

# REMBRANDT USER'S GUIDE

*Version 1.5*



NATIONAL<sup>®</sup>  
CANCER  
INSTITUTE

Center for Biomedical  
Informatics and Information

# CREDITS AND RESOURCES

<b>REMBRANDT Development and Management Teams</b>		
<b>Development</b>		<b>Product and Program Management</b>
Subha Madhavan	Dana Zhang	Subha Madhavan
Alex Jiang	Vesselina Bakalov	Anand Basu
Kevin Rosso	Nick Xiao	
Ryan Landy	Jean-Claude Zenklusen	
Himanso Sahni	Yuri Kotliarov	
David Bauer	Gregg Silk	
Huaitian Liu	Hangjiong Chen	
Michael Harris	Shine Jacob	
Ram Bhattaru	Nonna Rabinovich	<b>Documentation</b>
David Hall	Leonie Misquitta	Laura Jackel
Ye Wu	Mervi Heiskanen	Huaitian Liu
Ying Long	Hangjiong Chen	Jill Hadfield
Don Swan		

<b>Contacts and Support</b>	
Application Support	<a href="http://ncicb.nci.nih.gov/NCICB/support">http://ncicb.nci.nih.gov/NCICB/support</a> Telephone: 301-451-4384 Toll free: 888-478-4423



# TABLE OF CONTENTS

<b>Credits and Resources .....</b>	<b>i</b>
<b>About This Guide .....</b>	<b>1</b>
Purpose .....	1
Audience .....	1
Topics Covered .....	1
Text Conventions Used .....	2
<b>Chapter 1</b>	
<b>Getting Started with REMBRANDT .....</b>	<b>3</b>
About REMBRANDT .....	4
How to Cite REMBRANDT Data .....	4
Launching REMBRANDT .....	4
REMBRANDT's Opening Page .....	6
Logging In .....	7
Accepting REMBRANDT Provisions .....	7
REMBRANDT Menu .....	8
REMBRANDT Tabs .....	9
REMBRANDT Side Bar .....	9
Application Support .....	10
Logging Out .....	10
<b>Chapter 2</b>	
<b>Conducting a Simple Search .....</b>	<b>13</b>
Simple Search Overview .....	13
Gene Expression Simple Search .....	14
Eliminating Aliases .....	14
Understanding a Gene Expression Plot .....	15
K-M Gene Expression Simple Search .....	24
Redrawing the K-M Survival Plot for Gene Expression Data .....	25
Understanding K-M Survival Plot for Gene Expression Data .....	25
K-M Copy Number Simple Search .....	28

Redrawing the K-M Survival Plot for Copy Number Data .....	29
Understanding K-M Survival Plot for Copy Number Data .....	29
K-M Sample Search .....	31
Understanding K-M Survival Plot for Sample Data .....	31

## **Chapter 3**

### **Conducting Advanced Searches .....33**

Advanced Searches Overview .....	33
Gene Expression Advanced Search .....	34
Copy Number Advanced Search .....	39
Clinical Study Advanced Search .....	42
Managing Individual and Compound Queries .....	45
Refining a Query .....	46

## **Chapter 4**

### **High Order Analysis .....49**

High Order Analysis Overview .....	49
Performing a Class Comparison .....	50
Performing a Principal Component Analysis .....	52
Performing Hierarchical Clustering Analysis .....	54
Performing a GenePattern Analysis .....	55
Launching GenePattern .....	57

## **Chapter 5**

### **Viewing Results .....59**

Results Overview .....	59
Notification of Prolonged Queries .....	60
Clinical Reports .....	60
Viewing Clinical Plots .....	63
Advanced Search or Query Results .....	64
Gene Expression Sample Report .....	65
Copy Number Sample Report .....	71
High Order Analysis Results .....	73
Class Comparison Report .....	73
Principal Component Analysis Plot .....	77
Hierarchical Clustering Report .....	80
GenePattern Analysis Reports .....	80

## **Chapter 6**

### **Downloading Data .....87**

Data Download Overview .....	87
Downloading caArray Tools and Files .....	87
Downloading BRB Array Tools and Files .....	88

**Chapter 7**

**Managing Your Workspace .....91**

    Managing Lists Overview ..... 91

        Adding New Lists ..... 92

        Viewing the Data Items in a List ..... 95

        Removing Data Items ..... 96

        Deleting an Entire List ..... 97

    Organizing Existing Lists and Queries ..... 97

    Importing Lists and Queries ..... 98

    Exporting Lists and Queries ..... 99

**Appendix A**

**Glossary .....101**

**Index .....105**



# ABOUT THIS GUIDE

This section introduces you to the *REMBRANDT 1.5 User's Guide*. It includes the following topics:

- [Purpose](#) on page 1
- [Audience](#) on page 1
- [Topics Covered](#) on page 1
- [Text Conventions Used](#) on page 2

## Purpose

---

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

## Audience

---

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

## Topics Covered

---

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

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



- [Chapter 7](#) describes how to manage user-defined or study-defined patient ID, gene, and reporters lists.

## Text Conventions Used

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

<b>Convention</b>	<b>Description</b>	<b>Example</b>
<b>Bold</b>	Highlights names of option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons.	Click <b>Search</b> .
<a href="#">URL</a>	Indicates a Web address.	<a href="http://domain.com">http://domain.com</a>
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.
<i>Italics</i>	Highlights references to other documents, sections, figures, and tables.	See <i>Figure 4.5</i> .
<b><i>Italic boldface monospace type</i></b>	Represents text that you type.	In the <b>New Subset</b> text box, enter <b><i>Proprietary Proteins.</i></b>
<b>Note:</b>	Highlights information of particular importance	<b>Note:</b> This concept is used throughout the document.
{ }	Surrounds replaceable items.	Replace {last name, first name} with the Principal Investigator's name.

# GETTING STARTED WITH REMBRANDT

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

Topics in this chapter include:

- [\*About REMBRANDT\*](#) on page 4
- [\*Launching REMBRANDT\*](#) on page 4
- [\*REMBRANDT's Opening Page\*](#) on page 6
- [\*Provide the REMBRANDT team with any feedback about the product with the Provide us your feedback link.New User Registration\*](#) on page 6
- [\*Logging In\*](#) on page 7
- [\*REMBRANDT Menu\*](#) on page 8
- [\*REMBRANDT Tabs\*](#) on page 9
- [\*REMBRANDT Side Bar\*](#) on page 9
- [\*Application Support\*](#) on page 10
- [\*Logging Out\*](#) on page 10

## 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. Researchers can also perform various higher-order analyses on clinical and genomic datasets.

### How to Cite REMBRANDT Data

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

*For Example:*

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

## Launching REMBRANDT

---

To launch REMBRANDT, follow these steps:

1. Go to the REMBRANDT portal on the NCICB website:  
<http://rembrandt.nci.nih.gov/>.
2. Click the **REMBRANDT Application** button located in the lower left-hand blue column (*Figure 1.1*).

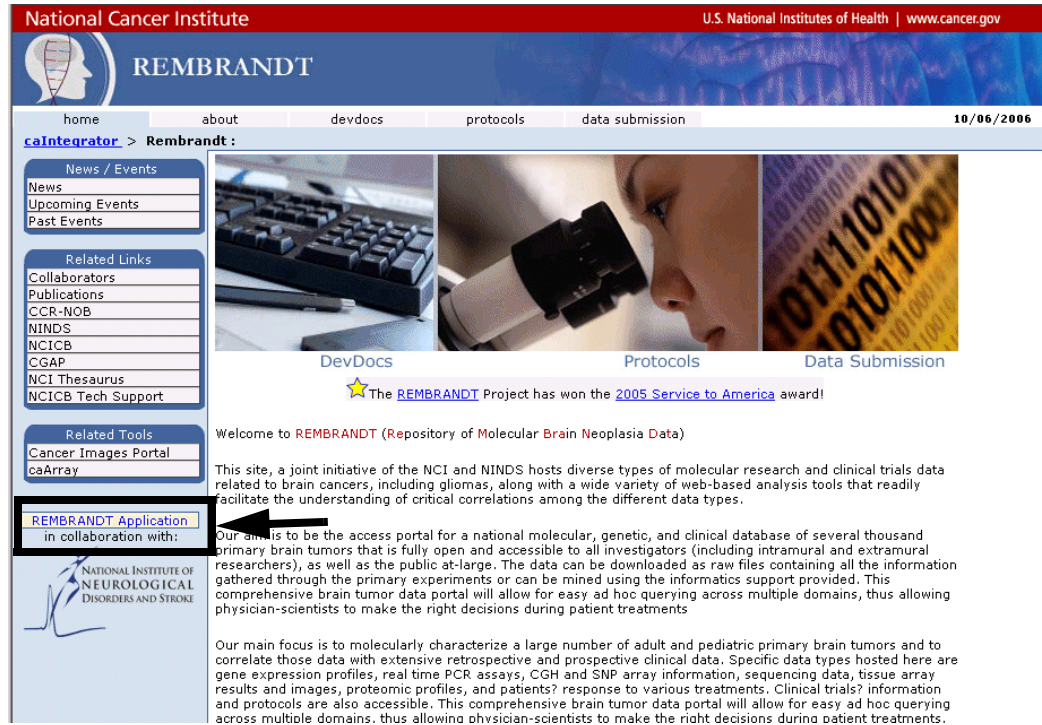


Figure 1.1 The REMBRANDT portal on the NCICB website

The REMBRANDT login page appears (Figure 1.2).

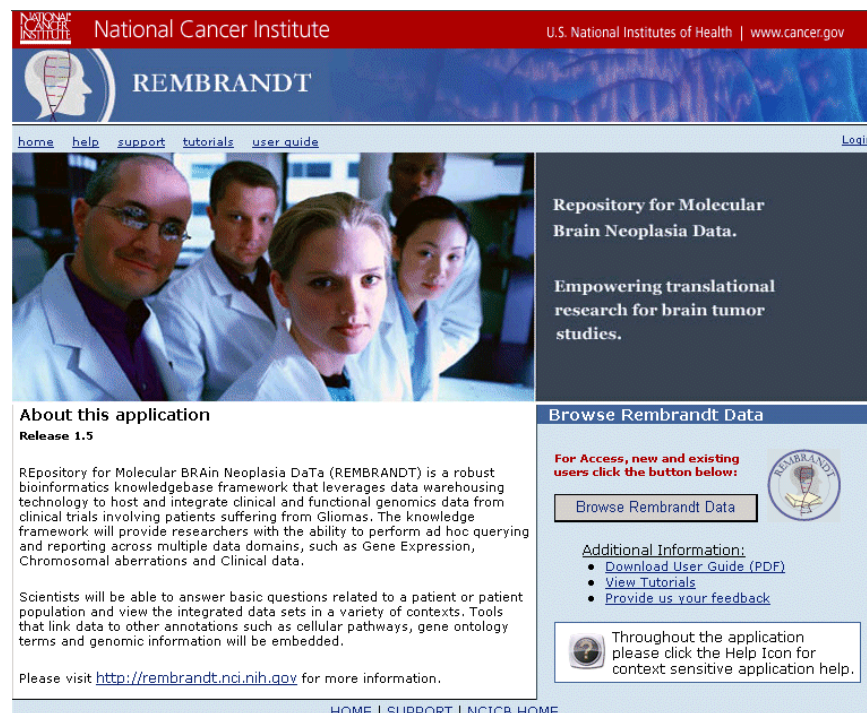


Figure 1.2 REMBRANDT opening page

## REMBRANDT's Opening Page

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

- Register new users and log in current users with the **Browse REMBRANDT Data** button.
- Download an online version of the REMBRANDT 1.5 User's Guide with the **Download User Guide (PDF)** link.
- Run REMBRANDT tutorials that will assist you in using new REMBRANDT options with the **View Tutorials** link.
- Provide the REMBRANDT team with any feedback about the product with the **Provide us your feedback** link.

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

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

**New Users**

Register for an account to gain instant access to public data  
Please enter the text displayed in the image again. Refresh the  
page if you would like another image or email  
ncicb@pop.nci.nih.gov for assistance

\* required field

**Name:**

First name\*:

Last name\*:

**Contact Information:**

Email\*:

- Your account information will be sent to this address

Phone\*:

Institution\*:

Department:

**Verification:**

Image\*:

- Please type the text below, displayed in the image above

Figure 1.3 New User registration

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

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

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

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

4. If you successfully register, the Logging In panel fills in with a temporary username and password that you can use immediately.
5. You should receive an e-mail registration confirmation and then an additional e-mail containing your new account information. Once you receive your username and password, do not use the temporary account.

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

## Logging In

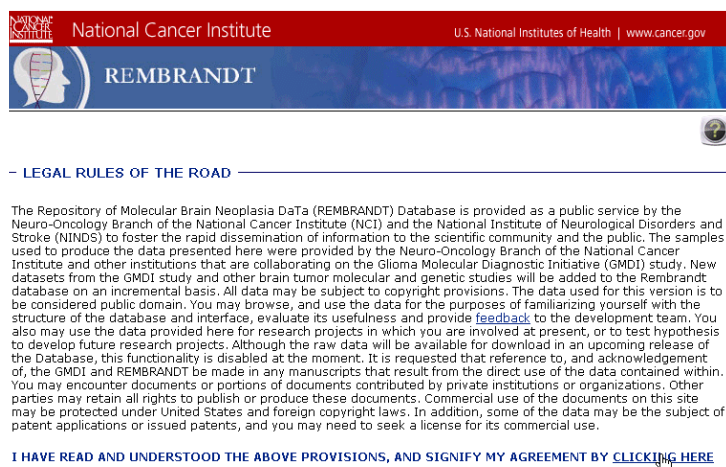
To log into REMBRANDT, you need a username and password. You should have received an e-mail with this information once you registered.

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

**Note:** From the Registration/Login page, you can join (or remove yourself from) the Rembrandt List Serve. Enter the appropriate information in the lower left corner of the page.

## Accepting REMBRANDT Provisions

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



*Figure 1.4 Legal Rules of the Road page*

The REMBRANDT workspace appears (*Figure 1.5*).

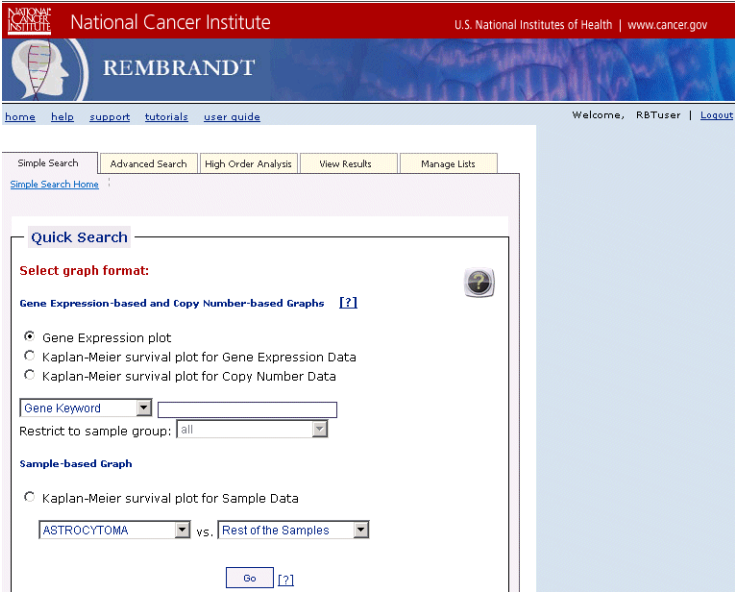


Figure 1.5 The REMBRANDT workspace

## REMBRANDT Menu

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

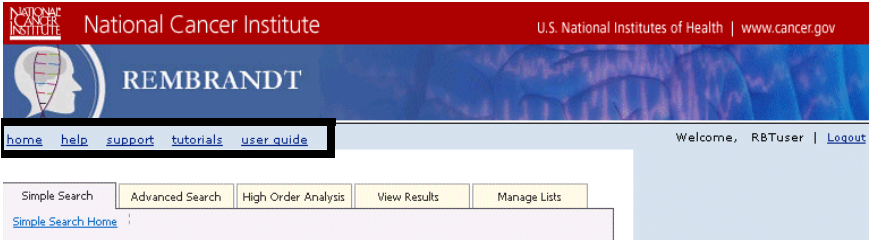


Figure 1.6 REMBRANDT's menu

Table 1.1 describes each item on the REMBRANDT menu.


Menu Option	Function
help	Click to access a complete version of online REMBRANDT help. For complete page-level help, click  on any REMBRANDT page. To open a help page to the field's information, click <a href="#">[?]</a>
support	Click to obtain support for REMBRANDT.
tutorials	Click to REMBRANDT access online tutorials to walk you through REMBRANDT.
user guide	Click to access a pdf version of the REMBRANDT 1.5 User's Guide.

Table 1.1 Getting help with REMBRANDT



## REMBRANDT Tabs

Users can perform a variety of tasks in REMBRANDT. [Table 1.2](#) describes each REMBRANDT tab on the workspace.

<b>Tab Name</b>	<b>Function</b>
Simple Search	Search the database and view the following search results: <ul style="list-style-type: none"> <li>• Gene Expression plots</li> <li>• Kaplan-Meier Survival plots</li> </ul> For more information, see <a href="#">Simple Search Overview</a> )
Advanced Search	Create the following types of queries and group them to generate results for compound queries: <ul style="list-style-type: none"> <li>• Gene Expression analysis</li> <li>• Copy Number Data analysis</li> <li>• Clinical Study analysis</li> </ul> For more information, see <a href="#">Advanced Searches Overview</a> .
High Order Analysis	Run higher order analyses, including class comparisons, hierarchical clustering, and principal component analyses. For more information, see <a href="#">High Order Analysis Overview</a> .
View Results	View Advanced Search and High Order Analysis results. Also download static, archive files for use in BRB-ArrayTools. For more information, see <a href="#">Results Overview</a> .
My Workspace	Manage user- or study-defined patient identifier, gene, or reporter lists. For more information, see <a href="#">Managing Lists Overview</a> on page 91).

Table 1.2 REMBRANDT tabs

## REMBRANDT Side Bar

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

<b>Information Displayed</b>	<b>Function</b>
Filter Settings	Displays the filter settings for the following: <ul style="list-style-type: none"> <li>• A Principal Component Analysis (see <a href="#">Performing a Principal Component Analysis</a> on page 52)</li> <li>• A Hierarchical Clustering Analysis (see <a href="#">Performing Hierarchical Clustering Analysis</a> on page 54)</li> </ul>
Queries	Lists queries created with the Advanced Queries function and enables you to add, copy, edit, and delete existing queries. See <a href="#">Managing Individual and Compound Queries</a> on page 45.

Table 1.3 Getting help with REMBRANDT



Information Displayed	Function
PatientDID List	Displays the default <a href="#">PatientDID lists</a> provided with REMBRANDT, and displays in red any PatientDID lists added to REMBRANDT. See <a href="#">Adding New Lists</a> on page 92.  To display the items in any type of list, hover over the name and a popup displays the data items. To export a list to a spreadsheet file, double-click the list name.
Gene List	Displays in red any <a href="#">Gene lists</a> added to REMBRANDT. See <a href="#">Adding New Lists</a> on page 92.
Reporter List	Displays in red any <a href="#">Reporter lists</a> added to REMBRANDT. See <a href="#">Adding New Lists</a> on page 92.

Table 1.3 Getting help with REMBRANDT

## Application Support

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

Email: <a href="mailto:ncicb@pop.nci.nih.gov">ncicb@pop.nci.nih.gov</a>	When submitting support requests via email, please include: <ul style="list-style-type: none"> <li>Your contact information, including your telephone number.</li> <li>The name of the application/tool you are using</li> <li>The URL if it is a Web-based application</li> <li>A description of the problem and steps to recreate it.</li> <li>The text of any error messages you have received</li> </ul>
Application Support URL	<a href="http://ncicb.nci.nih.gov/NCICB/support">http://ncicb.nci.nih.gov/NCICB/support</a>
Telephone: 301-451-4384 Toll free: 888-478-4423	Telephone support is available: Monday to Friday, 8 am – 8 pm Eastern Time, excluding government holidays.

## Logging Out

To log out of REMBRANDT, follow these steps.

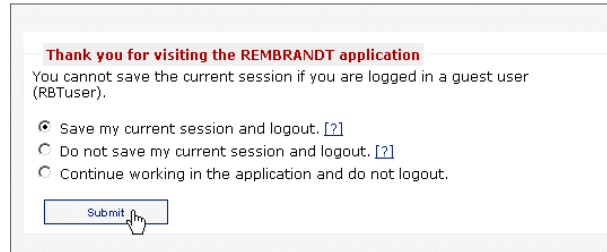
- On the REMBRANDT workspace, click the **logout** link in the upper right-hand corner.



Figure 1.7 Logout link

The Logout page appears.

### Logout



**Thank you for visiting the REMBRANDT application**

You cannot save the current session if you are logged in a guest user (RBTuser).

☒ Save my current session and logout. [?]

☐ Do not save my current session and logout. [?]

☐ Continue working in the application and do not logout.

Figure 1.8 Logout link

Select one of the following options:

- To return to REMBRANDT, select **Continue working in the application and do not logout.**
  - To log out of REMBRANDT *without* saving the session, select **Do not save my current session and logout.**
  - To log out and save your session, select **Save my current session and logout.**
2. To fill out a three-question survey, click **Click Here to take a quick feedback survey.** Answer the questions.
  3. Click the **Submit** button.



## CHAPTER 2

# CONDUCTING A SIMPLE SEARCH

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

Topics in this chapter include:

- [Simple Search Overview](#) on page 13
- [Gene Expression Simple Search](#) on page 14
- [K-M Gene Expression Simple Search](#) on page 24
- [K-M Copy Number Simple Search](#) on page 28
- [K-M Sample Search](#) on page 31

### 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.

---

**Note:** It is not possible to save queries launched as simple searches in Rembrandt. For information about saving advanced queries, see [Advanced Searches Overview](#) on page 33.

---

## 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 2.1](#)).

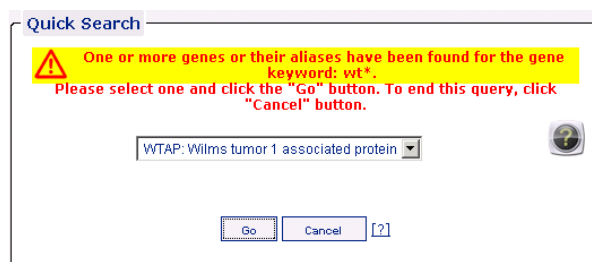


Figure 2.1 Alias message

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

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

# Understanding a Gene Expression Plot

When you perform a Gene Expression simple search, by default the **Median** Gene Expression Plot (*Figure 2.2*) appears.

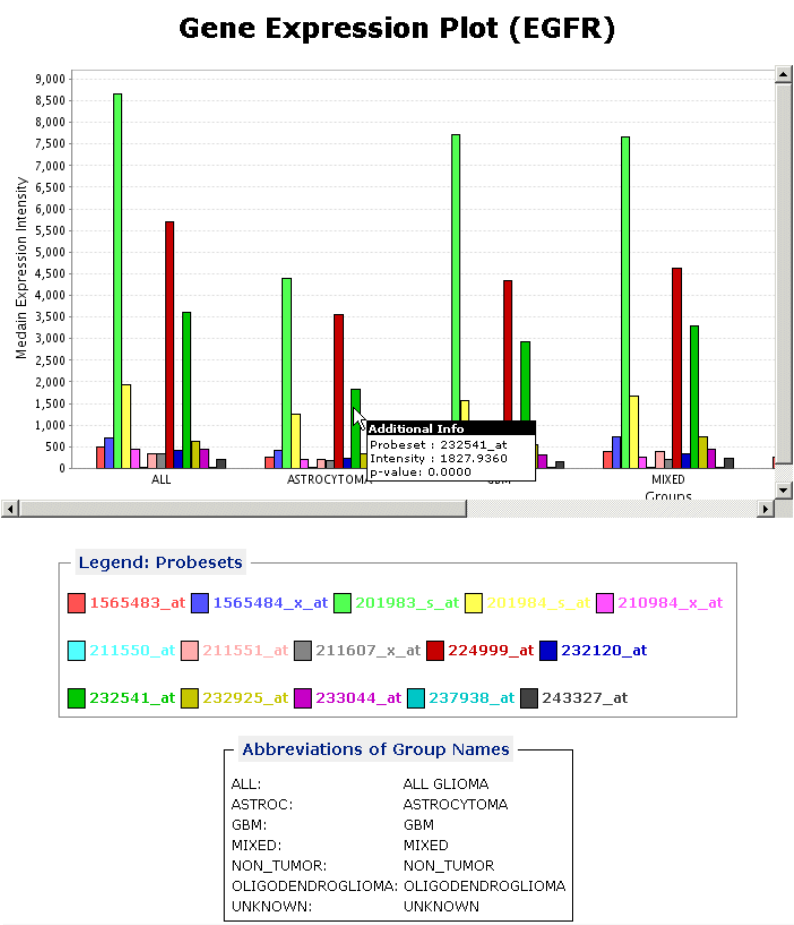


Figure 2.2 Median Gene Expression Plot page

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

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

Table 2.1 Understanding the Gene Expression Plot page

<b>Item</b>	<b>Special Instructions</b>
<b>Graph Type</b>	<p>Displays different versions of the Gene Expression Plot.</p> <ul style="list-style-type: none"> <li>• <b>Median</b> is the default graph shown when you perform a simple search. For additional graph details, see <a href="#">Median Plot Details</a>.</li> <li>• <b>Geometric Mean</b> displays mean expression intensity (Geometric mean) versus Groups. For additional graph details, see <a href="#">Geometric Mean Plot Details</a>.</li> <li>• <b>Log2 Intensity</b> displays average expression intensities for the gene of interest. For additional graph details, see <a href="#">Log2 Intensity Gene Expression Plot Details</a>.</li> <li>• <b>Box and Whisker Log2 Intensity</b> displays a Box and Whisker plot or box plot. For additional graph details, see <a href="#">Box and Whisker Log2 Intensity Gene Expression Plot Details</a>.</li> </ul>
<b>Click here to open plot in a new window</b>	<p>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.</p> <p>See <a href="#">Saving, Printing, and E-mailing a Gene Expression Plot</a>.</p>
<b>Legend Probesets</b>	Indicates the color for each probeset appearing in the graph.
<b>Abbreviations of Group Names</b>	Lists the complete name of each group abbreviation in the plot.
<b>Print this Graph</b>	Click to print the graph.

Table 2.1 Understanding the Gene Expression Plot page

## Median Plot Details

The **Median** Gene Expression Plot ([Figure 2.3](#)) displays the median expression versus Groups.

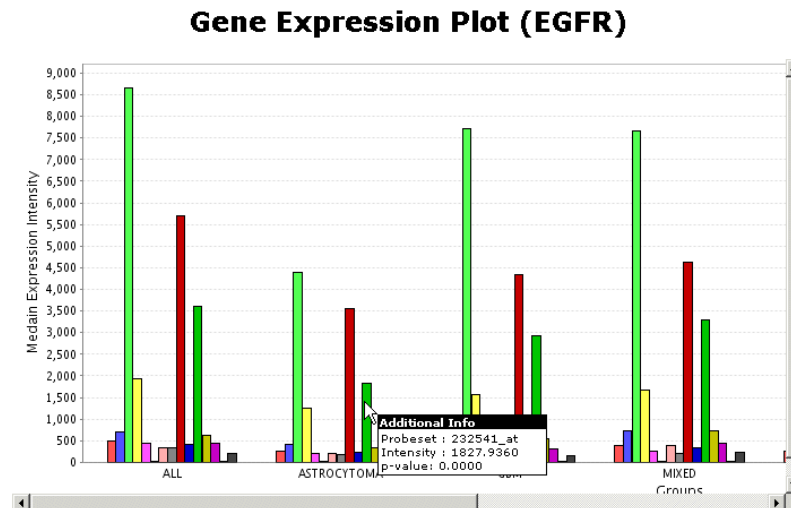


Figure 2.3 Median Gene Expression Plot

Mouse-over a bar on the graph to display Additional Information. [Table 2.3](#) describes Additional Information details.

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

Table 2.2 Median Gene Expression Plot Additional Information



## Geometric Mean Plot Details

The **Geometric Mean** Gene Expression Plot ([Figure 2.3](#)) displays mean expression intensity (Geometric mean) versus Groups.

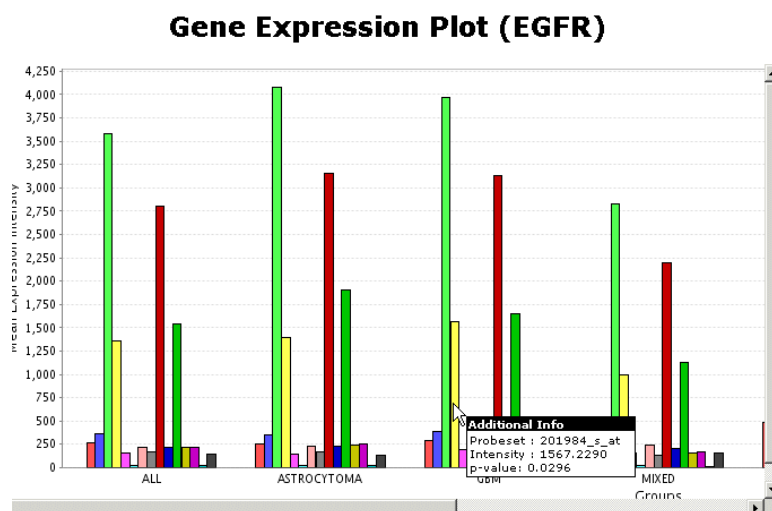


Figure 2.4 Geometric Mean Gene Expression Plot

Mouse-over a bar on the graph to display Additional Information. [Table 2.3](#) describes Additional Information details.

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

Table 2.3 Geometric Mean Gene Expression Plot Additional Information

## Log2 Intensity Gene Expression Plot Details

The **Log2 Intensity** Gene Expression Plot ([Figure 2.5](#)) displays average expression intensities for the gene of interest based on Affymetrix GeneChip arrays (U133 Plus 2.0 arrays). Multiple probesets (for some genes) are designed to measure the expression of the gene of interest. For more information on the probeset design strategy for human genes, go to <http://www.affymetrix.com>.

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

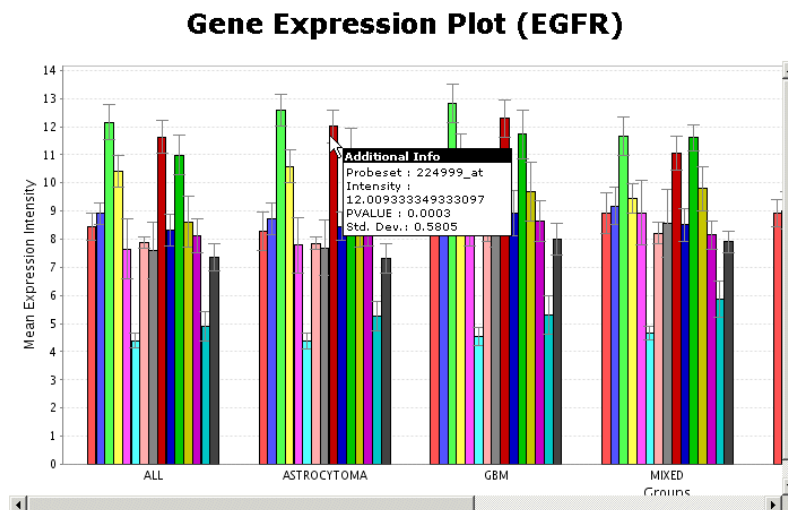


Figure 2.5 Log2 Intensity Gene Expression plot

Mouse-over a bar on the graph to display Additional Information. [Table 2.4](#) describes Additional Information details.

Item	Description
<b>Probeset</b>	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.
<b>Intensity</b>	The mean value calculated for each comparison group.
<b>p-value</b>	The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples.
<b>Std. Dev.</b>	The standard deviation value of a comparison group, such as GBM or Astrocytoma, for a particular probeset or gene. Standard deviation is a statistical measure of spread or variability.

Table 2.4 Log2 Intensity Gene Expression Plot Additional Information

## Box and Whisker Log2 Intensity Gene Expression Plot Details

The **Box and Whisker Log2 Intensity** Gene Expression Plot ([Figure 2.6](#)) displays a box plot without all the individual data points. Example uses of box and whisker plots include the following:

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

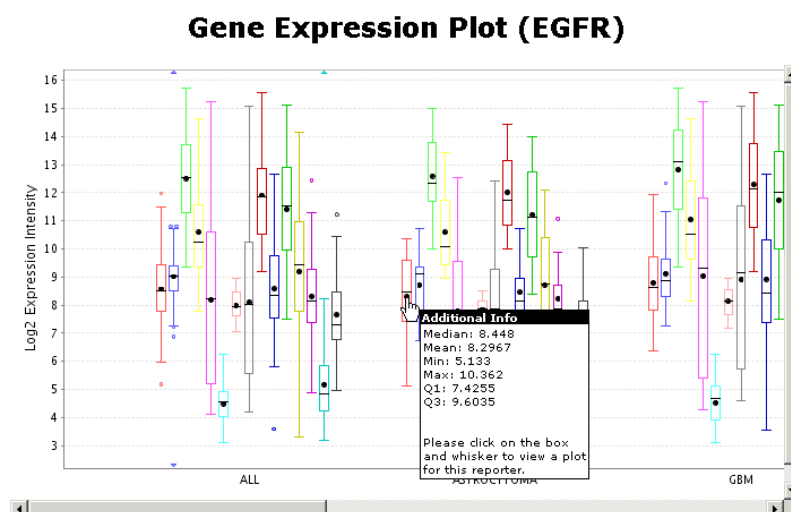


Figure 2.6 Box and Whisker Log2 Intensity Gene Expression plot

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

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

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

Item	Description
...plot	Represents the probeset name.

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

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

In the box-and-whisker plot, the individual probeset summary is represented as follows ([Figure 2.7](#)). Horizontal lines (the "whiskers") extend to, at the most, 1.5 times the box length (the interquartile range) from either or both ends of the box. They end at an observed value, thus connecting all the values outside the box that are not more than 1.5 times the box width away from the box.

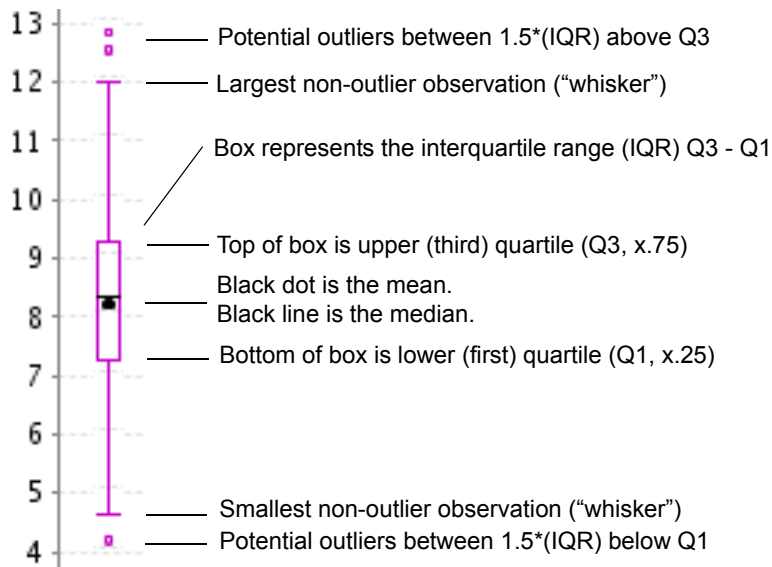


Figure 2.7 Box and Whisker Plot details

### Displaying a Coin Plot

A *coin plot* is box-and-whisker plot ([Figure 2.8](#)) with all individual data points. This enables you to obtain a diagram representing a statistical summary of the data without the disadvantage of concealing the real data.

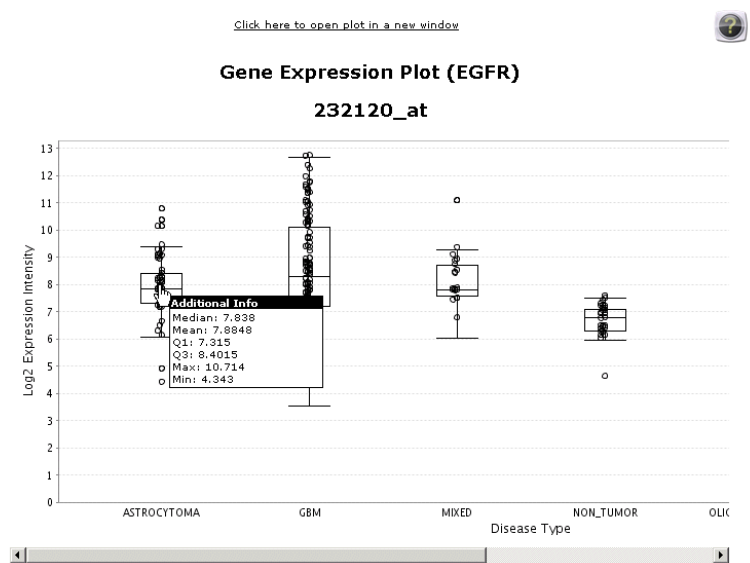


Figure 2.8 Coin Plot for a Probeset

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

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

Table 2.6 Coin Plot Additional Information

In the coin plot, the individual probeset summary is represented as follows ([Figure 2.9](#)). Horizontal lines (the "whiskers") extend to, at the most, 1.5 times the box length (the interquartile range) from either or both ends of the box. They end at an observed value, thus connecting all the values outside the box that are not more than 1.5 times the box width away from the box.

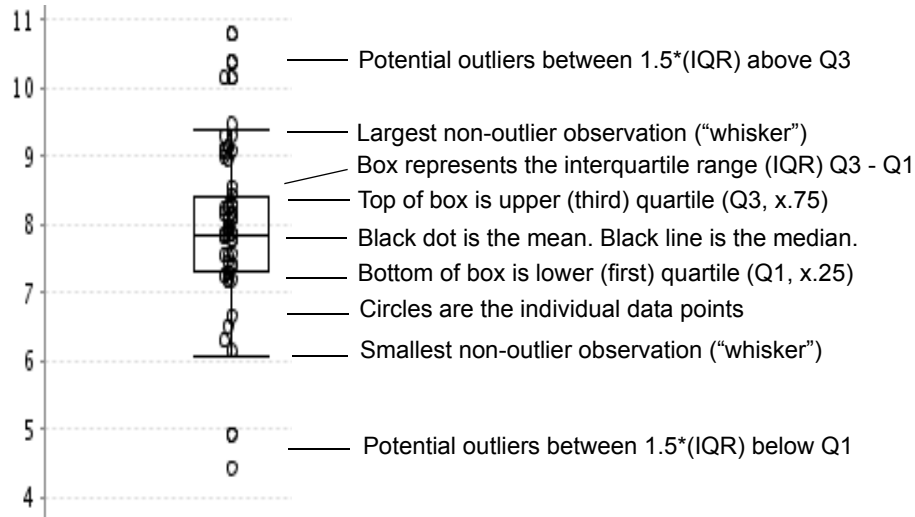


Figure 2.9 Coin Plot details

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

By opening a Gene Expression plot in a new window ([Figure 2.10](#)), you can perform a number of tasks with the Gene Expression plot.

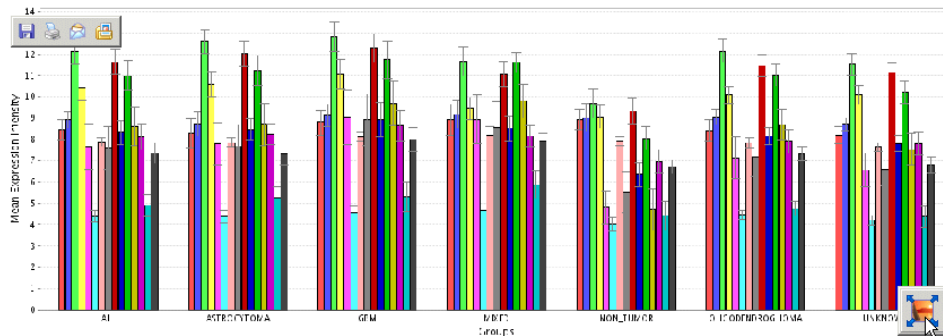




Figure 2.10 Displaying a Gene Expression Plot in a new window

[Table 2.7](#) describes the tasks you can perform when you open a Gene Expression plot in a new window.

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

*Table 2.7 A Gene Expression Plot in a new window*

## K-M Gene Expression Simple Search

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

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

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

3. From the **Restrict to Sample Group** drop-down list, select a saved sample group.
4. Click the **Go** button.

## Redrawing the K-M Survival Plot for Gene Expression Data

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

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

1. To dynamically modify the fold change thresholds and redraw the plot, adjust the **Up-Regulated** and **Down-Regulated** values.
2. Specify a **Unified** or a **Affymetrix Reporter Type**.
3. To visualize the K-M plot for the unified probeset, click the **Reporters** drop-down list ([Figure 2.11](#)).

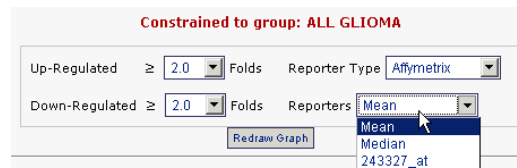


Figure 2.11 Redrawing a Kaplan-Meier Gene Expression data

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

## Understanding K-M Survival Plot for Gene Expression Data

A K-M Survival Plot for Gene Expression Data ([Figure 2.13](#)) 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.

**Note:** Samples for which there is no survival information and censor status are excluded from the Kaplan Meier Plot Analysis.

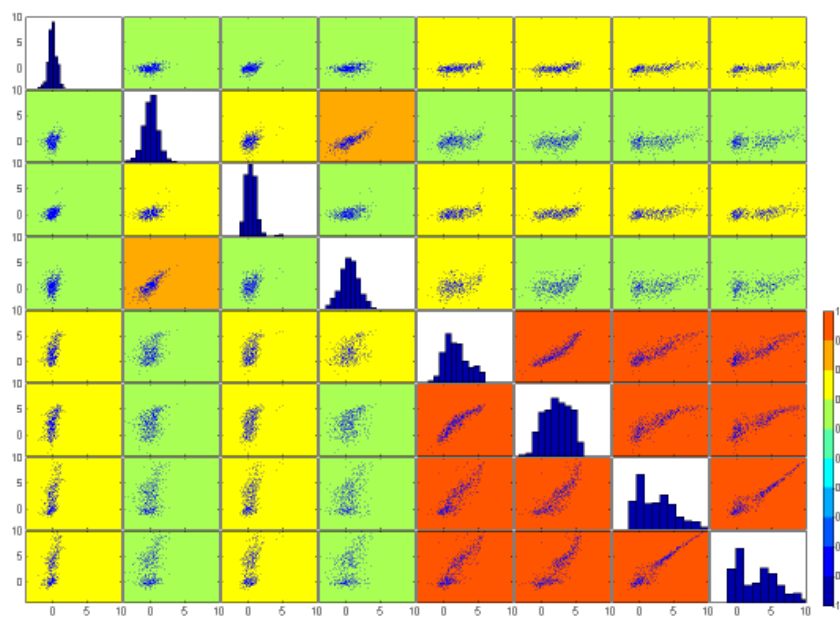
The following discussion summarizes the relationships between Affymetrix probe sets to genes, as annotated by NetAffx.



### Note on Probe Set to Gene Summarization

NetAffx annotation of Affymetrix expression probe sets contains a relationship between probe sets and genes, and many genes are represented by multiple probe sets. Often researchers need to calculate expression of a particular gene in a sample as a single value. Several methods are widely used: (1) selecting the “best” single probe set for each gene by either highest mean, or median, or maximum expression across all the samples in the dataset; (2) combining probes into a re-annotated sets and creating a custom CDF file; (3) simply averaging expression values of all probe sets associated with the same gene. The second method, while perhaps biologically more relevant, requires re-processing the whole data set starting from raw data in CEL files. The third method, although quite simple, should be avoided.

*Figure 2.12* shows pair-wise correlations between all probe sets associated with the EGFR gene. The histograms on the diagonal are distributions of expression values (log2-transformed) for each probe set. On the correlation plots, each dot represents a sample, and color corresponds to Pearson's correlation coefficient. The probe sets (eight in the case of the EGFR gene) are sorted by their mean expression across all the samples. Thus low expressing probe sets are located in the upper left corner of the figure, and high expressing in the lower right corner.



*Figure 2.12 Pair-wise correlations between all probe sets associated with the EGFR gene*

It is clear that different probe sets have very different expression ranges, and averaging them by either mean or median is (a) simply incorrect, because values do not belong to the same population (one-way ANOVA analysis gives zero p-value), and (b) results in lower values and larger noise, making any further analysis very difficult.

Another observation from the figure is that probe sets with highest expression are usually correlated with each other very well, in contrast to low expression probe sets, suggesting that a single probe set with relatively high expression can be used to represent the gene (method 1 above).

The EGFR gene has been chosen only because of its known wide range of expression in brain tumors. However, this pattern is very similar for the majority of genes represented by multiple probe sets on Affymetrix platform.

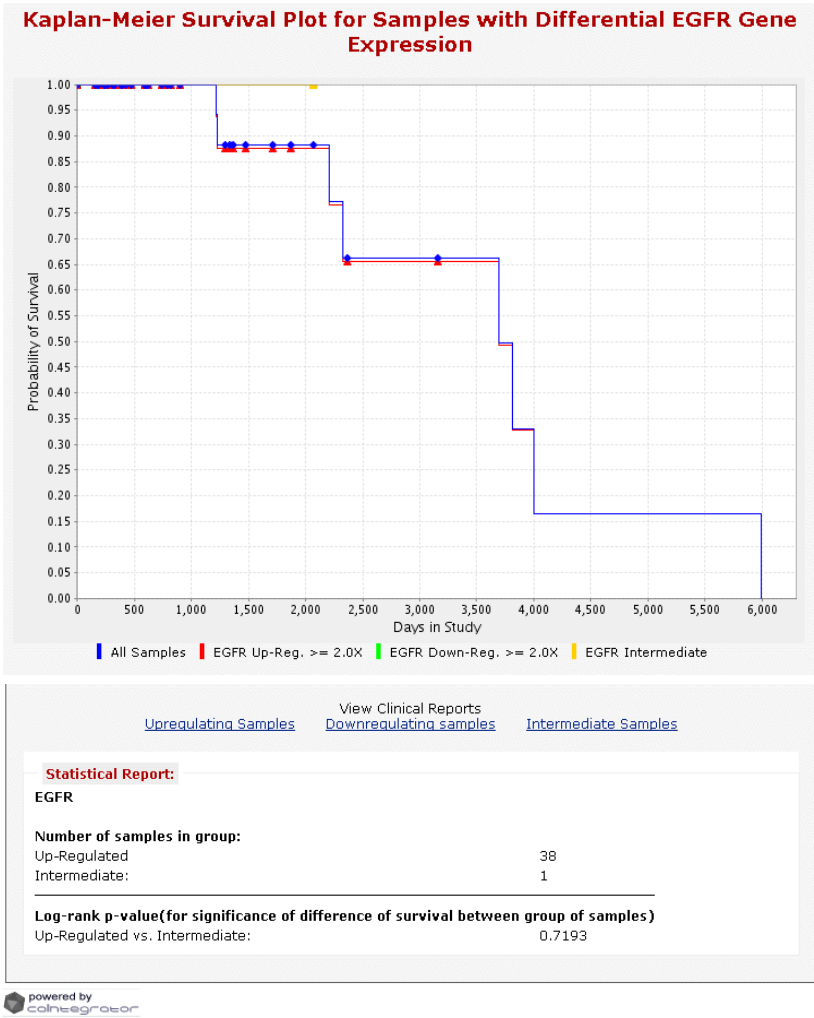


Figure 2.13 Kaplan-Meier Survival Plot for Gene Expression Data

[Table 2.8](#) describes areas on the Kaplan-Meier Survival Plot for Gene Expression data page.


Item	Special Instructions
	Mouse-over the graph, and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. See <a href="#">Saving, Printing, and E-mailing a Gene Expression Plot</a> on page 23.
<b>View Clinical Reports</b>	When you apply a gene expression filter, REMBRANDT provides links to display the gene expression for <b>Upregulating Samples</b> , <b>Downregulating Samples</b> , and <b>Intermediate Samples</b> . For more information, see <a href="#">Clinical Reports</a> .
<b>Statistical Report</b>	<ul style="list-style-type: none"> <li>Displays the gene keyword entered as search criteria for the plot.</li> <li>Displays the reporter selected for the plot.</li> <li><b>Number of Samples</b> specifies the number of Up-Regulated, Intermediate, Down-Regulated samples, if any.</li> <li><b>Log-rank <math>p</math>-Value</b> 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 <math>p</math>-value is calculated using Mantel-Haenszel procedure. The <math>p</math>-values are recalculated every time a new threshold is selected.</li> </ul>

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

## K-M Copy Number Simple Search

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

- From the Simple Search page, select **Create Kaplan-Meier survival plot for Copy Number Data**.
- 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 Affymetrix 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.
- To search on a sample group saved in a previous search, select a group from the **Restrict to Sample Group** list box.
- Click the **Go** button.

## Redrawing the K-M Survival Plot for Copy Number Data

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

---

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

---

1. Select the amplification and deletion criteria.
2. To visualize the K-M plot for the unified probeset, select a value from the **Reporters** drop-down list (*Figure 2.14*).

The screenshot shows a web interface for a search tool. At the top, there are tabs: 'Simple Search', 'Advanced Search', 'High Order Analysis', 'View Results', and 'Manage Lists'. Below the tabs, a link 'Simple Search Home' is visible. The main content area has a red header 'Constrained to group: ALL GLIOMA'. Below this, there are two rows of input fields. The first row is 'Amplified ≥ 2.0 Copies' and the second row is 'Deleted ≤ 2.0 Copies'. To the right of these is a 'Reporters' dropdown menu with 'SNP\_A-1675151' selected. At the bottom of the form is a 'Redraw Graph' button.

*Figure 2.14 Redrawing a Kaplan-Meier Survival Plot for Copy Number data page*

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

## Understanding K-M Survival Plot for Copy Number Data

A **gene keyword** search displays a plot (*Figure 2.15*) for each SNP probeset for samples with certain amplification/deletion characteristics (e.g., amplification of the cytochrome band 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.

---

**Note:** Samples for which there is no survival information and censor status are excluded from the Kaplan Meier Plot Analysis.

---

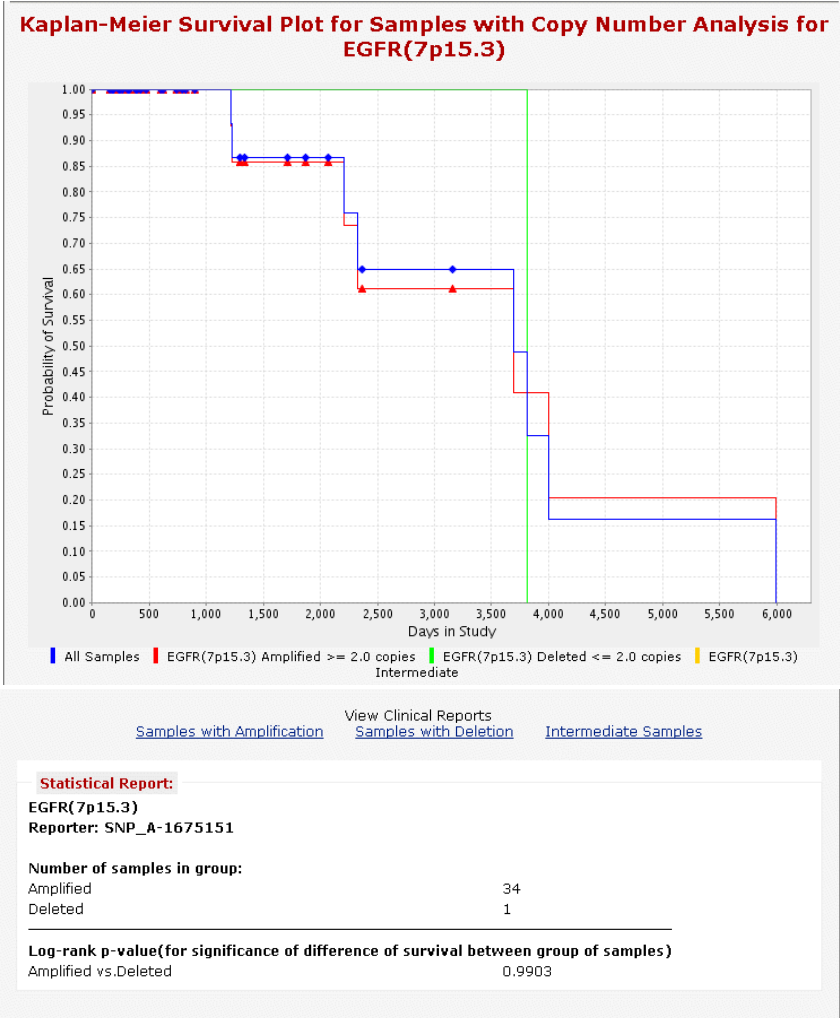


Figure 2.15 Kaplan-Meier Survival Plot for Copy Number Data

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


Item	Special Instructions
	Mouse-over the graph and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. <a href="#">Saving, Printing, and E-mailing a Gene Expression Plot</a> on page 23.
<b>View Clinical Reports</b>	When you apply a copy number filter, REMBRANDT provides links to display the copy number data for samples. For more information, see <a href="#">Clinical Reports</a> .
<b>Statistical Report</b>	<ul style="list-style-type: none"> <li>Displays the search criteria for the plot.</li> <li>Displays the reporter selected for the plot.</li> <li><b>Number of Samples</b> specifies the number of different types of samples, if any.</li> <li><b>Log-rank <math>p</math>-value</b> 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 <math>p</math>-value is calculated using Mantel-Haenszel procedure. The <math>p</math>-values are recalculated every time a new threshold is selected.</li> </ul>

Table 2.9 Kaplan-Meier Survival Plot for Copy Number Data

## K-M Sample Search

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

1. From the **Simple Search** tab, select **Create Kaplan-Meier survival plot for Sample Data**.
2. Select a sample from each drop-down list for comparison purposes.
3. Click the **Go** button. The Kaplan-Meier survival plot appears ([Figure 2.16](#)).

### Understanding K-M Survival Plot for Sample Data

A Kaplan-Meier Survival Plot for Sample Data ([Figure 2.16](#)) 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.

**Note:** Samples for which there is no survival information and censor status are excluded from the Kaplan Meier Plot Analysis.

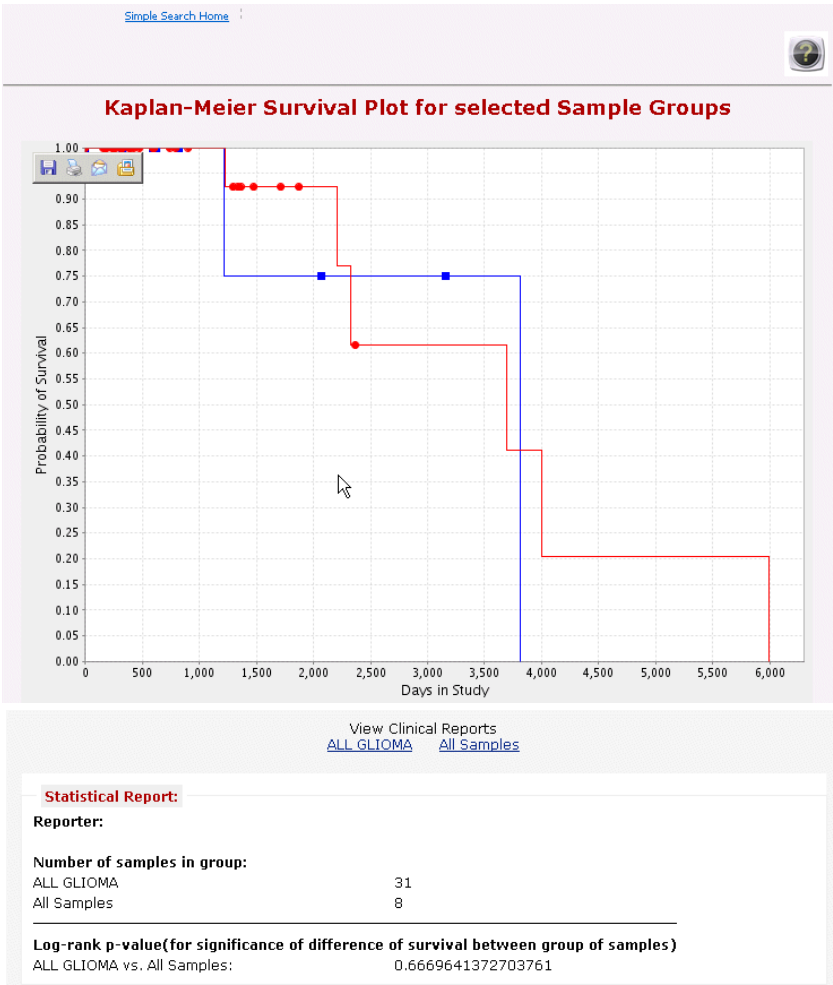


Figure 2.16 Kaplan-Meier Survival Plot for Sample Data

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

Item	Special Instructions
	Mouse-over the graph, and standard icons appear in the upper left-hand corner. Click the appropriate icon to save, print, or e-mail the graph. See <a href="#">Saving, Printing, and E-mailing a Gene Expression Plot</a> on page 23.
<b>View Clinical Reports</b>	To display clinical data for the selected sample groups, click the group link. For more information, see <a href="#">Clinical Reports</a> .
<b>Statistical Report</b>	<ul style="list-style-type: none"><li>• Displays the search criteria for the plot.</li><li>• Displays the reporter selected for the plot.</li><li>• <b>Number of Samples</b> specifies the number of Up-Regulated, Intermediate, Down-Regulated samples, if any.</li><li>• <b>Log-rank p-value</b> 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 <i>p</i>-value is calculated using Mantel-Haenszel procedure. The <i>p</i>-values are recalculated every time a new threshold is selected.</li></ul>

Table 2.10 Kaplan-Meier Survival Plot for Sample Data page

## CHAPTER 3

# CONDUCTING ADVANCED SEARCHES

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

Topics in this chapter include:

- [Advanced Searches Overview](#) on page 33
- [Gene Expression Advanced Search](#) on page 34
- [Copy Number Advanced Search](#) on page 39
- [Clinical Study Advanced Search](#) on page 42
- [Managing Individual and Compound Queries](#) on page 45
- [Refining a Query](#) on page 46

## Advanced Searches Overview

---

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

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



**Note:** To save queries, when you select to log out, Rembrandt prompts you to save the session. If you do so, when you log back in, advanced queries are saved and display in the right sidebar of the browser window. You cannot save the current session if you are logged in as a guest user.

## Gene Expression Advanced Search

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

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

### Gene Expression

Figure 3.1 Advanced Gene Expression page (top portion)

2. You are required to enter at least one search criteria for the query. [Table 3.1](#) lists the available search criteria:

Criteria	Item Name	Special Instructions
Gene	Type Genes	Select a gene identifier option ( <b>Name/Symbol</b> , <b>Locus Link ID</b> , or <b>GenBank AccNo.</b> ), and then enter or paste comma-delimited values for the genes to be searched.
	Choose a Saved Gene List	Drop down the list box and select a saved gene list. If you have not added a <a href="#">Gene List</a> with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).

Table 3.1 Advanced Gene Expression search criteria instructions



Criteria	Item Name	Special Instructions
	All Genes	Click if you do not wish to specify a list of genes but want to display data for all the genes analyzed. You must apply this option to a pre-existing result set, as described in Step 1 of <a href="#">Refining a Query</a> .
Region	Chromosome Number	Select the chromosomal region of interest ( <b>1-22, X or Y</b> ). Cytoband fills in based on the selected chromosome number.
	Cytoband	A context-sensitive list displays only the relevant cytobands for the selected chromosome. Select a cytoband range.
	Map Browser	Click to conduct a search of cytoband ranges.
	Base Pair Position	Enter the start and end base pair positions.
Clone Id/Probe Set ID	Type Reporters	Select an option ( <b>Probeset ID</b> or <b>IMAGE ID</b> ), and then enter or paste comma-delimited values for the identifiers to be searched. IMAGE identifiers must start with <b>IMAGE :</b> .
	Choose a Saved Reporters List	Drop down the list box and select a saved Reporters list. If you have not added a Reporters list with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).
Gene Ontology (GO) Classifications	(list box)	Enter a <a href="#">Gene Ontology (GO)</a> ID in the format GO:##### to search for one or more branches of the GO hierarchy. For example, enter GO:0005006 (epidermal growth factor receptor activity (12)).
	Go Browser	Click the button to search for and select a <a href="#">GO</a> classification. See <a href="#">Selecting a Gene Ontology (GO) Classification</a> .
Pathways		Click the button to search for and select a pathway. See <a href="#">Selecting a Pathway</a> .
		Click the button to search for and select a pathway. See <a href="#">Selecting a Pathway</a> .
	clear text area	Click the button to search for and select a pathway. See <a href="#">Selecting a Pathway</a> .
Clone Location	3' UTR	<i>Future Implementation</i>
	5' UTR	<i>Future Implementation</i>

Table 3.1 Advanced Gene Expression search criteria instructions

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

AND

Disease Type [?]

ALL GLIOMA  
ASTROCYTOMA  
GBM  
MIXED  
NON\_TUMOR  
OLIGODENDROGLIOMA  
UNKNOWN

Grade: All [?]

Mouseover disease types and any relevant sub-type will be displayed

Sample Identifier [?]

-or-

ASTROCYTOMA

Re-Resection Tumor Samples: [?]

☐ Exclude Re-Resection Tumor Samples

Fold Change [?]

☐ Up-regulation  $\geq$  2 fold(s)

☐ Down Regulation  $\geq$  2 fold(s)

☐ Up or Down

Up-regulation  $\geq$  2 fold(s)

Down-regulation  $\geq$  2 fold(s)

☐ Unchanged 0.8-to-1.2 fold(s)

Array Platform [?]

All

Figure 3.2 Advanced Gene Expression Disease Type

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

Criteria	Item Name	Special Instructions
Disease Type	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL+click the remaining disease types.  To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	Future Implementation
Sample Identifier	(list boxes)	To further filter the search, enter or paste comma-delimited sample identifiers to be searched.  OR  Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT My Workspace function, <b>none</b> appears (see <i>Managing Lists Overview</i> on page 91).

Table 3.2 Advanced Gene Expression disease type criteria instructions

<b>Criteria</b>	<b>Item Name</b>	<b>Special Instructions</b>
<b>Re-resection Tumor Samples</b>	checkbox	Select to exclude subsequent re-resection tumor samples from the query. Typically re-resection specimens labels are appended with letter "B" for second resection or "C" for third resection and so on.
<b>Fold Change</b>	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 <b>All Genes</b> query, you must select a fold change threshold of 4 or above.
<b>Array Platform</b>	(list box)	Select an array platform.

Table 3.2 Advanced Gene Expression disease type criteria instructions

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

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

You can also use the other buttons as follows:

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

**Tip:** A search initiated with this button cannot be saved.

### Selecting a Gene Ontology (GO) Classification

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



Figure 3.3 GO ID list

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

## Selecting a Pathway

To select a pathway of interest, follow these steps.

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

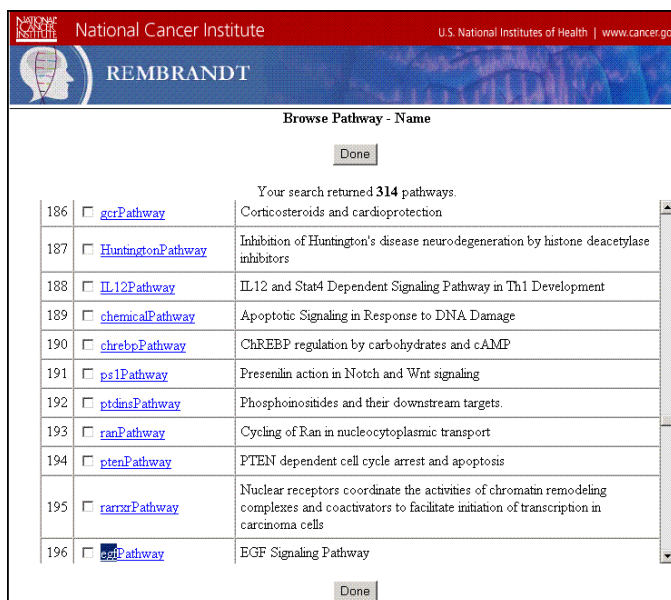


Figure 3.4 caBIO pop-up window

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

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

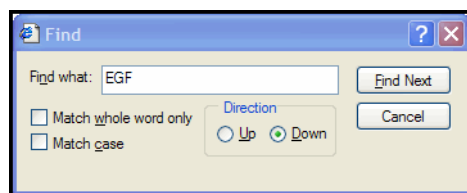


Figure 3.5 Search dialog box

3. Enter the pathway into the search text box, and click the **Find Next** button.  
The pathway(s) fulfilling the search criteria appear.
4. Select the check box next to the pathway of interest and click the **Done** button.  
The selected pathways are added to the query.

## Copy Number Advanced Search

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

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

### Copy Number Data

Figure 3.6 Copy Number Data page (top portion)

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

Criteria	Item Name	Special Instructions
Gene	Type Genes	Select a gene identifier option ( <b>Name/Symbol</b> , <b>Locus Link ID</b> , or <b>GenBank AccNo.</b> ), and then enter or paste comma-delimited values for the genes to be searched.

Table 3.3 Advanced Copy Number search criteria instructions

Criteria	Item Name	Special Instructions
	Choose a Saved Gene List	Drop down the list box and select a gene list. If you have not added a <a href="#">Gene List</a> with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).
	All Genes	Click if you do not wish to specify a list of genes but want to display data for all the genes analyzed. You must apply this option to a pre-existing result set, as described in Step 1 of <a href="#">Refining a Query</a> .
Region	Chromosome Number	Select the chromosomal region of interest ( <b>1-22, X or Y</b> ). Cytoband fills in based on the selected chromosome number.
	Cytoband	A context-sensitive list displays only the relevant cytobands for the selected chromosome. Select a cytoband range.
	Map Browser	Click to conduct a search of cytoband ranges.
	Base Pair Position (kb)	Enter the start and end base pair positions.
Geonomic Annotation Track	(text box)	<i>Future Implementation</i>
	Geonomic Browser	<i>Future Implementation</i>
SNP Id	Type SNPs	Select an SNP type identifier ( <b>dbSNP ID</b> or <b>SNP Probe Set ID</b> ), and then enter or paste comma-delimited SNP values to be searched.
	Choose a Saved SNP List	Drop down the list box and select a saved <a href="#">SNP</a> list. If you have not added an SNP list with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).
	Validated SNPs	Select one type of Validated SNPs: <b>All</b> , <b>Excluded</b> , <b>Included</b> , or <b>Only</b> .
Allele Frequency	Population Type	<i>Future Implementation</i>
Allele Frequency	Population Type	<i>Future Implementation</i>

Table 3.3 Advanced Copy Number search criteria instructions (Continued)

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

**Disease Type** [?]

ALL\_GLIOMA  
ASTROCYTOMA  
CELL\_LINE  
GBM  
MIXED  
NON\_TUMOR  
OLIGODENDROGLIOMA  
UNKNOWN

**MIXED Tumor Sub-types**

1. Mixed Glioma
2. Adult Brain Stem Mixed Glioma
3. Anaplastic Oligoastrocytoma
4. Mixed Astrocytoma-Ependymoma
5. Mixed Astrocytoma-Ependymoma-Oligodendroglioma
6. Oligoastrocytoma

Mouseover disease

**Sample Identifier**

-or- ASTROCYTOMA

**Specimen Type** [?]

**Copy Number** [?]

☐ Amplified  $\geq$  [ ] copies

☐ Deleted  $\leq$  [ ] copies

☐ Amplified or Deleted

Amplified  $\geq$  [ ] copies

Figure 3.7 Advanced Copy Number Disease Type

3. Optionally, you can combine a disease type with the query. [Table 3.4](#) lists the Disease Type items:

Criteria	Item Name	Special Instructions
Disease Type	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL+click the remaining disease types.  To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	<i>Future Implementation</i>
Sample Identifier	(list boxes)	To further filter the search, enter or paste comma-delimited sample identifiers.  OR  Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).
Specimen Type	Blood Tissue (brain)	Specify the specimen type. <ul style="list-style-type: none"> <li>• To return all the samples, leave the box blank.</li> <li>• <b>Blood</b> returns blood samples only.</li> <li>• <b>Tissue (Brain)</b> returns brain tissue samples only.</li> </ul>
Copy Number	Amplified Deleted Amplified or Deleted Unchanged	Specify the threshold for the copy number.  To create an <b>All Genes</b> query, you must select an amplification threshold greater than 10 or a deletion threshold less than 1.

Table 3.4 Advanced Copy Number disease type criteria instructions



Criteria	Item Name	Special Instructions
Array Platform	(list box)	Select the array platform.

Table 3.4 Advanced Copy Number disease type criteria instructions

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

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

You can also use the other buttons as follows:

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

## Clinical Study Advanced Search

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

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

### Clinical Data

Figure 3.8 Clinical Data page (top portion)

2. You must specify a disease type, and optionally complete the remaining information. [Table 3.5](#) lists the Disease Type items:

<b>Criteria</b>	<b>Item Name</b>	<b>Special Instructions</b>
<b>Disease Type</b>	(list box)	To select a disease, click on the name. To select more than one disease, click the first name and CTRL + click the remaining disease types.  To display the tumor sub-types for a disease type, mouse over the disease type name.
	Grade	<i>Future Implementation</i>
<b>Sample Identifier</b>	(list box)click	To further filter the search, enter or paste comma-delimited sample identifiers.  OR  Drop down the list box and select a saved sample identifier list. If you have not added a PatientDID List with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Managing Lists Overview</a> on page 91).
<b>Occurrence</b>	First Presentation Recurrence	<i>Future Implementation</i>
<b>Prior Therapy</b>	Radiation Radiation Type	Select <b>Radiation</b> and then select the type of radiation that the patient received prior to enrollment in the current study.
	Chemo Agent	Select <b>Chemo</b> and then select the agent that the patient received prior to enrollment in the current study.
	Surgery Title Outcome	Select <b>Surgery</b> 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.
<b>Onstudy Therapy</b>	Radiation Radiation Type	Select <b>Radiation</b> and then select the type of radiation that the patient received after enrollment in the current study.
	Chemo Agent	Select <b>Chemo</b> and then select the agent that the patient received after enrollment in the current study.
	Surgery Title Outcome	Select <b>Surgery</b> and then enter the name of the surgery that the patient had after enrollment in the current study and the outcome of the surgery
<b>Survival Range</b>	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.
<b>Age at Dx</b>	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.
<b>Gender</b>		Select the appropriate gender of the patient.
<b>Race</b>		Select the appropriate race of the patient.

Table 3.5 Advanced Clinical Data criteria instructions

Criteria	Item Name	Special Instructions
Clinical Evaluation	<i>Karnofsky</i>	Score from the Karnofsky Performance status scale, representing the functional capabilities of a person.
	<i>Lansky</i>	Score from an enumerated set of values representing performance status according to the Lansky scale. The Lansky scale is intended for use only with subjects under 12 years old.
	Neuro Exam	The participant's neurologic exam score. Score identifiers are the following: +2 DEFINITELY BETTER +1 POSSIBLY BETTER 0 STABLE -1 POSSIBLY WORSE -2 DEFINITELY WORSE
	MRI	Relates to the disease evaluation as measured by scan (MRI/CT). Score definitions are the following: +3 DISAPPEARANCE OF TUMOR (CR) +2 DEFINITELY BETTER (PR) +1 POSSIBLY BETTER 0 UNCHANGED -1 POSSIBLY WORSE -2 DEFINITELY WORSE (PD) -3 DEVELOPMENT OF A NEW LESION (PD)

Table 3.5 Advanced Clinical Data criteria instructions

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

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

You can also use the other buttons as follows:

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

## Managing Individual and Compound Queries

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

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

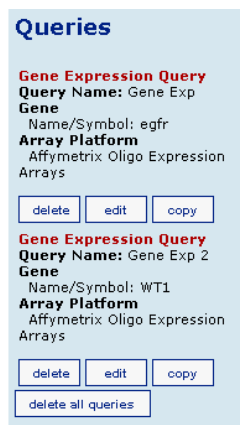


Figure 3.9 Modifying existing queries with the side bar

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

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

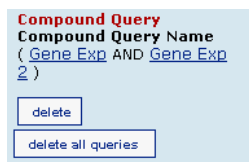


Figure 3.10 Deleting a compound query with the side bar

# Refining a Query

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

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

Step 1: Refine your result set

☒ Please refine your results by grouping the queries

Group Your Queries			
(	Query Name	)	and/or
<input type="text"/>	<input type="text" value="EGFR"/>	<input type="text"/>	<input type="text"/>

☐ Please select an "All Genes" query

Step 2: Select Result set (mandatory for "All Genes" queries)

Select Result set to apply the above Query:

Step 3: Validate your query

Step 4: Please select a View

Step 5: Please select data source(s) to filter the query by ( Optional )

Step 6: Run report or return to previous screen

Figure 3.11 Refine Query page

Table 3.6 lists the Refine Query options:

Item Name	Special Instructions
Step 1. Refine your result set	<p>You can group the queries to obtain a particular result set, or select all queries.</p> <p>To group queries click <b>Please refine your results by grouping queries</b>.</p> <ul style="list-style-type: none"><li>• Select the open parentheses, (.</li><li>• Select a <b>Query Name</b>.</li><li>• Select a closing parentheses ).</li><li>• Select an <b>and/or</b> operator at the end of a query row to enable the next row where you can select another query of interest.</li><li>• Repeat for each query name to be grouped. Go to Step 3.</li></ul> <p>– OR –</p> <p>To select all queries, click <b>Please select an All Genes query</b>. The drop-down list appears from which you can choose an All Genes query. Go to Step 2.</p>

Table 3.6 Refining Query instructions

<i>Item Name</i>	<i>Special Instructions</i>
<b>Step 2. Select result set (mandatory for “All Genes” queries)</b>	Select a previously saved result set to which to apply these queries. You will not see any result sets if you have not saved a sample set from a previous query, for example from a Clinical report page. The available sets are also listed in the side bar under PatientDID Lists in red type.
<b>Step 3. Validate your query</b>	REQUIRED. Click to validate the syntax of the query is correct.
<b>Step 4. Select a view</b>	Select a report from the drop-down list. The available reports vary based type of queries selected.
<b>Step 5. Select data source(s) to filter the query by (Optional)</b>	<p>Optionally select a datasource to filter the query by the institute providing data. You can select more than one institute.</p> <p><b>Note:</b> The Simple Search function and Preview assigns all the institutes to which you have access.</p>

*Table 3.6 Refining Query instructions*

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

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

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



# CHAPTER 4

## HIGH ORDER ANALYSIS

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

Topics in this chapter include:

- *High Order Analysis Overview* on page 49
- *Performing a Class Comparison* on page 50
- *Performing a Principal Component Analysis* on page 52
- *Performing Hierarchical Clustering Analysis* on page 54
- *Performing a GenePattern Analysis* on page 55
- *Launching GenePattern* on page 57

### High Order Analysis Overview

---

REMBRANDT stores preprocessed gene expression data (filtering and normalization). The *High Order Analysis* tab includes buttons, described below, to further analyze gene expression data.

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

Additionally, REMBRANDT can integrate with the GenePattern application in two ways, using its tools to analyze gene expression data.

- You can send gene expression data to GenePattern for up to four methods of analysis. GenePattern returns the analysis results to REMBRANDT where they can be reviewed on the View Results tab.



- You can launch GenePattern itself, giving you the option to perform your tasks right in the GenePattern application.

A high order analysis generates results that you can review on the View Results page. (See [High Order Analysis Results](#) on page 73).

## Performing a Class Comparison

To create a High Order Analysis with [Class Comparisons](#), follow these steps:

- On the High Order Analysis tab, click the **Class Comparison Analysis** button.
- The Class Comparison Analysis Form page ([Figure 4.1](#)) enables you to define the criteria to perform a class comparison.

**Note:** Clicking the plus (+) sign in Step 1 expands and displays the Advanced Statistic options.

### Class Comparison Analysis Form

The screenshot displays the 'Class Comparison Analysis Form' within the REMBRANDT 1.5 application. The interface features a top navigation bar with tabs: 'Simple Search', 'Advanced Search', 'High Order Analysis' (selected), 'View Results', and 'Manage Lists'. Below the navigation bar, the form is organized into four distinct steps:

- Step 1: Select Group\* [?] choose 2 groups**: This section contains two lists. The 'Existing Groups' list includes 'ASTROCYTOMA', 'GBM', 'MIXED', 'NON\_TUMOR', and 'OLIGODENDROGLIOMA'. The 'Selected Groups' list is currently empty. Navigation buttons '<<' and '>>' are positioned between the lists. Below the lists are 'T' and 'L' icons and a 'Baseline [?] : none' label.
- Step 2: Select Statistic [?]**: This section offers two radio button options: 'Default' (selected) and 'Advanced'. The 'Advanced' option is expanded, revealing sub-sections 'a)' and 'b)'. Sub-section 'a)' includes a 'Statistical Method' dropdown menu set to 'T-Test Two Sample Test' and a 'Multiple Comparison Adjustment' dropdown menu. Sub-section 'b) Select Constraint [?]' includes two radio button options: 'Fold Change ≥ 2' (selected) and 'Fold Change ≥' (with an empty input field).
- Step 3: Select Array Platform [?]**: This section features a single dropdown menu set to 'Oligo (Affymetrix U133 Plus 2.0)'.
- Step 4: Name Analysis Result \* [?]**: This section contains a text input field with the instruction '(should be unique)'.

At the bottom of the form, there are three buttons: 'clear', 'cancel', and 'submit'.

Figure 4.1 Class Comparison Analysis Form

You are required to complete at least one step for the class comparison. [Table 4.1](#) lists the available criteria:

Criteria	Item Name	Special Instructions
Step 1. Select Group	Existing Groups Selected Groups	<p>A <i>group</i> is a pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics (See <a href="#">Managing Lists Overview</a> on page 91). Note that user-defined PatientDID lists appear in red type in the side bar.</p> <p>Select two groups in the <b>Existing Groups</b> box and move them to the <b>Selected Groups</b> box.</p>
	Baseline	<p>To select a baseline, follow these steps:</p> <ul style="list-style-type: none"> <li>• Select a group in the <b>Selected Groups</b> box.</li> <li>• Use the <b>Baseline</b> up or down arrows to move the group to the bottom of the list.</li> <li>• Once you correctly select the baseline, <b>(baseline)</b> appears next to your selection.</li> </ul>
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.
	<ul style="list-style-type: none"> <li>• Statistical Method</li> </ul>	<p>Select the appropriate statistical method:</p> <ul style="list-style-type: none"> <li>• <b>T-test: Two Sample Test</b> to identify genes showing statistically significant differences between two samples.</li> <li>• <b>Wilcoxon Test: Man-Whitney Test</b> 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.</li> <li>• <b>F-test: One Way ANOVA</b> to identify genes showing statistically significant differences across two or more groups. <ol style="list-style-type: none"> <li>1. If there are three or more predefined groups, <b>F-test: One Way ANOVA</b> is the default statistical method.</li> <li>2. When you select the <b>F-test</b> option to test a hypothesis of the means of two or more populations, the technique is called the <i>Analysis of Variance (ANOVA)</i>. The ANOVA simplifies the F-test, where F-test is the mean square for each main effect and the interaction effect divided by the <i>within</i> variance. A one-way ANOVA or single factor ANOVA tests differences between the groups classified only on one independent variable.</li> <li>3. Using ANOVA instead of multiple t-tests reduces the probability of a type-I error.</li> </ol> </li> </ul>

Table 4.1 Class Comparison criteria instructions

Criteria	Item Name	Special Instructions
	<ul style="list-style-type: none"> <li>Multiple Comparison Adjustment</li> </ul>	<i>Family-wise Error Rate</i> (FWER): Bonferroni <i>False Discover Rate</i> (FDR): <i>Benjamini-Hochberg</i>
	<ul style="list-style-type: none"> <li>Select constraint</li> </ul>	<i>Future Implementation</i>
	<ul style="list-style-type: none"> <li>p-value</li> </ul>	<i>Future Implementation</i>
<b>Step 3. Select Array Platform</b>	Select Array Platform	Select the array platform.

Table 4.1 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. On the High Order Analysis tab, click the **Principal Component Analysis (PCA)** button.
2. The Principal Component Analysis (PCA) Form page (*Figure 4.2*) enables you to define criteria to perform a PCA. When you access the page, Current Filter Settings display in the side bar. To modify the filter settings, see the following table *Table 4.2*.

Principal Component Analysis (PCA) Form

Simple Search | Advanced Search | High Order Analysis | View Results | Manage Lists

[Analysis Home](#)

**Step 1: Select Group\* [?]**

☐ Show all samples

☒ Select samples

Existing Groups

GBM  
NON\_TUMOR  
OLIGODENDROGLIOMA  
UNKNOWN  
ALL\_GLIOMA

Selected Groups

ASTROCYTOMA  
MIXED

Current Filter Settings

Variance: 50  
Gene Set Name: none  
Reporter Set Name: none

[Manage Lists](#)

Queries

**Gene Expression Query**  
Query Name: Gene Exp  
Gene Name/Symbol: egfr  
Array Platform: all

Figure 4.2 Selecting Principal Component Analysis criteria

3. You are required to complete at least one step for the Principal Component analysis. [Table 4.2](#) lists the available criteria:

Criteria	Item Name	Special Instructions
<b>Step 1. Select Group</b>	Show all samples	Select to show all samples.
	Select samples	Select to specify the groups to include in the sample. A <i>group</i> is a pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics (See <a href="#">Managing Lists Overview</a> on page 91). Note that user-defined PatientDID lists appear in red type in the side bar.
	<ul style="list-style-type: none"> <li>Existing Groups</li> <li>Selected Groups</li> </ul>	Select at least two groups in the <b>Existing Groups</b> box and move them to the <b>Selected Groups</b> box.
<b>Step 2. Filter Genes/Reporters</b>	View Filter Settings	To use the default filter settings, continue to Step 3. Current settings display in the side bar.
	+ (-)	Click to access (and close) the advanced options.
	<ul style="list-style-type: none"> <li>Constrain reporters by variance (Gene Vector) percentile: %</li> </ul>	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.
	<ul style="list-style-type: none"> <li>Use differentially expressed genes</li> </ul>	Drop down the list box and select a saved list of differentially expressed genes identified by class comparison. If you have not added a <a href="#">Gene List</a> with the REMBRANDT My Workspace function, <b>none</b> appears (see <a href="#">Adding New Lists</a> on page 92).
	<ul style="list-style-type: none"> <li>Use differentially expressed reporters</li> </ul>	Drop down the list box and select a saved list of differentially expressed reporters identified by class comparison. If you have not added a reporter list with the Class Comparison report, <b>none</b> appears (see <a href="#">Adding New Lists</a> on page 92).
	<ul style="list-style-type: none"> <li>Set These Filters as Default</li> </ul>	Click to save the options as default filter settings.
<b>Step 3. Select Array Platform</b>	Select Array Platform	Select the array platform.

Table 4.2 Principal Comparison Analysis criteria instructions

4. 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.
5. To submit your criteria and create a Principal Comparison Analysis report, click the **Submit** button.
6. To clear the form, click **Clear**.
7. To cancel the analysis, click **Cancel**.

## Performing Hierarchical Clustering Analysis

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

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

Figure 4.3 Selecting Hierarchical Clustering criteria

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

Criteria	Item Name	Special Instructions
Step 1. Filter Genes/Reports	View Filter Settings	To use the default filter settings, continue to Step 3. Current settings display in the side bar.
	+ (-)	Click to access (and close) the advanced options.
	<ul style="list-style-type: none"> <li>Constrain reporters by variance (Gene Vector) percentile: %</li> </ul>	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.
	<ul style="list-style-type: none"> <li>Use differentially expressed genes</li> </ul>	Drop down the list box and select a saved list of differentially expressed genes identified by class comparison. If you have not added a <i>Gene List</i> with the REMBRANDT My Workspace function, <b>none</b> appears (see <i>Managing Lists Overview</i> on page 91).
	<ul style="list-style-type: none"> <li>Use differentially expressed reporters</li> </ul>	Drop down the list box and select a saved list of differentially expressed reporters identified by class comparison. If you have not added a reporter list with the REMBRANDT My Workspace function, <b>none</b> appears (see <i>Managing Lists Overview</i> on page 91).
	<ul style="list-style-type: none"> <li>Set These Filters as Default</li> </ul>	Click to save the options as default filter settings.

Table 4.3 Hierarchical Clustering criteria instructions

<b>Criteria</b>	<b>Item Name</b>	<b>Special Instructions</b>
<b>Step 2. Select Statistic</b>	Distance Matrix	Select a distance matrix option: <ul style="list-style-type: none"> <li>• <b>Pearson</b> 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.</li> <li>• <b>Euclidean</b> distance is the most common distance measure. It measures the absolute level of gene regulation.</li> </ul>
	Linkage Method	Select a linkage option to affect the shape of the resulting clusters: <ul style="list-style-type: none"> <li>• <b>Average linkage</b> is the average of all pair-wise distances between members of the two clusters.</li> <li>• <b>Single linkage</b> is the minimum distance between two clusters.</li> <li>• <b>Complete linkage</b> is the maximum distance between two clusters.</li> </ul>
<b>Step 3. Cluster By</b>	Cluster by	Leave the default to cluster on <b>Samples</b> or cluster by <b>Genes</b> .
<b>Step 4. Select Array</b>	Select Array Platform	Select the array platform.

Table 4.3 Hierarchical Clustering criteria instructions

4. 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.
5. To submit your criteria and create a Hierarchical Clustering Analysis report, click the **Submit** button.
6. To clear the form, click **Clear**.
7. To cancel the analysis, click **Cancel**.

## Performing a GenePattern Analysis

REMBRANDT provides two means to use GenePattern to analyze gene expression data.

- This section describes sending data from REMBRANDT to GenePattern for analyses whose results can then be viewed in REMBRANDT.
- The following section describes how to launch GenePattern itself so you can work within that application.

To perform a GenePattern analysis of gene expression data, follow these steps:

1. On the High Order Analysis tab, click the **Send Data to GenePattern** button.

- The GenePattern Integration Form page ([Figure 4.4](#)) that opens enables you to define the criteria to perform GenePattern data analyses.

#### Gene Pattern Integration Form

Figure 4.4 Selecting GenePattern query criteria

- Complete the form as described in [Table 4.3](#):

Criteria	Item Name	Special Instructions
<b>Step 1. Select Group</b>	Existing Groups Selected Groups	A <i>group</i> is a pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics (See <a href="#">Managing Lists Overview</a> on page 91). Note that user-defined PatientDID lists appear in red type in the side bar.  Select two groups in the <b>Existing Groups</b> box and move them to the <b>Selected Groups</b> box.
<b>Step 2. Select Array Platform</b>	Select Array Platform	Select the array platform. For the current version of REMBRANDT, Affymetrix is the sole option.

Table 4.3 GenePattern analysis criteria instructions

- 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.
- To submit your criteria and create a GenePattern Analysis report, click the **Submit** button.

Once the data has been sent to GenePattern, REMBRANDT directs you to the View Result page. For more information, see [GenePattern Analysis Reports](#) on page 80.

6. To clear the form, click **Clear**.
7. To cancel the analysis, click **Cancel**.

## Launching GenePattern

---

To launch a session in the GenePattern application, on the High Order Analysis page, click **Launch Gene Pattern Application**. You can perform many gene expression analyses working within GenePattern itself. For more information about using GenePattern, click the online help links within the application.





# CHAPTER 5 VIEWING RESULTS

This chaptersection 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 chaptersection include the following:

- [Results Overview](#) on page 59
- [Notification of Prolonged Queries](#) on page 60
- [Clinical Reports](#) on page 60
- [Advanced Search or Query Results](#) on page 64
- [High Order Analysis Results](#) on page 73
- [GenePattern Analysis Reports](#) on page 80

## Results Overview

---

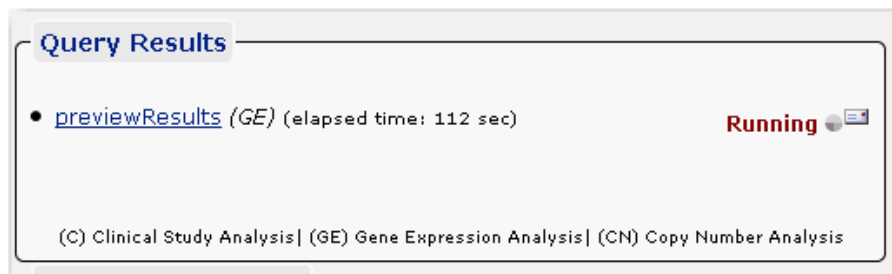
When you opens the View Results tab after launching an advanced query, the page displays queries in progress or newly completed. If you have opened this tab using the Preview, the Finalize Query or Refine Query buttons, you may see the icon that monitors the timing of the query.

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 or even in spreadsheet format. For example, you can compare clinical and gene expression reports with a set of patient samples. You can view Clinical reports (see [Clinical Reports](#)), results generated with the Advanced Search function (see [Advanced Search or Query Results](#)) and the High Order Analysis function (see [High Order Analysis Results](#)).


## Notification of Prolonged Queries

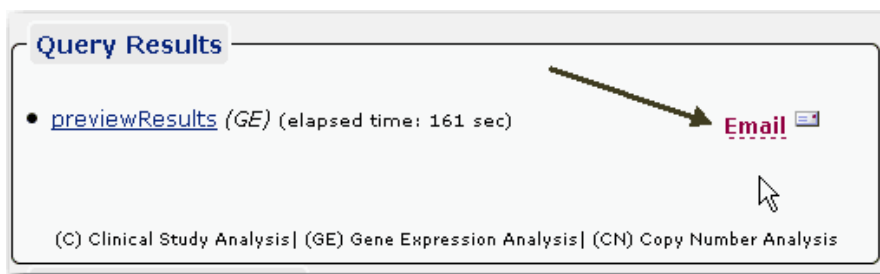
Once you execute an advanced query, either through the **Preview**, **Finalize Query** or **Refine Query** button, if the query takes longer than 45 seconds, REMBRANDT provides you the option of being notified by email when the query ends.

After you execute the search by clicking any of those buttons, the system opens the View Results tab which displays a monitor of the elapsed time for the query (*Figure 5.1*).



*Figure 5.1 Query Results preview monitor, showing email icon*

- The Checking Status monitor refreshes every 10 seconds. After 45 seconds, REMBRANDT displays a letter icon (  ). If you click the icon, a dialog box opens where you can enter your email address. The system first validates your email address, then displays an email link (*Figure 5.2*)..



*Figure 5.2 An email link notifies you that an email will be sent when the query ends.*

- If you mouse over the link, a popup indicates that you will be notified by email when the results are ready. The results are saved in REMBRANDT for 5 days.
- In the email you receive, you can click the link to the query. After you log into REMBRANDT, click the View Results tab to retrieve the query results.
- The query report is available only to the user who generates the report. Other logged in users will not be able to see the report.

---

**Note:** The query continues even if you log out of REMBRANDT.

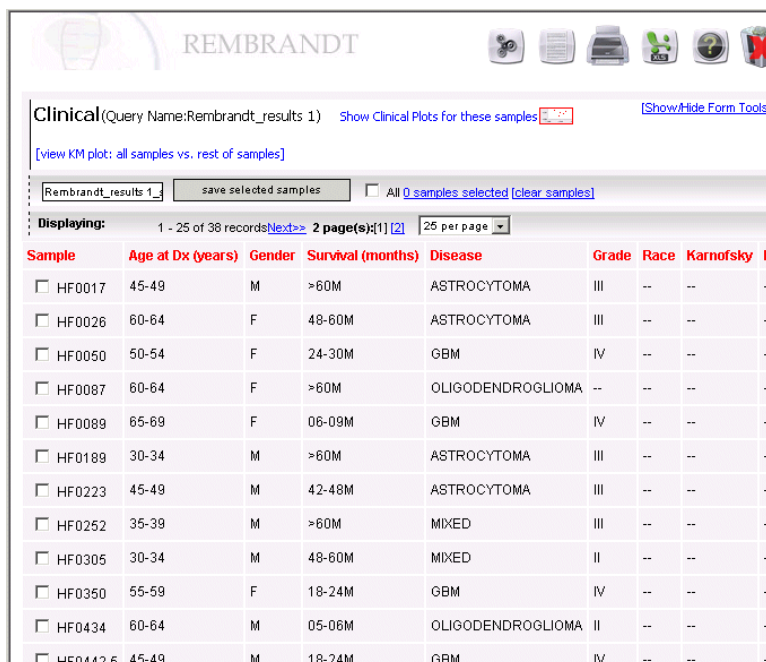
---

## Clinical Reports

A Clinical report displays patient demographics, therapy, and outcome data. This information is displayed in a single domain, such as gene expression, or in multiple domains that combine queries, such as gene expression, chromosomal aberrations,

and clinical areas ([Figure 5.3](#)). On the Clinical page, you can save samples to a [PatientDID list](#) stored in the My Workspace function. To save samples, follow these steps:

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



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

Figure 5.3 Clinical page

- There are two ways to select samples on the Clinical window:
  - To select an individual sample, select the box in the **Sample** column ([Figure 5.4](#)).

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

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

Figure 5.4 Checking the Sample column on the Clinical window

- To select all of the samples, select the **All** box. To display a list of the selected samples, click the **samples selected** link ([Figure 5.5](#)).

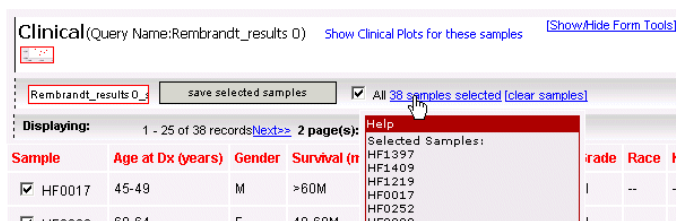


Figure 5.5 Selecting all of the samples on the Clinical window

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

2. To save the selected samples, enter a unique name for the PatientDID list next to **Select Samples**, or maintain the current name (*Figure 5.6*).

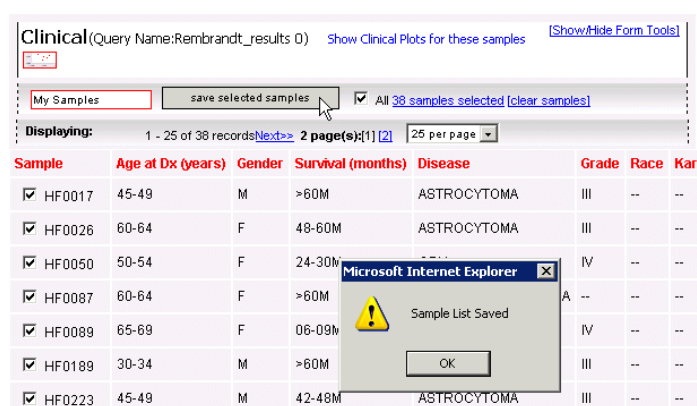



Figure 5.6 Saving Selected Samples on the Clinical page

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

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

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

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

6. To export only the samples you select to an Excel spreadsheet, select the samples and click the **Download for Excel** icon (  ). The spreadsheet opens displaying data for the selected samples ( [Figure 5.7](#) ).

	A	B	C	D	E	F	G	H	I	J	K
1	Sample	Age at Dx	Gender	Survival (n	Disease	Grade	Race	Institution	Karnofsky	Neurologic M	
2	E08021	30-34	F	>60M	MIXED	--	WHITE	UCSF	--	--	--
3	E09139	50-54	F	--	GBM	--	WHITE	NIH NEUF	--	--	--
4	E09151	20-24	M	--	UNKNOWN	--	WHITE	NIH NEUF	100  100	0  0  0	0
5	E09167	65-69	M	12-18M	GBM	--	WHITE	DANA-FA	60	--	-1

*Figure 5.7 Spreadsheet displaying selected samples and their metadata*

[Understanding K-M Survival Plot for Gene Expression Data](#)

[Understanding K-M Survival Plot for Copy Number Data](#)

[Understanding K-M Survival Plot for Sample Data](#)

[Gene Expression Sample Report](#)

[Copy Number Sample Report](#)

## Viewing Clinical Plots

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

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

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

To select the samples of interest, follow these steps:

1. Click and drag a rectangle around the samples to save for future use. A red rectangle appears around the samples, and the list of the samples appears on the right-hand side ( [Figure 5.8](#) ).

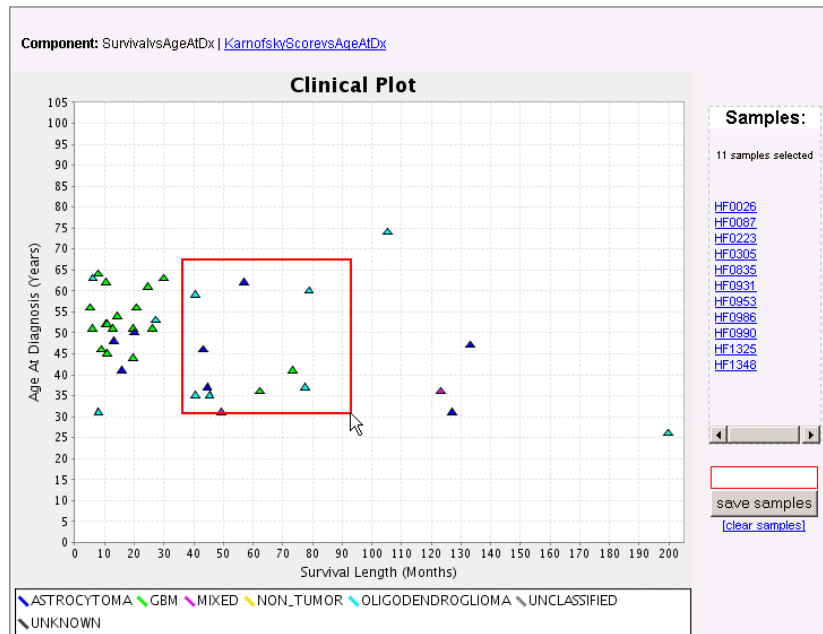


Figure 5.8 Sampling from a Clinical plot

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.

### Clinical Reports

[Understanding K-M Survival Plot for Gene Expression Data](#)

[Understanding K-M Survival Plot for Copy Number Data](#)

[Understanding K-M Survival Plot for Sample Data](#)

[Gene Expression Sample Report](#)

[Copy Number Sample Report](#)

## Advanced Search or Query Results

The following Advanced Search reports are generated:

- [Gene Expression Sample Report](#)
- [Copy Number Sample Report](#)

View Results ([Figure 5.9](#)) 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.

Query Results		
Compound Query	Institutions	View
Gene Exp 1		<a href="#">Clinical View</a>
		<a href="#">Gene Expression Data Per Sample View</a>
		<a href="#">Gene Expression Data Per Disease Group View</a>

Figure 5.9 Query Results

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

## Gene Expression Sample Report

The Gene Expression Sample report ([Figure 5.10](#)) displays gene expression ratios (between the tumor sample and the geometric mean of non-tumor samples) for each probeset (or IMAGE clone) for the genes selected in the queries. Each column represents a sample, and the samples are grouped by tumor sub-type. For Affymetrix probesets, the ratio of the absolute expression values of the tumor sample and the geometric mean of the expression value of the non-tumor samples displays. For each IMAGE clone, missing values are handled and the ratio of expression values between the tumor and geometric mean of the non-tumor group is calculated for each sample.

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

REMBRANDT

Gene Expression Sample (Query Name: gene expression query)

[\[Show/Hide Form Tools\]](#)

[\[View Clinical Report for all samples\]](#)

Filter: ☐ Show Only ☒ Hide Gene(s)

Highlight: highlight values

Select Samples:   [\[Check All\]](#) [\[Uncheck All\]](#) [\[?\]](#)

Show all Values:   [\[?\]](#)

Hide Diseases: ☐ ASTROCYTOMA ☐ GBM ☐ MIXED ☐ NON\_TUMOR ☐ OLIGODENDROGLIOMA ☐ UNCLASSIFIED ☐

UNKNOWN

Displaying: 1 - 15 of 15 records  119 samples returned.

Gene Reporter

ASTROCYTOMA Samples ☐

	<input type="checkbox"/> E09137	<input type="checkbox"/> E09214	<input type="checkbox"/> E09262	<input type="checkbox"/> HF0017	<input type="checkbox"/> HF0026	<input type="checkbox"/> HF0189	<input type="checkbox"/> HF0223	<input type="checkbox"/> HF0608	<input type="checkbox"/>
EGFR 1565483_at	0.3100	0.4000	0.1800	2.3200	5.3300	2.7700	2.5500	8.8900	0.2
EGFR 1565484_x_at	0.5400	1.2200	0.2000	1.6400	2.0600	0.9200	1.8100	4.0300	0.3
EGFR 201983_s_at	1.2500	6.8300	4.3500	12.0800	8.3400	6.6400	6.4600	31.9100	4.2
EGFR 237938_at	2.2500	0.6300	0.8000	0.8700	2.3900	2.7600	2.6500	1.0800	1.6
EGFR 243327_at	0.5200	0.8000	1.0400	3.2100	2.1100	2.0900	1.9600	14.2100	0.4

Figure 5.10 Gene Expression Data Per Sample Report

When two queries are included in the results, the results are separated by a double line and the query information is listed at the bottom. For more information about the report, see the Related Topics.

[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*

*Gene Expression Advanced Search*

## Filtering Results by Gene or Reporter (Filter Toolbar)

To filter a report, follow these steps:

1. From the **Filter** toolbar (*Figure 5.11*), select the filter mode **Show only** or **Hide**.
2. Select **Gene** or **Reporter** from the drop-down list, and enter gene or reporter to be filtered.

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

3. Click the **Filter** button.

The results are filtered based on your selections (*Figure 5.11*).

The screenshot shows the REMBRANDT software interface. At the top, there's a header with the REMBRANDT logo and several icons. Below the header, there's a section titled "Gene Expression Sample (Query Name: Gene Expression 2 filter report)". This section contains a "Filter:" toolbar with radio buttons for "Show Only" and "Hide", a dropdown menu for "Gene(s)", a text input field containing "WT1", and buttons for "Filter" and "Reset (show all)". Below the filter toolbar, there are sections for "Highlight:" (with a dropdown for "highlight values" and buttons for "Highlight" and "Clear Highlighting"), "Select Samples:" (with a dropdown for "Gene Expression 2 filter report" and buttons for "Save Samples", "Check All", "Uncheck All", and "?"), and "Show all Values:" (with buttons for "Show all values on this report" and "View Previous Report", and a "?"). Below these, there's a "Hide Diseases:" section with checkboxes for "ASTROCYTOMA", "GBM", "MIXED", "OLIGODENDROGLIOMA", "UNCLASSIFIED", and "UNKNOWN". A "Displaying:" section shows "1 - 2 of 2 records", "25 per page", and "91 samples returned". Below the filter toolbar, there's a table titled "ASTROCYTOMA Samples" with columns for "Gene" and "Reporter". The table has two rows of data. The first row shows "WT1" and "206067 s at" with values 0.6000, 1.0900, 0.8100, 0.3600, 1.1600, 2.2600, 1.9500, and 1.6200. The second row shows "WT1" and "216953 s at" with values 0.2600, 0.5200, 0.7200, 0.9200, 1.8300, 1.7400, 2.7100, and 1.9500. Below the table, there's a "Query: ( Gene Expression 1 AND Gene Expression 2 )" section with two columns for "Gene Expression Query". The first column shows "Query Name: Gene Expression 1", "Gene Name/Symbol: EGFR", and "Array Platform: Affymetrix Oligo Expression Arrays". The second column shows "Query Name: Gene Expression 2", "Gene Name/Symbol: WT1", and "Array Platform: Affymetrix Oligo Expression Arrays".

*Figure 5.11 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.

*Gene Expression Sample Report*

*Copy Number Sample Report*

## Highlighting Results By Value (Highlight Toolbar)

To highlight certain data, follow these steps:

1. From the **Highlight** toolbar, select an operator and a threshold value. (*Figure 5.12*).

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

2. Click the **Highlight** button.

The values that meet this criteria are highlighted in yellow (*Figure 5.12*).

The screenshot shows the REMBRANDT web interface for a Gene Expression Sample report. The 'Highlight' toolbar is highlighted with a black box, showing the 'highlight values' dropdown set to '<' and the threshold '1'. The 'Highlight' button is also highlighted. Below the toolbar, the 'ASTROCYTOMA Samples' table is displayed. The table has columns for Gene, Reporter, and various sample IDs. The values for WT1 206067 s at are highlighted in yellow: 0.6000, 1.0900, 0.8100, 0.3600, 1.1600, 2.2600, 1.9500, 1.6200. The values for WT1 216953 s at are also highlighted in yellow: 0.2600, 0.5200, 0.7200, 0.9200, 1.8300, 1.7400, 2.7100, 1.9500.

Gene	Reporter	ASTROCYTOMA Samples
WT1	206067 s at	0.6000, 1.0900, 0.8100, 0.3600, 1.1600, 2.2600, 1.9500, 1.6200
WT1	216953 s at	0.2600, 0.5200, 0.7200, 0.9200, 1.8300, 1.7400, 2.7100, 1.9500

Figure 5.12 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.

[Gene Expression Sample Report](#)

[Copy Number Sample Report](#)

[Class Comparison Report](#)

## Selecting and Saving Sample Results (Select Samples Toolbar)

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

- On the sample report, there are several ways to select samples.
  - To select all the listed samples on the **Select Samples** toolbar, click the **Check All** link. To deselect all the listed samples, click the **Uncheck All** link (*Figure 5.13* below).

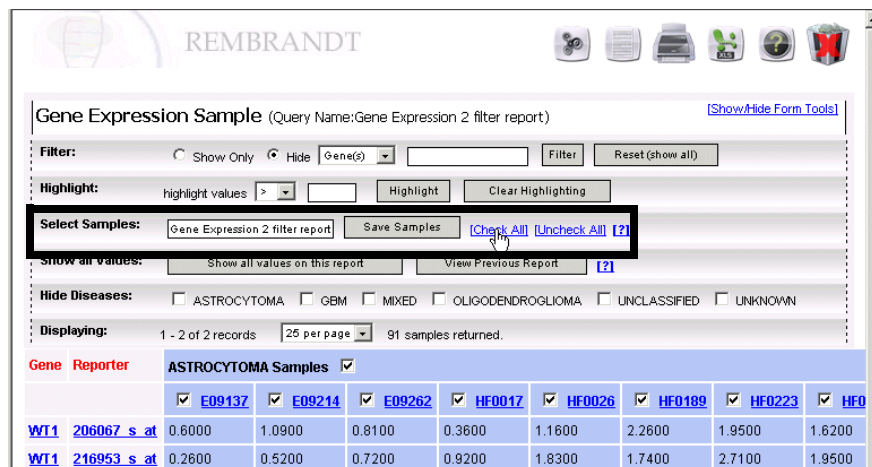


Figure 5.13 Selecting samples from the Select Samples toolbar

- To select (or deselect) all the samples in a sample group, click the box next to the sample group name, for example, the box next to **ASTROCYTOMA Samples**. All the samples in the group are selected (*Figure 5.14* below).
- To select (or deselect) an individual sample within a group, click the box in the column next to the sample name.

### Gene Expression Sample Report

### Copy Number Sample Report

Gene	Reporter	ASTROCYTOMA Samples <input checked="" type="checkbox"/>			
		<input checked="" type="checkbox"/> E09137	<input checked="" type="checkbox"/> E09214	<input checked="" type="checkbox"/> E09262	<input checked="" type="checkbox"/> HF0017
WT1	206067 s at	0.6000	1.0900	0.8100	0.3600
WT1	216953 s at	0.2600	0.5200	0.7200	0.9200

Figure 5.14 Selecting samples from the results

- To save the selected samples, enter a unique name for the PatientDID list next to **Select Samples**, or maintain the current name (*Figure 5.15*).

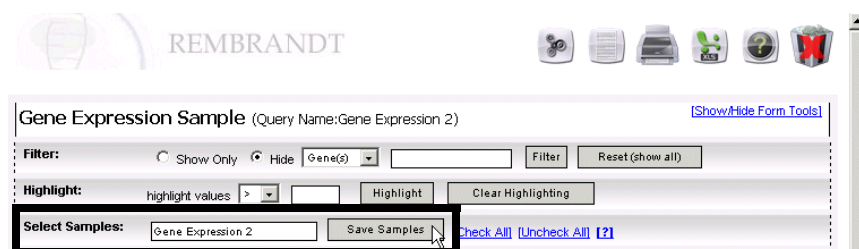


Figure 5.15 Saving the samples

- Click the **Save Samples** button.

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

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

*Gene Expression Sample Report*

*Copy Number Sample Report*

### Differentiating Data (Show All Values Toolbar)

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

- Click **Show All Values on this Report** on the **Show all Values** toolbar (*Figure 5.16*).

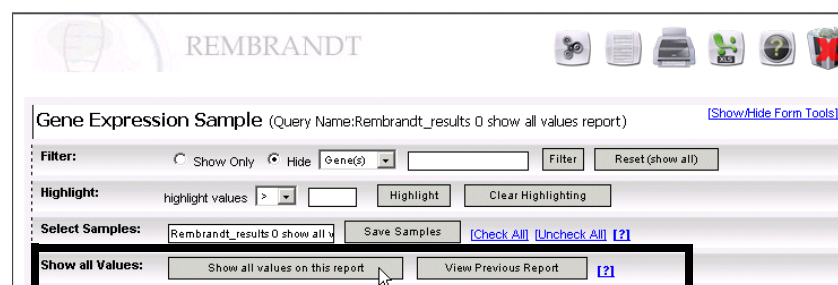


Figure 5.16 Show All Values toolbar

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

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

*Gene Expression Sample Report*

*Copy Number Sample Report*

### Removing Columns (Hide Diseases toolbar)

To remove a disease from the report, select the check box for the disease in the **Hide Diseases** toolbar (*Figure 5.17*). The checked disease is NOT included in the results.

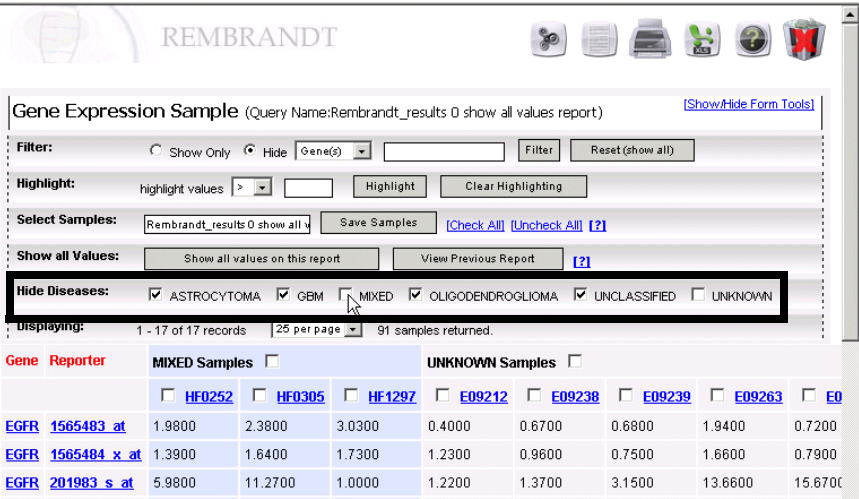


Figure 5.17 Hide Diseases toolbar

*Gene Expression Sample Report*

*Copy Number Sample Report*

### Showing Additional Information

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

For example, to display more information about a gene, click the gene symbol link (*Figure 5.18*).



Figure 5.18 The Gene column

The Cancer Genome Anatomy Project (CGAP) browser opens.

## Copy Number Sample Report

The Copy Number Data Per Sample report displays the copy number data from Affymetrix 100K SNP arrays. The CHP files from the Affymetrix Gene Chip Operating System were processed using the Affymetrix GDAS (GeneChip® DNA Analysis Software). Copy number data was collected for each mapping SNP reporter on the Chip, for all the tumor samples. Each column represents a sample, and the samples are grouped based on the tumor sub-type (*Figure 5.19*).

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

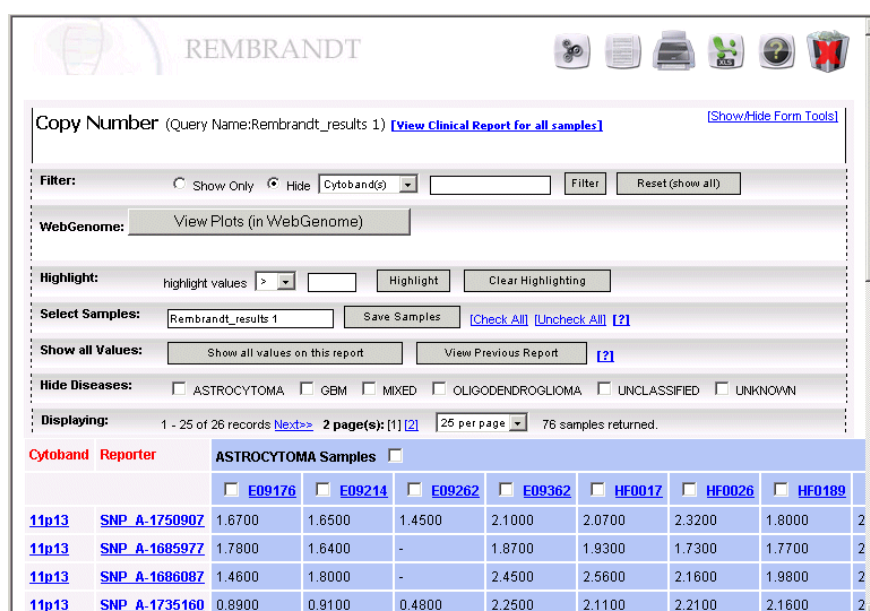


Figure 5.19 Copy Number Data for Sample report

For more information about the report, see the Related Topics.

*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*

*Copy Number Advanced Search*

## 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 data for selected reporter probes.

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

1. Click the **View Plots in webGenome** button (*Figure 5.20*).

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

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

Copy Number (Query Name: Rembrandt\_results 1) [\[View Clinical Report for all samples\]](#) [\[Show/Hide Form Tools\]](#)

Filter: ☐ Show Only ☒ Hide Cytoband(s)

WebGenome: **View Plots (in WebGenome)**

Highlight: highlight values

Select Samples: Rembrandt\_results 1  [\[Check All\]](#) [\[Uncheck All\]](#) [\[?\] 21](#)

Show all Values:   [\[?\] 21](#)

Hide Diseases: ☐ ASTROCYTOMA ☐ GBM ☐ MIXED ☐ OLIGODENDROGLIOMA ☐ UNCLASSIFIED ☐ UNKNOWN

Displaying: 1 - 25 of 26 records [Next >>](#) 2 page(s): [1] [2] 25 per page 76 samples returned.

Cytoband	Reporter	ASTROCYTOMA Samples							
		<input type="checkbox"/> E09176	<input type="checkbox"/> E09214	<input type="checkbox"/> E09262	<input type="checkbox"/> E09362	<input type="checkbox"/> HF0017	<input type="checkbox"/> HF0026	<input type="checkbox"/> HF0189	
11p13	SNP_A-1750907	1.6700	1.6500	1.4500	2.1000	2.0700	2.3200	1.8000	2
11p13	SNP_A-1685977	1.7800	1.6400	-	1.8700	1.9300	1.7300	1.7700	2

Figure 5.20 Copy Number Report



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

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

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

For more information, see [Copy Number Sample Report](#).

## High Order Analysis Results

The following High Order Analysis reports are generated:

- [Class Comparison Report](#) on this page
- [Principal Component Analysis Plot](#) on page 77
- [Hierarchical Clustering Report](#) on page 80
- [GenePattern Analysis Reports](#) on page 80

View Results ([Figure 5.9](#)) displays the query name and lists the output generated for the query.

Query Results		
Compound Query	Institutions	View
<a href="#">Gene Exp 1</a>		<a href="#">Clinical View</a>
		<a href="#">Gene Expression Data Per Sample View</a>
		<a href="#">Gene Expression Data Per Disease Group View</a>

Figure 5.21 Query Results

## Class Comparison Report

The Class Comparison report ([Figure 5.22](#)) displays group average, fold change, and  $p$ -value based on the Advanced Search parameters that you selected. The output varies based on the statistical method chosen.

For **T-test** or **Wilcoxon** Statistical Method analysis ([Figure 5.22](#)), the Class Comparison report is as follows.

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



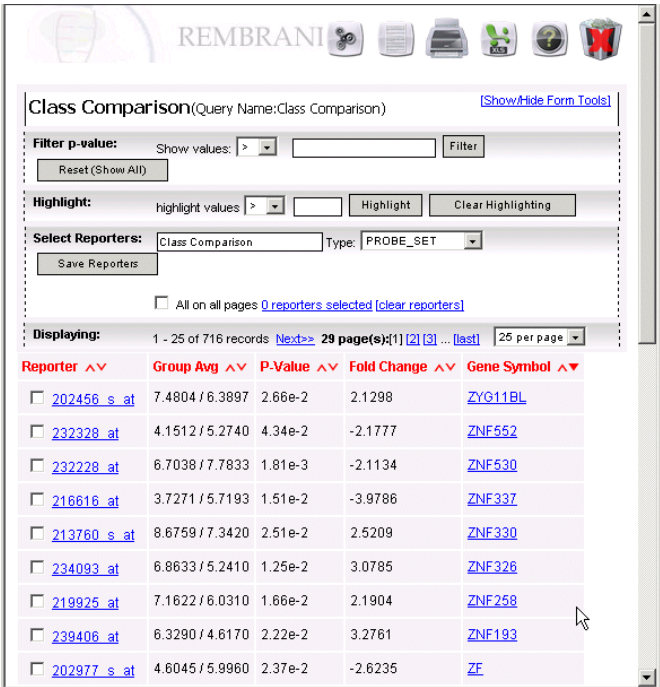


Figure 5.22 Class Comparison report

For an **F-test** Statistical Method analysis (Figure 5.23), the difference is that in the results there is a group average displayed for each group.

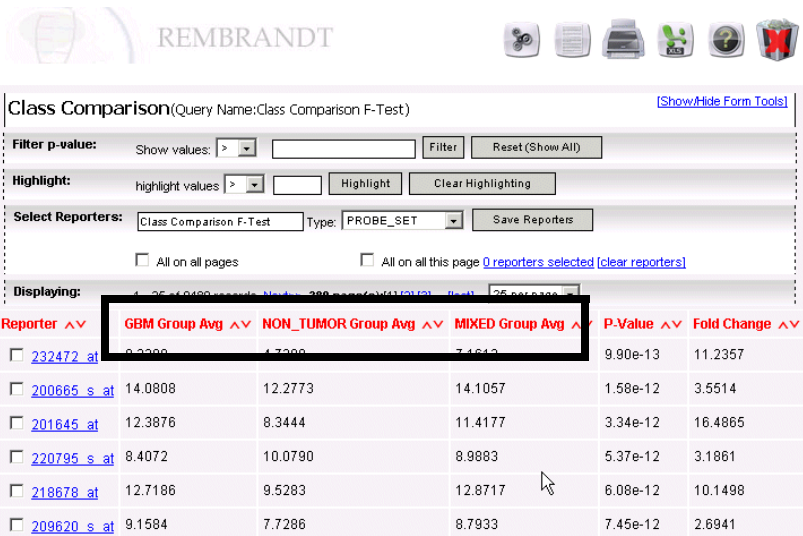


Figure 5.23 Class Comparison report - F-test

For more information, see [Related Topics](#).

[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](#)

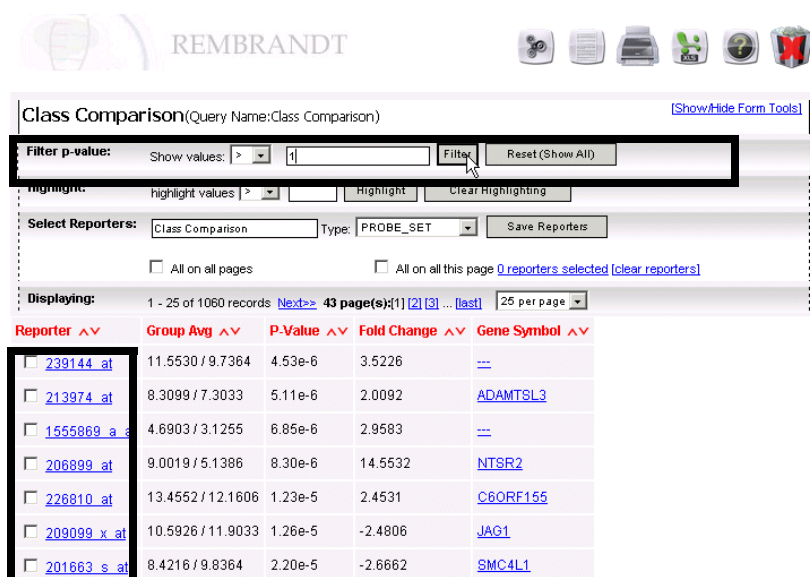
A high order analysis generates results that you can review on the View Results page. See [High Order Analysis Results](#) on page 73.

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

To filter a p-value, follow these steps:

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

The results are filtered based on your selections ([Figure 5.24](#)).



The screenshot shows the REMBRANDT software interface. At the top, there's a logo and several icons. Below that, the title 'Class Comparison (Query Name: Class Comparison)' is displayed. The 'Filter p-value' toolbar is highlighted with a black box. It includes a 'Show values' dropdown set to '>', a text input field containing '1', and a 'Filter' button. To the right of the 'Filter' button is a 'Reset (Show All)' button. Below the toolbar, there are sections for 'Highlight' and 'Select Reporters'. The 'Select Reporters' section shows 'Class Comparison' as the selected class and 'PROBE\_SET' as the type. Below this, there are checkboxes for 'All on all pages' and 'All on all this page', and a link to '0 reporters selected (clear reporters)'. The 'Displaying' section shows '1 - 25 of 1060 records', 'Next>>' and '43 page(s)' navigation links, and a '25 per page' dropdown. Below these sections is a table with columns: Reporter, Group Avg, P-Value, Fold Change, and Gene Symbol. The first six rows of the table are highlighted with a black box. The first row is: 239144\_at, 11.5530 / 9.7364, 4.53e-6, 3.5226, ---. The second row is: 213974\_at, 8.3099 / 7.3033, 5.11e-6, 2.0092, ADAMTSL3. The third row is: 1555869\_at, 4.6903 / 3.1255, 6.85e-6, 2.9583, ---. The fourth row is: 206899\_at, 9.0019 / 5.1386, 8.30e-6, 14.5532, NTSR2. The fifth row is: 226810\_at, 13.4552 / 12.1606, 1.23e-5, 2.4531, C6ORF155. The sixth row is: 209099\_x\_at, 10.5926 / 11.9033, 1.26e-5, -2.4806, JAG1. The seventh row is: 201663\_s\_at, 8.4216 / 9.8364, 2.20e-5, -2.6662, SMC4L1.

Reporter	Group Avg	P-Value	Fold Change	Gene Symbol
239144_at	11.5530 / 9.7364	4.53e-6	3.5226	---
213974_at	8.3099 / 7.3033	5.11e-6	2.0092	ADAMTSL3
1555869_at	4.6903 / 3.1255	6.85e-6	2.9583	---
206899_at	9.0019 / 5.1386	8.30e-6	14.5532	NTSR2
226810_at	13.4552 / 12.1606	1.23e-5	2.4531	C6ORF155
209099_x_at	10.5926 / 11.9033	1.26e-5	-2.4806	JAG1
201663_s_at	8.4216 / 9.8364	2.20e-5	-2.6662	SMC4L1

Figure 5.24 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.

For more information, see [Class Comparison Report](#).

## Selecting and Saving Reporters (Select Reporters toolbar)

You can save reporters to a Reporterlist stored in the My Workspace function. Reporter lists enable you to further filter advanced queries. To save reporters, follow these steps ([Figure 5.25](#)):

- 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.

Class Comparison (Query Name: Class Comparison) [Show/Hide Form Tools]

Filter p-value: Show values: > [ ] Filter Reset (Show All)

Highlight: highlight values > [ ] Highlight Clear Highlighting

Select Reporters: Class Comparison Type: AFFY\_HGU133PLUS2\_PROBE\_SET Save Reporters

☐ All on all pages ☐ All on all this page 3 reporters selected [clear reporters]

Displaying: 1 - 25 of 8113 records Next>> 325 page(s): [1] [2] [3] ... [last] 25 per page

Reporter ^v	Group Avg ^v	p-value ^v	Fold Change ^v	Gene Symbol ^v
<input checked="" type="checkbox"/> 202189_x_at	11.7264 / 9.4551	1.64e-13	4.8273	PTBP1
<input type="checkbox"/> 201645_at	12.3876 / 8.3444	3.09e-13	16.4865	TNC
<input checked="" type="checkbox"/> 209620_s_at	9.1584 / 7.7286	4.45e-13	2.6941	ABCB7
<input type="checkbox"/> 219246_s_at	7.5968 / 5.2874	5.43e-13	4.9568	FLJ13491
<input checked="" type="checkbox"/> 218679_at	12.7186 / 9.5283	6.59e-13	9.1281	NES
<input type="checkbox"/> 232472_at	8.2209 / 4.7309	7.55e-13	11.2357	---
<input type="checkbox"/> 216306_x_at	11.8665 / 9.7990	7.71e-13	4.1915	PTBP1

Figure 5.25 Selecting Reporters instructions

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

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

Once saved, the Reporters list appears in red type under Reporter Lists in the side bar.

- 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.

For more information, see [Class Comparison Report](#).

## Resorting Column Results

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

1. If a report column has red triangles pointing up and down next to the name, you can sort a column of numeric or alphabetical values ([Figure 5.26](#)).



Fold Change ▲▼
-8.1624
-6.7058
-5.6246
-5.3892
-5.3799
-5.1293
-4.8571

Figure 5.26 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.

For more information, see [Class Comparison Report](#).

## Principal Component Analysis Plot

The Principal Component Analysis plot ([Figure 5.27](#)) is a two-dimensional graph which plots the various principal components from the analyses. The following list describes the different areas of the plot:

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

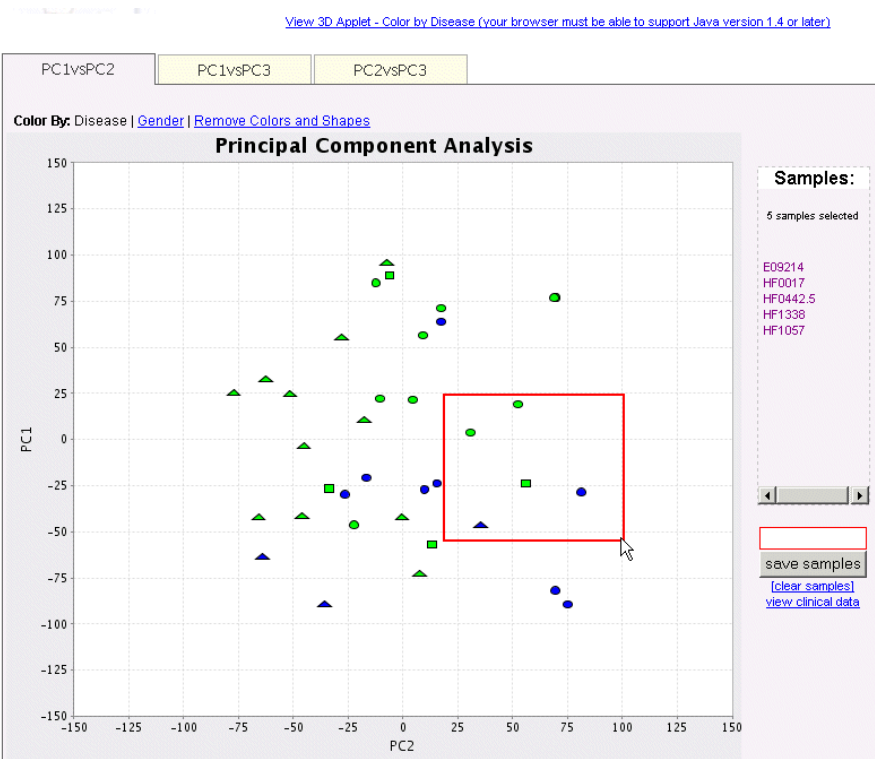


Figure 5.27 Principal Component Analysis report

[Viewing a Three-dimensional Principal Component Analysis \(PCA\)](#)

[Viewing Clinical Plots](#)

[Performing a Principal Component Analysis](#)

**Viewing a Three-dimensional Principal Component Analysis (PCA)**

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

[Table 5.1](#)The following table provides general tasks you can perform with the three-dimensional PCA view:

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

Table 5.1 Tasks for the Three-dimensional PCA view

Don't know which version you have installed (if any)? Need help installing the plugin? Visit this site: <http://www.java.com/en/download/help/testvm.xml> for details.

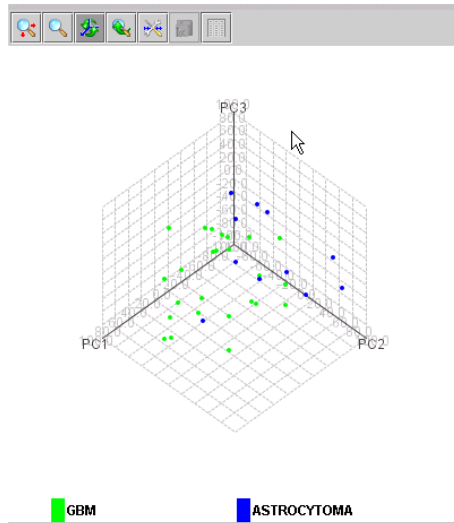


Figure 5.28 Three-dimensional PCA

**Table 5.2** The following table describes the icons:

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

Table 5.2 Three-dimensional PCA instructions




Icon	Special Instructions
	<p>Displays a spreadsheet view of the data with each dataset in a separate tab. Click the button, and the spreadsheet view appears. From the spreadsheet view, you can perform the following tasks:</p> <ul style="list-style-type: none"> <li>To remove a dataset, uncheck <input type="checkbox"/> <b>visible</b>. All data from the set is removed from the main plot.</li> <li>To copy the data to the clipboard, click .</li> <li>To change the color of a data group, click the color box next to <input checked="" type="checkbox"/> <b>visible</b> .</li> </ul>

Table 5.2 Three-dimensional PCA instructions

## Hierarchical Clustering Report

The Hierarchical Clustering report ([Figure 5.29](#)) displays the dendrogram from the hierarchical clustering analysis. To display the image at full resolution, click on **full size** at the top left-hand corner of the graph. Based on the cluster parameter that you select, the report displays either sample or reporter annotations beneath the dendrogram.

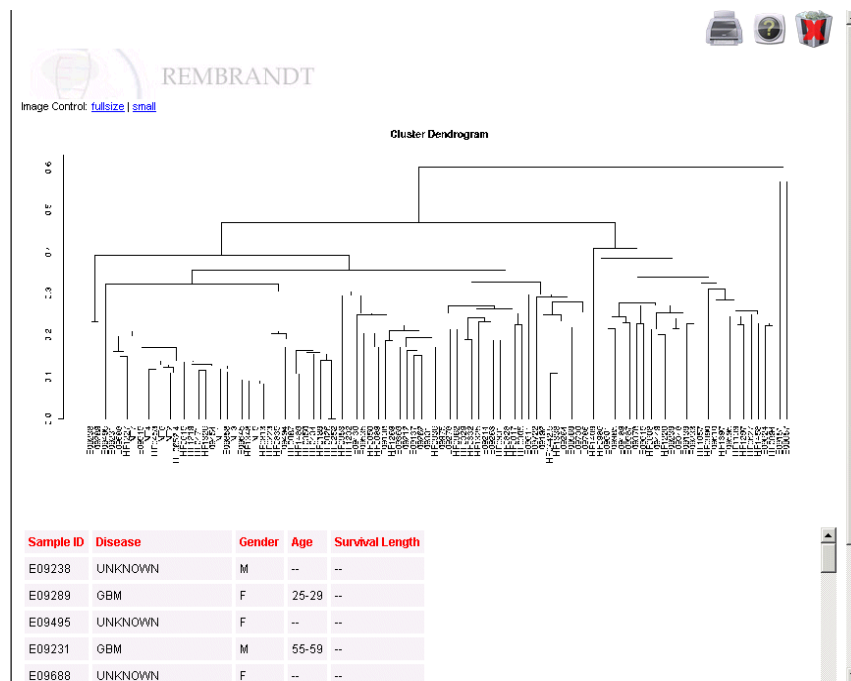


Figure 5.29 Hierarchical Clustering report

To perform a Hierarchical Clustering analysis, see [Performing Hierarchical Clustering Analysis](#).

## GenePattern Analysis Reports

See [High Order Analysis Overview](#).

If you have launched a GenePattern Analysis from the High Order Analysis tab, REMBRANDT directs you to the View Results tab where you can monitor the status of the task you have initiated. (*Figure 5.30*). From this tab, you can also launch GenePattern itself to perform tasks within that application. (See [step 5](#) on page 85.)

#### Rembrandt View GenePattern Job

The screenshot shows the 'View Results' tab in the Rembrandt application. At the top, there are navigation tabs: 'Simple Search', 'Advanced Search', 'High Order Analysis', 'View Results' (selected), 'Manage Lists', and 'Download'. Below these are links for 'Report Results' and 'GenePattern Job Results'. The main content area is divided into two sections. The top section, 'GenePattern Modules [?]', contains a form with 'GP job Number' and 'GP Process' dropdown menus, a 'Heat Map Viewer' dropdown, and a 'go' button. The bottom section, 'GenePattern Job Results [?]', displays the status of a job. It starts with 'Checking status...' in red. The text explains that the request has been sent to GenePattern and provides the job ID '3648'. It states that the data will be ready for analysis in GenePattern and that the approximate processing time is 2-3 minutes. Below this, a list shows 'GenePattern Job# 3648 (test ibh)' with a status of 'running' in red. At the bottom, it says 'All available GenePattern jobs' and provides instructions to click the link to launch GenePattern or a button to launch the Visualizer, and to refresh the page if the task does not appear in the sidebar.

*Figure 5.30 View Results tab displays the status of the GenePattern job you launched*

The GenePattern Job Results (lower panel) in the View Results tab indicates that your query has been sent to GenePattern. It displays the number assigned to the job and its current status. When the status shows **Completed**, the job number appears in the Gene Pattern Modules section in the upper portion of the page.

In the GenePattern Modules section, you can select your data and an analysis method by which GenePattern will analyze it.

1. Select a completed job by number. If you have more than one job in the queue, select from the drop down list the data to be analyzed.



- Select the analysis (described in [Table 5.3](#) the following table) that you want GenePattern to perform on your data.

<b>Analysis Method</b>	<b>Description</b>
<b>HeatMap Viewer</b>	Displays values in a heat map format where the largest values are displayed as the reddest (hot), the smallest values are displayed as the bluest (cool), and intermediate values are a lighter color of either blue or red.
<b>Hierarchical Clustering</b>	Genes or other expression data are clustered according to predetermined algorithms.
<b>K-Nearest Neighbors</b>	Classifies a sample by assigning it the label most frequently represented among the k nearest samples.
<b>Comparative Marker Selection</b>	Finds the genes in a dataset that are most closely correlated with the two phenotypes (ALL and AML).

*Table 5.3 Gene expression data analyses performed by GenePattern from REMBRANDT*

- Click **Go** to launch the analysis.

The Gene Pattern Modules section of the page displays a message telling you to be patient. Your results should be finished shortly ([Figure 5.31](#)).

**Gene Pattern Modules**

Please be patient, the viewer is loading...  
The viewer requires JVM 1.5 or above.

GP job Number    GP Process

3463    Heat Map Viewer    go

*Figure 5.31 GenePattern Modules analysis results monitor*

REMBRANDT displays the following results, based on your analysis selection:



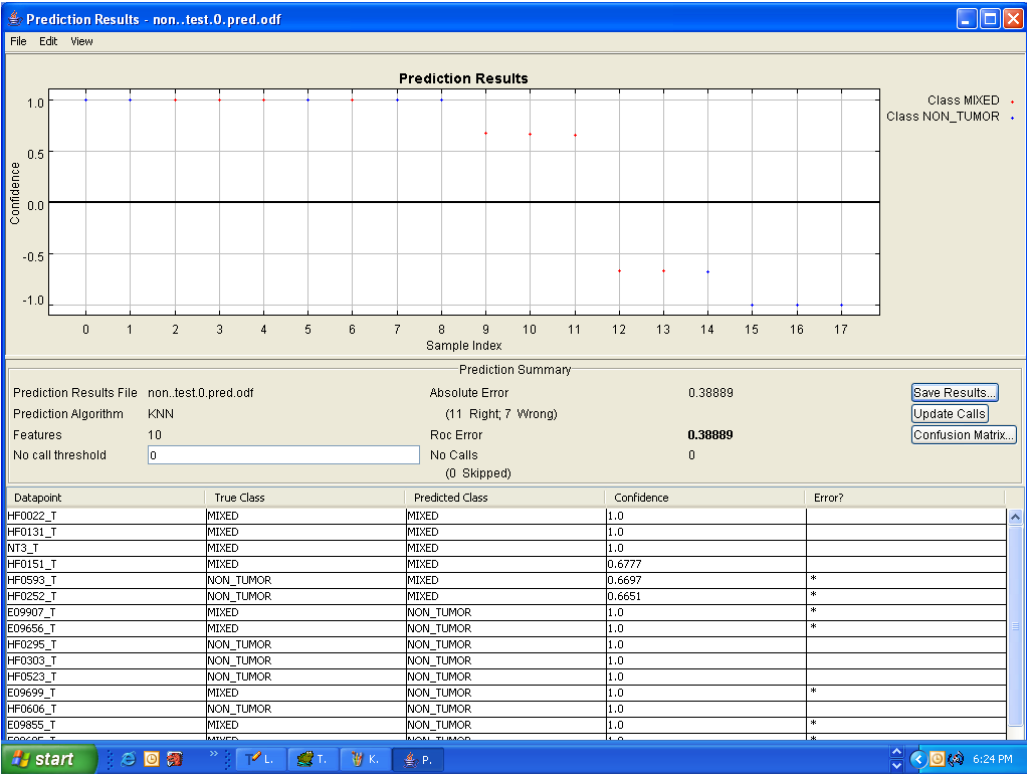


Figure 5.34 Example of K-nearest neighbors result displayed as a GenePattern analysis result

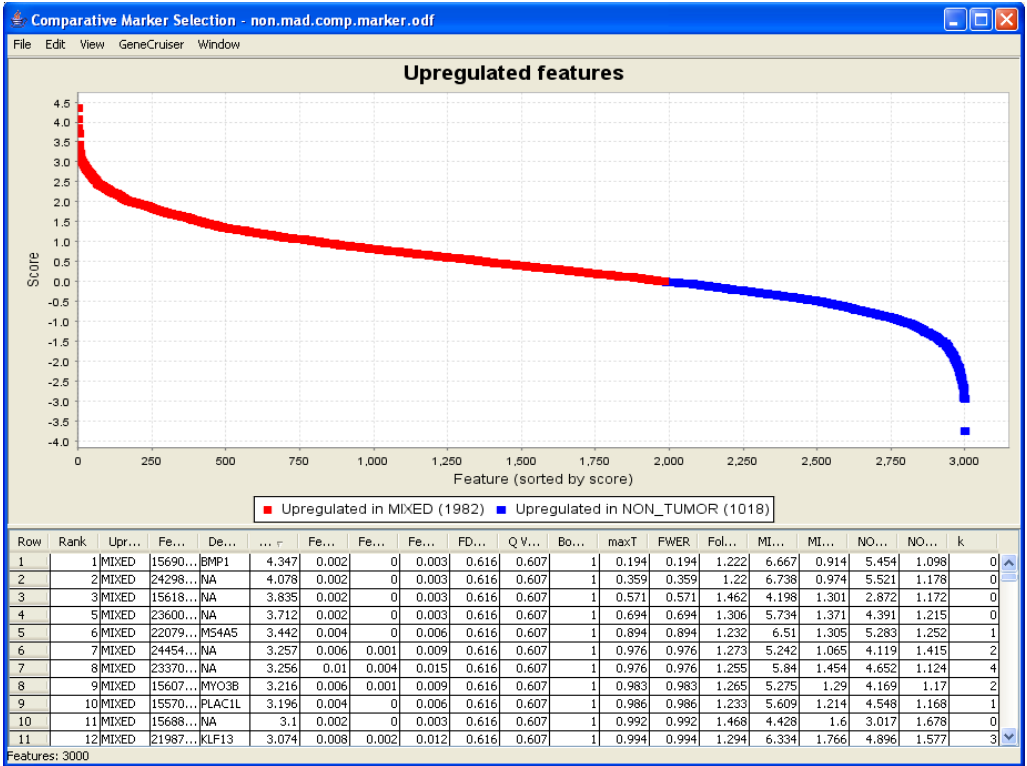



Figure 5.35 Example of Comparative Marker Selection (CMS) result displayed as a GenePattern analysis result

4. When your task is complete, the task will be ready for the visualizer. Launch the associated visualizer by clicking the (  ) icon next to the job link.
5. If you click on the job hyperlink, GenePattern is launched.



# CHAPTER 6 DOWNLOADING DATA

This chapter describes two options available in REMBRANDT for downloading data.

Topics in this chapter include the following:

- [Downloading caArray Tools and Files](#) on this page
- [Downloading BRB Array Tools and Files](#) on page 88

## Data Download Overview

---

REMBRANDT enables you to download data from the caArray application developed by the NCI Center for Biomedical Informatics and Information Technology. It also enables you to analyze data using BRB-ArrayTools created by the Biometric Research Branch of the National Cancer Institute.

### Downloading caArray Tools and Files

caArray guides the annotation and exchange of array data using a federated model of local installations whose results are sharable across the cancer Biomedical Informatics Grid (caBIG®).

To download caArray data files, follow these steps.

1. Select the **Download** tab.

The page that opens displays a section for downloading caArray data ([Table 6.1](#)).

The screenshot shows a web interface titled "caArray [?]". It contains three steps for downloading data:

- Step 1: Choose the saved List:** A dropdown menu with "ASTROCYTOMA" selected.
- Step 2: Choose Array Platform:** A dropdown menu with "Oligo (Affymetrix U133 Plus 2.0)" selected.
- Step 3: Select file type to download:** A dropdown menu with "CEL" selected.

At the bottom of the form is a "download" button.

Table 6.1 Download tab

2. Select the download criteria described in Table 6.1:

Criteria	Special Instructions
<b>Step 1. Select Data</b>	Select from the drop-down list the data set corresponding to the dataset you want to download from caArray
<b>Step 2. Select Array Platform</b>	Select the array platform. For the current version of REMBRANDT, Oligo U133 Plus 2.0 and 100 K SNP gene Chip are the two options.
<b>Step 3. Select File Type</b>	Select the file type to download. The options are .chp and .cel.

Table 6.2 Criteria for downloading data

3. Click the **Download** button.

The Download Results section of the page monitors the download status. You can stop the download process at any time by clicking the **Stop Updating** link.

Upon completion, REMBRANDT generates a hyperlink to a zip file which contains the downloaded file from caArray.

## Downloading BRB Array Tools and Files

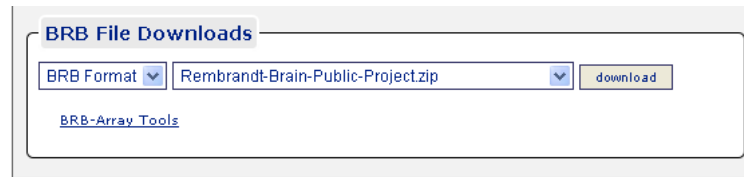
REMBRANDT enables you to analyze data using BRB-ArrayTools created by the Biometric Research Branch of the National Cancer Institute. BRB-ArrayTools is a comprehensive state-of-the-art statistical analysis system for microarray gene expression data. It is free for non-commercial purposes and can be licensed 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, select the **Download** tab.
2. The the lower portion of the page displays the **BRB File Downloads** section.



*Figure 6.1 Downloading BRB Array Tools*

The format and name of the available file(s) are predetermined.

3. Click the **BRB-Array Tools** link for more information about using the BRB website and tools. On the web site that appears, select and download the appropriate version of the product, and follow the prompts.
4. Once you have downloaded and installed BRB-Array Tools, select the files to download to analyze a dataset with BRB-Array Tools.
5. Click the **Download** button.
6. Unzip the REMBRANDT static BRB archive file(s).
7. Open the project worksheet in Excel on a Microsoft Windows PC.





## CHAPTER 7

# MANAGING YOUR WORKSPACE

This chapter describes how to manage your workspace by editing and or reorganizing existing lists and queries, adding new items or creating new custom lists from existing lists and queries. Additionally, this chapter describes import and export functions, allowing you to share lists and queries with others.

Topics in this chapter include:

- *Managing Lists Overview* on page 91
- *Organizing Existing Lists and Queries* on page 97
- *Importing Lists and Queries* on page 98
- *Exporting Lists and Queries* on page 99

## Managing Lists Overview

---

The REMBRANDT My Workspace tab centralizes all activities pertaining to the creation and management of user-defined, as well as study-defined, *PatientDID lists*, *Gene lists*, and *Reporter lists*, as well as queries that you have saved in Rembrandt. Lists and queries are managed in the same way. You can further refine queries or facilitate analysis. You can also organize, import and export lists and queries. When working with lists and queries on the My Workspace/Manage Lists page, you can minimize the number of lists displayed by clicking the PatientDID Lists, Gene Lists, and Reporters Lists heading.

---

**Tip:** 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.

---

## Adding New Lists

*Table 7.1* lists how to add each list type to REMBRANDT.

<b>List Type</b>	<b>How to Add a List</b>
PatientDID List	REMBRANDT provides a collection of default <a href="#">PatientDID Lists</a> . To create a new PatientDID list, see the following: <ul style="list-style-type: none"> <li>• <a href="#">Combining Existing Lists to Create a New List</a> on page 92</li> <li>• <a href="#">Uploading a List</a> on page 93</li> <li>• <a href="#">Manually Entering a List</a> on page 94</li> <li>• Save patients from a Clinical report. See <a href="#">Clinical Reports</a> on page 60.</li> <li>• Save patients from a Gene Expression Sample report. See <a href="#">Selecting and Saving Sample Results (Select Samples Toolbar)</a> on page 68.</li> </ul>
Gene List	To add a <a href="#">Gene List</a> see the following: <ul style="list-style-type: none"> <li>• <a href="#">Combining Existing Lists to Create a New List</a> on page 92</li> <li>• <a href="#">Uploading a List</a> on page 93</li> <li>• <a href="#">Manually Entering a List</a> on page 94</li> </ul>
Reporter List	To add a <a href="#">Reporter List</a> , see the following: <ul style="list-style-type: none"> <li>• <a href="#">Combining Existing Lists to Create a New List</a> on page 92</li> <li>• <a href="#">Uploading a List</a> on page 93</li> <li>• <a href="#">Manually Entering a List</a> on page 94</li> <li>• Save Reporters from a Class Comparison report. See <a href="#">Selecting and Saving Reporters (Select Reporters toolbar)</a> on page 76.</li> </ul>

*Table 7.1 Adding REMBRANDT lists*

### Combining Existing Lists to Create a New List

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

1. Select the **My Workspace** tab.  
On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.
2. Click the type of list you would like to view (**PatientDID List**, **Gene List**, **Reporter List**). The names for the lists appear.
3. Click the box next to the list name(s) to be combined to create a new list.  
**Note:** You cannot select more than two lists to use the **Difference** option.

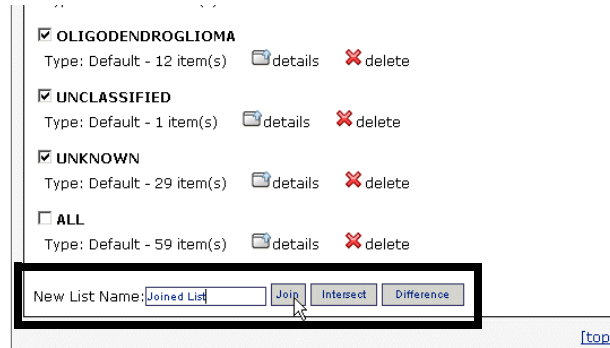


Figure 7.1 Combining existing lists

4. Enter a unique name for the new list you are creating, and then click the appropriate button:
  - **Join** combines two or more lists into one new list.
  - **Intersect** creates a new list from only the items that appear on more than one selected list.
  - **Difference** creates up to two lists each comprising items that appeared in one of the selected lists. For example, if you select **Astrocytoma** and **GBM**, the new lists are “Astrocytoma-GBM” comprising the items that appeared in the Astrocytoma list only and “GBM-Astrocytoma” comprising the items appearing in the GBM list only (*Figure 7.2*).

The new list appears on the My Workspace page and in the side bar in red (*Figure 7.2*).

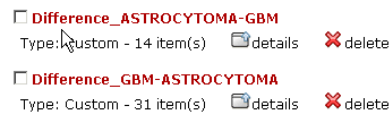


Figure 7.2 New Difference lists

## Uploading a List

You may add a new list type by *uploading* a list from your computer.

---

**Note:** This feature is for creating a new list from scratch, one item per line in a simple text file. This is in contrast to the the import file feature which imports files save in XML format. See *Importing Lists and Queries*.

---

To upload a list, follow these steps:

1. Select the **My Workspace** tab.

On the page, Rembrandt displays the following sub-categories as hyptertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.

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

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

3. Click **Upload List** at the top of the box (*Figure 7.3*).

*Figure 7.3 Uploading a list*

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

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

## Manually Entering a List

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

1. Select the **My Workspace** tab.

On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.

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

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

3. Click **Manually Type List** at the top of the box (*Figure 7.4*).

*Figure 7.4 Manually typing a list*

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

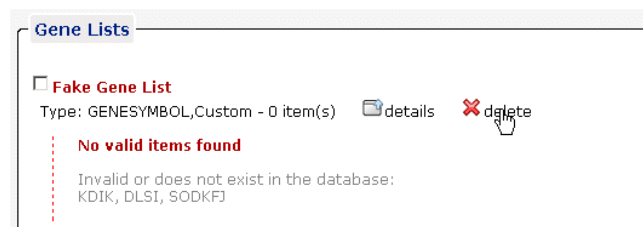
**Table 7.2** lists examples of correctly formatted codes for each list type.

List Type	Correctly Formatted Examples
PatientDID	CB160831 K03193
Gene-GENBANK_ACCESSION_NUMBER	AF125253 S75264
Gene-GENESYMBOL	BPIL2 IVL
Gene-LOCUS_LINK	10 100 10017
Reporter-AFFY_GHU133PLUS2_PROBE_SET	1007_s_at 1053_at
Reporter-IMAGE_CLONE	IMAGE:1407831 IMAGE:143995
Reporter-DBSNP	rs1000015 rs1000025
Reporter-AFFY_100K_SNP_PROBE_SET	SNP_A-1708471 SNP_A-1655302

*Table 7.2 List type code formats*

5. In the **Type Ids** box, enter items into the text block by typing them one to a line.
6. Enter a unique name for the list, and then click the **Add List** button.  
The new list name appears under the appropriate category on the My Workspace page and in the side bar in red.
7. To display the values in the list, click **Details**.

**Note:** If the format of the values entered in the **Type Ids** box was not correct, you must **Delete** the list and start again ([Figure 7.5](#)).



*Figure 7.5 Invalid list*

## Viewing the Data Items in a List

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

1. Select the **My Workspace** tab.

On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.

- At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID Lists**, **Gene Lists**, **Reporter Lists**).
- Find a list to be viewed, and click the **details** icon to display all of the items in the list (*Figure 7.6*).

### Rembrandt Manage Lists

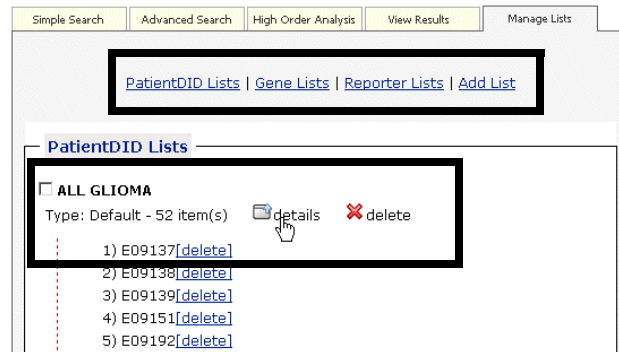


Figure 7.6 List types and Details

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

To export the list, see *Exporting Lists and Queries* on page 99.

## Removing Data Items

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

- Select the My Workspace tab. On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.
- At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List**, **Gene List**, **Reporter List**).
- Find the list you want to change, and click on the box next to the list name.
- Click the **details** icon to display all the items in the selected list.
- Click the **delete** link beside the item you want to delete. The item is removed from the list (*Figure 7.7*).



Figure 7.7 Deleting data items

Once you remove the items, you can view the new list or export the list to your computer. See [Exporting Lists and Queries](#) on page 99.

## Deleting an Entire List

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

1. Select the My Workspace tab. On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.
2. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List**, **Gene List**, **Reporter List**).
3. 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.
4. To delete the selected lists, click an **x delete** icon. The selected categories are removed ([Figure 7.8](#)).

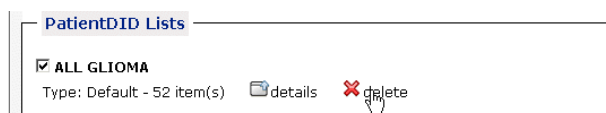


Figure 7.8 Deleting an entire list

The list(s) are deleted.

**Note:** This delete function is completely different than deleting lists and queries in the Organize function. For clarification, see [Organizing Existing Lists and Queries](#) on page 97.

## Organizing Existing Lists and Queries

To organize lists and queries, follow these steps:

1. Select the **My Workspace** tab.  
On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.



2. At the top of the Manage List page, click the **Organize** hypertext link. This opens the Organize Workspace page (*Figure 7.9*).

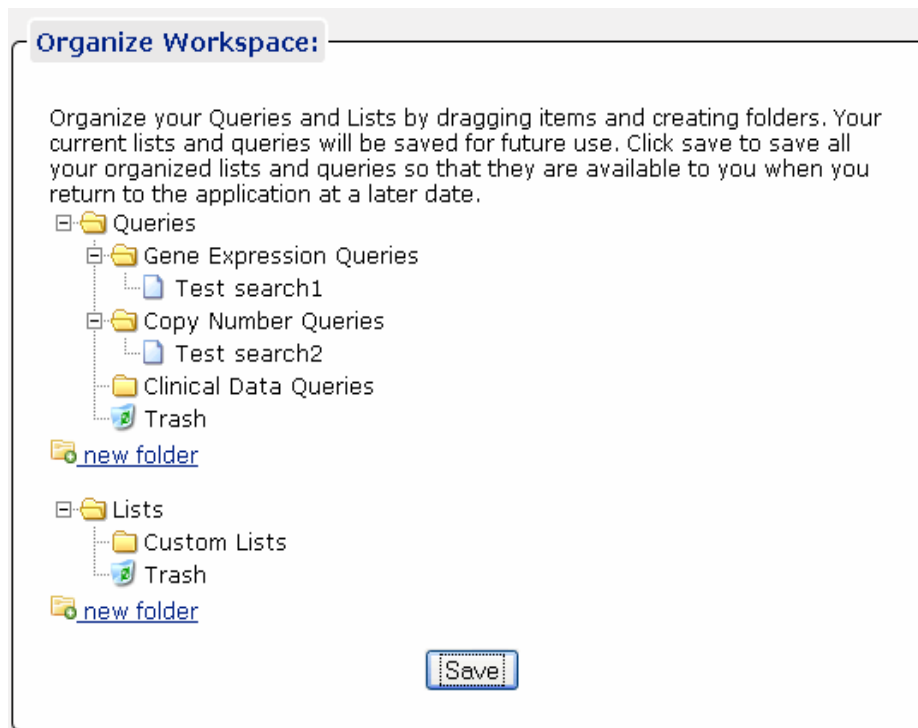


Figure 7.9 Organize Workspace page showing several folders in the hierarchy

3. The Organize Workspace text describes how you can organize lists and queries displayed on the page.
  - a. Create a new folder by clicking the **New Folder** link.
  - b. To name the folder, double click the folder. Enter the name in the text box.
  - c. Click and drag items to new locations, reorganizing the directory structure.
  - d. To delete items from the hierarchy, drag and drop them in the Trash bin.
4. Click **Save** to save the new hierarchy.
  - If you navigate to another location in Rembrandt without saving your “organized” lists and queries, Rembrandt prompts you to click **Save**. If you do not confirm that you want to save them, your organization will be lost.
  - When you click **Save**, the Trash bin empties, and any empty folders in the structure are deleted.

## Importing Lists and Queries

To import a list or a query that you have saved in XML format at an external location, follow these steps:

1. Select the **My Workspace** tab.

On the page that opens, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.

2. At the top of the Manage List page, click the **Import** hypertext link.  
The Import Workspace text describes how you can import lists and queries.
3. At the bottom of the page, select the file type you are importing, **Query** or **List**.
4. Click the **Browse** button to navigate for the file(s).
5. Click the **Import** button.

Rembrandt informs you if the file you are trying to import does not comply with acceptable format standards. If it does meet requirements, the system opens the Organize Workspace page where you will see the imported file displaying in a new “Imported” folder named with the date and time stamp. If a file with an identical name already exists in Rembrandt, the program renames the imported file (*Figure 7.10*).

Rembrandt Manage Workspace

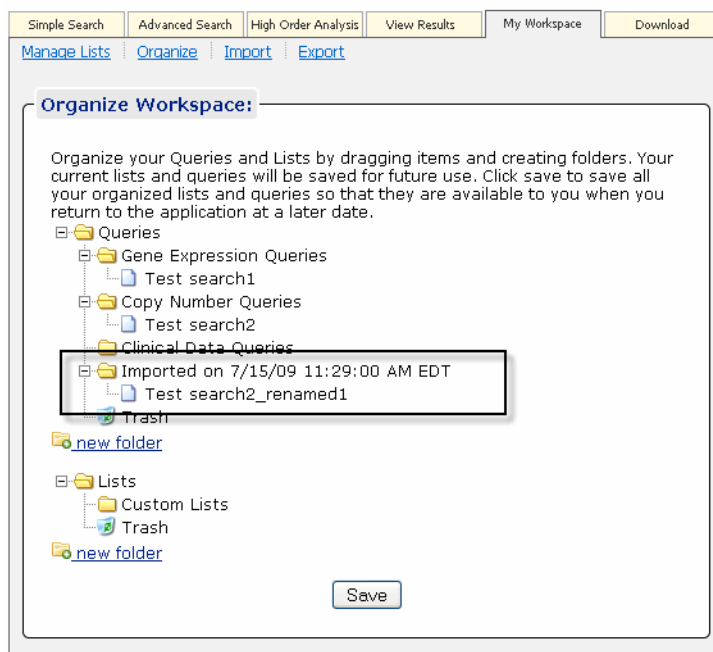


Figure 7.10 The Organize Workspace page displays imported file folders. Rembrandt renames files where duplicates exist in the database.

---

**Note:** You can import only one file at a time; import of a “batch” of files is not possible.

---

## Exporting Lists and Queries

---

To export a list, follow these steps:

1. Select the **My Workspace** tab.

On the page, Rembrandt displays the following sub-categories as hypertext links: Manage Lists, Organize, Import, and Export. By default, the page opens to the Manage Lists subcategory.

2. At the top of the Manage List page, click the **Export** hypertext link.
3. In the directory, locate and click on the query or list you want to export. You can select one file in an existing folder or you can select an entire folder. In each case, your selection is exported as an one XML file. If you are exporting a folder of multiple files, they will all be part of one large XML file.
4. This opens an Opening... dialog box which indicates the file and file type. Select the option to open or to save the file on your local drive.

---

**Note:** When you select to log out, Rembrandt will prompt you to save lists or queries you may have been working with.

---

# APPENDIX

# A

## GLOSSARY

Acronyms, objects, tools and other terms referred to in the chapters or appendixes of this REMBRANDT 1.5 User's Guide are described in this glossary.

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

*Table A.1 Glossary of REMBRANDT terms*

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

Table A.1 Glossary of REMBRANDT terms

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

Table A.1 Glossary of REMBRANDT terms

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

*Table A.1 Glossary of REMBRANDT terms*

# INDEX

## Symbols

- ( ) parentheses
  - using to group queries 46
- + or - sign
  - click for more options 50

## Numerics

- 3-D Principal Component Analysis plot 77, 78

## A

- adding
  - advanced search 45
  - compound query 45
  - list, combining 92
  - list, typing 94
  - list, uploading 93
  - new list (removing data items) 96
  - query, overview 33
  - simple search 13
- advanced searches 33
  - adding 45
  - clinical study 42
  - compound query 45
  - copying, editing, deleting 45
  - copy number 39
  - gene expression 34
  - overview 33
  - saving samples 68
- Analyses
  - Class Comparison 50
  - Hierarchical Clustering 54
  - High Order Analysis 49
  - Principal Component 52
- Application Support 10
- Average linkage, defined 55

## B

- Baseline, assigning 51
- Box and Whisker Coin plot, description 22
- Box and Whisker Log2 Intensity option 16

- Box and Whisker Log2 Intensity plot

- description 20
  - uses for 20

- BRB-ArrayTools, downloading files 87, 88

## C

- caArray
  - downloading files 87
  - glossary 101
- Choose a Saved
  - Gene List, defined 40
  - Reporters List, defined 35
  - SNP List, defined 40
- citing, REMBRANDT data 4
- Class Comparison
  - Analysis Form, description 50
  - High Order Analysis 50
  - report description (View Results) 73
- Clinical Data page, description 42
- clinical plot
  - saving samples of interest 64
  - selecting samples of interest 63
- clinical plot, description 63
- clinical report
  - description 60
  - saving samples 62
  - selecting samples 61
- Clinical Study, advanced search 42
- columns, hiding diseases 70
- combining lists 92
- Comparative Marker Selection
  - GenePattern results 83
- Complete linkage, defined 55
- compound query
  - creating 45
  - definition 33
  - deleting 45
- Constrain reporters by variance (Gene Vector)
  - percentile, defined 53, 54
- copying, individual query 45
- Copy Number



- advanced search [39](#)
- Data page, description [39](#)
- page description (View Results) [71](#)
- showing data in webGenome [72](#)

Copy Number Data [73](#)

Copy Number Data Per Sample report [71](#)

Copy Number Data Per Sample View [71](#)

creating

- compound query [45](#)
- list, combining [92](#)
- list, typing [94](#)
- list, uploading [93](#)
- new list (removing data items) [96](#)
- query, overview [33](#)
- user account [6](#)

## D

data, citing REMBRANDT datasets [4](#)

data analyses, GenePattern [55](#)

data items

- removing from a list [96](#)
- viewing in a list [95](#)

Delete All Queries [45](#)

deleting

- compound query [45](#)
- individual query [45](#)
- list [97](#)

Difference option, description [93](#)

displaying LOH data in webGenome [73](#)

Distance Matrix, defined [55](#)

downloading

- BRB files [87](#), [88](#)
- caArray files [87](#)
- samples to Excel [63](#)

## E

editing, individual query [45](#)

e-mailing, plots or graphs [24](#)

Euclidean distance, defined [55](#)

Excel, downloading samples to [63](#)

exporting

- list [99](#)
- samples to Excel [63](#)

## F

feedback about REMBRANDT [11](#)

filtering

- p-value (Filter p-value Toolbar) [75](#)
- reports [75](#)
- reports (Filter toolbar) [66](#)

Filter p-value Toolbar [75](#)

Filter Toolbar [66](#)

Finalize Query button, purpose [33](#)

F-test One Way ANOVA

- defined [51](#)
- output [74](#)

## G

Gene Expression

- advanced search [34](#)
- Data Per Sample View [65](#)
- Plot page, description [15](#)
- simple search [14](#)

Gene Expression plot, resizing [24](#)

Gene Expression Plot page

- Box and Whisker Log2 Intensity plot details [20](#)
- Geometric Mean plot details [18](#)
- Log2 Intensity plot details [19](#)
- Median plot details [17](#)

Gene Expression Sample page [70](#)

- description (View Results) [65](#)

Gene Expression Sample report [65](#)

GenePattern

- Comparative Marker Selection report [83](#)
- HeatMap Viewer report [83](#)
- hierarchical clustering report [83](#)
- initiating analyses from REMBRANDT [55](#)
- K-nearest neighbors report [83](#)
- launching [57](#)
- opening analysis report visualizer [85](#)

GenePattern analysis report

- description [81](#)

GenePattern Job report page [81](#)

Geometric Mean

- Gene Expression plot [18](#)
- option, defined [16](#)

Glossary [101](#)

Graph Type, defined [16](#)

grouping queries [46](#)

## H

HeatMap Viewer, GenePattern report [83](#)

help, obtaining [8](#)

Help link [8](#)

Hide Diseases toolbar [70](#)

Hierarchical Clustering

- Analysis Form, description [54](#)
- GenePattern results [83](#)
- High Order Analysis [54](#)
- page description (View Results) [80](#)
- report, description [80](#)

highlighting

- results by value (Highlight Toolbar) [67](#)

Highlight Toolbar [67](#)

High Order Analysis [49](#)

Class Comparison 50  
 Hierarchical Clustering 54  
 Principal Component 52

## I

importing lists 98  
 individual queries  
   adding 45  
   copy, editing, deleting 45  
 Intensity, defined 17, 18, 19  
 Intersect option, description 93

## J

Join option, description 93

## K

Kaplan-Meier for Copy Number-based Data  
   plot description 29  
   redrawing plot 29  
   searching steps 28  
 Kaplan-Meier for Gene Expression Data  
   plot description 25  
   redrawing plot 25  
   simple search 24  
 Kaplan-Meier for Sample Data  
   plot description 31  
   simple search 31  
 Karnofsky Clinical Evaluation, defined 44  
 Karnofsky score, defined 63  
 K-nearest neighbors GenePattern results 83

## L

Lansky Clinical Evaluation, defined 44  
 launching  
   GenePattern 57  
   REMBRANDT 4  
 Legal Rules of the Road page 7  
 Linkage Method, defined 55  
 lists  
   combining lists 92  
   creating a new list 96  
   deleting 97  
   exporting 99  
   importing 98  
   organizing 97  
   removing data items 96  
   typing a list 94  
   uploading a list 93  
   viewing data items 95  
 list serve, joining 7  
 list serve, leaving 7  
 Log2 Intensity

Gene Expression plot 19  
   option 16  
 logging in, REMBRANDT 7  
 logging out, REMBRANDT 10  
 Logout link 10  
 Logout page 11  
 Log-rank p-value, Statistical Report 31, 32  
 LOH data  
   display REMBRANDT data in webGenome 73  
 Loss of Heterozygosity data  
   display REMBRANDT data in web Genome 73

## M

managing lists, overview 91  
 Manually Type List options, description 94  
 mean, defined 20, 22  
 median  
   defined 20, 22  
   Gene Expression plot 17  
   option, defined 16  
 missing Run Report button 47

## N

NCICB Application Support 10

## O

organizing lists 97  
 overview  
   creating a query 33  
   high order analyses 49  
   managing lists 91  
   simple search 13  
   viewing output 59

## P

parentheses, using to group queries 46  
 PatientDID list, creating 60, 68  
 Pearson correlation, defined 55  
 Please select an All Genes query, defined 46  
 previewing results 37  
 Principal Component  
   Analysis Form, description 52  
   Analysis page, description (View Results) 77  
   High Order Analysis 52  
 Principal Component Analysis report  
   description 77  
   three-dimensional plot 77, 78  
 printing plots or graphs 24  
 Probeset, defined 17, 18, 19  
 purpose of the User's Guide 1  
 p-value  
   defined (for Geometric Mean plot) 17, 18

defined (for Log2 Intensity plot) 19  
Filtering (Filter p-value Toolbar) 75

## Q

Q1, defined 20, 22

queries

- grouping 46
- overview of process 33
- prolonged 60
- refining 46
- saving 13, 34

queries See also lists

## R

redrawing

- K-M Copy Number-based Data 29
- K-M Gene Expression plot 25

Refine Query

- description 46
- option, purpose 33
- page, description 46

REMBRANDT

- citing data 4
- description of 4
- displaying LOH data in webGenome 73
- feedback about application 11
- functions 9
- getting help 8
- launching 4
- logging in 7
- logging out 10
- menu 8
- registering 6
- stands for 4
- workspace 7

Rembrandt

- list serve 7

Rembrandt MyWorkspace page

- description 91

removing data items in a list 96

Reporter list, selecting and saving 76

reports

- Class Comparison 73
- Copy Number Sample 71
- Gene Expression Sample 65
- GenePattern Comparative Marker Selection 83
- GenePattern HeapMap Viewer 83
- GenePattern hierarchical clustering 83
- GenePattern Job 81
- GenePattern K-nearest neighbors 83
- Hierarchical Clustering 80
- Principal Component Analysis 77

resizing Gene Expression plot (in new window) 24

results See lists

Run Report >> button, missing 47

## S

sample

- saving (Select Samples Toolbar) 68
- saving on clinical report 62
- selecting on Clinical report 61

samples of interest

- saving on Clinical plot 64
- selecting on Clinical plot 63

saving

- a query 13, 34
- plots or graphs 24
- reporters (Select Reporters Toolbar) 76
- sample of interest on clinical plot 64
- samples on Clinical report 62

searches

- advanced overview 33
- simple overview 13

Select a data source, defined 47

selecting

- reporters (Select Reporters Toolbar) 76
- sample of interest on Clinical plot 63
- sample on Clinical report 61

Select Reporters toolbar 76

Select Samples Toolbar 68

Show All Values Toolbar 69

side bar

- manage list information 96
- managing queries 45

simple searches 13

- Gene Expression steps 14
- K-M Copy Number-based Data 28
- K-M Gene Expression Data 24
- K-M Sample Data 31

Single linkage, defined 55

sorting report columns 77

specimen type 41

Statistical Method, defined 51

Std. Dev., defined 19

submitting

- feedback about REMBRANDT 11

Support link 8

Survival vs Age at Dx, defined 63

## T

Three-dimensional PCA plot 77, 78

Triangles, top of columns to sort 77

T-test Two Sample Test

- defined 51
- output 73

Tutorials link 8

typing a list 94

## U

Unified link, defined 15

uploading a list 93

user's guide

- organization of 1
- purpose 1

user account, creating 6

user guide link 8

## V

Validate Query button

- defined 47
- required 47

View 3D Applet - Color by Disease link 77

viewing data items in a list 95

View Results page, description 59

views

- Copy Number Data Per Sample view 71
- Gene Expression Data Per Sample view 65

## W

webGenome

- display Copy Number data 72
- display LOH data 73

Wilcoxon Test Man-Whitney Test

- defined 51
- output 73

