REMBRANDT
USER'S GUIDE

Version 1.5

Center for Bioinformatics

October 13, 2006

Credits and Resources

National Cancer Imaging Archive (NCIA) Development and Management Teams		
Development	***Documentation***	***Product and Program Management***
Subha Madhavan	Subha Madhavan	Subha Madhavan
Alex Jiang	Neysa Narena	
Kevin Rosso	Jill Hadfield	
Ryan Landy	Ying Long	
Himanso Sahni	Huaitian Liu	
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		
[1] Science Application International Corporation (SAIC)	[2] National Cancer Institute Center for Bioinformatics (NCICB)	[3] National Cancer Institute, Cancer Imaging Program (CIP)

National Cancer Imaging Archive (NCIA) Development and Management Teams		
Development	***Documentation***	***Product and Program Management***
[4] University of Maryland, Department of Diagnostic Radiology	[5] VA Maryland Healthcare System	[6] MIRC Committee, Radiological Society of North America (RSNA)
[4] Northern Taiga Ventures, Inc.		

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

LISTSERV facilities pertinent to NCIA		
LISTSERV	***URL***	***Name***
NCIA Steering	**https://list.nih.gov/archives/ncia_steering.html**	NCIA Steering Discussion Forum
NCIA Developers	**https://list.nih.gov/archives/ncia_developers.html**	NCIA Developers Discussion Forum

TABLE OF CONTENTS

About This Guide v

Purpose v
Audience v
Topics Covered v
Text Conventions Used vi

Chapter 1
About REMBRANDT 1.5 1

About REMBRANDT 1
About REMBRANDT Functions 2

Chapter 2
Getting Started with REMBRANDT 1.5 3

Launching REMBRANDT 3
Creating a User Account 5
Logging In 5
Accepting REMBRANDT Provisions 5
Welcome to REMBRANDT 1.5 6
Getting Help 7
Application Support 7
Logging Out 8

Chapter 3
Conducting a Simple Search 9

Simple Search Overview 9
Gene Expression Simple Search 10
Eliminating Aliases 10
Understanding a Gene Expression Plot 11
K-M Gene Expression Simple Search 18
Redrawing the K-M Survival Plot for Gene Expression Data 18
Understanding K-M Survival Plot for Gene Expression Data 19
K-M Copy Number Simple Search 20

Redrawing the K-M Survival Plot for Copy Number Data 21
Understanding K-M Survival Plot for Copy Number Data 21
K-M Sample Search 23
Understanding K-M Survival Plot for Sample Data 23
Viewing the Clinical Reports 25
Viewing Clinical Plots 27

Chapter 4
Conducting Advanced Searches 29

Advanced Searches Overview 29
Gene Expression Advanced Search 30
Copy Number Advanced Search 35
Clinical Study Advanced Search 38
Managing Advanced Searches 40
Refining a Query 41

Chapter 5
High Order Analysis 43

High Order Analysis Overview 43
Performing a Class Comparison 44
Performing a Principal Component Analysis 46
Performing Hierarchical Clustering Analysis 48

Chapter 6
Viewing Results 51

Results Overview 51
Advanced Search or Query Results 51
Gene Expression Sample Report 52
Gene Expression Disease Report 58
Copy Number Sample Report 59
High Order Analysis Results 60
Class Comparison Report 61
Principal Component Analysis Plot 65
Hierarchical Clustering Report 66
Downloading BRB Array Tools and Files 66

Chapter 7
Managing Lists 69

Managing Lists Overview 69
Viewing the Data Items in a List 70
Removing Data Items to Create a New List 70
Combining Existing Lists to Create a New List 71
Deleting an Entire List 72

Adding a New "Custom" List 72

Glossary 75

About This Guide

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

- *Purpose* on page v
- *Audience* on page v
- *Topics Covered* on page v
- *Text Conventions Used* on page vi

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 an overview of the REMBRANDT program.
- *Chapter 2* provides instructions to start using REMBRANDT.
- *Chapter 3* describes how to search by gene keyword and Reporter ID and to create gene expression plots, Kaplan-Meier surival plots, and copy number-based graphs from those search results.
- *Chapter 4* describes how to create gene expression, copy number, and clinical queries.
- *Chapter 5* extends the basic knowledge of the previous chapters and shows you how to work with class comparisons, hierarchical clustering, and principal component analysis.

- *Chapter 6* describes how to view all the results generated from advanced searches and high order analyses.
- *Chapter 7* describes how to manage user-defined or study-defined patient ID, gene, and reporters lists.

Text Conventions Used

The following table 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.

Table Documentation conventions

Chapter 1

About REMBRANDT 1.5

This chapter introduces you to REMBRANDT and provides an overview of REMBRANDT functions.

Topics in this chapter include:

- *About REMBRANDT* on page 1
- *About REMBRANDT Functions* on page 2

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 Disorders 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.

About REMBRANDT Functions

Users can perform a variety of tasks in REMBRANDT. *Table 1.1* describes each REMBRANDT task.

Task	Description
Perform Simple Searches by Gene ID and SNP Probeset ID	Search the database and view the following search results: • Gene Expression plots • Kaplan-Meier Survival plots For more information, see *Simple Search Overview* on page 9)
Perform Advanced Adhoc Queries	Query the database and view search results for the following: • Gene Expression analysis • Copy Number Data analysis • Clinical Study analysis For more information, see *Advanced Searches Overview* on page 29.
Perform High Order Analyses	Run higher order analyses, including class comparisons, hierarchical clustering and principal component analyses. For more information, see *High Order Analysis Overview* on page 43.
Generate Reports	View Advanced Search and High Order Analysis results. Also download static, archive files for use in BRB-ArrayTools. For more information, see *Results Overview* on page 51.
Manage Lists	Manage user-defined or study-defined patient identifier, gene, or reporters lists. For more information, see *Managing Lists Overview* on page 69).

Table 1.1 REMBRANDT user tasks

CHAPTER

2

GETTING STARTED WITH REMBRANDT 1.5

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

Topics in this chapter include:

- *Launching REMBRANDT* on page 3
- *Creating a User Account* on page 5
- *Logging In* on page 5
- *Accepting REMBRANDT Provisions* on page 5
- *Welcome to REMBRANDT 1.5* on page 6
- *Getting Help* on page 7
- *Application SupportAppendix*
- *Logging Out* on page 8

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 2.1*).

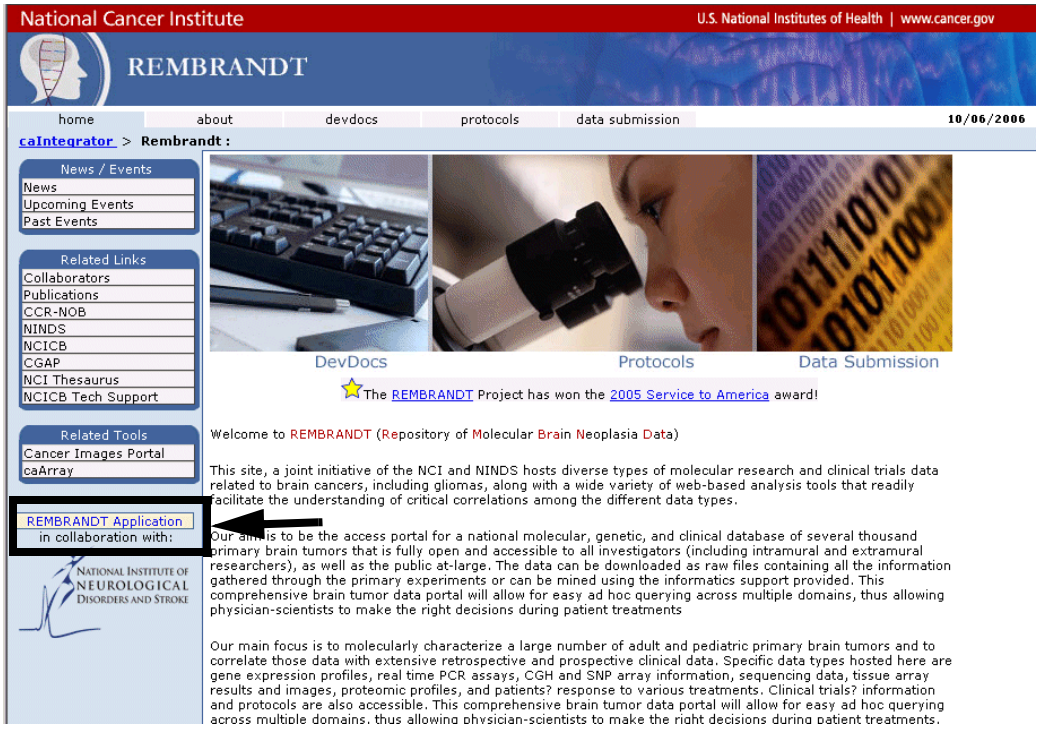


Figure 2.1 The REMBRANDT portal on the NCICB website

The REMBRANDT login page appears *(Figure 2.2).*

Figure 2.2 REMBRANDT login

Creating a User Account

Each REMBRANDT user is given a unique user name and password. The user name and password you are assigned determines your access rights for the software. To set up a user account, you must:

- Contact NCICB Application Support:
  - NCICB@pop.nci.nih.gov
  - 888-478-4423 (toll-free) or 301-451-4384 (local)

OR

- Go to the NCICB REMBRANDT login page and click the **request username/ password** link to send an e-mail requesting a username and password to NCICB Application Support.

Logging In

To log into REMBRANDT, you need your username and password assigned to you by the REMBRANDT Administrator.

1. On the login page, enter your **User Name** and **Password**.

   Note: If you would like to offer feedback via e-mail to the REMBRANDT development team, click the **feedback** link.

2. Click the **Submit** button. If your login is successful, the Legal Rules of the Road page appears *(Figure 2.3).*

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 2.3)* in the lower right-hand corner.

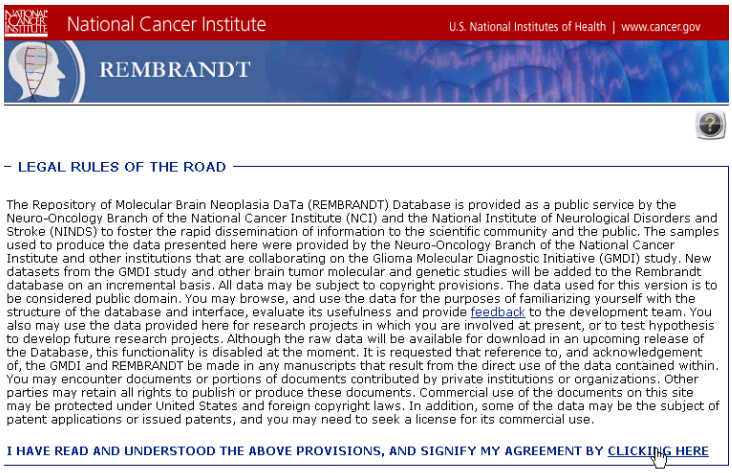


Figure 2.3 Legal Rules of the Road page

The REMBRANDT workspace appears (*Figure 2.4*).

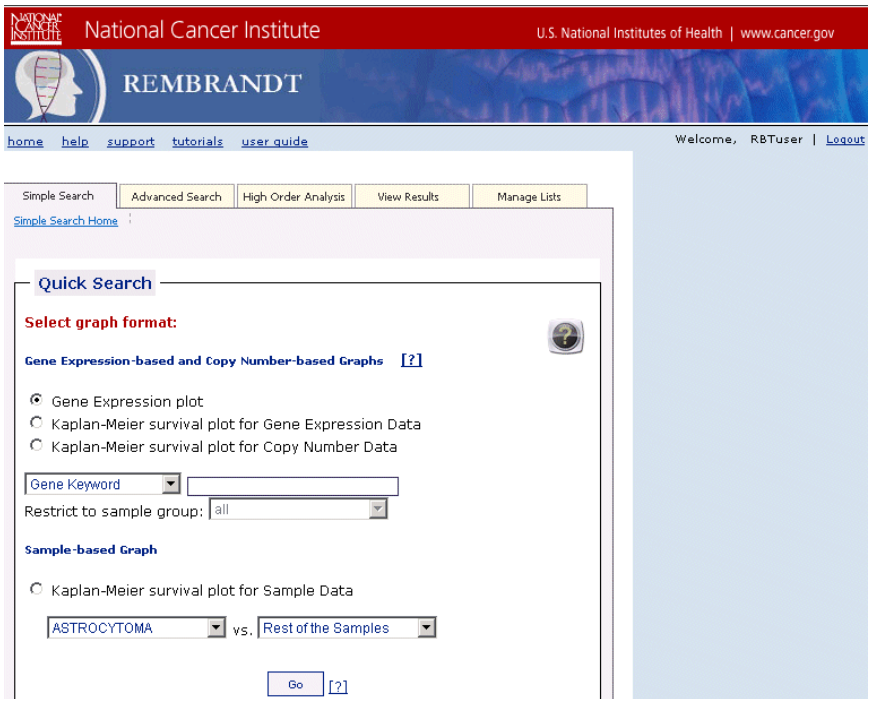


Figure 2.4 The Rembrandt workspace

Welcome to REMBRANDT 1.5

The REMBRANDT workspace comprises a set of five tabs, a blue panel, help links, and a logout link. The five tabs enable you to perform the following functions:

1. Perform a simple search
2. Create a complex queries
3. Perform higher order analyses
4. View results of searches
5. Manage lists

The blue panel may appear blank when you start using REMBRANDT, but once you start using REMBRANDT, information from various functions displays here:

- Queries defined with Advanced Search function
- Any filter settings defined in the High Order Analysis function
- List information managed in Manage Lists function

Getting Help

Information about how to use REMBRANDT is easily accessed from REMBRANDT's menu (*Figure 2.5*) in the top left of the REMBRANDT workspace.

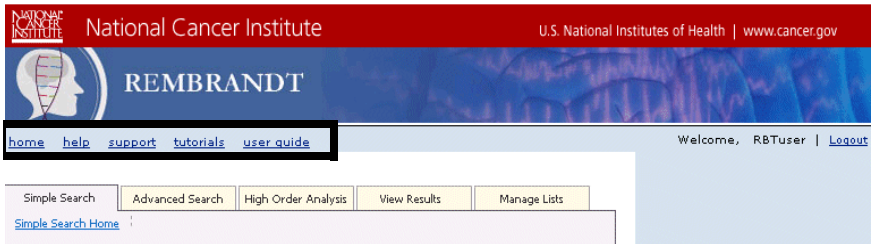


Figure 2.5 REMBRANDT's menu

Table 2.1 describes each item on the REMBRANDT toolbar.

Help	***How to Access***
Complete online help	To access the complete version of online REMBRANDT help, click the **help** link located under the REMBRANDT menu. For page-level help, click on any REMBRANDT page.
Application support	To obtain support for REMBRANDT, click the **support** link located under the REMBRANDT menu.
Tutorials	To access REMBRANDT tutorials, click the **tutorials** link located under the REMBRANDT menu.
User's Guide	To access a pdf version of the *REMBRANDT User's Guide*, click the **user guide** link located under the REMBRANDT menu.

Table 2.1 Getting help with REMBRANDT

Application Support

You can find additional support at the NCICB Applications Support Web site. To access the site, do the following:

Click the **support** link in the upper right-hand corner. The NCICB Applications Support Group page appears.

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.

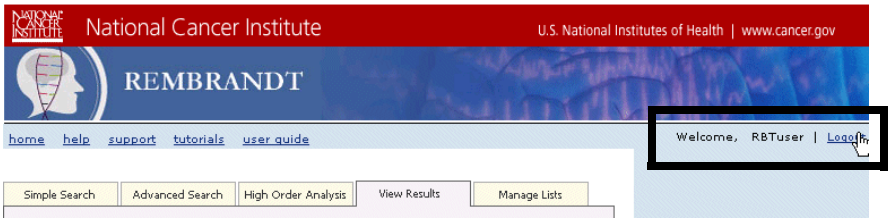


Figure 2.6 Logout link

The Logout page appears.

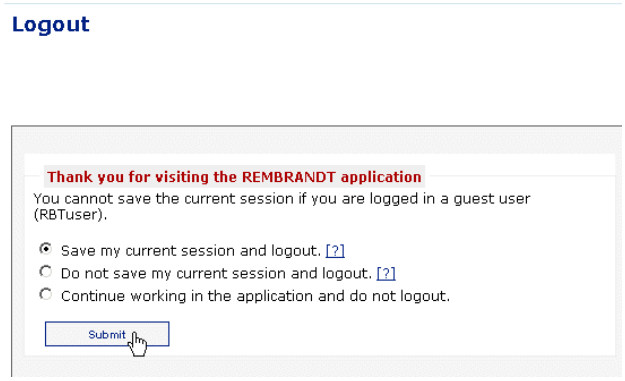


Figure 2.7 Logout link

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. Click the **Submit** button.

CHAPTER

3

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* on page 9
- *Gene Expression Simple Search* on page 10
- *K-M Gene Expression Simple Search* on page 18
- *K-M Copy Number Simple Search* on page 20
- *K-M Sample Search* on page 23
- *Viewing the Clinical Reports* on page 25
- *Viewing Clinical Plots* on page 27

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*).

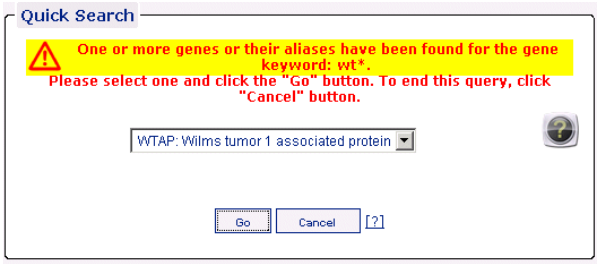


Figure 3.8 Alias message

   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 **Geometric Mean** Gene Expression Plot (*Figure 3.1*) appears.

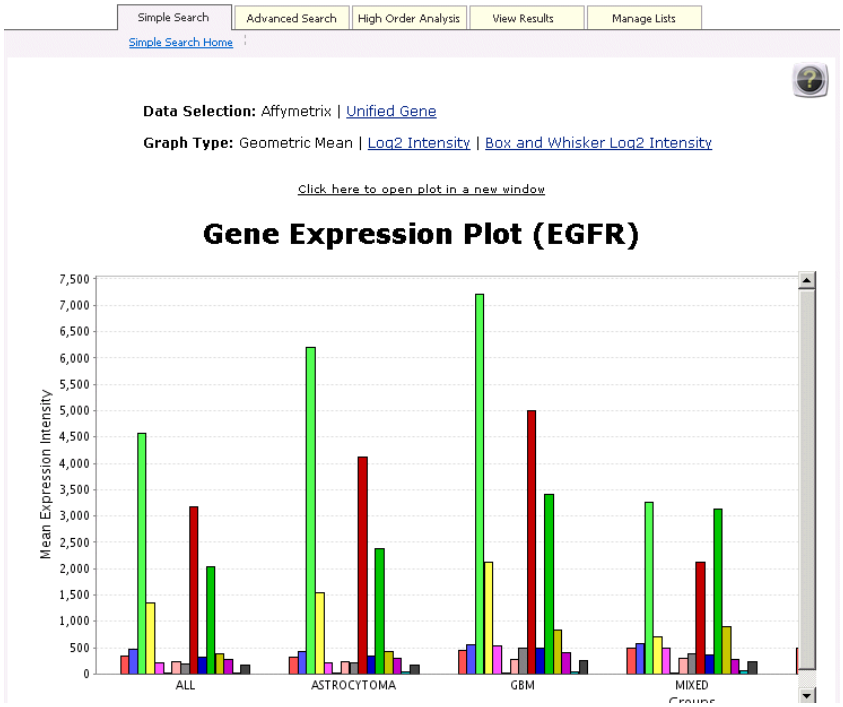


Figure 3.1 Gene Expression Plot page (1 of 2)

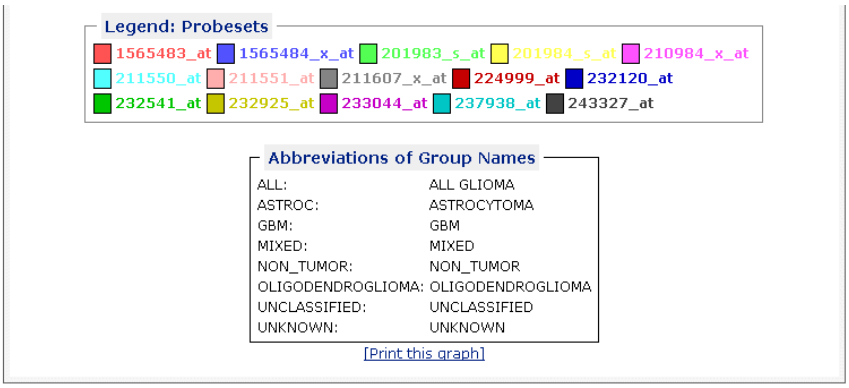


Figure 3.2 Gene Expression Plot page (2 of 2)

Table 3.1 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 gene-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.
Graph Type	Displays different versions of the Gene Expression Plot. **Note:** If you select the **Unified** Data Selection type, the **Box and Whisker Log2 Intensity** Graph Type is not available. • **Geometric Mean** is the default graph shown when you perform a simple search. 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 3.1 Understanding the Gene Expression Plot page

Geometric Mean Plot Details

By default, the **Geometric Mean** Gene Expression Plot (*Figure 3.3*) displays when you perform a simple search. The Geometric Mean Gene Expression Plot displays mean expression intensity (Geometric mean) versus Groups.

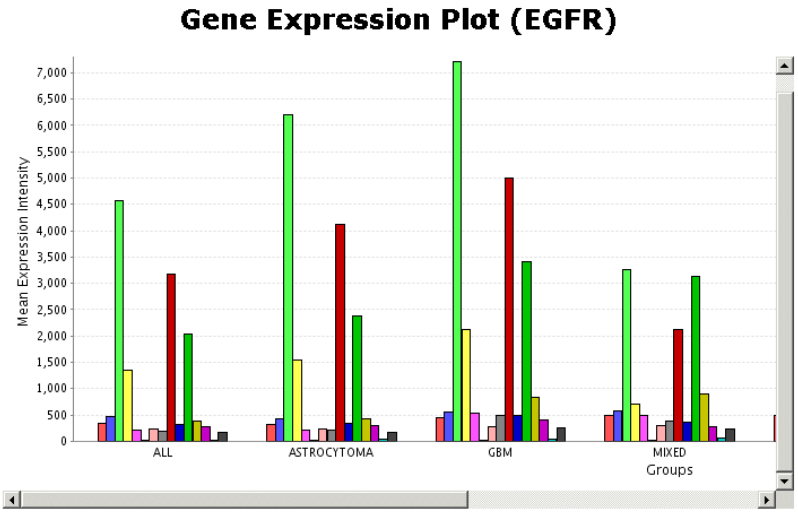


Figure 3.3 Geometric Mean Gene Expression Plot

Mouse-over a bar on the graph to display Additional Information. *Table 3.2* describes Additional Information details.

Item	***Special Instructions***
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 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 3.2 Geometric Mean Gene Expression Plot Additional Information

Log2 Intensity Gene Expression Plot Details

The **Log2 Intensity** Gene Expression Plot (*Figure 3.4*) 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 (samples average 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.

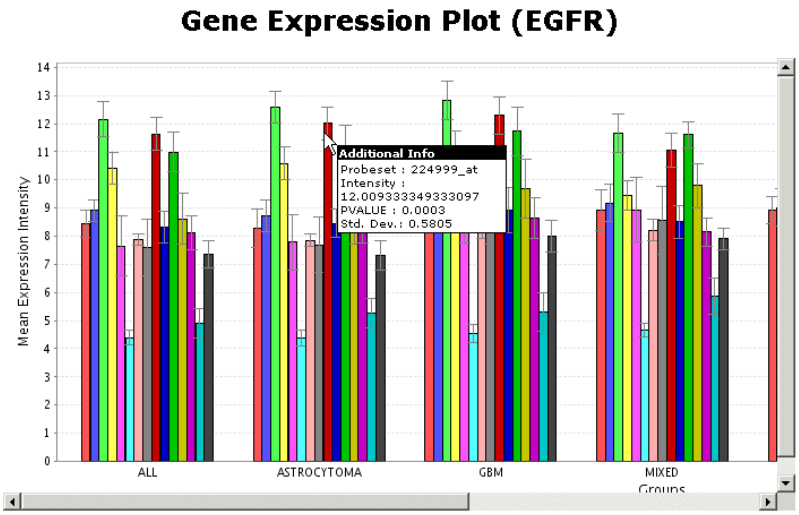


Figure 3.4 Log2 Intensity Gene Expression plot

Mouse-over a bar on the graph to display Additional Information. *Table 3.3* describes Additional Information details.

Item	***Special Instructions***
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 3.3 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 3.5*) displays a box plot without all the individual data points. A box and whisker plot or box plot is a graph that presents information from a five-number summary. Mouse over one probe-set on the plot to display additional information about the probeset for one group. The following values are listed together and ordered from lowest to highest:

- Most extreme values in the dataset (the maximum and minimum values)
- Lower and upper quartiles
- Median

The following items in the graph indicate the following:

- **Black dot in the box** indicates mean value.
- **Circles** are potential outliers.
- **Triangles** are outliers beyond the graph.

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.

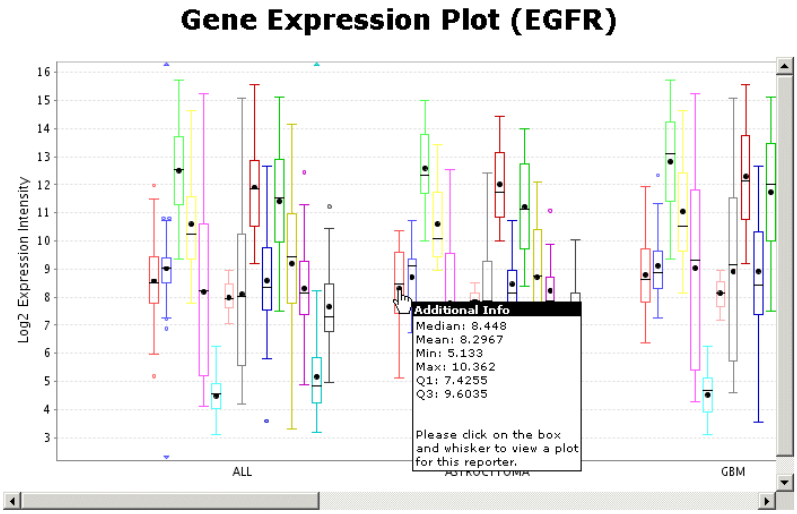


Figure 3.5 Box and Whisker Log2 Intensity Gene Expression plot

Mouse-over the bar within the box to display Additional Information. *Table 3.4* describes Additional Information details.

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

Item	*Special Instructions*
Median	Median value of log 2 (or ratio) gene expression values for particular probeset or unified gene.
Mean	Mean value of log 2 (or ratio) gene expression values for 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.
...plot	Represents the probeset name.

Table 3.4 Box and Whisker Log2 Intensity Gene Expression Plot Additional Information

Displaying a Coin Plot

A *coin plot* is box-and-whisker plot (*Figure 3.6*) 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. The following items in the graph indicate the following:

- **Circles** mean potential outliers.
- **Triangles** mean some outliers beyond the graph.

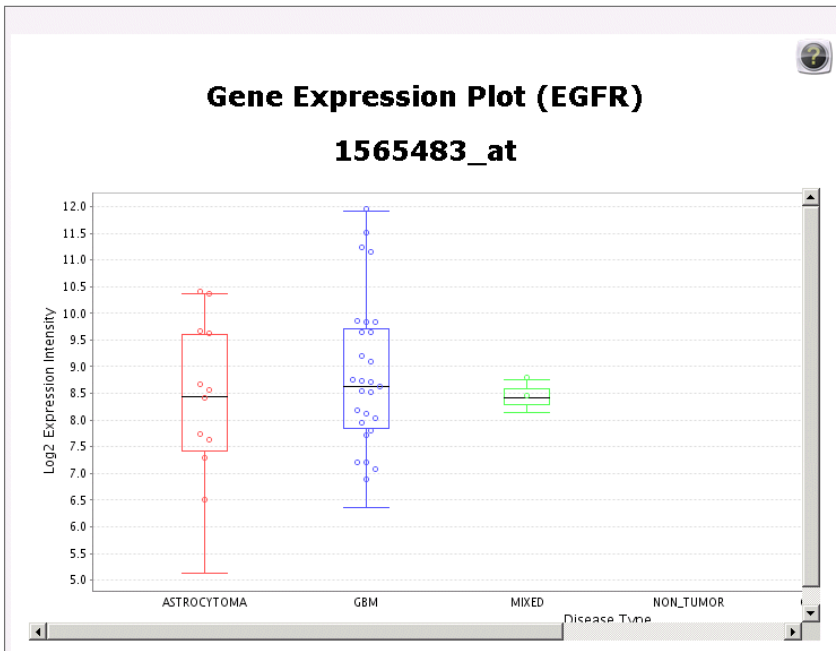


Figure 3.6 Coin Plot for a Probeset

Table 3.5 describes each area of the Gene Expression Coin Plot page.

Item	Special Instructions
Abbreviations of Group Names	Lists the complete name of each group abbreviation in the plot.
Print this Graph	Click to print the graph.

Table 3.5 Understanding the Gene Expression Plot Coin Plot page

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

By opening a Gene Expression plot in a new window (*Figure 3.7*), you can perform a number of tasks with the Gene Expression plot.

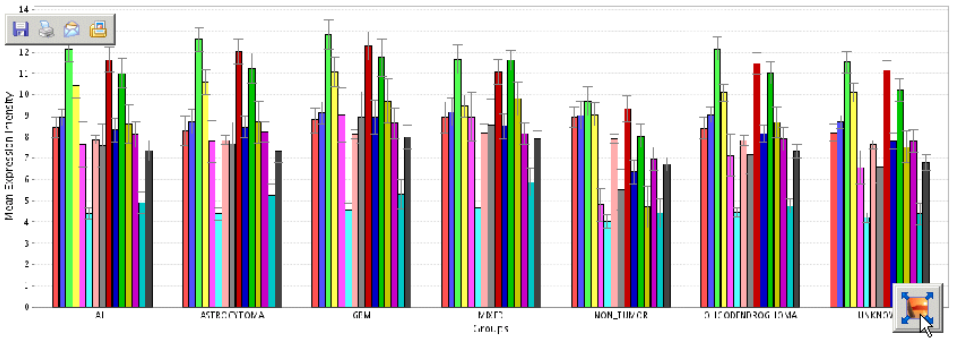


Figure 3.7 Displaying a Gene Expression Plot in a new window

Table 3.6 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 3.6 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. To visualize the K-M plot for the unified probeset, select a value from the **Reporters** drop-down list *(Figure 3.8).*

Figure 3.8 Redrawing a Kaplan-Meier Gene Expression data

3. Specify a **Unified** or a **Affymetrix Reporter Selection**.
4. Click the **Redraw Graph** button.

Understanding K-M Survival Plot for Gene Expression Data

A K-M Survival Plot for Gene Expression Data (*Figure 3.9*) 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.

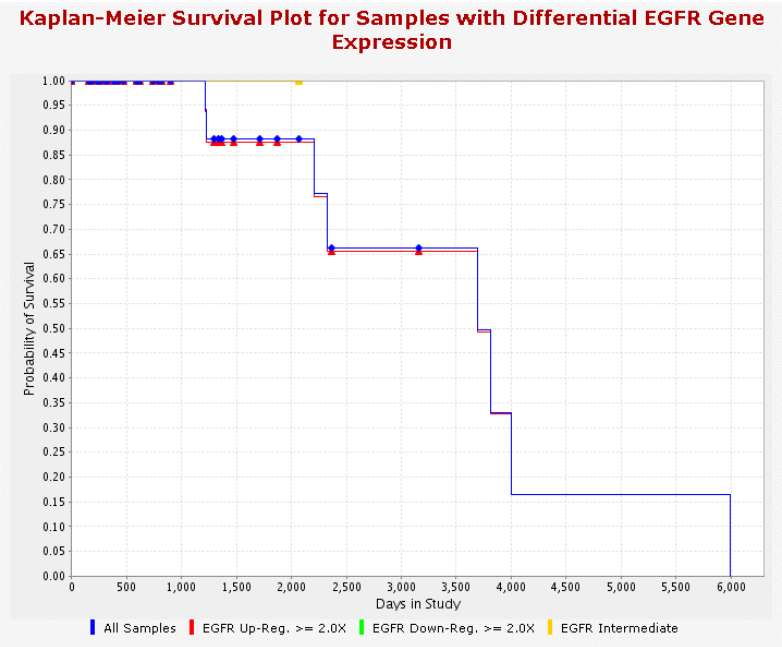


Figure 3.9 Kaplan-Meier Survival Plot for Gene Expression Data (1 of 2)

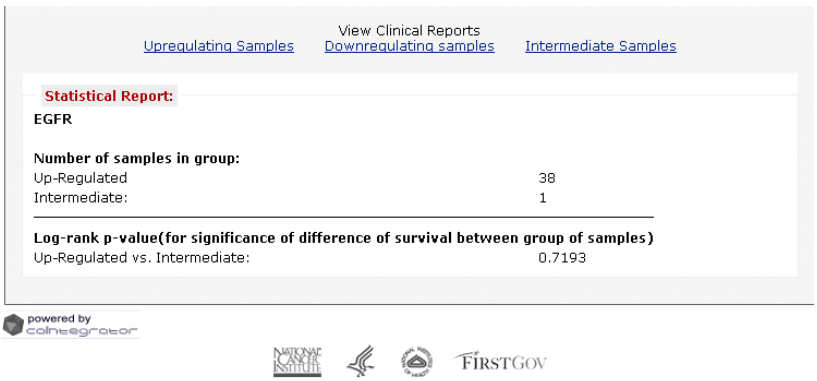


Figure 3.10 Kaplan-Meier Survival Plot for Gene Expression Data (2 of 2)

Table 3.7 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.
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 *Viewing the 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 3.7 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 3.11).*

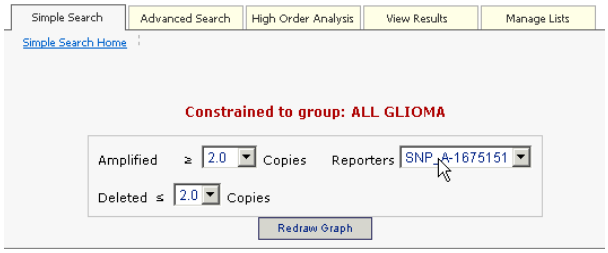


Figure 3.11 Redrawing a Kaplan-Meier Survival Plot for Copy Number data page

3. Click the **Redraw Graph** button.

Understanding K-M Survival Plot for Copy Number Data

A **gene keyword** search displays a plot (*Figure 3.12*) 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.

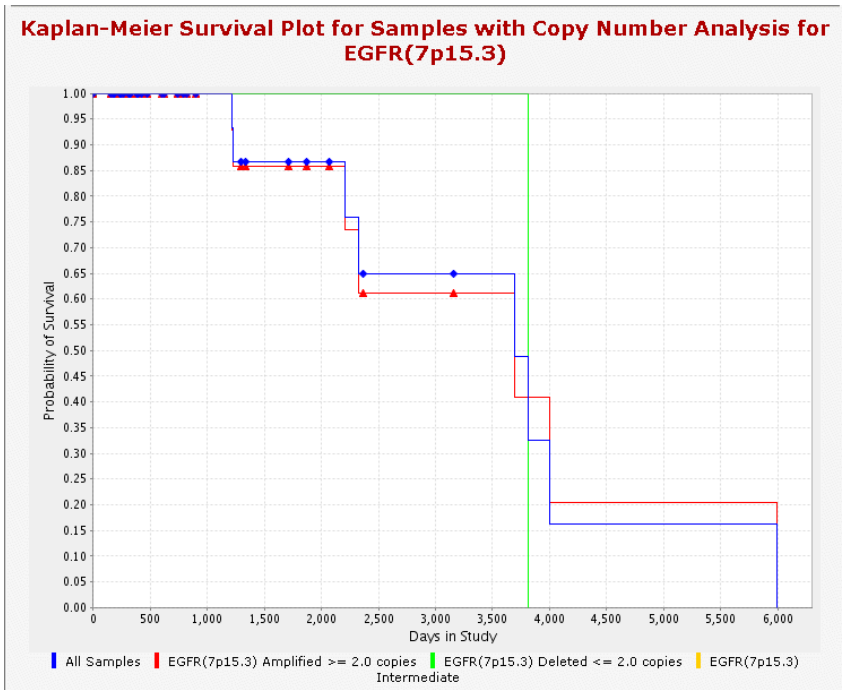


Figure 3.12 Kaplan-Meier Survival Plot for Copy Number Data (gene search) (1 of 2)

View Clinical Reports

Samples with Amplification | Samples with Deletion | Intermediate Samples

Statistical Report:

EGFR(7p15.3)
Reporter: SNP_A-1675151

Number of samples in group:

Amplified	34
Deleted	1

Log-rank p-value(for significance of difference of survival between group of samples)

Amplified vs.Deleted	0.9903

powered by caintegrator

Figure 3.13 Kaplan-Meier Survival Plot for Copy Number Data (gene search) (2 of 2)

Table 3.8 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.
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 *Viewing the 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 3.8 Kaplan-Meier Survival Plot for Copy Number Data (gene keyword search)

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 each drop-down list for comparison purposes.
3. Click the **Go** button. The Kaplan-Meier survival plot appears *(Figure 3.14).*

Understanding K-M Survival Plot for Sample Data

A Kaplan-Meier Survival Plot for Sample Data *(Figure 3.14)* 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. You can dynamically modify the fold change (up and down regulation) thresholds and redraw the plot.

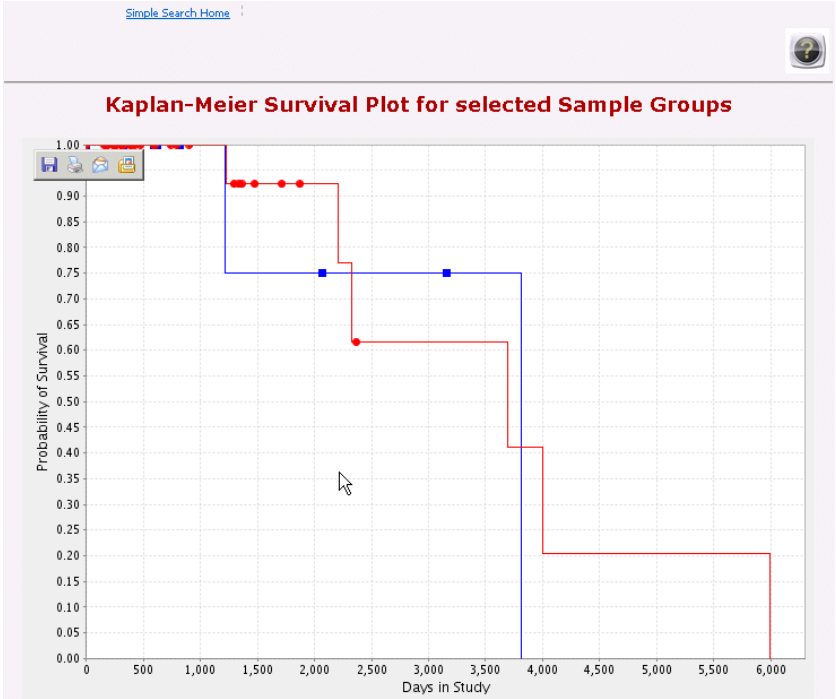


Figure 3.14 Kaplan-Meier Survival Plot Sample Data (1 of 2)

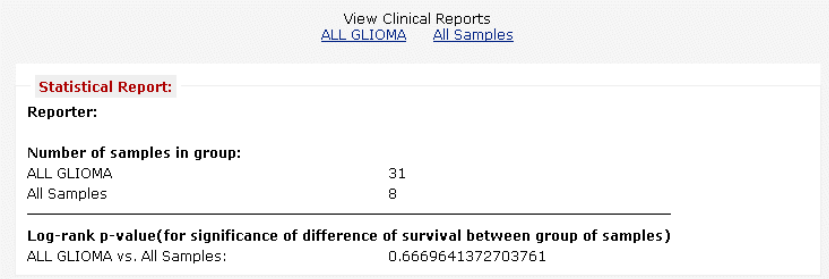


Figure 3.15 Kaplan-Meier Survival Plot for Sample Data (2 of 2)

Table 3.9 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.
View Clinical Reports	To display clinical data for the selected sample groups, click the group link. For more information, see *Viewing the Clinical Reports*.

Table 3.9 Kaplan-Meier Survival Plot for Sample Data page

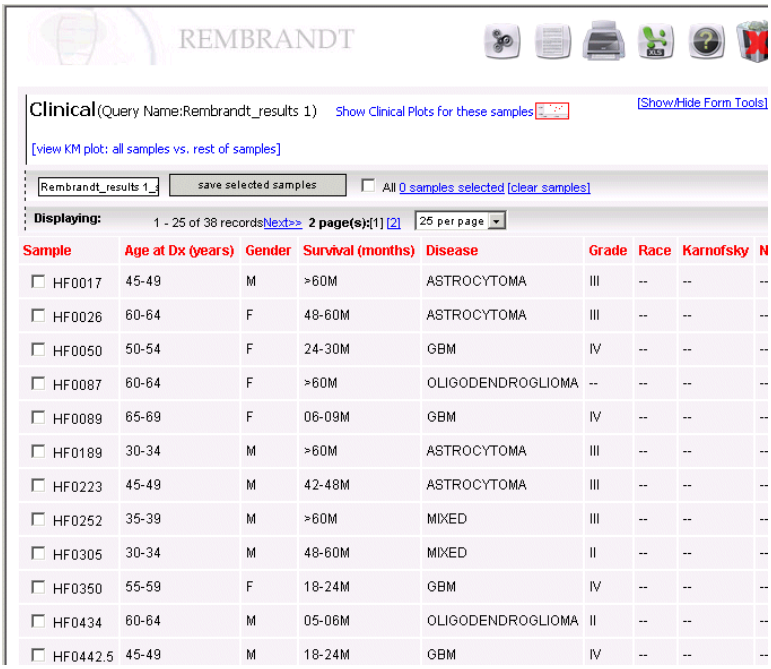
Item	Special Instructions
Statistical Report	• Displays the search criteria for the plot. • Displays the reporter selected for the plot. • **Number of Samples** specifies the number of Up-Regulated, Intermediate, Down-Regulated samples, if any. • **Log-rank *p*-value** indicates the significance of the 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 3.9 Kaplan-Meier Survival Plot for Sample Data page

Viewing the 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 3.16*). On the Clinical page, you can select and save samples or show clinical plots of the selected samples.

Note: When either a gene expression filter and/or a copy number filter are 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.

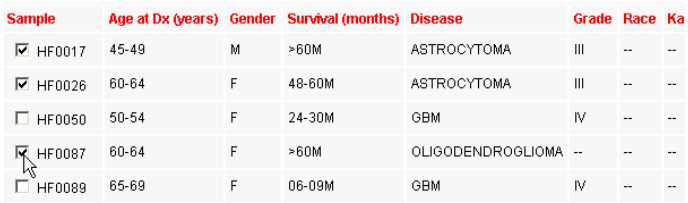


Sample	Age at Dx (years)	Gender	Survival (months)	Disease	Grade	Race	Karnofsky	N
HF0017	45-49	M	>60M	ASTROCYTOMA	III	--	--	--
HF0026	60-64	F	48-60M	ASTROCYTOMA	III	--	--	--
HF0050	50-54	F	24-30M	GBM	IV	--	--	--
HF0087	60-64	F	>60M	OLIGODENDROGLIOMA	--	--	--	--
HF0089	65-69	F	06-09M	GBM	IV	--	--	--
HF0189	30-34	M	>60M	ASTROCYTOMA	III	--	--	--
HF0223	45-49	M	42-48M	ASTROCYTOMA	III	--	--	--
HF0252	35-39	M	>60M	MIXED	III	--	--	--
HF0305	30-34	M	48-60M	MIXED	II	--	--	--
HF0350	55-59	F	18-24M	GBM	IV	--	--	--
HF0434	60-64	M	05-06M	OLIGODENDROGLIOMA	II	--	--	--
HF0442.5	45-49	M	18-24M	GBM	IV	--	--	--

Figure 3.16 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 3.17*).

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



Sample	Age at Dx (years)	Gender	Survival (months)	Disease	Grade	Race	Ka
☑ HF0017	45-49	M	>60M	ASTROCYTOMA	III	--	--
☑ HF0026	60-64	F	48-60M	ASTROCYTOMA	III	--	--
☐ HF0050	50-54	F	24-30M	GBM	IV	--	--
☑ HF0087	60-64	F	>60M	OLIGODENDROGLIOMA	--	--	--
☐ HF0089	65-69	F	06-09M	GBM	IV	--	--

Figure 3.17 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 3.18*).

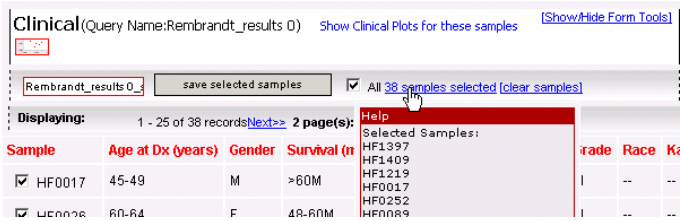


Figure 3.18 Selecting all of the samples on the Clinical window

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

2. To save the samples, enter a name for the sample (*Figure 3.19*).

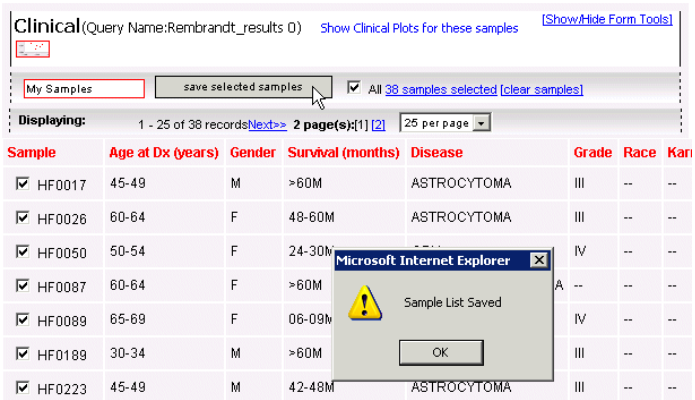


Figure 3.19 Saving Selected Samples on the Clinical page

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

Note: If you want to add the saved sample set to an Advanced Search query later, close the Clinical window. For more information, see Step 2 on the Refine Query page (see *Refining a Query* on page 41).

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*).
6. To show a Kaplan-Meier Survival Plot for Sample Groups, click View KM Plot (see *K-M Sample Search*).

Viewing 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 3.20*).

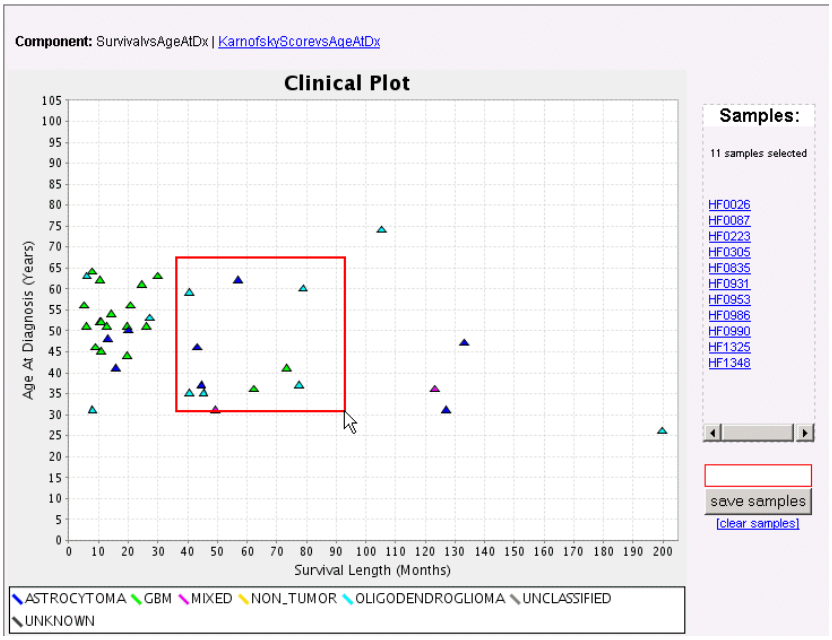


Figure 3.20 Sampling from a Clinical plot

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.

CHAPTER

4

CONDUCTING ADVANCED SEARCHES

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

Topics in this chapter include:

- *Advanced Searches Overview* on page 29
- *Gene Expression Advanced Search* on page 30
- *Copy Number Advanced Search* on page 35
- *Clinical Study Advanced Search* on page 38
- *Managing Advanced Searches* on page 40
- *Refining a Query* on page 41

Advanced Searches Overview

The Advanced Search function enables you to create multiple searches and then group the searches into a single, complex query from which you can generate reports. The following is an overview of this process.

1. The Advanced Search - Build Query page enables you to define advanced searches from three categories:
   - **Gene Expression Analysis**
   - **Copy Number Analysis**
   - **Clinical Analysis**
2. Once you create a search, the Advanced Search - Build Query page appears. You can add more searches or copy, edit and delete existing searches with the buttons on the blue panel .
3. Once the searches are complete, click the **Finalize Query** button or **Refine Query** option.

4. On the Refine Query page, group the searches with parentheses and AND/OR conditional items to create one complex query.
5. Validate the query and select and run a report. Report results are listed 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 4.1*).

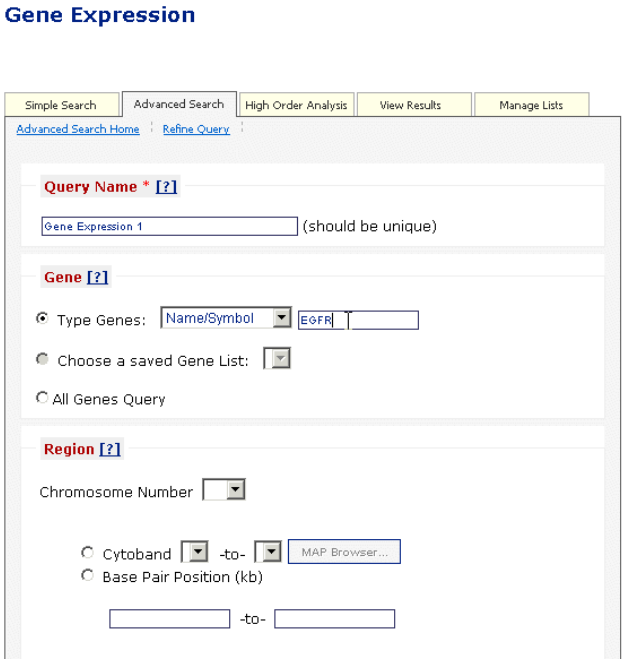


Figure 4.1 Advanced Gene Expression page (top portion)

2. You are required to enter at least one search criteria for the query. *Table 4.1* 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.**). Enter the corresponding comma delimited value or an identifier for the genes to be searched. OR Click the **Browse** button to upload a text file containing Gene identifiers. The file must have one entry per line and a return at the end of the file.
	Choose a Saved Gene List	Select a gene list created with the REMBRANDT Manage Lists function (see *Managing Lists Overview*) to further filter the search. If you have not added a gene list, this option is not available.
	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 *Refining a Query*.
Region	Chromosome Number	Select the chromosomal region of interest to search for by specifying a chromosome of interest (**1-22, X** or **Y**).
	Cytoband	Lists only the relevant cytobands for a particular 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 Clones	Enter or paste a comma-delimited IMAGE Clone ID/ Affymetrix probeset ID list to be searched. IMAGE Clone identifiers must start with `IMAGE:`. OR Click the **Browse** button to upload a text file containing identifiers. The file must have one entry per line and a return at the end of the file.
	Choose a Saved Clone List	Select a clone list created with the REMBRANDT Manage Lists function (see *Managing Lists Overview*) to further filter the search. If you have not added a clone list, this option is not available.
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)).

Table 4.1 Advanced Gene Expression search criteria instructions

Criteria	Item Name	Special Instructions
	Go Browser	Click the button to search for and select a GO classification. See *Selecting a Gene Ontology (GO) Classification*.
Pathways	browse caBIO	Click the button to search for and select a pathway. See *Selecting a Pathway*.
	KEGG	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 4.1 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 4.2*).

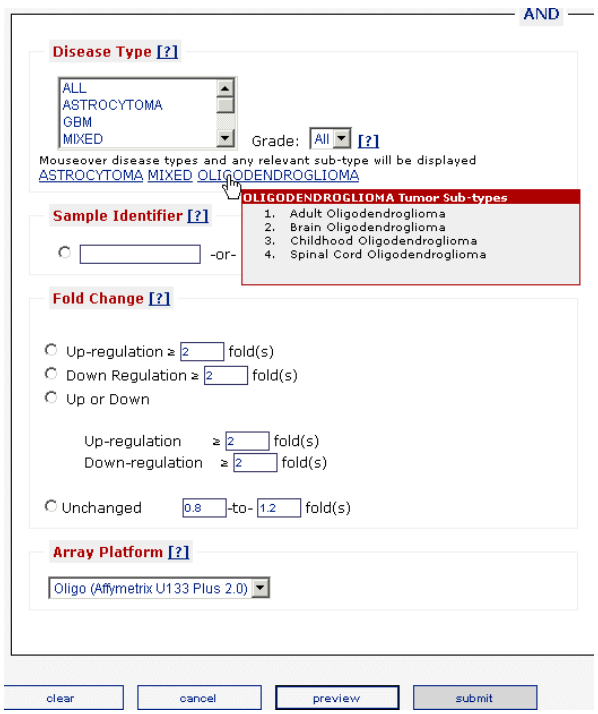


Figure 4.2 Advanced Gene Expression Disease Type

3. Optionally, you can combine a disease type with the query. *Table 4.2* lists the Disease Type items:

Criteria	*Item Name*	*Special Instructions*
Disease Type	(list box)	Select a disease. Mouse over the disease type to display the tumor sub-types for the selected disease type.
	Grade	*Future Implementation*
Sample Identifier	(list boxes	Enter or paste a comma-delimited sample identifier list to be searched. OR Click the **Browse** button to upload a text file containing identifiers. The file must have one entry per line and a return at the end of the file.
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 4.2 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 restore the report to its original state and clear any highlighting, click the **Clear** button.
   - To eliminate all data currently entered in the form and not save the search, 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 4.3*).

Figure 4.3 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 4.4*).

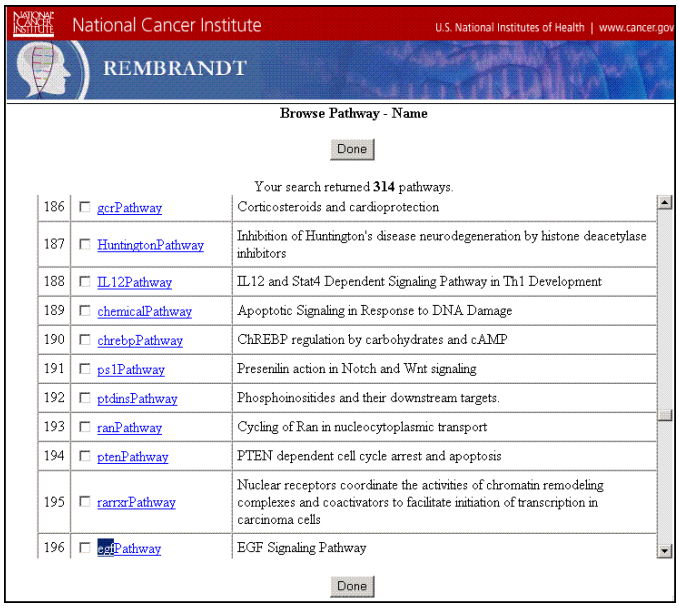


Figure 4.4 caBIO pop-up window

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 4.5).*

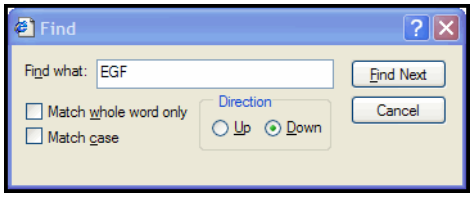


Figure 4.5 Search dialog box

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 4.6*).

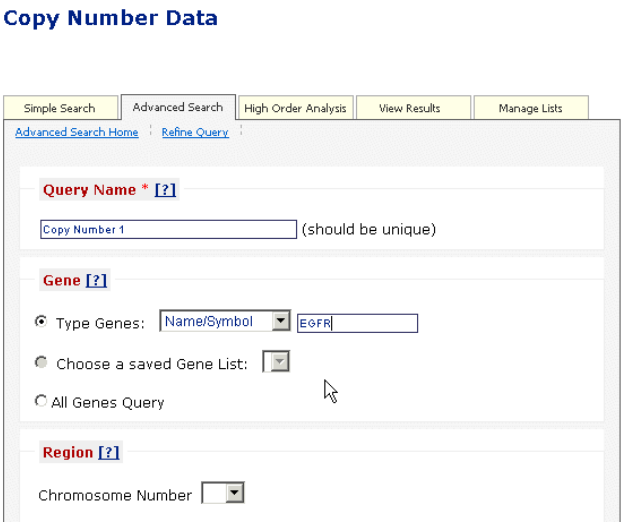


Figure 4.6 Copy Number Data page (top portion)

2. You are required to enter at least one search criteria for the copy number query. *Table 4.3* 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 the corresponding comma delimited value or identifiers for the genes to be searched.
	Choose a Saved Gene List	Select a gene list created with the REMBRANDT Manage Lists function (see *Managing Lists Overview*) to further filter the search. If you have not added a gene list, this option is not available.
	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 *Refining a Query*.
Region	Chromosome Number	Select the chromosomal region of interest to search for by specifying a chromosome of interest (**1-22, X** or **Y**).
	Cytoband	Lists only the relevant cytobands for a particular 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**). Enter or paste a comma-delimited SNP list to be searched. OR Click the **Browse** button to upload a text file containing identifiers. The file must have one entry per line and a return at the end of the file.
	Choose a Saved SNP List	Select an SNP list created with the REMBRANDT Manage Lists function (see *Managing Lists Overview*) to further filter the search. If you have not added a SNP list, this option is not available.
	Validated SNPs	Select one type of Validated SNPs: **All**, **Excluded**, **Included**, or **Only**.
Allele Frequency	Population Type	*Future Implementation*

Table 4.3 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 4.2*).

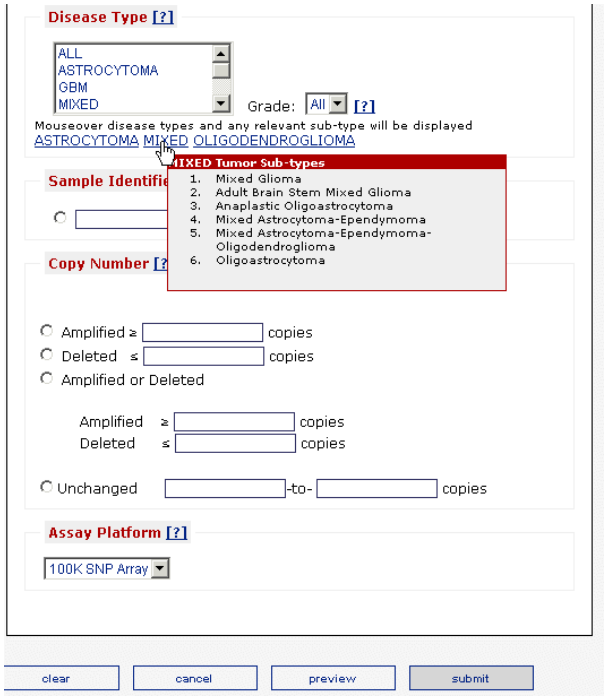


Figure 4.7 Advanced Copy Number Disease Type

3. Optionally, you can combine a disease type with the query. *Table 4.4* lists the Disease Type items:

Criteria	***Item Name***	***Special Instructions***
Disease Type	(list box)	Select a disease. Mouse over the disease type to display the tumor sub-types.
	Grade	*Future Implementation*
Sample Identifier	(list boxes	Enter or paste a comma-delimited sample identifier list to be searched. OR Click the **Browse** button to upload a text file containing identifiers. The file must have one entry per line and a return at the end of the file.
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 4.4 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 restore the report to its original state and clear any highlighting, click the **Clear** button.
   - To eliminate all data currently entered in the form and not save the search, 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 4.8*).

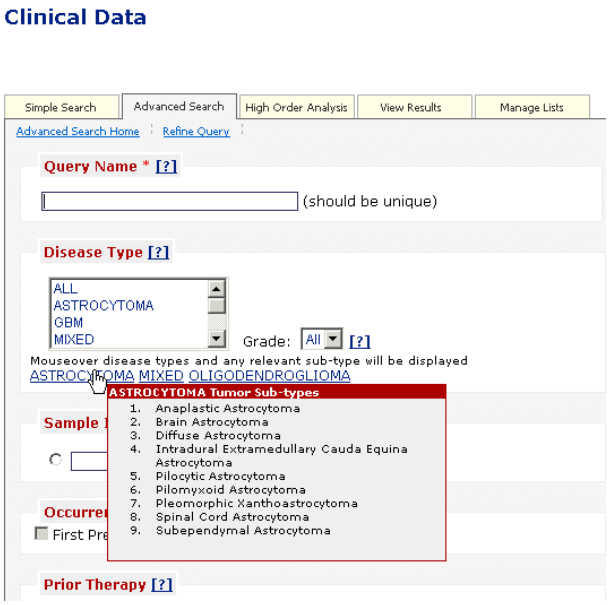


Figure 4.8 Clinical Data page (top portion)

2. You must specify a disease type, and optionally complete the remaining information. *Table 4.5* lists the Disease Type items:

Criteria	*Item Name*	*Special Instructions*
Disease Type	(list box)	Select a disease. Mouse over the disease type to display the tumor sub-types for the selected disease type.
	Grade	*Future Implementation*
Sample Identifier	(list box)click	Enter or paste a comma-delimited sample identifier list to be searched. OR Click the **Browse** button to upload a text file containing identifiers. The file must have one entry per line and a return at the end of the file.
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.

Table 4.5 Advanced Clinical Data criteria instructions

Criteria	Item Name	Special Instructions
	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
	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 4.5 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 restore the report to its original state and clear any highlighting, click the **Clear** button.
   - To eliminate all data currently entered in the form and not save the search, click the **Cancel** button.
   - To display a preview of the report generated by the search results, click the **Preview** button.

Managing Advanced Searches

Once you submit an advanced search, you are returned to the Advanced Search - Build Query page. Your search is added to the counter next to the appropriate Advanced Search button. You can perform one of the following tasks:

- Add more searches: Click the **Gene Expression Analysis**, **Copy Number Analysis**, or **Clinical Analysis** button.
- Copy, edit, or delete existing searches. Click the **Finalize Query** button or the **Refine Query** option under the Advanced Search tab, and see *Refining a Query*.
- Create a complex query from the submitted advanced searches. Click the **Finalize Query** button or the **Refine Query** option under the Advanced Search tab, and see *Refining a Query*.

Refining a Query

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

Note: The blue panel on the right side displays any previously-defined advanced searches. Using the **Copy**, **Edit**, and **Delete** buttons, you can add, modify, or remove existing searches.

1. Fill in the Refine Query criteria (*Figure 4.9*).

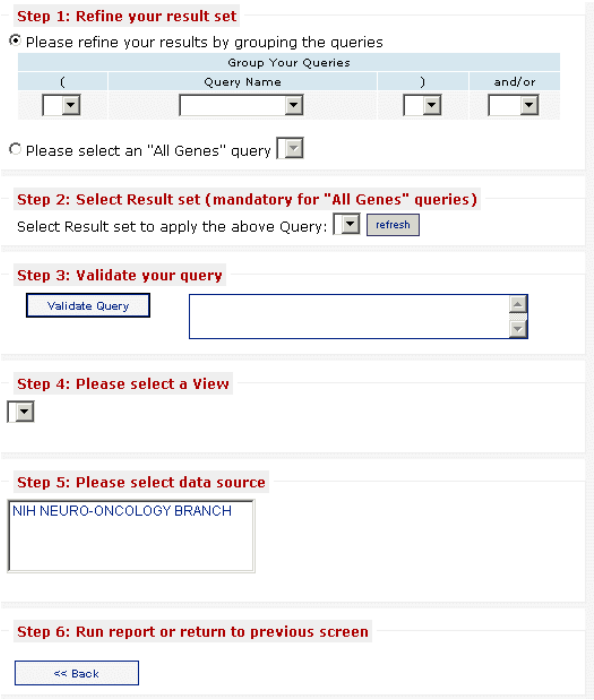


Figure 4.9 Refine Query page

Table 4.6 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. To ensure that the list is current, click the **Refresh** button.
Step 3. Validate your query	You must click to validate that the number of parentheses added to the query grouping and the name of your query appears.
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 a data source	Select a datasource to filter the query by the institute providing data. You can select more than one institute. The **NIH Neuro oncology branch** is the public dataset. **Note:** The Simple Search function and Preview assigns all the institutes to which you have access.

Table 4.6 Refining Query instructions

To return to the Advanced Search - Build Query page, 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 syntax of your query.

Chapter 5

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* on page 43
- *Performing a Class Comparison* on page 44
- *Performing a Principal Component Analysis* on page 46
- *Performing Hierarchical Clustering Analysis* on page 48

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**

The blue panel displays the filter settings selected for you 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 5.1)* 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.

Class Comparison Analysis Form

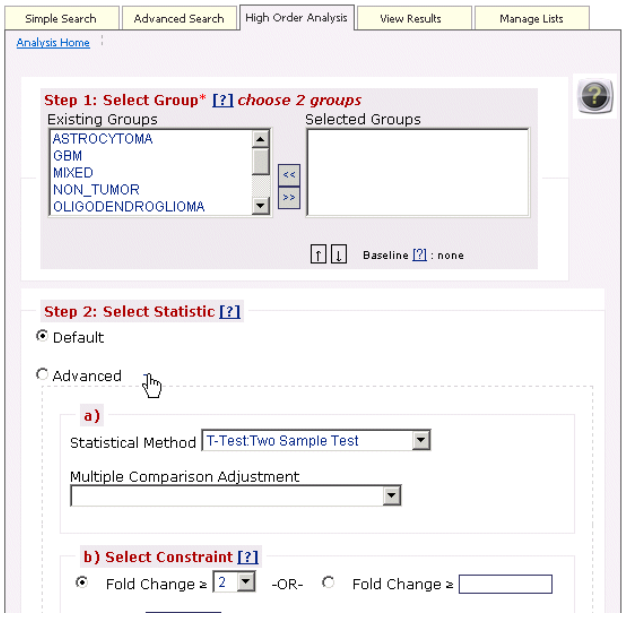


Figure 5.1 Class Comparison Analysis Form page

2. You are required to complete at least one step for the class comparison. *Table 5.1* lists the available criteria:

Criteria	*Item Name*	*Special Instructions*
Step 1. Select Group	Existing Groups Selected Groups	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. ● **Wilcoxin 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. a. If there are three or more predefined groups, **F-test: One Way ANOVA** is the default statistical method. b. 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. c. Using ANOVA instead of multiple T-tests reduces the probability of a type-I error.
	● Multiple Comparison Adjustment	Family-wise Error Rate (FWER): Bonferroni False Discover Rate (FDR): Benjamini-Hochberg

Table 5.1 Class Comparison criteria instructions

Criteria	Item Name	Special Instructions
	• Select constraint	*Future Implementation*
	• p-value	*Future Implementation*
Step 3. Select Array Platform	Select Array Platform	Select the array platform.

Table 5.1 Class Comparison 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 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 5.2)* enables you to define criteria to perform a PCA.

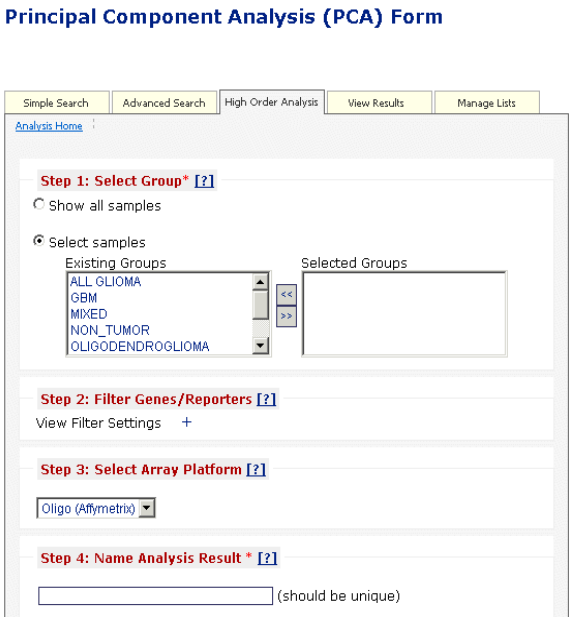


Figure 5.2 Selecting Principal Component Analysis criteria

2. You are required to complete at least one step for the Principal Component analysis. *Table 5.2* 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.
	• 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.
	+ (-)	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	Select saved differentially expressed genes identified by class comparison. Choose an option (**gene list1** or **gene list2**) or click the **Upload** link to upload a file.
	• Use differentially expressed reporters	Select saved differentially expressed reporters identified by class comparison. Choose an option (**reporter list1** or **reporter list2**) or click the **Upload** link to upload a file.
	• 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 5.2 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 5.3)* enables you to perform a clustering.

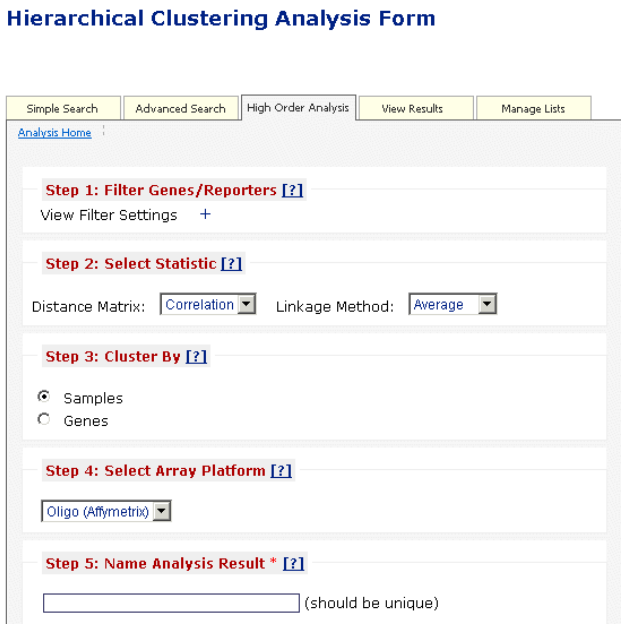


Figure 5.3 Selecting Hierarchical Clustering criteria

2. You are required to enter at least one step for the hierarchical clustering. *Table 5.3* 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.
	+ (-)	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	Select saved differentially expressed genes identified by class comparison. Choose an option (**gene list1** or **gene list2**) or click the **Upload** link to upload a file.
	• Use differentially expressed reporters	Select saved differentially expressed reporters identified by class comparison. Choose an option (**reporter list1** or **reporter list2**) or click the **Upload** link to upload a file.

Table 5.3 Hierarchical Clustering criteria instructions

Criteria	*Item Name*	*Special Instructions*
	• Set These Filters as Default	Click to save the options as default filter settings.
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 5.3 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 6

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* on page 51
- *Advanced Search or Query Results* on page 51
- *High Order Analysis Results* on page 60
- *Downloading BRB Array Tools and Files* on page 66

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 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.

Advanced Search or Query Results

The following Advanced Search reports are generated:

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

View Results (*Figure 6.1*) 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.

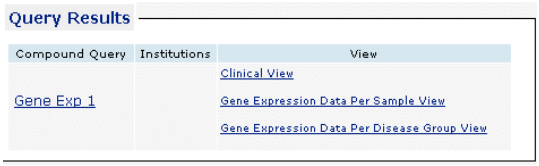


Figure 6.1 Query Results

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

Gene Expression Sample Report

The Gene Expression Sample report (*Figure 6.2*) 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.

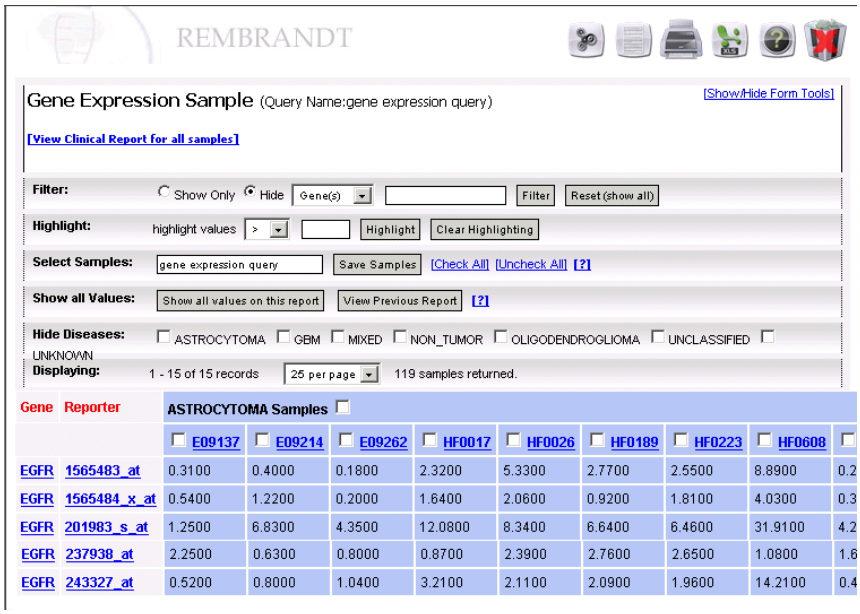


Figure 6.2 Gene Expression Data Per Sample Report

Note: 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.

For more information, see the following:

- *Filtering Results by Gene or Reporter (Filter Toolbar)*
- *Highlighting Results By Value (Highlight Toolbar)*
- *Selecting and Saving Sample Results (Select Samples Toolbar)*
- *Differentiating Data (Show All Values Toolbar)*
- *Removing Columns (Hide Diseases toolbar)*
- *Showing Additional Information*

Filtering Results by Gene or Reporter (Filter Toolbar)

To filter a report, follow these steps:

1. From the **Filter** toolbar *(Figure 6.3)*, 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 6.3)*.

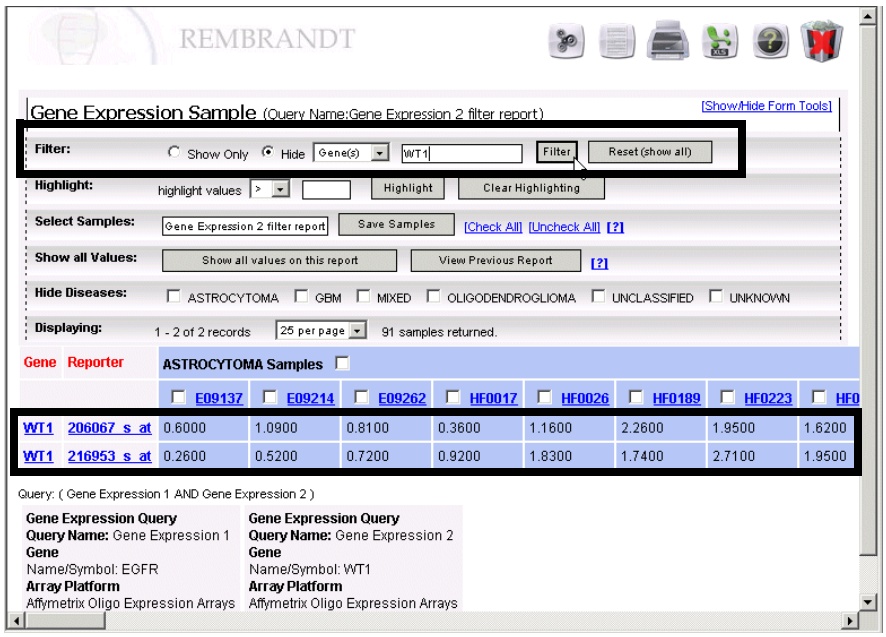


Figure 6.3 Filter toolbar

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 6.4).*

   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 6.4).*

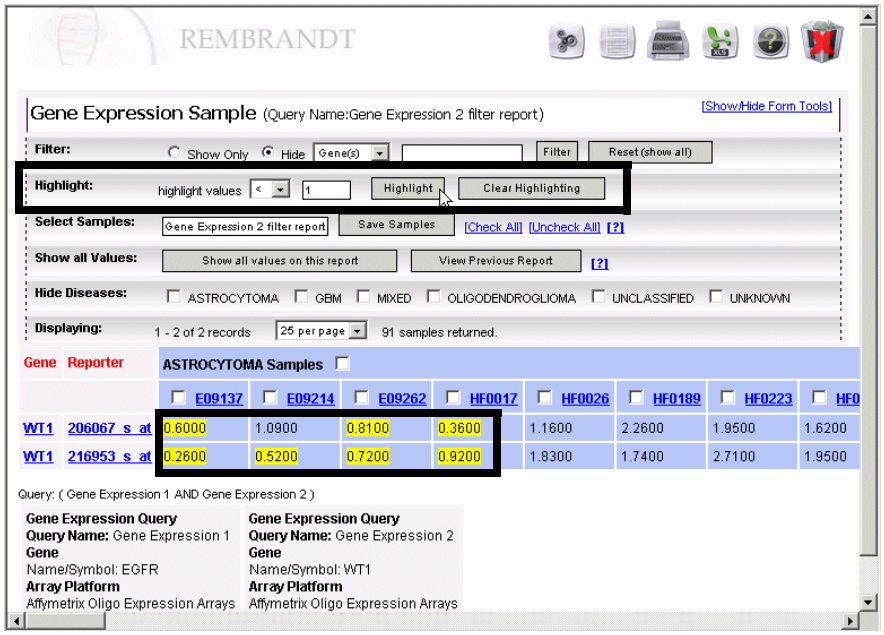


Figure 6.4 Highlight toolbar

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)

To save samples for use with an additional query, 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 6.5* below).

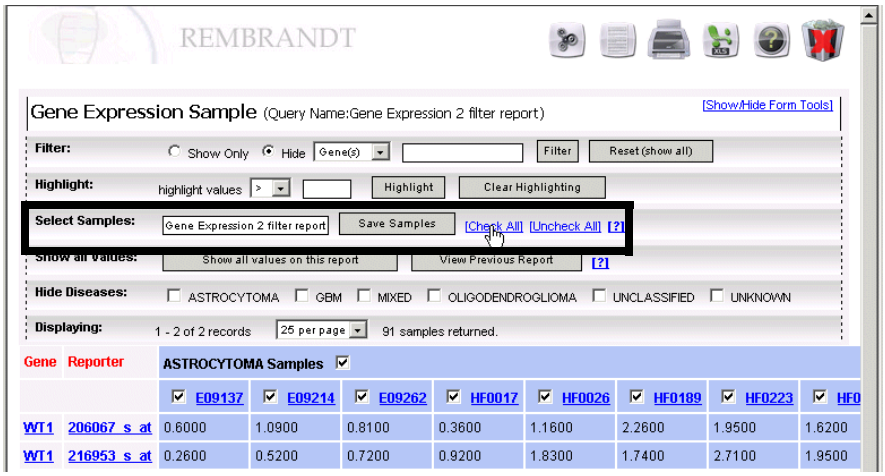


Figure 6.5 Selecting samples from the Select Samples toolbar

   - 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 6.6* below).
   - To select (or deselect) an individual sample within a group, click the box in the column next to the sample name.

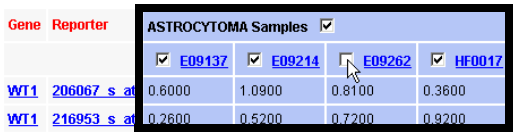


Figure 6.6 Selecting samples from the results

Once you create a sample set, you can save the selected samples to a file and later add the saved sample set to a query (*Figure 6.7*).

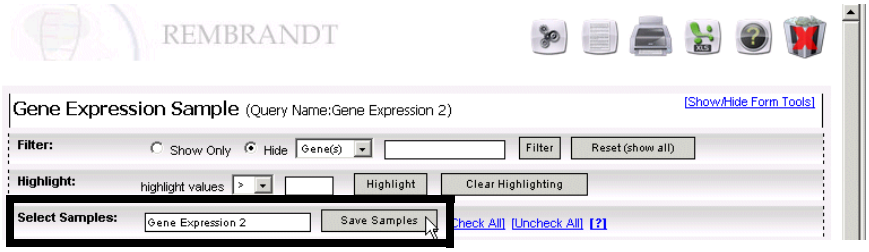


Figure 6.7 Saving the samples

2. Enter a unique name for the sample set. Once saved, this name will 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.
3. Click the **Save Samples** button.

   The samples are saved to on the Clinical window. For more information, see *Viewing the Clinical Reports*.

Differentiating Data (Show All Values Toolbar)

1. To differentiate between missing values in the array and data that did not meet your search criteria, click **Show All Values on this Report** on the **Show all Values** toolbar (*Figure 6.8*).

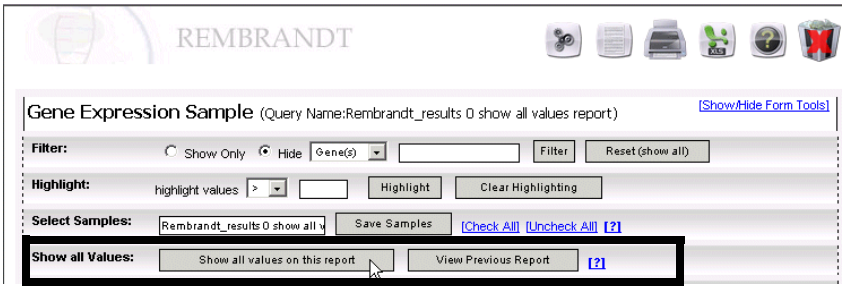


Figure 6.8 Show All Values toolbar

   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 6.9).* The checked disease is NOT included in the results.

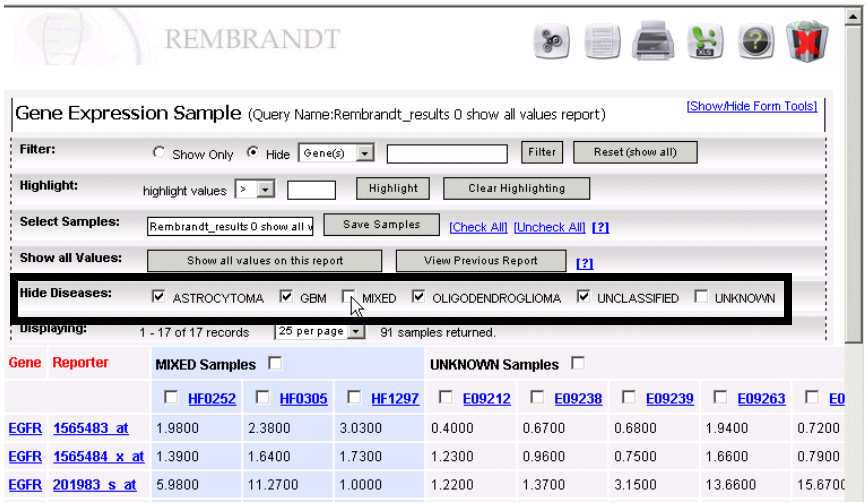


Figure 6.9 Hide Diseases toolbar

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 name link (*Figure 6.10*).

Figure 6.10 The Gene column

The Cancer Genome Anatomy Project (CGAP) browser opens.

Gene Expression Disease Report

The Gene Expression Data Per Disease Group report (*Figure 6.11*) displays the geometric mean of the gene expression (between the tumor group and the average of non-tumor samples) for each probeset (or IMAGE clone). Each column represents a sample group (tumor sub-type). Group average samples were based on tumor subtype categories.

Group average samples are also calculated for each probeset (or IMAGE clone). To indicate probabilities of obtaining the differences in expression values between tumor (or a sub-type of tumor) and non-tumor samples, a *p*-value is displayed within the parenthesis for each geometric value (or ratio).

REMBRANDT

Gene Expression Disease (Query Name:Rembrandt_results 0) [Show/Hide Form Tools]

Displaying: 1 - 17 of 17 records 25 per page

Gene	Reporter	ALL	ASTROCYTOMA	GBM	MIXED	OLIGODENDROGLIOMA	UNKNOWI
EGFR	1565483_at	1.4132 (0.3131)	1.5531 (0.3651)	1.0970 (0.8212)	1.0092 (0.9856)	1.4328 (0.3169)	1.6555 (0.
EGFR	1565484_x_at	1.0859 (0.7558)	1.2381 (0.5986)	1.0706 (0.8487)	1.1183 (0.8092)	1.0123 (0.9624)	1.2487 (0.
EGFR	201983_s_at	5.5714 (0.0001)	7.5474 (0.0001)	8.7763 (0.0001)	3.9829 (0.0197)	5.5562 (0.0001)	3.5853 (0.
EGFR	201984_s_at	2.5500 (0.0181)	2.8944 (0.0195)	3.9795 (0.0063)	1.3135 (0.4850)	2.0406 (0.0141)	2.0437 (0.
EGFR	210984_x_at	7.2062 (0.0096)	7.8788 (0.0076)	18.7115 (0.0022)	17.5672 (0.0073)	5.0133 (0.0264)	3.3919 (0.
EGFR	211550_at	1.2909 (0.1433)	1.2858 (0.2033)	1.4180 (0.1202)	1.5485 (0.0360)	1.3670 (0.0328)	1.1381 (0.
EGFR	211551_at	1.0401 (0.7745)	1.0625 (0.6978)	1.1472 (0.3481)	1.2100 (0.4893)	1.0700 (0.6837)	1.2285 (0.
EGFR	211607_x_at	4.2762 (0.0419)	4.5615 (0.0385)	10.6155 (0.0082)	8.3282 (0.0347)	3.1411 (0.0684)	2.1124 (0.
EGFR	224999_at	4.9085 (0.0002)	6.3872 (0.0003)	7.7482 (0.0001)	3.3047 (0.0215)	4.3614 (0.0004)	3.4496 (0.
EGFR	232120_at	3.8883 (0.0007)	4.2606 (0.0006)	5.9197 (0.0033)	4.4565 (0.0060)	3.4494 (0.0002)	2.7488 (0.
EGFR	232541_at	7.6712 (0.0001)	8.9912 (0.0003)	12.8184 (0.0002)	11.7786 (0.0001)	7.8187 (0.0000)	4.4595 (0.
EGFR	232925_at	15.1563 (0.0000)	16.1096 (0.0007)	31.8798 (0.0000)	34.5296 (0.0002)	15.9857 (0.0000)	7.0976 (0.
EGFR	233044_at	2.2203 (0.0567)	2.3999 (0.0212)	3.1773 (0.0313)	2.2554 (0.0483)	1.9095 (0.1018)	1.7710 (0.
EGFR	237938_at	1.3795 (0.3809)	1.7710 (0.1312)	1.8387 (0.2035)	2.6774 (0.0586)	1.2556 (0.3543)	1.0445 (0.
EGFR	243327_at	1.5840 (0.1665)	1.5354 (0.2638)	2.4703 (0.0288)	2.2994 (0.0118)	1.5303 (0.0656)	1.0717 (0

Figure 6.11 Gene Express Data Per Disease Group report

To show additional information about the resulting gene or reporter, see *Showing Additional Information*.

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 6.12*).

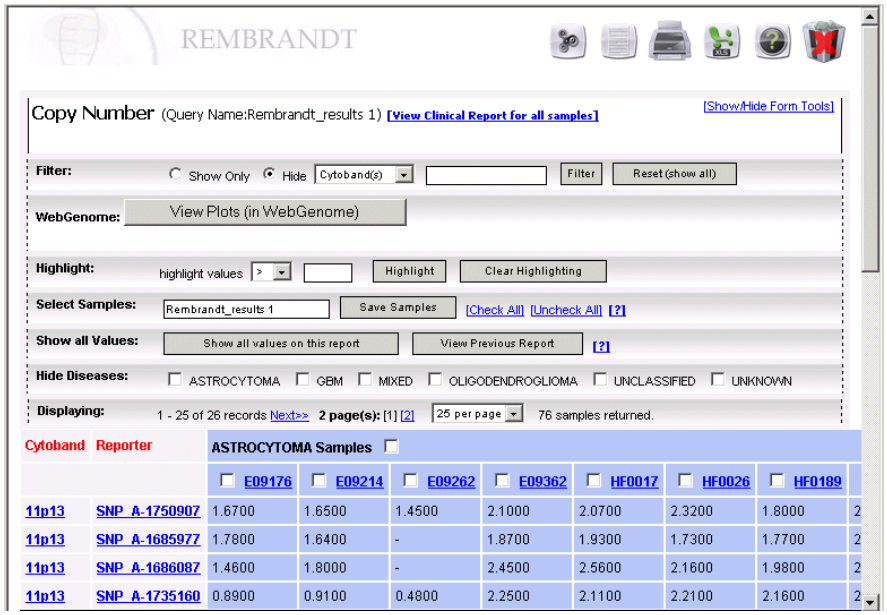


Figure 6.12 Copy Number Data for Sample report

For more information, see to the following:

- *Filtering Results by Gene or Reporter (Filter Toolbar)*
- *Displaying Copy Number Sample Data in webGenome*
- *Highlighting Results By Value (Highlight Toolbar)*
- *Selecting and Saving Sample Results (Select Samples Toolbar)*
- *Differentiating Data (Show All Values Toolbar)*
- *Removing Columns (Hide Diseases toolbar)*
- *Showing Additional Information*

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.
- Annotation Plots show DNA copy number measurements in relation to annotated genome feature, such as genes.
- Annotation Reports show annotation genome features in a tabular format.
- Ideogram Plots show chromosomal amplifications and deletions in relation to cytogenetic chromosome ideograms.
- Probe Plots show measured copy number for selected reporter probes.

Click the **View Plots in webGenome** button (*Figure 6.13*).

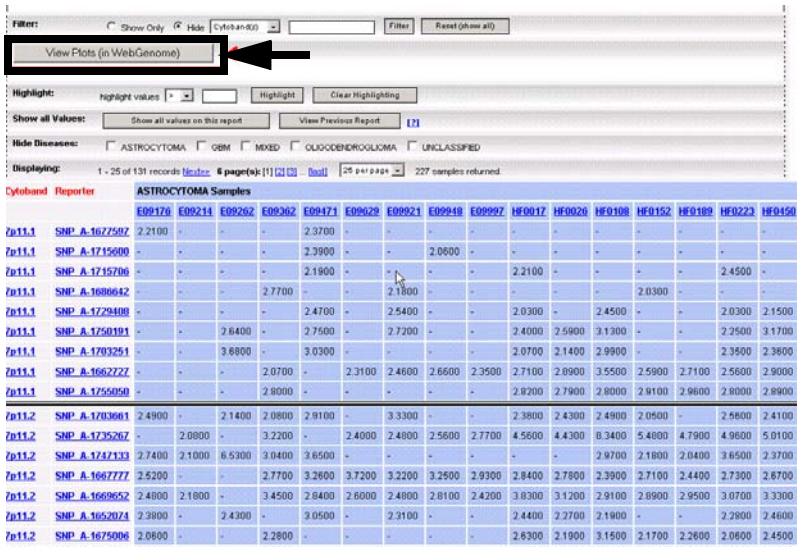


Figure 6.13 Copy Number Report

webGenome opens a new window plotting the Rembrandt Copy Number data.

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 6.1*) displays the query name and lists the output generated for the query.

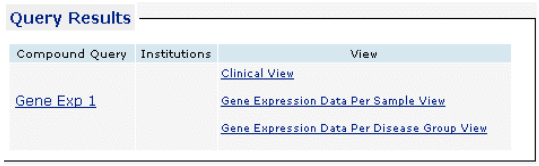


Figure 6.14 Query Results

Class Comparison Report

The Class Comparison report (*Figure 6.15*) 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 a **T-test** or **Wilcox** Statistical Method analysis (*Figure 6.15*), 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.

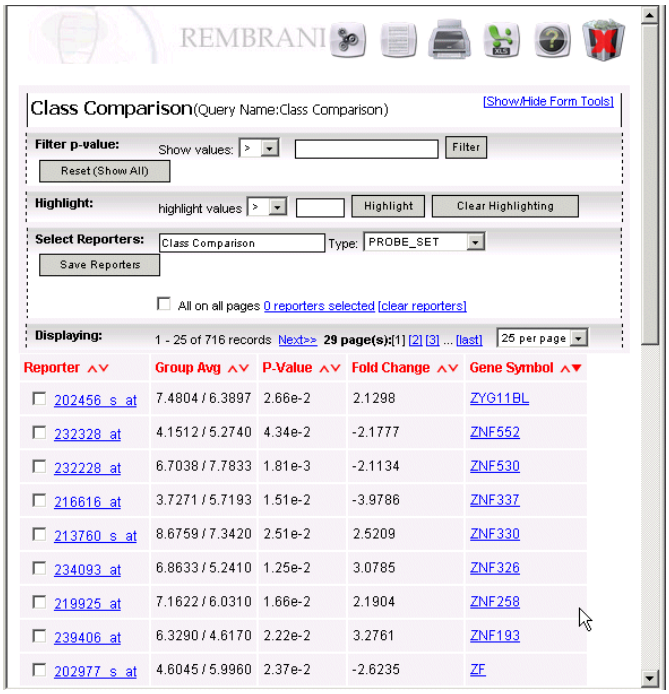


Figure 6.15 Class Comparison report

For an **F-test** Statistical Method analysis (*Figure 6.16*), the difference is that in the results there is a group average displayed for each group.

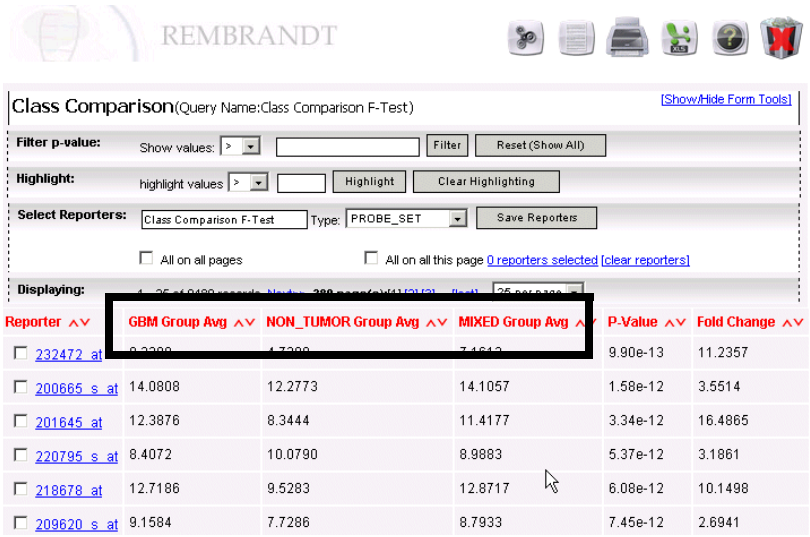


Figure 6.16 Class Comparison report - F-test

For more information, see the following:

- *Filtering a p-value (Filter p-value Toolbar)*
- *Highlighting Results By Value (Highlight Toolbar)*
- *Selecting and Saving Reporters (Select Reporters toolbar)*
- *Resorting Column Results*
- *Showing Additional Information*

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 6.17)*.

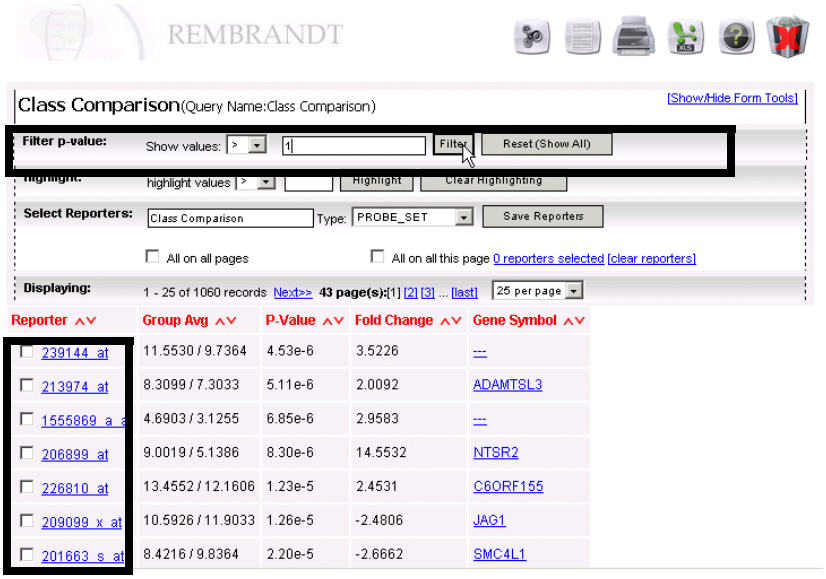


Figure 6.17 Filter toolbar

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)

To select reporters, follow these steps (*Figure 6.18*):

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.

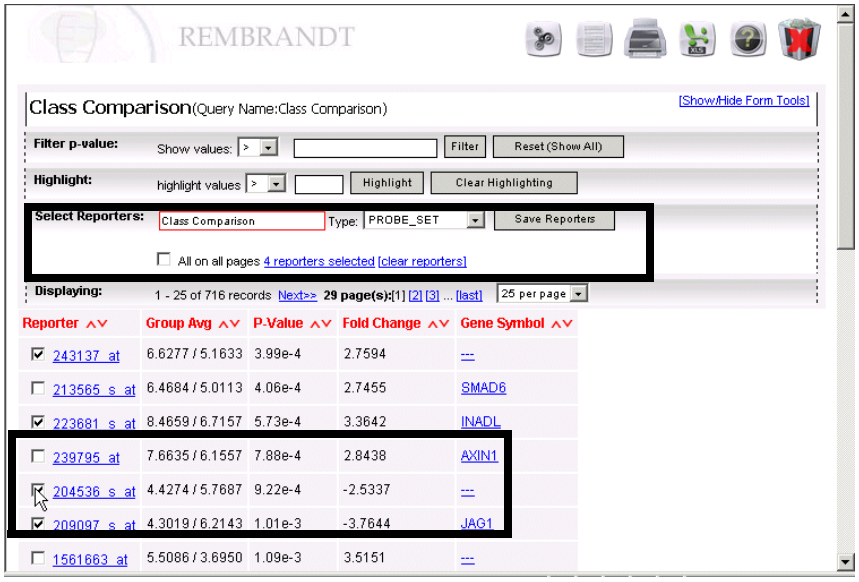


Figure 6.18 Selecting Reporters instructions

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

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

   The results are saved.
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 6.19*).

Figure 6.19 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 6.20*) is a two-dimensional graph which plots the various principal components from the analyses. You can click on the three tabs at the top of the graph to display the following:

- PC1 versus PC2
- PC1 versus PC3
- PC2 versus PC3

Each point on the graph represents a sample. By default, the samples are colored by **Disease Type**. 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.

To view clinical data for the selected sample, click the **view the clinical data** link (see *Viewing Clinical Plots*).

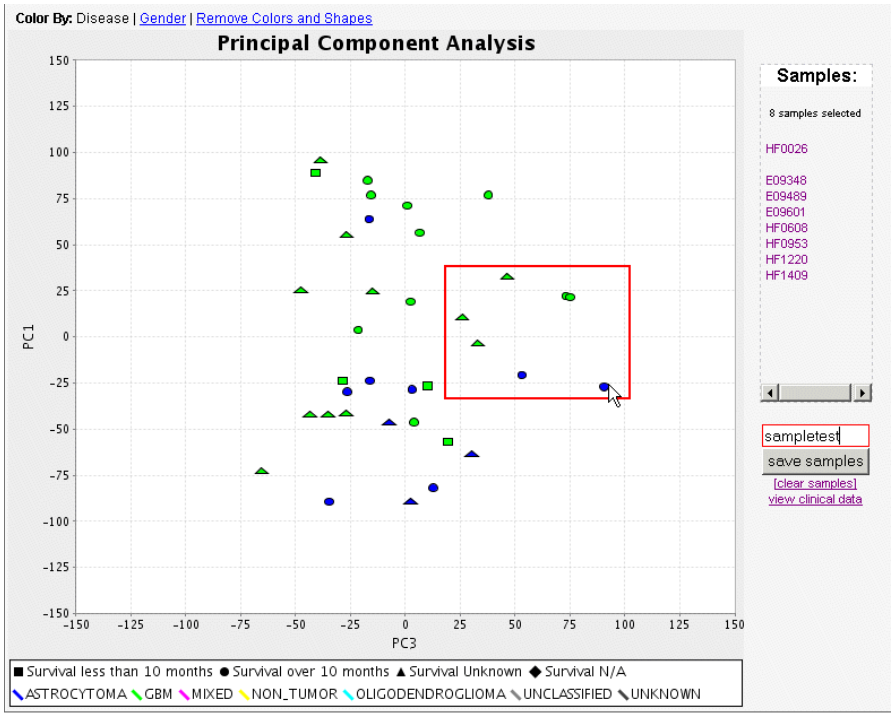


Figure 6.20 Principal Component Analysis report

Hierarchical Clustering Report

The Hierarchical Clustering report (*Figure 6.21*) 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.

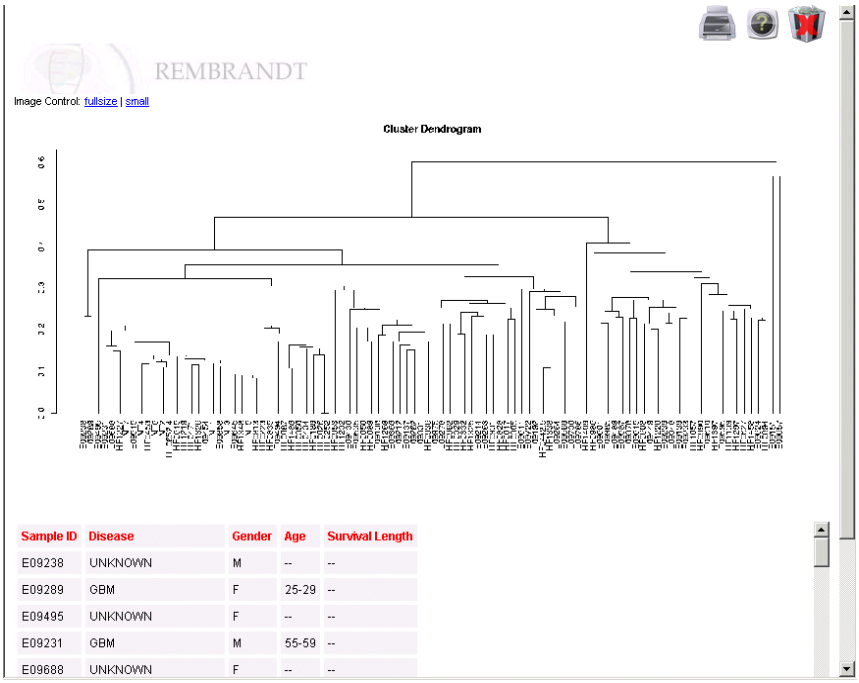


Figure 6.21 Hierarchical Clustering report

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 6.22 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 7

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* on page 69
- *Viewing the Data Items in a List* on page 70
- *Removing Data Items to Create a New List* on page 70
- *Deleting an Entire List* on page 72
- *Adding a New "Custom" List* on page 72
- *Combining Existing Lists to Create a New List* on page 71

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 **Reporters Lists**. With these lists, you can further refine queries or facilitate analysis.

Note: You can 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.

The blue panel displays each list type and the associated lists. You can mouse-over a list and display the data items. Using the Manage List function, you can perform the following functions:

- View data items in a list
- Create new lists from existing lists
- Delete lists
- Add lists by uploading them or typing them

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**, **Report Lists**).

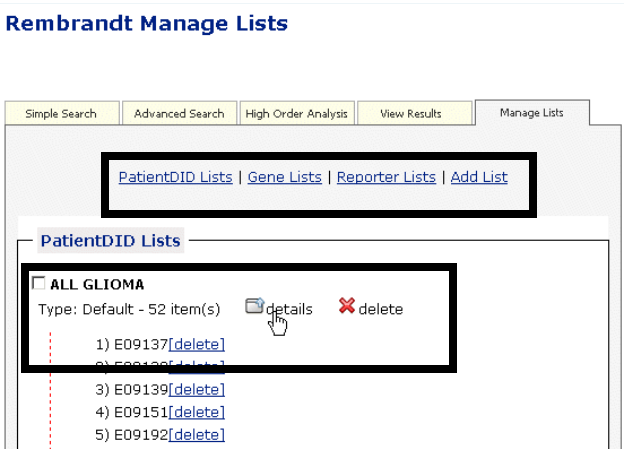


Figure 7.1 List types and Details

2. Find a list to be viewed, and click the **details** icon to display all of the items in the list.

   Note: You can also mouse-over the list name in the blue panel and the list's data item names appear.

Removing Data Items to Create a New List

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**, **Report 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.

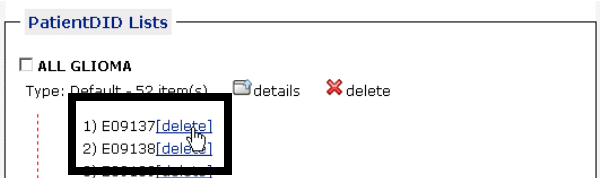


Figure 7.2 Deleting data items

4. Click the **delete** link beside the item you want to delete.The item is removed from the list.

Once you remove the items,you can view the new list or save the list to your computer.

5. Click the **export link** at the bottom of the items list to open and view the new list or save the list on your computer. Click **Open** or **Save**.

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**, **Report List**). The categories for the list appear.
2. Find the category for the new list, and click the box next to the category name. Click more than one box to select multiple categories.

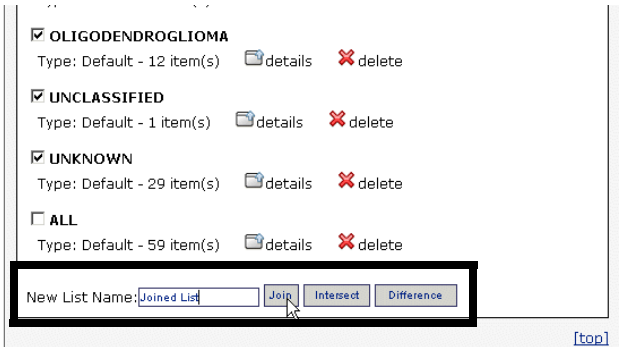


Figure 7.3 Combining existing lists

3. Enter a unique name for the new list you are creating, and then click the appropriate button:
   - **Join** combines two or more categories into a new list.
   - **Intersect** creates a new list from only the items that appear on more than one selected list category.
   - **Difference** creates a new list from items that appear only in one selected list category.

   The new list appears in the blue panel in red.

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**, **Report 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.

Figure 7.4 Deleting an entire list

3. To delete the selected lists, click an **x delete** icon. The selected categories are removed.

Adding a New "Custom" List

You may add a new list type by *uploading* a list from your computer or *manually creating* a list. To add a new 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. To upload a list, follow these steps:
   1. Click **Upload List** at the top of the box.

Figure 7.5 Uploading a list

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

3. To create and add a list manually, follow these steps:
   1. Click **Manually Type List** at the top of the box.

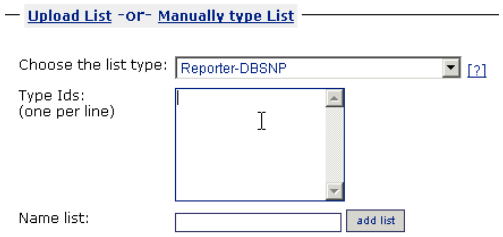


Figure 7.6 Manually typing a list

   2. From the **Choose the list type** drop-down list box, select the list to be uploaded.
   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 appears on the blue panel in red.
   5. To open and view the newly created list or save it to your computer, click on the list name in the blue panel. Click **Open** or **Save**.
6. To open and view the newly created list or save it to your computer, click on the list name in the blue panel. Click **Open** or **Save**.

GLOSSARY

Acronyms and other terms referred to in the chapters of this 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.
anaplastic	Cancer cells that divide rapidly and have little or no resemblance to normal cells.
Astrocytic tumors	Neoplasms of the brain and spinal cord derived from glial cells.
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.
ependymoma	A type of brain tumor that may arise in the ventricles of the brain or in the spinal cord. Also called an ependymal tumor.
False Discovery Rate (FDR)	The expected proportion of Type I errors among rejected hypotheses in simultaneous testing of multiple null hypotheses.
Family-wise Error Rate (FWER)	Denotes the probability of having at least one false significant test result within the set of tested hypotheses.

Table 8.1 Glossary of REMBRANDT terms

Term	Definition
fibrillary astrocytoma	Most frequent histological variant of Diffuse Astrocytoma; predominantly composed of fibrillary neoplastic astrocytes.
gemistocytic astrocytoma	Rare variant of Diffuse Astrocytoma. It is characterized by the presence of a conspicuous, though variable, fraction of gemistocytic neoplastic astrocytes.
Gene Ontology (GO) Classification	The Gene Ontology (GO) project is 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. GO provides three structured networks of defined terms, molecular function, biological process, and cellular component, to describe gene product attributes. (from http://www.geneontology.org)
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	Hierarchical cluster analysis is 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, including class comparison, class discovery and class prediction.

Table 8.1 Glossary of REMBRANDT terms

Term	Definition
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.
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.
myxopapillary ependymoma	Slow growing gliomas which generally occur in young adults and arise almost exclusively in the conus-cauda-filum terminale region. It generally has a favorable prognosis and is characterized histologically by tumor cells arranged in a papillary manner around vascularized mucoid stromal cores.

Table 8.1 Glossary of REMBRANDT terms

Term	Definition
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 begins in the oligodendrocytes (brain cells that provide support and nourishment for nerve cells). Also called an oligodendroglioma.
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	PCA is 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 (i.e. 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.
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.
subependymal giant cell astrocytoma;	Benign, slowly growing tumor (WHO grade I) typically arising in the wall of the lateral ventricles and composed of large ganglioid astrocytes. It is the most common CNS neoplasm in patients with Tuberous Sclerosis Complex and typically occurs during the first two decades of life.
Wilcoxin Test	Nonparametric statistics for testing hypotheses about whether two samples differ.

Table 8.1 Glossary of REMBRANDT terms

INDEX

Symbols

( ) parentheses
  using to group queries 42
+ or - sign
  click for more options 44

A

adding
  advanced search 40
  custom list (uploading or typing) 72
  new list (combining) 71
  new list (removing data items) 70
  query, overview 29
  simple search 9
  user account 5
advanced searches 29
  adding 40
  clinical study 38
  copy number 35
  gene expression 30
  overview 29
  saving samples 55
analyses
  Class Comparison 44
  Hierarchical Clustering 48
  High Order Analysis 43
  Principal Component 46
audience
  of the User's Guide v
Average linkage, defined 49

B

Baseline, assigning 45
blue panel
  copy edit delete searches or queries 41
  description 6
  high order analysis filters 43
  manage list information 69
Box and Whisker Coin plot
  circles, defined 16
  description 16
  triangles, defined 16
Box and Whisker Log2 Intensity option 12
Box and Whisker Log2 Intensity plot
  circles defined 15
  description 15
  triangles, defined 15
  uses for 15
BRB-ArrayTools
  downloading files 66

C

Choose a Saved
  Clone List, defined 31
  Gene List, defined 36
  SNP List, defined 36
circles
  on Box and Whisker Coin plot 16
  on Box and Whisker plot 15
Class Comparison
  Analysis Form,description 44
  High Order Analysis 44
  report description (View Results) 61
Clinical Data page, description 38
Clinical plot
  description 27
  saving samples of interest 28
  selecting samples of interest 27
Clinical report
  description 25
  saving samples 26
  selecting samples 26
Clinical Study
  advanced search 38
clone identifiers
  uploading 31
columns
  hiding diseases 57
combining lists 71
Complete linkage, defined 49
Constrain reporters by variance (Gene Vector)
  percentile, defined 47, 48

copying
queries (Note) 41
Copy Number
advanced search 35
Data page, description 35
page description (View Results) 59
showing data in webGenome 60
Copy Number Data Per Sample report 59
Copy Number Data Per Sample View 59
creating
custom list (uploading or typing) 72
new list (combining) 71
new list (removing data items) 70
query, overview 29
user account 5

D

data items
removing from a list 70
viewing in a list 70
deleting
list 72
queries (Note) 41
Difference option, description 71
Distance Matrix, defined 49
Downloading BRB Files
defined 66

E

editing
queries (Note) 41
e-mailing
NCICB Application Support 5
plots or graphs 17
Euclidean distance, defined 49

F

filtering
p-value (Filter p-value Toolbar) 62
reports 62
reports (Filter toolbar) 53
Filter p-value Toolbar 62
Filter Toolbar 53
Finalize Query button
purpose 29
F-test One Way ANOVA
defined 45
output 62

G

Gene Expression
advanced search 30
Data Per Disease Group View 58
Data Per Sample View 52
Plot page, description 11
simple search 10
Gene Expression Data Per Disease Group report 58
Gene Expression Disease page
description 58
Gene Expression plot
resizing 17
Gene Expression Plot page
Box and Whisker Log2 Intensity plot details 15
Geometric Mean plot details 13
Log2 Intensity plot details 14
Gene Expression Sample page 57
description (View Results) 52
Gene Expression Sample report 52
gene identifiers
uploading 31
Geometric Mean
Gene Expression plot 13
option, defined 12
Graph Type, defined 12
grouping queries 42

H

help
obtaining 7
help link 7
Hide Diseases toolbar 57
Hierarchical Clustering
Analysis Form,description 48
High Order Analysis 48
page description (View Results) 66
report, description 66
highlighting
results by value (Highlight Toolbar) 54
Highlight Toolbar 54
High Order Analysis 43
Class Comparison 44
Hierarchical Clustering 48
Principal Component 46

I

IMAGE clone identifiers
uploading 31
Intensity, defined 13, 14
Intersect option, description 71

J

Join option, description 71

K

Kaplan-Meier for Copy Number-based Data
plot description 21
redrawing plot 21
searching steps 20
Kaplan-Meier for Gene Expression Data
plot description 19
redrawing plot 18
simple search 18
Kaplan-Meier for Sample Data
plot description 23
simple search 23
Karnofsky Clinical Evaluation, defined 39
Karnofsky score, defined 27

L

Lansky Clinical Evaluation, defined 39
launching
REMBRANDT 3
Legal Rules of the Road page 5
Linkage Method, defined 49
lists
adding a custom list 72
combining lists 71
creating a new list 70, 71
deleting 72
removing data items 70
viewing data items 70
Log2 Intensity
Gene Expression plot 14
option 12
logging in
REMBRANDT 5
logging out
REMBRANDT 8
logout link 8
Logout page 8
Log-rank p-value, Statistical Report 23, 25

M

managing lists
overview 69
Manually Type List options, description 73
Mean, defined 16
Median, defined 16
missing
Run Report button 42

O

overview
creating a query 29
high order analyses 43
managing lists 69
simple search 9
viewing output 51

P

parentheses
using to group queries 42
Pearson correlation, defined 49
Please select an All Genes query, defined 42
Principal Component
Analysis Form,description 46
Analysis page, description (View Results) 65
High Order Analysis 46
Principal Component Analysis report, description 65
printing
Gene Expression plot (Simple Search) 17
plots or graphs 17
Probeset, defined 13, 14
purpose
of the User's Guide v
p-value
defined (for Geometric Mean plot) 13
defined (for Log2 Intensity plot) 14
filtering (Filter p-value Toolbar) 62

Q

Q1, defined 16
queries
grouping 42
overview of process 29
refining 41

R

redrawing
K-M Copy Number-based Data 21
K-M Gene Expression plot 18
Refine Query
option, purpose 29
page, description 41
Refine Query page
description 41
REMBRANDT
creating user account 5
description of 1
functions 2
getting help 7
launching 3
logging in 5
logging out 8
menu 7
stands for 1
workspace 6

Rembrandt Manage Lists page
description 69
removing
data items in a list 70
reporters
selecting and saving 63
reports
Class Comparison 61
Copy Number Sample 59
Gene Expression Disease 58
Gene Expression Sample 52
Hierarchical Clustering 66
Principal Component Analysis 65
resizing
Gene Expression plot (in new window) 17
Run Report >> button
missing 42

S

sample
saving (Select Sampes Toolbar 55
saving on Clinical report 26
selecting on Clinical report 26
sample identifiers
uploading 33, 37, 39
samples of interest
saving on Clinical plot 28
selecting on Clinical plot 27
saving
plots or graphs 17
reporters (Select Reporters Toolbar) 63
sample of interest on Clinical plot 28
samples on Clinical report 26
searches
advanced overview 29
simple overview 9
Select a data source, defined 42
selecting
reporters (Select Reporters Toolbar) 63
sample of interest on Clinical plot 27
sample on Clinical report 26
Select Reporters toolbar 63
Select Samples Toolbar 55
Show All Values Toolbar 56
simple searches 9
Gene Expression steps 10
K-M Copy Number-based Data 20
K-M Gene Expression Data 18
K-M Sample Data 23
Single linkage, defined 49
SNP identifiers
uploading 36
sorting
report columns 64
Statistical Method, defined 45
Std. Dev., defined 14
support link 7
Survival vs Age at Dx, defined 27

T

triangles
on Box and Whisker Coin plot 16
on Box and Whisker plot 15
top of columns to sort 64
T-test Two Sample Test
defined 45
output 61
tutorials link 7
typing
a list (Manage List function) 72

U

Unified link, defined 12
uploading
gene identifiers 31
IMAGE clone identifiers 31
list (Manage List function) 72
sample identifiers 33, 37, 39
SNP identifiers 36
User's Guide
audience v
organization of v
purpose v
text conventions vi
user account
creating 5
user guide link 7

V

Validate Query button
defined 42
required 42
viewing
data items in a list 70
View Results page
description 51
views
Copy Number Data Per Sample view 59
Gene Expression Data Per Disease Group view 58
Gene Expression Data Per Sample view 52

W

webGenome
display Copy Number data 60
Wilcoxin Test Man-Whitney Test

defined 45
output 61
workspace
description 6