

TIGR ExpressConverter

(Version 1.7)



January, 2005

Microarray Software Group

The Institute for Genomic Research

Table of Contents

Introduction	1
Interface Structure	1
Tool Bar	1
Working Place	2
How to Use	9
Convert GPR Files	9
Convert ImaGene Files	9
Convert ScanArray Files	10
Convert ArrayVision Files	10
Convert Agilent Files	10
Convert TAV Files	11
Convert MEV Files	11
Convert Customized Files	11
Delete the Selected Files	13
Edit Output File Names	13
Change Directory for Output Files	14
File Formats	15
TAV File	15
MEV File	15
Annotation File	16
Intensity Converting Algorithms	17

ExpressConverter is a file transformation tool that reads microarray data files in a variety of file formats and generates TIGR MultiExperiment Viewer file (.mev) or TIGR ArrayView file (.tav) as output so that the microarray data can be uploaded to databases with MADAM or analyzed with MIDAS and MEV.

The difference between this version and the previous ones is that the ScanArray page is modified to allow you to define the dye type (Cy3/Cy5) for the channels of input data file. In this way, it ensures that the output mev file always has IA with Cy3 and IB Cy5.

As always, your comments are always appreciated. If you have any questions, please contact Jerry Li (jli@tigr.org) at the Institute for Genomic Research.

1. Interface Structure

The interface of ExpressConverter has two parts as indicated in the Figure 1. At the top is the **Tool Bar** that contains menus and icons through which you can control the converter. The **Working Place** at the bottom shows you both names of the input and output files. The eight tabs in the working place presents the specific interfaces for converting GenePix, ImaGene, ScanArray, ArrayVision, Agilent, Tav, Mev, and Customized files respectively.

1.1 Tool Bar

The Tool Bar contains five menus: File, Edit, View, Ann File and Help (see Figure 1.). When you move mouse over the File and click, you can see a menu that allows you to select input files, start and stop converting, save and open a header file for customized conversion, set your preference, and exit the program. Using the Edit menu, you can delete the selected input files, change the output file names and directory, and clear the Working Place. If the converter does not work well, you can click View menu to open a Log window for error messages. Also from the View menu, you can change the interfaces to your favorite look and feel. Ann File menu allows you to set the number of annotation files to the mev file. When you have questions about the converter, you can get answers by click the Help menu (Mac OSx version does not have this function). The functions of icons are equivalent to the menus. The Output selection allows you to define what type (tav/mev) of file formats to be produced.

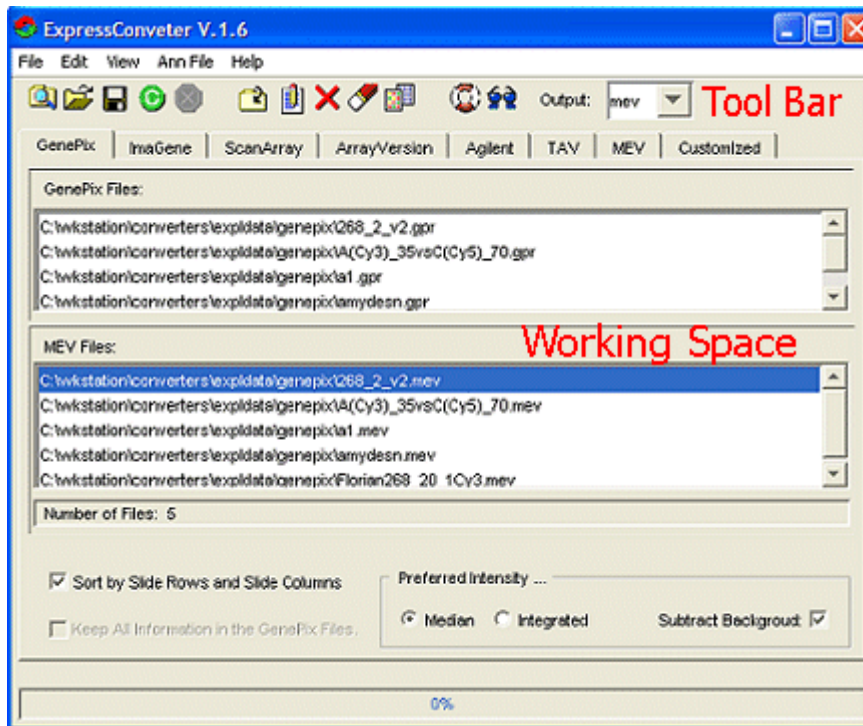




Figure 1. The working window of ExpressConverter.

1.2 Working Place

The Working Place shows you a file window and other controls. The file window consists of input file window at the top and the output window at the bottom. Below the windows are function controls that vary based on the types of input and output files selected (see Figure 1). The *mev* file contains integrated intensities and median intensities. You can select the Median or Integrated to place the either of them immediately after the UID. Checking the Subtract Background will remove the background from the intensities.

When you load the input files to the file window, the output file names will be the same as the input files and showed in the output file window. You can delete the selected file by clicking icon . If you want to change the output file names and paths, clicking icon  will open an edit window for the edition.

1.2.1 GenePix Interface

Other than the file window showing you the GenePix file (*gpr*) as input and the *tav/mev* files as output, there are selections at the bottom of the window (Figure 2). Un-checking the Sort by Slide Row sand Slide Columns will keep the order of data items (spots) in the *tav/mev* files as the same in the *gpr* files. If you want to attach all information in *gpr* file that is not used in the *tav* file, check the Keep All Information in the GenePix Files. This function is not applicable for converting *mev* file. The annotation information in *gpr* file will be saved to an annotation file with *.ann* as file extension automatically.

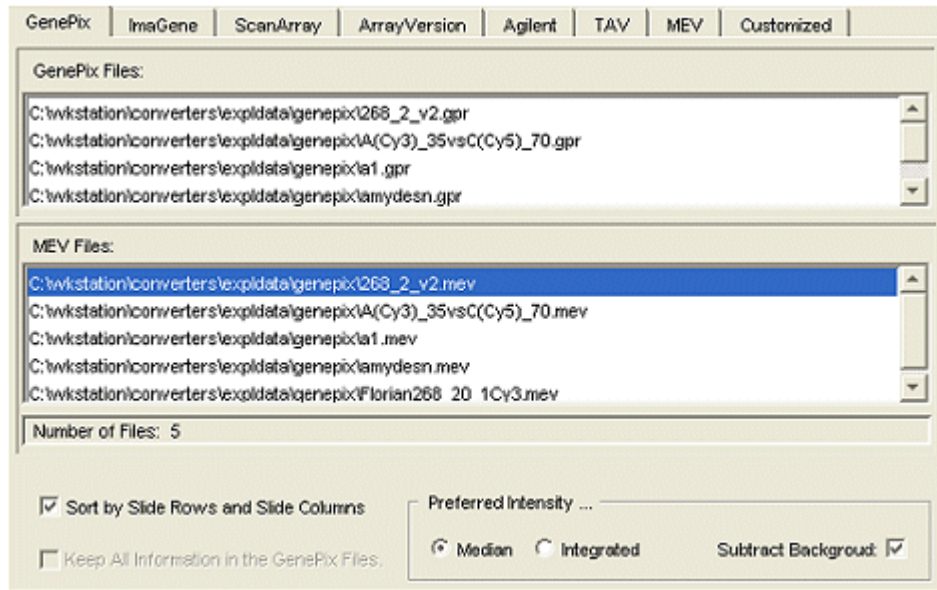


Figure 2. The GenePix interface showing the selections at the bottom.

1.2.2 ImaGene Interface

Converting the ImaGene files requires a pair of input files, each representing a channel. Thus, the input window contains two parts for Channel A and B respectively. You can load the input files to each Channel individually or to a channel and freely move them within or between the channel windows by clicking the four arrow buttons.

The two check boxes at the bottom allow you to sort the spots and attach the gene id in ImaGene files to the end of spots in the *tav* files. Un-checking the Sort by Slide Row and Slide Columns will keep the order of data items (spots) in the *tav/mev* files as the same in the *ImaGene* files. If you want to manually examine the converted *tav* file, you can check the Keep Gene ID for Manually Checking to attach the Gene ID in the input file at the end of each row of *tav* file.

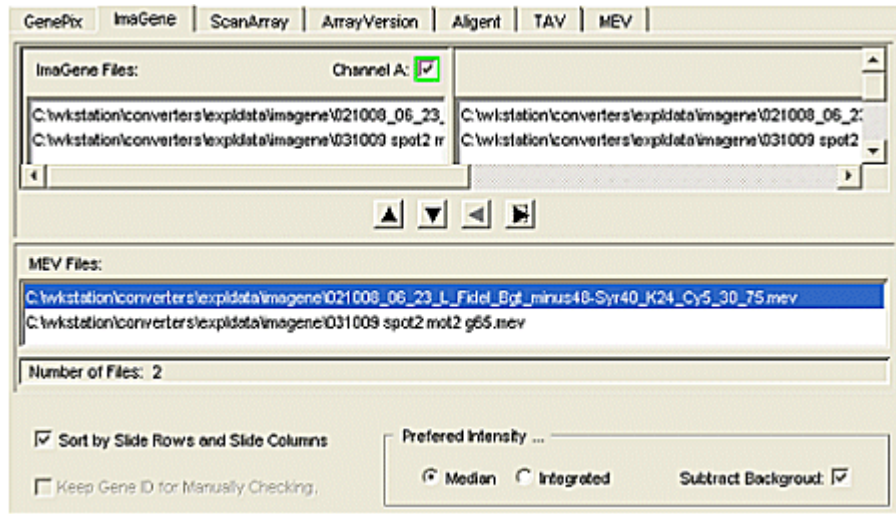


Figure 3. The ImaGene Interface

1.2.3 ScanArray Interface

Other than the file window showing you the ScanArray file (*txt*) as input and the *tav/mev* files as output, there are two check boxes. Un-checking the Sort by Slide Row and Slide Columns will keep the order of data items (spots) in the *tav/mev* files as the same in the **ScanArray** files. If you want to attach all information in **ScanArray** file that is not used in the *tav* file, check the Keep All Information in the ScanArray Files. This function is not applicable for converting *mev* file. The annotation information in *txt* file will be saved to an annotation file with *.ann* as file extension automatically.

The *mev* file always has IA with Cy3, but ScanArray file does not. It could have its channel 1 with Cy5. In this case, the Swap Channel check box should be checked to ensure *mev* file to have IA with Cy3.

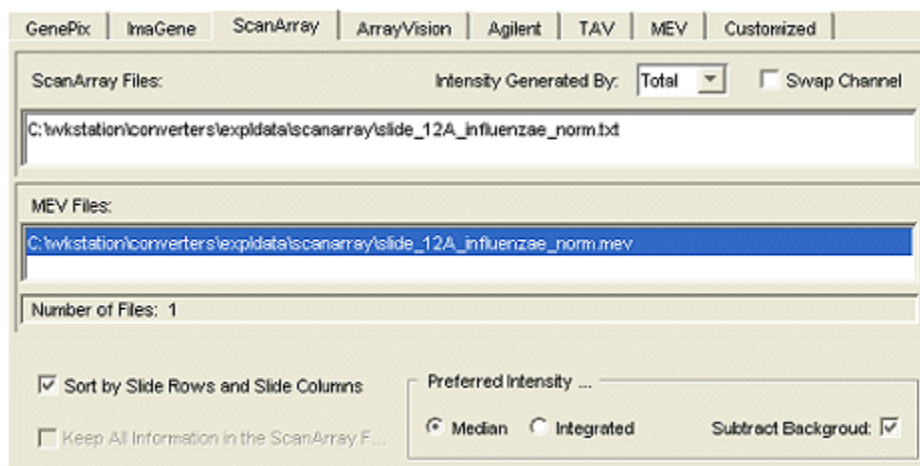


Figure 4. ScanArray Interface

1.2.4 ArrayVision Interface

The interface is similar to GenePix Interface (Figure 2), but the Sort by Slide Row and Slide Columns check-box is not applicable due to none of slide-row and slide-column information in the ArrayVision files.

1.2.5 Agilent Interface

The interface is similar to GenePix Interface (Figure 2). Other than the file window showing you the Agilent file (*txt*) as input and the *mev* files as output, there is a checkbox Sort by Slide Row and Slide Columns. Unchecking it will keep the order of data items (spots) in the *mev* files as the same in the *txt* files.

1.2.6 Tav Interface

As others, Tav interface has the input and output windows that show you the TAV file (*tav*) and the *mev* files respectively. At the bottom there is a check box, Sort by Slide Row and Slide Columns, which is defaulted checked. Unchecking it will keep the order of data items (spots) in the *mev* files as the same in the **TAV** files.

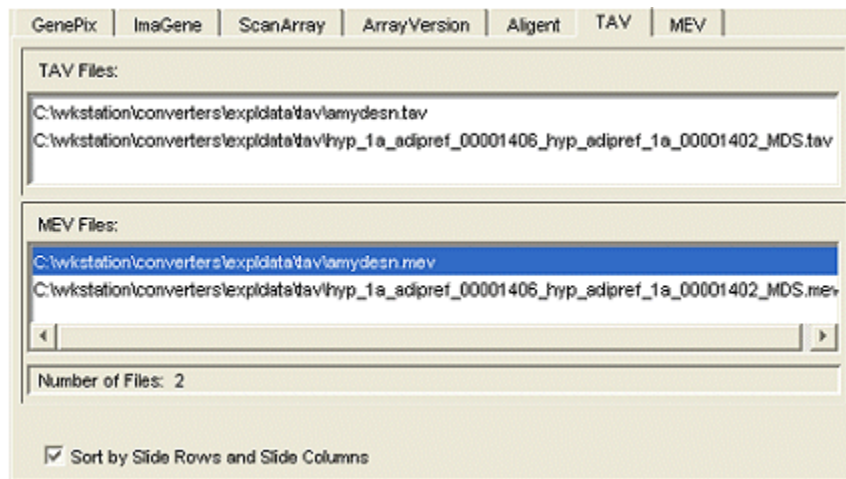


Figure 5. The Tav Interface.

1.2.7 Mev Interface

MEV interface is the same as the TAV interface except that there is one more check box, Keep Unique ID in the MEV File(s). Checking it will attach the unique id at the end of each spot in the *tav* files.

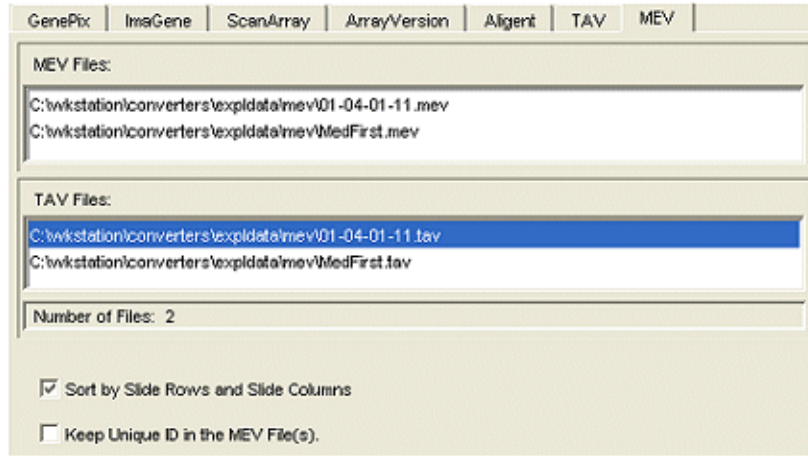


Figure 6. The Mev interface

1.2.8 Customized Interface

Different from other interfaces, the Customized has a header section, consisting of mev and input headers (Fig.7), that allows you to define which column of the input file is corresponding to which column of the *mev* file. Using the icons at the top of the headers, you can further set a formula for a *mev* column. At the bottom of the headers are input and output file windows.

The check box at the bottom gives you an option to sort the spots if you define the mev R and C columns. Un-checking the Sort by Slide Row sand Slide Columns will keep the order of data items (spots) in the *mev* files as the same in your input files.

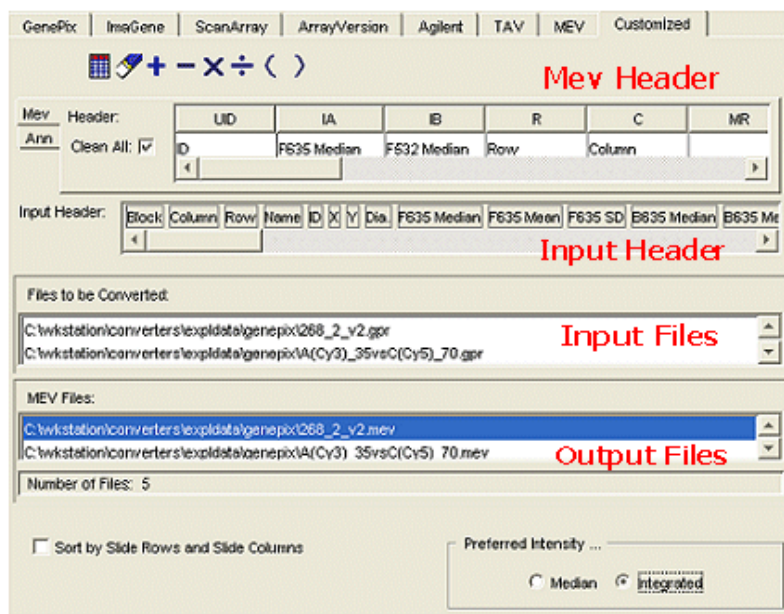



Figure 7. The Customized interface

1.2.9 Header Choose Window

The Header Choose Window (Fig. 8) shows you the first section of your input file so that you can select the header. If you cannot find the header after scrolling the window, it suggests that the buffer of the window (the number of lines of the input file) is small. You can increase the buffer by setting the number through the Preference Window (Fig. 9).

Also, the Header Choose Window can be opened by clicking the icon . It always shows the first file in the input file list.

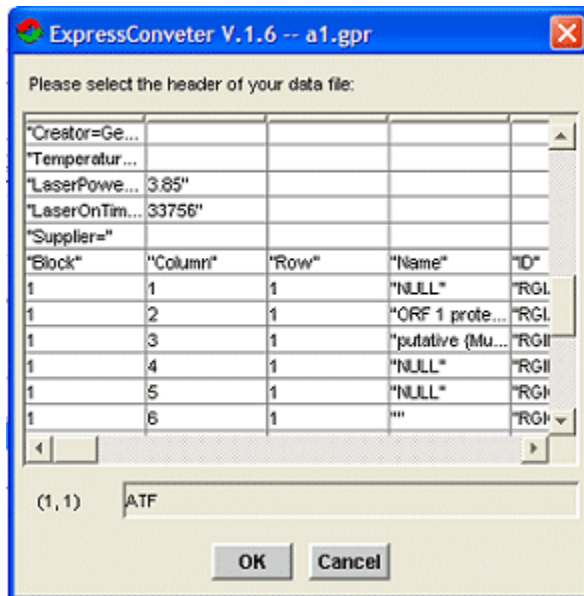


Figure 8. Header choose window.

1.2.10 Preference Window

The Preference Window (Fig. 9) allows you to add the file type of your input file to the list and to extend the buffer (number of lines) of the Header Chooser Window. By clicking the buttons on the right, you can add, edit, or remove the input file name extension (file type) to or from the list. To extend the buffer, just increase the Preview Lines.

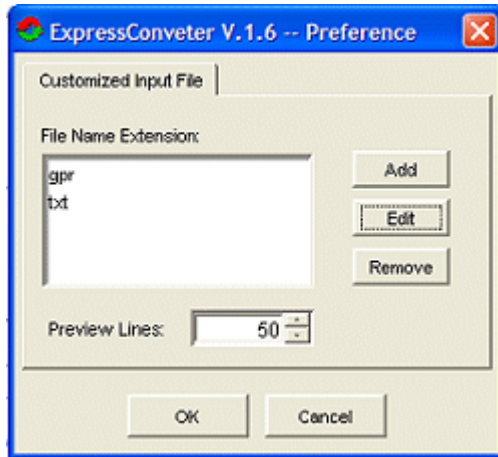






Figure 9. Preference window.


The ExpressConverter can take seven types of data formats as input. You can click the tabs for GenePix, ImaGene, ScanArray, ArrayVision, Agilent, TAV, and MEV types. If your data file is not in these seven types, click the Customized tab to define your converting algorithm. Although different types of inputs require different procedures, the common steps are.


- 1) Select the Output on the Tool Bar to define what type of file formats you want to produce. The Working Place will change based on your selection. You can select the Output at any time before starting the conversion. This function is not applicable for converting ArrayVision and Agilent files.
- 2) Click the "Select input file ..." from the File menu or the icon  to open a file choose window.
- 3) Select one or more input files. You can press and hold the Ctrl key to select multiple files. Click OK button to close the file choose window.
- 4) Check suitable check-box in the "Preferred Intensity ..." to indicate what kind of intensity will be placed immediately after UID in *mev* files.
- 5) If you do not want the background be subtracted, uncheck the Subtract Background radio button.
- 6) Click Start converting from the File menu or the icon  to process the conversion.

1. Convert GPR Files



- 1) Click GenePix tab.
- 2) Select the Output on the Tool Bar to define what type of file formats you want to produce.
- 3) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 4) Select one or more *gpr* files you want to convert. You can press and hold Ctrl key to select multiple files.
- 5) Click OK button to close the file choose window.
- 6) If you want the genes (spots) in the output files have the same order as your *gpr* files, uncheck the Sort by Slide Rows and Slide Columns.
- 7) Check suitable parameters for the "Preferred Intensity ...".
- 8) Click Start converting from the File menu or the icon  to process the conversion.

2. Convert ImaGene Files

- 1) Click ImaGene tab.
- 2) Select the Output on the Tool Bar to define what type of file formats you want to produce.
- 3) Check Channel A in the input window for selecting channel A data files.
- 4) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 5) Select one or more ImaGene files you want to convert. You can press and hold Ctrl key to select multiple files.
- 6) Click OK button to close the file choose window.



- 7) Check Channel B for selecting channel B data files and repeat steps 3-6. Alternatively, you can load all files (both channels) in one channel window and move them with following steps.
- 8) Click one of the file in Channel A window to check if the highlighted files are the pair of the files representing the two channels. If not correct, select the file and click the arrow buttons to move the file within or between the two channel windows. Checking one of the channel boxes indicates that you are manipulating the channel files.
- 9) If you want the genes (spots) in the output files have the same order as your ImaGene files, uncheck the Sort by Slide Rows and Slide Columns.
- 10) Check suitable parameters for the "Preferred Intensity ...".
- 11) Click Start converting from the File menu or the icon  to process the conversion.

3. Convert ScanArray Files

- 1) Click ScanArray tab.
- 2) Select the Output on the Tool Bar to define what type of file formats you want to produce.
- 3) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 4) Select one or more ScanArray files (*txt*) you want to convert. You can press and hold Ctrl key to select multiple files.
- 5) Click OK button to close the file choose window.
- 6) Select a method from the Intensity Generated By to indicate how your ScanArray files were produced.
- 7) If your input files have channel 1 with Cy5, you need to check the Swap Channel box to ensure the converted *mev* files to have IA with Cy3 and IB Cy5; otherwise, uncheck the box.
- 8) If you want the genes (spots) in the output files have the same order as your ScanArray files, uncheck the Sort by Slide Rows and Slide Columns.
- 9) Check suitable parameters for the "Preferred Intensity ...".
- 10) Click Start converting from the File menu or the icon  to process the conversion.



4. Convert ArrayVision Files

Only *mev* file can be converted as output.



- 1) Click ArrayVision tab.
- 2) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 3) Select one or more ArrayVision files (*txt*) you want to convert. You can press and hold Ctrl key to select multiple files.
- 4) Click OK button to close the file choose window.
- 5) Check suitable parameters for the "Preferred Intensity ...".
- 6) Click Start converting from the File menu or the icon  to process the conversion.

5. Convert Agilent Files



Only *mev* file can be converted as output.

- 1) Click Agilent tab.
- 2) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 3) Select one or more Agilent files (*txt*) you want to convert. You can press and hold the Ctrl key to select multiple files.
- 4) Click OK button to close the file choose window.
- 5) If you want the genes (spots) in the output files have the same order as your Agilent files, uncheck the Sort by Slide Rows and Slide Columns.
- 6) Check suitable parameters for the "Preferred Intensity ...".
- 7) Click Start converting from the File menu or the icon  to process the conversion.

6. Convert TAV Files

- 1) Click TAV tab.
- 2) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 3) Select one or more *tav* files you want to convert. You can press and hold the Ctrl key to select multiple files.
- 4) Click OK button to close the file choose window.
- 5) If you want the genes (spots) in the output files have the same order as your *tav* files, uncheck the Sort by Slide Rows and Slide Columns.
- 6) Click Start converting from the File menu or the icon  to process the conversion.





7. convert MEV Files

- 1) Click MEV tab.
- 2) Click the "Select input file ..." from the File menu or the icon  to open a file chooser window.
- 3) Select one or more *mev* files you want to convert. You can press and hold the Ctrl key to select multiple files.
- 4) Click OK button to close the file choose window.
- 5) If you want the genes (spots) in the output files have the same order as your *mev* files, uncheck the Sort by Slide Rows and Slide Columns.
- 6) If you do not want the UID to be attached at the end of *tav* file, uncheck the Keep UID in the MEV File.
- 7) Click Start converting from the File menu or the icon  to process the conversion.

8. convert Customized Files

Only *mev* file can be converted as output. If you are converting a set of files, the files are supposed to have the identical headers and lines of comments before the headers.

- 1) Click Customized tab.

- 2) Click "Select input file ..." from the File menu or the icon  to open a file chooser window. If you cannot find your files in the window, you need to add your input file type to the Files of Type list.
- 3) Select one or more files you want to convert. You can press and hold the Ctrl key to select multiple files.
- 4) Click OK button to have a header choose window. This window shows you the beginning of the first input file (if you select a set of files) as a table format.
- 5) Scroll the table to find the header and select it. If you cannot find the header, you need to extend the window buffer to hold more lines of the input file.
- 6) Click OK button to close the window. You will see the selected headers are showed in the Input Header
- 7) Select one of the *mev* columns by clicking it, then drag one of your input headers and drop it onto the *mev* column. This suggests that the selected input column is corresponding to the selected *mev* column. You can set a formula by dragging and dropping the math icons to the column.
- 8) Repeat the above step to define other *mev* columns. You do not need to define all of them, but IA and IB are required. Only they are defined, is the icon  activated.
- 9) Click Ann tab to define the columns for *ann* file.
- 10) If you make a mistake during defining *mev* the column, click icon  to clean it. You can save your header setting to a file. Also, you can change the input headers.
- 11) Check suitable parameters for the "Preferred Intensity ...".
- 12) Click Start converting from the File menu or the icon  to process the conversion.

8.1 Set Customer Parameters


If you cannot find the file name when you load the input file for the Customized conversion, you might need to add the file name extension to the file type list with preference window. This window also allows you to extend the buffer of the Header Choose Window.

- 1) Select the "Preference ..." from File menu to open the Preference Window (Fig. 9).
- 2) Click Add, Edit, or Remove button to add, edit, or remove the file type (file extension) of your input file to or from the File Name Extension list.
- 3) Increase the number of Preview Lines to extend the buffer of the Header Choose Window (Fig. 8).
- 4) Click OK button to close the window.

8.2 Save Defined Header



After setting up what columns of your input files should be converted to *mev* columns, you can save the defined columns into a file (header file). When you convert the same type of input files later, you can load the

header file into the header fields, instead of setting them again. To save the file:

- 1) Select the "Save header file ..." from File menu or click the icon  to open a file choose window.
- 2) Select a directory to which the file will be saved to.
- 3) Provide a file name (You do not need to give the file name extension).
- 4) Click OK button to close the window and the headers are saved.


8.3 Change Input Headers

If you make a mistake during selecting input file headers, you can re-open the Header Choose Window to select the header.


- 1) Click the icon  to open the Header Choose Window (Fig.8). The window will show the first section of the first file in the input file list. You can view the file name on the window title bar.
- 2) Select the header line by clicking it. If you cannot find the header after scrolling the table, you need to extend the window buffer to hold more lines of the file.
- 3) Click OK button to close the window.
- 4) Check the Clean All and click icon  to remove all pre-defined *mev* and *ann* columns.
- 5) Re-set the columns by dragging and dropping the input headers.

8.4 Open a Header File

If you are converting the same format of the input files as those previously converted and had the headers saved, you can load the header directly instead of re-defining them.


- 1) Select the "Open header file ..." from File menu or click the icon  to open a file choose window.
- 2) Find the header file name by browsing your system and select the name.
- 3) Click OK button to close the window and the header fields in the working place will be filled. It is ready to convert the files.

9. Delete the Selected Files

- 1) Select those input files you want to delete from the input file list. You cannot delete the *tav/mev* files. When you delete the input file, the corresponding *tav/mev* file will be removed.
- 2) Click Delete Files from the Edit menu or the icon .


10.Edit Output File Names

When you select the Output on the Tool Bar, the file extensions in the output file list will be changed to an appropriate one (*tav/mev*).

- 1) Select a *tav/mev* file from the output files list. You can edit only one name each time.
- 2) Click Change File Name from the Edit menu or the icon  to open an edit window.
- 3) Type new name and click the OK button to close the edit window.

11. Change Directory for Output Files

By default, the output files (*tav/mev*) are saved at the same directory as your input files. Though you can change the directories for one or some of the selected *tav/mev* files through Edit Output File Names, you are able to place all of the output files in a directory by following steps.

- 1) Click Change Directory from the Edit menu or the icon  to open a file choose window.
- 2) Select a directory. You can create a new directory with the file choose window.
- 3) Click the OK button to close the window and you will see the paths of the output files are changed.

1. TAV File

TIGR ArrayView needs certain type of file as its input. The type of file is called *tav* file format that contains a number of columns of data. Now the *tav* format is accessible by MADAM, MIDAS, and MEV. Although the number of the columns is variable, the first eight columns of data are constantly defined. This version of ExpressConverter generates 17 columns of data described as following:

- (1) slide_row
- (2) slide_column
- (3) meta_row.
- (4) meta_column.
- (5) sub_row.
- (6) sub_column.
- (7) intensity of channel A (Cy3).
- (8) intensity of channel B (Cy5).
- (9) spot area in number of pixels for channel A.
- (10) spot area in number of pixels for channel B.
- (11) *dummy* with 'null'.
- (12) *dummy* with 'null'.
- (13) *dummy* with 'null'.
- (14) background intensity of channel A.
- (15) background intensity of channel B.
- (16) flag of channel A.
- (17) flag of channel B.

2. MEV File

MEV file is a tab-delimited text file that contains coordinate and expression data for a single microarray experiment. The file can be divided into three parts: header, comments, and data. The header precedes the data in order to identify the columns below; the comments start with a '#' sign in each line and provides information about the file (i.e. how the file is generated). ExpressConverter generates the comments and header as following:

Comments:

version:	version number of MEV Generator in the Converter.
format_version:	version number of mev file format.
date:	date of file creation.
created by:	converter and its version number.
row_count:	number of rows of data items (genes).
converted from:	the file from which the <i>mev</i> file is generated.
background_correction:	ON if background subtracted.

Header Names (definition of columns)

UID:	Unique identifier for this spot.
IA:	Integrated intensity for channel 1 (Cy3).
IB:	Integrated intensity for channel 2 (Cy5).
R:	Row (slide_row).

C:	Column (slide column).
MR:	Meta-row (block row).
MC:	Meta-column (block column).
SR:	Sub-row (row in block).
SC:	Sub-column (column in block).
FlagA:	TIGR SpotFinder flag value for channel 1.
FlagB:	TIGR SpotFinder flag value for channel 2.
BkgA:	Background intensity for channel 1.
BkgB:	Background intensity for channel 2.
SAA:	Spot area for channel 1.
SAB:	Spot area for channel 2.
MedA:	Median Intensity for channel 1 (Cy3).
MedB:	Median Intensity for channel 2 (Cy5).

In the data section, all data items are arranged based on the column definition for each spot. The format of UID varies based on the type of inputs. The Customized page generates more columns. Besides the above columns, they are:

QC:	Combined quality control score from channels A and B.
QCA:	Quality control score in channel A.
QCB:	Quality control score in channel B.
SDA:	Standard deviation for spot pixels in channel A.
SDB:	Standard deviation for spot pixels in channel B.

3. Annotation File

An annotation file is a tab-delimited text file containing annotation data for a *mev* files. The ExpressConverter extracts the useful information from an input file and converts it to *mev* format. At the same time, it writes the annotation information to an annotation file. The annotation file name is automatically produced using the same name of the input file with different extension *.ann*. The association between the *mev* and the annotation files is set up by the unique id defined in the *mev* file. The comment of the *.ann* file is the same as that in *mev* file and the first three columns are:

UID:	Unique identifier for this spot.
R:	Row (slide_row).
C:	Column (slide column).

If you have a set of files that were generated with the same array design method, the annotation of each spot for the files should be the same. Thus, you can produce only one annotation file for the set of *mev* files by checking the Many to One under the Ann File menu.

Mev file contains both integrated and median intensities. Selecting Median or Integrated will place either of them immediately after UID. By default, the background is subtracted and they are calculated with following formulas:

1. GenePix to Mev

Integrated Intensity = $F_{\text{Pix}} \times (F_{\text{xxx}} \text{ Median} - B_{\text{xxx}} \text{ Median}) \times (1 - F_{\text{xxx}} \% \text{ Sat})$

Median Intensity = $F_{\text{xxx}} \text{ Median} - B_{\text{xxx}} \text{ Median}$

UID = consequential number during processing

where, xxx is the wave length.

2. ImaGene to Mev

Integrated Intensity = $(\text{Signal Median} - \text{Background Median}) \times \text{Signal Area}$

Median Intensity = $\text{Signal Median} - \text{Background Median}$

UID = Gene ID

3. ScanArray to Mev

If a ScanArray file is generated with the **Total** method,

Integrated Intensity = $(\text{ch}_x \text{ Intensity} - \text{ch}_x \text{ Background}) \times \text{ch}_x \text{ Area}$

Median Intensity = $\text{ch}_x \text{ Intensity} - \text{ch}_x \text{ Background}$

UID = consequential number during processing

otherwise,

Integrated Intensity = $\text{ch}_x \text{ Intensity} - \text{ch}_x \text{ Background}$

There is no Median Intensity converted.

where, x is the channel variable, 1 or 2.

4. ArrayVision to Mev

Integrated Intensity = $\text{VOL} - \text{Bkgd}$

Median Intensity = $(\text{VOL} - \text{Bkgd}) / \text{Area}$

UID = Spot labels

5. Agilent to Mev

Integrated Intensity = $({}_x \text{ MedianSignal} - {}_x \text{ BGMedianSignal}) \times {}_x \text{ NumPix}$

Median Intensity = ${}_x \text{ MedianSignal} - {}_x \text{ BGMedianSignal}$

UID = FeatureNum

where, x is the channel variable, r or g .