applied biosystems

Expression Console™ Software v1.4

USER GUIDE

Publication Number MAN0017729

Revision A.0

For Research Use Only. Not for use in diagnostic procedures.

ThermoFisher SCIENTIFIC

Manufacturer: Affymetrix Inc. | 3450 Central Expressway | Santa Clara, CA 95051 | USA

The information in this guide is subject to change without notice.

DISCLAIMER: TO THE EXTENT ALLOWED BY LAW, THERMO FISHER SCIENTIFIC INC. AND/OR ITS AFFILIATE(S) WILL NOT BE LIABLE FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING YOUR USE OF IT.

Revision history: Pub. No. MAN0017729

Revision	Date	Description
A.0	10 May 2018	Initial release in Thermo Fisher Scientific document control system. Supersedes legacy Affymetrix publication number 702387. Updated to the current document template, with associated updates to trademarks, logos, licensing, and warranty.

Contents

1 Welcome

Welcome to the Applied Biosystems™ Expression Console™ Software (v.1.4) User Guide. The software provides signal estimation and QC functionality for the GeneChip™ Expression Arrays.

The Expression Console™ Software allows users to:

- Generate probe set summarization (CHP) files from feature intensity (CEL) files for:
  - 3' Expression Arrays
  - Gene Arrays
  - Exon Arrays
  - miRNA Arrays
  - Human Transcriptome Arrays
- Capture a standard set of metrics for evaluating the success of the individual hybridizations for Expression Arrays
- Identify outlier samples in the data set

The Expression Console™ Software is targeted for research personnel (such as laboratory technicians, research associates, and scientists) analyzing GeneChip™ data.

Introduction

The Expression Console™ Software provides an easy way to create summarized expression values (CHP files) for individual files or collections of Expression Array feature intensity (CEL) files. In addition to CHP writing, the Expression Console™ Software application also produces a collection of QC metrics for evaluating the success of hybridizations. The user defines thresholds for these metrics and the software highlights the metrics that do not meet the defined thresholds. In addition, the CHP files are highlighted in the study table, if they have any metrics outside of the defined thresholds. Individual QC metrics for each labeling technique are discussed in Chapter 9, “Viewing QC and analysis results data“.

The Expression Console™ Software application contains graphic capabilities for visual inspection of the hybridization results. To identify outliers, the application is designed to display:

- Line graphs for individual or collections of metrics or probe sets
- Box plots for signal distributions before or after normalization
- MvA plots for signal distributions
- Heat maps for correlation matrices

Note: The Expression Console™ Software application is not a secondary analysis package. However, it does create the CHP files required for Transcriptome Analysis Console (TAC), and secondary analysis packages from the GeneChip™ Compatible Program.

New features in Expression Console™ Software v1.4

New features in this version of Expression Console™ Software v1.4 include:

- New graphs:
  - Signal (CHP) scatter plots and signal scatter plot thumbnails
  - Probe cell intensity histogram
  - Probe intensity thumbnails
  - Principle component analysis of probe cell intensity (CEL files) or signal (CHP files)
- The Sample Attribute Editor which provides tools for adding or editing sample attributes
- Access to the Transcriptome Analysis Console from within Expression Console™ Software

Workflow diagram

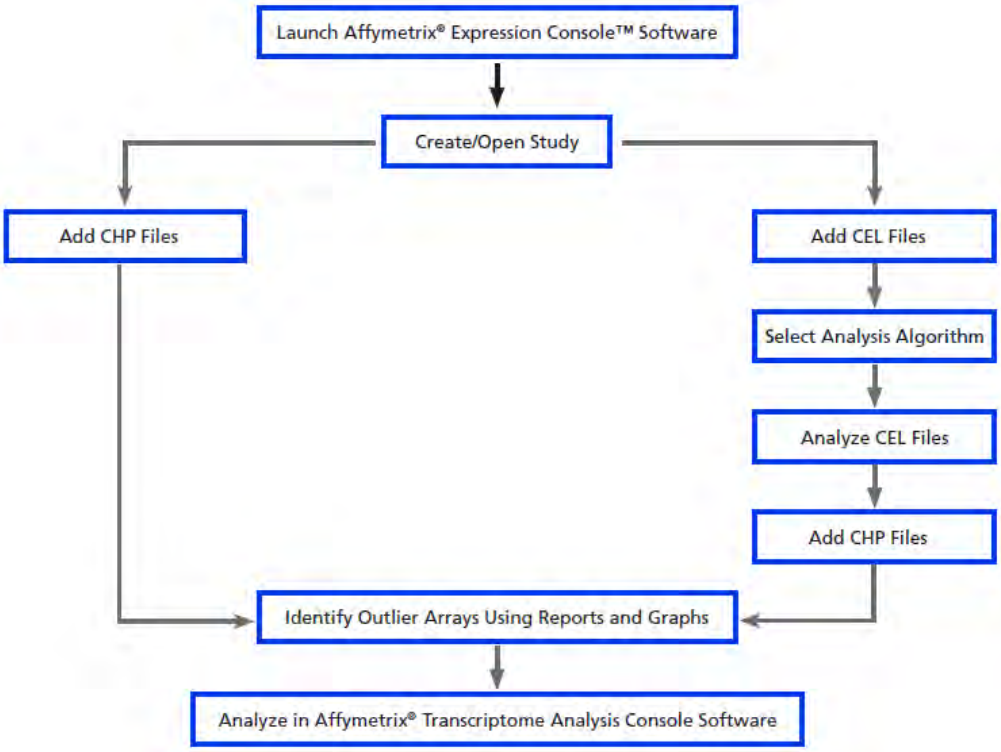


Figure 1 Expression Console™ Software workflow.

Note: Signal (CHP files) can also be analyzed using GeneChip™ Compatible Partner software.

FAQS

A list of Frequently Asked Questions (FAQS) about the Expression Console™ Software can be found on our **website**.

2 Installation and setup

The Expression Console™ Software is a stand-alone application. It can be installed on computers that have GeneChip™ Operating System (GCOS) software, GeneChip™ Command Console™ Software, or neither.

Software and hardware requirements

The table below (Table 1) show the operating systems that the Expression Console™ Software is supported on and the recommended minimum requirements. The larger data file size associated with Exon Array data should be taken into account when calculating the necessary available disk space requirement.

Table 1 Supported 64-bit operating systems and recommended requirements for the Expression Console™ Software.

64-bit Operating System	Speed	Memory (RAM)	Available Disk Space	Web Browser
Microsoft™ 7 (64 bit) Professional with Service Pack 1	2.83 GHz Intel™ Pentium®Quad Core Processor	16 GB RAM	150 GB HD + data storage	IE 8.0 and above
Microsoft™ 8.1 (64 bit) Professional	2.83 GHz Intel™ Pentium® Quad Core Processor	16 GB RAM	150 GB HD + data storage	IE 8.0 and above

Installation instructions

To install the Expression Console™ Software:

1. Go to our web site and download the software.
2. Unzip the downloaded software package.
3. Double-click ECx64.exe to install the software.
4. Follow the directions provided by the installer.

Note: If there is a previous version of Expression Console™ Software installed, the installation software prompts the user to remove it before installing the new version.

The software needs to be configured before it can be used. Configuration requires:

- "Create a profile" on page 13
- "Using library files" on page 17
- "Using annotation files" on page 21

Setting up a profile

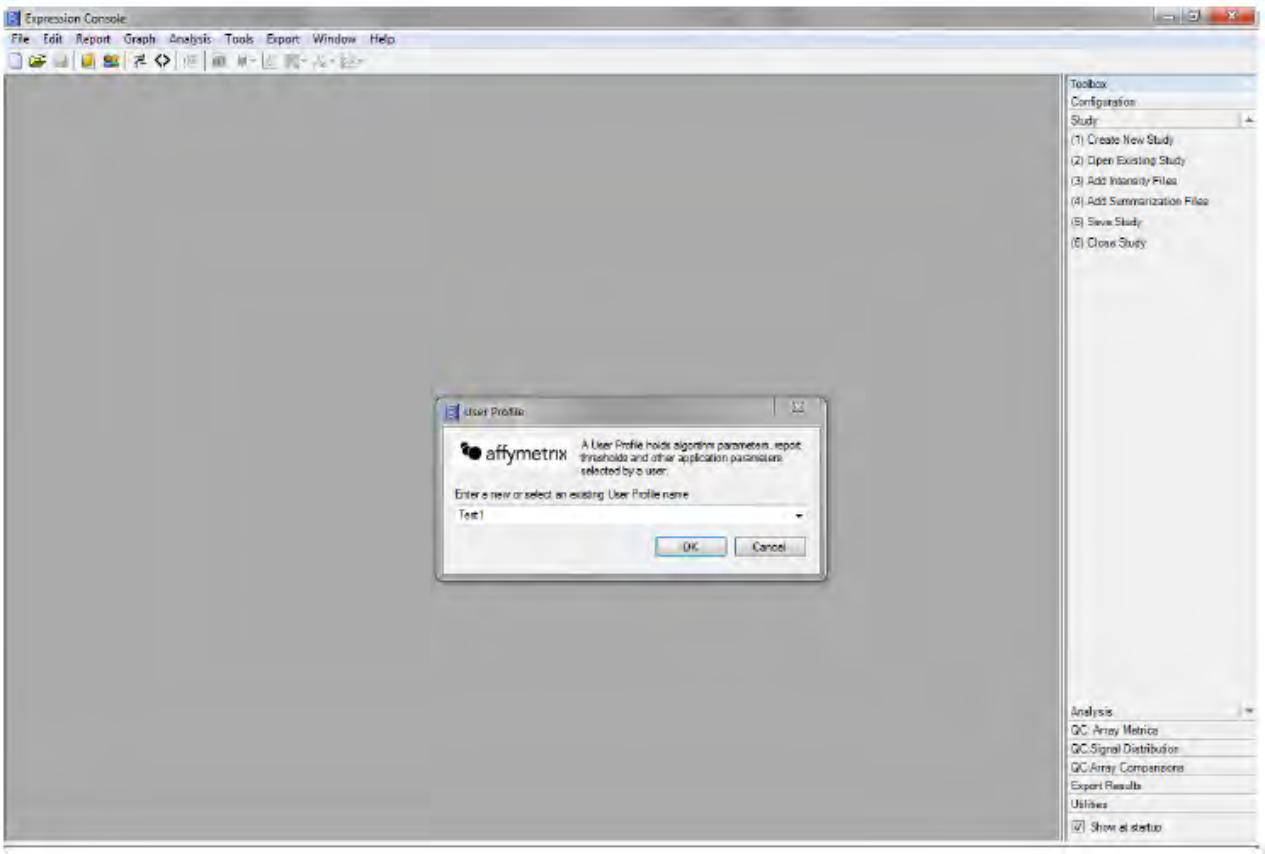
A profile is used as a method to group options and parameters so that those parameters can be used again. The Expression Console™ Software provides tools for:

- "Create a profile" on page 13
- "Change or add a user profile" on page 13
- "Delete profiles no longer needed" on page 14

Create a profile

1. Open the Expression Console™ Software application by selecting **Start ▸ Programs ▸ Affymetrix ▸ Expression Console™**.

   The Expression Console™ Software window opens with the User Profile dialog box displayed.



2. Type in a name for your profile and click OK.

   By entering a profile name, you retain the specific analysis settings and quality control thresholds for each profile entered.

A list of previously created profiles, if any, are found in the drop-down menu on the Profile Information dialog box.

Note: You can select a different profile without terminating the program, but the current study must be closed to open another profile

Change or add a user profile

1. Do one of the following:
   - Click the Select user profile button.
   - From the Edit menu, Select Change User Profile.
   - In the Toolbox, select **Configuration ▸ Specify User Profile**.

The User Profile dialog box opens. (Figure 2)

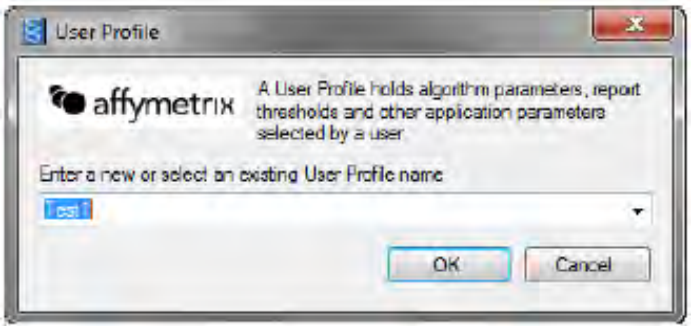


Figure 2 User Profile.

2. Select a profile name from the drop-down list, or enter a new profile name.
3. Click OK.
   The selected or created user profile is used by the software.

Delete profiles no longer needed

The list of previously created profiles are found in the drop-down menu on the Profile Information dialog box.

1. Close the study window, if open.
2. Select **File ▸ Utilities**.
3. Select the User Profile Management tab (Figure 3)

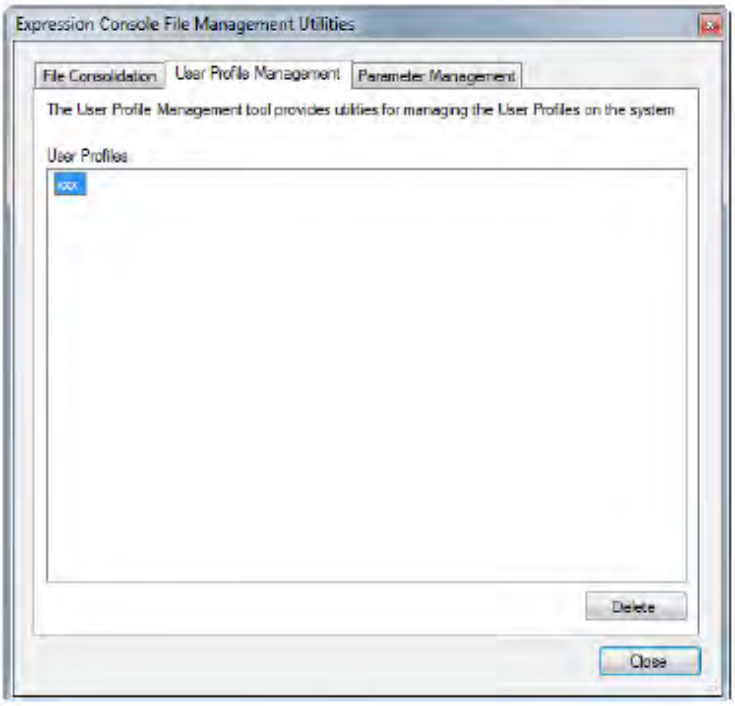


Figure 3 User Profile Management tab.

4. Highlight the profile to be removed.
5. Click Delete.

Parts of the Expression Console™ Software window

After creating or selecting the user profile, the Expression Console™ Software opens (Figure 4). The components of the software interface are introduced below.

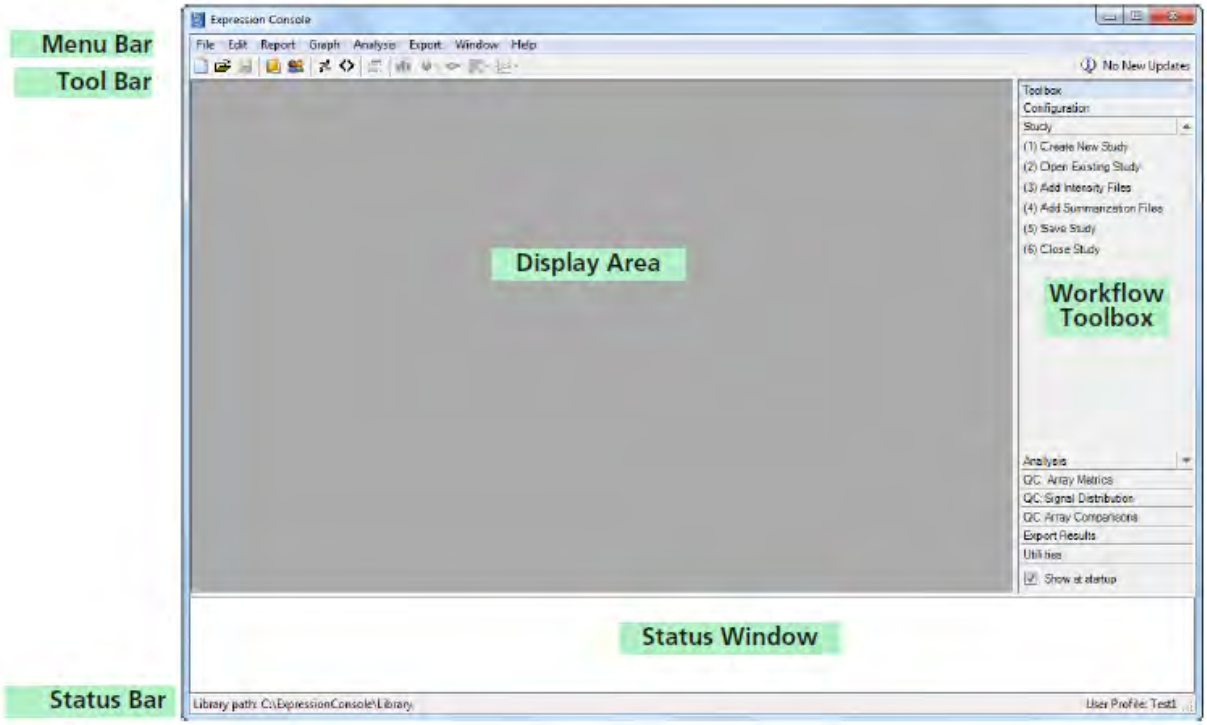


Figure 4 Expression Console™ Software window.

Menu bar and tool bar

The menu bar and tool bar provide quick access to the EC functions.

Display area

The display area is where the following Expression Console™ Software components are displayed:

- Study dialog box, used to select data for analysis and display.
  See "The Study dialog box" on page 25.
- QC Tables and Graphs, used to review the quality of the data and analysis.
  See Chapter 9, "Viewing QC and analysis results data".
- Analysis Results tables, with the probe summarization data.
  See "Displaying analysis results" on page 84.

Workflow toolbox

The Workflow Toolbox (Figure 5) provides a convenient way to access the functions of the Expression Console™ Software.

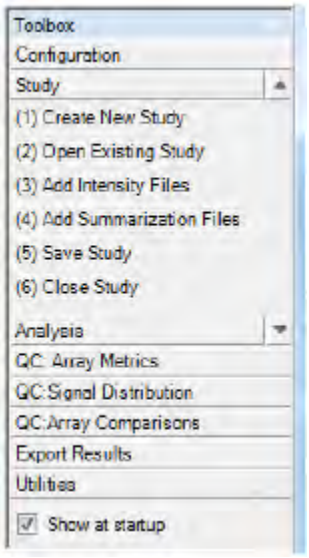


Figure 5 Workflow toolbox.

To disable this view, go to the Window menu and select Hide Workflow Toolbox (Figure 6).

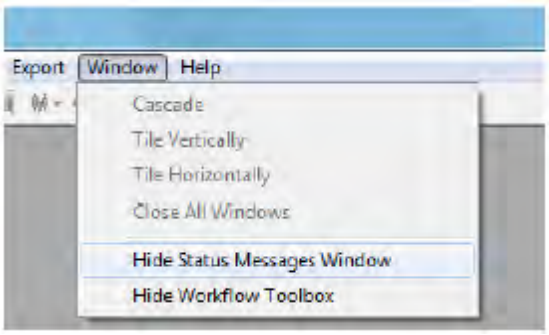


Figure 6 Enabling/disabling the workflow toolbox display.

Status window

The Status window displays all status and algorithm progress information (Figure 7).

Figure 7 Status window.

To disable this view, go to the Window menu and select Hide Status Messages Window (Figure 8).

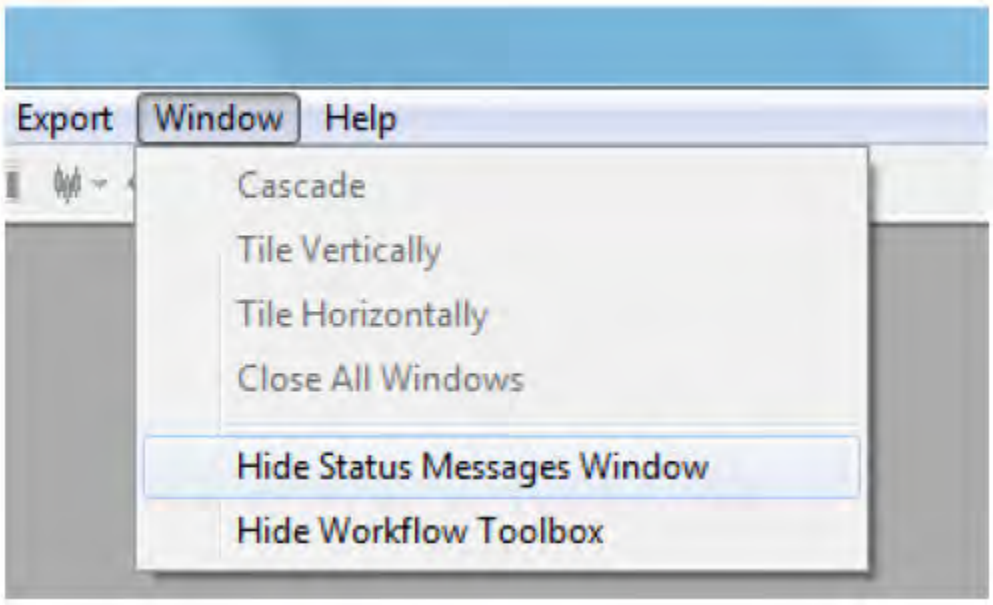


Figure 8 Enabling/disabling the status window display.

Status bar

The Status bar at the bottom of the Expression Console™ Software window (Figure 9) displays information on the path to library files and the user profile.

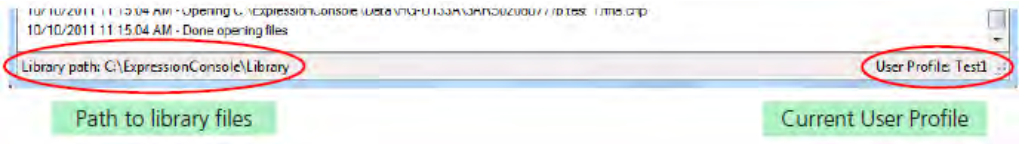


Figure 9 Status bar shows the library path and current user profile.

Using library files

The Expression Console™ software requires information stored in library files (array types) to analyze the CEL files generated by GCOS or GeneChip™ Command Console™ (AGCC) software. These files are available and can be downloaded within the Expression Console™ application.

Expression Console™ provides tools for:

- "Choosing the library files folder location" on page 18
- "Download library files" on page 19
- "Copy library files manually" on page 21

Choosing the library files folder location

When you click OK in the User Profile Information window the first time (Figure 10), a dialog box opens asking you to direct the software to your GeneChip™ library files folder (Figure 10). You can select any location for the library files folder.

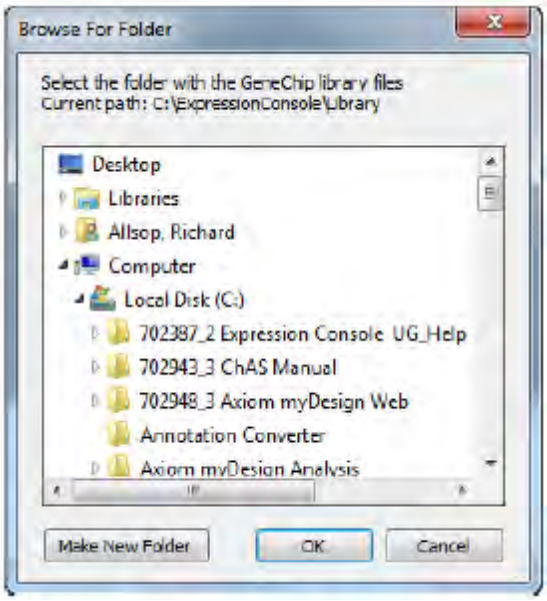


Figure 10 Browse for folder window to locate GeneChip™ library files folder.

Note: You can select any location for the library files folder. However, once you direct the software to the folder location, do not place any library files in a subfolder. The Expression Console™ Software application cannot find library files in a subfolder!

Note: If the GeneChip™ Command Console™ Software is installed on your system, the Expression Console™ Software application defaults to the files in C:\Command_Console\Library.

Note: Expression Console™ downloads library files from NetAffx™ Analysis Center, but these files are not registered with the Command Console™ Software and are not sufficient to scan arrays.

The Browse for Folder dialog box can also be opened to designate a new library folder:

- Select **Edit ▸ Set library path**.
  The Browse for Folder dialog box opens (Figure 10).

Download library files

Library files can be downloaded from the **NetAffx™ Analysis Center**.

1. Select **File ▸ Download Library Files**.
   A dialog box opens requesting your account information for the **NetAffx™ Analysis Center** (Figure 11).
   To obtain a **NetAffx™ Analysis Center** account, contact **thermofisher.com/support**.

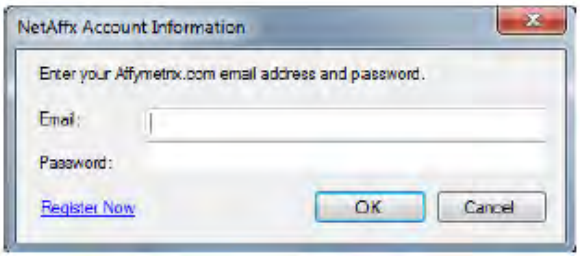


Figure 11 NetAffx™ Analysis Center Account Information dialog box.

2. Enter your registered email address and password.

   The NetAffx™ Library Files window opens (Figure 12). This window contains a complete list of library files that can be downloaded. Library files previously downloaded are marked currently installed.

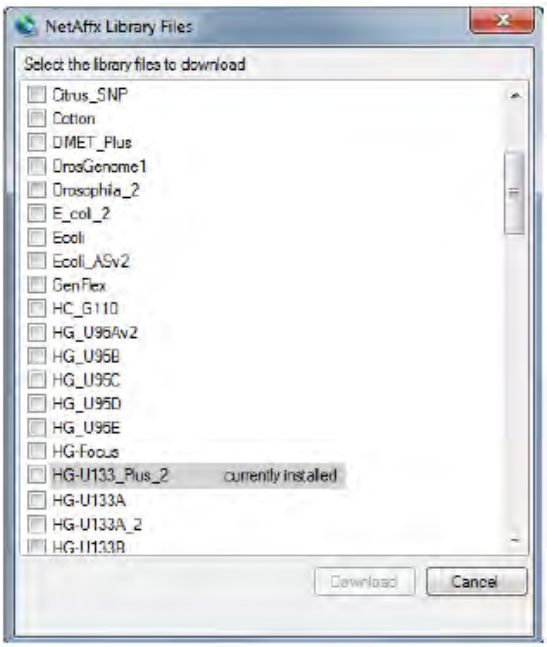


Figure 12 NetAffx™ Library Files window.

3. Check the library files needed and click the Download button.

   The appropriate files are downloaded to the folder you previously indicated. The green status bar at the bottom of the window highlights during the download process.

   If the Abort button is selected, the download process stops with a message indicating which library files failed to download.

Missing library files

When you create a new study and add CEL/CHP files that do not have corresponding library files loaded in the correct folder, you will be prompted to download the appropriate library files.

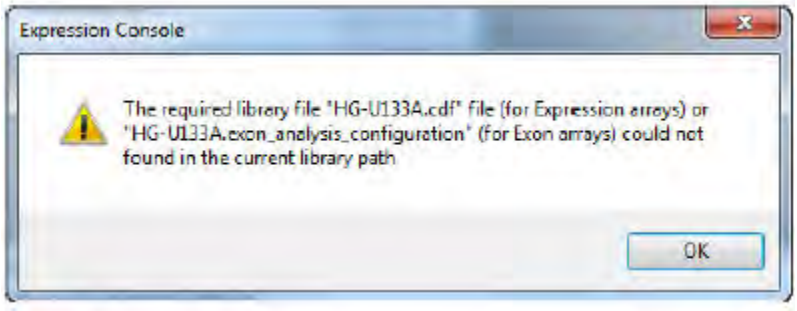


Figure 13 Notice of missing library files.

The status window at the bottom of the Expression Console™ Software application window lists the missing library files.

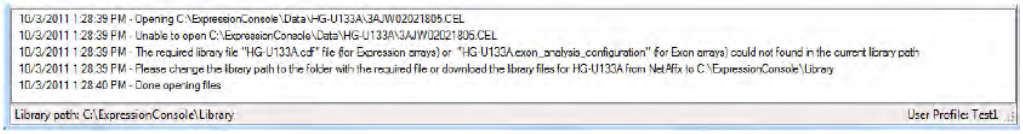


Figure 14 Status window - indicating missing library file.

For more information on changing the library file path and obtaining library files, see "Using library files" on page 17.

Copy library files manually

For computers that are not connected to the internet and therefore cannot take advantage of the library file download option, it is possible to manually copy the necessary files to the computer with the Expression Console™ Software:

1. Create a folder on the computer to hold the library files for the Expression Console™ Software application.
2. Copy the necessary files from the CD or other removable media to the library file folder. In order for the default report controls to be identified for the array type, this should be done with the application closed.
   - For 3' Expression Arrays, only the *.psi and *.cdf files need to be copied to the directory.
   - For Exon Arrays, the *.exon_analysis_configuration, *.ps, *.bgp, *.clf, *.mps, *.pgf, and *.qcc files need to be copied to the directory.
   - For Gene Arrays, the *.exon_analysis_configuration, *.ps, *.bgp, *.clf, *.mps, *.pgf, and *.qcc files need to be copied to the directory.
   - For miRNA Arrays, the *.mirna_analysis_configuration, *.ps, *.bgp, *.clf, *.pgf, and *.qcc files need to be copied to the directory.
   - For Human Transcriptome Arrays, the *.exon_analysis_configuration, *.ps, *.bgp, *.clf, *.mps, *.pgf, and *.qcc files need to be copied to the directory

Note: Do not create subdirectories within the library file folder. The Expression Console™ Software does not look at subdirectories.

Using annotation files

To aid analysis of the signal estimates produced by the Expression Console™ Software information contained in comma-separated files can be combined with the signal values into a single tab-delimited file. These annotation CSV files are available for commercial arrays and can be downloaded within the Expression Console™ Software application.

The annotation files will be stored in the default library file directory.

Note: You can select any location for the library files folder. However, once you direct the software to the folder location, do not place any library or annotation files in a subfolder. The Expression Console™ Software application cannot find files in a subfolder!

The Expression Console™ Software provides three options for obtaining annotation files:

- "Download annotation files" on page 22
- "Copy annotation files manually" on page 23
- "Create custom annotation files" on page 24

To merge the annotation data with the signal values, you will need to create a Merge file, as described in Exporting Analysis Results with Annotation Information.

Download annotation files

Annotation files can be downloaded from the NetAffx™ Analysis Center.

1. Select **File ▸ Download Annotation Files**.

   A dialog box (Figure 15) opens requesting your account information for the NetAffx™ Analysis Center.

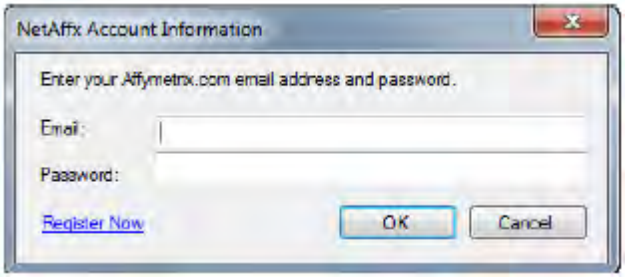


Figure 15 NetAffx™ Account Information dialog box.

To obtain a NetAffx™ Analysis Center account, please contact **thermofisher.com/support**.

2. Enter your registered email address and password.

   The NetAffx™ Library Files window opens (Figure 16). This window contains a complete list of library files that can be downloaded. Library files previously downloaded are marked currently installed.

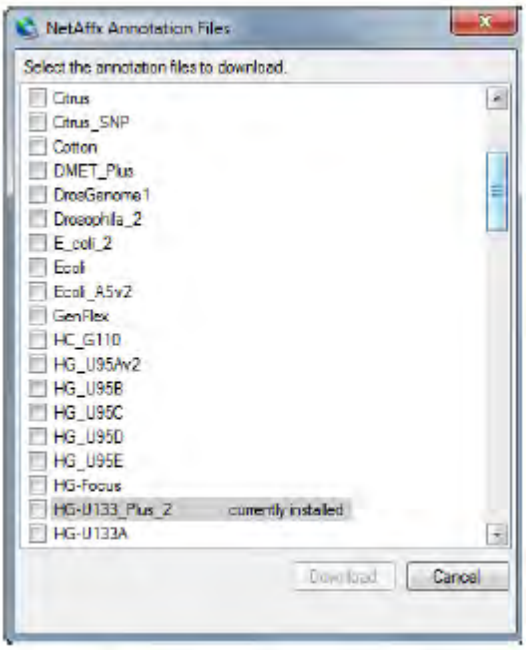


Figure 16 NetAffx™ Annotation Files window.

3. Check the annotation files needed and click the Download button.

   The appropriate files are downloaded to the folder you previously indicated. The green status bar at the bottom of the window highlights during the download process.

   If the Abort button is selected, the download process stops with a message indicating which library files failed to download.

Copy annotation files manually

For computers that are not connected to the internet, or for those desiring to use custom annotation files and therefore cannot take advantage of the library file download option, it is possible to manually copy the necessary files to the computer with the Expression Console™ Software:

1. Create a folder on the computer to hold the library files for the Expression Console™ Software application.

2. Copy the necessary files from the CD or other removable media to the library file folder.

   Note: Do not create subdirectories within the library file folder. The Expression Console™ Software does not look at subdirectories.

Create custom annotation files

The Expression Console™ Software will allow the use of custom annotation files in the annotation merge.

1. Create a comma-separated value file (CSV) containing the annotation information.
   a. The first column must contain the probe identifiers that must match the probe ids used on the corresponding array. This column will be used to merge the annotation in the custom file with the signal data.
   b. Each column must have a column header that will be used to select the annotation information when creating the annotation merge files (see "Exporting analysis results with annotation information" on page 124).
   c. Each row must have the same number of columns. If a row does not have a value for the column it must be padded or the annotation file can not be used.
   d. The data in the data columns will be appended to the probe set summarization results table.
2. Place the CSV file in the library file directory used by the Expression Console™ Software.

3 Creating a study

This chapter describes how to create a study with probe level intensity files (CEL) and probe level summarization files (CHP) in the Expression Console™ Software.

The Study dialog box

To get started using the Expression Console™ software, you create a study consisting of a collection of probe cell intensity files (CEL) and/or probe level summarization files (CHP) and their associated sample information for analysis and examination.

A study is created and changed using the Study dialog box.

Opening the study dialog box

Do one of the following:

- Click the New Study button.
- From the File menu, select New Study.
- In the Toolbox, select Study → Create New Study.

The Study window opens (Figure 17).

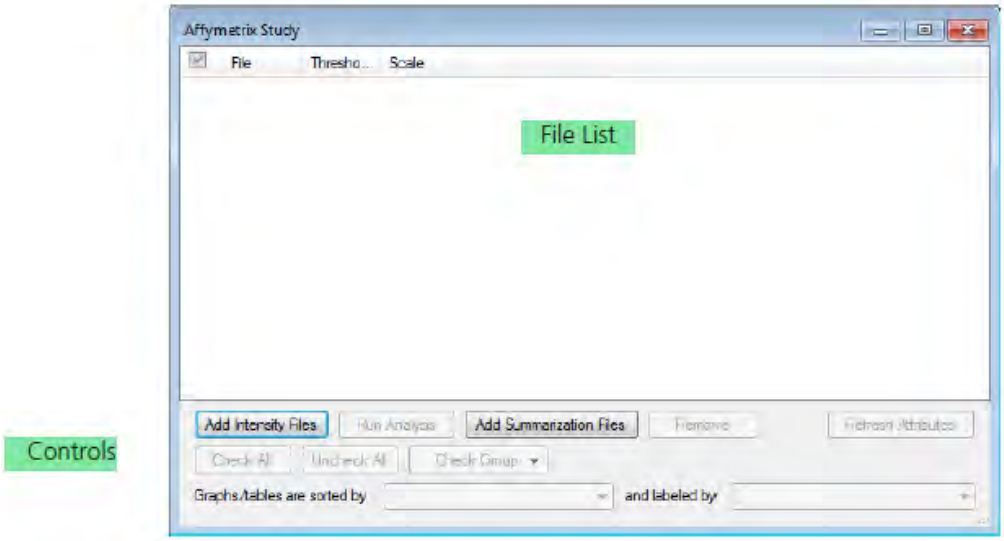


Figure 17 Study dialog box.

The File List displays (Figure 18).

- The CEL files selected for the study
- The CHP files generated by analysis

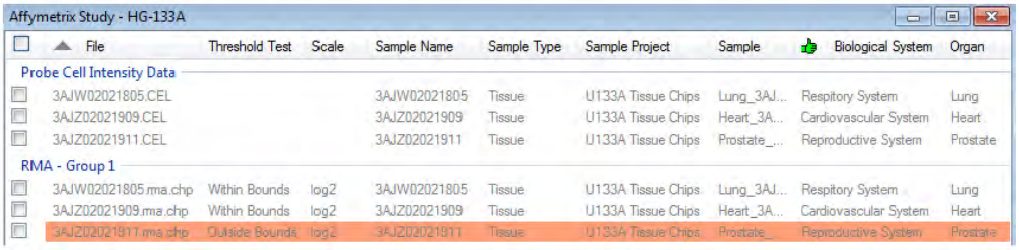


Figure 18 File List in the study dialog box.

The File List contains the following items (Table 2).

Table 2 Items in the file list.

Item	Description
Selection check box	Check box to select the file for analysis or removal
File	Name of the CEL or CHP file
Threshold Test	An indication of whether or not CHP files pass the user-defined tolerances for the array metrics. See Chapter 10, "Controls and thresholds". CHP files that do not meet tolerance are also highlighted in the list.

Item	Description
Scale (CHP files only)	Whether or not the results are on a linear or log scale By default, the algorithms PLIER and MAS 5.0 are linear; the RMA algorithm is in the log scale. Under the edit menu, the scale can be changed, but this does not change the scale of the data in the CHP file. See Appendix A, "Algorithms".
Sample Attributes	Sample attribute information from either the .xml files for GCOS (requires the use of DTT to transfer the data from the GCOS database) or .arr files (sample/array) from the GeneChip™ Console Center.

The controls (Figure 19) are at the bottom of the dialog box.

Figure 19 Study dialog box controls.

The controls provide functions for:

- "Creating, saving and opening a study" on page 27
- "Adding and removing files" on page 30
- "Selecting files for analysis, display, export, or removal" on page 38
- "Displaying files and data" on page 39

Creating, saving and opening a study

Refer to the sub-sections below to work with a study and analyze CEL files or view CHP file data.

Create a new study

Open the Study dialog box by doing one of the following:

- Click the New Study button .
- From the File menu, select New Study.
- From the Toolbox, select **Study ▸ Create New Study**.

The Study window opens (Figure 20).

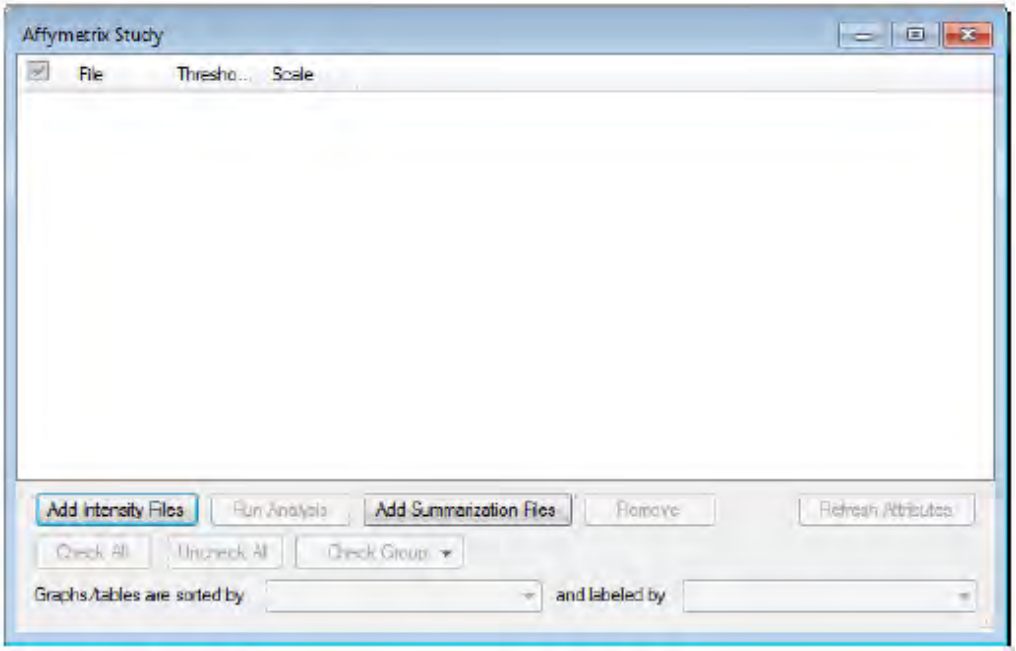


Figure 20 Study window.

You can now add data files to the study as described in "Adding and removing files" on page 30.

Create a new study

1. Do one of the following:
   - From the File menu, select Save Study.
   - In the Toolbar, click the Save Study button .
   - In the workflow toolbox, select **Study ▸ Save Study**.

The Save As dialog box opens (Figure 21).

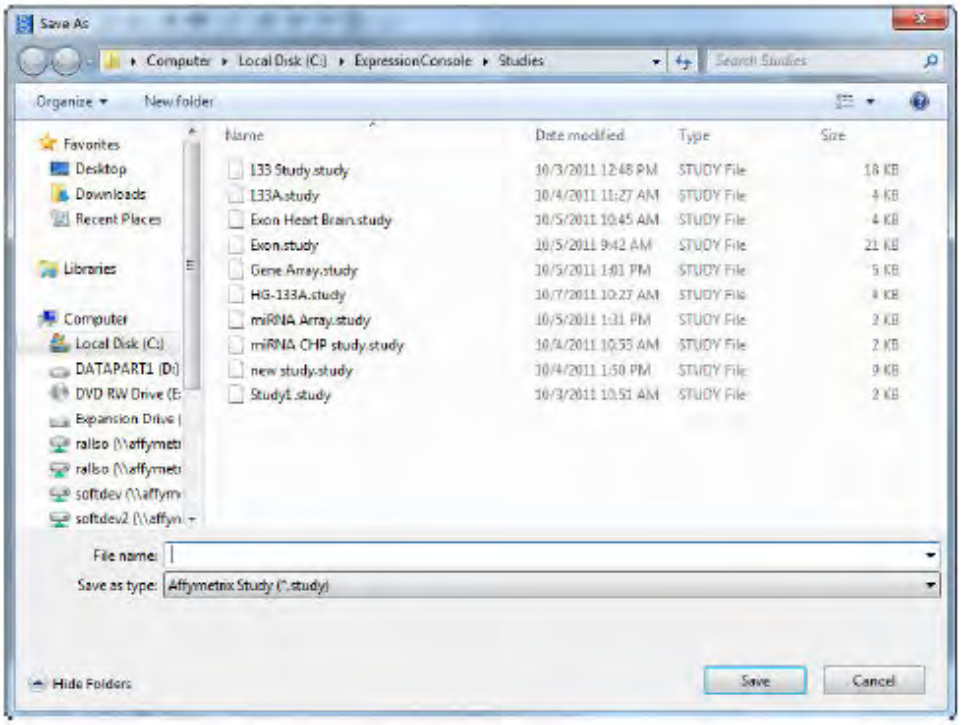


Figure 21 Save as dialog box.

2. Navigate to the folder where you wish to save the study.
3. Enter a name for the study and click Save.
   The study is saved.

   Note: You will also be prompted to save a new study when you attempt to close the Study dialog box.

Open an existing study

Do one of the following:

- Select the Open Study button in the toolbar.
- From the File menu, select Open Study.
- In the Toolbox, select **Study ▸ Open Existing Study**.

The Open dialog box opens (Figure 22).

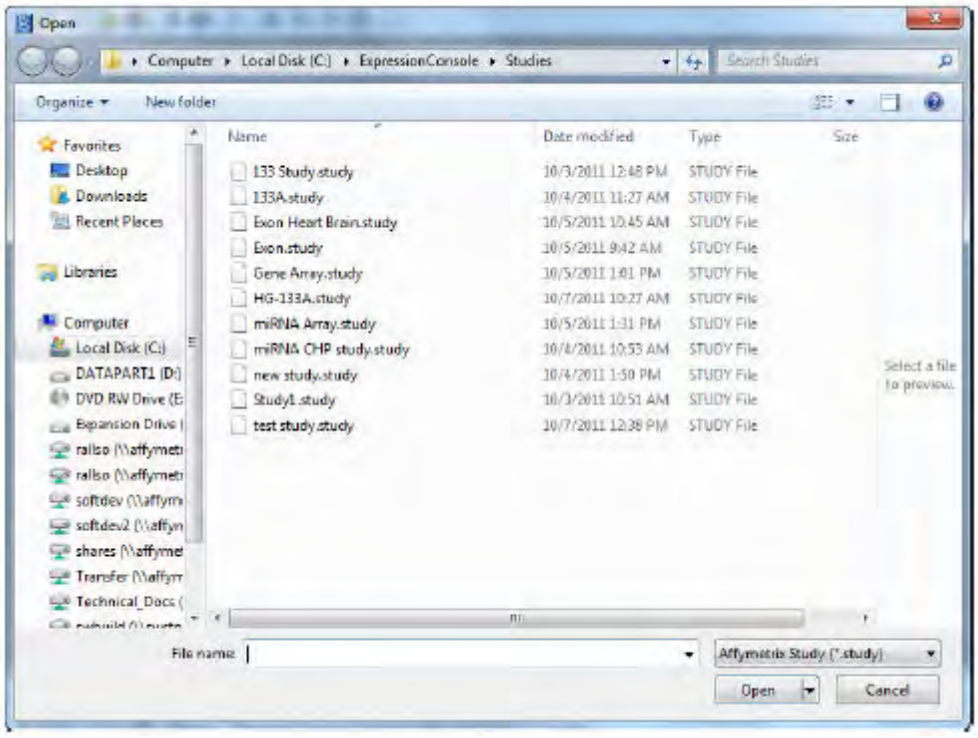


Figure 22 Open dialog box.

Adding and removing files

The Expression Console™ Software provides tools to:

- "Add CEL files from your data set of interest to a study" on page 30
- "Add CHP files to a study" on page 34
- "Remove files from a study" on page 36
- "Consolidate files" on page 37

Add CEL files from your data set of interest to a study

1. Open the Study dialog box.
2. Do one of the following:
   - Click the Add Intensity Files button
   - From the Files menu, select **Add to Study ▸ Probe Cell Intensity Files**

- In the workflow toolbox, select **Study ▸ Add Intensity Files**.
  The Select Probe Cell Intensity Files dialog box opens (Figure 23).

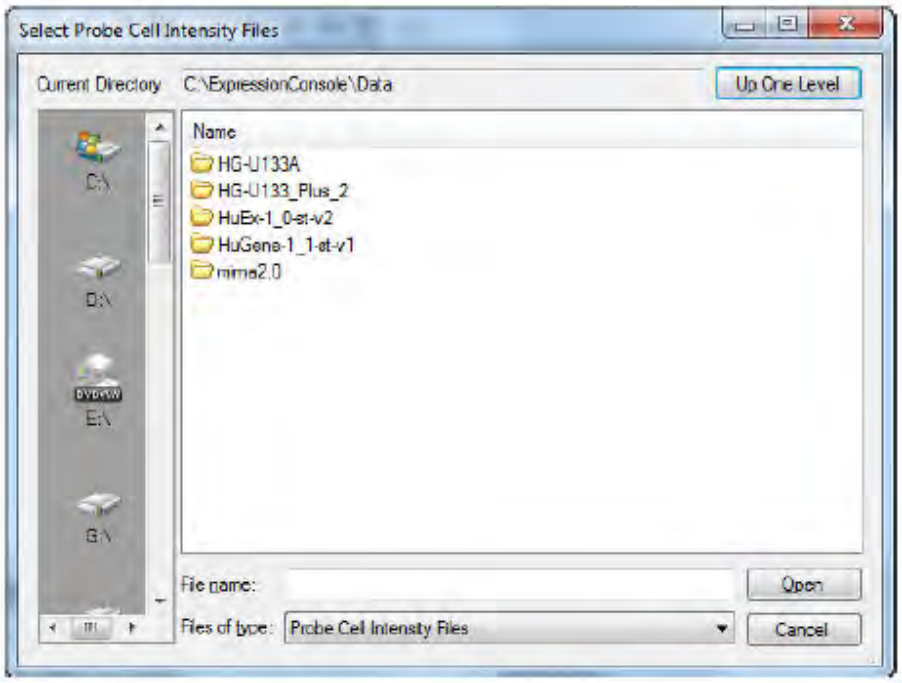


Figure 23 Select probe cell intensity files dialog box.

3. Locate the directory with the CEL files you wish to add.

   Sample attributes from .arr or .xml files are displayed for the associated CEL files, if available (Figure 24).

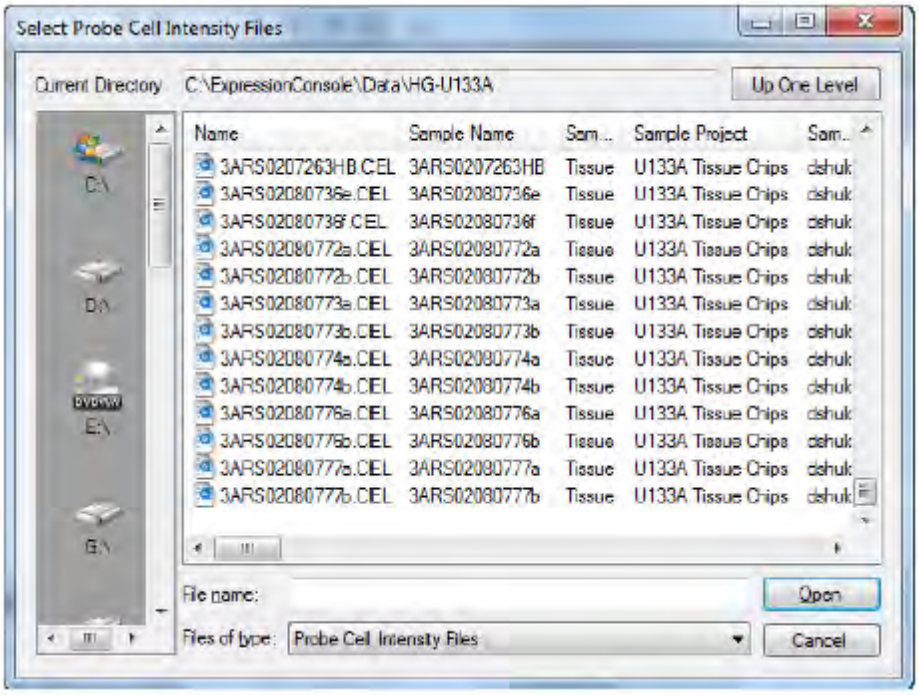


Figure 24 CEL files with attributes displayed.

STOPPING POINT You can sort the files based on an attribute by clicking the column heading for that attribute.

4. Select the CEL files to be added and click Open in the Select Probe Cell Intensity Files dialog box. The selected files are added to the Study (Figure 25).

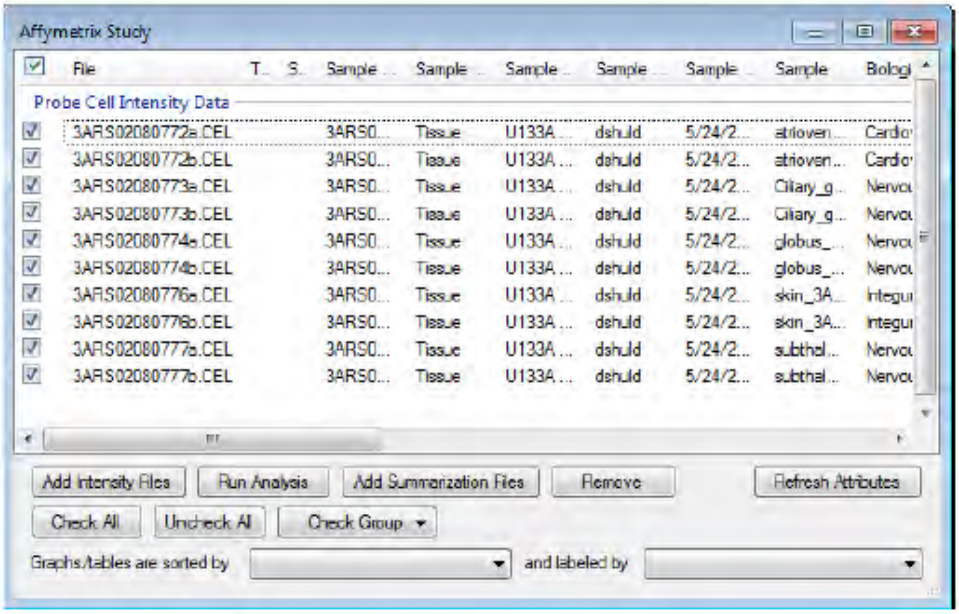


Figure 25 Study dialog box with CEL files added.

Missing library files

When you create a new study and add CEL/CHP files that do not have corresponding library files loaded in the correct folder, you will be prompted to download the appropriate library files (Figure 26).

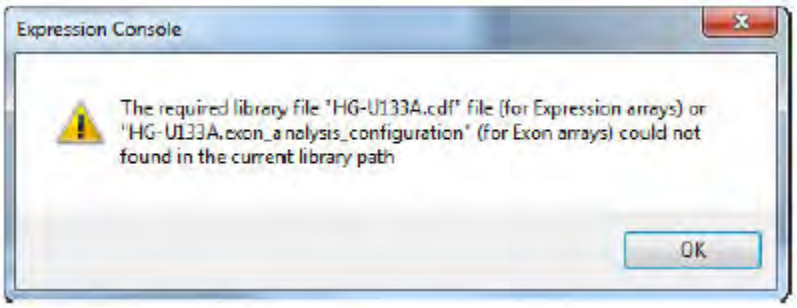


Figure 26 Notice of missing library files.

The status window at the bottom of the Expression Console™ Software application window lists the missing library files (Figure 27).

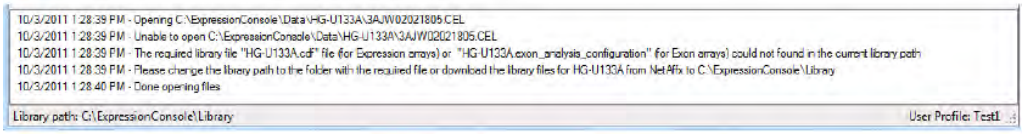


Figure 27 Status window - indicating missing library file.

For more information on changing the library file path and obtaining library files, see "Using library files" on page 17.

Add CHP files to a study

1. Do one of the following:
   - Click the Add Summarization Files button.
   - From the File menu, **select Add to Study ▸ Probe Level Summarization Files**.
   - In the Toolbox, select **Study ▸ Add Summarization Files**.

   The Select Probe Level Summarization Files window opens (Figure 28).

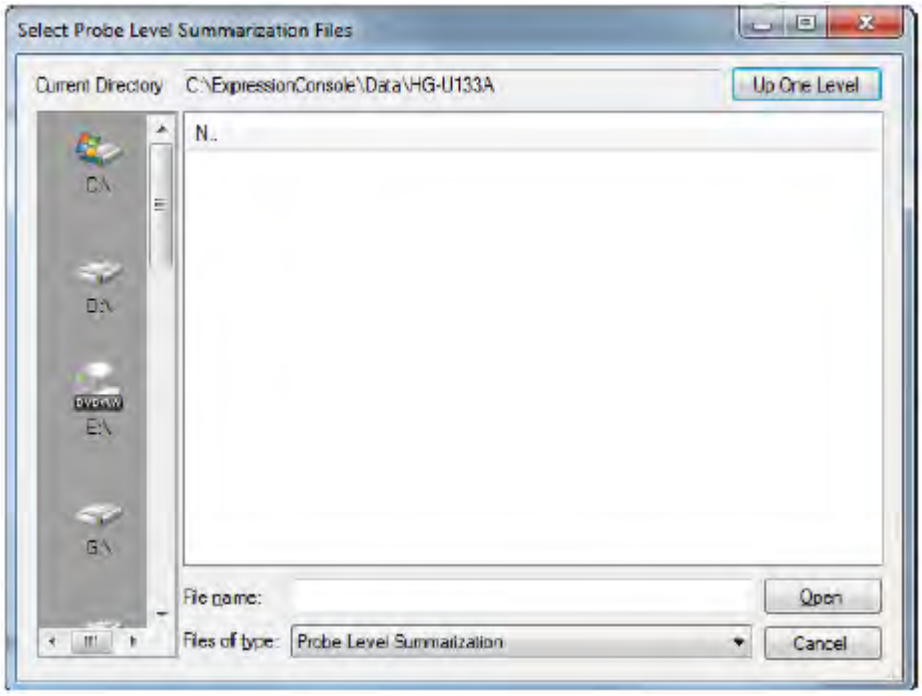


Figure 28 Select Probe Level Summarization Files window.

2. Navigate to the directory with the CHP files you wish to add.

   A list of the available CEL files with sample attributes, if available, are displayed in the Select Probe Level Summarization Files dialog box (Figure 29).

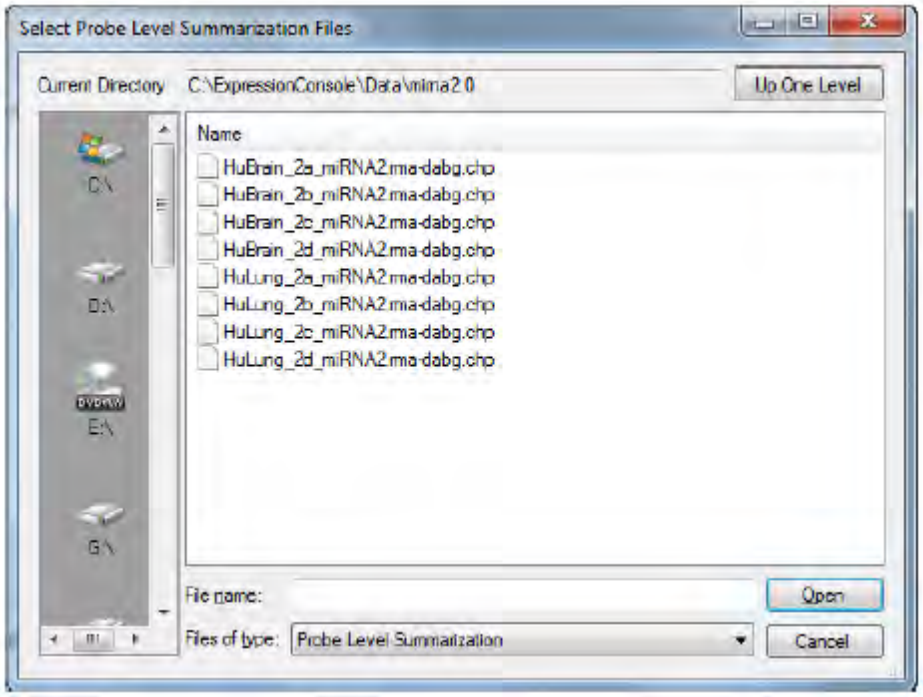


Figure 29 CHP files displayed in dialog box.

Note: Use the Files of type menu selection at the bottom of the window to filter the CHP files based on the selected algorithm type.

STOPPING POINT You can sort the files based on an attribute by clicking the column heading for that attribute.

3. Select the files to be added and click Open to add the selected CEL/CHP files to the study window. Figure 30

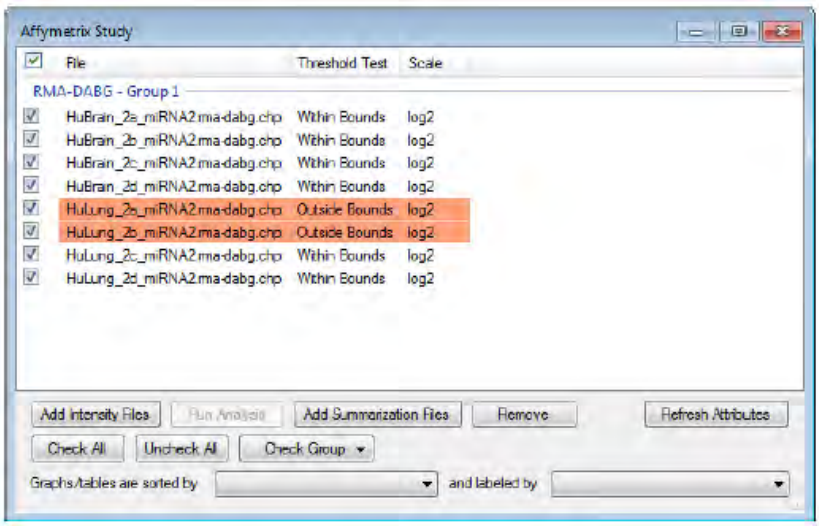


Figure 30 CHP files added to the study.

After CHP files are added to the study, either at the end of an analysis or directly by clicking the Add Summarization Files button, the metrics and controls associated with that algorithm run are compared against the user defined thresholds.

Any files that fail to meet any one of the criteria are highlighted.

Remove files from a study

1. Select the files to be removed.

   See “Selecting files for analysis, display, export, or removal“ on page 38.

2. Click Remove.

   The Confirm Removal notice appears (Figure 31).

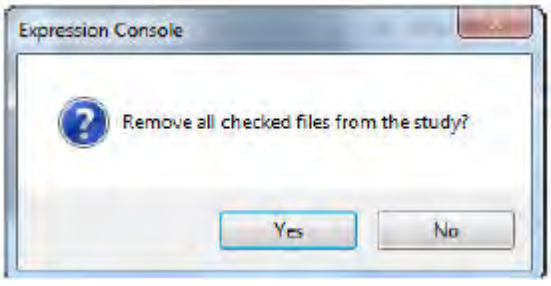


Figure 31 Confirm removal notice.

3. Click Yes.

   The selected files are removed from the study.

Consolidate files

A study can contain files that are located in multiple directories on the computer. To aid in file management, the Expression Console™ Software application can move all of the files associated with a study to a single directory. Studies cannot be consolidated when a study is open.

1. From the File menu, select Utilities.
2. Select the File Consolidation Tab (Figure 32).

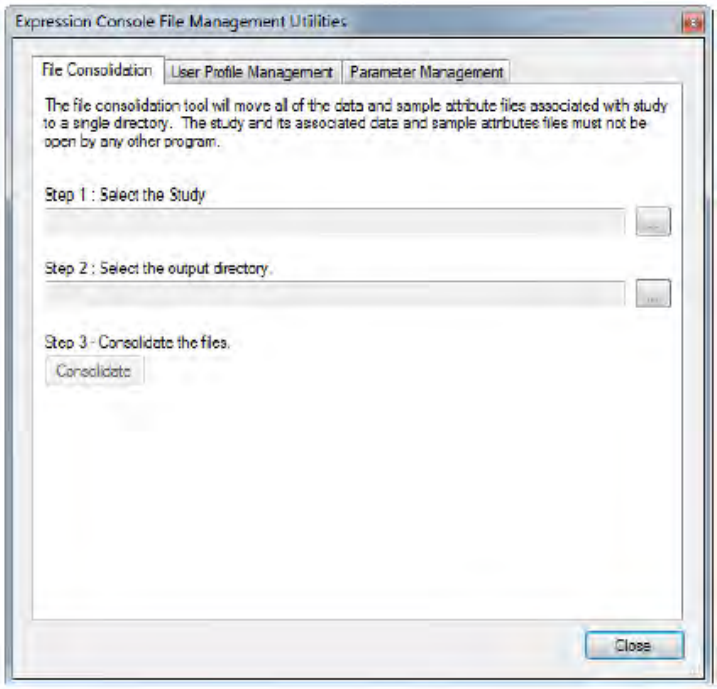


Figure 32 File consolidation tab.

Browse [...] to find the study to be consolidated.

3. Select an output directory to hold all of the associated files.
4. Click Consolidate and then Close.

   Note: File consolidation moves the files to the selected directory.

Selecting files for analysis, display, export, or removal

To select individual files for analysis, display, or removal, do one of the following:

- Select the check box next to the file name. (Figure 33)

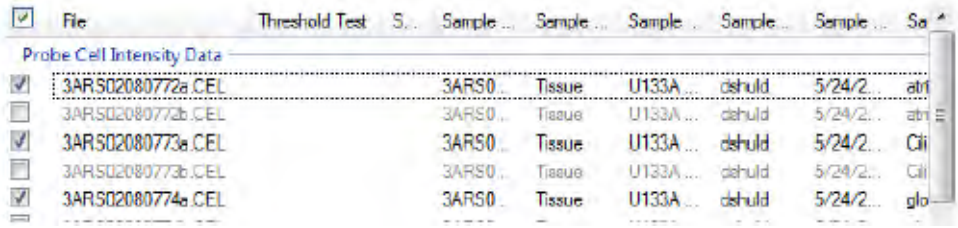


Figure 33 Selecting individual checkboxes and files.

- Click Check All [Check All] to select all files.
- Click Uncheck All [Uncheck All] to deselect all files.
- Select a group of files from the Select Group drop down list (Figure 34).

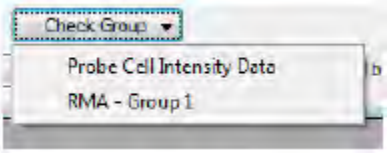


Figure 34 Check group dropdown list.

- Right-click selected CHP files and select Highlight associated CEL files on the shortcut menu (Figure 35).

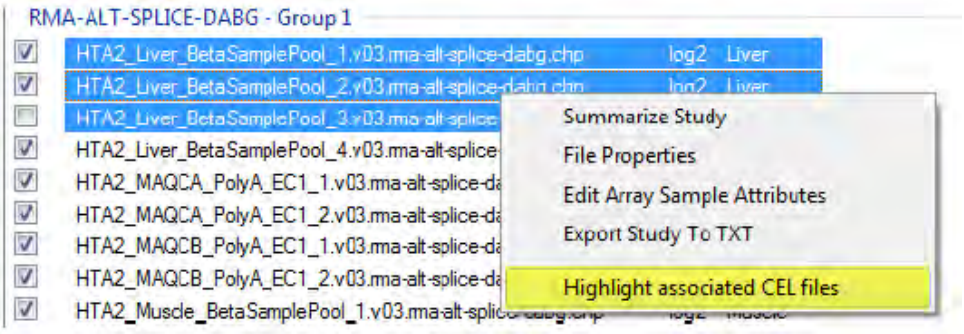


Figure 35 Selecting and highlighting CEL files

Note: If outlier signals are identified, use this feature to quickly locate the related CEL files, uncheck the CEL files and repeat the analysis.

For more information about analysis, see:

- Chapter 4, “3’ expression array analysis“
- Chapter 5, “Exon array analysis“
- Chapter 6, “Gene Array analysis“
- Chapter 7, “miRNA Array analysis“

For more information about displaying file data, see Chapter 9, “Viewing QC and analysis results data“.

For more information about exporting data, see Chapter 11, "Exporting data".

Displaying files and data

The order that files are displayed in the Study dialog box determines:

- The order they are displayed in tables and graphs
- The order in which the data is exported

By default the files are displayed in order of the file name (Figure 36).

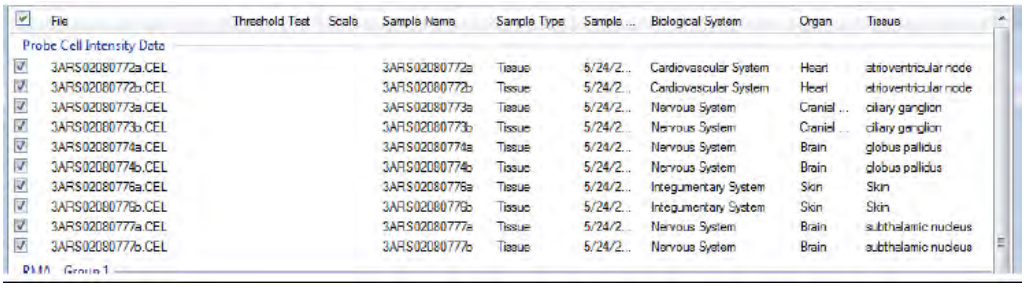


Figure 36 CEL files displayed in default order.

You can sort files by a selected attribute and append an attribute to the file name as it appears in tables and graphs, and some export files.

To sort files in a study by a selected attribute, do one of the following:

- Clicking in the header for the attribute column (Figure 37).

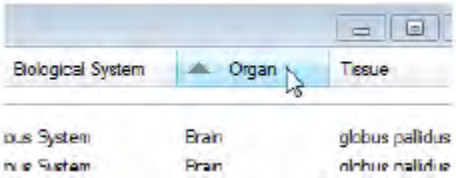


Figure 37 Sort by clicking in column header.

- Selecting the attribute from the Graphs/Tables are sorted by drop down list (Figure 38).

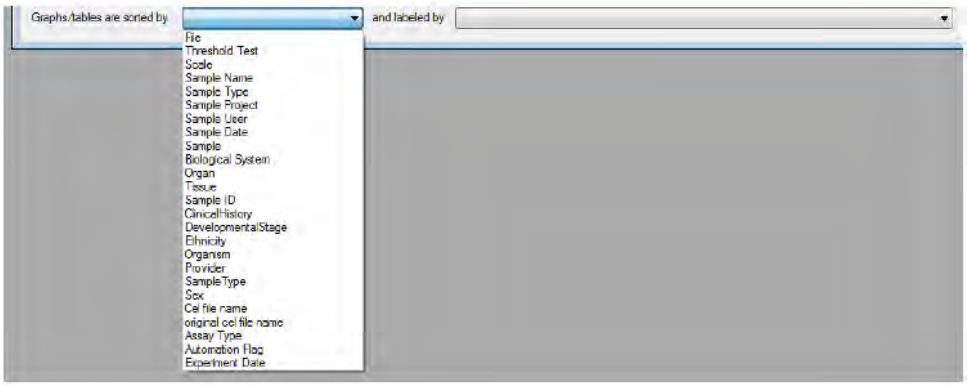


Figure 38 Selecting sort attribute from dropdown list.

The files are sorted in the new order.

An arrow in the column header indicates the sort order (Figure 39).

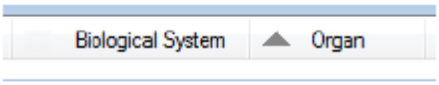


Figure 39 Column sort indicators.

If you sort by additional attributes, the first attributes selected are indicated by a green thumb in the column header.

To append an attribute prefix to the file name in graphs and tables:

- Select the attribute from the Labeled by drop-down list (Figure 40).

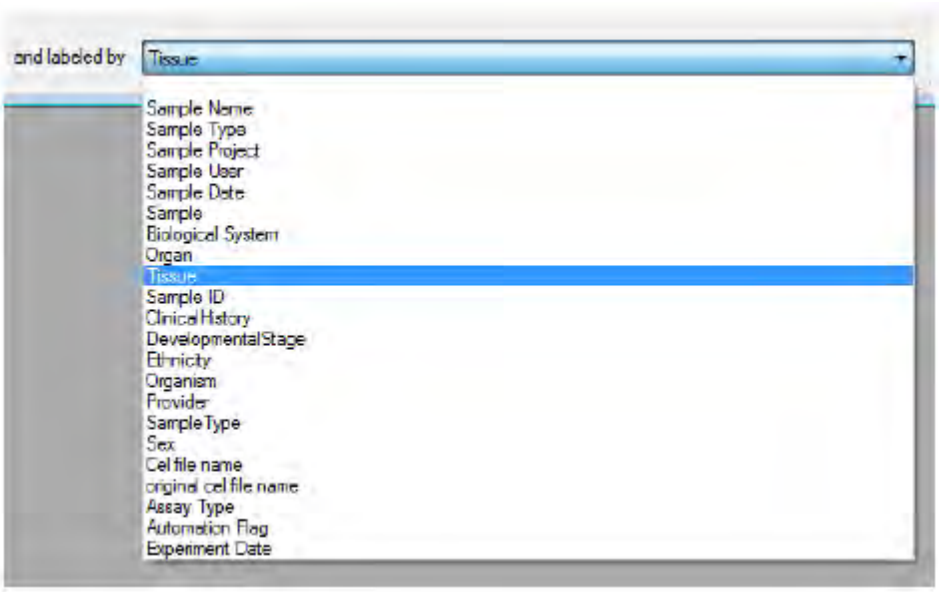


Figure 40 Labeled by dropdown list.

Files in tables and graphs are displayed in the selected order with the attribute appended to the file label (Figure 41, Figure 42).

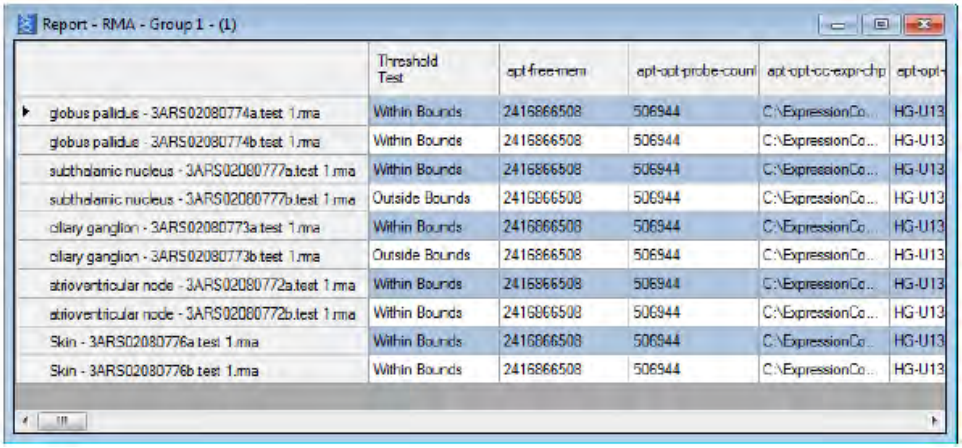


Figure 41 Full Report with attribute appended to file name.

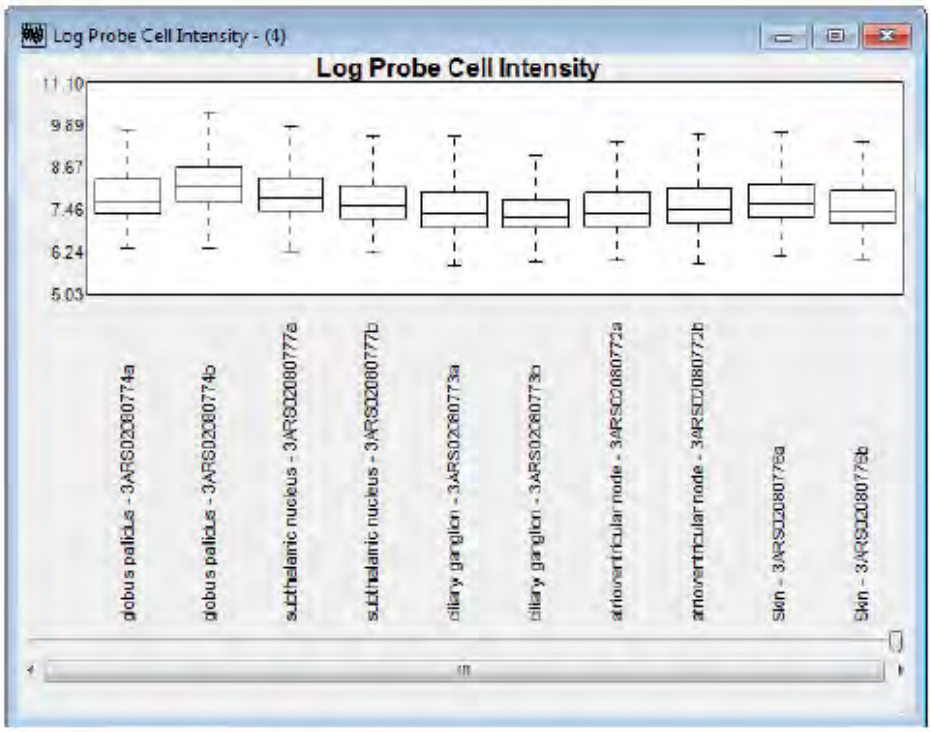


Figure 42 Lob Probe Cell Intensity graph with attribute appended to the file name.

Refreshing attributes

The attributes displayed in the Study window are read from the ARR file when the files are first added to the study. If the ARR files are edited after adding to the study, click Refresh Attributes to update the attributes in the Study window (Figure 43).

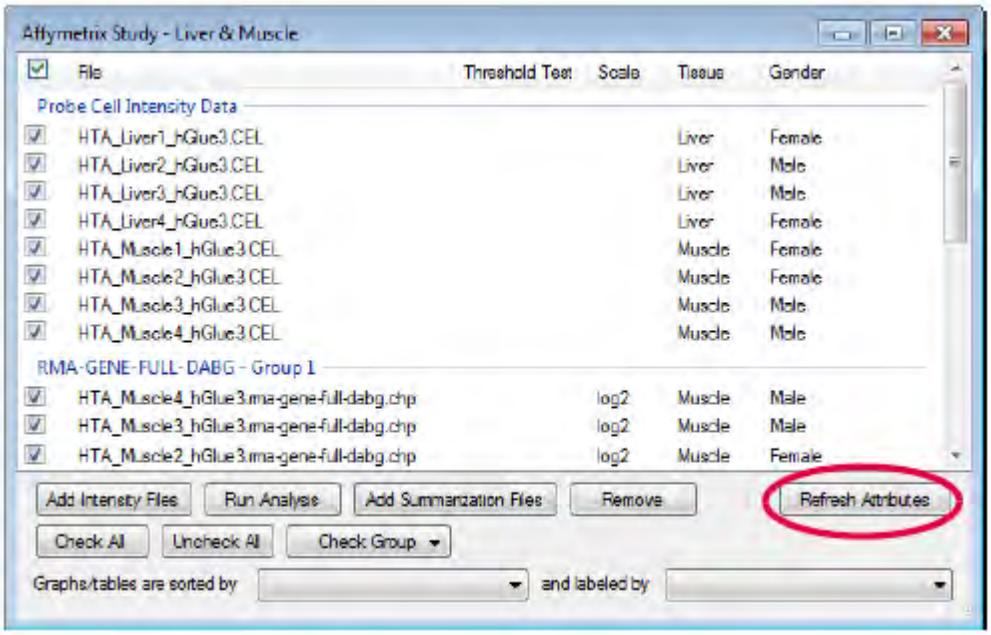


Figure 43 Study window.

Viewing study summary/attributes

To view a study summary, select **File ▸ Summarize Study**. Alternatively, right-click the Study window and select Summarize Study from the shortcut menu (Figure 44.

The study summary includes information about:

- Probe cell intensity data (CEL):
  - Number Files – The number of CEL files in the Study window.
  - Number Checked – The number of CEL files selected for analysis or display.
- Gene or exon signal (CHP):
  - Number Files – The number of CHP files in the Study window.
  - Number Checked – The number of CHP files selected for analysis or display.
  - Number Within Bounds – The number of analysis results that are within the report controls thresholds.
  - Number Outside Bounds – The number of analysis results that are outside the report controls thresholds.

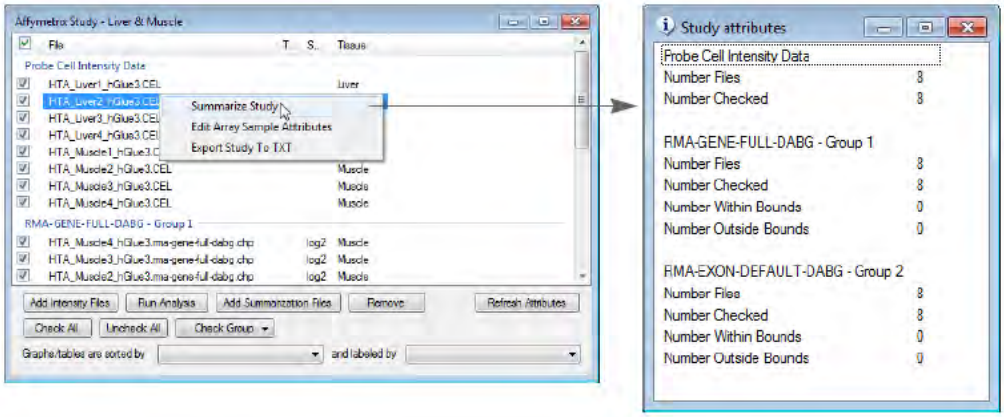


Figure 44 Viewing the study summary.

Editing sample attributes

The Sample Attribute Editor enables you to edit or add attributes to the sample files (ARR). Do either of the following to open the Sample Attribute Editor:

- Right-click data in the Study window and select Edit Array Sample Attributes on the shortcut menu.

  or

- Select **Tools ▸ Sample Attribute Editor** from the menu bar.

  The Sample Attribute Editor opens and contains the sample files for all of the data in the Study window (Figure 45).

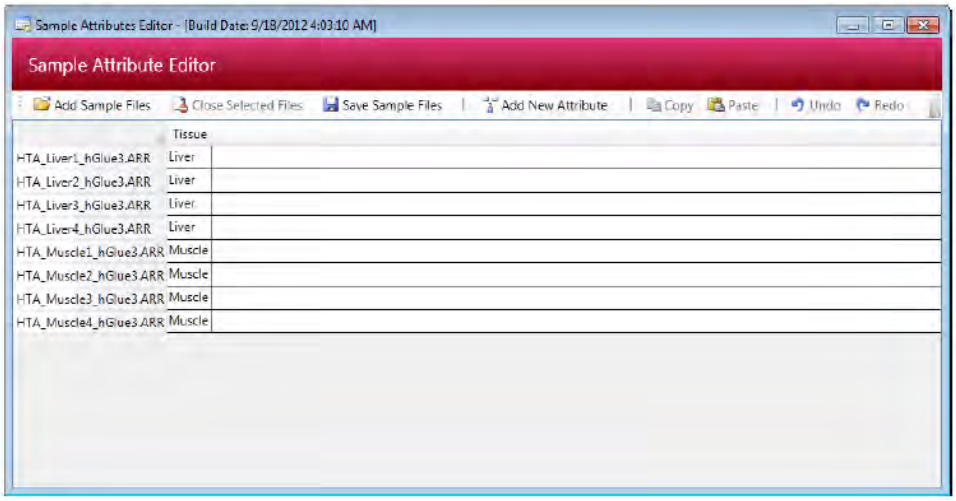


Figure 45 Sample attribute editor.

or

- Select **File ▸ Open Sample/Array Attribute File** and choose one or more sample files (ARR) in the dialog box that appears.

  The Sample Attribute Editor opens and contains the selected files.

Table 3 Sample attribute editor commands.

Item	Description
Add Sample Files	Opens a dialog box that enables you to select sample files (ARR) to add to the Sample Attribute Editor.
Close Selected Files	Removes one or more selected ARR files from the Sample Attribute Editor.
Save Sample Files	After editing sample information, click Save Sample Files to save the changes to the ARR files.
Add New Attribute	Enables you to create a new column in the Sample Attribute Editor.

Item	Description
Copy	Copies selected text to the system clipboard. For example: • Copy a cell and paste it to multiple selected cells • Copy multiple cells and paste to multiple cells.
Paste	Pastes text from the system clipboard.
Undo	Erases the previous change made in the Sample Attribute Editor (can be performed multiple times).
Redo	Reverses the undo command (can be performed multiple times).

Add sample files from the sample attribute editor

1. Click Add Sample Files.
2. Select a sample file(s) (ARR) from the dialog box that appears.

Remove sample files from the sample attribute editor

1. Select a file(s) in the Sample Attribute Editor.
2. Click Close Selected Files.

Adding, editing, or removing attributes

Add a new attribute

1. Click Add New Attribute. Alternatively, right-click an attribute column header and select New Attribute on the shortcut menu.
2. Enter a name for the attribute in the box that appears and click OK.
   The attribute is added to the Sample Attribute Editor.

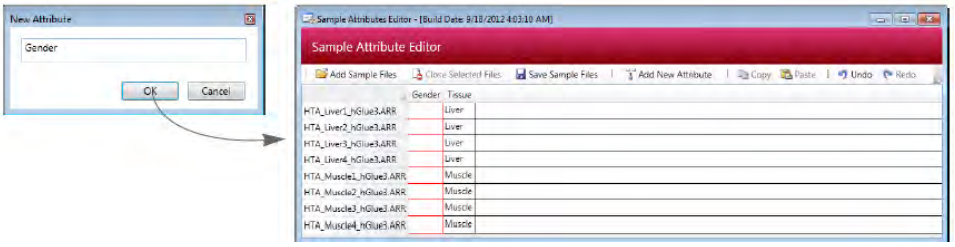


Figure 46 New attribute box.

3. To enter attribute values:
   a. Select one or more cells.

b. Enter text in the last cell in the selection and press Enter.
The same value is added to all of the selected cells.

4. Click Save Sample Files.

Copy and paste values from a spreadsheet

1. Copy multiple cells in the spreadsheet.

2. In the Sample Attributes Editor, click the first cell to receive the values and click Paste (Figure 47).

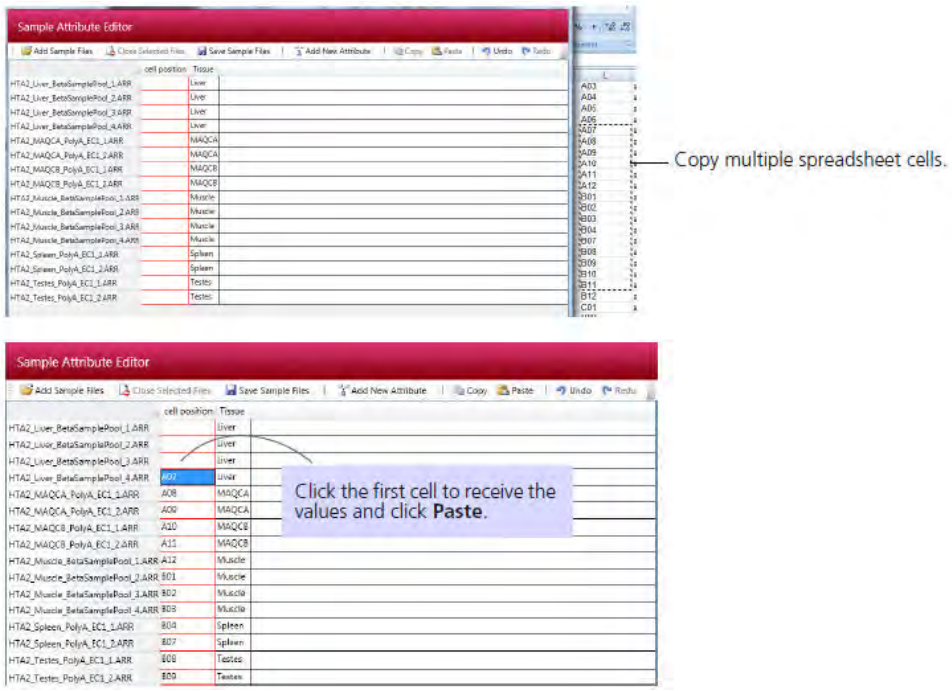


Figure 47 Copying values from Excel™ to the sample attributes editor.

3. Click Save Sample Files.

Edit an attribute value

1. Double-click the cell.

2. Enter text or right-click the cell to view a shortcut menu of edit commands (Cut, Copy, Paste).

Sort attributes

1. Click the column header
Attributes are sorted in ascending alphanumeric order.

2. Click the column header again to sort in descending alphanumeric order.

Remove an attribute

1. Right-click the attribute column header and select Delete Attribute on the shortcut menu.
2. Click Save Sample Files.
   The Study window shows the revised attributes.

Zipping a study

Zipping a study provides a convenient way to share data.

Zip a study

1. Open a study.
2. Select **File ▸ ZIP Study** on the main bar.
   The study data are compressed to a zip file with the same name as the study. The zip file is saved in the folder that contains the study (Figure 48).

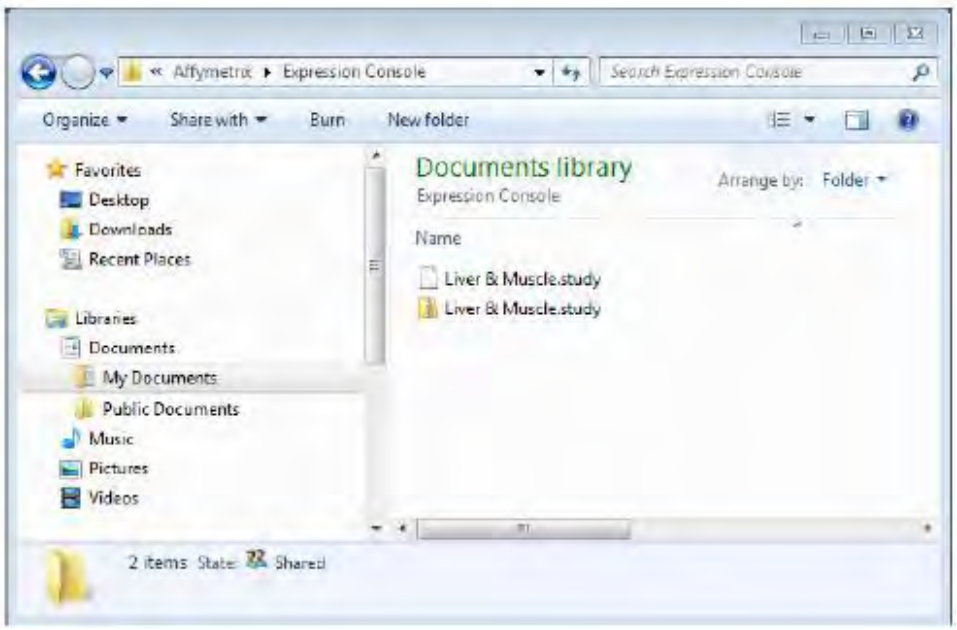


Figure 48 Zipped study.

Unzip a study

1. Select **File ▸ unZIP Study** on the menu bar.
2. Select a zip file in the dialog box that appears and click Open.

4 3’ expression array analysis

Probe cell intensity data from 3' Expression Array analyses is further analyzed in the Expression Console™ Software application using the MAS5, RMA, and PLIER algorithms to create CHP files. Follow the instructions below to analyze 3' Expression Array data.

IMPORTANT! You must have the necessary library files that are downloaded and installed to perform 3’ expression array analysis. See “Using library files“ on page 17 for more information.

Analysis controls

Our microarrays contain the hybridization, labeling and housekeeping controls that help determine the success of the hybridizations. For more information on the interpretation of these controls, see the white paper, Data Analysis Fundamentals, available on our web site.

To aid in the examination of these controls, the Expression Console™ Software application displays summarized probe information in tabular format. In order for this feature to function properly, the controls must be identified prior to analysis. Once the controls have been identified for a particular GeneChip™ (probe array type), they are saved and only need to be updated if the user wishes to modify the controls.

For most standard GeneChip™ Arrays, a standard set of defaults has been provided; however, it is recommended that the user should verify all controls before initiating their analysis. For details on defining the controls and their thresholds, refer to Chapter 10, “Controls and thresholds“.

Analysis algorithms

By default, the 3' Expression Array workflows are set as follows:

- PLIER workflow is Quantile normalization with PM-MM
- RMA workflow is Quantile normalization and has a general background correction
- MAS 5.0 workflow is set so that all probe sets are scaled to TGT = 500

For more details about the algorithms, see Appendix A, "Algorithms" and Appendix B, "Algorithm parameters and outputs".

Processing data using the Expression Console™ Software

Analyze 3' expression array data

1. Create a new study or open an existing study. See Chapter 3, "Creating a study".
2. In the Study dialog box (Figure 49), check the CEL files for Analysis.

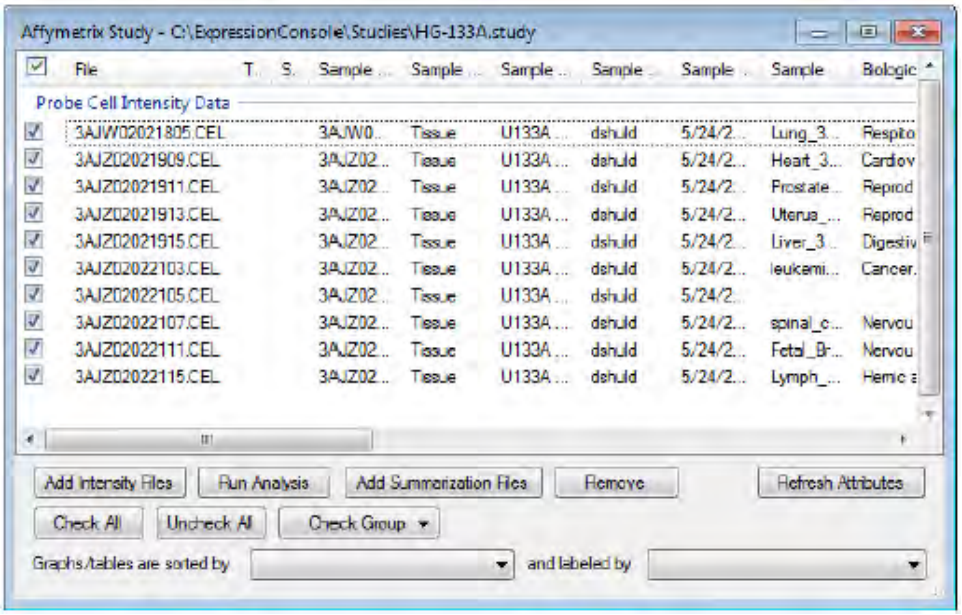


Figure 49 Study dialog box with CEL files selected.

3. Select the Analysis Workflow by doing one of the following:
   - From the Analysis Menu (Figure 50):
     a. Select **Analysis ▸ 3' Expression Arrays** and select the appropriate workflow:
        – MAS5
        – RMA
        – PLIER

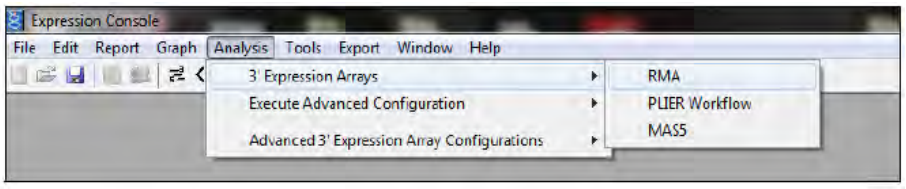


Figure 50 Analysis menu options for 3' expression arrays.

Note: For information on using the Advanced Configuration, see Chapter 12, "Advanced analysis".

Or:

- In the Study dialog box (Figure 51):

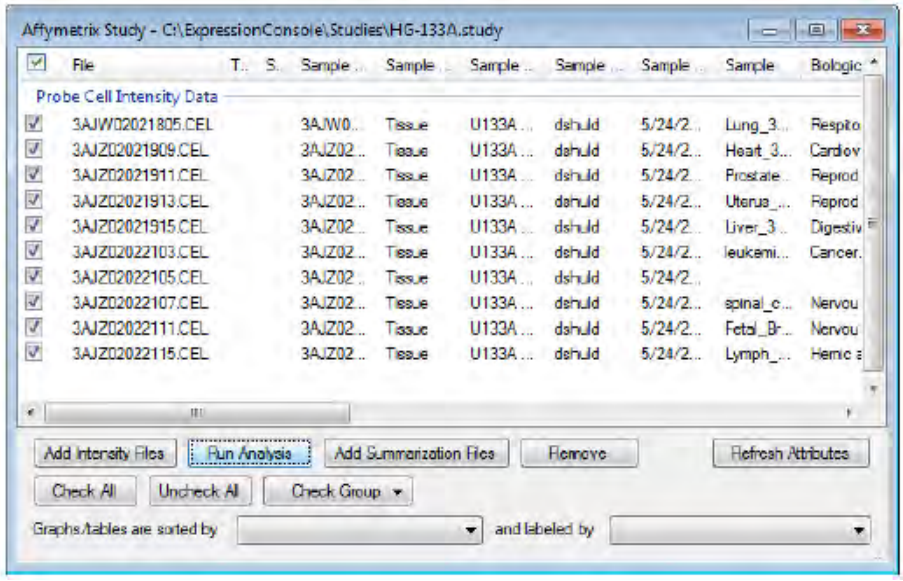


Figure 51 Study dialog box.

a. Click Run Analysis in the Study dialog box.
The Available Analysis dialog box opens (Figure 52).

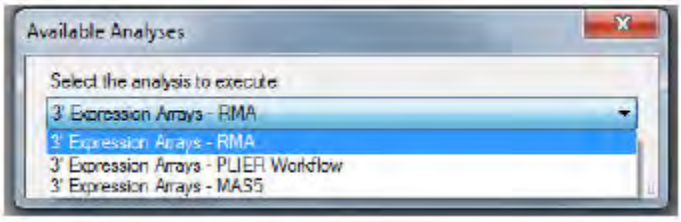


Figure 52 Available Analysis dialog box.

b. Select the analysis you wish to run.
c. Click OK in the Available Analysis dialog box.

The suffix dialog box for the selected workflow opens (Figure 53).

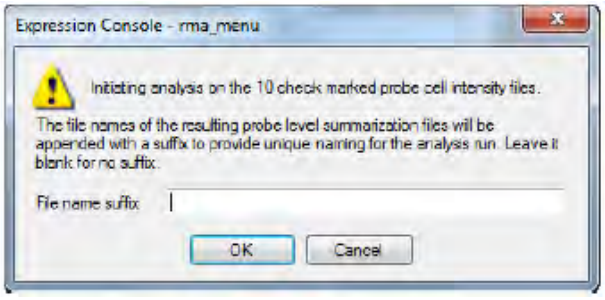


Figure 53 File name suffix dialog box.

Note: You can add a suffix to the file name in order to further identify your samples, or you can leave it blank. The summarization method is automatically included in the file name.

4. Enter a suffix to identify the specific samples after analysis.

5. Click OK to begin analysis.
   The progress of analysis is displayed in the Status window (Figure 54, Figure 55).

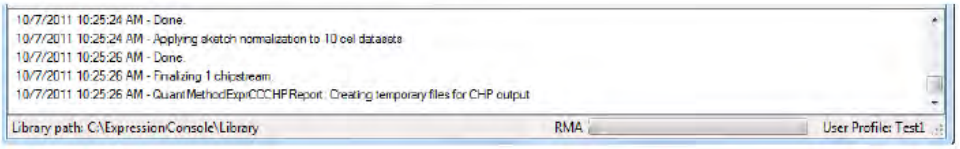


Figure 54 Status window with analysis running.

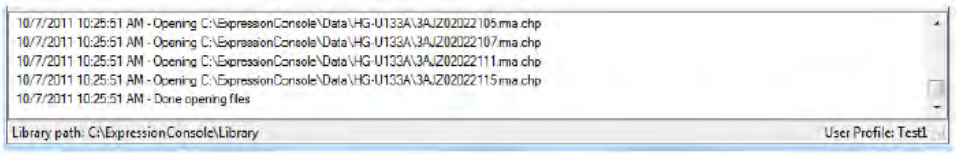


Figure 55 Status window with finished analysis displaying CHP files.
The progress bars provide an indication of continuing activity.

The Study window (Figure 61) is populated with the corresponding CHP files.

Stop an analysis run

From the Analysis menu, select Stop (Figure 56).

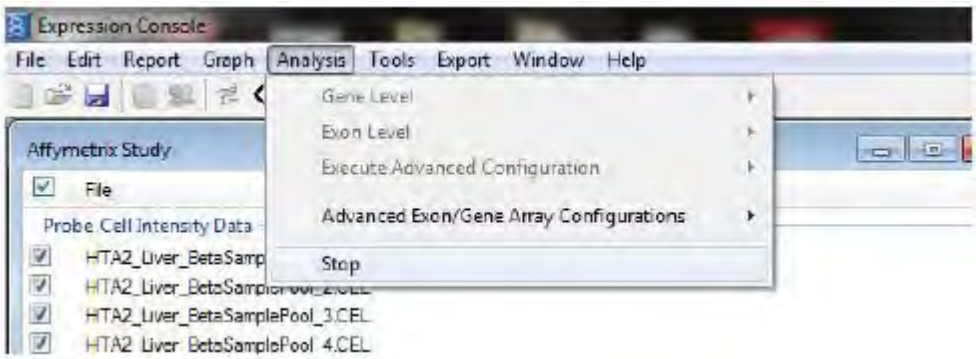


Figure 56 Canceling an analysis run.

A notice will pop-up indicating the application is stopping the current analysis (Figure 57).

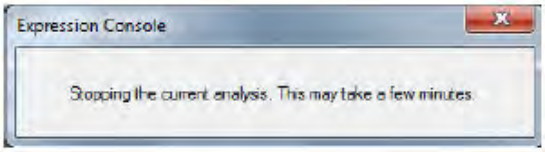


Figure 57 Notice of stopped analysis.

In some cases it may not be possible to cancel the process immediately, but within a couple of minutes the operation will be cancelled.

After cancellation of the analysis run the user is notified that the analysis failed. Depending upon the stage where the analysis was cancelled, different messages appear (Figure 58, Figure 59, Figure 60).

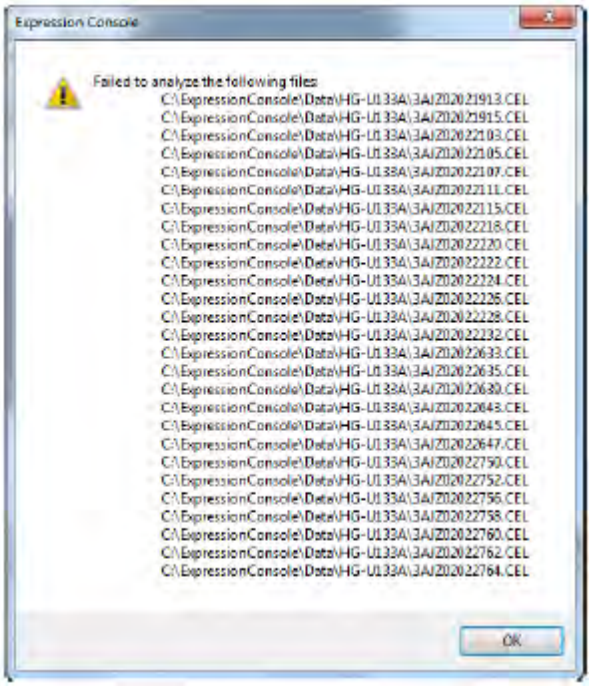


Figure 58 Message after canceling an analysis run.

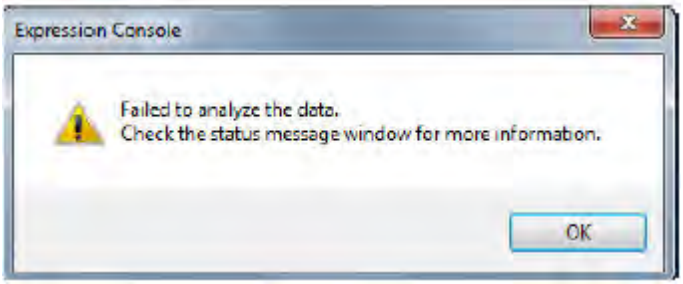


Figure 59 Message after canceling an analysis run.

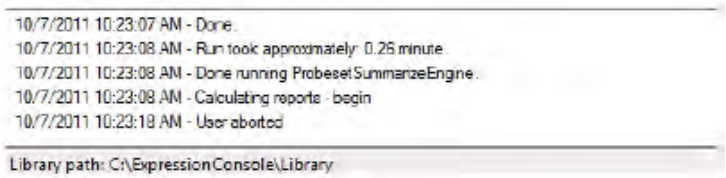


Figure 60 Status window after stopping analysis.

Results of 3' expression array analysis

Several minutes may be required for processing results, depending on the number of arrays selected. Once processing is complete, the study window shows a new group containing the newly summarized data (CHP files). The new group is named with the type of algorithm selected.

The files are listed in the status window at the bottom of your screen as they are generated, and the Study window (Figure 61) is populated with the corresponding CHP files.

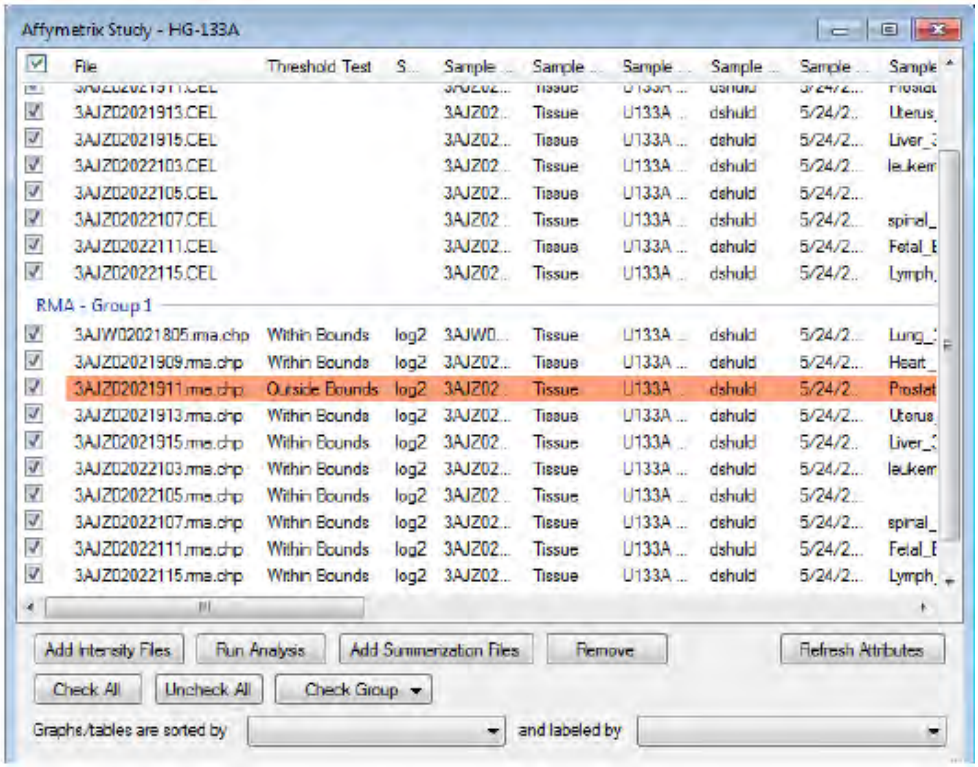


Figure 61 Study window - RMA analysis results.

Highlighted CHP files contain metrics outside of defined thresholds.

For information about viewing analysis results and QC data from the CHP files, see Chapter 9, "Viewing QC and analysis results data".

For information about Exporting Data, see Chapter 11, "Exporting data".

Note: You can further analyze results (CHP) that pass QC using Transcriptome Analysis Console. Select Tools → Transcriptome Analysis Console to start the software.

5 Exon array analysis

Probe cell intensity data (CEL) from GeneChip™ exon arrays are analyzed in the Expression Console™ Software. For exon arrays, Expression Console™ Software uses the RMA-sketch workflow for both exon and gene level analyses to create CHP files. Follow the instructions below to analyze exon array data.

IMPORTANT! You must have the necessary library files that are downloaded and installed to perform exon array analysis. See "Using library files" on page 17 for more information.

Analysis options for Exon Arrays

Expression Console™ Software provides different analysis options for Exon Arrays.

- "Gene level analysis" on page 54
- "Exon level analysis" on page 55
- "Control analysis" on page 55

The Gene and Exon Level analyses use different groupings of probe sets for analysis. and corresponds to different biology. The gene level case looks at the expression of the gene overall, and the exon level case looks at parts of the gene.

Control Analysis looks at a subset of probes, enabling you to quickly obtain a cursory look at the success of hybridization for different cell files.

These options can be selected during Exon Array analysis, as described in "Running an analysis" on page 55.

Gene level analysis

Gene Level analysis detects the expression of the gene overall. There are two transcript structure confidence levels for Gene Level analysis, described below.

These describe transcript structure confidence levels (for the annotation used to define the transcript) that are annotations broadly defined as follows:

- Core: limits analysis to exon-level probe sets that map to BLAT alignments of mRNA with annotated full-length CDS regions.
- Extended: limits analysis to transcripts that are defined by exon-level probe sets that map to cDNA alignments and their annotations based on cDNA alignments, in addition to transcripts defined by the core category.

  For further explanation of confidence levels, download "White Papers: Exon Probeset Annotations and Transcript Cluster Groupings v1.0." from our **website**.

Exon level analysis

Exon Level analysis detects the expression of parts of the gene. These are gene confidence levels, which are input transcript annotations broadly defined as follows:

- Core: limits analysis to exons that consist of BLAT alignments of mRNA with annotated full-length CDS regions.
- Extended: limits analysis to cDNA alignments and their annotations based on cDNA alignments, in addition to core defined exons.
- Full: analysis uses exons derived from sets of ab-initio gene predictions in addition to the core and extended exons.
- All: consists of all three confidence levels plus probe sets that map to more than one gene or to genes that do not align to the genome.

  For further explanation of confidence levels, download "White Papers: Exon Probeset Annotations and Transcript Cluster Groupings v1.0." from our **website**.

Control analysis

A subset of probes, on all arrays, are identified as controls. This option summarizes expression values for controls only.

IMPORTANT! You can look at the controls alone to obtain a cursory look at the success of the hybridization. This allows you to remove the obvious outliers prior to the time investment of analyzing the full array.

Running an analysis

Note: To use custom library files for an analysis, see Chapter 12, “Advanced analysis“.

Analyze data from the Exon arrays

1. Create a new study or open an existing study. See Chapter 3, “Creating a study“.
2. In the Study dialog box, check the CEL files for analysis (Figure 62).

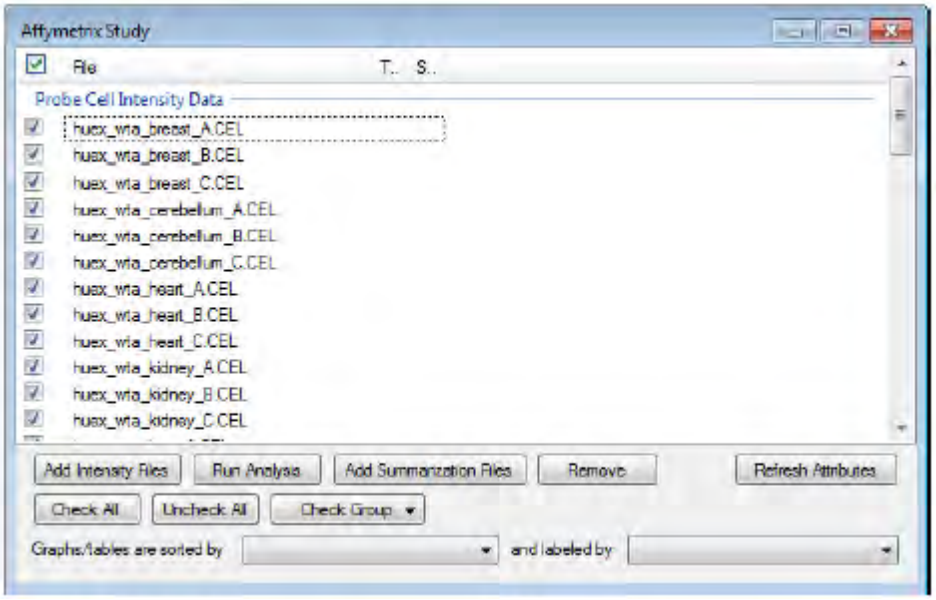


Figure 62 Study dialog box.

You can choose from three different analysis options:

- Gene Level
- Exon Level
- Controls Only

  You can select from different confidence levels for Gene and Exon Level analyses.

  See "Analysis options for Exon Arrays" on page 54 for more information.

3. Select the analysis and confidence levels by doing one of the following:
   - From the Analysis Menu (Figure 63), select the appropriate workflow.

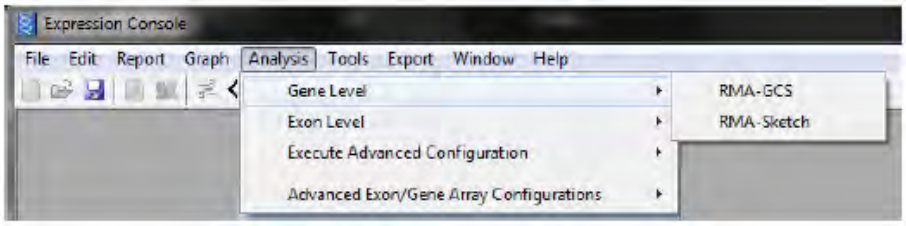


Figure 63 Analysis menu for exon arrays.

The analysis options are described in "Analysis options for Exon Arrays" on page 54.

Note: For information on using the Advanced Configuration option, see "Advanced configuration Exon/Gene analysis" on page 140.

Or:

- In the Study dialog box (Figure 64):

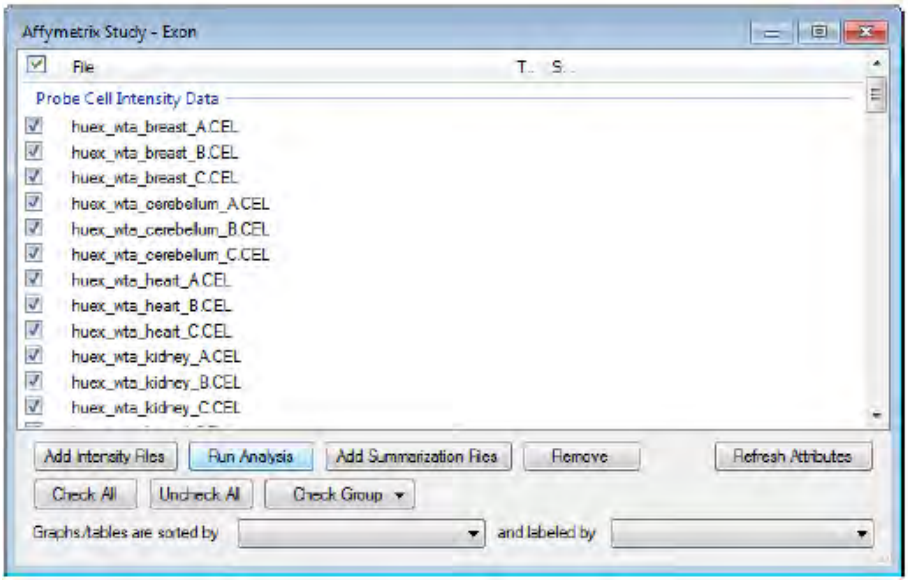


Figure 64 Study dialog box.

a. Click Run Analysis in the Study dialog box.
The Available Analyses dialog box opens (Figure 65).

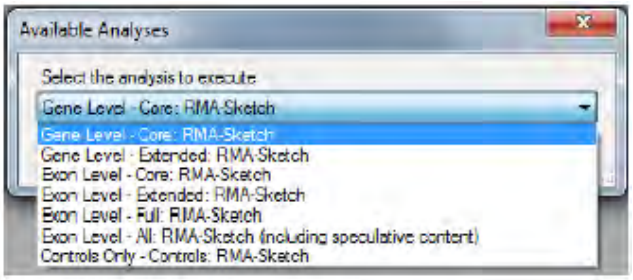


Figure 65 Available analysis dialog box.

b. Select the analysis you wish to run.

Note: The drop-down list displays all of the Exon Array analysis options for:

- Gene Level
- Exon Level
- Controls Only

The options are described in "Analysis options for Exon Arrays" on page 54.

c. Click OK in the Available Analysis dialog box.
The suffix dialog box for the selected workflow opens (Figure 66).

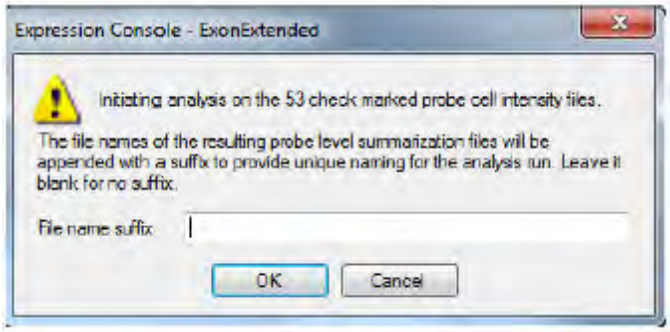


Figure 66 File name suffix dialog box.

Note: You can add a suffix to the file name in order to further identify your samples, or you can leave it blank. The summarization method is automatically included in the file name.

d. Enter a suffix to identify the specific samples after analysis.

e. Click OK; analysis begins.

IMPORTANT! The processing time for the CEL files depends on a number of factors, which include the number of CEL files, level of analysis, number of probes under consideration, amount of available RAM, and computer processor speed. See "Software and hardware requirements" on page 11.

The progress of the analysis can be tracked in the status window (Figure 67, Figure 68). Since Exon Arrays are processed as a batch, the CHP files are not added to the study until the completion of the analysis. The files are listed in the status window as they are generated; then the Study window (Figure 61) is populated with the corresponding CHP files.

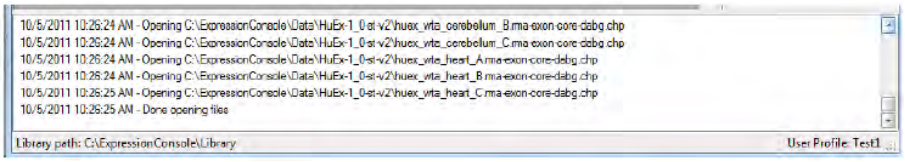


Figure 67 Exon core level analysis files in the status window.

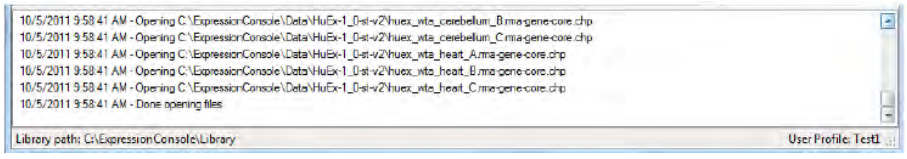


Figure 68 Gene core level analysis files in the status window.

Stop an analysis run

From the Analysis menu, select Stop (Figure 69).

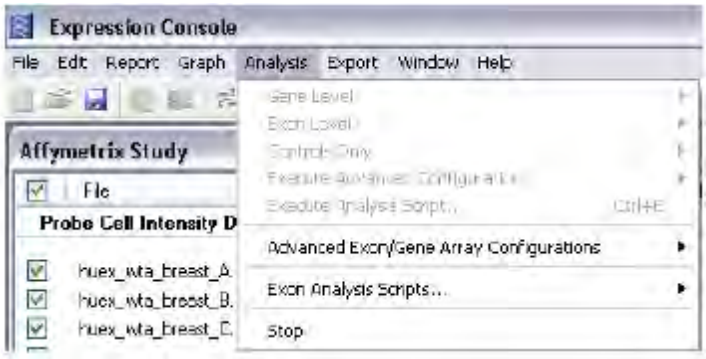


Figure 69 Canceling an analysis run.

A message will pop-up indicating the application is stopping the current analysis (Figure 70). In some cases it may not be possible to cancel the process immediately, but within a couple of minutes the operation will be cancelled.

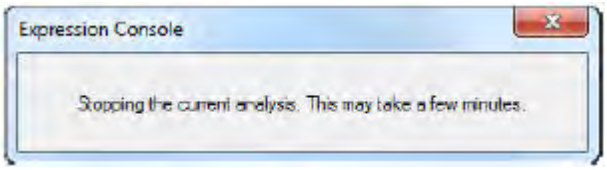


Figure 70 Notice.

After cancellation of the analysis run the user is notified that the analysis failed (Figure 71) and to check the status window (Figure 72).

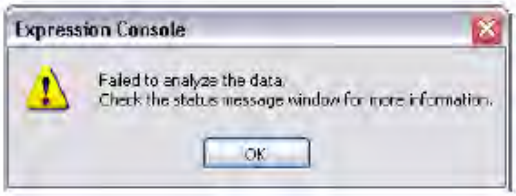


Figure 71 Message after canceling an analysis run.

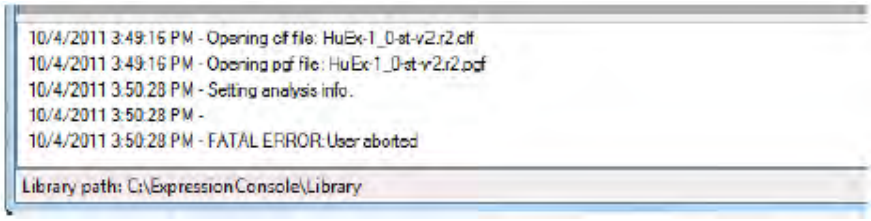


Figure 72 Status window after canceling an analysis run.

Results of Exon Array analysis

The CHP files are ready for QC analysis within the Expression Console™ software or other compatible software. CHP files are located in the same directory as the first CEL files in your input list. Outliers, samples outside the threshold boundaries, are highlighted in orange (Figure 73).

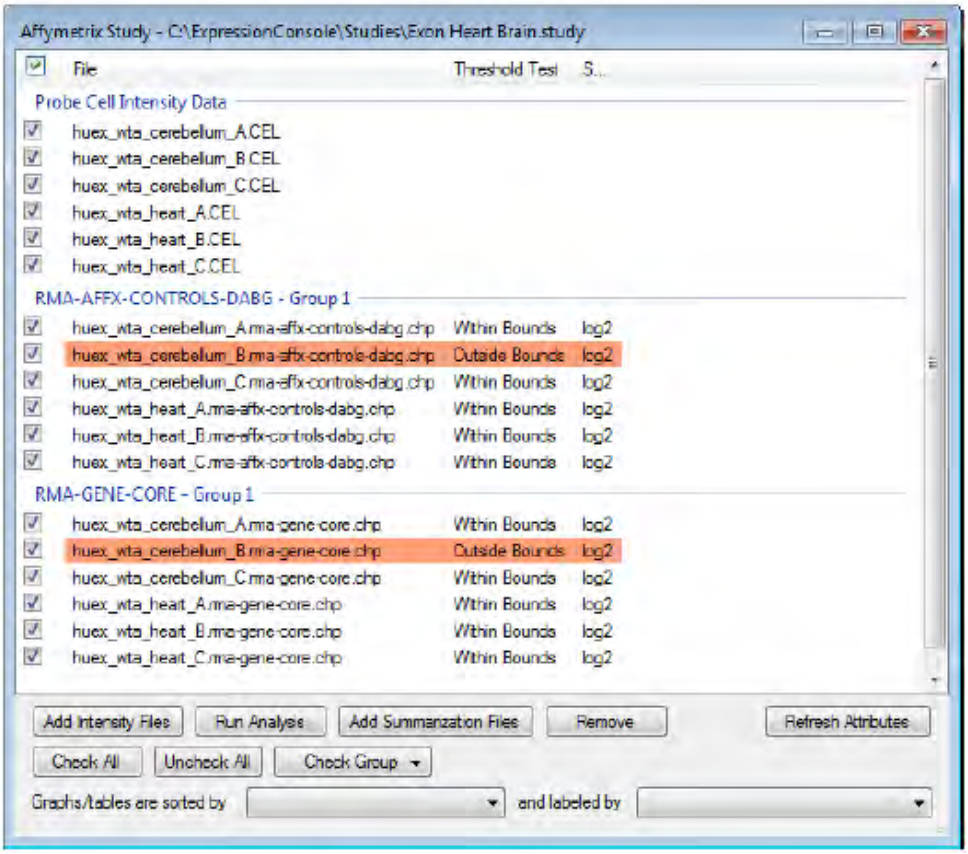


Figure 73 Study window - analysis results with outliers.

For information about viewing analysis results and QC data from the CHP files, see Chapter 9, “Viewing QC and analysis results data“.

For information about exporting data, see Chapter 11, “Exporting data“.

Note: You can further analyze results (CHP) that pass QC using Transcriptome Analysis Console. Select Tools → Transcriptome Analysis Console to start the software.

6 Gene Array analysis

Probe cell intensity data (CEL) from GeneChip™ Gene Arrays are analyzed in the Expression Console™ Software. The application uses the RMA-sketch workflow as the default analysis to create CHP files. Follow the instructions below to analyze Gene Array data.

IMPORTANT! You must have the necessary library files downloaded and installed to perform Gene Array analysis. See "Using library files" on page 17 for more information.

Analyze data from the Gene Arrays

1. Create a new study or open an existing study. See Chapter 3, "Creating a study".
2. In the Study dialog box (Figure 74), check the CEL files for analysis.

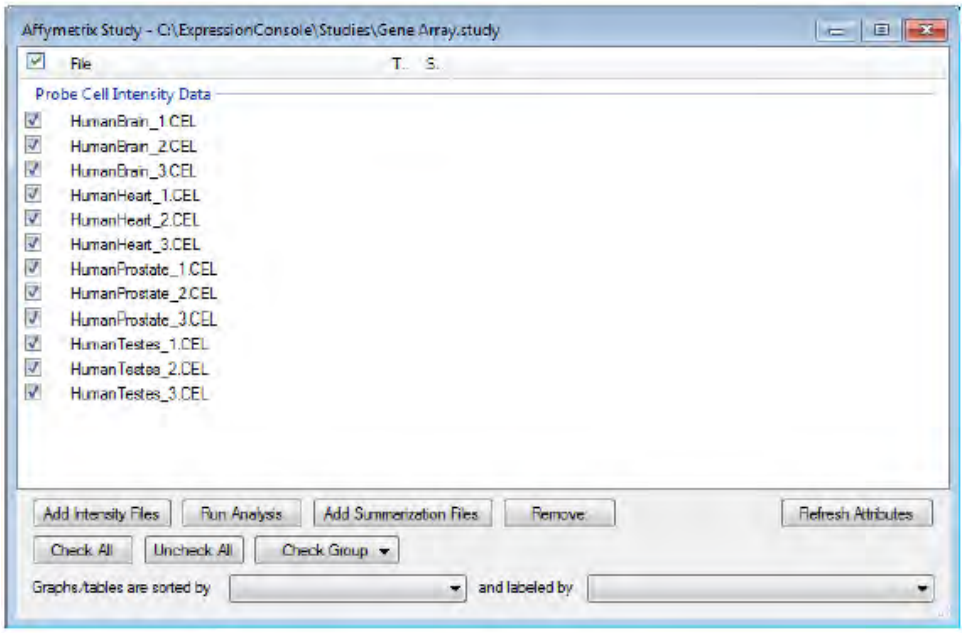


Figure 74 Study dialog box.

You can choose from several different analysis options, some of which have different options themselves.

- Gene Level
- Exon Level

The Gene and Exon Level analyses use different groupings of probe sets for analysis. and corresponds to different biology. The gene level case looks at the expression of the gene overall, and the exon level case looks at parts of the gene.

3. Select the analysis level by doing one of the following:
   - Select the appropriate workflow from the Analysis menu: Gene Level or Exon Level (Figure 75).

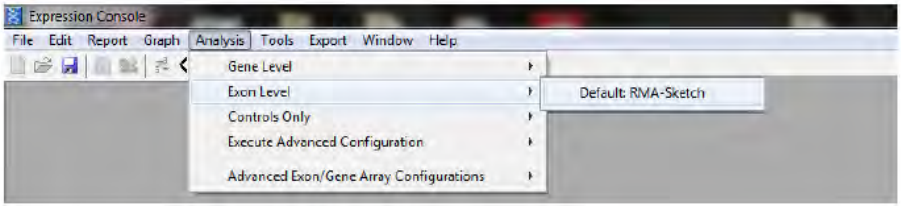


Figure 75 Analysis menu for gene arrays, exon level workflow selected.

Note: For information on using the Advanced Configuration option, see “Advanced configuration Exon/Gene analysis“ on page 140.

Or:

- In the Study dialog box (Figure 76):

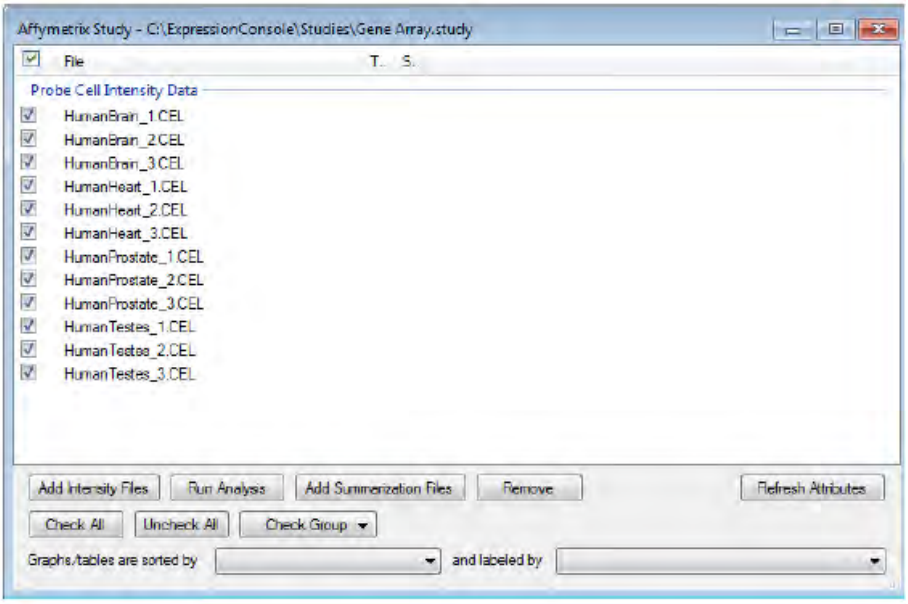


Figure 76 Study dialog box.

a. Click Run Analysis in the Study dialog box.
The Available Analysis dialog box opens (Figure 77).

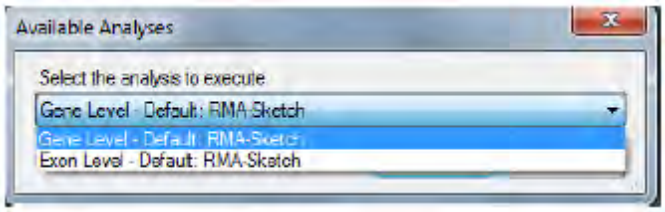


Figure 77 Available analysis dialog box.

b. Select the analysis level you wish to run.

c. Click OK in the Available Analysis dialog box.
The suffix dialog box for the selected workflow opens (Figure 78).

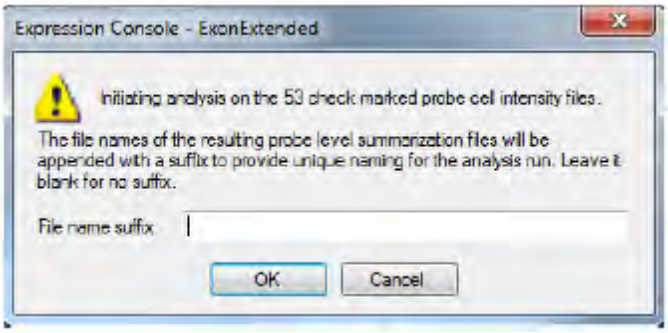


Figure 78 File name suffix dialog box.

Note: You can add a suffix to the file name in order to further identify your samples, or you can leave it blank. The summarization method is automatically included in the file name.

4. Enter a suffix to identify the specific samples after analysis.

5. Click OK; analysis begins.

IMPORTANT! The processing time for the CEL files depends on a number of factors, which include the number of CEL files, level of analysis, number of probes under consideration, amount of available RAM, and computer processor speed. See "Software and hardware requirements" on page 11.

The progress of the analysis can be tracked in the status window (Figure 79, Figure 80). Since Gene Arrays are processed as a batch, the CHP files are not added to the study until the completion of the analysis. The files are listed in the status window as they are generated; then the Study window (Figure 85) is populated with the corresponding CHP files.

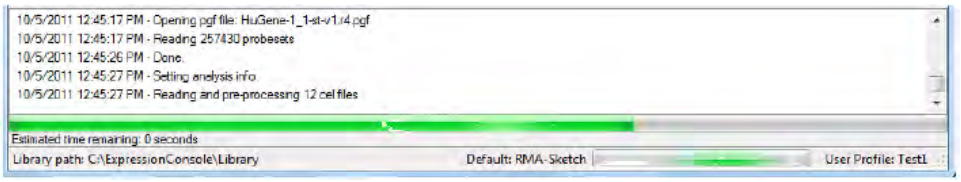


Figure 79 Status window with gene analysis running.

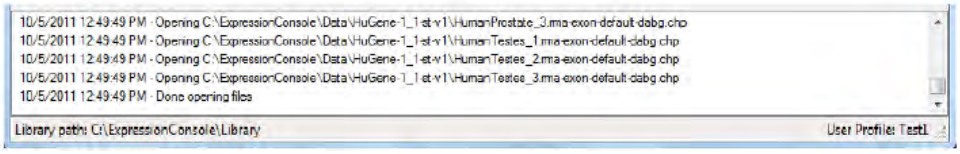


Figure 80 Exon level analysis files in the status window.

Stop an analysis run

From the Analysis menu, select Stop (Figure 81).

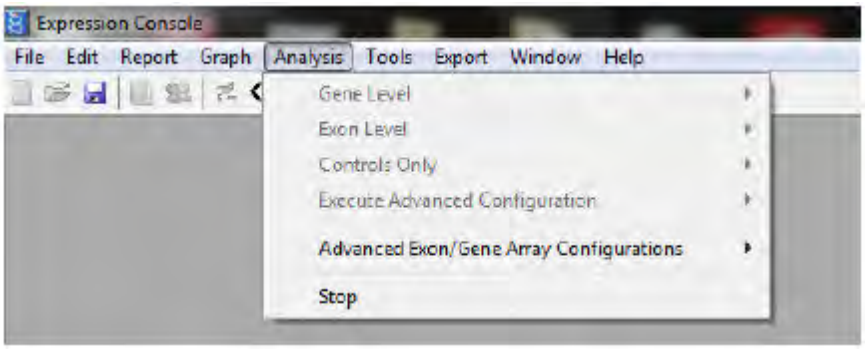


Figure 81 Canceling an analysis run.

A message will pop-up indicating the application is stopping the current analysis. In some cases it may not be possible to cancel the process immediately, but within a couple of minutes the operation will be cancelled.

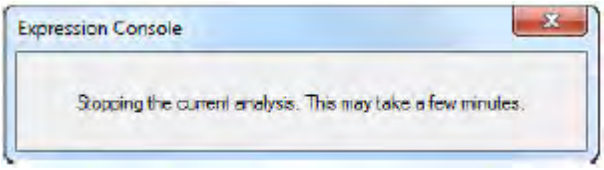


Figure 82 Notice.

After cancellation of the analysis run the user is notified that the analysis failed (Figure 83) and to check the status window (Figure 84).

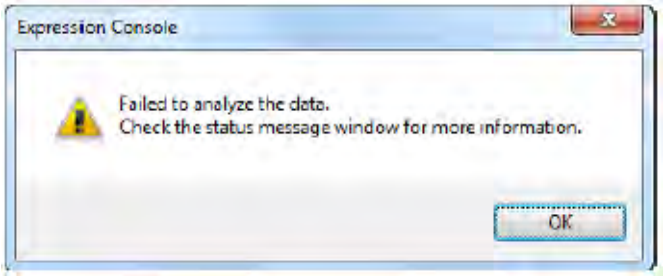


Figure 83 Message after cancelling an analysis run.

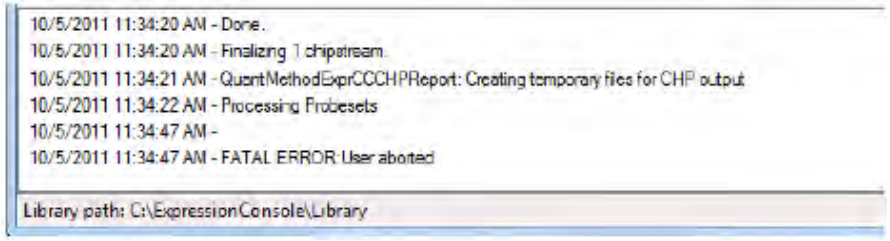


Figure 84 Status window after cancelling an analysis run.

Results of Gene Array analysis

After the analysis completes the CHP files will be added to the Study window (Figure 85), and CHP files are ready for QC analysis within the Expression Console™ Software or other compatible software. The CHP files are located in the same directory as the first CEL files in your input list.

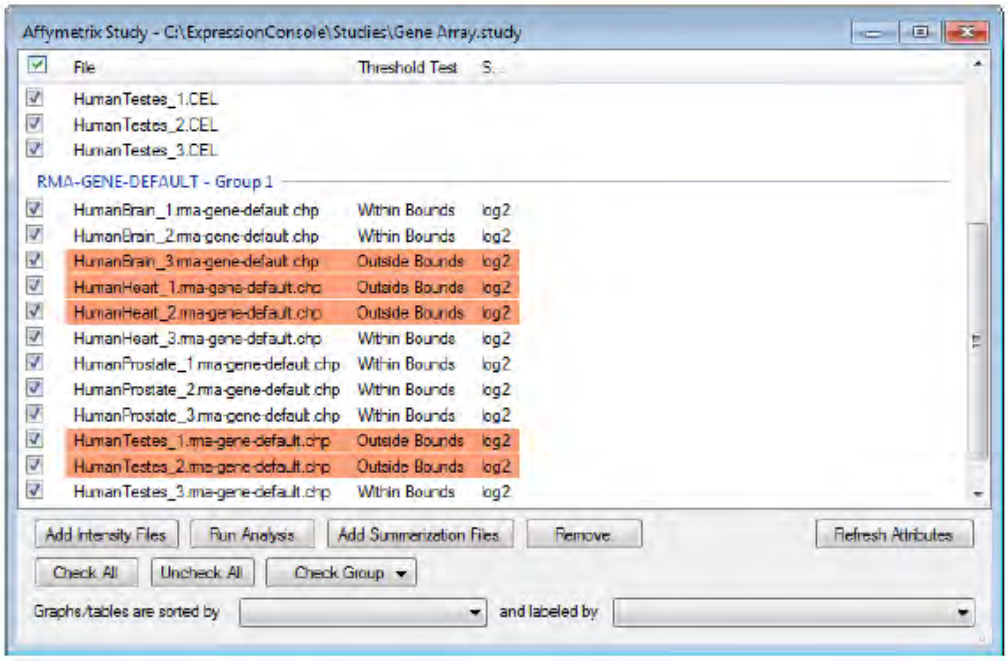


Figure 85 Study window - gene array analysis results with outliers.

Outliers, samples outside the threshold boundaries, are highlighted in orange.

For information about viewing analysis results and QC data from the CHP files, see Chapter 9, “Viewing QC and analysis results data“.

For information about exporting data, see Chapter 11, “Exporting data“.

Note: You can further analyze results (CHP) that pass QC using Transcriptome Analysis Console. Select Tools → Transcriptome Analysis Console to start the software.

7 miRNA Array analysis

Probe cell intensity data (CEL) from GeneChip™ miRNA Arrays are analyzed in the Expression Console™ Software. The application uses the RMA + DABG analysis as the default analysis to create CHP files. Follow the instructions below to analyze miRNA Array data.

IMPORTANT! You must have the necessary library files downloaded and installed to perform miRNA Array analysis. See "Using library files" on page 17 for more information.

Analyze data from miRNA arrays

1. Create a new study or open an existing study. See Chapter 3, "Creating a study".
2. In the Study dialog box (Figure 86), check the CEL files for analysis.

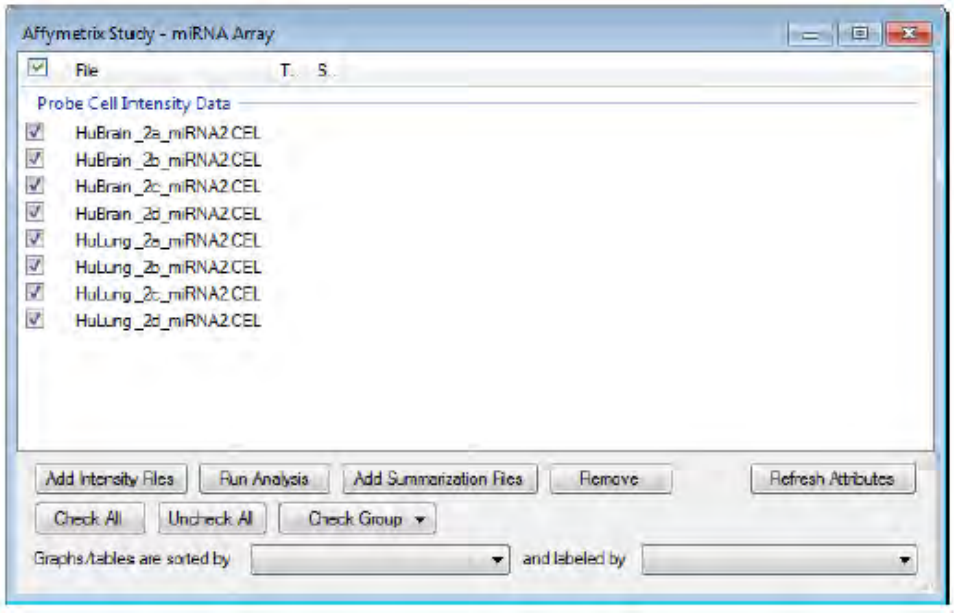


Figure 86 Study dialog box.

3. Select the analysis by doing one of the following:
   - Select the RMA + DABG workflow from the Analysis menu (Figure 87):

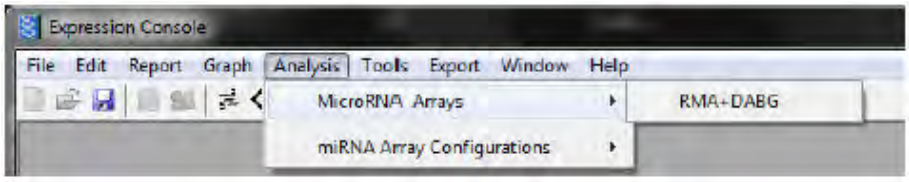


Figure 87 Analysis menu for miRNA arrays.

Note: For more details on viewing miRNA Configuration information, see "miRNA array configurations" on page 145.

Or:

   - In the Study dialog box (Figure 88):

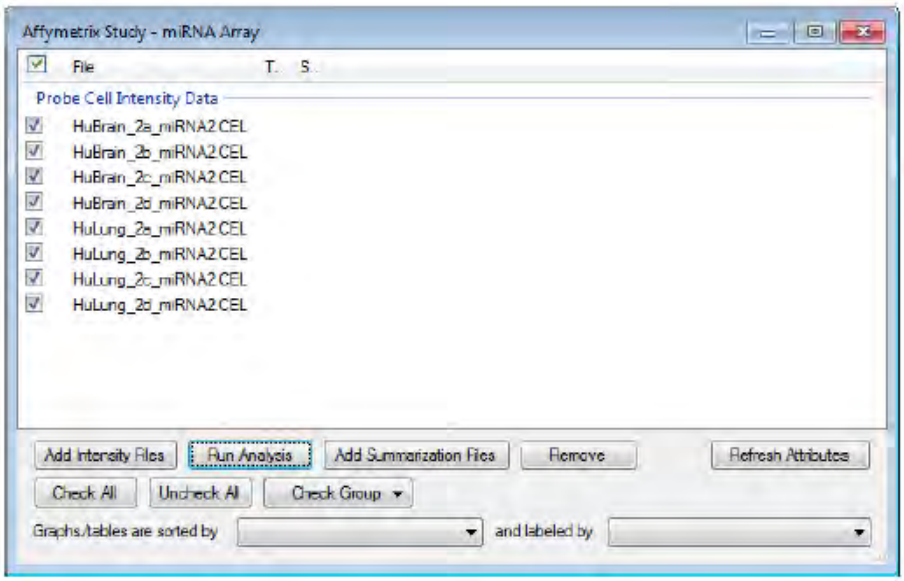


Figure 88 Study dialog box.

   a. Click Run Analysis in the Study dialog box.
   The Available Analysis dialog box opens (Figure 89).

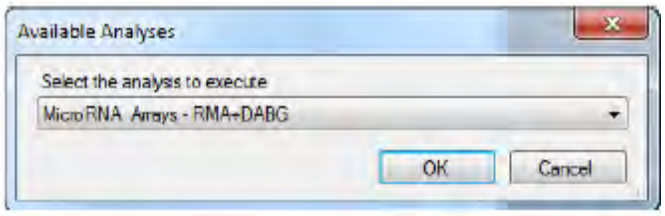


Figure 89 Available analysis dialog box.

   b. Select the MicroRNA Arrays RMA + DABG workflow.

c. Click OK in the Available Analysis dialog box.
The suffix dialog box for the workflow opens (Figure 90).

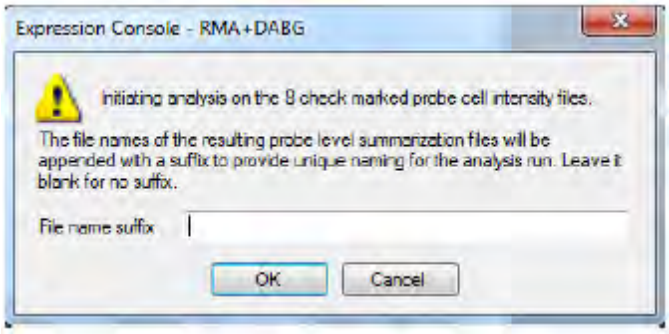


Figure 90 Expression Console™ Software - MicroRNA array RMA + DABG workflow-file name suffix dialog box.

Note: You can add a suffix to the file name in order to further identify your samples, or you can leave it blank. The summarization method is automatically included in the file name.

4. Enter a suffix to identify the specific samples after analysis.

5. Click OK.

IMPORTANT! The processing time for the CEL files depends on a number of factors, including the number of CEL files, level of analysis, number of probes under consideration, amount of available RAM, and computer processor speed. See "Software and hardware requirements" on page 11.

The progress of the analysis can be tracked in the status window (Figure 91, Figure 92). Since miRNA Arrays are processed as a batch, the CHP files are not added to the study until the completion of the analysis. The files are listed in the status window as they are generated; then the Study window is populated with the corresponding CHP files (Figure 97).

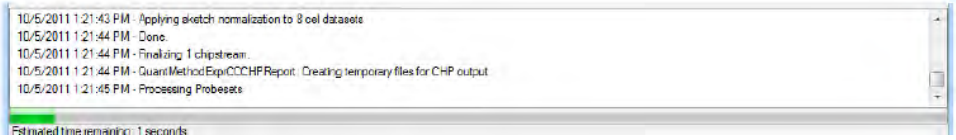


Figure 91 Status window with MicroRNA array analysis running.

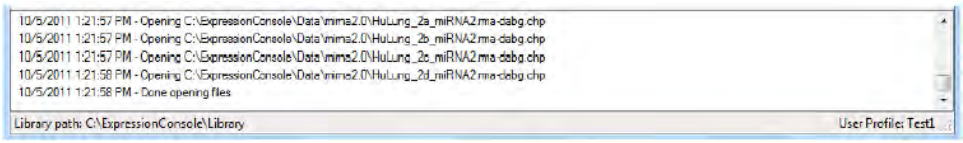


Figure 92 MicroRNA array analysis files in the status window.

Stop an analysis run

From the Analysis menu, select Stop (Figure 93).

Figure 93 Canceling an analysis run.

A message will pop-up indicating the application is stopping the current analysis (Figure 94). In some cases it may not be possible to cancel the process immediately, but within a couple of minutes the operation will be cancelled.

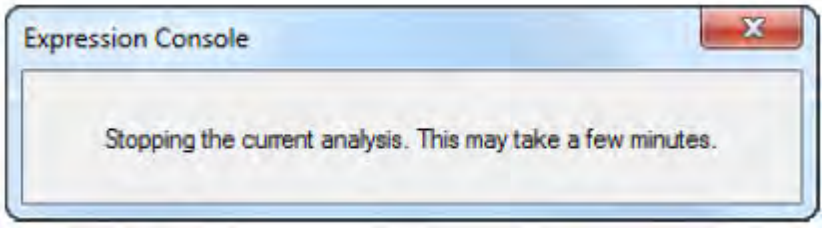


Figure 94 Stopping analysis notice.

After cancellation of the analysis run the user is notified that the analysis failed (Figure 95) and to check the status window (Figure 96).

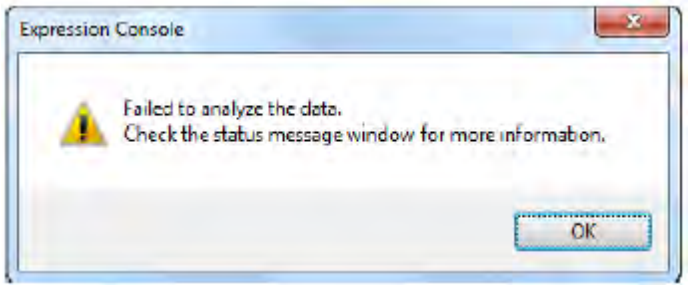


Figure 95 Message after canceling an analysis run.

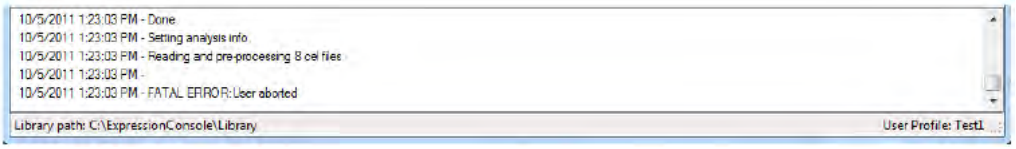


Figure 96 Status window after canceling an analysis run.

Results of miRNA analysis

After the analysis completes the CHP files will be added to the Study window, and CHP files are ready for QC analysis within theExpression Console™ Software or other compatible software. The CHP files are located in the same directory as the first CEL files in your input list.

Outliers, samples outside the threshold boundaries, are highlighted in orange (Figure 97).

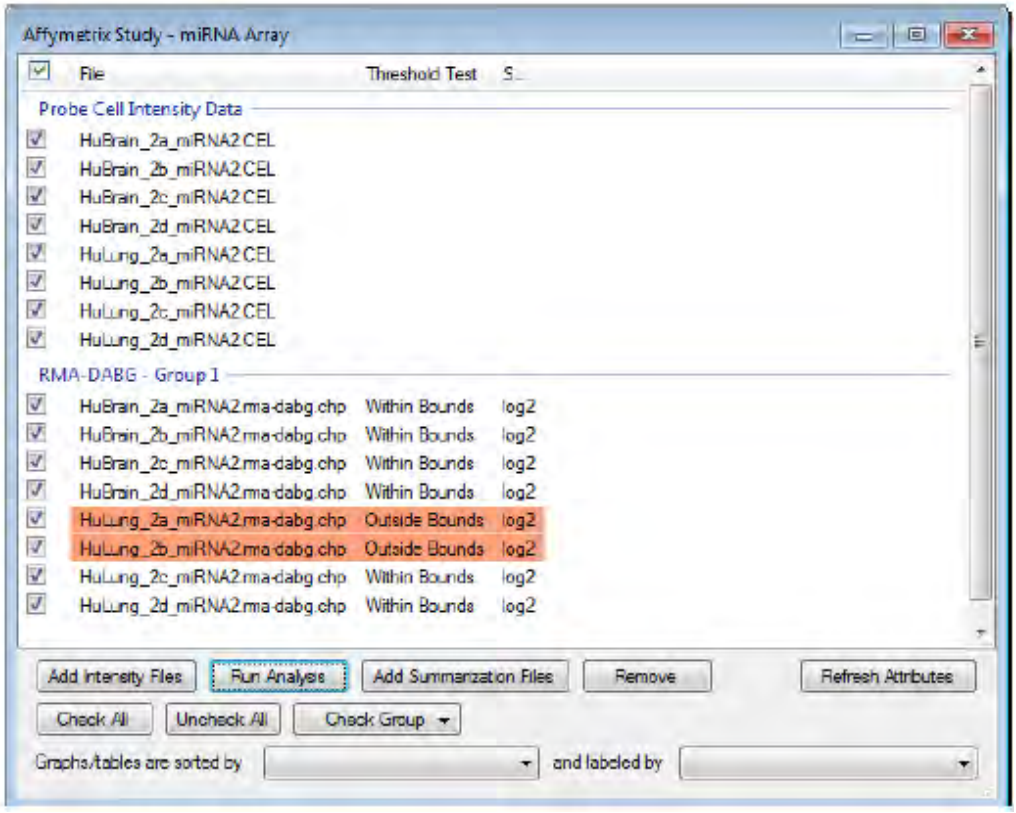


Figure 97 Study window - miRNA array analysis results with outliers.

For information about viewing analysis results and QC data from the CHP files, see Chapter 9, “Viewing QC and analysis results data“.

For information about Exporting Data, see Chapter 11, “Exporting data“.

Note: You can further analyze results (CHP) that pass QC using Transcriptome Analysis Console. Select Tools → Transcriptome Analysis Console to start the software.

8 Human Transcriptome Array analysis

Probe cell intensity data (CEL) from GeneChip™ Human Transcriptome Arrays are analyzed in the Expression Console™ Software. The application uses the RMA analysis to create CHP files. Follow the instructions below to analyze Human Transcriptome Array data.

IMPORTANT! You must have the necessary library files downloaded and installed to perform Human Transcriptome Array analysis. See "Using library files" on page 17 for more information.

Analyze data from Human Transcriptome Arrays

1. Create a new study or open an existing study. See Chapter 3, "Creating a study".
2. In the Study dialog box (Figure 98), check the CEL files for analysis.

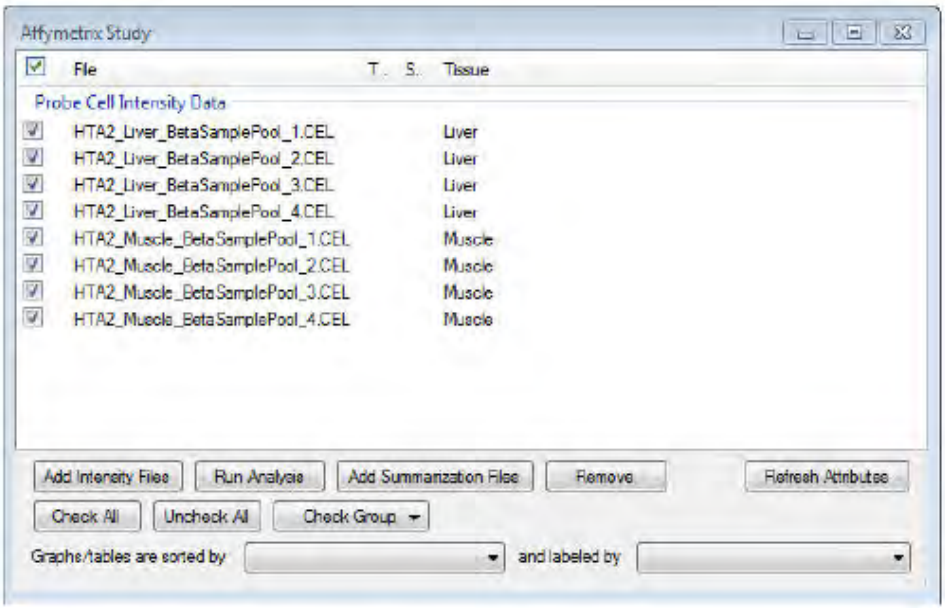


Figure 98 Study dialog box.

3. Select the analysis by doing one of the following:
   - From the Analysis menu, choose the **Gene Level ▸ SST-RMA** workflow or the **Exon Level ▸ SST-Alt Splice Analysis** workflow. (Figure 99)

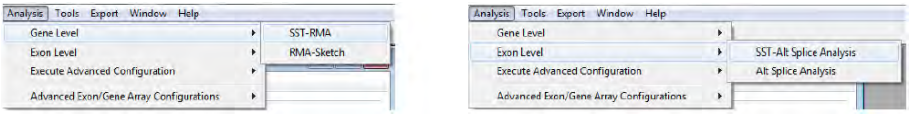


Figure 99 Analysis menu for Human Transcriptome Arrays.

Or:

   - In the Study dialog box (Figure 100):

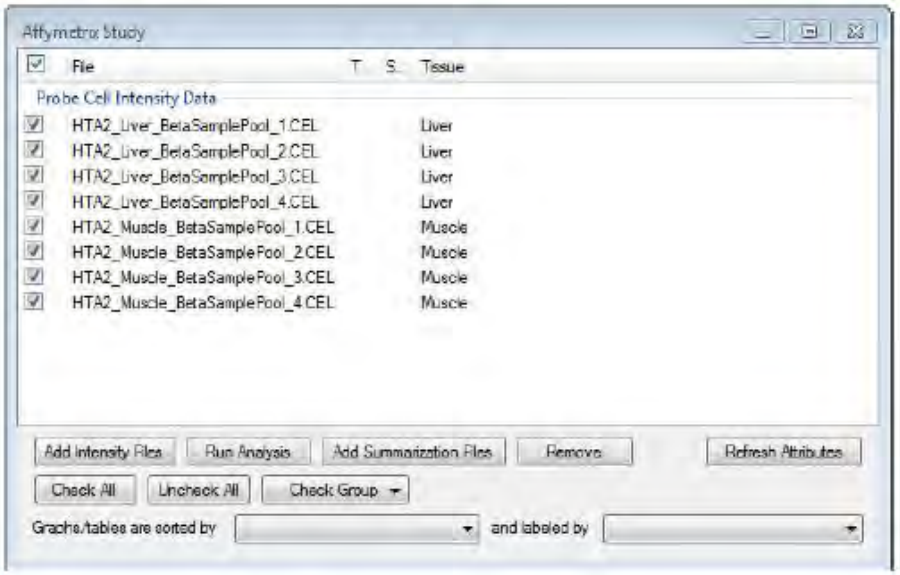


Figure 100 Study dialog box.

   a. Click Run Analysis in the Study dialog box.
   The Available Analysis dialog box opens (Figure 101).

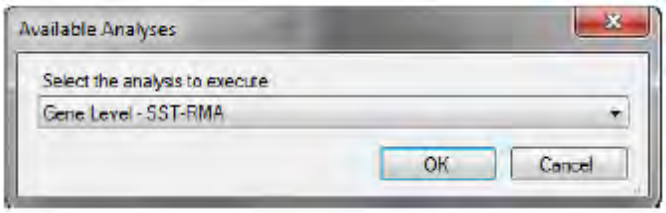


Figure 101 Available analysis dialog box.

b. Select the default Gene Level - SST-RMA or Exon Level - SST-Alt Splice Analysis workflow, then click OK.

The suffix dialog box for the workflow opens (Figure 102).

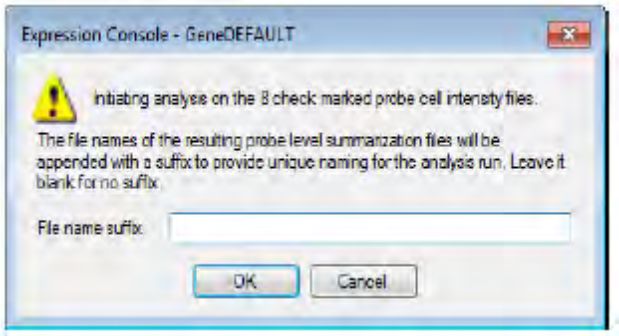


Figure 102 Expression Console™ Software - Human Transcriptome Array RMA + DABG workflow-file name suffix dialog box.

Note: You can add a suffix to the file name in order to further identify your samples, or you can leave it blank. The summarization method is automatically included in the file name.

c. Enter a suffix to identify the specific samples after analysis.

d. Click OK; analysis begins.

IMPORTANT! The processing time for the CEL files depends on a number of factors, which include the number of CEL files, level of analysis, number of probes under consideration, amount of available RAM, and computer processor speed. See "Software and hardware requirements" on page 11.

The progress of the analysis can be tracked in the status window (Figure 103, Figure 104). Because Human Transcriptome Arrays are processed as a batch, the CHP files are not added to the study until the completion of the analysis. The files are listed in the status window as they are generated; then the Study window is populated with the corresponding CHP files (Figure 109).

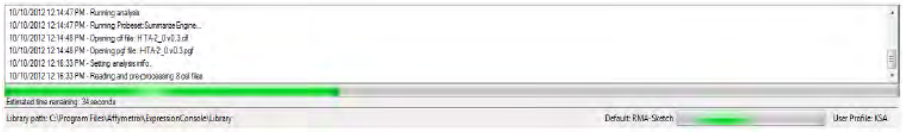


Figure 103 Status window with Human Transcriptome Array analysis running.

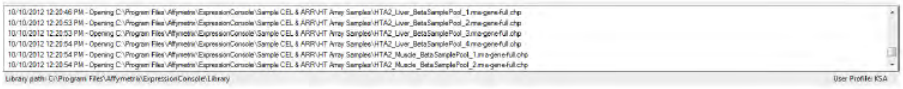


Figure 104 Human Transcriptome Array analysis files in the status window.

Stop an analysis run

From the Analysis menu, select Stop. (Figure 105)

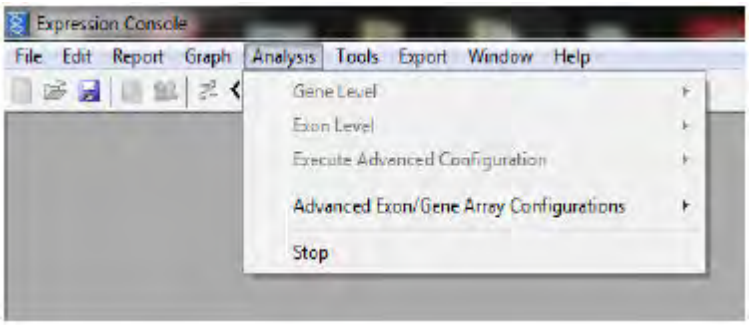


Figure 105 Canceling an analysis run.

A message will pop-up indicating the application is stopping the current analysis (Figure 106). In some cases it may not be possible to cancel the process immediately, but within a couple of minutes the operation will be cancelled.

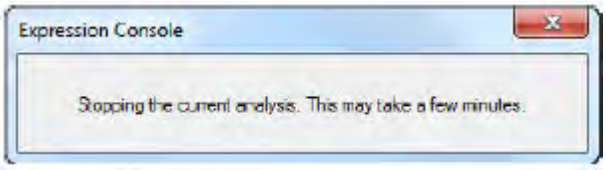


Figure 106 Stopping analysis notice.

After cancellation of the analysis run the user is notified that the analysis failed (Figure 107) and to check the status window (Figure 108).

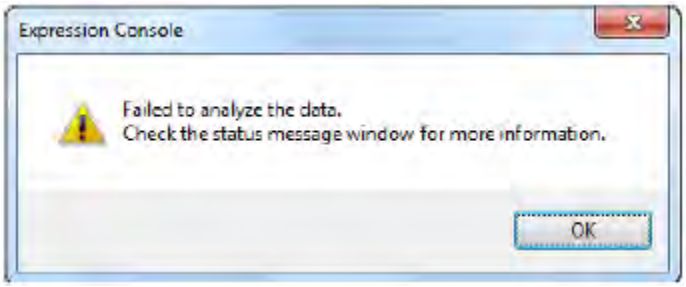


Figure 107 Message after canceling an analysis run.

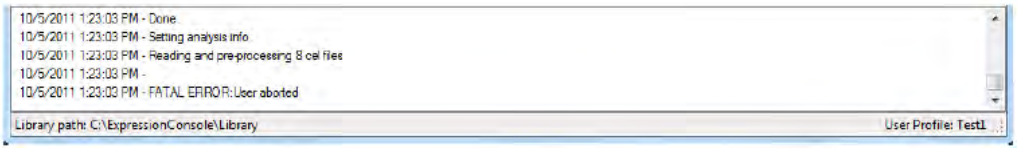


Figure 108 Status window after canceling an analysis run.

Results of Human Transcription Array

After the analysis completes the CHP files will be added to the Study window, and CHP files are ready for QC analysis within theExpression Console™ Software or other compatible software. The CHP files are located in the same directory as the first CEL files in your input list.

Outliers, samples outside the threshold boundaries, are highlighted in orange (Figure 109).

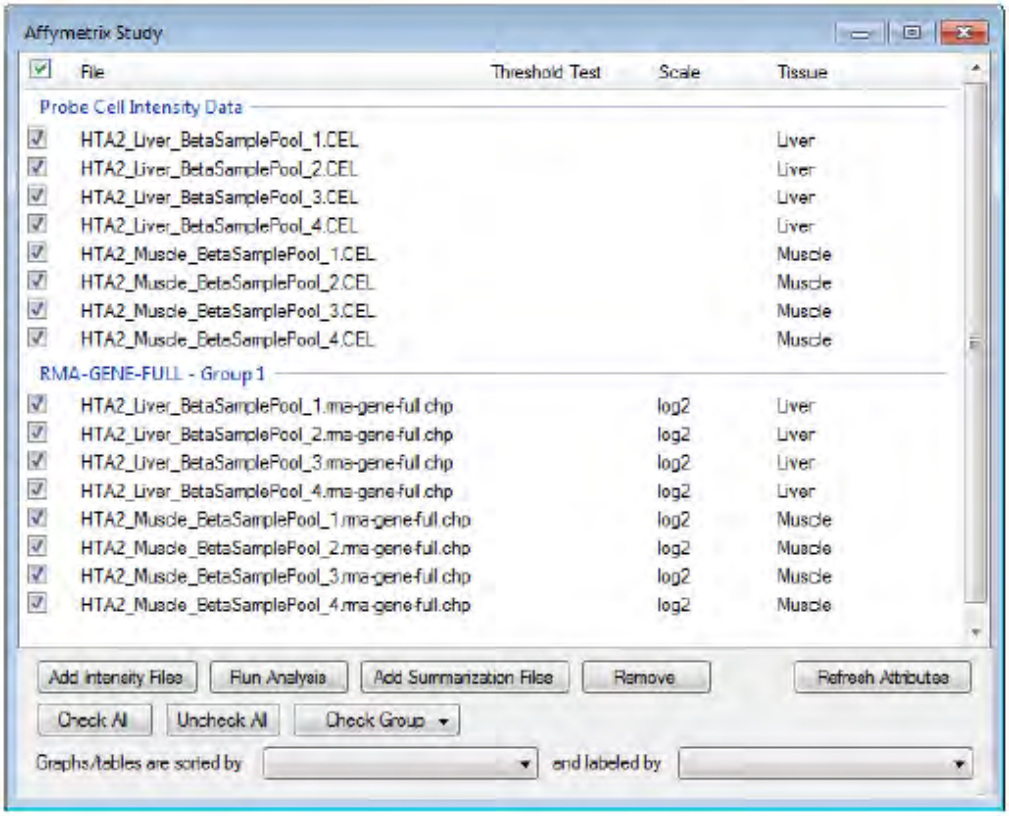


Figure 109 Study window - Human Transcriptome Array analysis results.

For information about viewing analysis results and QC data from the CHP files, see Chapter 9, “Viewing QC and analysis results data“.

For information about Exporting Data, see Chapter 11, “Exporting data“.

9 Viewing QC and analysis results data

Expression Console™ Software provides several types of reports in table format and graphs for viewing QC metrics and other data.

QC reports

Expression Console™ Software provides a variety of QC metrics to evaluate results and tools for viewing the metrics. These metrics and tools are described in:

- "Introduction to QC metrics" on page 77
- "Displaying a report" on page 78
- "Custom Report" on page 80

Note: You can further analyze results (CHP) that pass QC using Transcriptome Analysis Console. Select **Tools ▸ Transcriptome Analysis Console** to start the software.

Introduction to QC metrics

To help researchers establish quality control processes for gene expression analyses, we have developed several controls. Researchers are encouraged to monitor these controls on a regular basis to assess assay data quality. Many of these control metrics are generated as a product of the primary analysis by the respective analysis algorithms. To help with monitoring these values, a collection of tools for viewing and graphing metrics associated with the array and the analysis are provided in the Expression Console™ Software. These metrics include but are not limited to:

- Hybridization controls
- Labeling controls
- Internal control genes (Housekeeping controls)
- Global array metrics
- Algorithm parameters

This section describes how to generate graphs and tables allowing the quality of individual hybridizations included in a single study to be easily assessed and gives some introductory guidance on their interpretation. For more detailed information about graph and table interpretation, download "White Paper: Quality Assessment of Exon Arrays", on our **website**.

In general, we highly encourage users to create a running log of the parameters to monitor quality and potentially flag outlier samples. Evaluation of particular samples should be based on the examination of all sample and array performance metrics in

light of this history. A good, general rule for the examination of these types of quality control data is to look for outliers when compared to other highly related samples. For example, tissue A may have an overall low level of gene expression so that the percent of probes detected may normally be between 10–15%. Therefore, a %Present = 11 would not be an indication of a problem. However, if tissue B normally has a %Present between 45–50, and one sample has %Present = 11, that is an indication of a problem. Examination of which metrics or controls are outliers will provide insight into the source of the problem and possible solutions.

The Expression Console™ Software has the ability to store user defined thresholds for the QC metrics and highlight, in the reports and study tables, those metrics outside the thresholds and arrays that contain them. The Expression Console™ Software comes with an initial set of defaults for both metrics and thresholds, which the user can modify according to individual preferences. For more information, see Chapter 12, "Advanced analysis".

Reports available in the Expression Console™ Software provide a tabular view of all QC metrics, including:

- Algorithm parameters (See Appendix B, "Algorithm parameters and outputs" for a list of parameters, outputs, and their definitions.)
- Global analysis metrics
- Corner +, Corner -, Center +, and Center -, if available
- Count and percentages of detection calls
- Average signal value for each detection call type
- Signals, detections, and 3'/5' ratio for spike and housekeeping controls

Note that these metrics are algorithm specific. For details about the MAS 5.0 algorithm parameters, see the GeneChip™ Operating System Software (GCOS) manual (Cat. No. 701439) and the Statistical Algorithms Reference Guide. For details about the PLIER algorithm parameters, see the Guide to Probe Logarithmic Intensity Error (PLIER) Estimation technical note and the Quality Assessment of Exon and Gene Arrays white paper on our web site.

The Expression Console™ Software provides these reports:

- "Display a full report" on page 78
- "Display a transpose report" on page 79
- "Custom Report" on page 80

Displaying a report

Display a full report

The Full Report displays a list of the selected CHP files with the full set of appropriate algorithm parameters and array QC metrics.

1. Select the CHP files you wish to review in the Study dialog box.

   STOPPING POINT You can use the Study dialog box controls to control the order that files are displayed in the report. See "Displaying files and data" on page 39

2. From the Report menu, select View Full Report, or
   From the Toolbox, select QC: **Array Metrics ▸ Tabular Report**.

The Full Report table opens (Figure 110).

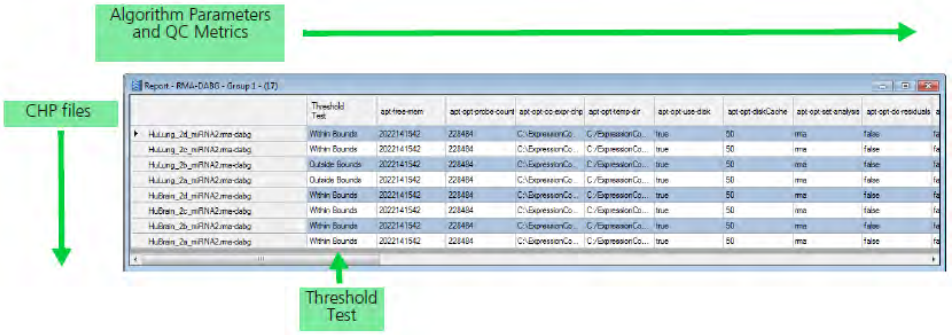


Figure 110 Full report - miRNA with RMA-DBAG.

The table displays a list of the selected CHP files with all available metrics displayed.

The Threshold Test column displays whether or not all the metrics are within or one or more metrics are outside of the user-specified thresholds.

Metrics outside of the user-specified thresholds are highlighted (Figure 111).

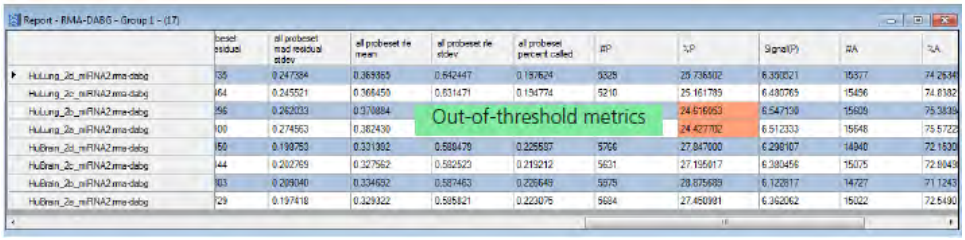


Figure 111 Out-of-threshold metrics highlighted.

Display a transpose report

The Transpose Report displays a list of the full set of appropriate algorithm parameters and array QC metrics for the selected CHP files.

1. Select the CHP files you wish to review in the Study dialog box.

STOPPING POINT You can use the Study dialog box controls to control the order that files are displayed in the report. See "Displaying files and data" on page 39.

2. From the Report menu, select View Report (transposed), or

3. In the Toolbox, select **QC: Array Metric ▸ Tabular Report (transposed)**.
   The Pivot Table Report table opens (Figure 112).

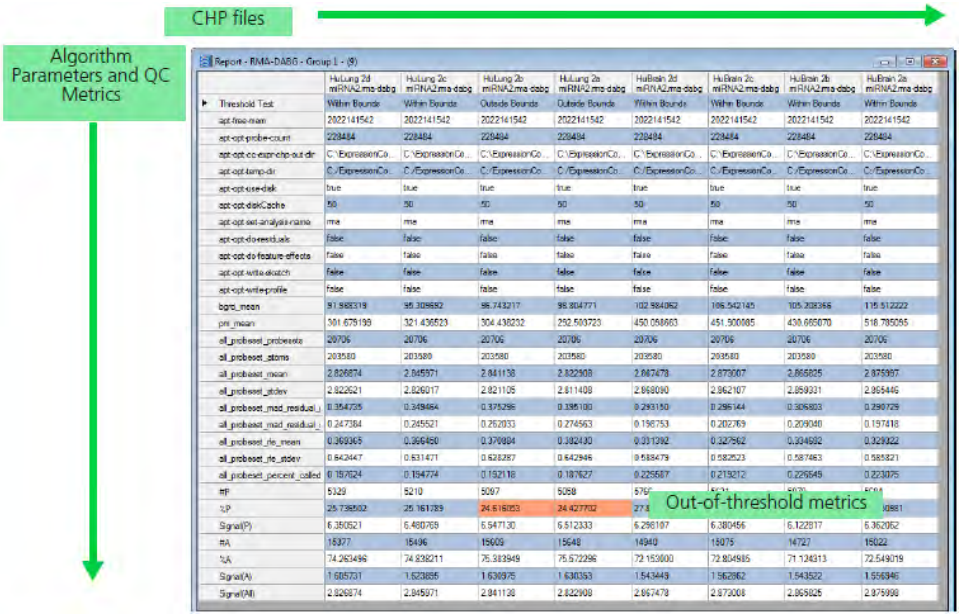


Figure 112 Pivot table report.

Custom Report

You can use the Custom Report feature to define a subset of columns (report metrics) for the tabular display. Using Expression Console™ Software, you can:

- "Create a new report format" on page 81
- "Create a custom report" on page 82
- "Delete a custom report format" on page 84

Create a new report format

1. From the Report menu, select New Report.
   The Report dialog box opens (Figure 113).

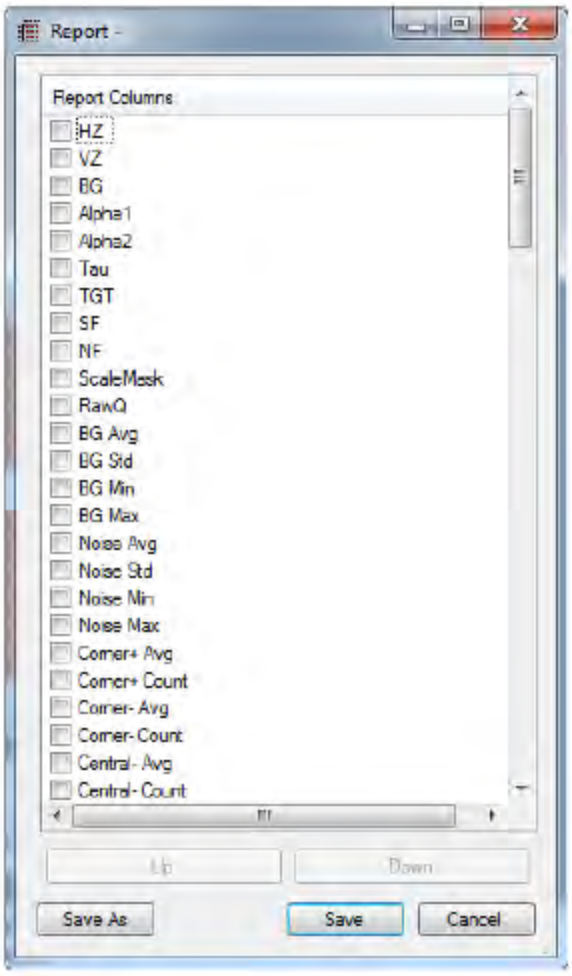


Figure 113 Report dialog box.

The Report dialog box displays a list of the metrics available for the particular array and algorithms in the selected CHP files.

Note: The list of available metrics to plot is based on the CHP files in the study. For example, if you have MAS 5.0 CHP files open, then parameters associated with PLIER are not visible.

IMPORTANT! The categories available for selection in the Report window depend on the expression array selected and the algorithms used to generate the CHP files in the study.

---

2. Click in the checkbox next to a metric name to select the metric for the custom report.

3. Use the Up and Down buttons to change the position of a metric in the list. This will change the position that the metric is displayed in the report.

4. Click Save or Save as.
   The Save dialog box opens (Figure 114).

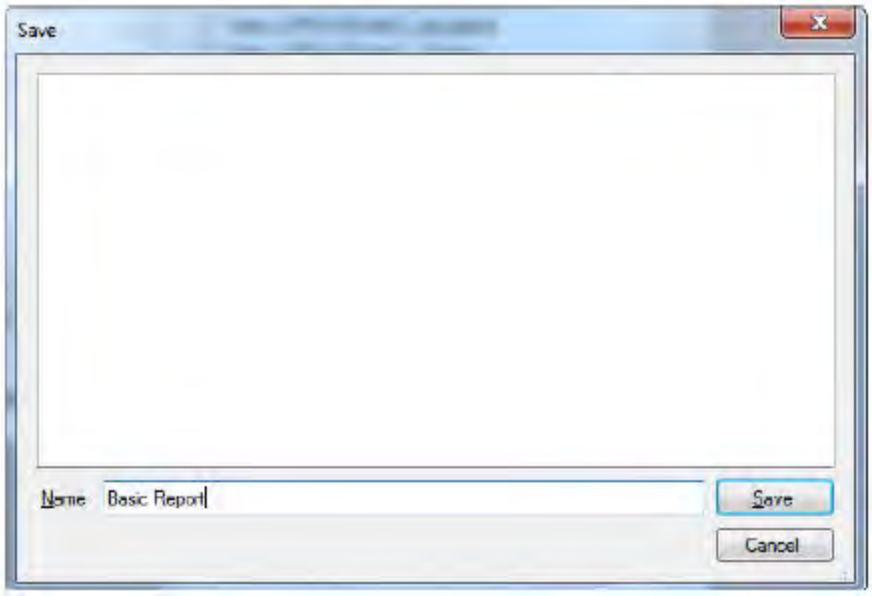


Figure 114 Save dialog box.

5. Enter a name for the format in the Name box or select a previously created format name.

6. Click Save.
   If you previously clicked Save As in the Report dialog box, the Save dialog box closes but the Report dialog box remains open.
   If you previously clicked Save in the Report dialog box, both dialog boxes close.

Create a custom report

The report format can be used to create a custom report.

1. Select the CHP files you wish to review in the Study dialog box.

STOPPING POINT You can use the Study dialog box controls to control the order that files are displayed in the report. See "Displaying files and data" on page 39.

2. From the Report menu, select View report.
   The Select Report dialog box opens (Figure 115).

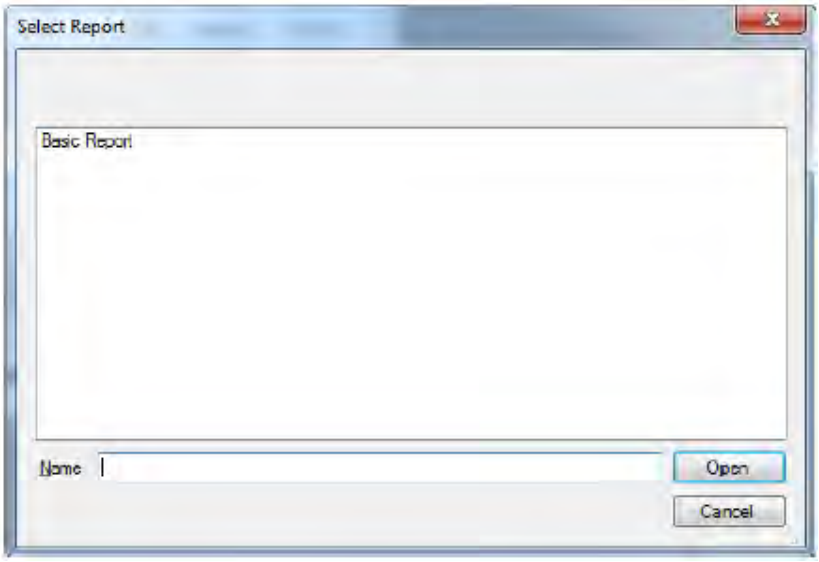


Figure 115 Select Report dialog box.

3. Select the report format you wish to use from the list and click Open.
   The Custom Report opens for the selected CHP files (Figure 116).

Figure 116 Custom Report.

Delete a custom report format

1. From the File menu, select Utilities.
   The File Management Utilities dialog box opens.
2. Select the Parameter Management tab (Figure 117).

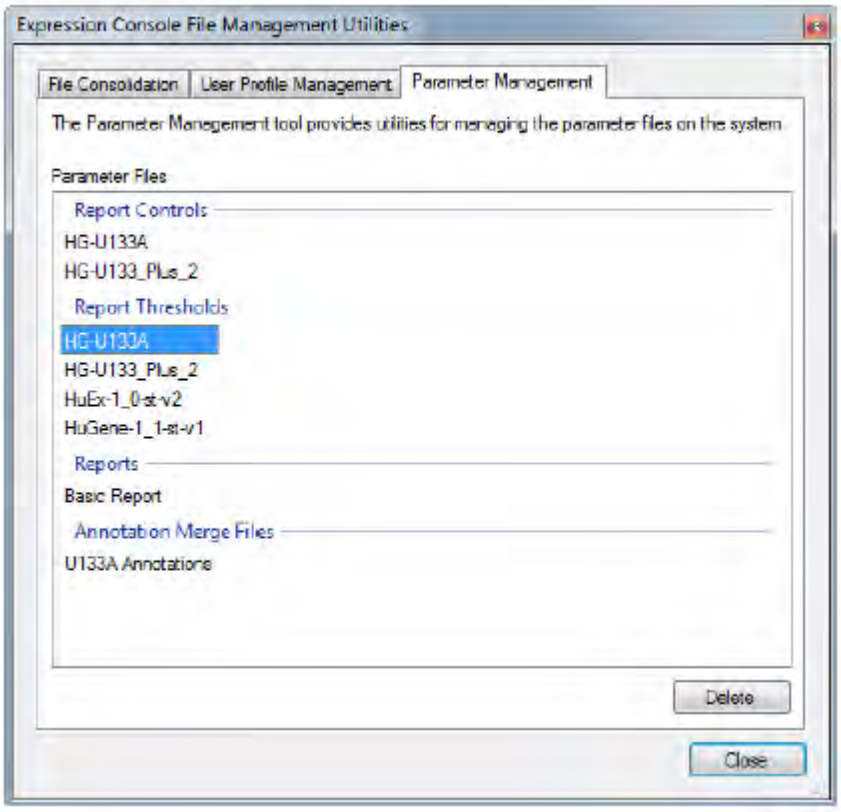


Figure 117 Paramater Management tab.

3. In the Custom Reports section, select the report to be deleted.
4. Select Delete and then Close.

Displaying analysis results

The Probe Set Results table (Figure 118) displays the probe set results from the CHP files in the study.

Display the probe level summarization report

1. Select the CHP files you wish to display in the Study dialog box.

STOPPING POINT You can use the Study dialog box controls to control the order that files are displayed in the report. See “Displaying files and data“ on page 39.

2. From the Reports menu, select View Probe Level Summarization Report.
   The Probe Set Results table opens (Figure 118).

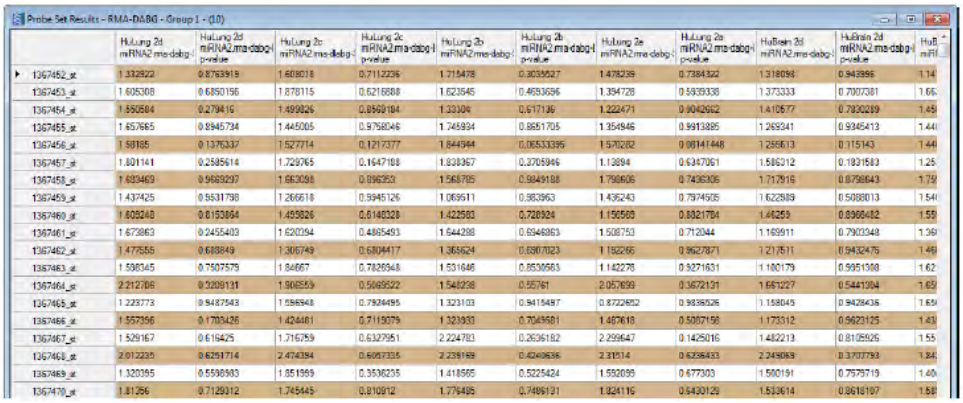


Figure 118 Probe set results table.

Each row in the table displays values for a particular probe set. Each column displays a selected value for a probe set from a CHP file.

More than one value and one column may be displayed for a single CHP file.

Change the values displayed

1. From the Edit menu, select Probe Level Summarization Report Options.
   The Probe Level Summarization Report Options dialog box opens.
   This dialog box is also used to determine the values exported.

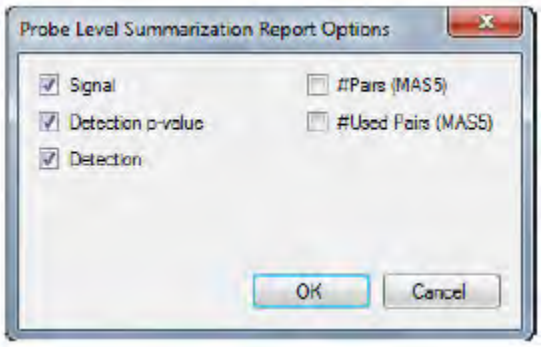


Figure 119 Probe level summarization report options.

2. Select the options to be displayed in the probe set results:
   - Signal
   - Detection p-value
   - Detection
   - #Pairs (MAS5)
   - #Used Pairs (MAS5)

Graphs

In addition to the tabular display of the metrics, the Expression Console™ Software supports several different graphical views, including signal histograms, box plots, line graphs, and heat maps depicting the correlation between arrays. The graphical displays, especially the box plots and heat maps, are useful in identifying outlier samples.

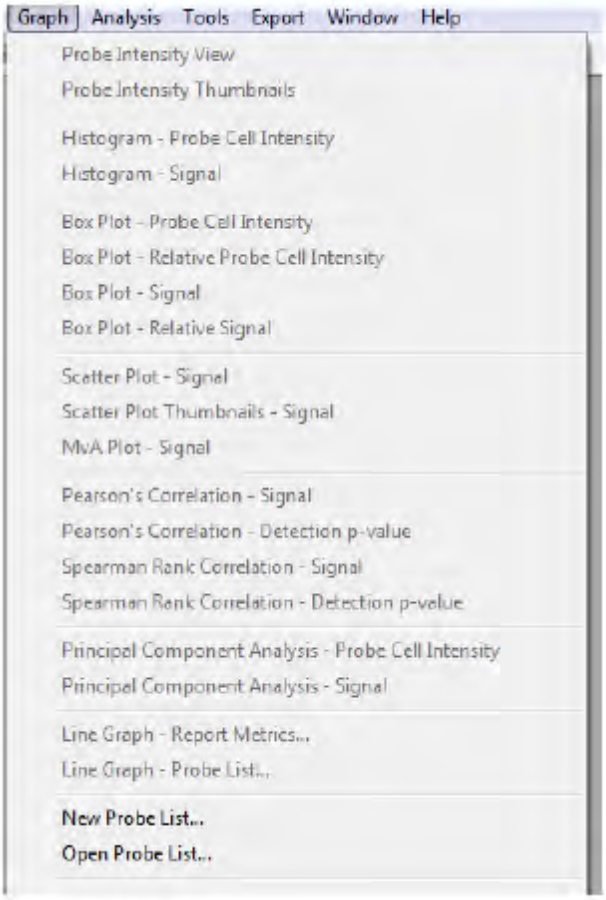


Figure 120 Graph menu items.

The different types of graphs include:

- “Probe intensity views“ on page 87
- “Histograms“ on page 89
- “Box plot“ on page 90
- “Scatter plot“ on page 94
- “MvA plot“ on page 98
- “Correlations“ on page 100
- “Line graphs“ on page 103
- “Principle component analysis“ on page 107

Probe intensity views

The Probe Cell Intensity View (**Graph ▸ Probe Cell Intensity View**) is a view of the CEL intensities arranged by physical position on the array. This view is used to quickly determine if there is any major issue with the image. For in-depth image inspection and gridding evaluation, it is recommended that customers use GCOS or the manual grid application available within the GeneChip™ Command Console™ Software package.

Display probe cell intensity views

1. Select probe cell intensity data (CEL) in the study window.
2. Select **Graph ▸ Probe Intensity View** from the menu bar.
   An image of probe cell intensities is displayed for each CEL file.

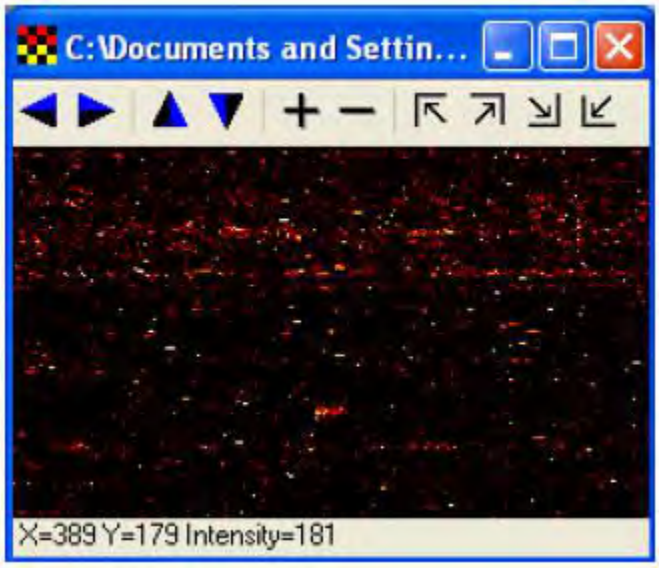


Figure 121 Probe cell intensity view.

The display area can be adjusted using the arrow keys, while the zoom level is controlled by the + and - buttons. Moving the mouse over a cell in the image displays the features XY coordinate and intensity values. Selecting a region (click and drag the mouse) zooms the display to show the selected region. Right clicking on the image enables the following options:

- Select a gray scale palette (Gray)
- Select a blue scale palette
- Select a color scale palette that goes from black to red to yellow to white (Heat)
- Adjust the range for the color scale (Scale)
- Save the image in PNG format
- Copy the image to the clipboard (Copy to clipboard)

Note: A particular file may appear solid black when you first open the Probe Cell Intensity View. Use the View → Zoom menu commands to zoom in on a particular area of investigation.

Display probe cell intensity thumbnails

1. Select probe cell intensity data (CEL) in the study window.
2. Select **Graph ▸ Probe intensity Thumbnails** on the menu bar.

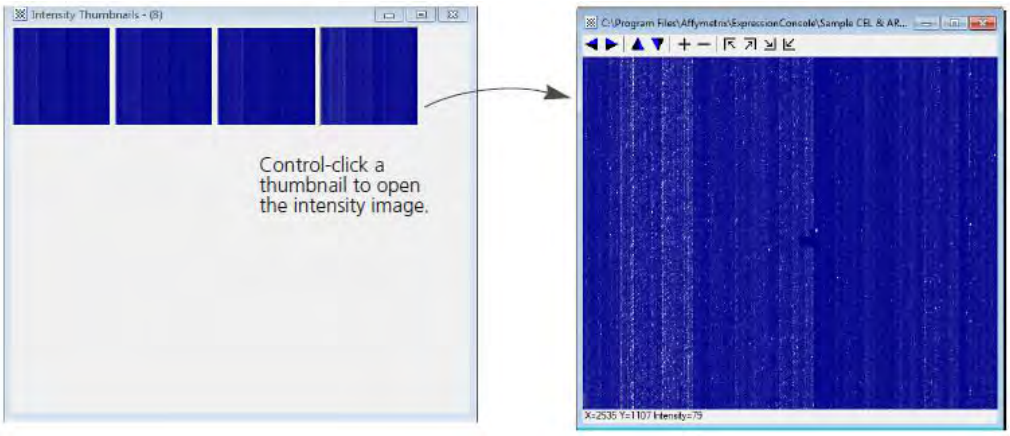


Figure 122 Probe cell intensity thumbnails.

Right-clicking a thumbnail enables the following options:

- Adjust the range for the color scale (Scale)

  Note: If an image is enlarged (Ctrl-click a thumbnail), auto-scaling is applied. If you change scale after enlarging an image, that scale is only applied to that particular image, not to all images
- Save the image in PNG format
- Copy the image to the clipboard (Copy to clipboard)

Histograms

Signal histogram

The signal histogram (**Graph ▸ Histogram - Signal**) is enabled when the CHP files are selected in the study window (Figure 123). This graph displays a histogram plot of the signal values for the selected CHP files in a box or line graph. A legend of the input CHP files is displayed in the upper right corner of the histogram.

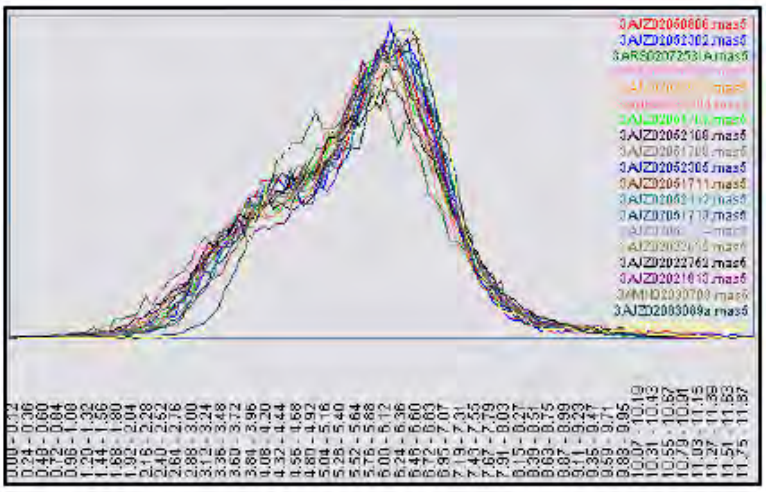


Figure 123 Signal histogram - 3' IVT with MAS5 algorithm.

A right click popup menu provides the following options for the user:

- Save the graphic as a PNG file (Save as PNG)
- Save the underlying data used from the histogram to a TXT file. (Save data as TXT)
- Copy the image to the clipboard (Copy to clipboard)
- Use bars in the histogram to display the values (Bars only)
- Use lines in the histogram to display the values (Lines only)
- Use bars and lines to display the values (Bars + Lines)
- Edit the scale of the Y axis (Edit Scale)

A slider just below the histogram allows you to adjust the zoom factor of the X axis and a scroll bar allows you to adjust the visible region of the graph.

Probe cell intensity histogram

A histogram of the Cel Intensity Data is similar to the signal histogram function (Figure 124).

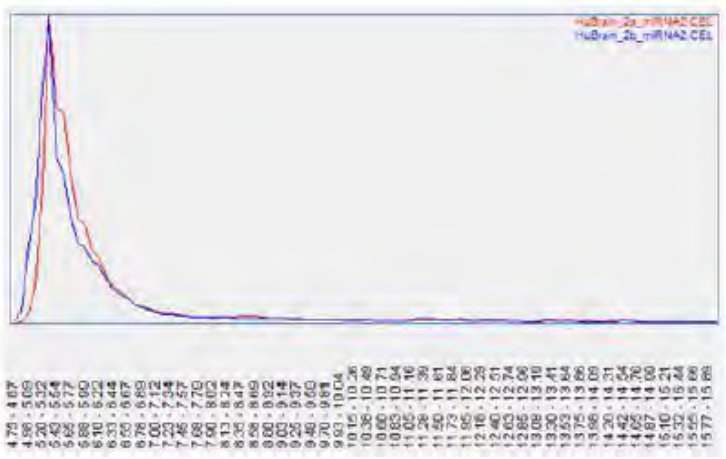


Figure 124 Probe cell intensity histogram.

Box plot

The box plot graph shows a standard box plot of data from either CHP or CEL files. A general rule to use when examining box plots is to look for individual arrays that are dramatically different from the others and most importantly from other replicates in the same group.

A right-click popup menu is provided for all of the box plots containing:

- Save the graphic as a PNG file (Save as PNG)
- Save the data to a TXT file (Save data as TXT)
- Copy the image to the clipboard (Copy to clipboard)

A slider is provided on the X axis to adjust the zoom factor. This slider adjusts the number of files to display on the X axis.

Box plot – probe cell intensity

The probe cell intensity box plots are generated from CEL file probe cell intensity values (Figure 125). This graph is generated by selecting Graph → Box Plot - Probe Cell Intensity. The probe cell intensity creates a box plot of the probe intensity values for each array. Probe cell intensities are prior to analysis/summarization and have not been normalized; therefore, some differences in the distributions are to be expected. At the feature intensity level, the Exon Arrays have about 6 million probes; some delays in the generation of these graphs are expected.

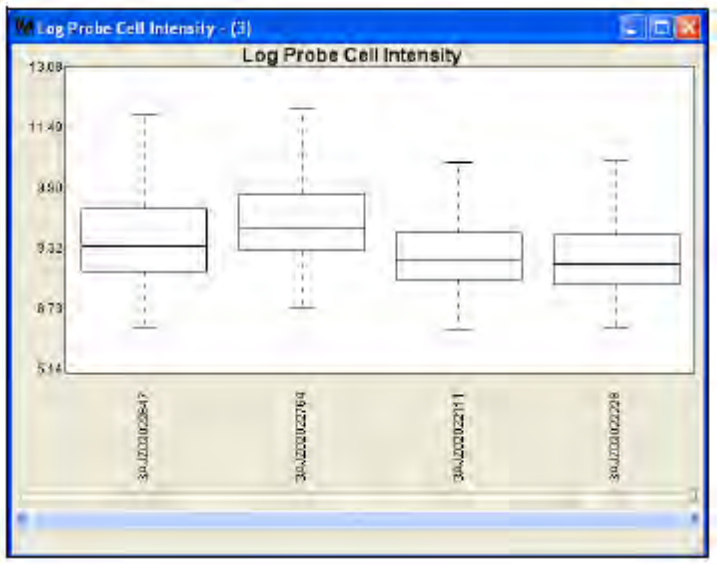


Figure 125 Box plot - probe cell intensity - 3' IVT expression array.

Relative probe cell intensity

The distribution of the ratio of the intensity of each probe to the median probe intensity across all of the selected arrays is summarized in the Box Plot - Relative Probe Cell Intensity (Figure 126). Therefore, the plot compares the distribution of intensities on each array to the median probe intensity value for the group. As such, it is a good way to identify arrays with divergent probe intensity distributions relative to the other arrays in the study (Figure 127).

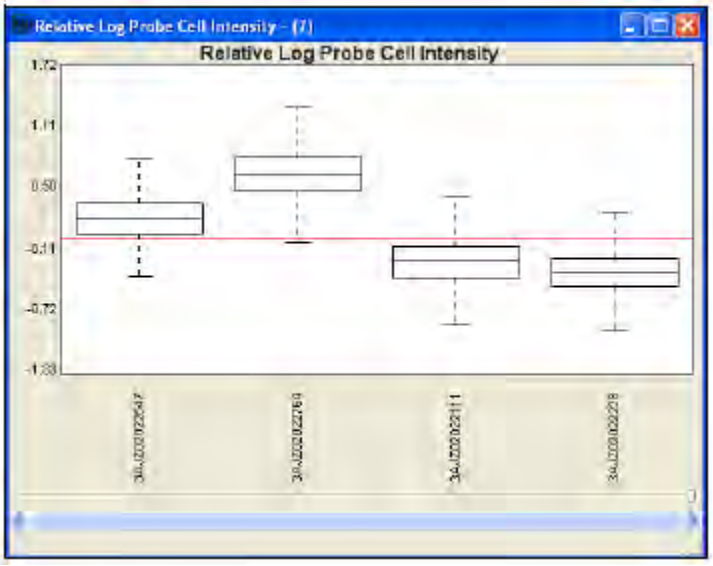


Figure 126 Box plot - relative probe cell intensity - 3' IVT expression array.

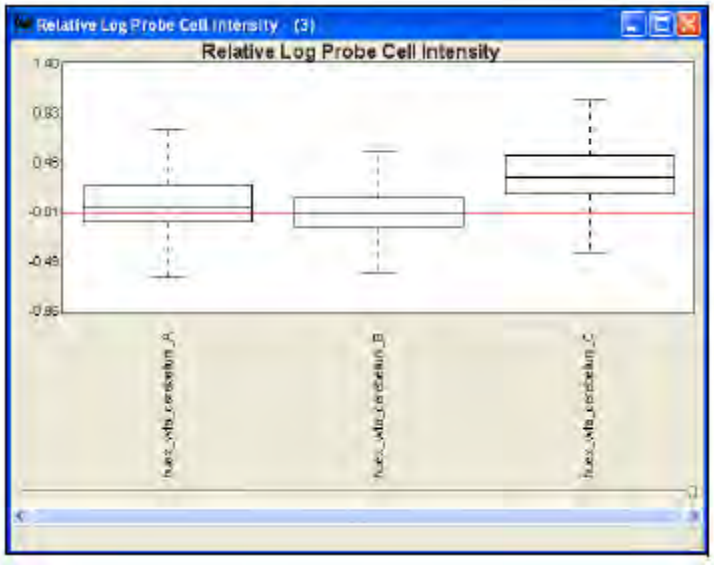


Figure 127 Box plot - relative probe cell intensity - exon array.

Box plot – signal

The Signal Box Plot is generated from summarized probe set signal values in CHP files (Figure 128, Figure 129). This graph is generated by selecting the **Graph ▸ Box Plot - Signal** menu item.

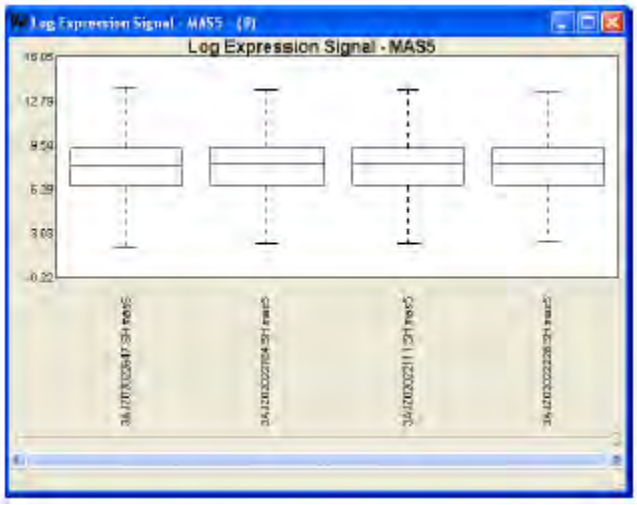


Figure 128 Box plot - signal - 3' IVT expression array.

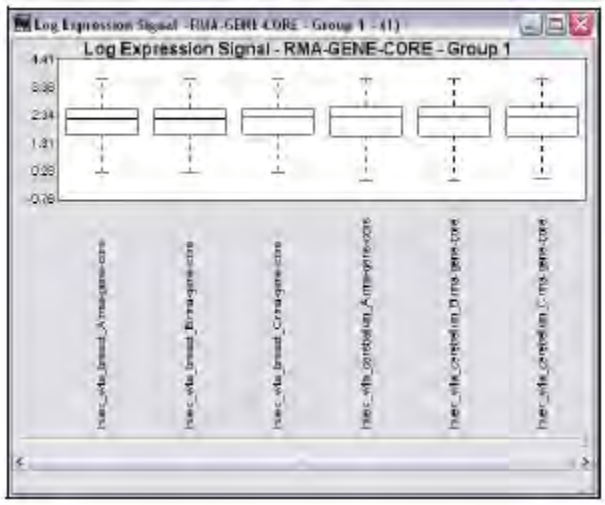


Figure 129 Box plot - signal - exon array.

Box plot – relative signal

The Relative Signal Box Plot summarizes the distribution of the ratio of signal for each probe set to the median probe set signal across all the selected arrays (Figure 130). Therefore, the plot compares the distribution of probe set signal values on each array to the median array for the group. As a result, it is a good way to identify arrays with divergent signal distributions relative to the other arrays in the study.

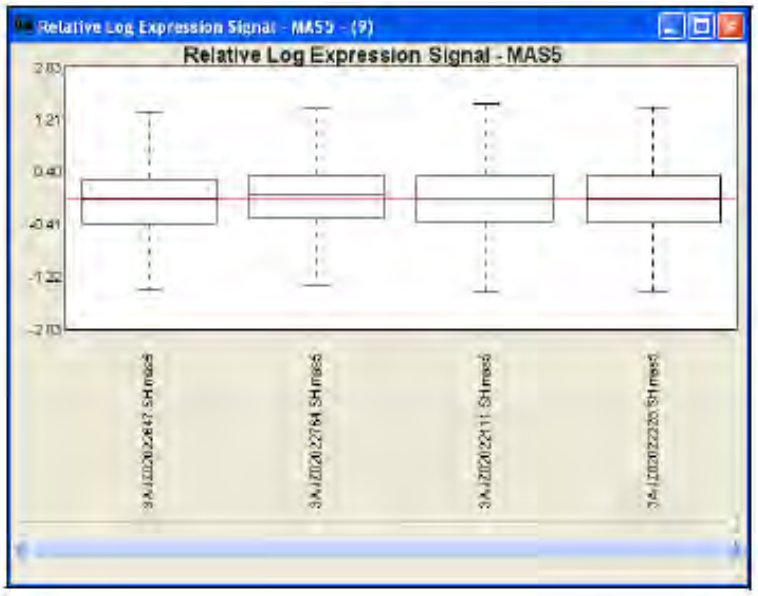


Figure 130 Box plot - relative signal - 3' IVT expression array.

Scatter plot

Generate scatter plots

1. Check two or more CHP files in the Study window.

> **WARNING!** The software generates a scatter plot for each pair of CHP files that are checked in the Study window. If many CHP files are checked, it may take a long time to generate the scatter plots.

2. Do either of the following:
   - To view thumbnails of the scatter plots, click the Scatter Graph toolbar button or select **Graph ▸ Scatter Plot Thumbnails - Signal** from the menu bar.
     One window displays all of the scatter plot thumbnails (Figure 131).

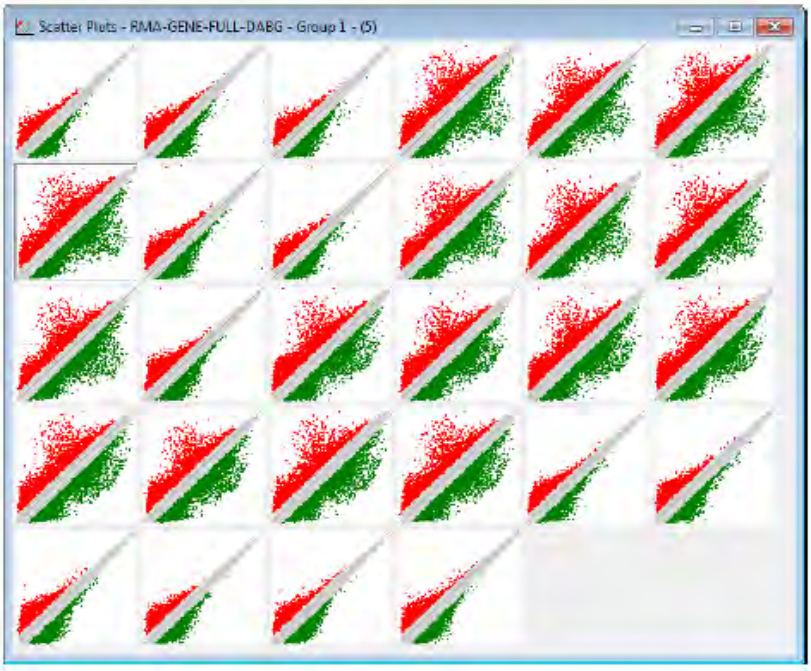


Figure 131 Scatter plot thumbnails.

   - To view each scatter plot in a separate window, select **Graph ▸ Scatter Plot - Signal** from the menu bar and click OK in the prompt that appears.
     Each scatter plot is displayed in a separate window (Figure 132).

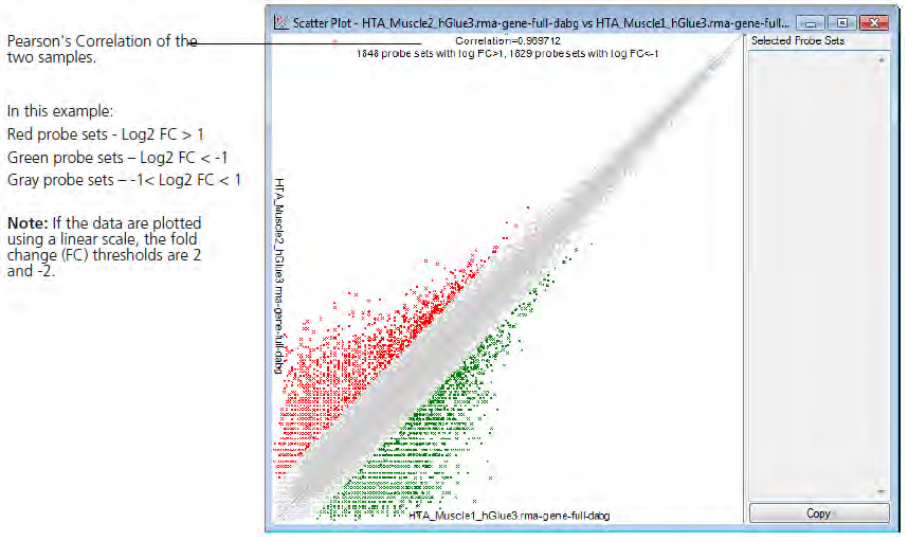


Figure 132 Scatter plot.

Display a scatter plot

1. Put the mouse arrow over a thumbnail to display a tooltip of the CHP file names.
2. Click a thumbnail to highlight the corresponding CHP files in the Study window (Figure 133).

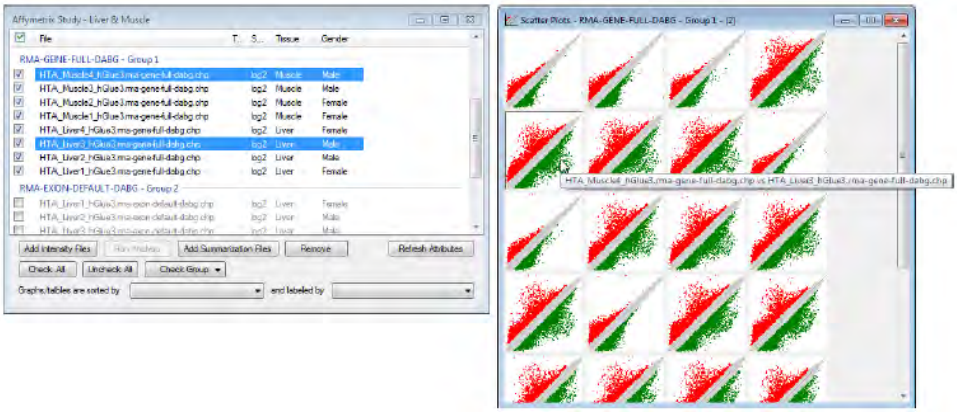


Figure 133 Tooltip and Study Windows.

3. Press the Ctrl key and click a thumbnail to open the graph in a new window.

Create a probe set list

Do one of the following to copy a probe set list to the system clipboard:

1. Right-click the graph and select: Copy up regulated probe sets to clipboard or Copy down regulated probe sets to clipboard.

   or

2. Lasso points in the Scatter Plot (drag the cursor to draw a closed shape around them). Click Copy after the list of selected probe set names appears (Figure 134).

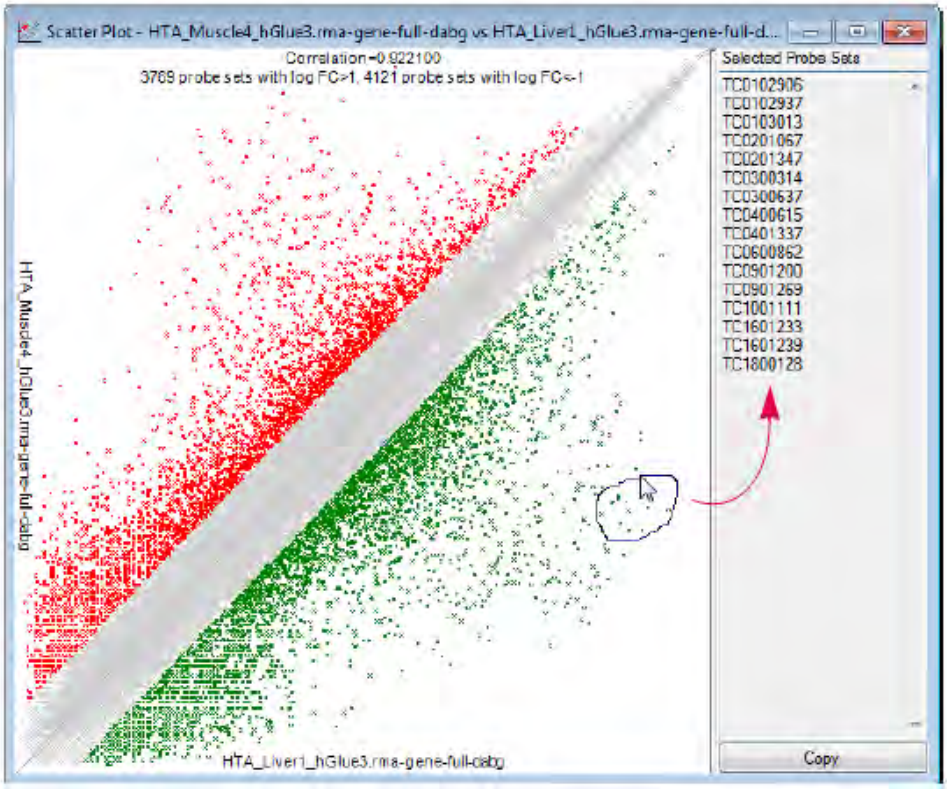


Figure 134 Scatter Plot.

MvA plot

The MvA plot (Figure 135) is a comparison plot comparing M (magnitude of change) on the Y-axis versus A on the X-axis, where M = log(Signal array1) - log(Signal array2) and A (average log(Signal)). The Y-axis is displayed on a log2 scale with green threshold lines for +/- two-fold changes. The color coding of the plot indicates the density of probes represented by that data point.

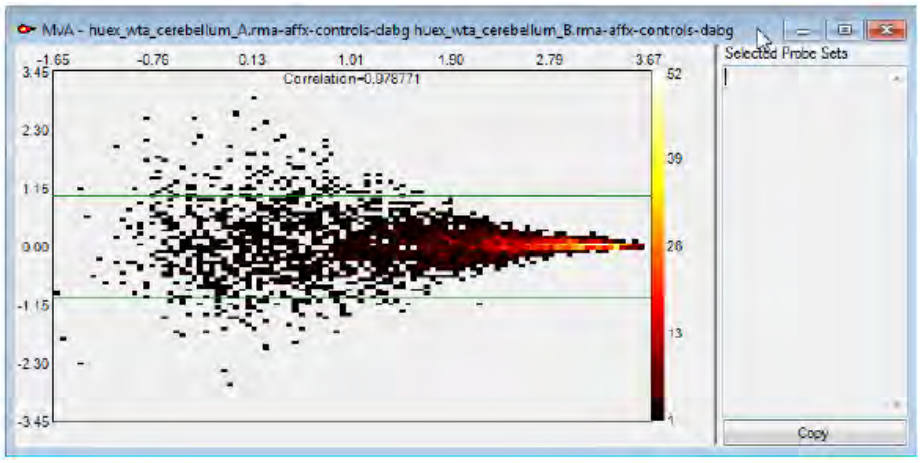


Figure 135 MvA plot - 3' expression array with MAS5 algorithm.

The following are displayed:

- The title of the graph, which indicates the two CHP files used to create the MvA plot
- The Pearson's correlation (r^2) value

You can select probe set points in the graph and use the list to create probe list sets.

Select probe sets

Drag the cursor around the group of probe sets to draw a closed shape around them (Figure 136).

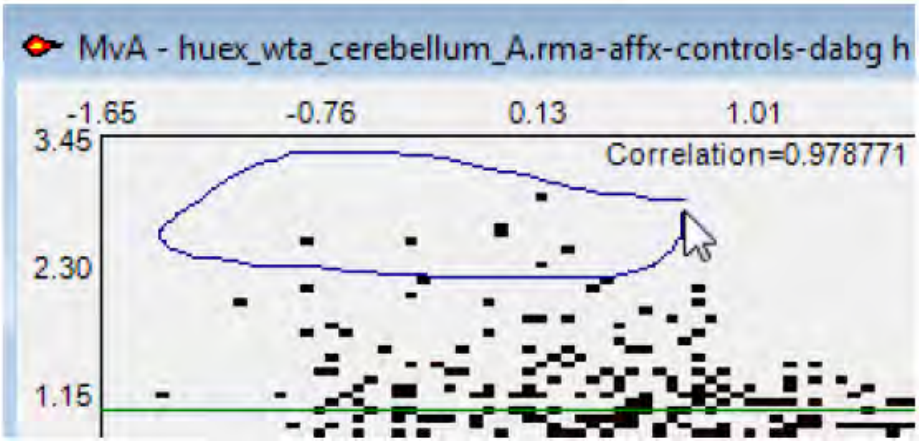


Figure 136 Selecting probe sets.

The lasso function automatically draws a straight line to the starting point if you release the mouse button.

The probe sets are selected when you release the button (Figure 137). A list of the selected probe sets appears in the Selected Probe Sets list.

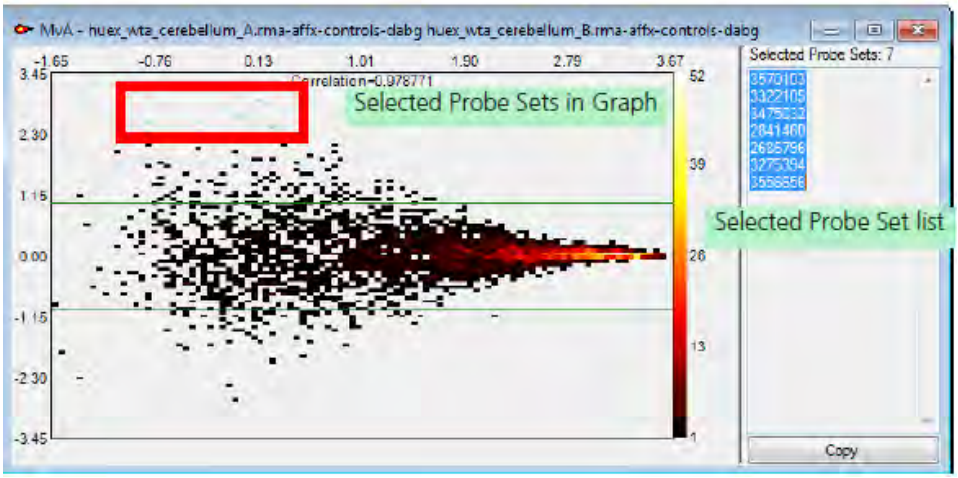


Figure 137 Selected probe sets.

Click the Copy button to copy the list for pasting into:

- A text file
- A Probe List (see Probe List)

Correlations

The relationship between two variables is described by their correlation. Two standard statistical measures of linear correlation (Pearson's and Spearman) are provided in the Expression Console™ Software. These are used to compare the signal estimates or detection p-values between two arrays. To simplify the interpretation of this data, the correlation values are presented as a heat map (Figure 138).

Note: The Spearman test is used with fewer samples since it is a rank-based test. Pearson's test is used when there are more data, which allows normalcy to be confirmed.

For all of the heat maps, a right-click popup menu enables the following:

- The graphic to be saved in PNG format (Save as PNG)
- The underlying r^2 values to be saved as a TXT file (Save data as TXT)
- The image to be copied to the clipboard (Copy to clipboard)
- A dialog box to adjust the min/max values to use to map values to a color scale (Edit Scale)

In addition, a slider is provided on either axis to adjust the zoom factor controlling the number of files to show on a given axis. A scroll bar is provided to adjust the visible region of the graph. Clicking on a cell in the graph displays the associated files in the lower left part of the window. The display includes the file names and the value.

Pearson's correlation (signal)

Signal concordance is evaluated using the Pearson's correlation coefficient (r^2) value to compare the signal values in two CHP files. The heat map contains a pairwise comparison of the signal values from all the selected CHP files, where the r^2 values have been converted into a pseudocolor scale (Figure 138).

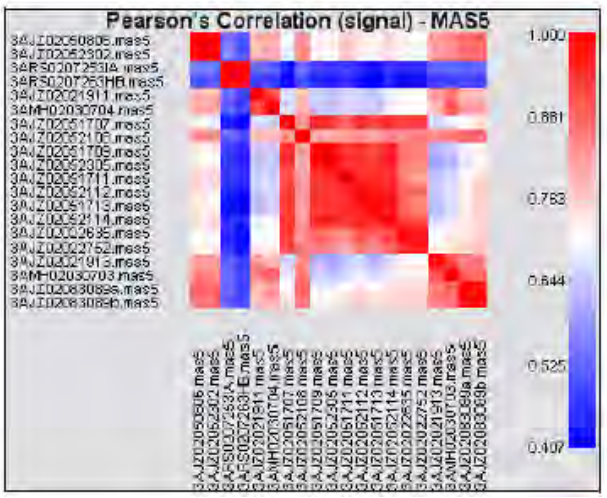


Figure 138 Pearson's correlation (signal) - 3' IVT expression array.

Pearson's correlation (detection p-value)

P-value concordance is evaluated using the Pearson's correlation coefficient (r^2) to compare the p-values from two CHP files. The resulting heat map contains a pairwise comparison of the detection p-values from all of the selected CHP files, where the r^2 values have been converted into a pseudocolor scale (Figure 139).

Note: Only MAS5 and Exon level analyses calculate detection p-values to generate the graph associated with them.

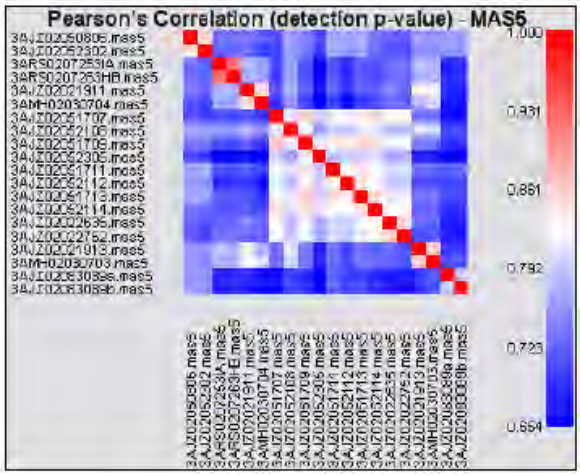


Figure 139 Pearson's correlation (detection p-value) - 3' IVT expression array.

Spearman rank correlation (signal)

This is the Spearman rank correlation of the signal values between two CHP files. The heat map contains a pairwise comparison of the signal values from all the selected CHP files, where the Spearman r^2 values have been converted into a pseudocolor scale (Figure 140).

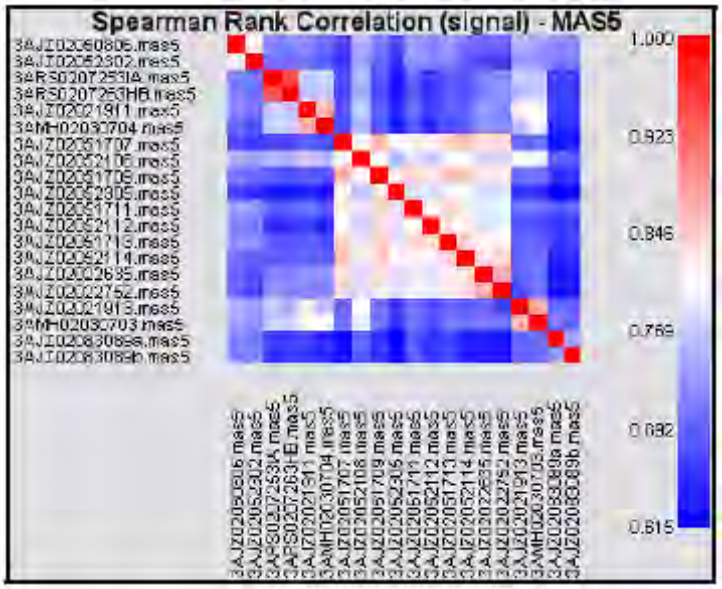


Figure 140 Spearman rank correlation (signal) - 3' IVT expression array.

Spearman rank correlation (detection p-value)

This is the Spearman rank correlation of the p-values between two CHP files. P-values are only available in MAS5 CHP files and Exon Arrays analyzed at the exon level. The heat map contains a pairwise comparison of the detection p-values from all the selected CHP files, where the Spearman r^2 values have been converted into a pseudocolor scale (Figure 141).

Note: Only MAS5 and Exon level analyses calculate detection p-values to generate the graph associated with them.

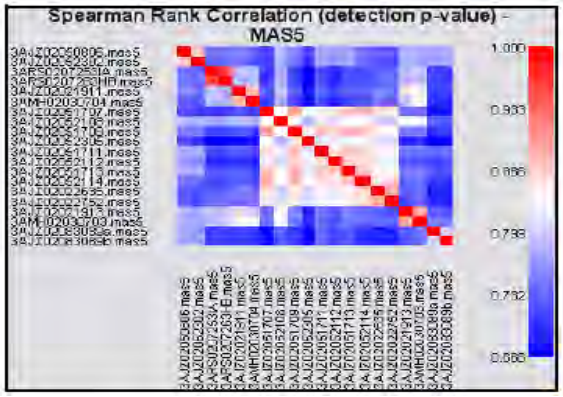


Figure 141 Spearman rank correlation (detection p-value) - 3' expression array.

Line graphs

Line graphs allow the user to graph metrics or signal values for specified groups of probe sets for selected CHP files.

Report metrics

The report metrics is a line graph of the selected metrics (**Graph ▸ Line Graph - Report Metrics**) calculated and stored in the header of the CHP (Figure 142). These include the algorithm parameters, global array metrics, and spike-in and housekeeping control values that are stored in the CHP file. The available metrics for graphing depend on the metrics available for the CHP files in the current study. Metrics not present in the CHP file cannot be graphed.

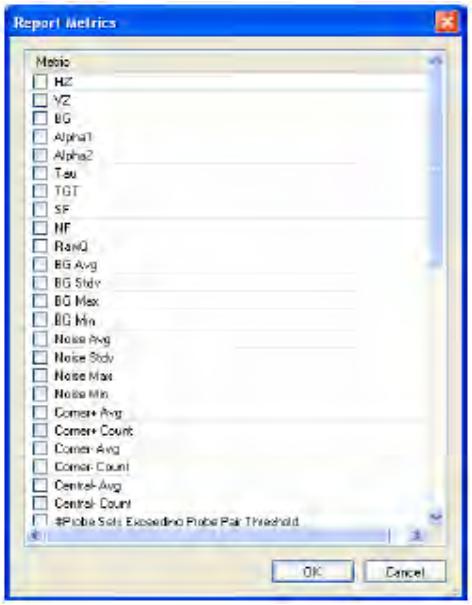


Figure 142 Report metrics - 3' expression array analyzed with MAS5.

Probe List

The Probe List is a line graph of probe sets defined in a probe list file. Select **Graph ▸ Line Graph - Probe List**, the Select the probe list window opens (Figure 143). When you select the probe list of interest, Expression Console™ Software creates a line graph of the probe sets (Figure 144).

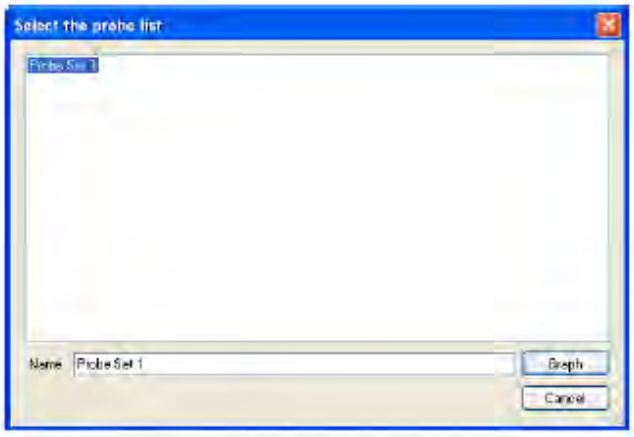


Figure 143 Select the probe list window.

Figure 144 Probe List line graph - exon array.

New Probe List

The New Probe List window (Figure 145, Figure 146) (**Graph ▸ New Probe List**) enables users to create their own probe list by selecting the probe sets they want to investigate.

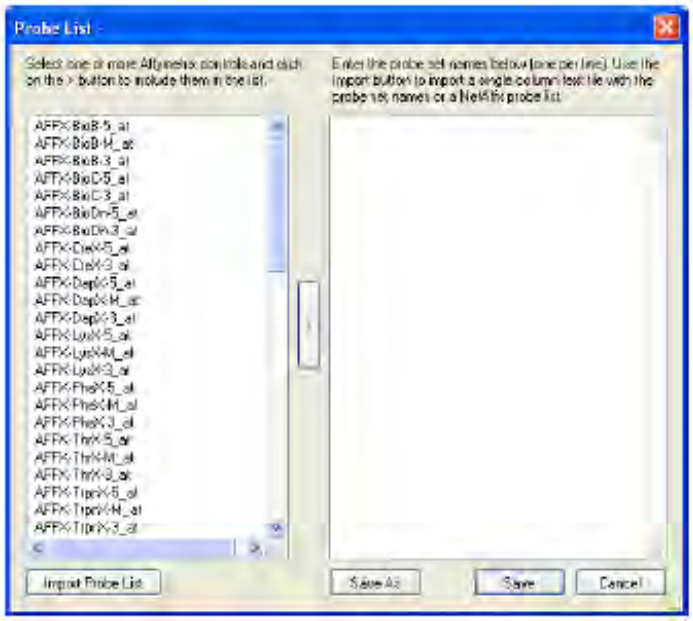


Figure 145 New Probe List - 3' expression array.

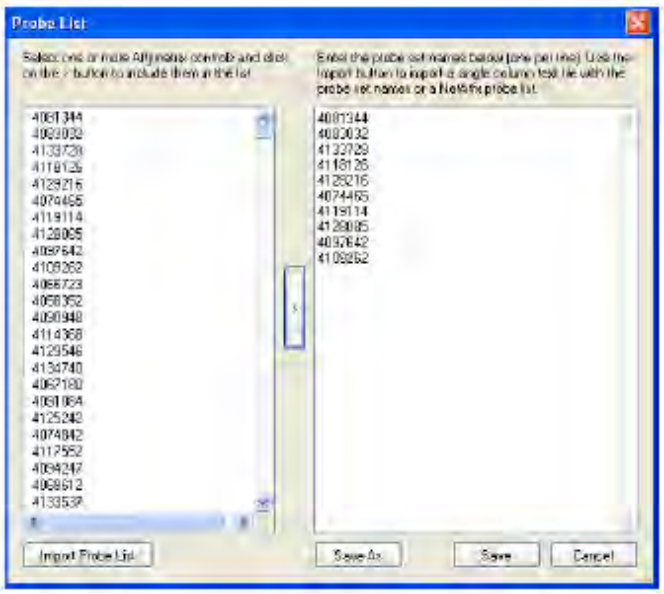


Figure 146 New Probe List - exon array.

Open a Probe List

1. Select **Graph ▸ Open Probe List** (Figure 147), and select the probe list.

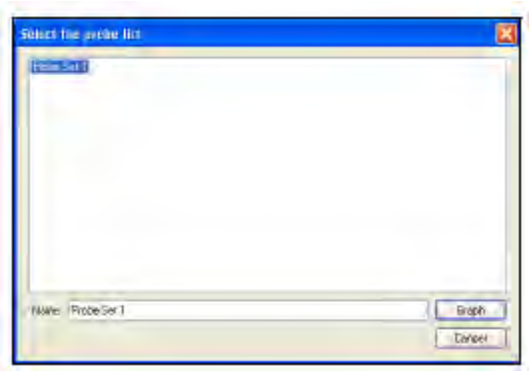


Figure 147 Select the probe list.

2. Click Open.
   The Probe List of interest opens for viewing (Figure 148).

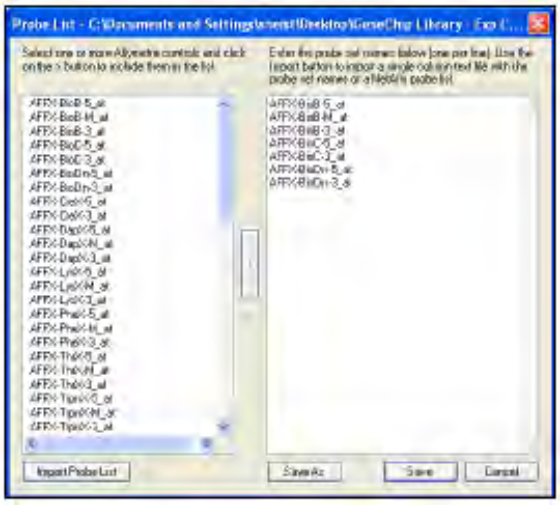


Figure 148 View the Probe List.

Delete a Probe List

1. Select **File ▸ Utilities**.
2. Select the Parameter Management tab (Figure 149).

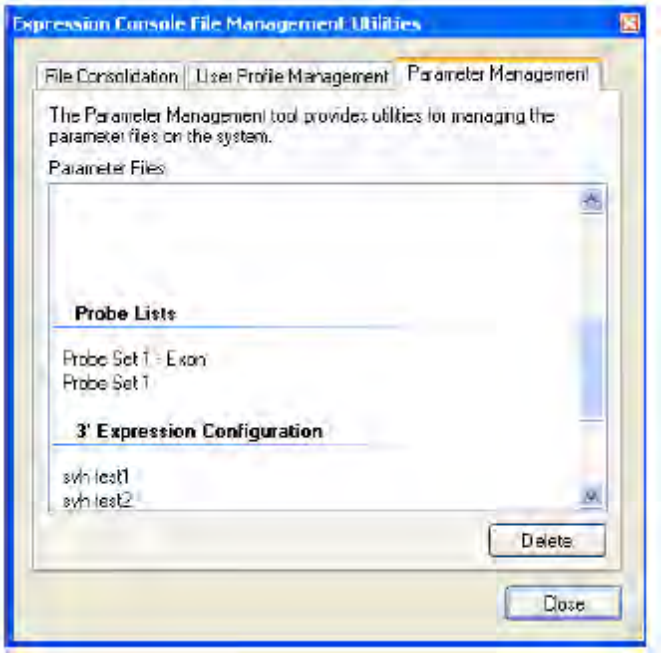


Figure 149 Parameter Management tab - delete a probe list.

3. In the Probe List section, select the probe list to be deleted.
4. Click Delete and then Close.

Principle component analysis

Expression Console™ Software enables you to perform Principle Component Analysis (PCA) on probe cell intensity (CEL) or signal (CHP) data. PCA identifies a new set of variables (PCA1, PCA2, PCA3) that account for a majority of the variance in the original data set.

The first principal component (PCA1) captures as much variability in the data as possible. PCA2 captures as much of the remaining variability (not accounted for by PCA1) as possible. PCA3 captures as much of the remaining variability (not accounted for by PCA2) as possible.

Probe cell intensity data are pre-processed for PCA analysis due to memory restrictions. A non-random sampling method selects 50,000 probe sets for analysis.

Perform PCA

1. Select the data for PCA (CEL or CHP files) in the Study Window.
2. Do either of the following:
   - Click the PCA graph toolbar button and select a graph type: PCA - Probe Cell Intensity or PCA - Signal.
   - Select **Graph ▸ Principle Component Analysis - Probe Cell Intensity or Principle Component Analysis - Signal** from the menu bar.

   The 3-dimensional PCA graph(s) appear (Figure 150).
   The graph axes represent the top three variables (PCA1, PCA2, and PCA3) that account for the majority of the variability among the samples.

Sample display options

Samples can be labeled in the PCA graph by color (up to 10 different colors) and shape (up to five shapes) with information in the ARR files.

Note: Use the Sample Attribute Editor to edit or add new attributes to an existing sample attribute file (ARR). For more information, see "Editing sample attributes" on page 43.

- Select attributes for display by color and by shape from the drop-down lists (Figure 150).

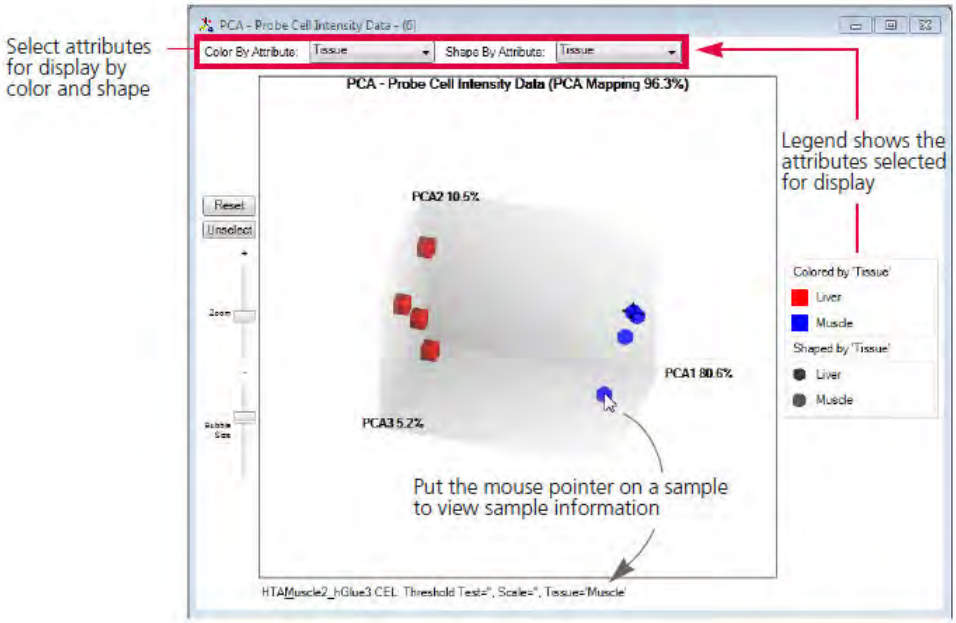


Figure 150 Example PCA - Probe Cell Intensity (CEL) Data.

- Put the mouse pointer on a sample to display sample information in the graph window (Figure 150).

- Click one or more outlier samples in the PCA window to highlight them in the Study window (Figure 151). Click Unselect in the PCA window to remove the selection.

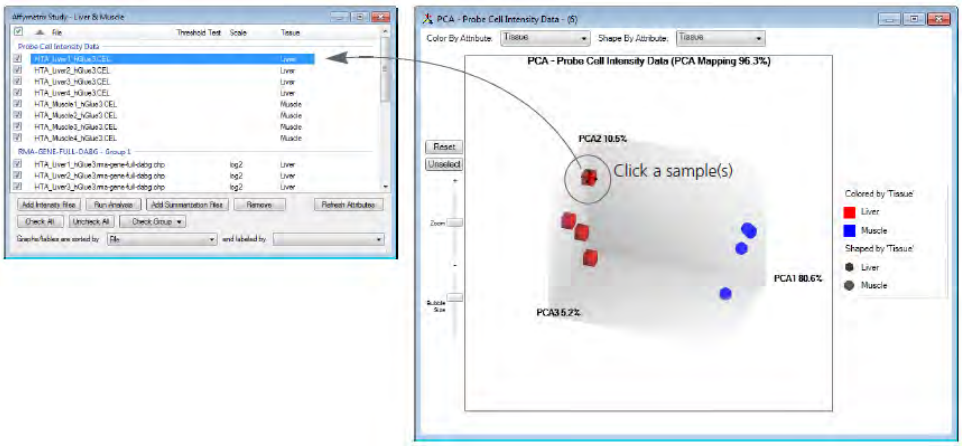
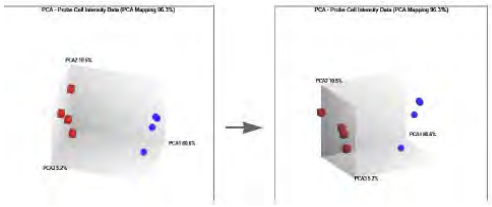
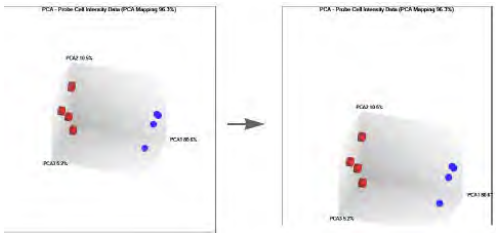
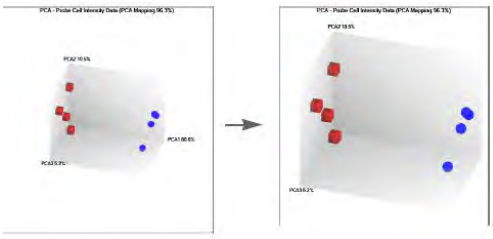
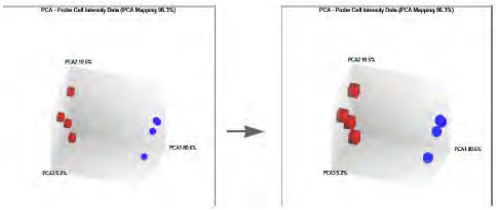


Figure 151 Samples selected in the PCA window are highlighted in the Study window.

PCA graph display options

Adjusting the zoom and sample bubble size may help you identify the outliers among samples more clearly in the graph.

To...	Do This
Rotate the graph and change the view perspective. 	Click and drag the graph.
Move the graph without changing the view perspective. 	Right-click and drag the graph.

To...	Do This
Increase or decrease the graph size. 	Move the Zoom slider up or down.
Increase or decrease the sample bubble size. 	Move the Bubble Size slider up or down.
Return to the default PCA graph display.	Click Reset.

Export/Copy options

Right-click the PCA window and select Save as PNG or Copy to clipboard on the shortcut menu.

10 Controls and thresholds

This section describes:

- How to identify and modify the controls for 3' Expression Arrays (Currently, modification of the controls for Gene and Exon Arrays is not supported with the Expression Console™ Software.)
- How to set user-defined thresholds for Gene, Exon, miRNA, and 3' Expression Arrays.

For guidance on how to set thresholds for individual metrics for 3' Expression Arrays, refer to the white paper Data Analysis Fundamentals, and for Gene and Exon Arrays, refer to the white paper Quality Assessment of Exon and Gene Arrays.

Report controls

To help researchers establish quality control processes for gene expression analyses, Several controls are available which allow researchers to monitor assay data quality. These include but are not limited to:

- hybridization controls
- labeling controls
- internal control genes
- algorithm parameters
- algorithm outputs

The Expression Console™ Software provides functionality to display and highlight individual metrics and the arrays containing metrics outside of user-defined tolerances.

In general, we highly encourage users to create a running log of these parameters to monitor quality and potentially flag outlier samples. The QC functionality built into the Expression Analysis Software can help with this. Evaluation of particular samples should be based on the examination of all sample and array performance metrics in light of the history of the metrics performance in an individual tissue and array type.

When analysis results are viewed within the software, the metrics for the report controls are compared to a set of user definable thresholds. Any results identified as being outside of the selected thresholds are tagged as Outside of Bounds.

Defining 3' Expression Array controls

For 3' Expression Arrays, the Report Controls window (Figure 152) enables the user to modify the group of probe sets that constitute the spike and housekeeping controls. The probe array type is selected from the drop-down menu at the top of the window to modify the control probes for that probe array type only. The list of control probe sets on the array are displayed on the left side of the window. By default, the probe sets are filtered to show only the AFFX controls using the Filter button at the bottom. The filter is case sensitive.

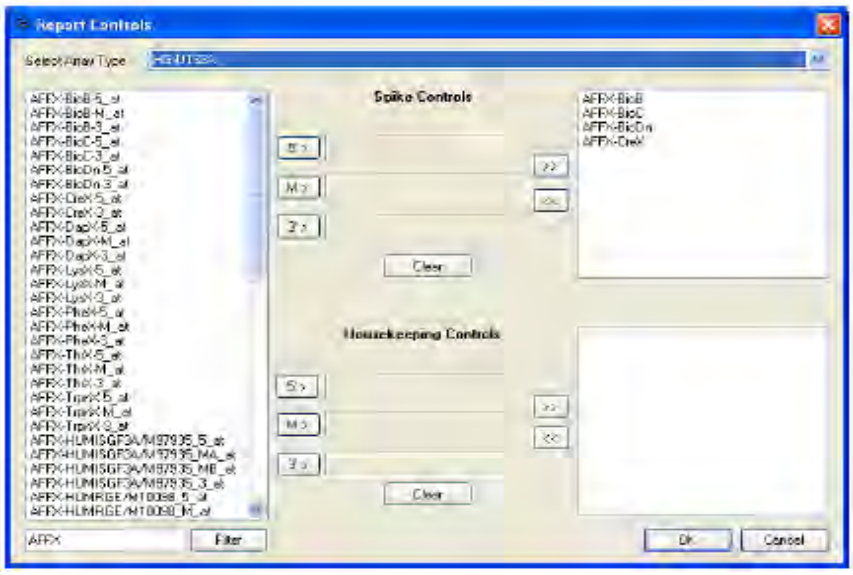


Figure 152 3' IVT Expression Array report controls window.

Note: In order for the report to contain control information, the controls must be defined before the analysis is run. If additional or different controls are required, the analysis must be rerun following redefining the controls.

Spike controls (Eukaryotic Hybridization Controls)

The 20x Eukaryotic Hybridization Controls are spiked into the hybridization cocktail, independent of RNA sample preparation, and are therefore used to evaluate sample hybridization efficiency on gene expression arrays. The default spike controls are listed as:

- AFFX-r2-Ec-BioB
- AFFX-r2-Ec-BioC
- AFFX-r2-Ec-BioD
- AFFX-r2-P1-Cre

Note: BioB is at the level of assay sensitivity (1:100,000 complexity ratio) and should be called Present at least 70% of the time. *BioC*, *BioD*, and *cre* should always be called Present with increasing signal values, reflecting their increasing relative concentrations.

Internal control genes (housekeeping genes)

Internal control genes, or housekeeping genes, are gene transcripts that are constitutively expressed in most samples. These transcripts serve as internal controls, are useful for monitoring the quality of the starting sample, and are subject to any variability in the labeling of the sample and hybridization to the array, for 3' Expression Arrays. For Human, Mouse, and Rat 3' Expression Array types, β-actin and GAPDH are used to assess RNA sample and assay quality. Specifically, the signal values of the 3' probe sets for actin and GAPDH are most informative and, therefore, as a general recommendation, these should be compared to the signal values of the corresponding 5' probe sets. The ratio of the 3' probe set to the 5' probe set should generally be less than 3 for the One Cycle Labeling protocol. For more details on interpreting the housekeeping genes, see the white paper, Data Analysis Fundamentals on our web site. The Housekeeping controls are:

- GAPDH
- β-Actin

Note: Control probe set names are unique to each array design.

Labeling controls

Poly-A RNA controls can be used to monitor the entire target labeling process. Each eukaryotic GeneChip™ probe array contains probe sets from several B. subtilis genes that are absent in eukaryotic samples (lys, phe, thr, and dap). These Poly-A RNA controls are in vitro synthesized, and the polyadenylated transcripts for the B. subtilis genes are premixed at staggered concentrations. The Poly-A controls can be spiked into a complex RNA sample, carried through the sample preparation process, and evaluated like the internal control genes. The GeneChip™ Poly-A RNA Control Kit (P/N 900433) contains the following four exogenous, premixed control spikes:

- Lys: AFFX-r2-Bs-lys (1:100,000)
- Phe: AFFX-r2-Bs-phe (1:50,000)
- Thr: AFFX-r2-Bs-thr (1:25,000)
- Dap: AFFX-r2-Bs-dap (1:6,667)

All of the Poly-A controls should be called Present with increasing signal values in the order of lys, phe, thr, dap.

Note: Using the Poly-A RNA control and ERCC controls in the same sample is not recommended. The ERCC controls also contain lys; therefore, if both types of controls are used, the lys signal value will be increased. It will not be the lowest labeling control signal value.

ERCC controls

The ERCC RNA Spike-In Control Mixes (Cat. Nos.4456740, 4456739) contain pre-formulated sets of 92 polyadenylated transcripts from the External RNA Controls Consortium (ERCC) plasmid reference library. There are two formulations, Spike-In Control Mix 1 and Spike-In Control Mix 2. The transcripts in Spike-In Mix 1 and Spike-In Mix 2 are present at defined Mix 1:Mix 2 molar concentration ratios of four subgroups. Each subgroup contains 23 transcripts of similar size distribution and GC content at concentrations across a 106-fold range.

Use the ERCC RNA Spike-In Mix (Cat. Nos. 4456740, contains Spike-In Mix 1 only) to assess the dynamic range of an experiment and the lower limit of detection.

Use ERCC ExFold RNA Spike-In Mixes (Cat. Nos. 4456739, contains Spike-In Mix 1 and Spike-In Mix 2) to assess the accuracy of differential gene expression measurements.

Download the ERCC RNA Spike-In Control Mixes User Guide (Pub. No. 4455352) for detailed instructions.

Identify/remove controls

Most 3' Expression Array types have a set of controls identified by default. Follow the steps below to change or set the report controls for the specified array type.

Select , or Edit → 3′ Expression Report Controls.

The Report Controls dialog box opens (Figure 153).

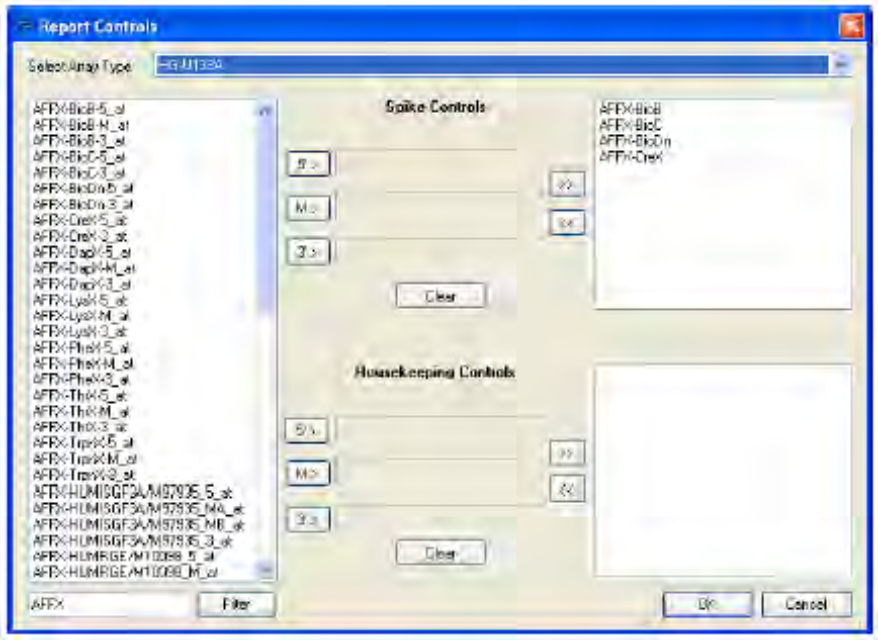


Figure 153 3′ expression report controls dialog box.

1. Choose the array type from the Select Array Type drop-down menu at the top of the dialog box.

2. In the probeset box, located on the left side of the window, highlight a probe set ID and select the 5'> to make that probe set the 5' probe, M> the middle probe, or 3'> the 3' probe.

   The upper box is used to define the spike-in control probe set IDs and the lower box the housekeeping control probe set IDs. For each probe set ID, a minimum of two probe sets must be identified (for example, a 3' and a 5').

3. After you identify the 3', M, or 5' probe sets, select the >> button to load it into the control list.

4. Add or remove probe set IDs by clicking on the appropriate button >> or <<.

   Control metrics for the probe set IDs contained in the two boxes of the right hand side of the window will be included in the report for any subsequent analyses performed.

   Note: To view all available probe sets, clear the text box next to the Filter button, then click the Filter button. A list of all the probe sets on the array are displayed.

   Or, to view specific probeset controls, type the identifying letters and/or numbers in the text field and click the Filter button.

Delete all of the controls for an array type

1. Select **File ▸ Utilities**.
2. Select the Parameter Management tab. (Figure 154)

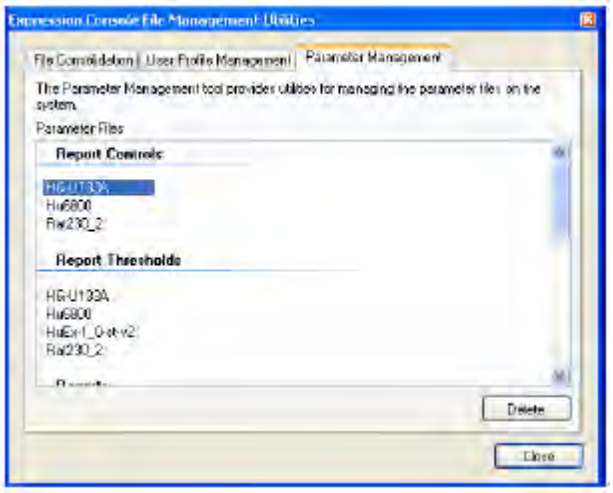


Figure 154 Parameter Management tab - delete report thresholds.

3. In the Report Controls section, select the array type.
4. Click Delete to remove the controls for that array type.

Report Thresholds

The Report Thresholds (Figure 155) define boundary conditions against a set of metrics that are computed during the analysis step. These metrics include report controls and statistics based on probe set Signal and Detection calls.

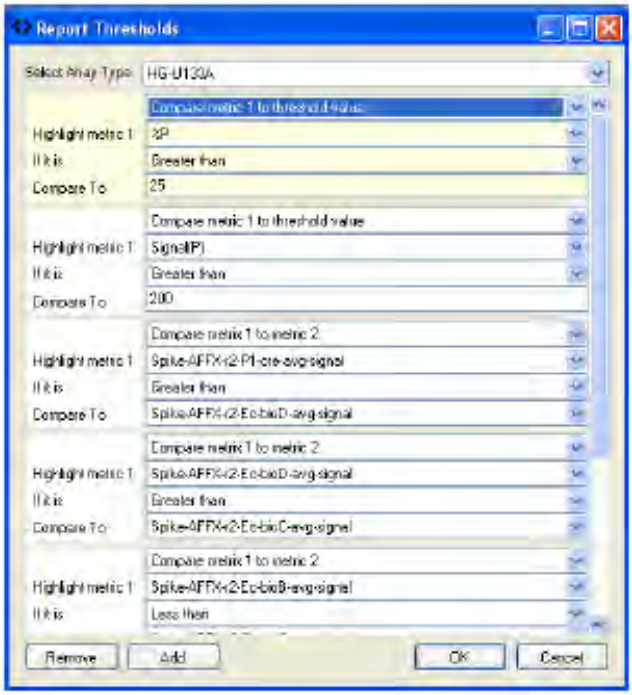


Figure 155 Report Thresholds window.

Fixed thresholds are applied to selected metrics in one of three ways:

- Compare metric 1 to threshold value – Example, compare the Percent Present calls to a fixed value to confirm that the metric passes a minimum or maximum value.
- Compare metric 1 to metric 2 – Example, compare two items, such as the average signal value for the Cre control against the average signal value for the bioD control to confirm that the detected signals are consistent with the known relative abundance in the sample.
- Compare metric 1 to average of metric 1 across arrays by standard deviation – Example, compare the spread of values such as scaling factors across all selected arrays of the same type. The calculation uses the mean and standard deviation to describe the spread of the data being compared. The final in-bounds range is then determined as the mean plus and minus a user defined factor times the standard deviation times the user selected multiplier.

Defining/modifying Report Thresholds

The Report Thresholds listed are default settings and can be used without further modification on the part of the user. Report Thresholds can be changed at anytime and the tables will be dynamically updated to reflect the changes once they have been saved.

Note: If a desired metric is not present in the report for a 3' Expression Array type, the 3' Expression Controls must first be modified and then the analysis must be run.

Set Report Thresholds

1. Select **Edit ▸ Report Thresholds**,  , or **Toolbox ▸ Configurations ▸ Specify report thresholds** to open the window.
2. Select the comparison type:
   - Compare metric 1 to threshold value
   - Compare metric 1 to metric 2
   - Compare metric 1 to average of metric 1 across arrays by standard deviation
3. Select the comparison operator – less than, greater than, or equal to.
4. Select the item/value/range multiplier to compare to. The availability of this option is dependent on the comparison type you selected.
5. To delete a threshold item, click in the box so that it is highlighted in yellow, then click the Remove button.

Note: To set thresholds for algorithm metrics, a CHP file analyzed with that algorithm must be loaded into the study window.
If a CHP is not selected, the software is unable to determine the metrics that are available for thresholding; therefore, one cannot add thresholds for new metrics without first selecting a CHP file.

Delete Report Thresholds

1. Select **File ▸ Utilities**.
2. Select the Parameter Management tab (Figure 156).

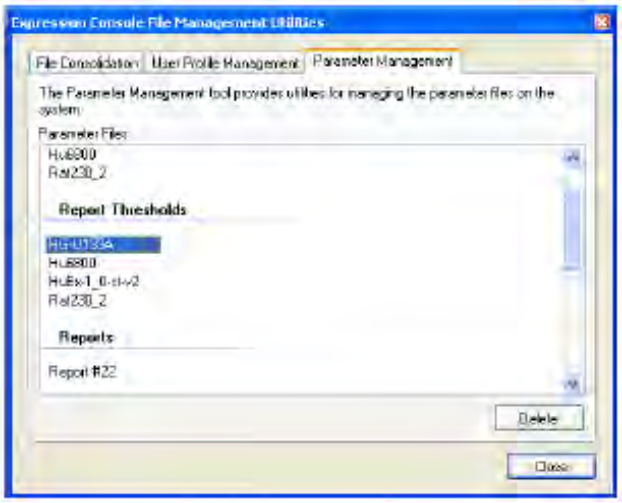


Figure 156 Parameter Management tab - delete Report Thresholds.

3. In the Report Thresholds section, select the array type.
4. Click Delete to remove the Report Thresholds for that array type.

11 Exporting data

The Expression Console™ Software provides options that allow you to export the data in tables and graphs to various text or pdf documents.

Printing displayed data

The currently active Table or Graph can be printed.

- Select **File** ▸ **Print** or ctrl-P to bring up the printer dialog.
- Select **File** ▸ **Print Preview** to display a preview of what will be printed.

Exporting data

Expression Console™ Software provides several options for exporting data:

- "Tables and graphs to PDF" on page 120
- "A report to GCOS RPT file" on page 120
- "A study to TXT file" on page 121
- "A table as TXT file" on page 121

Tables and graphs to PDF

All of the currently opened/displayed tables and graphs can be exported to a PDF format (Figure 157). Select **Export ▸ Export All Tables/Graphs To PDF** and browse to the chosen folder.

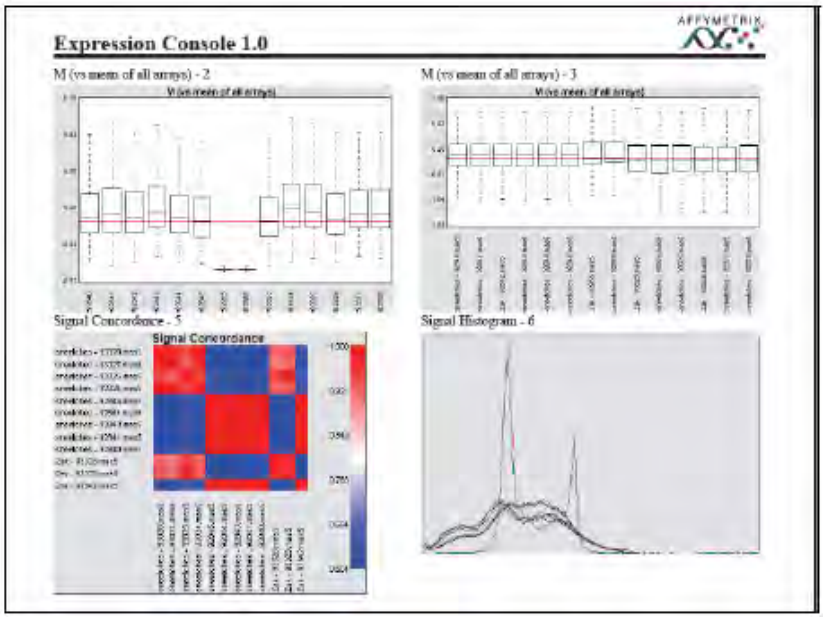


Figure 157 Tables and graphs in PDF format

A report to GCOS RPT file

The Expression Console™ Software is capable of creating GCOS formatted RPT files for MAS 5.0 CHP files. However, they cannot be generated by the other applications. The summarization workflows offered (i.e., RMA and PLIER-based workflows) are created with the Expression Console™ Software, but not other applications.

Use the **Export ▸ Export Report to GCOS RPT File** menu item to save the CHP file information for all checked files in the current study to RPT files (Figure 158). Each CHP file creates a RPT file in the same directory using the same base file name.

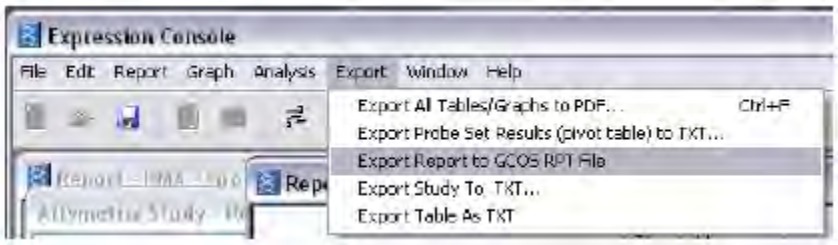


Figure 158 Export report to GCOS RPT file.

A study to TXT file

To export the Expression Console™ Software study table to TXT, right-click on the study window and select Figure 159.

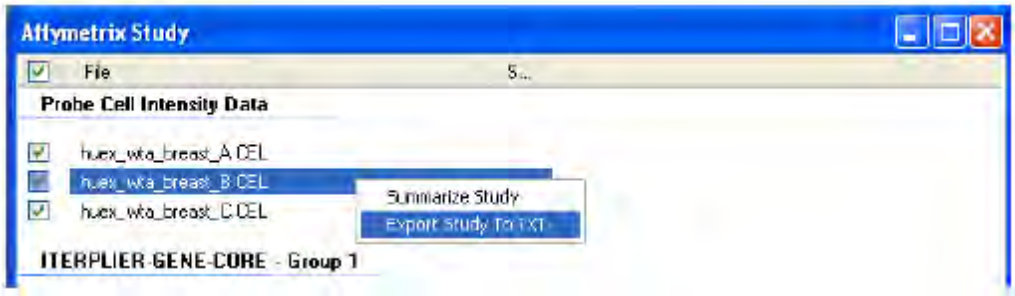


Figure 159 Export study to TXT.

A table as TXT file

After running an analysis, select **Report ▸ View Full Report** (or alternatively, a user customized report selection). When the report displays (Figure 160), right-click on the report and select Export Table As TXT.

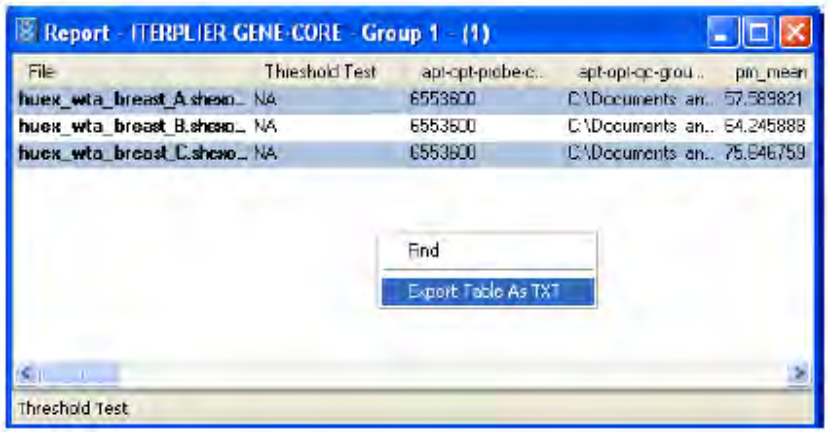


Figure 160 Export table as text.

Exporting analysis results data

Probe level summarization data can be exported from Expression Console™ Software for further downstream analysis.

Export analysis results can be exported in two formats:

- "Export analysis results without annotation information" on page 122
- "Exporting analysis results with annotation information" on page 124

Export analysis results without annotation information

1. Select **Edit ▸ Probe Level Summarizations Report Options**.
   The Probe Level Summarization Report Options dialog box opens (Figure 161).

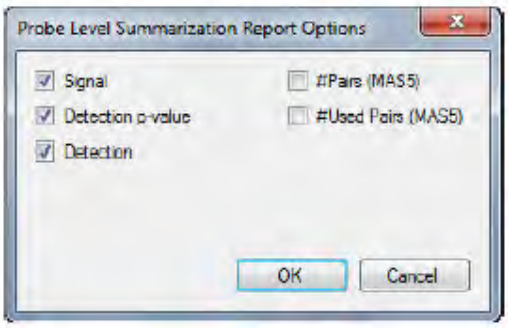


Figure 161 Probe Level Summarization Report Options.

2. Select the metrics to be exported:
   - Signal
   - Detection p-value
   - Detection
   - #Pairs (MAS5)
   - #Used Pairs (MAS5)

   Note: P-values are only available when using MAS5 summarization for 3' Expression Arrays and Exon Arrays analyzed at the exon level. No p-values are available for the gene level analysis irrespective of the summarization method chosen.

3. From the Export menu, select Export probe set results (pivot table) to TXT (Figure 162).

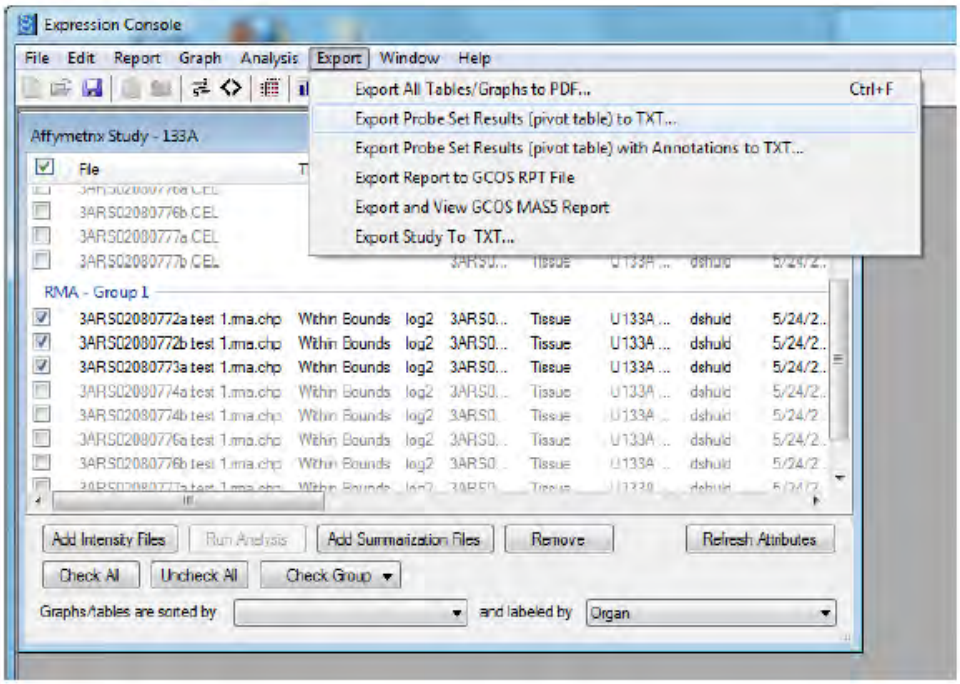


Figure 162 Export probe set results (pivot table) to TXT.

The Export probe set results (pivot table) dialog box opens (Figure 163).

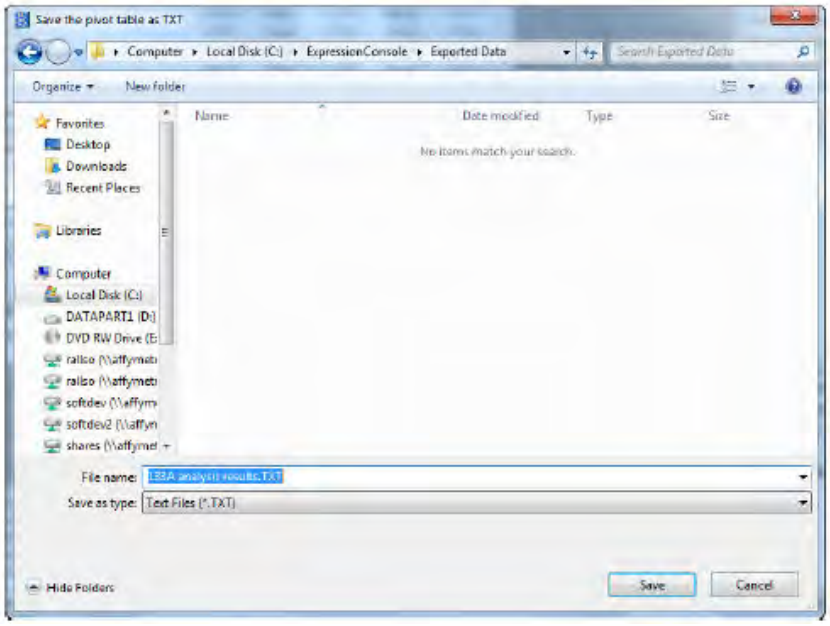


Figure 163 Save the pivot table as TXT dialog box.

4. Specify a name and location for the TXT file and click Save.
   The data is exported.

Exporting analysis results with annotation information

To aid the analysis of the expression data, annotation information contained in comma-separated files can combined with the signal values into a single tab-delimited file. Merging annotation information into a probe set results table is a one step process after a short setup procedure.

The setup procedure has two steps:

- Step 1: download the annotation file of interest.
- Step 2: create an annotation merge file by selecting the annotation information of interest.

Perform the initial set up

1. Download the necessary annotation files for your array type if they are not already in the library file directory. For more information see "Using annotation files" on page 21.

2. From the Edit menu (Figure 164), select Create Annotation Merge File, and then select one of the types of annotation merge file to create.

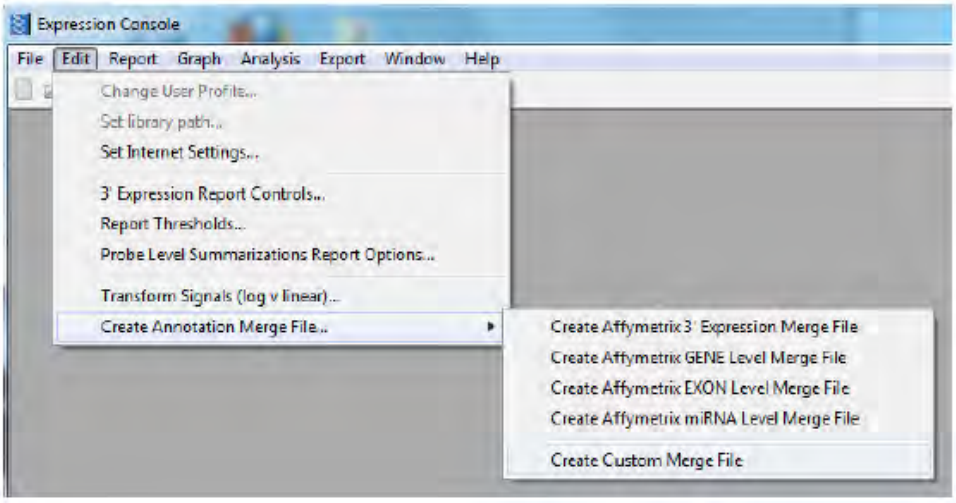


Figure 164 Create annotation merge file option.

- Create 3' Expression Merge File
- Create Gene Level Merge File
  Separate Gene Level Merge files are required for:
  - Exon Arrays
  - Gene Arrays
- Create Exon Level Merge File
  Separate Exon Level Merge files are required for:
  - Exon Arrays
  - Gene Arrays
- Create miRNA Level Merge File
- Create Custom Merge File

Note: The first column of the custom CSV file must contain the probe set identifiers to be used to match-up the records during the merge, See "Create custom annotation files" on page 24 for more information.

The Custom Annotation Configuration for Expression Reports dialog box opens (Figure 165).

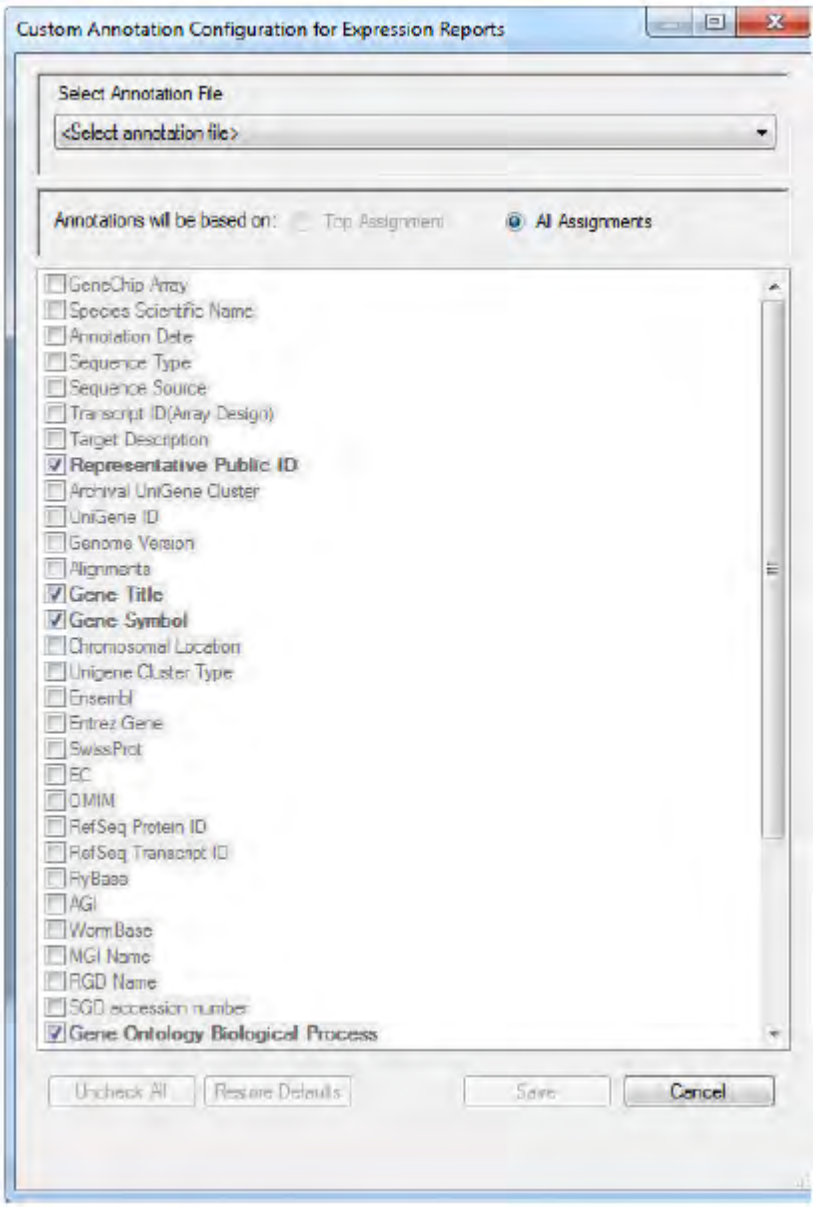


Figure 165 Custom Annotation Configuration for Expression Reports dialog box.

3. Select the annotation file from the drop-down list at the top of the dialog box (Figure 166).

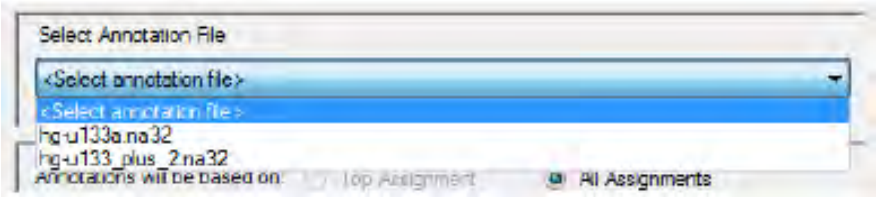


Figure 166 Select annotation file.

4. Select the level of the annotations to be used in the merge file (Figure 167). This option is only supported for Supplied Annotation files.

Figure 167 Annotations based options.

- Top Assignment Only: Annotations will be based on the single transcript to which the probe set best aligns. Selecting this option results in a single value for the annotation fields at the expense of information.
- All Assignments: The annotation will be based on all of the transcripts contained in the annotation file.

5. Select the annotation fields of interest:

   Bold indicates the default selections (This is only supported for Supplied Annotation files).

   The selected fields will each appear in their own column in the probe set summarization results table.

6. Select Save to save the annotation merge file.

   The Save dialog box opens (Figure 168).

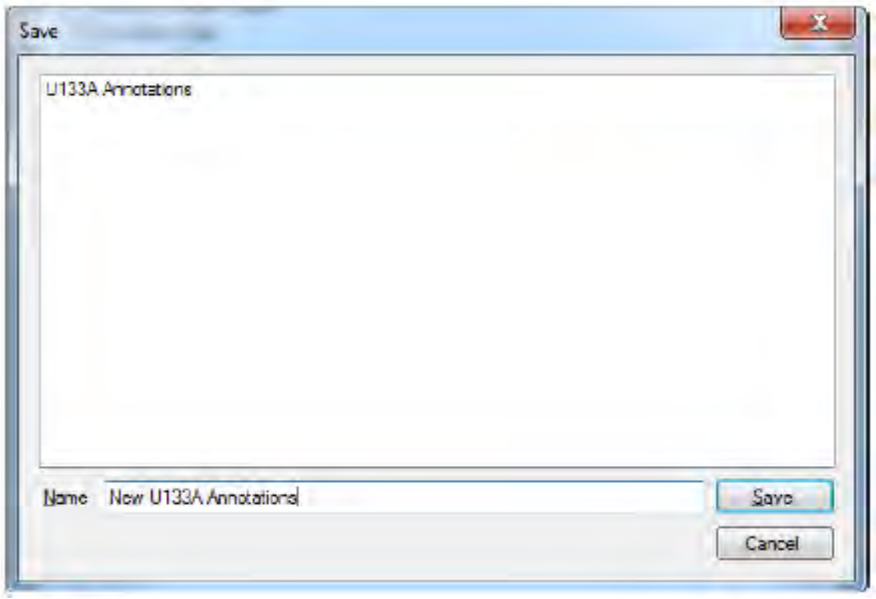


Figure 168 Save dialog box.

7. Enter a name for the annotation merge file when prompted. The name of the CSV file used to create the merge file will automatically be appended to the name you enter.

   Depending on the size of the CSV file and the number of probe sets in the file several minutes may be required for processing.

Merging annotation

Note: This functionality is only enabled if a study containing CHP files is open.

1. Select the CHP files for export in the Study dialog box.
2. From the Export menu (Figure 169), select Export Probe Set Results (pivot table) with Annotations to TXT.

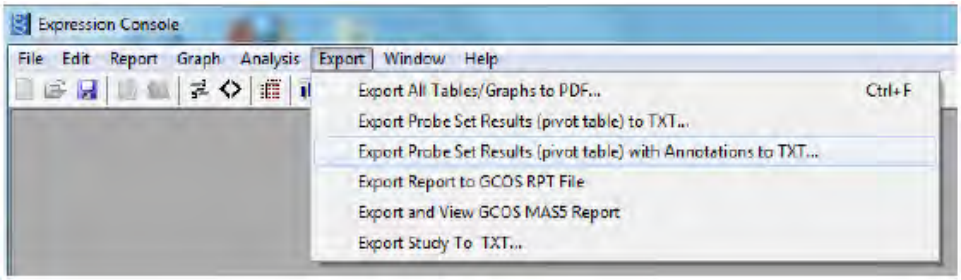


Figure 169 Export menu.

If you have selected files for an array or workflow that have no merge file available, the following notice appears (Figure 170).

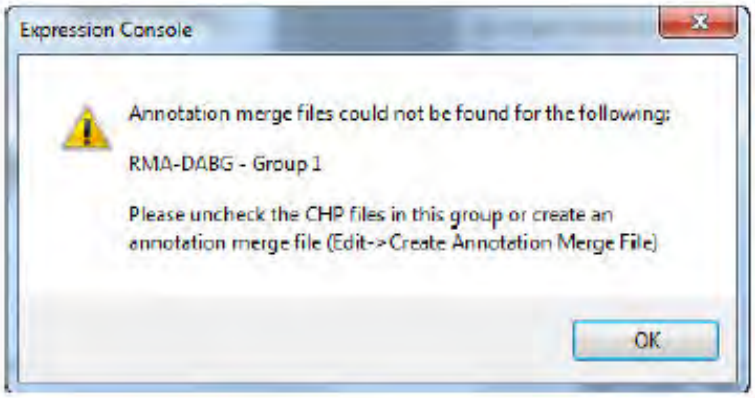


Figure 170 No merge file notice.

- Follow the instructions in the notice.

The Annotation Selection dialog box opens (Figure 171).

Figure 171 Annotation Selection dialog box.

The dialog box enables you to select an annotation merge file for each of group of CHP files.

3. For each group of CHP files:
   a. Click the Browse button for the group for which you wish to select a merge file.

      The Select an annotation merge file dialog box appears (Figure 172)

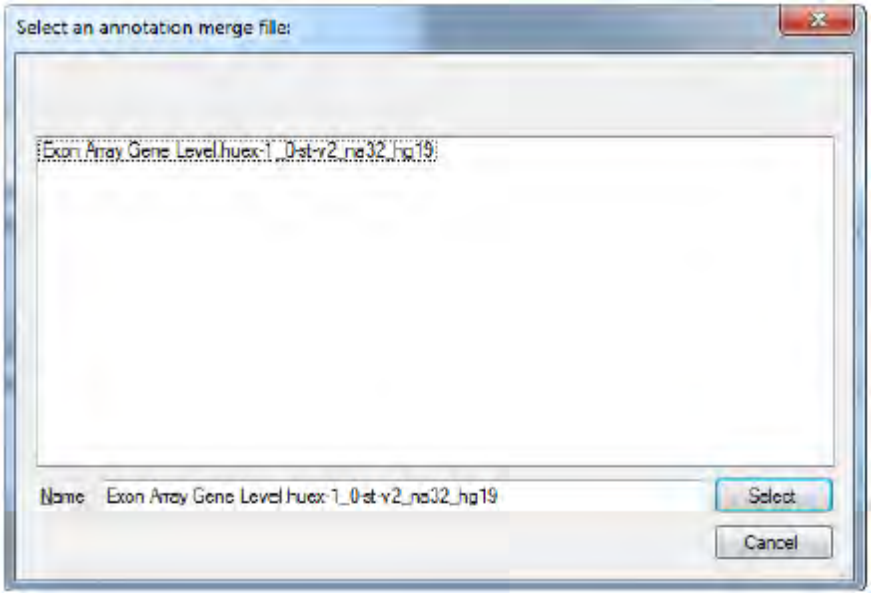


Figure 172 Select an annotation merge file dialog box.

   b. Select the merge file you wish to use and click the Select button.

      The selected merge file appears in the Annotation Selection dialog box (Figure 173).

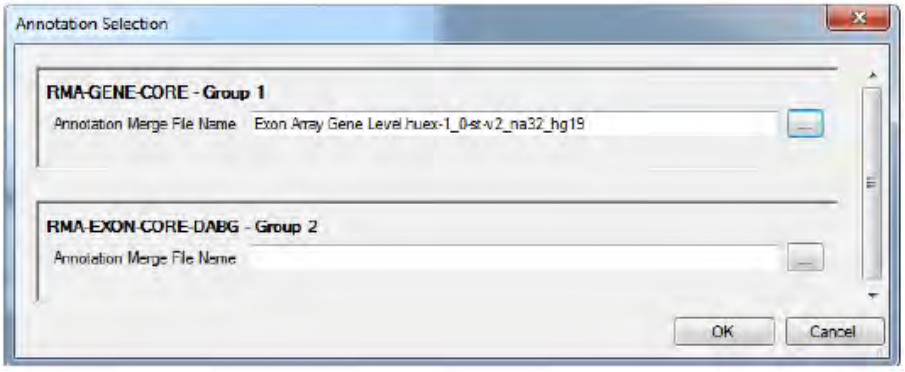


Figure 173 Merge file selected.

4. When you have finished selecting merge files for the groups, click OK in the Annotation Selection dialog box

   The Save the pivot table as TXT dialog box opens (Figure 174).

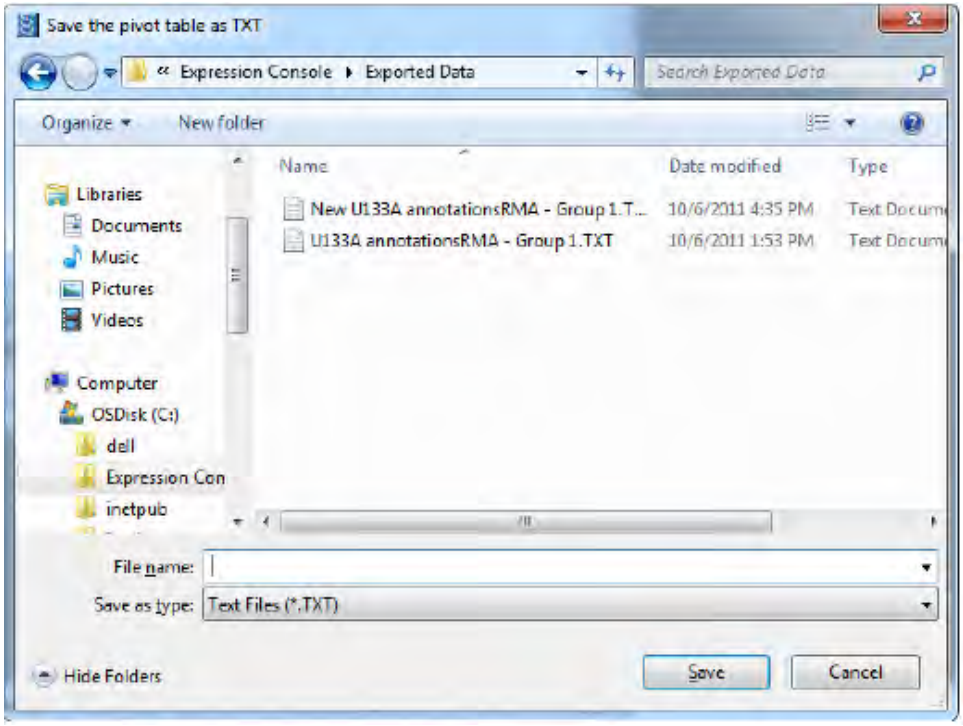


Figure 174 Save the pivot table as TXT dialog box.

5. Specify a name and location for the TXT file and click Save.

   The data is exported.

Saving images

Images of the graphs and plots in Expression Console™ Software can be exported for use in other applications two ways:

- The image can be copied to the clipboard by right-clicking the image and selecting Copy to clipboard.
- Save the image as PNG by right-clicking on the graph and selecting Save As PNG or select **File ▸ Save As PNG**. The application prompts the user for a name and location to save the file.

  The Expression Console™ Software allows you to save a graph as a PNG (Portable Network Graphic), which is an image format that enables the graphic to be easily moved into another application or document without loss of quality (Figure 175).

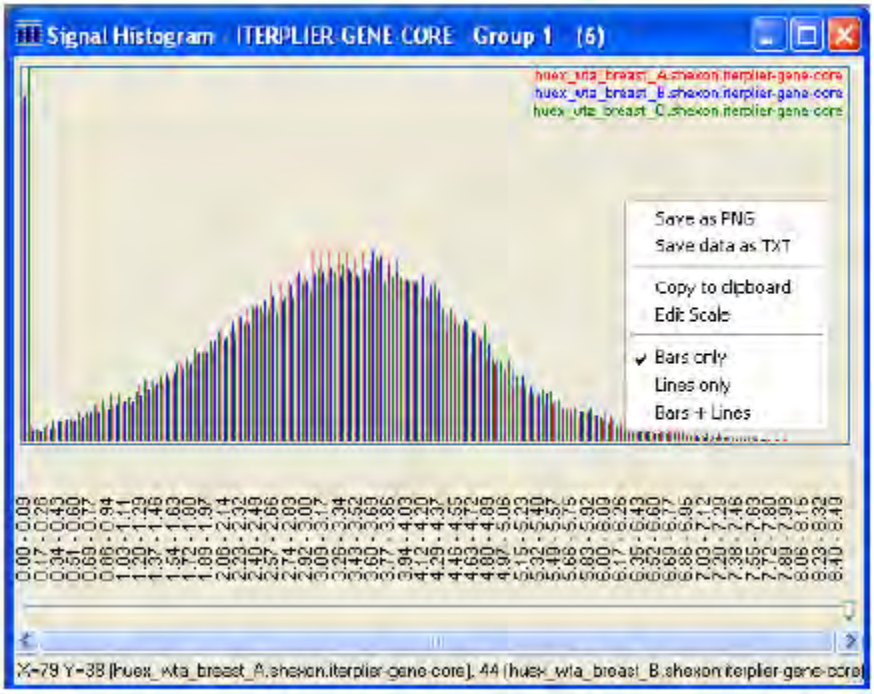


Figure 175 Save the histogram as a PNG.

12 Advanced analysis

Each analysis workflow provided in the Expression Console™ Software application has several parameters, which can be changed from the defaults provided. This section describes how to change the parameters and save them to a configuration file. The file is then saved to the user profile for use in later analyses. Because Exon and 3' Expression Arrays have different algorithm options available, each array type has its own custom analysis configuration setup dialog.

Note: Power Tools is a command line application that provides the ability to further customize algorithm parameters and is intended for use by more experienced users. Power Tools is also available as source code and binaries for several different computing platforms. These can be found at the same location on our web site.

3' Expression Array configurations

This section describes how to set the algorithm parameters for MAS 5.0, PLIER, and RMA for use with 3' Expression Arrays. For additional information about setting MAS 5.0 parameters, see the white paper, Data Analysis Fundamentals and the Statistical Algorithms Reference Guide, on our web site. For explanations of the different normalization and background correction strategies for PLIER and RMA, see Appendix A, "Algorithms".

Create a new configuration – MAS5

1. Select **Analysis ▸ Advanced Expression Configurations**.
   The Custom Analysis Algorithms dialog box opens (Figure 176).

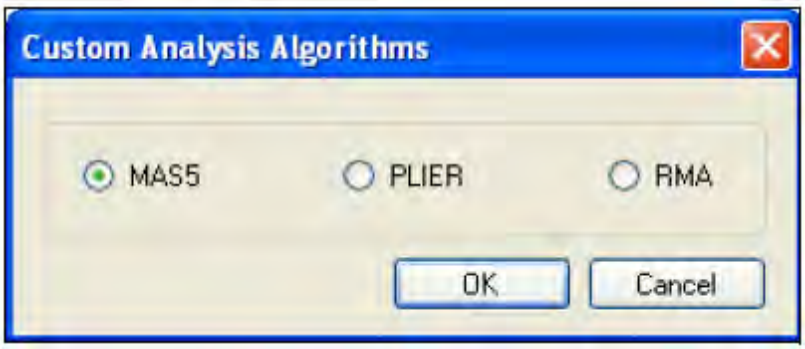


Figure 176 Advanced analysis - MAS5 custom analysis algorithms dialog box.

2. Select the MAS5 algorithm and click OK.
   The New MAS5 Parameter settings dialog box opens (Figure 177).

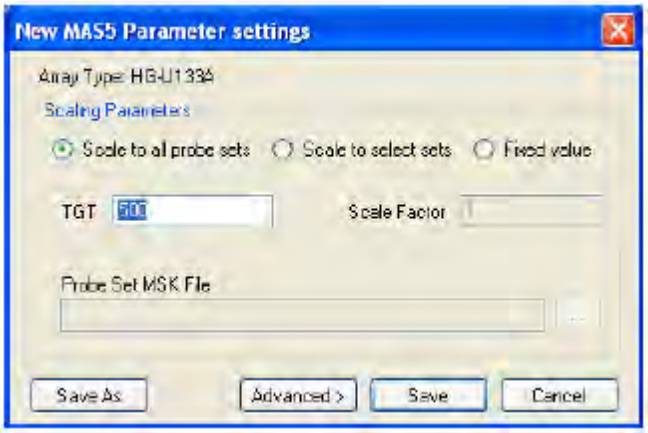


Figure 177 New MAS5 Parameter settings dialog box.

3. Select the settings and click Save.

4. To change the detection parameters or mask out a subset of probes, click the Advanced button Figure 177.
   The window expands to include Detection Parameters (Figure 178).

Figure 178 New MAS5 Parameter settings - advanced window.

5. Enter new parameters and click Save.

   The Save window opens.

   If you select the Scale to select sets radio button, the Probe Set MSK file field is highlighted (Figure 179). Use the browse button [...] to find the appropriate MSK file.

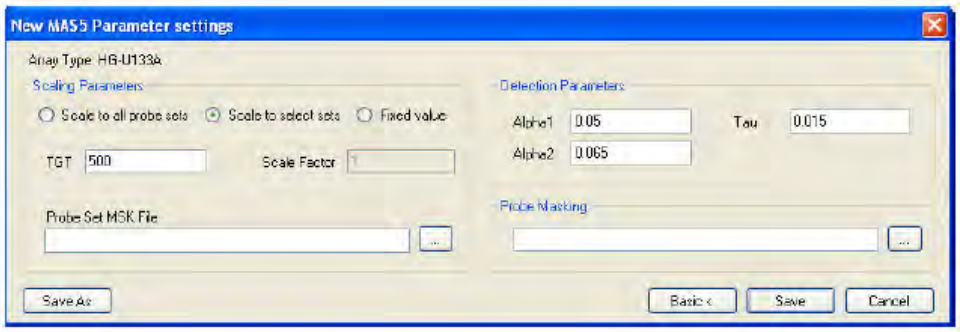


Figure 179 Scale to select sets.

Note: To create a new MSK file, see "Mask (MSK) files" on page 137.

6. Enter a name for this collection of parameters (Figure 180), and click Save.

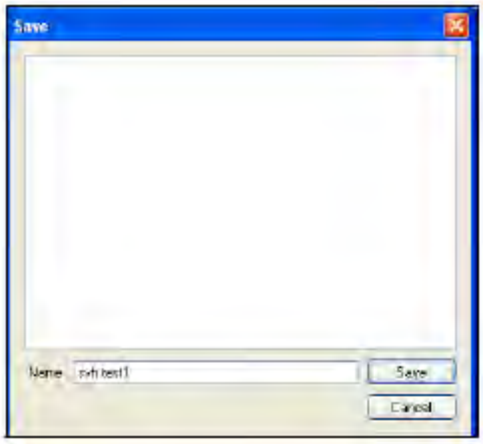


Figure 180 Name the new configuration.

7. To analyze array data with the new parameters, select Analysis → Execute Advanced Configuration, and select the appropriate file name (Figure 181).

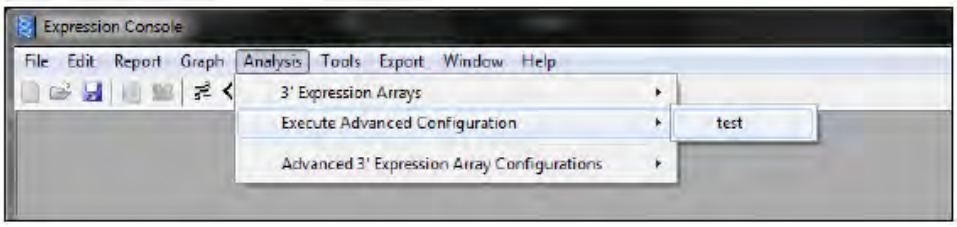


Figure 181 Execute Advanced Configuration.

A dialog box opens to announce analysis initiation (Figure 182).

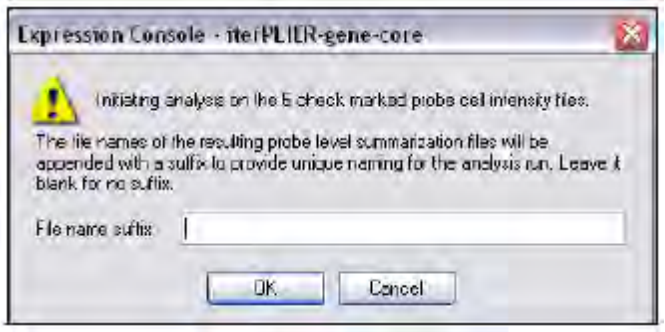


Figure 182 Advanced analysis suffix identifier.

8. Enter a suffix for run identification, if needed, and click OK to start the analysis.

Create a new configuration – PLIER workflow

1. Select **Analysis ▸ Advanced Expression Configurations**.
   The Custom Analysis Algorithms dialog box opens (Figure 183).

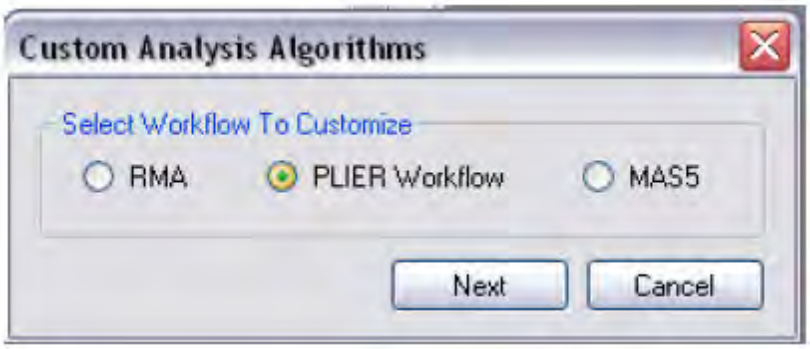


Figure 183 Advanced analysis - PLIER Custom Analysis Algorithms dialog box.

2. Select the PLIER workflow and click OK.
   The New custom workflow for PLIER dialog box opens (Figure 184).

Figure 184 Advanced configuration for PLIER.

3. Select the settings and click Save.
   A dialog box opens to announce analysis initiation (Figure 185).

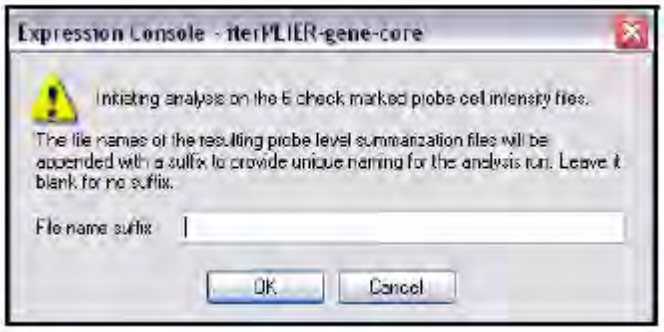


Figure 185 Advanced analysis suffix identifier.

4. Enter a suffix for run identification, if needed, and click OK to start the analysis.

Create a new configuration – RMA

1. Select **Analysis ▸ Advanced Expression Configurations**.
   The Custom Analysis Algorithms dialog box opens

2. Select the RMA algorithm and click OK.
   The New custom configuration for RMA dialog box opens ("Create a new configuration – RMA" on page 135).

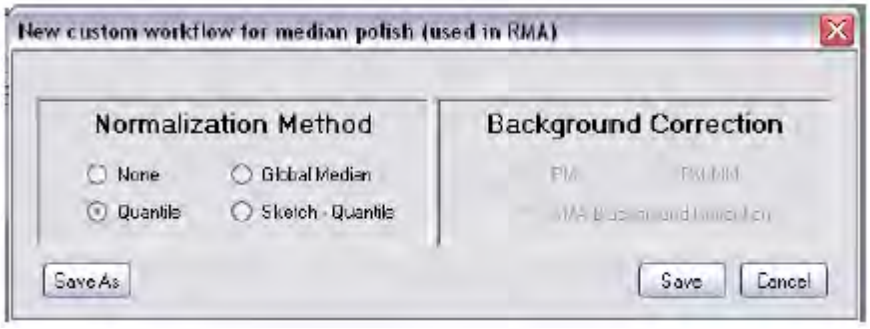


Figure 186 Advanced configuration for RMA.
The RMA Background Correction field is automatically applied to the PM-only probes when selecting RMA.

Note: Since RMA has its own background correction algorithm, Expression Console™ Software does not allow this to be modified.

3. Select the settings and click Save.
   A dialog box opens to announce analysis initiation (Figure 187).

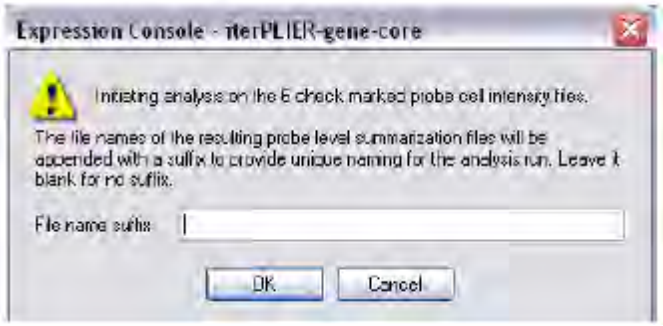


Figure 187 Advanced analysis suffix identifier.

4. Enter a suffix for run identification, if needed, and click OK to start the analysis.

Delete an advanced configuration

1. Select **File ▸ Utilities**.
2. Select the Parameter Management tab (Figure 188).

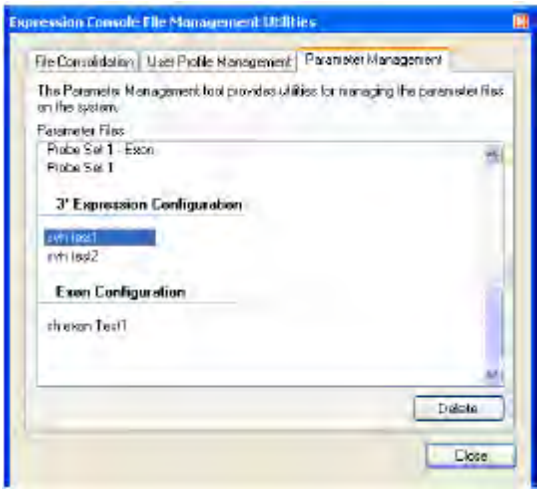


Figure 188 Parameter Management tab - delete configuration.

3. In the appropriate Configuration section, select the configuration to be deleted.
4. Click Delete.

Mask (MSK) files

The Expression Console™ Software application uses MSK files for either setting the scaling factor or excluding or masking certain user-selected probe pairs from an expression analysis. The Selected Probe Sets scaling option, in the advanced configuration for the MAS5 algorithm, adjusts the trimmed mean signal of the selected probe sets on a probe array to the user-specified target signal value. Expression Console™ Software software utilizes user-selected probe sets (specified by a Scale Factor mask file) to calculate the trimmed mean signal and derive the scale factor for the probe array so that Target Signal = Scale Factor x Trimmed Mean Signal of selected probe sets. For a selected set of arrays, predefined MSK files are available from our web site. The other use for mask files is to allow user-selected probe pairs to be excluded or masked from an expression analysis.

Note: For human, mouse, and rat, a list of 100 normalization genes is provided to the user.

Create a Mask file

1. Open **Analysis ▸ Advanced Expression Configurations and select New Mask (MSK) File** (Figure 189).

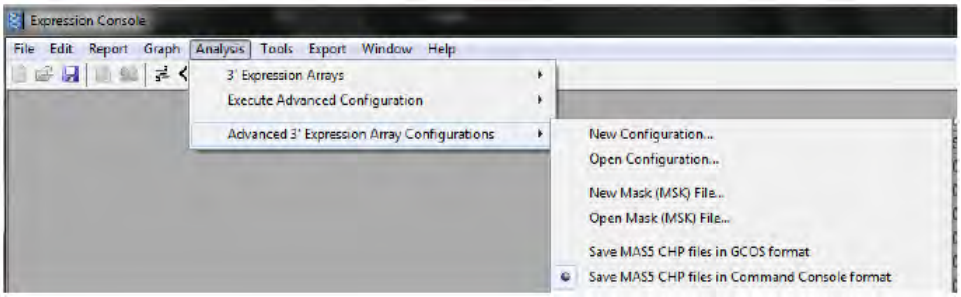


Figure 189 Advanced Expression Array Configurations - New Mask File.

The Select the Array Type dialog box opens (Figure 190).

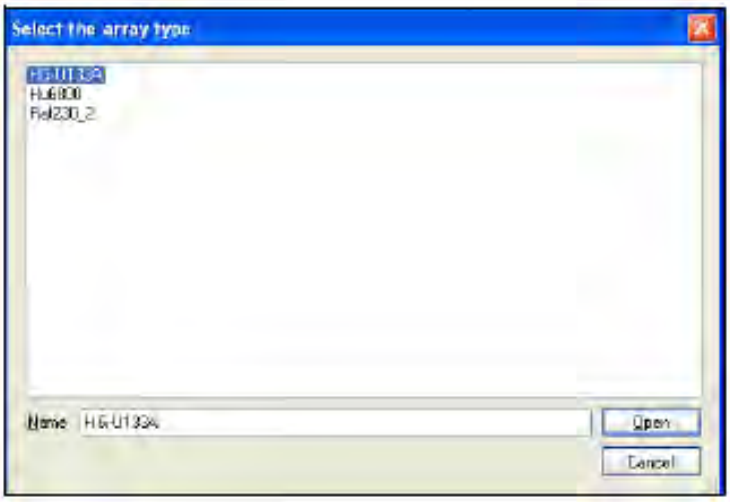


Figure 190 Select the array type.

2. Select the array type and click Open (Figure 191).

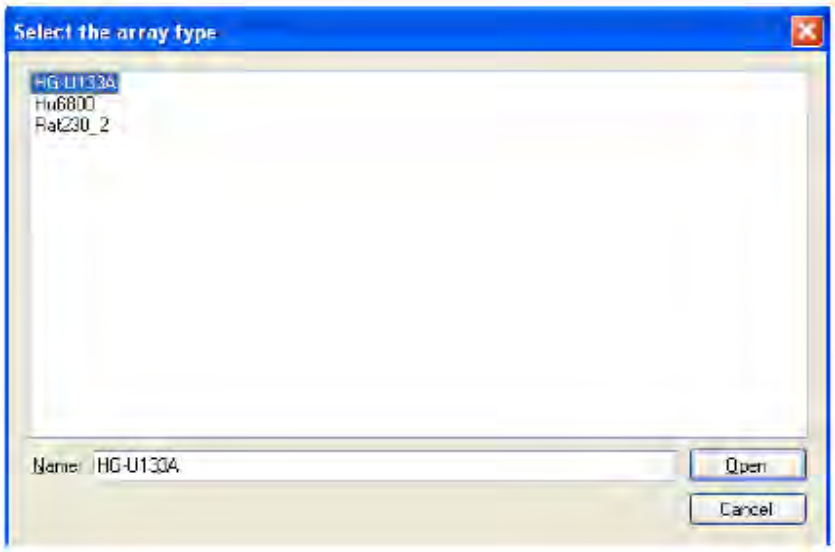


Figure 191 New MASK (MSK) file dialog box.

A window opens displaying all probe sets for that array type (Figure 192).

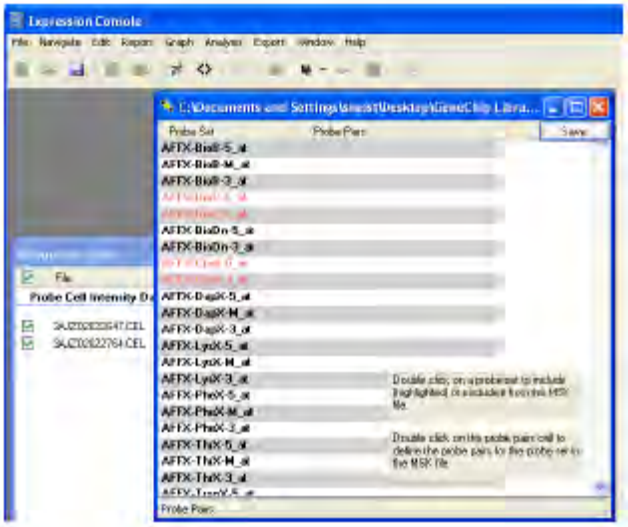


Figure 192 Probe sets for exon array types.

3. Double-click on the probe sets of interest to highlight each set. When you click Save, an MSK file of the selected probe sets is created.

4. Make changes if necessary and click Save, or click the box to exit.

Opening a Mask (MSK) file.

Mask files are specific for the probe array type. The Expression Console™ Software will not open a Mask file that is incompatible with the currently selected probe array type.

View or edit an existing Mask (MSK) file

1. Select **Analysis ▸ Advanced Expression Configurations ▸ Open Mask (MSK) File**.

   The Select the MSK file dialog box opens (Figure 193).

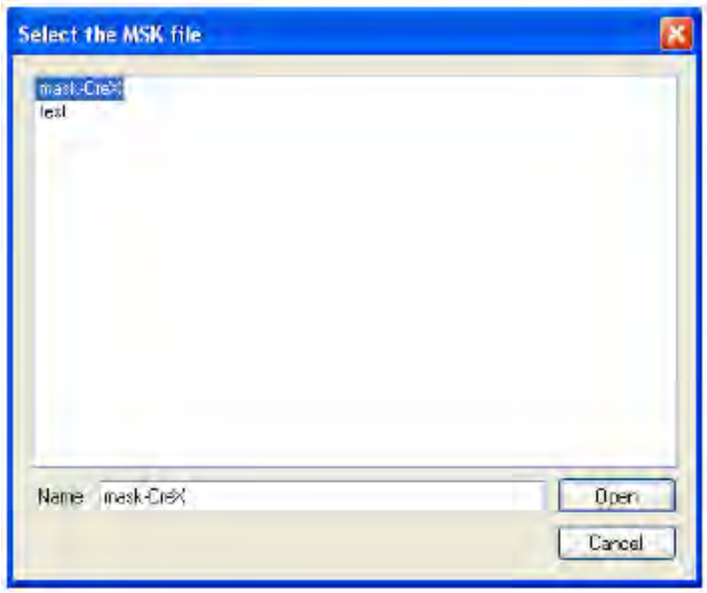


Figure 193 Select the MSK File dialog box.

2. Select the MSK file and click Open.

   The selected MSK file opens displaying the selected probe sets, which are highlighted.

3. Make changes if necessary and click Save, or click the box to exit.

Exporting MAS 5.0 files

Expression Console™ Software provides two options for exporting MAS 5.0 files:

- "MAS 5.0 CHP files in the Expression Console™ Software format" on page 139
- "MAS 5.0 CHP files in the GCOS format" on page 140

MAS 5.0 CHP files in the Expression Console™ Software format

By default, the Expression Console™ Software stores the MAS5 analysis results (Signal, Detection, Detection P-Value, # of Probe Pairs, and # of Probe Pairs Used) in a CHP file using the new GeneChip™ Command Console™ Software format. This new format provides additional features such as:

- An embedded unique file identifier
- A copy of the CEL and DAT headers (if the input CEL/DAT files are also in the GeneChip™ Command Console™ file format). The new file features provide the ability to trace the file's lineage independent of the file name.

To configure the software to write the MAS5 CHP files in the newer GeneChip™ Command Console™ format:

- Select **Analysis ▸ Advanced Expression Configurations ▸ Save MAS5 CHP files in Command Console™ format** (Figure 194).

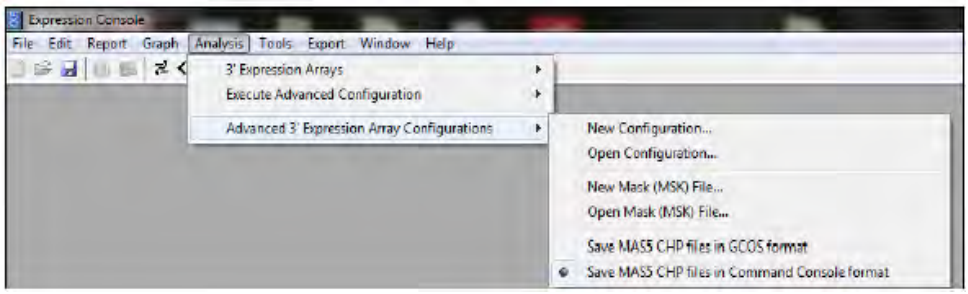


Figure 194 Save MAS5 CHP files in Expression Console™ Software format.

MAS 5.0 CHP files in the GCOS format

To support the transition of software applications to use the newer file parsers (those capable of reading both the older CHP file format as well as the newer AGCC CHP file format), the Expression Console™ software has the ability to write the MAS5 analysis results to the older GCOS format CHP file.

To configure the software to write MAS5 CHP files in the older GCOS format:

- Select **Analysis ▸ Advanced Expression Configurations ▸ Save MAS5 CHP files in GCOS format** (Figure 195).

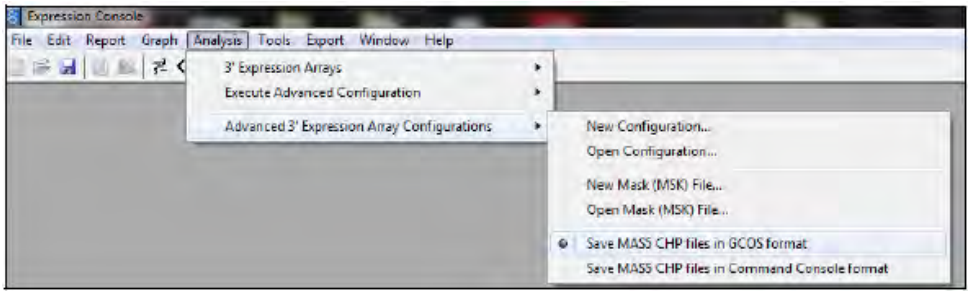


Figure 195 Save MAS5 CHP files in GCOS format

For more information about the Command Console™ Software file format and C++/Java parsers that read both GCOS and Command Console™ files, visit our web site.

Note: Analysis results from PLIER and RMA cannot be stored in GCOS formatted CHP files.

Advanced configuration Exon/Gene analysis

You can create advanced Exon/Gene level configurations for both Exon and Gene Arrays. Not all options are available for all array and levels

Create an advanced configuration for exon or gene arrays

1. From the **Analysis menu, select ▸ Advanced Exon/Gene Array Configurations ▸ New Configuration.**Figure 196

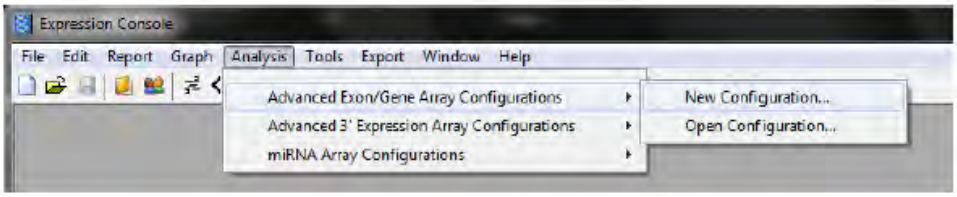


Figure 196 Advanced Exon/Gene Array Configurations - New Configuration.

The Select the probe array dialog box opens (Figure 198).

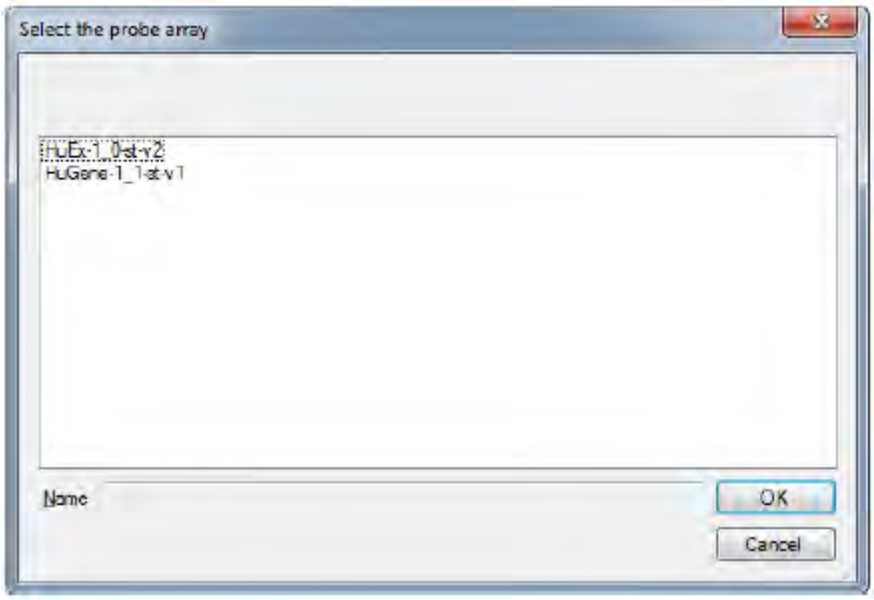


Figure 197 Select the probe array.

The dialog box displays a list of the Exon and Gene Arrays on the system.

2. Select the Array for which you want to create an advanced configuration and click OK.

   The Advanced Analysis Configuration dialog box opens (Figure 198).

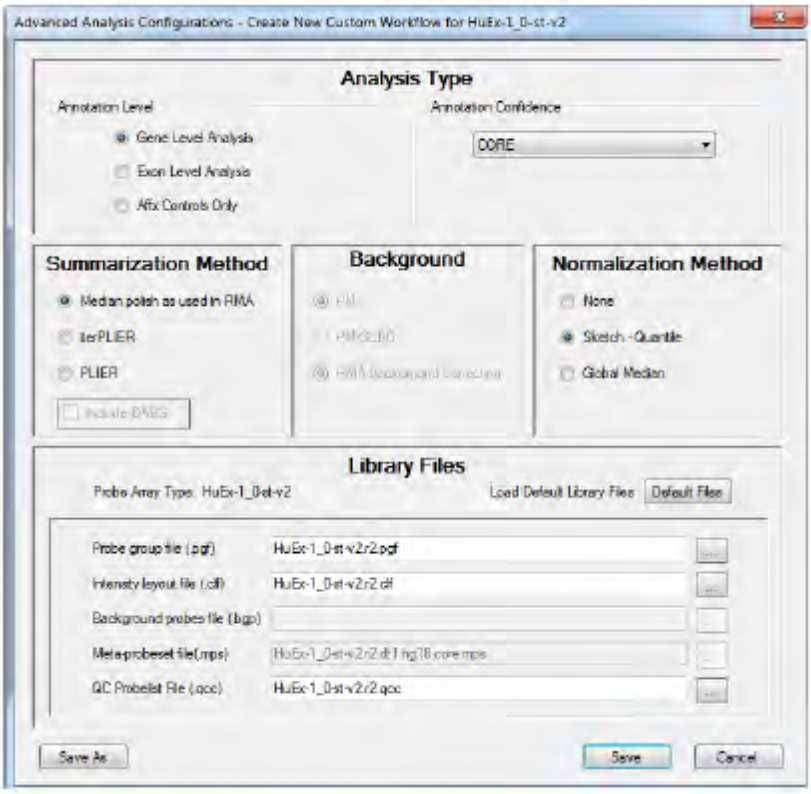


Figure 198 Advanced exon setup dialog box.

3. Select the configuration options:

   a. Select the Analysis Type:

      The setup box in the upper left-hand corner enables users to select their analysis level for the configuration. For details on the different analysis levels, see the white paper, *Exon Probeset Annotations and Transcript Cluster Groupings*, available from our website.

      In the Expression Console™ Software, each level is paired with a default set of library files for that analysis and array type provided. To use other versions of the probe group file, the intensity layout file, background probes, or the QC probeset file, browse to the desired file on the appropriate line of the library file section at the bottom of the screen.

      It is always possible to reload the default files for an analysis by clicking the Default Files button. To use a different meta-probeset file for Gene level analysis or probeset file for Exon analysis, select Other and browse to the file you wish to use in the library file section.

   b. Select the Summarization Method. For more information about the summarization methods, see the algorithm details in *Appendix A, "Algorithms"* and the white papers, *Gene Signal Estimates from Exon Arrays* and *Guide to Probe Logarithmic Intensity Error (PLIER) Estimation*, available on our web site.

c. Select the Background Correction. For more information about background correction, see the white paper, *Exon Background Correction*, available on our web site.

d. Select the Normalization Method. For more information about normalization, see the white paper, *Gene Signal Estimates from Exon Arrays*, available on our web site.

e. Confirm the library file selections in the bottom portion of the window. Note that selecting a confidence level in the Analysis Type automatically selects the appropriate file.

   If you want to use your own custom meta-probeset file or probe list file, select Other in the analysis level to activate the browse feature for that line.

4. Save.

   The Save dialog box opens (Figure 199).

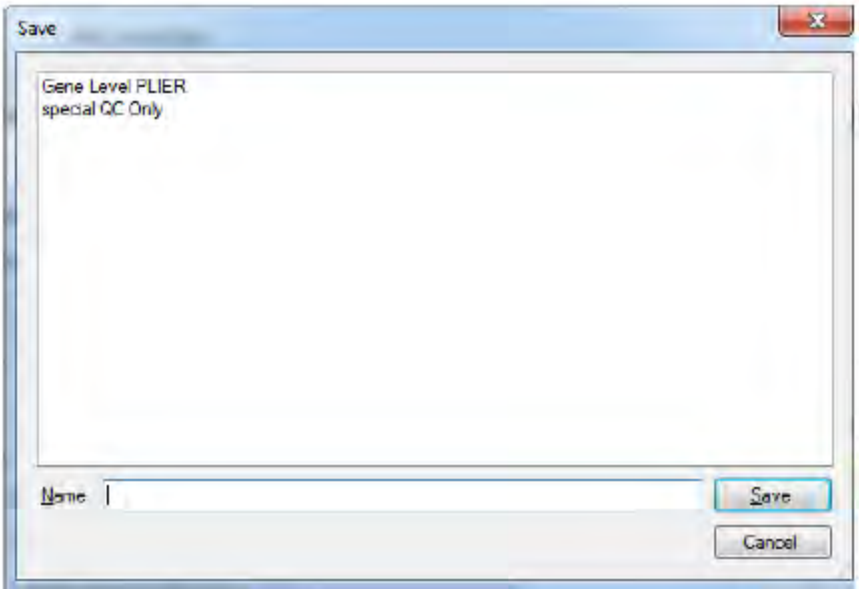


Figure 199 Name the new configuration.

5. Enter a name for this configuration (Figure 199) and click Save.

Analyze array data with the new parameters

1. Select a study and group of CEL files.
2. From the Analysis menu, select **Execute Advanced Configuration ▸ [Configuration File name]** (Figure 200).

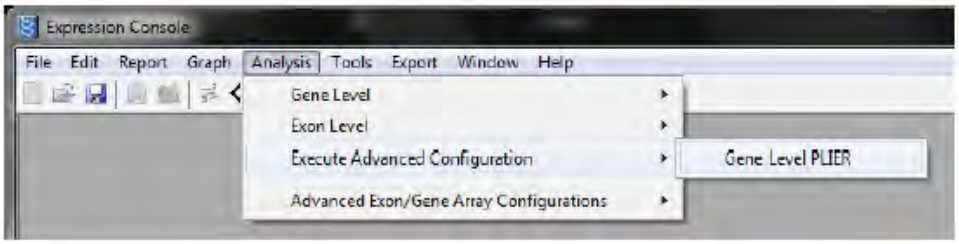


Figure 200 Execute Advanced Configuration.

A dialog box opens to announce analysis initiation (Figure 201).

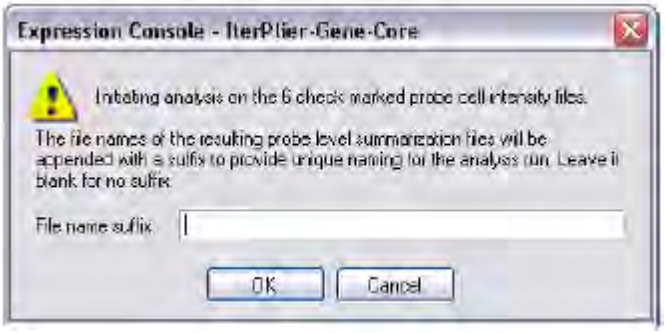


Figure 201 Suffix dialog box.

3. Enter a Suffix for run identification, if needed, and click OK to start the analysis.

Delete an advanced configuration

1. Select **File ▸ Utilities**.
2. Select the Parameter Management tab (Figure 202).

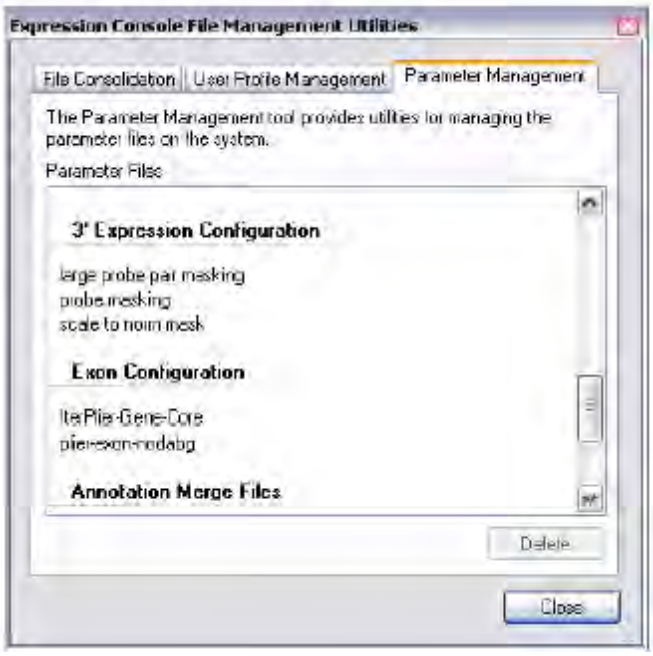


Figure 202 Parameter Management tab - delete configuration.

3. In the appropriate Configuration section, select the configuration to be deleted.
4. Click Delete.

miRNA array configurations

You cannot create a new configuration and change configuration parameters for miRNA arrays in Expression Console™ Software 1.4.

View miRNA array configuration information

1. From the Analysis menu (Figure 203), select **miRNA Array Configurations ▸ Open Configuration....**

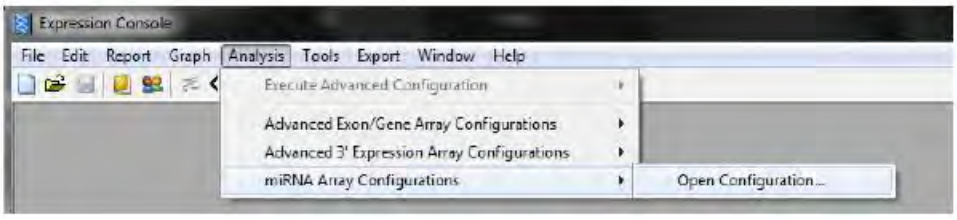


Figure 203 miRNA analysis menu options.

The Select the Probe Array dialog box opens (Figure 204).

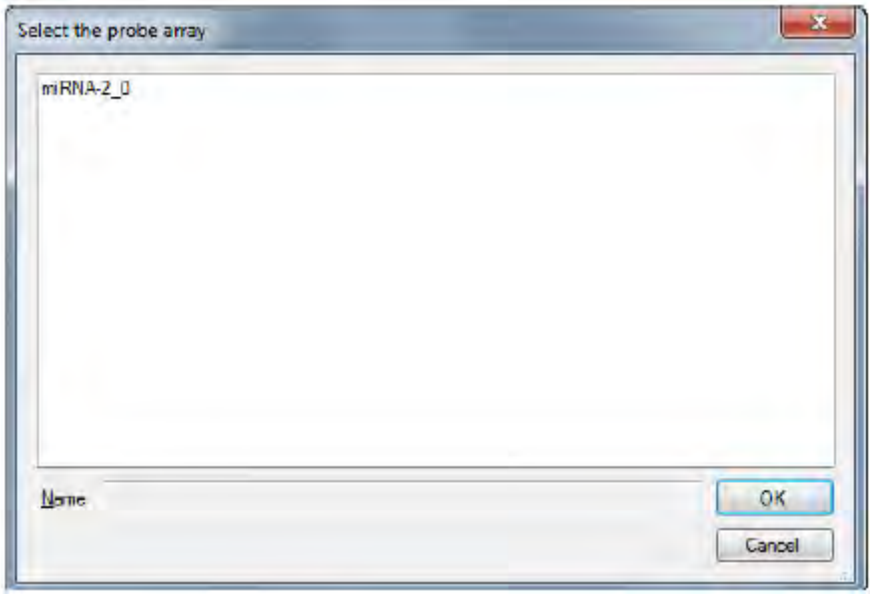


Figure 204 Select the probe array dialog box.

2. Select the probe array in the list and click OK.
   The MicroRNA Parameters dialog box opens (Figure 205).

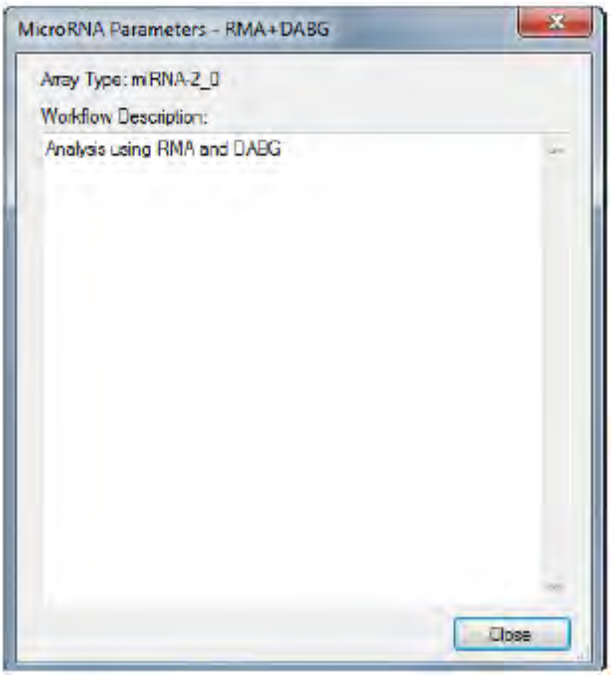


Figure 205 MicroRNA Parameters dialog box.
The dialog box displays basic information about the algorithm.

3. Click Close to close the dialog box

Algorithms

This appendix briefly describes the MAS 5.0, RMA, and PLIER algorithms as they are used to analyze data in the Expression Console™ Software. References and links to various publications, which describe the MAS 5.0, RMA, and PLIER algorithms in detail, are given. See a "Comparison of algorithms" on page 149 for a side-by-side comparison of the assumptions, advantages and disadvantages of using each individual algorithm.

MAS 5.0 algorithm

The MAS 5.0 algorithm uses the Tukey's biweight estimator to provide a robust mean Signal value and the Wilcoxon's rank test to calculate a significance or p-value and Detection call for each probe set. Background estimation is provided by a weighted average of the lowest 2% of the feature intensities. Mismatch probes are utilized to adjust the perfect match (PM) intensity. Linear scaling of the feature level intensity values, using the trimmed mean, is the default to make the means equal for all arrays being analyzed.

The MAS 5.0 algorithm (also known as the Statistical Algorithm) analyzes each array independently. As a result, individual probe-specific affinities can not be considered and the ability to detect small changes between experiment and control samples is reduced in comparison to either RMA or PLIER.

The primary use of the MAS 5.0 algorithm is to obtain a quick report regarding the performance of the arrays and to identify any obvious problems before submitting the final set of arrays to one of the multichip analysis methods (RMA, PLIER).

For a more detailed description of the MAS 5.0 algorithm, see the *Statistical Algorithms Reference Guide* available on our website.

RMA algorithm

The Robust Multichip Analysis (RMA) algorithm fits a robust linear model at the probe level to minimize the effect of probe-specific affinity differences. This approach:

- Increases sensitivity to small changes between experiment and control samples.
- Minimizes variance across the dynamic range, but does compress calculated fold change values.

RMA consists of three steps:

- Background adjustment
- Quantile normalization
- Summarization

This is a multi-chip analysis approach. Therefore, all arrays intended for comparison should be included together in the summarization step.

For a more detailed description of the RMA algorithm, see the publication *Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data*, Biostatistics, April 2003; Vol. 4; Number 2: 249–264.

SST-RMA algorithim

The SST-RMA method is to reduce background and to increase fold change. The SST-RMA method includes GC4 (GC Correction Version 4) background reduction and SST (Signal Space Transformation) intensity normalization to the expression data processing workflow. The GC4 background reduction is achieved through GC count leveling. SST Intensity normalization is an adjustment of probe intensity levels. For detailed information related to GC4 background reduction and SST intensity normalization, download "White Papers: Microarray normalization using SST and GCCN" from our **website**.

PLIER algorithm

The Probe Logarithmic Intensity Error Estimation (PLIER) algorithm method produces a signal by accounting for experimentally observed patterns in probe behavior and handling error at the appropriately low and high signal values.

Similar to RMA, the PLIER algorithm also utilizes data across all arrays submitted for analysis to minimize the effect of probe-specific affinity differences. Unlike RMA, which uses a simple global background correction approach, the standard PLIER configuration uses a probe-specific background providing a higher degree of accuracy at the cost of increased signal variance. Exon Arrays use a more advanced and efficient background correction method utilizing surrogate mismatched probes. (See the technical note, *GeneChip™ Exon Array Design* on our **website**. Signal variance can be addressed following the PLIER analysis by applying a variance-stabilizing data transformation. A simplistic but effective approach is to add a value of 16 to each and every signal value (PLIER + 16 algorithm).

This is a multi-chip analysis approach. Therefore, all arrays intended for comparison should be included together in the summarization step.

For a more detailed description of the PLIER algorithm, see *Guide to Probe Logarithmic Intensity Error (PLIER) Estimation* on our **website**.

Comparison of algorithms

Table 4 A comparison of the MAS5, RMA, and PLIER algorithms used with the Expression Console™ Software application

Algorithm	Advantages	Disadvantages	Assumptions
MAS5	Single-array algorithm is independent of other data in the data set:	Not as sensitive as either RMA or PLIER to small changes in target abundance:	
	• Less computationally intensive than RMA or PLIER • Conservative • Smooth down-weighting of outliers • Positive output values • Minimal bias	• Limited ability to adjust for probe-specific affinity differences • Unstable variance at low end • Lower precision in signal calculation	• Single-array analysis • Multiplicative error • Signal is adjusted by the mismatch probes (PM– MM) • Background imputed to handle negative differences
RMA	Minimizes the variance seen across the arrays	Not as sensitive as PLIER in the ability to detect small fold changes	
	• Higher reproducibility of signal over single array analyses • Good differential change detection • Variance stable on log scale	• In cases where feature intensities disagree, may have more than one solution (mitigated by median polish) • Positive bias contributed to signal values • Compresses fold changes for low intensity probe sets	• Multiple-array analysis • Multiplicative error • PM (perfect match) only • Single background is used to adjust each intensity
PLIER	Ability to detect small fold changes:	More variance in individual signals than seen with RMA	

Algorithm	Advantages	Disadvantages	Assumptions
	• Higher reproducibility of signal (lower coefficient of variation) without loss of accuracy • Higher sensitivity to changes in abundance for targets near background • Dynamic weighting of the most informative probes in an experiment to determine signal • High degree of accuracy for signal and fold change calculations • Lack of bias	• Computationally intensive • In cases where feature intensities disagree, may have more than one solution (mitigated by median polish) • Performance relative to amount of model data provided • Variance not stable on log scale	• Multiple-array analysis • Mixed error model PM– MM, PM only, etc. • Multiple background options • Smoothly handles intensities below background

B Algorithm parameters and outputs

The Expression Console™ Software displays column headings for individual parameters and outputs in the Report window (Report → View Full Report), for each of the 3' Expression Array algorithms (MAS 5.0, RMA, PLIER). This appendix gives column heading definitions.

MAS 5.0 column headings

Column heading	Description
Threshold Test	The minimum number of probe pairs a probe set must have in order for the probe set data to be included in the calculation of the report statistics.
HZ	Number of horizontal zones used in background subtraction.
VZ	Number of vertical zones used in background subtraction.
BG	Minimum, maximum, average, and standard deviation of the background intensity calculated for the probe array.
α1 (α1)	Significance level for the detection p-value in an analysis – α1 is a user-modifiable parameter that is set in the New MAS5 Parameter Settings dialog box (**Analysis ▸ Advanced Expression Configurations ▸ New Configuration**). If the probe set detection p-value is < α1, the call is present.
α2 (α2)	Second significance level for the detection p-value in an analysis and a user-modifiable parameter set in the same dialog box as α1 – If the probe set detection p-value is greater than or equal to α2, the call is absent. If α1 is less than or equal to the Detection p-value < α2, the call is marginal.
τ (τ)	A user-modifiable parameter, ideally set to a value that is a little larger than the median of the discrimination scores of the probe sets whose targets are absent, to avoid false detected calls.
Noise (Raw Q)	The degree of pixel-to-pixel variation among the probe cells used to calculate the background.
Scale Factor (SF)	The scale factor specified in the Scaling tab of the Expression Analysis Settings dialog box or computed by the algorithm.

Column heading	Description
Norm Factor (NF)	The normalization factor computed by the algorithm.
Scale Mask	Contains the name and location of the mask file if one was used for the analysis.
RawQ	Baseline Noise. The degree of pixel-to-pixel variation among the probe cells used to calculate the background in the baseline probe array.
BG Avg	The average background intensity calculated for the probe array.
BG Std	The standard deviation background intensity calculated for the probe array.
BG Max	The maximum background intensity calculated for the probe array.
BG Min	The minimum background intensity calculated for the probe array Noise Avg. The average noise calculated for the probe array.
Noise Std	The standard deviation noise calculated for the probe array.
Noise Max	The maximum noise calculated for the probe array.
Noise Min	The minimum noise calculated for the probe array.
Corner+ Avg	The average cell intensity for the sense probe cells used in the grid alignment process.
Corner+ Count	The corner count for the sense probe cells used in the grid alignment process.
Corner-Avg	The average cell intensity of the antisense probe cells used in the grid alignment process.
Corner-Count	The corner count of the antisense probe cells used in the grid alignment process.
Central-Avg	The average cell intensity for the nine probe cells that comprise the cross at the center of an antisense probe array.
Central-Count	The central count for the nine probe cells that comprise the cross at the center of an antisense probe array.
#Probe Sets Exceeding Probe Pair Threshold	The number of probe sets that exceed the probe pair threshold.
Probe Pair Threshold	The minimum number of probe pairs a probe set must have in order for the probe set data to be included in the calculation of the report statistics.
Control Direction	The direction (sense or antisense) of the target (sample).
#P	The number of probe sets present.
%P	The percent of probe sets present.
Signal(P)	The average signal for the probe sets defined as present.

Column heading	Description
#M	The number of probe sets whose detection is marginal.
%M	The percent of probe sets whose detection call is marginal.
Signal(M)	The average signal for probe sets whose detection call is marginal.
#A	The number of probe sets absent.
%A	The percent of probe sets called present that are absent.
Signal(A)	The average signal for the probe sets defined as absent.
Signal(All)	The average signal for all probe sets on the array.

Column headings for RMA and PLIER

Metrics are dependent on the analysis algorithm, array type, and the level; therefore, not all of the metrics are always present in the Full Report.

Column heading	Description
Raw Corner+ Avg	The average cell intensity prior to any background for the sense probe cells adjustment used in the grid alignment process.
Raw Corner+Count	The count of the number of sense probe cells used in grid alignment.
Raw Corner-Avg	The average cell intensity prior to any background for the antisense probe cells adjustment used in the grid alignment process.
Raw Corner-Count	The count of the number of antisense probe cells used in grid alignment.
Raw Central-Avg	The average cell intensity before background adjustment for the sense probe cells that comprise the cross at the center of the array.
Raw Central-Count	The average cell intensity before background adjustment for the anti sense probe cells that comprise the cross at the center of the array.
Spike-probeID-signal	The signal for the probe sets that correspond to the labeling and hybridization spike controls.
pm_mean	The mean signal for all of the PM probes on the array.
mm_mean	The mean signal for all of the MM probes on the array.
bgrd_mean	The average signal for the probes used to calculate the background.

Column heading	Description
pos_vs_neg_auc	The area under the curve (AUC) for a receiver operator curve (ROC) comparing the intron controls to the exon controls by applying a threshold to the probe set summary. The ROC curve is generated by evaluating how well the probe set summary separates the positive controls from the negative controls (e.g., exon from intron). The assumption (which is only valid in part) is that the negative controls are a measure of false positives and the positive controls are a measure of true positives. An AUC of 1 reflects perfect separation whereas as an AUC value of 0.5 would reflect no separation. Note that the AUC of the ROC curve is equivalent to a rank sum statistic used to test for differences in the center of two distributions.
Signal(A)	The average signal for the probe sets defined as absent (Exon-level DABG > 0.01).
Signal(All)	The average signal for all probe sets on the array.
apt-opt-cdf-file	The cdf file used for the analysis.
apt-opt-probe-count	The total number of probes (features) on the array whether or not the probes are used.
apt-opt-qc-groups-file	The file used to identify the control probe and probe sets for the analysis run.
#P	The number of probe sets present. For exon level analysis, the DABG p-value is less than or equal to 0.01.
%P	The percent of probe sets present. For the exon level analysis, this is defined as the DABG probe level p-value is less than or equal to 0.01.
Signal(P)	The average signal for the probe sets defined as present (Exon-level DABG ≤0.01).
#A	The number of probe sets absent, For exon level analysis, the DABG p-value is greater than 0.01.
%A	The percent of probe sets called present absent, For exon level analysis, the DABG p-value is greater than 0.01.
Signal(A)	The average signal for the probe sets defined as absent (Exon-level DABG >0.01).
Signal(All)	The average signal for all probe sets on the array.

Probe set suffixes

A number of standard metrics are created for particular groups of probe sets on the arrays. The key below describes the suffix and descriptions for the following groups of probe sets:

- *all_probeset* (all probe sets within a group)
- *bac_spike* (bacterial spikes and hybridization controls)
- *polya_spike* (labeling controls)
- *neg_control* (negative control probes)
- *pos_control* (positive control probes)

Table 5 Probe set suffixes.

KEY: Suffix for output	Definition
_probesets	From the QC Report, the number of probe sets actually analyzed.
_atoms	From the QC Report, the number of PM–MM or PM–GCBG probe pairs actually analyzed.
_mean	The mean signal value for all the probe sets.
_stdev	The standard deviation for all of the probe sets analyzed.
_mad_residual_mean	The mean absolute deviation of the residual for a chip versus all chips in the data set.
_mad_residual_stdev	The standard deviation of the residual for a chip versus all chips in the data set.
_rle_mean	The mean absolute relative log expression (RLE). This metric is generated by taking the probe set summary for a given chip and calculating the difference in log base 2 from the median value of that probeset over all the chips. The mean is then computed from the absolute RLE for all the probe sets for a given CEL file.
_rle_stdev	The standard deviation of the relative log expression (RLE). This metric is generated by taking the probe set summary for a given chip and calculating the difference in log base 2 from the median value of that probeset over all the chips. The standard deviation is then computed from the absolute RLE for all the probe sets for a given CEL file.
_percent_called	The percent of probe sets from the exon analysis called present (DABG ⩽0.01).

Documentation and support

Customer and technical support

Visit **thermofisher.com/support** for the latest in services and support, including:

- Worldwide contact telephone numbers
- Product support, including:
  - Product FAQs
  - Software, patches, and updates
  - Training for many applications and instruments
- Order and web support
- Product documentation, including:
  - User guides, manuals, and protocols
  - Certificates of Analysis
  - Safety Data Sheets (SDSs; also known as MSDSs)

    Note: For SDSs for reagents and chemicals from other manufacturers, contact the manufacturer.

Limited product warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale found on Life Technologies' website at **www.thermofisher.com/us/en/home/global/terms-and-conditions.html**. If you have any questions, please contact Life Technologies at **www.thermofisher.com/support**.