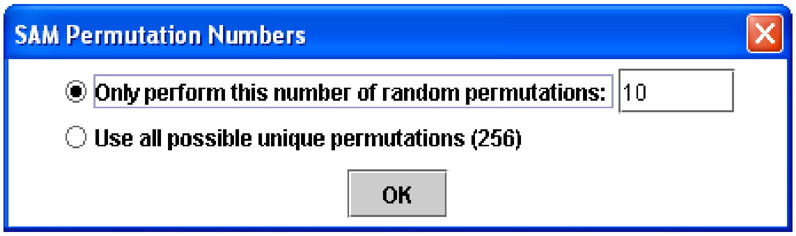


Parameter	Value
Hypothetical Non-significant Log Ratio	0.0
Number of Permutations	10 random
S0 Calculation Method	5th percentile
Missing Data Imputation Method	10 Nearest Neighbors

- **Hypo Non-significant Log Ratio**
  Hypothetical non-significant log-ratio is the criteria that user sets to determine if a gene should be believed to be expressed significantly. This parameter is set to 0 by default.

- **Number of Permutations**
  This parameter is used to indicate the permutation numbers that MIDAS will use to evaluate the expected d-values. There are two options:

  Only perform this number of random permutations
  The specified number of random permutations will be performed.
  Use all possible unique permutations (N)
  MIDAS will perform all possible N permutations, $N = 2^{\text{number of selected files}}$



SAM Permutation Numbers

● Only perform this number of random permutations: 10

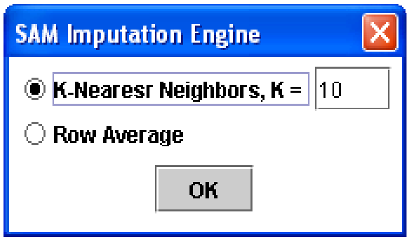
○ Use all possible unique permutations (256)

- **S0 Calculation Method**
  This parameter allows user to specify different methods for the S_0 calculation. There are five options:

  Tusher's method
  5th percentile
  50th percentile
  90th percentile
  Use minimum S value

- **Missing Data Imputation Method**
  This parameter allows user to specify different methods for missing data imputation. There are two options:

  K Nearest Neighbors
  MIDAS will use the specified K most similar genes (using Euclidean distance) to the gene with a missing value to impute the missing value.
  Row Average
  MIDAS will replace the missing expression measurements with the mean expression of a row (gene) across all columns (experiments).

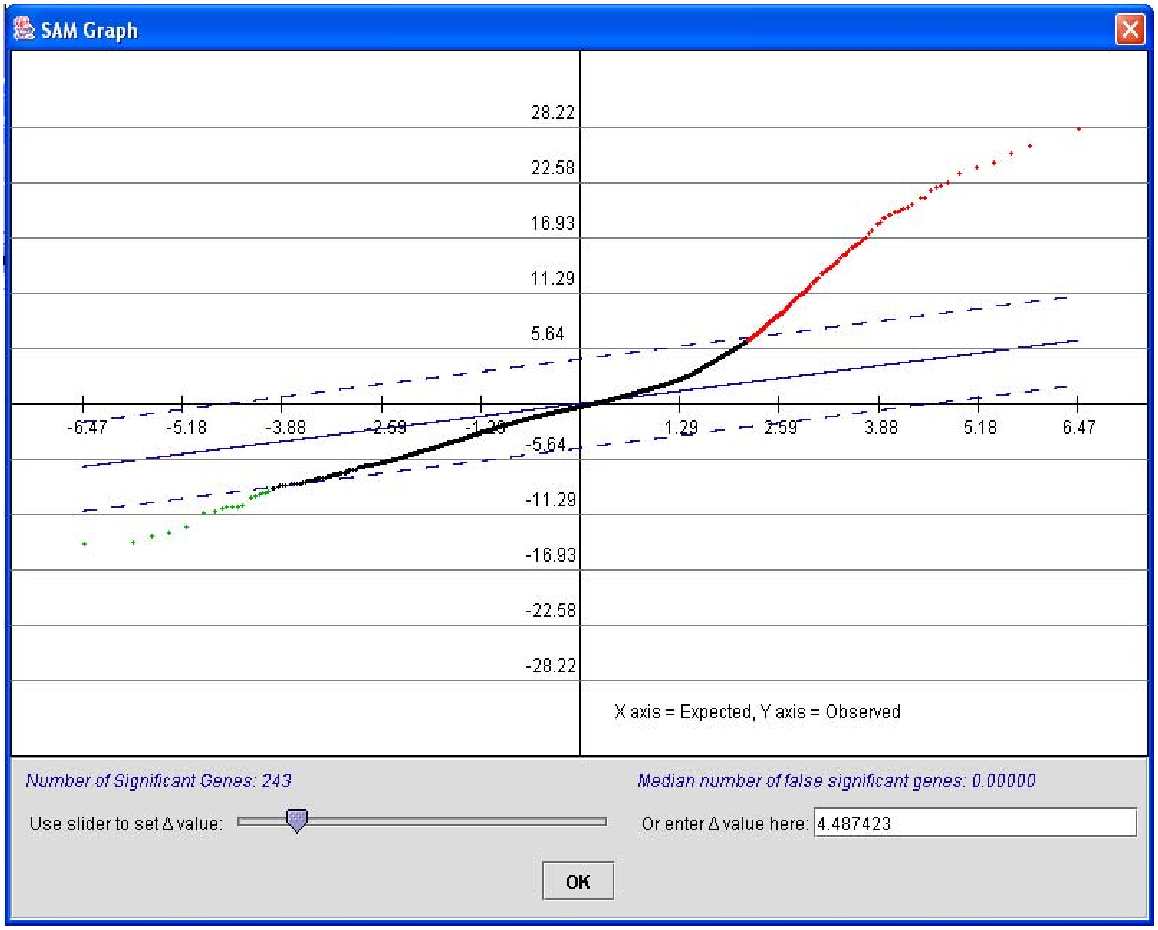


SAM Imputation Engine

● K-Nearesr Neighbors, K = 10

○ Row Average

SAM is an interactive analysis procedure that requires user's inputs through SAM plot to identify the significant genes.



Number of Significant Genes: 243

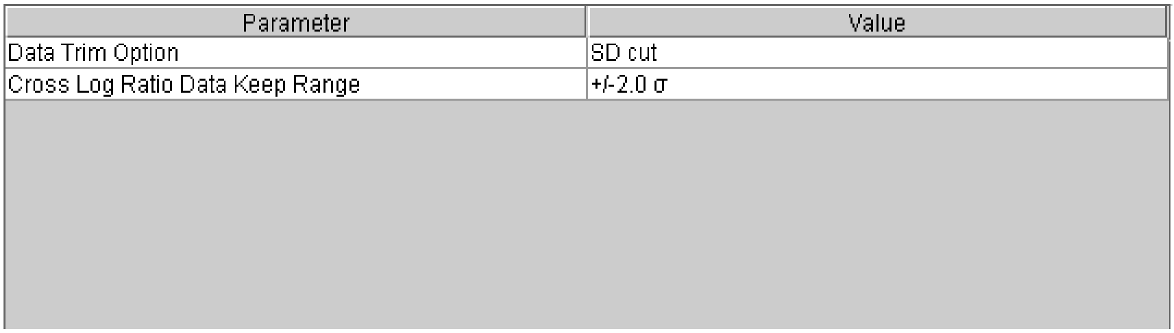
Median number of false significant genes: 0.00000

Use slider to set Δ value:

Or enter Δ value here: 4.487423

Δ value can be adjusted through moving the slider or directly entered at the bottom of the plot. The Number of significant genes and the Median number of false significant genes are calculated and shown as the Δ value changes. The blue dash lines have slope of 1, which define the cases when genes' expected *d* values agrees with the observed *d* values. The genes beyond the blue dash line are significant genes candidates, with positive significant genes marked red and negative significant genes marked green. The goal is picking a set of differentially expressed genes with a satisfied false significant genes number.

5.15 Flip dye consistency checking parameters [1], [5]



Parameter	Value
Data Trim Option	SD cut
Cross Log Ratio Data Keep Range	+/-2.0 σ

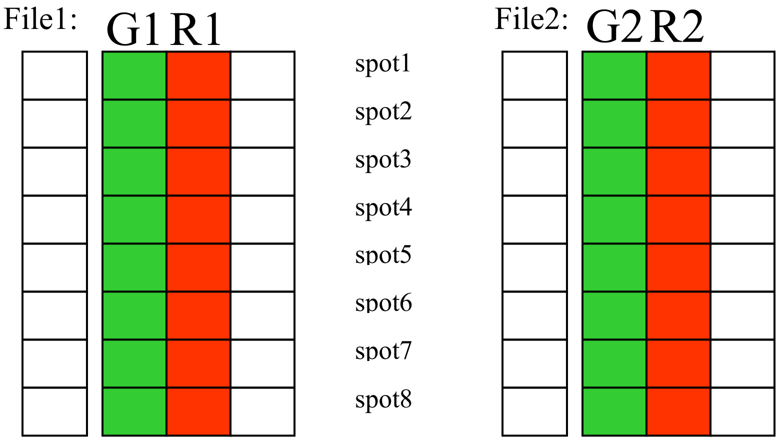
Flip dye consistency checking function allows filtering out spots that show expression level inconsistency between a pair of flip-dye replicates. The following steps are performed when this function is applied.

(a) Calculate the cross-log-ratios of the data file pair for each spot.

$$A \equiv \log_2 \frac{R1 * R2}{G1 * G2}$$

R1 is the channelB intensity of the spot in file1;
R2 is the channelB intensity value in file2;
G1 is the channelA intensity value in file1;
G2 is the channelA intensity value in file2.

If any of R1, R2, G1 and G2 is 0, set A = 0.



(b) Calculate the mean $\overline{A}$ and standard deviation $\sigma(A)$ of all none-zero A

(c) For each spot:

If user uses *SD cut* for Data Trim Option, compare *A* with *σ(A)*N*, where *N* is the number of fold of standard deviation that user specified for *Cross Log Data Keep Range* parameter. The spots with theirs *A* values having

$$A \le \overline{A} - \sigma(A) * N \text{ Or } A \ge \overline{A} + \sigma(A) * N$$

are marked as "flip-dye inconsistent" and will be excluded from downstream analysis. Their channel A and channelB intensities in both files will set to 0s.

If user uses *threshold cut* for "Data Trim Option", compare *A* with *K*, where *K* is the value that user specified for *Cross Log Data Keep Range* parameter. The spots with their A values having

$$A \le \overline{A} - K \text{ Or } A \ge \overline{A} + K$$

are marked as "flip-dye inconsistent" and will be excluded from downstream analysis. Their channel A and channelB intensities in both files will set to 0s.

For those spots passed the flip-dye consistency checking, the channelA and channelB intensities from both files will be merged and presented in a merged output file. The merged intensities will be the geometric means of the flip-dye replicates.

Let *I(A)'* denotes the averaged channelA intensity of a flip-dye spot pair;

I(B)' denotes the averaged channelB intensity of a flip-dye spot pair;

The following equations are used to average the replicated spots' intensities:

$$I(A)' = \sqrt{G1 \cdot R2}\,, \qquad I(B)' = \sqrt{R1 \cdot G2}$$

For those spots failed the flip-dye consistency checking, the channelA and channelB intensities of the spots will be set to 0s. They will be marked as "bad" and excluded from downstream analysis.

In the merged file, spot information other than channelA and channelB intensities will be copied from the first file in the pair.

(d) Two factors describing the overall consistency level between the flip-dye pair are calculated as written to report file(s). The two factors are *Confidence Factor* and *Dispersion Factor*.

$$ConfidenceFactor = \frac{NumberOfSpotsThatSatisfyTheConsistencyCriteria}{TotalNonZeroIntensitySpotsNumber}$$

$$DispersionFactor = \frac{NumberOfSpotsThatHavingInconsistentRatioBeyond2Folds}{TotalNonZeroIntensitySpotsNumber}$$

If the *A* values as defined in (a) follow normal distribution closely, the Confidence Factor for SD cut range of ±2SD should be very close to 95%.

Dispersion Factor describes the percentage of spots which are having their *A* values NOT between -1 and 1. Different from Confidence Factor, when calculating Dispersion Factor, the distribution of *A* values is not taken into consideration. From the definition of *A*, all the spots that contribute to the Dispersion Factor should satisfy:

$$\log_2 \frac{R1*R2}{G1*G2} \le -1 \qquad \log_2 \frac{R1*R2}{G1*G2} \ge 1$$

$$\log_2 \left[\frac{\left(\frac{R1}{G1}\right)}{\left(\frac{G2}{R2}\right)} \right] \le -1 \quad \text{Or} \quad \log_2 \left[\frac{\left(\frac{R1}{G1}\right)}{\left(\frac{G2}{R2}\right)} \right] \ge 1$$

$$\frac{\left(\frac{R1}{G1}\right)}{\left(\frac{G2}{R2}\right)} \le \frac{1}{2} \qquad \frac{\left(\frac{R1}{G1}\right)}{\left(\frac{G2}{R2}\right)} \ge 2$$

so Dispersion Factor measures the percentage of spots which are having more than 2-folds ratio inconsistencies between the replicates pair.

- **Data Trim Option**
  There are two filtering options that can be used to filter out inconsistent expression spots between a flip-dye pair: *SD cut* or *threshold cut*. This parameter can be selected from the drop-down list.

  SD cut filters spots based on the *A* values statistical distribution. The spots with their *A* value beyond a user-specified number of folds of standard deviation will be marked as "bad" and excluded from downstream analysis.

  Threshold cut filters spots by comparing the *A* value with a user-specified threshold value. The spots with their *A* value beyond the threshold will be marked as "bad" and excluded from downstream analysis, their channelA and channelB intensities will be set to 0s in the merged outputs.

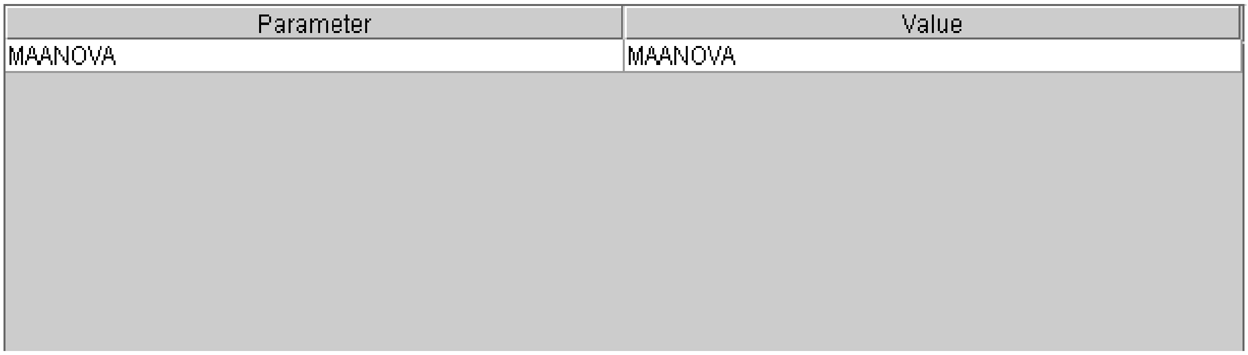
  This parameter is set to *SD cut* by default and the corresponding default Cross Log Ratio Data Keep Range is ±2.0SD.

- **Cross Log Ratio Data Keep Range**
  This parameter sets the criteria of inconsistency for filtering.

When *SD cut* is selected as the Data Trim Option, it sets the threshold number of folds for *A* values standard deviation.

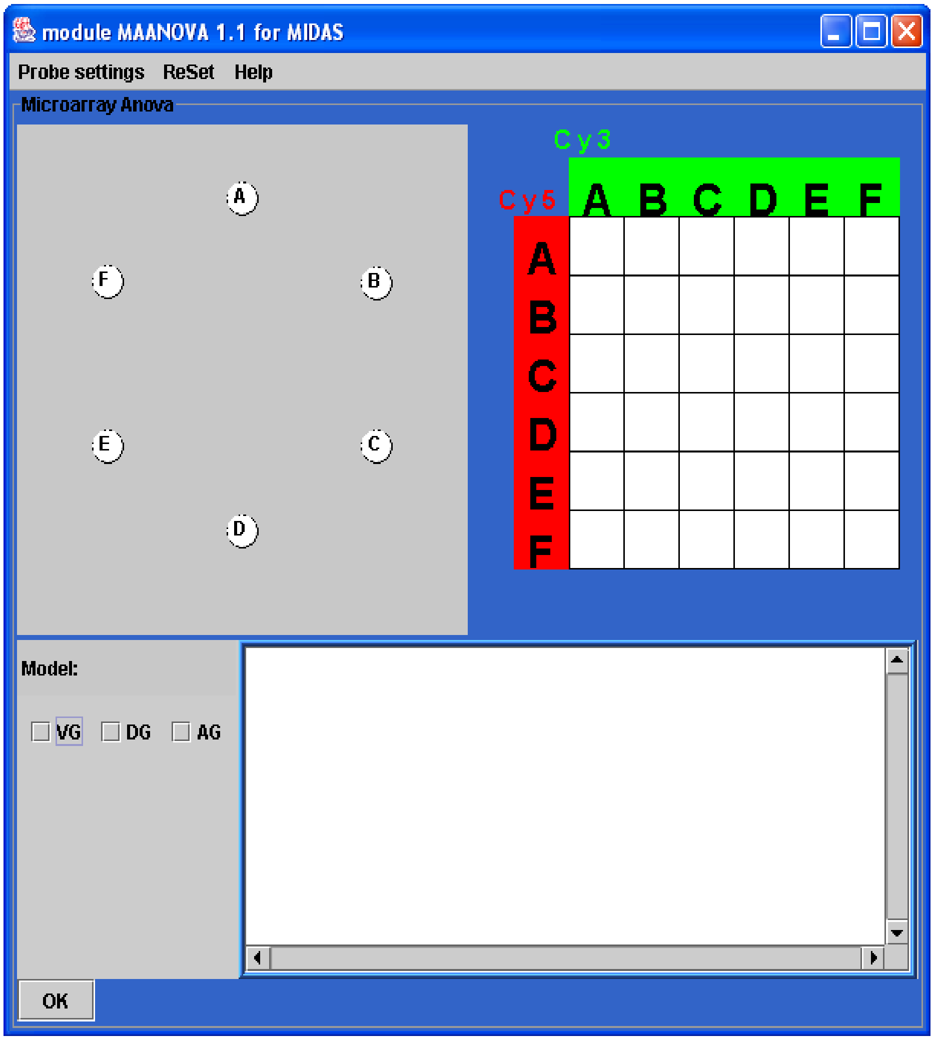
When *threshold cut* is selected as the Data Trim Option, it sets the threshold number for *A* values.

5.16 MAANOVA parameters [10]



Parameter	Value
MAANOVA	MAANOVA

- **MAANOVA**
  Clicking the MAANIVA value cell launches the MAANOVA module. A "module MAANOVA 1.1 for MIDAS" is popped up for parameters setting.



MAANOVA is a JAVA implementation of Jackson Lab's MAANOVA MATLAB tool.

MAANOVA takes two or more data files and process them in accordance with the model and experimental design specified by a user.

Three experimental designs are available: flip-dye design, reference design and loop design.

Three models are available: VG for variety-gene effects, DG for dye-gene effects and AG for array-gene effects. Depending on the analysis purpose, different combinations of these models may be applied. Please refer to Reference [8] for details about the MAANOVA algorithm.

The MAANOVA dialog consists of three major controls: interactive *loop view* with six default nodes from A to F; interactive *experimental design matrix* with default size of 6 by 6 (from A to F); and non-interactive *list view* for displaying selected files and nodes. In addition, three checkbox model selection controls allow a user to select appropriate model: AG, DG and VG. Either clicking a node in the loop view or a cell in the experimental design matrix will pop up a file browser for selecting an input file.

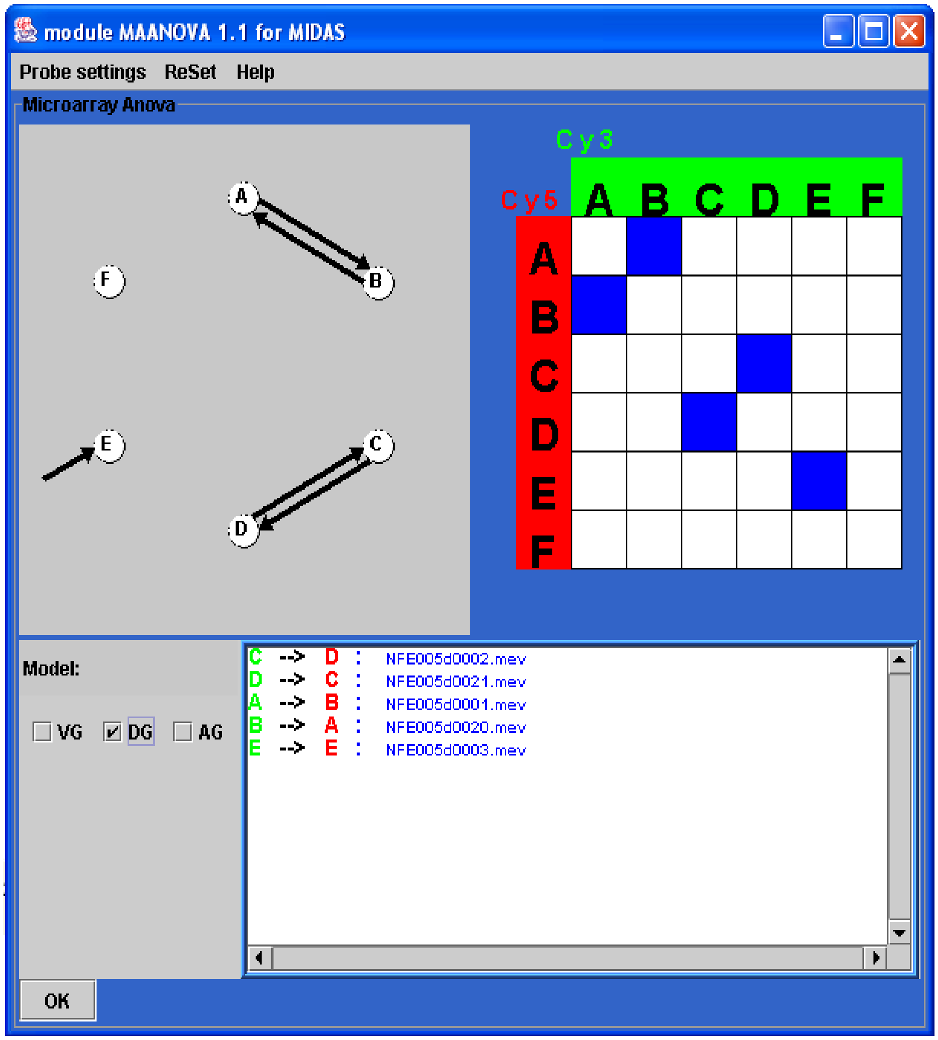
When working in the experimental design matrix, each cell represents a designed relation between the 2 probes (Cy3 and Cy5) in an experiment or a data file. The column-coordinate of the cell denotes the first probe; the row-coordinate of the cell denotes the second probe. Clicking the cell launches a file browser for file selection. The designed relation is also displayed in the loop view.

When working in the loop view, each pair of nodes connected by a headed arrow represents a designed relation between 2 probes (Cy3 and Cy5) in an experiment or a data file. To construct a designed relation, click a node representing the first probe, the node's color will turn yellow; and then click the second node representing the second probe, a file browser will pop up for file selection. After a file is selected an arrow will be displayed pointing from the first node to the second node. The designed relation is also displayed in the experimental design matrix.

Note: Be careful to start selecting probe nodes from A, then B, etc. Probe nodes have to be selected in consecutive alphabetical order.

Besides the experimental design matrix and the loop view, the designed relations are also shown in the list view.

The screenshot below shows 5 designed relations or experiments. Among them, A->B and B->A form a flip dye experiment pair; C->D and D->C for a flip dye experiment pair; E is a self-referenced experiment.



Cy5 \ Cy3	A	B	C	D	E	F
A		■				
B	■					
C				■		
D			■			
E					■	
F						

Model:

☐ VG ☑ DG ☐ AG

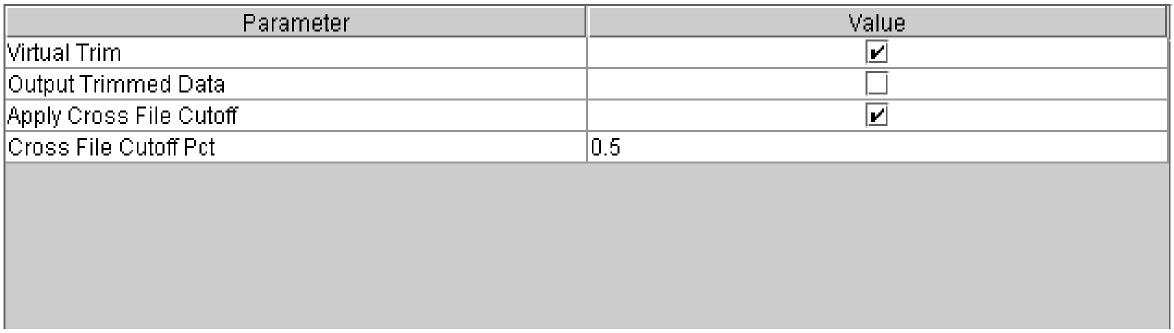
```
C  -->  D  :  NFE005d0002.mev
D  -->  C  :  NFE005d0021.mev
A  -->  B  :  NFE005d0001.mev
B  -->  A  :  NFE005d0020.mev
E  -->  E  :  NFE005d0003.mev
```

A designed relation or an experiment can be removed by mouse clicking on the corresponding experimental design matrix cell or end node in the loop view. The number of nodes and dimension of the experimental design matrix can be adjusted by selecting “Probe Settings” in the menu bar. “ReSet” in the menu bar allows the shapes and sizes of the arrows be adjusted.

Note: Always select ‘DG’ model for a flip-dye experiment design. AG may also be applied to flip-dye experiment design.

After the MAANOVA parameter settings, click OK button in the bottom-left corner or the ‘X’ button in the upper-right corner to register the settings into MIDAS and dismiss the MAANOVA parameter setting window.

5.17 Write parameters



Parameter	Value
Virtual Trim	☑
Output Trimmed Data	□
Apply Cross File Cutoff	☑
Cross File Cutoff Pct	0.5

- **Virtual Trim**
  When generating the output file(s), those spots that are marked as “bad” from the upstream analysis can be chosen either not to write to the output file(s) or still write to the outputs, but present their channelA and channelB intensities as 0s to distinguish them from other “good” spots in the output file(s).

  Check this checkbox when “write to the outputs, but present their channelA and channelB intensities as 0s” is desired. To have a Virtual Trim outputs is useful when keeping the original spot numbers in the inputs is a concern.

  This parameter is set checked by default.

- **Output Trimmed Data**
  If user prefers having a record of those spots that are filtered from the input file(s), this checkbox should be checked on. MIDAS will output “Trimmed Data” file(s) in the MIDAS project folder.

  Assuming the input data file is called “sample.xxx”, where ‘xxx’ can be either ‘tav’ or ‘mev’, the “Trimmed Data” file will be named “sample_MDSTrim.xxx”.

- **Apply Cross File Cutoff**
  This parameter is only available when inputs are multiple data files.

  Cross File Cutoff is a function allowing filtering spots that show inconsistencies among multiple data files within a directory.

  The *Cross File Trim* Algorithm can applied under the following situations

  i. Inputs are a directory containing multiple data files.
  ii. There are N ($N \geq 1$) numbers of data files in the directory.
  iii. There are M ($M \geq 1$) numbers of spots within each data file.
  iv. User is interested in investigating if the i^{th} ($i=1...M$) spot in any of the N files show consistencies comparing with the i^{th} spot in the other $N-1$ files.

  MIDAS checks the i^{th} across N files and counts how many of them are marked as “good” in the upstream analysis. A Cross File Trim percentage *CFT%* is calculated as

$$CFT\% = \frac{"good"\, SpotsNumberAcrossFiles}{N}$$

The *CFT%* of each spot is then compared with user-specified percentage threshold, if *CFT%* ≥ threshold, the spot's data will be written out in all *N* output files; otherwise the spot will be either not written out or Virtual Trimmed in the *N* output files.

This parameter is set checked by default.

- **Cross File Cutoff Percentage**
  This parameter is set by the user to specify the *CFT%* threshold used for *Cross File Trim.*

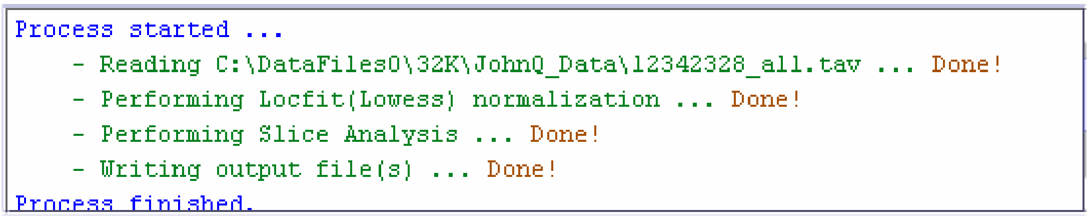
  This parameter is set to 0.5 or 50% by default.

  Note: Only when "Apply Cross File Cutoff" checked, the parameter will be used.

6 Process status window

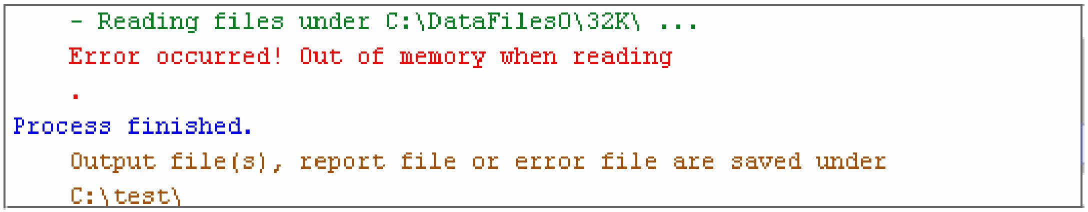
Once the analysis process sequence and parameters for each step are defined in the Work Flow window and Parameters window, pressing the “Execution” button will start the execution of the process. The running status will be display in the Process Status window. The Process Status window includes a Text status window and a Progress animation indicator.

- **Text status window**



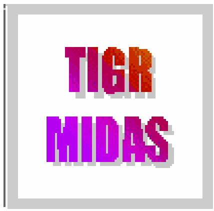
```
Process started ...
    - Reading C:\DataFiles0\32K\JohnQ_Data\12342328_all.tav ... Done!
    - Performing Locfit(Lowess) normalization ... Done!
    - Performing Slice Analysis ... Done!
    - Writing output file(s) ... Done!
Process finished.
```

This window tells user the process status in color-coded text; it also displays the location where user can find the final results and error message if errors occurred during the execution.

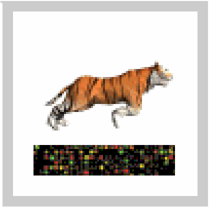


```
    - Reading files under C:\DataFiles0\32K\ ...
    Error occurred! Out of memory when reading
    .
Process finished.
    Output file(s), report file or error file are saved under
    C:\test\
```

- **Progress animation indicator**



indicates MIDAS is in stand-by status.



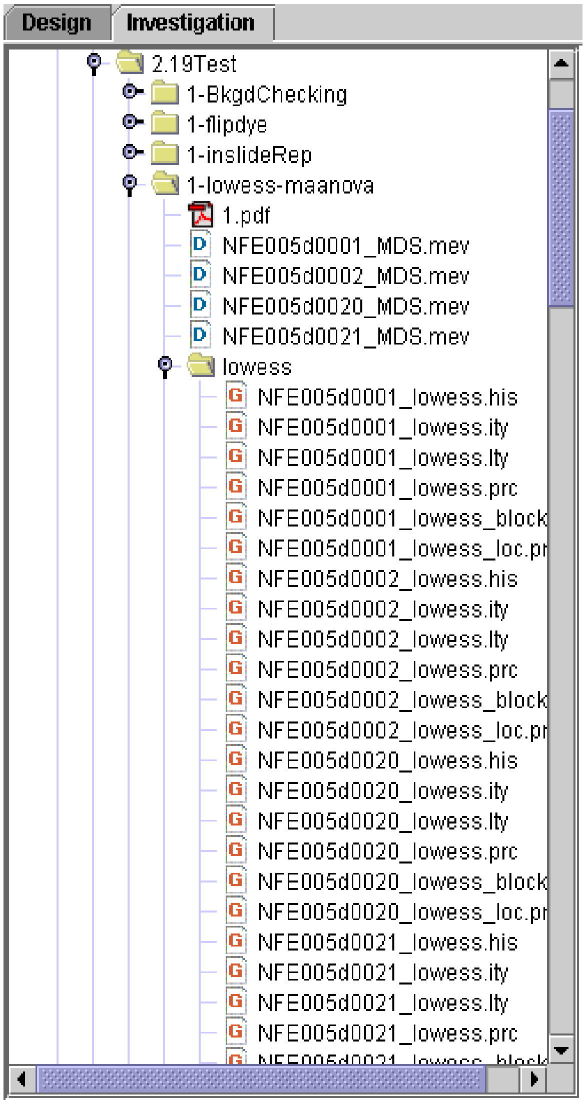
indicates MIDAS is processing data.

7 Investigation panel

Investigation panel allows a user to view raw and processed data file in a spreadsheet (TIGR-MEV format), draw a variety of analysis graphs, as well as display PDF reports.

Clicking the Investigation tab under the graphical tool bar brings this panel to the front. This panel is composed of two windows: Explorer and Viewer.

7.1 Explorer



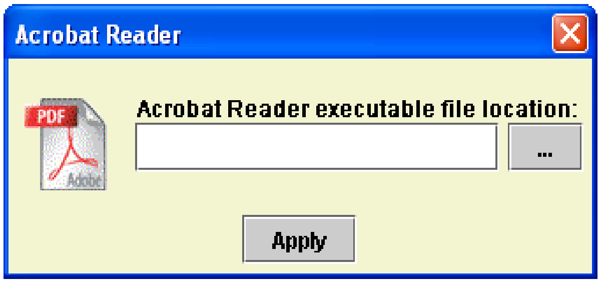
```
2.19Test
    1-BkgdChecking
    1-flipdye
    1-inslideRep
    1-lowess-maanova
        1.pdf
        NFE005d0001_MDS.mev
        NFE005d0002_MDS.mev
        NFE005d0020_MDS.mev
        NFE005d0021_MDS.mev
        lowess
            NFE005d0001_lowess.his
            NFE005d0001_lowess.ity
            NFE005d0001_lowess.lty
            NFE005d0001_lowess.prc
            NFE005d0001_lowess_block
            NFE005d0001_lowess_loc.pr
            NFE005d0002_lowess.his
            NFE005d0002_lowess.ity
            NFE005d0002_lowess.lty
            NFE005d0002_lowess.prc
            NFE005d0002_lowess_block
            NFE005d0002_lowess_loc.pr
            NFE005d0020_lowess.his
            NFE005d0020_lowess.ity
            NFE005d0020_lowess.lty
            NFE005d0020_lowess.prc
            NFE005d0020_lowess_block
            NFE005d0020_lowess_loc.pr
            NFE005d0021_lowess.his
            NFE005d0021_lowess.ity
            NFE005d0021_lowess.lty
            NFE005d0021_lowess.prc
            NFE005d0021_lowess_block
```

Explorer is a file browser which allows a user to navigate through the file system and select a MIDAS data file (TIGR-MEV format), or a MIDAS plot files, or a PDF report file. The selected file will be displayed or drawn in the Viewer window.

Each MIDAS data file name (TIGR-MEV format) will be led by icon D. Highlighting the selected data file, right-clicking it and them pressing "Plot/View" in the popup will display the file in spreadsheet format in the Data Viewer window.

Each MIDAS plot file name will be led by icon G. Highlighting the selected plot file, right-clicking it and then pressing the "Plot/View" in the popup will draw the plot file in the Graph Viewer window.

Each MIDAS PDF report file will be led by icon. Highlighting the selected plot file, right-clicking it and then pressing the "Plot/View" in the popup will launch default PDF reader software, such as Adobe Acrobat Reader, installed in user's computer to display the report. If the default PDF reader cannot be located by MIDAS, a dialog window will be displayed to require user to manually specify the PDF reader software's location in the file system.

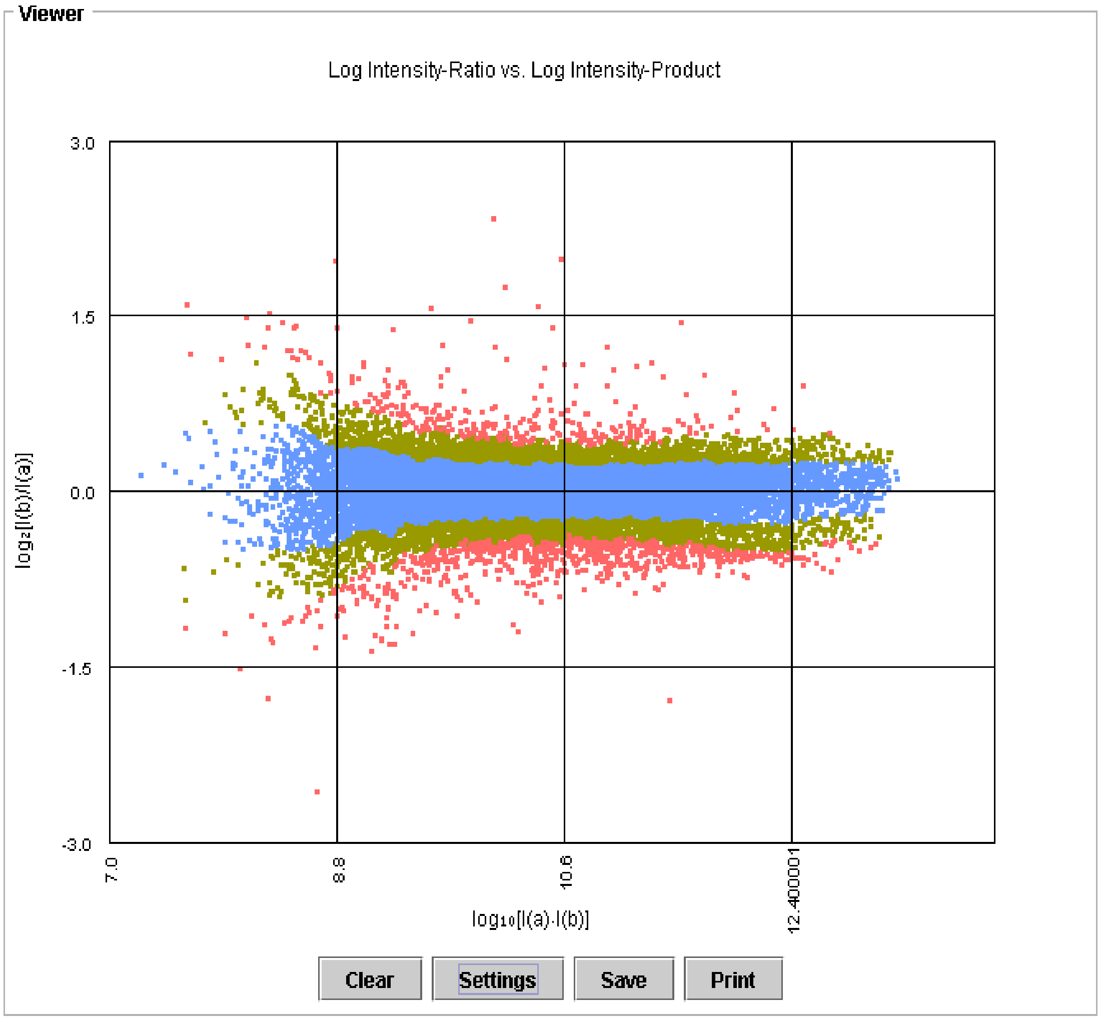


7.2 Graph Viewer

Graphing is an intuitive way for visualizing microarray data analysis results. MIDAS graphing feature allows a variety of analysis graphs to be plotted by reading MIDAS plot files. The plot files are prepared during the execution of analysis processes that user defined and are saved within the corresponding sub-directories under MIDAS *project folder* (Section 3.1).

The available graphs available for plotting and their corresponding plot file types in this release of MIDAS include:

- **RI graph (.*prc* file)**

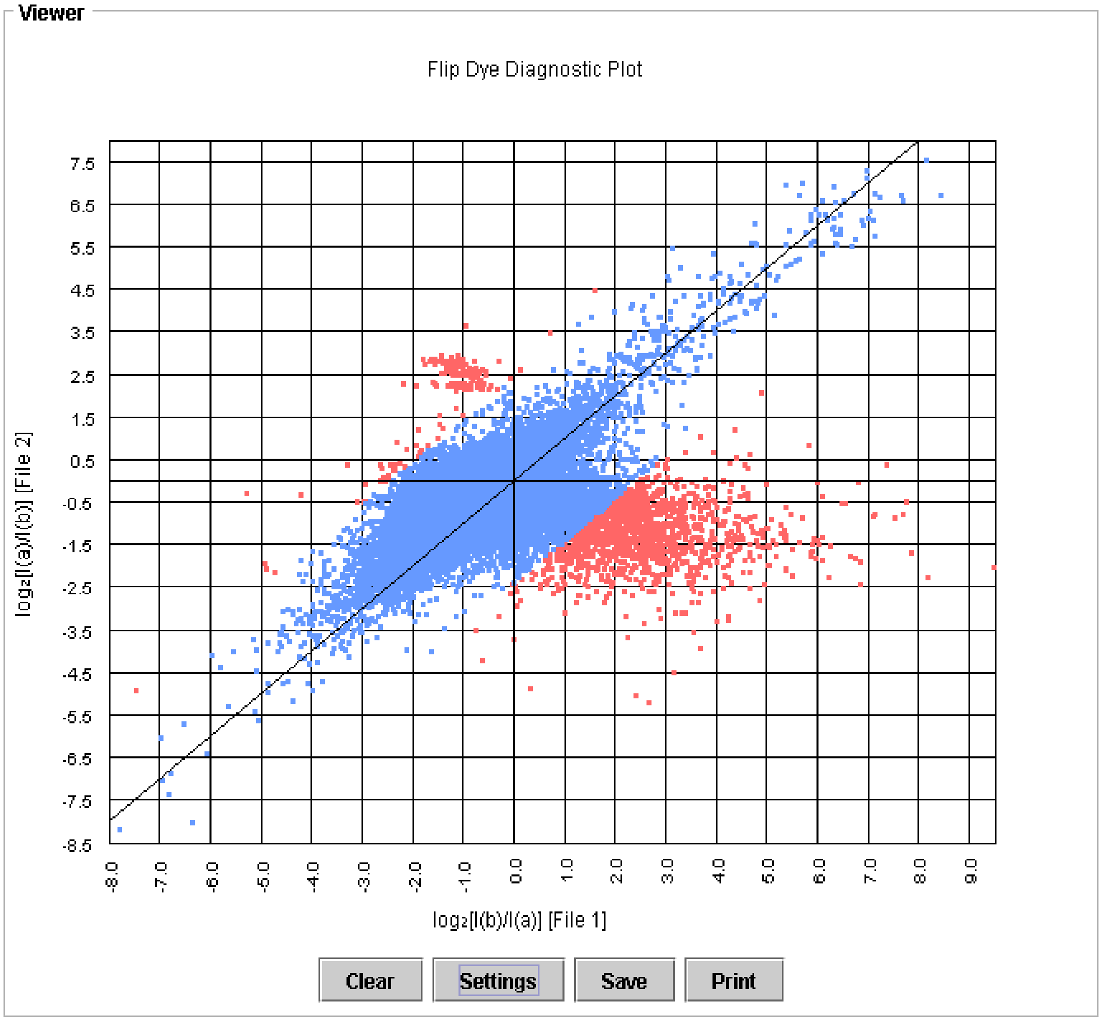


RI graph is plotted based on $\log_2 \frac{I(B)}{I(A)}$ vs. $\log_{10}(I(A) \cdot I(B))$ data sets of a slide, where,

I(A) denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

The corresponding plot files for RI graph are generated with extension *.prc*.

- **Flip Dye Diagnostic graph**



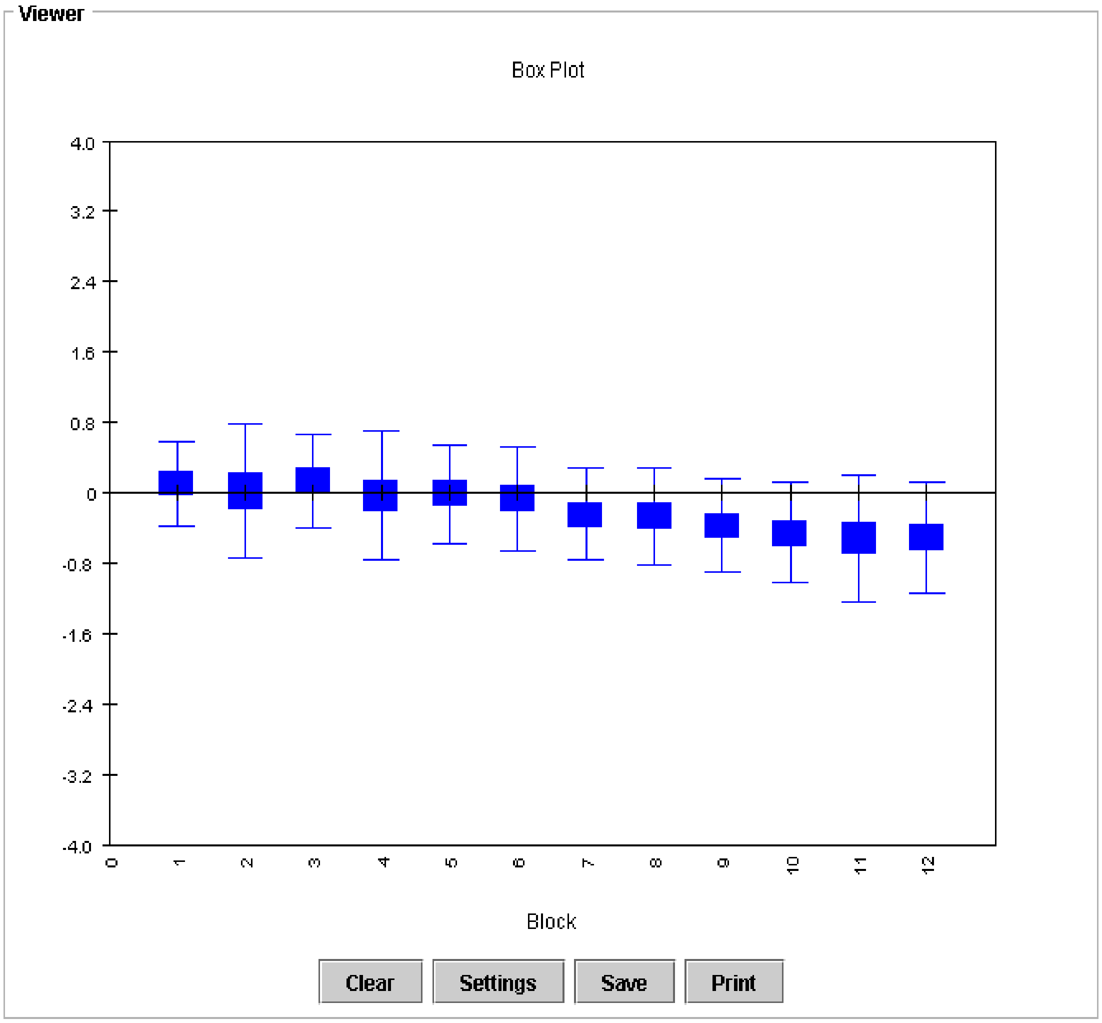
Flip Dye Diagnostic graph is plotted based on $\log_2 \frac{I(B)_1}{I(A)_1}$ vs. $\log_2 \frac{I(A)_2}{I(B)_2}$ data sets of a pair of flip dye slides, where,

I(A)$_1$ denotes the channelA intensity of each spot in the slide1 of the flip dye pair;
I(B)$_1$ denotes the channelB intensity of each spot in the slide1 of the flip dye pair;
I(A)$_2$ denotes the channelA intensity of each spot in the slide2 of the flip dye pair;
I(B)$_2$ denotes the channelB intensity of each spot in the slide2 of the flip dye pair;

Flip Dye Diagnostic graph provides an excellent visual summary of how consistent the gene expressions are between a pair of Flip-dye experiments (Section 5.16) works.

The corresponding plot files for Flip Dye Diagnostic graph are generated with extension *.rrc*.

- **Box graph**

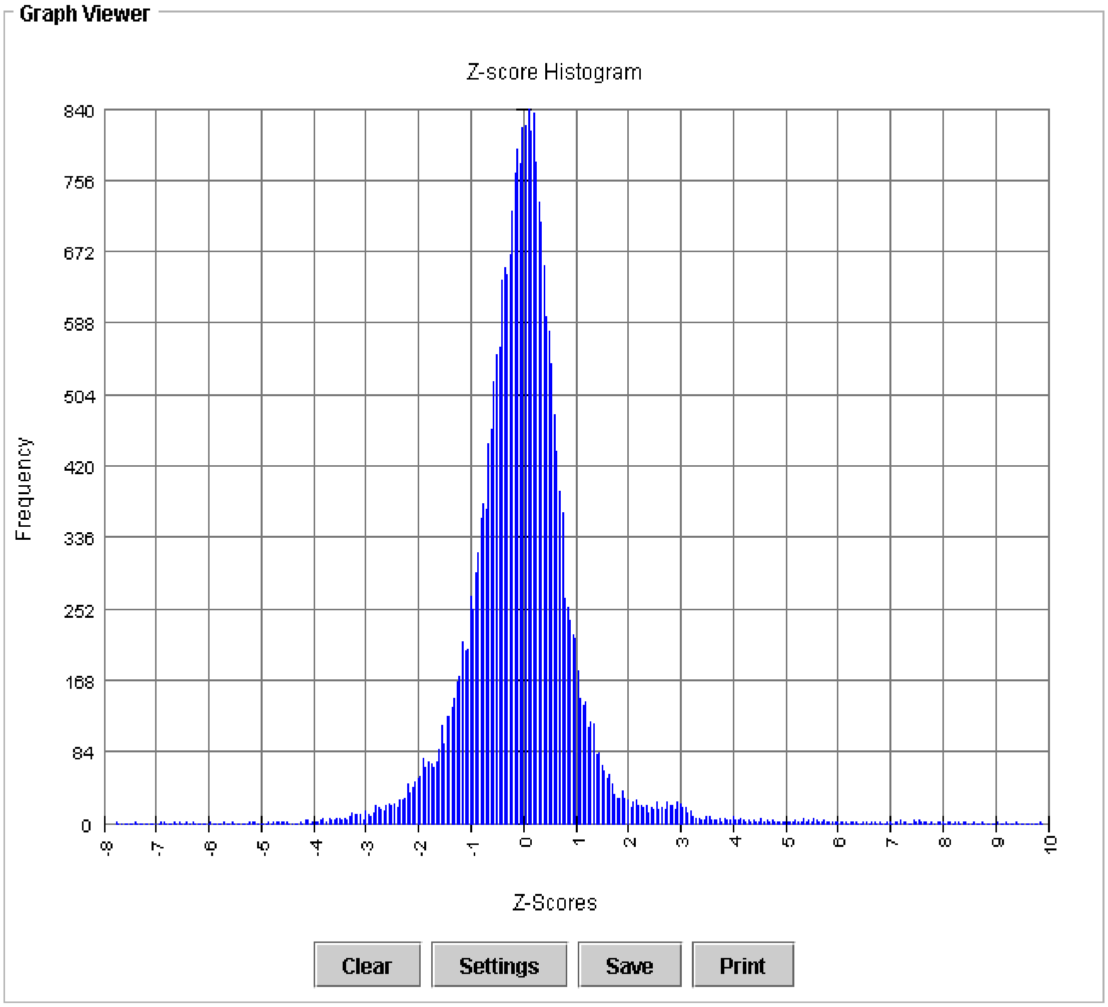


Box graph depicts how the distribution of expression values, or $\log_2 \frac{I(B)}{I(A)}$ values, of spots vary among different blocks within a slide or among different slides, where *I(A)* denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

Box graph provides an excellent visual summary of how Standard Deviation Regularization (Section 5.9) works.

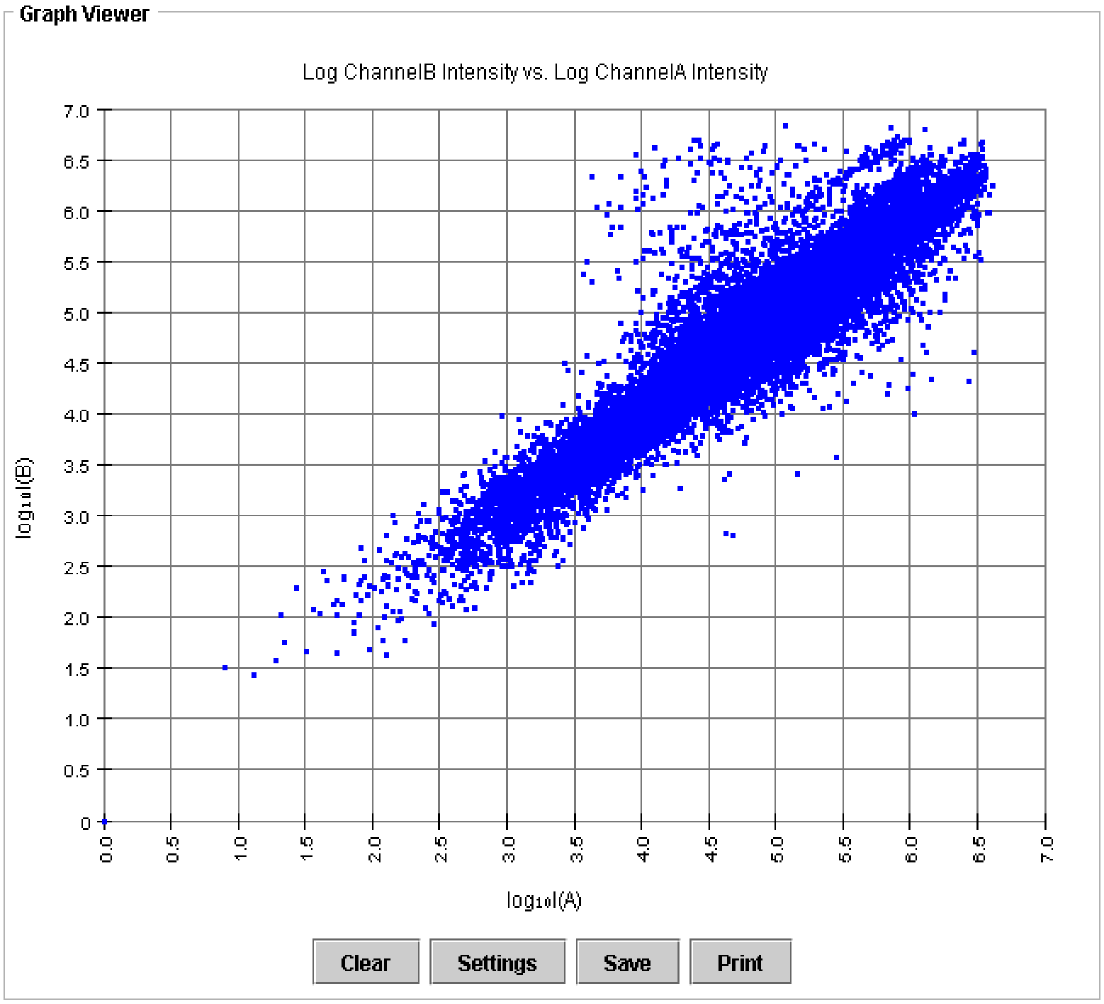
The corresponding plot files for Box graph are generated with extension *.box*.

- **Histogram graph**



Z-score histogram graph depicts how Z scores of spots within a slide distribute.

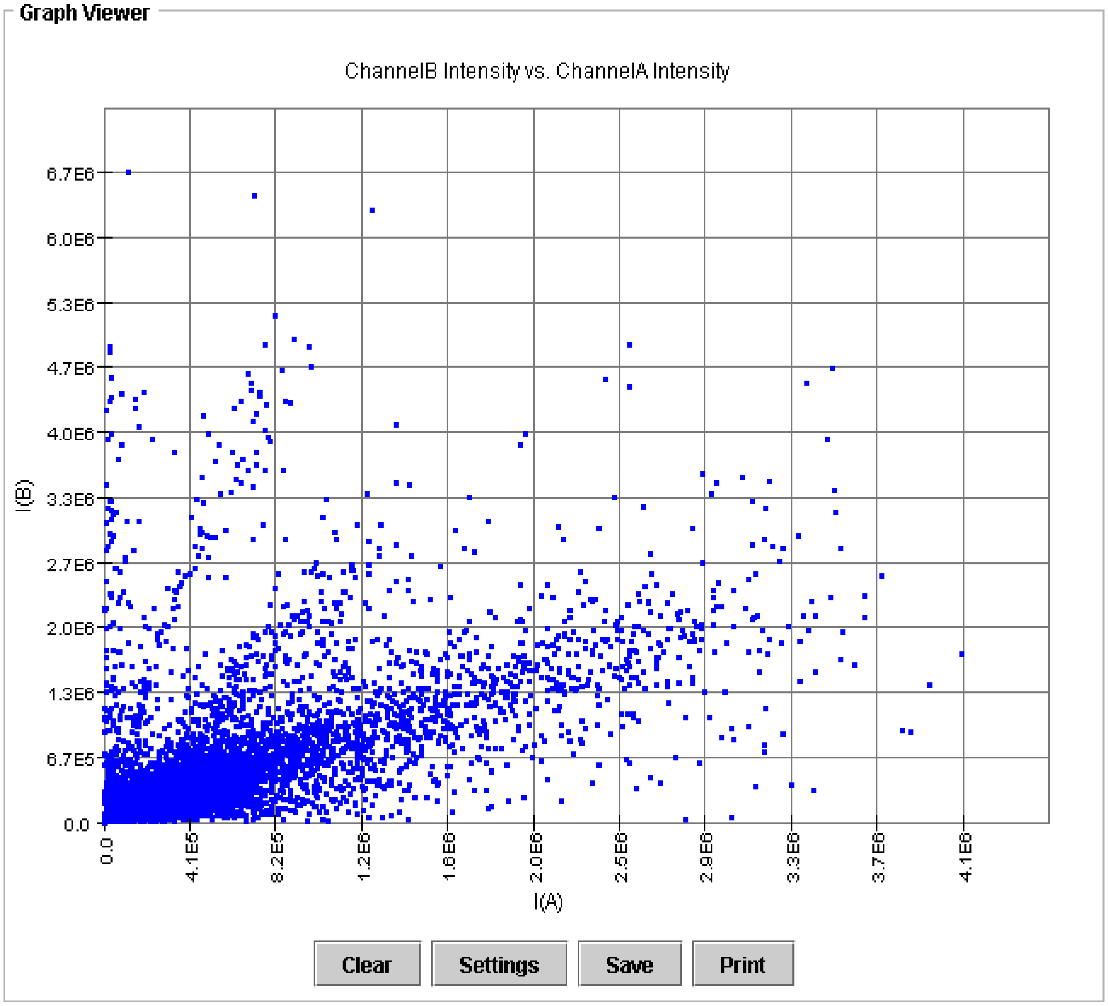
- **Log Intensity graph**



Log Intensity graph shows the relations between the $\log_{10}$(intensity)s of two channels for each spot in a slide.

The corresponding plot files for Intensity graph are generated with extension *.lty*.

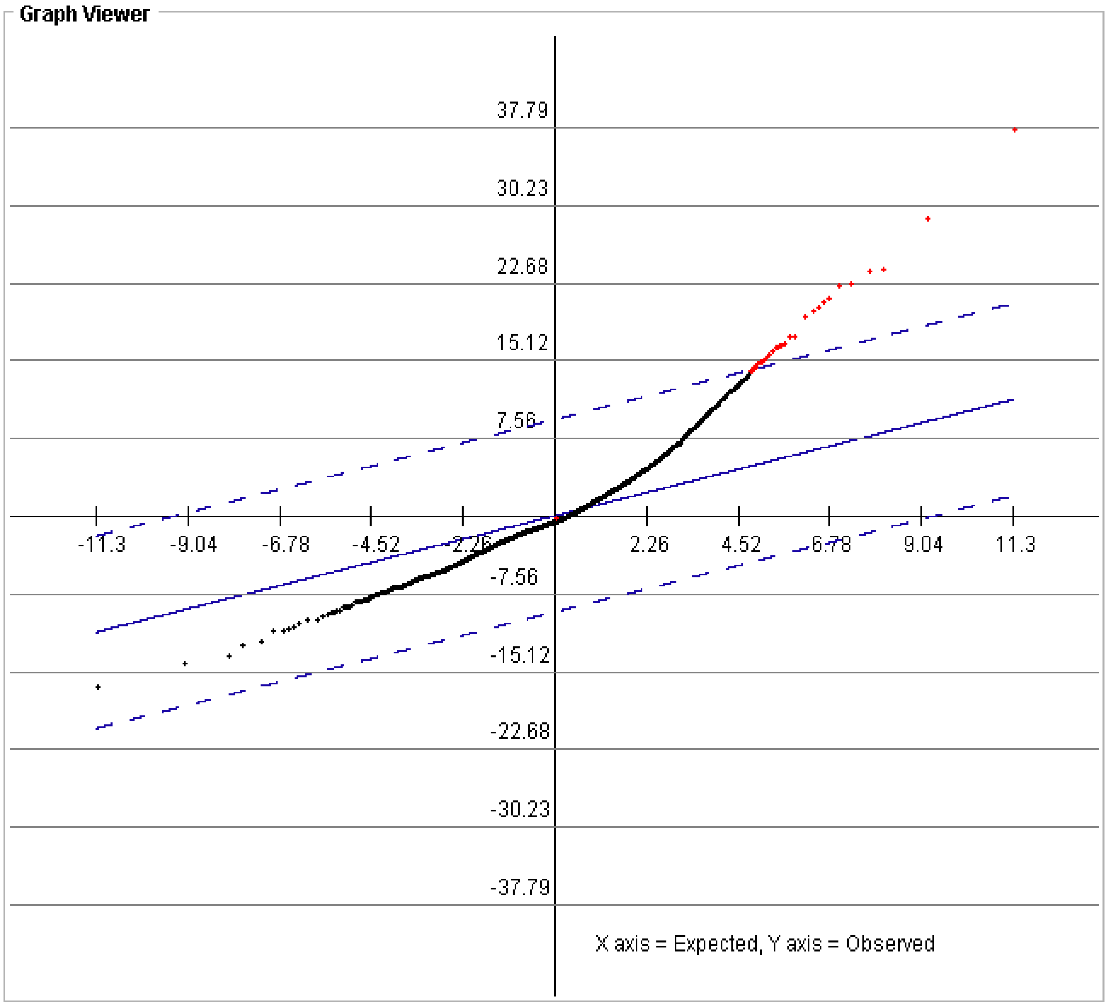
- **Intensity graph**



Intensity graph shows the relations between the intensities of two channels for each spot in a slide.

The corresponding plot files for Intensity graph are generated with extension *.ity*.

- **SAM graph**



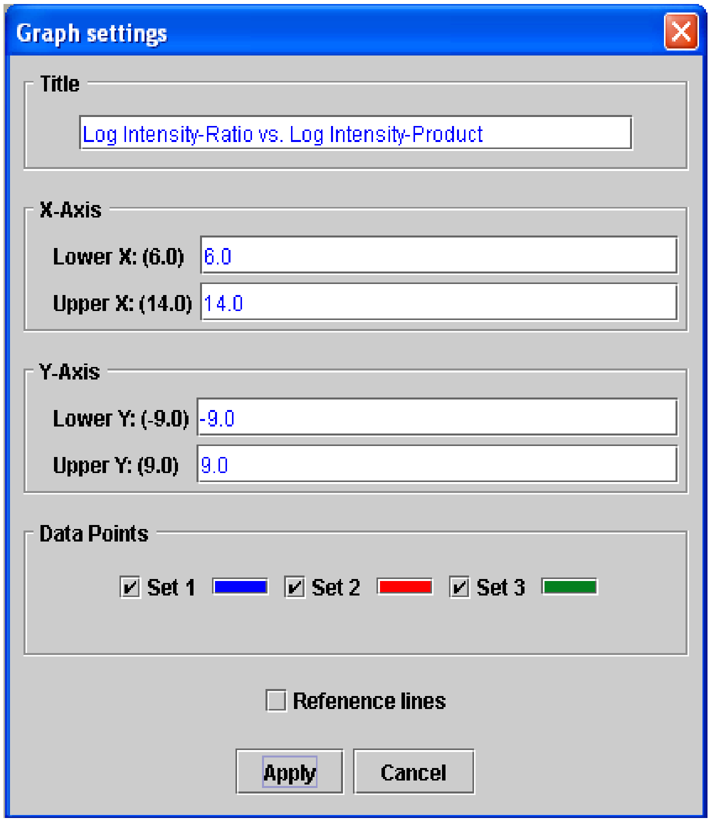
SAM graph shows a snapshot of the interactive SAM plot that user used to identify the significant genes during the SAM analysis.

The corresponding plot files for Intensity graph are generated with extension *.sam*.

When graphs (except SAM graph) are displayed in the Viewer, the button controls below the graph allow a user to perform the following operations.

- **Clear**
  Clear all graphs in the Viewer.

- **Settings**



Title

Log Intensity-Ratio vs. Log Intensity-Product

X-Axis

Lower X: (6.0) 6.0

Upper X: (14.0) 14.0

Y-Axis

Lower Y: (-9.0) -9.0

Upper Y: (9.0) 9.0

Data Points

☑ Set 1 🟦 ☑ Set 2 🟥 ☑ Set 3 🟩

☐ Reference lines

- Title
  Changes the graph title displayed.

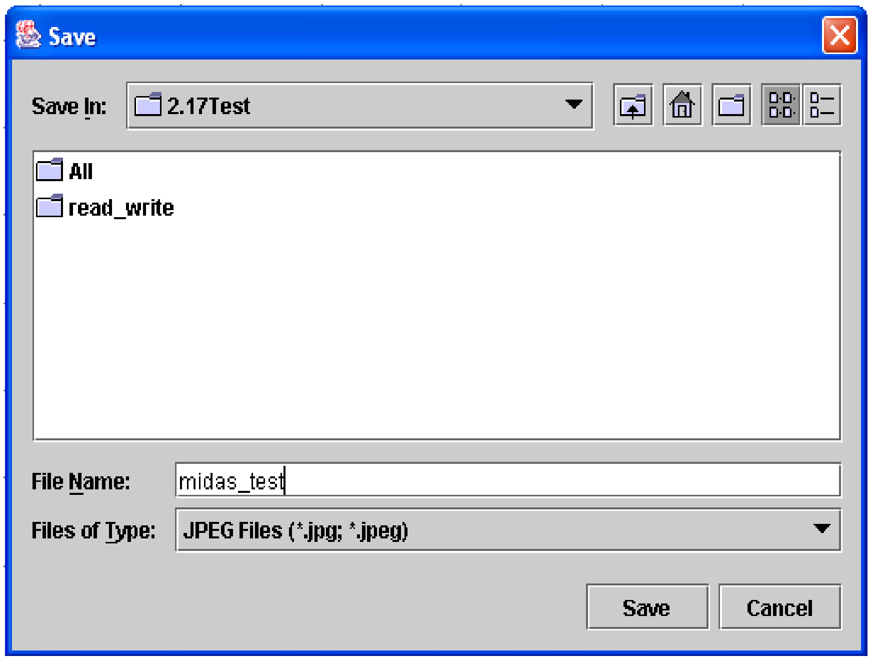
- X-Axis
  Changes the minimum and maximum values of the X-axis.

- Y-Axis
  Changes the minimum and maximum values of the Y-axis.

- Data Points
  Changes the colors of each data set in the graph. The checkbox in front of each data set allows a user to toggle this data set on and off.

- Reference Lines
  Add or remove reference lines to the graph.

- **Save**
  Save the graph with a user-specified image format.

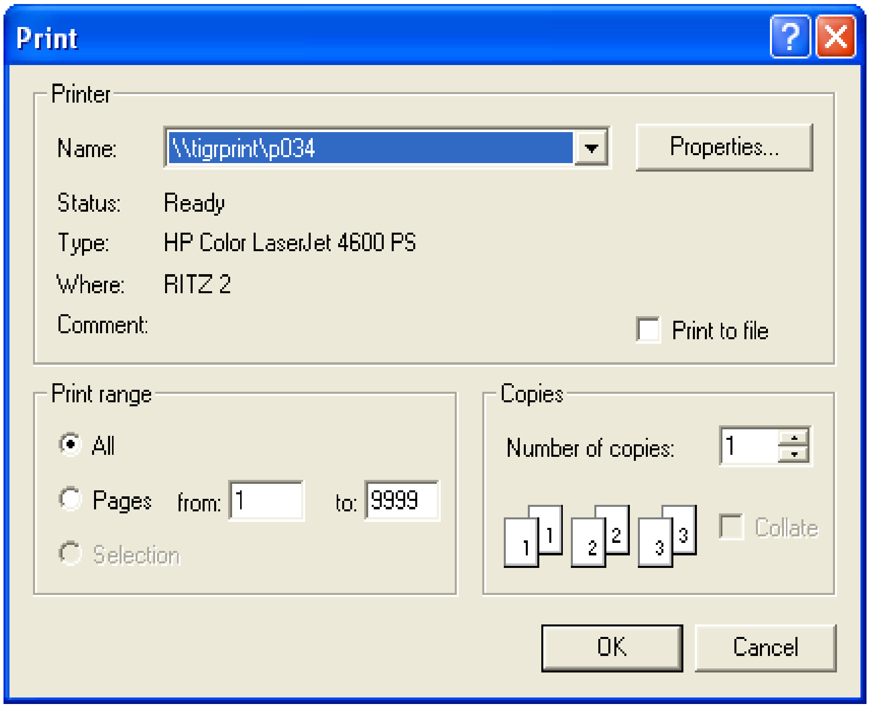


Save In: 2.17Test

File Name: midas_test

Files of Type: JPEG Files (*.jpg; *.jpeg)

- **Print**
  Print the graph.



Name: \\tigrprint\p034

Status: Ready

Type: HP Color LaserJet 4600 PS

Where: RITZ 2

Comment:

□ Print to file

● All

○ Pages from: 1 to: 9999

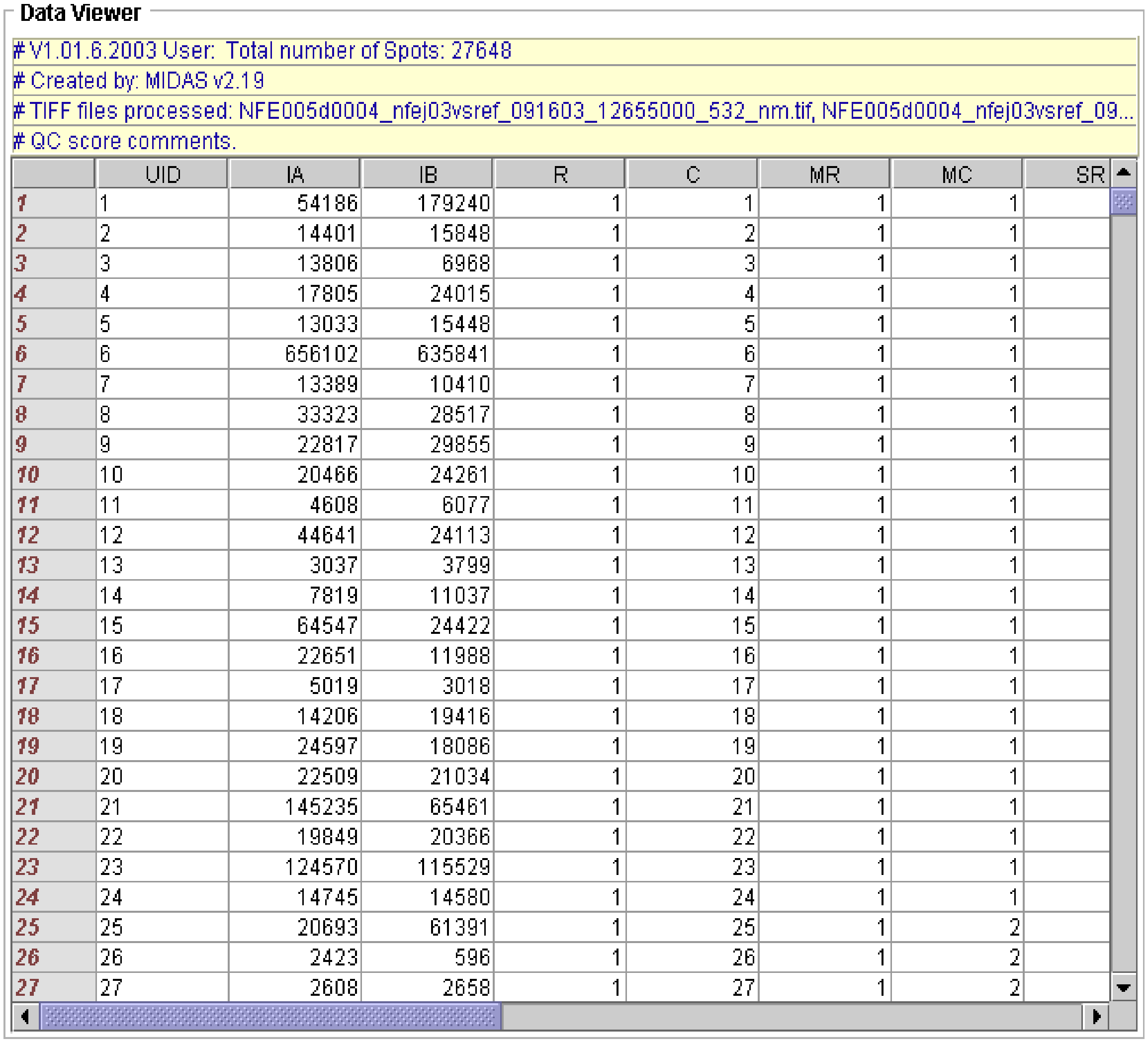
○ Selection

Number of copies: 1

□ Collate

7.3 Data Viewer

A data file in TIGR-MEV format can be view as spreadsheet in the Data Viewer window.



Data Viewer

```
# V1.01.6.2003 User:  Total number of Spots: 27648
# Created by: MIDAS v2.19
# TIFF files processed: NFE005d0004_nfej03vsref_091603_12655000_532_nm.tif, NFE005d0004_nfej03vsref_09...
# QC score comments.
```

	UID	IA	IB	R	C	MR	MC	SR
1	1	54186	179240	1	1	1	1	
2	2	14401	15848	1	2	1	1	
3	3	13806	6968	1	3	1	1	
4	4	17805	24015	1	4	1	1	
5	5	13033	15448	1	5	1	1	
6	6	656102	635841	1	6	1	1	
7	7	13389	10410	1	7	1	1	
8	8	33323	28517	1	8	1	1	
9	9	22817	29855	1	9	1	1	
10	10	20466	24261	1	10	1	1	
11	11	4608	6077	1	11	1	1	
12	12	44641	24113	1	12	1	1	
13	13	3037	3799	1	13	1	1	
14	14	7819	11037	1	14	1	1	
15	15	64547	24422	1	15	1	1	
16	16	22651	11988	1	16	1	1	
17	17	5019	3018	1	17	1	1	
18	18	14206	19416	1	18	1	1	
19	19	24597	18086	1	19	1	1	
20	20	22509	21034	1	20	1	1	
21	21	145235	65461	1	21	1	1	
22	22	19849	20366	1	22	1	1	
23	23	124570	115529	1	23	1	1	
24	24	14745	14580	1	24	1	1	
25	25	20693	61391	1	25	1	2	
26	26	2423	596	1	26	1	2	
27	27	2608	2658	1	27	1	2	

The first several lines highlighted in yellow are comments. The comment lines are followed by a header line and then data body. Each row in the spreadsheet records data about a spot on the slide. The column can be ascending sorted by single-clicking, or double-clicking the column header name. When multiple column sorting is desired, first sort the first column, hold CTR key and sort the second column, and so on.

8 MIDAS outputs

After the execution of a MIDAS analysis process is finished, a series of output files are generated to store process results or errors under the MIDAS *project folder* (Section 3.1).

Assuming *sample.xxx* is the file name when input is a single data file, or *sample1.xxx* and *sample2.xxx* are the file names when inputs are a flip dye file pair, where *xxx* is either *tav* or *mev* representing TIGR-TAV or TIGR-MEV data format, MIDAS output files may include:

- *project.prj*

- *project.rpt*
- *project.pdf*
- *project.err*
- *sample_MDS.xxx*
- *sample1_sample2_MDS.xxx*
- *sample_MDSTrim.xxx*
- *sample1_sample2_MDSTrim.xxx*
- *sample_[module].prc*
- *sampl_[module]e.box*
- *sampl_[module]e.ity*
- *sampl_[module]e.lty*
- *sampl_[module]e.his*
- *sample1_sample2_[module].his.rrc*
- *project.sam*

They are explained in details below.

- *project.prj*
  This is the MIDAS project file (Section 3.1), which stores the user-defined analysis process sequence and the parameters set for each step.

- *project.rpt* and *project.pdf*
  These are MIDAS text report file and PDF-format report file. The report files record the analysis process sequence and all the parameter settings as well as text or graphical results of each analysis step.

  By default, only the text report is prepared for each MIDAS project. User can demand the PDF report also be prepared by checking the **Report/Create PDF report** checkbox before executing the analysis.

  Note: Since PDF report contains analysis graphs, it takes longer CPU time and requires more system resources (RAM and hard disk) to prepare.

- *project.err*
  This is the error file, which records error message thrown out by MIDAS id error occurs during the execution.

- *sample_MDS.xxx* and *sample1_sample2_MDS.xxx*
  After MIDAS finishes a user-defined processe, *sample_MDS.xxx* stores the processed data when input is a single file and *sample1_sample2_MDS.xxx* store the processed data when inputs are a flip dye file pair. They follow the same data format as the input, in either TIGR-TAV or TIGR-MEV format.

If "Virtual Trim" (Section 5.14) is applied, the file will have the same row number as the input data file but have zero reset for channelA and channelB intensities for those spots marked as "bad" during the analysis process.

If "Virtual Trim" (Section 5.14) is NOT applied, the file will have a reduced row number than the input data file. Those spots marked as "bad" during the analysis process are trimmed
For flip dye consistency checking process, the output file will be named sample1_sample2_MDS.tav, given sample1.tav and sample2.tav are the two input file names in a pair.

- *sample_MDSTrim.xxx* and *sample1_sample2_MDSTrim.xxx*
  These are files recording those spots that are marked as "bad" and trimmed during the execution of an analysis process. Only if "Output Trimmed Data" (Section 5.14) is applied, MIDAS will generate these file(s) under the MIDAS project folder.

- *sample_[module].prc*
  These are the plot files for MIDAS to make RI graph (Section 7.2), which are prepared during executing the analysis module.

  There are three columns of data in this file:
  column1 – sorted $\log_{10}(I(A) \cdot I(B))$
  column2 – $\log_2 \frac{I(B)}{I(A)}$
  column3 – original row index in the input file.

  I(A) denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

- *sample_[module].box,*
  These are the plot files for MIDAS to make Box graph (Section 7.2), which are prepared during executing the analysis module.

  The row number of this file equals to the number of *block* in the input data file. *Block* is defined by *meta-row* and *meta-column* data of each spot in the input data file. Spots within the same block should have the same meta-row and meta-column data. In TIGR-TAV format file, meta-row data is stored in column3; meta-column data is stored in column4. In TIGR-MEV format file, meta-row and meta-column are stored under the column with header MR and MC.

  Within each row, $\log_2 \frac{I(B)}{I(A)}$ for each spot is recorded, where *I(A)* denotes the channelA intensity of each spot and *I(B)* denotes the channelB intensity of each spot.

- *sample_[module].ity*

These are the plot files for MIDAS to make Intensity graph (Section 7.2), which are prepared during executing the analysis module.

There are two columns of data in this file:
column1 – *I(A)*
column2 – *I(B)*

I(A) denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

- *sample_[module].lty*
  These are the plot files for MIDAS to make Intensity graph (Section 7.2), which are prepared during executing the analysis module.

  There are two columns of data in this file:
  column1 – $\log_{10}(I(A))$ values
  column2 – $\log_{10}(I(B))$ values

  I(A) denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

- *sample_[module].his*
  These are the plot files for MIDAS to make z-score histogram graph which are prepared during executing the analysis module.

  There are eight columns of data in this file:
  column1 – sorted $\log_{10}(I(A)\cdot I(B))$
  column2 – $\log_2 \dfrac{I(B)}{I(A)}$
  column3 – $\log_2 \dfrac{I(B)}{I(A)}$ mean
  column4 – $\log_2 \dfrac{I(B)}{I(A)}$ standard deviation
  column5 – Z-score
  column6 – T-distribution p-value of Z-score
  column7 – $\log_{10}$(T-distribution p-value of Z-score)
  column8 – original row index in the input file.

  I(A) denotes the channelA intensity of each spot in the slide and *I(B)* denotes the channelB intensity of each spot in the slide.

- *sample1_sample2_[module].rrc*
  These are the plot files for MIDAS to make Flip Dye Diagnostic graph (Section 7.2), which are prepared during executing the analysis module.

There are two columns of data in this file:

column1 – $\log_2 \frac{I(B)_1}{I(A)_1}$

column2 – $\log_2 \frac{I(A)_2}{I(B)_2}$

I(A)$_1$ denotes the channelA intensity of each spot in the slide1 of the flip dye pair;
I(B)$_1$ denotes the channelB intensity of each spot in the slide1 of the flip dye pair;
I(A)$_2$ denotes the channelA intensity of each spot in the slide2 of the flip dye pair;
I(B)$_2$ denotes the channelB intensity of each spot in the slide2 of the flip dye pair;

Note: *.rrc* plot file(s) are only available when the inputs are flip dye file pair(s) and *Flip dye consistency checking* (Section 5.15) is defined in the analysis process sequence.

- *project.sam*
  These are the plot files for MIDAS to make SAM graph (Section 5.14), which are prepared during executing the SAM analysis module.

9 License

This software is OSI Certified Open Source Software.
OSI Certified is a certification mark of the Open Source Initiative.

Please view the license (Artistic_License.pdf) in the root MIDAS directory.

10 Appendix

10.1 TIGR TAV file format descriptions

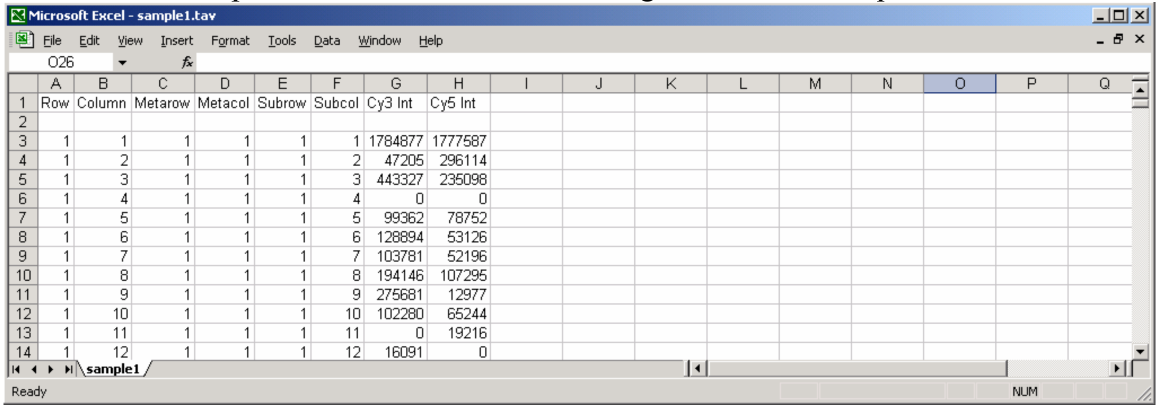
The original TAV (TIGR Array Viewer) file type was an eight-column, tab-delimited text format developed at TIGR for the purposes of storing the intensity values of the spots on a single slide. It is written out by the program TIGR Spotfinder and contains one row for each spot. The first six columns of the file contain positional data for the spots and are followed by two columns of intensity data.

These eight columns are required by MIDAS for normalization and filtering operations. Optional columns can contain flags, background intensities, annotation, genbank numbers, feat name, etc. Appropriate optional columns are required when certain operations such as flag checking, background checking or In-slide replicates analysis are applied.

A flag is simply a letter code corresponding to a description of the spot:

A – 0 non-saturated pixels in the spot
B – 0-50 non-saturated pixels in the spot
C – 50 or more non-saturated pixels in the spot
X – spot is rejected, due to spot shape and intensity relative to background
Y – background is higher than spot intensity
Z – spot not detected by Spotfinder.

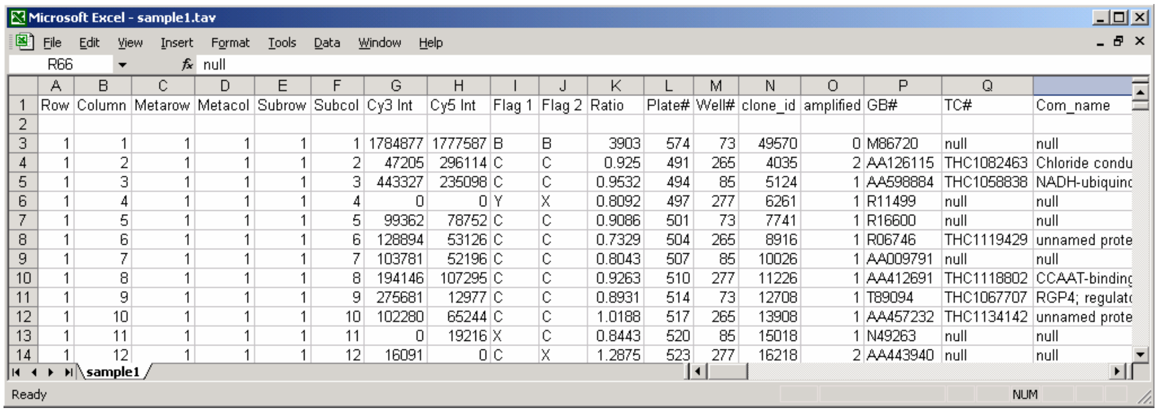
Below is an example of TIGR TAV file containing the minimum required fields.



	A	B	C	D	E	F	G	H
1	Row	Column	Metarow	Metacol	Subrow	Subcol	Cy3 Int	Cy5 Int
2								
3	1	1	1	1	1	1	1784877	1777587
4	1	2	1	1	1	2	47205	296114
5	1	3	1	1	1	3	443327	235098
6	1	4	1	1	1	4	0	0
7	1	5	1	1	1	5	99362	78752
8	1	6	1	1	1	6	128894	53126
9	1	7	1	1	1	7	103781	52196
10	1	8	1	1	1	8	194146	107295
11	1	9	1	1	1	9	275681	12977
12	1	10	1	1	1	10	102280	65244
13	1	11	1	1	1	11	0	19216
14	1	12	1	1	1	12	16091	0

Below is an example of TIGR TAV file containing extra fields besides the minimum required fields.

Note: the first header row is for demonstration purpose and not included in the real TIGR TAV file.



Microsoft Excel - sample1.tav
File Edit View Insert Format Tools Data Window Help
R66
null
sample1
Ready
NUM

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	Row	Column	Metarow	Metacol	Subrow	Subcol	Cy3 Int	Cy5 Int	Flag 1	Flag 2	Ratio	Plate#	Well#	clone_id	amplified	GB#	TC#	Com_name
2																		
3	1	1	1	1	1	1	1784877	1777587	B	B	3903	574	73	49570	0	M86720	null	null
4	1	2	1	1	1	2	47205	296114	C	C	0.925	491	265	4035	2	AA126115	THC1082463	Chloride condu
5	1	3	1	1	1	3	443327	235098	C	C	0.9532	494	85	5124	1	AA598884	THC1058838	NADH-ubiquinc
6	1	4	1	1	1	4	0	0	Y	X	0.8092	497	277	6261	1	R11499	null	null
7	1	5	1	1	1	5	99362	78752	C	C	0.9086	501	73	7741	1	R16600	null	null
8	1	6	1	1	1	6	128894	53126	C	C	0.7329	504	265	8916	1	R06746	THC1119429	unnamed prote
9	1	7	1	1	1	7	103781	52196	C	C	0.8043	507	85	10026	1	AA009791	null	null
10	1	8	1	1	1	8	194146	107295	C	C	0.9263	510	277	11226	1	AA412691	THC1118802	CCAAT-binding
11	1	9	1	1	1	9	275681	12977	C	C	0.8931	514	73	12708	1	T89094	THC1067707	RGP4; regulatc
12	1	10	1	1	1	10	102280	65244	C	C	1.0188	517	265	13908	1	AA457232	THC1134142	unnamed prote
13	1	11	1	1	1	11	0	19216	X	C	0.8443	520	85	15018	1	N49263	null	null
14	1	12	1	1	1	12	16091	0	C	X	1.2875	523	277	16218	2	AA443940	null	null

Note: the first header row is for demonstration purpose and not included in the real TIGR TAV file.

10.2 TIGR MEV file format descriptions

Evolved yet different from TIGR-TAV format, TIGR-MEV file format stores array expression data and annotation information into two associated files: an MEV file and an annotation file.

- **MEV file**

  A MultiExperimentViewer or mev file is a tab-delimited text file that contains coordinate and expression data for a single microarray experiment. A single header row is required to precede the expression data in order to identify the columns below. With the exception of optional comment lines, each remaining row of the file stores data for a particular spot on the array.

  MIDAS and other TM4 software tools will consider comment lines non-computational. A comment line must start with the pound symbol '#', and can be included anywhere in the file. If the pound symbol is the first character on a line, the entire line (up to the new line character '\n') will be ignored by the software tool.

  An example of the leading comments:
  # Version: V1.0 \t Date: 11/06/2002 \t Creator: aisaeed
  # Analysis_id: 10579 \t Slide_type: IASCAG1 \t Row Count: 32448
  # Created by: TIGR Spotfinder 2.2.2
  # TIFF files processed: gpc30025a_532_nm.tif, gpc30025a_635_nm.tif
  # Description: Tumor type comparison

  The header row consists of the field names for each subsequent row in this file (with the exception of comment lines). A minimum of seven columns must be present, and these must use a set of specifically named headers. Any number of additional columns may be included. The seven required column headers are:

UID	Unique identifier for this spot
IA	Intensity value in channel A

IB	Intensity value in channel B
R	Row (slide row)
C	Column (slide column)
MR	Meta-row (block row)
MC	Meta-column (block column)

An example *.mev* file created by Spotfinder may look like this:

UID \t IA \t IB \t R \t C \t MR \t MC \t SR \t SC \t FlagA \t FlagB \t SA \t SF \t QCScore \t QCA \t QCB \t BkgA \t BkgB \t SDIA \t SDIB \t SDBkgA \t SDBkgB \t IAmedian \t IBmedian

UID	Unique identifier for this spot
IA	Intensity value in channel A
IB	Intensity value in channel B
R	Row (slide row)
C	Column (slide column)
MR	Meta-row (block row)
MC	Meta-column (block column)
SR	Sub-row
SC	Sub-column
FlagA	*TIGR Spotfinder* flag value in channel A
FlagB	*TIGR Spotfinder* flag value in channel B
SA	Actual spot area (in pixels)
SF	Saturation factor
QCScore	Cumulative quality control score
QCA	Quality control score in channel A
QCB	Quality control score in channel B
BGA	Background value in channel A
BGB	Background value in channel B
SDIA	Standard deviation of the intensity value in channel A
SDIB	Standard deviation of the intensity value in channel B
SDBkgA	Standard deviation of the background value in channel A
SDBkgB	Standard deviation of the background value in channel B
IAmedian	Median intensity value in channel A
IBmedian	Median intensity value in channel B

The first seven fields (UID, IA, IB, R, C, MR and MC) are required as specified above. This flexible format allows users to track slide-specific data of interest, such as background, spot size and alternate intensities without requiring them of all users or adopting a limited ‘vocabulary’ of field names. This header row serves to identify the required and additional data columns. UID must be the left-most column in the mev file. Other columns do not need to be present in a fixed order.

mev files are required to end with the extension ‘.mev’. At this time there are no further naming conventions for mev files.

Example MeV File:
Based on this format description, the first few rows of a mev file created at TIGR might look like:

```
# Version: V1.0 \t Date: 11/06/2002 \t Creator: aisaeed
# Analysis_id: 10579 \t Slide_type: IASCAG1 \t Row Count: 32448
# Created by: TIGR Spotfinder 2.2.2
# TIFF files processed: gpc30025a_532_nm.tif, gpc30025a_635_nm.tif
# Description: Tumor type comparison
# This is the 4th experiment in a series of 20 to identify tissue-specific genes.
UID \t IA \t IB \t R \t C \t MR \t MC \t SR \t SC \t FlagA \t FlagB \t SA
cage:1043 \t 20934 \t 390823 \t 1 \t 1 \t 1 \t 1 \t 1 \t 1 \t C \t C \t 215
cage:1044 \t 298734 \t 90823 \t 1 \t 2 \t 1 \t 1 \t 1 \t 2 \t C \t C \t 198
cage:1045 \t 789435 \t 713952 \t 1 \t 3 \t 1 \t 1 \t 1 \t 3 \t C \t C \t 255
```

To the software, the same file would appear to be:

```
UID \t IA \t IB \t R \t C \t MR \t MC \t SR \t SC \t FlagA \t FlagB \t SA
cage:1043 \t 20934 \t 390823 \t 1 \t 1 \t 1 \t 1 \t 1 \t 1 \t C \t C \t 215
cage:1044 \t 298734 \t 90823 \t 1 \t 2 \t 1 \t 1 \t 1 \t 2 \t C \t C \t 198
cage:1045 \t 789435 \t 713952 \t 1 \t 3 \t 1 \t 1 \t 1 \t 3 \t C \t C \t 255
```

- **Annotation File**
  An annotation file is a tab-delimited text file containing annotation data for a specific array design. Mev files can be associated with an annotation file only if both types of files are based on the same array design. The keys to this association are the unique ids in both files. Rows of mev and annotation files can be associated with each other if the unique ids are identical. A single header row is required to precede the annotation data in order to identify the columns below. Each remaining row of the file stores annotation data for a particular spot on the array.

  Annotation files may contain any number of non-computational comment lines. These lines, starting with '#', will be treated identically to comment lines in mev files, and should precede the header row.

  The header row consists of headers that identify each column of data. Each subsequent row of the file stores data for a particular spot on the array.

  An example of the leading comments:

  ```
  # Version: V3.0 \t Date: 04/20/2003 \t Creator: jwhite
  # GI_Version: 3.0 \t Slide_type: IASCAG1 \t Row Count: 32448
  # Description: Standard annotation file
  ```

The header row consists of the field names for each subsequent row in this file. Only the UID field is required. It must be the first field present and it must be named 'UID'. Any number of additional fields may be included.

Some varieties of annotation files follow. The format may vary depending on the purpose of the file:

UID \t R \t C \t feat_name \t GB# \t TC# \t common_name \t …

UID \t R \t C \t Gene \t Reaction \t Pathway \t …

UID \t R \t C \t Feat_name \t End5 \t End3 \t Chromosome \t …

Of course, it would be possible to combine the fields of these files, or add fields that have not been mentioned here. The goal is to keep the annotation flexible and the processing seamless.

There are not any naming conventions for annotation files at this time. If such a standard is introduced in the future, it will be detailed here.

Example Annotation File:
Based on this format description, the first few rows of an annotation file created at TIGR might look like:

```
# Version: V3.0 \t Date: 04/20/2003 \t Creator: jwhite
# GI_Version: 3.0 \t Slide_type: IASCAG1 \t Row Count: 32448
# Description: Standard annotation file
UID \t R \t C \t clone_name \t gb# \t TC# \t putative:guess
cage:1043 \t 1 \t 1 \t A.t.RCA \t M86720 \t null \t null
cage:1044 \t 1 \t 2 \t Image:511428 \t AA126115 \t THC1324489 \t TC: FXYD domain-containing ion transport regulator 3 precursor
cage:1045 \t 1 \t 3 \t Image:897987 \t AA598884 \t THC1286273 \t TC: NADH-ubiquinone oxidoreductase 39kDa subunit {Homo sapiens}
```

To the software, the same file would appear to be:

```
UID \t R \t C \t clone_name \t gb# \t TC# \t putative:guess
cage:1043 \t 1 \t 1 \t A.t.RCA \t M86720 \t null \t null
cage:1044 \t 1 \t 2 \t Image:511428 \t AA126115 \t THC1324489 \t TC: FXYD domain-containing ion transport regulator 3 precursor
cage:1045 \t 1 \t 3 \t Image:897987 \t AA598884 \t THC1286273 \t TC: NADH-ubiquinone oxidoreductase 39kDa subunit {Homo sapiens}
```

11 References

[1] Quackenbush, J. Microarray data normalization and transformation. Nature Genetics. Vol.32 supplement pp496-501 (2002).

[2] Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. J. Amer. Stat. Assoc. 74, 829-836 (1979).

[3] Finkelstein, D., Gollub, J., etc. Iterative linear regression by sector: renormalization of cDNA microarray data and cluster analysis weighted by cross homology. (2002).

[4] Chen, Y., Dougherty, E.R. & Bittner, M.L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics 2, 364-374 (1997).

[5] Yang, Y.H. et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30, e15 (2002).

[6] Microarray Gene Expression Data Analysis -- A beginner's Guide, ISBN 1-40510-682-4, page55-56

[7] Yang, I.V. et al. Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol. 3, research0062.1-0062.12 (2002).

[8] Pan, W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 18: 546-554. (2002).

[9] Tusher, V.G., R. Tibshirani and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences USA 98: 5116-5121 (2001).

[10] Kerr, Martin, and Churchill. Analysis of Variance for Gene Expression Microarrays. Journal of Computational Biology, 7:819-837 (2000).