REMBRANDT
USER'S GUIDE

Version 1.5.1

NATIONAL® CANCER INSTITUTE

Center for Bioinformatics

December 9, 2007

Credits and Resources

REMBRANDT Development and Management Teams		
Development	***Documentation***	***Product and Program Management***
Subha Madhavan	Laura Jackel	Subha Madhavan
Alex Jiang	Huaitian Liu	
Kevin Rosso	Jill Hadfield	
Ryan Landy		
Himanso Sahni		
David Bauer		
Huaitian Liu		
Michael Harris		
Ram Bhattaru		
David Hall		
Ye Wu		
Ying Long		
Don Swan		
Dana Zhang		
Vesselina Bakalov		
Nick Xiao		
Jean-Claude Zenklusen		
Yuri Kotliarov		
Gregg Silk		

Contacts and Support	
Application Support	http://ncicb.nci.nih.gov/NCICB/support Telephone: 301-451-4384 Toll free: 888-478-4423

TABLE OF CONTENTS

About This Guide

This section introduces you to the *REMBRANDT 1.5.1 User's Guide*. It includes the following topics:

Purpose

This guide provides an overview of REMBRANDT. This book is organized into chapters that parallel REMBRANDT's workflow.

Audience

This guide is designed for researchers who want to perform ad hoc querying and reporting across multiple domains, such as gene expression, chromosomal aberrations, and clinical data.

Topics Covered

If you are new to REMBRANDT, read this brief overview, which explains what you will find in each chapter.

- *Chapter 1* provides instructions to start using REMBRANDT.
- *Chapter 2* describes how to search by gene keyword and reporter identifier and to create gene expression plots, Kaplan-Meier surival plots, and copy number-based graphs from those search results.
- *Chapter 3* describes how to add gene expression, copy number, and clinical queries to REMBRANDT, and group them to create and generate results for compound queries.
- *Chapter 4* extends the basic knowledge of the previous chapters and shows you how to work with class comparisons, hierarchical clustering, and principal component analysis.
- *Chapter 5* describes how to view all the results generated from advanced searches and high order analyses.

- *Chapter 6* describes how to manage user-defined or study-defined patient ID, gene, and reporters lists.

Text Conventions Used

This section explains conventions used in this guide. The various typefaces represent interface components, keyboard shortcuts, toolbar buttons, dialog box options, and text that you type.

Convention	***Description***	***Example***
Bold	Highlights names of option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons.	Click **Search**.
URL	Indicates a Web address.	http://domain.com
text in SMALL CAPS	Indicates a keyboard shortcut.	Press ENTER.
text in SMALL CAPS + text in SMALL CAPS	Indicates keys that are pressed simultaneously.	Press SHIFT + CTRL.
Italics	Highlights references to other documents, sections, figures, and tables.	See *Figure 4.5*.
Italic boldface monospace type	Represents text that you type.	In the **New Subset** text box, enter ***Proprietary Proteins.***
Note:	Highlights information of particular importance	**Note:** This concept is used throughout the document.
{ }	Surrounds replaceable items.	Replace {last name, first name} with the Principal Investigator's name.

CHAPTER

1

GETTING STARTED WITH REMBRANDT

This chapter introduces you to REMBRANDT, REMBRANDT interfaces, and navigation, and common features used on REMBRANDT pages.

Topics in this chapter include:

About REMBRANDT

REMBRANDT (REpository for Molecular BRAin Neoplasia DaTa) is a joint initiative of NIH's National Cancer Institute (NCI) and the National Institute of Neurological Disorder

s and Stroke (NINDS). REMBRANDT provides a bioinformatics knowledge base framework that leverages data warehousing technology to host and integrate clinical and functional genomics data from clinical trials involving patients suffering from gliomas (tumors).

Researchers can use REMBRANDT to answer questions related to a patient or patient population and view integrated datasets in a variety of contexts. REMBRANDT also includes tools that link data to other annotations, such as cellular pathways, gene ontology terms, and genomic information. Researches can also perform various higher-order analyses on clinical and genomic datasets.

How to Cite REMBRANDT Data

When referencing the REMBRANDT data set, please cite National Cancer Institute as the source, including year of first production release (2005), the REMBRANDT website (http://rembrandt.nci.nih.gov) and the accessed date.

For Example:

National Cancer Institute. 2005. REMBRANDT home page. <http://rembrandt.nci.nih.gov>. Accessed 2007 September 24

Launching REMBRANDT

To launch REMBRANDT, follow these steps:

1. Go to the REMBRANDT portal on the NCICB website:

   http://rembrandt.nci.nih.gov/.

2. Click the **REMBRANDT Application** button located in the lower left-hand blue column (*Figure 1.1*).

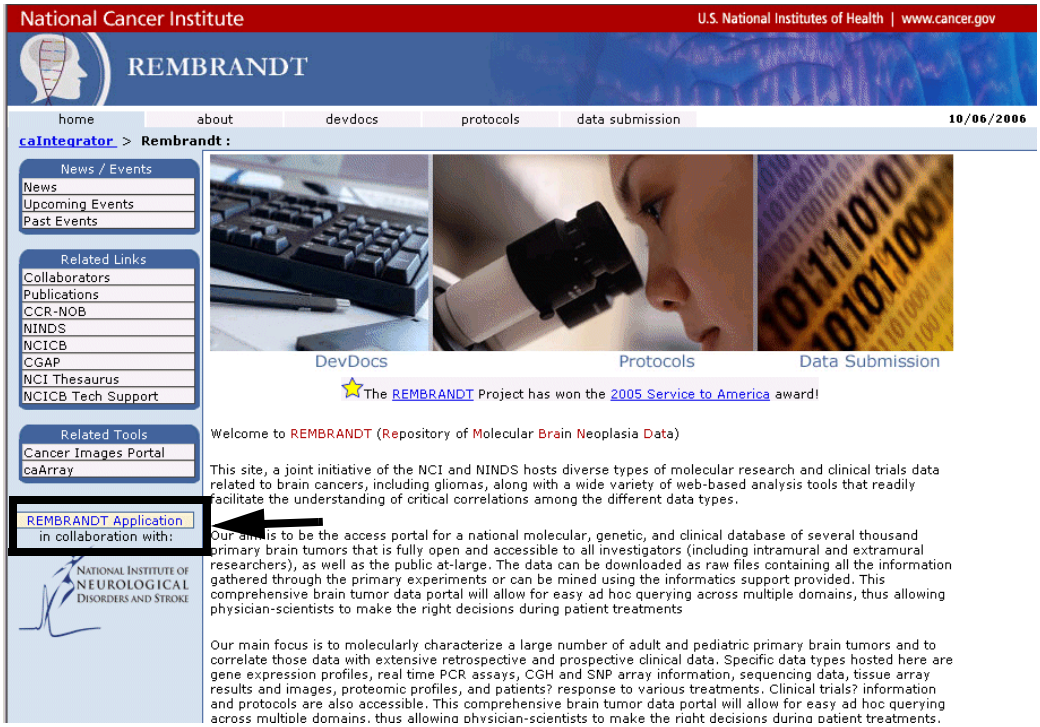


Figure 1.1 The REMBRANDT portal on the NCICB website

The REMBRANDT login page appears *(Figure 1.2)*.

Figure 1.2 REMBRANDT opening page

REMBRANDT's Opening Page

REMBRANDT's opening page enables you to perform the following tasks:

- Register new users and log in current users with the **Browse REMBRANDT Data** button.
- Download an online version of the *REMBRANDT 1.5.1 User's Guide* with the **Download User Guide (PDF)** link.
- Run REMBRANDT tutorials that will assist you in using new REMBRANDT options with the **View Tutorials** link.
- Provide the REMBRANDT team with any feedback about the product with the **Provide us your feedback** link.

New User Registration

First-time REMBRANDT users need to register first to obtain a username and password. To register in REMBRANDT, follow these steps:

1. Fill in the the **Name** and **Contact** information (*Figure 1.3*). Department is optional.

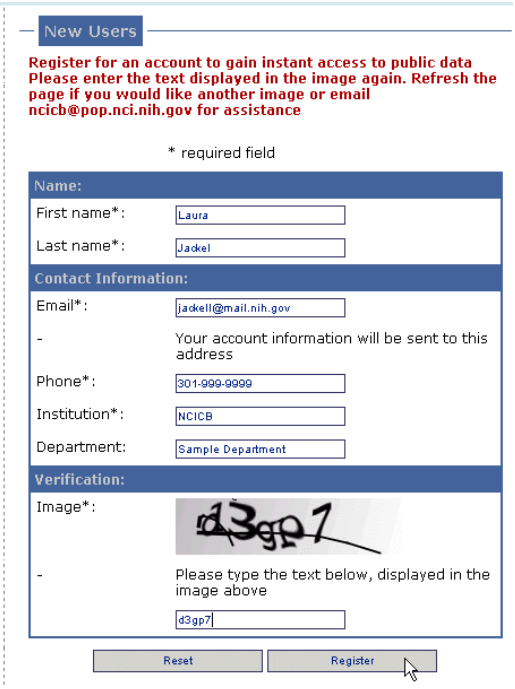


Figure 1.3 New User registration

2. For security purposes, in the **Verification** text box, type the alpha-numeric characters that you see in the image. The entry is case-sensitive.

   To clear all of the fields and start again, click the **Reset** button.

3. To submit your registration, click the **Register** button.

   Note: If you do not enter the information properly, a message appears at the top. Correct the information and continue.

4. If you successfully register, the Logging In panel fills in with a temporary username and password that you can use immediately.

5. You should receive an e-mail registration confirmation and then an additional e-mail containing your new account information. Once you receive your username and password, do not use the temporary account.

   Note: If you have any problems with the form, click the **support** link at the top of the REMBRANDT window.

Logging In

To log into REMBRANDT, you need a username and password. You should have received an e-mail with this information once you registered.

1. On the login panel, enter your **Username** and **Password**.
2. Click the **Login** button. If your login is successful, the Legal Rules of the Road page appears *(Figure 1.4).*

Accepting REMBRANDT Provisions

Once you log in, the Legal Rules of the Road page appears. After reading the provisions, click the **CLICKING HERE** link *(Figure 1.4)* in the lower right-hand corner.

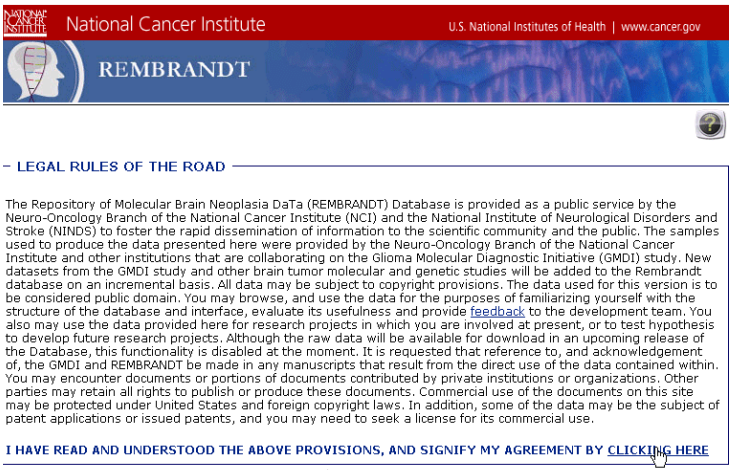


Figure 1.4 Legal Rules of the Road page

The REMBRANDT workspace appears (*Figure 1.5*).

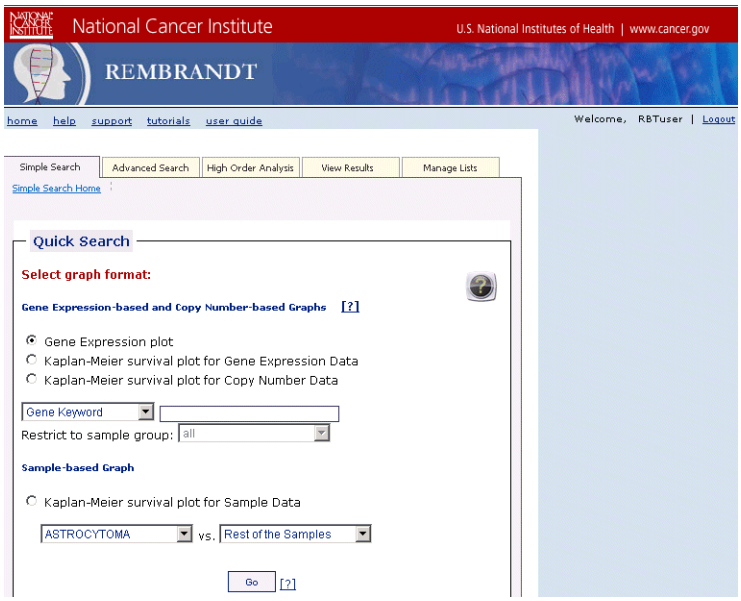


Figure 1.5 The REMBRANDT workspace

REMBRANDT Menu

Information about how to use REMBRANDT is easily accessed from REMBRANDT's menu (*Figure 1.6*) in the top left of the REMBRANDT workspace.

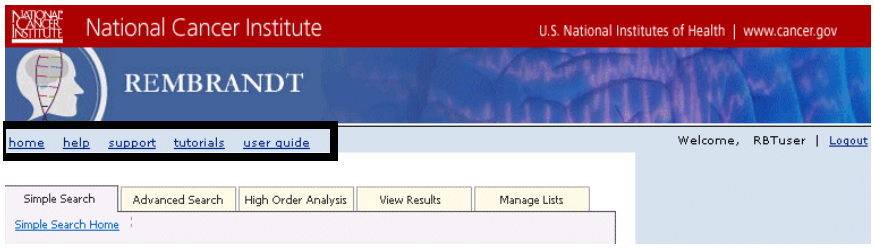


Figure 1.6 REMBRANDT's menu

Table 1.1 describes each item on the REMBRANDT menu.

Menu Option	***Function***
help	Click to access a complete version of online REMBRANDT help. For complete page-level help, click  on any REMBRANDT page. To open a help page to the field's information, click [?]
support	Click to obtain support for REMBRANDT.
tutorials	Click to REMBRANDT access online tuturials to walk you through REMBRANDT.
user guide	Click to access a pdf version of the *REMBRANDT 1.5.1 User's Guide*.

Table 1.1 Getting help with REMBRANDT

REMBRANDT Tabs

Users can perform a variety of tasks in REMBRANDT. *Table 1.2* describes each REMBRANDT tab on the workspace.

Tab Name	***Function***
Simple Search	Search the database and view the following search results: • Gene Expression plots • Kaplan-Meier Survival plots For more information, see *Simple Search Overview*)
Advanced Search	Create the following types of queries and group them to generate results for compound queries: • Gene Expression analysis • Copy Number Data analysis • Clinical Study analysis For more information, see *Advanced Searches Overview*.

Table 1.2 REMBRANDT tabs

Tab Name	Function
High Order Analysis	Run higher order analyses, including class comparisons, hierarchical clustering, and principal component analyses. For more information, see *High Order Analysis Overview*.
View Results	View Advanced Search and High Order Analysis results. Also download static, archive files for use in BRB-ArrayTools. For more information, see *Results Overview*.
Manage Lists	Manage user- or study-defined patient identifier, gene, or reporter lists. For more information, see *Managing Lists Overview*).

Table 1.2 REMBRANDT tabs

REMBRANDT Side Bar

The side bar appears on the right side of the REMBRANDT workspace.*Table 1.3* provides an overview of the information that may appear as you use additional REMBRANDT functions.

Information Displayed	Function
Filter Settings	Displays the filter settings for the following: • A Principal Component Analysis (see *Performing a Principal Component Analysis* on page 50 • A Hierarchical Clustering Analysis (see *Performing Hierarchical Clustering Analysis* on page 52
Queries	Lists queries created with the Advanced Queries function and enables you to add, copy, edit, and delete existing queries. See *Managing Individual and Compound Queries* on page 43.
PatientDID List	Displays the default *PatientDID lists* provided with REMBRANDT, and displays in red any PatientDID lists added to REMBRANDT. See *Adding New Lists* on page 78. To display the items in any type of list, hover over the name and a popup displays the data items. To export a list to a spreadsheet file, double-click the list name.
Gene List	Displays in red any *Gene lists* added to REMBRANDT. See *Adding New Lists* on page 78.
Reporter List	Displays in red any *Reporter lists* added to REMBRANDT. See *Adding New Lists* on page 78.

Table 1.3 Getting help with REMBRANDT

Application Support

For any general information about the application, application support or to report a bug, contact NCICB Application Support.

Email: ncicb@pop.nci.nih.gov	When submitting support requests via email, please include: ● Your contact information, including your telephone number. ● The name of the application/tool you are using ● The URL if it is a Web-based application ● A description of the problem and steps to recreate it. ● The text of any error messages you have received
Application Support URL	http://ncicb.nci.nih.gov/NCICB/support
Telephone: 301-451-4384 Toll free: 888-478-4423	Telephone support is available: Monday to Friday, 8 am – 8 pm Eastern Time, excluding government holidays.

Logging Out

To log out of REMBRANDT, follow these steps.

1. On the REMBRANDT workspace, click the **logout** link in the upper right-hand corner.

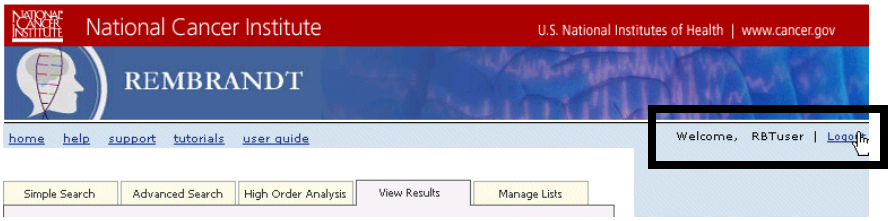


Figure 1.7 Logout link

The Logout page appears.

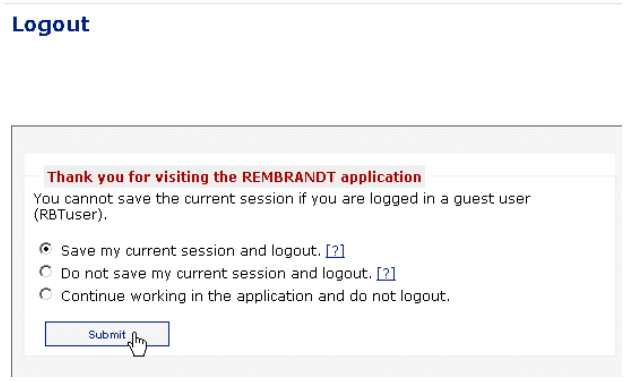


Figure 1.8 Logout link

Select one of the following options:

- To return to REMBRANDT, select **Continue working in the application and do not logout**.
- To log out of REMBRANDT *without* saving the session, select **Do not save my current session and logout**.
- To log out and save your session, select **Save my current session and logout**.

2. To fill out a three-question survey, click **Click Here to take a quick feedback survey**. Answer the questions.
3. Click the **Submit** button.

CHAPTER

2

CONDUCTING A SIMPLE SEARCH

This chapter describes how to use REMBRANDT to conduct simple searches of the REMBRANDT repository and create graphs from the results obtained.

Topics in this chapter include:

Simple Search Overview

The Simple Search page enables you to perform the following types of searches:

- Gene Expression search
- *Kaplan-Meier* survival plot for the following:
  - Gene Expression Data search
  - Copy Number Data search
  - Sample Data search

Results are generated for each search. The Kaplan-Meier survival plots also create Clinical reports and plots.

Gene Expression Simple Search

To create a gene expression plot, follow these steps:

1. From the Simple Search page, select **Gene Keyword**.

   Note: If you do not enter a valid gene symbol, the following message appears: *The gene you entered is either invalid, or not in the database. Please select another.* Close the message window, and enter another gene symbol.

2. Enter a gene keyword, for example, enter a *HUGO* gene symbol such as `EGFR` or `WT1`, to plot a gene expression profile based on the expression of your gene of interest.

3. Click the **Go** button.

Eliminating Aliases

If a message indicates that one or more genes or their aliases have been found, follow these steps:

1. Select the appropriate option from the drop-down list (*Figure 3.8*).

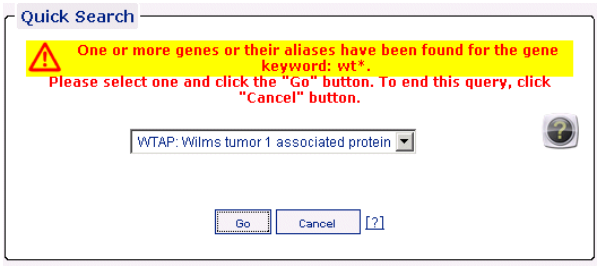


Figure 2.8 Alias message

   To end the search, click **Cancel** button.

2. To continue, click the **Go** button.

Understanding a Gene Expression Plot

When you perform a Gene Expression simple search, by default the **Median** Gene Expression Plot (*Figure 2.1*) appears.

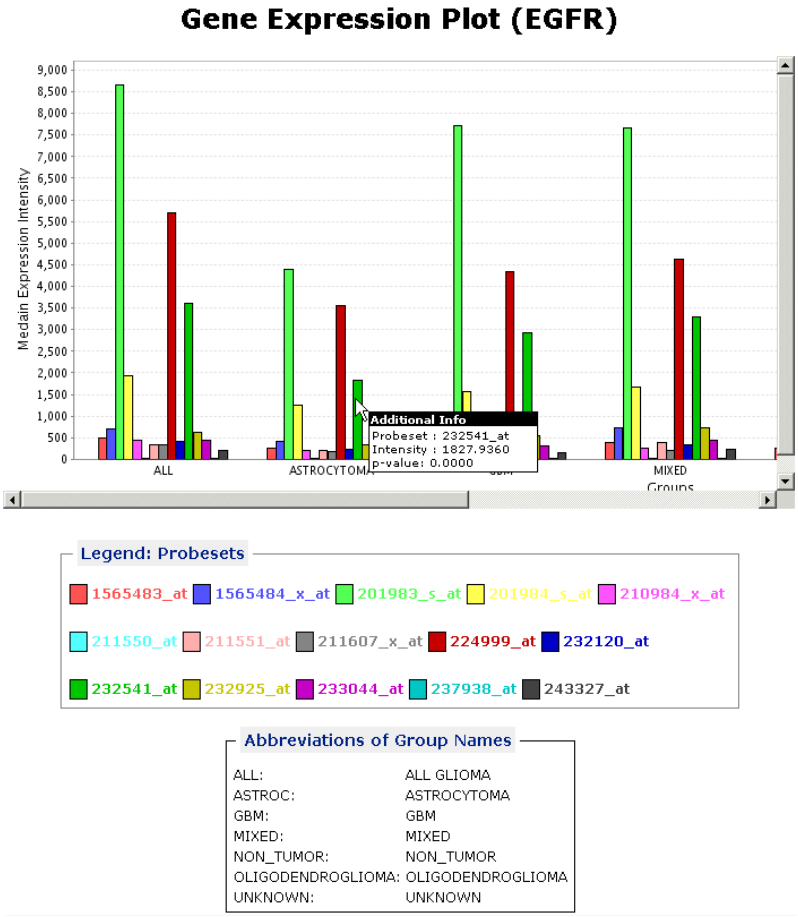


Figure 2.1 Median Gene Expression Plot page

Table 2.1 describes each area of the Gene Expression Plot page.

Item	***Special Instructions***
Data Selection	Select the **Affymetrix** link to repaint the graph. Select the **Unified** link to view a unified gene expression with lesser reporters. This displays a gene-based view of the expression data. To obtain the unified gene expression values, the probe-level data is processed with custom CDF (Chip Definition Files) that rearranges Affymetrix probes into splice-form based probesets. Probes mapped to alternatively spliced exons are grouped into a distinct probeset. The most 3` probes are selected for processing. Non-specific probes are masked before processing.

Table 2.1 Understanding the Gene Expression Plot page

Item	*Special Instructions*
Graph Type	Displays different versions of the Gene Expression Plot. • **Median** is the default graph shown when you perform a simple search. For additional graph details, see *Median Plot Details*. • **Geometric Mean** displays mean expression intensity (Geometric mean) versus Groups. For additional graph details, see *Geometric Mean Plot Details*. • **Log2 Intensity** displays average expression intensities for the gene of interest. For additional graph details, see *Log2 Intensity Gene Expression Plot Details*. • **Box and Whisker Log2 Intensity** displays a Box and Whisker plot or box plot. For additional graph details, see *Box and Whisker Log2 Intensity Gene Expression Plot Details*.
Click here to open plot in a new window	Click the link to open the current graph in a new window and adjust the display. You can then save, print, and e-mail the graph. See *Saving, Printing, and E-mailing a Gene Expression Plot*.
Legend Probesets	Indicates the color for each probeset appearing in the graph.
Abbreviations of Group Names	Lists the complete name of each group abbreviation in the plot.
Print this Graph	Click to print the graph.

Table 2.1 Understanding the Gene Expression Plot page

Median Plot Details

The **Median** Gene Expression Plot (*Figure 2.2*) displays the median expression versus Groups.

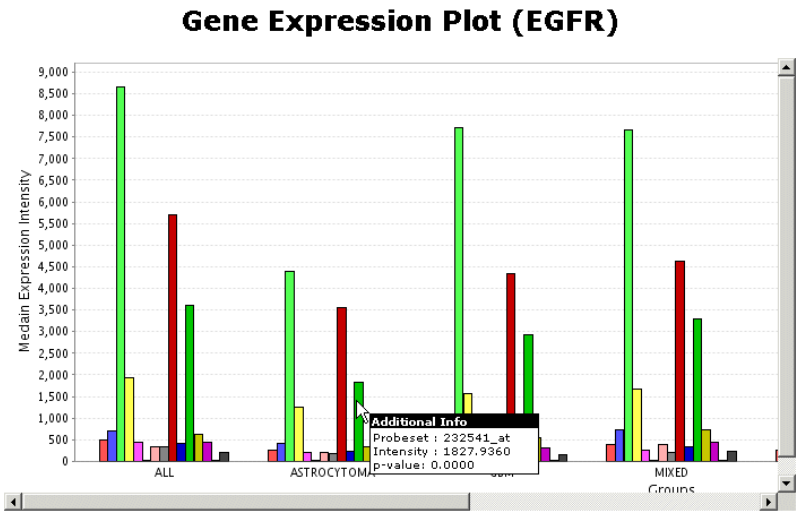


Figure 2.2 Median Gene Expression Plot

Mouse-over a bar on the graph to display Additional Information. *Table 2.3* describes Additional Information details.

Item	***Special Instructions***
Probeset	Each probeset contains multiple probe pairs. Each probe pair consists of two groups of probes—one called a perfect match (PM) and the other called a mismatch (MM). The perfect match is a set of oligonucleotide whose sequence exactly matches the gene of interest; the mismatch differs from the perfect match at one base position in the middle of the sequence.
Intensity	The median value calculated for each comparison group.
***p*-value**	The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples.

Table 2.2 Median Gene Expression Plot Additional Information

Geometric Mean Plot Details

The **Geometric Mean** Gene Expression Plot (*Figure 2.2*) displays mean expression intensity (Geometric mean) versus Groups.

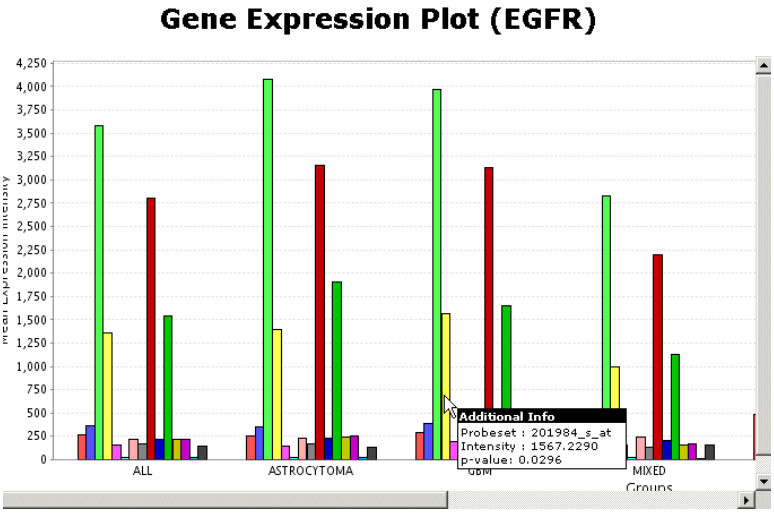


Figure 2.3 Geometric Mean Gene Expression Plot

Mouse-over a bar on the graph to display Additional Information. *Table 2.3* describes Additional Information details.

Item	***Special Instructions***
Probeset	Each probeset contains multiple probe pairs. Each probe pair consists of two groups of probes—one called a perfect match (PM) and the other called a mismatch (MM). The perfect match is a set of oligonucleotide whose sequence exactly matches the gene of interest; the mismatch differs from the perfect match at one base position in the middle of the sequence.
Intensity	The geometric mean value calculated for each comparison group.
***p*-value**	The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples.

Table 2.3 Geometric Mean Gene Expression Plot Additional Information

Log2 Intensity Gene Expression Plot Details

The **Log2 Intensity** Gene Expression Plot (*Figure 2.4*) displays average expression intensities for the gene of interest based on Affymetrix GeneChip arrays (U133 Plus 2.0 arrays). Multiple probesets (for some genes) are designed to measure the expression of the gene of interest. For more information on the probeset design strategy for human genes, go to http://www.affymetrix.com.

Group average (sample averages based on tumor subtypes in six categories, Glioblastoma Multiforme, Oligodendroglioma, Astrocytoma, Mixed, Unclassified, and Unknown tumors) is calculated for each probeset and is plotted on the Y-axis for each tumor type.

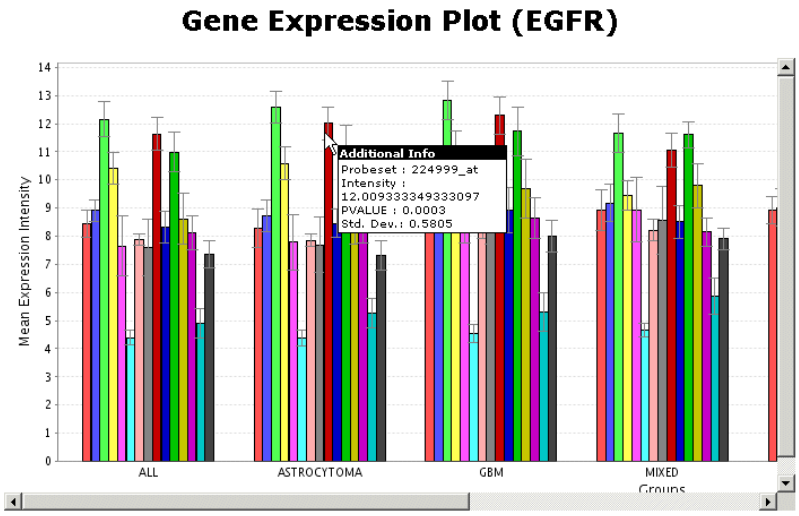


Figure 2.4 Log2 Intensity Gene Expression plot

Mouse-over a bar on the graph to display Additional Information. *Table 2.4* describes Additional Information details.

Item	***Description***
Probeset	Each probeset contains multiple probe pairs. Each probe pair consists of two probes—one called a perfect match (PM) and the other called a mismatch (MM). The perfect match is an oligonucleotide whose sequence exactly matches the gene of interest; the mismatch differs from the perfect match at one base position in the middle of the sequence.
Intensity	The mean value calculated for each comparison group.
***p*-value**	The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples.
Std. Dev.	The standard deviation value of a comparison group, such as GBM or Astrocytoma, for a particular probeset or gene. Standard deviation is a statistical measure of spread or variability.

Table 2.4 Log2 Intensity Gene Expression Plot Additional Information

Box and Whisker Log2 Intensity Gene Expression Plot Details

The **Box and Whisker Log2 Intensity** Gene Expression Plot (*Figure 2.5*) displays a box plot without all the individual data points. Example uses of box and whisker plots include the following:

- Indicate whether a distribution is skewed and whether there are potential unusual observations (outliers) in the dataset.
- Perform a large number of observations.
- Compare two or more datasets.
- Compare distributions because the centre, spread, and overall range are immediately apparent.

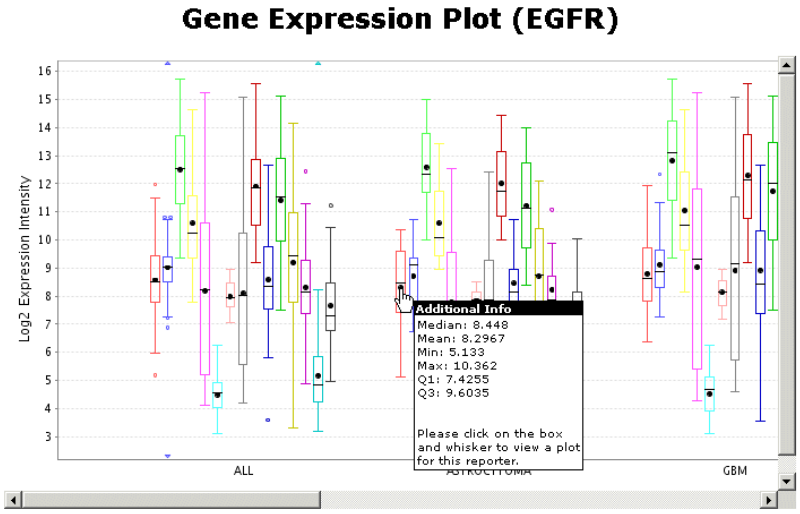


Figure 2.5 Box and Whisker Log2 Intensity Gene Expression plot

A box and whisker plot or box plot is a graph that presents information from a five-number summary. To display the summary about a probeset for one group, mouse over the probe-set on the plot to display the **Additional Information**. *Table 2.5* describes Additional Information details.

Item	***Description***
Median	Median value of log 2 (or ratio) gene expression values for a particular probeset or unified gene.
Mean	Mean value of log 2 (or ratio) gene expression values for a particular probeset or unified gene.
Min.	The minimum value.
Max.	The maximum value.
Q1	The bottom of the box. The first quartile is the median of the lower part of the data.
Q3	The top of the box. The third quartile is the median of the upper part of the data.

Table 2.5 Box and Whisker Log2 Intensity Gene Expression Plot Additional Information

Item	Description
...plot	Represents the probeset name.

Table 2.5 Box and Whisker Log2 Intensity Gene Expression Plot Additional Information

Note: To display a coin plot for the reporter, *click* in the box. A *coin plot* is a box-and-whisker plot with all the individual data points (see *Displaying a Coin Plot*).

In the box-and-whisker plot, the individual probeset summary is represented as follows (*Figure 2.6*). Horizontal lines (the "whiskers") extend to, at the most, 1.5 times the box length (the interquartile range) from either or both ends of the box. They end at an observed value, thus connecting all the values outside the box that are not more than 1.5 times the box width away from the box.

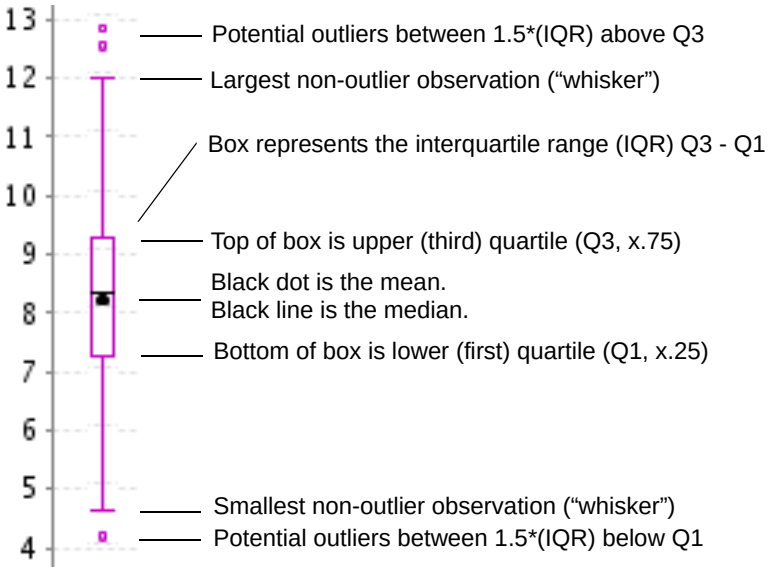


Figure 2.6 Box and Whisker Plot details

Displaying a Coin Plot

A *coin plot* is box-and-whisker plot (*Figure 2.7*) with all individual data points. This enables you to obtain a diagram representing a statistical summary of the data without the disadvantage of concealing the real data.

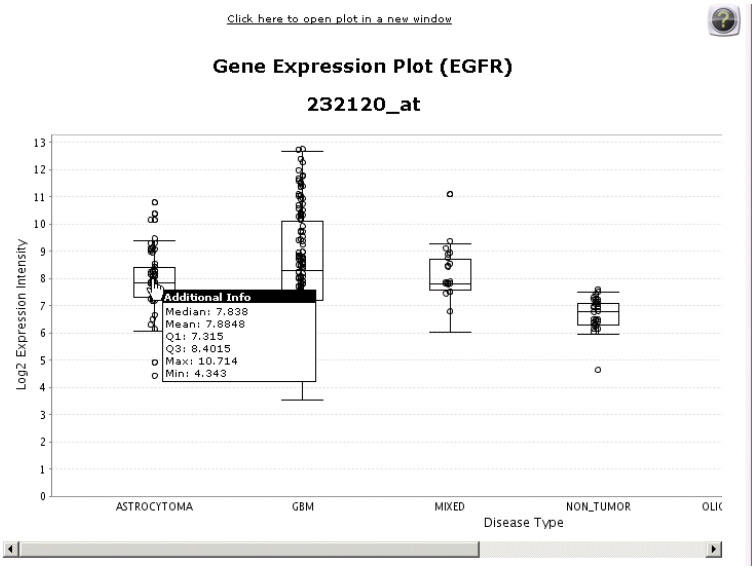


Figure 2.7 Coin Plot for a Probeset

The coin plot is a graph that presents information from a five-number summary. To display the summary about a probeset for one group, mouse over the probe-set on the plot to display the **Additional Information**. *Table 2.5* describes Additional Information details.

Item	***Description***
Median	Median value of log 2 (or ratio) gene expression values for particular a probeset or unified gene.
Mean	Mean value of log 2 (or ratio) gene expression values for particular a probeset or unified gene.
Min.	The minimum value.
Max.	The maximum value.
Q1	The bottom of the box. The first quartile is the median of the lower part of the data.
Q3	The top of the box. The third quartile is the median of the upper part of the data.
...plot	Represents the probeset name.

Table 2.6 Coin Plot Additional Information

In the coin plot, the individual probeset summary is represented as follows (*Figure 2.8*). Horizontal lines (the "whiskers") extend to, at the most, 1.5 times the box length (the interquartile range) from either or both ends of the box. They end at an observed value, thus connecting all the values outside the box that are not more than 1.5 times the box width away from the box.

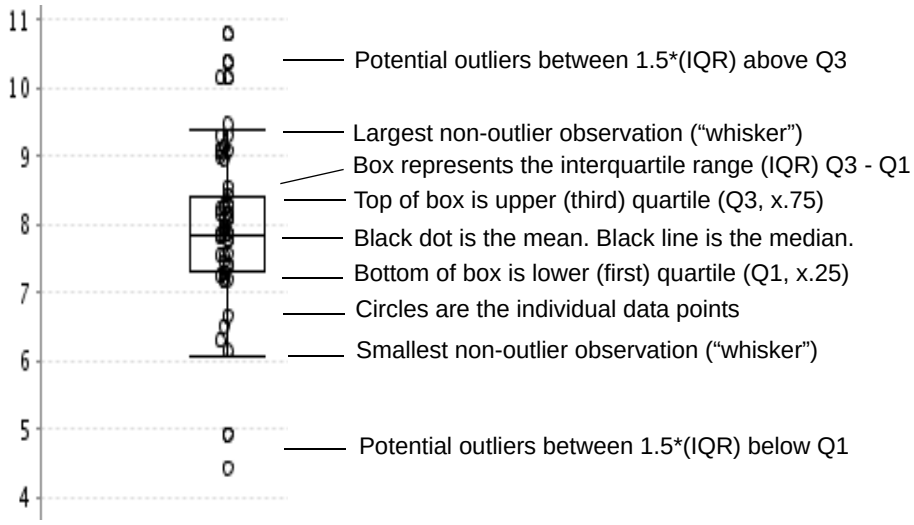


Figure 2.8 Coin Plot details

Saving, Printing, and E-mailing a Gene Expression Plot

By opening a Gene Expression plot in a new window (*Figure 2.9*), you can perform a number of tasks with the Gene Expression plot.

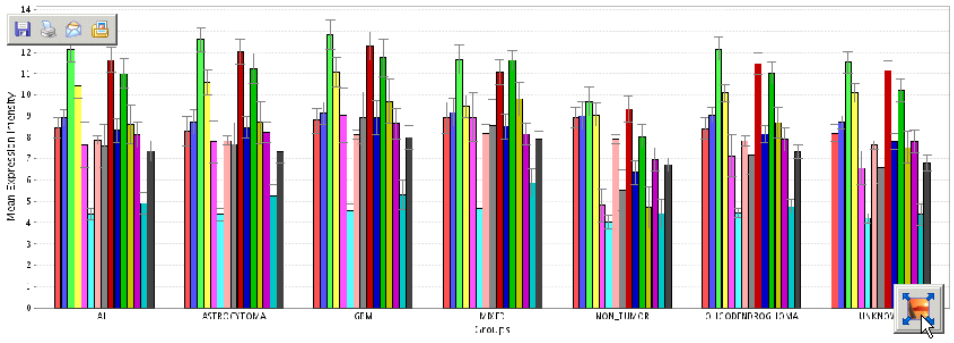


Figure 2.9 Displaying a Gene Expression Plot in a new window

Table 2.7 describes the tasks you can perform when you open a Gene Expression plot in a new window.

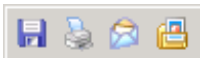
Icon	*Special Instructions*
	Mouse-over the graph, and the icon appears in the lower right-hand corner. Click the icon to enlarge or reduce the size of the graph.
	Mouse-over the graph, and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph.

Table 2.7 A Gene Expression Plot in a new window

K-M Gene Expression Simple Search

To create a *Kaplan-Meier* survival plot for gene expression data (K-M Gene Expression), follow these steps:

1. From the Simple Search page, select **Create Kaplan-Meier survival plot for Gene Expression Data**.
2. Enter a gene keyword, for example, enter a *HUGO* gene symbol, such as `EGFR` or `WT1`, to plot a gene expression profile based on the expression of your gene of interest.

   Note: If you do not enter a valid gene symbol, the following message appears: *The gene you entered is either invalid, or not in the database. Please select another.* Close the message window, and enter another gene symbol.
3. From the **Restrict to Sample Group** drop-down list, select a saved sample group.
4. Click the **Go** button.

Redrawing the K-M Survival Plot for Gene Expression Data

To redraw a K-M Gene Expression data, follow these steps:

Note: If you restricted the search to a group, **Constrained to group** appears at the top.

1. To dynamically modify the fold change thresholds and redraw the plot, adjust the **Up-Regulated** and **Down-Regulated** values.
2. Specify a **Unified** or a **Affymetrix Reporter Type**.
3. To visualize the K-M plot for the unified probeset, click the **Reporters** drop-down list *(Figure 2.10).*

Figure 2.10 Redrawing a Kaplan-Meier Gene Expression data

4. Select an individual reporter or one of the following options:
   - **Median** is the median value of all Reporters (default).
   - **Mean** is the mean value of all Reporters.
5. Click the **Redraw Graph** button.

Understanding K-M Survival Plot for Gene Expression Data

A K-M Survival Plot for Gene Expression Data (*Figure 2.11*) displays the survival rate at each time point for samples with certain expression characteristics (e.g., EGFR expression levels in tumor samples greater than those in the non-tumor samples by 3 fold or higher). Kaplan-Meier estimates are calculated based on the last follow-up time and the censor status (0=alive, 1=dead) from the samples of interest.The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph. You can dynamically modify the fold change (up and down regulation) thresholds and redraw the plot.

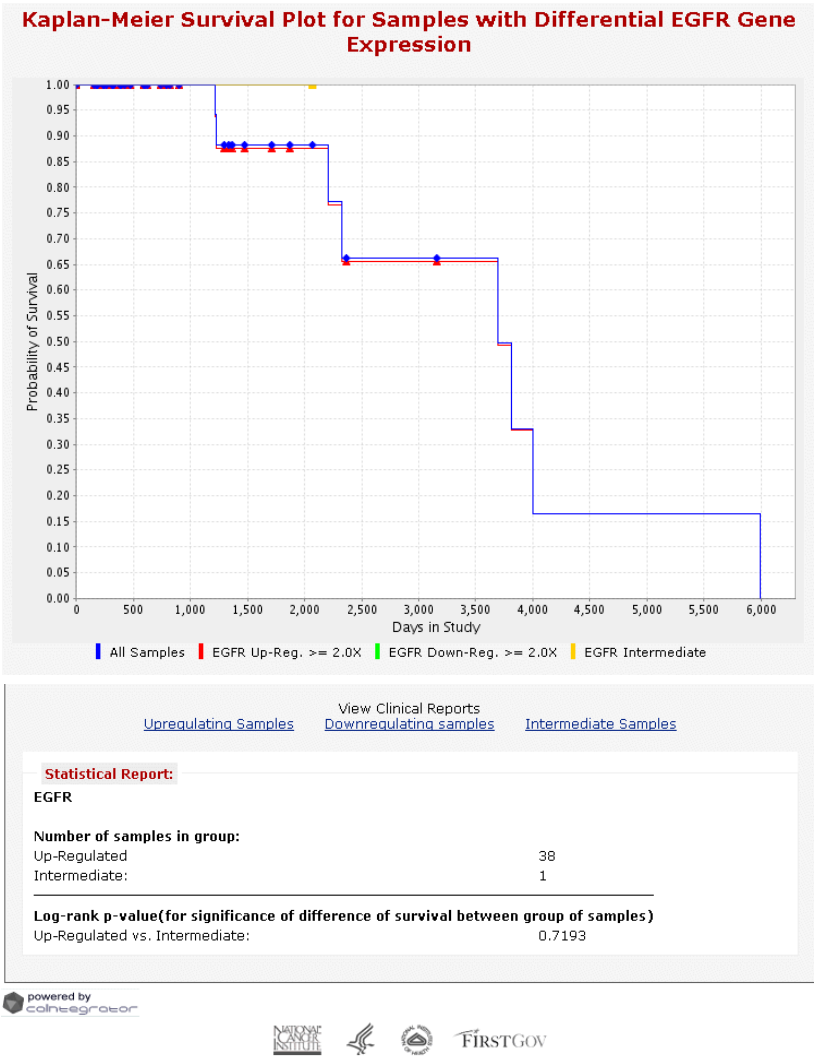


Figure 2.11 Kaplan-Meier Survival Plot for Gene Expression Data

Table 2.8 describes areas on the Kaplan-Meier Survival Plot for Gene Expression data page.

Item	*Special Instructions*
	Mouse-over the graph, and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. See *Saving, Printing, and E-mailing a Gene Expression Plot* on page 21.
View Clinical Reports	When you apply a gene expression filter, REMBRANDT provides links to display the gene expression for **Upregulating Samples**, **Downregulating Samples**, and **Intermediate Samples**. For more information, see *Clinical Reports*.
Statistical Report	• Displays the gene keyword entered as search criteria for the plot. • Displays the reporter selected for the plot. • **Number of Samples** specifies the number of Up-Regulated, Intermediate, Down-Regulated samples, if any. • **Log-rank *p*-Value** indicates the significance of the difference in survival between any two groups of samples segregated based on gene expression of the gene of interest. The log rank *p*-value is calculated using Mantel-Haenszel procedure. The *p*-values are recalculated every time a new threshold is selected.

Table 2.8 Kaplan-Meier Survival Plot for Gene Expression Data page

K-M Copy Number Simple Search

To create a *Kaplan-Meier* survival plot for copy number-based data (KM Copy Number), follow these steps:

1. From the Simple Search page, select **Create Kaplan-Meier survival plot for Copy Number Data**.
2. Enter one type of search criteria:
   - Select **Gene Keyword** and enter a *HUGO* gene symbol, such as `EGFR` or `WT1`, to plot a Kaplan-Meier survival plot based on the expression of your gene of interest.
   - Select **SNP Probe Set ID** and enter an *SNP* array vendor specific identifier. For example, `SNP_A-1650833` is one of the SNP probeset for Affymetrics Human Mapping 100K Set chip.

   Note: If you do not enter a valid ID, the following message appears: *The gene you entered is either invalid, or not in the database. Please select another.* Close the message window, and enter another ID.
3. To search on a sample group saved in a previous search, select a group from the **Restrict to Sample Group** list box.
4. Click the **Go** button.

Redrawing the K-M Survival Plot for Copy Number Data

To redraw a KM Copy Number graph, follow these steps:

Note: If you restricted the search to a group, **Constrained to group** appears at the top.

1. Select the amplification and deletion criteria.
2. To visualize the K-M plot for the unified probeset, select a value from the **Reporters** drop-down list *(Figure 2.12).*

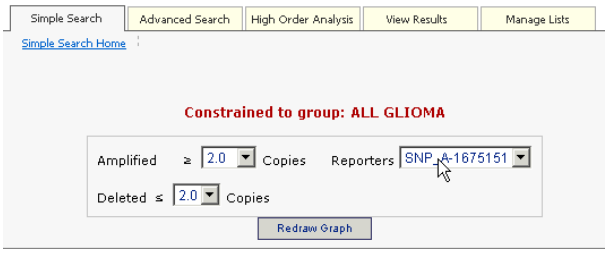


Figure 2.12 Redrawing a Kaplan-Meier Survival Plot for Copy Number data page

3. Select an individual reporter or one of the following options:
   - **Median** is the median value of all Reporters (default)
   - **Mean** is the mean value of all Reporters.
4. Click the **Redraw Graph** button.

Understanding K-M Survival Plot for Copy Number Data

A **gene keyword** search displays a plot (*Figure 2.13*) for each SNP probeset for samples with certain amplification/deletion characteristics (e.g., amplification of the cytoband that EGFR maps to 7p11.2). Each SNP probeset is associated with the gene of interest to show the survival rate at each time point. Kaplan-Meier estimates are calculated based on the last follow-up time and the censor status (0=alive, 1=dead) from the samples of interest. The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph. You can dynamically modify the fold change (up and down regulation) thresholds and redraw the plot.

An **SNP Probeset ID** search displays a plot showing the survival rate at each time point for samples with certain expression characteristics (e.g. EGFR expression levels in tumor samples are greater than those in the non-tumor samples by 3 fold or higher). Kaplan-Meier estimates are calculated based on the last follow-up time and the censor status (0=alive, 1=dead) from the samples of interest. The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph. You can dynamically modify the fold change (up and down regulation) thresholds and redraw the plot.

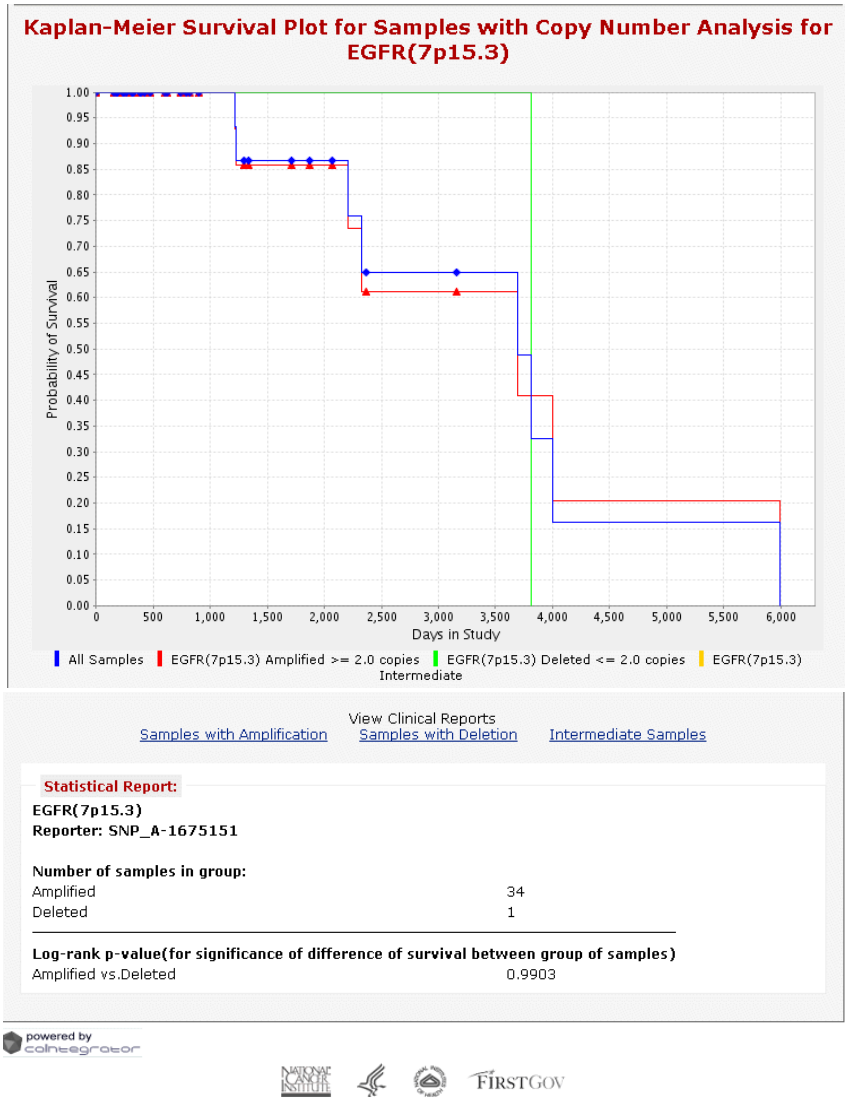


Figure 2.13 Kaplan-Meier Survival Plot for Copy Number Data

Table 2.9 describes areas on the Copy Number-based Plot page.

Item	***Special Instructions***
	Mouse-over the graph and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. *Saving, Printing, and E-mailing a Gene Expression Plot* on page 21.
View Clinical Reports	When you apply a copy number filter, REMBRANDT provides links to display the copy number data for samples. For more information, see *Clinical Reports*.
Statistical Report	• Displays the search criteria for the plot. • Displays the reporter selected for the plot. • **Number of Samples** specifies the number of different types of samples, if any. • **Log-rank *p*-value** indicates the significance of the difference in survival between any two groups of samples segregated based on gene expression of the gene of interest. The log rank *p*-value is calculated using Mantel-Haenszel procedure. The *p*-values are recalculated every time a new threshold is selected.

Table 2.9 Kaplan-Meier Survival Plot for Copy Number Data

K-M Sample Search

To create a *Kaplan-Meier* survival plot for sample data (K-M Sample), follow these steps:

1. From the **Simple Search** tab, select **Create Kaplan-Meier survival plot for Sample Data**.
2. Select a sample from each drop-down list for comparison purposes.
3. Click the **Go** button. The Kaplan-Meier survival plot appears *(Figure 2.14).*

Understanding K-M Survival Plot for Sample Data

A Kaplan-Meier Survival Plot for Sample Data *(Figure 2.14)* shows the survival rate at each time point for samples with certain expression characteristics (e.g. EGFR expression levels in tumor samples are greater than those in the non-tumor samples by 3 fold or higher). Kaplan-Meier estimates are calculated based on the last follow-up time and the censor status (0=alive, 1=dead) from the samples of interest. The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph.

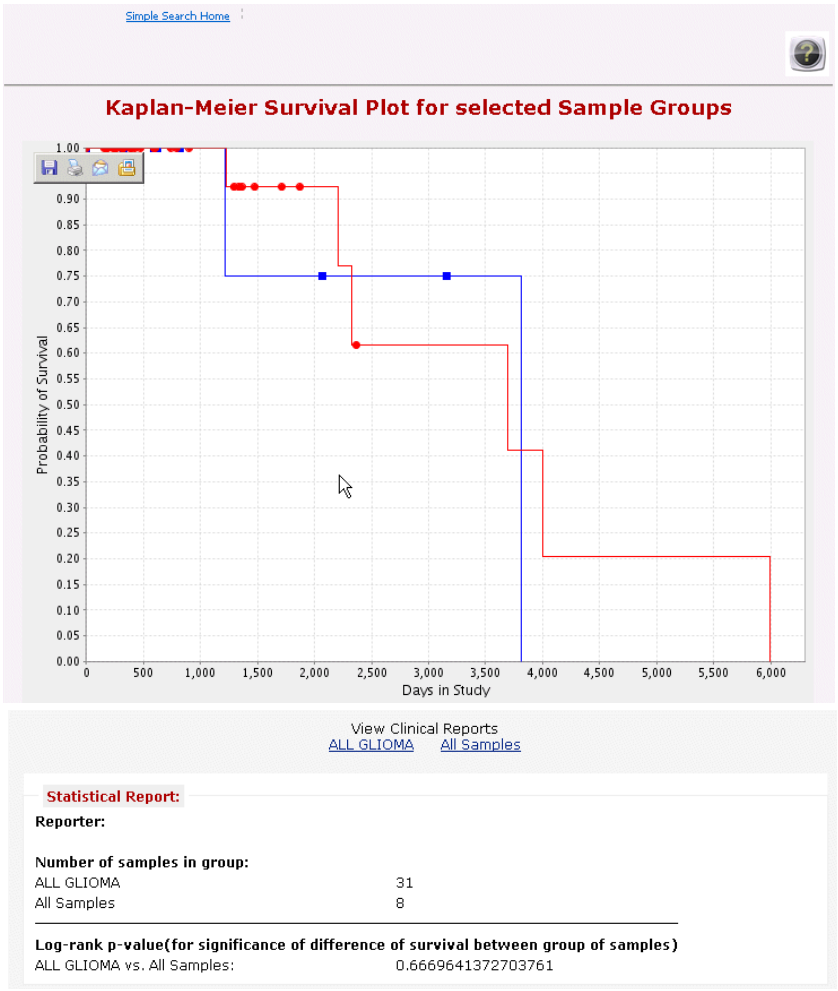


Figure 2.14 Kaplan-Meier Survival Plot for Sample Data

Table 2.10 describes areas on the Kaplan-Meier Survival Plot for Sample Data page.

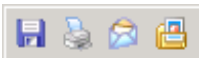
Item	***Special Instructions***
	Mouse-over the graph, and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. See *Saving, Printing, and E-mailing a Gene Expression Plot* on page 21.
View Clinical Reports	To display clinical data for the selected sample groups, click the group link. For more information, see *Clinical Reports*.

Table 2.10 Kaplan-Meier Survival Plot for Sample Data page

Item	*Special Instructions*
Statistical Report	• Displays the search criteria for the plot. • Displays the reporter selected for the plot. • **Number of Samples** specifies the number of Up-Regulated, Intermediate, Down-Regulated samples, if any. • **Log-rank *p*-value** indicates the significance of the difference in survival between any two groups of samples segregated based on gene expression of the gene of interest. The log rank *p*-value is calculated using Mantel-Haenszel procedure. The *p*-values are recalculated every time a new threshold is selected.

Table 2.10 Kaplan-Meier Survival Plot for Sample Data page

CHAPTER

3

CONDUCTING ADVANCED SEARCHES

This chapter describes how to create compound queries to generate graphs.

Topics in this chapter include:

Advanced Searches Overview

The Advanced Search function enables you to add individual queries to REMBRANDT and then group the queries to create and generate results for a *compound query*. The following is an overview.

1. The Advanced Search - Build Query page enables you to define advanced searches in three categories:
    - Gene Expression Analysis
    - Copy Number Analysis
    - Clinical Analysis
2. Once you create a query, you can add, copy, edit, and delete queries from the side bar.
3. To create a compound query, click the **Finalize Query** button or the **Refine Query** option.
4. Validate the compound query and generate results on the View Results page.

Gene Expression Advanced Search

To define an advanced gene expression search, follow these steps:

1. On the Gene Expression page, in the **Query Name** box, you are required to enter a name for the gene expression query. The name must be unique among all the queries in the current session (*Figure 3.1*).

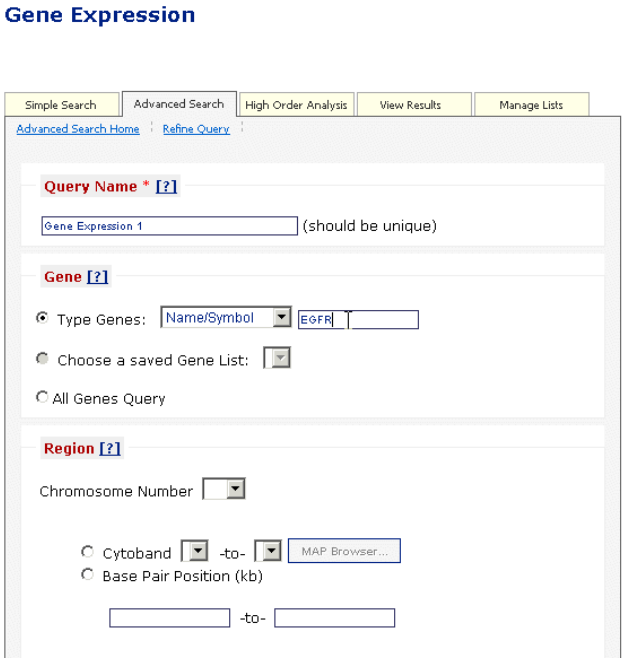


Figure 3.1 Advanced Gene Expression page (top portion)

2. You are required to enter at least one search criteria for the query. *Table 3.1* lists the available search criteria:

Criteria	***Item Name***	***Special Instructions***
Gene	Type Genes	Select a gene identifier option (**Name/Symbol**, **Locus Link ID**, or **GenBank AccNo.**), and then enter or paste comma-delimited values for the genes to be searched.
	Choose a Saved Gene List	Drop down the list box and select a saved gene list. If you have not added a *Gene List* with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
	All Genes	Click if you do not wish to specify a list of genes but want to display data for all the genes analyzed. You must apply this option to a pre-existing result set, as described in Step 1 of *Refining a Query*.

Table 3.1 Advanced Gene Expression search criteria instructions

Criteria	*Item Name*	*Special Instructions*
Region	Chromosome Number	Select the chromosomal region of interest (**1-22, X** or **Y**). Cytoband fills in based on the selected chromosome number.
	Cytoband	A context-sensitive list displays only the relevant cytobands for the selected chromosome. Select a cytoband range.
	Map Browser	Click to conduct a search of cytoband ranges.
	Base Pair Position (kb)	Enter the start and end base pair positions.
Clone Id/Probe Set ID	Type Reporters	Select an option (**Probeset ID** or **IMAGE ID**), and then enter or paste comma-delimited values for the identifiers to be searched. IMAGE identifiers must start with `IMAGE:`.
	Choose a Saved Reporters List	Drop down the list box and select a saved Reporters list. If you have not added a Reporters list with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
Gene Ontology (GO) Classifications	(list box)	Enter a *Gene Ontology (GO)* ID in the format `GO:#######` to search for one or more branches of the GO hierarchy. For example, enter `GO:0005006` (epidermal growth factor receptor activity (12)).
	Go Browser	Click the button to search for and select a *GO* classification. See *Selecting a Gene Ontology (GO) Classification*.
Pathways		Click the button to search for and select a pathway. See *Selecting a Pathway*.
		Click the button to search for and select a pathway. See *Selecting a Pathway*.
	clear text area	Click the link to remove the selected pathway(s).
Clone Location	3’ UTR	*Future Implementation*
	5’ UTR	*Future Implementation*

Table 3.1 Advanced Gene Expression search criteria instructions

At the bottom of the Gene Expression page, you can optionally add disease type criteria to the search (*Figure 3.2*).

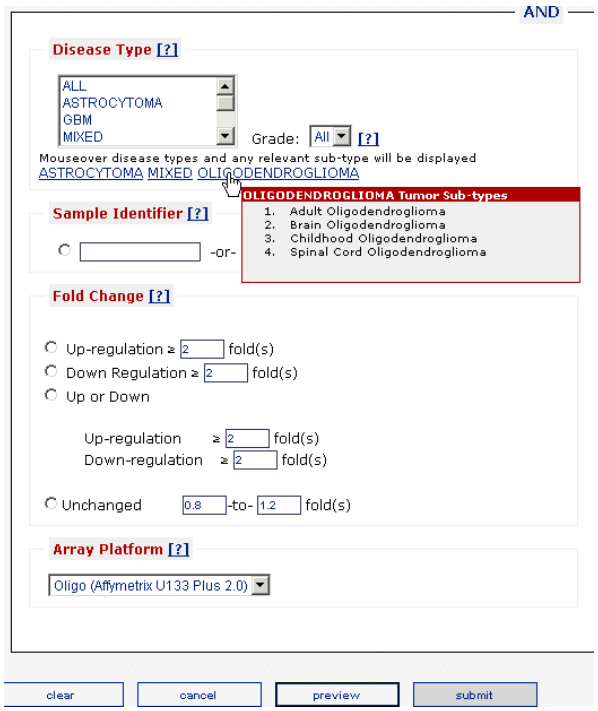


Figure 3.2 Advanced Gene Expression Disease Type

3. Optionally, you can combine a disease type with the query. *Table 3.2* lists the Disease Type items:

Criteria	***Item Name***	***Special Instructions***
Disease Type	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL+click the remaining disease types. To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	*Future Implementation*
Sample Identifier	(list boxes	To further filter the search, enter or paste comma-delimited sample identifiers to be searched. OR Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).

Table 3.2 Advanced Gene Expression disease type criteria instructions

Criteria	*Item Name*	*Special Instructions*
Fold Change	Up-Regulation Down-Regulation Unchanged	Specify the threshold for the differential regulation. This returns differential expression ratios between tumor and non-tumor samples for a particular reporter. To create an **All Genes** query, you must select a fold change threshold of 4 or above.
Array Platform	(list box)	Select an array platform.

Table 3.2 Advanced Gene Expression disease type criteria instructions

For the last step, you submit the search. Submitting the search saves the search criteria, and the Advanced Search - Build Query page appears. No results are generated until you create a query from your saved searches.

4. To save the search and return to the Advanced Search tab, click the **Submit** button.

   You can also use the other buttons as follows:

   - To clear the values entered on the page and enter new values, click the **Clear** button.
   - To clear the values and return to the Advanced Search page, click the **Cancel** button.
   - To display a preview of the report generated by the search results, click the **Preview** button.

Selecting a Gene Ontology (GO) Classification

Once you select the **GO Browser** button on the Gene Expression page, a list of *GO IDs* appears (*Figure 3.3*).

Figure 3.3 GO ID list

To add the GO ID to your advanced search, click on the appropriate GO term. The GO ID is added to the Gene Expression page.

Selecting a Pathway

To select a pathway of interest, follow these steps.

1. Browse the pathway list, and check the pathways of interest (*Figure 3.4*).

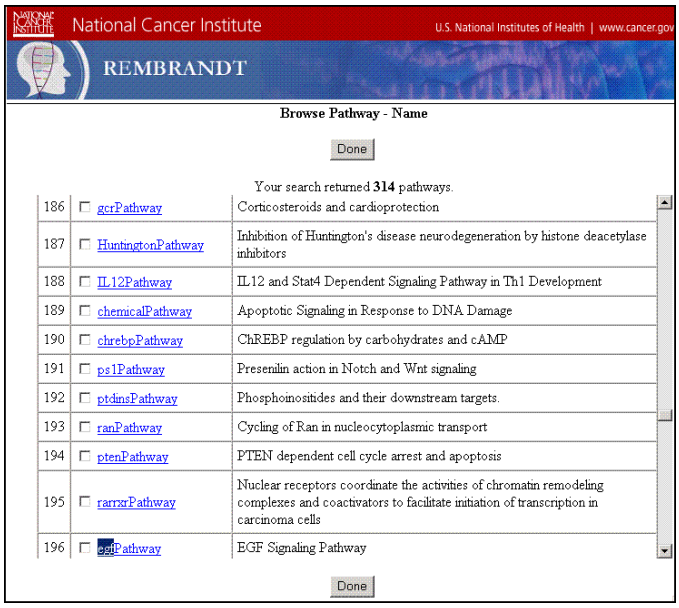


Figure 3.4 caBIO pop-up window

Note: For more information about a pathway, click the pathway name.

2. To quickly find the pathway of interest, type CTR-F. The Search dialog box appears *(Figure 3.5).*

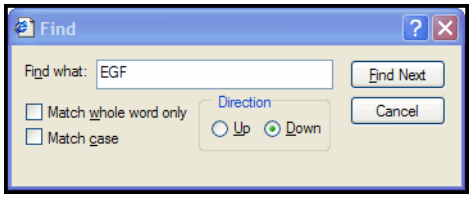


Figure 3.5 Search dialog box

3. Enter the pathway into the search text box, and click the **Find Next** button.

   The pathway(s) fulfilling the search criteria appear.

4. Select the check box next to the pathway of interest and click the **Done** button. The selected pathways are added to the query.

Copy Number Advanced Search

To add a copy number data analysis to a query, follow these steps:

1. On the Copy Number Data page, in the **Query Name** box, you are required to enter a name for the query. The name must be unique among all the queries in the current session (*Figure 3.6*).

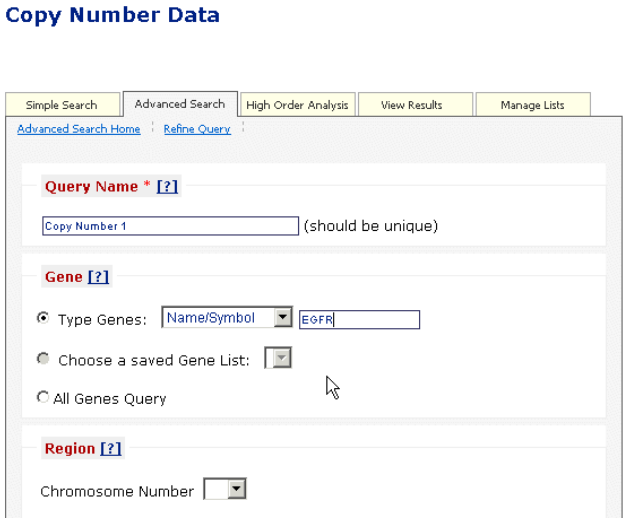


Figure 3.6 Copy Number Data page (top portion)

2. You are required to enter at least one search criteria for the copy number query. *Table 3.3* lists the available search criteria:

Criteria	*Item Name*	*Special Instructions*
Gene	Type Genes	Select a gene identifier option (**Name/Symbol**, **Locus Link ID**, or **GenBank AccNo.**), and then enter or paste comma-delimited values for the genes to be searched.
	Choose a Saved Gene List	Drop down the list box and select a gene list. If you have not added a *Gene List* with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
	All Genes	Click if you do not wish to specify a list of genes but want to display data for all the genes analyzed. You must apply this option to a pre-existing result set, as described in Step 1 of *Refining a Query*.

Table 3.3 Advanced Copy Number search criteria instructions

Criteria	*Item Name*	*Special Instructions*
Region	Chromosome Number	Select the chromosomal region of interest (**1-22, X** or **Y**). Cytoband fills in based on the selected chromosome number.
	Cytoband	A context-sensitive list displays only the relevant cytobands for the selected chromosome. Select a cytoband range.
	Map Browser	Click to conduct a search of cytoband ranges.
	Base Pair Position (kb)	Enter the start and end base pair positions.
Geonomic Annotation Track	(text box)	*Future Implementation*
	Geonomic Browser	*Future Implementation*
SNP Id	Type SNPs	Select an SNP type identifier (**dbSNP ID** or **SNP Probe Set ID**), and then enter or paste comma-delimited SNP values to be searched.
	Choose a Saved SNP List	Drop down the list box and select a saved *SNP* list. If you have not added an SNP list with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
	Validated SNPs	Select one type of Validated SNPs: **All**, **Excluded**, **Included**, or **Only**.
Allele Frequency	Population Type	*Future Implementation*

Table 3.3 Advanced Copy Number search criteria instructions

At the bottom of the Advanced Copy Number page, you can add disease type criteria to the search (*Figure 3.2*).

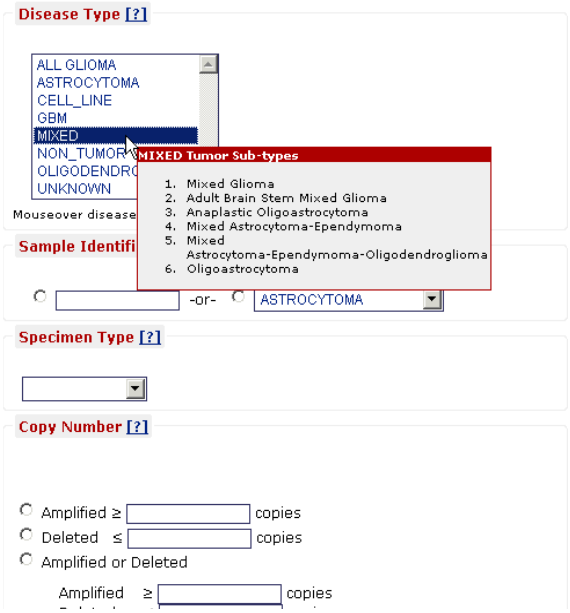


Figure 3.7 Advanced Copy Number Disease Type

3. Optionally, you can combine a disease type with the query. *Table 3.4* lists the Disease Type items:

Criteria	*Item Name*	*Special Instructions*
Disease Type	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL+click the remaining disease types. To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	*Future Implementation*
Sample Identifier	(list boxes)	To further filter the search, enter or paste comma-delimited sample identifiers. OR Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
Specimen Type	Blood Tissue (brain)	Specify the specimen type. • To return all the samples, leave the box blank. • **Blood** returns blood samples only. • **Tissue (Brain)** returns brain tissue samples only.
Copy Number	Amplified Deleted Amplified or Deleted Unchanged	Specify the threshold for the copy number. To create an **All Genes** query, you must select an amplification threshold greater than 10 or a deletion threshold less than 1.
Array Platform	(list box)	Select the array platform.

Table 3.4 Advanced Copy Number disease type criteria instructions

For the last step, you submit the search. Submitting the search saves the search criteria, and the Advanced Search - Build Query page appears. No results are generated until you create a query from your saved searches.

4. To save the search and return to the Advanced Search tab, click the **Submit** button.

   You can also use the other buttons as follows:

   - To clear the values entered on the page and enter new values, click the **Clear** button.
   - To clear the values and return to the Advanced Search page, click the **Cancel** button.
   - To display a preview of the report generated by the search results, click the **Preview** button.

Clinical Study Advanced Search

To add a clinical data analysis to a query, follow these steps:

1. On the Clinical Data page, in the **Query Name** box, you are required to enter a name for the query. The name must be unique among all the queries in the current session (*Figure 3.8*).

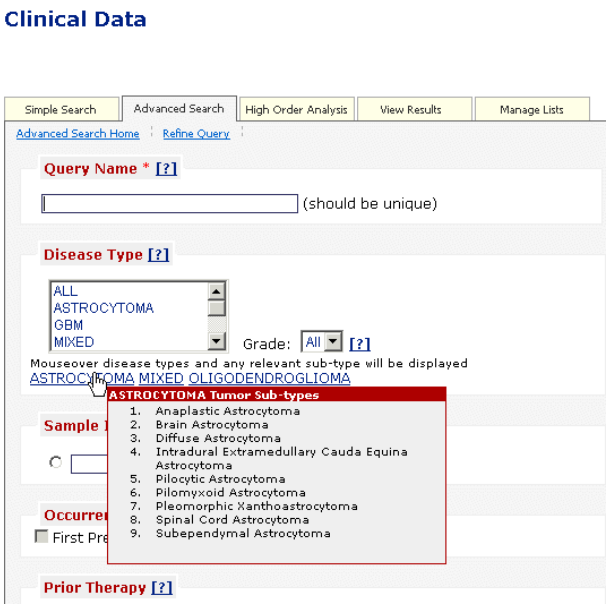


Figure 3.8 Clinical Data page (top portion)

2. You must specify a disease type, and optionally complete the remaining information. *Table 3.5* lists the Disease Type items:

Criteria	***Item Name***	***Special Instructions***
Disease Type	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL + click the remaining disease types. To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	*Future Implementation*
Sample Identifier	(list box)click	To further filter the search, enter or paste comma-delimited sample identifiers. OR Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).

Table 3.5 Advanced Clinical Data criteria instructions

Criteria	*Item Name*	*Special Instructions*
Occurence	First Presentation Recurrence	*Future Implementation*
Prior Therapy	Radiation Radiation Type	Select **Radiation** and then select the type of radiation that the patient received prior to enrollment in the current study.
	Chemo Agent	Select **Chemo** and then select the agent that the patient received prior to enrollment in the current study.
	Surgery Title Outcome	Select **Surgery** and then enter the name of the surgery that the patient had prior to enrollment in the current study and the outcome of the surgery.
Onstudy Therapy	Radiation Radiation Type	Select **Radiation** and then select the type of radiation that the patient received after enrollment in the current study.
	Chemo Agent	Select **Chemo** and then select the agent that the patient received after enrollment in the current study.
	Surgery Title Outcome	Select **Surgery** and then enter the name of the surgery that the patient had after enrollment in the current study and the outcome of the surgery
Survival Range	Lower Upper	Specify the upper and lower limits (in months) for filtering the clinical data based on the age (in years) at which a patient was diagnosed.
Age at Dx	Lower Upper	Specify the upper and lower limits for filtering the clinical data based on the age at which a patient was diagnosed with the disease.
Gender		Select the appropriate gender of the patient.
Race		Select the appropriate race of the patient.
Clinical Evaluation	*Karnofsky*	Score from the Karnofsky Performance status scale, representing the functional capabilities of a person.
	Lansky	Score from an enumerated set of values representing performance status according to the Lansky scale. The Lansky scale is intended for use only with subjects under 12 years old.
	Neuro Exam	The participant's neurologic exam score. Score identifiers are the following: +2 DEFINITELY BETTER +1 POSSIBLY BETTER 0 STABLE -1 POSSIBLY WORSE -2 DEFINITELY WORSE

Table 3.5 Advanced Clinical Data criteria instructions

Criteria	Item Name	Special Instructions
	MRI	Relates to the disease evaluation as measured by scan (MRI/CT). Score definitions are the following: +3 DISAPPEARANCE OF TUMOR (CR) +2 DEFINITELY BETTER (PR) +1 POSSIBLY BETTER 0 UNCHANGED -1 POSSIBLY WORSE -2 DEFINITELY WORSE (PD) -3 DEVELOPMENT OF A NEW LESION (PD)

Table 3.5 Advanced Clinical Data criteria instructions

For the last step, you submit the search. Submitting the search saves the search criteria, and the Advanced Search - Build Query appears. No results are generated until you create a query from your saved searches.

3. To save the search and return to the Advanced Search - Build Query, click the **Submit** button.

   You can also use the other buttons as follows:

   - To clear the values entered on the page and enter new values, click the **Clear** button.
   - To clear the values and return to the Advanced Search page, click the **Cancel** button.
   - To display a preview of the report generated by the search results, click the **Preview** button.

Managing Individual and Compound Queries

Once you submit an individual query, you are returned to the Advanced Search - Build a Query page. Your search is added to the counter next to the Analysis button. The following list describes how to manage your individual and compound queries (if defined):

- Add more individual queries: Click the **Gene Expression Analysis**, **Copy Number Analysis**, or **Clinical Analysis** button on the Advanced Search - Build a Query page.
- Copy, edit, or delete existing individual queries. Find the query listed in the side bar (*Figure 3.9*) and use the **Copy**, **Edit**, and **Delete** buttons.

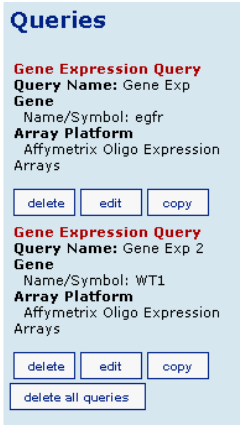


Figure 3.9 Modifying existing queries with the side bar

- Create a compound query. Click the **Finalize Query** button or the **Refine Query** option on the Advanced Search page, and see *Refining a Query*.
- Delete a compound query. Find the compound query listed in the side bar and click the **Delete** button.

  Note: The **Delete All Queries** button deletes all compound *and* individual queries listed in the side bar.

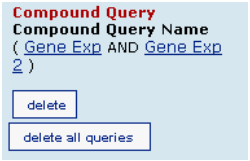


Figure 3.10 Deleting a compound query with the side bar

Refining a Query

The Refine Query page enables you to group multiple queries into a single, compound query. You must validate the query to generate a result.

1. Fill in the Refine Query criteria (*Figure 3.11*).

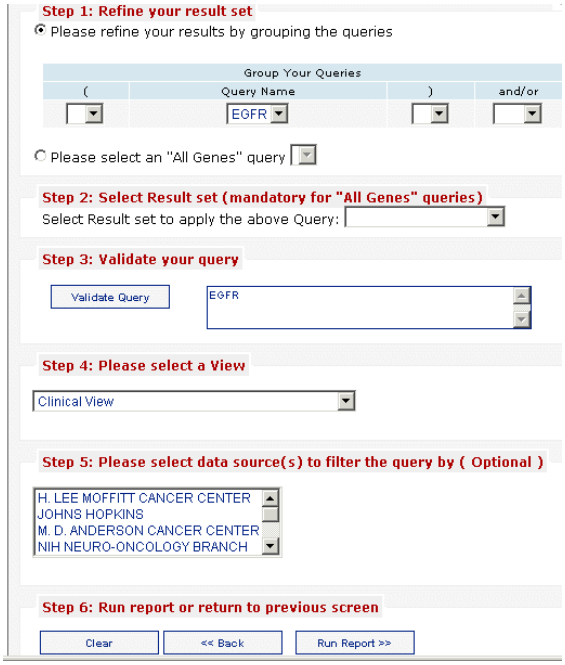


Figure 3.11 Refine Query page

Table 3.6 lists the Refine Query items:

Item Name	*Special Instructions*
Step 1. Refine your result set	You can group the queries to obtain a particular result set, or select all queries. To group queries click **Please refine your results by grouping queries**. • Select the open parentheses, **(**. • Select a **Query Name**. • Select a closing parentheses **)**. • Select an **and/or** operator at the end of a query row to enable the next row where you can select another query of interest. • Repeat for each query name to be grouped. Go to Step 3. OR To select all queries, click **Please select an All Genes query**. The drop-down list appears from which you can choose an All Genes query. Go to Step 2.
Step 2. Select result set (mandatory for "All Genes" queries)	Select a previously saved result set to which to apply these queries. You will not see any result sets if you have not saved a sample set from a previous query, for example from a Clinical report page. The available sets are also listed in the side bar under PatientDID Lists in red type.
Step 3. Validate your query	REQUIRED. Click to validate the syntax of the query is correct.
Step 4. Select a view	Select a report from the drop-down list. The available reports vary based type of queries selected.
Step 5. Select data source(s) to filter the query by (Optional)	Optionally select a datasource to filter the query by the institute providing data. You can select more than one institute. **Note:** The Simple Search function and Preview assigns all the institutes to which you have access.

Table 3.6 Refining Query instructions

To return to the Advanced Search - Build Query page and not save the information, click the **<< Back** button.

2. To generate a report from the defined criteria, click the **Run Report >>** button.

   Note: If the **Run Report** button does not appear, click the **Validate Query** button first to check the query syntax.

CHAPTER

4

HIGH ORDER ANALYSIS

This chapter describes how to use REMBRANDT to run higher order analyses, including class comparisons, hierarchical clustering, and principal component analyses.

Topics in this chapter include:

High Order Analysis Overview

REMBRANDT stores preprocessed gene expression data (filtering and normalization). Click one of the following buttons on the *High Order Analysis* page to further analyze gene expression data.

- Class Comparison Analysis
- Principal Component Analysis (PCA)
- Heiarchical Clustering Analysis

A high order analysis generates results that you can review on the View Results page.

Performing a Class Comparison

To create a High Order Analysis with *Class Comparisons*, follow these steps:

1. The Class Comparison Analysis Form page *(Figure 4.1)* enables you to define the criteria to perform a class comparison.

   Note: Clicking the plus (+) sign in Step 2 expands and displays the Advanced Statistic options.

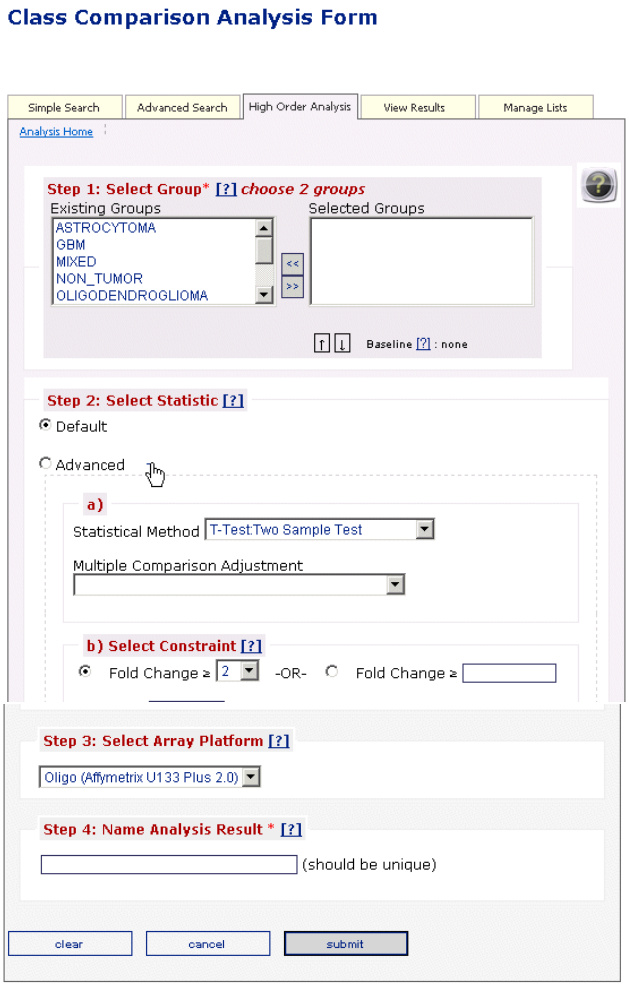


Figure 4.1 Class Comparison Analysis Form page

You are required to complete at least one step for the class comparison. *Table 4.1* lists the available criteria:

Criteria	*Item Name*	*Special Instructions*
Step 1. Select Group	Existing Groups Selected Groups	A *group* is a pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics (See *Managing Lists Overview*). Note that user-defined PatientDID lists appear in red type in the side bar. Select two groups in the **Existing Groups** box and move them to the **Selected Groups** box.
	Baseline	To select a baseline, follow these steps: • Select a group in the **Selected Groups** box. • Use the **Baseline** up or down arrows to move the group to the bottom of the list. • Once you correctly select the baseline, **(baseline)** appears next to your selection.
Step 2. Select Statistic	Default	Select to perform a default statistical analysis.
	Advanced	Select to define additional statistical analysis options.
	+ (-)	Click to access (and close) the advanced options.
	• Statistical Method	Select the appropriate statistical method: • **T-test: Two Sample Test** to identify genes showing statistically significant differences between two samples. • **Wilcoxon Test: Man-Whitney Test** is the non-parametric test analog to the independent two-sample t-test. This test is used in place of a two-sample t-test when the populations being compared are not normal. • **F-test: One Way ANOVA** to identify genes showing statistically significant differences across two or more groups. 1. If there are three or more predefined groups, **F-test: One Way ANOVA** is the default statistical method. 2. When you select the **F-test** option to test a hypothesis of the means of two or more populations, the technique is called the *Analysis of Variance (ANOVA)*. The ANOVA simplifies the F-test, where F-test is the mean square for each main effect and the interaction effect divided by the *within* variance. A one-way ANOVA or single factor ANOVA tests differences between the groups classified only on one independent variable. 3. Using ANOVA instead of multiple t-tests reduces the probability of a type-I error.

Table 4.1 Class Comparison criteria instructions

Criteria	Item Name	Special Instructions
	• Multiple Comparison Adjustment	*Family-wise Error Rate* (FWER): Bonferroni *False Discover Rate* (FDR): *Benjamini-Hochberg*
	• Select constraint	*Future Implementation*
	• p-value	*Future Implementation*
Step 3. Select Array Platform	Select Array Platform	Select the array platform.

Table 4.1 Class Comparison criteria instructions

2. You must enter a title/name for this analysis in the **Name Analysis Result** text box. This name must be unique among all your queries in this session.
3. To submit your criteria and create a Class Comparison report, click the **Submit** button.

Performing a Principal Component Analysis

To create a High Order Analysis with *Principal Component Analysis*, follow these steps:

1. The Principal Component Analysis (PCA) Form page *(Figure 4.2)* enables you to define criteria to perform a PCA. When you access the page, Current Filter Settings display in the side bar. To modify the filter settings, see the following table *Table 4.2*.

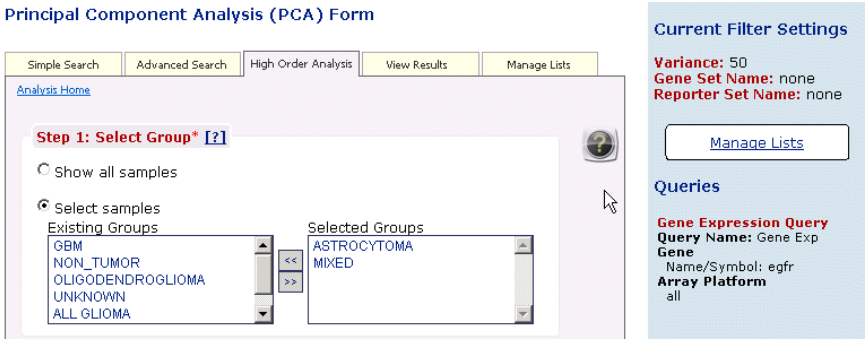


Figure 4.2 Selecting Principal Component Analysis criteria

2. You are required to complete at least one step for the Principal Component analysis. *Table 4.2* lists the available criteria:

Criteria	*Item Name*	*Special Instructions*
Step 1. Select Group	Show all samples	Select to show all samples.
	Select samples	Select to specify the groups to include in the sample. A *group* is a pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics (See *Managing Lists Overview*). Note that user-defined PatientDID lists appear in red type in the side bar.
	• Existing Groups • Selected Groups	Select at least two groups in the **Existing Groups** box and move them to the **Selected Groups** box.
Step 2. Filter Genes/ Reporters	View Filter Settings	To use the default filter settings, continue to Step 3. Current settings display in the side bar.
	+ (-)	Click to access (and close) the advanced options.
	• Constrain reporters by variance (Gene Vector) percentile: %	Enter a percentage which selects the reporters whose variances of the log ratio (or Log2 signals) across all experiments were among the top percentile of variance of all reports identified. For example, 70% chooses reporters with the top 30 (100 - 70) percentile of variance.
	• Use differentially expressed genes	Drop down the list box and select a saved list of differentially expressed genes identified by class comparison. If you have not added a *Gene List* with the REMBRANDT Manage Lists function, **none** appears (see *Adding New Lists*).
	• Use differentially expressed reporters	Drop down the list box and select a saved list of differentially expressed reporters identified by class comparison. If you have not added a reporter list with the Class Comparison report, **none** appears (see *Adding New Lists*).
	• Set These Filters as Default	Click to save the options as default filter settings.
Step 3. Select Array Platform	Select Array Platform	Select the array platform.

Table 4.2 Principal Comparison Analysis criteria instructions

3. You must enter a title/name for this analysis in the **Name Analysis Result** text box. This name must be unique among all your queries in this session.
4. To submit your criteria and create a Principal Comparison Analysis report, click the **Submit** button.

Performing Hierarchical Clustering Analysis

To create a High Order Analysis with *Hierarchical Clustering*, follow these steps:

1. The Hierarchical Clustering Analysis Form *(Figure 4.3)* enables you to perform a clustering. When you access the page, Current Filter Settings display in the side bar. To modify the filter settings, see the following table *Table 4.3*.

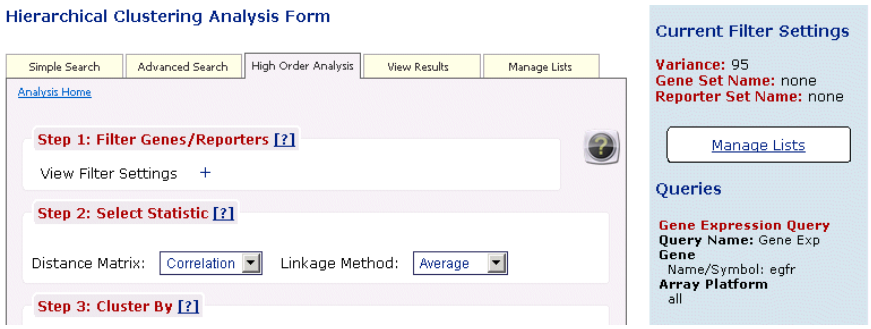


Figure 4.3 Selecting Hierarchical Clustering criteria

2. You are required to enter at least one step for the hierarchical clustering. *Table 4.3* lists the available criteria:

Criteria	***Item Name***	***Special Instructions***
Step 1. Filter Genes/ Reports	View Filter Settings	To use the default filter settings, continue to Step 3. Current settings display in the side bar.
	+ (-)	Click to access (and close) the advanced options.
	• Constrain reporters by variance (Gene Vector) percentile: %	Enter a percentage which selects the reporters whose variances of the log ratio (or Log2 signals) across all experiments were among the top percentile of variance of all reports identified. For example, 70% chooses reporters with the top 30 (100 - 70) percentile of variance.
	• Use differentially expressed genes	Drop down the list box and select a saved list of differentially expressed genes identified by class comparison. If you have not added a *Gene List* with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
	• Use differentially expressed reporters	Drop down the list box and select a saved list of differentially expressed reporters identified by class comparison. If you have not added a reporter list with the REMBRANDT Manage Lists function, **none** appears (see *Managing Lists Overview*).
	• Set These Filters as Default	Click to save the options as default filter settings.

Table 4.3 Hierarchical Clustering criteria instructions

Criteria	*Item Name*	*Special Instructions*
Step 2. Select Statistic	Distance Matrix	Select a distance matrix option: • **Pearson** correlation measures the relative shape of the gene regulations rather than the absolute levels. This is a natural choice, because it is widely used to measure gene correlations. • **Euclidean** distance is the most common distance measure. It measures the absolute level of gene regulation.
	Linkage Method	Select a linkage option to affect the shape of the resulting clusters: • **Average linkage** is the average of all pair-wise distances between members of the two clusters. • **Single linkage** is the minimum distance between two clusters. • **Complete linkage** is the maximum distance between two clusters.
Step 3. Cluster By	Cluster by	Leave the default to cluster on **Samples** or cluster by **Genes**.
Step 4. Select Array	Select Array Platform	Select the array platform.

Table 4.3 Hierarchical Clustering criteria instructions

3. You must enter a title/name for this analysis in the **Name Analysis Result** text box. This name must be unique among all your queries in this session.
4. To submit your criteria and create a Hierarchical Clustering Analysis report, click the **Submit** button.

CHAPTER

5

VIEWING RESULTS

This chapter describes reports and search results that REMBRANDT returns after advanced searches and high order analyses. You can also download the product BRB Array Tools and the static BRB-ArrayTools archive files.

Topics in this chapter include the following:

Results Overview

The View Results page shows a collection of reports previously viewed in a particular user session. This allows you to compare reports by opening them in separate windows. For example, you can compare clinical and gene expression reports with a set of patient samples. You can view Clinical reports , results generated with the Advanced Search function and the High Order Analysis function.

The View Results page also enables you to download BRB Array tools and the static, archive files for use in BRB-ArrayTools.

Clinical Reports

A Clinical report displays patient demographics, therapy, and outcome data. This information is displayed in a single domain, such as gene expression, or in multiple domains that combine queries, such as gene expression, chromosomal aberrations, and clinical areas (*Figure 5.1*). On the Clinical page, you can save samples to a *PatientDID list* stored in the Manage Lists function. To save samples, follow these steps:

Note: When either a gene expression filter and/or a copy number filter is applied with the Advanced Search function, hyperlinks are provided in this report to display the gene expression and/or copy number data for a particular sample.

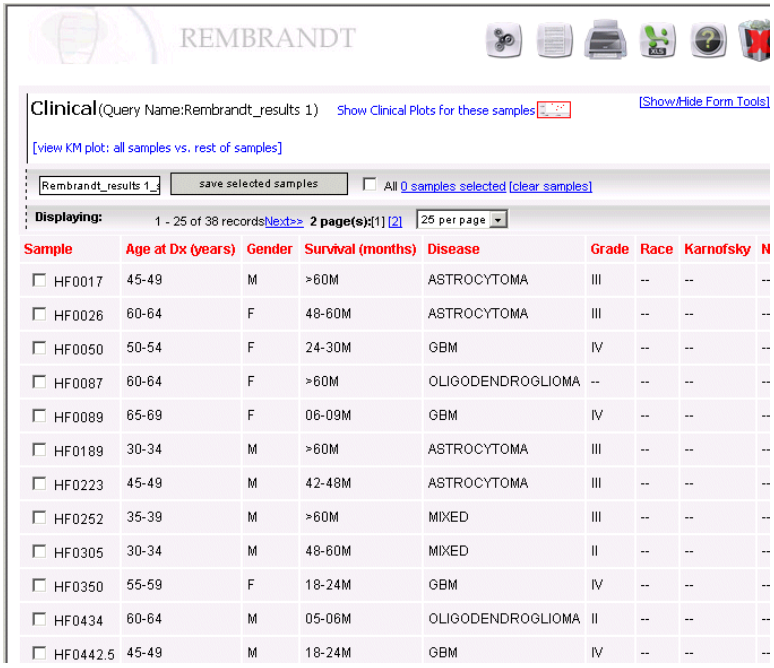


Figure 5.1 Clinical page

1. There are two ways to select samples on the Clinical window:
   - To select an individual sample, select the box in the **Sample** column (*Figure 5.2*).

   Note: Selecting individual items in the list may not be available for all Clinical reports.

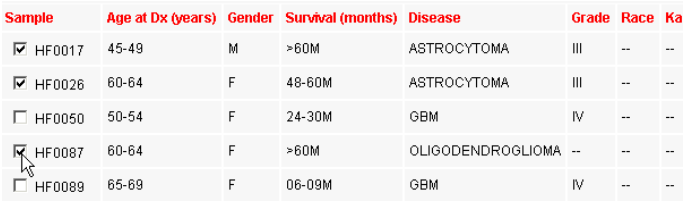


Figure 5.2 Checking the Sample column on the Clinical window

- To select all of the samples, select the **All** box. To display a list of the selected samples, click the **samples selected** link (*Figure 5.3*).

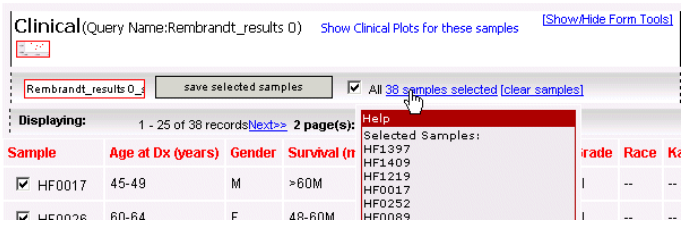


Figure 5.3 Selecting all of the samples on the Clinical window

To clear all of the samples, click the **clear samples** link.

2. To save the selected samples, enter a unique name for the PatientDID list next to **Select Samples**, or maintain the current name (*Figure 5.4*).

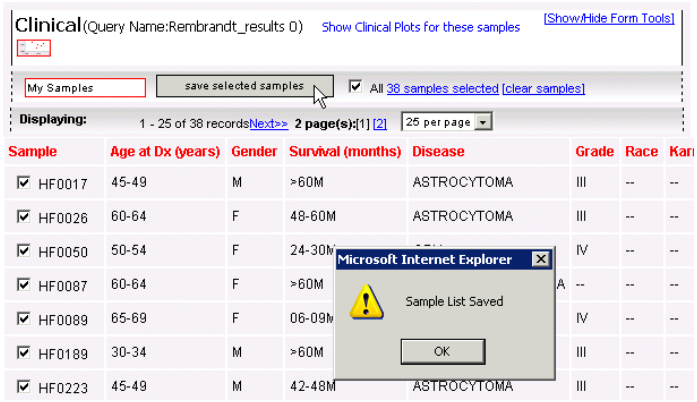


Figure 5.4 Saving Selected Samples on the Clinical page

3. Click the **Save Selected Samples** button. *Sample List Saved* appears.
4. Click the **OK** button.

   Once saved, the sample set is listed in red type under PatientDID List in the side bar.

   Note: The sample set name will also appear on the Refine Query page, in the **Select the Result set to apply the above query** drop-down list. This enables you to add the saved sample set to another query.

5. To show clinical plots for the selected samples, click **Show Clinical Plots for these Samples** at the top of the window (see *Viewing Clinical Plots*).

Viewing Clinical Plots

On certain clinical plots, you can display two kinds of clinical plots:

- **Survival vs Age at Dx** (diagnosis in years) indicates the survival or number of months versus the age at diagnoses in years. The data points are colored by disease type.
- **Karnofsky score** (Neurological assessment) **Vs Age at Dx** (diagnosis in years) indicates the *Karnofsy score* or neurological assessment versus the age at diagnosis in years. The data points are colored by disease type.

To toggle between the different types of plots, click the **SurvivalvsAgeatDx** link or the **KarnofskyscoreVsAgeatDx** link.

To select the samples of interest, follow these steps:

1. Click and drag a rectangle around the samples to save for future use. A red rectangle appears around the samples, and the list of the samples appears on the right-hand side (*Figure 5.5*).

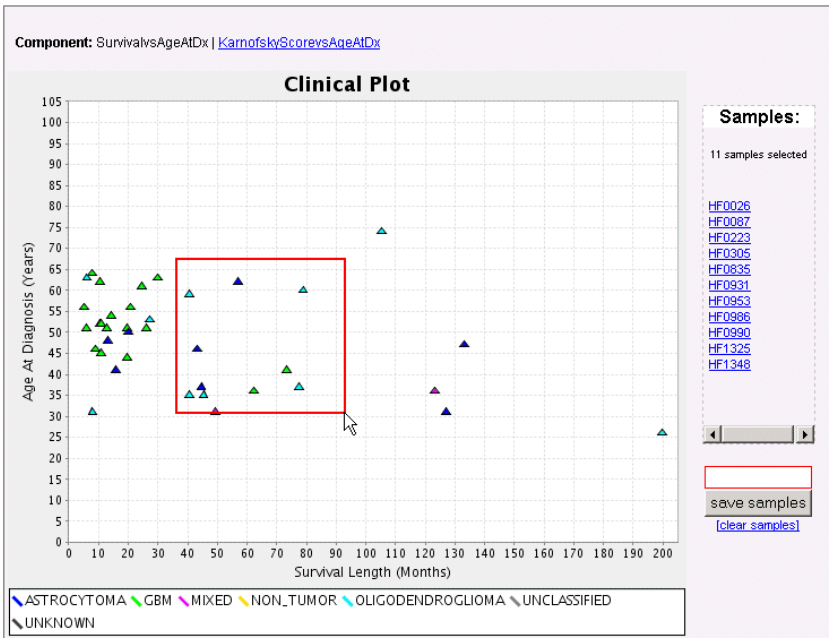


Figure 5.5 Sampling from a Clinical plot

To clear the samples, and select another group, click the **clear samples** link under the **Save Samples** button.

2. To help lasso the points on the plot and identify the location of these points, mouse over a sample name in the list. A yellow circle appears on the plot where the sample is located.
3. To save the selected samples to a file, enter a name for the samples in the text box, and click the **Save Samples** button.

Advanced Search or Query Results

The following Advanced Search reports are generated:

- *Gene Expression Sample Report*
- *Copy Number Sample Report*

View Results (*Figure 5.6*) displays the query name and the output generated for the query. To view the report, click the report name and the file opens in a new window.

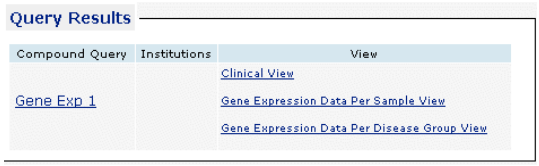


Figure 5.6 Query Results

All Advanced Search options (Gene Expression, Copy Number Data, and Clinical) generate a Clinical report. For more information about Clinical Reports, see *Clinical Reports*.

Gene Expression Sample Report

The Gene Expression Sample report (*Figure 5.7*) displays gene expression ratios (between the tumor sample and the geometric mean of non-tumor samples) for each probeset (or IMAGE clone) for the genes selected in the queries. Each column represents a sample, and the samples are grouped by tumor sub-type. For Affymetrix probesets, the ratio of the absolute expression values of the tumor sample and the geometric mean of the expression value of the non-tumor samples displays. For each IMAGE clone, missing values are handled and the ratio of expression values between the tumor and geometric mean of the non-tumor group is calculated for each sample.

Note: To display a clinical report for all samples, click the **View Clinical Report for All Samples** link. For more information, see *Clinical Reports*.

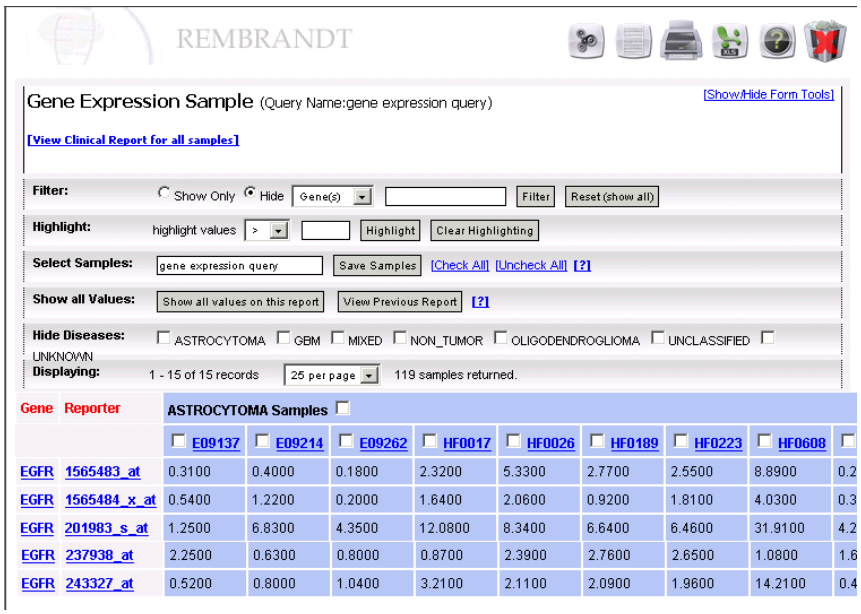


Figure 5.7 Gene Expression Data Per Sample Report

When two queries are included in the results, the results are separated by a double line and the query information is listed at the bottom.

Filtering Results by Gene or Reporter (Filter Toolbar)

To filter a report, follow these steps:

1. From the **Filter** toolbar *(Figure 5.8),* select the filter mode **Show only** or **Hide**.
2. Select **Gene** or **Reporter** from the drop-down list, and enter gene or reporter to be filtered.

   For example, if you click **Show Only**, select **Gene**, and enter `WT1`. Only WT1 samples appear in the list.
3. Click the **Filter** button.

The results are filtered based on your selections *(Figure 5.8)*.

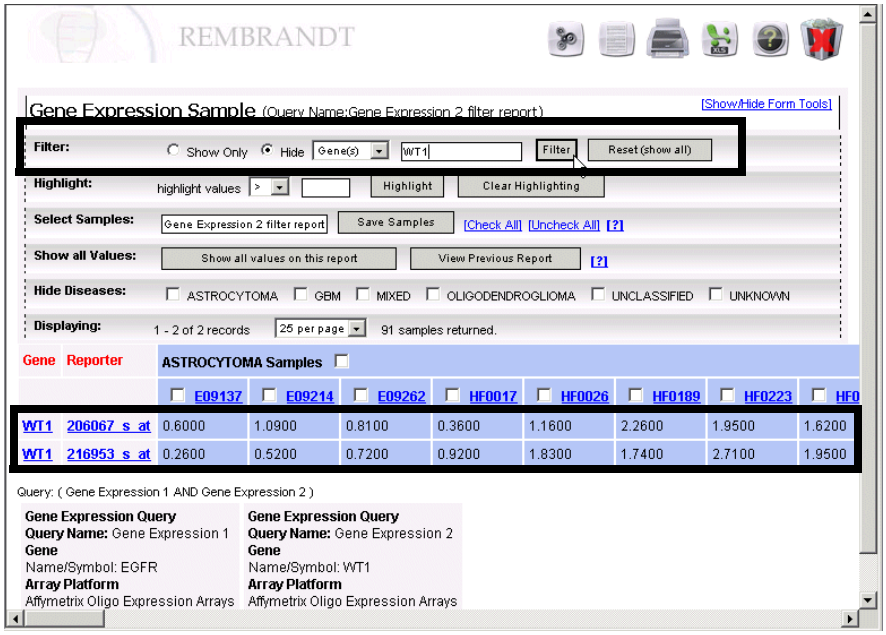


Figure 5.8 Filter toolbar

4. To show all of the samples again, click the **Reset (show all)** button.

   Note: You can use more than one toolbar to limit the samples shown in the results. For example, you can filter the results and then highlight certain filtered results.

Highlighting Results By Value (Highlight Toolbar)

To highlight certain data, follow these steps:

1. From the **Highlight** toolbar, select an operator and a threshold value. *(Figure 5.9).*

   For example, select **< 5** to highlight all values less than 5.

2. Click the **Highlight** button.

The values that meet this criteria are highlighted in yellow *(Figure 5.9).*

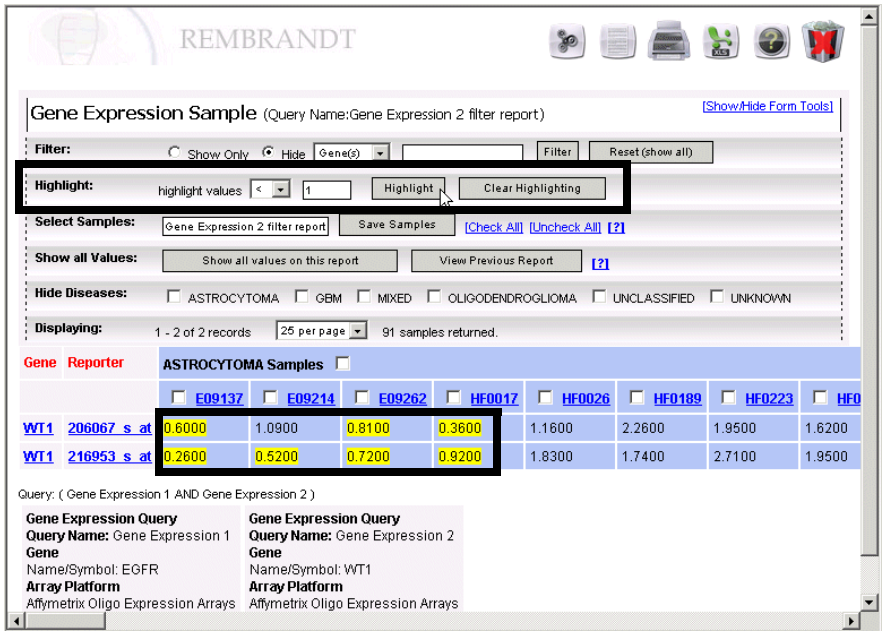


Figure 5.9 Highlight toolbar

3. To restore the report to its original state and clear the highlighting, click the **Clear Highlighting** button.

   Note: You can use more than one toolbar to limit the samples shown in the results. For example, you can filter the results and then highlight certain filtered results.

Selecting and Saving Sample Results (Select Samples Toolbar)

You can save samples to a PatientDID list stored in the Manage Lists function. PatientDID lists enable you to further filter advanced queries. To save samples, follow these steps:

1. On the sample report, there are several ways to select samples.
   - To select all the listed samples on the **Select Samples** toolbar, click the **Check All** link. To deselect all the listed samples, click the **Uncheck All** link (*Figure 5.10* below).

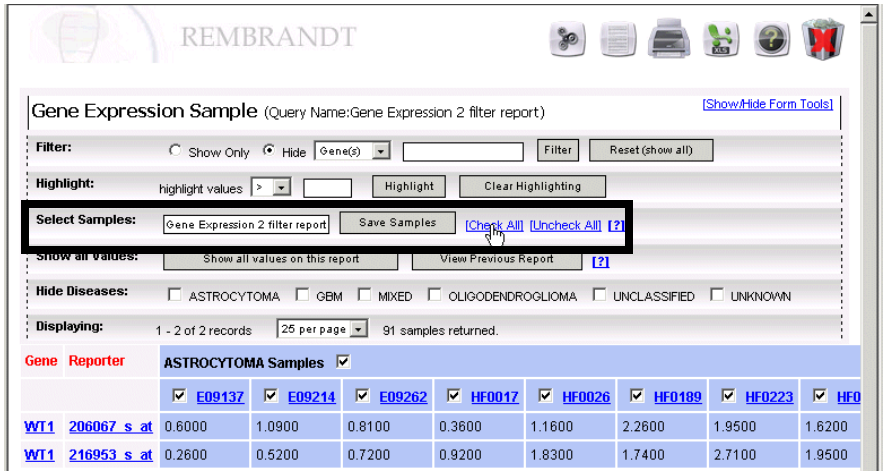


Figure 5.10 Selecting samples from the Select Samples toolbar

   - To select (or deselect) all the samples in a sample group, click the box next to the sample group name, for example, the box next to **ASTROCYTOMA Samples**. All the samples in the group are selected (*Figure 5.11* below).
   - To select (or deselect) an individual sample within a group, click the box in the column next to the sample name.

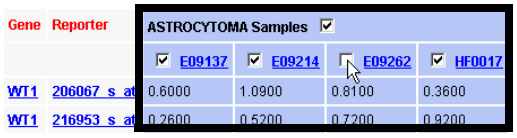


Figure 5.11 Selecting samples from the results

2. To save the selected samples, enter a unique name for the PatientDID list next to **Select Samples**, or maintain the current name (*Figure 5.12*).

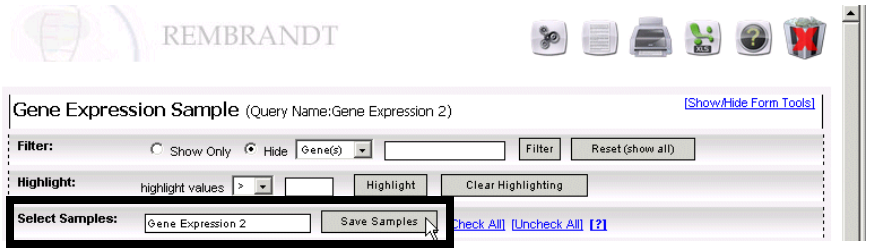


Figure 5.12 Saving the samples

3. Click the **Save Samples** button.

   Once saved, the sample set is listed in red type under PatientDID List in the side bar.

   Note: The sample set name will also appear on the Refine Query page, in the **Select the Result set to apply the above query** drop-down list. This enables you to add the saved sample set to another query.

Differentiating Data (Show All Values Toolbar)

To differentiate between missing values in the array and data that did not meet your search criteria, follow these steps:

1. Click **Show All Values on this Report** on the **Show all Values** toolbar (*Figure 5.13*).

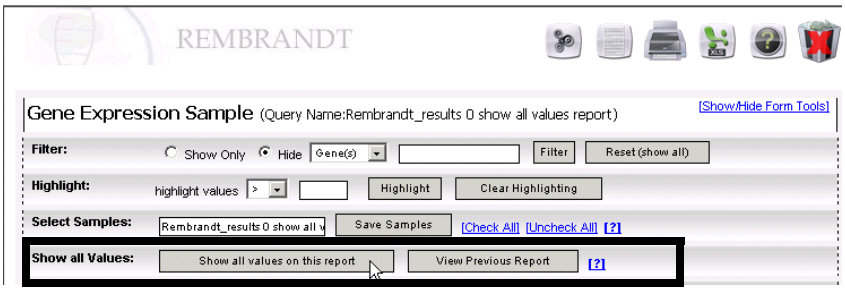


Figure 5.13 Show All Values toolbar

   The samples that did not meet your criteria, appear in gray. A value of **Null** indicates a missing value for that reporter.

2. To display the previous report before you clicked the **Show All Values on this Report** button, click the **View Previous Report** button.

Removing Columns (Hide Diseases toolbar)

To remove a disease from the report, select the check box for the disease in the **Hide Diseases** toolbar *(Figure 5.14)*. The checked disease is NOT included in the results.

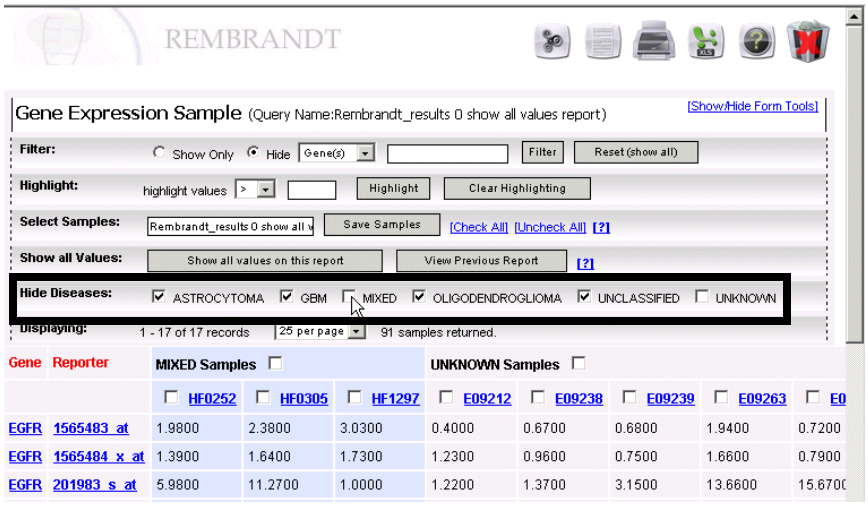


Figure 5.14 Hide Diseases toolbar

Showing Additional Information

When results are listed in a report, row or column items may appear as links. These links may be for genes, reporters, cytobands, or sample names. Click the link to display additional information about the item.

For example, to display more information about a gene, click the gene symbol link (*Figure 5.15*).

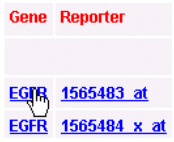


Figure 5.15 The Gene column

The Cancer Genome Anatomy Project (CGAP) browser opens.

Copy Number Sample Report

The Copy Number Data Per Sample report displays the copy number data from Affymetrix 100K SNP arrays. The CHP files from the Affymetrix Gene Chip Operating System were processed using the Affymetrix GDAS (GeneChip® DNA Analysis Software). Copy number data was collected for each mapping SNP reporter on the Chip, for all the tumor samples. Each column represents a sample, and the samples are grouped based on the tumor sub-type (*Figure 5.16*).

Note: To display a clinical report for all samples, click the **View Clinical Report for All Samples** link. For more information, see *Clinical Reports*.

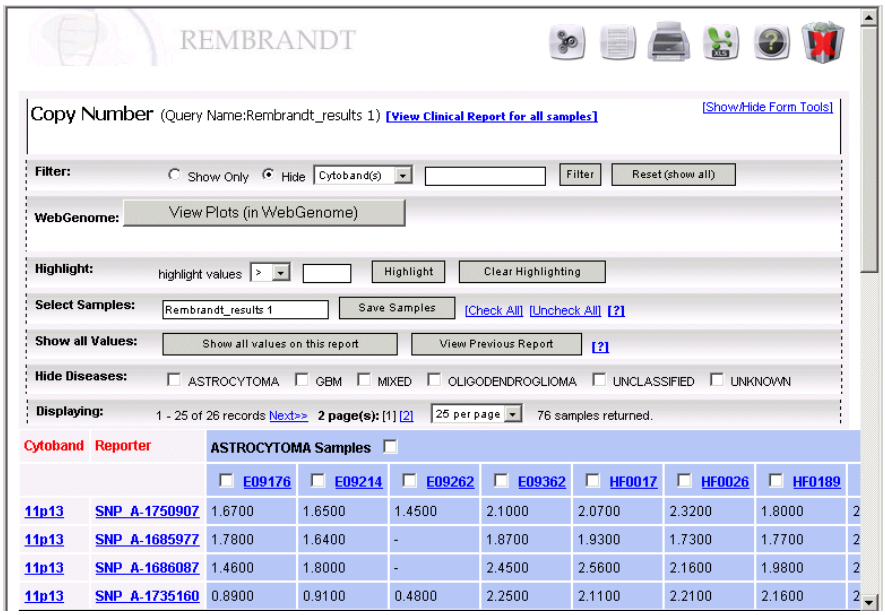


Figure 5.16 Copy Number Data for Sample report

Displaying Copy Number Sample Data in webGenome

You can display the Copy Number for Data Sample report data as a graphic in the application webGenome. webGenome is a web-based application for plotting and visualizing microarray data, especially comparative genome hybridization (CGH) data. In webGenome, you can select microarray datasets from public, as well as private, database areas. You can also perform preliminary filtering, smoothing, and normalization of data prior to plotting. The system supports several types of plots:

- Scatter Plots plot DNA copy number measurements across the genome, chromosome, or chromosomal interval.
- Ideogram Plots show chromosomal amplifications and deletions in relation to cytogenetic chromosome ideograms.

To plot the REMBRANDT copy number data in webGenome, follow these steps:

1. Click the **View Plots in webGenome** button (*Figure 5.17*).

   REMBRANDT passes copy number data for only those regions selected in the query. Although only the selected region data is passed to webGenome, all the reporters associated with the genome are passed.

   Note: If there are more than 50 results, the following message appears below the **View Plots in webGenome** button: *Note: Your query request contains more than 50 samples. The resultset may cause WebGenome request to timeout.*

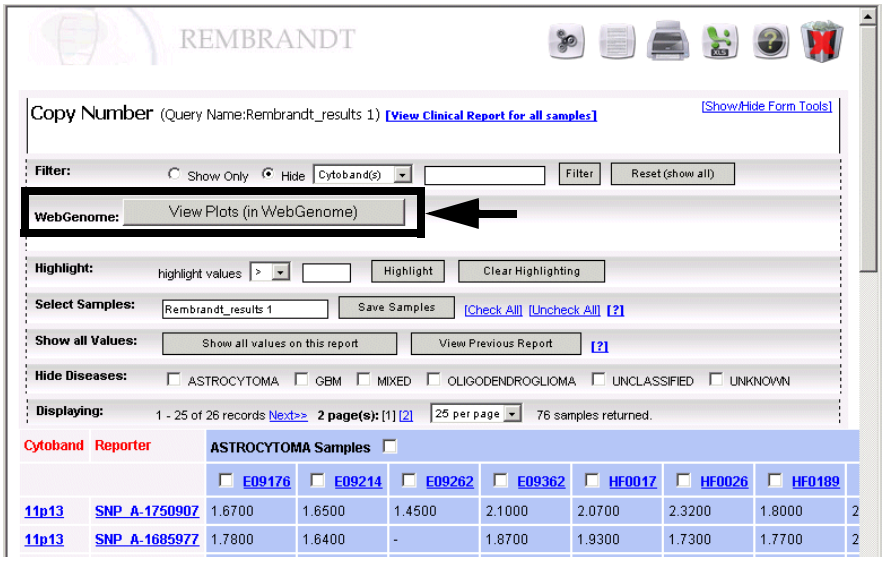


Figure 5.17 Copy Number Report

2. webGenome opens a new window plotting the REMBRANDT Copy Number data.

   Note: Functions performed in webGenome do not affect data in REMBRANDT.

3. Once you review the copy number plot, you can display LOH (Loss of Heterozygosity) data using webGenome. For more information, see webGenome online help.

High Order Analysis Results

The following High Order Analysis reports are generated:

- *Class Comparison Report*
- *Principal Component Analysis Plot*
- *Hierarchical Clustering Report*

View Results (*Figure 5.6*) displays the query name and lists the output generated for the query.

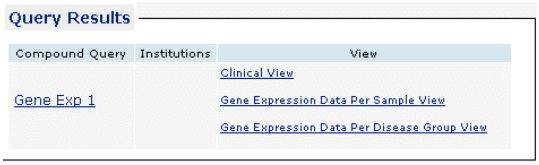


Figure 5.18 Query Results

Class Comparison Report

The Class Comparison report (*Figure 5.19*) displays group average, fold change, and *p*-value based on the Advanced Search parameters that you selected. The output varies based on the statistical method chosen.

For **T-test** or **Wilcoxon** Statistical Method analysis (*Figure 5.19*), the Class Comparison report is as follows.

- The report displays the group average, where the numerator is the mean of log(base 2) expression signals from the samples in the first group. The denominator is the mean of log(base 2) expression signals from the samples in the second group.
- The fold change for the reporter between the selected groups appears along with *p*-value.
- Gene symbol annotations appear for each reporter. To obtain extensive annotations, click the Excel icon on the upper right-hand corner of the report.

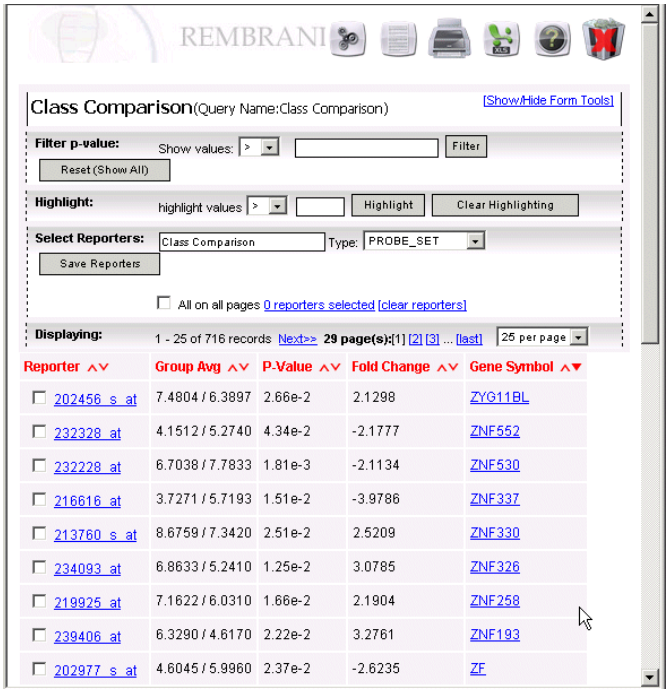


Figure 5.19 Class Comparison report

For an **F-test** Statistical Method analysis (*Figure 5.20*), the difference is that in the results there is a group average displayed for each group.

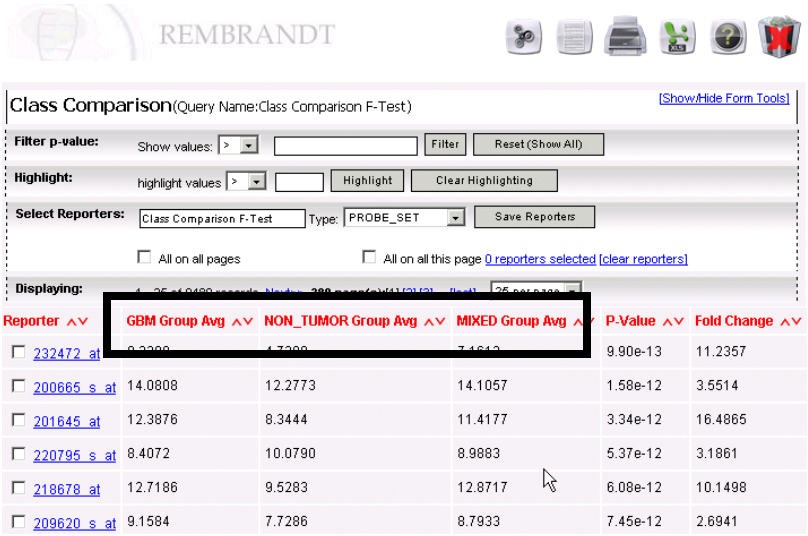


Figure 5.20 Class Comparison report - F-test

Filtering a *p*-value (Filter p-value Toolbar)

To filter a *p*-value, follow these steps:

1. From the **Filter p-value** toolbar, select an operator from **Show Values** and enter a threshold value.
2. Click the **Filter** button.

   The results are filtered based on your selections *(Figure 5.21)*.

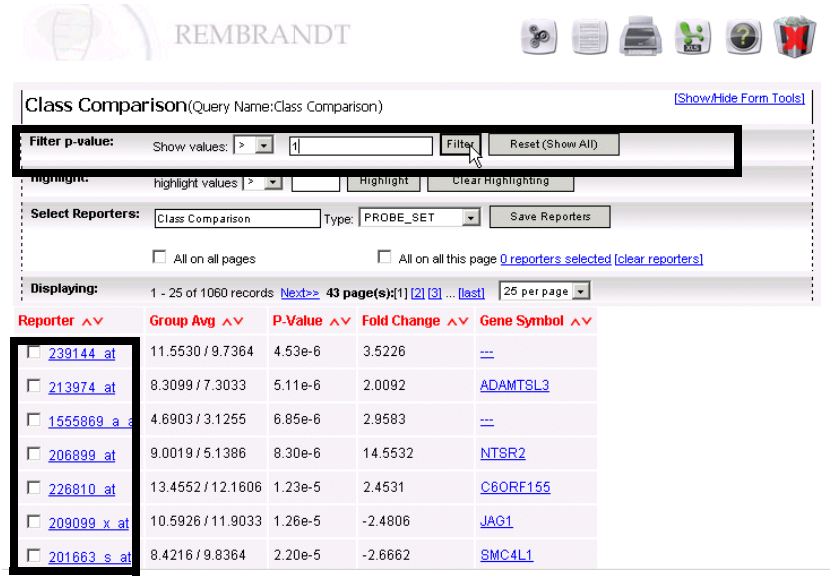


Figure 5.21 Filter toolbar

3. To show all of the samples again, click the **Reset (show all)** button.

Note: You can use more than one toolbar to limit the samples shown in the results. For example, you can filter the results and then highlight certain filtered results.

Selecting and Saving Reporters (Select Reporters toolbar)

You can save reporters to a Reporter list stored in the Manage Lists function. Reporter lists enable you to further filter advanced queries. To save reporters, follow these steps (*Figure 5.22*):

1. There are several ways to select reporters in the result list:
   - From the **Select Reporters** toolbar, select a reporter type from the **Type** drop-down list. The reporters with the selected criteria are displayed.
   - To select all of the results, click the **All on all pages** box.
   - To select one row of results at a time, click the box on the left side of the result row.

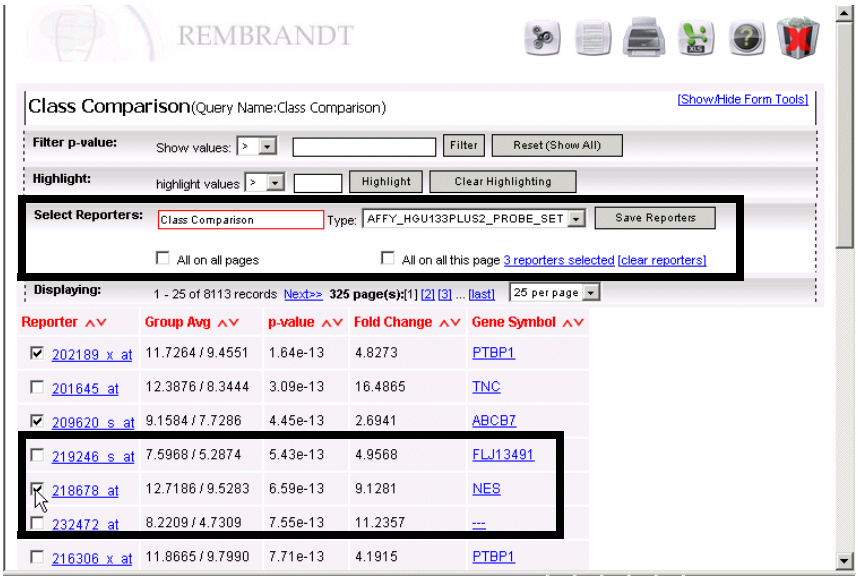


Figure 5.22 Selecting Reporters instructions

To clear the selected reporters, click the **clear reporters** link.

2. To save the selected reporters, enter a unique name for the Reporters list next to **Select Reporters**, or maintain the current name. The name varies based on the type of Statistical Method selected for the analysis.
3. Click the **Save Reporters** button.

   Once saved, the Reporters list appears in red type under Reporter Lists in the side bar.
4. Click the **OK** button.

Note: You can use more than one toolbar to limit the samples shown in the results. For example, you can filter the results and then highlight certain filtered results.

Resorting Column Results

To sort a column in a report, follow these steps:

1. If a report column has red triangles pointing up and down next to the name, you can sort a column of numeric or alphabetical values (*Figure 5.23*).

Figure 5.23 Sorting column results

2. To sort a column in ascending order, select the red triangle pointing up. To sort a column in descending order, select the red triangle pointing down.

Principal Component Analysis Plot

The Principal Component Analysis plot (*Figure 5.24*) is a two-dimensional graph which plots the various principal components from the analyses. The following list desribes the different areas of the plot:

- To review a three-dimensional version of the color by disease PCA, click the **View 3D Applet - Color by Disease** link at the top of the page (see *Viewing a Three-dimensional Principal Component Analysis (PCA)*).
- The three tabs at the top of the page enable you to display **PC1 versus PC2**, **PC1 versus PC3**, or **PC2 versus PC3**.
- Each point on the graph represents a sample. By default, the samples are colored by **Disease**. To color by gender, click the **Gender** link or **Remove Colors and Shapes**.
- At the bottom of the graph, there is a legend defining how the different shapes on the graph indicate different survival ranges for patients.
- The Samples area enables you to select, review, and save samples in the plot (see *Viewing Clinical Plots*). You can also display clinical data for the PCA analysis, by clicking the **view clinical data** link.

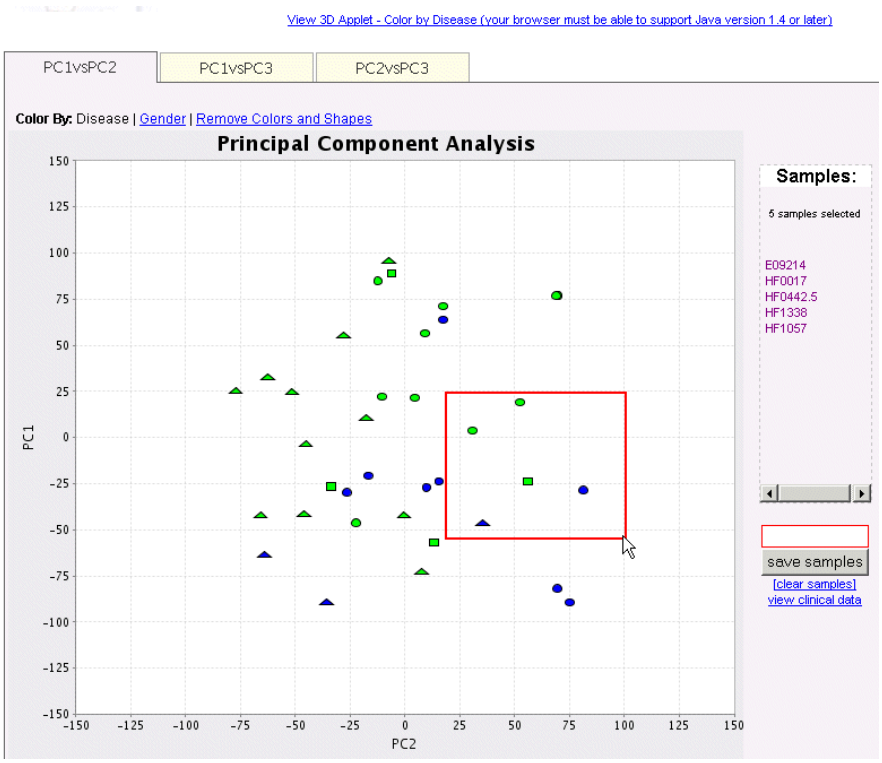


Figure 5.24 Principal Component Analysis report

Viewing a Three-dimensional Principal Component Analysis (PCA)

The three-dimensional Principal Component Analysis enables you to view PCA plot data in three dimensions. For more information about the PCA plot, see *Principal Component Analysis Plot*. The applet supports Java Plugin versions 1.4.x and 1.5.x. For assistance with the plugin, refer to the links at the top of the page.

Table 5.1 provides general tasks you can perform with the three-dimensional PCA view:

Task	***Instructions***
Highlight all points in a data set (in black)	Click on a legend label or click on a single point in the plot.
Rename a dataset	Double-click on a legend label and enter a new name.
Show the values for a single point	Right-click on a point in the plot. Right-click again to remove the values.
Open the points of a dataset into a spreadsheet view	Double-click on a point on the plot.

Table 5.1 Tasks for the Three-dimensional PCA view

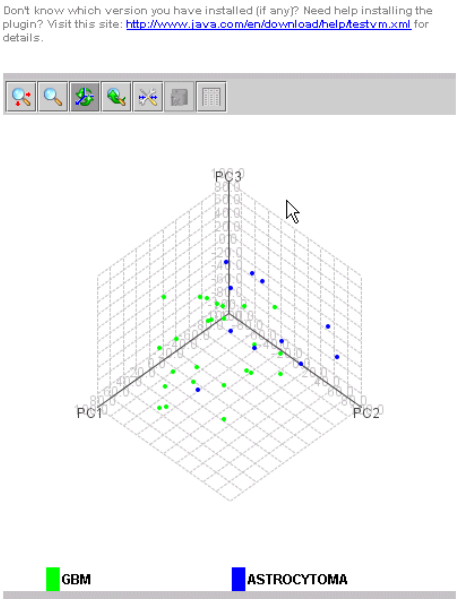


Figure 5.25 Three-dimensional PCA

Table 5.2 describes the icons:

Icon	***Special Instructions***
	Moves the plot around the page. Click the button, and click and drag the plot. To return to the original 3-D view, click  .
	Magnifies a selected area on the plot. Click the button, and click and drag the box around an area on the plot. The selected area is magnified. To move within the area, click  . To return to the original 3-D view, click  .
	Moves and spins the plot, so you can display clusters. Default tool selected when the page displays. Click the button, and click and drag the plot to rotate the plot in the 3-D space.
	Returns to the original 3-D view. Click the button to cancel a zoom or return to the original axes or placement on the page.
	Not applicable to a Principal Component Analysis plot.
	Not available.

Table 5.2 Three-dimensional PCA instructions

Icon	Special Instructions
	Displays a spreadsheet view of the data with each dataset in a separate tab. Click the button, and the spreadsheet view appears. From the spreadsheet view, you can perform the following tasks: • To remove a dataset, uncheck  . All data from the set is removed from the main plot. • To copy the data to the clipboard, click  . • To change the color of a data group, click the color box next to  .

Table 5.2 Three-dimensional PCA instructions

Hierarchical Clustering Report

The Hierarchical Clustering report (*Figure 5.26*) displays the dendrogram from the hierarchical clustering analysis. To display the image at full resolution, click on **full size** at the top left-hand corner of the graph. Based on the cluster parameter that you select, the report displays either sample or reporter annotations beneath the dendrogram.

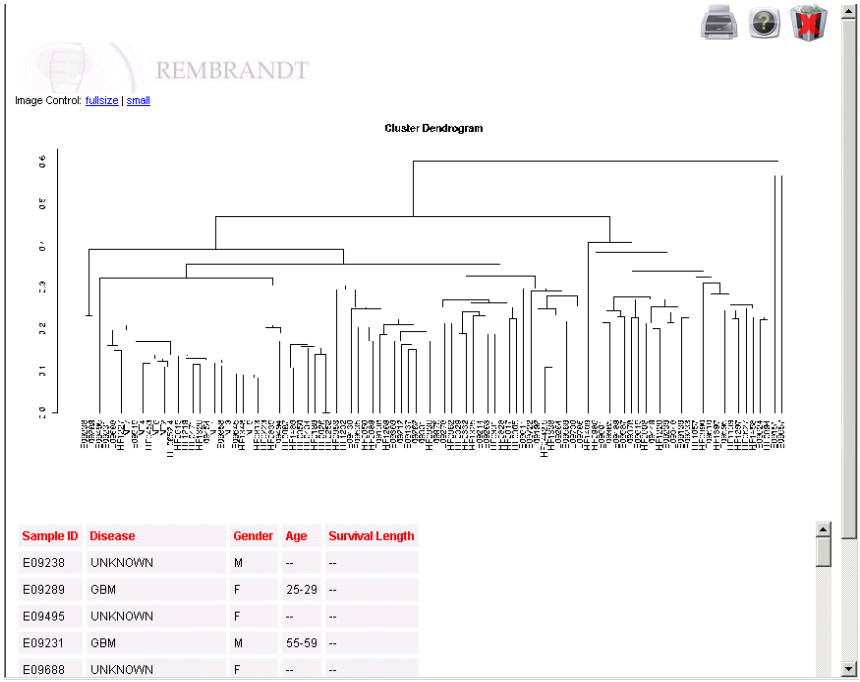


Figure 5.26 Hierarchical Clustering report

To perform a Hierarchical Clustering analysis, see *Performing Hierarchical Clustering Analysis*.

Downloading BRB Array Tools and Files

REMBRANDT enables you to analyze data using BRB-ArrayTools created by the Briometric Research Branch of the National Cancer Institute. BRB-ArrayTools is a comprehensive state-of-the-art statistical analysis system for microarray gene expression data. It is free for non-commercial purposes and can be licensed for commercial purposes from the NIH. BRB-ArrayTools installs as an Excel plug-in using a self-installer.

The Bulk Downloads drop-down list box enables you to download the static BRB archive files which enable you to obtain the appropriate files based on your *user role* as follows:

- *Public users* receive the Public dataset file (`Rembrandt-Brain-Public-Project.zip`)
- *Institution users* receive the institution data and public dataset files.
- *Super users* receive all dataset files.

To download BRB-Array Tools and the appropriate files, follow these steps.

1. To download BRB-ArrayTools, click the **BRB_Array Tools** link.

Figure 5.27 Downloading BRB Array Tools

   The web site appears. Download the appropriate version of the product, and follow the prompts.

2. Once you have downloaded and installed BRB-Array Tools, select the files to download to analyze a dataset with BRB-Array Tools.
3. Click the **Download** button.
4. Unzip the REMBRANDT static BRB archive file(s).
5. Open the project worksheet in Excel on a Microsoft Windows PC.

CHAPTER

6

MANAGING LISTS

This chapter describes how to manage lists by editing existing lists, adding new lists or creating new custom lists from existing lists.

Topics in this chapter include:

Managing Lists Overview

The REMBRANDT Manage Lists function centralizes all activities pertaining to the creation and management of user-defined, as well as study-defined, *PatientDID lists*, *Gene lists,* and *Reporter lists*. With these lists, you can further refine queries or facilitate analysis. When working with lists on the Manage Lists page, you can minimize the number of lists displayed by clicking the PatientDID Lists, Gene Lists, and Reporters Lists heading.

Note: To add a saved gene or IMAGE clone list to an advanced gene expression search. You can also add a saved Gene List or probeset list to an advanced copy number search.

Adding New Lists

Table 6.1 lists how to add each list type to REMBRANDT.

List Type	How to Add a List
PatientDID List	REMBRANDT provides a collection of default *PatientDID Lists*. To create a new PatientDID list, see the following: • *Combining Existing Lists to Create a New List* on page 78 • *Uploading a List* on page 80 • *Manually Entering a List* on page 80 • Save patients from a Clinical report. See *Clinical Reports* on page 56. • Save patients from a Gene Expression Sample report. See *Selecting and Saving Sample Results (Select Samples Toolbar)* on page 63.
Gene List	To add a *Gene List* see the following: • *Combining Existing Lists to Create a New List* on page 78 • *Uploading a List* on page 80 • *Manually Entering a List* on page 80
Reporter List	To add a *Reporter List*, see the following: • *Combining Existing Lists to Create a New List* on page 78 • *Uploading a List* on page 80 • *Manually Entering a List* on page 80 • Save Reporters from a Class Comparison report. See *Selecting and Saving Reporters (Select Reporters toolbar)* on page 71.

Table 6.1 Adding REMBRANDT lists

Combining Existing Lists to Create a New List

You may create new lists from existing lists. To create a custom list from existing lists, follow these steps:

1. At the top of the Manage List page, click on the type of list you would like to view (**PatientDID List**, **Gene List**, **Reporter List**). The names for the lists appear.

2. Click the box next to the list name(s) to be combined to create a new list.

   Note: You cannot select more than two lists to use the **Difference** option.

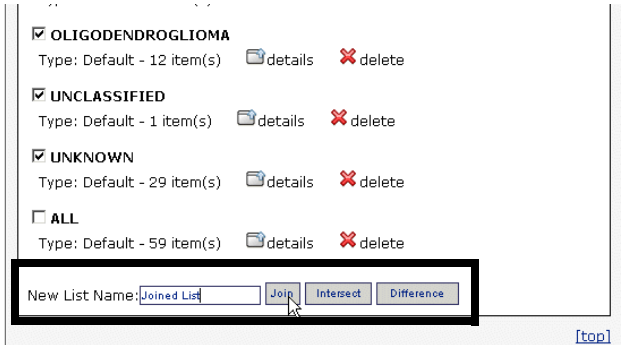


Figure 6.1 Combining existing lists

3. Enter a unique name for the new list you are creating, and then click the appropriate button:

   - **Join** combines two or more lists into one new list.
   - **Intersect** creates a new list from only the items that appear on more than one selected list.
   - **Difference** creates up to two lists each comprising items that appeared in one of the selected lists. For example, if you select **Astrocytoma** and **GBM**, the new lists are “Astrocytoma-GBM” comprising the items that appeared in the Astrocytoma list only and “GBM-Astrocytoma” comprising the items appearing in the GBM list only (*Figure 6.2*).

   The new list appears on the Manage Lists page and in the side bar in red (*Figure 6.2*).

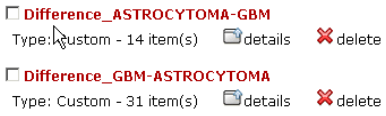


Figure 6.2 New Difference lists

Uploading a List

You may add a new list type by *uploading* a list from your computer. To upload a list, follow these steps:

1. At the top of the Manage List page, click **Add List**.

   The **Upload List or Manually type List** block appears.

2. Click **Upload List** at the top of the box (*Figure 6.3*).

Figure 6.3 Uploading a list

3. From the **Choose the list type** drop-down list box, select the list to be uploaded.
4. Click the **Browse** button beside the **Upload file** box. Navigate to and select the file on your computer that you would like to upload.
5. Enter a unique name for the list, and then click the **Add List** button.

   The name of the list appears on the Manage Lists page or in the side bar under the appropriate list type.

Manually Entering a List

You may create a new list type by *manually typing or entering* a list. To enter a list manually, follow these steps:

1. At the top of the Manage List page, click **Add List**.

   The **Upload List or Manually type List** block appears.

   1. Click **Manually Type List** at the top of the box (*Figure 6.4*).

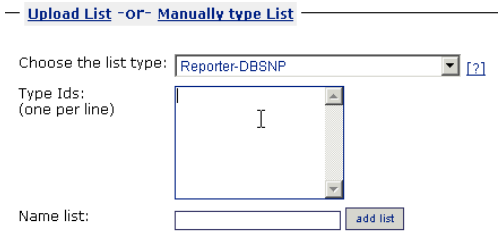


Figure 6.4 Manually typing a list

   2. From the **Choose the list type** drop-down list box, select the type of list to be entered.

Table 6.2 lists examples of correctly formatted codes for each list type.

List Type	***Correctly Formatted Examples***
PatientDID	`CB160831` `K03193`
Gene-GENBANK_ACCESSION_NUMBER	`AF125253` `S75264`
Gene-GENESYMBOL	`BPIL2` `IVL`

Gene-LOCUS_LINK	`10`
	`100`
	`10017`

Reporter-AFFY_GHU133PLUS2_PROBE_SET	`1007_s_at` `1053_at`
Reporter-IMAGE_CLONE	`IMAGE:1407831` `IMAGE:143995`
Reporter-DBSNP	`rs1000015` `rs1000025`
Reporter-AFFY_100K_SNP_PROBE_SET	`SNP_A-1708471` `SNP_A-1655302`

Table 6.2 List type code formats

3. In the **Type Ids** box, enter items into the text block by typing them one to a line.
4. Enter a unique name for the list, and then click the **Add List** button.

   The new list name appears under the appropriate category on the Manage Lists page and in the side bar in red.
5. To display the values in the list, click **Details**.

   Note: If the format of the values entered in the **Type Ids** box was not correct, you must **Delete** the list and start again (*Figure 6.5*).

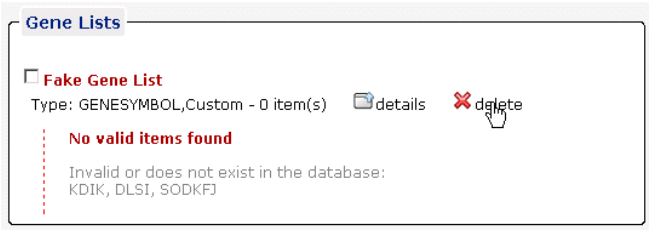


Figure 6.5 Invalid list

Viewing the Data Items in a List

To view the individual data items on a list, follow these steps:

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID Lists**, **Gene Lists**, **Reporter Lists**).
2. Find a list to be viewed, and click the **details** icon to display all of the items in the list (*Figure 6.6*).

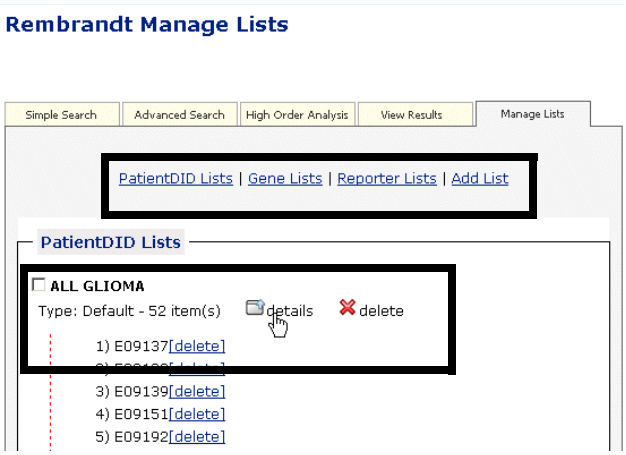


Figure 6.6 List types and Details

Note: The side bar displays each list type and the associated lists. You can mouse-over a list and display the data items.

To export the list, see *Exporting a List* on page 83.

Removing Data Items

You may delete items from an existing list, then view the new list or save the list on your computer. Follow these steps:

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List**, **Gene List**, **Reporter List**).
2. Find the list you want to change, and click on the box next to the list name.
3. Click the **details** icon to display all the items in the selected list.
4. Click the **delete** link beside the item you want to delete.The item is removed from the list (*Figure 6.7*).

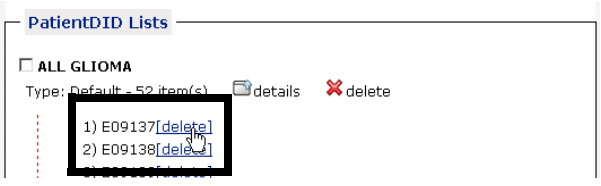


Figure 6.7 Deleting data items

Once you remove the items, you can view the new list or export the list to your computer. See *Exporting a List* on page 83.

Deleting an Entire List

To delete one or more lists from a list type, follow these steps:

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List**, **Gene List**, **Reporter List**).
2. Find the list you want to delete, and click the box next to the list name. Click more than one box to select multiple lists for deletion.
3. To delete the selected lists, click an **x delete** icon. The selected categories are removed (*Figure 6.8*).

Figure 6.8 Deleting an entire list

The list(s) are deleted.

Exporting a List

To export a list, follow these steps:

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List**, **Gene List**, **Reporter List**).
2. Find a list to be exported, and click the **details** icon to display all of the items in the list.

   Note: The side bar displays each list type and the associated lists. You can mouse-over a list and display the data items.
3. Click the **export list** link (*Figure 6.9*).

Figure 6.9 Exporting a list

The list is exported to a spreadsheet on your computer.

APPENDIX A
GLOSSARY

Acronyms, objects, tools and other terms referred to in the chapters or appendixes of this *REMBRANDT 1.5.1 User's Guide* are described in this glossary.

Term	***Definition***
allele	Mutually exclusive alternative forms of the same gene occupying the same locus on homologous chromosomes, differing in DNA sequence and governing the same biochemical and developmental process.
Astrocytic tumors: Astrocytoma	Neoplasms of the brain and spinal cord derived from glial cells. Also called an astrocytoma.
Benjamini-Hochberg Multiple Testing Correction	The concept of False Discovery Rate (FDR) was introduced in multiple testing by Benjamini and Hochberg (1995).
CCR	Center of Cancer Research
CCR-NOB	CCR Neuro-Oncology Branch
CGAP	Cancer Genome Anatomy Project
Class Comparison	Differential gene expression across the tumor types will be evaluated by calculating the typical *t*-statistic for each reporter. Both parametric and non-parametric *p*-value will be computed.
False Discovery Rate (FDR)	The expected proportion of Type I errors among rejected hypotheses in simultaneous testing of multiple null hypotheses.

Table A.1 Glossary of REMBRANDT terms

Term	***Definition***
Family-wise Error Rate (FWER)	Denotes the probability of having at least one false significant test result within the set of tested hypotheses.
Gene List	A pre-defined or user-defined list in REMBRANDT comprising genes with a set of characteristics. Used to filter a query.
Gene Ontology (GO) Classification	A collaborative effort to address the need for consistent descriptions of gene products in different databases. The goal of the Gene Ontology project is to produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing.
Glioblastoma	Malignant form of astrocytoma histologically characterized by pleomorphism of cells, nuclear atypia, microhemorrhage, and necrosis.
Gliomas	Any of the largest group of primary tumors of the brain, composed of malignant glial cells. Kinds of gliomas are astrocytoma, ependymoma, glioblastoma multiforme, medulloblastoma, and oligodendroglioma.
Hierarchical Clustering	A statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left.
High Order Analysis	After data preprocessing (filtering and normalization), further statistical analysis of gene expression data are performed.
HUGO [gene symbol] [Human Genome Organisation]	HUGO is an international organization of scientists involved in human genetics. Established in 1989 by a collection of the world's leading human geneticists, the primary ethos of the Human Genome Organisation is to promote and sustain international collaboration in the field of human genetics.
Kaplan-Maier	The Kaplan Maier method is used for survival analysis. Kaplan-Maier curves are used to estimate survival probability as a function of time, and survival differences are analyzed by the log-rank test.

Table A.1 Glossary of REMBRANDT terms

Term	***Definition***
Karnofsky Performance Status	A standard way of measuring the ability of cancer patients to perform ordinary tasks. The scores range from 0 to 100, with a higher score indicating a better ability to carry out daily activities. KPS may be used to determine a patient's prognosis, to measure changes in functioning, or to decide if a patient could be included in a clinical trial.
Lansky Play-Performance Status	The play-performance scale for children is a parent-rated instrument which records usual play activity as the index of performance. It is similar to the Karnofsky Performance Scale for adults.
Mann-Whitney Test	A nonparametric test (distribution-free) used to compare two independent groups of sampled data. Unlike the parametric *t*-test, this non-parametric makes no assumptions about the distribution of the data (e.g., normality).
Multiple Comparison Adjustment	Since tens of thousands of genes are compared, many genes can be false positives. However, genes are not all independent and genes in the same pathway could have similar *t*-statistics or *p*-values. Multiple-comparison adjusted *p*-values have been proposed to handle the multiple comparison issues in the context of microarray data.
NCIA	National Cancer Imaging Archive
NCI	National Cancer Institute
NCICB	National Cancer Institute Center for Bioinformatics
NINDS	National Institute of Neurological Disorders and Stroke
Oligodendroglial tumor: Oligodendroglioma	Rare, slow-growing tumor that grows in the oligodendrocytes (brain cells that provide support and nourishment for nerve cells). Also called an oligodendroglioma.
PatientDID List	A pre-defined or user-defined list in REMBRANDT comprising patients with a set of characteristics. Used to filter a query.

Table A.1 Glossary of REMBRANDT terms

Term	***Definition***
Pleomorphic Xanthoastrocytoma	Astrocytic tumor with a relatively favorable prognosis (WHO grade II) and is typically encountered in children and young adults. It has a superficial location in the cerebral hemispheres and involvement of the meninges.
Principal Component Analysis	Commonly used in microarray research as a tool. It is designed to capture the variance in a dataset in terms of principle components. In effect, one is trying to reduce the dimensionality of the data to summarize the most important, for example defining parts, while simultaneously filtering out noise.
Protoplasmic [astrocytoma]	Rare variant of Diffuse Astrocytoma. It is predominantly composed of neoplastic astrocytes showing a small cell body with few, flaccid processes with a low content of glial filaments and scant GFAP expression.
Reporter List	A pre-defined or user-defined list in REMBRANDT comprising reporters with a set of characteristics. Used to filter a query.
SNP	Single nucleotide polymorphisms or SNPs (pronounced "snips") are DNA sequence variations that occur when a single nucleotide (A,T,C or G) in the genome sequence is altered.
Wilcoxin Test	Nonparametric statistics for testing hypotheses about whether two samples differ.

Table A.1 Glossary of REMBRANDT terms

INDEX

D

E

F

G

H

I

V

W