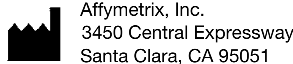
applied biosystems

Transcriptome Analysis Console (TAC) 4.0.2

USER GUIDE

Publication Number 703150
Revision 8

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


Affymetrix, Inc.
3450 Central Expressway
Santa Clara, CA 95051

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

DISCLAIMER

TO THE EXTENT ALLOWED BY LAW, LIFE TECHNOLOGIES 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. 703150

Revision	Date	Description
8	January 2019	4.0.2 release. Update WikiPathways support; Bug fixes related to Q_probeset-summarize and analysis.
7	December 2017	4.0.1 release. Bug fixes related to LIMMA data analysis and foreign language Operating Systems.
6	May 2017	4.0 release includes, integration of Expression Console and Transcriptome Analysis Software into a single application; and implementation of R/LIMMA and event pointer for data analysis.
5	December 2015	3.1 release includes, 2-way ANOVA and wikipathway p-value calculations.
4	December 2014	3.0 release includes, updates related to ANOVA, hierarchical clustering graph, wikipathway, and AED exporting compatibility with the Chromosome Analysis Suite (ChAS) application.
3	March 2014	2.0 release includes, alternative splicing event identification and miRNA-mRNA interaction network capabilities.
2	March 2013	1.0.1 release. Minor revision to system requirement information. Revised table column definitions.
1	January 2013	Initial 1.0 release.

Important Licensing Information

This product may be covered by one or more Limited Use Label Licenses. By use of this product, you accept the terms and conditions of all applicable Limited Use Label Licenses.

Legal entity

Affymetrix, Inc. | Santa Clara, CA 95051 USA | Toll Free in USA 1 800 955 6288

TRADEMARKS

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

©2019 Thermo Fisher Scientific Inc. All rights reserved.

Contents

CHAPTER 1 Introduction . . . 14

Overview . . . 14
Features . . . 15
Recommended hardware . . . 15
Minimum requirements . . . 15
Setting up . . . 16
  Assigning a library folder/path . . . 17
  Downloading library files . . . 17
  Installed array types . . . 18
    Customizing an Installed Array's Algorithm Settings . . . 18
  Activating automatic library file update checks . . . 21
  Setting up custom proxy settings . . . 21
Open Existing Result window tab . . . 22
  Browsing for existing analysis result . . . 22

CHAPTER 2 Analysis setup . . . 23

Setting up a new analysis . . . 23
  Importing CEL files . . . 23
  Importing CHP files . . . 24
  Assigning custom sample attributes . . . 24
    ARR Files . . . 24
    Tab-delimited text files . . . 24
    Direct entry . . . 25
    Editing a sample attribute . . . 26
    Importing array physical attributes . . . 26
    Exporting sample attributes . . . 26
    Sorting columns . . . 26
Showing QC graphs . . . 27
  Principal Component Analysis (PCA) graph . . . 27
  Using the PCA graph window . . . 28
  Interacting with the PCA graph . . . 29
    Assigning mouse button functions . . . 29
    Zooming and resizing . . . 29
  PCA graph options . . . 30
    Save as PNG . . . 30
    Print . . . 30
    PCA Options . . . 30
    Export Coordinates . . . 31
    Show Legend . . . 31
    Clear Selection . . . 31

Contents

Signal Box Plot 32
Signal Plot graph options 32
Save as PNG 32
Print 32
Scale Settings 33
Clear Selections 33
Show ToolTip 33
Viewing and Labeling 33
Probe Intensities 34
Probe Intensities graph options 34
Save as PNG 34
Print 34
Color Settings 34
Removing listed files before analysis 35

CHAPTER 3 Algorithm and summarization selection 36

Supported arrays 36
Gene Level + Alt-Splice analysis 36
Selecting an analysis type to run 36
Gene Level analysis 37
Selecting an analysis type to run 37
3'IVT Expression arrays 38
Selecting an analysis type to run 38
miRNA arrays 39
Selecting an analysis type to run 39
Setting up a comparison analysis 40
New Analysis window tab 40
Setting up One-Way ANOVA comparisons 40
Setting up Two-Way ANOVA comparisons in the New Analysis window tab 41
Setting up a comparison analysis using repeated measure 42
Utilizing Batch Effect in a comparison analysis 43
Assigning a Normalization Group in a comparison analysis 44
Random Factor (Advanced) 44
Using the Random Factor 45
Real Covariate (Advanced) 45
F-tests 45
Comparison Wizard 46
Comparison questions 46
Setting up a Two Way ANOVA 50

Contents

■ **CHAPTER 4** Analysis results and QC review 51

Viewing options 52
Changing the default view 52
Viewing a completed analysis 53
Browsing for existing analysis result not displayed 54
Introduction to QC Metrics 55
Spike Controls (Eukaryotic Hybridization Controls) 55
Internal Control Genes (Housekeeping Genes) 56
Labeling Controls 56
ERCC Controls 56
Quality control metrics for CHP files 57
Sample Table Views 57
Customizing views 57
Removing a custom view 57
Sample Table filters 58
Exporting the Sample Table 59
Adding custom user attributes 59
Reanalyzing samples from the Sample table 60
Sample QC Graphs 62
PCA plot 62
Sample display options 62
PCA Graph display options 63
Selecting data points in the PCA Plot 65
Exporting and PCA options 65
Save as PNG 65
Print 65
PCA Options 65
Export Coordinates 66
Show Legend 66
Clear Selection 66
Labeling Controls 67
Labeling Controls Graph Options 68
Save as PNG 68
Print 68
Scale Settings 68
Clear Selections 68
Transpose Graph 68
Hybridization Controls 68
Spike Controls 68

Contents

Hyb Controls Graph options 69
- Save as PNG 69
- Print 69
- Scale Settings 69
- Clear Selections 69
- Transpose Graph 69

Positive vs Negative AUC 70

Pos vs Neg AUC Graph options 71
- Save as PNG 71
- Print 71
- Scale Settings 71
- Clear Selections 71
- Transpose Graph 71

Signal Box plot 72

Signal Box plot graph options 73
- Save as PNG 73
- Print 73
- Scale Settings 73
- Clear Selections 73
- Show ToolTip 73

Creating a custom line graph 74
- ERCC Controls 74

Line Graph options 76
- Save as PNG 76
- Print 76
- Scale Settings 76
- Clear Selections 77
- Transpose Graph 77
- Show ToolTip 77

Distance Matrix Graph (EGA analysis only) 77

Using the Distance Matrix Graph’s tool bar 79

Selecting samples in a cluster 79

Distance Matrix graph options 80
- Save as PNG 80
- Print 80
- Color Settings 80
- Show Legend 81
- Show ToolTip 81
- Show Cluster Points 81
- Clear Selections 81

Contents

CHAPTER 5 Analysis summary view, Gene lists and Venn diagram. . 82

Analysis summary view 82
Gene lists and Venn diagrams 83
Gene list tool bar 83
Viewing the contents of a Gene list 86
Adding a column to the table 87
Adding/Removing selected row(s) in a Gene list 87
Adding from right-click menu 87
Removing from right-click menu 87
Copying selected rows and IDs 88
Copying selected Row(s) 88
Copying selected Cell(s) 88
Copying DNA sequence (5’ to 3’) 88
Accessing external databases 89
Searching the NetAffx website 89
Venn Diagram 90
Exporting a Gene list within the Venn Diagram. 91

CHAPTER 6 Gene view 92

Viewing options 92
Changing the default view 92
Common table functions 93
Using the comparison feature 94
Customizing views 95
Removing a custom view 96
Working with column headers 96
Rearranging default columns 96
Showing or hiding columns 97
Sorting columns 98
Double-Click Sorting Method 98
Adding a column to the table 98
Copying column data to your clipboard 98
Filtering column data 99
Editing filtering properties 99
Managing filters 100
Table filters 100
Clearing filters 101
Table options 101
Copying selected row(s) 101
Copying selected cell(s) 101

Contents

Copying DNA sequence (5’ to 3’) 101
Adding/removing selected row(s) in a Gene list 102
Adding from right-click menu 102
Removing from right-click menu 102
From Gene list drop-down 102
Saving the current table as a Gene list 102
Adding selected row(s) to a Gene list 103
Removing selected row(s) from a Gene list 103
Filtering table by Gene list 103
Accessing external databases 103
Searching the NetAffx website 104
Searching keywords 104
Exporting options 104
Exporting the current table 104
Exporting all rows and columns 104
Exporting AED 105
Exporting sample signals 105
Custom export 105
Customizing annotations 106
Selecting an annotation file 108
Selecting an Annotation Assignment 108
Adding Annotation Columns 108
Check/Uncheck All 108
Gene Level Differential Expression analysis graphs 109
Scatter Plot graph 109
Lassoing Genes of Interest within the Scatter Plot 110
Obtaining information related to individual probe sets 111
Scatter Plot graph options 111
Save as PNG 111
Print 111
Scale Settings 111
Color Settings 112
Show Legend 112
Show ToolTip 113
Clear Selections 113
Volcano Plot graph 113
Lassoing genes of interest within the Volcano Plot 114
Obtaining Information Related to Individual Probe Sets 114
Volcano Plot graph options 115
Save as PNG 115
Print 115
Scale Settings 115
Color Settings 116

Show Legend 116
Show ToolTip 116
Clear Selections 116
Hierarchical Clustering graph 117
Labeling the hierarchical clustering graph 118
Selecting interesting probe sets 118
Method 1: Zoom Slider 118
Method 2: Mouse Zoom 118
Restoring Full View 118
Refreshing the View 118
Obtaining information related to individual TC IDs 120
Selecting Items in a cluster 120
Hierarchical Clustering graph options. 121
Save as PNG 121
Print 121
Colors Settings 121
Show Legend 121
Show ToolTip 121
Show Cluster Points 122
Clear Selections 122
WikiPathways 122
Overview of the WikiPathways Window 123
Updating an outdated pathway 124
Changing the order of the Histogram panel 125
Changing the up/down regulated and significance colors 126
Adjusting the Histogram panel size 126
Adjusting the WikiPathways graph size 127
Using the scroll wheel 127
Panning the WikiPathways graph 127
Using the table and WikiPathways graph 128
WikiPathways graph tool bar 128
Download/Update Pathways 128
Export Pathway Metrics 129
Show Pathway Comments 129
Save as Image 130
Print 131
WikiPathways Trademark/Legal Page 131
Sample Signals 132
Signal Sample Bar graph options 133
Save as PNG 133
Print 133
Scale Settings 133
Show ToolTip 133
Clear Selections 133

Contents

Chromosome Summary graph 134
Changing graph colors 135
Selected Chromosome view 136
Exporting options 136
Save as PNG 136
Print 136
Interaction Network graph 137
Using the Graph Option 138
Using the Tool Bar Buttons 140
Identifying miRNA Data Nodes, Genes, and Interaction Network Relationships 140
Using the Interaction Network Target table 144
Showing or Hiding Interaction Network Target Table Columns 145
Full miRNA Table 146
Searching Keywords 146
Gene View summary 147
Gene View Sample table 148

CHAPTER 7 Alternative splicing analysis 150

Viewing options 151
Changing the default view 151
Common table functions 151
Using the comparison feature 152
Customizing views 152
Removing a custom view 153
Working with column headers 153
Rearranging default columns 153
Showing or hiding columns 153
Sorting columns 154
Double-click sorting method 154
Adding a column to the table 154
Filtering column data 154
Editing filtering properties 155
Copying column data to your clipboard 155
Managing filters 156
Table filters 156
Clearing filters 157
Copying selected rows and IDs 157
Copying selected row(s) 157
Copying selected cell(s) 157
Copying DNA sequence (5' to 3') 157
Accessing external databases 158
Searching the NetAffx website 158
Searching keywords 159

Contents

Exporting options 159
Exporting the current table 159
Exporting all rows and columns 159
Exporting AED. 160
Exporting sample signals 160
Custom export 160
Using the Gene table in Alt-Splice view 162
Viewing options for analyses with multiple conditions 163
Showing or hiding columns in the Alt-Splice Gene table. 164
Customizing annotations 167
Selecting an annotation file 168
Selecting an annotation assignment 168
Adding annotation columns 168
Check/uncheck all 168
Adding/removing selected row(s) in a Gene list 168
Adding from right-click menu 168
Removing from right-click menu 168
From Gene list drop-down 169
Saving the current table 169
Adding a selected row(s) 169
Removing a selected row(s) 169
Filter table by Gene list 169
Using the Alt-Splice PSR/JUC table 170
Showing or hiding columns in the Alt-Splice PSR/JUC table 170
Copying selected cell(s) 172
Copying selected row(s) 172
Adding a column in the Alt-Splice table 172
Exporting options 173
Exporting the current table 173
Exporting all rows and columns 173
Exporting sample signals 173
Custom export 174
Using the Alt-Splice table and viewer 176
Visualization tab features 176
Changing factory set scale limits for Splicing Index and Signal Scale 177
Setting new scale ranges 177
Changing scale colors 177
Alt-Splicing visualization options 178
To save the graph as a PNG 178
Printing the currently displayed graph 178
Opening the Genomic Region in the Alt Splice Visualization in an external browser . 178
Changing the Color Settings 178

Contents

Copying Gene Position or Gene Info 178
Showing/Hiding the ToolTip 178
To clear selected items in the graphs 178
Transcript cluster details in the Alt Splice Visualization graph 179
Changing the Alt Splice Visualization display 179
Structure Vview 180
Genomic view (truncated) 181
Genomic view 181
Visualization display options 182
Displaying the PSR intensities for each condition 183
Isoform display 184
Isoform display options 184
Isoforms 184
Sort Isoforms 184
Color By 185
3D line plot 185
3D Line plot tool bar 186
3D Line plot graph options 186
Save as PNG 186
Print 186
Scale Settings 186
Showing/Hiding the ToolTip 186
Turning Graph Legend On/Offf 186
Clearing 3D Line Plots 187
Working with the 3D Line plot 187
Alt-Splice view summary table 188
How TAC 4.0.2 assigns Locus types 189
Alt-Splice View sample table 191

APPENDIX A Algorithms 193

Exploratory Grouping Analysis (EGA) module 193
EGA parameters 195
EGA parameter option definitions 195
Batch Effects module 197
Limma Differential Expression analysis 199
Anova 201
eBayes 201
Alternative Splicing Analysis 202
Event Pointer Alternative Splicing Analysis 202
Splicing Index (SI) algorithm 204
Performing an Alternative Splicing Analysis 204

Contents

Splicing Event estimation score 206
Splicing Event Estimation 206
Event Estimation Score 209
Event Pointer Algorithm 210
Event score 210
MAS 5.0, RMA and PLIER 211
MAS 5.0 211
MAS 5.0 Column Headers and Definitions 212
RMA 212
RMA and PLIER Column Headers and Definitions 213
SST-RMA 213
PLIER 213
Comparison of algorithms. 214
Benjamini-Hochberg Step-Up FDR-controlling Procedure 215
Fold change 215
Hierarchical clustering 215
Probe set suffixes 216
Chromosome naming scheme 217
References 218

1 Introduction

Overview

Transcriptome Analysis Console (TAC) combines the CEL file analysis and QC features of Expression Console and the statistical analysis of TAC into a single streamlined software application.

This newly combined application provides an easy way to create and QC summarized values (CHP files) of Expression Array feature intensity (CEL) files followed by downstream statistical analysis, including extensive graphs and tables.

TAC serves as a friendly graphical interface to a set of R and Bioconductor modules including:

- **Limma** [1]
- **EventPointer** [2]
- **Dbscan** [3]
- **Rtsne** [4]
- **Apcluster** [5]

References:

1. **Limma :** Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.
2. **EventPointer :** Juan Pablo Romero and Angel Rubio (2015). EventPointer: EventPointer: Analysis of Alternative Splicing Using Microarray Data. R package version 1.0.
3. **Dbscan:** Michael Hahsler (2016). dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 0.9-8. URL: https://CRAN.R-project.org/package=dbscan
4. **Rtsne:** L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9(Nov):2579-2605, 2008. L.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct):3221-3245, 2014. Jesse H. Krijthe (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation, URL: https://github.com/jkrijthe/Rtsne
5. **Apcluster**: Ulrich Bodenhofer, Andreas Kothmeier, and Sepp Hochreiter (2011) APCluster: an R package for affinity propagation clustering Bioinformatics 27:2463-2464. DOI: 10.1093/bioinformatics/btr406. Brendan J. Frey and Delbert Dueck (2007). Clustering by passing messages between data points Science 315:972-977. DOI: 10.1126/science.1136800.

Features

- Capabilities of Expression Console 1.4 and TAC 3.1 combined in a single easy to use application.
- QC an individual or collection of sample files using tables, graphs and box plots both before and after analysis.
- Use the Comparison Wizard to easily set up comparisons within your experiment.
- Ability to run Gene or Alternative Splicing Analysis.
- Ability to import Sample Attribute from ARR files or a tab-delimited TXT file.
- Use Principal Component Analysis based on signal intensity to compare samples by Sample Attribute.
- Use Venn Diagrams based on Gene Lists to obtain common genes based on multiple analyses.
- Analyze up to 1000 sample files in a single run.

Recommended hardware

Operating System	CPU	Memory (RAM)	Hard Drive Space	Browser
Microsoft Windows 10 (64 bit) Professional	Intel Pentium 4X 2.83 GHz (Quad Core processor)	16 GB	150 GB HD + Data storage	Internet Explorer 11 (or greater) and Microsoft Edge

Minimum requirements

Operating System	CPU	Memory (RAM)	Hard Drive Space	Browser
Microsoft Windows 7 and10 (64 bit) Professional	Intel Pentium 4X 2.83 GHz (Quad Core processor)	8 GB	150 GB HD + Data storage	Internet Explorer 11 (or greater) and Microsoft Edge

IMPORTANT! Larger data file sizes associated with Whole Transcriptome arrays should be taken into account when calculating the necessary available disk space requirement.

IMPORTANT! TAC 4.0.2 is not an upgrade, but can co-exist on workstations with previously installed versions of TAC.

1. Go to: thermofisher.com
2. Locate and download the TAC 4.0.2 software package.
3. Unzip the file, then double-click **TAC4.exe** to install the software.

4. Follow the directions provided by the installer.
5. After the installation is complete, double-click on the TAC 4.0.2 Desktop shortcut icon or click **Start** >**All Programs** > **Thermo Fisher Scientific** > **Transcriptome Analysis Console**

   The application opens.

Note: If you are prompted to assign your Library File path (immediately after the application opens), go to "Assigning a library folder/path" on page 17.

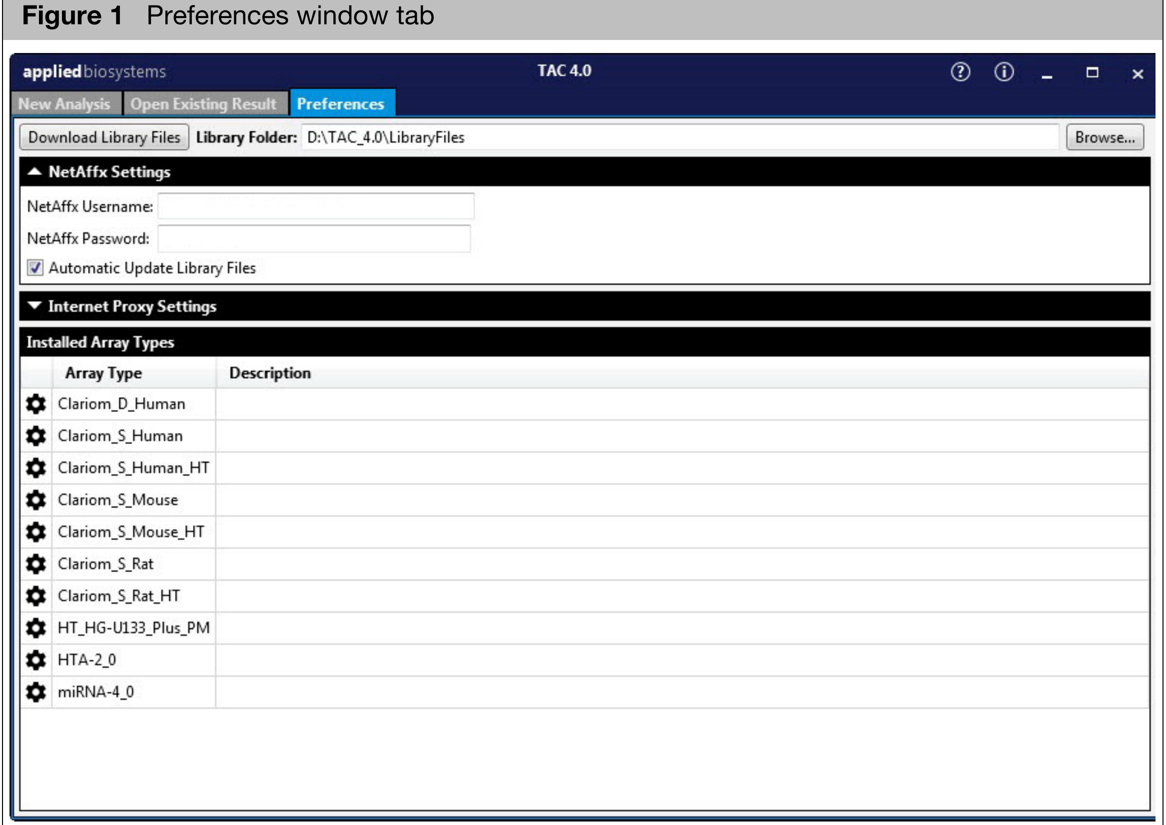
Setting up

Before using TAC, you must install the required library and annotation files, assign a library folder path, and set your Proxy server and algorithm analysis values.

Note: When you install TAC for the very first time, you are prompted to assign a library file folder. Acknowledge the message, then either create a new folder or select an existing folder. After you have designated a library folder, continue to step 1.

1. Click on the **Preferences** tab. (Figure 1).

Figure 1 Preferences window tab



Assigning a library folder/path

1. Click **Browse...** (far right of the Library path field).

   The **Select Library Folder** window appears.
2. Navigate to the folder you want to use or click **New Folder** to designate a new library folder.
3. Click **Select Folder**.

Downloading library files

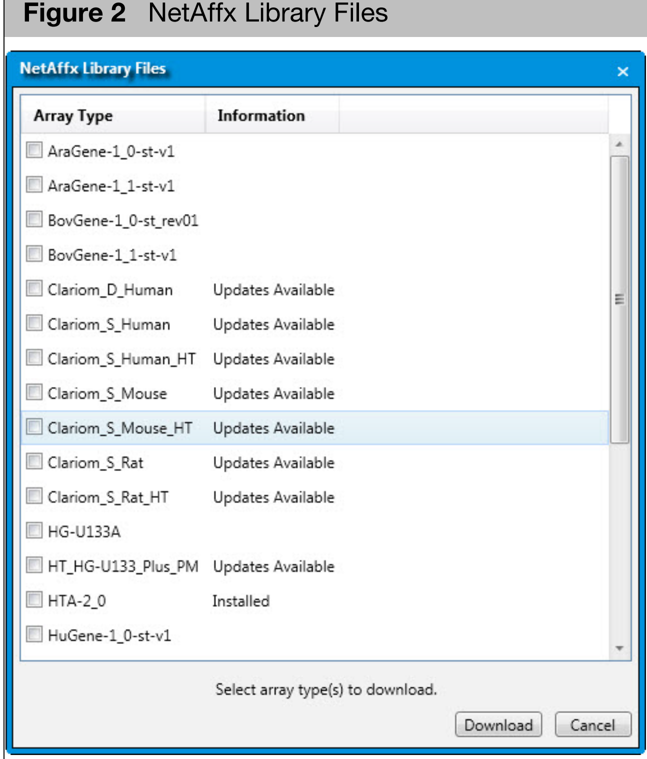
1. Click Download Library Files.

   The NetAffx Account Login window appears.
2. Enter your NetAffx account email and password, then click **OK** or click the **Register Now** link to create a new account.

Note: Your NetAffx username and password credentials are auto-saved, therefore no log in is required for future Library File downloads and/or updates.

The NetAffx Library Files window appears. (Figure 2)

Figure 2 NetAffx Library Files



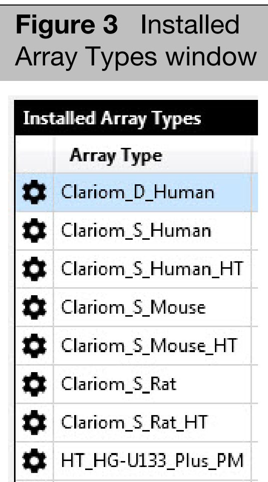
3. Click the check box next to the library file(s) you want to download, then click **Download**.

   Downloading progress bars appears to show the download status of each selected file. When an update to an installed library/annotation file becomes available, you will be prompted to download it.

Installed array types

The Installed Array Types window pane (Figure 3) displays the array types that reside in your library folder.

Figure 3 Installed Array Types window



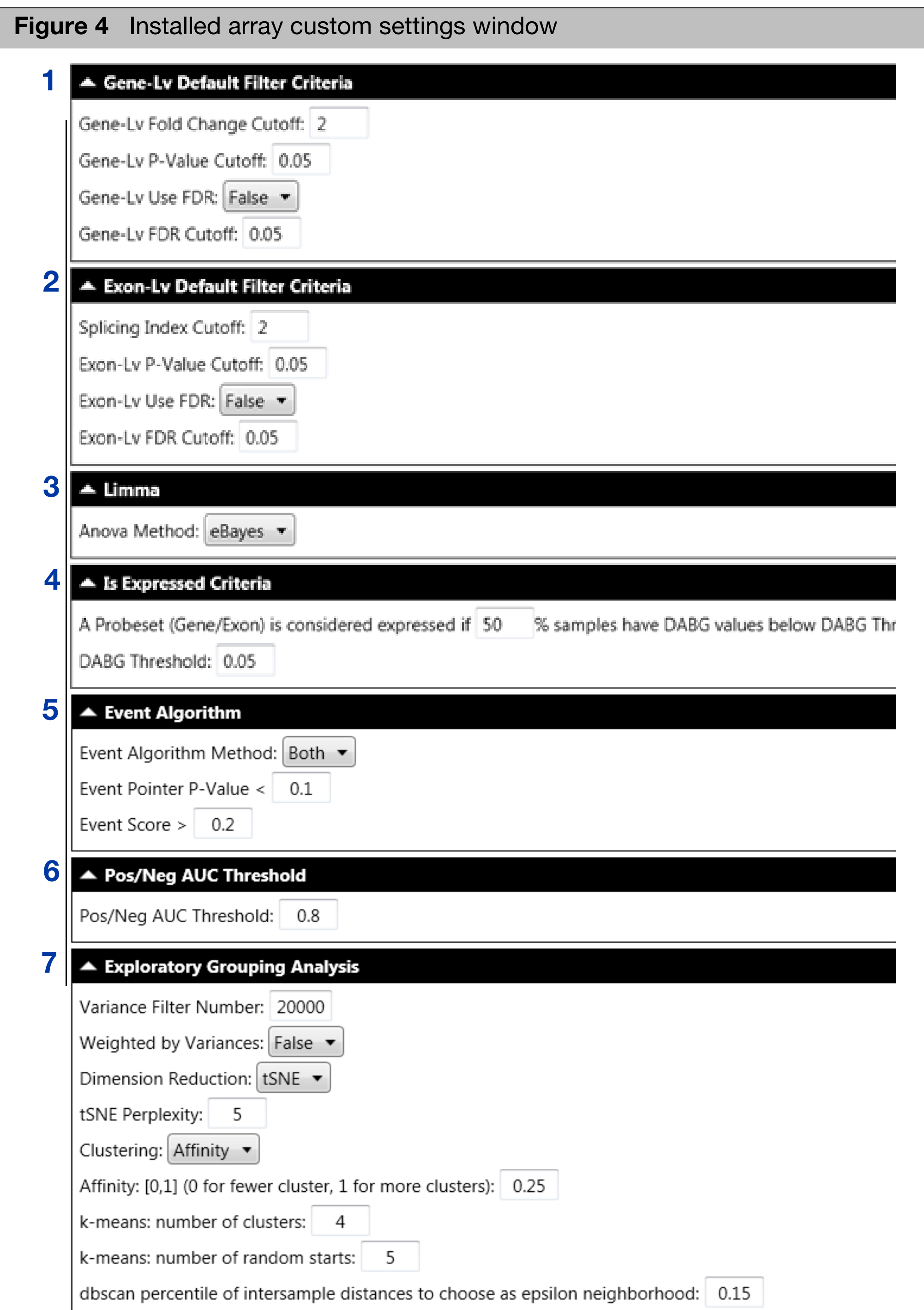
Customizing an Installed Array's Algorithm Settings

1. Click on the Array's adjacent button.

   The Settings window appears. (Figure 4)

   Note: The content displayed in the Custom Settings window varies with each selected array type.

Figure 4 Installed array custom settings window



See the Table 1 on page 20 for a description of each custom setting (1-7).

Table 1 Custom Setting Descriptions (1-7)

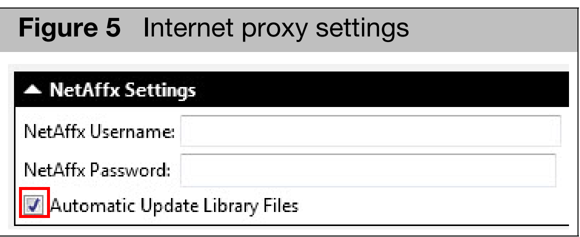
1 Gene-Level Default Filter Criteria	Results displayed in the Gene Table will be filtered based on these criteria. In the above example, only probesets with a fold change > 2 or with a p-value < 0.05. The FDR pvalue has been set to false so probesets will not be filtered by this p-value in this case.
2 Exon-Level Default Filter Criteria	Results displayed in the Alt Splice Table will be filtered based on these criteria. In the above example, only probesets with a fold change > 2 or < -2 and with a p-value < 0.05. The FDR p-value has been set to false so probesets will not be filtered by this p-value in this case.
3 Limma	Choose the ANOVA method (eBayes or ANOVA). For details see "Anova" and "eBayes" on page 201.
4 Is Expressed Criteria	Defines if a probeset is considered expressed. In the Figure 4 example above, a probeset is considered to be expressed if 50% of the samples in the dataset have DABG (Detected Above Background) values below the DABG threshold. The default DABG is set to 0.05.
5 Event Pointer	Use the Event Pointer drop-down menu to select the Event Pointer or Event Estimation or combination of both splicing event algorithms. • Set the Event Pointer p-value threshold. • Set the Event Score threshold. For more information, see "Event score" on page 210.
6 Pos/Neg AUC Threshold	Recent Whole Transcriptome Arrays contain a set of control probesets for normal genes that were empirically selected as housekeeping genes across many different samples. "Pos" controls are exon-level probesets and "Neg" controls are the intron probesets for the same genes. By looking at the separation of these two distributions, you can generate a pos/neg metric from which a QC threshold can be applied. The metric is the area under the curve (AUC) for a receiver operator characteristic curve (ROC) describing the separation of these two distributions. An AUC of 1 reflects perfect separation whereas as an AUC value of 0.5 reflects no separation. Typically customers have used a threshold of 0.7 to flag less desirable results. However, this threshold is arbitrary and is only used as a guideline. More difficult samples will naturally have poorer AUCs and can often be tolerated and used for downstream results.
7 Exploratory Group Analysis	Exploratory Grouping Analysis (EGA) is a tool for analyzing the relationships among a group of samples. Such an analysis can prove helpful in identifying potential groups in the sample set for later differential expression analysis. Moreover, EGA can highlight possible batch effects that can be removed in later analysis. EGA provides a set of tools so that samples can be examined in multiple ways. You can use the highest variance probesets, or select a set of probesets associated with a gene list of interest. Next, the large number of probeset signals per sample is reduced to three for viewing, either linearly using Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE). Finally, a clustering analysis applies a label to each sample indicating a grouping. The clustering analysis mechanisms include k-means, DBSCAN, and affinity propagation clustering. See "Exploratory Grouping Analysis (EGA) module" on page 193 for parameter setting details.

2. Use the provided drop-down menus and text fields to customize the Array's algorithm settings.
3. Click **Save**.
4. Optional: Click **Default** to return all values to their factory defaults. Click **Cancel** to return to the Preferences window tab.

Activating automatic library file update checks

Enter your NetAffx Username and Password, then click to check the **Automatic Update Library Files** check box (Recommended), as shown in Figure 5.

Figure 5 Internet proxy settings

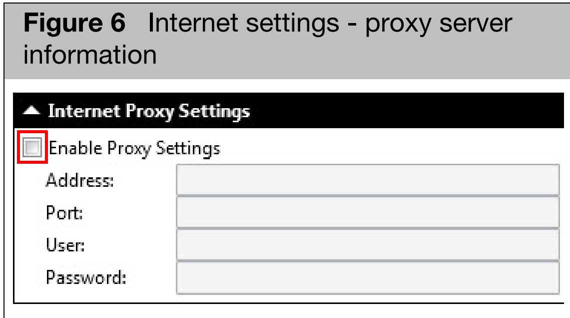


Setting up custom proxy settings

Follow the steps below if your system has to pass through a Proxy Server before it can access the NetAffx server.

1. Click the **Enable Proxy Settings** check box (Figure 6), then complete the required fields.

Figure 6 Internet settings - proxy server information



Note: The proxy user ID and password is NOT the same ID and password used to connect to the NetAffx server.

2. Enter the **Address**, **Port**, **User**, and **Password**. If you do not know what the proxy settings are, contact your IT department.

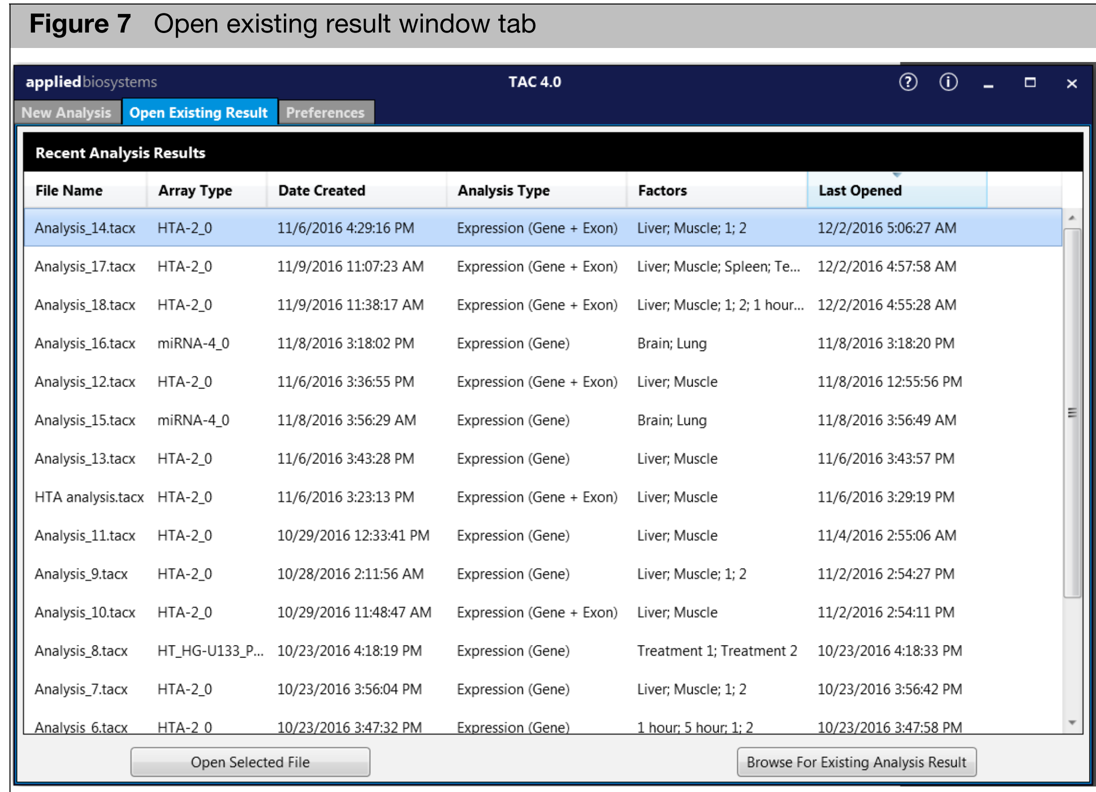
Open Existing Result window tab

TAC auto-saves your studies enabling you to view past analysis results as needed.

Note: Previous analysis results generated in earlier versions of TAC are not supported in TAC 4.0.2 and higher. Also, analysis results generated in TAC 4.0.2 are not backward compatible with earlier versions of TAC.

1. Click on the **Open Existing Result** tab. (Figure 7)

Figure 7 Open existing result window tab



Note: Click any of the window's header columns to sort your recent studies in either ascending (A-Z) or descending (Z-A) order.

2. Double-click on a recent study or single-click on it, then click Open Selected File.

   After a few moments, your recent analysis opens (in the same state - you last left it).

Browsing for existing analysis result

1. TAC stores up to 20 recent results. If you still cannot locate a past analysis result, click Browse For Existing Analysis Result.

   An Analysis Results window opens.

2. Locate your analysis, click to highlight it, then click **Open**.

   After a few moments, your analysis opens as you last left it.

2

Analysis setup

Setting up a new analysis

An analysis can be initiated using either CEL or CHP files.

- To start an analysis from CEL files, see "Importing CEL files".
- To start an analysis from CHP files, see "Importing CHP files" on page 24.

Note: Starting from CHP files limits the selection of available analysis options, as described in Chapter 3, "Algorithm and summarization selection" on page 36.

Importing CEL files

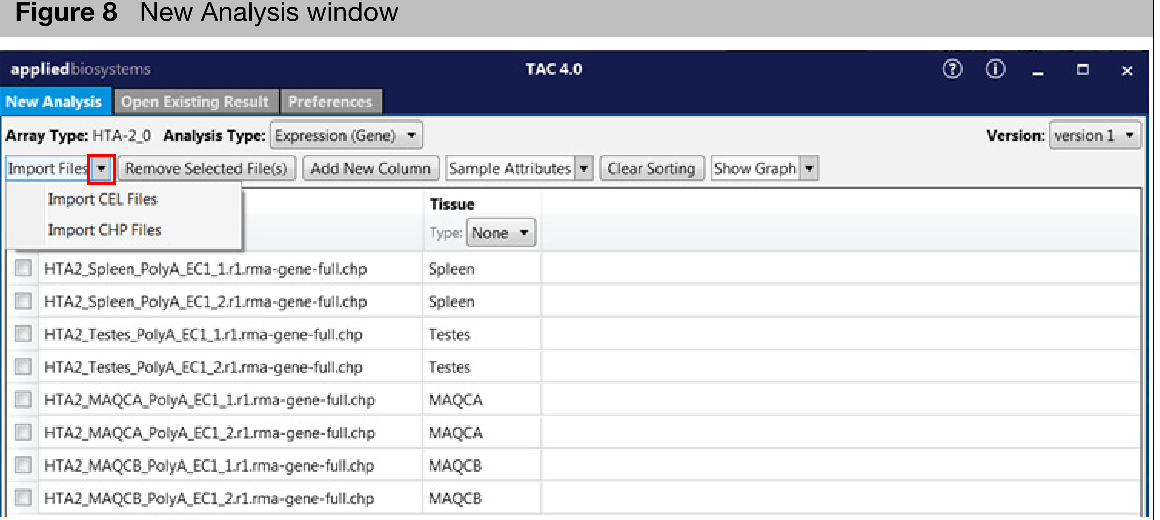
1. At the **New Analysis** window tab, click **Import Files** > **Import CEL Files**.

   The Import CEL Files window appears.
2. Single click, Ctrl click, or Shift click to select multiple files.
3. Click **Open**.

   If the required library files associated with your selected CEL files are not found, a message prompting you to download them appears. Click **OK** to acknowledge the message, then go to "Downloading library files" on page 17.

   After your selected CEL files have been successfully loaded into the Analysis Window (Figure 8), the Array Type is auto-detected and displayed (upper left).

Figure 8 New Analysis window



Importing CHP files

1. At the **New Analysis** window tab, click **Import Files** > **Import CHP Files**.

   The Import CHP Files window appears.
2. Single click, Ctrl click, or Shift click to select multiple files.
3. Click **Open**.

   If the required library files associated with your selected CHP files are not found, a message prompting you to download them appears. Click **OK** to acknowledge the message, then go to "Downloading library files" on page 17.

   After your selected CHP files have been successfully loaded into the Analysis Window (Figure 8), the Array Type is auto-detected and displayed (upper left).

   Note: Not all CHP files from previous versions of TAC can be loaded into a New Analysis. For example: **rma-alt-splice-dabg.chp**. There are three types of CHP files used in previous versions of TAC:

   – **.*exon*.chp**: Not supported in TAC 4.0.2
   – **.*alt-splice*.chp**: Not supported in TAC 4.0.2, but Alt-splice analysis can be run from the CEL files.
   – Any other **.chp** files are imported as gene level CHP files, therefore Gene level options are enabled.

Assigning custom sample attributes

A Sample Attribute column in TAC 4.0.2 is known as **Factor**. The contents within the Sample Attribute column are known as **Factor Values**.

You can assign custom sample attributes (also known as Factors) using any of the three following methods:

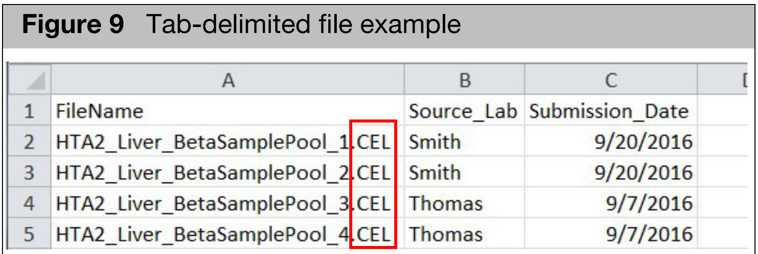
ARR Files

- Create or use a Template to assign attribute(s) to ARR files. Attributes can be assigned upon registering your samples in Affymetrix GeneChip Command Console (AGCC). Refer to the AGCC User Manual (P/N 702569) for details on templates and assigning attributes.
- When importing CEL files, place the ARR files from AGCC into the same folder as their corresponding CEL files to automatically display the attributes you designated in the ARR file(s).

Tab-delimited text files

- Create a tab-delimited TXT file containing Sample Attributes for your CEL files to be analyzed. **Column A** must contain the full filename including the ".CEL", as shown in Figure 9.

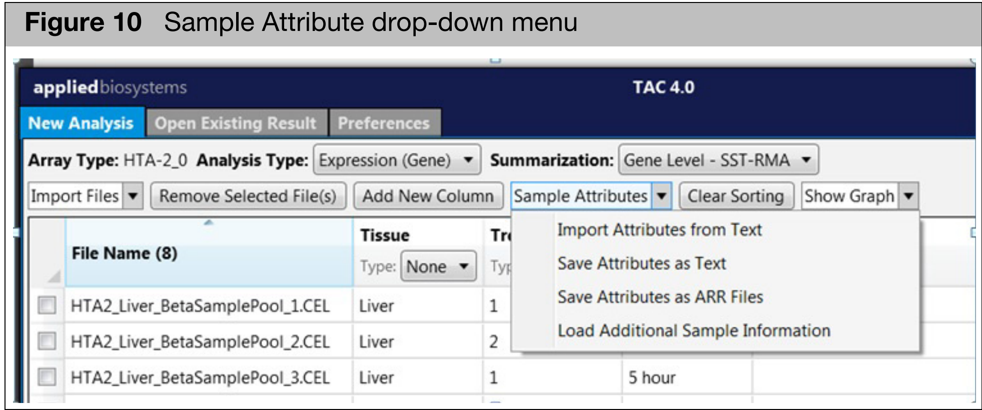
Figure 9 Tab-delimited file example



	A	B	C
1	FileName	Source_Lab	Submission_Date
2	HTA2_Liver_BetaSamplePool_1.CEL	Smith	9/20/2016
3	HTA2_Liver_BetaSamplePool_2.CEL	Smith	9/20/2016
4	HTA2_Liver_BetaSamplePool_3.CEL	Thomas	9/7/2016
5	HTA2_Liver_BetaSamplePool_4.CEL	Thomas	9/7/2016

1. Click the **Sample Attribute** drop-down (Figure 10) to select **Import Attribute from Text**.

Figure 10 Sample Attribute drop-down menu

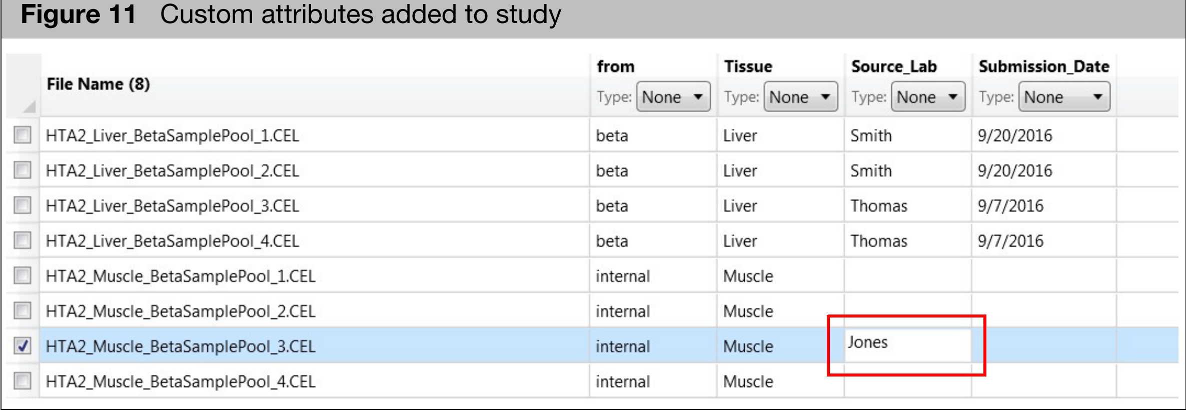


A **Select Attributes Text File** window appears.

2. Navigate to your previously saved attribute file, click to highlight it, then click **Open**.

   Your custom attributes are now added. (Figure 11).

Figure 11 Custom attributes added to study



File Name (8)	from	Tissue	Source_Lab	Submission_Date
	Type: None	Type: None	Type: None	Type: None
HTA2_Liver_BetaSamplePool_1.CEL	beta	Liver	Smith	9/20/2016
HTA2_Liver_BetaSamplePool_2.CEL	beta	Liver	Smith	9/20/2016
HTA2_Liver_BetaSamplePool_3.CEL	beta	Liver	Thomas	9/7/2016
HTA2_Liver_BetaSamplePool_4.CEL	beta	Liver	Thomas	9/7/2016
HTA2_Muscle_BetaSamplePool_1.CEL	internal	Muscle		
HTA2_Muscle_BetaSamplePool_2.CEL	internal	Muscle		
HTA2_Muscle_BetaSamplePool_3.CEL	internal	Muscle	Jones	
HTA2_Muscle_BetaSamplePool_4.CEL	internal	Muscle		

Direct entry

1. Click Add New Column.

   An **Add New Column** window appears.

2. Enter a name, then click **OK**.
3. Click inside the cell(s) under your newly created column.

   A cursor appears.

4. Enter the attribute name you want to add, as shown in Figure 11.
5. Press **Enter** or click outside the cell to update your field entry.

   Note: You must click **Run Analysis** to save any attributes that have been added. Sample Attributes will be lost if the New Analysis window has been closed prior to running the analysis.

Editing a sample attribute

Any sample attribute can be edited.

1. Click in the cell of the attribute to edit.
2. Update the information in the cell.
3. Click out of the cell to view the updated information.

   Note: You must click **Run Analysis** to save any attributes that have been added. Sample Attributes will be lost if the New Analysis window has been closed prior to running the analysis.

Importing array physical attributes

1. Click the **Sample Attributes** drop-down (Figure 10), then select **Load Additional Sample Information**.

   The following columns may be populated (depending on array type) with information obtained from the array:

   – Scanner Type
   – Scanner ID
   – Scan Date
   – File Creator
   – Scanner Serial Number
   – Plate Barcode
   – Fluidics Serial Number
   – Scanner Operator

Note: If the ARR files are located in the same directory as the CEL file, any sample attributes captured in the ARR file are loaded into the table.

Exporting sample attributes

To save the New Analysis Table as a text file:

1. Click the **Sample Attribute** drop-down (Figure 10 on page 25), then select **Save Attributes as text**.
2. Select the location and filename to save the export sample attributes, then click **Save**.

To save the New Analysis Table as ARR files:

1. Click the **Sample Attribute** drop-down (Figure 10 on page 25), then select **Save Attributes as ARR files**.

   The ARR files (for each of the samples) in the New Analysis Table are saved and placed into your CEL files folder (accessed in Step 1 on page 23).

Sorting columns

1. Click any column header to sort it.

   A tiny arrow graphic appears on the header to indicate an ascending or descending sorted order.

   Click Clear Sorting to restore a column back to its original sorted state.

Showing QC graphs

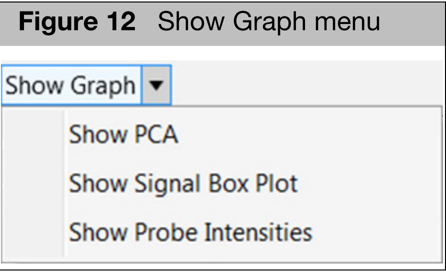
At anytime you can review the QC metrics of your loaded CEL files BEFORE running an analysis. QC Graphs interact with the table in the New Analysis Tab. Any sample(s) highlighted in a graph are also highlighted in the table.

If any of the sample attributes in the table are updated through directly editing the information in a CEL or editing the corresponding ARR or TXT file, you must update the graph. To do this, click Refresh.

Before running your analysis, you may want to review your listed samples in a visual format. **Note:** Each graph is generated prior to analysis/summarization and have not been normalized, therefore some differences should be expected.

1. Click the **Show Graph** drop-down arrow (Figure 12), then click on the graph you want to view.

Figure 12 Show Graph menu

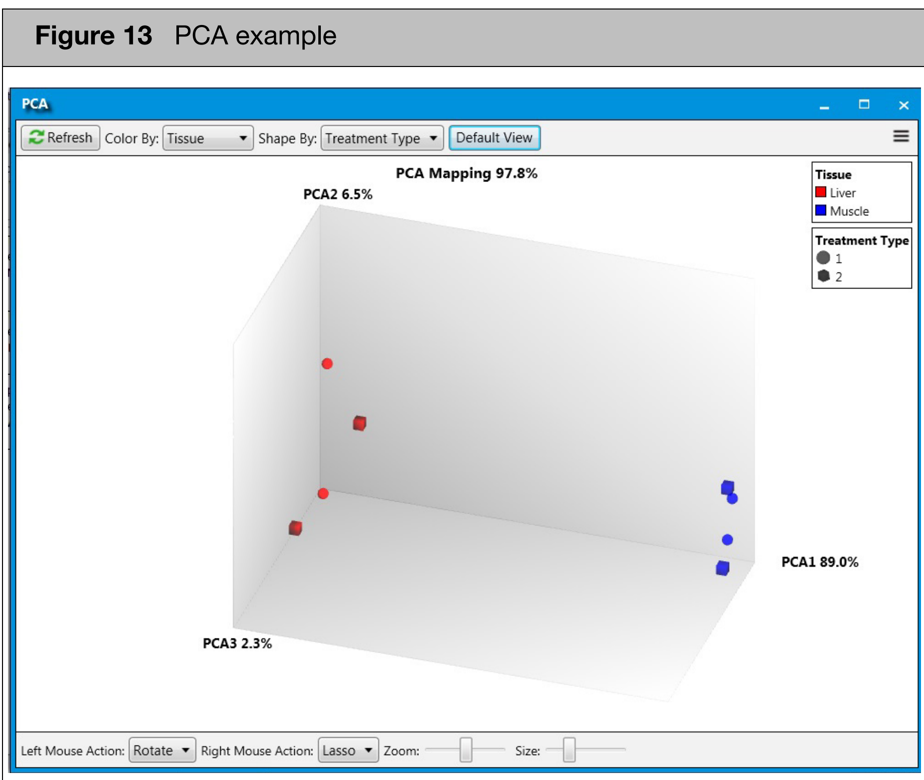


Principal Component Analysis (PCA) graph

1. Click **Show PCA**.

   The PCA window appears. (Figure 13)

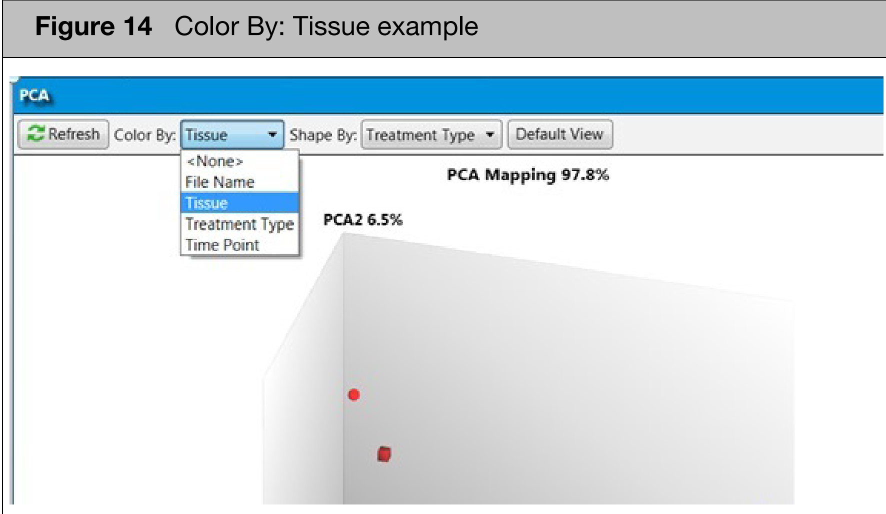
Figure 13 PCA example



Using the PCA graph window

1. Click on the **Color By** drop-down menu to select the view you want, as shown in Figure 14.

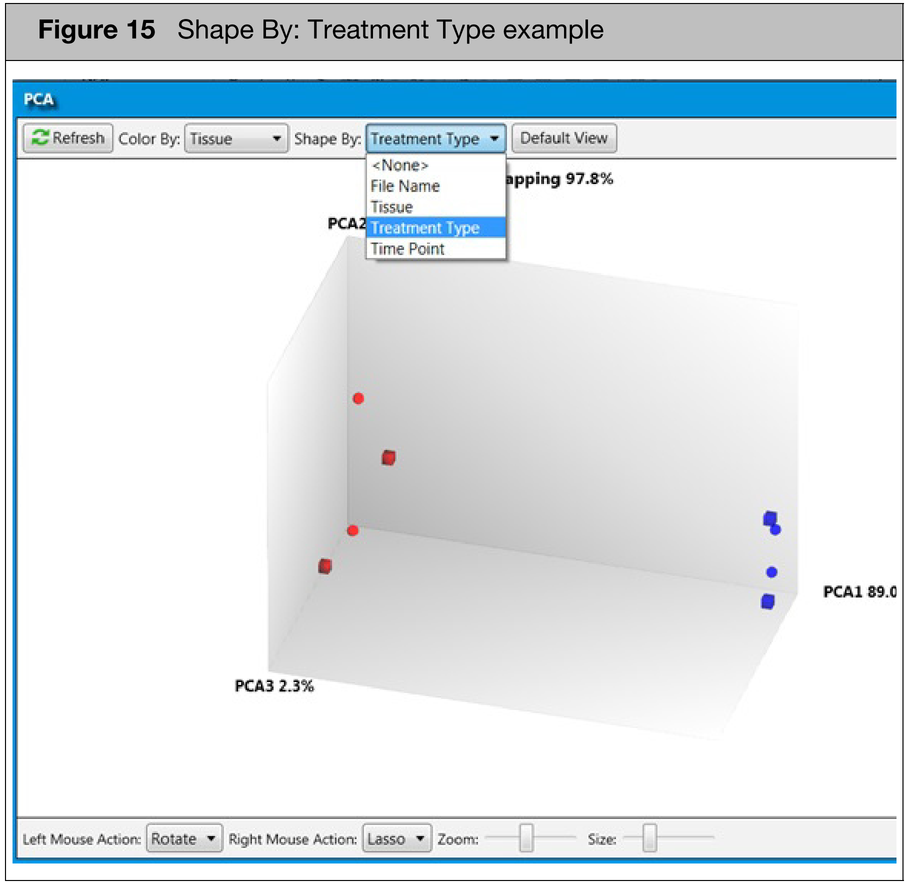
Figure 14 Color By: Tissue example



Note: The available Color By and Shape By drop-down entries are generated from the attributes in the New Analysis Table.

2. Click on the **Shape By** drop-down menu to select the view you want, as shown in Figure 15.

Figure 15 Shape By: Treatment Type example



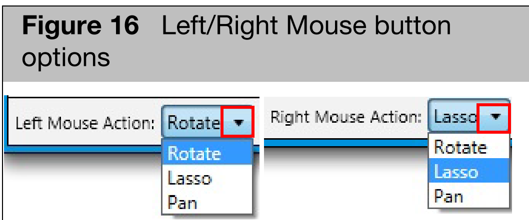
Interacting with the PCA graph

Use your mouse's buttons to rotate or pan the graph and lasso (isolate) genes of interest. The slider bars enable you to zoom in and out or change the scale of the displayed sample shape(s).

Assigning mouse button functions

1. Click on either Mouse Action drop-down (Figure 16), to assign a specific mouse button function.

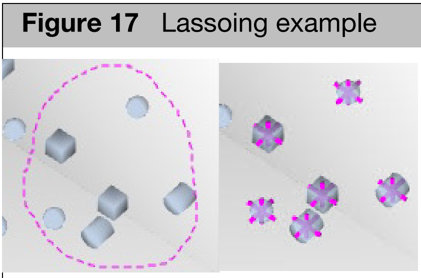
Figure 16 Left/Right Mouse button options



- **Rotate** - Place your mouse cursor on the graph, hold down the assigned button, then move the mouse to rotate the graph on its fixed axis.

  Lasso - Hold down the assigned Lasso mouse button, then use the cross-hair cursor to encircle the gene(s) of interest. Release the mouse button to capture. (Figure 17) Right-click on your lassoed selection to deselect it.

Figure 17 Lassoing example

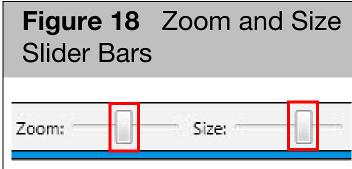


- **Pan**- Place your mouse cursor on the graph, hold down the assigned Pan button, then move the mouse to re-position the graph.

Zooming and resizing

Use the slider bars (Figure 18) to zoom in and out and/or change the scale of the displayed sample shape(s).

Figure 18 Zoom and Size Slider Bars



- Click Default View to return the graph to its default view.
- Click Refresh to refresh attribute information that was edited.

PCA graph options

1. Click the Options button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

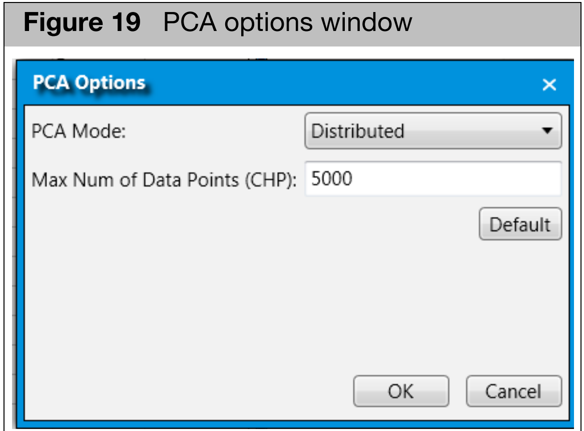
Prints the currently displayed graph.

PCA Options

1. Click **PCA Options**. (Figure 19)

   The PCA Options window appears. (Figure 19)

Figure 19 PCA options window



2. From the **PCA Mode** drop-down menu select either:

- **Distributed** - TAC 4.0.2 auto-selects N evenly distributed probes/probesets from CEL file when computing PCA.
  - **Max Num of Data Points (CEL)** - Use the provided text field to enter the maximum number of data points when using CEL files.
- **Variance** - TAC 4.0.2 auto-selects N probes/probesets that contain the highest variance from CEL file when computing PCA.
  - **Increase Sensitivity** - Click this check box if you want to:
    - Weight probes/probsets by variance.
    - Multiply probes/probesets signals by their variance to increase sensitivity.
    - Reduce background noise when smaller number of probes/probesets are present.
  - **Max Num of Data Points (CEL)** - Use the provided text field to enter the maximum number of data points when using CEL files.

Export Coordinates

Exports coordinates as a TXT file.

Show Legend

Click the **Show Legend** check box to show the Graph's Legend. Uncheck to hide it.

Clear Selection

Click **Clear Selection** to clear a lassoed or highlighted selection/area.

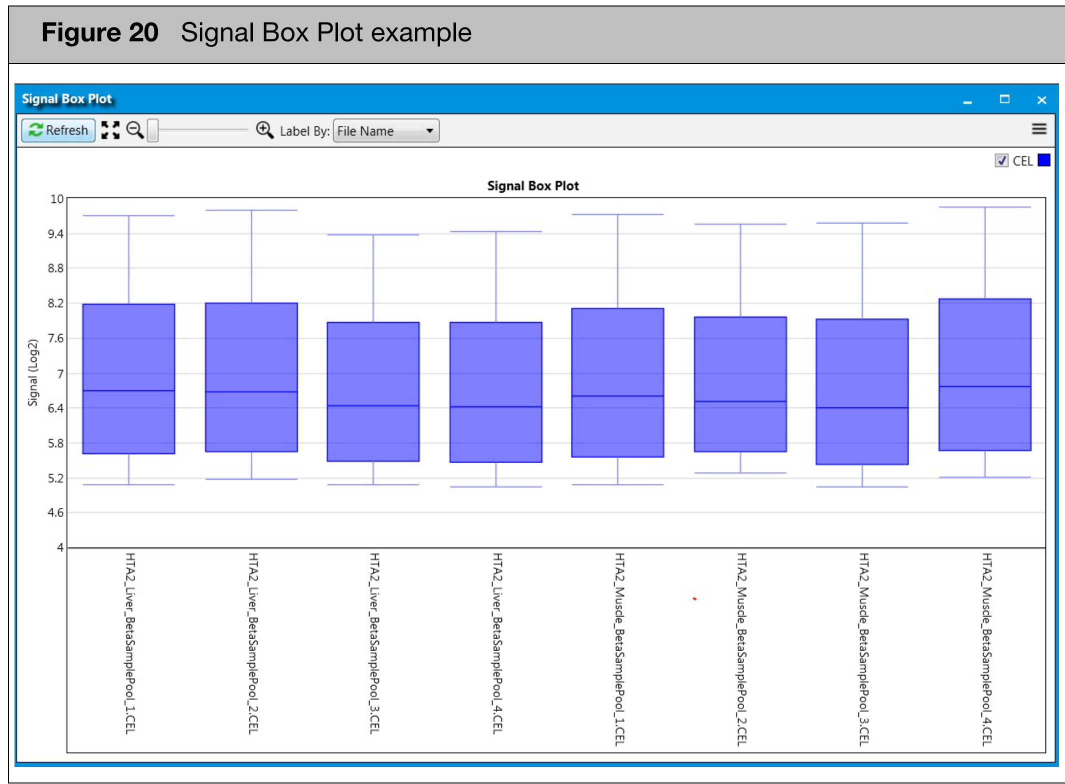
Signal Box Plot

The Signal Box Plot is based on your raw data BEFORE normalization.

1. Click **Show Signal Box Plot**.

   The Signal Box Plot window appears. (Figure 20)

Figure 20 Signal Box Plot example



Note: Click on any signal box plot to see its summary statistics.

Signal Plot graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

Prints the currently displayed graph.

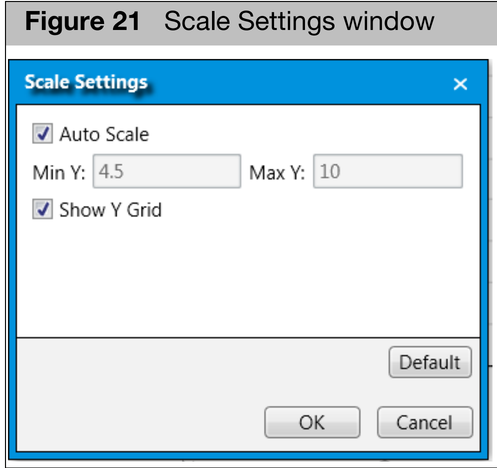
Scale Settings

1. Click **Scale Settings**.

   The Scale Settings window appears.

2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box (Figure 21) to enable both the X and Y Min and Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display X/Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Figure 21 Scale Settings window



Clear Selections

1. Click **Clear Selection**.

   All previously selected items are now cleared form the graphs.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

- Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Viewing and Labeling

- Click on the slider bar. Hold down the left mouse button, then move the bar right to magnify point(s) of interest. Click to return to the default display view.
- Click the **Label By** drop-down to select a different Label By display.
- Mouse over (position the cursor over) a Signal Bar to reveal its values.
- Un-check the (upper right) check box CEL to remove the box plots from the window.

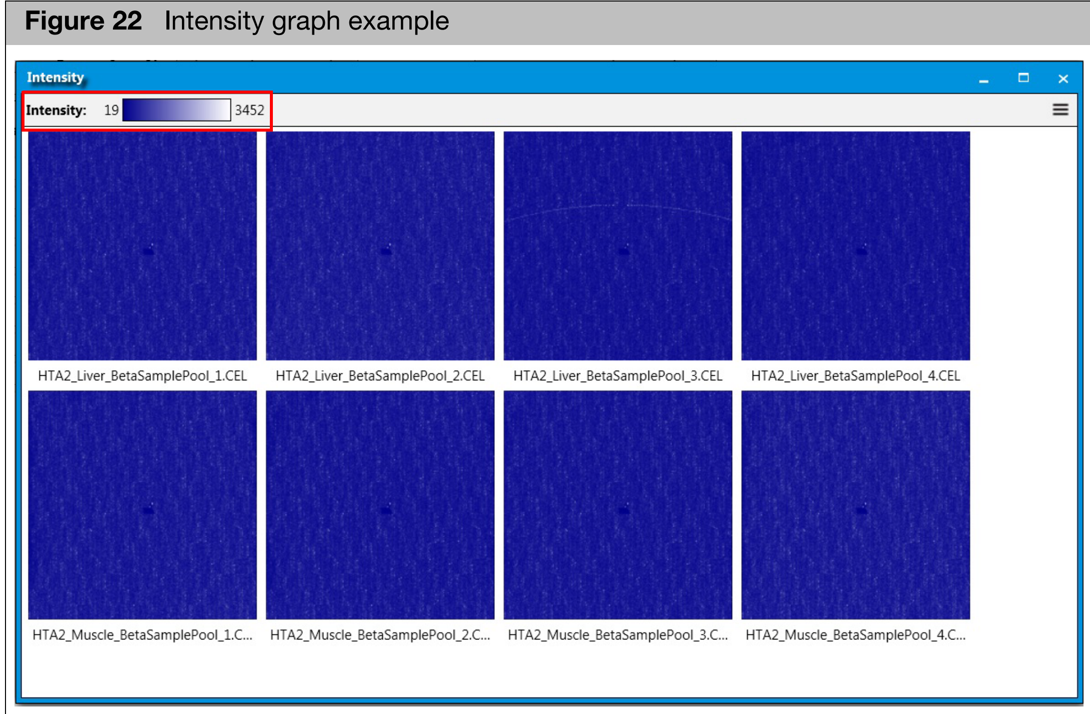
Probe Intensities

The Show Probe Cell Intensity option produces a thumbnail image of the probe intensity values for each array. **Note:** Thumbnail images are not available when starting an analysis with CHP files.

1. Click **Show Probe Intensities**.

   The Intensity window appears. **Note:** The Intensity Scale Key (Figure 22) provides the signal values based on color in the Intensity graph.

Figure 22 Intensity graph example



Probe Intensities graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

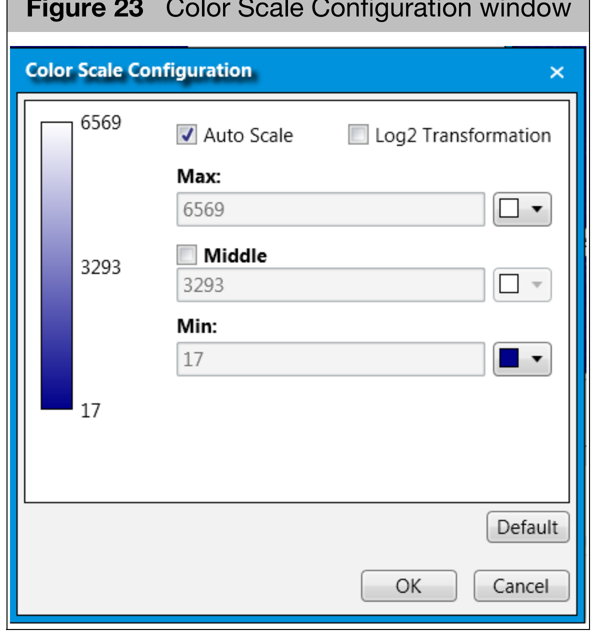
Prints the currently displayed graph.

Color Settings

1. Click the **Color Settings** button.

The Color Scale Configuration window appears. (Figure 23)

Figure 23 Color Scale Configuration window



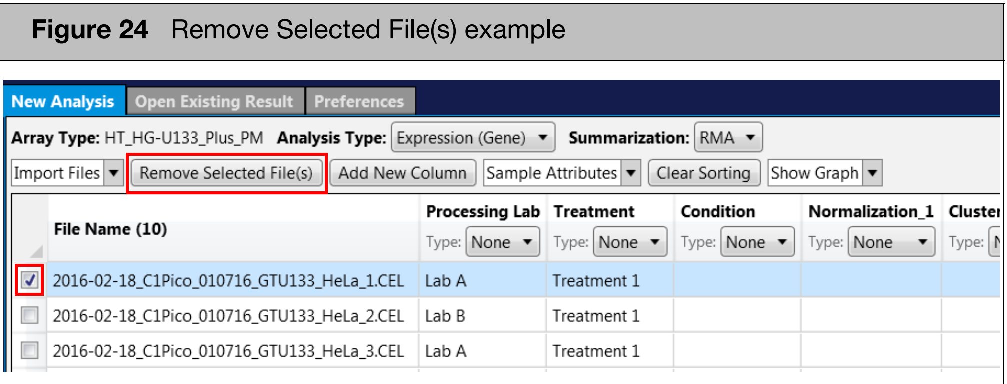
2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation. **Note:** Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.
3. Use the color drop-down menus to assign new colors for your data points.

Click **OK** to save your changes or click **Default** to return to the factory settings.

Removing listed files before analysis

1. Click the check box adjacent to the CEL file(s) you want to remove from the list, then click **Remove Selected File(s)**. (Figure 24)

Figure 24 Remove Selected File(s) example



Algorithm and summarization selection

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 "Recommended hardware" on page 15.

Supported arrays

IMPORTANT! Before you can perform an analysis, you must have the required library files downloaded and installed. See "Downloading library files" on page 17 for more information.

Gene Level + Alt-Splice analysis

TAC 4.0.2 provides either Gene Level only or Gene Level plus Alternative Splicing Analysis depending on the array type. The Gene ST arrays and Clariom S provide Gene Level analysis, while the Transcriptome Arrays (Human (HTA), Mouse (MTA), Rat (RTA), and Clariom D provide both Gene Level and Alt Splice analyses. For details on these algorithms, see Appendix A, "Algorithms" on page 193

Note: The Limma Bioconductor package (implemented in TAC 4.0.2), provides analysis for expression data based on linear models. Limma has been the most commonly used method in the Bioconductor ecosystem for the analysis of differential expression measures when using microarrays. Limma provides two important capabilities. First, it enables the analysis of complex experimental designs, including combinations of the following: comparisons across multiple attributes, interactions, blocking or batch factors, random factors, and real covariates. Second, it adds an empirical Bayesian correction that improves the estimate of the standard deviation by using information gleaned from a multi-probeset analysis. This empirical Bayesian approach is especially useful when analyzing smaller numbers of samples whose estimated standard deviation is less accurate without the correction.

Selecting an analysis type to run

The Array Type is auto-detected, as it is based on your imported CEL files. The Analysis Type drop-down menu selections are generated based on the array type.

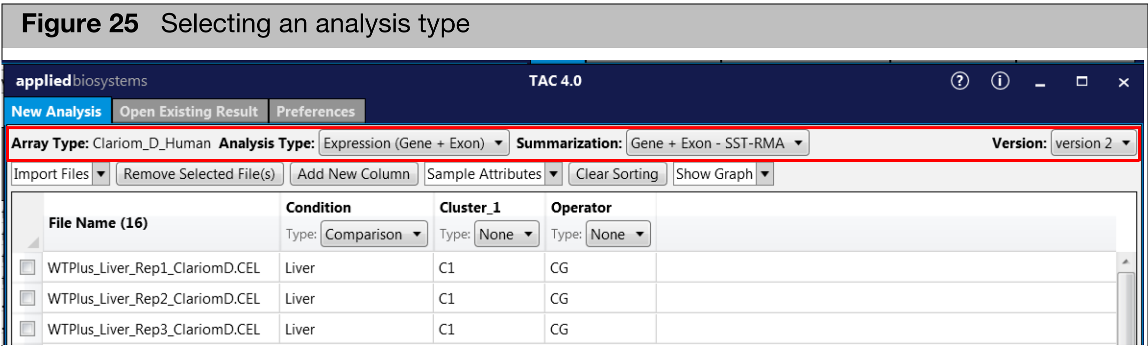
1. Click the **Analysis Type** drop-down menu (Figure 25) to select the analysis you want to run.
   - **Expression (Gene + Exon)** - Provides both Gene Level and Alt Splice analysis for supported array types.
   - **Exploratory Grouping Analysis (EGA) (Gene + Exon)** This workflow involves a two step process. The first step generates clusters from the data followed by expression analysis to generate fold change and p-values. This analysis option enables you to choose to run with or without a specific Gene list. For more information, see page 193.

- **Summarization Only** - Performs normalization and summarization in the data. This option summarizes the probes into a single signal for each probeset. This process converts CEL files to CHP files without performing a differential expression analysis.

Note: Algorithm Settings can be customized by selecting the Algorithm Settings button at the bottom of the window. See "Customizing an Installed Array's Algorithm Settings" on page 18.

2. Click the **Summarization** drop-down menu to select the summarization you want to use.
   - **Gene - SST RMA or Gene - RMA** is used for generating Gene Level analysis
   - **Gene + Exon - SST-RMA or Gene + Exon - RMA** is used for generating both Gene Level and Alt Splice analysis.
3. Click the **Version** drop-down menu to select a different version of library and annotation files (if available).

Figure 25 Selecting an analysis type



After selecting your Algorithm and Summarization preferences, go to "Setting up a comparison analysis" on page 40.

Gene Level analysis

TAC supports both SST-RMA and RMA summarization methods (depends on array type selected) for Gene Level analysis. For details on these algorithms, see "Algorithms" on page 193.

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

Selecting an analysis type to run

The Array Type is auto-detected, as it is based on your imported CEL files. The Analysis Type drop-down menu selections are generated based on the array type.

1. Click the **Analysis Type** drop-down menu to select the analysis you want to run.
   - **Expression (Gene)** - Provides gene level analysis.
   - **Exploratory Grouping Analysis (EGA) (Gene)** - Provides Gene Level analysis from small starting material. This workflow involves a two step process. The first step generates clusters from the data followed by expression analysis to generate fold change and p-values. This analysis option enables you to choose to run with or without a specific Gene list. For more information, see page 193.
   - **Summarization Only** - Performs normalization and summarization in the data. This option summarizes the probes into a single signal for each probeset. This process converts CEL files to CHP files without performing a differential expression analysis.

2. Click the **Summarization** drop-down to select the summarization method.
3. Click the Version drop-down menu to select a different version of library and annotation files (if available).

   After selecting your Algorithm and Summarization preferences, go to "Setting up a comparison analysis" on page 40.

3'IVT Expression arrays

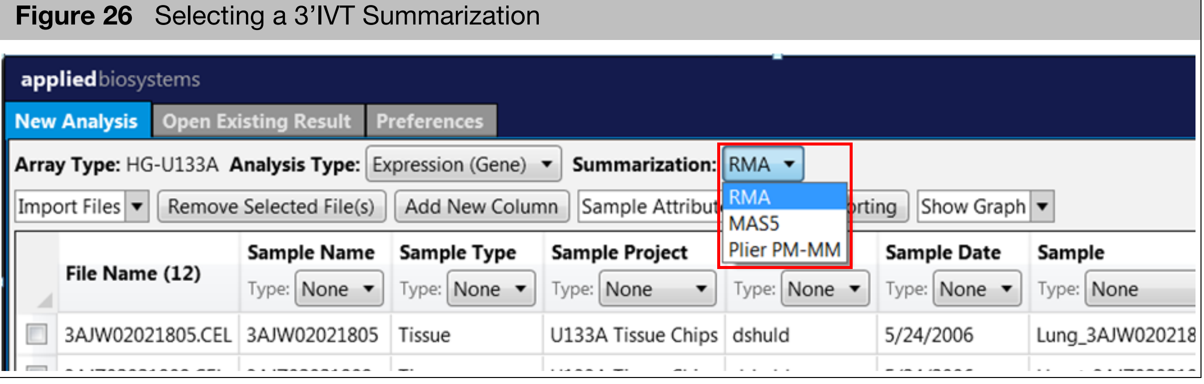
3' IVT Expression arrays can be analyzed using any of the following algorithms: RMA, MAS 5.0, PLIER. For details on these algorithms, see page 211.

Selecting an analysis type to run

The Array Type is auto-detected, as it is based on your imported CEL files. The Analysis Type drop-down menu selections are generated based on the array type.

1. Click the **Analysis Type** drop-down menu to select the analysis you want to run.
   – **Expression (Gene)** - Provides gene level analysis.
   – **Exploratory Grouping Analysis (EGA) (Gene)** - Provides Gene Level analysis from small starting material. This workflow involves a two step process. The first step generates clusters from the data followed by expression analysis to generate fold change and p-values. This analysis option enables you to choose to run with or without a specific Gene list. For more information, see page 193.
   – **Summarization Only** - Performs normalization and summarization in the data. This option summarizes the probes into a single signal for each probeset. This process converts CEL files to CHP files without performing a differential expression analysis.
2. Click the **Summarization** drop-down menu to select **RMA, MAS5,** or **Plier PM-MM**, as shown in Figure 26.

Figure 26 Selecting a 3'IVT Summarization



After selecting your Algorithm and Summarization preferences, go to "Setting up a comparison analysis" on page 40.

miRNA arrays

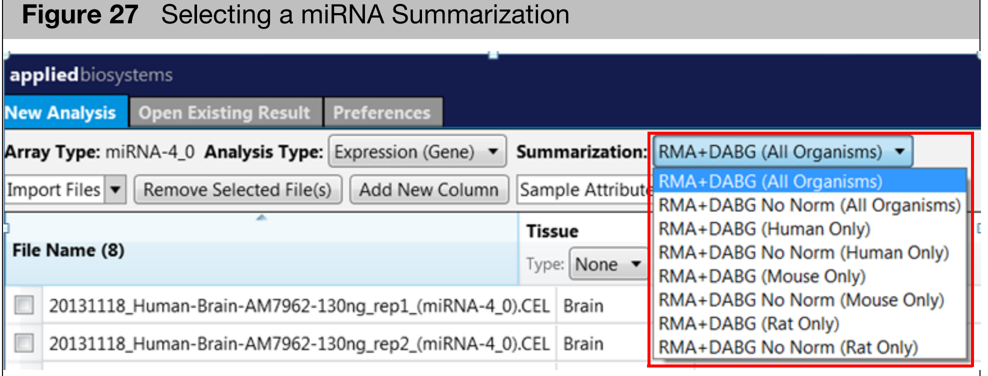
TAC uses the RMA+DABG algorithm for miRNA arrays. Follow the instructions below to analyze miRNA arrays. For details on these algorithms, see Appendix A, "Algorithms" on page 193.

Selecting an analysis type to run

The Array Type is auto-detected, as it is based on your imported CEL files. The Analysis Type drop-down menu selections are generated based on the array type.

1. Click the **Analysis Type** drop-down menu to select the analysis you want to run.
   - **Expression (Gene)** - Provides gene level analysis.
   - **Exploratory Grouping Analysis (EGA) (Gene)** - Provides Gene Level analysis from small starting material. This workflow involves a two step process. The first step generates clusters from the data followed by expression analysis to generate fold change and p-values. This analysis option enables you to choose to run with or without a specific Gene list. For more information, see page 193.
   - **Summarization Only** - Performs normalization and summarization in the data. This option summarizes the probes into a single signal for each probeset. This process converts CEL files to CHP files without performing a differential expression analysis.
2. Click the **Summarization** drop-down (Figure 27), then select the summarization you want to use in your analysis.

Figure 27 Selecting a miRNA Summarization



3. Click the Version drop-down menu to select a different version of library and annotation files (if available).

After selecting your Algorithm and Summarization preferences, go to "Setting up a comparison analysis" on page 40.

Setting up a comparison analysis

A comparison attribute indicates that the levels of this attribute will be compared during the differential expression analysis.

There are two ways to set up a comparison analysis:

- "New Analysis window tab"
- "Comparison Wizard" on page 46

New Analysis window tab

Setting up One-Way ANOVA comparisons

IMPORTANT! At least two factor values in an Attribute column must be imported in order to run a comparison analysis.

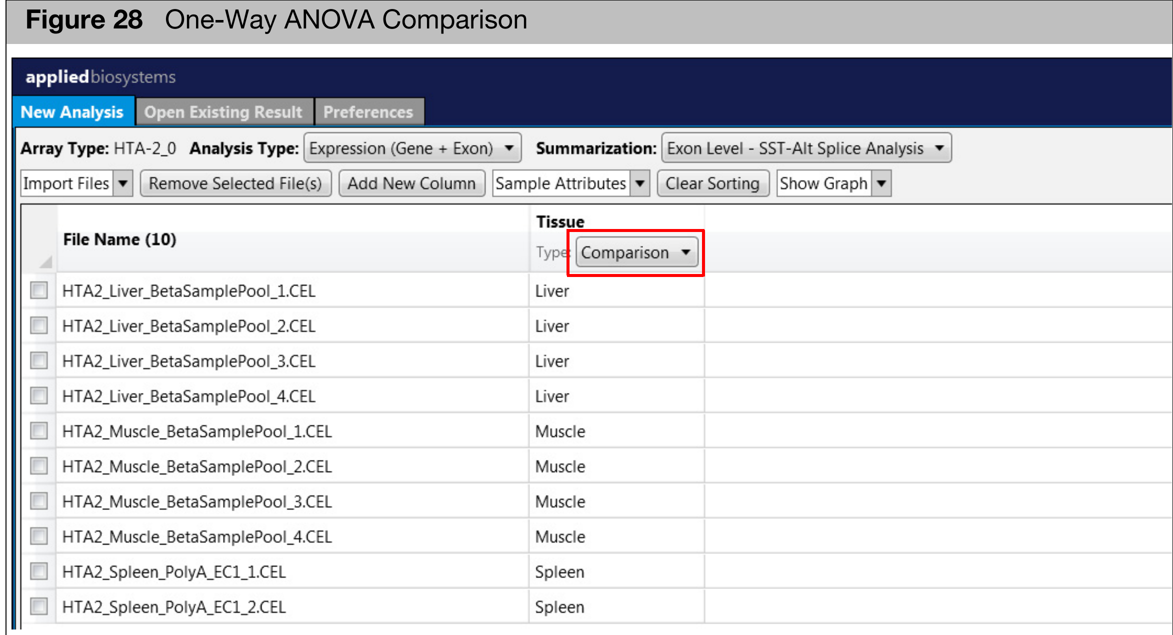
1. From the drop-down menu (in the header of the column you would like to compare), select **Comparison**.

   If there are more than two factor values in the attribute column which is selected for Comparison, each factor value will be compared to the others in the column.

   (For example: Factor Values X, Y and Z. In the Comparison, Factor Value X will be compared to Y and also to Z. Factor Value Y will also be compared to z.) In Figure 28, a comparison will be performed on Liver vs Muscle, Liver vs Spleen, and Muscle vs Spleen.

IMPORTANT! When selecting Comparison for a column, all attributes in that column must be populated.

Figure 28 One-Way ANOVA Comparison



2. Enter a Result Name (the name of the batch folder for this analysis run).
3. Click the Output Folder Browse button to select where you want the Results Name (batch) folder saved.
4. Click **Run Analysis**.

A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

When the analysis is complete, it automatically opens in the TAC 4.0.2 Viewer.

Setting up Two-Way ANOVA comparisons in the New Analysis window tab

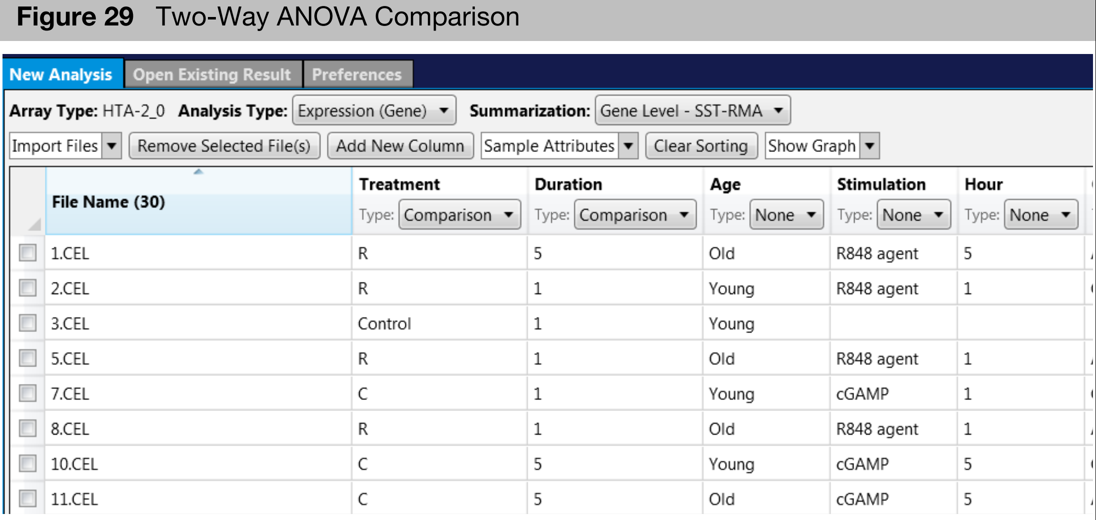
IMPORTANT! At least two factors must be imported in order to run a comparison analysis.

1. From the drop-down menu (in the headers of the column you would like to compare), select **Comparison**.

   If there are more than two factors selected for a comparison, then a comparison is performed between all attributes. A comparison is performed on the Treatment and Duration, as shown in Figure 29.

IMPORTANT! When selecting a Comparison for a column, all contents in that column must be populated.

Figure 29 Two-Way ANOVA Comparison



2. Enter a Result Name (the name of the batch folder for this analysis run).
3. Click the Output Folder's **Browse** button to select where you want the Results Name (batch) folder saved.
4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC 4.0.2 Viewer.

   Note: Use the "Comparison Wizard" on page 46 to automatically set up comparisons on all your defined conditions.

Setting up a comparison analysis using repeated measure

In a repeated measure experimental design, multiple measurements are performed on the same set of subjects. The interest lies in differences between the measurements, so the repeated measure removes the effect of the subjects. TAC 3.0 referred to the two-measurement case as pair-wise.

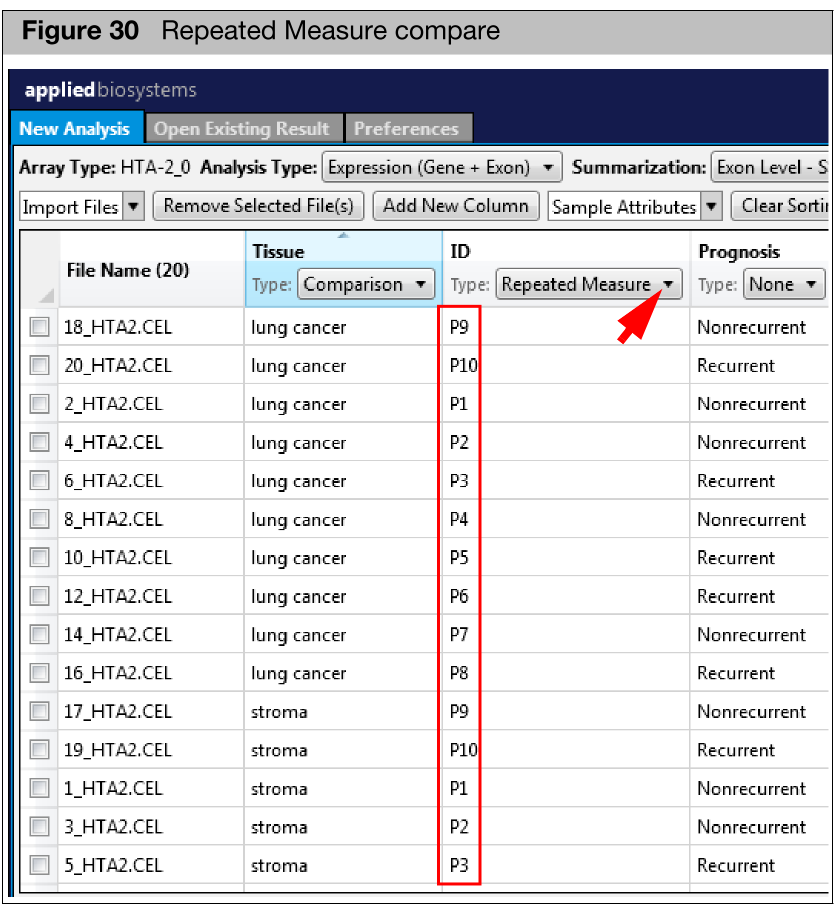
Repeated measure is equivalent to what was known as pairwise analysis in TAC 3.0/3.1. Assuming that you are interested in the difference between the tissues and that the patients represent a “batch” effect that you wish to remove so as to improve the tissue comparison, ID is the repeated measure and Tissue is the comparison. In Figure 30, the ID is the repeated measure (patient) and Tissue is the comparison (stroma vs lung cancer).

1. From the drop-down menu (in the headers of the column you would like to compare), select **Comparison**.

IMPORTANT! When selecting Comparison for a column, all factor values in that column must be populated.

2. For the column with the related groups you want to compare, click **Repeated Measure**. This setting will compare two tissue types over various related samples (P1-P10), as shown in Figure 30.

Figure 30 Repeated Measure compare



3. Enter a Result Name (the name of the batch folder for this analysis run).
4. Click the Output Folder Browse button to select where you want the Results Name (batch) folder saved.

5. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC Viewer.

Utilizing Batch Effect in a comparison analysis

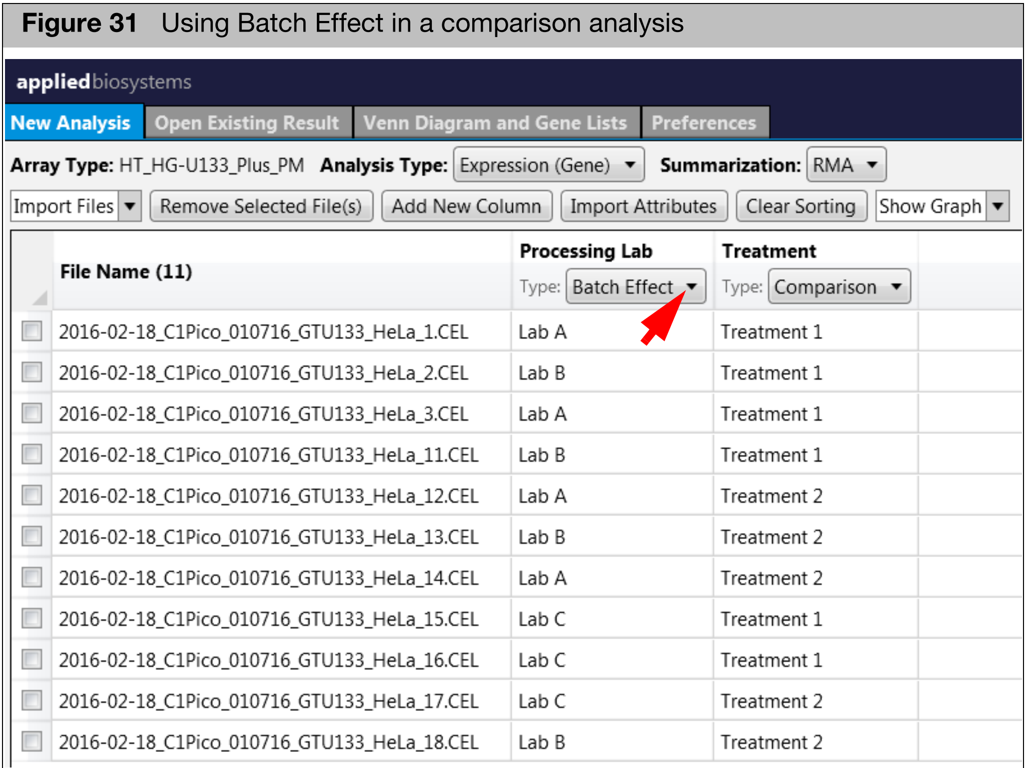
Expression experiments should be designed to minimize any variability other than the condition(s) being studied. Setting an attribute to be a batch factor will remove the effect of this attribute from the differential expression results. The level structure of the batch factors must be different than the level structure of the comparison factors otherwise removal of the batch effect will also remove the differential expression associated with the comparisons of interest.

1. From the drop-down menu (in the headers of the column you would like to compare), select **Comparison**.

IMPORTANT! When selecting Comparison for a column, all attributes in that column must be populated.

For the column the most potential source of unintentional variability, use its drop-down menu to click **Batch Effect**, as shown in Figure 31.

Figure 31 Using Batch Effect in a comparison analysis



2. Enter a Result Name (the name of the batch folder for this analysis run).
3. Click the Output Folder Browse button to select where you want the Results Name (batch) folder saved.
4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC Viewer.

Assigning a Normalization Group in a comparison analysis

This factor controls the transformation of the CEL files into CHP files. It specifies which samples should be normalized and summarized as a group. When the samples represent different conditions that result in signal distributions that are significantly different, these conditions should be normalized and summarized separately as it is most likely inappropriate to force them to a common distribution. For example, if using two different tissue types in an experiment, these sample types should be normalized separately prior to the comparison.

1. Select Normalization Group from the drop down in the header of the column you would like to compare.

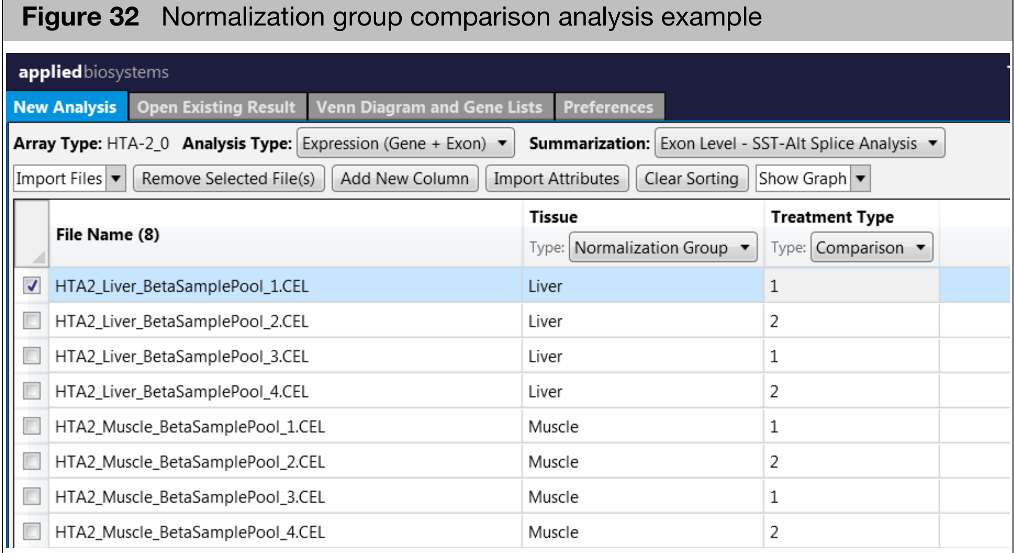
   Note: When selecting any option from the drop down for a column, all attributes in that column must be populated.

2. Name the Result Name. This will be the name of the batch folder of this analysis.
3. Select the path for the Results Name folder to be saved. (Figure 32)
4. Click **Run Analysis**.

   A status message appears.

5. Click **Cancel** to cancel the analysis in process, then click **Close** to return to the New Analysis Tab. When the analysis completes, it automatically opens for viewing.

Figure 32 Normalization group comparison analysis example



Random Factor (Advanced)

Limma has an approximation for the random effects ANOVA model, where all of the probesets share a common correlation. The addition of a random effects model can facilitate analysis of multi-level and split-plot experimental designs.

The random factor addresses the case of a repeated measure

See "Multi-level experiments" (page 48) within the following link:

http://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf

Using the Random Factor

The first column is **Type** (**N**ormal or **C**ancer), the second column is **Pt** (Patient 1, 2, 3, 4, 5), and the third column is **Treatment** (**U**ntreated and **T**reated). For this example, the Random Factor is Patient.

Type	Pt	Tr
N	1	U
N	1	T
N	2	U
N	2	T
C	3	U
C	3	T
C	4	U
C	4	T
C	5	U
C	5	T

Real Covariate (Advanced)

A Real Covariate is a continuous value whose effect needs to be removed from the analysis. Examples of a real covariate include age or tumor size.

The Real Covariate feature allows for a blocking factor that is continuous rather than groups of samples. The standard examples would be removing the effect of age or tumor size from the analysis. Without this feature, continuous blocking factor into groups must manually deciphered. Example: Small, Medium, and Large.

F-tests

TAC 4.0.2 includes F-tests for attributes as a way to analyze whether a given attribute's levels are associated with changes in expression. While a standard statistical F-test is independent of the order of the samples, Limma computes the F-test using the specified contrasts. Contrasts are created for the purpose of computing the per-attribute F-tests, but it is not possible to define such contrasts symmetric to permutations in the sample order. Changes in the sample order can lead to ~1% differences in F-test p-values that have no practical effect on interpretation of the results.

Comparison Wizard

IMPORTANT! At least one factor must be imported in order to run a comparison analysis.

If you have already selected **Comparison** as column header(s) in the New Analysis window tab, the comparisons are automatically displayed in the Comparison Setup Wizard.

You can customize the comparisons you would like to run using the Comparison Wizard. The Comparison Wizard will auto-generate all the possible combinations based on your selections in the New Analysis Tab. If you are only interested in certain combinations, you can remove unwanted comparison. If you are unsure how to set up your analysis, you can use the four provided templates to assist you in setting up comparisons to answer the question being asked in the experiment.

Use the pre-populated drop-down menus to select your values, then click the **Add** button to place it in the Comparisons queue (lower pane). **Note:** After you click **Run Analysis**, the comparisons will run in a serial order.

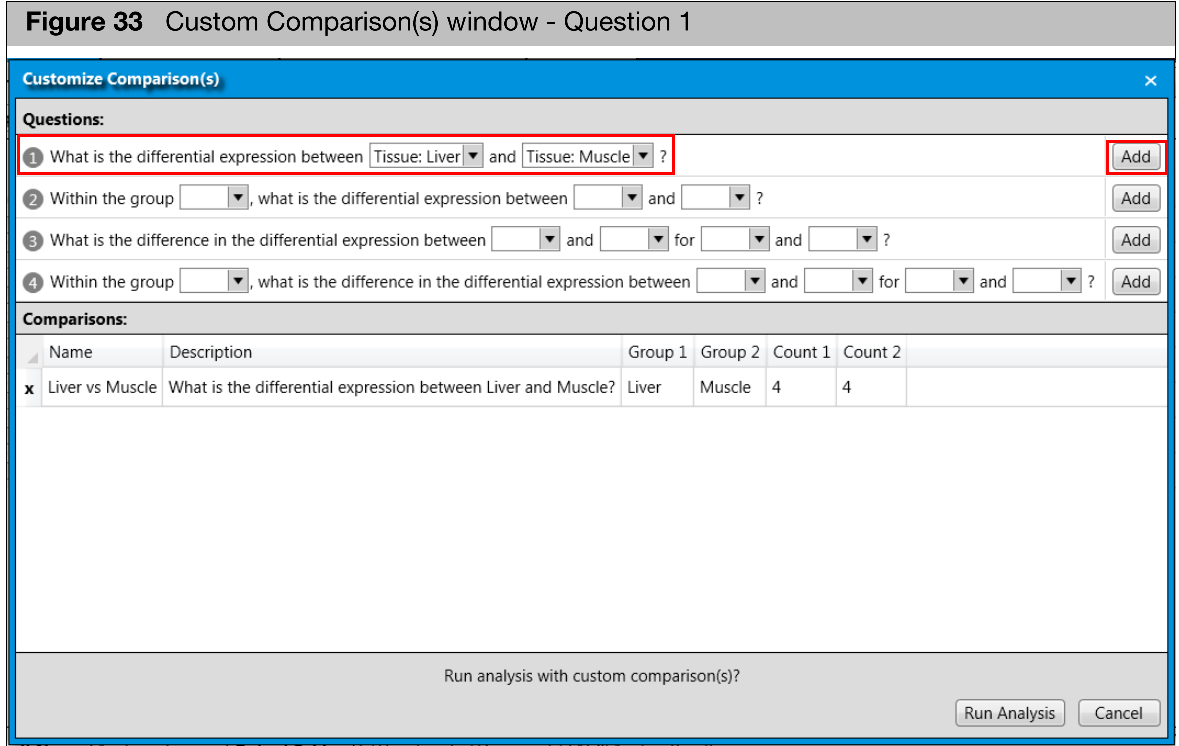
Comparison questions

Q1: Comparison Analysis between Two Factor Values

1. From the New Analysis window tab, click **Comparison Setup Wizard** button (bottom right). **Note:** This button only appears when performing an Analysis Type: Expression (Gene) or Expression (Gene + Exon)

   The Custom Comparison(s) window appears. (Figure 33)

Figure 33 Custom Comparison(s) window - Question 1



2. Use Question 1's drop-down menus to set up the comparison.
3. Click **Add**.

   The comparison is added to the window's lower pane, as shown in Figure 33.

4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC Viewer.

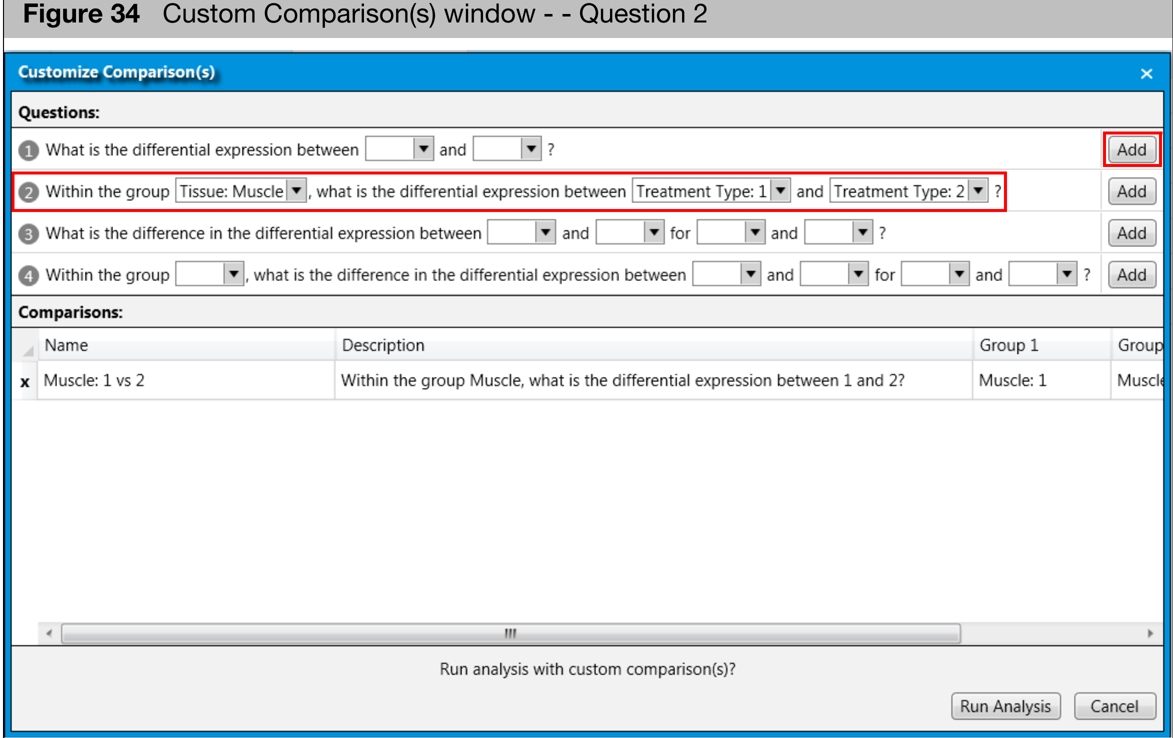
Q2: Comparison Analysis between Two Factor Values (within the same attribute column)

Note: A comparison within a single group is only available using the Comparison Wizard.

1. From the New Analysis window tab, click **Comparison Setup Wizard** (bottom right).

   The Custom Comparison(s) window appears. (Figure 34)

Figure 34 Custom Comparison(s) window - - Question 2



2. Use Question 2’s drop-down menus to set up the comparison.
3. Click **Add**.

   The comparison is added to the window’s lower pane, as shown in Figure 34.
4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

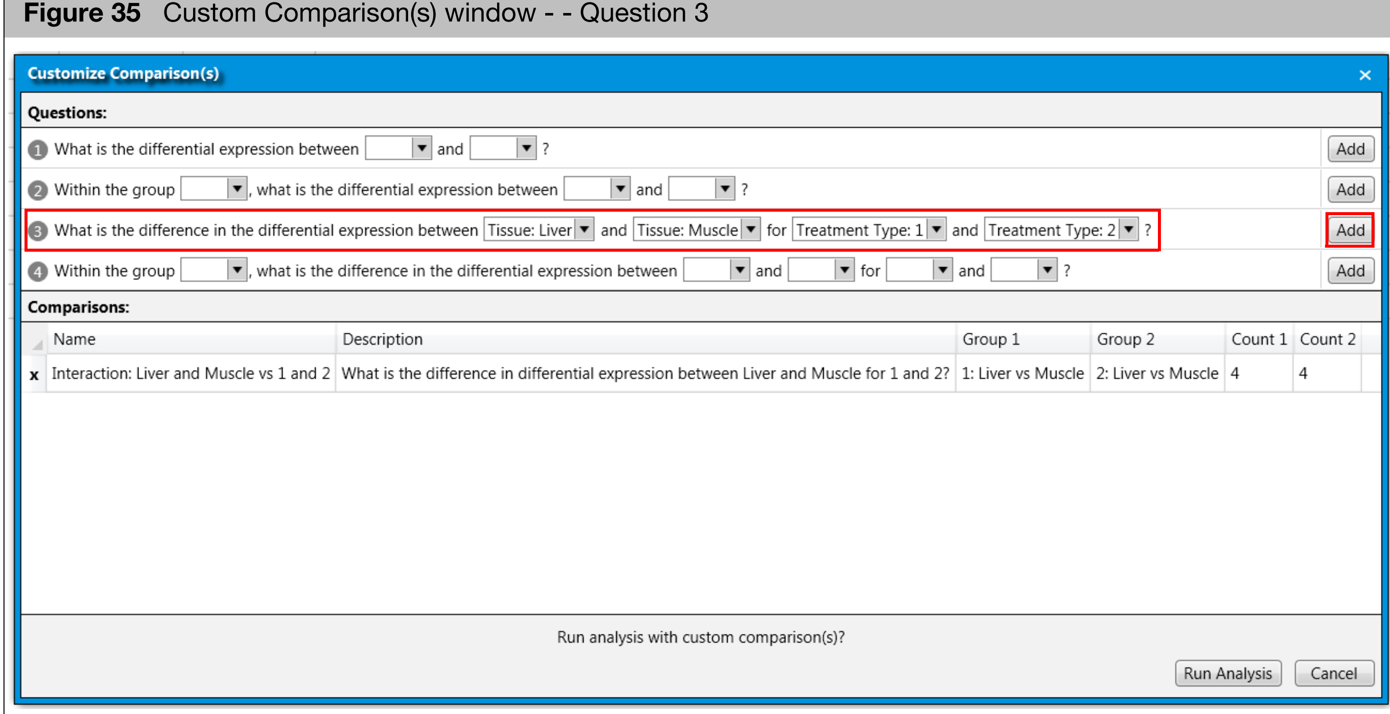
   When the analysis is complete, it automatically opens in the TAC Viewer.

Q3: Interaction Analysis between Two Factors

1. From the New Analysis window tab, click **Comparison Setup Wizard** (bottom right).

   The Custom Comparison(s) window appears. (Figure 35)

Figure 35 Custom Comparison(s) window - - Question 3



2. Use Question 3's drop-down menus to set up the comparison.

   Note: When drop-down selections are not allowed, they are highlighted in red indicating the selection needs attention.

   Figure 35 is an example of finding the expression difference between Treatment 1 and Treatment 2 for the 2 Tissue groups: Muscle and Liver.

3. Click **Add**.

   The comparison is added to the window's lower pane.

4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC Viewer.

Q4: Interaction Analysis between Two Factors (within the same attribute column)

1. From the New Analysis window tab, click **Comparison Setup Wizard** (bottom right).

   The Custom Comparison(s) window appears. (Figure 36)

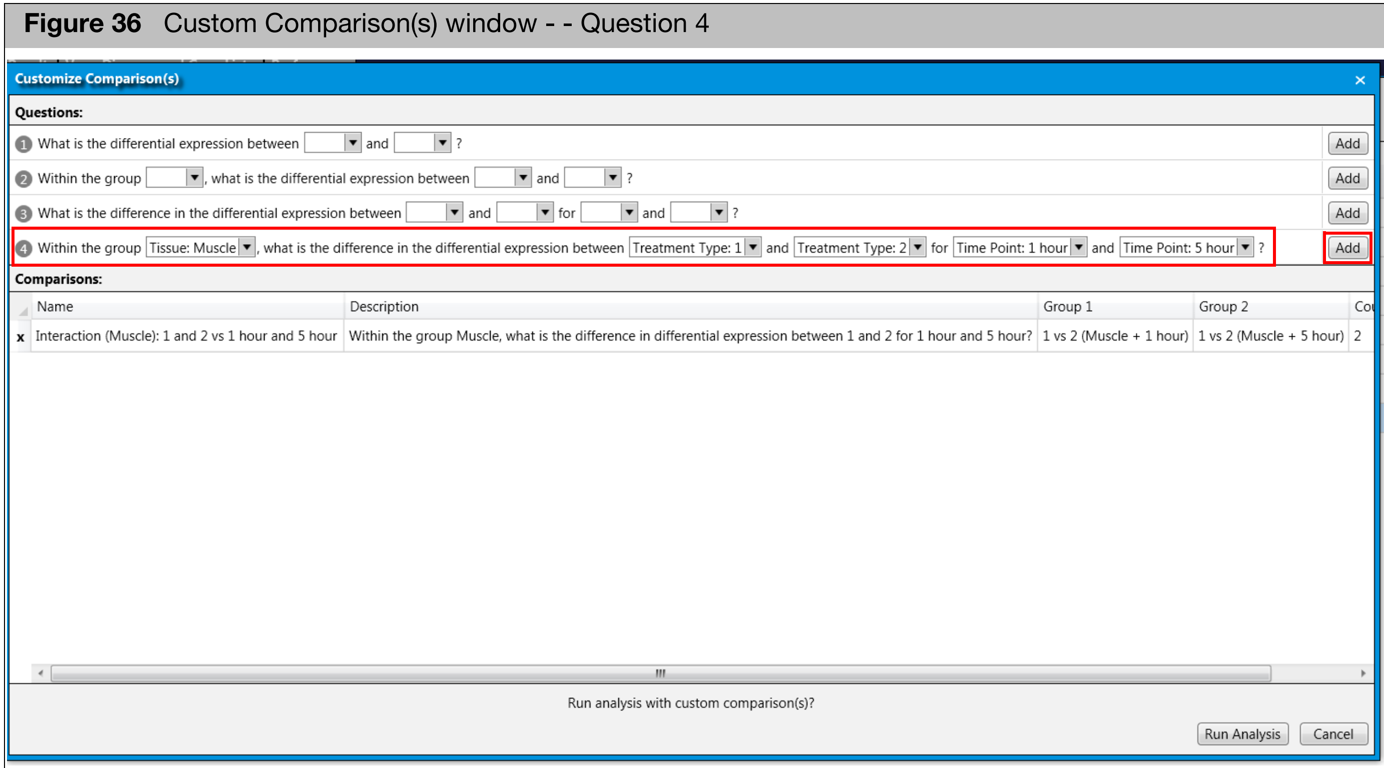


Figure 36 Custom Comparison(s) window - - Question 4

2. Use Question 4's drop-down menus to set up the comparison. Note: When drop-down selections are not allowed, they are highlighted in red indicating the selection needs attention.

   Figure 36 is an example query setup to show the intersection of Treatment 1 and Treatment 2 for the time points of 1 hour and 5 hours (within the Muscle Tissue Group).

3. Click **Add**.

   The comparison is added to the window's lower pane.

4. Click **Run Analysis**.

   A status message appears. To cancel the analysis in progress, click **Cancel**, then click **Close** to return to the New Analysis window tab.

   When the analysis is complete, it automatically opens in the TAC Viewer.

Setting up a Two Way ANOVA

Using the information from the New Analysis Tab set up in Figure 29 on page 41, click the **Comparison Wizard** button to automatically generate comparisons across all conditions, as shown in the Figure 37 example.

Figure 37 Comparisons across all conditions example

Comparisons:

	Name	Description	Group 1	Group 2	Count 1	Count 2
x	C vs Control	What is the differential expression between C and Control?	C	Control	12	6
x	C vs R	What is the differential expression between C and R?	C	R	12	12
x	Control vs R	What is the differential expression between Control and R?	Control	R	6	12
x	1 vs 5	What is the differential expression between 1 and 5?	1	5	18	12
x	1: C vs Control	Within the group 1, what is the differential expression between C and Control?	1: C	1: Control	6	6
x	1: C vs R	Within the group 1, what is the differential expression between C and R?	1: C	1: R	6	6
x	5: C vs R	Within the group 5, what is the differential expression between C and R?	5: C	5: R	6	6
x	Interaction: C and R vs 1 and 5	What is the difference in differential expression between C and R for 1 and 5?	1: C vs R	5: C vs R	12	12
x	1: Control vs R	Within the group 1, what is the differential expression between Control and R?	1: Control	1: R	6	6

Run analysis with custom comparison(s)?

Run Analysis | Cancel

Analysis results and QC review

After the analysis is complete, the results automatically open in the TAC Viewer.

By default, the Sample Table and Graph panes appear in a side by side configuration, as shown in Figure 38.

Figure 38 Default Viewer display

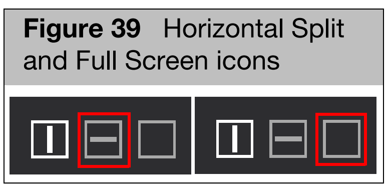
File Name (count: 16)	Labeling Controls Threshold	Hybridization Controls Threshold	Pos vs Neg AUC Threshold	Condition
WTPlus_Liver_Rep1_ClariomD.sst-r...	Pass	Pass	Pass	Liver
WTPlus_Liver_Rep2_ClariomD.sst-r...	Pass	Pass	Pass	Liver
WTPlus_Liver_Rep3_ClariomD.sst-r...	Pass	Pass	Pass	Liver
WTPlus_Liver_Rep4_ClariomD.sst-r...	Pass	Pass	Pass	Liver
WTPlus_Muscle_Rep1_ClariomD.sst...	Pass	Pass	Pass	Muscle
WTPlus_Muscle_Rep2_ClariomD.sst...	Pass	Pass	Pass	Muscle
WTPlus_Muscle_Rep3_ClariomD.sst...	Pass	Pass	Pass	Muscle
WTPlus_Muscle_Rep4_ClariomD.sst...	Pass	Pass	Pass	Muscle
WTPlus_Spleen_Rep1_ClariomD.sst...	Pass	Pass	Pass	Spleen
WTPlus_Spleen_Rep2_ClariomD.sst...	Pass	Pass	Pass	Spleen
WTPlus_Spleen_Rep3_ClariomD.sst...	Pass	Pass	Pass	Spleen
WTPlus_Spleen_Rep4_ClariomD.sst...	Pass	Pass	Pass	Spleen
WTPlus_Testes_Rep1_ClariomD.sst...	Pass	Pass	Pass	Testes
WTPlus_Testes_Rep2_ClariomD.sst...	Pass	Pass	Pass	Testes
WTPlus_Testes_Rep3_ClariomD.sst...	Pass	Pass	Pass	Testes
WTPlus_Testes_Rep4_ClariomD.sst...	Pass	Pass	Pass	Testes

Viewing options

Changing the default view

- Click the **Horizontal Split** icon for a top and bottom configuration. (Figure 39)
- Click the **Full Screen** icon for individual window panes.

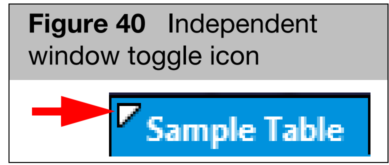
Figure 39 Horizontal Split and Full Screen icons



To toggle a tab window to full screen:

- Click on the tab's white triangle graphic. (Figure 40)

Figure 40 Independent window toggle icon



To toggle a full screen window BACK to its default tab window:

- Double-click anywhere along the top of the window.

To change the size of a window pane:

- Click, hold, then drag the edge of the window pane to resize it.

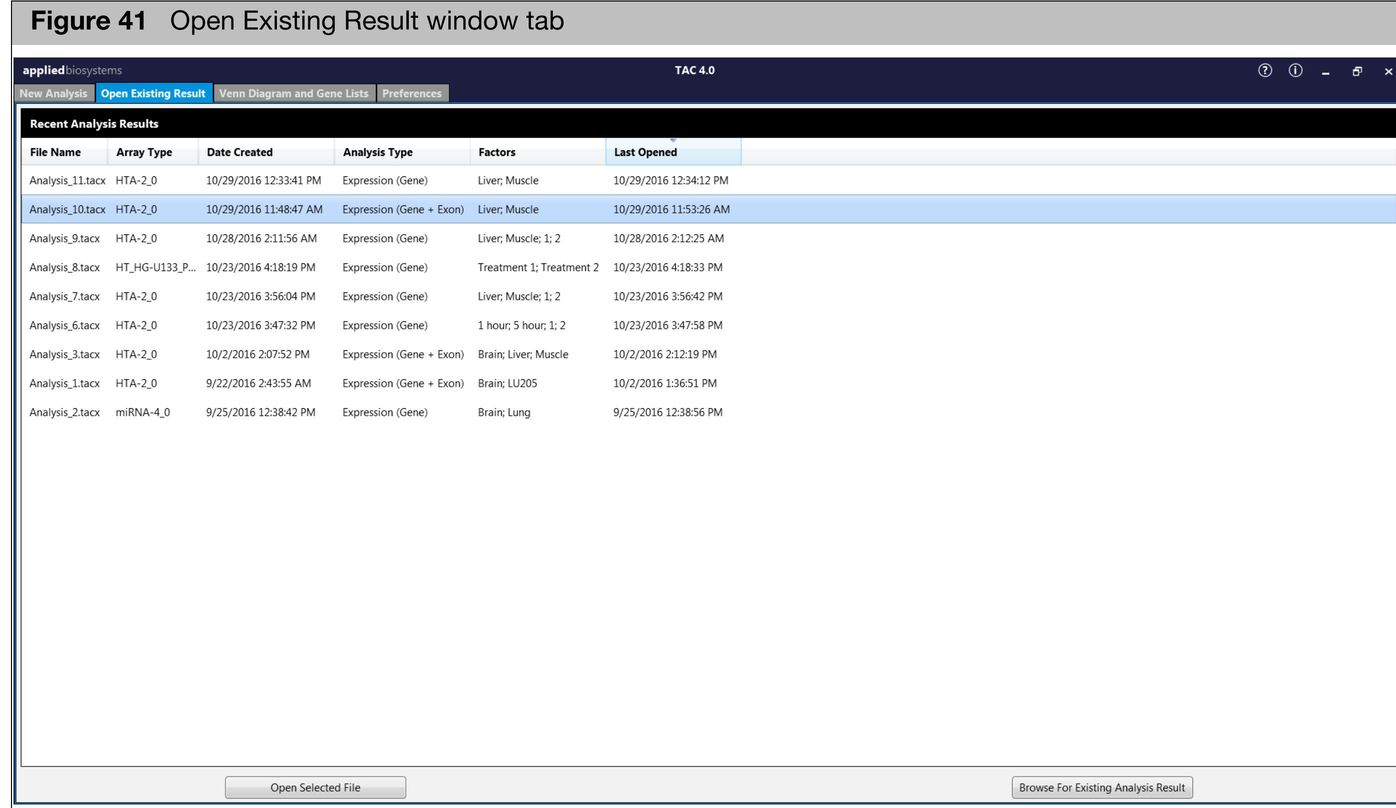
Viewing a completed analysis

TAC auto-saves and displays up to 20 of your most recent analyses.

1. Click on the **Open Existing Result** tab.

   The Open Existing Result window tab appears and displays your most recent results. (Figure 41)

Figure 41 Open Existing Result window tab



2. Double-click on a recent study or single-click on it, then click Open Selected File.

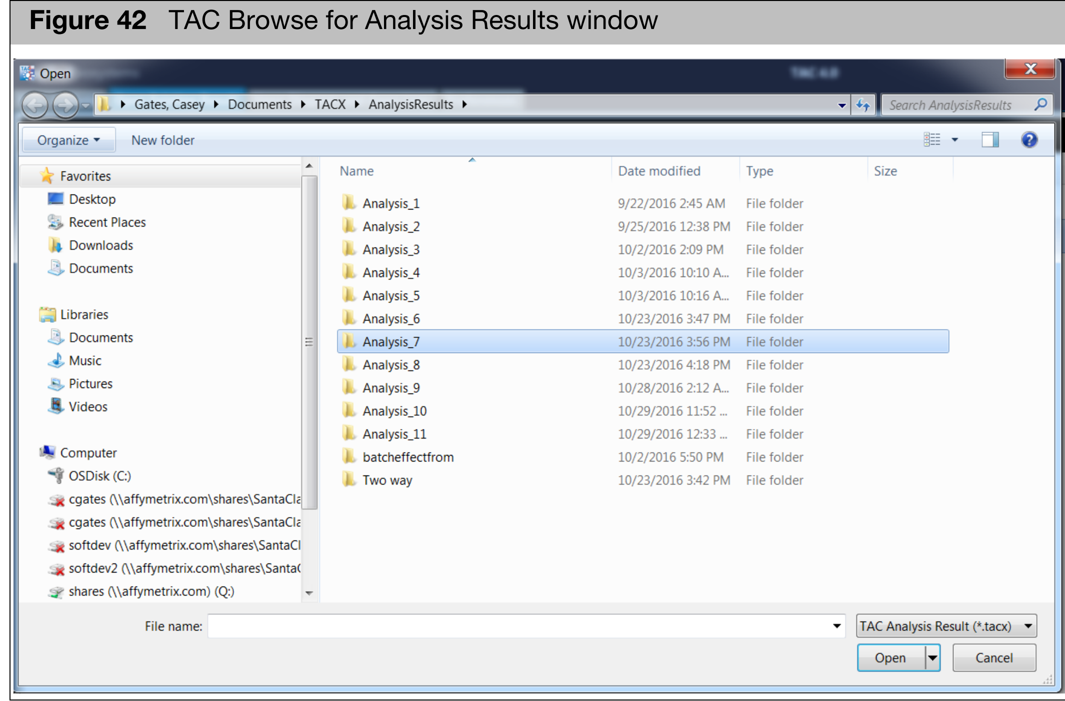
   Note: Click any of the window's header columns to sort your recent studies by Ascending (A-Z) or Sort By Descending (Z-A).

   After a few moments, your study opens in the TAC Viewer (in the same state you last left it).

Browsing for existing analysis result not displayed

1. Click Browse For Existing Analysis Result.
   The Open window appears. (Figure 42)

Figure 42 TAC Browse for Analysis Results window



2. Single-click on the study you want, then click **Open**.
   After a few moments, your study opens in the TAC Viewer (in the same state you last left it).

Introduction to QC Metrics

To help researchers establish quality control processes for gene expression analyses, Thermo Fisher Scientific has 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 TAC 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, refer to the white papers, Data Analysis Fundamentals and Quality Assessment of Exon and Gene Arrays, at thermofisher.com.

In general, Thermo Fisher highly encourages 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.

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

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, b-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 more details on interpreting the housekeeping genes, see the white paper, Data Analysis Fundamentals at thermofisher.com. The Housekeeping controls are:

- GAPDH
- b-Actin

Note: Control probe set names are unique to each array design. Internal Control Genes reside here: **Gene View** > **Gene Table**.

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 Ambion® ERCC RNA Spike-In Control Mixes (Life Technologies 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 (Life Technologies PN 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 (Life Technologies PN 4456739, contains Spike-In Mix 1 and Spike-In Mix 2) to assess the accuracy of differential gene expression measurements.

See the ERCC RNA Spike-In Control Mixes User Guide (Life Technologies Publication Number 4455352) for detailed instructions.

Quality control metrics for CHP files

Sample Table Views

The Sample QC View opens automatically to the Default View. The Sample QC View contains all the information on various quality control metrics depending upon with algorithm was used during analysis. For columns definitions, see Appendix A, "Algorithms" on page 193.

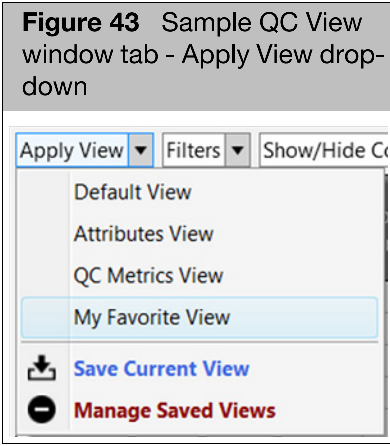
Customizing views

1. Select the **Show/Hide Columns** drop-down to select the columns you wish to view in the Sample Table. You can save this view as a Custom View for future use by clicking on **Save Current View**, choose a name for the view (ex. My Favorite View), then click **OK**.

   This view will now be available in the **Apply View** drop-down, as shown in Figure 43.

   Note: Custom Views will be available in other analysis for the same array type. Example: a custom view saved when analyzing Clariom D arrays will be available for future Clariom D analysis.

Figure 43 Sample QC View window tab - Apply View drop-down

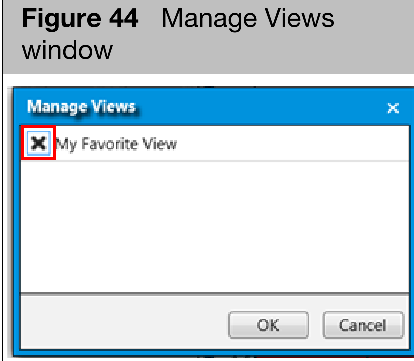


Removing a custom view

1. Click **Manage Saved Views**.

   The Manage Views window appears. (Figure 44)

Figure 44 Manage Views window



2. Click the **X** adjacent to the View name, then click **OK**.

The Custom View is now removed from the Apply View drop-down menu.

Sample Table filters

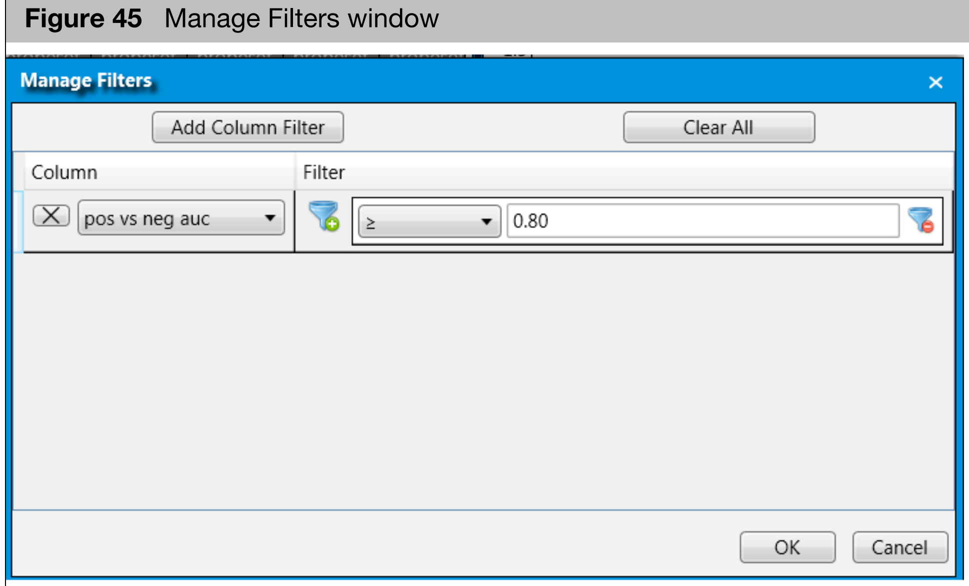
Filters can be applied to hide samples that do not meet desired QC thresholds.

1. From the Filters drop-down menu, click **Manage Filters**.

   The Manage Filters window appears. (Figure 45)
2. Click **Add Column Filter**, then choose the metric you want to apply a filter on.

   For example, pos vs neg AUC >= 0.80 will show samples with a pos vs neg AUC value greater than or equal to 0.80 and will filter out samples with a pos vs neg auc value less than 0.80.
3. Click **Add Column Filter** to filter by more metrics.
4. Click **OK** to apply the filters.

Figure 45 Manage Filters window

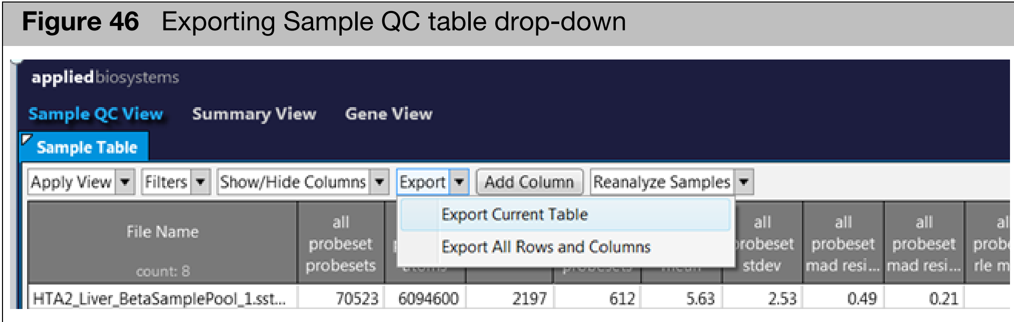


- To list all samples, including those filtered and not filtered, un-check the Sample Table's **Show Filtered Only** check box (bottom right).
- To remove filters, click **Clear All** or click **Clear Current Filter(s)** from the Filter drop-down.
- To show only the samples that have been highlighted in the table or graph, click **Filters** -> **Filter Table By Selection**.
- To show only the samples that are NOT highlighted/selected in the table or graph, click **Filters** -> **Filter Table by Selection (Exclude)**.

Exporting the Sample Table

1. Click on the **Export** drop-down (Figure 46), then choose to export either:
   a. **Export Current Table** - The current columns and rows displayed in the table.
   b. **Export All Rows and Columns** - All columns and all rows in the table whether they are currently being displayed or not.

Figure 46 Exporting Sample QC table drop-down



Adding custom user attributes

Additional User Attributes can be added to the Sample Table to be used in QC analysis. For example: A column can be added for RNA Prep Method and its QC of the data can then be accessed (based on this attribute), as well as other QC metrics.

1. Click **Add Column**.

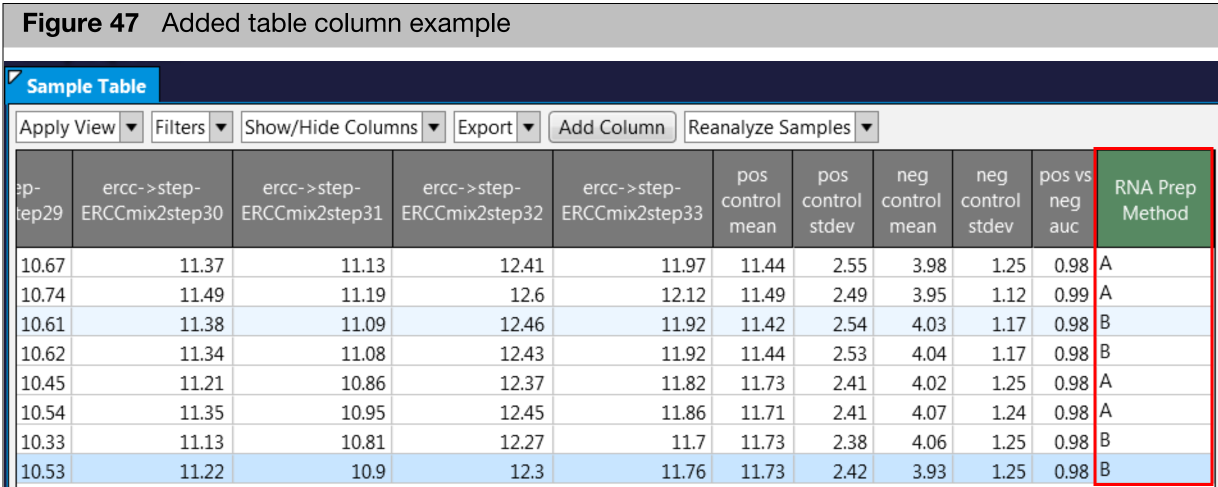
   A New Column window appears.

2. Name the column, then click **OK**.

   The new column is added to the far right of the Sample Table, as shown in Figure 47.

   Optional: Click and drag the column to another position in the table.

Figure 47 Added table column example



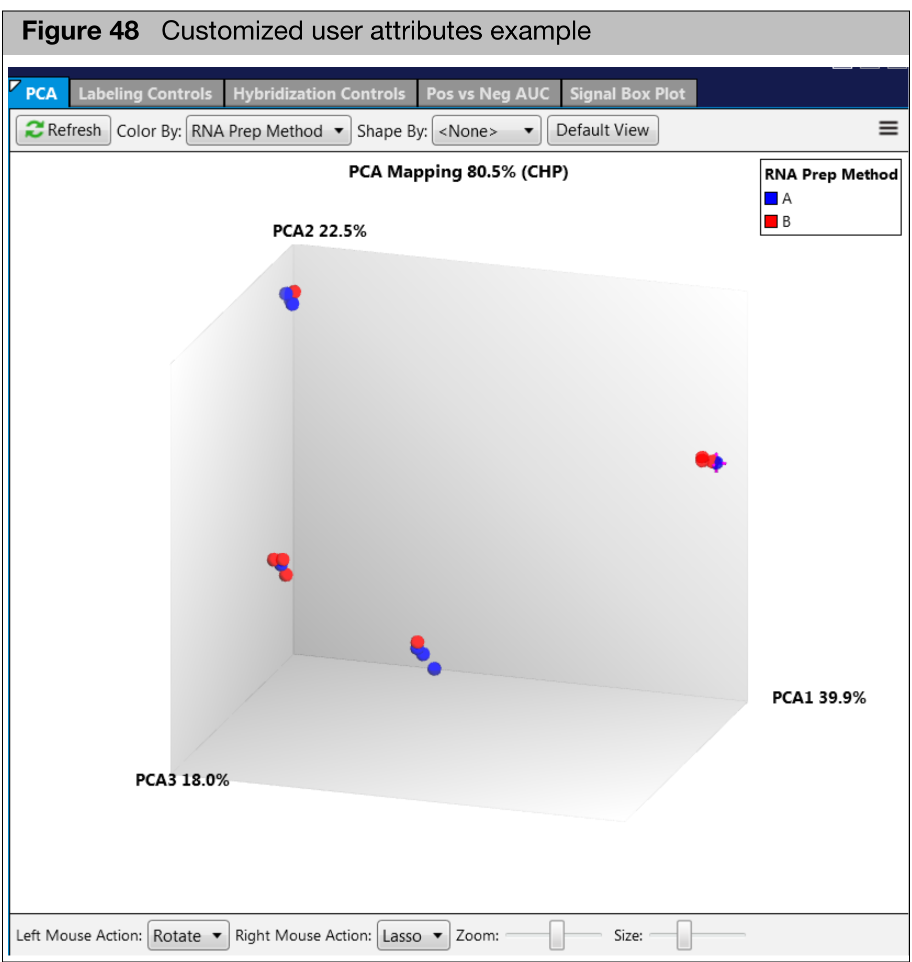
ep-tep29	ercc->step-ERCCmix2step30	ercc->step-ERCCmix2step31	ercc->step-ERCCmix2step32	ercc->step-ERCCmix2step33	pos control mean	pos control stdev	neg control mean	neg control stdev	pos vs neg auc	RNA Prep Method
10.67	11.37	11.13	12.41	11.97	11.44	2.55	3.98	1.25	0.98	A
10.74	11.49	11.19	12.6	12.12	11.49	2.49	3.95	1.12	0.99	A
10.61	11.38	11.09	12.46	11.92	11.42	2.54	4.03	1.17	0.98	B
10.62	11.34	11.08	12.43	11.92	11.44	2.53	4.04	1.17	0.98	B
10.45	11.21	10.86	12.37	11.82	11.73	2.41	4.02	1.25	0.98	A
10.54	11.35	10.95	12.45	11.86	11.71	2.41	4.07	1.24	0.98	A
10.33	11.13	10.81	12.27	11.7	11.73	2.38	4.06	1.25	0.98	B
10.53	11.22	10.9	12.3	11.76	11.73	2.42	3.93	1.25	0.98	B

3. Enter the attributes for each sample.

   Your entered parameter is now available in the Sample QC drop-down menu. This user defined attribute is now available for use in the graph views to the right.

   Figure 48 shows the PCA plot colored by a custom user attribute. When adding or editing a user attribute in the table, click the **Refresh** button to synchronize the graphs with your new updates.

Figure 48 Customized user attributes example



Reanalyzing samples from the Sample table

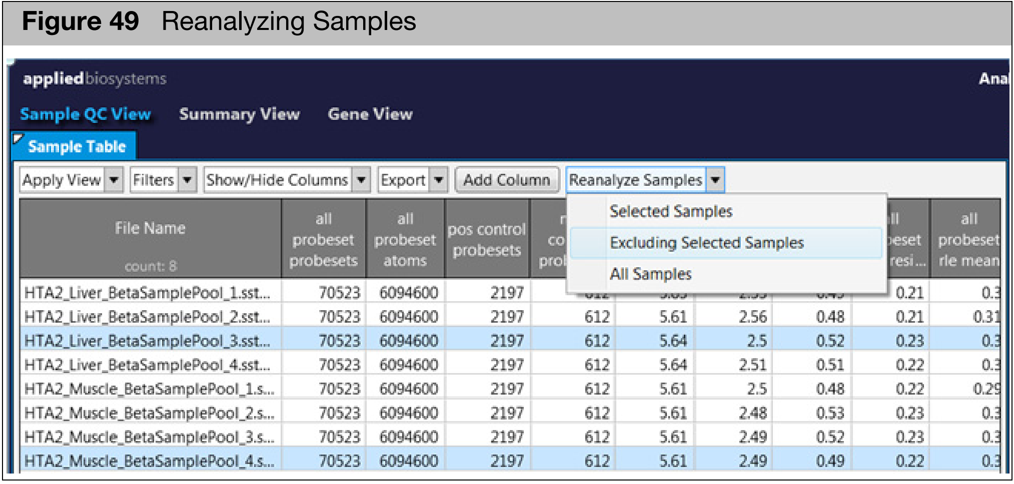
Upon reviewing the QC metrics of the samples in an analysis, if an outlying sample(s) has been identified, the analysis can be re-run from this Sample Table.

- **Method 1**: Reanalyze the Selected samples only. The analysis is re-run including ONLY the highlighted samples.
- **Method 2**: Reanalyze but exclude the selected sample. The analysis is re-run including samples that are NOT highlighted.
- **Method 3**: All samples can be reanalyzed using different attributes or comparison.

1. Select the sample(s) by highlighting the samples to be selected based on the analysis you plan to use (Method 1 or Method 2). Multiple samples can be selected using Shift+left click.
2. Click on the **Reanalyze Samples** drop-down (Figure 49), then select the analysis type you want to re-run.

3. Choose if you want to start the analysis from CEL or CHP files.

Figure 49 Reanalyzing Samples

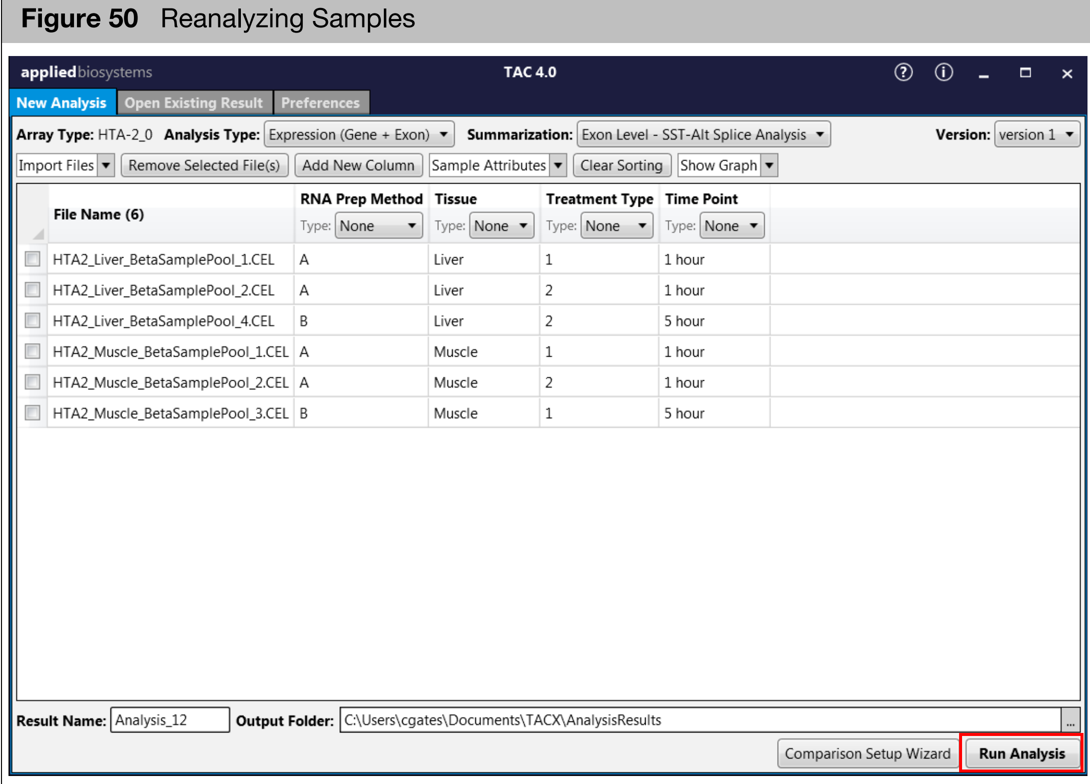


A Select a file type to reanalyze message appears.

– Click **CEL** to start from CEL files.
– Click **CHP** to start from CHP files.

The New Analysis window (Figure 50) automatically opens and loads the CEL or CHP file names. See "Algorithm and summarization selection" on page 36 for available analysis options.

Figure 50 Reanalyzing Samples



4. Click **Run Analysis** to start the reanalysis.

Sample QC Graphs

PCA plot

TAC 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.

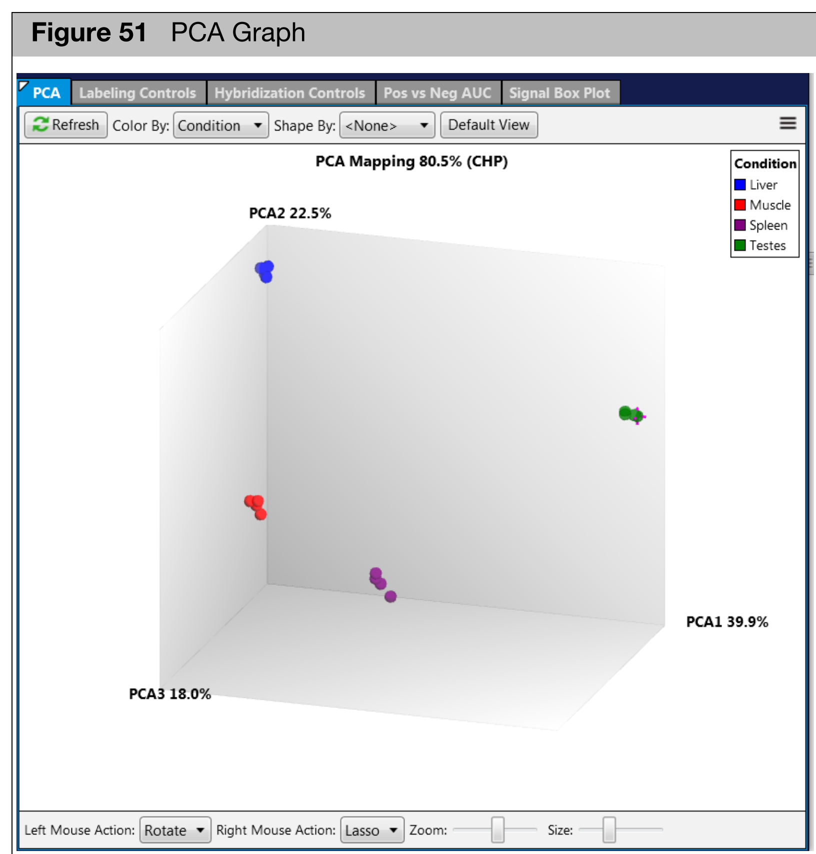
- The first principle component (PCA1) captures as much variability in the data as possible.
- PCA2 captures as much of the remaining variability (not accounted for by PCA1) as possible.
- PCA3 captures as much of the remaining variability (not accounted for by PCA2) as possible.
- Probe cell intensity data are pre-processed for PCA analysis due to memory restrictions. A non-random sampling method selects 50,000 probe sets for analysis.

Sample display options

Samples can be labeled in the PCA graph by color (up to 14 different colors) and shape (up to five shapes) with information in the ARR files. (Figure 51)

- – Select attributes for display by color and by shape from the drop-down lists.
- – Place the mouse pointer on a sample to display its sample information in the graph window.
- – Click one or more outlier samples in the PCA window to highlight them in the Sample Table.

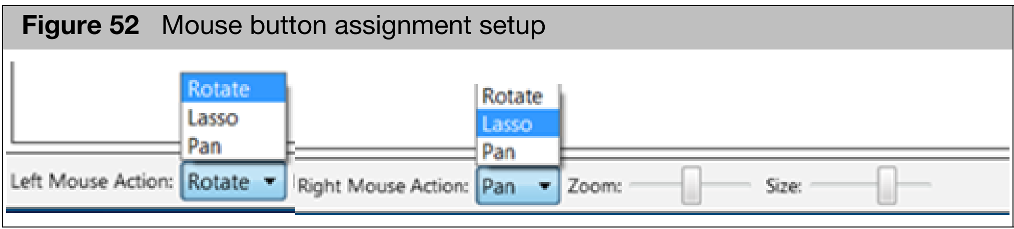
Figure 51 PCA Graph



PCA Graph display options

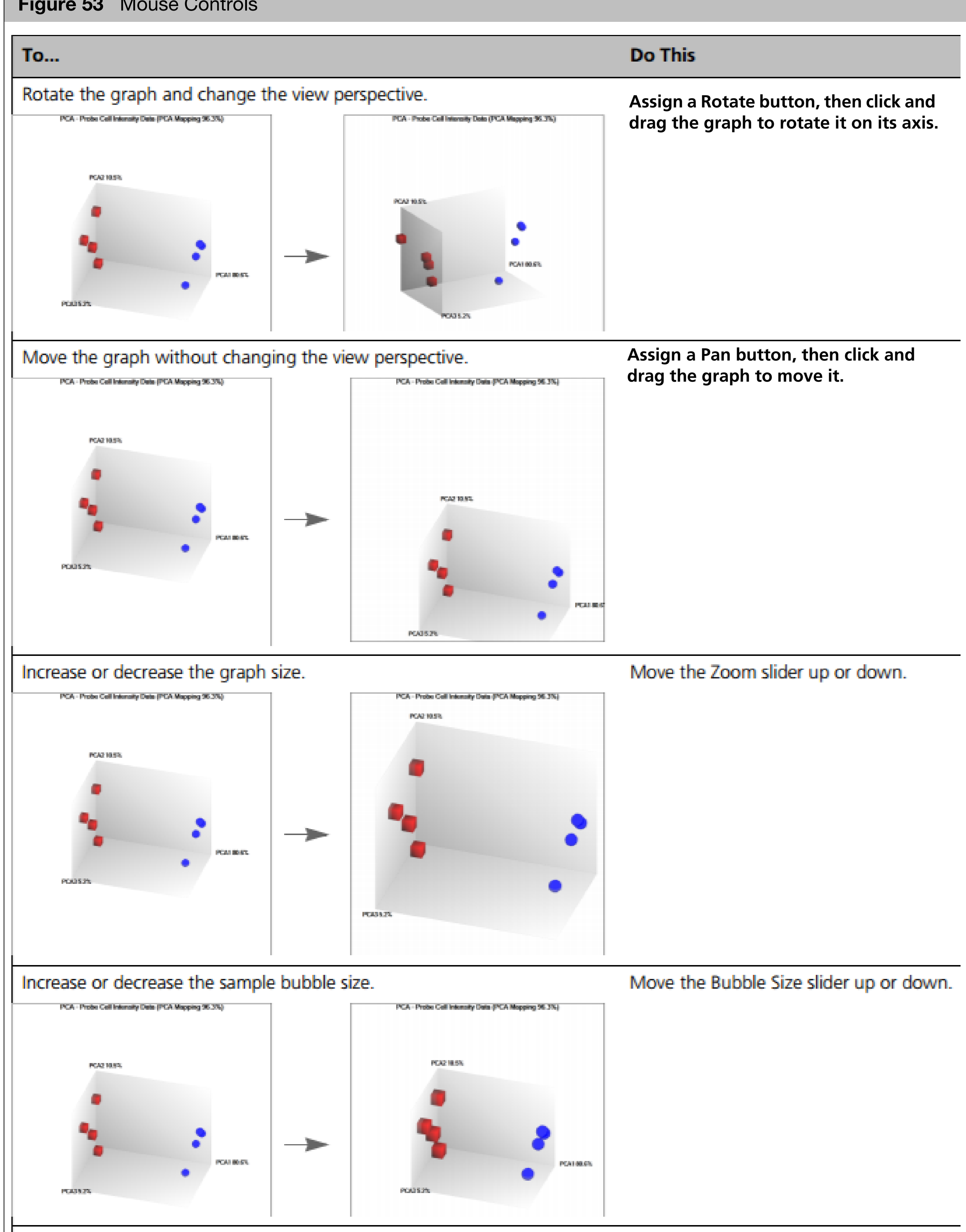
- Assign the Left and Right mouse buttons to either Rotate, Lasso, or Pan the PCA plot. (Figure 52)
- Adjust the zoom and sample shape size to help identify the outliers (among the samples) more clearly.

Figure 52 Mouse button assignment setup



See Figure 53 on page 64 for mouse controls.

Figure 53 Mouse Controls



To...	Do This
Rotate the graph and change the view perspective.	**Assign a Rotate button, then click and drag the graph to rotate it on its axis.**
Move the graph without changing the view perspective.	**Assign a Pan button, then click and drag the graph to move it.**
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 Default.**

Selecting data points in the PCA Plot

1. Assign **Lasso** as either a left or right mouse action.
2. Click and drag a lasso around the data points you want to isolate.
   Your selected samples are now highlighted in the Sample Table.
   Click Clear Selection (upper right menu) to deselect the samples.

Exporting and PCA options

1. Click the Options ☰ button (upper right).
   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

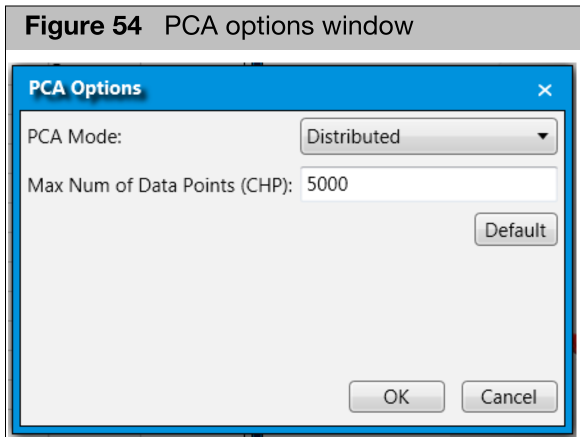
Print

Prints the currently displayed graph.

PCA Options

1. Click **PCA Options**. (Figure 54)
   The PCA Options window appears. (Figure 54)

Figure 54 PCA options window



2. From the **PCA Mode** drop-down menu select either:

- **Distributed** - TAC 4.0.2 auto-selects N evenly distributed probes/probesets from CHP file when computing PCA.
  - **Max Num of Data Points (CHP)** Use the provided text field to enter the maximum number of data points when using CHP files.
- **Variance** - TAC 4.0.2 auto-selects N probes/probesets that contain the highest variance from CHP file when computing PCA.
  - **Increase Sensitivity** - Click this check box if you want to:

  - Weight probes/probsets by variance.
  - Multiply probes/probesets signals by their variance to increase sensitivity.
  - Reduce background noise when smaller number of probes/probesets are present.
- **Max Num of Data Points (CHP)** Use the provided text field to enter the maximum number of data points when using CHP files.

Export Coordinates

Exports coordinates as a TXT file.

Show Legend

Click the **Show Legend** check box to show the Graph's Legend. Uncheck to hide it.

Clear Selection

Click **Clear Selection** to clear a lassoed or highlighted selection/area.

Labeling Controls

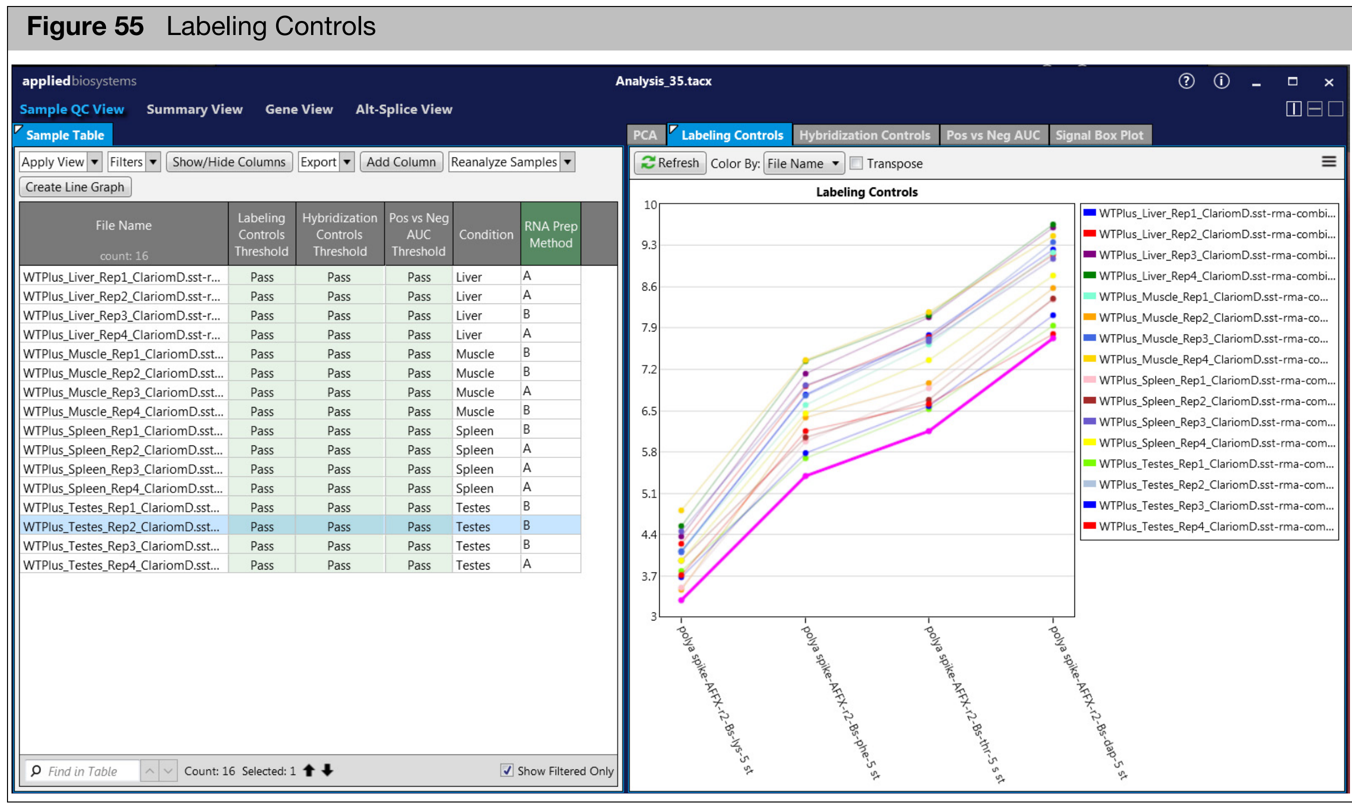
Poly-A RNA controls can be used to monitor the entire target labeling process. Each eukaryotic GeneChip probe array contains probe sets from several B. subtilis genes that are absent in eukaryotic samples (lys, phe, thr, and dap). These Poly-A RNA controls are in vitro synthesized, and the polyadenylated 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.

Labeling controls can be viewed in a line graph. Monitor consistency across samples. Outlier samples can be clicked on (which in turn) highlights them in the Sample Table, as shown in Figure 55.

Samples can be colored by the File Name (Figure 55) or any sample attribute imported for the sample. See "Using the PCA graph window" on page 28.

Figure 55 Labeling Controls



To revert the current X and Y axis view, click the Transpose check box.

For more graph options, see "Labeling Controls Graph Options" on page 68.

Labeling Controls Graph Options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Y Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Clear Selections

1. Click **Clear Selection**.

   All previously selected items are now cleared form the graphs.

Transpose Graph

1. To transpose the Graph, click on the **Transpose Graph** check mark.

Hybridization 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, BioC, BioD, and Cre should always have increasing signal values, reflecting their increasing relative concentrations.

Hybridization Controls can be viewed in a line graph to monitor consistency across samples as well as a relative increase in signal from BioB to Cre . Sample(s) can be clicked on (which in turn) highlights them in the Sample Table.

Samples can be colored by the File Name or any sample attribute imported for the sample. See page 28.

Hyb Controls Graph options

1. Click the Options button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Y Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Clear Selections

1. Click **Clear Selection**.

   All previously selected items are now cleared form the graphs.

Transpose Graph

1. To transpose the Graph, click on the **Transpose Graph** check mark.

Positive vs Negative AUC

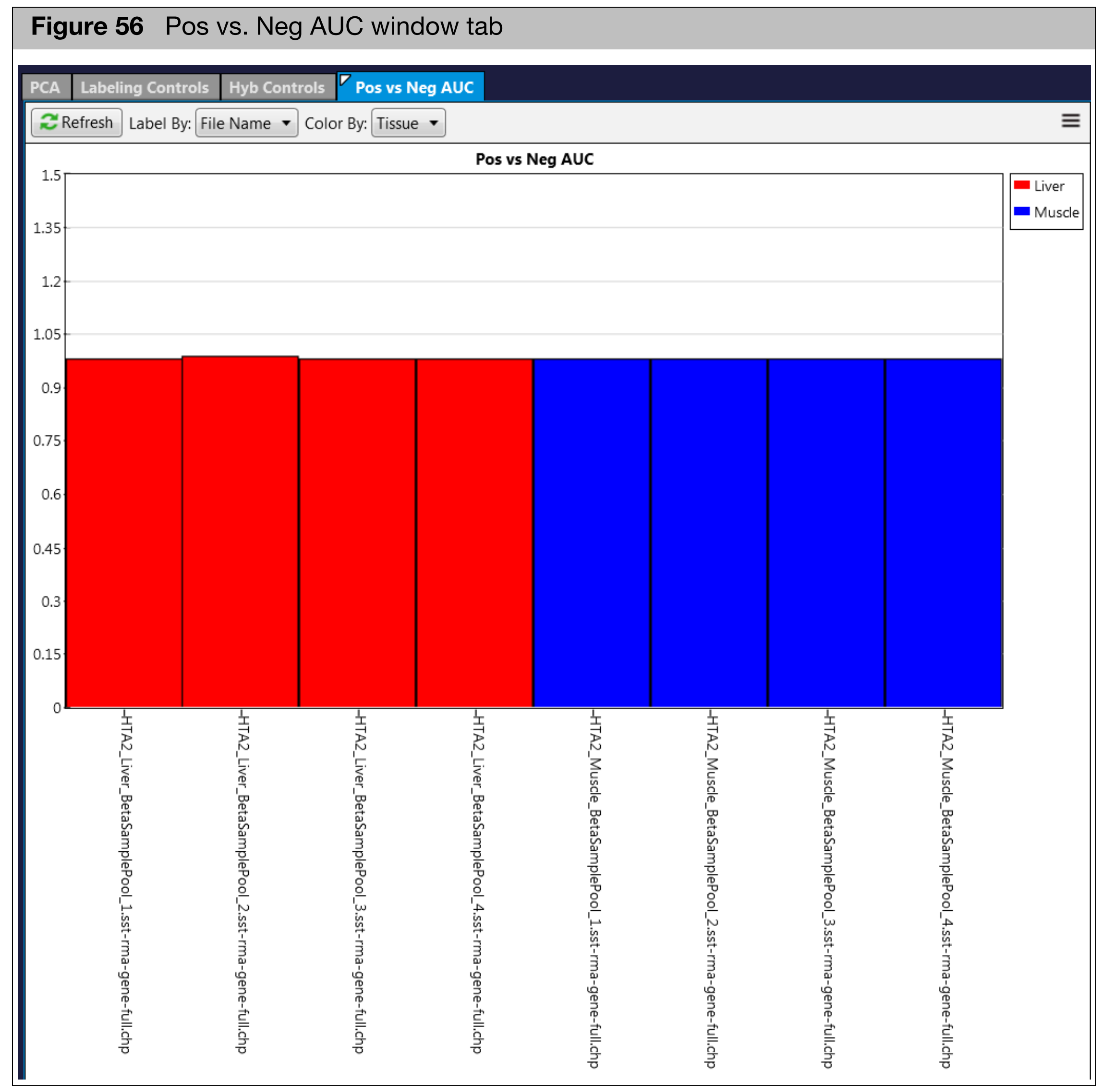
The area under the curve (AUC) for a receiver operator curve (ROC) comparing the intron controls to the exon controls by applying a threshold to the probe set summary

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. If no separation is detected, the "pos vs neg auc" value is 0.5 and the "pos control" and the "neg control" column values is 0. 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.

The Pos vs Neg AUC controls (Figure 56) can be viewed as a bar graph to monitor consistency across samples. Sample(s) can be clicked on (which in turn) highlights them in the Sample Table.

Figure 56 Pos vs. Neg AUC window tab



Pos vs Neg AUC Graph options

1. Click the Options button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Y Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Clear Selections

1. Click **Clear Selection**.

   All previously selected items are now cleared form the graphs.

Transpose Graph

1. To transpose the Graph, click on the **Transpose Graph** check mark.

Signal Box plot

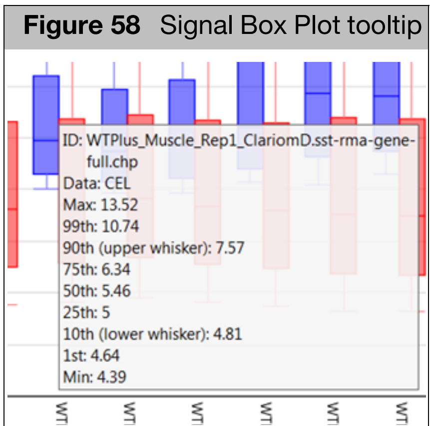
The Signal Box Plot (Figure 57) graph shows a standard box plot of data from either CHP or CEL files. A general rule to use when examining box plots is to look for individual arrays that are dramatically different from the others and most importantly from other replicates in the same group.

Figure 57 Signal Box Plot window tab

- View the signal box plot for CEL files by checking the CEL check box. (Figure 57)
- View the signal box plot for CHP files by checking the CHP check box. (Figure 57)

– Mouse over the box plot to view additional QC information (tooltip) about a specific CEL or CHP file. (Figure 58) Optional: The tooltip feature can be turned off in the graph's Settings menu (upper right).

Figure 58 Signal Box Plot tooltip



Signal Box plot graph options

1. Click the Options button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

- Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.

   The Scale Settings window appears.

2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Y Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Clear Selections

1. To deselect the samples, click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Creating a custom line graph

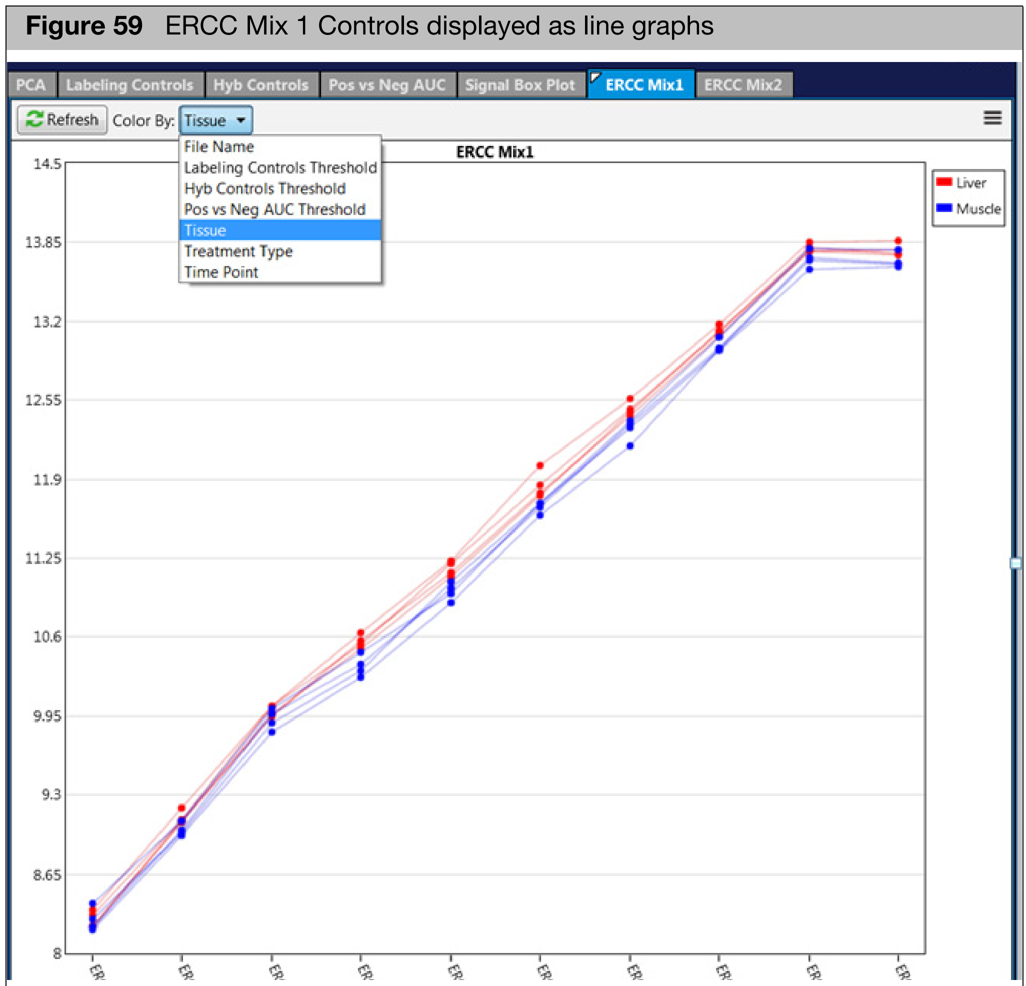
Display metrics or controls in a Custom Line Graph. The ERCC Controls can be viewed in the Custom Line Graph View.

ERCC Controls

The Ambion® ERCC RNA Spike-In Control Mixes (Life Technologies 4456740, 4456739) contain preformulated 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 (Life Technologies PN 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 (Life Technologies PN 4456739, contains Spike-In Mix 1 and Spike-In Mix 2) to assess the accuracy of differential gene expression measurements. See the ERCC RNA Spike-In Control Mixes User Guide (Life Technologies Publication Number 4455352) for detailed instructions.

ERCC controls are displayed as line graphs. (Figure 59) Samples can be colored by various attributes using the Color By drop down. **Note:** If a user attribute is added or edited in the Table View, click the **Refresh** button to update the Color By field in the Line Graph window tab.

Figure 59 ERCC Mix 1 Controls displayed as line graphs



To Create a Custom Line Graph:

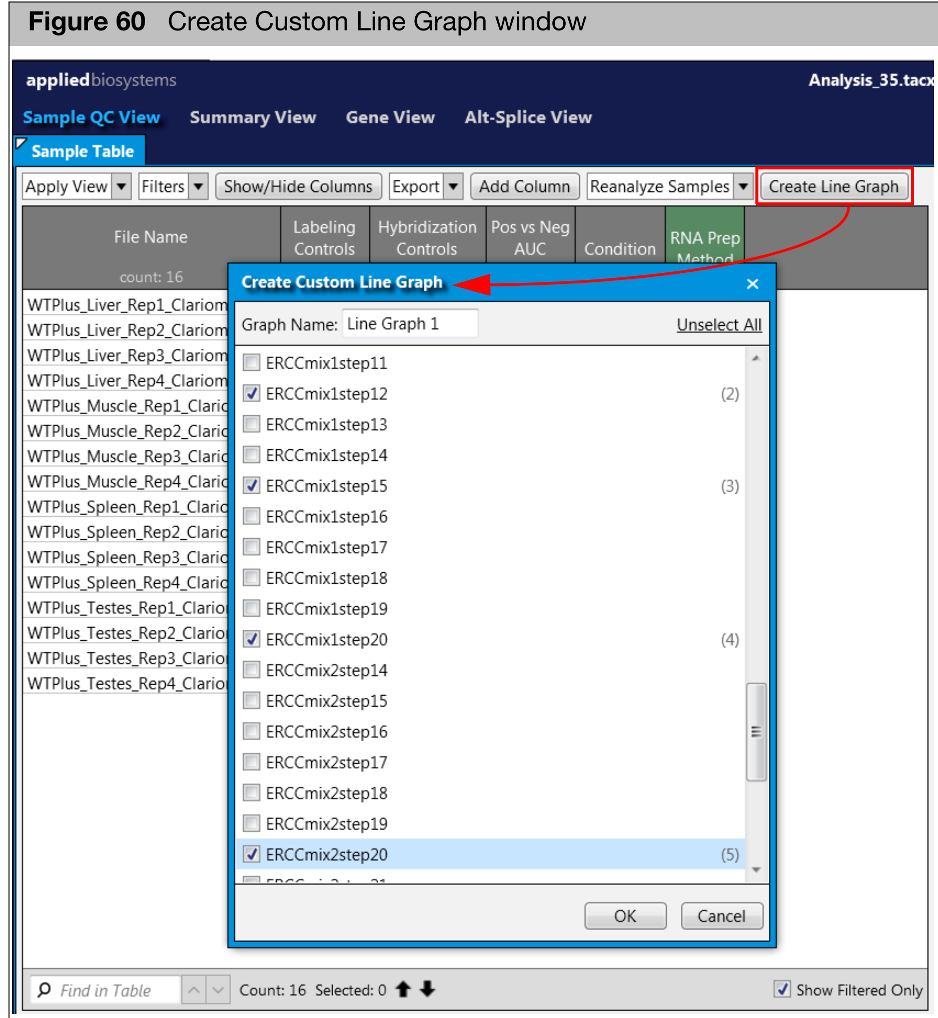
1. Click on the **Create Line Graph** button. (Figure 60)

   The Create Custom Line Graph window appears.

2. Enter a Name (optional) for the custom Line Graph.
3. Check the appropriate box(es) of the controls you want to display in your custom line graph.
4. Click **OK**.

   Your line graph and future custom line graphs are added as their own window tab, as shown in Figure 61 on page 76.

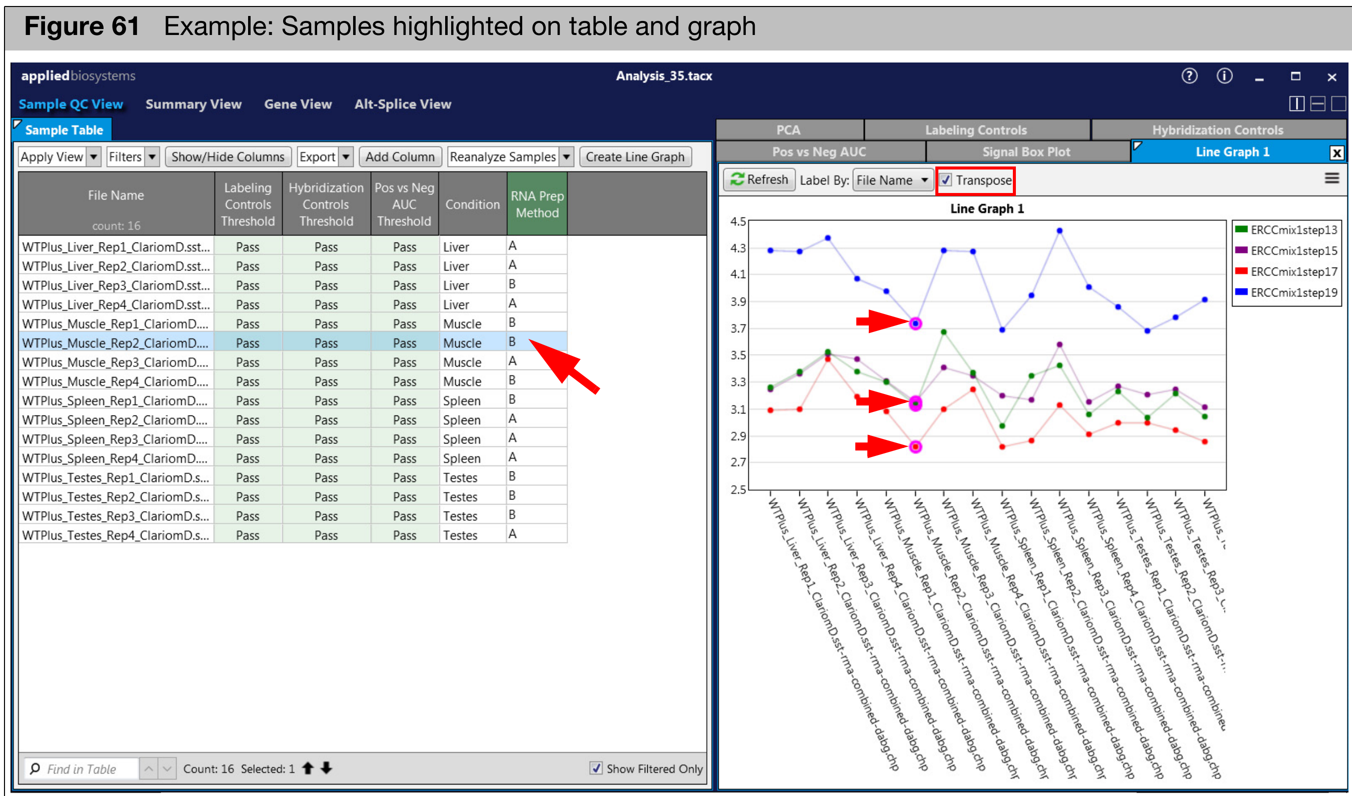
Figure 60 Create Custom Line Graph window



5. Optional: Click the **Transpose** check box to switch the X and Y axis

Note: Samples highlighted in the Graph View are also highlighted in the Table view, as shown in Figure 61.

Figure 61 Example: Samples highlighted on table and graph



Line Graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

- Prints the currently displayed graph.

Scale Settings

1. To change the Y axis scale and grid line appearance, click **Scale Settings**.

   The Scale Settings window appears.

2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Y Max value boxes, then enter the appropriate Min and Max values.

4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Clear Selections

1. To deselect the samples, click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

Transpose Graph

1. To transpose the Graph, click on the **Transpose Graph** check mark.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Distance Matrix Graph (EGA analysis only)

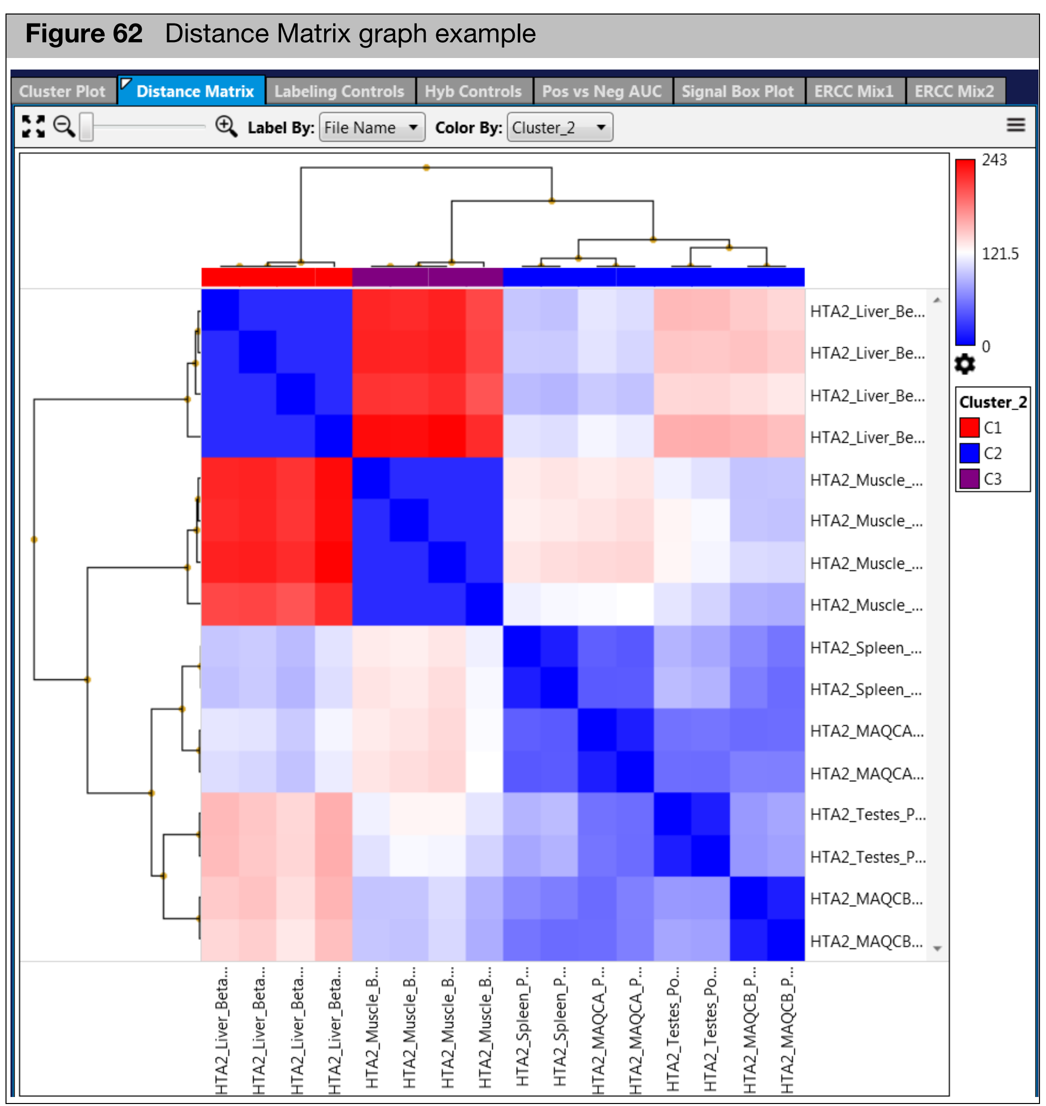
An unlabeled cluster plot provides a qualitative perspective of the relationships among a group of samples. Clustering applies specific labels to these samples. The distance matrix represents the quantitative basis for the clustering algorithm. As the name implies, the distance matrix provides the distance between the samples. Each row and column is labeled by a sample, and each element indicates the distance between the samples associated with the row and column. In TAC 4.0.2, the distance between two samples is computed using the Euclidean metric, that is, the distance is the square root of the sum of squares of the differences between equivalent signals on the samples.

The Distance Matrix graph illustrates how closely related are the samples when examined pairwise. Samples that are more closely related have a lower distance value (in our example blue is set to the lower end of the scale).

When comparing a sample on the right to itself on the bottom, the color block is represented as dark blue. This indicates there is no distance between the samples (because they are the same sample).

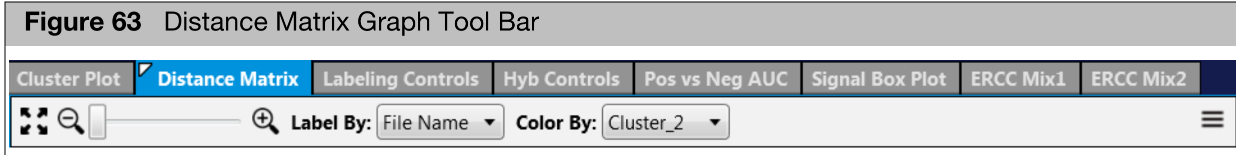
Samples that have different expression profiles are colored in red, as shown in the Figure 62 example. When comparing the HTA2_Liver samples to the HTA2_Muscle samples, the color block is represented as bright red. This denotes the samples are quite different.

Figure 62 Distance Matrix graph example



Using the Distance Matrix Graph's tool bar

Figure 63 Distance Matrix Graph Tool Bar



Note: After adding or editing a user attribute within the table, click on the **Refresh** button to synchronize the graphs with your new updates.

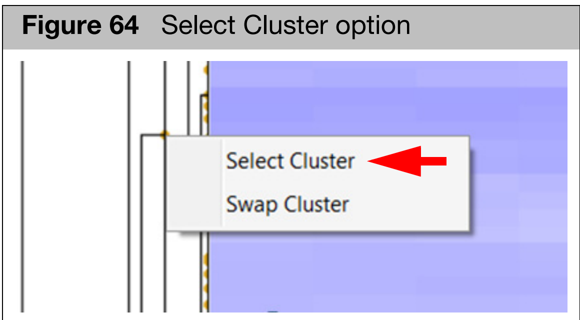
- Click the Zoom slider bar (Figure 63), then hold down the left mouse button and move the bar right to magnify point(s) of interest.
- Click [icon] to return to the default display.
- Click the **Color By** drop-down to select a different Color by display. In the example above, the graph is Colored By an assigned Factor Value named Cluster 2. The key on the right of the Distance Matrix shows the factor value of "Cluster_2" assigned prior to analysis. The top of the matrix is a color block showing which samples were user defined as C1 (red), C2 (blue), and C3 (purple). For assigned user attributes or factor values, "Assigning custom sample attributes" on page 24.

Selecting samples in a cluster

If you are interested in viewing which samples are in a particular branch of the graph, right click on a yellow node on the left side of the graph. (These steps can also be performed on the yellow nodes at the top of the graph to select items based on the **Label By** drop-down). The samples are also highlighted in the Table.

1. In the **Filters** drop-down menu select **Filter Table by Selection** to view only those selected samples in the table.
2. Click on **Select Cluster** to highlight the gene (listed on the right) in that cluster. (Figure 64)

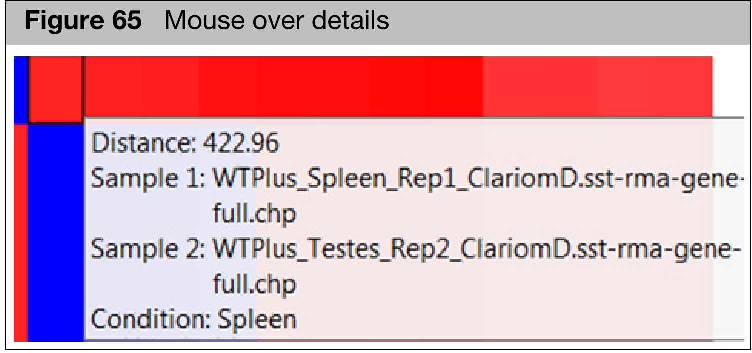
Figure 64 Select Cluster option



3. Mouse over a desired region in the Distance Matrix.

   Mousing over a block provides details on the comparison, as shown in Figure 65.

Figure 65 Mouse over details



– **Signal** - The distance between the samples.
– **Sample and ID** - The names of the files in the comparison.
– **Cluster** - The cluster number that has been determined by the EGA algorithm.

Distance Matrix graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

Print

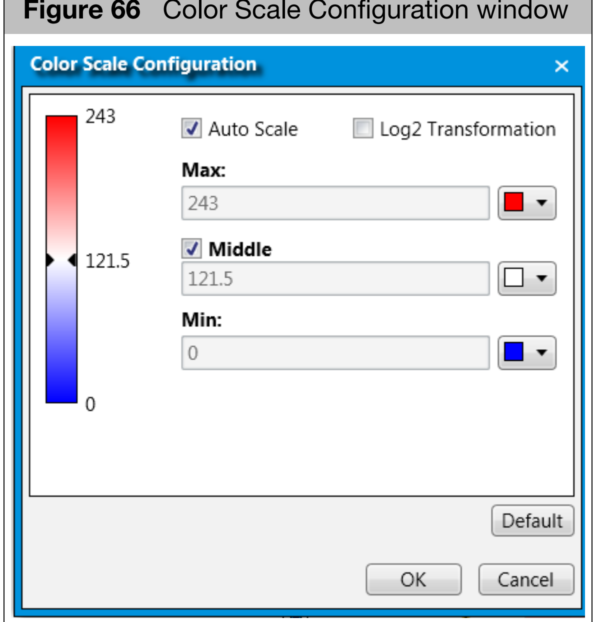
1. Click **Print**, then print the currently displayed graph using your configured printer as you normally would.

Color Settings

1. Click **Color Settings**. Alternatively, click on the Color Scale Key Legend (right side of graph).

The Color Scale Configuration window appears: (Figure 66)

Figure 66 Color Scale Configuration window



2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation.

   Note: Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.

3. Use the color drop-down menus to assign new colors for your data points.
4. Click **OK** to save your changes or click **Default** to return to the factory settings.

Show Legend

1. Click **Show Legend** check box to display the Legend. Uncheck to turn it off.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Show Cluster Points

1. Click **Show Cluster Points** to make the yellow nodes visible in the branches of the graph. To turn off the nodes (cluster points), uncheck Show Cluster Points.

Clear Selections

1. Click **Clear Selection**.

   All previously selected items are now cleared form the graphs.

Analysis summary view, Gene lists and Venn diagram

Analysis summary view

Figure 67 Summary Table example

Comparison	Group 1	Group 2	Count 1	Count 2	Up	Down
Liver vs Muscle	Liver	Muscle	4	4	5940	6165
Liver vs Spleen	Liver	Spleen	4	4	6412	11157
Liver vs Testes	Liver	Testes	4	4	12245	11950
Muscle vs Spleen	Muscle	Spleen	4	4	6919	11517
Muscle vs Testes	Muscle	Testes	4	4	12391	11704
Spleen vs Testes	Spleen	Testes	4	4	11594	9661

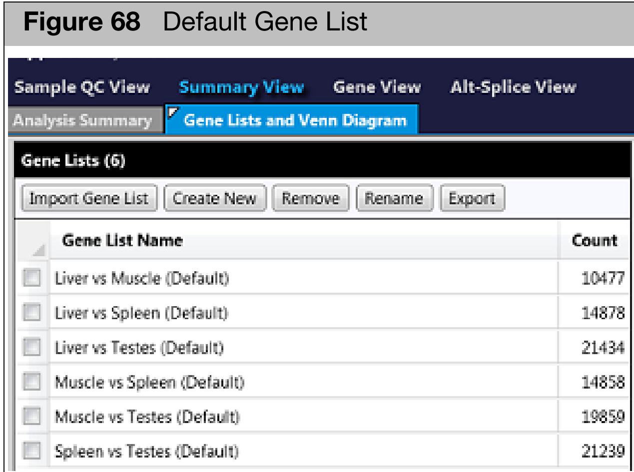
[1]The source of variation is a simple way to determine the fraction of the total variation of the samples can be explained by a given attribute. First, the variance of each probeset is computed, and the 1000 probesets having the highest variance are retained. Second, the total sum of squares () is accumulated for each attribute. Third, the residual sum of squares (, where the sum over j represents the sum over samples within the attribute level) is accumulated. Fourth, the fraction of variation explained for the probeset is. Finally, the fraction of variance explained for the attribute is the mean of the fraction explained over all of the probesets. This calculation is independent for each attribute, so the sum of the fractions associated with each of the attributes will not likely sum to 1. If the fraction of variance explained is large for a particular batch variable, it might improve the differential expression results to add this batch variable to the differential expression analysis.

Gene lists and Venn diagrams

A Default gene list is generated based on the comparison(s) run in the analysis.

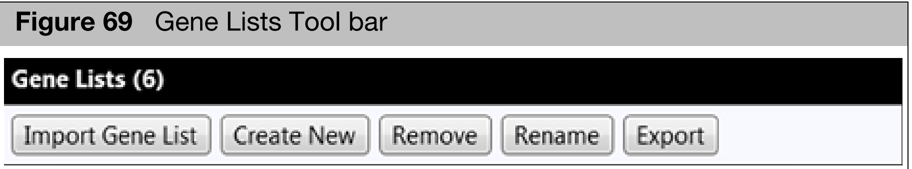
Figure 68 displays the default genes lists generated when comparing four tissue types. The count column indicates the number of genes differentially expressed in the comparison.

Figure 68 Default Gene List



Gene list tool bar

Figure 69 Gene Lists Tool bar



- **Import Gene List** - Enables you to import a previously generated gene list. Gene Lists should be saved as ***.gene_list.txt** files and contain a **probeset_id** and a **gene_symbol** column, as shown in Figure 70.

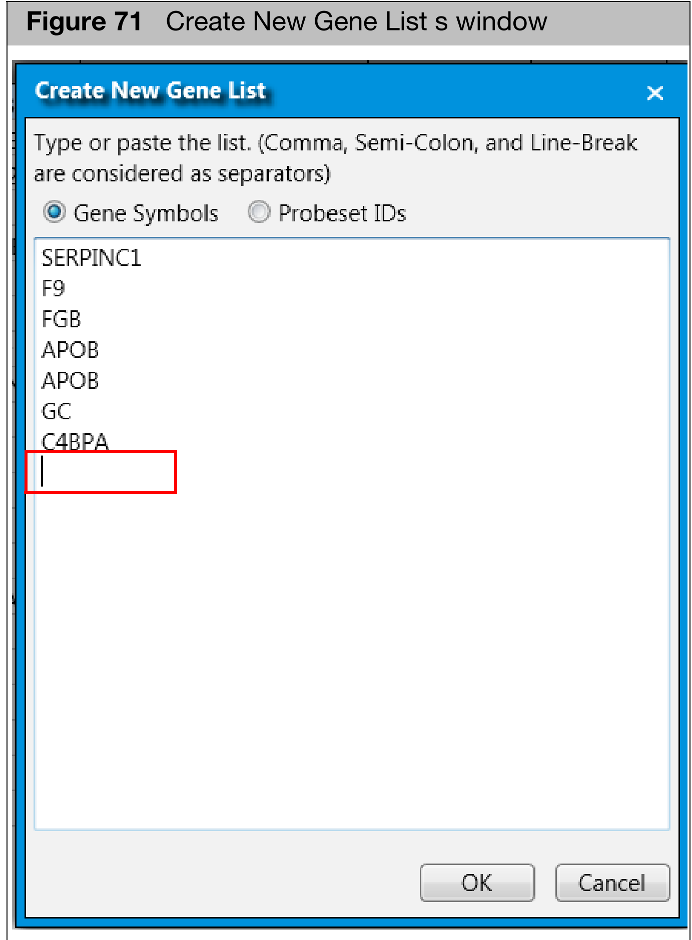
.

Figure 70 Probeset ID and Gene Symbol columns

probeset_id	gene_symbol
TC19000571.hg.1	CYP2B6
TC11002891.hg.1	APOC3
TC01006295.hg.1	CFHR2
TC02001628.hg.1	APOB
TC09001421.hg.1	ALDOB

- **Create New** - Enables you to create a new Gene List by directly typing or copying/ pasting either the Gene Symbol or the Probeset IDs into the Create New Gene List window (Figure 71).

Figure 71 Create New Gene List s window



1. Click either the **Gene Symbols** or **Probeset IDs** radio button, then enter your new Gene List name(s) directly under the last entry, as shown in Figure 71.
2. Click **OK**.

Your new entry or entries are now added to the Gene List Table, as **NewGeneList** or **NewGeneList_***.

- **Rename** - Enables you to rename any current Gene List.
  a. Highlight/Check the box for the Gene list to rename.
  b. Click **Rename**, the enter the new name in the window.
  c. Click **OK. A** message appears indicating the Gene List has been renamed. Acknowledge the message, then click **OK**.
- **Remove** - Enables you to remove a Gene List.
  a. Highlight/Check the box for the Gene list you want to remove.
  b. Click **Remove** to remove the Gene List.
- **Export** - Enables you to export a Gene List as a TXT file.
  a. Highlight/Check the box for the Gene List to be exported
  b. Click **Export**.
  c. Save the Gene List to the same Library folder you assigned earlier (in the Preferences Tab), then click **OK**.

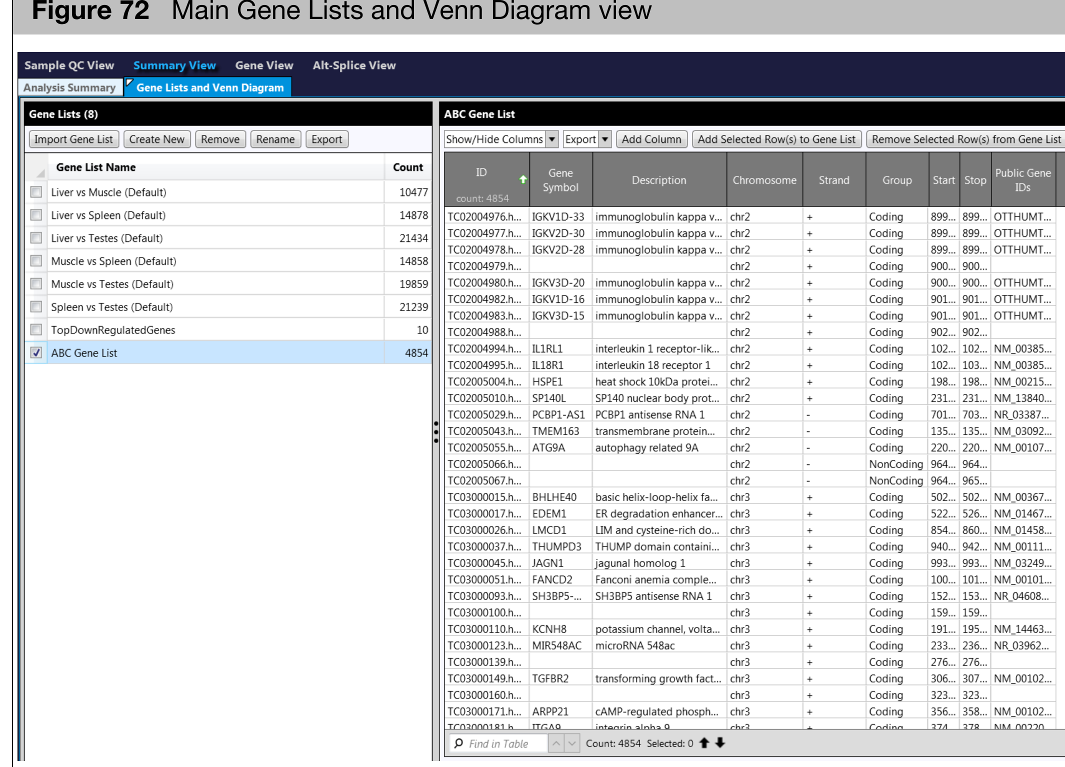
     The Gene List is now available (as an option) when running an EGA analysis.

For methods on creating Gene Lists in the Gene View or Alt-Splice View, see "Adding/ removing selected row(s) in a Gene list" on page 168.

Viewing the contents of a Gene list

Highlighting or selecting the check box of a single Gene List in the Gene List table displays the contents within the Gene List in the right hand panel, as shown Figure 72.

Figure 72 Main Gene Lists and Venn Diagram view



The Gene List displays the following columns. Use the Show/Hide Column drop down to display or hide any of these columns.

- TC ID
- Gene Symbol
- Description
- Chromosome
- Strand
- Group
- Start
- Stop
- Public Gene IDs

Adding a column to the table

The contents added to a new column in this table will only exist for this analysis in TAC. When setting up a new analysis for this same dataset, the contents in these new columns will not exist. Only New Columns added in the New Analysis Tab will be available for subsequent TAC analysis runs.

3. Click **Add Column**.

   A New Column window appears.

4. Name the column, then click **OK**.

   The new column is added to the far right of the Table.

   Optional: Click and drag the column to another position in the table. The column can be removed by right-clicking on the column header and selecting **Delete Column**.

5. Optional: Enter any custom annotation you want into each cell.

Adding/Removing selected row(s) in a Gene list

Adding from right-click menu

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Add selected row(s) to a Gene List** or click on the Gene List drop-down to select **Add Selected row(s) to a Gene List**.

   The Add to Gene List window appears.

3. Select a Gene List to add the selected row(s) from the drop-down or enter in the name of a new Gene list.

Removing from right-click menu

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Remove selected row(s) from a Gene List** or click on the Gene List drop-down to select **Remove Selected row(s) from a Gene List**.

   A Remove Gene List window appears.

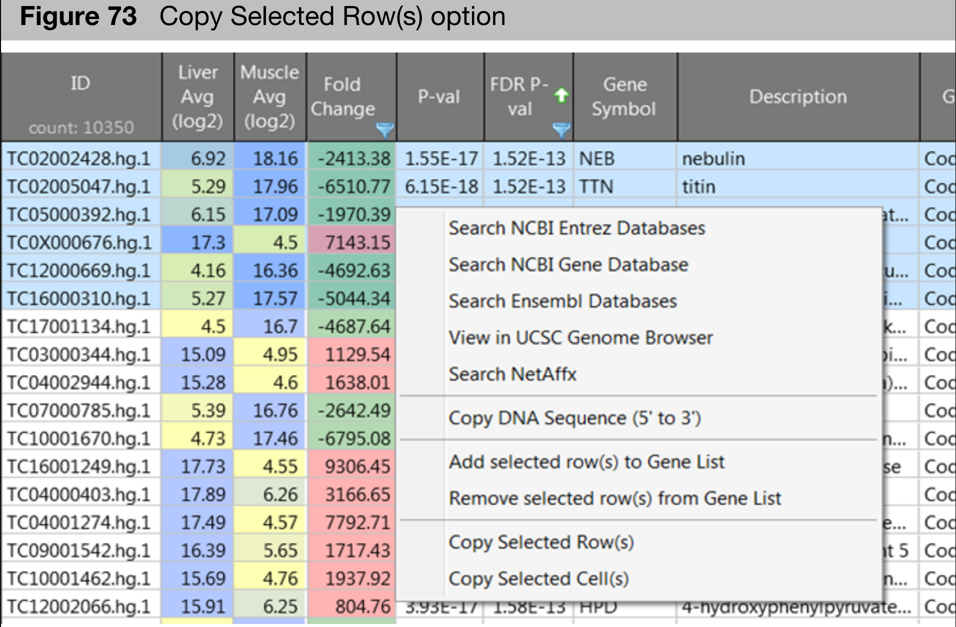
3. Use the drop-down to select the name of the Gene List (you want to remove the rows from), then click **OK**.

Copying selected rows and IDs

Copying selected Row(s)

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click to select **Copy Selected Row(s)**. (Figure 73)

Figure 73 Copy Selected Row(s) option



The selected gene level information is now copied to the Windows Clipboard for pasting.

Copying selected Cell(s)

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Right-click, then click **Copy Selected Cell(s)**.

   The selected cells are now copied to the Windows Clipboard for pasting.

Copying DNA sequence (5' to 3')

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Click **Copy DNA Sequence (5′ to 3′)**.

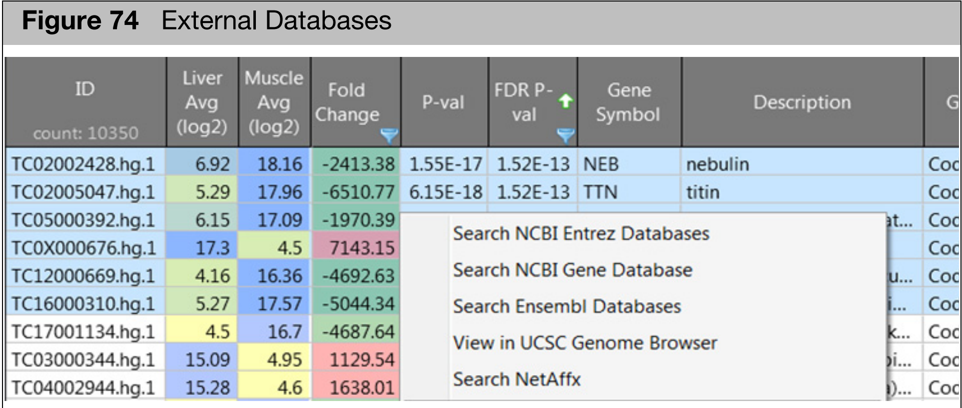
   The selected DNA Sequence (5′ to 3′) is now copied to the Windows Clipboard for pasting.

   Note: DNA sequences are obtained from server using start and stop positions. Sequences are represented from 5′ to 3′ for genes/exons/PSRs on both positive strand and negative strand.

Accessing external databases

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. To link out to various external databases, right-click on a TC ID of interest.
3. Click to select the external database you want to visit. (Figure 74)

Figure 74 External Databases



Your Internet browser opens to the appropriate website.

Searching the NetAffx website

1. Click **Search NetAffx**.

   The Internet browser opens to the NetAffx Customer Login window.

2. Enter your NetAffx **Email ID** and **Password**, then click **Submit**.

   Your Internet browser opens to the **NetAffx Query Center** and displays information about your gene of interest

   Note: The NetAffx Query Center is compatible with Windows Internet Explorer and Firefox. Chrome is not supported at this time. If a Probe Set or Transcript Cluster is not available, an appropriate message appears.

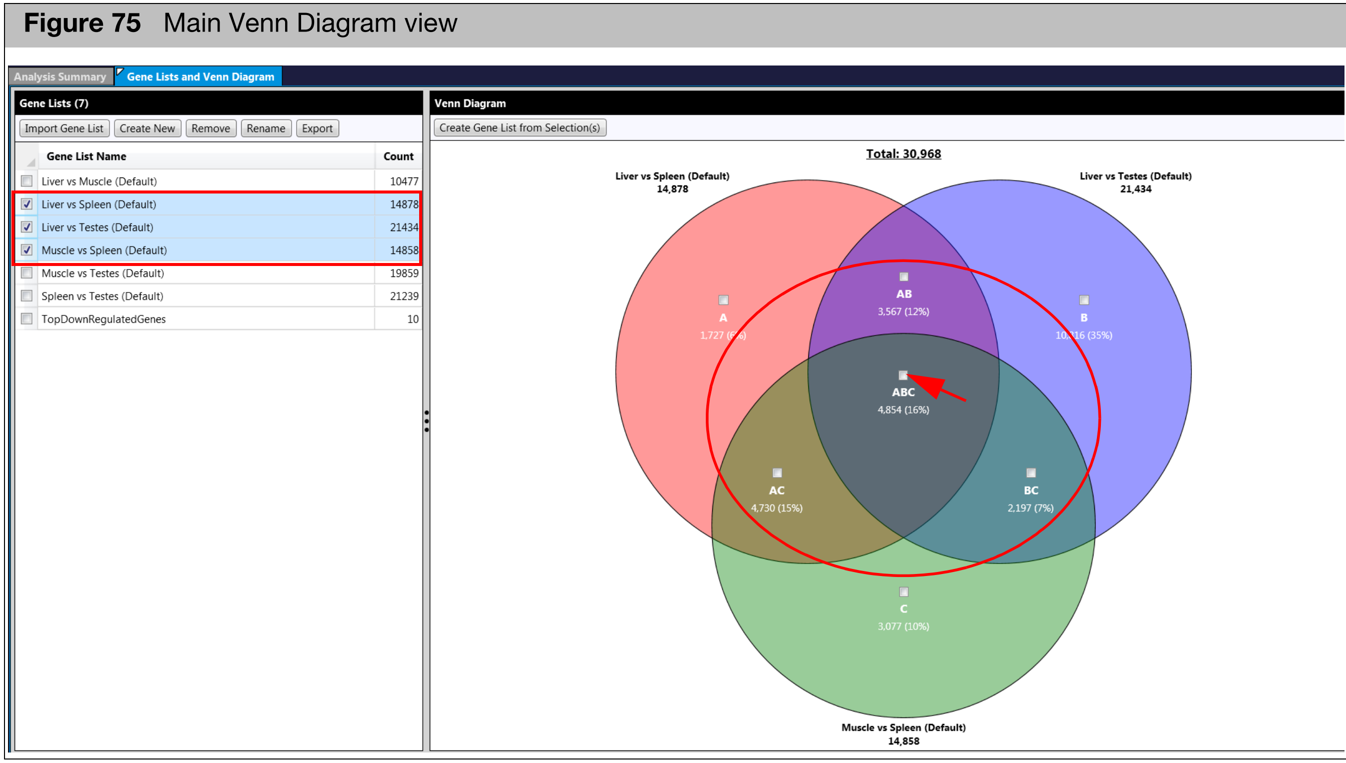
Venn Diagram

The Venn Diagram graph is a convenient way to see the intersection of genes between your comparison(s).

1. From the Gene Lists (left pane), click (2-3) Gene Lists check box(es) that you want to see intersected.

   The intersected Gene Lists appear in the Venn Diagram, as shown in Figure 75.

Figure 75 Main Venn Diagram view

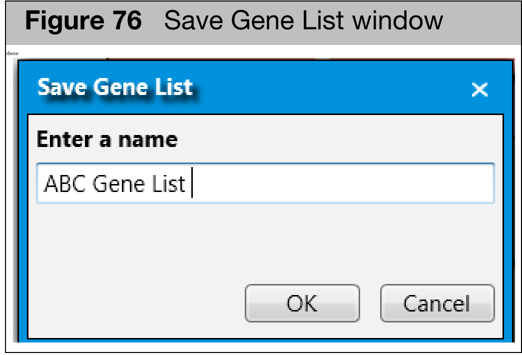


Exporting a Gene list within the Venn Diagram

1. Click the check box in the region of the Diagram that you would like to export. For example, If you want to export the list of genes common between all three comparisons, check the **ABC** check box (Figure 75).
2. Click the Venn Diagram's **Create Gene List from Selection(s)**.

   A Save Gene List window appears. (Figure 76)

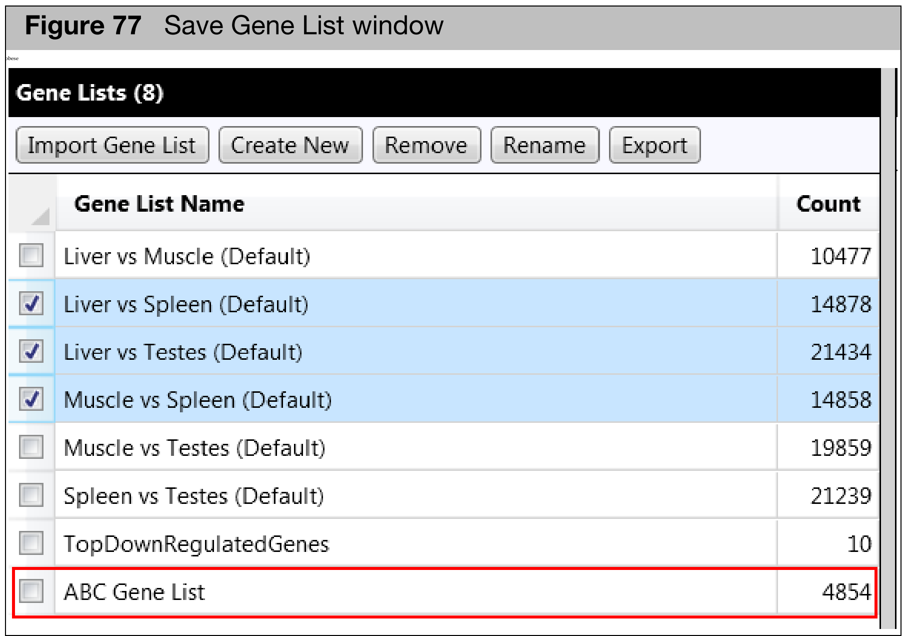
Figure 76 Save Gene List window



3. Enter a Gene List name, then click **OK**.

The Gene List (example ABC) is now displayed in the Gene List table, as shown in Figure 77.

Figure 77 Save Gene List window



Gene Lists (8)

Import Gene List | Create New | Remove | Rename | Export

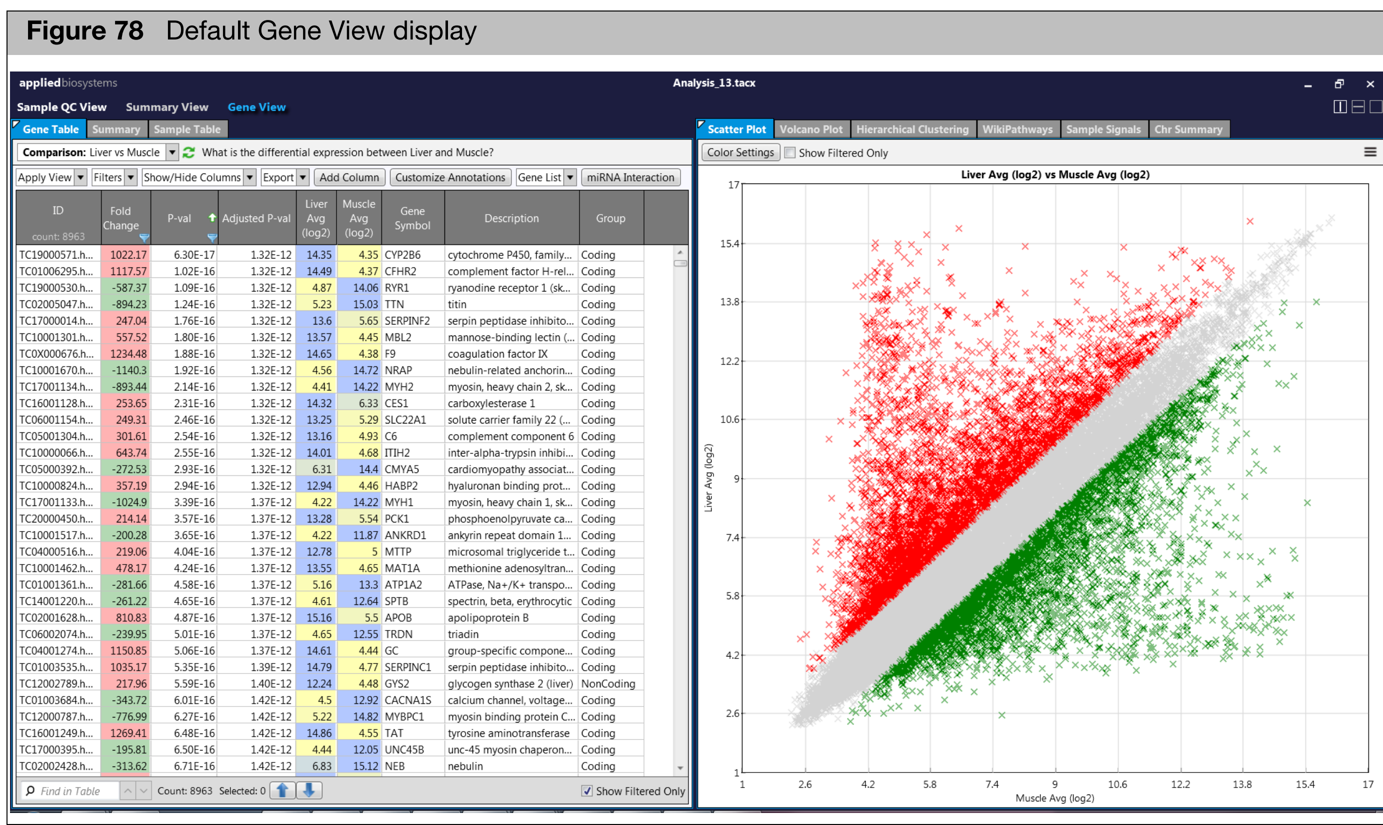
	Gene List Name	Count
☐	Liver vs Muscle (Default)	10477
☑	Liver vs Spleen (Default)	14878
☑	Liver vs Testes (Default)	21434
☑	Muscle vs Spleen (Default)	14858
☐	Muscle vs Testes (Default)	19859
☐	Spleen vs Testes (Default)	21239
☐	TopDownRegulatedGenes	10
☐	ABC Gene List	4854

Gene Lists created in this Summary View are also available in the Gene View and Alt-Splice View for filtering Gene Tables by Gene List.

Gene view

By default, the Gene Table and Scatter Plot panes appear in a side by side configuration, as shown in Figure 78.

Figure 78 Default Gene View display

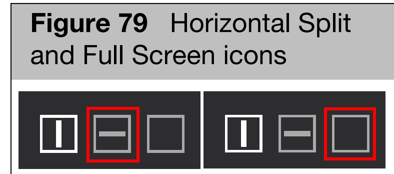


Viewing options

Changing the default view

- Click the **Horizontal Split** icon for a top and bottom configuration. (Figure 79)
- Click the **Full Screen** icon for individual window panes.

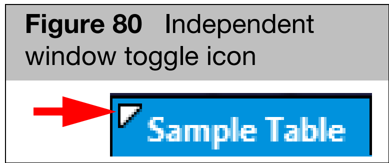
Figure 79 Horizontal Split and Full Screen icons



To toggle a tab window to full screen:

- Click on the tab's white triangle graphic. (Figure 80)

Figure 80 Independent window toggle icon



To toggle a full screen window BACK to its default tab window:

- Click on the window's **X** (as if you were to close the window).
  The window reverts back into the TAC Viewer's main window.

To change the size of a window pane:

- Click, hold, then drag the edge of the window pane to resize it.

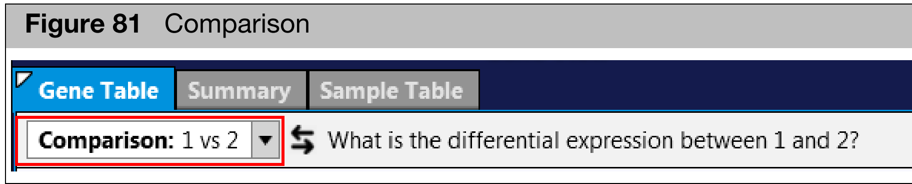
Common table functions

"Using the comparison feature" on page 94
"Customizing views" on page 95
"Working with column headers" on page 96
"Filtering column data" on page 99
"Managing filters" on page 100
"Table options" on page 101
"Accessing external databases" on page 103
"Searching the NetAffx website" on page 104
"Searching keywords" on page 104
"Exporting options" on page 104

Using the comparison feature

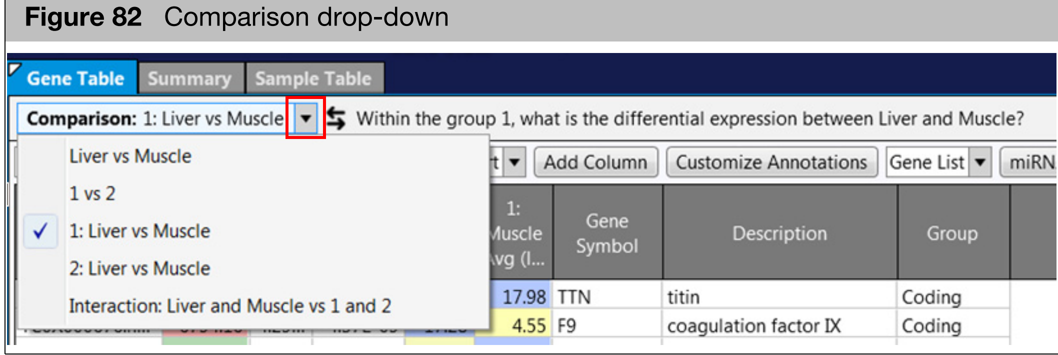
The comparison that was used in your completed analysis run is displayed at the top of the Gene Table pane, as shown in Figure 81.

Figure 81 Comparison



- For analyses that were run with multiple comparisons, click the **Comparison** drop-down. (Figure 82)

Figure 82 Comparison drop-down



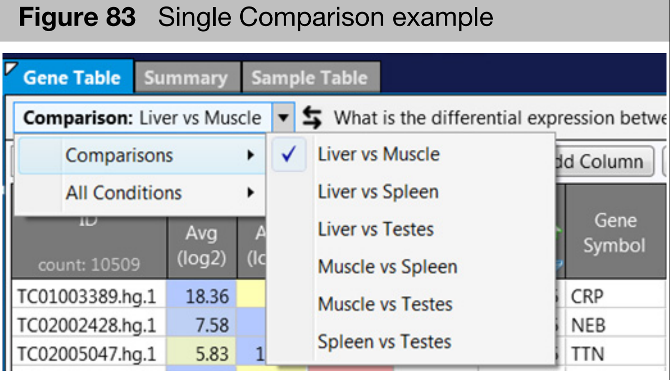
- For a Comparison analysis with more than two factor values (i.e. comparing 4 different tissue types), select to view one of the comparisons, as show in Figure 83.

Or

- Select **All Conditions** with the option to set one as the baseline, as shown in Figure 84.

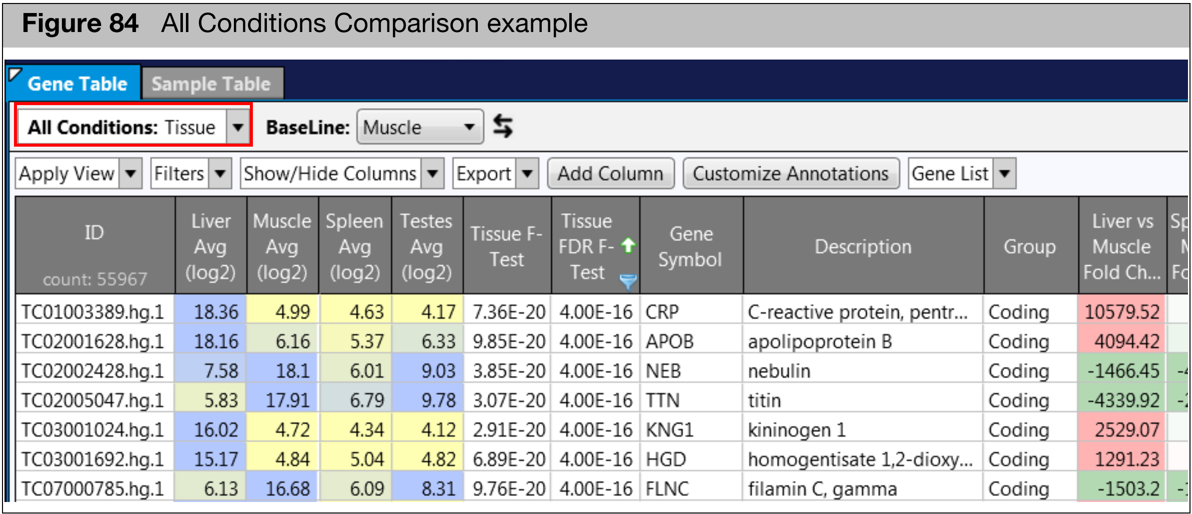
Note: Graphs displayed will depend on the Comparison selection. For Example when selecting All Conditions, the Graph options change to Signal Summary and Hierarchical Clustering. For details see "Hierarchical Clustering graph" on page 117 and "Sample Signals" on page 132.

Figure 83 Single Comparison example



- Click ⇆ to reverse the displayed conditions.

Figure 84 All Conditions Comparison example



ID (count: 55967)	Liver Avg (log2)	Muscle Avg (log2)	Spleen Avg (log2)	Testes Avg (log2)	Tissue F-Test	Tissue FDR F-Test	Gene Symbol	Description	Group	Liver vs Muscle Fold Ch...
TC01003389.hg.1	18.36	4.99	4.63	4.17	7.36E-20	4.00E-16	CRP	C-reactive protein, pentr...	Coding	10579.52
TC02001628.hg.1	18.16	6.16	5.37	6.33	9.85E-20	4.00E-16	APOB	apolipoprotein B	Coding	4094.42
TC02002428.hg.1	7.58	18.1	6.01	9.03	3.85E-20	4.00E-16	NEB	nebulin	Coding	-1466.45
TC02005047.hg.1	5.83	17.91	6.79	9.78	3.07E-20	4.00E-16	TTN	titin	Coding	-4339.92
TC03001024.hg.1	16.02	4.72	4.34	4.12	2.91E-20	4.00E-16	KNG1	kininogen 1	Coding	2529.07
TC03001692.hg.1	15.17	4.84	5.04	4.82	6.89E-20	4.00E-16	HGD	homogentisate 1,2-dioxy...	Coding	1291.23
TC07000785.hg.1	6.13	16.68	6.09	8.31	9.76E-20	4.00E-16	FLNC	filamin C, gamma	Coding	-1503.2

All conditions (like the Tissue type example in the Figure 84) are listed in the Gene Table.

In Figure 84, one Condition (Muscle) is set as the baseline which provides the fold change columns for the remaining conditions in comparison with the baseline.

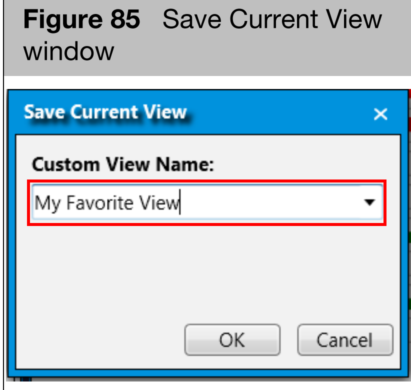
Customizing views

Custom Views can be created to quickly access the columns in the order you want to see them using the filter parameters you prefer.

1. Select the columns you would like in your custom view
2. Select your custom view's column order.
3. Set any Filter parameters.
4. Click the Apply View drop-down, then select Save Current View.

   The Save Current View window appears.
5. Type in a new name for the Custom View, as shown in Figure 85, then click **OK**.

Figure 85 Save Current View window



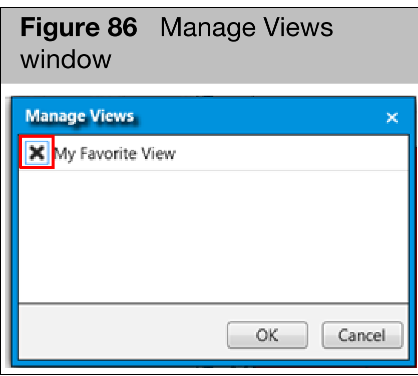
Your custom view is now available within the Table's **Apply View** drop-down menu.

Removing a custom view

1. Click **Manage Saved Views**.

   The Manage Views window appears. (Figure 86)

Figure 86 Manage Views window



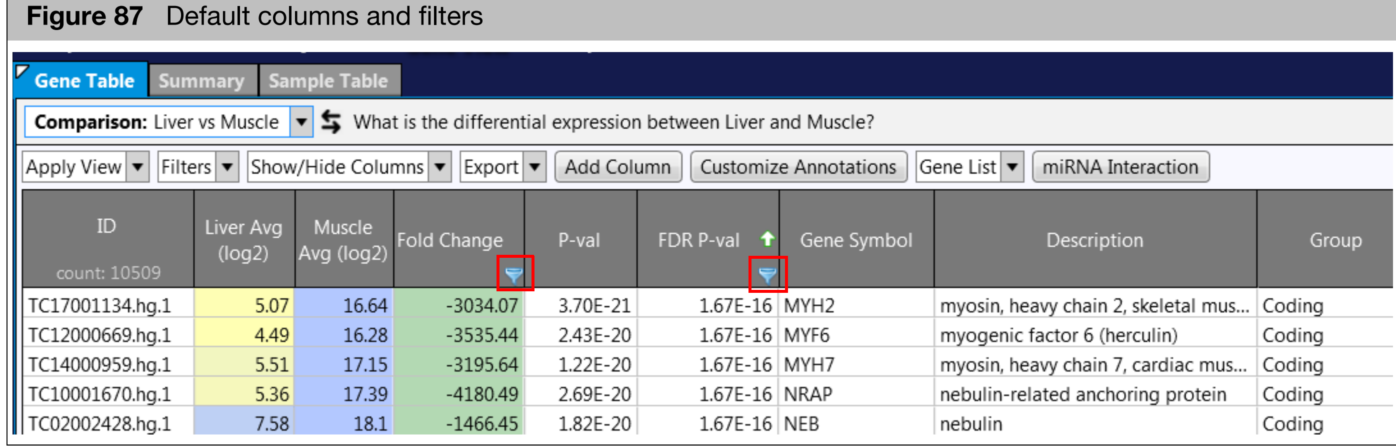
2. Click the **X** adjacent to the View name, then click **OK**.

   The Custom View is now removed from the Apply View drop-down menu.

Working with column headers

The factory default columns and preset filters for the Gene Table are as shown in Figure 87. For the definitions of these columns, see the table under "Showing or hiding columns" on page 97.

Figure 87 Default columns and filters



Rearranging default columns

1. Click on a column you want to move, then drag it (left or right) to its new location.
2. Release the mouse button.

   The column is now in its new position.

Showing or hiding columns

1. Click the **Show/Hide Columns** drop-down menu.
2. Click the check box adjacent to the column you want to show in the table. Uncheck the adjacent check box to the column you want to hide.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

Available Columns	Description
ID	ID of Transcript Cluster (TC)
Fold Change	Displays the fold change (in linear space) of Condition 1 vs Condition 2
Delta Fold Change	Change in the component fold changes associated with the interaction. If you convert the displayed component fold changes to a fractional value (fractional_fold_change = displayed_fold_change if displayed_fold_change >= 1 and 1 / |displayed_fold_change| if displayed_fold_change <= -1), the fractional_delta_fold_change = fractional_fold_change(condition 1) / fractional_fold_change(condition 2). The displayed_delta_fold_change = fractional_delta_fold_change if fractional_delta_fold_change >=1 and -1 / fractional_delta_fold_change if fractional_delta_fold_change <1. **Note:** This column is only shown for Interaction Comparisons.
P-val	Probability that there is no differential expression for this PSR/JUC between these conditions.
Adjusted P-val	FDR adjusted version of preceding where FDR only includes cases where TC is expressed in both conditions and PSR/JUC is expressed in at least one of the conditions.
Condition 1 Expressed	True or false indication of whether this PSR/JUC is expressed in condition 1 (based on percentage of samples that meets the DABG cutoff specified).
Condition 2 Expressed	True or false indication of whether this PSR/JUC is expressed in condition 2 (based on percentage of samples that meets the DABG cutoff specified).
Condition 1 Avg (log2)	Tukey biweight average of the expression levels of this PSR/JUC for samples from condition 1.
Condition 2 Avg (log2)	Tukey biweight average of the expression levels of this PSR/JUC for samples from condition 2.
Condition 1 Standard Deviation	Standard deviation of the expression levels of this PSR/JUC for samples from condition 1.
Condition 2 Standard Deviation	Standard deviation of the expression levels of this PSR/JUC for samples from condition 2.
Public Gene IDs	Public Gene IDs for this TC.
Gene Symbol	Gene symbol for this transcript cluster. **Note:** RefSeq gene symbol is listed as the first gene symbol (if there are more than 1 gene symbol). Also, a TC with no gene symbol, may be auto-assigned a public gene ID.
Description	Gene Description for this Transcript Cluster.
Chromosome	Chromosome for this transcript cluster. See Chromosome Naming Scheme[3] for a detailed description.
Strand	Describes the (+/-) strand location of the transcript cluster in the version of the genome assembly used at the annotation time.
Group	Whether this TC is coding or non-coding. See Table 2 on page 189 for the different options within Group and their definitions.

Available Columns	Description
Start	The beginning genomic position of the transcript cluster.
Stop	The ending genomic position of the transcript cluster.
Gene Expressed in Both Conditions	True (T) or False (F) indicator as to whether the Gene is expressed in both Conditions.
Target Data Range	After the Interaction Network tab/window tab is generated, a Target Data Range column appears in both the main (left) table and the target table (bottom right). The Target Data Range column displays all the genes or miRNAs that passed the filter criteria associated with a particular miRNA or mRNA and how many have a fold change within each of the 9 specified fold change ranges. For more information, see Figure 151 on page 144.

Sorting columns

1. Select a column, then right-click on it.

   A menu appears.
2. Click to select either **Sort By Ascending** (A-Z) or **Sort By Descending** (Z-A).

Double-Click Sorting Method

1. Double-click on a column header to sort its data in an ascending order. Double-click on the same column header to sort its data in a descending order.

Adding a column to the table

1. Click **Add Column**.

   A New Column window appears.
2. Name the column, then click **OK**.

   The new column is added to the far right of the Table.

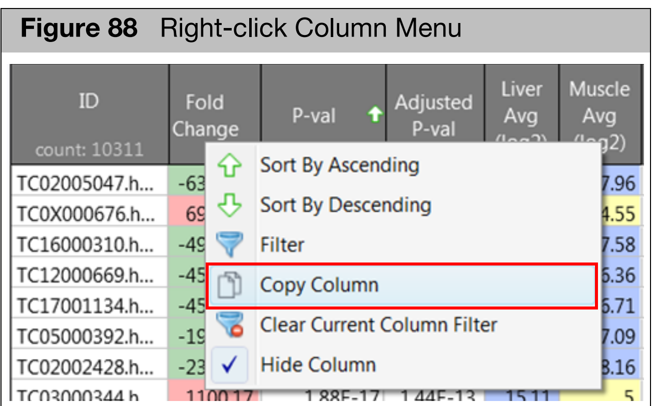
   Optional: Click and drag the column to another position in the table. The column can be removed by right-clicking on the column header and selecting **Delete Column**.
3. Optional: Enter any custom annotation you want into each cell.

Copying column data to your clipboard

1. Select a column you want to copy to a clipboard, then right-click on it.

   The following menu appears: (Figure 88)

Figure 88 Right-click Column Menu



2. Click **Copy Column**.

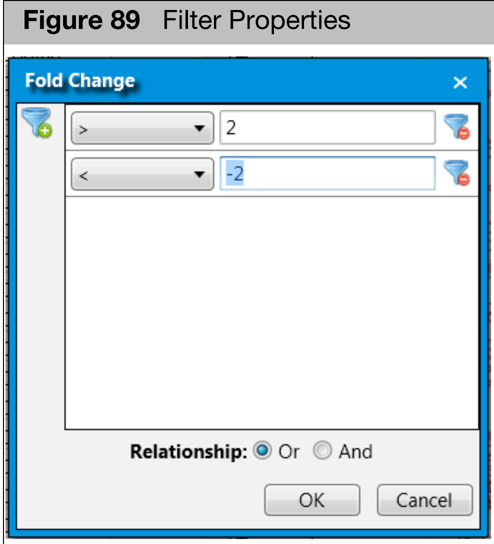
   The column data is now ready for pasting (Ctrl v).

Filtering column data

All tables can be filtered.

1. Select a column, then right-click on it.
   A menu appears.
2. Click **Filter**.
   The following window appears (Fold Change column example shown): Figure 89

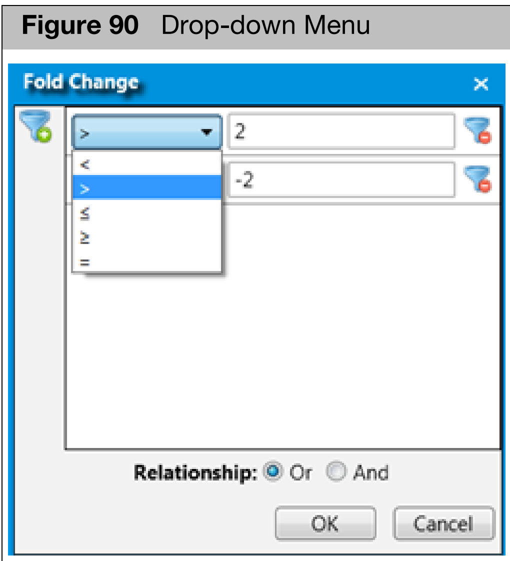
Figure 89 Filter Properties



Editing filtering properties

1. Click the **Or** or **And** button to choose **Or** or **AND** logic.
2. Click the symbol drop-down menu(s) to select new symbol(s), as shown in Figure 90.

Figure 90 Drop-down Menu



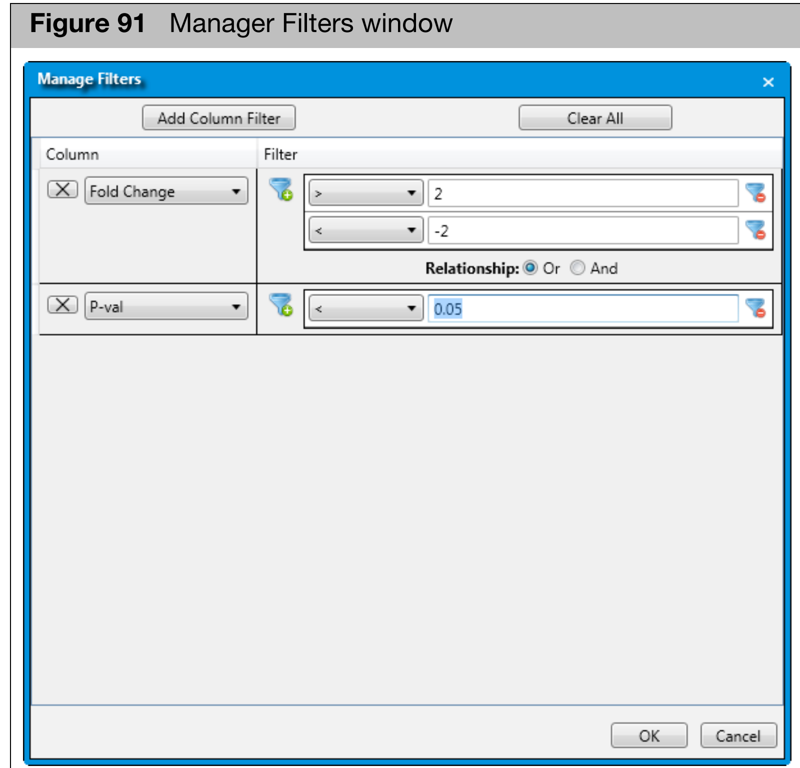
3. Click inside the numbering field(s) 2 to enter new value(s).
4. Click to add filter(s).
5. Click to remove filter(s).

Managing filters

1. Click the Filters drop-down, then select **Manage Filters**.

   The Manage Filters window appears. (Figure 91)

Figure 91 Manager Filters window



2. Click **Add Column Filter**.

   A new column drop-down field appears.
3. Select a name from the drop-down.
4. Select the operator and enter the value to filter on.
5. Optional: Click to add filter(s).

   Click the **Or** or **And** radio button to choose Or or AND logic.
6. Optional: Click to remove filter(s) or click on the **X** to the left of the row to remove that filter.
7. Click **OK**.

   To remove all displayed filters in the Manage Filters window, click **Clear All** or click **Clear All Filter(s)** from the Table's Filters drop-down.

Table filters

The Table can be filtered to either hide or show highlighted rows.

- From the Filters drop down:
  - Select **Filter Table by Selection** to display only highlighted rows in the table
  - Select **Filter Table by Selection (Exclude)** to hide highlighted rows while displayed all unselected rows.

Clearing filters

1. Right-click on the filtered column you want to clear.

   A menu appears.

2. Click **Clear Current Column Filters**.

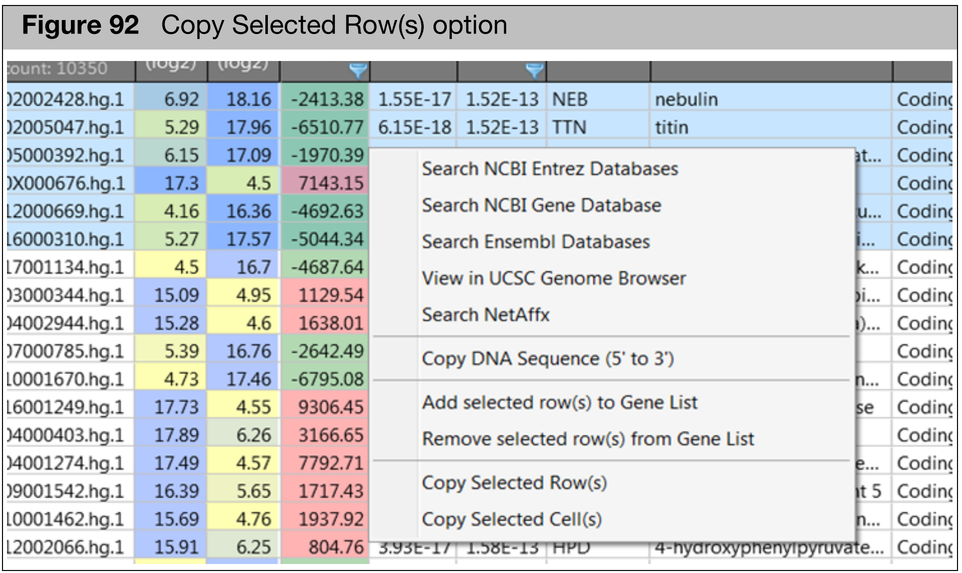
   The filter is removed.

Table options

Copying selected row(s)

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click to select **Copy Selected Row(s)**. (Figure 92)

Figure 92 Copy Selected Row(s) option



The selected gene level information is now copied to the Windows Clipboard for pasting.

Copying selected cell(s)

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Right-click, then click **Copy Selected Cell(s)**.

   The selected cells are now copied to the Windows Clipboard for pasting.

Copying DNA sequence (5’ to 3’)

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Click **Copy DNA Sequence (5′ to 3′)**.

   The selected DNA Sequence (5′ to 3′) is now copied to the Windows Clipboard for pasting.

   Note: DNA sequences are obtained from server using start and stop positions. Sequences are represented from 5′ to 3′ for genes/exons/PSRs on both positive strand and negative strand.

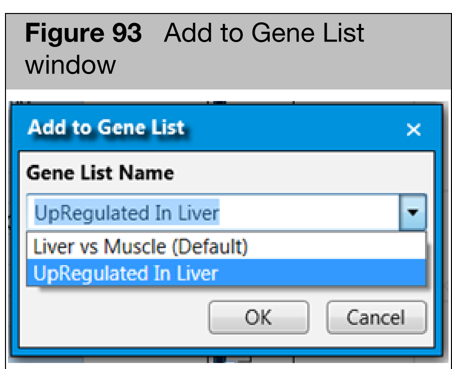
Adding/removing selected row(s) in a Gene list

Adding from right-click menu

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Add selected row(s) to Gene List**.

   The Add to Gene List window appears. (Figure 93)

Figure 93 Add to Gene List window



3. Select a Gene List to add the selected row(s) from the drop-down or enter in the name of a new Gene list.

Removing from right-click menu

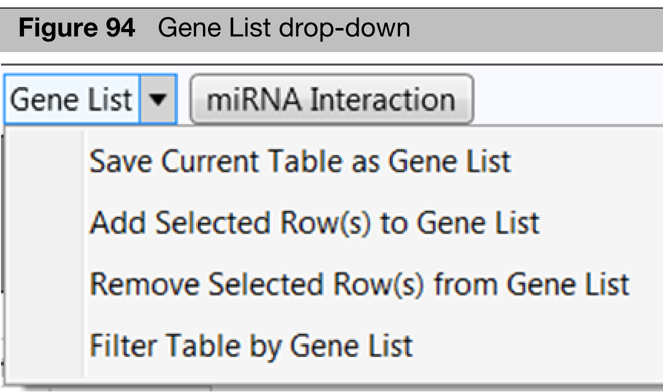
1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Remove Selected row(s) from a Gene List** or click on the Gene List drop-down to select **Remove Selected row(s) from a Gene List**.

   A Remove Gene List window appears.
3. Use the drop-down to select the name of the Gene List (you want to remove the rows from), then click **OK**.

From Gene list drop-down

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Click the **Gene List** drop-down, as shown in Figure 94.

Figure 94 Gene List drop-down



Saving the current table as a Gene list

1. Click **Save the Current Table as Gene List**.

   A Save Gene List window appears.
2. Enter a name, then click **OK**.

Adding selected row(s) to a Gene list

1. Click **Add Selected Row(s) to Gene List**.

   An Add to Gene List window appears.
2. Use the drop-down to select an existing name or enter a new name in the provided text field.
3. Click **OK**.

Removing selected row(s) from a Gene list

1. Click **Remove Selected Row(s) from Gene List**.

   A Remove from a Gene List window appears.
2. Use the drop-down to select the name you want to remove, then click **OK**.

Filtering table by Gene list

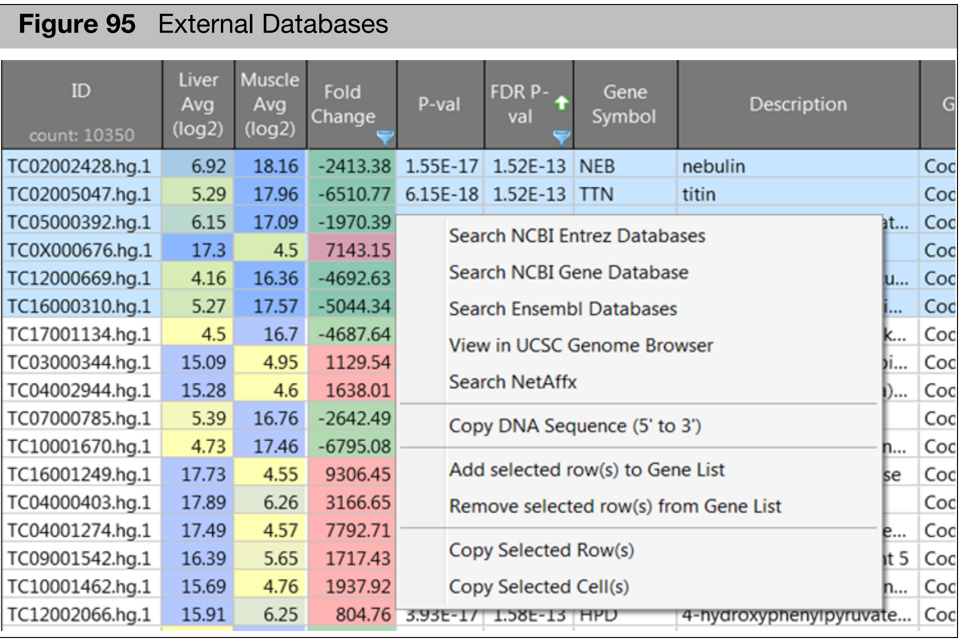
To show information pertaining only to rows in a Gene List:

1. Click **Filter Table by Gene List**.
2. Select the Gene List you want to filter from the Filter Table by Gene List window, then click **OK**.

Accessing external databases

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. To link out to various external databases, right-click on a TC ID of interest.
3. Click to select the external database you want to visit. (Figure 95)

Figure 95 External Databases



Your Internet browser opens to the appropriate website.

Searching the NetAffx website

1. Click **Search NetAffx**.

   The Internet browser opens to the NetAffx Customer Login window.

2. Enter your NetAffx **Email ID** and **Password**, then click **Submit**.

   Your Internet browser opens to the **NetAffx Query Center** and displays information about your gene of interest

   Note: The NetAffx Query Center is compatible with Windows Internet Explorer and Firefox. Chrome is not supported at this time. If a Probe Set or Transcript Cluster is not available, an appropriate message appears.

Searching keywords

By default, the search tool returns matches that contain your (case insensitive) search inquiry. Use wild-card (*) characters to aid in your search. Example: ABC* = Any string that begins with "ABC". Use (") to search for exact cell content matches. Example: "ABC"

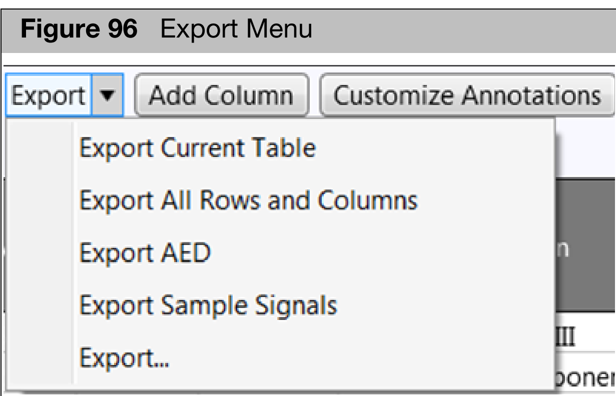
1. Click inside the **Find in Table** field, then type your keyword.
2. Click the [down] or [up] buttons to search.

Exporting options

If you want to export (Save) your table, click the **Export** drop-down.

The Export menu options appears. (Figure 96)

Figure 96 Export Menu



Exporting the current table

1. Click **Export Current Table** to export only the data currently shown in the Gene Table.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

Exporting all rows and columns

1. Click **Export All Rows and Columns** to export all the data in the Gene Table including hidden data.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

Exporting AED

Affymetrix Extensible Data (AED) files contain data that annotate positions on a genome and can be edited using Chromosome Analysis Suite (ChAS) software. For more information on AED files, see the ChAS User Manual (P/N 702943).

1. Click **Export AED**.
2. The Save As window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the file, then click Save. The table is now saved as a AED file.

Exporting sample signals

To export the signal for each transcript cluster ID per sample as a txt file:

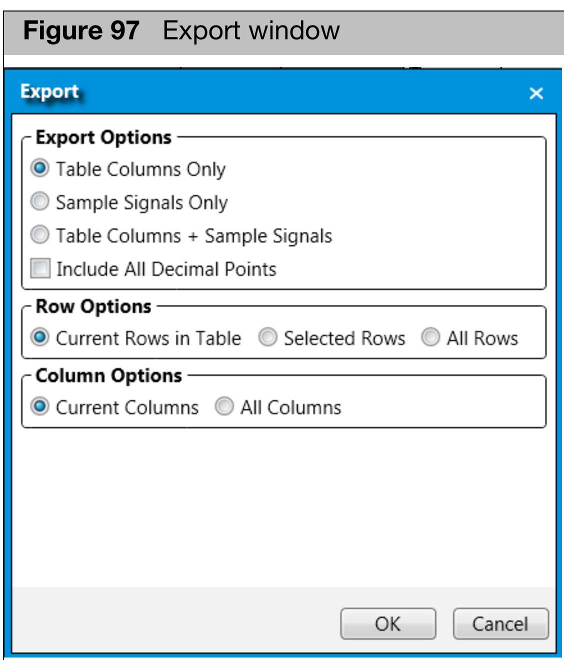
1. Click **Export Sample Signals**.

   The Save as window appears.
2. Click on an existing folder or click **New Folder** to choose a new save location.
3. Type a filename for the TXT file, click **Save**.

Custom export

1. To customize the export based on either current or all rows and/or columns click **Export…**
2. The Export window appears. (Figure 97)

Figure 97 Export window



3. Select a **Export Options** radio button:
   – **Table Columns Only**: Exports the currently visible table columns.
   – **Sample Signals Only**: Exports the currently visible samples.
   – **Table Columns + Sample Signals**: Exports currently visible table columns and samples.
4. Optional: Click the **Include All Decimal Points** check box to export all decimal points available for each metric.

   Note: If unchecked, decimal values appear as two digits.

5. Select a **Row Options** radio button:
   - **Current Rows in Table**: Exports the currently visible table rows.
   - **Selected Rows**: Exports the currently selected (highlighted) rows.
   - **All Rows**: Exports all rows visible and highlighted.
6. Select a **Column Options** radio button:
   - **Current Columns**: Exports currently displayed columns.
   - **All Columns**: Exports all columns (displayed or not).
7. Click **OK**.

   The Save as window appears.
8. Click on an existing folder or click **New Folder** to choose a new save location.
9. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

   Optional: Click and drag the column to another position in the table. The column can be removed by right-clicking on the column header and selecting **Delete Column**.

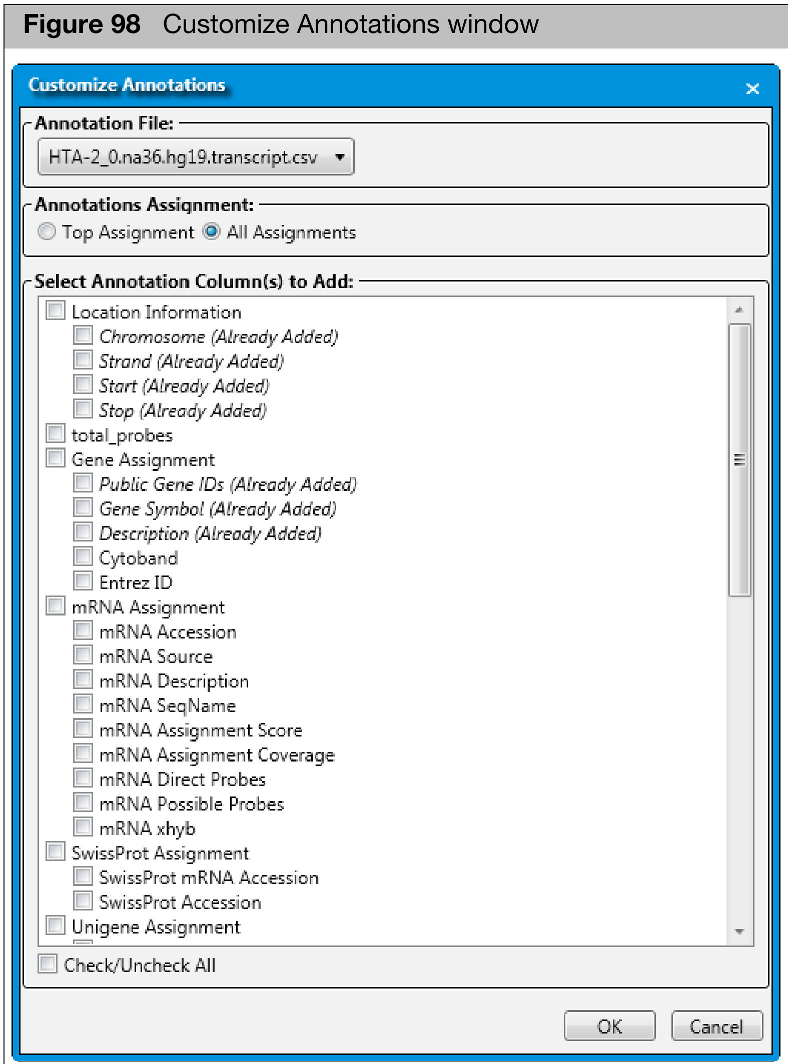
   Note: The Gene Table's default columns can NOT be deleted.

Customizing annotations

1. To customize the Gene Table's annotation column(s), click Customize Annotations .

The Customize Annotations window appears. (Figure 98)

Figure 98 Customize Annotations window



Selecting an annotation file

Click the drop-down button to view and select available Annotation Files.

Selecting an Annotation Assignment

1. Click either the Top Assignment or All Assignments button to select your table/column display preference.
   – **Top Assignment** is the first subfield in an annotation column.
   – **All Assignments** displays all the subfields in an annotation column.

Adding Annotation Columns

1. Click the check box next to the annotation column(s) you want to add. Uncheck the check box to remove a specific column(s).

   Note: Column descriptions marked as ***Already Added*** denote the column already exists in your current table. No action (check mark) is required.

Check/Uncheck All

1. Click the Check/Uncheck All check box to select or deselect all listed Annotations.

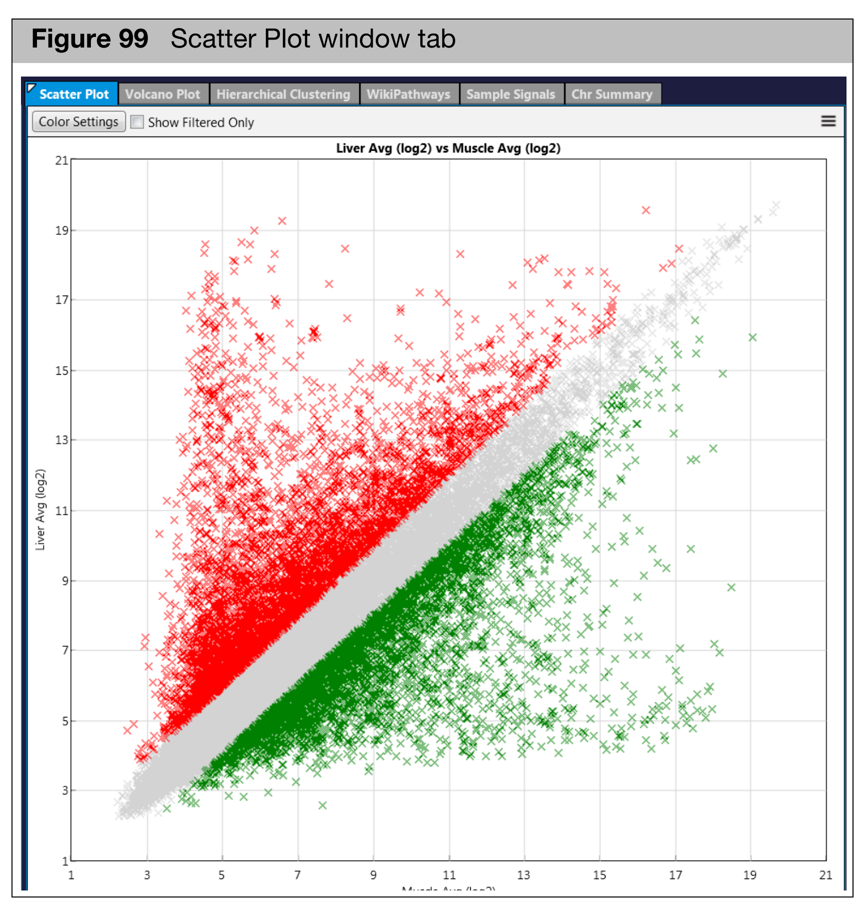
Gene Level Differential Expression analysis graphs

- **Condition Pairings** generate six graphs:
  - Scatter Plot
  - Volcano Plot
  - Hierarchical Clustering
  - Wiki Pathways
  - Sample Signal
  - Chromosome Summary (for select arrays)
- An **All Conditions Comparison** generates two graphs:
  - Sample Signals
  - Hierarchical Clustering

Scatter Plot graph

- The Scatter Plot (Figure 99) is a standard scatter plot graph of your current condition pairing.
- The signal is log2 based data. The scale is also log2 based scale for signals.
- X axis is second condition. The Y axis is first condition on the top left of the table.
- The gray TCs are the ones filtered out by the table. The green TCs are the ones down-regulated in Liver (Condition1) vs. Muscle (Condition2). The red TCs are the ones up-regulated in Liver (Condition1) vs. Muscle (Condition2).
- If you switch the condition pair or the filter criteria in the table, the data in the graph changes accordingly.

Figure 99 Scatter Plot window tab



Lassoing Genes of Interest within the Scatter Plot

The Lasso tool enables you to select genes of interest inside the Scatter Plot.

To lasso a gene or a group of genes, hold down the left mouse button and use the cross hair cursor to encircle the gene(s) of interest.

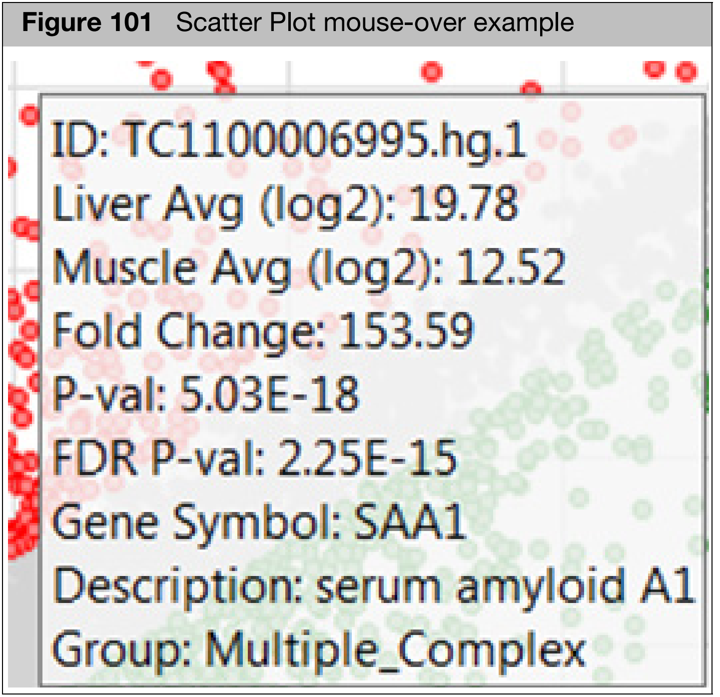
After closing the loop around your samples of interest, the genes of interest are highlighted in magenta and the data for these genes are also highlighted (blue) inside the table view, as shown in Figure 100.

Figure 100 Scatter Plot with lassoed genes example

Obtaining information related to individual probe sets

1. Mouse over (position the cursor over) a probe set to displays its details, as shown in Figure 101.

Figure 101 Scatter Plot mouse-over example



Scatter Plot graph options

1. Click the **Options** button (upper right).

   The Options menu appears.

Save as PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

Print

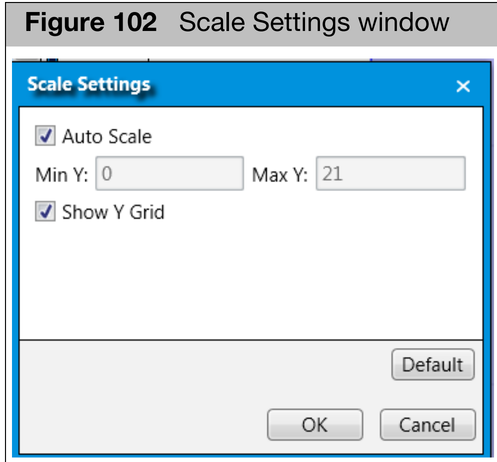
1. Click **Print**, then print the currently displayed graph using your configured printer as you normally would.

Scale Settings

1. Click **Scale Settings**.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box (Figure 102) to enable both the X and Y Min and Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show X Grid** and/or **Show Y Grid** check box(es) to display X/Y grid lines in the graph.

5. Optional: Click **Default** to return to the factory default settings.

Figure 102 Scale Settings window

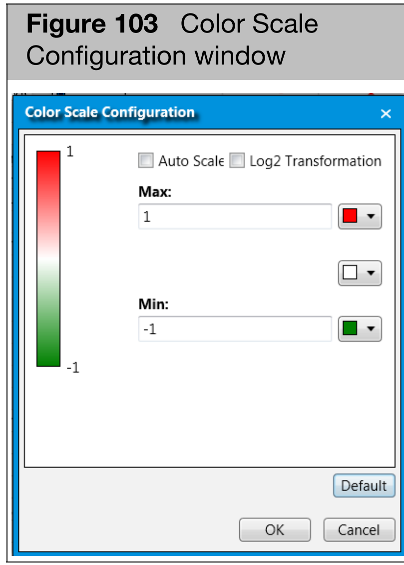


Color Settings

1. Click the **Color Settings**.

   The Color Scale Configuration window appears. (Figure 103)

Figure 103 Color Scale Configuration window



2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation.

   Note: Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.

3. Use the color drop-down menus to assign new colors for your data points.
4. Click **OK** to save your changes or click **Default** to return to the factory settings.

Show Legend

1. Click **Show Legend** check box to display the Legend. Uncheck to turn it off.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Clear Selections

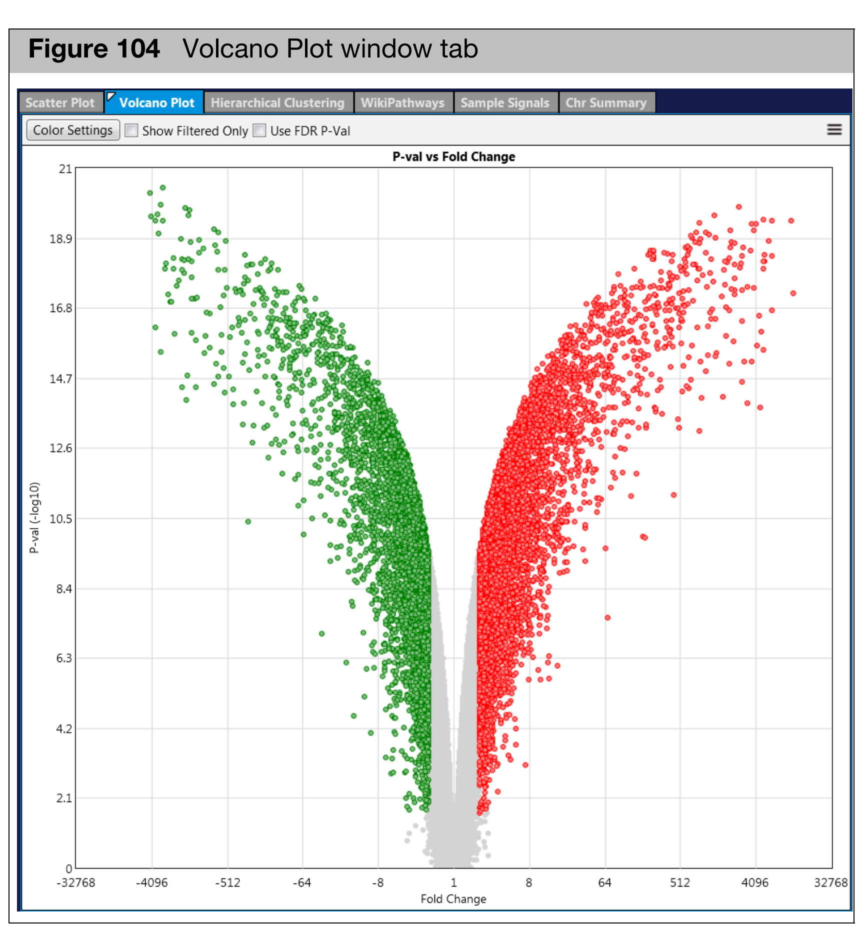
1. Click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

Volcano Plot graph

- The Volcano Plot graph (Figure 104) is a type of scatter plot that is used to quickly identify changes in large datasets.
- It plots significance versus fold-change on the y- and x-axes, respectively.
- X axis is the linear fold change from current condition pair; Y axis is -log10 p-value of the ANOVA p-values.
- In the example below (Figure 104), The gray TCs are the ones filtered out by the table. The green TCs are down-regulated in Liver (Condition1) vs. Muscle (Condition2), while the red TCs are represented as up-regulated in Liver (Condition1 vs. Muscle (Condition2).
- If you switch the condition pair or the filter criteria in the table, the data in the graph will change accordingly.

Figure 104 Volcano Plot window tab



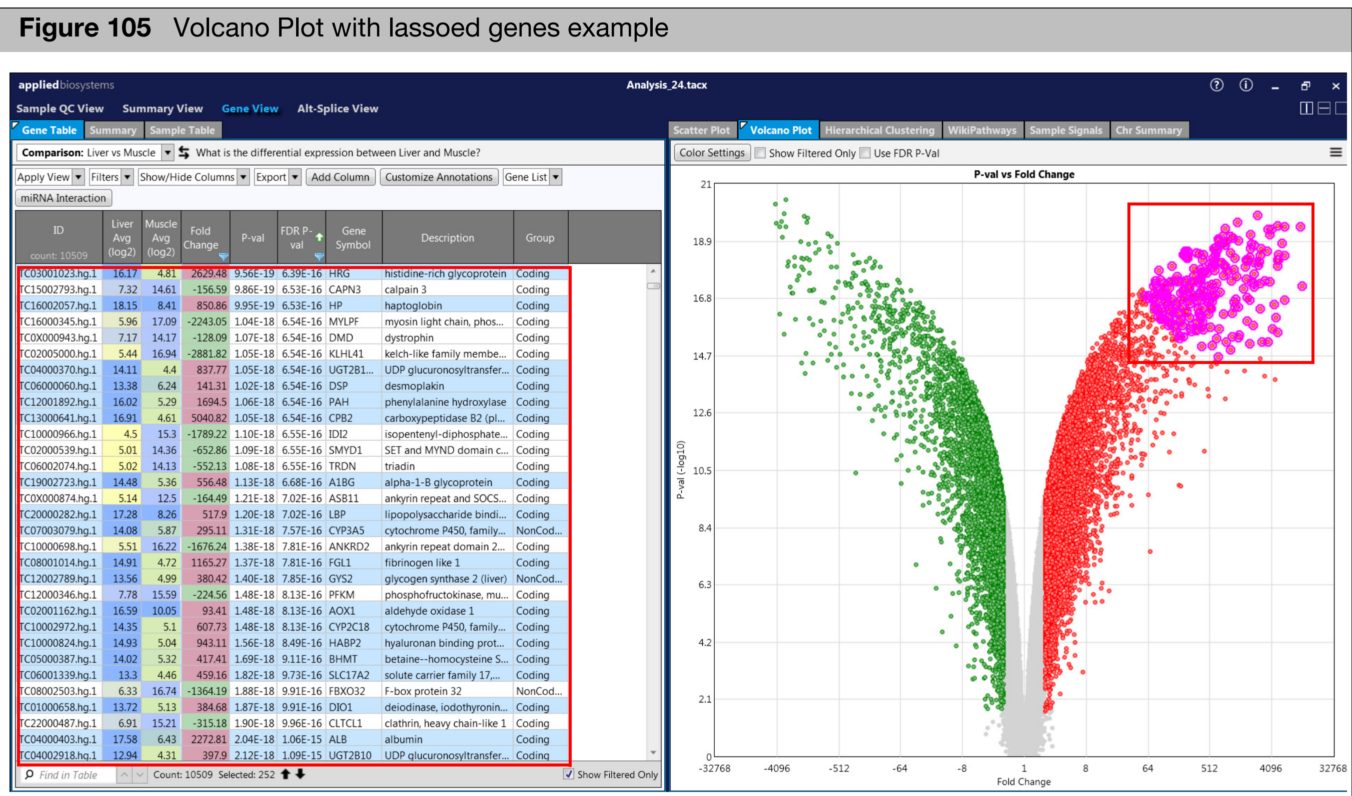
Lassoing genes of interest within the Volcano Plot

The Lasso tool enables you to select genes of interest inside the Volcano Plot.

To lasso a gene or a group of genes, hold down the left mouse button and use the cross hair cursor to encircle the gene(s) of interest.

After closing the loop around your samples of interest, the genes of interest are highlighted in blue and the data for these genes are also highlighted (blue) inside the table view, as shown in Figure 105.

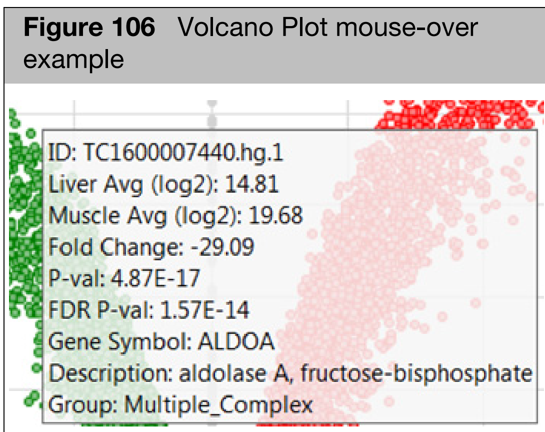
Figure 105 Volcano Plot with lassoed genes example



Obtaining Information Related to Individual Probe Sets

1. Mouse over (position the cursor over) a probe set to displays its details, as shown in Figure 106.

Figure 106 Volcano Plot mouse-over example



Volcano Plot graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

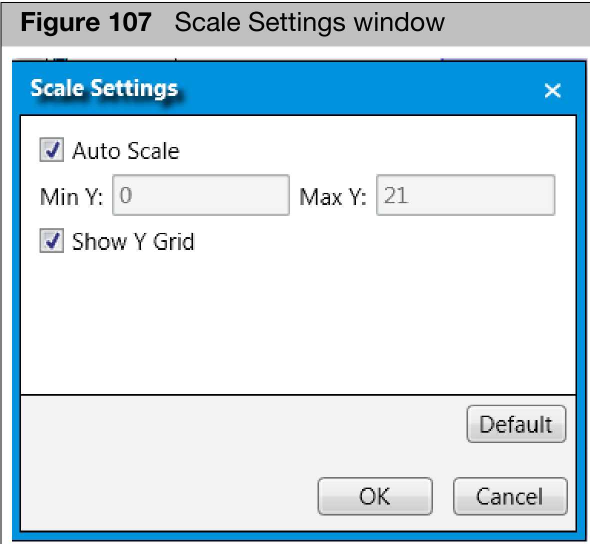
Print

Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box (Figure 107) to enable both the X and Y Min and Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show X Grid** and/or **Show Y Grid** check box(es) to display X/Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Figure 107 Scale Settings window

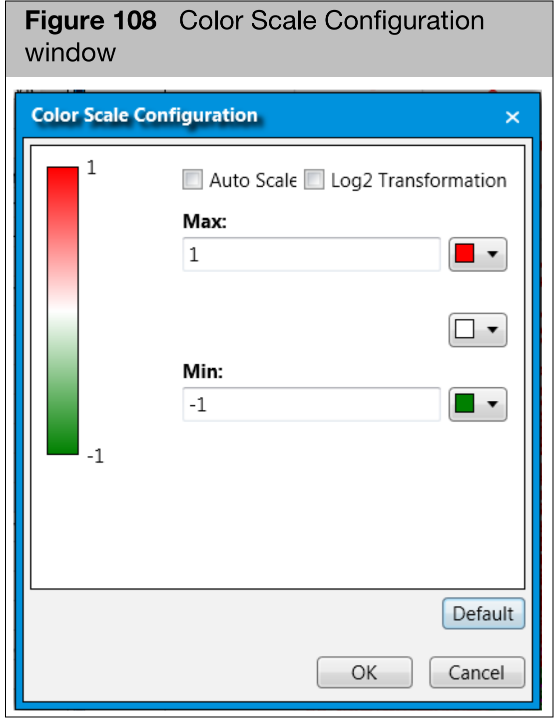


Color Settings

1. Click the **Color Settings**.

   The Color Scale Configuration window appears. (Figure 108)

Figure 108 Color Scale Configuration window



2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation.

   Note: Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.

3. Use the color drop-down menus to assign new colors for your data points.
4. Click **OK** to save your changes or click **Default** to return to the factory settings.

Show Legend

1. Click **Show Legend** check box to display the Legend. Uncheck to turn it off.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn on ToolTips. Uncheck to turn off.

Clear Selections

1. Click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

Hierarchical Clustering graph

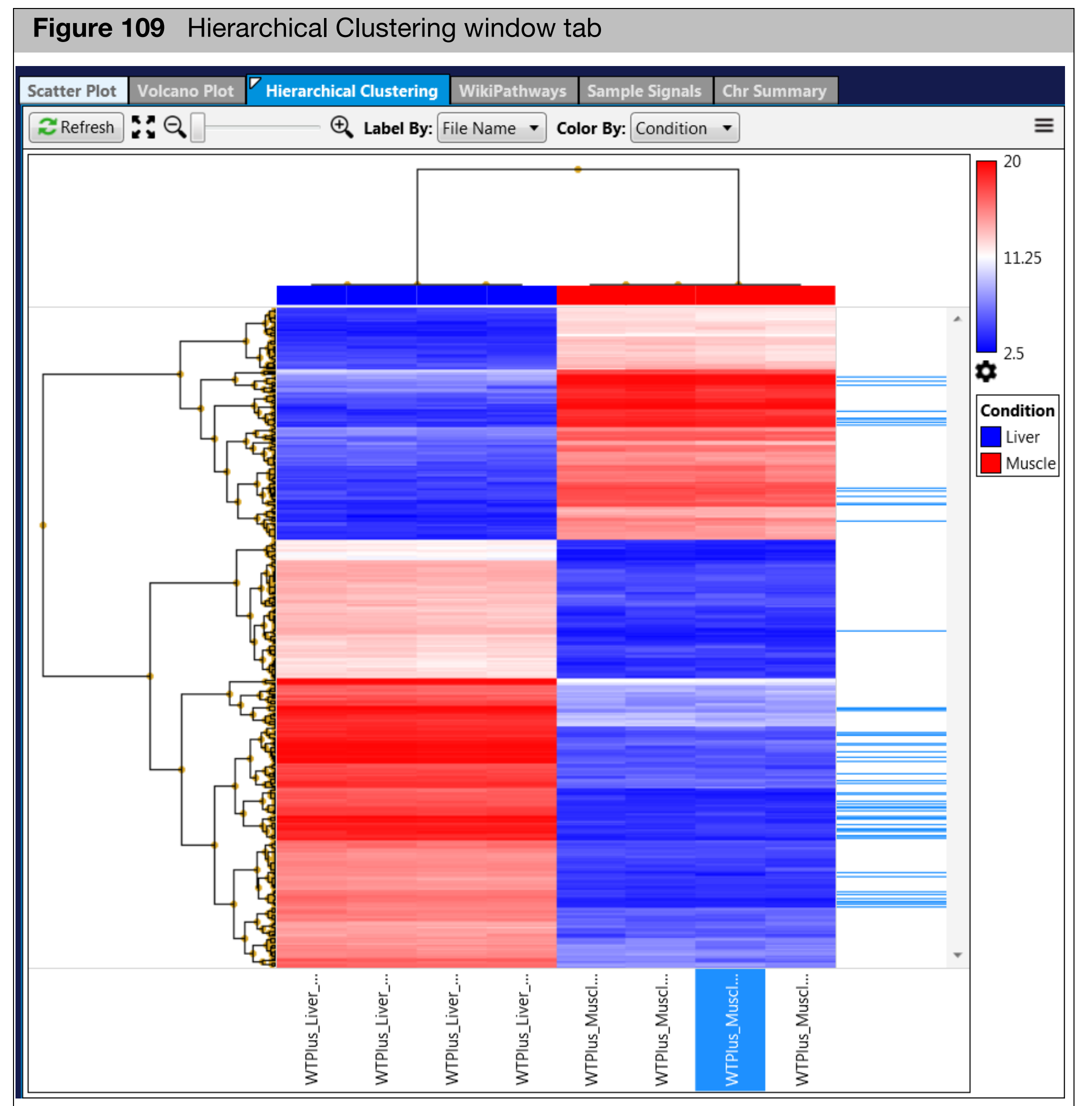
Hierarchical Clustering is a method of cluster analysis which seeks to build a hierarchy of clusters for use as a data mining tool.

1. Click on the Hierarchical Clustering tab. (Figure 109)

   The Hierarchical Clustering Graph as a 5000 gene limit. If the message, **Too much data to cluster. Apply a stricter filter to reduce the number of genes** appears, click on the Gene Table's, Filters drop- down menu to apply a stricter filter. "Managing filters" on page 100.

   After the filter is set to under 5000 results, a **Perform Clustering** button appears. Click this button to generate the graph.

Figure 109 Hierarchical Clustering window tab



IMPORTANT! The processing limit of the Hierarchical Clustering Graph is 5000 genes.

Labeling the hierarchical clustering graph

1. Click the **Label By** drop down to select how you want to label the Graph.

   The Label By selection you choose is displayed at the bottom of the graph, as shown in Figure 109. It shows the Label By is defined as Filename.

2. Click the **Color By** drop-down to select the attribute to color the graph by.

   The example in Figure 109 shows the Color By is defined by Tissue. The color legend key (far right) shows the Liver Tissue samples are red (top of the graph), while the Muscle Tissue samples are blue (top of the graph).

Selecting interesting probe sets

There are two methods to select interesting probe sets.

Method 1: Zoom Slider

1. Click the Zoom slider bar, then hold down the left mouse button and move the bar down right to magnify point(s) of interest.

Method 2: Mouse Zoom

1. To zoom in on a region of interest, left-click, hold, then move the mouse to frame your selection.
2. Release the mouse button.

Restoring Full View

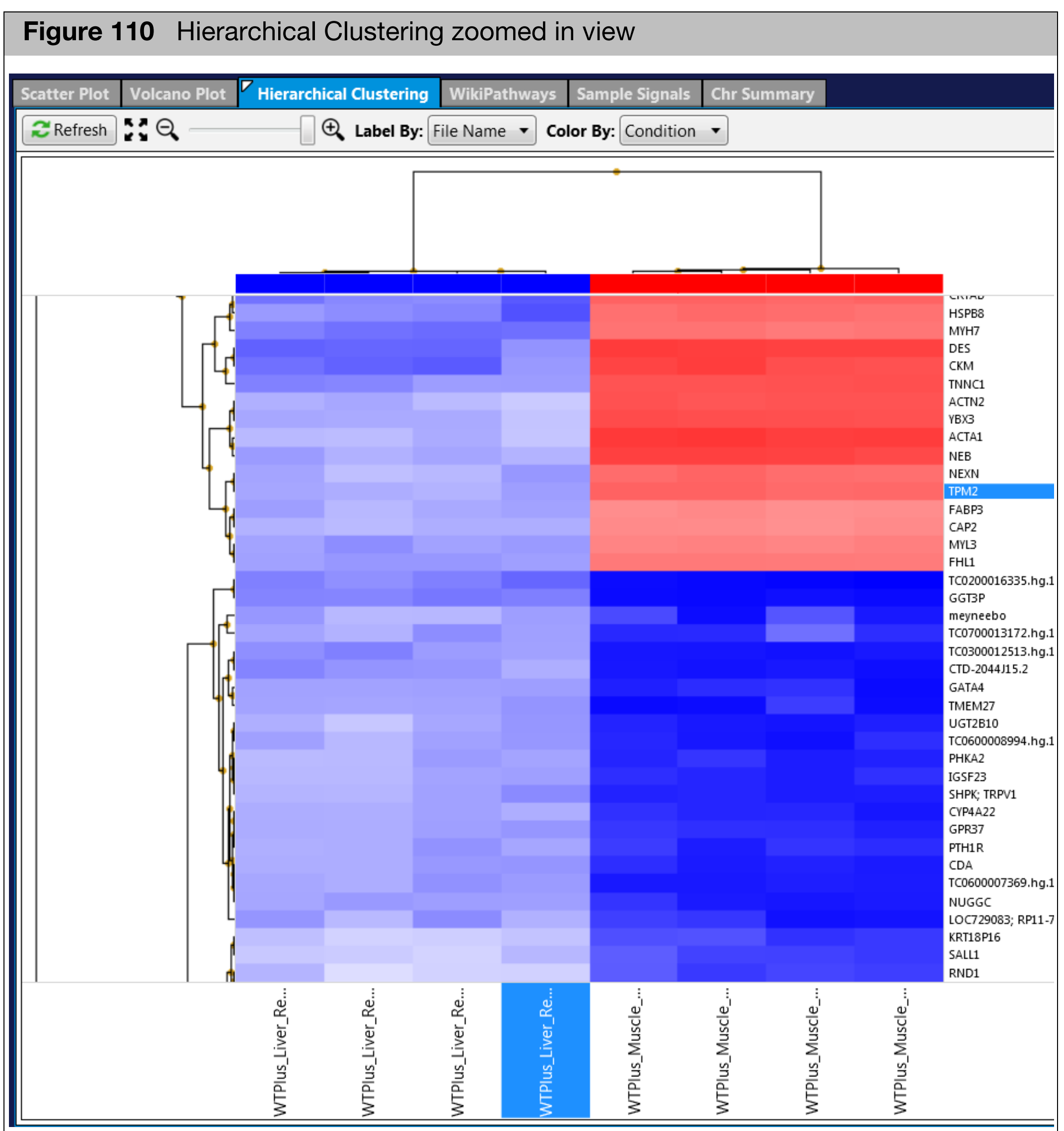
1. Click the four arrow icon or move the slider to its left stopping point.

Refreshing the View

After adding or editing a user attribute in the table, make sure you click the **Refresh** button to synchronize the graphs with your new updates.

Your region of interest is now magnified revealing the Gene Symbols (right column), as shown in Figure 110.

Figure 110 Hierarchical Clustering zoomed in view

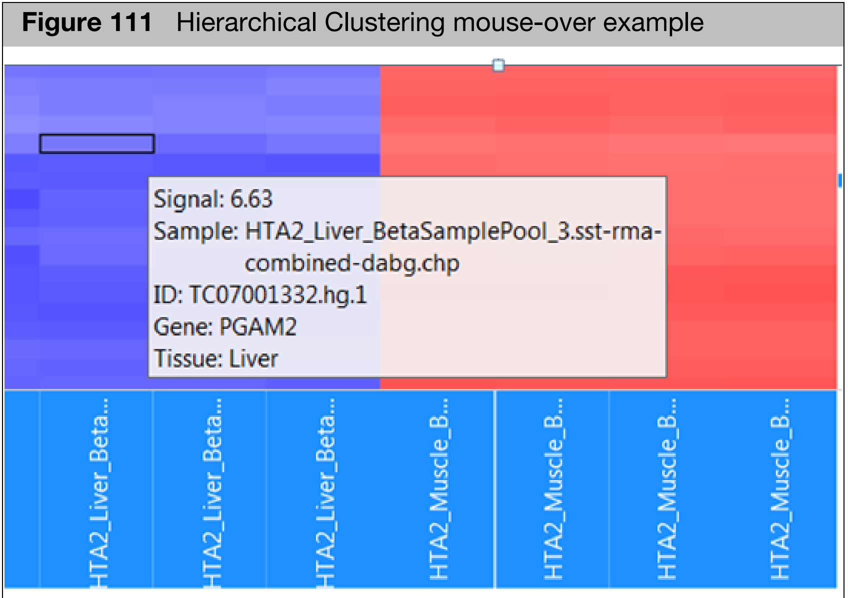


2. Click to highlight the Gene Symbol(s) to highlight the TC ID information within the Gene Table.

Obtaining information related to individual TC IDs

1. Mouse over (position the cursor over) a TC ID of interest to displays its details, as shown in Figure 111.

Figure 111 Hierarchical Clustering mouse-over example



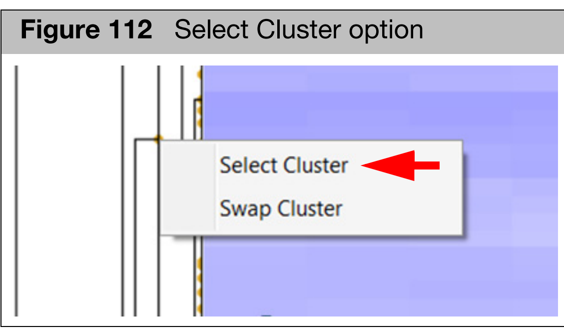
Selecting Items in a cluster

If you are interested in viewing which genes are in a particular branch of the graph, right click on a yellow node on the left side of the graph. (These steps can also be performed on the yellow nodes at the top of the graph to select items based on the **Label By** drop-down).

These TC IDs are also highlighted in the Table.

1. Click on **Select Cluster** to highlight the gene (listed on the right) in that cluster. (Figure 112)

Figure 112 Select Cluster option



Hierarchical Clustering graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

Print

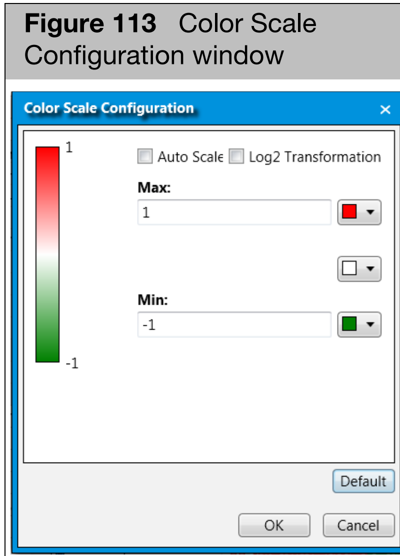
1. Click **Print**, then print the currently displayed graph using your configured printer as you normally would.

Colors Settings

1. Click the **Color Settings**.

   The Color Scale Configuration window appears. (Figure 113)

Figure 113 Color Scale Configuration window



2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation. **Note:** Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.
3. Use the color drop-down menus to assign new colors for your data points.
4. Click **OK** to save your changes or click **Default** to return to the factory settings.

Show Legend

1. Click **Show Legend** check box to display the Legend. Uncheck to turn it off.

Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Show Cluster Points

1. Click **Show Cluster Points** to make the yellow nodes visible in the branches of the graph. To turn off the nodes (cluster points), click **Show Cluster Points** to disable this feature.

Clear Selections

1. Click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

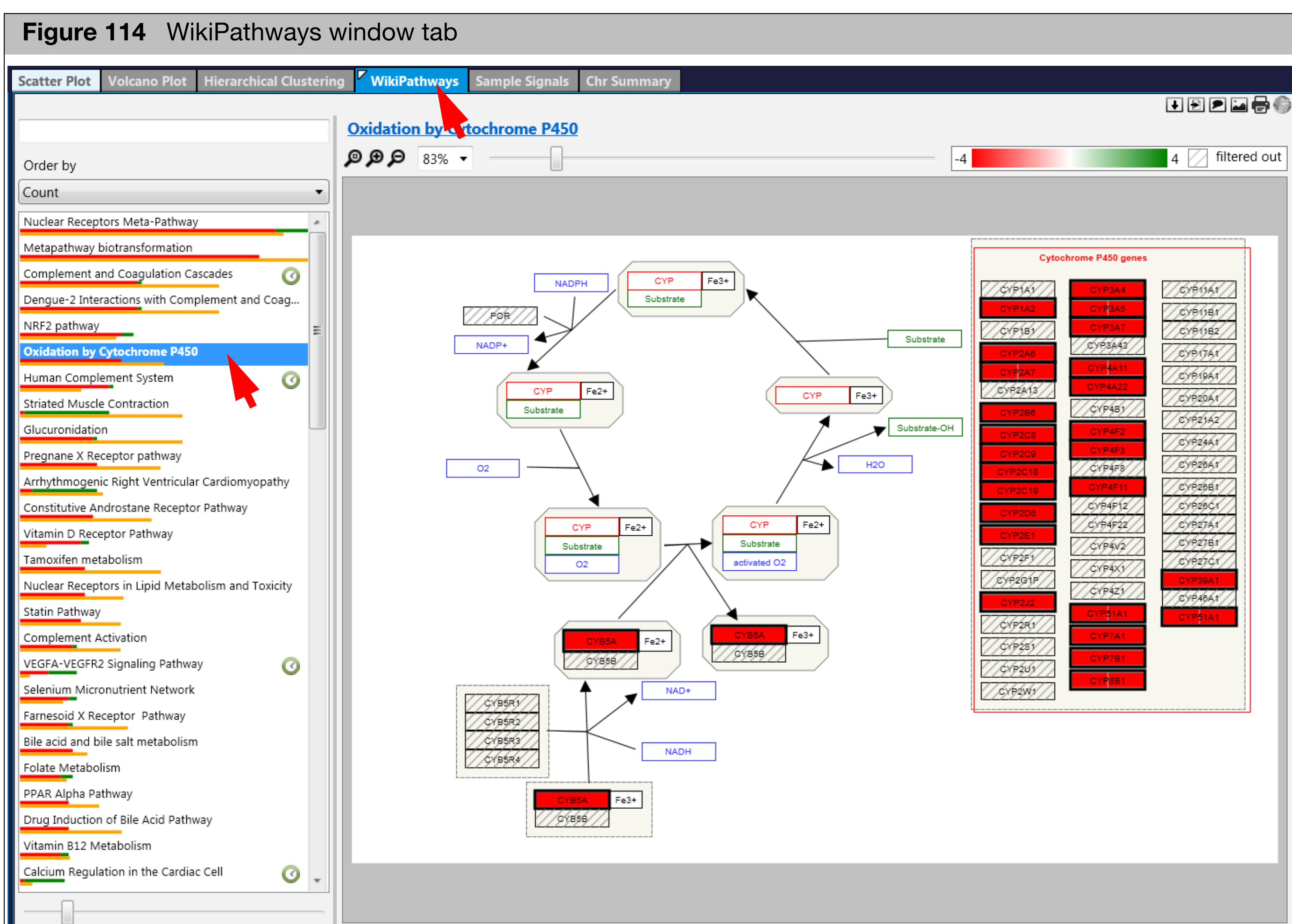
WikiPathways

WikiPathways (wikipathways.org) is a biology community maintained website that displays biological pathways. Each pathway has its own web page that contains diagrams, descriptions, references, version histories, component genes, protein lists, etc.

You must have ab active Internet connection before using WikiPathways.

1. Click on the **WikiPathways** window tab. (Figure 114)

Figure 114 WikiPathways window tab

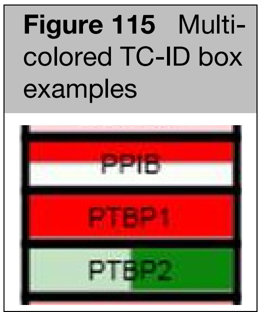


2. At the Histogram panel (left pane), double-click on a Pathway, as shown in Figure 114.

   The appropriate WikiPathways graph appears.

Note: In some Wikipathways cases, when a gene symbol is shared by several TC-IDs, the box is divided in several smaller boxes each with the relevant color for the TC-ID, as shown in Figure 115. Up-regulated is displayed at the top, while down-regulated is displayed on the bottom. Instances of multiple up/down regulated are displayed as split horizontal boxes. within the box.

Figure 115 Multi-colored TC-ID box examples



Overview of the WikiPathways Window

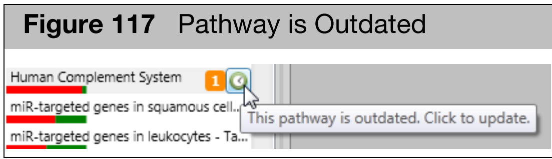
Figure 116 WikiPathways window tab

Updating an outdated pathway

A clock icon to the right of a pathway (as shown in Figure 117), indicates the software has detected an outdated Pathway.

1. Click the Clock icon to start the update process.

Figure 117 Pathway is Outdated



The Clock icon becomes an Hour-Glass while the update is downloading. (Figure 118)

Figure 118 Pathway Update In Progress

If this Warning icon appears (Figure 119), check your Internet connection and try the update again.

Figure 119 Pathway Failed to load

After downloading is complete, a NEW icon appears (Figure 120) confirming the update was successful.

Figure 120 Pathway Update Download Complete

To update ALL currently displayed Pathways, see "Download/Update Pathways" on page 128.

Changing the order of the Histogram panel

1. Use the **Order by** drop-down menu to change the histogram panel display from **Count** (default) to **Significance**, as shown in Figure 121.

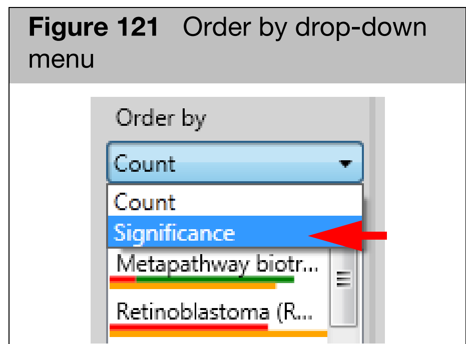
   Note: **Significance** is calculated using a 2x2 contingency in a Fisher's Exact Test (Two Sided). **Fisher's Exact Test** is a statistical significance test used in the analysis of contingency table(s).

A	B
C	D

   - **A** = Number of overlapping genes in this pathway (genes that passed the filtering criteria for a given pathway)
   - **B** = Number of non-overlapping genes in this pathway (genes that do not pass the filtering criteria for a given pathway)
   - **C** = Number of genes in current table that passed the filtering criteria but do not belong to this pathway.
   - **D** = (Total number of genes that did not pass the filtering criteria) - B

   After a p-value is established using Fisher's Exact Test, it is converted to -log10. The result of this conversion is the Significance.

Figure 121 Order by drop-down menu

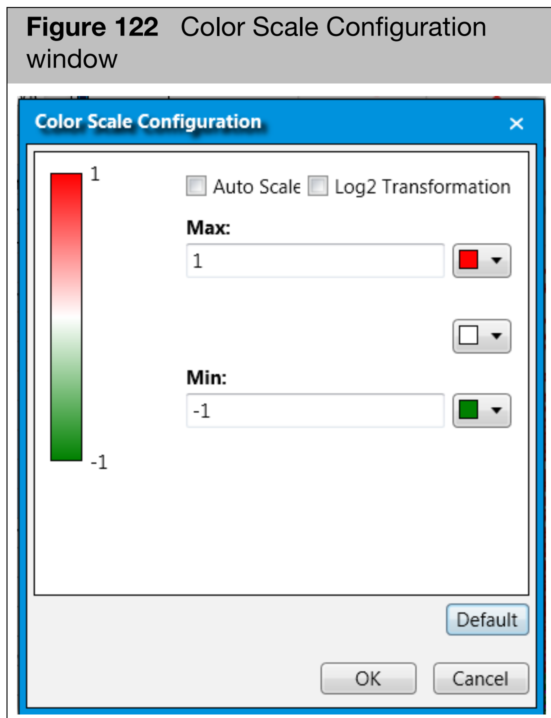


Changing the up/down regulated and significance colors

1. Click the Graph Legend color bar.

   The Color Scale Configuration window appears. (Figure 122)

Figure 122 Color Scale Configuration window



2. Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation.

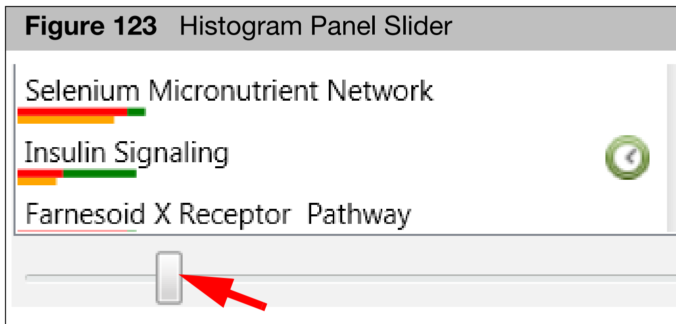
   Note: Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.

3. Use the color drop-down menus to assign new colors for your data points.
4. Click **OK** to save your changes or click **Default** to return to the factory settings.

Adjusting the Histogram panel size

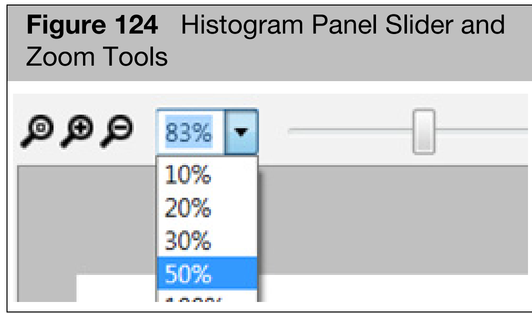
1. Move the slider left or right to change the size of the Histogram panel display. (Figure 123)

Figure 123 Histogram Panel Slider



Adjusting the WikiPathways graph size

Figure 124 Histogram Panel Slider and Zoom Tools



- Drag the slider left or right to change the size of the graph. (Figure 124)
- Click the drop-down menu, then click to select the viewing percentage you want.
- Click to increase the graph's size by 5%.
- Click to decrease the graph's size by 5%.
- Click to revert the graph's size to its default (fit to screen) size.

Using the scroll wheel

- Roll the scroll wheel forwards to zoom in.
- Roll the scroll wheel backwards to zoom out.

Panning the WikiPathways graph

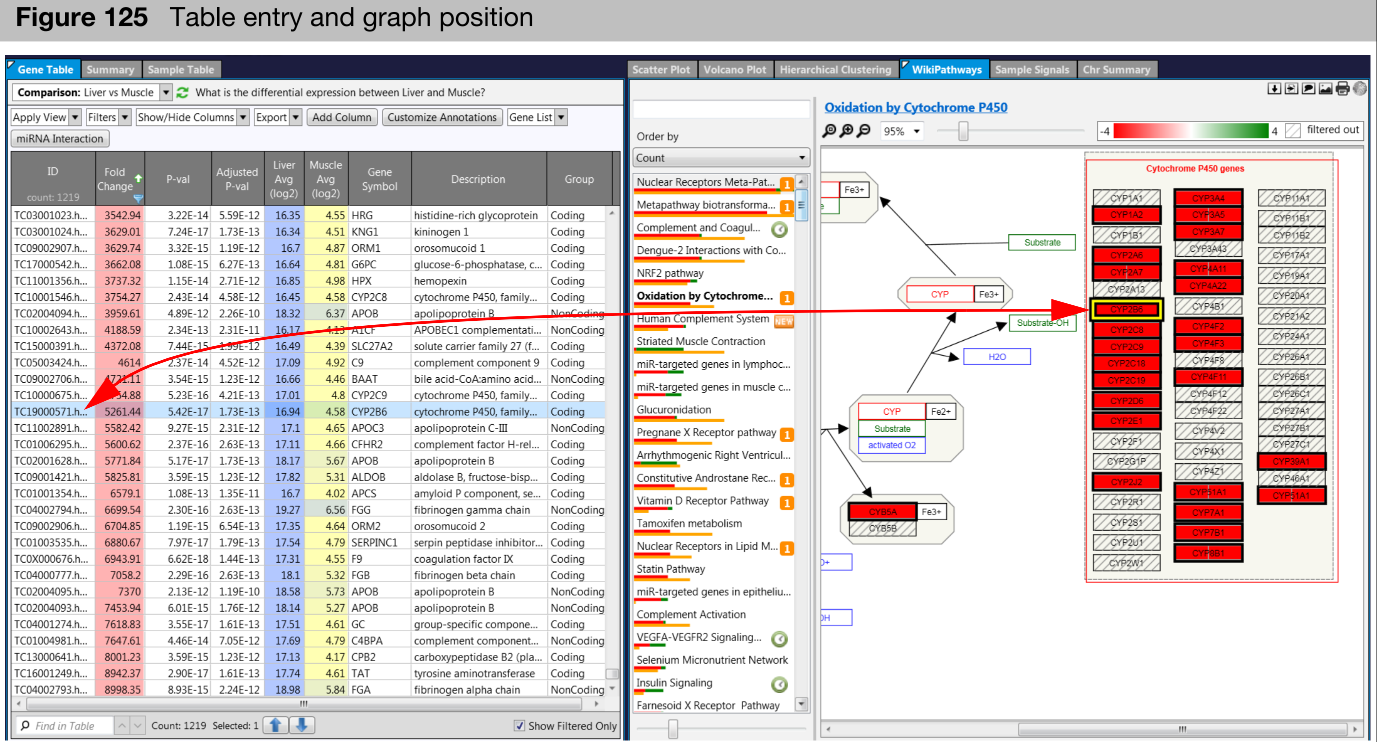
1. Click and hold down the left mouse button.
   A cursor with four points appears.
2. Move the mouse to pan the graph.
3. Release the mouse button to disable panning mode.

Using the table and WikiPathways graph

1. Single-click on a table entry to view it within the pathway. Your selection is highlighted with a yellow frame around it, as shown in Figure 125.

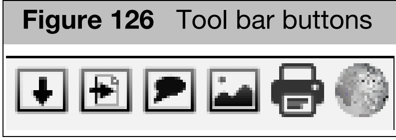
   For information on how to change column filtering properties, see "Filtering column data" on page 99.

Figure 125 Table entry and graph position



WikiPathways graph tool bar

Figure 126 Tool bar buttons



Download/Update Pathways

(Described left to right)

1. If outdated Pathways are detected, the [download icon] button is enabled (as shown in Figure 126). Click this button to download/update all the latest available pathways.

Note: If no Internet connection is detected, the message No pathways found appears inside the Pathways left window pane. Check your Internet connection or consult your IT department for help with establishing Internet access.

Export Pathway Metrics

1. To Export Pathway Metrics, click the button.

   An Explorer window appears.
2. Navigate to a export location, then click **Save**.

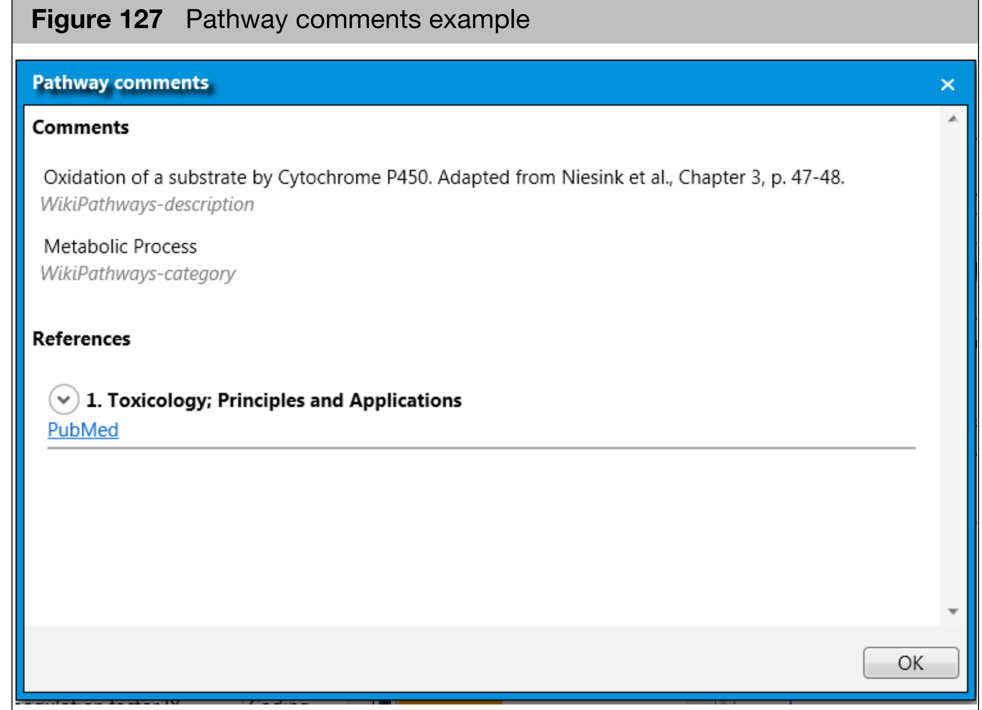
   Your pathway metrics are now saved as a tab-delimited TXT file.

Show Pathway Comments

1. Click the button.

   A window appears that contains any comments associated with the selected (highlighted) pathway. (Figure 127)

Figure 127 Pathway comments example

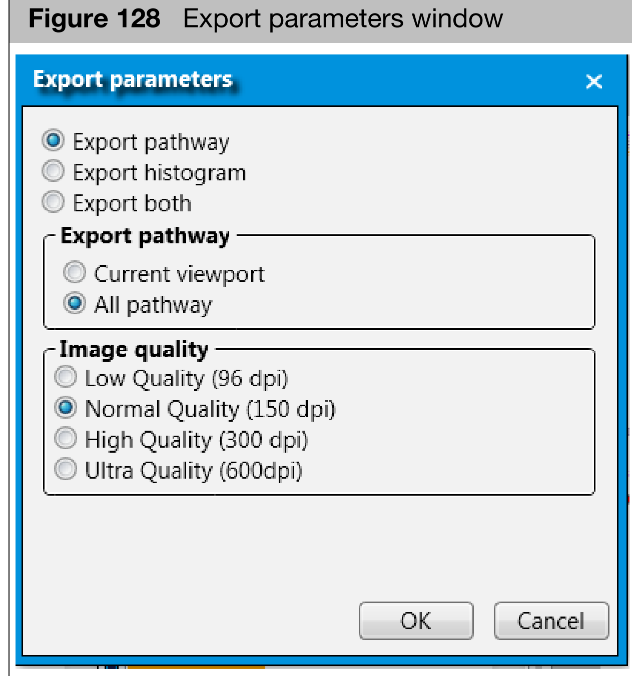


Save as Image

1. Click the button.

   An **Export parameters** window appears. (Figure 128)

Figure 128 Export parameters window



2. Click the appropriate radio buttons, then click **OK**.

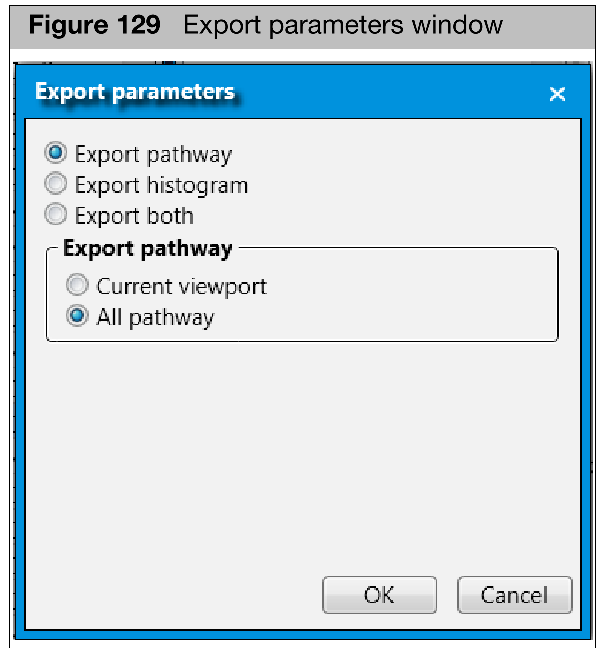
   An Explorer window appear.
3. Navigate to a save location.

   The default format is PNG, but you can click the **Save as Type** drop-down to select a different graphic format.
4. Enter a filename, then click **Save**.

Print

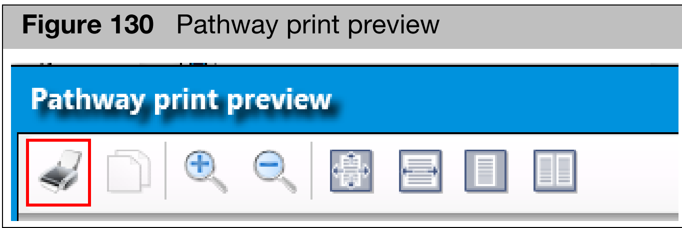
1. Click the button.
   An **Export parameters** window appears. (Figure 129)

Figure 129 Export parameters window



2. Click the appropriate radio button, then click **OK**.
   Pathway print preview window appears. (Figure 130)

Figure 130 Pathway print preview



3. Double-click on the window's header to expand the window to full size.
4. Optional: Use the tool bar buttons to adjust the preview window's size.
5. To print (to a pre-configured printer), click the **Print** icon. (Figure 130)
   The Print window appears.
6. Select the printing options as you normally would, then click **OK**.

WikiPathways Trademark/Legal Page

1. Click the button to review and understand the Terms and Usage Agreement.

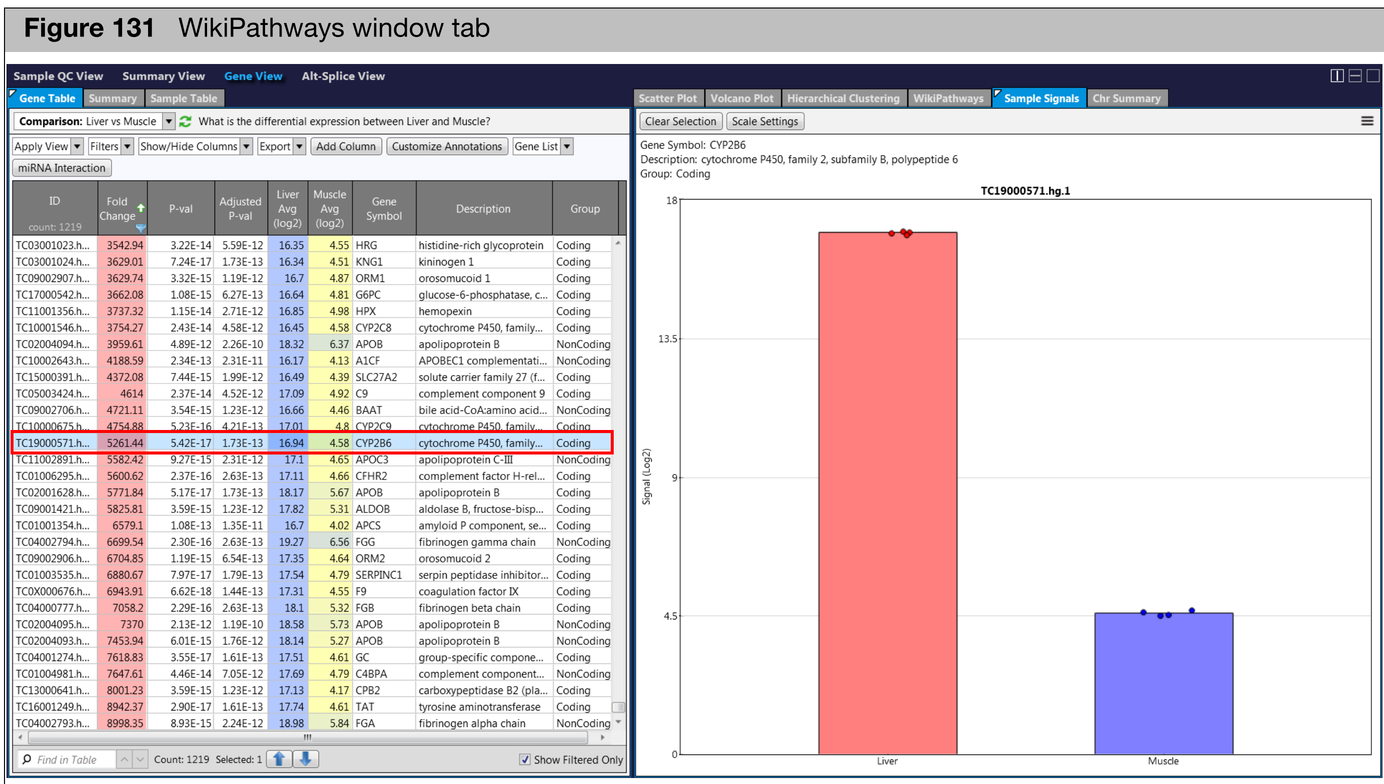
Sample Signals

The Signal Bar Graph is a graphical representation of each Transcript Cluster ID's signal intensities and standard deviations among conditions.

1. Highlight a row in the Sample Table to display the signals for that TC ID (for each condition) in the Sample Signals graph, as shown in Figure 131.

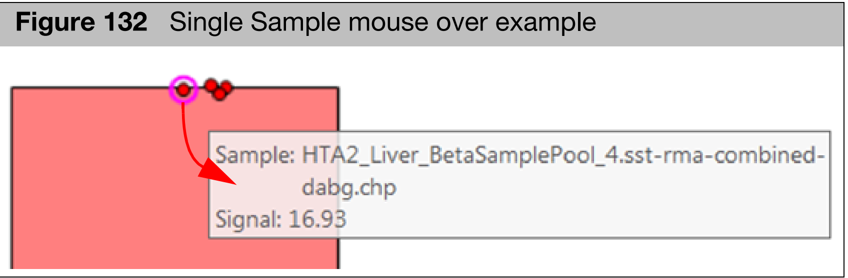
Note: The height of the bar represents the average signal of each comparison. Each dot that lies on the top of the bar represents the signal (from a sample for that specific gene).

Figure 131 WikiPathways window tab



2. Click on a dot to select a sample, then mouse over the sample to display the sample name and signal, as shown in Figure 132.

Figure 132 Single Sample mouse over example



This selected sample is also highlighted in the Sample Table. For details on how to use the Sample Table, see "Common table functions" on page 93.

Signal Sample Bar graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

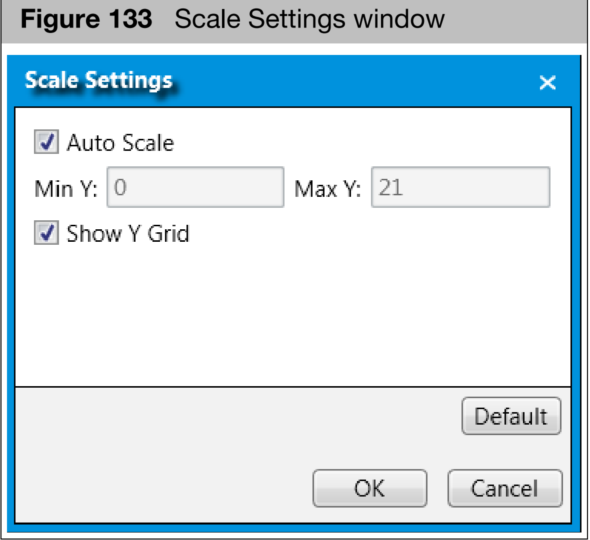
Print

1. Click **Print**, then print the currently displayed graph using your configured printer as you normally would.

Scale Settings

1. Click **Scale Settings** or click Scale Settings .
2. Select the Auto Scale check box to set the Y axis automatically based on the data selected
3. Deselect the Auto Scale check box (Figure 133) to enable the Y Min and Max value boxes. Select a Min and a Max Y axis value.
4. Select the **Show Y grid** the check box to display grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings.

Figure 133 Scale Settings window



Show ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn ToolTips on. Uncheck to turn ToolTips off.

Clear Selections

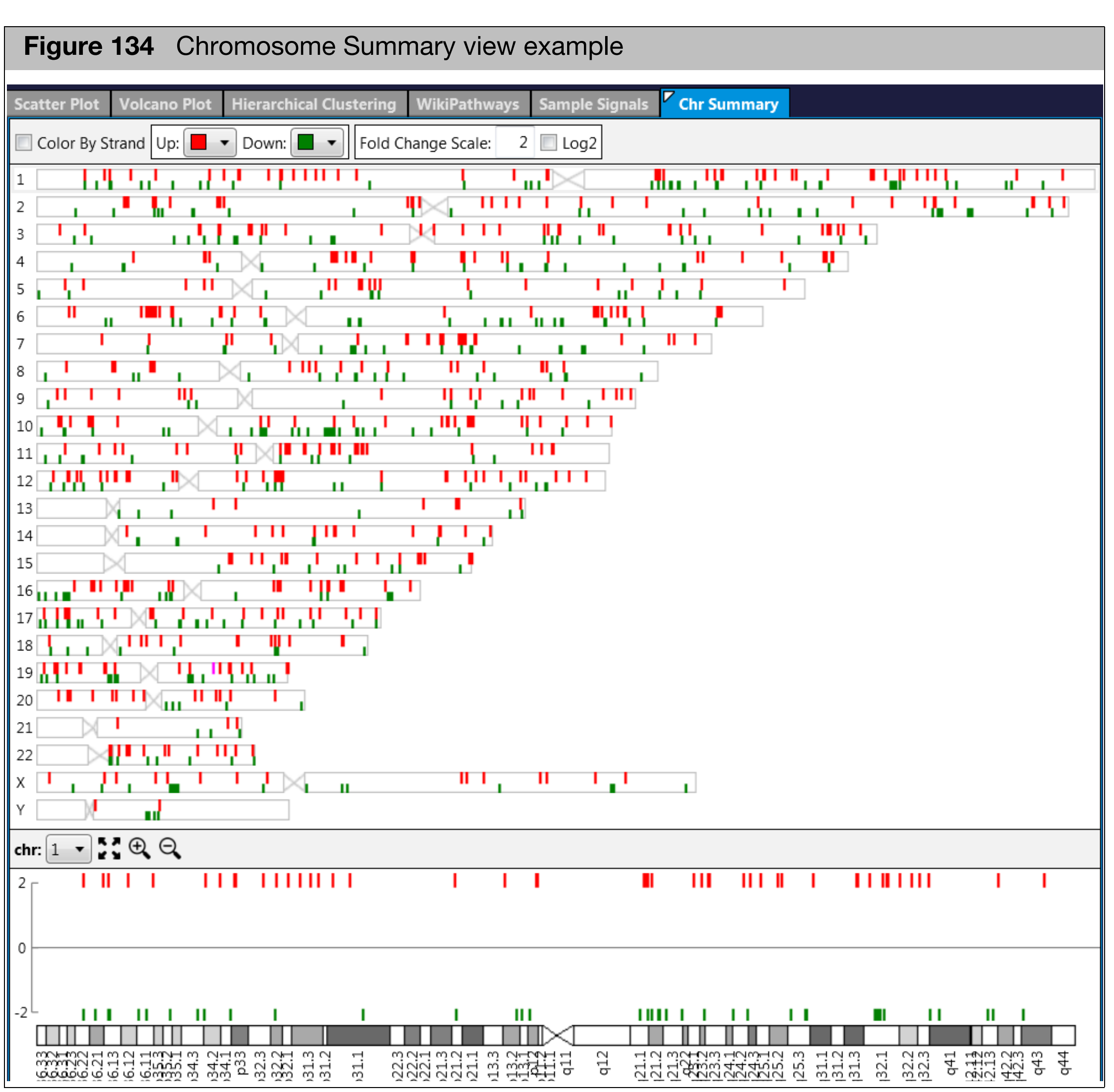
1. Click on **Clear Selection** or click Clear Selection .

   All previously selected items are now cleared from the graphs.

Chromosome Summary graph

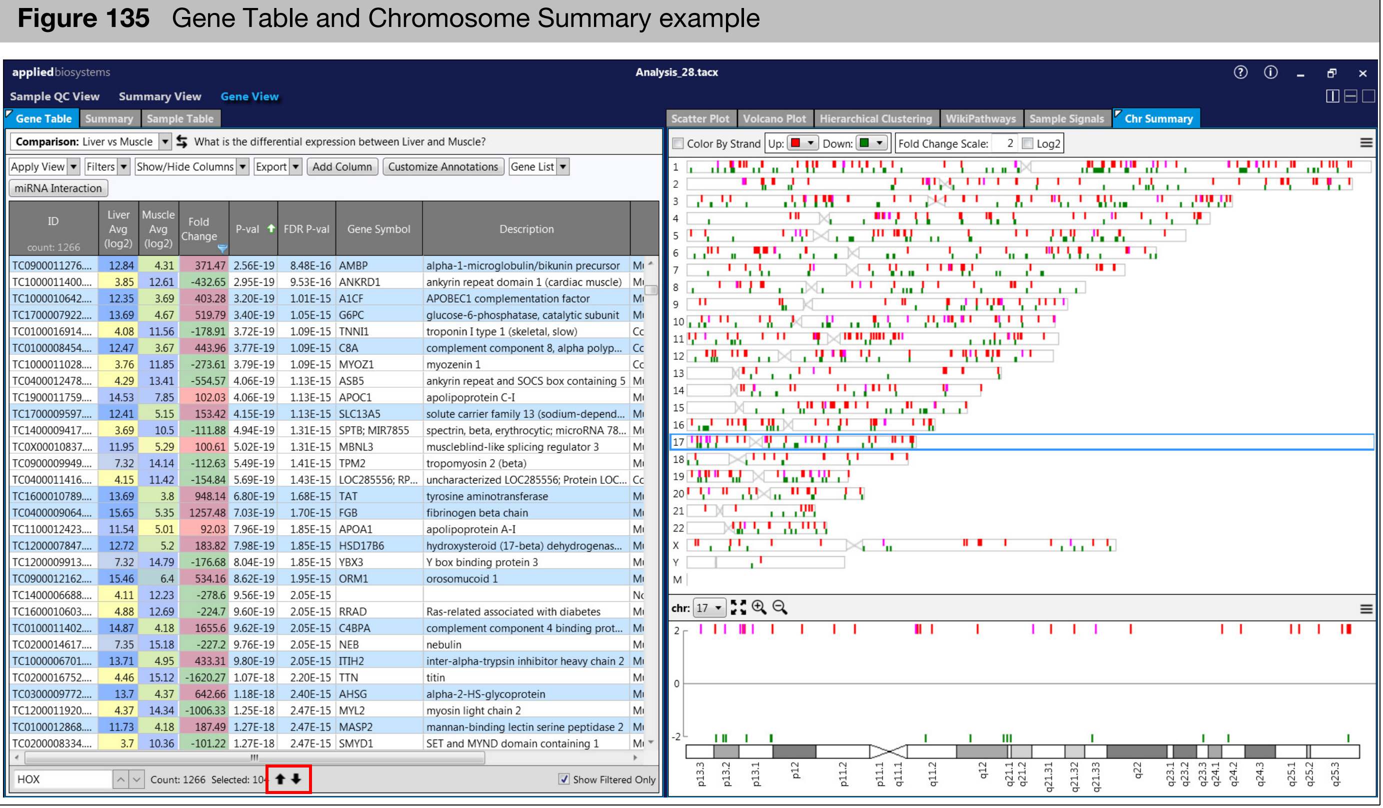
- The Chromosome Summary graph is a visual summary of your results on chromosomes.
- Probe Sets are plots based on their chromosomal positions, however Probe Sets without chromosomal positions are not plotted on the graph. Probe Sets with non-standard chromosome assignments are also not plotted. Keep this in mind, as you may see less probe sets in the chromosome summary graph than the scatter and volcano plots.
- In Figure 134, up-regulated probe sets in Liver (Condition1) vs. Muscle (Condition2) are plotted on the top of a chromosome in red. The down-regulated probe sets in Liver (Condition1) vs. Muscle (Condition2) are plotted at the bottom of a chromosome in green.
- The size of each color block in the chromosome summary graph represents its gene size, therefore, the larger the square, the larger the gene. In some instances, a square appears larger, because it may contain more than one probe set (due to the minimal pixel limitation). Also, a square might contain multiple probe sets due to minimal pixel limitation.
- Any previously lassoed genes are reflected in the Chromosome Summary graph and are represented in magenta.

Figure 134 Chromosome Summary view example



1. Click the black up or down arrows (Figure 135) to navigate through each highlighted gene or right click in the table to get the gene symbols/IDs of these lassoed genes.

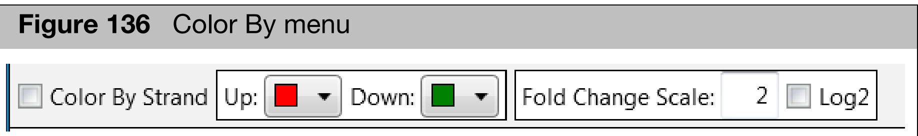
Figure 135 Gene Table and Chromosome Summary example



Changing graph colors

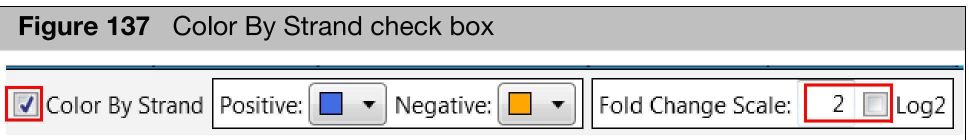
- Use the drop-down menus to select up and down regulated probe set graph colors. (Figure 136)

Figure 136 Color By menu



- Click the Color By Strand check box to color the graph by Positive and Negative strand. (Figure 137)

Figure 137 Color By Strand check box

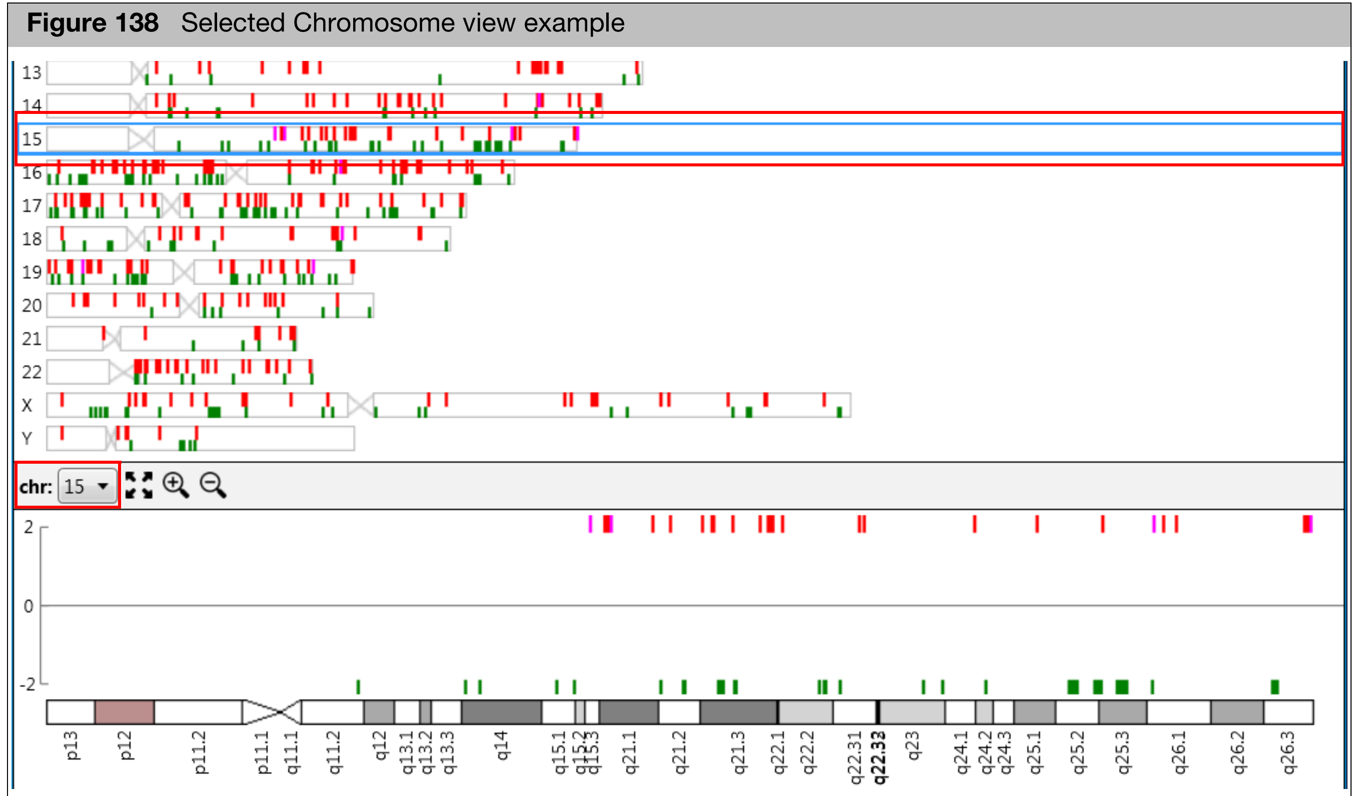


- Change the fold change filter by entering a new **Fold Change Scale** number (Figure 137), then click the **Log 2** check box if you want to change the scale to log2.

Selected Chromosome view

Click on any Chromosome in the Chromosome Summary graph to have it appear at the bottom of the graph in the Selected Chromosome View, as shown in Figure 138.

Figure 138 Selected Chromosome view example



- Click the **chr:** drop-down to select a different chromosome to view in the Selected Chromosome view.
- Use the Zoom tools as needed:
  - Click to zoom in.
  - Click to zoom out.
  - Click to return to the default display.

Exporting options

Save as PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

Print

1. Click **Print**, then print the currently displayed graph using your configured printer as you normally would.

Interaction Network graph

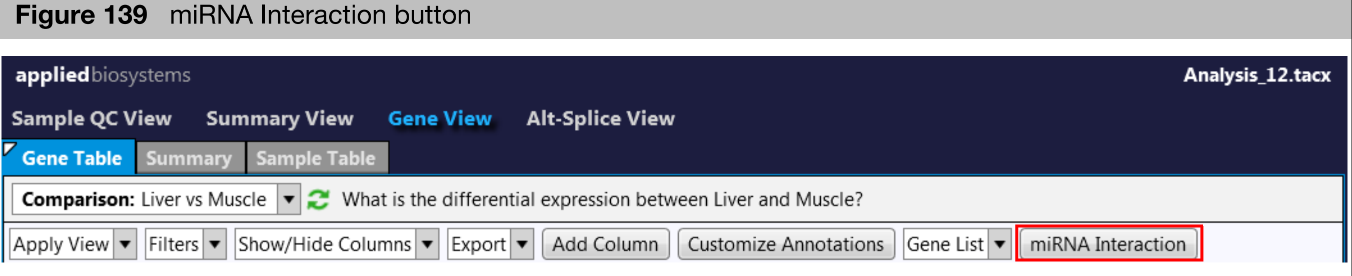
Use this feature if you want to view the interaction between miRNA and mRNA data.

IMPORTANT! Before you can use this Interaction Network feature, you must have analysis results from both the miRNA and mRNA arrays. If the current analysis displayed is miRNA data, you must select a mRNA Analysis from the Analysis Results window.

If the current analysis displayed is mRNA data, you must select a miRNA Analysis from the Analysis Results window.

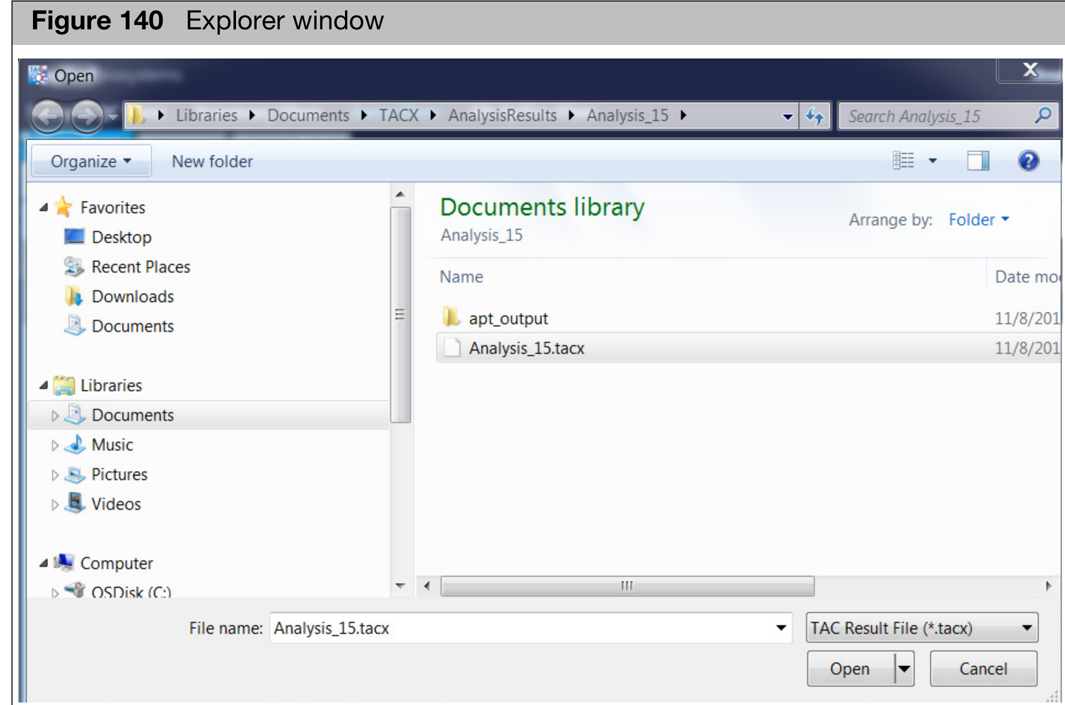
1. Click **miRNA Interaction**. (Figure 139)

Figure 139 miRNA Interaction button



An Explorer window appears. (Figure 140)

Figure 140 Explorer window



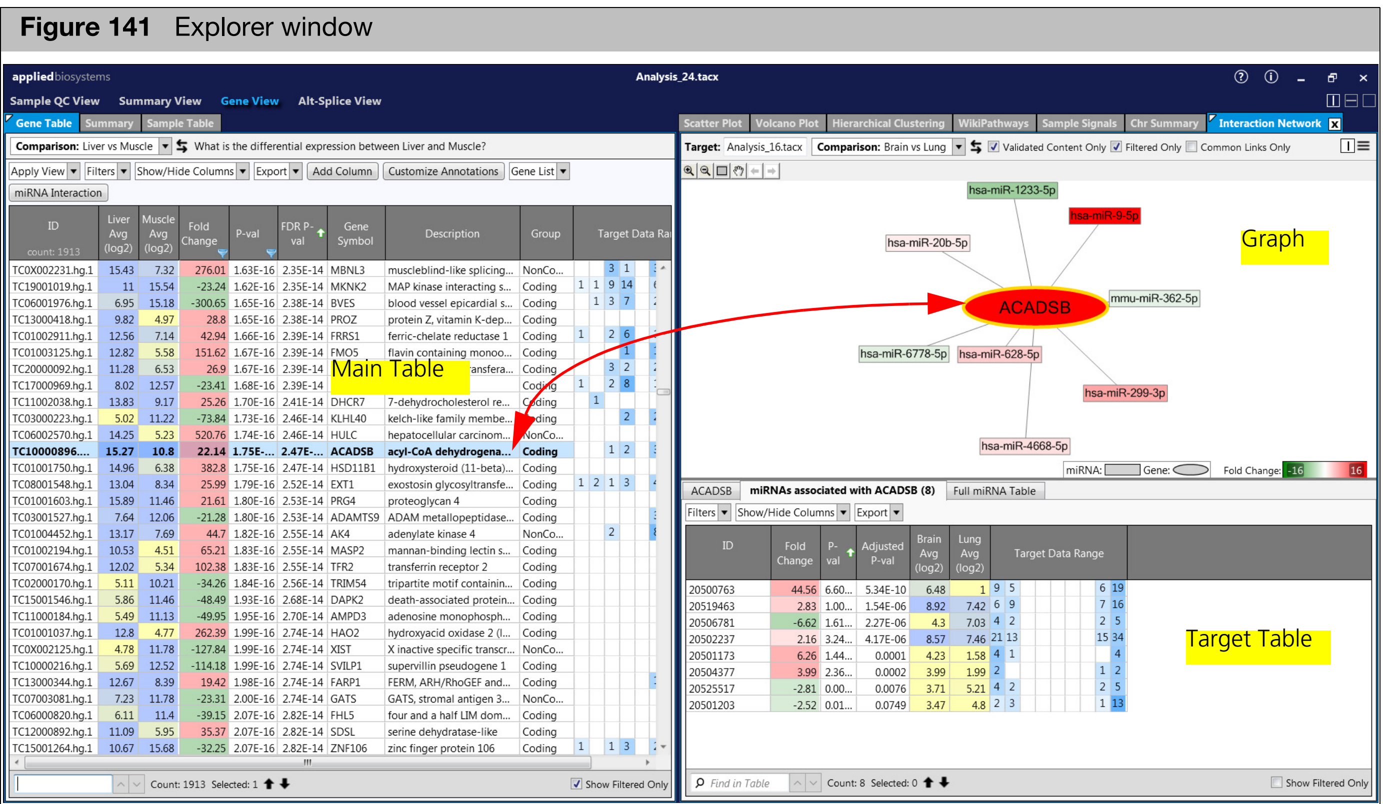
2. Locate your analysis, then click to highlight it.
3. Click **Open**.

   A Please Wait... message appears, then an Interaction Network window tab appears (Figure 141) displaying an interaction network graph (top) and its accompanying target table (bottom).

The Network graph is a representation of the targets associated with a mRNA or miRNA data displayed in the main (left) table.

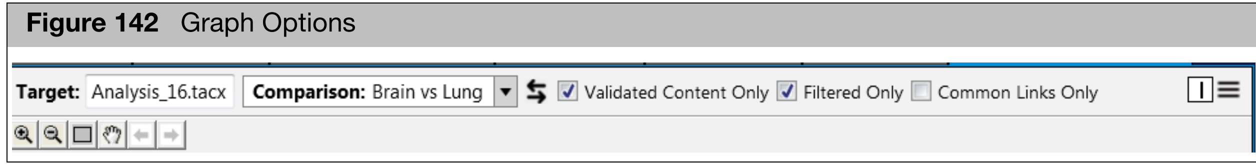
IMPORTANT! The Network Interaction feature only accepts miRNA data vs. mRNA data or mRNA data vs. miRNA data.

Figure 141 Explorer window



Using the Graph Option

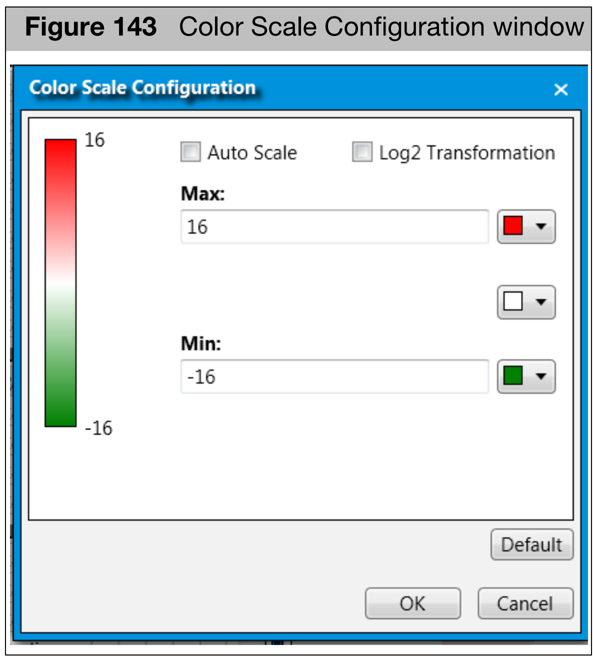
Figure 142 Graph Options



- **Target** - Displays the currently selected miRNA or mRNA TAC analysis result file name.
- **Comparison** - Use the drop-down menu to change your condition pairs.
- **Validated Content Only** - This check box is checked by default. Keep this check box checked, if you only want to see the interactions between miRNA and its validated targets in the interaction network.
- **Filtered Only** - This check box is checked by default. While checked, only filtered data is displayed. See "Filtering column data" on page 99 for more information. Uncheck this box to show data that did not pass the previous filter criteria.
- **Common Links Only** - Click the check box to show only common links between the data.

- **Enable/Disable Horizontal Split** - Click [icon] to toggle between an up and down or side by side view.
- **Save as PNG** - Click [icon] to save the graph as a Windows PNG graphic file.
- **Print** - Click [icon], then follow the on-screen instructions to print the current graph view.
- **Shape** - Shows the two graph shapes currently used.
- **Fold Change** - Displays the current color scale used. Click on the Fold Change color legend scale to customize the colors and scale of the Interaction Network. (Figure 143)

Figure 143 Color Scale Configuration window



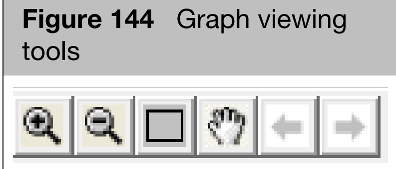
- Optional: Select **Log 2 Transformation** to color the fold change data points based on log transformation.

  Note: Leaving both check boxes unselected (Log 2 Transformation and Auto Scale) automatically colors the fold change based on linear data.
- Use the color drop-down menus to assign new colors.
- Click **OK** to save your changes or click **Default** to return to the factory settings.

Using the Tool Bar Buttons

Figure 144 (Left to right)

Figure 144 Graph viewing tools



- **Zoom In** - Click this button to zoom in.
- **Zoom Out** - Click this button to zoom out.
- **Zoom by Framing** - Click to depress this button, then use your mouse to draw a rectangle around an area of interest you want to zoom in on.
- **Pan** - Click to depress this button, then use your mouse to click, hold, then move your magnified area of interest.
- **Undo** - Click to undo graph layout changes.
- **Redo** - Click to redo (restore) graph layout changes.

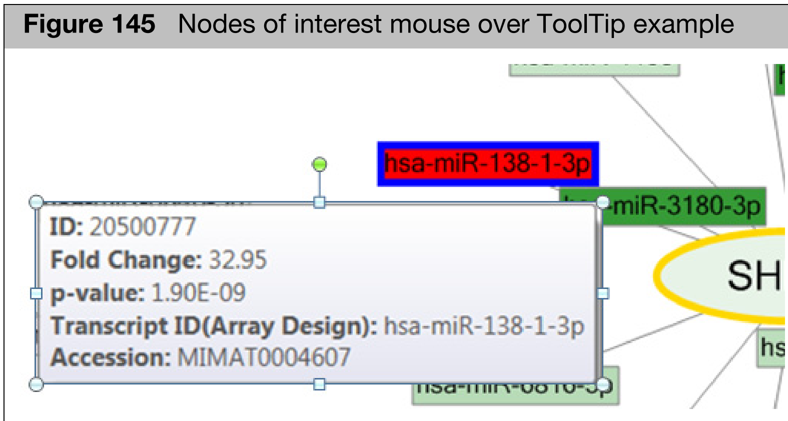
Identifying miRNA Data Nodes, Genes, and Interaction Network Relationships

To identify a miRNA nodes:

1. Mouse over a nodes of interest.

   A ToolTip window appears showing the miRNA's properties, as shown in Figure 145.

Figure 145 Nodes of interest mouse over ToolTip example

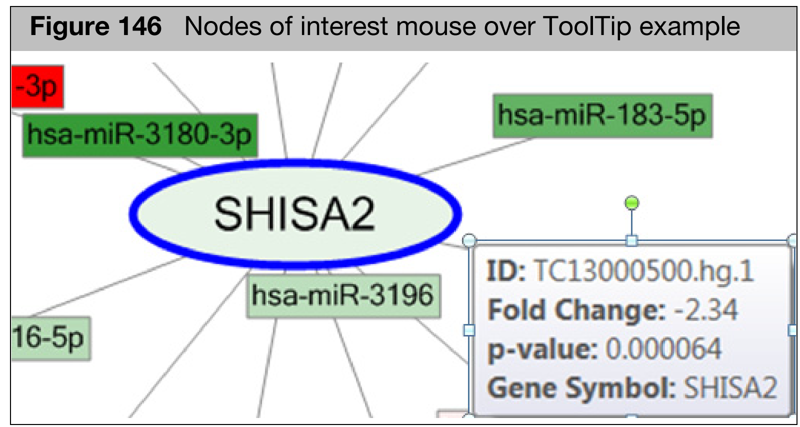


To identify a Gene:

1. Mouse over a gene of interest.

   A ToolTip window appears showing the Gene's properties, as shown in Figure 145.

Figure 146 Nodes of interest mouse over ToolTip example

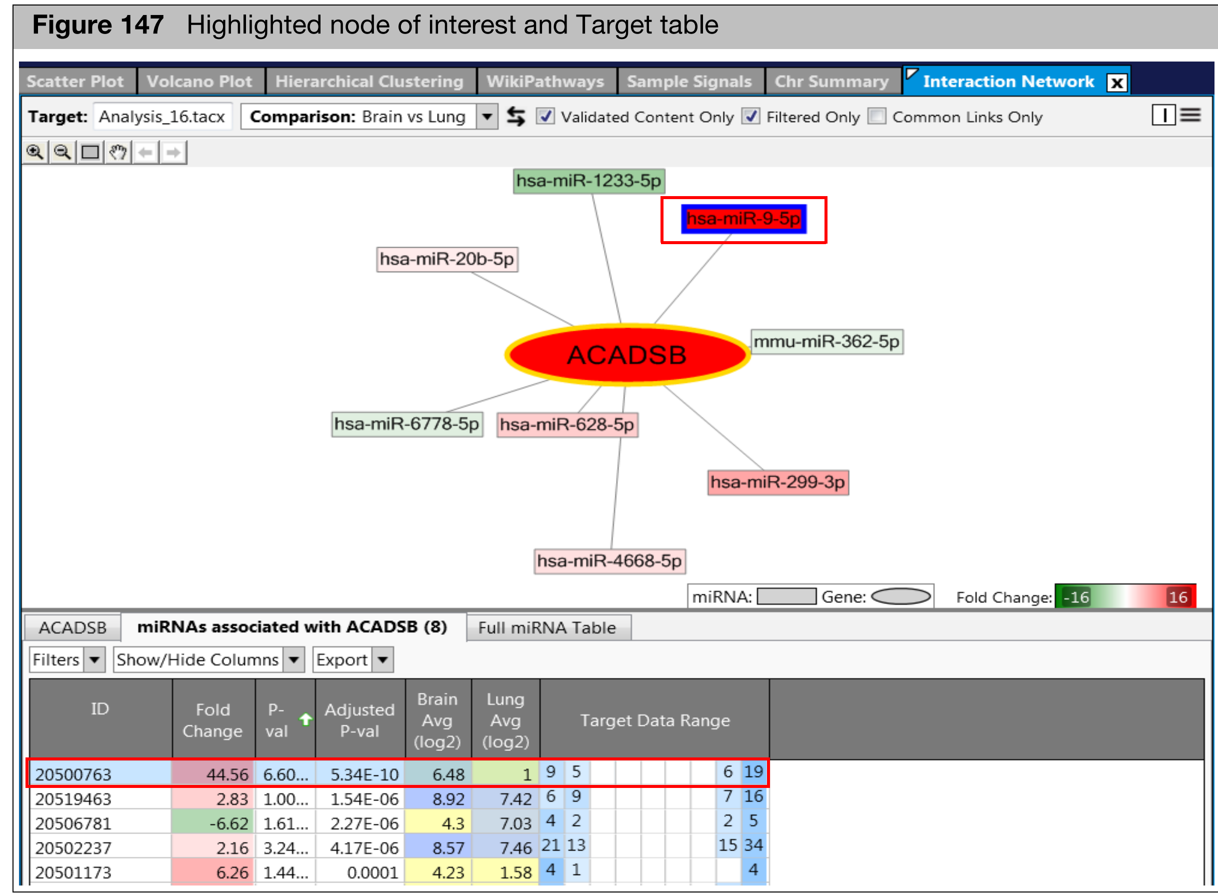


To select a miRNA node:

1. Single-click on a miRNA node of interest.

   The selected miRNA's node is highlighted, the ToolTip window appears, and its row of data is highlighted in the target table, as shown in Figure 147.

Figure 147 Highlighted node of interest and Target table



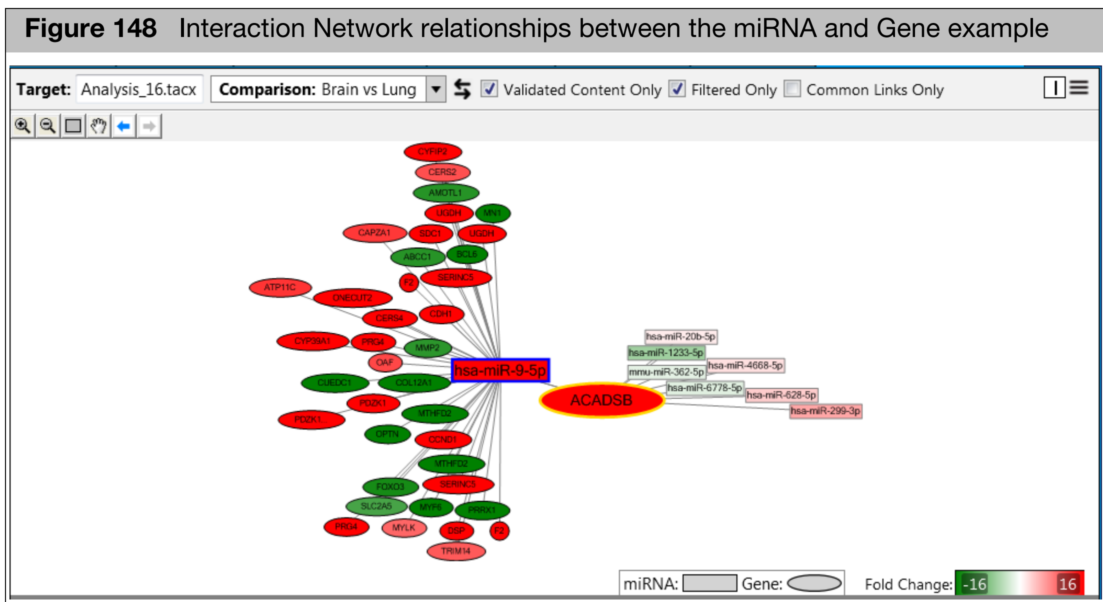
ID	Fold Change	P-val	Adjusted P-val	Brain Avg (log2)	Lung Avg (log2)	Target Data Range			
20500763	44.56	6.60...	5.34E-10	6.48	1	9	5	6	19
20519463	2.83	1.00...	1.54E-06	8.92	7.42	6	9	7	16
20506781	-6.62	1.61...	2.27E-06	4.3	7.03	4	2	2	5
20502237	2.16	3.24...	4.17E-06	8.57	7.46	21	13	15	34
20501173	6.26	1.44...	0.0001	4.23	1.58	4	1		4

To view the Interaction Network relationships between the miRNA and Gene:

1. Double-click on a miRNA node of interest.

   The following graph appears. (Figure 148).

Figure 148 Interaction Network relationships between the miRNA and Gene example



- Double-click (as needed) on a different miRNA or gene of interest for deeper relationship views (within the Interaction Network graph).
- Use the Graph's viewing tools and/or the full screen window option to closer examine the relationship findings within the graph.
- If needed, click, hold, then drag a miRNA or gene to move it.
- Click the Tool bar's blue back arrow to reset the shapes (after moving them back to their original locations on the graph.
- At any time, double-click on the highlighted table row entry (left table) to return the graph to its original state.

To view only common links within the Interaction Network:

1. From the Gene Table, Alt click, Shift click, or Ctrl click to select (highlight) multiple entries, then press **Enter**.

   The Interactive Network graph populates, as shown in Figure 149 on page 143.

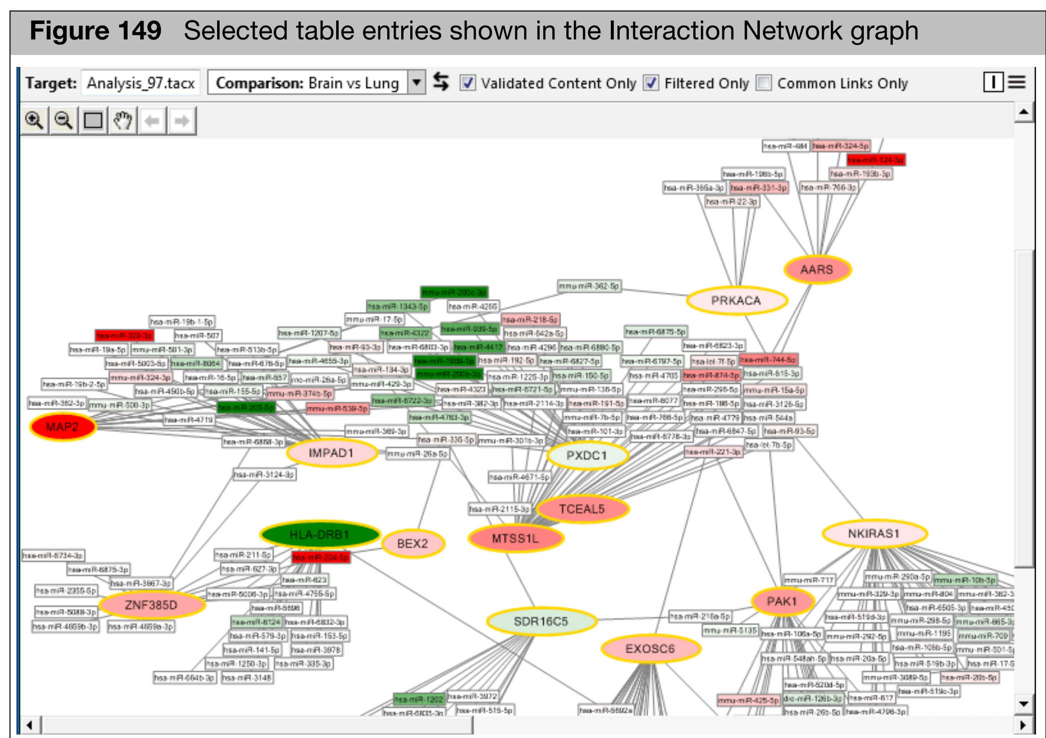


Figure 149 Selected table entries shown in the Interaction Network graph

2. Click the ☑ Common Links Only check box (upper right).

   Only common links are displayed, as shown in Figure 150.

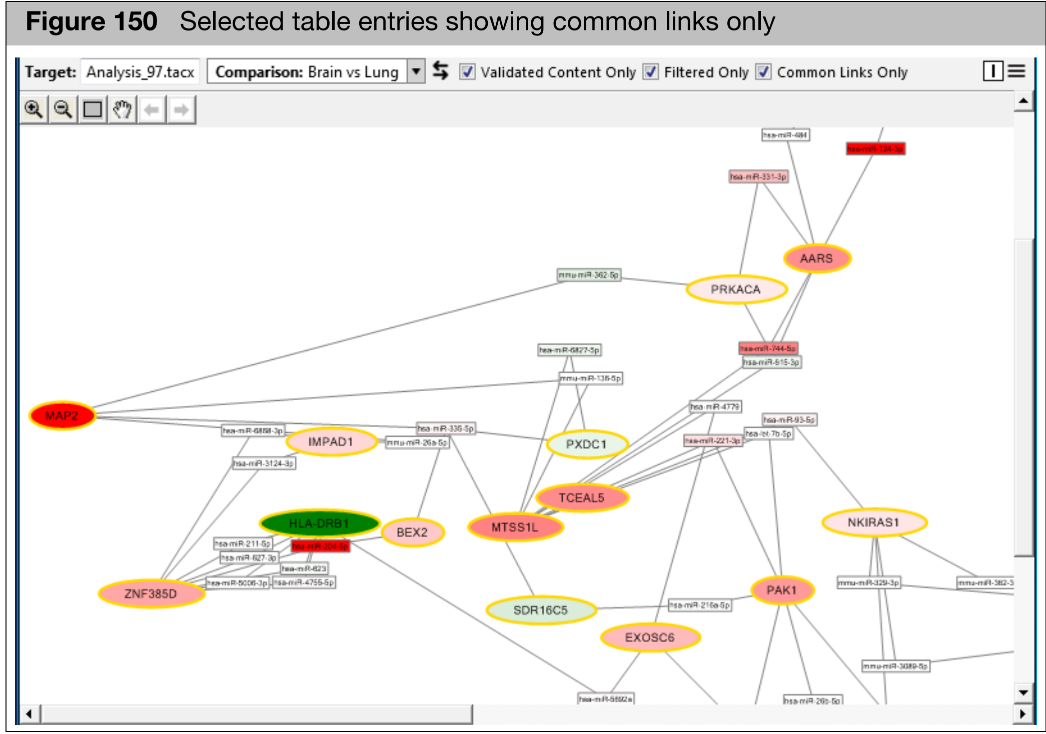
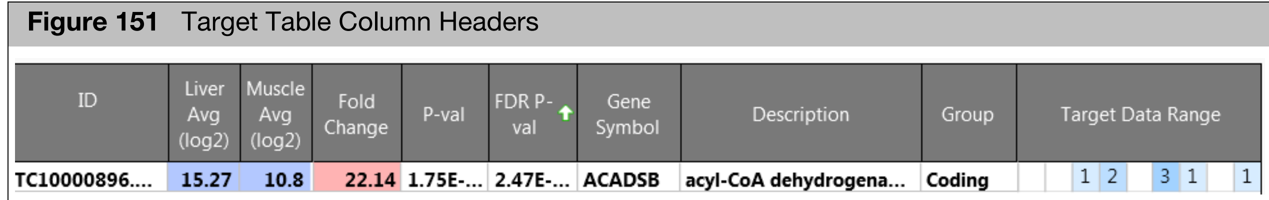


Figure 150 Selected table entries showing common links only

Using the Interaction Network Target table

The table columns displayed in the Interaction Network target table (Figure 151) are the same used in the main (left) table, however, after the Interaction Network tab/ window is generated, a **Target Data Range** column appears in both the main (left) table and the target table (bottom right).

Figure 151 Target Table Column Headers



ID	Liver Avg (log2)	Muscle Avg (log2)	Fold Change	P-val	FDR P-val	Gene Symbol	Description	Group	Target Data Range
TC10000896....	15.27	10.8	22.14	1.75E-...	2.47E-...	ACADSB	acyl-CoA dehydrogena...	Coding	1 2 3 1 1

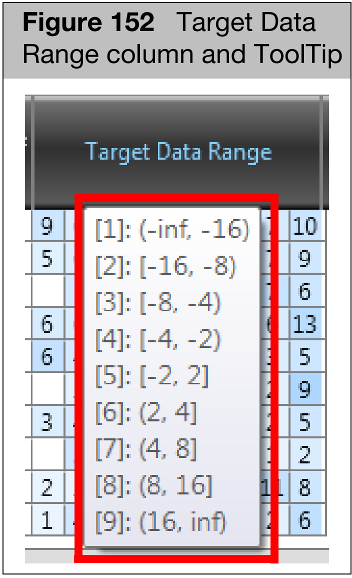
Note: Columns and their definitions are as follows. **BOLD** entries denote defaults.

Available Columns	Description
ID	ID of Transcript Cluster (TC)
(Condition1) Bi-weight Avg Signal (log2) **(Condition2) Bi-weight Avg Signal (log2)**	Tukey's Bi-weight[1] average of exon intensity of all the samples in a condition: Bi-weight average of (sample 1 exon1 intensity + sample 2 exon1 intensity +…+ sample N exon1 intensity.
(Condition1) Standard Deviation (Condition2) Standard Deviation	Standard Deviation[2] of exon intensities from all samples in a condition: STDEV of (sample 1 exon1 intensity + sample 2 exon1 intensity +…+ sample N exon1 intensity)
Fold Change (linear) (Condition1 vs. Condition2)	This shows the fold change (in linear space) of Condition1 vs. Condition2. $2^{[\text{Condition1 Bi-weight Avg Signal (log2) - Condition2 Bi-weight Avg Signal (log2)}]}$
P-value (Condition1 vs. Condition2)	One-Way Between-Subject p-value (Condition 1 vs. Condition2)[2]
FDR p-value (Condition1 vs. Condition2)	FDR adjusted p-value based on Benjamini-Hochberg Step-Up FDR-controlling Procedure[5] **IMPORTANT:** Only ANOVA p-values from PSR and Junctions that are expressed in at least one condition and its gene must be expressed in both conditions in order to be sent to FDR for correction.
Gene Symbol	This column is displayed as default column in miRNA result table. It shows the gene symbols that are associated with particular miRNA.
Description	Gene Description for this TC.
Chromosome	Chromosome for this transcript cluster. See Chromosome Naming Scheme[3] for a detailed description.
Genomic Position	Genomic Start/Stop position for this transcript cluster.
Public Gene IDs	Public Gene IDs for this TC.
Group	Whether this TC is coding or non-coding. See Table 2 on page 189 for the different options within Group and their definitions.
TC Comment	Click inside the Comment field to enter your own PSR related comment or note. Note: Entered comments are sortable.

The **Target Data Range** column (Figure 152) displays all the genes or miRNAs that

passed the filter criteria and are associated with a particular miRNA or mRNA and how many of them have a fold change within each of the nine specified fold change ranges.

Figure 152 Target Data Range column and ToolTip



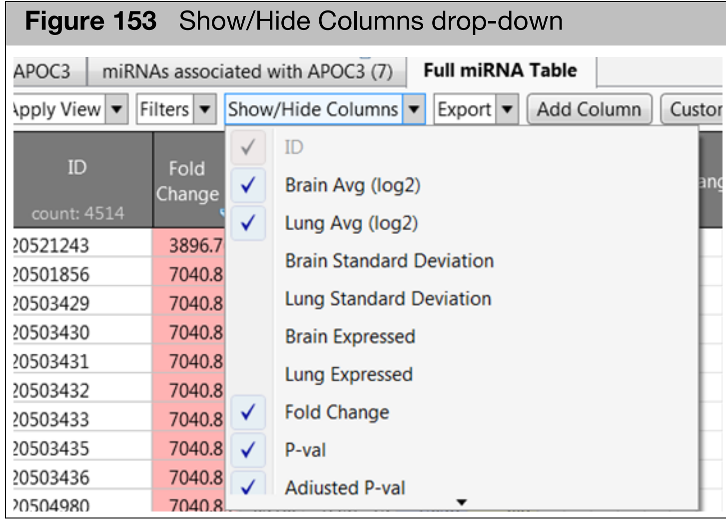
Mouse over this column header to view its ToolTip (Target Data Range legend). (Figure 152)

The number displayed inside each range is the count of genes with a fold change within that range. The range with the darkest blue shade is the range with most genes. The middle range is **-2<=** Fold Change **<=2**. A blank range is displayed if the genes have been filtered out of your selected mRNA result table. To populate a blank default range, change the default filter in your currently selected mRNA result file, then reload the mRNA result file.

Showing or Hiding Interaction Network Target Table Columns

1. Click the **Show/Hide Columns** drop-down arrow, as shown in Figure 153.

Figure 153 Show/Hide Columns drop-down

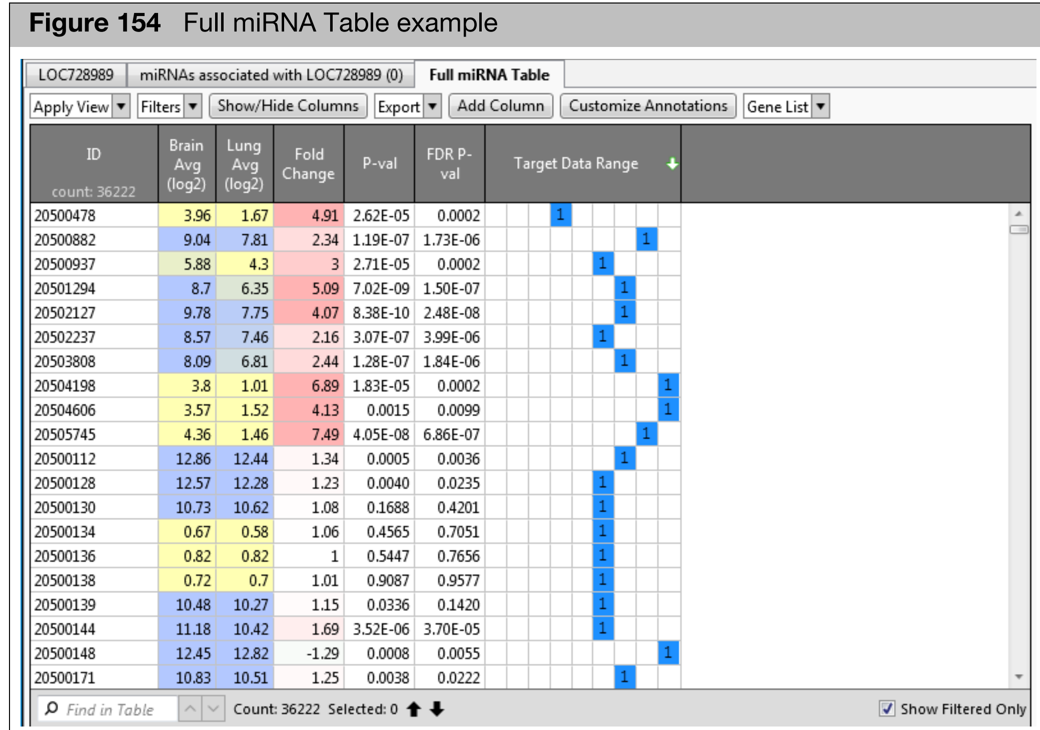


To change the current column filters, see "Table filters" on page 100.

Full miRNA Table

Selected entries from the Gene Table also appear in the **Full miRNA Table** window tab, as shown in Figure 154.

Figure 154 Full miRNA Table example



ID	Brain Avg (log2)	Lung Avg (log2)	Fold Change	P-val	FDR P-val	Target Data Range
20500478	3.96	1.67	4.91	2.62E-05	0.0002	1
20500882	9.04	7.81	2.34	1.19E-07	1.73E-06	1
20500937	5.88	4.3	3	2.71E-05	0.0002	1
20501294	8.7	6.35	5.09	7.02E-09	1.50E-07	1
20502127	9.78	7.75	4.07	8.38E-10	2.48E-08	1
20502237	8.57	7.46	2.16	3.07E-07	3.99E-06	1
20503808	8.09	6.81	2.44	1.28E-07	1.84E-06	1
20504198	3.8	1.01	6.89	1.83E-05	0.0002	1
20504606	3.57	1.52	4.13	0.0015	0.0099	1
20505745	4.36	1.46	7.49	4.05E-08	6.86E-07	1
20500112	12.86	12.44	1.34	0.0005	0.0036	1
20500128	12.57	12.28	1.23	0.0040	0.0235	1
20500130	10.73	10.62	1.08	0.1688	0.4201	1
20500134	0.67	0.58	1.06	0.4565	0.7051	1
20500136	0.82	0.82	1	0.5447	0.7656	1
20500138	0.72	0.7	1.01	0.9087	0.9577	1
20500139	10.48	10.27	1.15	0.0336	0.1420	1
20500144	11.18	10.42	1.69	3.52E-06	3.70E-05	1
20500148	12.45	12.82	-1.29	0.0008	0.0055	1
20500171	10.83	10.51	1.25	0.0038	0.0222	1

Unlike previous versions of TAC, TAC 4.0.2 enables filtering from this table. For instructions on how to use table filters, see "Filtering column data" on page 99.

IMPORTANT! Each time you apply a new filter, you must click on the **Full miRNA Table** window tab to view the newly filtered results.

Note: The adjacent Interaction Network graph and Gene Table are also updated with each newly applied Full miRNA Table filter.

To list all samples, including those filtered and not filtered, un-check the Full miRNA Table's **Show Filtered Only** check box (bottom right).

Searching Keywords

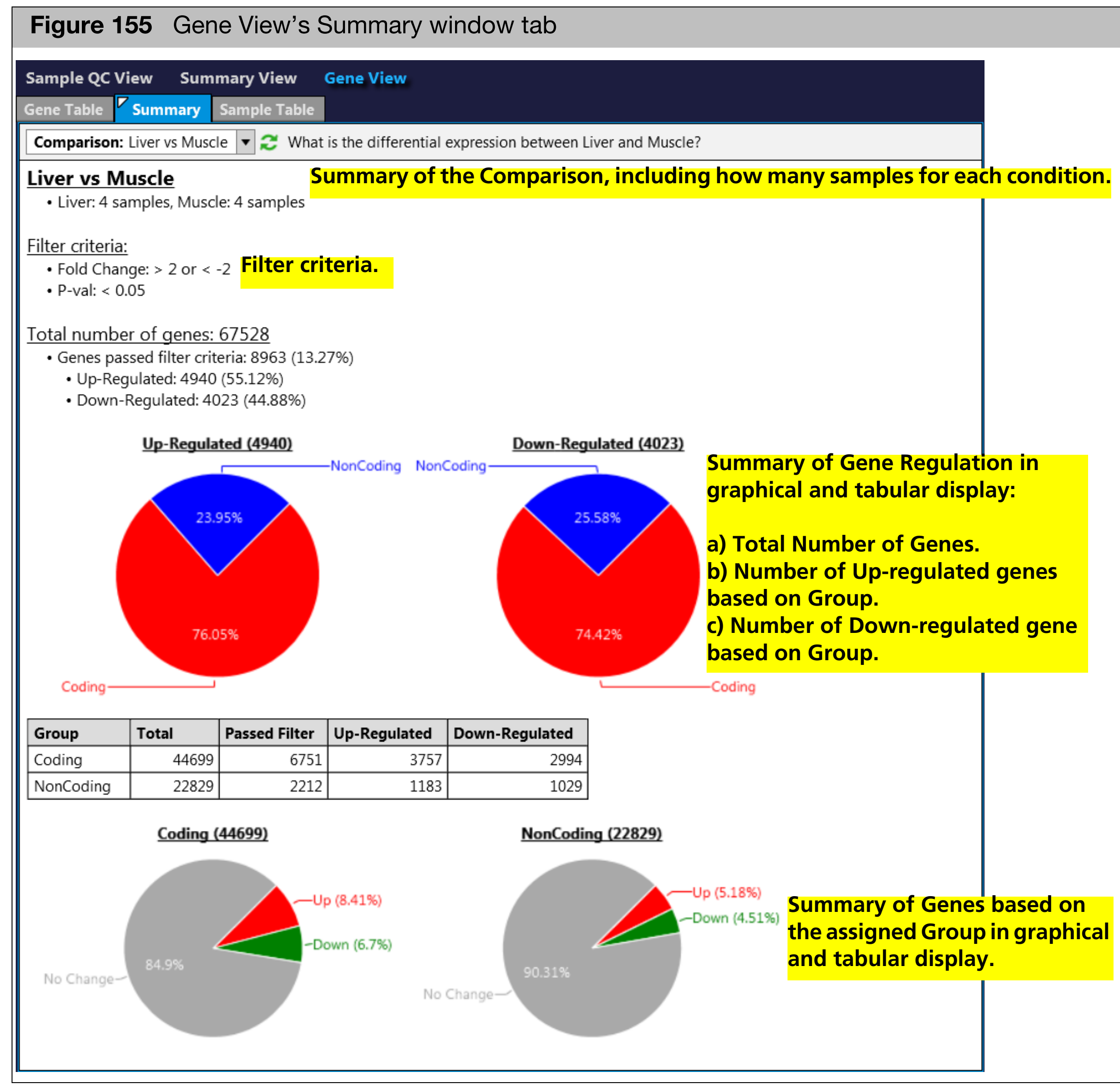
Note: By default, the search tool locates matches that contain your (case insensitive) search inquiry. Use wild-card (*) characters to aid in your search. Example: ABC* = Any string that begins with "ABC". Use (") to search for exact cell content matches. Example: "ABC"

1. Click inside the **Find in Table** field, then type your keyword, then click the ⌄ or ⌃ buttons to search.

Gene View summary

Click on the Gene View's Summary window tab (Figure 155) for a summary of the gene analysis.

Figure 155 Gene View's Summary window tab

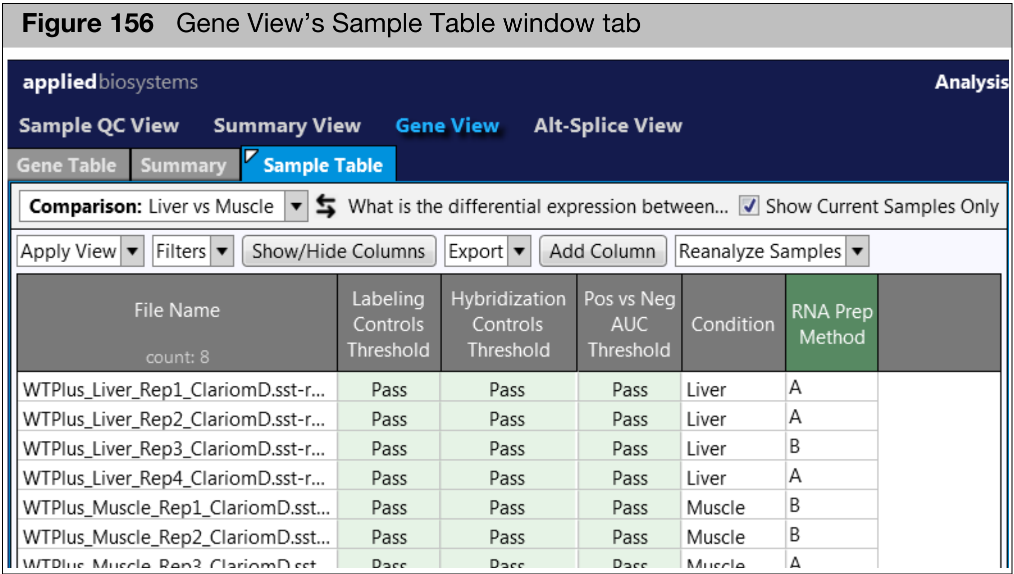


Group	Total	Passed Filter	Up-Regulated	Down-Regulated
Coding	44699	6751	3757	2994
NonCoding	22829	2212	1183	1029

Gene View Sample table

Click on the Gene View's Sample Table window tab (Figure 156) for a summary of the gene analysis.

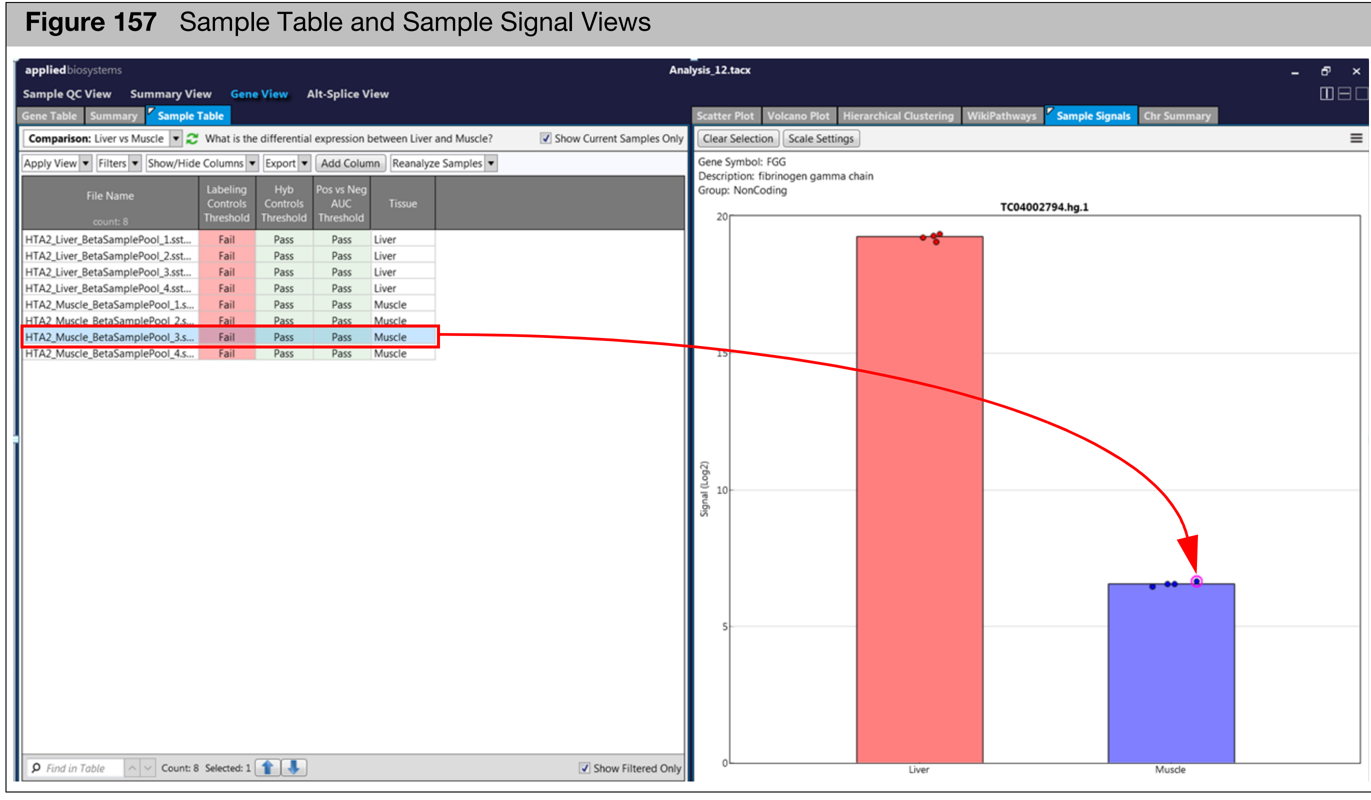
Figure 156 Gene View's Sample Table window tab



For instructions on how to use the Sample Table, see "Common table functions" on page 93.

1. Highlight a row in the Sample Table to highlight that same sample in the Sample Signal graph,, as shown in Figure 157.

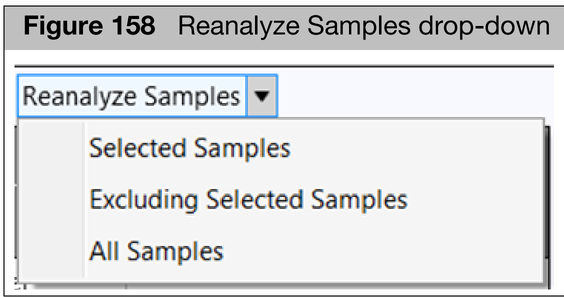
Figure 157 Sample Table and Sample Signal Views



If an outlier sample has been identified, the analysis can be re-run from the Sample Table resulting in the removal of that sample.

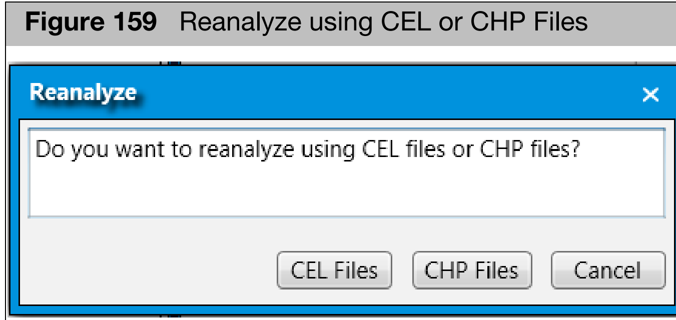
1. Click on the **Reanalyze Samples** drop-down menu. (Figure 158)

Figure 158 Reanalyze Samples drop-down



2. Select the appropriate Analysis.
   – Click **Selected Samples** to re-run analysis using the highlighted samples
   – Click **Excluding Selected Samples** to rerun the analysis on samples that are NOT highlighted
   – Click **All Samples** to re-run the analysis with all sample but with the option to choose different criteria.
3. After selecting an analysis, choose to reanalyze from CEL or CHP files. (Figure 159)

Figure 159 Reanalyze using CEL or CHP Files



The New Analysis Window opens and includes the CEL or CHP files based on your analysis selection (steps 2 and 3).

4. Select the options for the New Analysis, then click **Submit**.

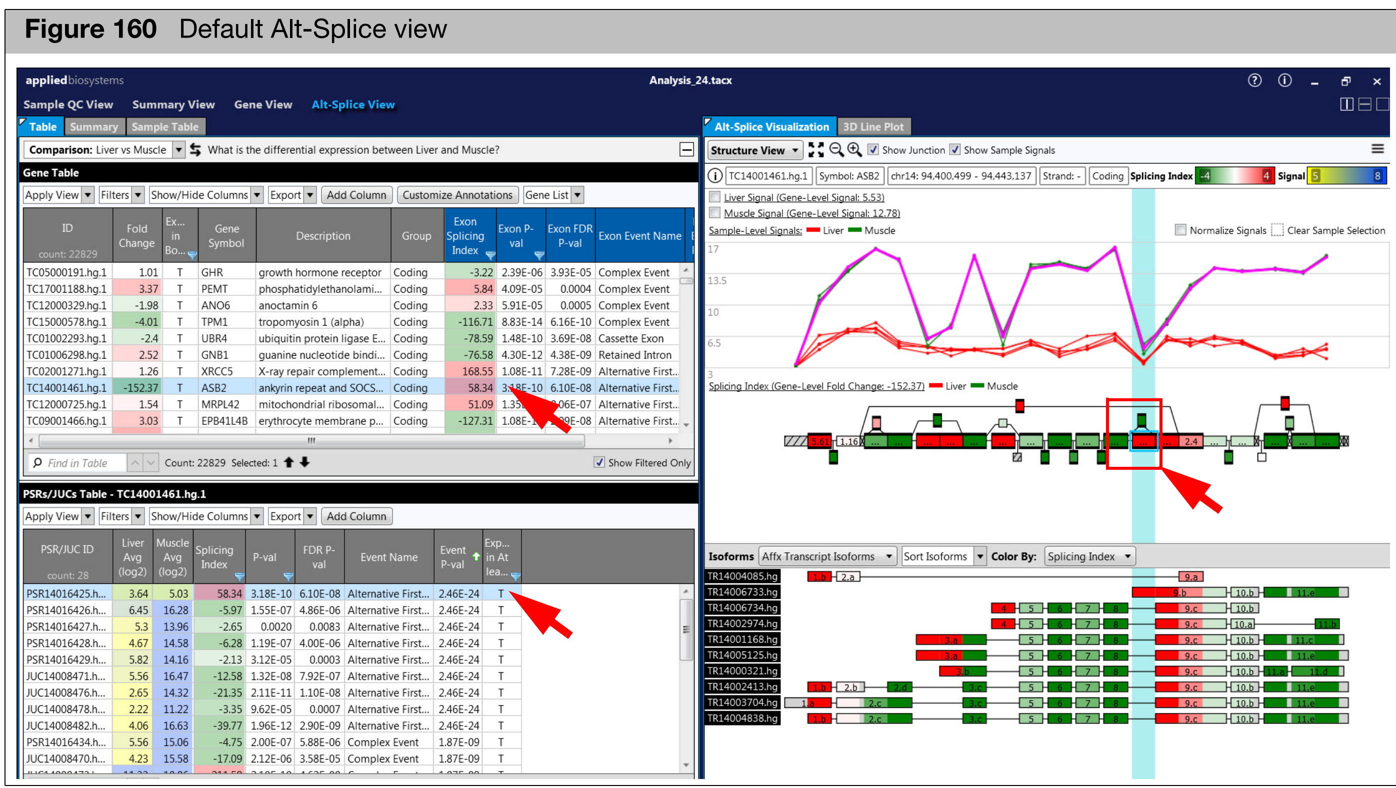
Alternative splicing analysis

IMPORTANT! Alt Splice Analysis Results are only available for Whole Transcriptome arrays that were analyzed using the Expression (Gene + Exon) Analysis.

The Alt-Splice Table is split into two tables: (Figure 160)

- Gene (Top)
- PSR (Bottom)

Figure 160 Default Alt-Splice view



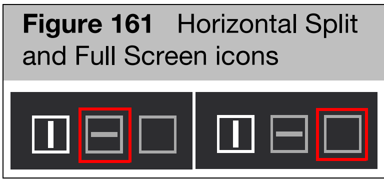
Highlighting a row(s) in the Alt Splice Gene Table will update the PSR Table with the relevant PSRs to the highlighted TC IDs. The Alt Splice Visualization Graphs and 3-D plots on the right will also update so show the PSRs for the selected TCID in a graph, as shown in Figure 160.

Viewing options

Changing the default view

- Click the **Horizontal Split** icon for a top and bottom configuration. (Figure 161)
- Click the **Full Screen** icon for individual window panes.

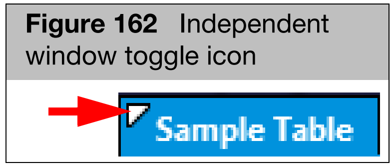
Figure 161 Horizontal Split and Full Screen icons



To toggle a tab window to full screen:

- Click on the tab's white triangle graphic. (Figure 162)

Figure 162 Independent window toggle icon



To toggle a full screen window BACK to its default tab window:

- Click on the window's **X** (as if you were to close the window).

  The window reverts back into the TAC Viewer's main window.

To change the size of a window pane:

- Click, hold, then drag the edge of the window pane to resize it.

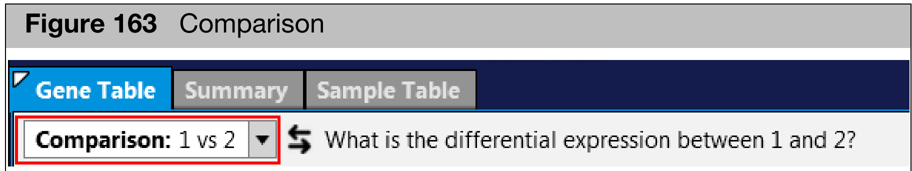
Common table functions

"Using the comparison feature"
"Customizing views" on page 152
"Working with column headers" on page 153
"Filtering column data" on page 154
"Managing filters" on page 156
"Copying selected rows and IDs" on page 157
"Accessing external databases" on page 158
"Searching the NetAffx website" on page 158
"Searching keywords" on page 159

Using the comparison feature

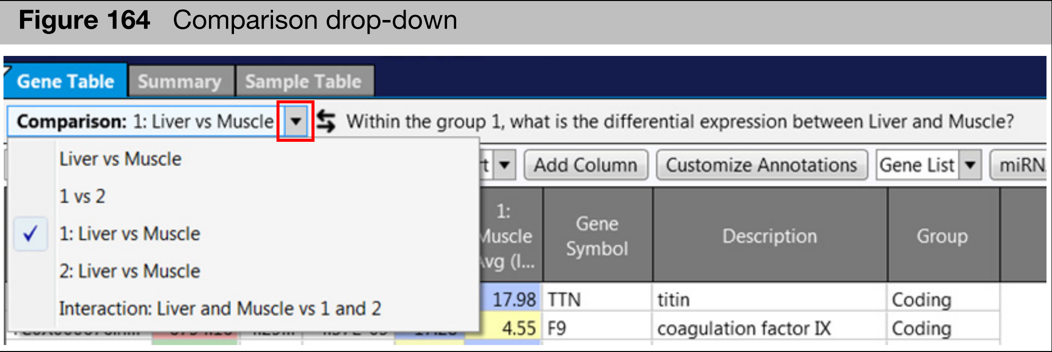
The comparison that was used in your completed analysis run is displayed at the top of the Gene Table pane, as shown in Figure 163.

Figure 163 Comparison



- For analyses that were run with multiple comparisons, click the Comparison drop-down. (Figure 164)

Figure 164 Comparison drop-down



- Click ⇆ to reverse the displayed conditions.

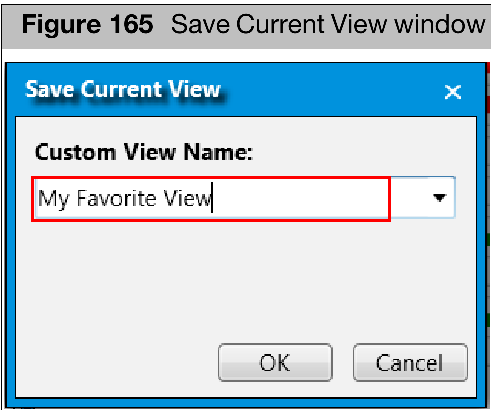
Customizing views

Custom Views created will be available for other TAC analyses using the same array type.

1. Select the columns you would like in your custom view
2. Select your custom view's column order.
3. Set any Filter parameters.
4. Click the **Apply View** drop-down, then select Save Current View.

   The Save Current View window appears.
5. Type in a new for the Custom View, as shown in Figure 165, then click **OK**.

Figure 165 Save Current View window



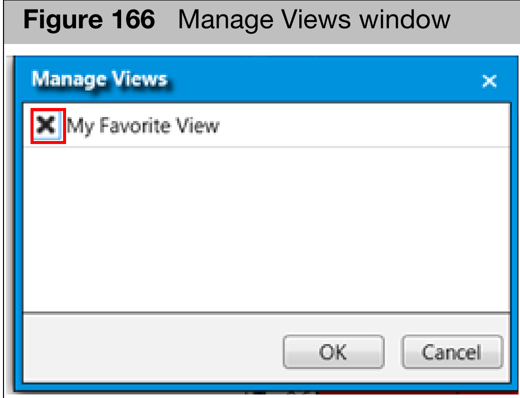
Your custom view is now available within the Table's **Apply View** drop-down menu.

Removing a custom view

1. Click **Manage Saved Views**.

   The Manage Views window appears. (Figure 166)

Figure 166 Manage Views window



2. Click the **X** adjacent to the View name, then click **OK**.

   The Custom View is now removed from the Apply View drop-down menu.

Working with column headers

The factory default columns and preset filters for the Alt-Splice table are as shown in Figure 167. For the definitions of these columns, see "Showing or hiding columns in the Alt-Splice PSR/JUC table" on page 170.

Figure 167 Default columns and filters

PSR/JUC ID count: 28	Liver Avg (log2)	Muscle Avg (log2)	Splicing Index	P-val	FDR P-val	Event Name	Event P-val	Expressed in At least One Condition
PSR14016425.h...	3.64	5.03	58.34	3.18E-10	6.10E-08	Alternative First...	2.46E-24	T
PSR14016426.h...	6.45	16.28	-5.97	1.55E-07	4.86E-06	Alternative First...	2.46E-24	T
PSR14016427.h...	5.3	13.96	-2.65	0.0020	0.0083	Alternative First...	2.46E-24	T
PSR14016428.h...	4.67	14.58	-6.28	1.19E-07	4.00E-06	Alternative First...	2.46E-24	T
PSR14016429.h...	5.82	14.16	-2.13	3.12E-05	0.0003	Alternative First...	2.46E-24	T

Rearranging default columns

1. Click on a column you want to move, then drag it (left or right) to its new location.
2. Release the mouse button.

   The column is now in its new position.

Showing or hiding columns

1. Click the **Show/Hide Columns** drop-down menu.
2. Click the check box adjacent to the column you want to show in the table. Uncheck the adjacent check box to the column you want to hide.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

Sorting columns

1. Select a column, then right-click on it.

   A menu appears.

2. Click to select either **Sort By Ascending** (A-Z) or **Sort By Descending** (Z-A).

Double-click sorting method

1. Double-click on a column header to sort its data in an ascending order. Double-click on the same column header to sort its data in a descending order.

Adding a column to the table

The contents added to a new column in this table will only exist for this analysis in TAC. When setting up a new analysis for this same dataset, the contents in these new columns will not exist. Only New Columns added in the New Analysis Tab will be available for subsequent TAC analysis runs.

1. Click **Add Column**.

   A New Column window appears.

2. Name the column, then click **OK**.

   The new column is added to the far right of the Table.

   Optional: Click and drag the column to another position in the table. The column can be removed by right-clicking on the column header and selecting **Delete Column**.

Filtering column data

Note: All table columns are filterable.

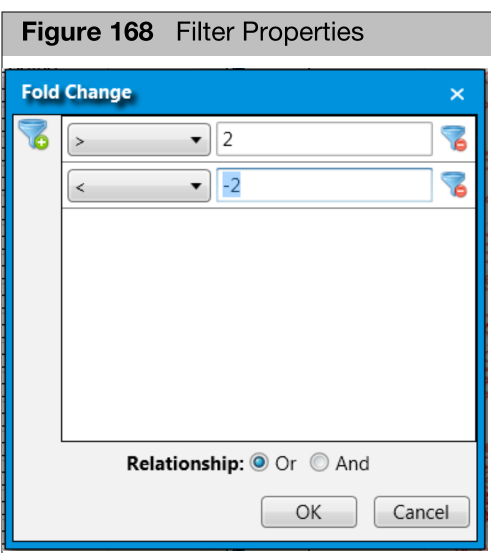
1. Select a column, then right-click on it.

   A menu appears.

2. Click **Filter**.

   The following window appears (Fold Change column example shown): (Figure 168)

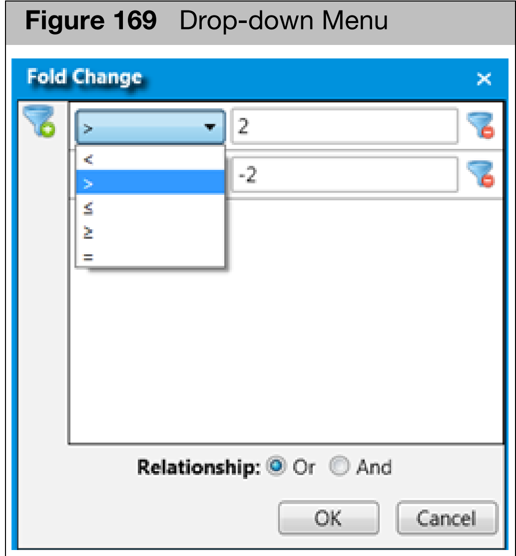
Figure 168 Filter Properties



Editing filtering properties

1. Click the **Or** or **And** button to choose **Or** or **AND** logic.
2. Click the symbol drop-down menu(s) to select new symbol(s), as shown in Figure 169.

Figure 169 Drop-down Menu



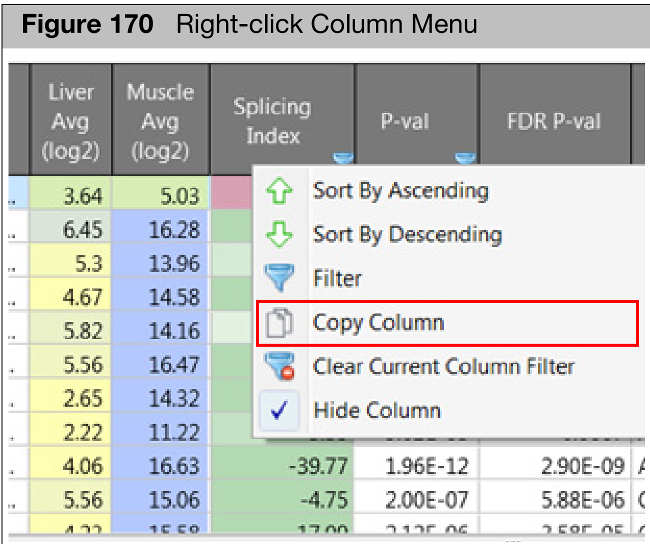
3. Click inside the numbering field(s) [2] to enter new value(s).
4. Click to add filter(s).
5. Click to remove filter(s).

Copying column data to your clipboard

1. Select a column you want to copy to a clipboard, then right-click on it.

   The following menu appears: (Figure 170)

Figure 170 Right-click Column Menu



2. Click **Copy Column**.

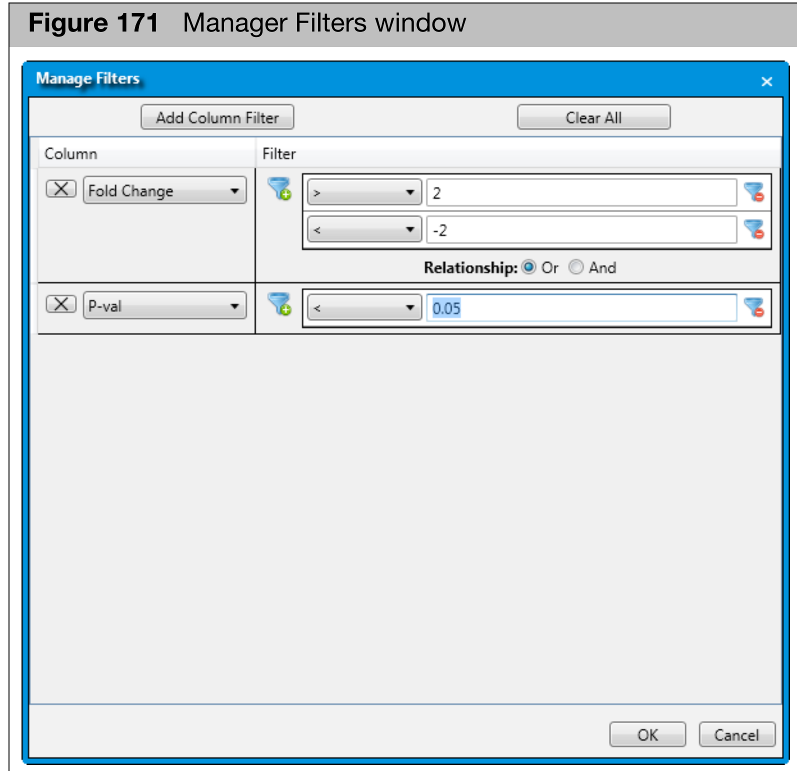
The column data is now ready for pasting (Ctrl v).

Managing filters

1. Click the Filters drop-down, then select Manage Filters.

   The Manage Filters window appears. (Figure 171)

Figure 171 Manager Filters window



2. Click **Add Column Filter**.

   A new column drop-down field appears.
3. Select a name from the drop-down.

   Click the **Or** or **And** radio button to choose Or or AND logic.
4. Select the operator and enter the value to filter on.
5. Optional: Click to add filter(s).
6. Optional: Click to remove filter(s).
7. Click **OK**.

   To remove all displayed filters in the Manage Filters window, click **Clear All** or click **Clear Current Filter(s)** from the Table's Filters drop-down.

Table filters

The Table can be filtered to either hide or show highlighted rows.

- From the Filters drop down:
  - Select **Filter Table by Selection** to display only highlighted rows in the table
  - Select **Filter Table by Selection (Exclude)** to hide highlighted rows while displayed all unselected rows.

Clearing filters

1. Right-click on the filtered column you want to clear.

   A menu appears.

2. Click **Clear Current Column Filter**.

   The filter is removed.

Copying selected rows and IDs

Copying selected row(s)

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click to select **Copy Selected Row(s)**. (Figure 172)

Figure 172 Copy Selected Row(s) option

The selected gene level information is now copied to the Windows Clipboard for pasting.

Copying selected cell(s)

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Right-click, then click **Copy Selected Cell(s)**.

   The selected cells are now copied to the Windows Clipboard for pasting.

Copying DNA sequence (5' to 3')

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Click **Copy DNA Sequence (5′ to 3′)**.

   The selected DNA Sequence (5′ to 3′) is now copied to the Windows Clipboard for pasting.

   Note: DNA sequences are obtained from server using start and stop positions. Sequences are represented from 5′ to 3′ for genes/exons/PSRs on both positive strand and negative strand.

Accessing external databases

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. To link out to various external databases, right-click on a TC ID of interest.
3. Click to select the external database you want to visit. (Figure 173)

Figure 173 External Databases

Your Internet browser opens to the appropriate website.

Searching the NetAffx website

1. Click **Search NetAffx**.

   The Internet browser opens to the NetAffx Customer Login window.
2. Enter your NetAffx **Email ID** and **Password**, then click **Submit**.

   Your Internet browser opens to the **NetAffx Query Center** and displays information about your gene of interest

   Note: The NetAffx Query Center is compatible with Windows Internet Explorer and Firefox. Chrome is not supported at this time. If a Probe Set or Transcript Cluster is not available, an appropriate message appears.

Searching keywords

By default, the search tool returns matches that contain your (case insensitive) search inquiry. Use wild-card (*) characters to aid in your search.

– Example: ABC* = Any string that begins with 'ABC'. Use double-quotes ["] to search for EXACT cell content matches.
– Example: The search "C3P1" will only return results that match exactly with "C3P1".

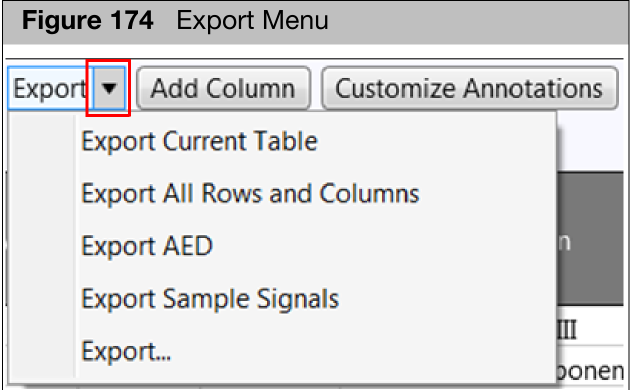
1. Click inside the **Find in Table** field, then type your keyword.
2. Click the ⌄ or ⌃ buttons to search.

Exporting options

1. To save your table, click **Export**. (Figure 174)

   The Export menu appears.

Figure 174 Export Menu



Exporting the current table

1. Click **Export Current Table** to export only the data currently shown in the Gene Table.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click Save.

   The table is now saved as a TXT file.

Exporting all rows and columns

1. Click **Export All Rows and Columns** to export all the data in the Gene Table including hidden data.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click Save.

   The table is now saved as a TXT file.

Exporting AED

Affymetrix Extensible Data (AED) files contain data that annotate positions on a genome and can be edited using Chromosome Analysis Suite (ChAS) software. For more information on AED files, see the ChAS User Manual (P/N 702943).

1. Click **Export AED**.
2. The Save As window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click Save. The table is now saved as a AED file.

Exporting sample signals

To export the signal for each transcript cluster ID per sample as a txt file:

1. Click **Export Sample Signals**.

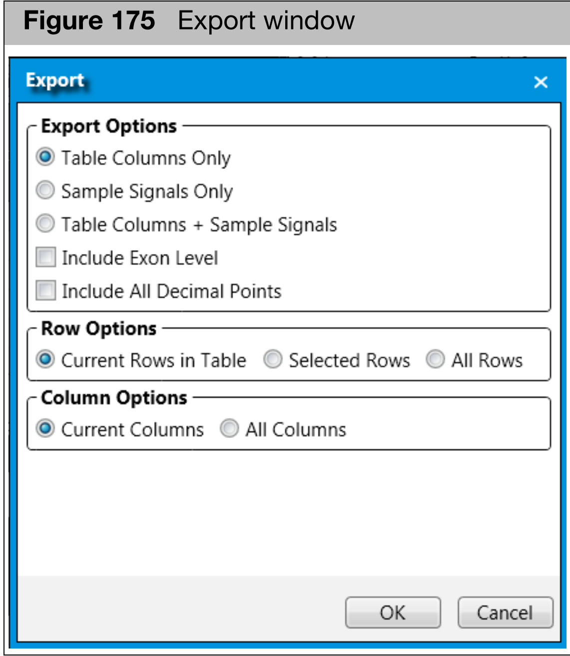
   The Save as window appears.
2. Click on an existing folder or click **New Folder** to choose a new save location.

Type a filename for the TXT file, click **Save**.

Custom export

1. To customize the export based on either current or all rows and/or columns click **Export...**
2. The Export window appears. (Figure 175)

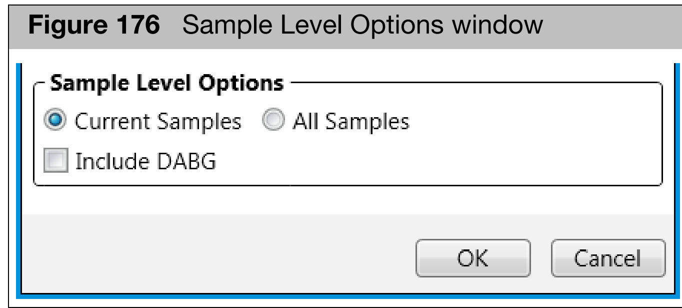
Figure 175 Export window



3. Select a **Export Options** radio button:
   – **Table Columns Only**: Exports the currently visible table columns.
   – **Sample Signals Only**: Exports the currently visible samples.
   – **Table Columns + Sample Signals**: Exports currently visible table columns and sample signals.
   – **Include Exon Level**: Click this check box to include Exon data.

Note: When selecting either **Sample Signals Only** or **Table Columns + Sample Signal**, a Sample Level Options window appears. (Figure 176)

Figure 176 Sample Level Options window



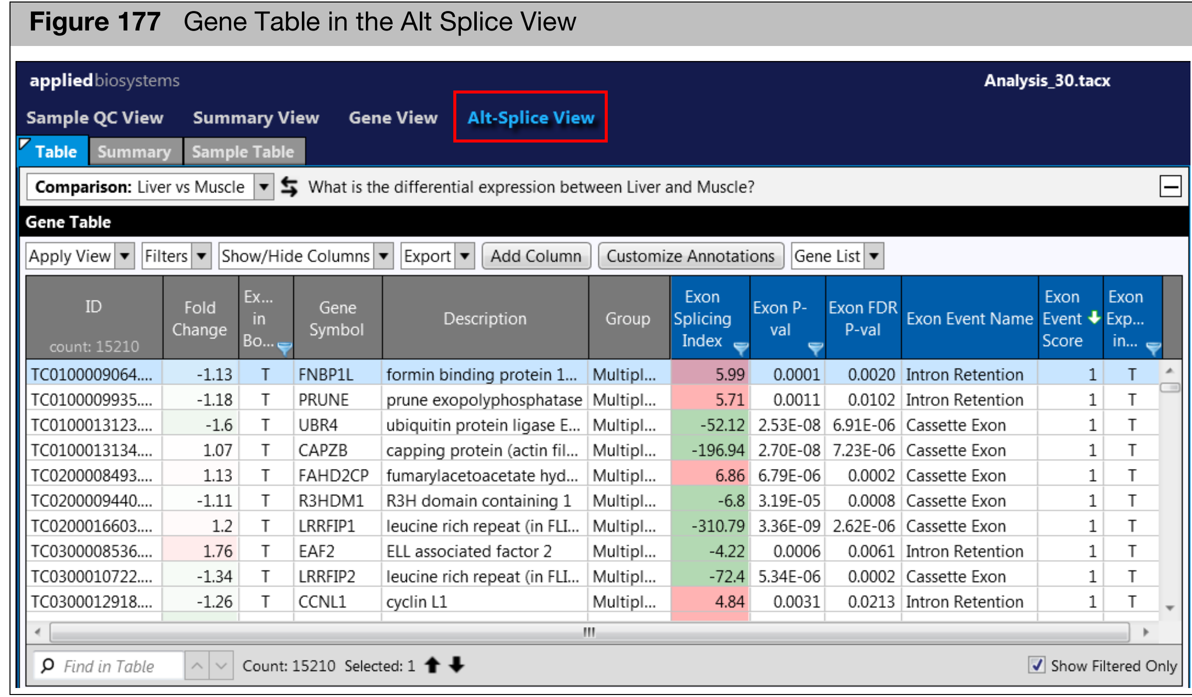
4. Optional: To export the visible samples, click the **Current Samples** radio button. To export all samples, click the **All Samples** radio button. To include DABG values in the export, Click the **Include DABG** check box, then click **OK**.
5. Optional: Click the **Include All Decimal Points** check box to export all decimal points available for each metric.

   Note: If unchecked, decimal values appear as two digits.
6. Select a **Row Options** radio button:
   - **Current Rows in Table**: Exports the currently visible table rows.
   - **Selected Rows**: Exports the currently selected (highlighted) rows.
   - **All Rows**: Exports all rows visible and hidden.
7. Select a **Column Options** radio button:
   - **Current Columns**: Exports currently displayed columns.
   - **All Columns**: Exports all columns (displayed or not).
8. Click **OK**.

   The Save as window appears.
9. Click on an existing folder or click **New Folder** to choose a new save location.
10. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

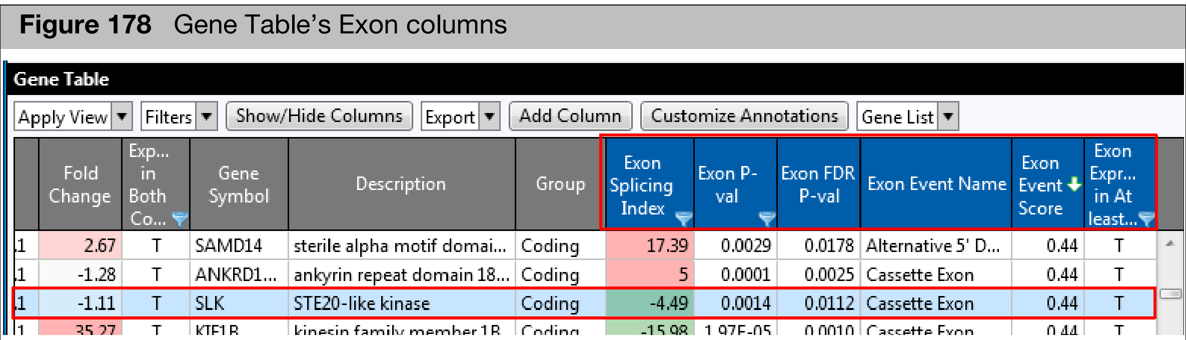
Using the Gene table in Alt-Splice view

Figure 177 Gene Table in the Alt Splice View



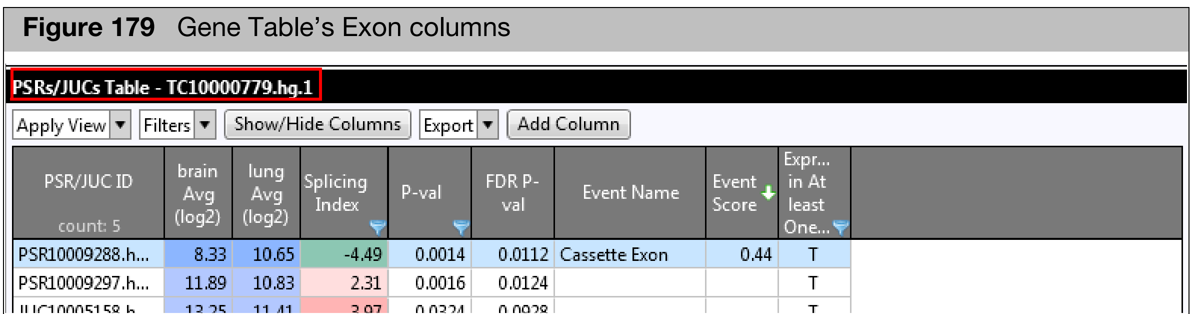
The Gene Table's blue Exon column headers are based on the **PSRs/JUCs** within the gene, as shown in Figure 177. **Note:** Changing the Gene Table's sorting and filtering criteria, may alter the displayed content of the columns.

Figure 178 Gene Table's Exon columns



The lower PSRs/JUCs Table (Figure 179) is a reflection of the highlighted entry in the upper Gene Table.

Figure 179 Gene Table's Exon columns



Viewing options for analyses with multiple conditions

1. Click the Comparison drop-down to select the comparison you want to change, as shown in Figure 180.

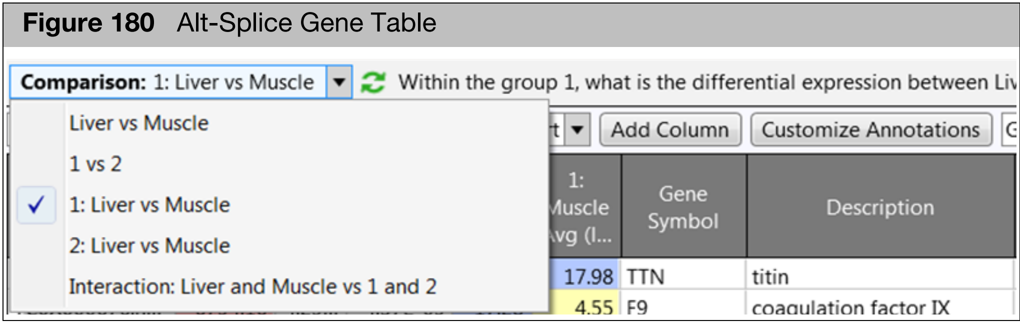


Figure 180 Alt-Splice Gene Table

For a Comparison analysis with more than two factor values, for example, comparing four different tissue types, select to view one of the comparisons (Figure 181) or choose All Conditions with the option to set one as the baseline, as shown in Figure 182.

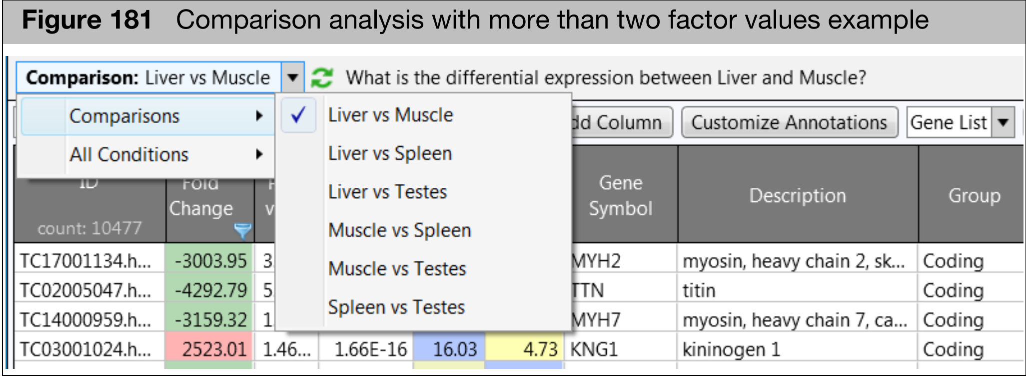


Figure 181 Comparison analysis with more than two factor values example

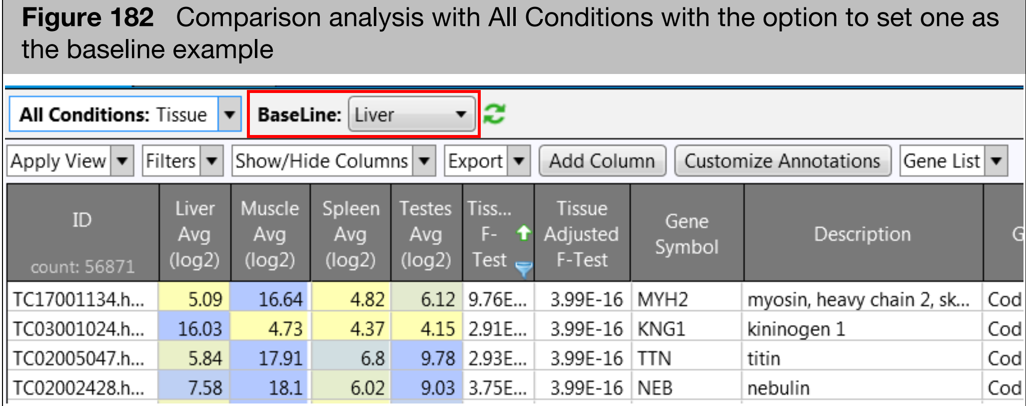


Figure 182 Comparison analysis with All Conditions with the option to set one as the baseline example

All conditions (tissue type in the example above) are listed. One Condition (Liver) is set as the baseline which provides the fold change columns for the remaining conditions in comparison with the baseline.

Showing or hiding columns in the Alt-Splice Gene table

1. Click the **Show/Hide Columns** drop-down menu.
2. Click the check box adjacent to the column you want to show in the table. Uncheck the check box to the column you want to hide.

   See the table below for column definitions.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

Available Columns	Description
ID	ID of Transcript Cluster (TC)
Fold Change	Transcript cluster's fold change (in linear space) for Condition 1 vs Condition 2.
Delta Splicing Index	Change in the component splicing indices associated with the interaction. If you convert the displayed component splicing indices to a fractional value (fractional_splicing_index = displayed_splicing_index if displayed_splicing_index >= 1 and 1 / |displayed_splicing_index| if displayed_splicing_index <= -1), the fractional_delta_splicing_index = fractional_splicing_index(condition 1) / fractional_splicing_index(condition 2). The displayed_delta_splicing_index = fractional_delta_splicing_index if fractional_delta_splicing_index >=1 and -1 / fractional_delta_splicing_index if fractional_delta_splicing_index <1. **Note:** This column is only shown for Interaction Comparisons.
Delta Fold Change	Change in the component fold changes associated with the interaction. If you convert the displayed component fold changes to a fractional value (fractional_fold_change = displayed_fold_change if displayed_fold_change >= 1 and 1 / |displayed_fold_change| if displayed_fold_change <= -1), the fractional_delta_fold_change = fractional_fold_change(condition 1) / fractional_fold_change(condition 2). The displayed_delta_fold_change = fractional_delta_fold_change if fractional_delta_fold_change >=1 and -1 / fractional_delta_fold_change if fractional_delta_fold_change <1. **Note:** This column is only shown for Interaction Comparisons.
P-val	The probability that there is no actual difference in this transcript cluster's expression and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference.
FDR P-val	False discovery rate corrected p-value. The multiple testing correction is based on the approach of Benjamini, Hochberg, and Yekutieli. The computation of the correction includes all transcript clusters.
Condition 1 Expressed	Whether this transcript cluster is expressed in a condition (based on percentage of samples that meet the specified DABG cutoff).
Condition 2 Expressed	Whether this transcript cluster is expressed in a condition (based on percentage of samples that meet the specified DABG cutoff).
Condition 1 Avg (log2)	Tukey-biweight robust average of this transcript cluster's expression level for the samples in condition 1.

Available Columns	Description
Condition 2 Avg (log2)	Tukey-biweight robust average of this transcript cluster's expression level for the samples in condition 2.
Condition 1 Standard Deviation	Standard deviation of this transcript cluster's expression level for the samples in condition 1.
Condition 2 Standard Deviation	Standard deviation of this transcript cluster's expression level for the samples in condition 2.
Public Gene IDs	Public Gene IDs for this transcript cluster.
Gene Symbol	Gene symbol for this transcript cluster. **Note:** RefSeq gene symbol is listed as the first gene symbol (if there are more than 1 gene symbol). Also, a TC with no gene symbol, may be auto-assigned a public gene ID.
Description	Gene Description for this Transcript Cluster.
Chromosome	Chromosome for this transcript cluster. See Chromosome Naming Scheme[3] for a detailed description.
Strand	Describes the (+/-) strand location of the transcript cluster in the version of the genome assembly used at the annotation time.
Group	Whether this TC is coding or non-coding.
Start	The beginning genomic position of the transcript cluster.
Stop	The ending genomic position of the transcript cluster.
PSR/JUC ID	ID of the Probe Selection Region (PSR) or junction probeset selected from the PSRs/JUCs Table. The filtering and sorting settings of the PSRs/JUCs Table affect this entry.
Exon Fold Change	Fold change (in linear space) for Condition 1 vs. Condition 2 for the PSR or junction selected from the PSRs/JUCs Table. The filtering and sorting settings of the PSRs/JUCs Table affect this entry.
Exon P-val	The probability that there is no actual difference in the expression for the PSR or junction selected from the PSRs/JUCs Table and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference.
Exon FDR P-val	The probability that there is no actual difference in the expression for the PSR or junction selected from the PSRs/JUCs Table and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference.
Exon Splicing Index	Fold change (in linear space) of normalized expression levels for Condition 1 vs. Condition 2 of the PSR or junction selected from the PSRs/JUCs Table. The PSR or junction's expression level is normalized to the gene's expression level.
Exon Event Pointer Name	EventPointer alternative splicing event type. Possible types include cassette exon, alternative 5' donor site, alternative 3' acceptor site, intron retention, alternative first exon, alternative last exon, mutually exclusive exons, and complex event selected from the PSRs/JUCs Table.

Available Columns	Description
Exon Event Pointer P-val	EventPointer alternative splicing event p-value. This p-value is the probability that the splicing event is not differentially expressed and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference for this alternative splicing event selected from the PSRs/JUCs Table.
Exon Event Pointer ID	EventPointer index for the alternative splicing event associated with this PSR or junction (or most likely event if multiple events are associated). This index can help to identify multiple PSRs and junctions that are associated with a single splicing event selected from the PSRs/JUCs Table.
Exon Event Estimation Name	EventPointer alternative splicing event type. Possible types include cassette exon, alternative 5' donor site, alternative 3' acceptor site, intron retention, alternative first exon, alternative last exon, mutually exclusive exons, and complex event selected from the PSRs/JUCs Table.
Exon Event Estimation Score	Event estimation alternative splicing score. The score is a value from 0 to 1, where 1 represents a high likelihood for the event selected from the PSRs/JUCs Table.
Exon Event Estimation ID	Index for the alternative splicing event estimation associated with this PSR or junction selected from the PSRs/JUCs Table.
Condition 1 Exon Expressed	Whether the PSR or junction selected from the PSRs/JUCs Table is expressed in Condition 1 (based on percentage of samples that meet the specified DABG cutoff).
Condition 2 Exon Expressed	Whether the PSR or junction selected from the PSRs/JUCs Table is expressed in a Condition 2 (based on percentage of samples that meet the specified DABG cutoff).
Condition 1 Exon Avg (log2)	Tukey-biweight robust average of the expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 1.
Condition 2 Exon Avg (log2)	Tukey-biweight robust average of the expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 2.
Condition 1 Exon Normalized Avg (log2)	Normalized expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 1. The normalized expression level is the Tukey-biweight robust average of the PSR or junction's expression level minus the Tukey-biweight robust average of the corresponding gene's expression level.
Condition 2 Exon Normalized Avg (1og2)	Normalized expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 2. The normalized expression level is the Tukey-biweight robust average of the PSR or junction's expression level minus the Tukey-biweight robust average of the corresponding gene's expression level.
Condition 1 Exon Standard Deviation	Standard deviation of the expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 1.
Condition 2 Exon Standard Deviation	Standard deviation of the expression level of the PSR or junction selected from the PSRs/JUCs Table for the samples in Condition 2.
Exon Event Name	Consensus alternative splicing event type. Possible types include cassette exon, alternative 5' donor site, alternative 3' acceptor site, intron retention, alternative first exon, alternative last exon, mutually exclusive exons, and complex event selected from the PSRs/JUCs Table.
Exon Event Score	Consensus alternative splicing event score. The score is a value from 0 to 1, where 1 represents a high likelihood for the event selected from the PSRs/JUCs Table.

Available Columns	Description
Exon Event ID	Consensus index for the alternative splicing event associated with this PSR or junction (or most likely event if multiple events are associated). This index can help to identify multiple PSRs and junctions that are associated with a single splicing event selected from the PSRs/JUCs Table.
Gene Expressed in Both Conditions	Whether this transcript cluster is expressed in both conditions.
Exon Expressed in at least One Condition	Whether the PSR or junction selected from the PSRs/JUCs Table is expressed in at least one of the conditions.

Customizing annotations

1. To customize the Alt-Splice Gene Table's annotation column(s), click Customize Annotations.

   The Customize Annotations window appears. (Figure 183)

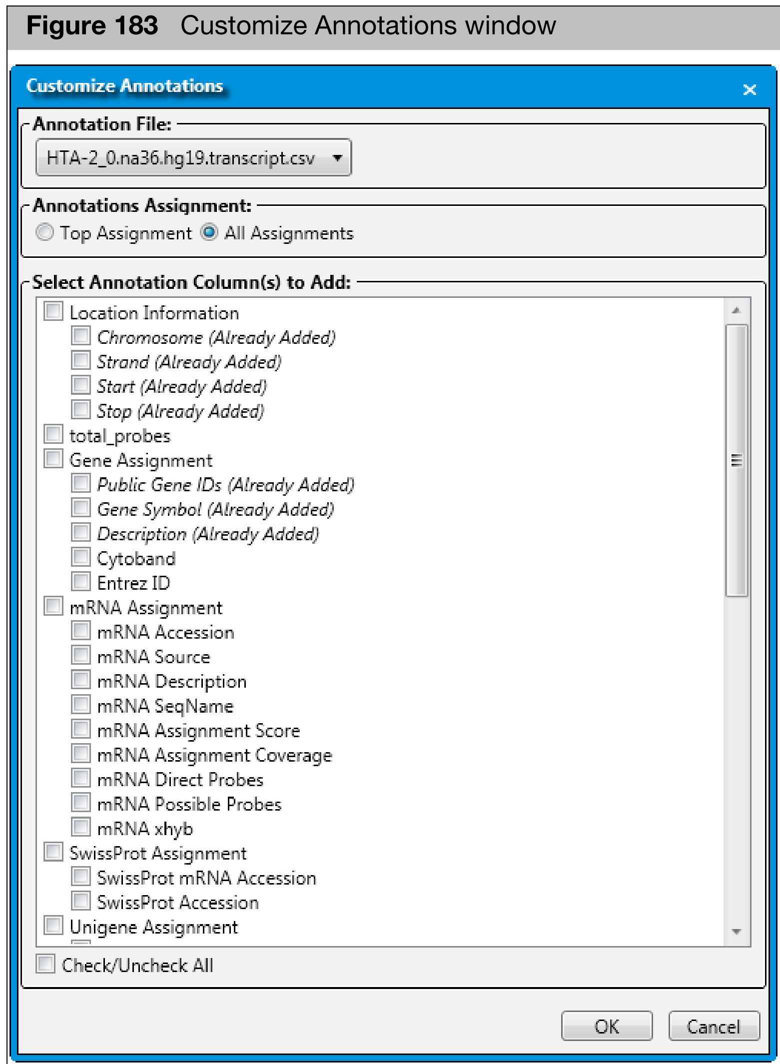


Figure 183 Customize Annotations window

Selecting an annotation file

Click the drop-down button to view and select available Annotation Files.

Selecting an annotation assignment

1. Click either the Top Assignment or All Assignments button to select your table/column display preference.
   - **Top Assignment** is the first subfield in an annotation column.
   - **All Assignments** displays all the subfields in an annotation column.

Adding annotation columns

1. Click the check box next to the annotation column(s) you want to add. Uncheck the check box to hide a specific column(s).

   Note: Column descriptions marked as **Already Added** denote the column already exists in your current table. No action (check mark) is required.

Check/uncheck all

1. Click the Check/Uncheck All check box to select or deselect all listed Annotations.

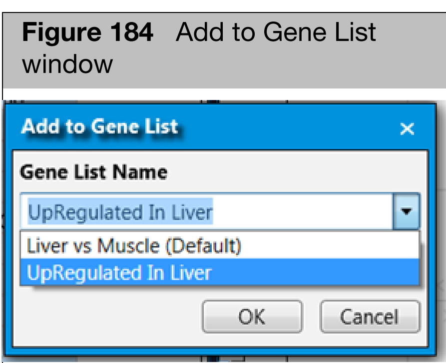
Adding/removing selected row(s) in a Gene list

Adding from right-click menu

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Add Selected Row(s) to a Gene List**.

   The Add to Gene List window appears. (Figure 184)

Figure 184 Add to Gene List window



3. Select a Gene List to add the selected row(s) from the drop-down or enter in the name of a new Gene list.

Removing from right-click menu

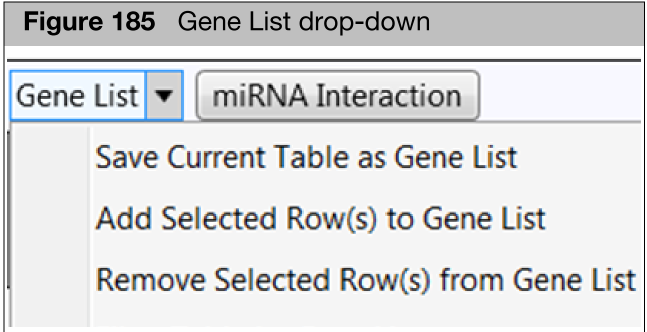
1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Remove Selected Row(s) from a Gene List**.

   A Remove Gene List window appears.
3. Use the drop-down to select the name you want to remove, then click **OK**.

From Gene list drop-down

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Click the **Gene List** drop-down, as shown in Figure 185.

Figure 185 Gene List drop-down



Saving the current table

1. Click **Save the Current Table as Gene List**.

   A Save Gene List window appears.
2. Enter a name, then click **OK**.

Adding a selected row(s)

1. Click to highlight a table row(s), then right-click.

   A menu appears.
2. Click **Add Selected Row(s) to Gene List**.

   An Add Gene List window appears.
3. Use the drop-down to select an existing name or enter a new name in the provided text field.
4. Click **OK**.

Removing a selected row(s)

1. Click to highlight a table row(s), then right-click.

   A menu appears.
2. Click **Remove Selected Row(s) from Gene List**.

   A Remove Gene List window appears.
3. Use the drop-down to select the name you want to remove, then click **OK**.

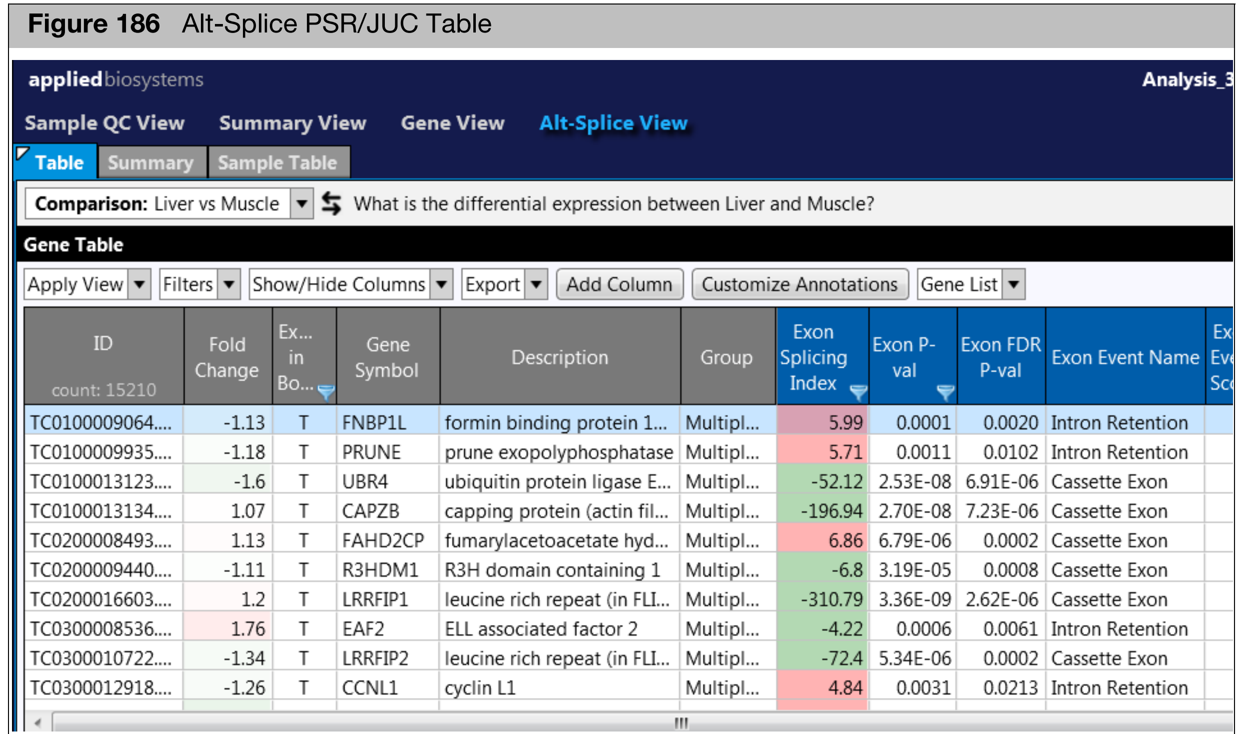
Filter table by Gene list

Use this filtering feature to show only the Gene List's row information.

1. Click on **Filter Table by Gene List**.
2. Select the Gene List you want from the Filter Table by Gene List window.
3. Click **OK**.

Using the Alt-Splice PSR/JUC table

Figure 186 Alt-Splice PSR/JUC Table



Showing or hiding columns in the Alt-Splice PSR/JUC table

1. Click the **Show/Hide Columns** drop-down menu.
2. Click the check box adjacent to the column you want to show in the table. Uncheck the check box to the column you want to hide.

   See the table below for column definitions.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

Available Columns	Description
PSR/JUC ID	ID of the Probe Selection Region (PSR) and Junction Probesets.
P-val	The probability that there is no actual difference in this PSR or junction's expression and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference.
FDR P-val	False discovery rate corrected p-value. The multiple testing correction is based on the approach of Benjamini, Hochberg, and Yekutieli.
Event Name	The content for the Event Name column is decided by the method that has the higher score. If event pointer has higher score than event estimation, then the column populates with results from the Event Pointer Name column, otherwise the column populates with results from Event Estimation Name.
Event Score	Consensus alternative splicing event score. The score is a value from 0 to 1, where 1 represents a high likelihood for the event.

Available Columns	Description
Event ID	Consensus index for the alternative splicing event associated with this PSR or junction (or most likely event if multiple events are associated). This index can help to identify multiple PSRs and junctions that are associated with a single splicing event. [KL: How is consensus ID determined?]
Overall F-test	The probability that there is no actual difference in this PSR or junction's expression for any of the comparisons and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference in at least one of the comparisons.
Overall FDR F-test	False discovery rate corrected F-test p-value. The multiple testing correction is based on the approach of Benjamini, Hochberg, and Yekutieli. The computation of the correction includes all PSRs and junctions, which, for any of the comparisons, are expressed in at least one of the conditions whose genes are expressed in both conditions.
Event Pointer P-val	EventPointer alternative splicing event p-value. This p-value is the probability that the splicing event is not differentially expressed and that any observed difference is due to variance in the measurement. Smaller values denote a more significant likelihood of expression difference for this alternative splicing event.
Exon Event Pointer Name	EventPointer alternative splicing event type. Possible types include cassette exon, alternative 5' donor site, alternative 3' acceptor site, intron retention, alternative first exon, alternative last exon, mutually exclusive exons, and complex event.
Exon Event Pointer ID	EventPointer index for the alternative splicing event associated with this PSR or junction (or most likely event if multiple events are associated). This index can help to identify multiple PSRs and junctions that are associated with a single splicing event.
Exon Event Estimation Name	Event estimation alternative splicing event type. Possible types include cassette exon, alternative 5' donor site, alternative 3' acceptor site, intron retention, alternative first exon, alternative last exon, mutually exclusive exons, and complex event.
Exon Event Estimation Score	Event estimation alternative splicing score. The score is a value from 0 to 1, where 1 represents a high likelihood for the event.
Exon Event Estimation ID	Index for the alternative splicing event estimation associated with this PSR or junction.
Fold Change	PSR or junction's fold change (in linear space) for Condition 1 vs. Condition 2.
Splicing Index	PSR or junction's fold change (in linear space) of normalized expression levels for Condition 1 vs. Condition 2. The PSR or junction's expression level is normalized to corresponding gene's expression level.
Condition 1 Expressed	Whether this PSR or junction is expressed in a condition (based on percentage of samples that meet the specified DABG cutoff).
Condition 2 Expressed	Whether this PSR or junction is expressed in a condition (based on percentage of samples that meet the specified DABG cutoff).
Condition 1 Avg (log2)	Tukey-biweight robust average of this PSR or junction's expression level for the samples in condition 1.
Condition 2 Avg (log2)	Tukey-biweight robust average of this PSR or junction's expression level for the samples in condition 2.
Condition 1 Standard Deviation	Standard deviation of this PSR or junction's expression level for the samples in condition 1.

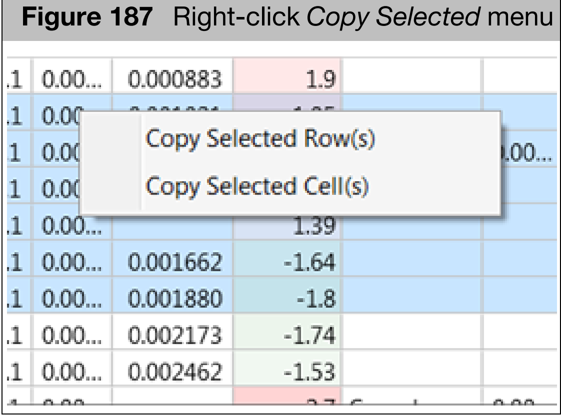
Available Columns	Description
Condition 2 Standard Deviation	Standard deviation of this PSR or junction's expression level for the samples in condition 2.
Condition 1 Normalized Avg (log2)	This PSR or junction's normalized expression level for the samples in condition 1. The normalized expression level is the Tukey-biweight robust average of the PSR or junction's expression level minus the Tukey-biweight robust average of the corresponding gene's expression level.
Condition 2 Normalized Avg (log2)	This PSR or junction's normalized expression level for the samples in condition 2. The normalized expression level is the Tukey-biweight robust average of the PSR or junction's expression level minus the Tukey-biweight robust average of the corresponding gene's expression level.
Expressed in at least One Condition	Whether this PSR or junction is expressed in at least one of the conditions.

Copying selected cell(s)

1. Click to highlight (light blue) a cell or Ctrl left-click to highlight multiple rows.
2. Right-click, then click **Copy Selected Cell(s)**. (Figure 187)

   The selected cells are now copied to the Windows Clipboard for pasting.

Figure 187 Right-click *Copy Selected* menu



Copying selected row(s)

1. Click to highlight (light blue) a row or **Ctrl** left-click to highlight multiple rows.
2. Right-click, then click **Copy Selected Row(s)**. (Figure 187)

Your selected rows are now copied to the Windows Clipboard for pasting.

Adding a column in the Alt-Splice table

1. Click **Add Column**.

   A New Column window appears.
2. Name the column, then click **OK**.

   The new column is added to the far right of the Table.

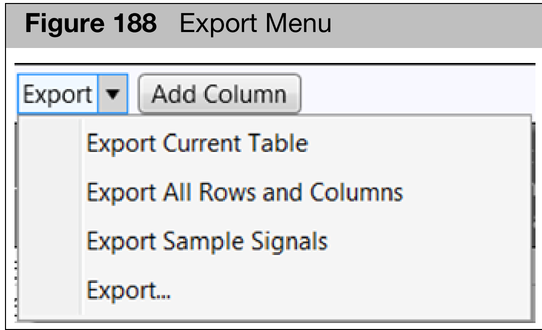
   Optional: Click and drag the column to another position in the table. The column can be removed by right-clicking on the column header and selecting **Delete Column**.
3. Optional: Enter any custom annotation you want into each cell.

Exporting options

If you want to export (Save) your table, click the **Export** drop-down.

The Export menu options appears. (Figure 188)

Figure 188 Export Menu



Exporting the current table

1. Click **Export Current Table** to export only the data currently shown in the PCS/JUC Table.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

Exporting all rows and columns

1. Click **Export All Rows and Columns** to export all the data in the PCS/JUC Table including hidden data.
2. The Save as window appears, click on an existing folder or click New Folder to choose a new save location.
3. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

Exporting sample signals

1. Click **Export Sample Signals** to export the signal data from PSR/JUC IDs table for all samples.

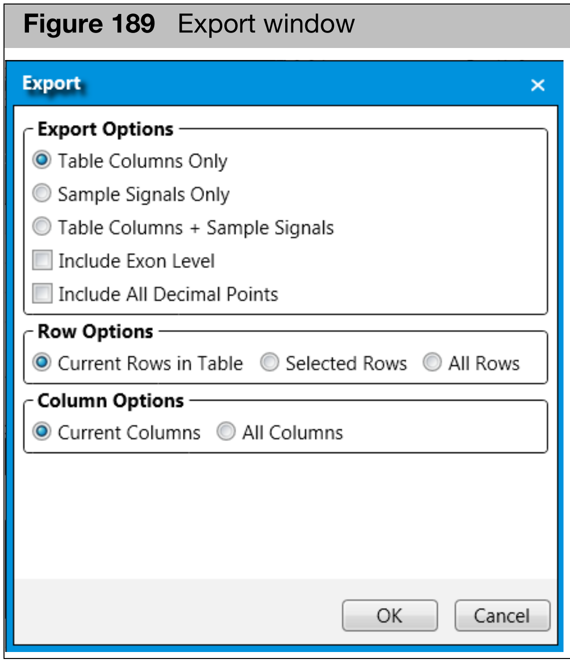
   The Save as window appears.
2. Click on an existing folder or click **New Folder** to choose a new save location.
3. Type a filename for the table, click **Save**.

   The table is now saved as a TXT file.

Custom export

1. To customize the export based on either current or all rows and/or columns click **Export…**
2. The Export window appears. (Figure 175)

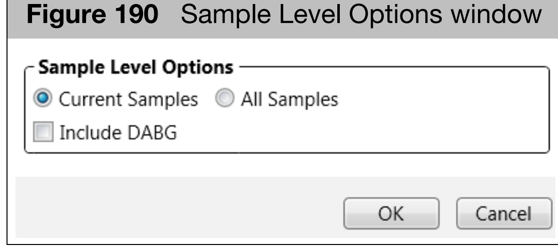
Figure 189 Export window



3. Select a **Export Options** radio button:
   - **Table Columns Only**: Exports the currently visible table columns.
   - **Sample Signals Only**: Exports the currently visible samples.
   - **Table Columns + Sample Signals**: Exports currently visible table columns and sample signals.
   - **Include Exon Level**: Click this check box to include Exon data.

Note: When selecting either **Sample Signals Only** or **Table Columns + Sample Signal**, a Sample Level Options window appears. (Figure 176)

Figure 190 Sample Level Options window



4. Optional: To export the visible samples, click the **Current Samples** radio button. To export all samples, click the **All Samples** radio button. To include DABG values in the export, Click the **Include DABG** check box, then click **OK**.
5. Optional: Click the **Include All Decimal Points** check box to export all decimal points available for each metric. **Note:** If unchecked, decimal values appear as two digits.
6. Select a **Row Options** radio button:

- **Current Rows in Table**: Exports the currently visible table rows.
- **Selected Rows**: Exports the currently selected (highlighted) rows.
- **All Rows**: Exports all rows visible and hidden.

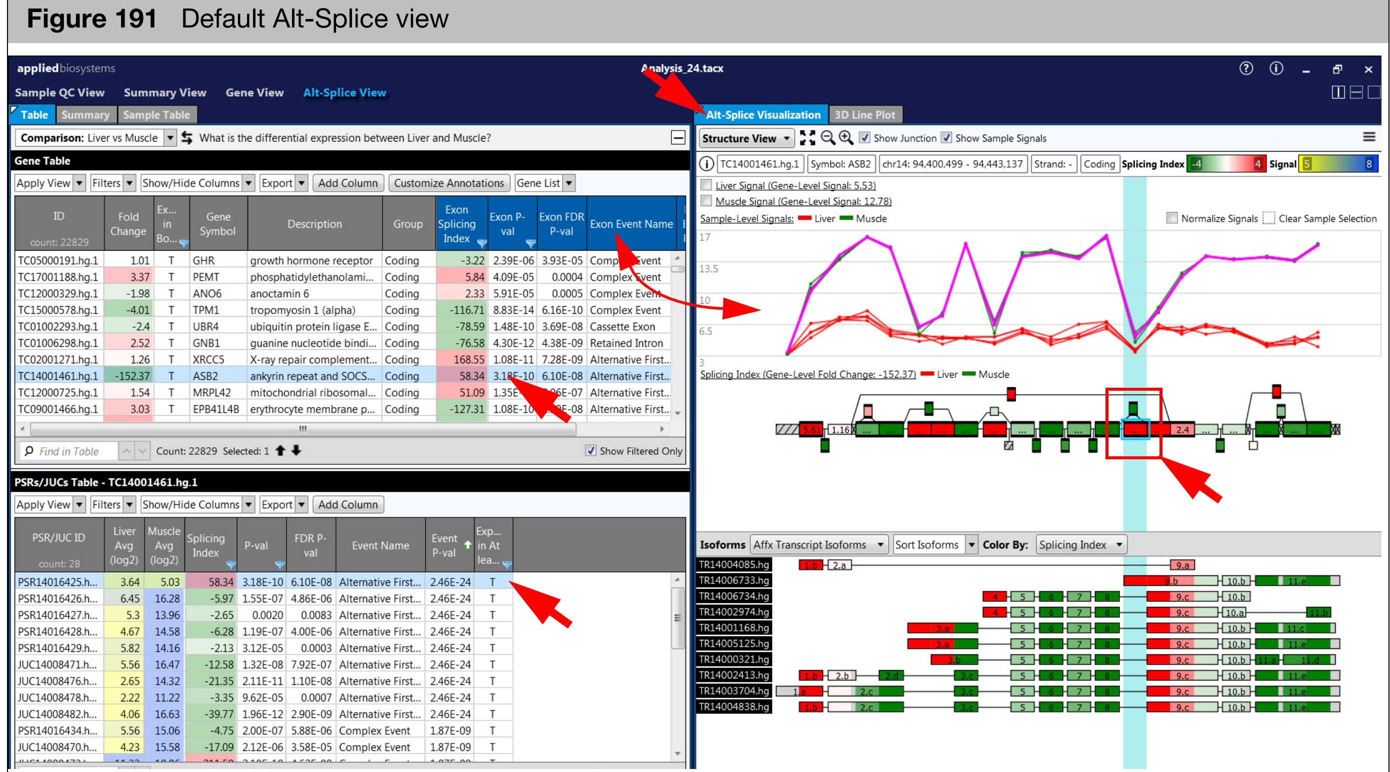
7. Select a **Column Options** radio button:
   - **Current Columns**: Exports currently displayed columns.
   - **All Columns**: Exports all columns (displayed or not).
8. Click **OK**.

   The Save as window appears.
9. Click on an existing folder or click **New Folder** to choose a new save location.
10. Type a filename for the table, then click **Save**. The table is now saved as a TXT file.

Using the Alt-Splice table and viewer

1. Single-click on any row(s) to highlight the PSR/JUC in the **Alt Splicing Visualization** window tab, as shown in Figure 191.

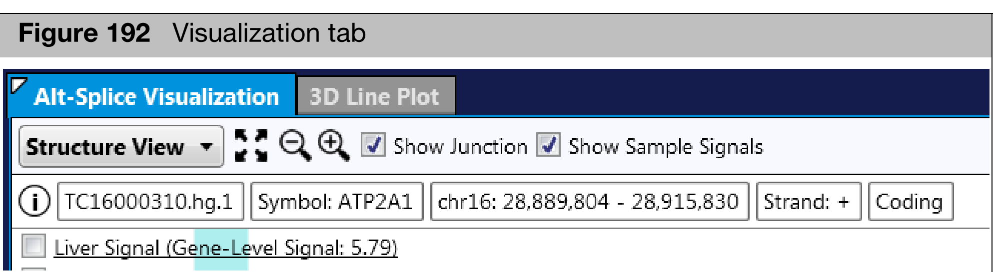
Figure 191 Default Alt-Splice view



Visualization tab features

- Splicing Index Scale
- Intensity Scale
- Vertical and Horizontal Split View buttons
- Save as PNG
- Print

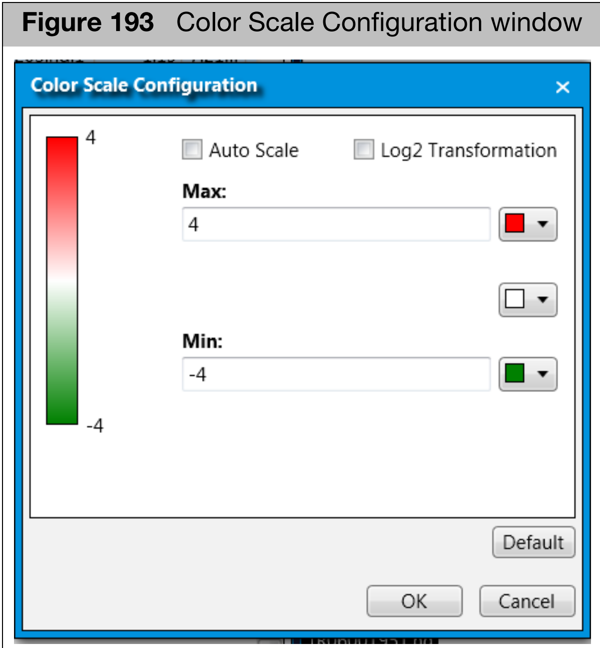
Figure 192 Visualization tab



Changing factory set scale limits for Splicing Index and Signal Scale

1. Click on the Splicing Index color bar (or Signal Color bar) to open the Configuration window.

Figure 193 Color Scale Configuration window



Setting new scale ranges

1. Click to check the **Auto Scale** check box to use the minimum (lower bound) and maximum (upper bound) splicing index values/intensities as the min/max scale for display. (Figure 193)
2. Click to un-check the Auto Scale check box to set a fixed scale, then enter your min and max number scales. This newly fixed scale is now saved for use with other TCs and genes.

   Note: For the Splicing Index scale the lower bound cannot be higher than -1. The upper bound cannot be lower than 1.

Changing scale colors

1. Use the drop-down color menus to change scale color properties.

Alt-Splicing visualization options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

To save the graph as a PNG

1. Click **Save as PNG**.
2. Choose a resolution.
3. Name, then select a location to save the PNG file.

Printing the currently displayed graph

1. Click **Print**, then print the graph using your configured printer as you normally would.

Opening the Genomic Region in the Alt Splice Visualization in an external browser

1. Click **Link to…**
2. Select UCSC or Ensemble browser.

   The default Internet browser will now open to your assigned link (based on the genomic location in Alt Splice Visualization).

Changing the Color Settings

1. Click **Color Settings**.
2. Select either Splicing Index or Signal.

   The Color Scale Configuration appears. See "Changing factory set scale limits for Splicing Index and Signal Scale" on page 177.

Copying Gene Position or Gene Info

1. Click **Copy…**
2. Select to copy either the Gene Position or Gene Info,

Showing/Hiding the ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn tooltips on. Uncheck to turn tooltips off.

To clear selected items in the graphs

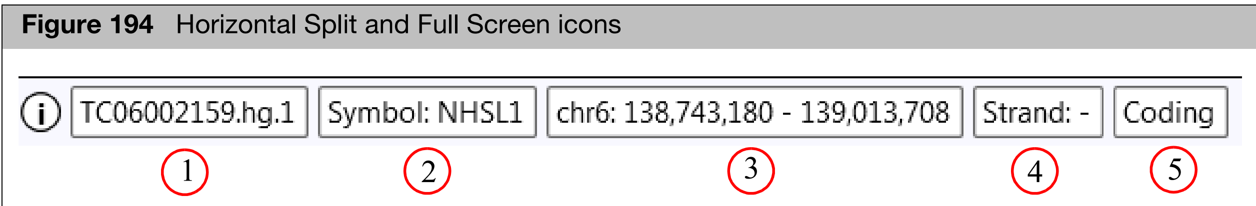
1. Click on **Clear Selection**.

   All previously selected items are now cleared from the graphs.

Transcript cluster details in the Alt Splice Visualization graph

The details for a selected Transcript Cluster ID in the Alt-Splice Gene Table are displayed at the top of the graph. (Figure 194)

Figure 194 Horizontal Split and Full Screen icons

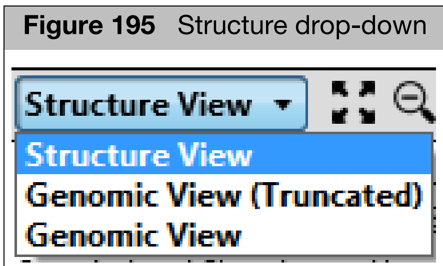


1. Transcript Cluster ID
2. Gene Symbol
3. Genomic position
4. Strand (+/-)
5. Group (See Table 2 on page 189 for more details)

Changing the Alt Splice Visualization display

1. Click the **Structure View** drop-down, as show in Figure 195 to change the display.

Figure 195 Structure drop-down



Structure Vview

The Structure View (Figure 196) displays gene structure. All PSRs and Junctions are represented in the structure view with boxes that have same size.

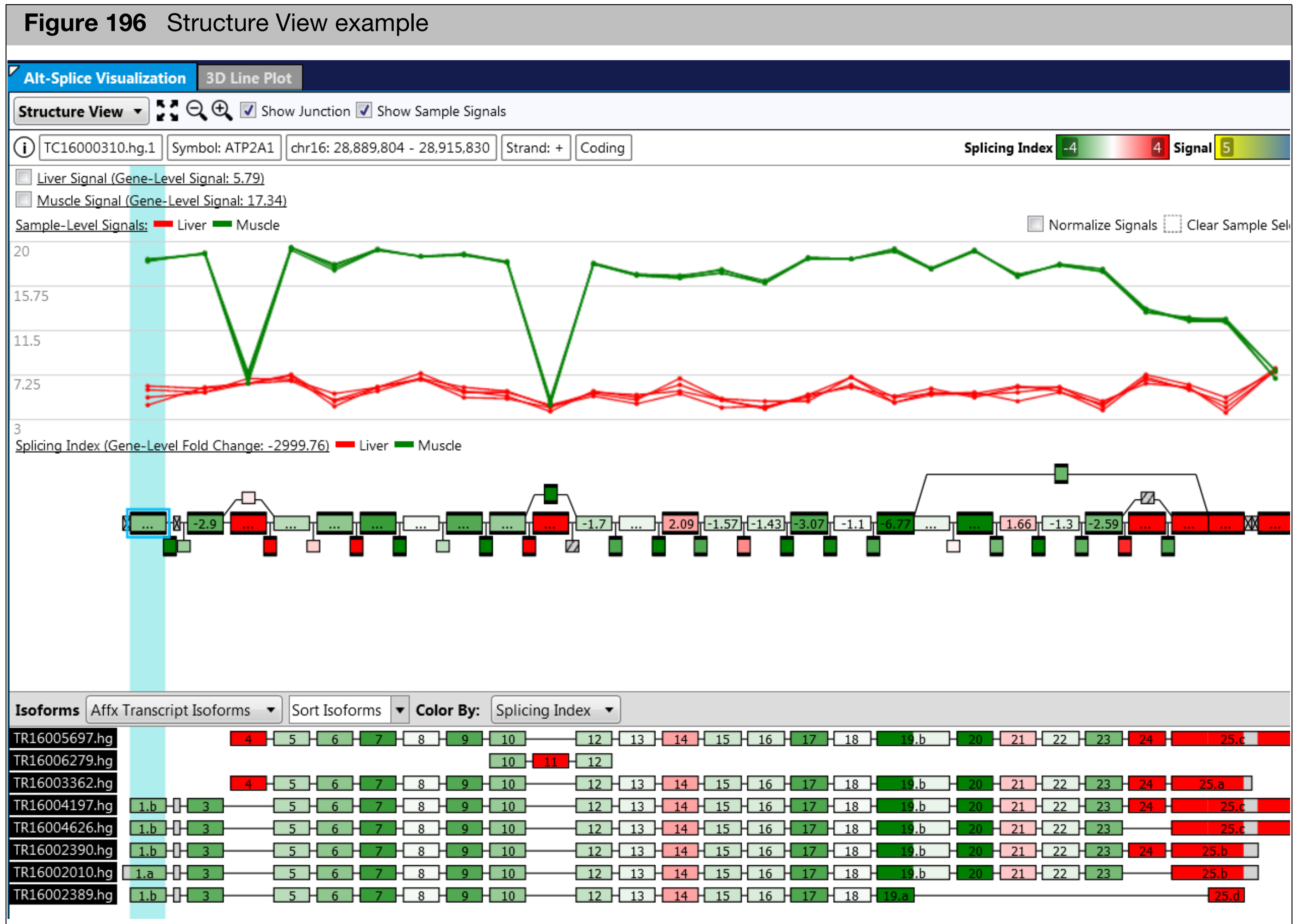


Figure 196 Structure View example

- An Inclusion junction detects 2 neighboring PSRs. The PSRs detected by an inclusion junction are linked and graphically represented as dotted lines when you mouse over or click that junction.
- An Exclusion junction detects PSRs that are apart from each other. The PSRs detected by an exclusion junction are linked and graphically represented as dotted lines when you mouse over or click that junction.
- A crossed-out box represents a PSR/Junction that does not contain data.
- A diagonally-crossed-out box represents a PSR/Junction that is not expressed in either condition.
- (bold top and bottom borders) represents a PSR/Junction that has passed through the current table's filtering criteria. This only applies to PSRs/Junctions currently filtered in the table.
- A **Spacer** represents a transcript cluster (TC) probe selection region (PSR) where a selection of probes is not possible. **Note:** Spacer is typically a region with less than 25 bases. Occasionally, some of them can exceed 25 bases.

Genomic view (truncated)

The Genomic View (Truncated) (Figure 197) maintains the relative size of the exons/ PSRs, but have spliced out intron regions that do not have any PSRs.

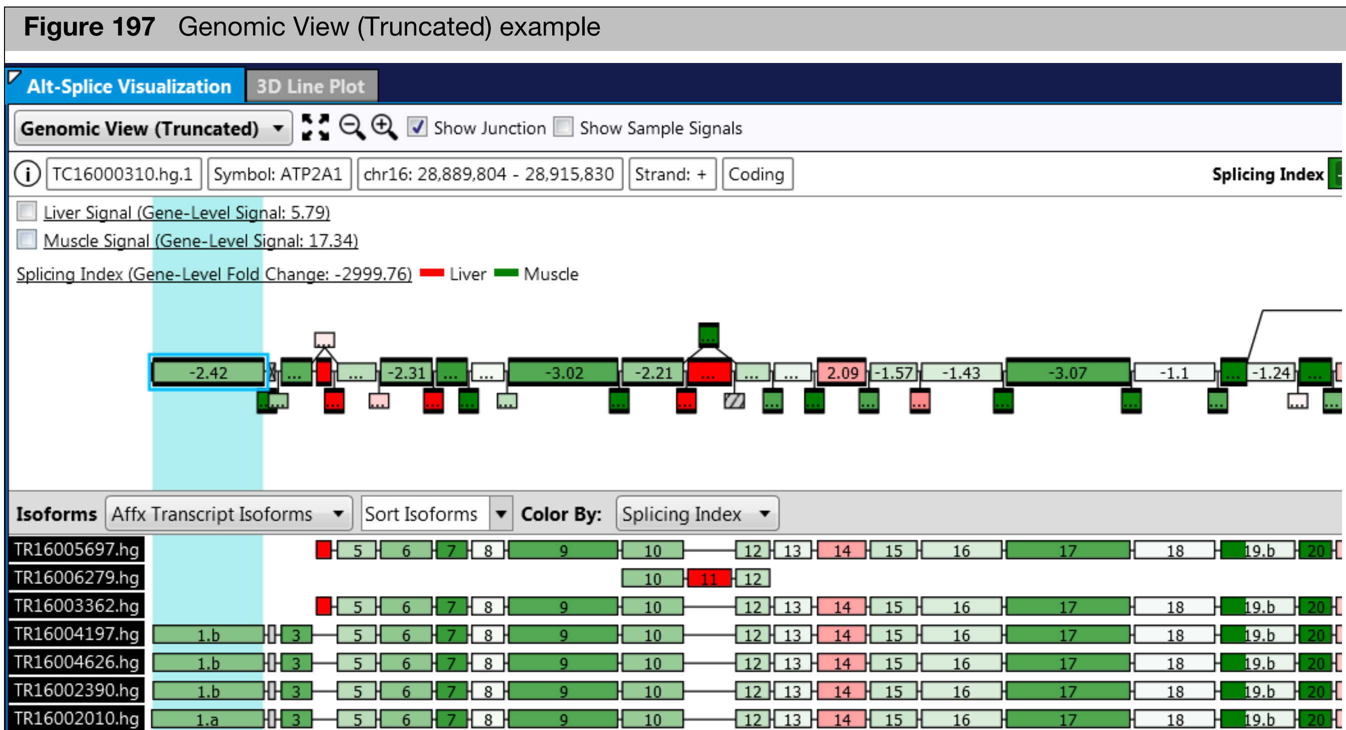


Figure 197 Genomic View (Truncated) example

Genomic view

The Genomic View (Figure 198) displays gene structure in a context of real genomic scale.

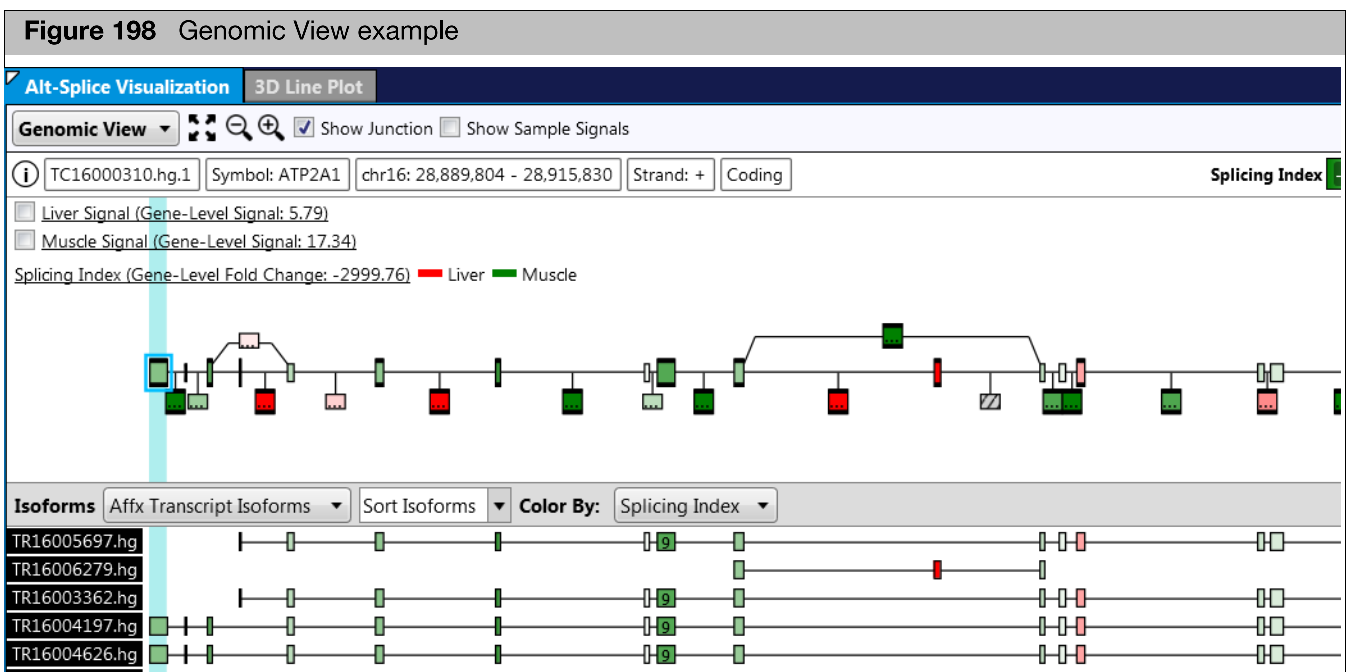
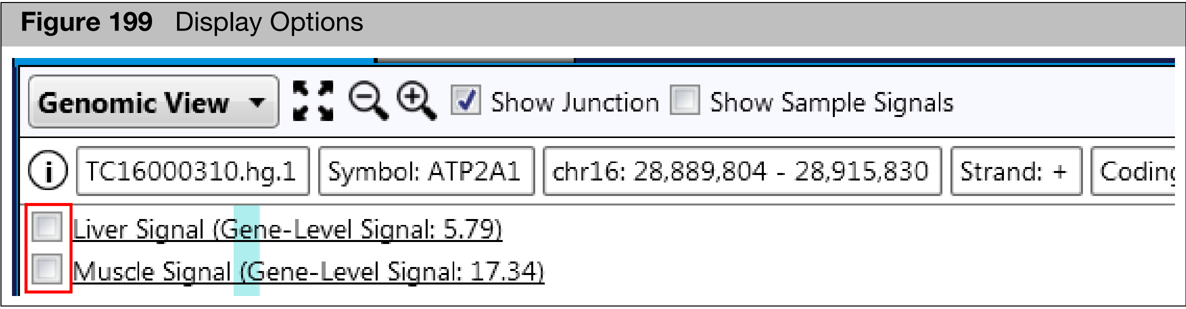


Figure 198 Genomic View example

- Each PSR/Junction probeset fold change direction is based on color and amount of fold change is based on the intensity of the color. These options are selected in the Genomic View's Splicing Index Color By drop-down menu.
- Below each gene structure, transcript isoforms are displayed in order. The order starts with the transcript isoforms that most likely exist in Condition1 to transcript isoforms that most likely exist in Condition2.
- The transcript isoforms that fit Condition1 are auto-sorted to the top. Transcript isoforms that fit Condition2 are auto-sorted to the bottom.

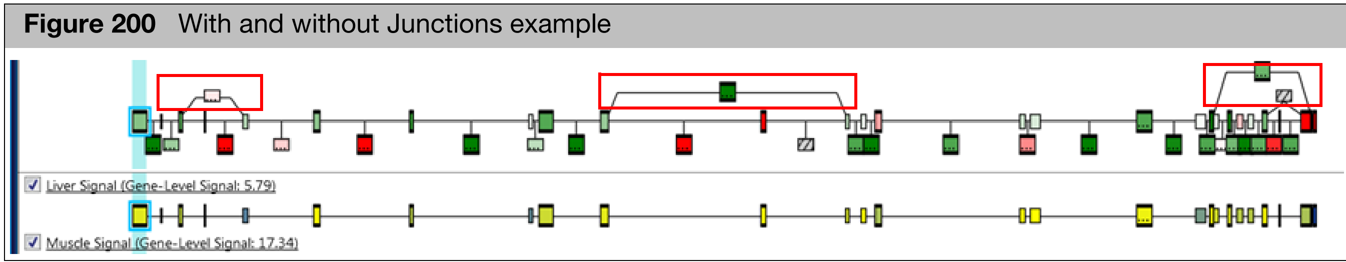
Visualization display options

Figure 199 Display Options



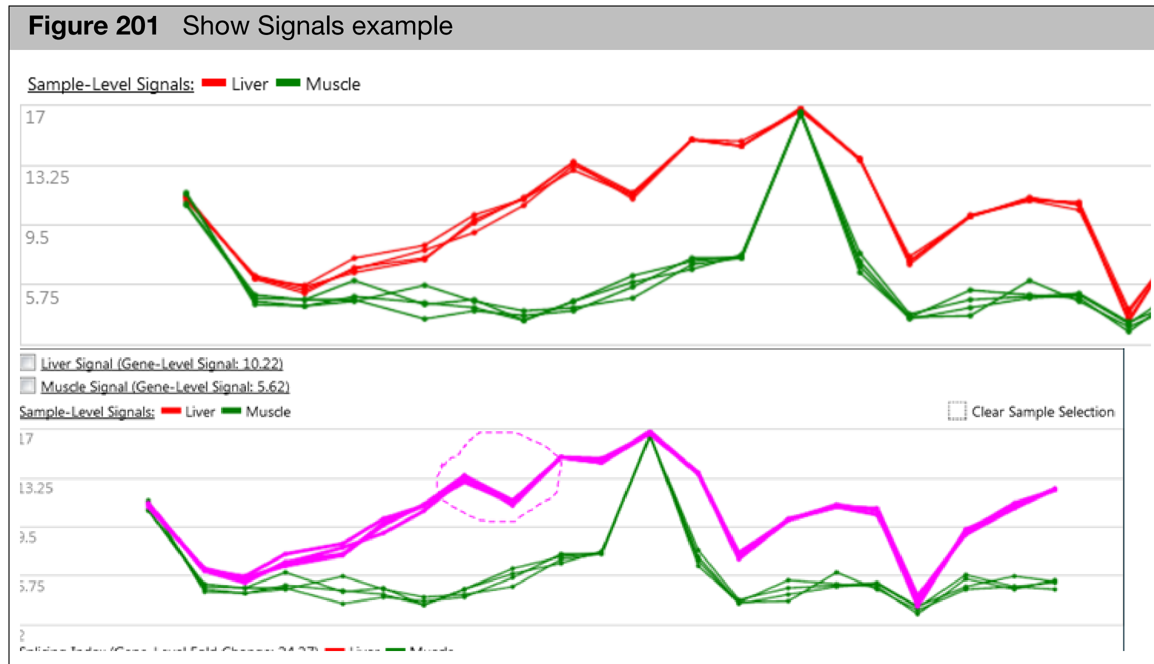
- Click to zoom in.
- Click to zoom out.
- Click to return to the default display.
- Click the **Show Junction** check box to show or hide Junctions. (Figure 200)
- Click the check box beside the Sample Signal name to show the sample signals with or without junctions. (Figure 200)

Figure 200 With and without Junctions example



- Click the **Show Sample Signals** check box to show or hide Signals.
- Left-click, then drag the mouse to lasso a selection of samples, as shown in Figure 200.

Figure 201 Show Signals example



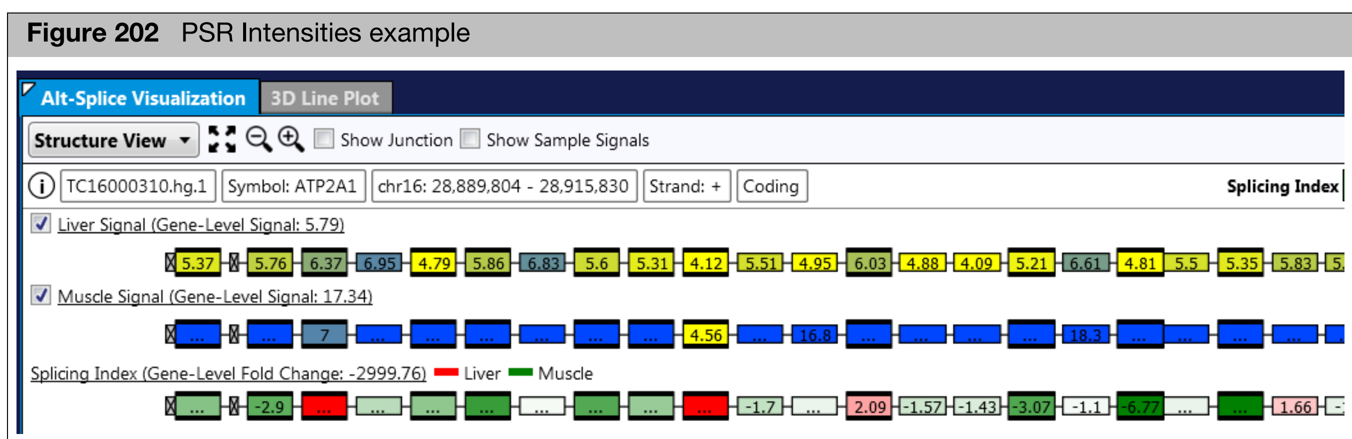
- Click **Clear Sample Selection** (upper right) to deselect all lassoed samples.
- Click the **Show Normalized Signals** check box to display the normalized Sample Signals.

  Note: To display the normalized signals, the Show Sample Signals must be checked.

Displaying the PSR intensities for each condition

- Select the Condition 1 Signal check box (example, Liver Signal in Figure 202) to display the intensities for the PSRs for the samples assigned to Condition 1. Note that the Show Junction is unchecked.
- Select the Condition 2 Signal check box (example, Muscle Signal in Figure 202) to display the intensities for the PSRs for the samples assigned to Condition 2. Note that the Show Junction unchecked.

Figure 202 PSR Intensities example



- Mouse over a PSR Splicing Index entry to reveal its tooltip (statistics), as shown in Figure 203.

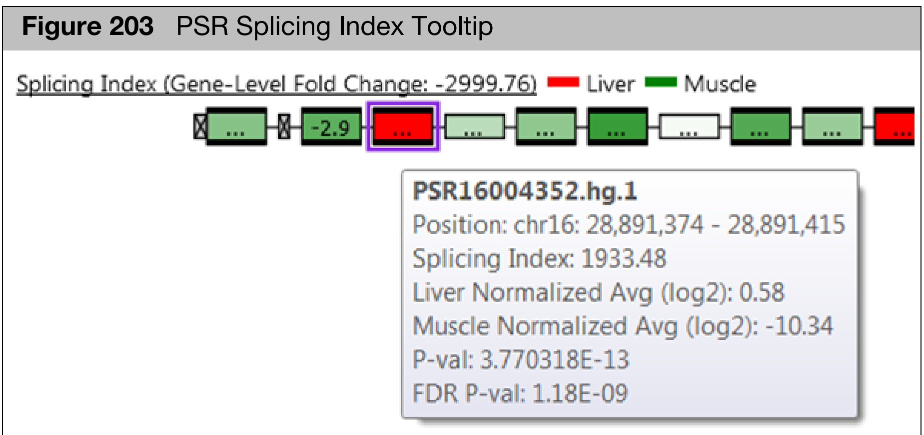


Figure 203 PSR Splicing Index Tooltip

Isoform display

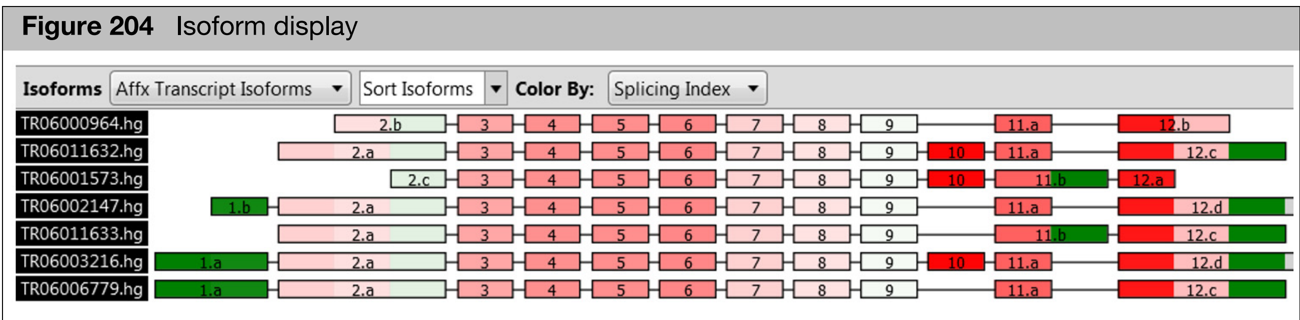


Figure 204 Isoform display

Isoform display options

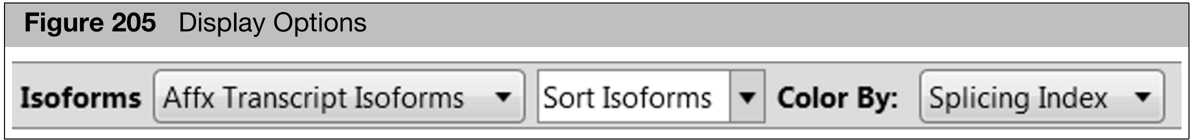


Figure 205 Display Options

Isoforms

Select a Transcript Isoform source

- Affx Transcript Isoforms
- Ensembl Transcripts
- RefSeq Genes
- Other

Sort Isoforms

- Up-Regulated on Top
- Down-regulated on Top

Color By

- Splicing Index
- Condition 1 Signal
- Condition 2 Signal

3D line plot

Click on the 3D Line Plot window tab. (Figure 206)

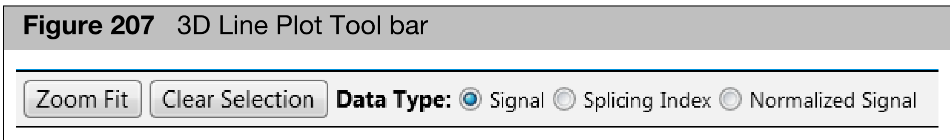
Figure 206 3D Line Plot view

The 3D Line Plot's header (upper left) displays the selected Transcript Cluster ID and corresponding PSR/JUC IDs in a 3-D line graph, including:

- TC ID
- Gene Symbol
- Description
- Chromosome
- Group: Coding, Non-Coding, etc.
- A legend (upper right) identifies the color assigned to each condition in the graph.
- Customized columns from the Gene Table.

3D Line plot tool bar

Figure 207 3D Line Plot Tool bar



- Click the **Zoom Fit** button (Figure 207) to Fit the graph to the window size
- Click the **Clear Selection** button to clear any selected PCR/JUC ID
- Click on the **Data Type** radio button view either:
  - **Signal**
  - **Splicing Index**
  - **Normalized Signals**

3D Line plot graph options

1. Click the Options ☰ button (upper right).

   The Options menu appears.

Save as PNG

Saves the currently displayed graph as a PNG. Select the resolution you want for your exported PNG.

- Low Quality (96dpi)
- Normal Quality (150dpi)
- High Quality (300dpi)
- Ultra Quality (600dpi)

Print

Prints the currently displayed graph.

Scale Settings

1. Click **Scale Settings**.

   The Scale Settings window appears.
2. Select the **Auto Scale** check box to set each axis automatically (based on the data selected).
3. Deselect the **Auto Scale** check box to enable the Y Min and Max value boxes, then enter the appropriate Min and Max values.
4. Click the **Show Y Grid** check box to display Y grid lines in the graph.
5. Optional: Click **Default** to return to the factory default settings

Showing/Hiding the ToolTip

The ToolTip feature enables you to mouse over a point of interest and view its details.

1. Click **Show ToolTip** check box to turn on. Uncheck to turn off.

Turning Graph Legend On/Offf

Click **Show Legend** check box to display the Legend. Uncheck to turn it off.

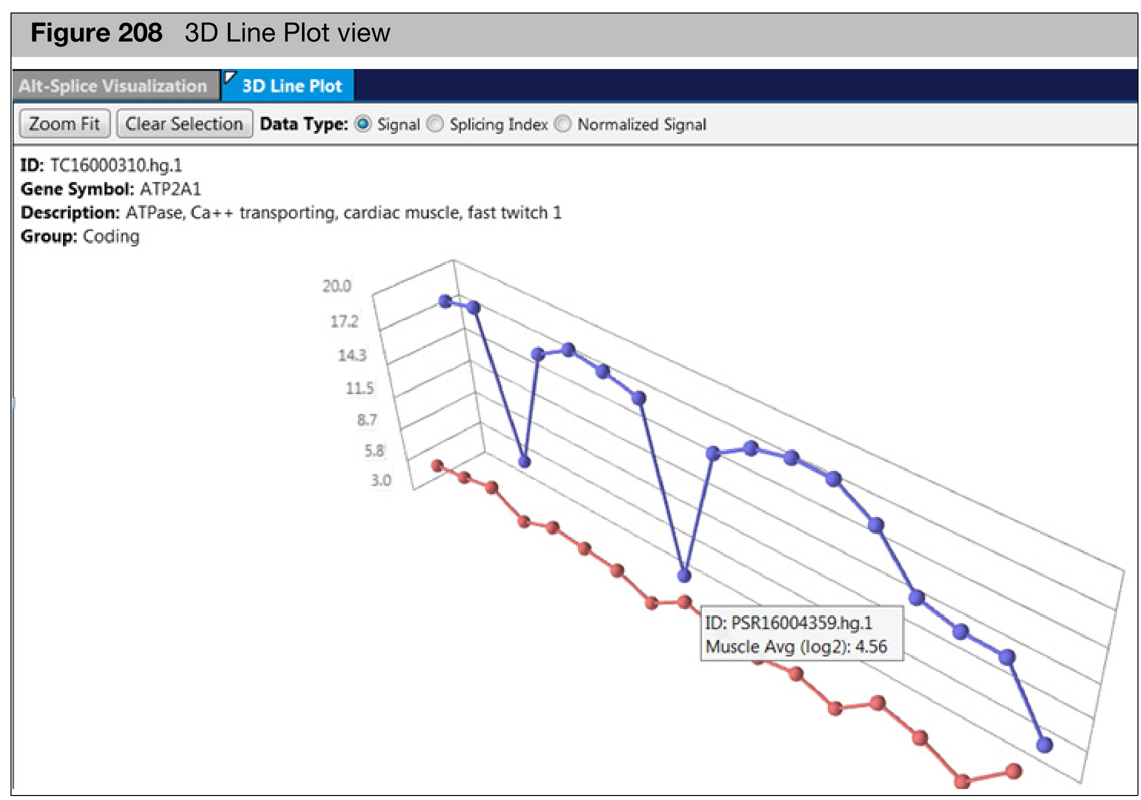
Clearing 3D Line Plots

1. Click on **Clear Selection**.

   All previously selected 3D Line Plots are now cleared from the graphs.

Working with the 3D Line plot

Figure 208 3D Line Plot view

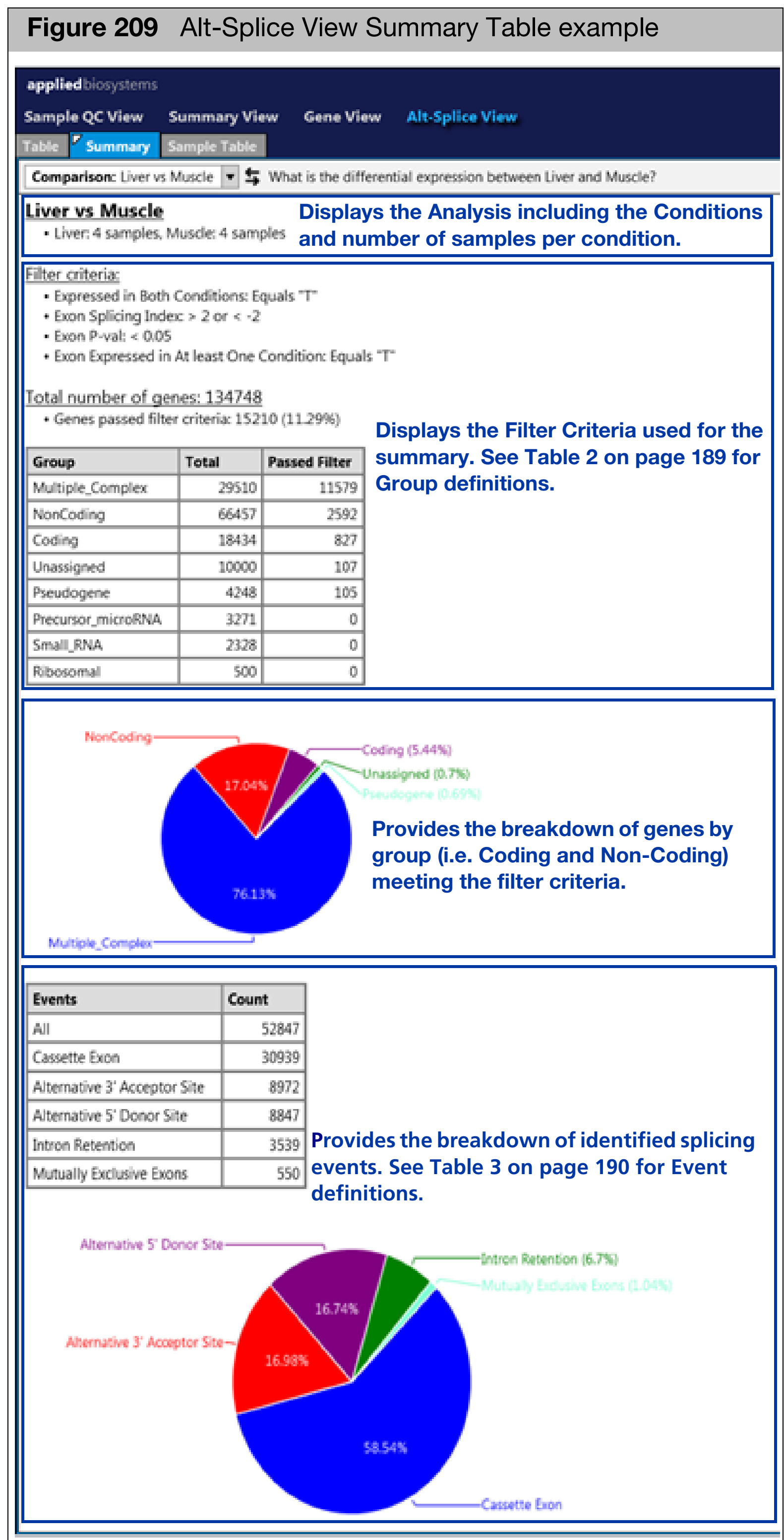


- To rotate the 3D Line Plot graph, hold down the left mouse button, then drag the mouse left or right.
- To pan the graph, hold down the right mouse button, then move the graph to the desired location.
- Mouse over a PSR region on the plot to display its details.

  Note: Show Tooltips must be enabled.
- Selecting a PSR data point highlights that region in the Alt-Splice visualization graph and Alt Splice Gene and PSR/JUC tables.

Alt-Splice view summary table

Figure 209 Alt-Splice View Summary Table example



How TAC 4.0.2 assigns Locus types

Step 1: Assign a locus type to every transcript in the gene. The locus types can be extracted using transcript-specific information from the data source as is done with RefSeq or can be applied to the entire data source if no transcript-specific information is available as is done with NONCODE. Some prioritization is done with respect to transcript-specific information compared to general data source information. For example, if RefSeq reports transcript-specific locus type as "tRNA" and NONCODE is generally considered "NonCoding", the transcript locus type will be set to "tRNA" since we have transcript-specific information from RefSeq and only general information from NONCODE.

Step 2: Assign a locus type for the gene based on the locus types from the individual transcripts. If more than one locus type is reported, then the locus type for the gene is assigned "Multiple Complex" except for the following conditions:

- One of the locus types is "Ribosomal" and the only other locus type is "NonCoding". In this case, the locus type becomes "Ribosomal". This is done because Ribosomal content tends to be highly expressed in every sample. These genes are highlighted appropriately.
- One of the locus types is "Unassigned". In this case, only the other locus type is reported.

Table 2 Group Filter Criteria used for the Summary

Group	Definition
Coding	Gene contains only protein-coding transcripts.
NonCoding	Gene contains only non-protein-coding transcripts. Most of these are long non-coding genes, but occasionally they may be small non-coding genes where the biotype cannot be determined.
Precursor-microRNA	Gene contains only microRNA precursors.
Pseudogene	Gene contains only represents pseudogenes. Note that some pseudogenes are transcribed and therefore can produce signal on probe sets.
Ribosomal	Gene contains only ribosomal RNAs, aka rRNA.
Small_RNA	Gene contains only small RNAs. These are small non-coding RNAs made up from several biotypes. Our software can only handle a limited number of different locus types as it becomes difficult to display or filter when many locus types are represented, therefore following biotypes are grouped into "Small_RNA", as follows: a. sRNA b. snRNA c. snoRNA d. scaRNA e. Y RNA f. guide RNA g. vault RNA h. ribozyme Ribosomal, tRNA, and Precursor-microRNA are not included here, because the frequency of each subset in Small_RNA tends to be much less than the separated ones. The idea here is that these small RNAs (including Ribosomal, Precursor-microRNA, tRNA) may not be captured/amplified very well with the assay, therefore the data from these probe sets should be treated with caution.
tRNA	Gene contains only transfer RNAs.
Multiple Complex	Gene contains more than one locus type above.
Unassigned	Locus type could not be determined based on source data.

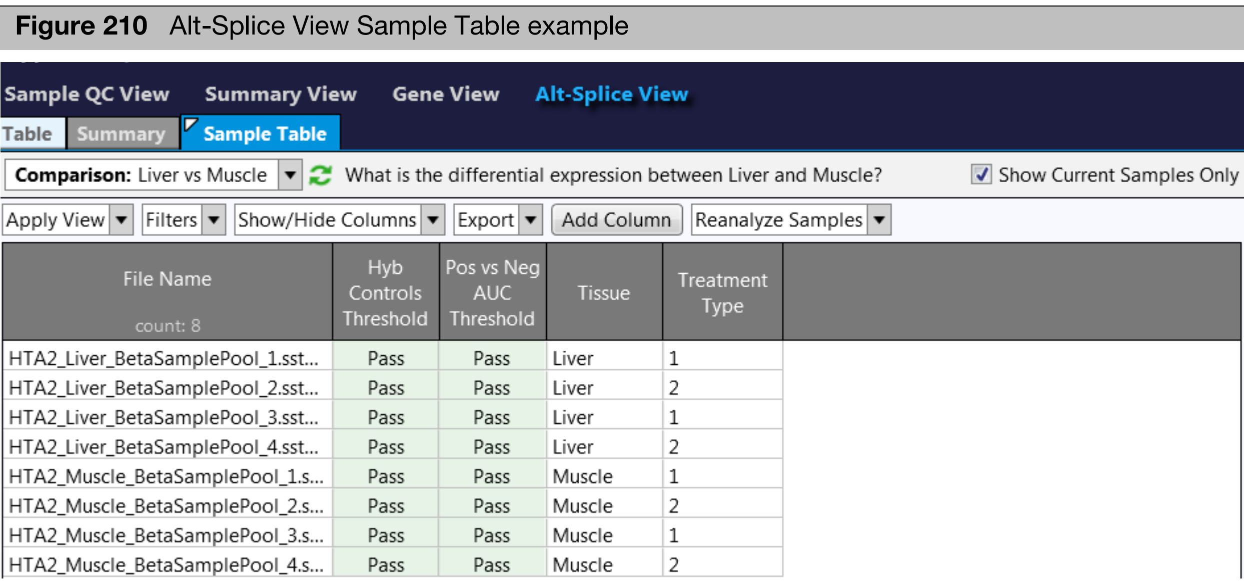
Table 3 Events and their descriptions

Event	Definition
Cassette Exon (CE) (Skipped Exon)	One exon is spliced out of the primary transcript together with its flanking introns. 5′ 3′
Intron Retention (IR)	A sequence is spliced out as an intron or remains in the mature mRNA transcript. 5′ 3′
Mutually Exclusive Exons (MXE)	Refer to a case in which multiple cassette exons are used in a mutually exclusive manner. In the simplest case: two consecutive exons that are never both included in the mature mRNA transcript. 5′ 3′
Alternative 3' Sites (A3SS)	Also called alternatively acceptor sites. Two or more splice sites are recognized at the 5' end of an exon. An alternative 3' splice junction (acceptor site) is used, changing the 5' boundary of the downstream exon. 5′ 3′
Alternative 5' Sites (A5SS)	Also called alternative donor sites. Two or more splice sites are recognized at the 3' end of an exon. An alternative 5' splice junction (donor site) is used, changing the 3' boundary of the upstream exon. 5′ 3′
Alternative First Exon (AFE)	Second exons of each variant have identical boundaries, but first exons are mutually exclusive. This is to annotate possible alternative promoter usage.[1] 5′ 3′
Alternative Last Exon (ALE)	Penultimate exons of each splice variant have identical boundaries, but last exons are mutually exclusive. This is to allow annotation of possible alternative polyadenylation usage.[1] 5′ 3′

[1] AFE and ALE event types are difficult to interpret and additional biological evidence is needed to support alternative promoter or polyadenylation usage, respectively.

Alt-Splice View sample table

Figure 210 Alt-Splice View Sample Table example



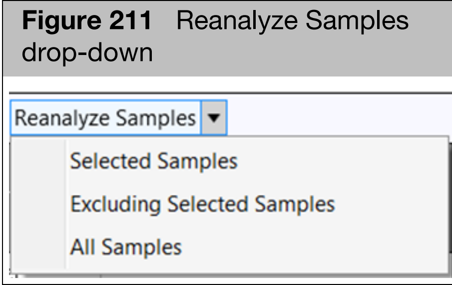
File Name (count: 8)	Hyb Controls Threshold	Pos vs Neg AUC Threshold	Tissue	Treatment Type
HTA2_Liver_BetaSamplePool_1.sst...	Pass	Pass	Liver	1
HTA2_Liver_BetaSamplePool_2.sst...	Pass	Pass	Liver	2
HTA2_Liver_BetaSamplePool_3.sst...	Pass	Pass	Liver	1
HTA2_Liver_BetaSamplePool_4.sst...	Pass	Pass	Liver	2
HTA2_Muscle_BetaSamplePool_1.s...	Pass	Pass	Muscle	1
HTA2_Muscle_BetaSamplePool_2.s...	Pass	Pass	Muscle	2
HTA2_Muscle_BetaSamplePool_3.s...	Pass	Pass	Muscle	1
HTA2_Muscle_BetaSamplePool_4.s...	Pass	Pass	Muscle	2

- For instructions on how to use the Sample Table's functions, see "Common table functions" on page 151.

If an outlier sample has been identified, the analysis can be re-run from the Sample Table resulting in the removal of that sample.

1. Click on the **Reanalyze Samples** drop-down menu. (Figure 211)

Figure 211 Reanalyze Samples drop-down

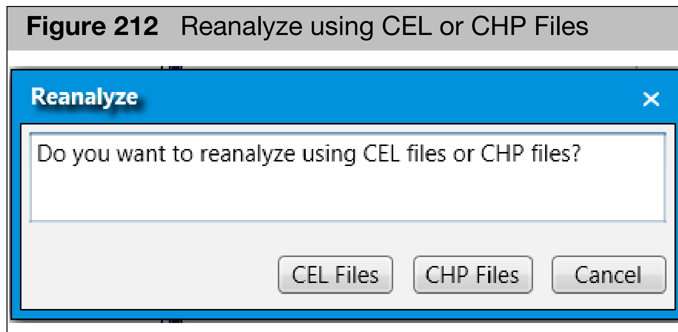


2. Select the appropriate Analysis.

- Click **Selected Samples** to re-run analysis using the highlighted samples
- Click **Excluding Selected Samples** to rerun the analysis on samples that are NOT highlighted
- Click **All Samples** to re-run the analysis with all sample but with the option to choose different criteria.

3. After selecting an analysis, choose to reanalyze from CEL or CHP files. (Figure 212)

Figure 212 Reanalyze using CEL or CHP Files

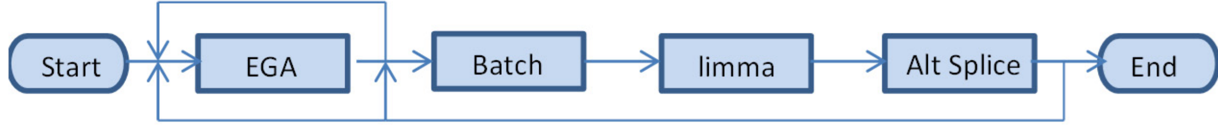


The New Analysis Window opens and includes the CEL or CHP files based on your analysis selection (steps 2 and 3).

4. Select the options for the New Analysis, then click **Run Analysis**.

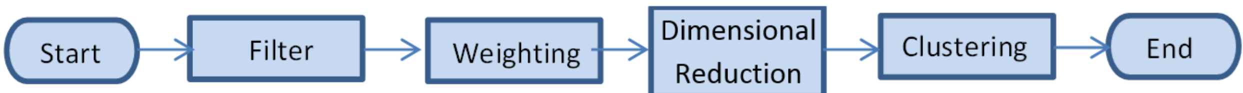
Algorithms

This appendix describes the algorithms provided by TAC 4.0.2. Most are implemented with a call to an R or Bioconductor package. While the default analysis flow involves limma and alternative splicing analysis for exon-level microarrays, several additional analysis modules enable additional ways to examine the inter-sample relationships and perform batch correction.



The preceding flow chart illustrates the overall algorithmic workflow in TAC 4.0.2 following the pre-processing of the CEL files to CHP files. The first optional step is the exploratory grouping analysis (EGA), which can be iterated. Next, batch effects module and limma differential expression analysis are performed. Finally, in the case of microarrays that include exon-level data, the alternative splicing analysis is run. After examining the results, the user can then go back and iterate the previous steps.

Exploratory Grouping Analysis (EGA) module



The Exploratory Grouping Analysis (EGA) module provides a variety of tools that enable the user to examine the relationships between the samples included in the experiment. Clearly, the experimental design suggests certain hypothesized relationships among the samples, but assay conditions can create outliers and batch effects. Furthermore, clinical samples can offer surprising differences and sub-clusters. The knowledge of the sample relationships can guide subsequent analysis; for example, outliers can be removed, batch effects corrected, and unexpected sub-clusters can be separated for analytical inspection. EGA supports a workflow with the following stages: gene filtering, gene weighting, dimensional reduction, and clustering. The results are presented as a 3-dimensional visualization and a set of cluster assignments. The Euclidean distance function is used throughout.

The gene filtering has several available controls. First, the user can identify a list of genes for use in the analysis. This approach is especially helpful if attention is focused on a certain pathway or genes of special interest. Second, the user can specify the number of maximally variant genes to be considered. Either way, the maximum

number of genes the EGA module can process is 20,000. The genes with the lowest variance tend to contribute noise to the computation that tends to damp the effect of the more strongly varying genes, so the maximum was chosen as a compromise that removes the noise-generating genes in the low variance tail.

Gene weighting involves the assignment of weight proportional to the square of the variance to each gene. The advantage of using such a weighting is that it focuses on the most variant genes without having to identify them directly or to specify their number.

Dimensional reduction involves projecting objects in a high-dimensional space into a reduced dimensional space that is easier for human perception. In this case, the intensities associated with the many genes in each sample represent the high-dimensional space. As the number of genes is typically thousands, humans could only look at a few genetic dimensions at a time. Dimensional reduction is a mechanism that merges the high-dimensional components (in this case, the gene intensities) into a smaller number of aggregate components (in this case three) that represent the highest variation present in the samples. EGA includes two algorithms for dimensional reduction: Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE).

PCA computes a set of linearly uncorrelated variables each of which is an orthogonal linear combination of the gene components. These variables are sorted so as to have decreasing variance. Retaining the top few components yields a reduced dimensional view of the samples that represents most of the variance among the samples. PCA is simple to use and requires no control parameters. TAC 4.0.2 calls the R prcomp function to compute the PCA.

The tSNE algorithm uses a stochastic non-linear dimensional reduction approach that offers more flexibility than PCA at the cost of some additional complexity. The perplexity parameter can roughly be considered to adjust the effective number of nearest neighbors and can range from 1 to nearly the number of samples / 3. Values of perplexity near the ends of the range can generate deceptive results, so values in the central portion of the range generally lead to the best results. The results for several perplexity values should be compared to find the preferred clustering. The distances between samples in the reduced coordinates do not have meaning, so the scatter plot should be viewed more qualitatively. The seed for the random number generator is fixed so as to generate consistent results for a given data set and perplexity value. TAC 4.0.2 calls the Rtsne function in the R package Rtsne.

Starting from the dimensionally reduced coordinates, clustering assigns each sample to a group or cluster. EGA offers three algorithms for clustering: k-means, affinity propagation clustering (APC), and density-based spatial clustering of applications with noise (DBSCAN). The distance matrix, which provides the distance between any two samples, is also returned.

The well-known k-means clustering algorithm requires the input of the number of clusters. K-means randomly assigns each sample to one of the clusters, finds the center for each cluster, and then reassigns samples to the nearest cluster center. New cluster centers are determined, and the process iterates to convergence. Depending on the random assignment, the cluster assignment can converge to different results. The k-means algorithm includes a parameter specifying the number of random starts. The final result of these multiple random starts is the result that has the minimum sum of samples to their respective centers. The random number generator's seed is fixed so as to produce consistent results. TAC 4.0.2 calls the R kmeans function to compute the k-means clustering.

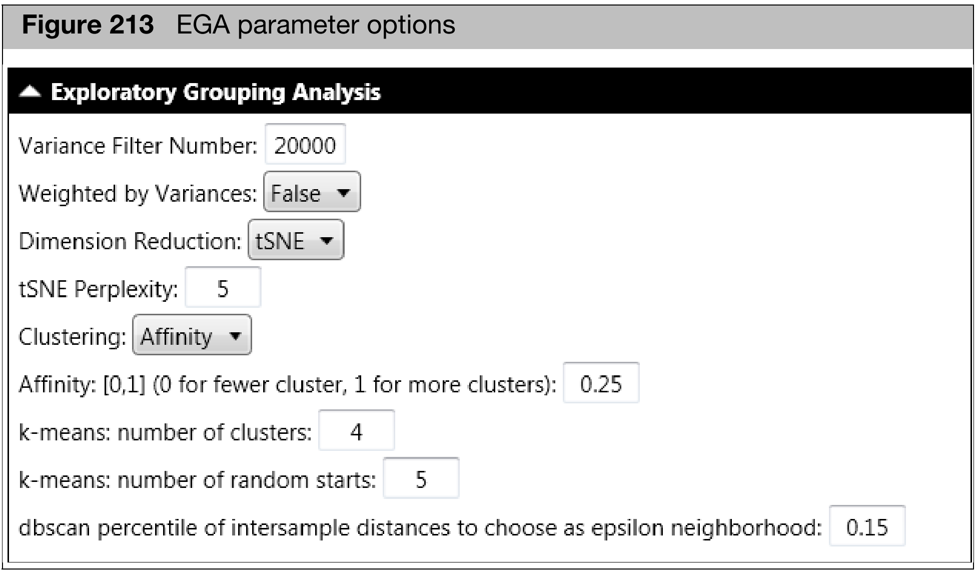
APC derives from the idea of passing messages between the data points. Unlike k-means, APC determines the appropriate number of clusters. The APC message passing

is used to identify exemplars, which are data points that represent the nuclei of clusters, and to associate other data points to each of the clusters. The affinity parameter ranges from 0 to 1 and controls the degree of aggregation. Smaller values yield fewer clusters. TAC 4.0.2 calls apcluster function in the R apcluster package.

DBSCAN is a density-based method of clustering that uses the number of other samples in the neighborhood of each sample to determine the sample's cluster label. The number of clusters is determined by the algorithm, and the clusters can have an arbitrary shape. There is a control parameter that sets the size of the neighborhood based on the percentile of all the distances between the samples. DBSCAN is different from k-means and affinity propagation clustering in that it may not be able to assign a group to all samples in which case some samples are unclustered. TAC 4.0.2 calls the dbscan function in the R dbscan package. The minPts parameter is fixed at 4.

EGA parameters

Figure 213 EGA parameter options



EGA parameter option definitions

Weighted by Variances - This mode enhances the clustering of the most strongly varying probesets by transforming the expression levels as follows: subtract the mean expression level for the probeset and multiply the result by the variance of the probeset.

Dimensional Reduction - A procedure that converts a high dimensional space, in this case the expression of the thousands of probesets in each sample, to a lower dimensional space by creating weighted combinations of the probeset expression levels. TAC4 supports the linear PCA and the nonlinear tSNE algorithms.

tSNE Perplexity - Control parameter for tSNE that roughly represents the effective number of near neighbors.

tSNE Perplexity Mode - When mode is set to auto, a reasonable value for perplexity parameter is generated for the user. In manual mode, the user inputs the perplexity value for greater control.

Similarity - Should not appear in the interface.

Clustering - Identifies the algorithm used to label the samples. The choices include affinity propagation clustering, k-means, and DBSCAN.

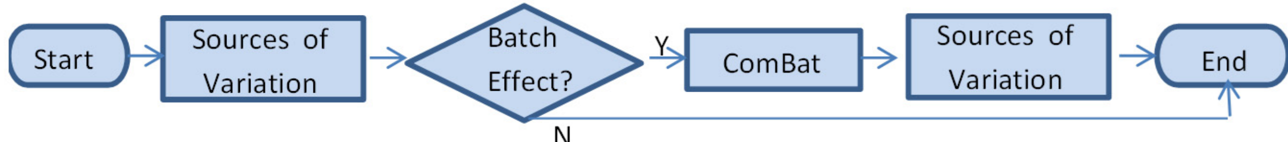
Affinity (Only active when clustering chooses affinity propagation) - A control parameter for the affinity propagation algorithm. The number of clusters is determined by this algorithm, and this parameter controls the relative propensity to generate a larger number of smaller clusters with a range of values [0,1]. Smaller values of affinity have fewer clusters and larger values have more clusters.

k-means number of clusters (only active when clustering chooses k-means) - The number of clusters.

k-means: number of random starts (only active when clustering chooses k-means) - K-means starts with a random assignment of the samples to one of the clusters. It then iteratively reassigns clusters until convergence. Different random staring cluster assignments can yield different final assignments. This option will run k-means the indicated number of times and merge the results.

DBSCAN - Size of epsilon neighborhood - DBSCAN is a density-based clustering algorithm, which does not require specification of the number of clusters and the clusters may have arbitrary shapes. The epsilon neighborhood controls how the samples are added to a cluster. Smaller values lead to larger number of clusters and larger values fewer clusters.

Batch Effects module



The TAC 4.0.2 batch effects module offers two functions: source of variation analysis and batch effect removal. This analysis is computed at the gene-level and the exon-level, so in the following, component refers to gene or PSR/JUC as appropriate. First, the variance is computed for the components, and the 1000 components with the highest variance are retained. Next, for each attribute in the sample attribute table (including columns of type "None"), percentage of variance explained is determined as follows:

$$TSS = total\ sum\ of\ squares = \sum_{all\ samples} (e_i - \langle e_i \rangle)^2$$

$$RSS = residual\ sum\ of\ squares = \sum_{\{all\ attr\ levels\}} \sum_{\{samples\ in\ level\}} (e_i - \langle e_i \rangle)^2$$

$$percentage\ of\ variance\ explained = \langle \frac{TSS - RSS}{TSS} \rangle$$

For each retained component, compute the total sum of squares (TSS) as the sum over samples of squares of the expression level for a sample minus the average expression level for the component (shown using angle brackets). The residual sum of squares (RSS) represents a similar sum of squares but the average expression level subtracted is the average over samples in the same attribute level. Only attribute levels with at least two samples contribute to this sum. Finally, the percentage of variance explained is the average across all the retained components of the ratio of the difference of the TSS and the RSS to the TSS.

This computation looks at each attribute independently, so the sum of the percentages is unconstrained. Generally, one expects the percentage of variance explained to be larger for the experimental attributes under investigation, while other secondary attributes such as date or operator should explain relatively smaller percentages of the variance. Of course, attributes that are unique for each sample will trivially explain 100% of the variance. Secondary attributes that explain relatively high levels of variance are candidate batch effects (assuming that they do not confound the experimental attributes of interest). The percentage of variance explained is computed both before and after the ComBat_Affy* correction so as to provide an indication as to the success of the batch effect removal.

TAC 4.0.2 batch effect removal operates on known batch effects. The batch effect is discovered either through inspection using the EGA or from the PCA plot and sources

of variation analysis generated during standard analysis. The user indicates the existence of a batch effect by marking an attribute column in the sample attribute table as a batch effect. TAC 4.0.2 provides two separate mechanisms for batch effect removal that are used at different stages of the analysis. Combat_Affy batch effect removal generates the transformed expression levels, group averages, group standard deviations, fold changes, and splice indexes. Limma handles the batch effect during the differential expression computation so as to correctly account for the ANOVA degrees of freedom. Limma is discussed in the next section.

Combat_Affy models the batches using their mean values and standard deviation per gene. An empirical Bayesian correction uses pooled information across genes so as to improve the model parameter estimates, which is especially useful for the cases when the number of samples in the comparison groups and batches is small. Finally, the modeled batch effect is removed from the expression signals to yield the corrected result. TAC 4.0.2 calls a C++ accelerated port of the SVA Combat_Affy code. The call uses the parametric empirical prior option.

Warning. Users should use some caution when setting up the batch effect. Batch effect attribute levels should not be correlated with the attribute levels representing the biological information of interest so as to retain the ability to remove the batch effects. During experimental design, one should attempt to randomize the biological sample types across potential batch effect attributes.

***** ComBat_Affy, ComBat_Affy_Gene, and ComBat_Affy_Exon are modified versions of the ComBat function from the Surrogate Variable Analysis package in Bioconductor **[**[11] and [12]**]**.

The modified version includes the following changes:

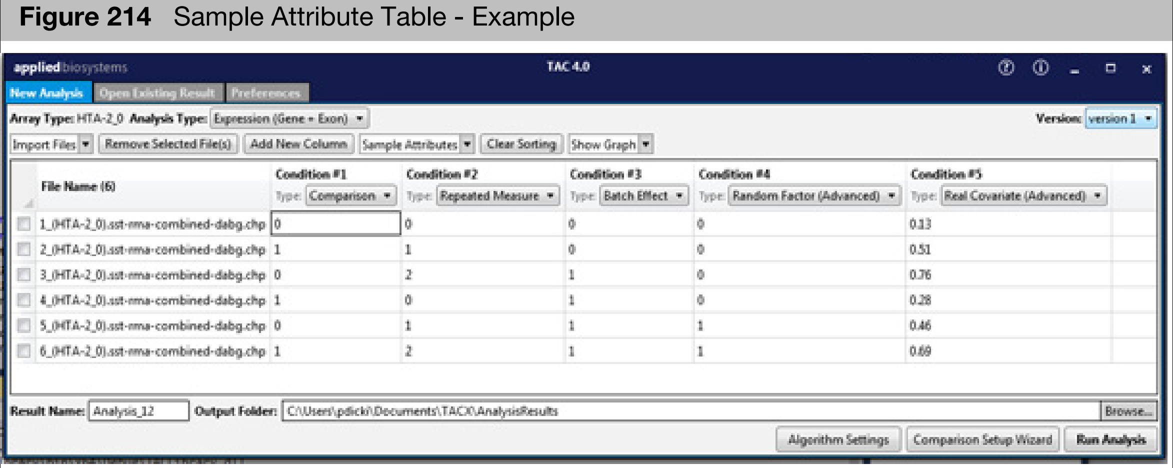
- The original R code was translated to C++.
- Three versions have been created:
  - ComBat_Affy loads all of the microarray data into memory for the processing (like the original code).
  - ComBat_Affy_Gene buffers the microarray data in files for gene-level analysis so that the processing can run in a smaller memory footprint.
  - ComBat_Affy_Exon buffers the microarray data in files for exon-level analysis so that the processing can run in a smaller memory footprint.
- The batch and design indicators are directly loaded through the interface rather than making the batch and design factors and indicators internally.
- If the variance of a gene or exon is <= 1e-8, the standardized mean is returned.
- If the hyperparameters cannot be computed, switch to mean_only mode.

Limma Differential Expression analysis

TAC 4.0.2 calls the limma differential expression portion of the Bioconductor package. Limma provides an extremely flexible framework to describe the experimental structure and desired differential comparisons through the use of the design and contrast matrices. Most TAC users, however, may not have experience with this statistical machinery, so TAC 4.0.2 implements a simpler interface that automatically populates the design and contrast matrices based on the sample attributes table. Moreover, there is a natural language interface called the comparison wizard that enables the user to customize the desired comparisons.

Consider the following sample attribute table (Figure 214) that has been artificially generated to show its attribute column types.

Figure 214 Sample Attribute Table - Example



Each attribute column in the sample attribute table can be set to one of six types: comparison, repeated measure, batch effect, random factor, real covariate, and none. The normalization group attribute type is separate and does not directly affect differential expression analysis. It provides a mechanism to preprocess the CEL files in subgroups rather than as a single group; the subgrouping allows for separate normalization when the samples differ so much that normalizing to a common distribution is not appropriate.

The comparison type specifies fixed attribute levels among which the differential expression can be computed. Moreover, interactions between attributes can be computed. At least one attribute column must be designated as a comparison type. The comparison attributes levels represent the biological factors of interest that the experiment was designed to study. All other attribute types represent various nuisance effects that the analysis will remove so as to enhance the effect of the comparison type attributes.

The repeated measure type indicates that the attribute marks experimental units (e.g. patients) that have multiple conditions associated with them. Differential expression between the conditions on the experimental units is the point of interest; the repeated measure type indicates that the specific effects of the experimental units should be removed.

The batch effect type indicates that the attribute levels represent fixed substrata in the data that represent nuisance factors in the differential expression analysis. Examples of

batch effects include assay date, scanner, operator, and so forth. Inclusion of an attribute as a batch effect has the effect of removing it. As in the case of Combat_Affy, batch effect attribute levels should not be well-correlated with the comparison attribute levels.

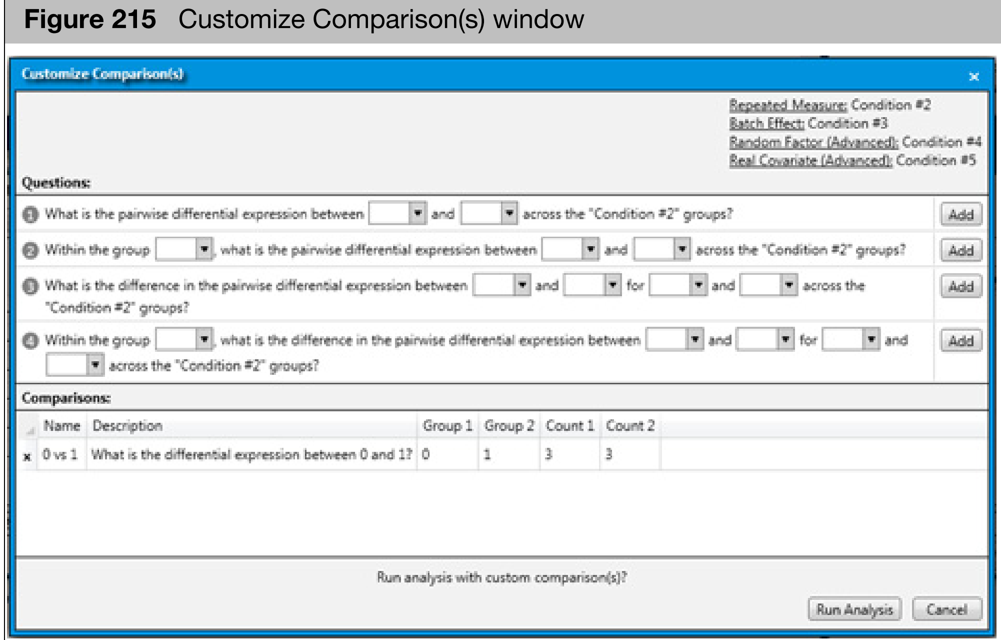
The random factor type represents random factors as supported by limma; that is, the attribute is a special kind of batch effect and the variance components of the random factor are not available. The application of random effects is somewhat subtle and generally a tool for more advanced statistical analysis. A common usage in limma involves comparisons both within and between experimental subjects in which the subject becomes the random factor. Furthermore, the correlation computation for a random factor is expensive, so the use of random factors can significantly extend analysis time.

The real covariate represents a mechanism to incorporate a continuously varying attribute into the analysis. Examples include age, tumor size, degree of contamination, and so forth. Limma will remove the effect of this real covariate on the differential expression results.

The fixed attribute types (comparison, repeated measure, and batch effects) are merged into a compound type. If each of the levels of the resulting type have a single sample, the variance cannot be computed, so all of the resulting comparison p-values for the contrasts will be 1.

The contrast matrix is automatically populated all reasonable comparisons and interactions based on the comparison type attributes. For simpler experiments, this default contrast set should be adequate. For more complex experimental designs, however, it can be helpful to use the comparison wizard to customize the analysis. Generally, this customization can reduce the number of contrasts to simplify the results, as well as reduce the computation time.

Figure 215 Customize Comparison(s) window



In the Custom Comparison(s) window (Figure 215), the default contrasts initially appear in the lower half. The upper half enables the creation of contrasts. The first two questions create contrasts that query differential expression between two groups; the second question specializes to the question of the differential expression between two

groups, both of which reside in a subgroup. The third and fourth questions query interactions, where the fourth question specializes to a subgroup analogously to question 2.

TAC 4.0.2 calls the lmFit, contrasts.fit, duplicateCorrelation, ebayes, and topTable funtions in the limma Bioconductor package. The user can selct whether to do a simple ANOVA analysis or to add the eBayes empirical parameter improvement(default).

Anova

Analysis of variance or ANOVA analysis fits a linear model to each probeset independently of the others.

eBayes

The eBayes analysis corrects the variance of the ANOVA analysis with an empirical Bayes approach that uses the information from all the probesets to yield an improved estimate for the variance. The eBayes correction is especially important when the number of samples being analyzed is small.

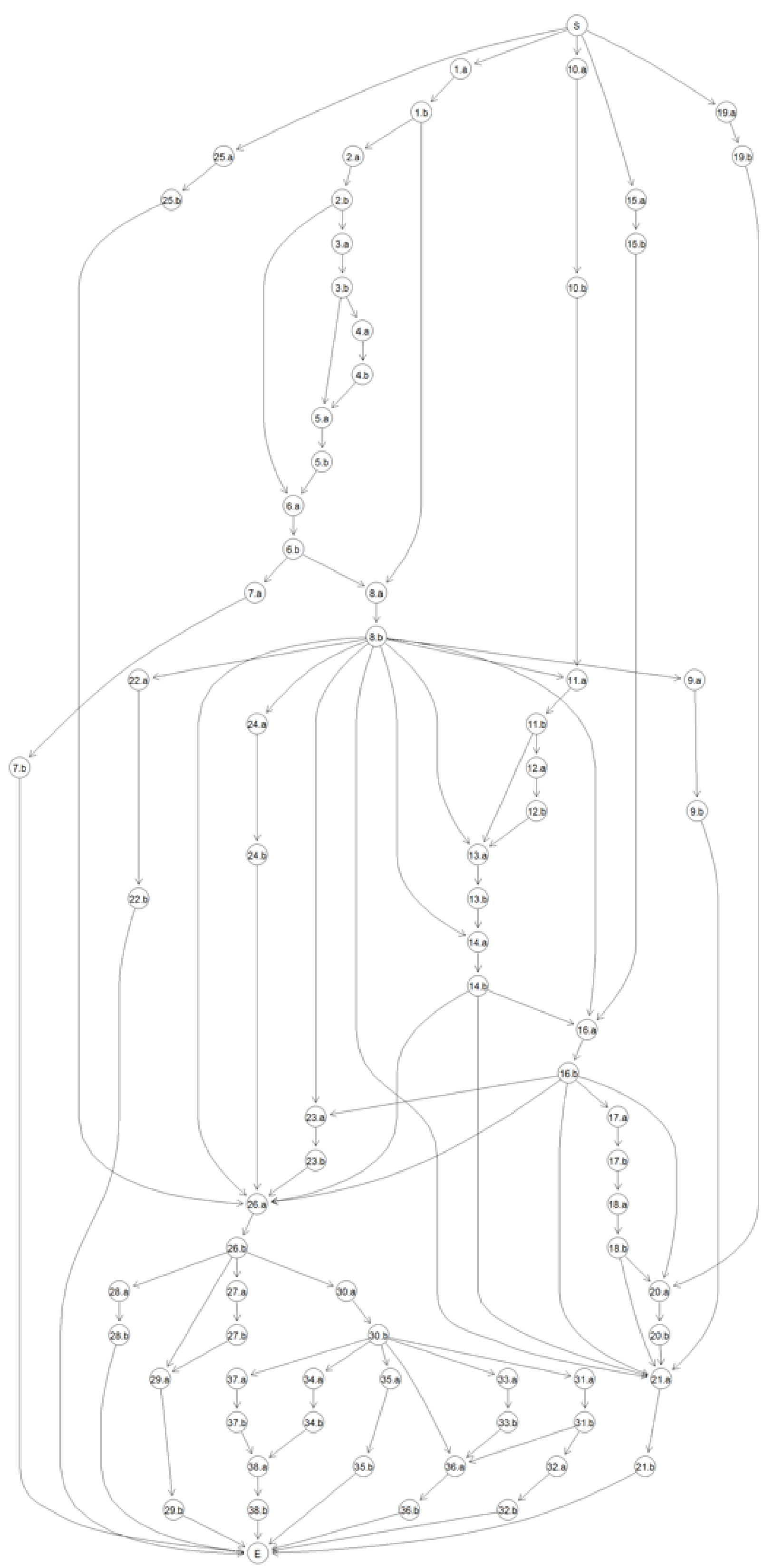
Alternative Splicing Analysis

Event Pointer Alternative Splicing Analysis

In Affymetrix microarray designs that enable the inspection of genes for potential alternative splicing events, the genes are composed of probe selection regions (PSRs) that represent exons and junctions (JUCs) that straddle the inter-exon connections. Event Pointer is a statistical algorithm for computing the probability of differential alternative splicing events in genes based on the summary expression signals of the PSRs and JUCs. Event Pointer starts with the set of all known isoforms for a gene from which it creates the splicing graph. Alternative splicing events are extracted from the splicing graph in the form of PSR/JUC signal sets for a reference and a set of associated paths. An extended limma analysis of these signals yields a probability that the differential alternative splicing event occurs between two conditions. The differential alternative splicing events are examined for the same contrasts that are examined for the ordinary limma analysis.

Splicing graph for the NCALD gene appears on the left. At the top, the start node is marked "S". At the bottom, the end node is marked "E". Paths from the start to the end represent isoforms. Nodes or combination of nodes represent different alternative splicing events. A flow through the splicing graph starting at "S" and ending at "E", and at each intermediate node, the flow randomly splits. An event is defined such that the sum of the flows in a set of paths is equal to the flow in some reference. For example, the flows through (3b-5a) + (3b-4a-4b-5a) = (2b-3a-3b, 5a-5b-6a). The event defined here is a cassette event. This type of analysis is done for every gene generating a list of potential events. The extended limma analysis computes the probability of these events.

TAC 4.0.2 calls a customized version of Event Pointer with a few changes. First, the number of paths can have an arbitrary number, and not just two. The extended limma analysis is generalized to allow for the arbitrary number of paths. Second, the signal quality of JUC probes is somewhat lower than PSR probes due to the constrained positioning of a junction probe, so JUC probes are down weighted in the analysis. Third, the p-value returned is the average of the one associated with the isoform fold change and the relative concentration change, relative to what is described in the paper [9] on page 218.


S
1.a
10.a
19.a
1.b
25.a
2.a
19.b
25.b
2.b
15.a
3.a
15.b
3.b
10.b
4.a
4.b
5.a
5.b
6.a
6.b
7.a
8.a
8.b
22.a
11.a
9.a
24.a
11.b
7.b
12.a
12.b
9.b
24.b
13.a
22.b
13.b
14.a
14.b
16.a
16.b
23.a
17.a
23.b
17.b
26.a
18.a
26.b
18.b
28.a
27.a
30.a
20.a
28.b
27.b
30.b
20.b
29.a
37.a
34.a
35.a
33.a
31.a
21.a
37.b
34.b
33.b
31.b
38.a
35.b
36.a
32.a
21.b
29.b
38.b
36.b
32.b
E

The advantages of Event Pointer are that it is a fully statistical analysis that uses information from both the PSRs and he JUCs. It has the full splicing graph information, so it recognizes the standard alternative splicing events (i.e. cassette exon, mutually exclusive exons, alternate 5' donor site, alternate 3' acceptor site, intron retention, alternative first exon, and alternative last exon) as well as more complex events that do not match these categories. The disadvantages of Event Pointer include that it can miss certain events due to the complexity of the splicing graphs in some genes.

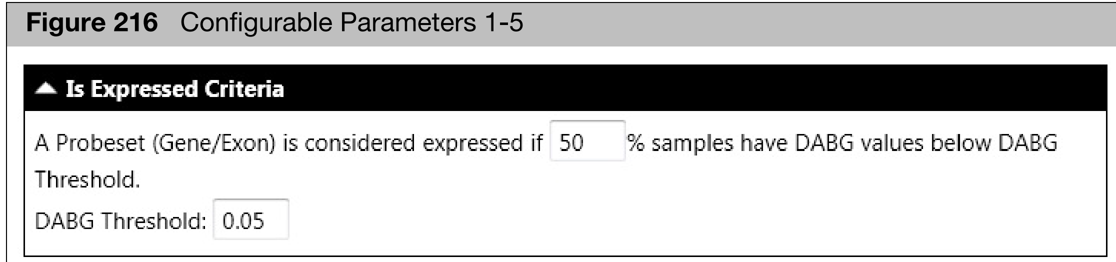
Splicing Index (SI) algorithm

Splicing Index algorithm is a way to measure of how much exon specific expression differs between two conditions after excluding gene level influences. The algorithm first normalizes the exon and junction expression values by the level of gene expression and creates a ratio of normalized signal estimates from one condition relative to another. **Note:** The Splicing Index algorithm in TAC 4.0.2 is different than in previous versions of TAC. TAC 4.0.2 (by default) runs the Event Pointer and Event Estimation simultaneously. For details, see "Event Pointer Algorithm" on page 210.

Performing an Alternative Splicing Analysis

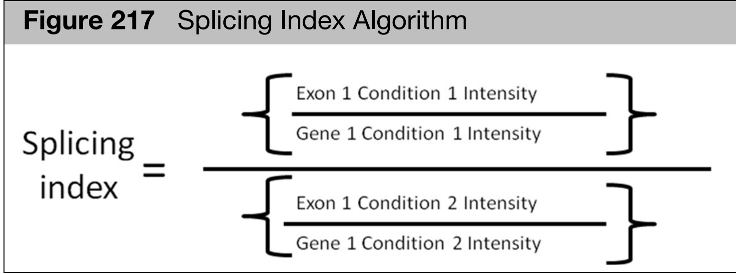
In the **Preference** window you can customize algorithm parameters by typing values in the text boxes. (Figure 216) **Note:** Alternative Splicing Analysis is only available for certain arrays.

Figure 216 Configurable Parameters 1-5



The Splicing Index algorithm compares normalized signal estimates from one condition to another. See the equation below: (Figure 217)

Figure 217 Splicing Index Algorithm



In order for Transcriptome Analysis Console to perform Splicing Index, two key criteria must be met. They are as follows:

1. **Criteria 1:** A Transcript Cluster gene must be expressed in both conditions, therefore for each condition, you must determine whether a gene is expressed or not.

   i. A gene is expressed in a sample if its gene level DABG <0.05 (configurable parameter)

   ii. A condition is expressed if >=50% (configurable parameter) samples are expressed

2. **Criteria 2:** A PSR or Junction can only be analyzed by Splicing Index if it expresses in at least one condition.

- To decide whether a PSR or junction is expressed in a condition, you need (as an example) to check the DABG p-values (from all samples in that condition) to see whether>=50% of samples have DABG p < 0.05. For more information, go to: thermofisher.com

Note: PSR/JUC FDR p-value calculation is only based on PSR/JUC that passed the two criteria above.

Figure 218 Criteria passing examples

	cond1 sample1	cond1 sample2	cond1 sample3	cond1 sample4	cond1 summary - whether >=50% samples have that PSR/JUC as expressed	cond2 sample1	cond2 sample2	cond2 sample3	cond2 sample4	cond2 summary - whether >=50% samples have that PSR/JUC as expressed	A PSR or Junction can only be analyzed by Splicing Index if it expresses in at least one condition
PSR/JUC DABG p-value	0.2871392	0.378386	0.272059	0.393614	NO	0.0206628	0.020067	0.021189	0.021041	YES	**YES**
PSR/JUC DABG p-value	0.04889926	0.031045	0.048693	0.055893	YES	0.066104	0.078026	0.08502	0.081922	no	**YES**
PSR/JUC DABG p-value	0.001298487	0.001298	0.001771	0.001653	YES	0.000177	0.000207	0.000148	0.000177	YES	**YES**
PSR/JUC DABG p-value	0.1558461	0.156082	0.157794	0.134274	no	0.015552	0.017323	0.021277	0.016437	YES	**YES**
PSR/JUC DABG p-value	0.1366051	0.092723	0.099304	0.168447	no	0.021159	0.021661	0.01992	0.020894	YES	**YES**
PSR/JUC DABG p-value	0.1534852	0.126306	0.120964	0.176976	no	0.017529	0.018887	0.020776	0.023225	YES	**YES**
PSR/JUC DABG	0.06194299	0.096116	0.065838	0.08856	no	0.03458	0.040754	0.034882	0.037154	YES	**YES**

Note: For each PSR/JUC, the normalized signal is calculated by "PSR/JUC signal"/ "gene signal". For example, if the PSR signal is 5.46, gene signal is 9.09, the normalized signal will be -3.63 (this is log scale).

Figure 219 Normalized Signal calculation example

normalized signals	cond1 sample1	cond1 sample2	cond1 sample3	cond1 sample4	cond2 sample1	cond2 sample2	cond2 sample3	cond2 sample4
PSR1	-3.63	-4.1	-2.83	-3.3	-0.97	-0.94	-1.25	-0.89
biweight avg of cond1 for PSR1	-3.47							
biweight avg of cond2 for PSR1	-0.94							
Splicing index of PSR1	**-2.53**							

Splicing Event estimation score

IMPORTANT! The Splicing Event Estimation method is used as a classification tool and not intended for a de novo event discovery.

Splicing Event Estimation

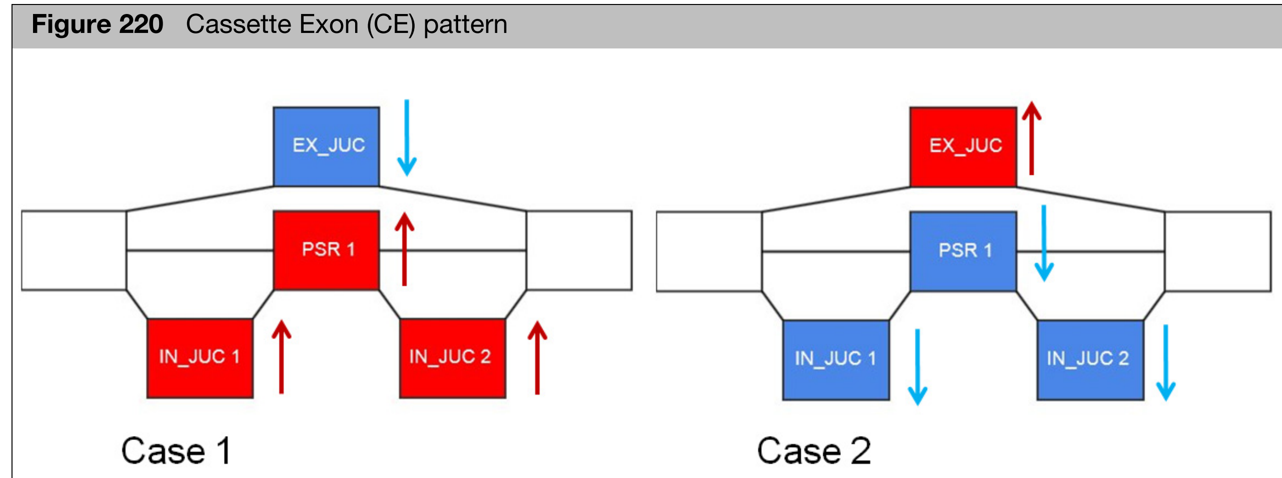
The Splicing Event Estimation is generated by detecting changes between two conditions using SI and p-values. It investigates each mode and finds specific patterns that meet each mode's definition.

The Splicing Event Estimations are based on five following classic modes of alternative splicing:

- **Cassette Exon (CE)** - From a Primary Transcript, one exon is spliced out along with its partnering introns.

  For CE, the Splicing Event Estimation is based on PSR and its two inclusion junction patterns moving in the same direction, while its exclusion junction moves in the opposite direction, as shown in Figure 220. The PSR direction influences the direction of the junction.

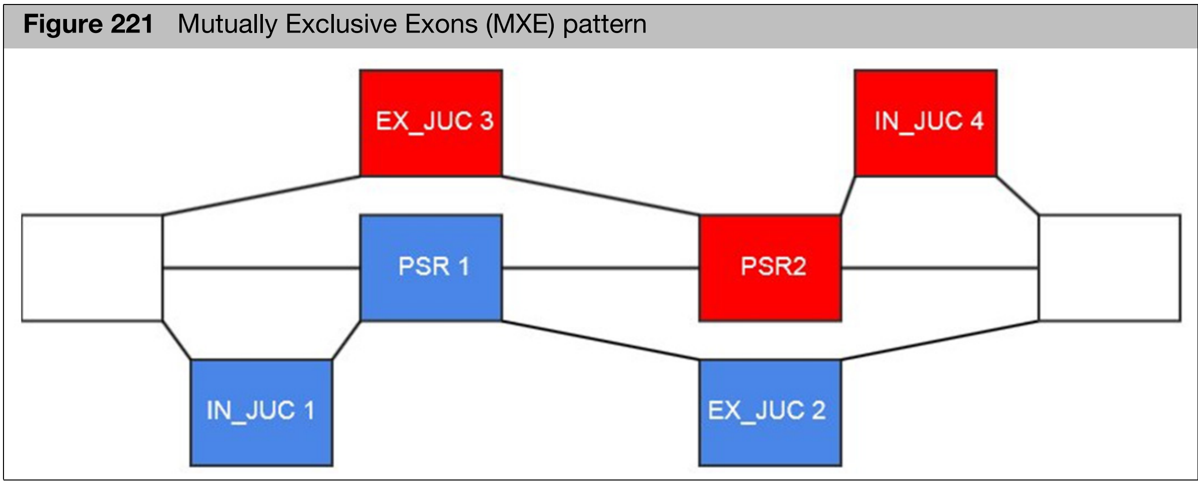
Figure 220 Cassette Exon (CE) pattern



- **Mutually Exclusive Exons (MXE)** - 2 exons (back to back) that are excluded from the mature mRNA transcript.

  For MXE, the Splicing Event Estimate is based on PSR and its inclusion and exclusion junctions moving in the opposite direction of another set, as shown in Figure 221. The PSR direction influences the direction of the junction.

Figure 221 Mutually Exclusive Exons (MXE) pattern

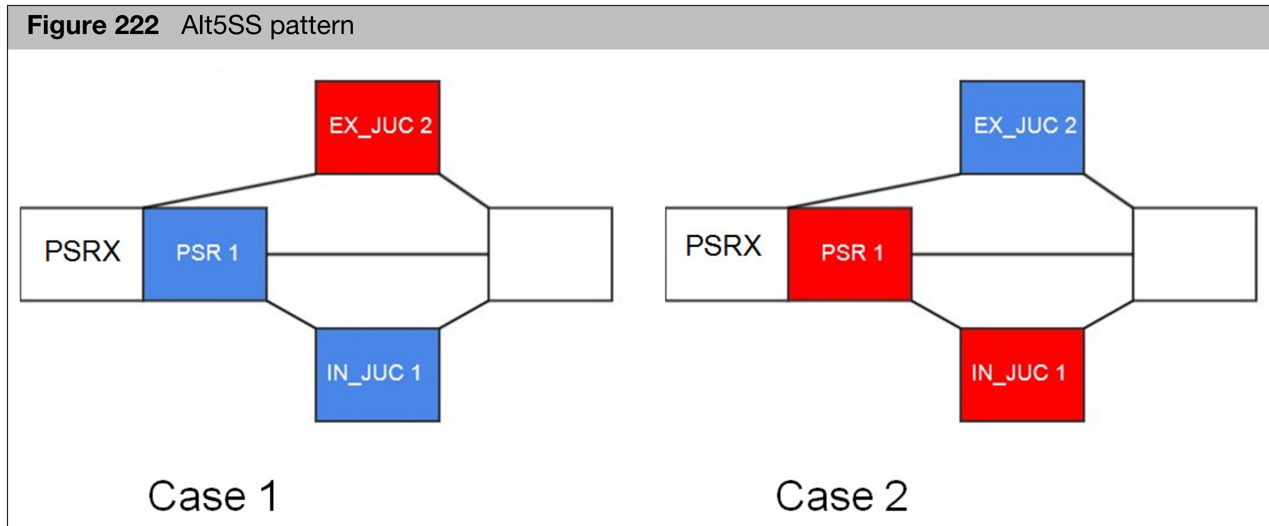


- **Alternative 5' Sites (Alt 5SS)** - 2 or more splices are recognized at the 3' end of the exon. At this point, an alternative 5' splice junction (donor site) is used. This changes the 3' boundary of the upstream exon.

  For Alt 5SS, the Splicing Event Estimate is based on PSRs and their inclusion junctions changing in the same direction, while their exclusion junctions move in

the opposite direction, as shown in Figure 222. PSRX and PSR1 are different, but are genomic neighbors. The PSR direction influences the direction of the junction; PSR1's direction decides PSRX's direction.

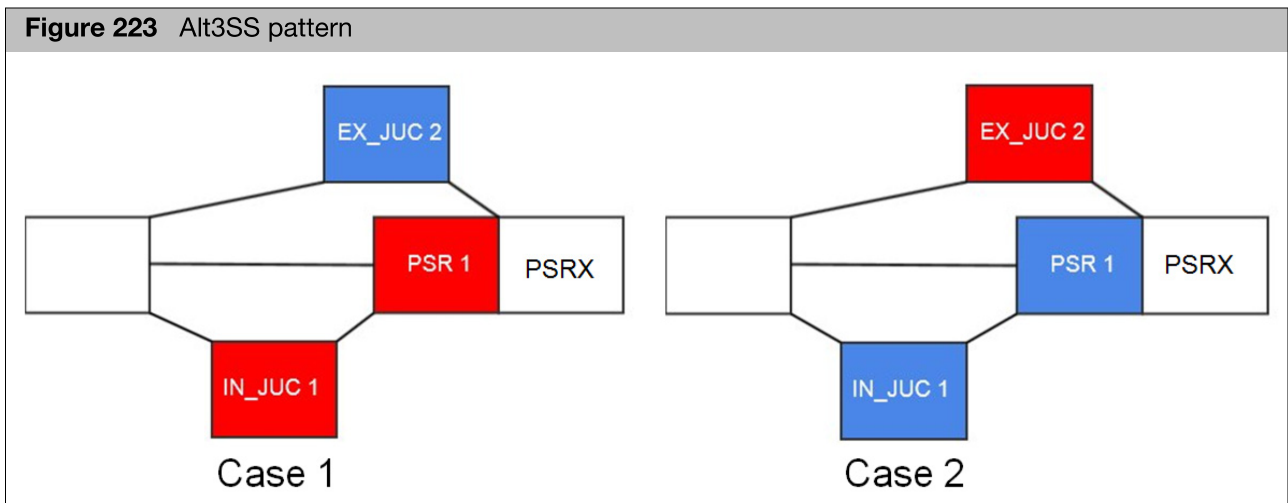
Figure 222 Alt5SS pattern



- **Alternative 3' Sites (Alt 3SS)** - 2 or more splices are recognized at the 5' end of the exon. At this point, an alternative 3' splice junction (acceptor site) is used. This changes the 5' boundary of the downstream exon.

  For Alt 3SS, the Splicing Event Estimate is based on PSRs and their inclusion junctions changing in the same direction, while their exclusion junctions move in the opposite direction, as shown in Figure 223. PSRX and PSR1 are different, but are genomic neighbors. The PSR direction influences the direction of the junction; PSR1's direction decides PSRX's direction.

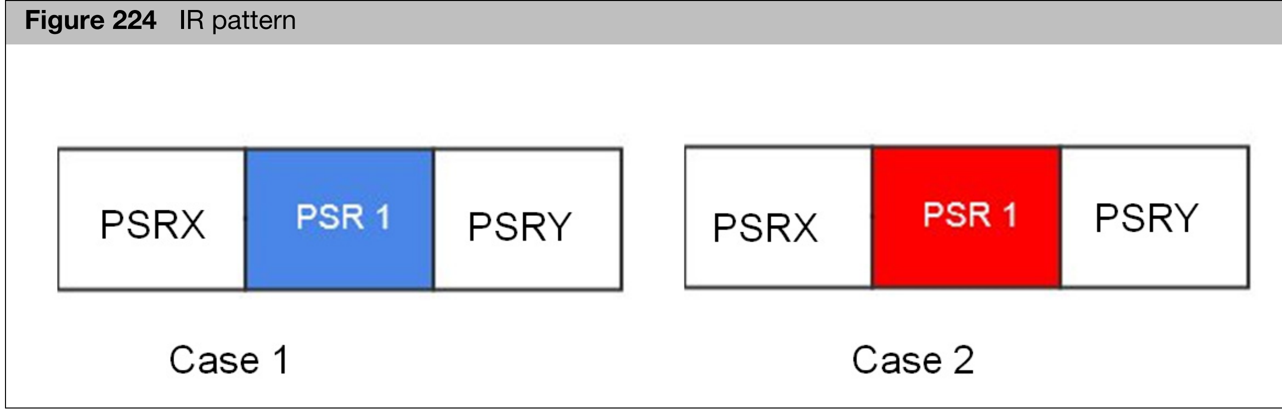
Figure 223 Alt3SS pattern



- **Intron Retention (IR)** - A sequence is spliced out as an intron or is retained within the mature mRNA transcript.

  For IR, the Splicing Event Estimate is based on the pattern of the PSR versus its genomic-related PSRX and PSRY neighbors, as shown in Figure 224.

Figure 224 IR pattern



Event Estimation Score

- The Event Estimation Score is between 0 and 1 (1 is the highest possible score).
- The Event Estimation Score is based on how well your data fits into pre-defined splicing patterns. PSRs and their related junctions all contribute to an event score.
- PSRs and Junctions can contribute positively or negatively to the event score depending on their specific pattern directions, as described earlier.
- Only PSRs are assigned Event Estimation Score
- PSRs with more than 1 pattern, receive an event estimation score for each of the possible modes of alternative splicing, but the final estimate of the PSR is based on the highest event score.

IMPORTANT! Real-time PCR is routinely used to verify microarray results due to its high sensitivity and wide dynamic range.

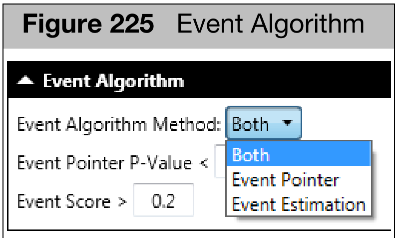
It is recommended to verify the splicing events by using the USB First-Strand cDNA Synthesis Kit [PN 75780] followed by real-time PCR with USB VeriQuest Probe or SYBR® Green qPCR master mixes [PN#75650/75600].

Detailed information can be found in the Data Sheet. Concordance of Affymetrix GeneChip® Human Transcriptome Array 3.0 and real-time PCR results using USB® VeriQuest® qPCR master mixes.

Event Pointer Algorithm

The Event Pointer is a new option for the Splicing Index (SI) algorithm and is run when Event Algorithm Method is set to **Both** or **Event Pointer**, as shown in Figure 225. For more information on customizing algorithms, see "Customizing an Installed Array's Algorithm Settings" on page 18.

Figure 225 Event Algorithm



Event score

TAC 4.0.2 includes two algorithms for identifying alternative splicing events: the TAC 2.0 algorithm and the new Event Pointer in TAC 4.0.2. The new Event Score leverages both previous TAC 2.0 event estimation score and Event Pointer p-value and sorts the most likely alternative splicing events to the top. The Event Score is computed as follows:

EventScore = (MIN(1, Log10(Event Pointer_P)/Log10(1E-20)) + Event Estimation Score) * 0.5

Then adjust the event score slightly by SI.

EventScore = EventScore - 0.0001 + 0.000005 * Log2(|SplicingIndex|)

If EventScore is greater than 1, then it will be capped at 1.

Events given high scores by both algorithms tend to be the best events scored by human inspection. The results from each constituent algorithm are available through examination of the hidden columns.

MAS 5.0, RMA and PLIER

The following are brief descriptions of the MAS 5.0, RMA, and PLIER algorithms. References and links to various publications, describe the MAS 5.0, RMA, and PLIER algorithms in detail. See "Comparison of algorithms" on page 214 for a side-by-side comparison of the assumptions, advantages and disadvantages of using each individual algorithm.

MAS 5.0

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.

Tukey's Bi-weight average is a method to determine a robust average unaffected by outliers.

- **Step 1:** The median is determined to define the center of the data.
- **Step 2:** The distance of each data point from the median is determined. This distance is then used to determine how much each value should contribute to the average. For example, outliers that are far away from the median should contribute less to the average.
- All end result values represented as a Bi-weight average, are shown in a log2 scale (if the CHP files are summarized using RMA)
- For more information, go to: thermofisher.com

The MAS 5.0 algorithm (also known as the Statistical Algorithm) analyzes each array independently. As a result, individual probe-specific affinities cannot 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 multi-chip analysis methods (RMA, PLIER).

For a more detailed description of the MAS 5.0 algorithm, see the Statistical Algorithms Reference Guide at thermofisher.com.

The TAC software displays column headings for individual parameters and outputs in the Sample QC View for each of the 3' Expression algorithms. See the following table for column header definitions.

MAS 5.0 Column Headers and Definitions

Column Header	Definition
#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.
#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.

RMA

The Robust Multi-chip 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:

1. Background adjustment
2. Quantile normalization
3. 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*.

The TAC software displays column headings for individual parameters and outputs in the Sample QC View. See the following table for column header definitions.

RMA and PLIER Column Headers and Definitions

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 Header	Definition
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.
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:** The AUC of the ROC curve is equivalent to a rank sum statistic used to test for differences in the center of two distributions.

SST-RMA

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, please refer to white paper available at thermofisher.com.

For column header definitions, refer to the table above ("RMA and PLIER Column Headers and Definitions" on page 213).

PLIER

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 at thermofisher.com.

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 at thermofisher.com.

For column header definitions, refer to the table above.

Comparison of algorithms

The table below is a comparison of the MAS 5.0, RMA, and PLIER algorithms.

Algorithm	Advantages	Disadvantages	Assumptions
MAS 5.0	Single-array algorithm is independent of other data in the data set: ▪ Less computationally intensive than RMA or PLIER ▪ Conservative ▪ Smooth down-weighting of outliers ▪ Positive output values ▪ Minimal bias	Not as sensitive as either RMA or PLIER to small changes in target abundance: ▪ 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 ▪ Higher reproducibility of signal over single array analyses ▪ Good differential change detection ▪ Variance stable on log scale	Not as sensitive as PLIER in the ability to detect small fold changes ▪ 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: ▪ 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	More variance in individual signals than seen with RMA ▪ 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

Benjamini-Hochberg Step-Up FDR-controlling Procedure

FDR control is a statistical method used in multiple hypothesis testing to correct for multiple comparisons.

- FDR is controlled at certain alpha level (default 0.05 in all result tables). This means that the expected proportion of rejections that are in error is less than alpha. Alpha level can be customized in the result tables by changing the default filtering criteria in the FDR p-value column (0.05) to something else.
- Benjamini, Yoav; Hochberg, Yosef (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". Journal of the Royal Statistical Society, Series B (Methodological) 57 (1): 289-300. MR 1325392.

Fold change

- Fold change is a number describing how much the signal changes from an initial condition group to a final condition group.
- Fold changes are represented in linear space.

Hierarchical clustering

- Clustering is performed on both probe sets and CHP signals.
- Distance metric used between objects is the Euclidean distance.
- Distances between clusters of objects are computed using the complete linkage method (maximum distance between a pair of objects in the two clusters).
- Results are displayed in a heat map and dendrogram.

Probe set suffixes

A number of standard metrics are created for particular groups of probe sets on the arrays. The table 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)

Suffix for Output	Definition
_probesets	From the QC Report, the number of probe sets 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).

Chromosome naming scheme

- In addition to the “regular” chromosomes, the hg19 browser contains nine haplotype chromosomes and 59 unplaced contigs. If an unplaced contig is localized to a chromosome, the contig name is appended to the regular chromosome name, as in chr1_gl000191_random. If the chromosome is unknown, the contig is represented with the name “chrUn” followed by the contig identifier, as in chrUn_gl000211. Note that the chrUn contigs are no longer placed in a single, artificial chromosome as they have been in previous UCSC assemblies. See the sequences page for a complete list of hg19 chromosome names.
- The 9 haplotype chromosomes are as follows:

Name	Accession	UCSC chr Name
HSCHR6_MHC_APD_CTG1	GL000250.1	chr6_apd_hap1
HSCHR6_MHC_COX_CTG1	GL000251.1	chr6_cox_hap2
HSCHR6_MHC_DBB_CTG1	GL000252.1	chr6_dbb_hap3
HSCHR6_MHC_MANN_CTG1	GL000253.1	chr6_mann_hap4
HSCHR6_MHC_MCF_CTG1	GL000254.1	chr6_mcf_hap5
HSCHR6_MHC_QBL_CTG1	GL000255.1	chr6_qbl_hap6
HSCHR6_MHC_SSTO_CTG1	GL000256.1	chr6_ssto_hap7
HSCHR4_1_CTG9	GL000257.1	chr4_ctg9_hap1
HSCHR17_1_CTG5	GL000258.1	chr17_ctg5_hap1

References

- **[1]** Jesse H. Krijthe (2015). Rtsne: T-distributed stochastic neighbor embedding using a Barnes-Hut implementation, URL: https://github.com/jkrijthe/Rtsne.
- **[2]** Michael Hahsler (2016). dbscan: Density based clustering of applications with noise (DBSCAN) and Related Algorithms. R package version 0.9-8. https://CRAN.R-project.org/package=dbscan
- **[3]** Ulrich Bodenhofer, Andreas Kothmeier, and Sepp Hochreiter (2011) APCluster: an R package for affinity propagation clustering Bioinformatics 27:2463-2464.
- **[4]** Brendan J. Frey and Delbert Dueck (2007). Clustering by passing messages between data points. Science 315:972-977.
- **[5]** Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.
- **[6]** Phipson, B, Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Annals of Applied Statistics 10(2), 946-963.
- **[7]** Smyth, G. K., Michaud, J., and Scott, H. (2005). The use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067- 2075.
- **[8]** Limma User's Guide, https://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf.
- **[9]** Juan P. Romero et al, "Event Pointer: an effective identification of alternative splicing events using junction arrays", BMC Genomics 2016 17:467.
- **[10]** Juan Pablo Romero and Angel Rubio (2015). Event Pointer: analysis of alternative splicing using microarray data. R package version 1.0.
- **[11]** W.E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray data using empirical bayes methods. Biostatistics, 8(1):118-127, 2007.
- **[12]** Jeffrey T. Leek, W. Evan Johnson, Hilary S. Parker, Elana J. Fertig, Andrew E. Jaffe and John D. Storey (2016). sva: Surrogate Variable Analysis. R package version 3.20.0.).