CANCER MOLECULAR ANALYSIS PORTAL

User's Guide

January 7, 2009

TABLE OF CONTENTS

About This Guide

This chapter introduces you to the *Cancer Molecular Analysis Portal User's Guide*. It includes the following topics:

Purpose

This guide provides an overview of the Cancer Molecular Analysis Portal (CMA Portal) and instructions for using its tools and resources for querying and analyzing patient tissue *sample* data. This book is organized into chapters that parallel the CMA Portal'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 CMA Portal, read this brief overview, which explains what you will find in each chapter.

- *Chapter 1* provides information for launching, logging in to, and navigating the CMA Portal.
- *Chapter 2* describes how to use the CMA Portal to research gene expression, copy number, SNP, LOH, and pathway data.
- *Chapter 3* describes how to use the CMA Portal to review pathways and the anomalies associated with them.
- *Chapter 4* describes how to use the CMA Portal to investigate chromosomal regions of amplification, deletion, and over/under-expression.

- *Chapter 5* describes how to use the CMA Portal to study clinical data and explore the relationships between clinical and molecular study data.
- *Chapter 6* describes how to use the CMA Portal to perform principal component and gene pattern analyses, and to access associated research tools.
- *Chapter 7* describes how to select, create, and manage patient and/or gene lists.
- *Glossary* defines terms in this guide.

Text Conventions Used

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

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

Credits and Resources

The following people contributed to the development of this document.

Cancer Molecular Analysis Portal Development and Management Teams		
Development	***Documentation***	***Project and Product Management***
Alex Jiang[5]	Lauren Anthone [2]	Anand Basu[1]
David Nassau[3]		Daniela Gerhard[1]
Erin Hickey[3]		Subha Madhavan[1]
Huaitien Liu[5]		Carl Schaefer[1]
Mark Lewis[5]		Shine Jacobs[4]
Michael Harris[5]		Jianghui Zhang[1]
Mike Edmondson[1]		
Richard Finney[1]		
Robert Sfeir[3]		
Ryan Landy[5]		
Ying Long[5]		
[1] National Cancer Institute Center for Biomedical Informatics and Information Technology (NCI-CBIIT)	[2] Lockheed Martin [3] SRA International, Inc.	[4] Enterprise Solutions And Consulting (ESAC) [5] Science Application International Corporation (SAIC)

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

Application Support

For any general information about the application, application support, or to report a bug, contact National Cancer Institute Center for Biomedical Informatics and Information Technology (NCI-CBIIT) (formerly, NCICB) Application Support

Email: ncicb@pop.nci.nih.gov	When submitting support requests via email, please include: • Your contact information, including your telephone number • The name of the application/tool you are using • The URL of the application • A description of the problem and steps to recreate it • The text of any error messages you have received

Application Support URL	http://ncicb.nci.nih.gov/NCICB/support
Telephone: 301-451-4384 Toll free: 888-478-4423	Telephone support is available from: Monday to Friday, 8 am – 8 pm Eastern Time, excluding government holidays.

CHAPTER

1

GETTING STARTED WITH THE CANCER MOLECULAR ANALYSIS PORTAL

This chapter provides information for launching, logging in to, and navigating the Cancer Molecular Analysis Portal.

Topics in this chapter include:

About the Cancer Molecular Analysis Portal

The CMA Portal provides access to query and analysis tools; and to data in The Cancer Genome Atlas (TCGA) and REpository for Molecular BRAin Neoplasia DaTa (REMBRANDT).

TCGA is designed to allow the integration of clinical, genomic characterization, and large-scale genomic sequencing data of all samples. TCGA researchers characterize and sequence the genomes of up to 500 samples of both tumor and matched normal tissues.

REMBRANDT is a robust 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. The knowledge framework provides researchers with the ability to perform ad hoc querying and

reporting across multiple data domains, such as Gene Expression, Chromosomal aberrations, and Clinical data.

For more information about the available data, refer to the following websites:

- The Cancer Genome Atlas' (TCGA's) website at: http://cancergenome.nih.gov/components/dmbca.asp
- REpository for Molecular BRAin Neoplasia DaTa (REMBRANDT) at: http://caintegrator-info.nci.nih.gov/REMBRANDT

Launching the Cancer Molecular Analysis Portal

The Cancer Molecular Analysis Portal is a web application.

How to Launch the Cancer Molecular Analysis Portal

1. In Internet Explorer or other browser, type the following CMA Portal URL in the address field: https://cma.nci.nih.gov/cma/index.jsp

   The CMA Portal home page appears *(Figure 1.1)* and displays the Gene View menu by default.

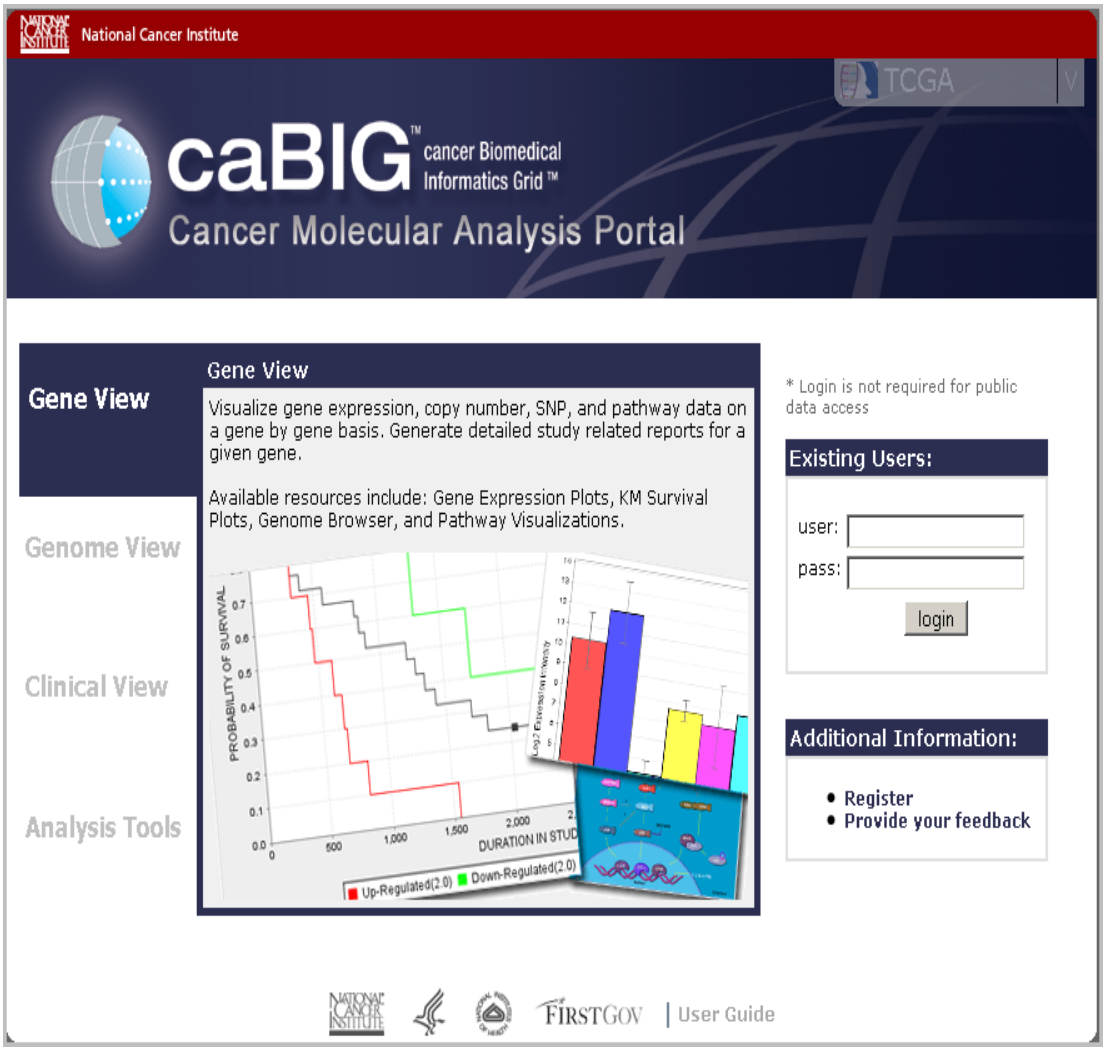


Figure 1.1 Cancer Molecular Analysis Portal – Home Page

Requesting a Username and Password

If you intend to conduct your research using *open-access data* only, you do not need a user account to use CMA Portal resources. However, to take advantage of *controlled-access data* in the Cancer Genome Atlas (TCGA), you must provide a username and password.

How to Request a Username and Password

1. In the **Additional Information** section of the side bar, click **Register**.

   The New Users page appears (*Figure 1.2*).

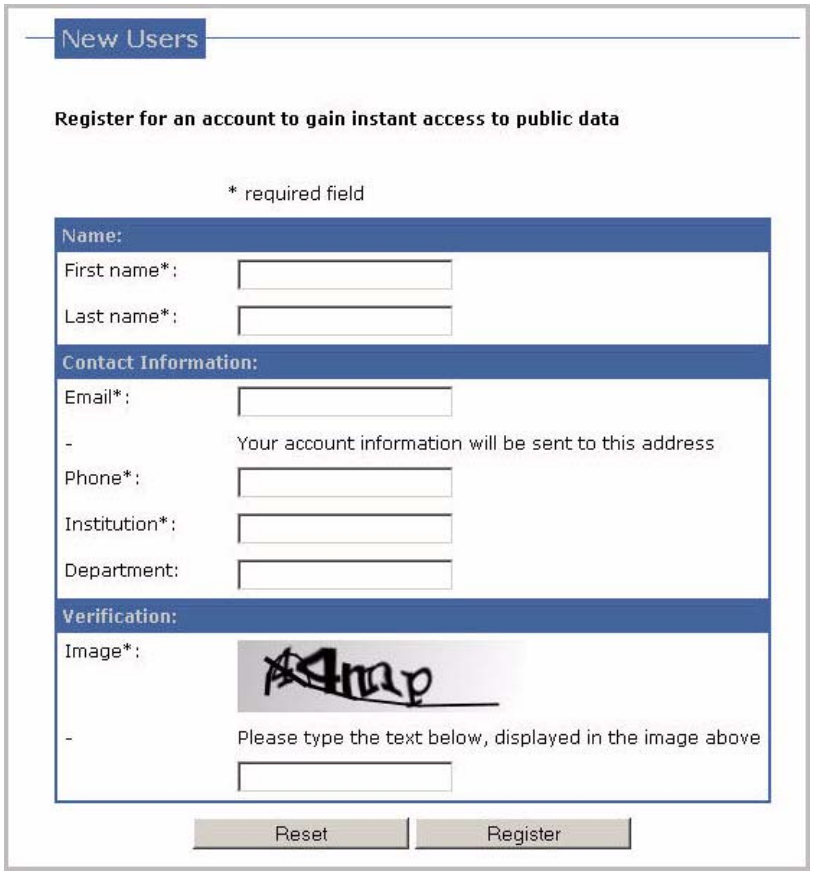


Figure 1.2 New Users Page

2. In the **Name** area, type your first and last names in the text boxes provided.
3. In the **Contact Information** area, type your email address and other contact information in the text boxes provided.

   Note: The system will send your account information to this email address.
4. In the **Verification** area, type the series of letters and numbers in the text box exactly as displayed in the verification image.
5. Click **Register**, or, to clear all text boxes, click **Reset**.

   Note: All registration information requested is required.
6. If you clicked **Register**, the system displays a message indicating that your account information will be sent to you by email.
7. When you receive your account information, follow the instructions in the email message to apply for a Data User Certificate (DUC). For further information about data access and DUCs, see http://cancergenome.nih.gov/dataportal/data/access/closed/duc/.

Related Topics:

- *Logging In to the Cancer Molecular Analysis Portal*
- *Log In to the CMA Portal*

- *Navigating the Cancer Molecular Analysis Portal* on page 10

Logging In to the Cancer Molecular Analysis Portal

If you intend to conduct your research using *open-access data* only, you do not need to log in to the CMA Portal. However, to take advantage of *controlled-access data,* you must provide a username and password. See *Requesting a Username and Password* on page 7.

How to Log In to the CMA Portal

1. In the **Existing Users** side bar on the Home page, type your username and password in the text boxes provided.
2. Click **Log In**.

   A message confirms that you have logged in successfully.

Caution: The system logs you out automatically after ten minutes of inactivity. Because the CMA Portal does not save your custom lists from one session to the next, you will not be able to retrieve your lists after you have been logged out.

Related Topics:

- *Logging Out*
- *Navigating the Cancer Molecular Analysis Portal*

Accepting the Cancer Molecular Analysis Portal Provisions

The Legal Rules of the Road page appears after you have logged in (*Figure 1.3*). This page also appears if you click any of the links on the menu pane (left side of the page) without having logged in first.

Figure 1.3 Legal Rules of the Road Page

Read the provisions, and then click **CLICKING HERE** (*Figure 1.3*).

Navigating the Cancer Molecular Analysis Portal

The panes on the CMA Portal Home page enable quick access to all the workspaces. Each workspace has tools and resources that allow you to investigate genomic data.

Related Topics:

- *Navigating the Cancer Molecular Analysis Portal Home Page*
- *Navigating the Cancer Molecular Analysis Portal Workspaces*
- *Menu Pane Links*
- *Overview Pane Features*
- *User Access Pane Features*
- *Tabs Features*
- *Side Bar Features*
- *Administration Section Features*
- *News Section Features*
- *Accessing TCGA Data*

Navigating the Cancer Molecular Analysis Portal Home Page

The Cancer Molecular Analysis Portal home page (*Figure 1.4*) provides links and other information that enable you to navigate through CMA Portal workspaces.

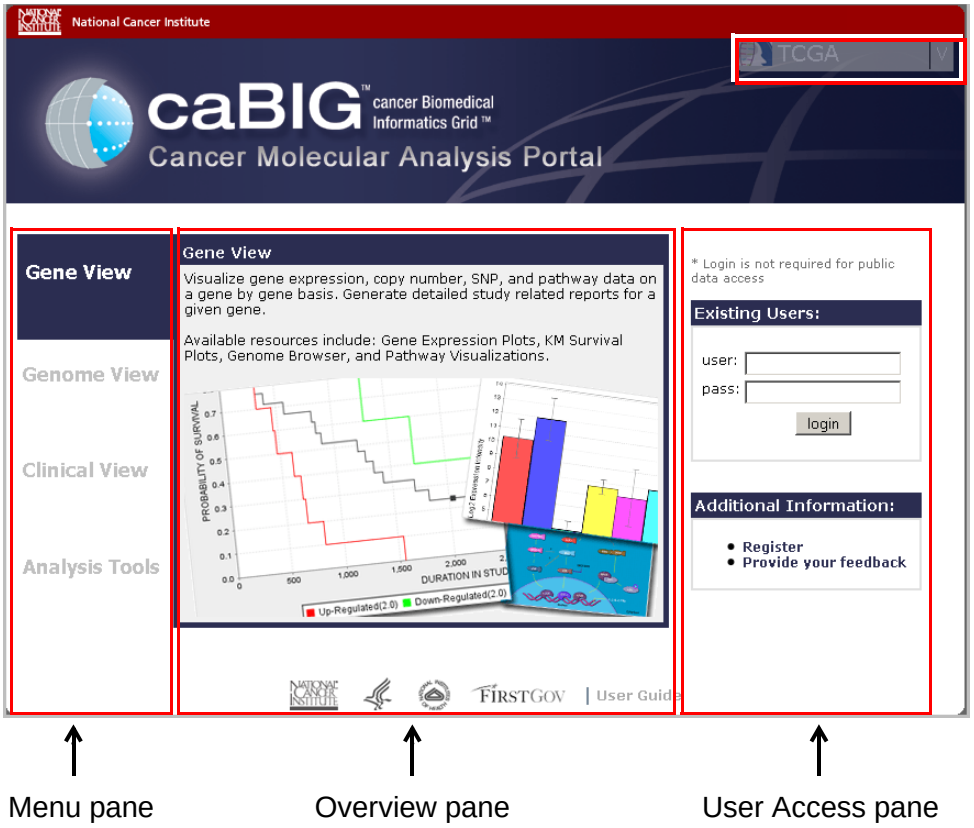


Figure 1.4 The Cancer Molecular Analysis Portal Home Page

Table 1.1 describes each of the panes on the Home page, and the Context drop-down list at the upper right side of the page.

Area	***Description/Function***
Menu pane	Provides links to query and analysis workspaces
Overview pane	Provides a brief overview of each of the workspaces
User Access pane	Login page for existing users. Provides links to register for an account and to provide feedback and request CMAP support.
Context	Indicates the current dataset used for your queries. Currently data from TCGA and REMBRANDT is available for research.

Table 1.1 Cancer Molecular Analysis Portal – Home page features

Menu Pane Links

Each section of the Menu pane acts as link to their respective workspaces. Click a link to access the tools and resources associated with each menu item. *Figure 1.5* illustrates the links to the workspace in the Menu pane.

Figure 1.5 Menu Pane Links

Table 1.2 describes each item on the Menu pane.

Callout Number	***Description/Function***
1	**Gene View** link – Provides access to the gene analysis workspace where you can visualize gene expression, copy number, SNP, and pathway data on a gene by gene basis; and generate detailed study related reports for a given gene.
2	**Genome View** link – Provides access to the genome workspace where you can explore all of the study data in one genome level visualization, and investigate chromosomal regions of amplification, deletion and over expression.
3	**Clinical View** link – Provides access to the clinical analysis workspace where you can investigate and study clinical data, and explore the relationships between clinical and molecular study data.
4	**Analysis Tools** link – Provides access to the analysis tools workspace, where you can analyze the study data using analysis tools such as GenePattern, Principal Component Analysis, and the Cancer Genome Workbench.

Table 1.2 Cancer Molecular Analysis Portal – Menu pane features

Overview Pane Features

The Overview pane provides a brief introduction to each workspace and the tools and resources available for research.

Figure 1.6 is an example of the overview and resource information that is provided for each workspace.

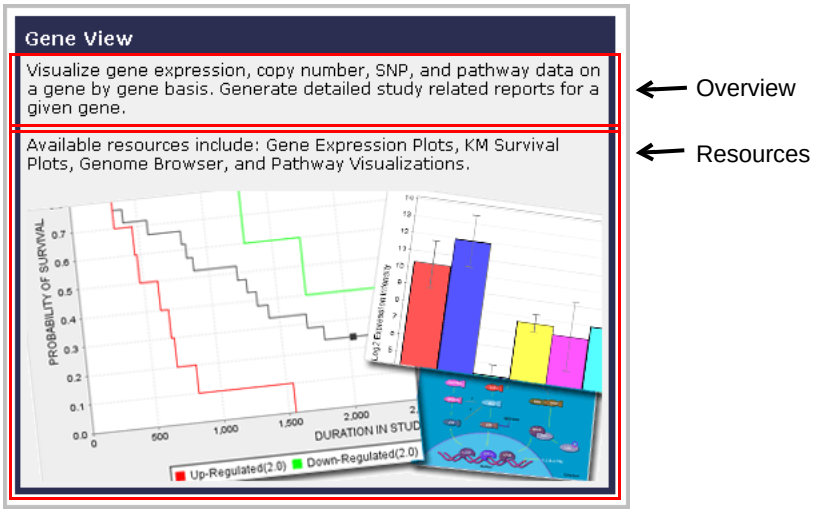


Figure 1.6 Overview Pane – Gene View

User Access Pane Features

The Access pane provides access to *controlled-access data* for users who have registered for an account for TCGA (The Cancer Genome Analysis) data (*Figure 1.7*). New users can register for an account in the Additional Information section.

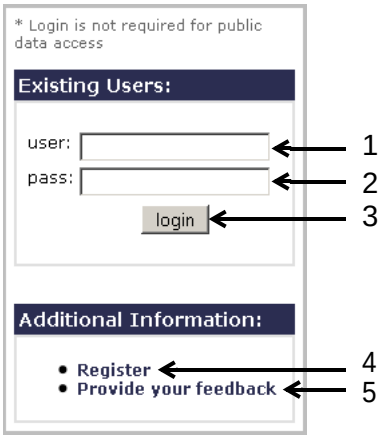


Figure 1.7 User Access Pane

Table 1.3 describes each item on the User Access Pane.

Callout Number	***Description/Function***
1	**User** – Type your username here to log in and obtain complete access to Cancer Molecular Analysis Portal
2	**Pass** – Type your password
3	**login** – Click to log in to the portal
4	**Register** – Click the link to register for access to the controlled data sets

Table 1.3 CMA Portal home page – User Access side bar features

Callout Number	Description/Function
5	**Provide Your Feedback** – Click the link to display the NCI Center for Bioinformatics Support page in a new browser window

Table 1.3 CMA Portal home page – User Access side bar features (Continued) (Continued)

Navigating the Cancer Molecular Analysis Portal Workspaces

Each of the workspaces, Gene View, Genome View, Clinical View, and the Analysis Tools has its own workspace which is displayed whenever you select an item on the Home page menu bar. Each workspace contains tabbed pages, a side bar, and unique tools and resources. The tabs and sidebar are the same for all workspaces.

Tabs Features

Tabs run across the top of each workspace (*Figure 1.8*). They allow you to access all CMAP workspaces.

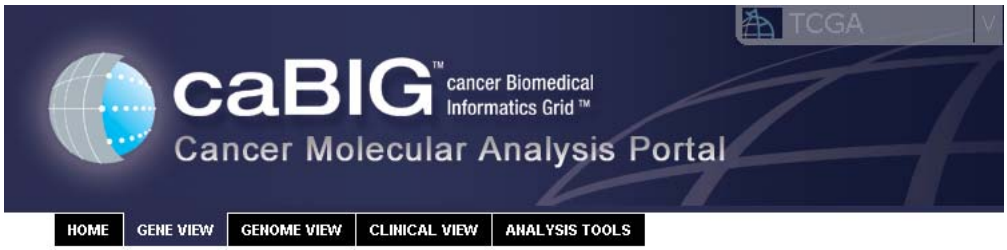


Figure 1.8 Workspace Tabs

Side Bar Features

The side bar appears on the right side of the of each workspace except the Home page (*Figure 1.9*).

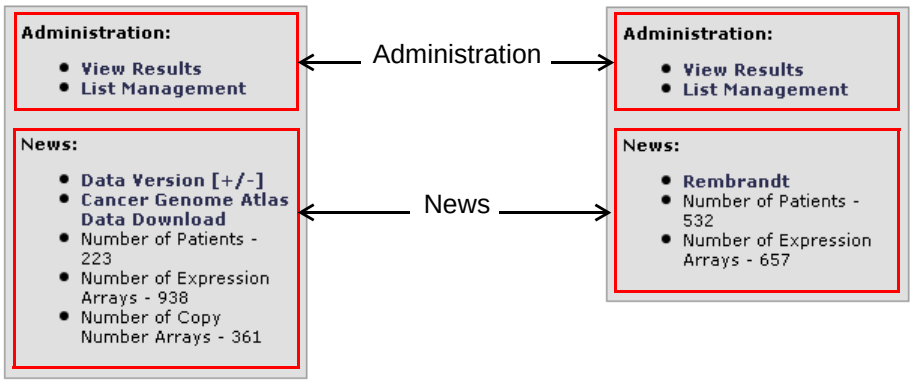


Figure 1.9 Side Bars – TCGA Context (L) and REMBRANDT Context (R)

Administration Section Features

Table 1.4 describes each feature in the Administration section.

Feature	Description/Function
View Results	Provides access to your saved queries.
List Management	Provides access to the default and any custom lists you may have created

Table 1.4 Side Bar – Administration side bar features

News Section Features

Table 1.5 describes each feature in the News section.

Feature	Description/Function
Data Version (TCGA context only)	Provides information about the date and source of the data available for research and a link to a TXT file that contains a list of data source files.
Cancer Genome Atlas Data Download (TCGA context only)	Links to The Cancer Genome Atlas Data (TCGA) portal from which you can search for and download data from TCGA datastores. Also provides access to the Data Access Matrix. See *Accessing TCGA Data* on page 15.
REMBRANDT (REMBRANDT context only)	Launches the REMBRANDT portal, which provides access to diverse types of molecular research and clinical trials data related to brain cancers; and to a variety of web-based analysis tools.
Number of Patients	Number of patients sampled for the dataset
Number of Expression Arrays	Number of arrays used to obtain the available expression data
Number of Copy Number Arrays (TCGA data only)	Number of arrays used to obtain the available copy number data

Table 1.5 Side Bar – News section features

Note: The numbers of patients sampled and arrays used are updated continuously as new data sets from TCGA become available through the Data Coordination Center (DCC). For further information, see TCGA website at: http://cancergenome.nih.gov/components/dmbca.asp.

Accessing TCGA Data

You can access TCGA (The Cancer Genome Atlas) data via the Data Access Matrix (the Matrix) application, which enables you to select results of individual samples from multiple centers, platforms, and data types; or you can search for and download entire data archives via TCGA Portal tools.

How to Access TCGA Data

1. On the sidebar, click the **Cancer Genome Atlas Data Download** link, or, open a new browser and type http://tcga-data.nci.nih.gov/tcga/homepage.htm in the address field.

   The Cancer Genome Atlas Data Portal appears in a new browser window.

2. Do one of the following to access data from in the main **Get TCGA Data** pane:

   - To search for and download complete data archives, click the **search by archive** link.

   - **or** -

   - To select results of individual samples from multiple centers, platforms, and data types, select the disease type and data type(s) of interest from the **Disease Type** drop-down list, and then click **Go to the Data Access Matrix**.

   Tip: You can also click anywhere in the Data Access Matrix image to access the application.

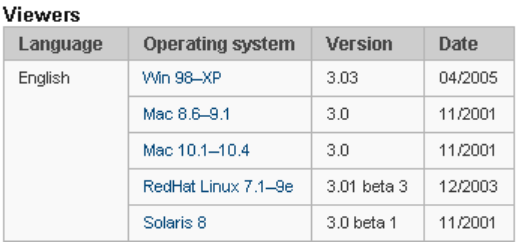
3. Follow the instructions in TCGA Portal and Data Access Matrix documentation for selecting and downloading data.

Installing the SVG Plugin

You must install the Adobe *SVG Plugin* to display the pathway diagrams generated in CMA Portal. If you are an Internet Explorer user, Adobe's installer will add the plug-in automatically into Internet Explorer.

How to Install the SVG Plugin if You Are a Mozilla Firefox User

1. Navigate to http://www.adobe.com/svg/viewer/install/main.html and scroll down the page to the Installing Adobe SVG Viewer section (*Figure 1.10*).



Viewers

Language	Operating system	Version	Date
English	Win 98–XP	3.03	04/2005
	Mac 8.6–9.1	3.0	11/2001
	Mac 10.1–10.4	3.0	11/2001
	RedHat Linux 7.1–9e	3.01 beta 3	12/2003
	Solaris 8	3.0 beta 1	11/2001

Figure 1.10 Adobe SVG Viewer Download Links

2. Click the link that is appropriate to your language and operating system.

   The Opening SVG Viewer dialog box appears (*Figure 1.11*).

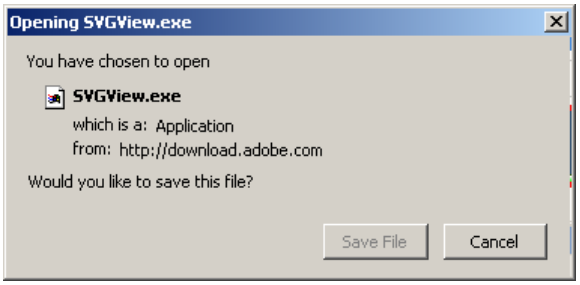


Figure 1.11 Opening SVG Viewer Dialog Box

3. Click **Save File**.
4. Using Windows Explorer browse to the **`C:\Program Files\Common Files\Adobe`** folder.
5. Click the **Plugins** folder and copy the **`NPSVG3.dll`**and **`NPSVG3.zip`**. files.
6. Paste both files into your Firefox plugins folder (the default Firefox plugins folder is `C:\Program Files\Mozilla Firefox\plugins`).
7. In the Firefox browser address field, type ***`about:config`***.

   A list of configurable files appears (*Figure 1.12*).

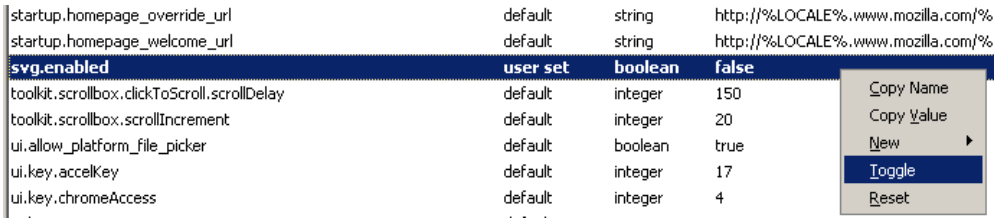


Figure 1.12 Plugins

8. Scroll down the list to **svg**.**enabled** to configure the setting.
9. To change the value from **true** to **false**, double click the row, or right-click the row, and select **Toggle**.
10. Close all browser windows and restart Firefox.

Logging Out

There are two ways to log out of the Cancer Molecular Analysis Portal.

To log out of the Cancer Molecular Analysis Portal, do one of the following:

- If you are not on the **Home** page, click the **LOGOUT** link beside your user name just below the tabs.

- **or** -

- On the Home page, in the **User Access** pane, click **LOGOUT**.

CHAPTER

2

WORKING WITH GENE-BASED VIEWS

This chapter describes how to use the Cancer Molecular Analysis Portal to research gene expression, copy number, SNP, and pathway data.

Topics in this chapter include:

Overview

The Gene View workspace in the CMA Portal enables researchers to analyze and visualize *gene expression*, *copy number*, *SNP*, *LOH*, and pathway data on a gene-by-gene basis, and to generate detailed study-related reports for a given gene.

This workspace provides access to gene expression plots, *Kaplan-Meier* survival plots, genome browser views, and pathway visualizations.

The CMA Portal analyzes data and categorizes gene anomalies (e.g., *overexpression*, *amplification*, etc.) by anomaly type. *Table 2.1* provides the categories assigned to different data types.

Data Type	***Category***	
Gene expression	Overexpressed	Underexpressed
Copy number	Amplified	Deleted

Table 2.1 Categories of data analysis

Data Type	Category	
Mutation	Mutated	NA

Table 2.1 Categories of data analysis (Continued)

Note: CMA Portal also lists any agents (chemical compounds) that are associated with a genetic anomaly. The Cancer Gene Index Project is the source for gene-to-drug relationship.

Related Topics:

- *Creating Gene Expression Plots*
- *Creating Gene Expression-Based Kaplan-Meier Plots*
- *Viewing Clinical Reports*
- *Viewing Mutation and Copy Number Changes*
- *Visualizing Pathways*

Creating Gene Expression Plots

Gene views enable you to study Geometric Mean, Log2 Intensity, and Box and Whisker Log2 Intensity, Kaplan-Meier Survival plots, and, for TCGA data, mutations and copy number changes.

How to Create a Gene Expression Plot

1. Do one of the following to navigate to the Gene View workspace:
   - From the Home page, click **Gene View**.

     - **or** -

   - From any page in the portal, click the **Gene View** tab.

   The Gene View workspace appears (*Figure 2.1*).

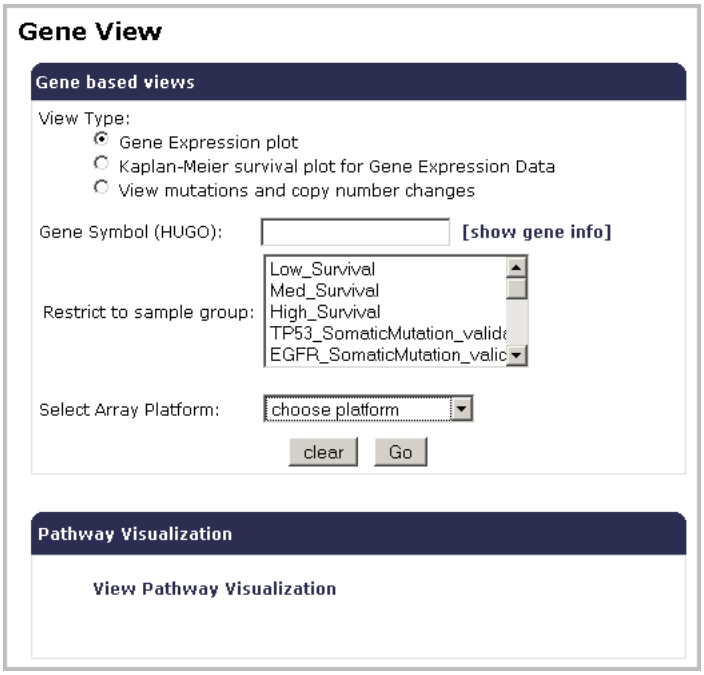


Figure 2.1 Gene View – Default Workspace

Note: The **View mutations and copy number changes** and **Pathway Visualization** features are not available in the REMBRANDT context.

2. Under **View Type**, select the type of plot you want to generate.

Table 2.2 lists the view type options.

View Type	***Description***
Gene Expression Plot	Generates the following plots: • **Geometric Mean** – displays mean expression intensity (Geometric mean) versus Groups. For additional gene expression plot details, see *Understanding Geometric Mean Gene Expression Plots* on page 23. • **Log2 Intensity** – displays average expression intensities for the gene of interest. For additional graph details, see *Log2 Intensity Gene Expression Plot Details* on page 26. • **Box and Whisker Log2 Intensity** – displays a Box and Whisker plot or box plot. For additional plot details, see *Box and Whisker Log2 Intensity Gene Expression Plot Details* on page 27.
Kaplan-Meier Survival Plot	The Kaplan-Meier method is used for survival analysis. Kaplan-Meier curves are used to estimate survival probability for the user-defined set of criteria as a function of time and survival differences as analyzed by the log-rank test.
View mutations and copy number changes (TCGA context only)	Generates heatmaps and launches the Cancer Genome Workbench. For additional information, see *Viewing Mutation and Copy Number Changes* on page 36.

Table 2.2 Types of gene-based views

3. In the Gene Symbol (*HUGO*) box, type the gene symbol of interest, for example, *EGFR* or *WT1*, for which you want to generate an expression plot.
4. To view details about the gene of interest in the Cancer Genome Anatomy Project website, click **show gene info**, otherwise, skip to Step 6

   The Cancer Genome Anatomy Project website appears in a new browser window (*Figure 2.1*).

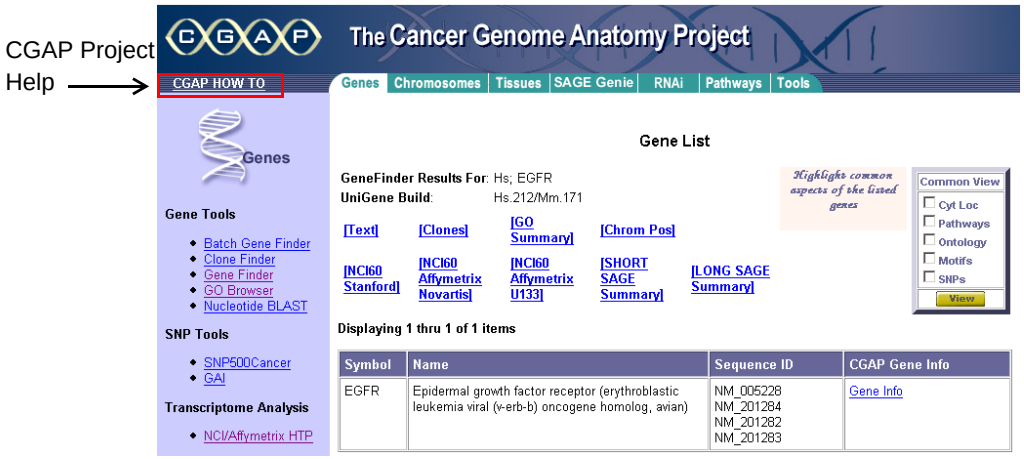


Figure 2.2 Cancer Genome Anatomy Project Web Page–EGFR

5. For help with using the Cancer Genome Anatomy Project, click **CGAP How To**.
6. To restrict your research to a specific *sample* group, from the **Restrict to sample group** drop-down list, select a sample group of interest. Custom lists you may have created earlier in your current session are listed as well.

   Note: The sample groups available in the REMBRANDT context differ from those in TCGA context. For more information about the available sample groups, refer to The Cancer Genome Atlas' (TCGA's) website at http://cancergenome.nih.gov/components/dmbca.asp, and the REMBRANDT website at: http://caintegrator-info.nci.nih.gov/REMBRANDT

   Tip: To select multiple sample groups for comparison, press and hold the CTRL key and click each of the groups of interest.
7. From the **Select Array Platform** drop-down list, select an array platform.

   Note: In the REMBRANDT context, Affy HT Human Genome U133 Plus 2.0 is the only array available.

Table 2.3 provides a description of each of the available arrays.

Array Platform	***Description***
(GBM: Broad) Affy HT Human Genome U133 Plus 2.0 (P2)	Includes Affy HT Human Genome U133 set plus 6,500 additional genes for analysis of over 47,000 transcripts. (For REMBRANDT data only)

Table 2.3 Description of available array platforms

Array Platform	*Description*
(GBM: Broad) Affy HT Human Genome U133A	Analyzes the expression level of 18,400 transcripts and variants, including 14,500 well-characterized human genes. Comprises more than 22,000 probe sets and 500,000 distinct oligonucleotide features.
(GBM:UNC) Agilent Whole Human Genome	High-density profiling analysis tool that covers over 41,000 unique human genes and transcripts
(GBM:LBL) Affy HE (Human Exon 1.0)	Contains approximately one million predicted and confirmed exons.
Agilent 8 x 15K Human miRNA-specific Microarray	Contains probes for 723 human and 76 human viral microRNAs from the Sanger database v.10.1.
Affy_SNP6	DNA analysis array that covers 1.8 million genetic markers, including more than 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation.

Table 2.3 Description of available array platforms (Continued)

8. Click **Go**, or, to clear the entries on the page and start over, click **Clear**.

The geometric mean gene expression plot appears.

Note: The CMA Portal does not generate an expression plot if any of your criteria are incompatible, for example, if the gene you selected is not applicable for the array platform you selected. In such cases, a message at the top of the page is displayed to alert you to the incompatibility issue.

For information about the Gene Expression Plot, see *Understanding Geometric Mean Gene Expression Plots*.

Related Topics:

- *Understanding Geometric Mean Gene Expression Plots*
- *Log2 Intensity Gene Expression Plot Details*
- *Box and Whisker Log2 Intensity Gene Expression Plot Details*
- *Visualizing Probe Sets In the Legend*
- *Displaying Coin Plots*
- *Creating Gene Expression-Based Kaplan-Meier Plots*
- *Viewing Clinical Reports*
- *Viewing Mutation and Copy Number Changes*
- *Visualizing Pathways*

Understanding Geometric Mean Gene Expression Plots

The geometric mean gene expression plot appears by default when you perform a gene expression search (*Figure 2.3*).

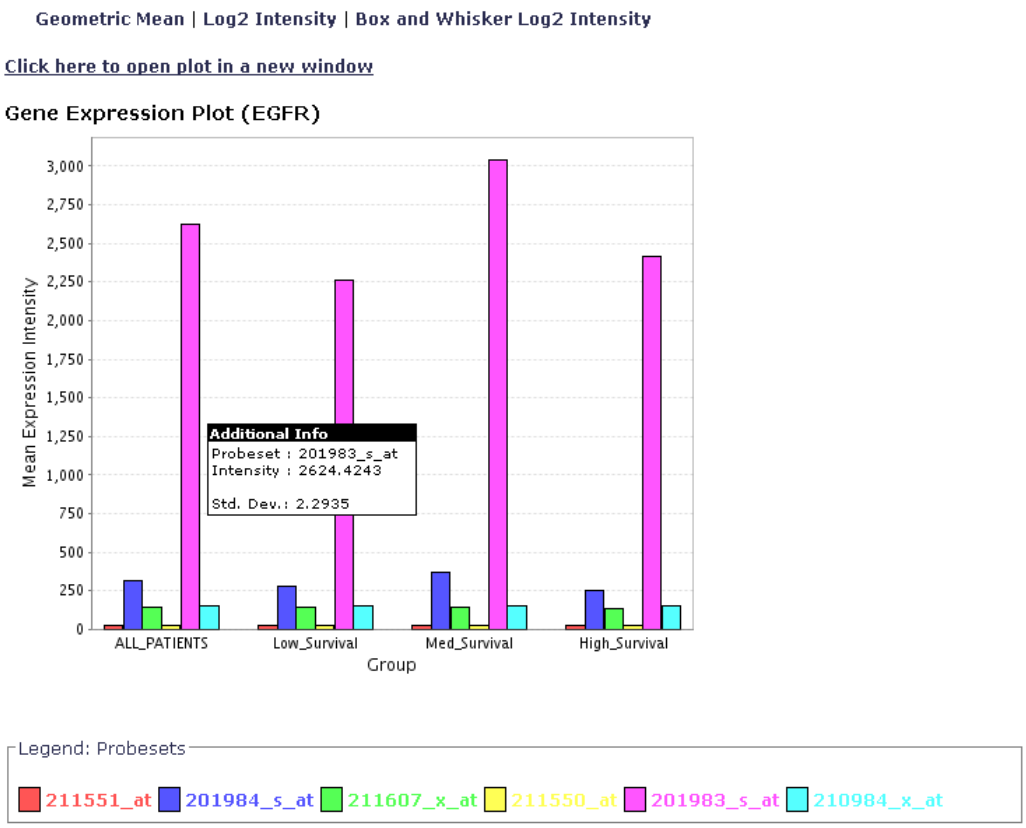


Figure 2.3 Geometric Mean Gene Expression Plot

Table 2.4 describes each area of the Gene Expression Plot page.

Item	*Special Instructions*
Plot Type Selection	Click a plot type name to display it. • **Geometric Mean** displays mean expression intensity (geometric mean) versus Groups. For additional graph details, see *Understanding Gene Expression-Based Kaplan-Meier Plots* on page 32. • **Log2 Intensity** displays average expression intensities (log2 values) for the reporter gene of interest. For additional graph details, see *Log2 Intensity Gene Expression Plot Details*. • **Box and Whisker Log2 Intensity** displays a Box and Whisker plot or box plot. For additional graph details, see *Box and Whisker Log2 Intensity Gene Expression Plot Details*.
Click here to open plot in a new window	Click the link to open the current plot in a new window and adjust the display.
Legend Probe Sets	**Reporter Gene information** – Indicates the color for each probe set appearing in the plot. To display reporter gene information, click a probe set number in the Legend.
Additional Info	To display additional information about each probe set, hover your mouse cursor over a bar in the graph. The pop-up window displays the following information: • **Probe Set** – Each probe set 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 oligonucleotides whose sequence exactly matches the gene of interest; the mismatch differs from the perfect match at one base position in the middle of the sequence. • **Intensity** – The geometric mean value calculated for each comparison group. • **Standard deviation** – The standard deviation value of a comparison group, such as GBM, for a particular probe set or gene. Standard deviation is a statistical measure of spread or variability.

Table 2.4 Understanding the Gene Expression Plot workspace

Visualizing Probe Sets In the Legend

Detailed information for each *probe set* in the gene expression plot legend is available via the CMA Reporter.

To view probe set details, click the probe set of interest.

The CMA Viewer displays the details (*Figure 2.4*).

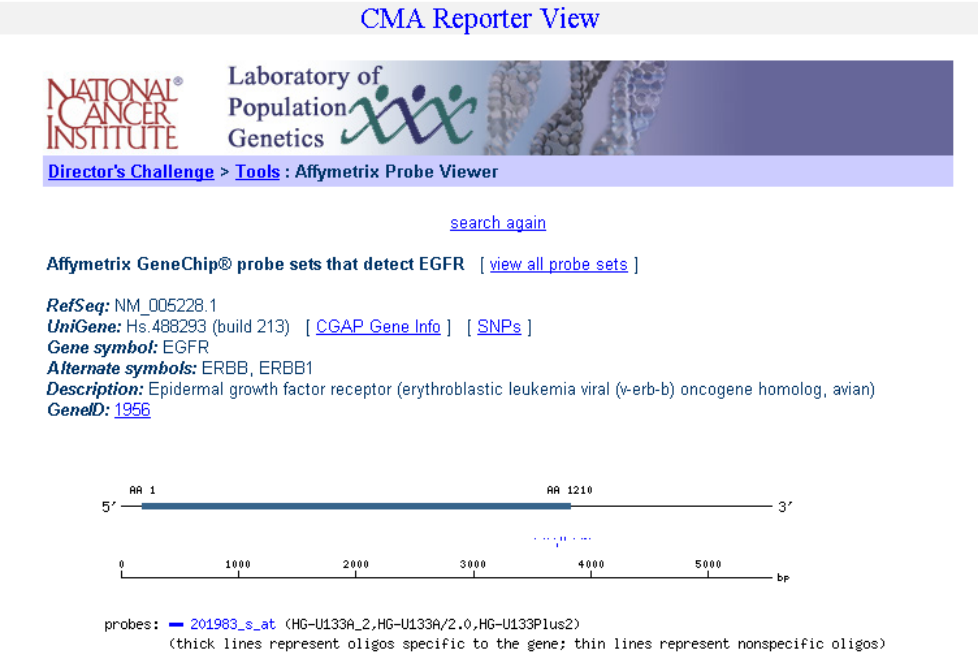


Figure 2.4 Section of a CMA Reporter View – Probe Sets That Detect EGFR

Log2 Intensity Gene Expression Plot Details

The Log2 Intensity Gene Expression Plot displays average expression intensities for the gene of interest based on Affymetrix GeneChip arrays (U133A arrays) (*Figure 2.5*). Multiple *probe sets* (for some genes) are designed to measure the expression of the gene of interest. For more information on the probe set design strategy for human genes, see the Affymetrix website at http://www.affymetrix.com.

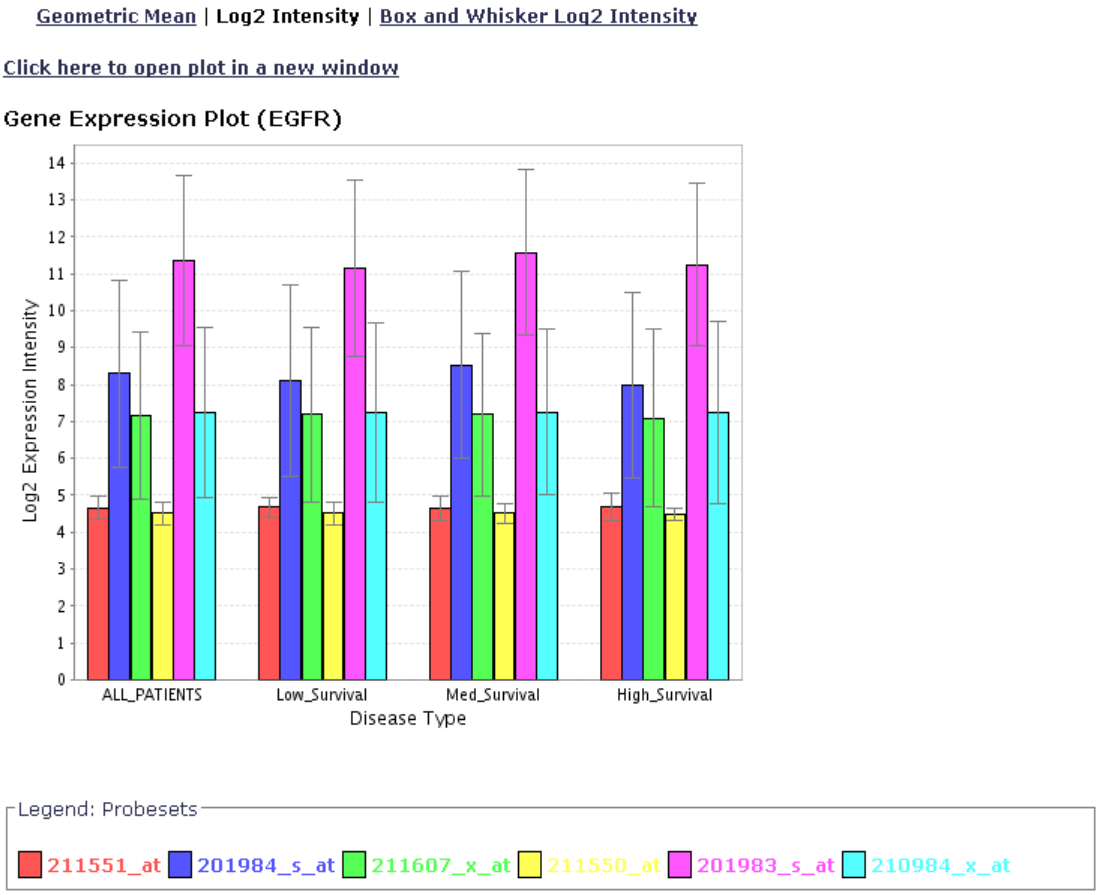


Figure 2.5 Log2 Intensity Gene Expression plot

Box and Whisker Log2 Intensity Gene Expression Plot Details

The Box and Whisker Log2 Intensity Gene Expression Plot displays a box plot without all the individual data points for each *sample* (*Figure 2.6*). Examples of 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
- Compare two or more datasets
- Compare distributions; the center, spread, and overall range are immediately apparent.

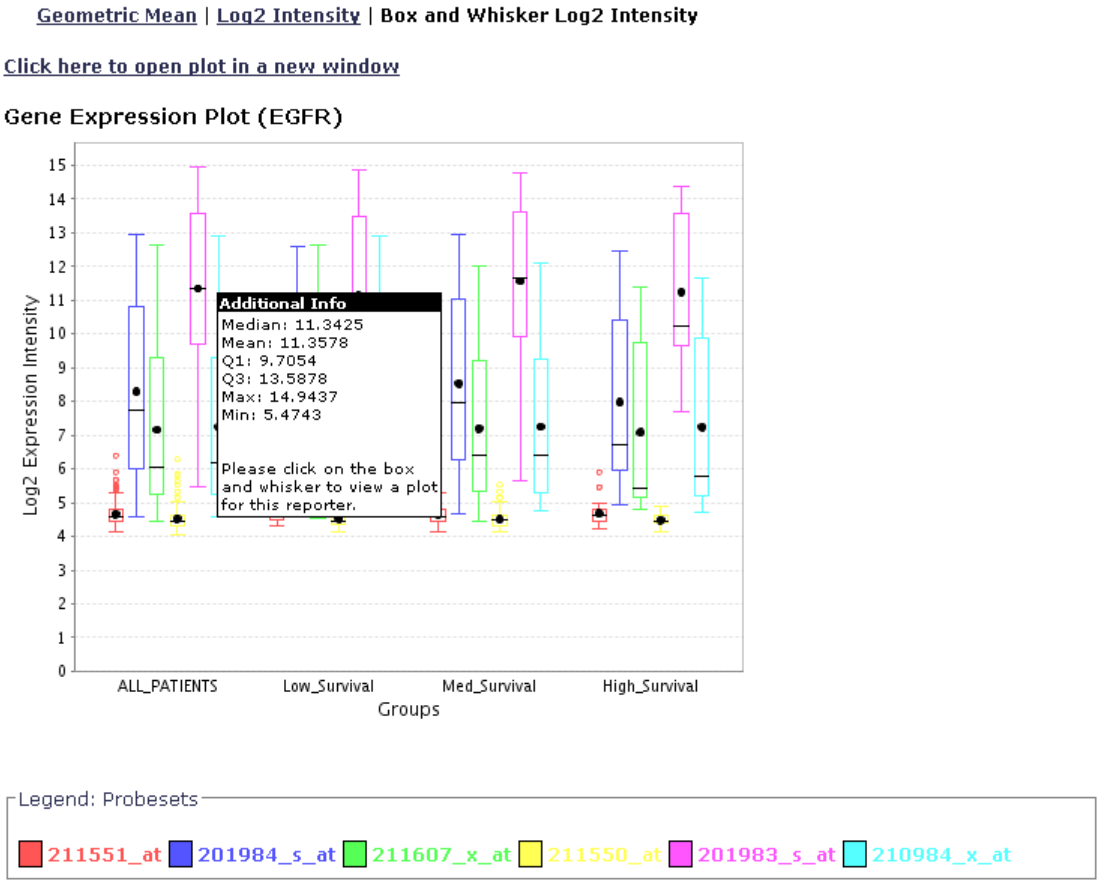


Figure 2.6 Box and Whisker Log2 Intensity Gene Expression plot

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 probe set for one group, hover your cursor over the probe set on the plot to display the Additional Information.

Table 2.5 describes Additional Information details.

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

Table 2.5 Box and whisker log2 intensity gene expression plot information

Tip: To display a coin plot for the reporter gene, click anywhere in the box. A *coin plot* is a box-and-whisker plot with all the individual data points (see *Displaying Coin Plots* on page 29).

In the box-and-whisker plot, the individual probe set summary is represented as illustrated in *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.

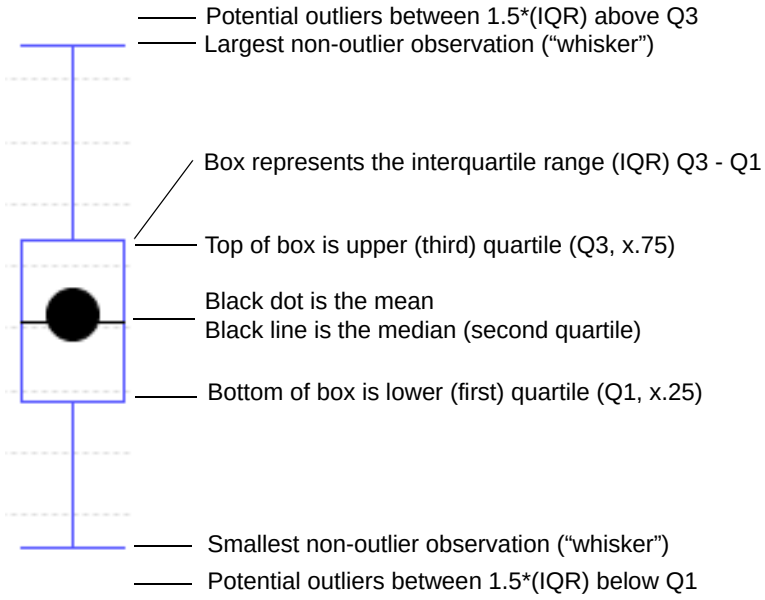


Figure 2.7 Box and Whisker Plot Details

Displaying Coin Plots

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

How to Display a Coin Plot

In the Box and Whisker Log 2 Intensity plot, click inside the box associated with the reporter gene of interest.

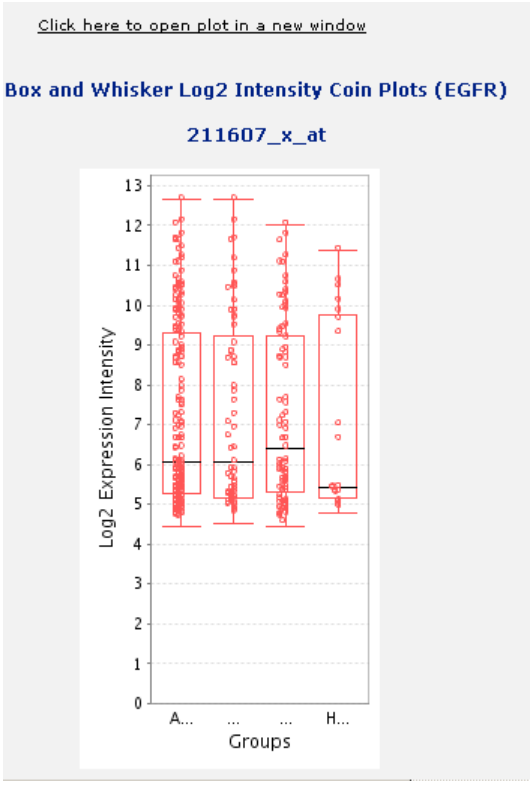


Figure 2.8 Coin Plot for a Probe Set

In the coin plot, the individual *probe set* summary is represented as illustrated in *Figure 2.9*. Horizontal lines (the “whiskers”) extend to a maximum of 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.

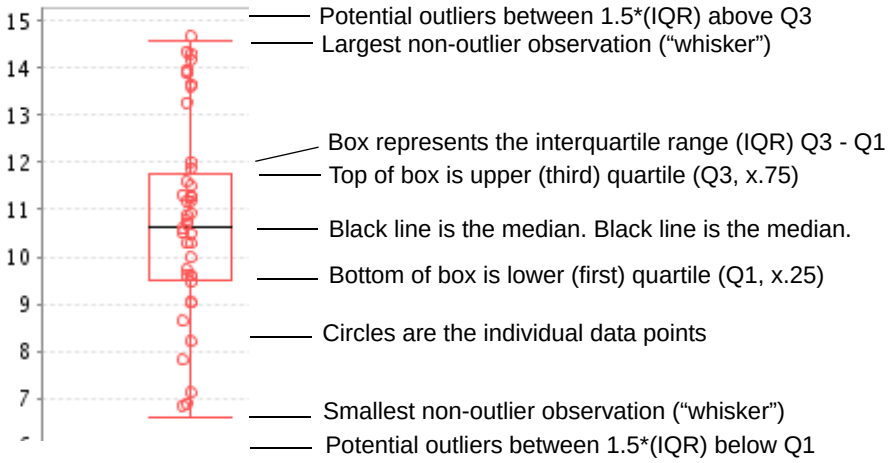


Figure 2.9 Coin Plot Details

Creating Gene Expression-Based Kaplan-Meier Plots

Gene views enable you to study Kaplan-Meier Survival plots.

How to Create a Gene Expression-Based Kaplan-Meier (KM) Plot

1. Do one of the following to navigate to the Gene View workspace:
   - From the Home page, click **Gene View**.

     - **or** -

   - From any page in the portal, click the **Gene View** tab.

   The Gene View workspace appears (*Figure 2.10*).

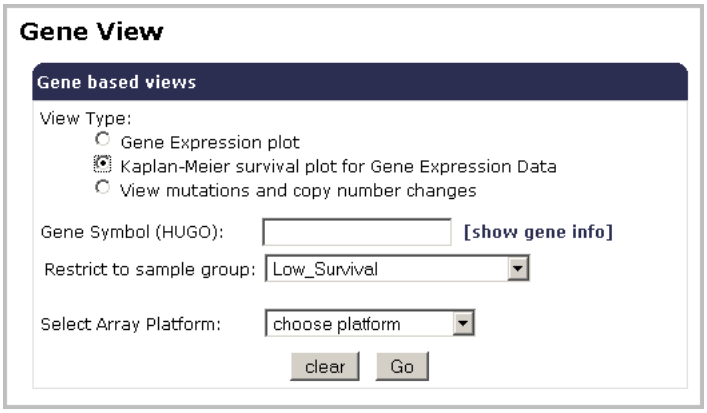


Figure 2.10 Gene View Workspace – Gene-based Views Section

Note: The "View mutations and copy number" feature is not available in the REMBRANDT context.

2. In the Gene Symbol (*HUGO*) box, type the gene symbol of interest, for example, *EGFR* or *WT1*, for which you want to generate an expression plot.
3. To view details about the gene of interest, click **show gene info**.
4. From the **Restrict to sample group** drop-down list, select a sample group. Custom patient lists you have saved appear in red text.

   Note: The sample groups available in the REMBRANDT context differ from those in TCGA context.
5. From the **Select Array Platform** drop-down list, select an *array platform*.

   Table 2.3 on page 22 provides a description of each of the available arrays.
6. Click **Go**, or, to clear the entries on the page and start over, click **Clear**.

   The Gene Expression-based *Kaplan-Meier Survival* Plot appears.

   If the system does not generate an expression plot, check for messages above the Gene-based View section.

For information about the Gene Expression Plot, see *Understanding Gene Expression-Based Kaplan-Meier Plots*.

Related Topics:

- *Understanding Gene Expression-Based Kaplan-Meier Plots*
- *Creating and Filtering Gene Expression Queries*

Understanding Gene Expression-Based Kaplan-Meier Plots

A Gene Expression-Based *Kaplan-Meier* plot 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) (*Figure 2.11*). Kaplan-Meier estimates are calculated based on the last follow-up (FU) time and the censor status (0=alive, 1=dead) from the samples of interest.The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph. You can dynamically modify the fold change (up and down regulation) thresholds and redraw the plot.

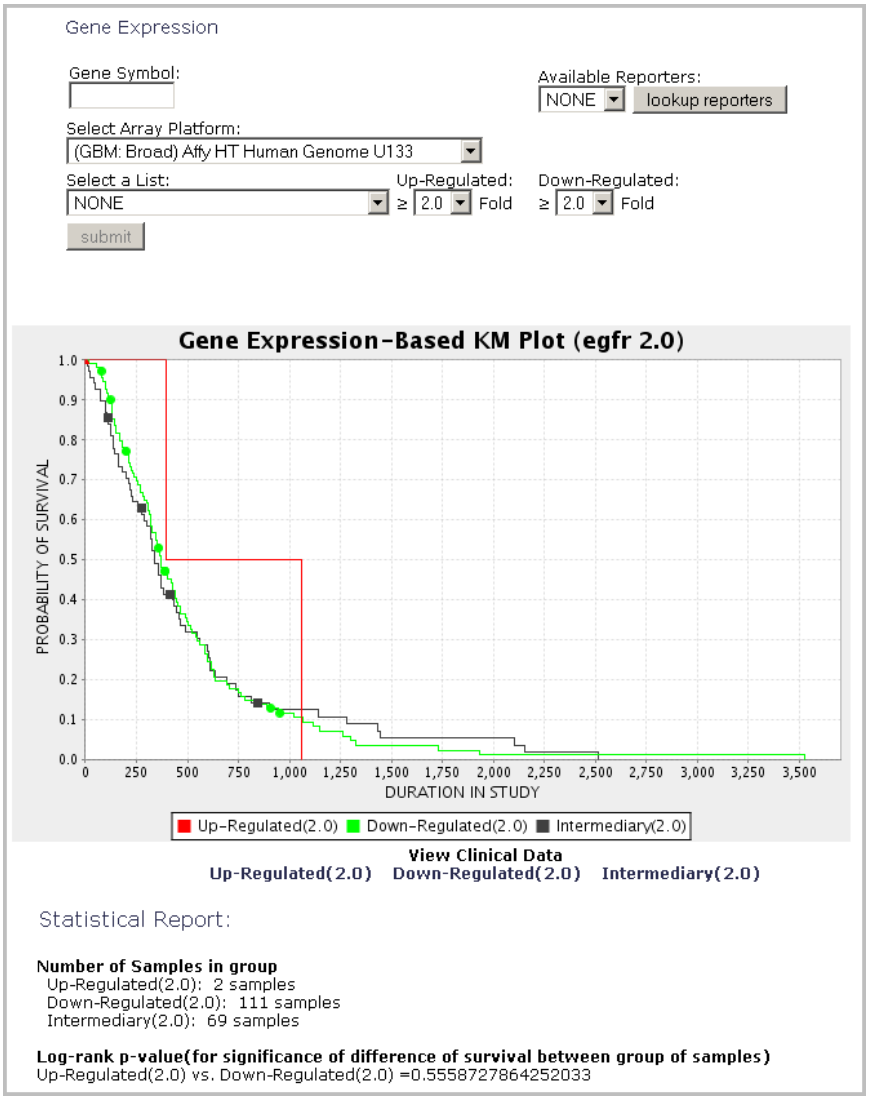


Figure 2.11 Gene Expression-Based KM Plot

Table 2.6 describes areas on the *Kaplan-Meier Survival Plot* for Gene Expression data page.

Item	***Function/Description***
Gene Expression/Copy Number Filter	When you apply a copy number filter, the CMA Portal provides links to display the copy number data for samples
View Clinical Reports	When you apply a gene expression filter, the CMA Portal provides links to display the gene expression for **Up-regulated 2.0**, **Down-regulated 2.0**, and **Intermediary2.0** samples. To generate a clinical report, click the appropriate link. For more information, see *Viewing Clinical Reports* on page 35.

Table 2.6 Gene Expression-Based KM Plot page

Item	Function/Description
Statistical Report	• **Number of Samples** specifies the number of Up-Regulated2.0, Intermediary2.0, Down-Regulated2.0 samples, if any. • **Log-rank *p*-Value** indicates the significance of the difference in survival between any two groups of samples segregated based on gene expression of the gene of interest. The log rank *p*-value is calculated using Mantel-Haenszel method. The *p*-values are recalculated every time a new threshold is selected.

Table 2.6 Gene Expression-Based KM Plot page (Continued)

Creating and Filtering Gene Expression Queries

The criteria selection boxes on the top of the Gene Expression-based *Kaplan-Meier plot* allows you to create and filter queries, and to modify fold change thresholds dynamically.

Figure 2.12 shows query criteria selection boxes for gene expression plots.

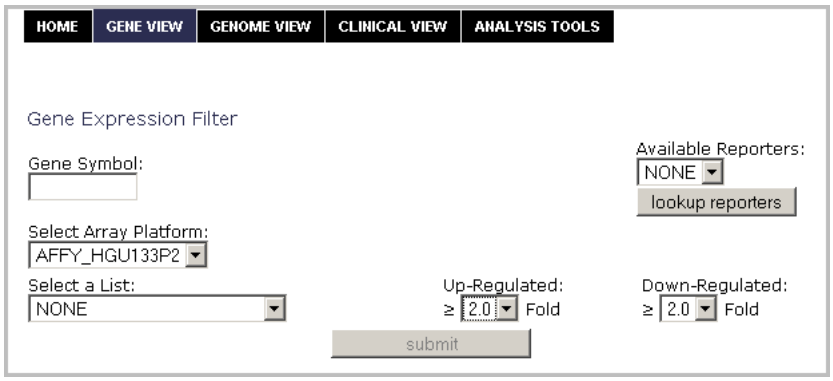


Figure 2.12 Gene Expression Filter

How to Create a Gene Expression Plot

1. In the **Gene Symbol** box, type the gene symbol of interest, for example, *EGFR*, for which you want to generate the plot.
2. From the **Select Array Platform** drop-down list, select an array platform.

   Table 2.3 on page 22 provides a description of each of the available arrays.
3. From the **Select a List** drop-down list, select a *sample* group. Patient lists you have saved appear in red.
4. To generate a list of reporter genes, click **look up reporters**.

   The list of available reporter genes is displayed in the **Available Reporters** box (*Figure 2.13*).

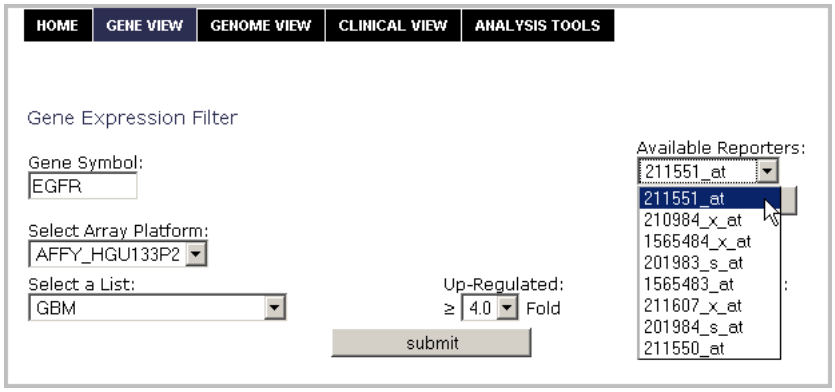


Figure 2.13 Gene Expression Filter – Reporters List

5. Select a reporter gene from the drop-down list.
6. In the **Up-Regulated** criterion box, select the "greater than or equal to" fold change value, and then do the same in the **Down-Regulated** criterion box.
7. Click **Submit**.

   The KM plot is redrawn with the criteria you selected.

Viewing Clinical Reports

Once you have generated a KM survival plot, you can view and download clinical data items for the selected patient list (*Figure 2.15*). On the Clinical Report page, you can save your *sample* group selection to a custom *PatientDID list*.

How to View a Clinical Report

1. Generate a gene expression-based Kaplain-Meier plot. For detailed instructions, see *Creating Gene Expression-Based Kaplan-Meier Plots* on page 31.

   The KM plot appears, with the clinical report links at the bottom of the graph (*Figure 2.14*).

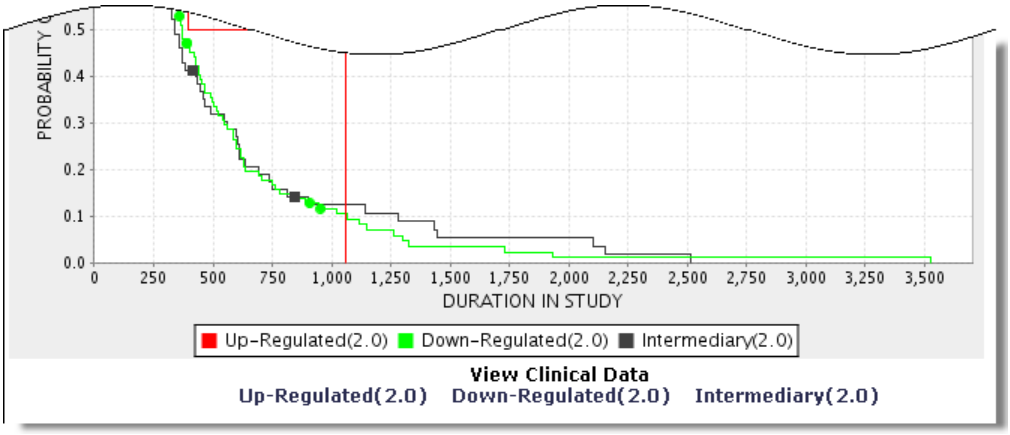


Figure 2.14 Gene Expression-Based KM Plot – View Clinical Data

2. Under **View Clinical Data**, select the link for the sample group data of interest, for example, **Down-Regulated(2.0)**.

   The clinical report appears in a new browser window (*Figure 2.15*).

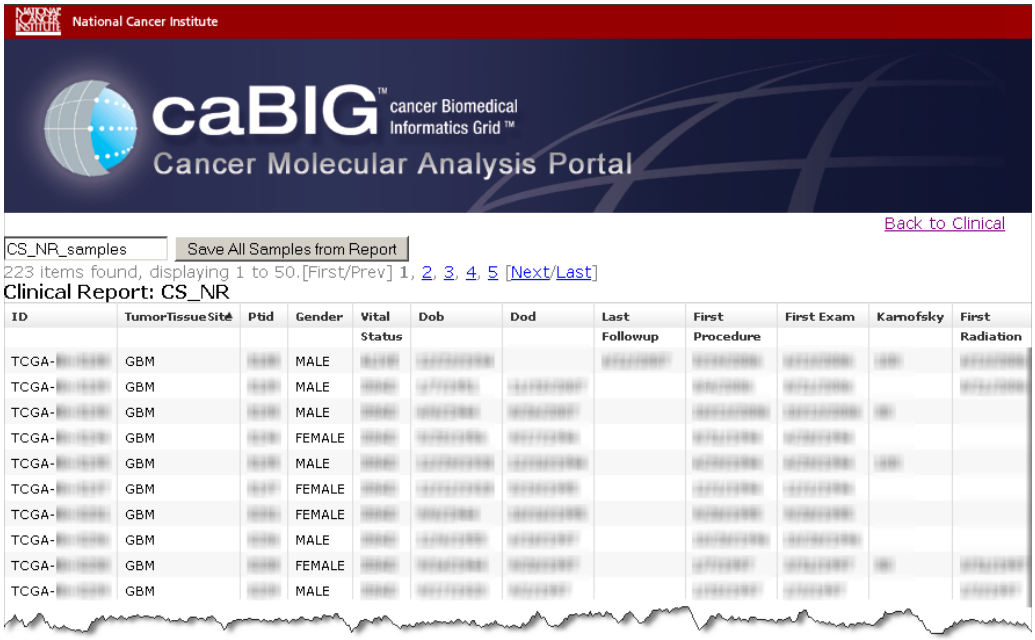


Figure 2.15 Clinical Report Page (partially redacted)

3. To sort the data, click the heading row for the column you want to sort. See *Navigating Clinical Report Table Results* on page 61 for details.
4. Optionally, to rename the report, type a unique name in the report name field.
5. To save the entire list of patients, click **Save All Samples from Report**.

Viewing Mutation and Copy Number Changes

The Gene View workspace provides a link to the Cancer Genome Workbench (CGWB), an application that allows you to view and drill down into copy number and mutation data of a particular sequence (*Figure 2.16*). The CGWB integrates clinical tumor mutation profiles with the reference human genome.

How to View Mutation and Copy Number Changes

1. In the **Gene Symbol** box, type the gene symbol of interest, for example, *EGFR*, for which you want to view mutation and copy number data.
2. Click **Go**, or, to clear the entries on the page and start over, click **Clear**.

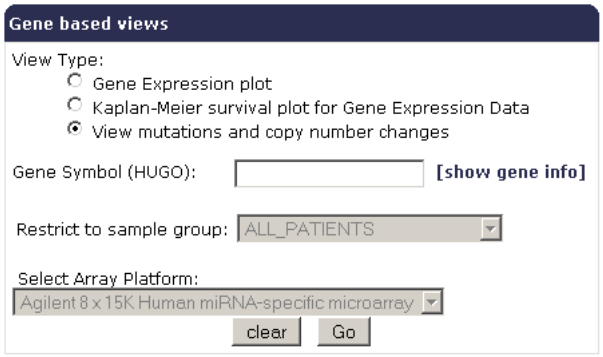


Figure 2.16 Gene-Based Views – View Mutation and Copy Number Changes

For instructions on using the CGWB features, refer to the online help in that application.

Visualizing Pathways

The pathway visualization link in the Gene View workspace provides access to graphic depictions of the pathways of interest and provides detailed information about associated genes, agents, and anomalies. See *Working With Pathways and Associated Anomalies* on page 39.

CHAPTER

3

WORKING WITH PATHWAYS AND ASSOCIATED ANOMALIES

This chapter describes how to use the Cancer Molecular Analysis Portal to review pathways and the anomalies associated with them.

Topics in this chapter include:

Overview

The pathway visualization section of the Gene View workspace provides access to graphic depictions of pathways of interest and provides detailed information about associated genes, agents, and anomalies.

Generating Pathway Diagrams

Pathway diagrams enable you to investigate genetic anomalies at a cellular level.

How to Generate a Pathway Diagram

1. Do one of the following to navigate to the Gene View workspace:
   - From the Home page, click **Gene View**.

     - **or** -

   - From any page in the portal, click the **Gene View** tab.

The Gene View workspace appears (*Figure 3.1*).

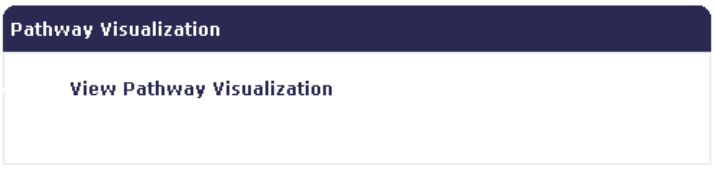
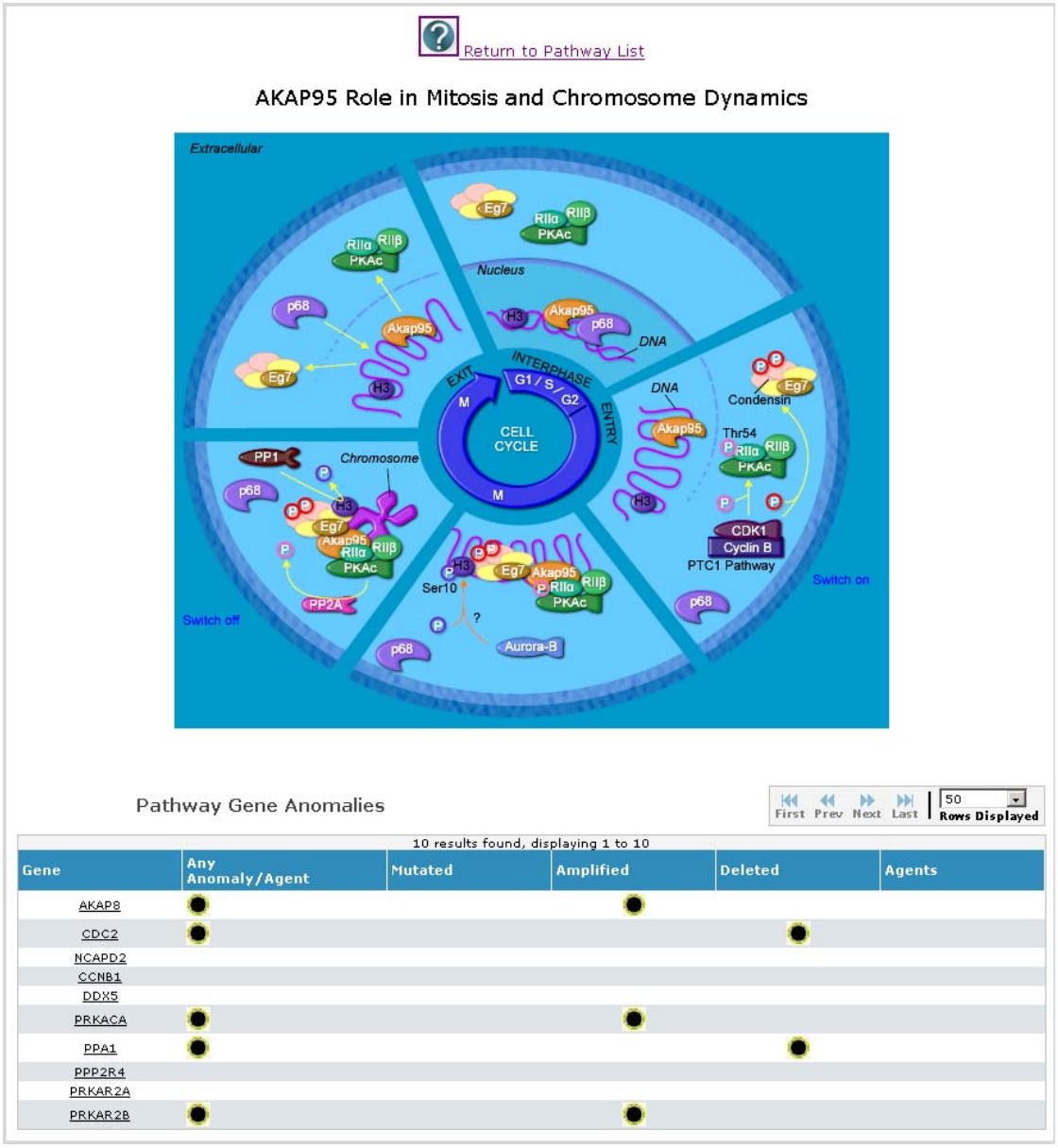


Figure 3.1 Clinical View – Sample-based Kaplan-Meier Graph Section

2. Under **Pathway Visualization**, select the pathway of interest from the **Select Pathway** drop-down list.
3. Click **Go**. Or, to return to the top of the pathway list, click **Clear**.

The Pathway Visualization and Pathway Gene Anomalies page appears in a new browser (*Figure 3.2*).



Gene	Any Anomaly/Agent	Mutated	Amplified	Deleted	Agents
AKAP8	●		●		
CDC2	●			●	
NCAPD2					
CCNB1					
DDX5					
PRKACA	●		●		
PPA1	●			●	
PPP2R4					
PRKAR2A					
PRKAR2B	●		●		

Figure 3.2 Pathway Visualization and Pathway Gene Anomalies Page

Note: If the pathway diagram is not displayed, install the *SVG plugin* on your computer. For more information about installing the SVG plugin, see *Installing the SVG Plugin* on page 16.

The Pathway Visualization and Pathway Gene Anomalies page is divided into 2 sections as follows:

- An interactive diagram of the pathway of interest appears at the top of the page. The basis for the visualization is a BioCarta pathway diagram. The selected pathway is depicted as if projected onto a morphological illustration of a cell. It depicts gene-to-gene and other molecular interactions in a graphical interface.

- A Pathway Genes Anomalies table appears at the bottom of the page. It lists the genes for the selected pathway and the anomalies associated with each of them.

Tip: Click the **Help** icon at the top of the page to access online help for the Pathway Viewer.

Investigating Genes Via Pathway Diagrams

The pathway diagram contains hyperlinks that allow you to investigate each gene associated with the pathway (*Figure 3.3*). The graphic representations of the genes link to, and launch, the UCSC Genome Bioinformatics website's Genome Browser, which provides detailed information about the genes of interest.

Additionally, the pathway diagram interacts with the Gene Pathways Anomalies table at the bottom of the page so that you can isolate genes that are of interest to you.

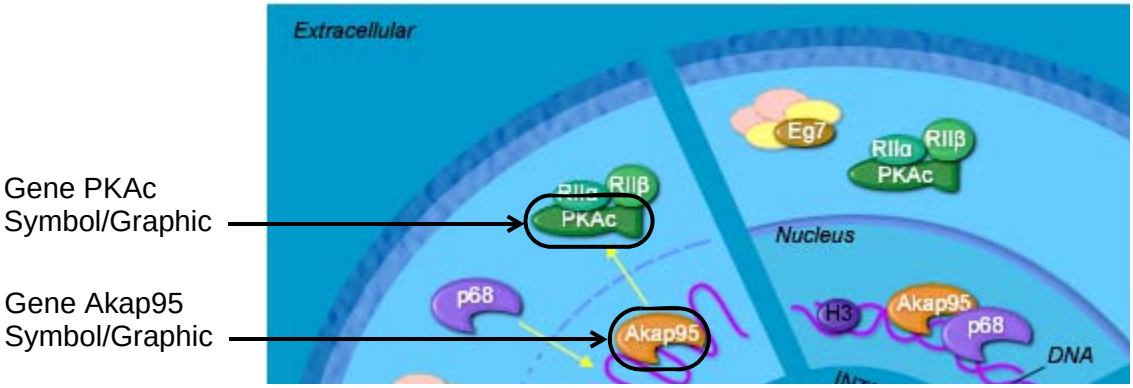


Figure 3.3 Pathway Diagram – Close-up View of Gene Symbol PKAc

How to Display Information About a Given Gene

1. To investigate a specific gene via the UCSC Genome Browser, click its symbol/graphic on the diagram. For example, to investigate the PKAc gene, click any of the graphic symbols labeled **PKAc**.

   The UCSC Genome browser opens and displays information about the selected gene. For complete documentation, click the **Help** button on the UCSC Genome Bioinformatics website.

2. To highlight the genes associated with a given column of data in the Gene Pathways Anomalies table, scroll down to the table and click the column heading. For example, to isolate the *amplified* genes on the Biocarta diagram, click the **Amplified** column heading (*Figure 3.4*).

   All amplified genes are highlighted in the diagram. Compare *Figure 3.4* with *Figure 3.2* on page 41.

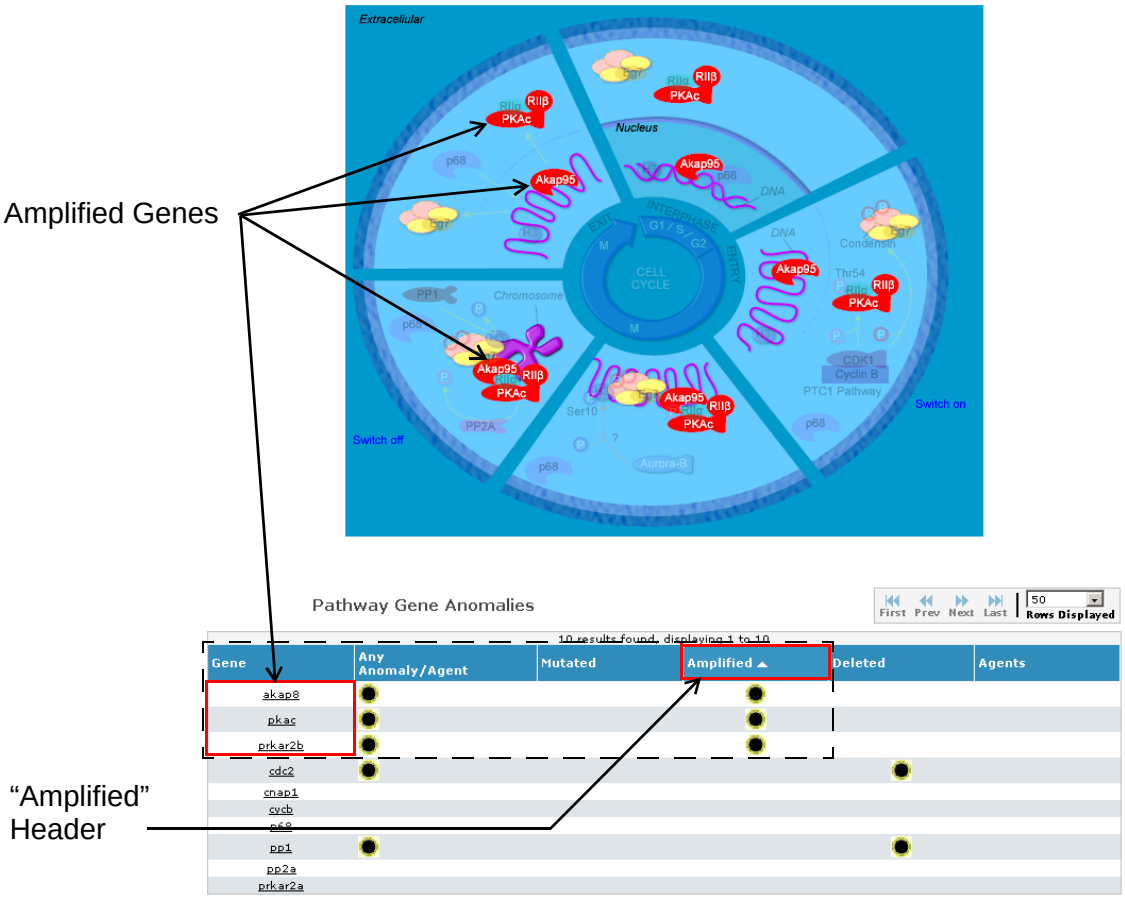


Figure 3.4 Pathway Diagram – Isolated Amplified Genes

3. To review genes for the selected pathway and the anomalies, mutations, and agents associated with them, scroll down to the Pathway Gene and Anomalies table, and follow the instructions in *Investigating Genes Via Pathway Gene Anomalies Tables* on page 43.

Investigating Genes Via Pathway Gene Anomalies Tables

The Pathway Gene Anomalies table (*Figure 3.5*) lists the genes for the selected pathway alphabetically, and indicates the anomalies associated with each gene with a black dot. Any agents associated with a gene are listed in the Agents column. They are hyperlinked to the NCI Thesaurus.

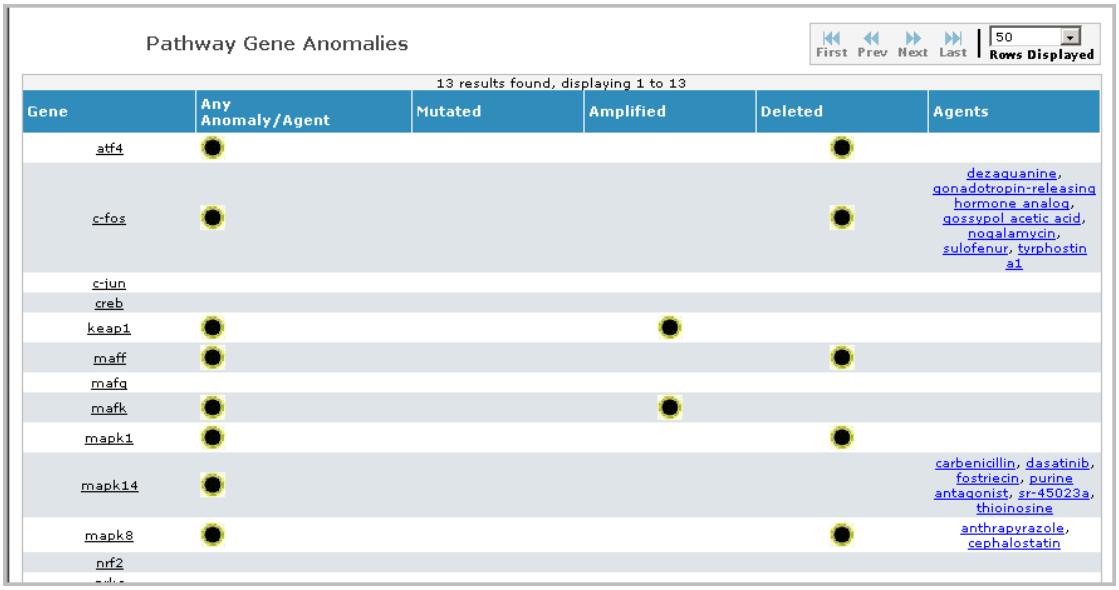


Figure 3.5 Pathway Gene Anomalies table

Table 3.1 describes each column on the Pathways and Associated Anomalies table.

Column Name	***Description***
Gene	Genes that are associated with a given pathway
Any Anomaly/Agent	A black dot appears when a gene within the pathway has an anomaly and/or drug agent associated with it.
Mutated	A black dot appears when genes within the pathway are categorized as *mutated*.
Amplified	A black dot appears when the computed *copy number* of the gene is greater than, or equal to, 2.5.
Deleted	A black dot appears when genes within the pathway have been *deleted*.
Overexpressed	A black dot appears when genes within the pathway are categorized as *overexpressed*.
Underexpressed	A black dot appears when genes within the pathway are categorized as *underexpressed*.
Agents	A black dot appears when genes within the pathway are associated with a drug agent.

Table 3.1 Column descriptions for Pathway Genes and Anomalies

How to Display Detailed Information About Genes and Agents

1. Scroll to the bottom of the **Pathway Visualization and Pathway Gene Anomalies** page.
2. For detailed gene information, click the gene name hyperlink.

   A new window displays the NCBI Entrez Gene page with detailed information about the selected gene.

3. For detailed agent information, click the agent name hyperlink.

   A new window displays the NCI Thesaurus page with detailed information about the selected agent.

Tip: To view the pathway list in TCGA, click **Return to Pathway List** at the top of the page.

Related Topics:

- *Navigating the Gene Anomalies Table*
- *Investigating Genes Via Pathway Diagrams*
- *Generating Pathway Diagrams*

Navigating the Gene Anomalies Table

The Gene Anomalies table provides several mechanisms for viewing and sorting the information provided.

By default, the table is sorted alphabetically by gene. Black dots in a Gene row indicate the anomaly(ies) with which a given gene is associated (*Figure 3.6*).

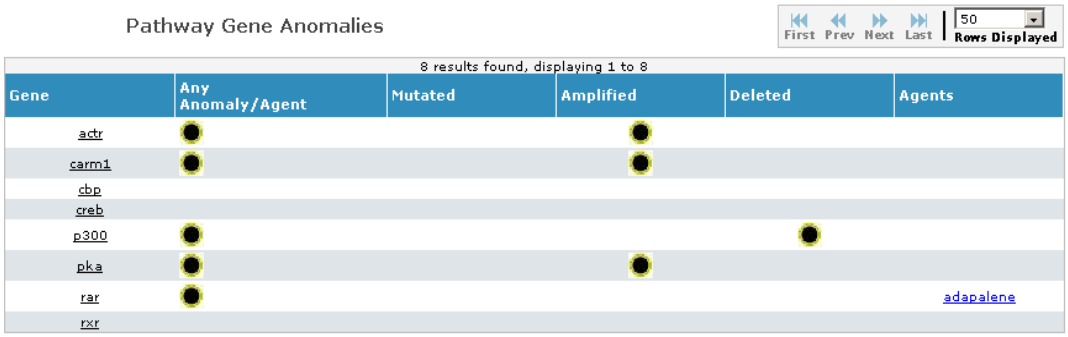


Figure 3.6 Gene Anomalies Table – Sorted Alphabetically

How to Change the Sort Order and View of the Table

1. To sort the table by anomaly, click the column heading for the anomaly of interest. For example, click the **Amplified** column heading (*Figure 3.7*).

   An “up” arrow appears next to the column name and the table is re-sorted such that the genes that have been amplified appear at the top of the Gene column.

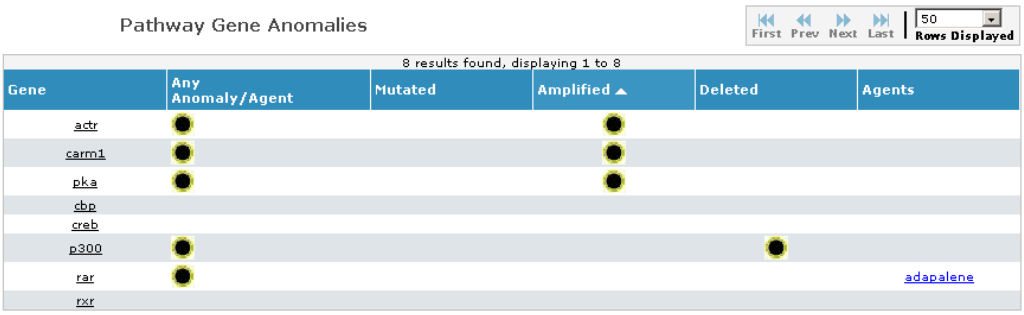


Figure 3.7 Gene Anomalies Table – Sorted by Amplification

2. To move all genes with anomalies to the top of the column, click the **Any Anomaly/Agent** column name.
3. To move all genes without anomalies to the top of the column, click the **Any Anomaly/Agent** column name again.
4. To display a longer list of items on one page, click the **Rows Displayed** list box just above the table on the right side. Select the number of items to display at once, either **50** or **100** (*Figure 3.8*).

Figure 3.8 Pathway Gene Anomalies Table Navigation Toolbar

The number of genes associated with a given pathway can exceed the limit that can be displayed on one page. The navigation arrows become active when the gene list exceeds one page, and allow you to move from page to page.

5. To move to another page of the list, click one of the arrow buttons next to the **Rows Displayed** list. You can move directly to the first or last page, or move to a previous page or subsequent page.

CHAPTER

4

WORKING WITH GENOME VIEWS

This chapter describes how to use the Cancer Molecular Analysis Portal to investigate chromosomal regions of amplification, deletion, and over-expression.

Topics in this chapter include:

About Genomic Views of Data

The CMA Portal enables you to explore data in one genome-level visualization and to investigate chromosomal regions of *amplification*, *deletion*, and *over-expression* via the integrated Heatmap Viewer.

Characterizations include the following types of data:

- Copy number
- Gene expression (not available for REMBRANDT data)
- Methylation (not available for REMBRANDT data)
- Mutation (not available for REMBRANDT data)

Note: A Java plugin is required for viewing heatmaps.

How to View Characterization Data

1. Do one of the following to navigate to the Genome View workspace:
   - From the Home workspace, click **Genome View**.

   - **or** -

   - From any workspace in the portal, click the **Genome View** tab.

   The Genome View workspace appears (*Figure 4.1*).

Figure 4.1 Genome View – Default Workspace, TCGA Context

2. To view the data, click the CGCC/platform link of interest.

Viewing Mutation Data

The Genomic View workspace provides a direct link to the Cancer Genome Workbench browser. For assistance with using the Cancer Genome Workbench, refer to the documentation provided in the Genome Workbench browser at: https://cgwb.nci.nih.gov/goldenPath/help/hgTracksHelp.html

Viewing Copy Number Data

Differences in the number of copies of certain genes contributes to genetic variability. Variability can be caused by deletions of some genes on only one chromosome, or multiple copies of some genes.

Copy number data in the CMA Portal is provided by the following sources:

- Broad Institute (Broad)
- Harvard Medical School (HMS)
- Memorial Sloan-Kettering Cancer Center (MSKCC)
- HudsonAlpha Institute for Biotechnology (HAIB)

Note: Copy numbers are also known as copy number variants (CNVs) and copy number polymorphisms (CNPs)

For more info on characterization sources, refer to TCGA site at: http://cancergenome.nih.gov/data/types/genomic/

Table 4.1 lists the CGCCs and the platforms used to derive their genome copy number data.

Broad	***HMS***	***MSKCC***	***SUSM***	***Combined Data***
Affymetrix	Agilent	Agilent	Illumina	Combined
Paired Affymetrix	Paired Agilent	Paired Agilent	Paired Illumina	Combined Paired
Unpaired Affymetrix	Unpaired Agilent	Unpaired Agilent	Unpaired Illumina	Combined Unpaired
Normal Affymetrix	Normal Agilent	Normal Agilent		

Table 4.1 Genome copy number data sources and platforms

Table 4.2 lists the CGCCs and the platforms used to derive their gene copy number data

Broad	***HMS***	***MSKCC***	***SUSM***
Affymetrix	Agilent	Agilent	Illumina
Paired Affymetrix	Paired Agilent	Paired Agilent	Paired Illumina
Unpaired Affymetrix	Unpaired Agilent	Unpaired Agilent	Unpaired Illumina

Table 4.2 Gene copy number data sources and platforms

To view copy number data, in the Genome workspace, click a CGCC/platform link.

The Heatmap Viewer displays a heatmap of the selected genes (*Figure 4.2*).

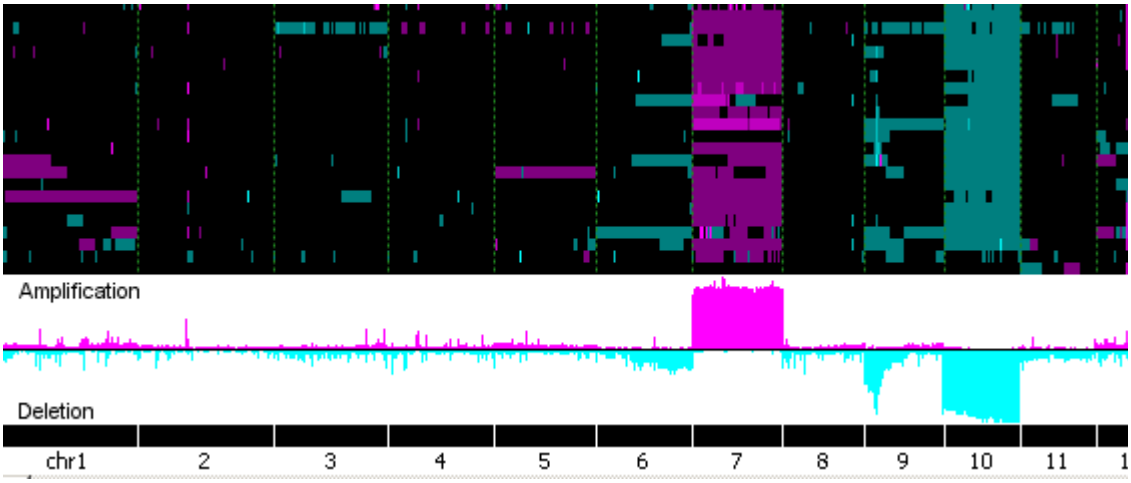


Figure 4.2 Heatmap Viewer – Section of Copy Number Heatmap

For Heatmap Viewer instructions, see *Navigating the Heatmap Viewer* on page 51.

Viewing Gene Expression Data

Gene expression data in CMA Portal is derived from the following sources:

- Gene expression of 188 GBM samples using Affymetrix U133. Submitted by the Broad Institute
- Gene expression of 188 GBM samples using Affymetrix U133 organized by 1000 genes with variable expression

To view gene expression data, click the CGCC/platform link of interest under **Genomic View of Expression Data**.

The Heatmap Viewer displays a heatmap of the selected genes (*Figure 4.3*).

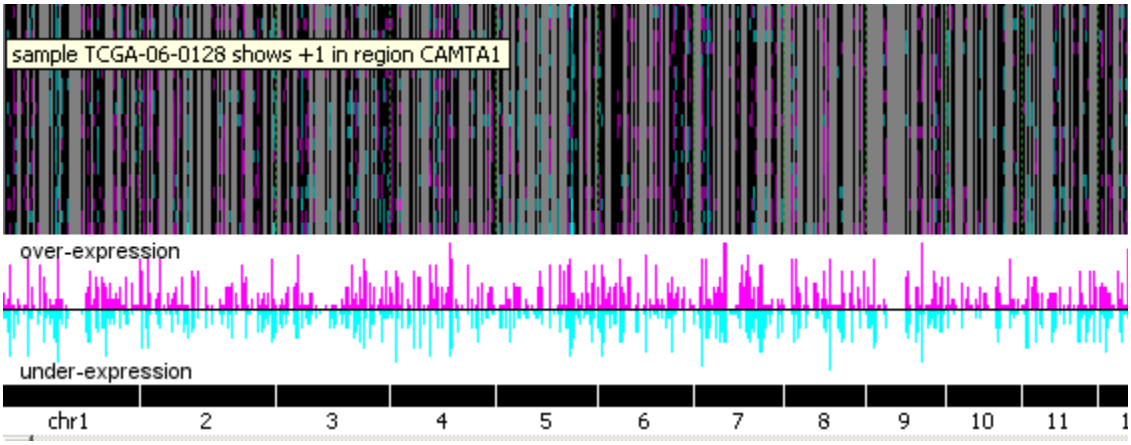


Figure 4.3 Heatmap Viewer – Section of Gene Expression Heatmap

For Heatmap Viewer instructions, see *Navigating the Heatmap Viewer* on page 51.

Viewing Methylation Data

Methylation data in CMA Portal is derived from the following sources:

- DNA methylation of 169 GBM samples in 1961 genes using Illumina. Submitted by the Sidney Kimmel Comprehensive Cancer Center At Johns Hopkins University (JHU-USC)
- DNA methylation of 169 GBM samples in 2178 genes using Illumina. Submitted by the Sidney Kimmel Comprehensive Cancer Center At Johns Hopkins University (JHU-USC)

To view gene methylation data, click the link of interest under **Genomic View of Methylation Data**.

The Heatmap Viewer displays a heatmap of the selected genes (*Figure 4.4*).

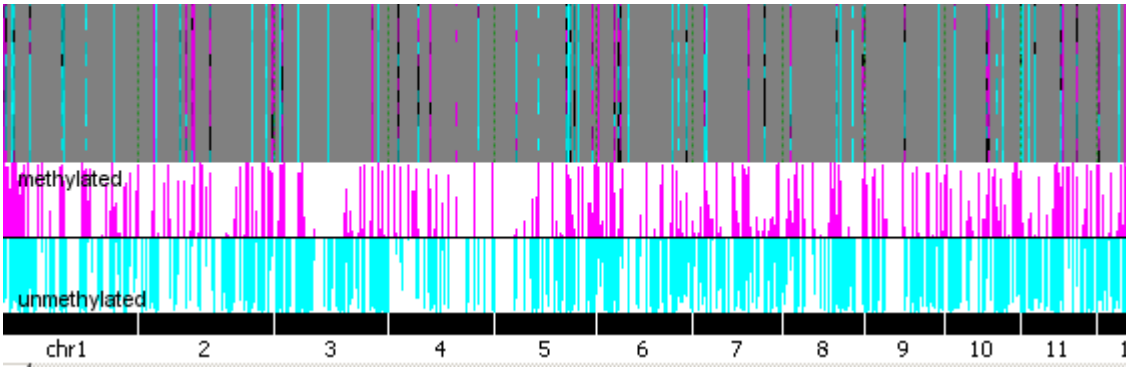


Figure 4.4 Heatmap Viewer – Section of Methylation Heatmap

For Heatmap Viewer instructions, see *Navigating the Heatmap Viewer* on page 51.

Navigating the Heatmap Viewer

Controls in the Heatmap Viewer allow you to zoom in to any location, determine the heatmap's color and contrast, and to research your data further through links to the Cancer Genome Workbench.

To access the Heatmap Viewer documentation, click the **Help** button at the top of the Heatmap Viewer, and select **Documentation**.

CHAPTER

5

WORKING WITH CLINICAL VIEWS

This chapter describes how to use the Cancer Molecular Analysis Portal to study clinical data and explore the relationships between clinical and molecular study data.

Topics in this chapter include:

Specifying Clinical Search Criteria

CMAP provides several tools for querying, visualizing, and downloading patient and *sample* data. You can search for data directly from the Clinical View workspace or via the Data Access Matrix. For instructions on accessing the Data Access Matrix, see *Accessing TCGA Data* on page 15.

The Clinical View workspace provides different search criteria depending on the context you are working with, TCGA or REMBRANDT. See *Searching for TCGA Clinical Data* on page 54 and *Searching for REMBRANDT Clinical Data* on page 57.

Related Topics:

- *Searching Open-Access Data*
- *Searching for TCGA Clinical Data*
- *Searching Open-Access Data*
- *Searching Controlled-Access Data*
- *Searching for REMBRANDT Clinical Data*
- *Navigating to the Query Results Page*
- *Navigating Clinical Report Table Results*

- *Understanding Sample-Based KM Plots*

Searching for TCGA Clinical Data

Open-access data groups are available to all CMA Portal users. For instructions on searching for TCGA open-access data, see *Searching for TCGA Clinical Data* on page 54. Other search criteria are available to registered viewers only. This serves to protect patient privacy. For instructions on searching for TCGA *controlled-access data*, see *Searching for TCGA Clinical Data* on page 54

For more information about patient privacy and data access policy, see TCGA Program Components page at:
http://cancergenome.nih.gov/components/hsp.asp

Searching Open-Access Data

The types of criteria available for searching for open-access TCGA data are more restricted than they are for controlled-access data.

How to Search Open-Access TCGA Data

1. Do one of the following to navigate to the Clinical View workspace:
   - From the Home workspace, click **Clinical View**.

   - **or** -

   - From any workspace in the portal, click the **Clinical View** tab.

   The Clinical View workspace appears (*Figure 5.1*).

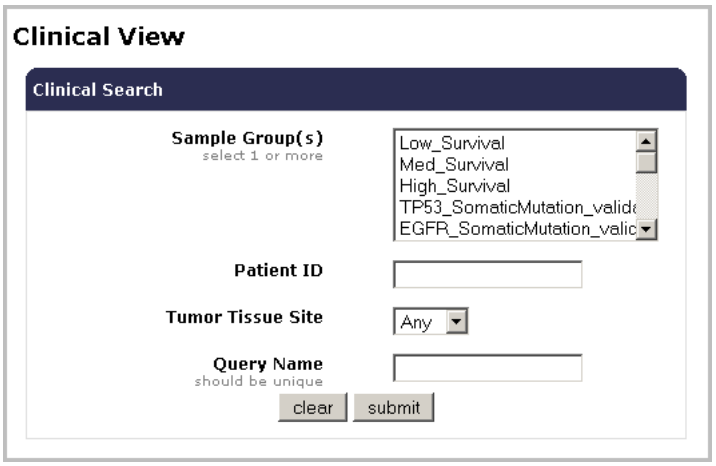


Figure 5.1 Clinical View – Clinical Search, TCGA Open-access Data

2. Select the groups to query from the **Sample Groups** list by doing any of the following:
   - To select a single group, click the group name.

   - To select several contiguous groups, click the first group, press and hold the SHIFT key, and select the last group in the series. Or, drag your cursor from the first group to the last group in the series.

- To select several discontinuous groups, click one group and CTRL + click additional groups.

3. To filter your query by patient ID, type the ID in the field provided.
4. To filter your query by the location of the tumor tissue, select the tissue of interest from the **Tumor Tissue Site** drop-down list.
5. In the **Query Name** box, type a unique name for the query.

   Caution: To prevent loss of data, do not type a name that you used for a previous query in your current session. The CMA Portal stores queries for the duration of your current CMA Portal session only.
6. To clear the values entered on the page and enter new values, click **Clear**.
7. To submit the query, click **Submit**.

   The CMA Portal processes your query and displays results in a new browser window.

For information about the results, see *Viewing Clinical Reports* on page 35.

Searching Controlled-Access Data

The types of criteria available for searching for controlled-access TCGA data enable you to search for more patient-related data than is possible with open-access data.

How to Search Controlled-Access TCGA Data

1. Do one of the following to navigate to the Clinical View workspace:

   - From the Home workspace, click **Clinical View**.

   - **or** -

   - From any workspace in the portal, click the **Clinical View** tab.

   The Clinical View workspace appears (*Figure 5.2*).

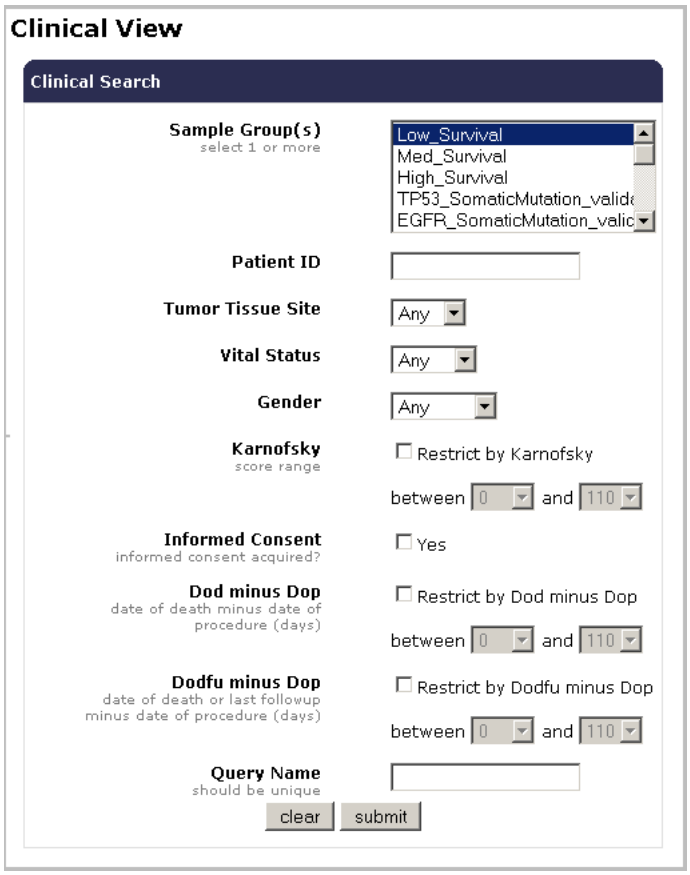


Figure 5.2 Clinical View – Clinical Search, TCGA Controlled Data

2. Select the groups to query from the **Sample Groups** list by doing any of the following:
   - To select a single group, click the group name.
   - To select several contiguous groups, click the first group, press and hold the SHIFT key, and select the last group in the series. Or, drag your cursor from the first group to the last group in the series.
   - To select several discontinuous groups, click one group and CTRL + click additional groups.
3. To filter by patient ID, type the ID in the field provided.
4. To filter by the location of the tumor tissue, select the tissue of interest from the **Tumor Tissue Site** drop-down list.
5. To filter by patient vital status, select **Alive** or **Dead** from the **Vital Status** drop-down list.
6. To filter by patient gender, select **M** (male), **F** (female), **O** (other) from the **Gender** drop-down list.

   Note: "O" appears as an option for REMBRANDT data only. It indicates that metadata for gender does not exist or has not been recorded.

7. To filter by patient functional capabilities, select the **Restrict by Karnofsky** check box, and then select the lower and upper scores from the drop-down lists.
8. To filter by those patients from whom informed consent was acquired, select the **Yes** check box.
9. To filter by the number of days between the date of death (DOD) and the date of the procedure (DOP), select the **Restrict by Dod minus Dop** check box, and then select the lower and upper limits from the drop-down lists.
10. To filter by the number of days between the date of death or the last followup (DODFU) and the date of the procedure (DOP), select the **Restrict by Dodfu minus Dop** check box, and then select the lower and upper limits from the drop-down lists.
11. In the **Query Name** box, type a unique name for the query.

    Caution: To prevent loss of data, do not type a name that you used for a previous query in your current session. The CMA Portal stores queries for the duration of your current CMA Portal session only.
12. To clear the values entered on the page and enter new values, click **Clear**.
13. To submit the query, click **Submit**.

    The CMA Portal processes your query and displays results in a new browser window.

For information about the results, see *Viewing Clinical Reports* on page 35.

Searching for REMBRANDT Clinical Data

Because all patient-related data in REMBRANDT is open-access, clinical views are more restricted than their controlled-access counterparts.

How to Search for REMBRANDT Patient Data in the Clinical View Workspace

1. Do one of the following to navigate to the Clinical View workspace:
   - From the Home workspace, click **Clinical View**.

   - **or** -

   - From any workspace in the portal, click the **Clinical View** tab.

   The Clinical View workspace appears (*Figure 5.3*).

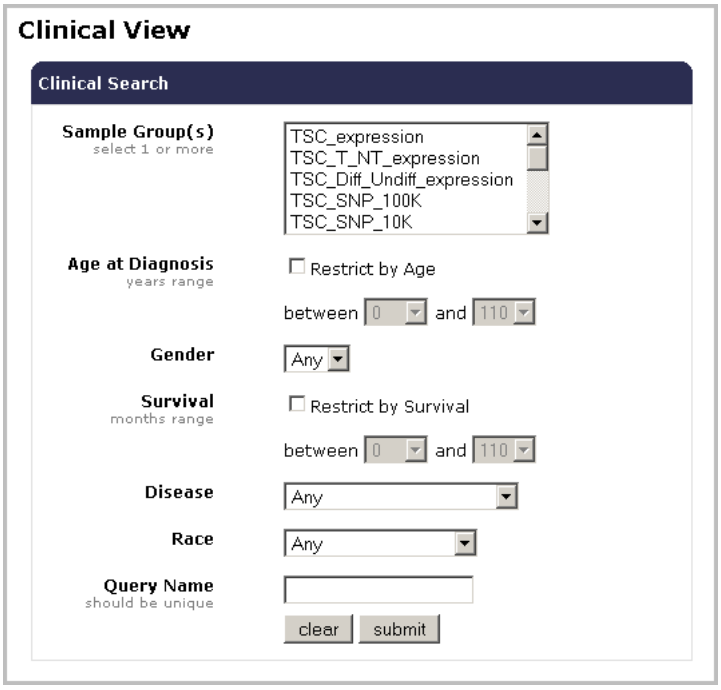


Figure 5.3 Clinical View – Clinical Search, REMBRANDT Context

2. Select the groups to query from the **Sample Groups** list by doing any of the following:
   - To select a single group, click the group name.
   - To select several contiguous groups, click the first group, press and hold the SHIFT key, and select the last group in the series. Or, drag your cursor from the first group to the last group in the series.
   - To select several discontinuous groups, click one group and CTRL + click additional groups.

Note: Steps 3 through 7 below are optional

3. To filter the query by patients age at diagnosis, select the **Restrict by Age** check box, and then select the lower and upper age limits from the drop-down lists provided.
4. To filter the query by patient gender, select **M** (male), **F** (female), **O** (other) from the **Gender** drop-down list.
5. To filter the query by the number of months a patient survived from the date if diagnosis, select the **Restrict by Survival** check box, and then select the lower and upper limits from the drop-down lists provided.
6. To filter the query by disease type, select the disease of interest from the **Disease** drop-down list.
7. To filter the query by patient race, select the race of interest from the Race drop-down list.
8. In the **Query Name** field, type a unique name for the query.

Caution: To prevent loss of data, do not type a name that you used for a previous query in your current session. The CMA Portal stores queries for the duration of your current CMA Portal session only.

9. To clear the values entered on the page and enter new values, click **Clear**.
10. To submit the query, click **Submit**.

    The CMA Portal processes your query and displays results in a new browser window.

For information about the results, see *Viewing Clinical Reports* on page 35.

Navigating to the Query Results Page

The system generates your custom query results and displays “completed” on the View Results page when the results are available for viewing. All queries you submitted and saved during your current CMA Portal session are listed on the View Results page as well.You can return to the Report Results page at any time to check on the processing status of a new query or to access results from a saved query.

To view the Report Results page, on the sidebar, click **View Results**. See *Side Bar Features* on page 14.

Related Topics:

- *Specifying Clinical Search Criteria*
- *Viewing Clinical Report*

Working With Clinical Reports

Results of your sample-based query appear in table format in a new browser on the Clinical Report page (*Figure 5.4*).

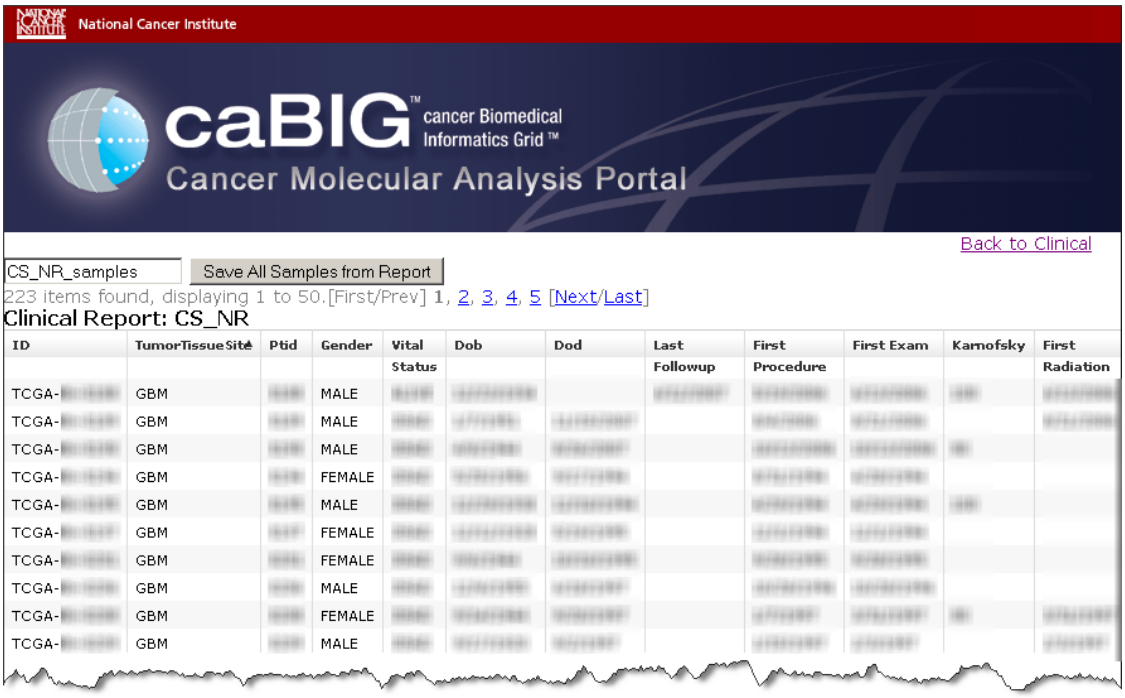


Figure 5.4 Clinical Report Page (partially redacted)

The types of data displayed in the report table reflect the criteria and filters you selected for your query. If you did not restrict your query by any of the criteria available, the report displays all columns in the table. The table column names displayed also reflect the context you have chosen and whether you are working with *controlled-access* or *open-access* data.

Table 5.1 lists the types of patient information available for each *sample* in your query. Some columns appear on both TCGA and REMBRANDT reports. Those columns that are unique to one or the other contexts are indicated.

Column Header	***Definition/Description***
Age at Diagnosis (REMBRANDT context)	Patient's age when first diagnosed
Disease (REMBRANDT context)	Tumor tissue type
DOB (TCGA context)	Date of birth – Calendar date of a patient's birth
DOD (TCGA context)	Date of death – Calendar date of a patient's death

Table 5.1 Clinical Report patient information for TCGA data

Column Header	*Definition/Description*
DODFU Minus DOP (TCGA context)	Date of death or last followup minus the date of the procedure – Number of days between the date of the patient's first procedure and date of death
DOD Minus DOP (TCGA context)	Date of death minus the date of the procedure – Number of days between the date of the patient's first procedure and date of death
First Exam	Date of the patient's first examination
First Procedure	Date of the patient's first procedure
First Radiation	Date of the patient's first radiation treatment
ID (REMBRANDT context)	REMBRANDT patient ID
Gender (REMBRANDT context)	Patient's gender
Grade (REMBRANDT context)	System for classifying cancer cells in terms of how abnormal they appear when examined under a microscope
Informed Consent Acquired? (TCGA context)	Indicates whether or not the patient's consent to provide samples was acquired
Karnofsky Score	Karnofsky Performance status scale – represents the patient's functional capabilities
Last Followup	Date of the patient's last followup examination
PatientDID (TCGA context)	Patient de-identified ID – Data that has been disassociated from a patient's personally identifiable information (PII). See *About PatientDID Lists* on page 78.
Patient_ID (TCGA context)	Patient identifier – Number associated with a patient as per TCGA code standards
PTID (TCGA context)	Patient identifier – Number associated with a patient without the TCGA prefix
Tumor Tissue Site (TCGA context)	Location of the tumor
Vital Status (TCGA context)	Indicates whether the patient is alive or dead

Table 5.1 Clinical Report patient information for TCGA data (Continued)

Note: REMBRANDT reports display many more columns of data than indicated in the table above. Refer to the REMBRANDT application for details at https://caintegrator.nci.nih.gov/rembrandt/.

Navigating Clinical Report Table Results

To assist you in selecting result records and analyzing the data, you can do the following:

- View multiple pages of records
- Sort the records

- Save all samples listed in the report
- Create a *PatientDID* list to use in other queries

Figure 5.5 displays the tools for accomplishing these tasks.

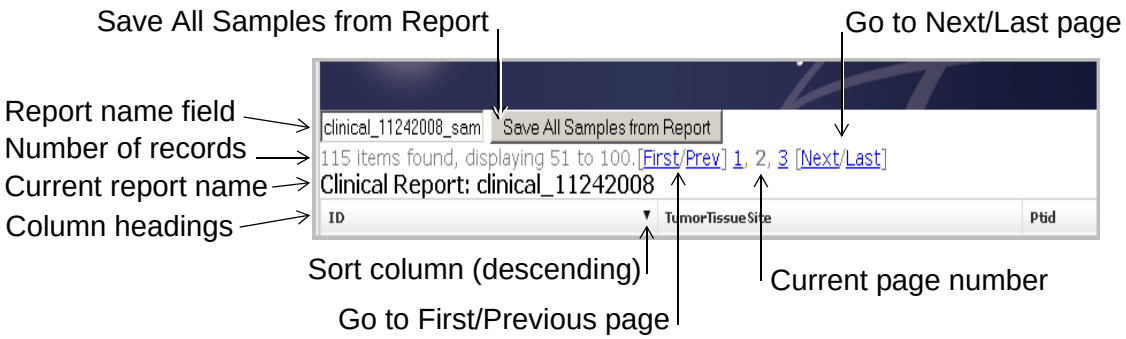


Figure 5.5 Clinical Report Table – Navigation

How to View Multiple Pages of Records

To view a subsequent page of the results table, click the **Next** link.

To view the last page of the report, click the **Last** link.

To view a previous page of the report, click the **Prev** link.

To view the first page of the report, click the **Prev** link.

How to Sort Records

You can sort the results table by any of the columns displayed.

- To sort the table by a column of data, click the column name.
- To change the sort order (e.g. ascending to descending order), click the arrow that appears in the column header by which you want to sort the table. Click it again to return to the previous order.

How to Create a PatientDID List From a Sample-based Clinical Report

1. Optionally, to rename the report, in the **Report Name** field (see *Figure 5.5* on page 62), type a unique name.

   Caution: To prevent loss of data, do not type a name that you used for a previous query in your current session. The CMA Portal stores queries for the duration of your current CMA Portal session only.

2. To save all the samples, click **Save All Samples from Report**.

   The query name appears on the Manage Lists page as a custom *PatientDID* list.

   You can modify any of the PatientDID lists on the Manage Lists page. For further information, see *Chapter 7, Managing Lists,* on page 77.

Creating Sample-Based KM Plots

How to Create a Sample-Based *Kaplan-meier* Plot

1. Do one of the following to navigate to the Clinical View page:
   - From the Home page, click **Clinical View**.

     - **or** -

   - From any page in the portal, click the **Clinical View** tab.

   The Clinical View page displays the Kaplan-Meier parameters at the bottom of the page (*Figure 5.6*).

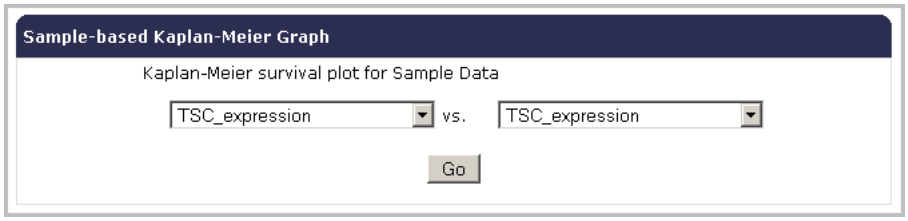


Figure 5.6 Clinical View – Sample-based Kaplan-Meier Graph Criteria

2. From the drop-down lists, select *sample* patient groups for comparison purposes.
3. Click **Go**.

The Sample-Based Kaplan-Meier survival plot appears (*Figure 5.7*).

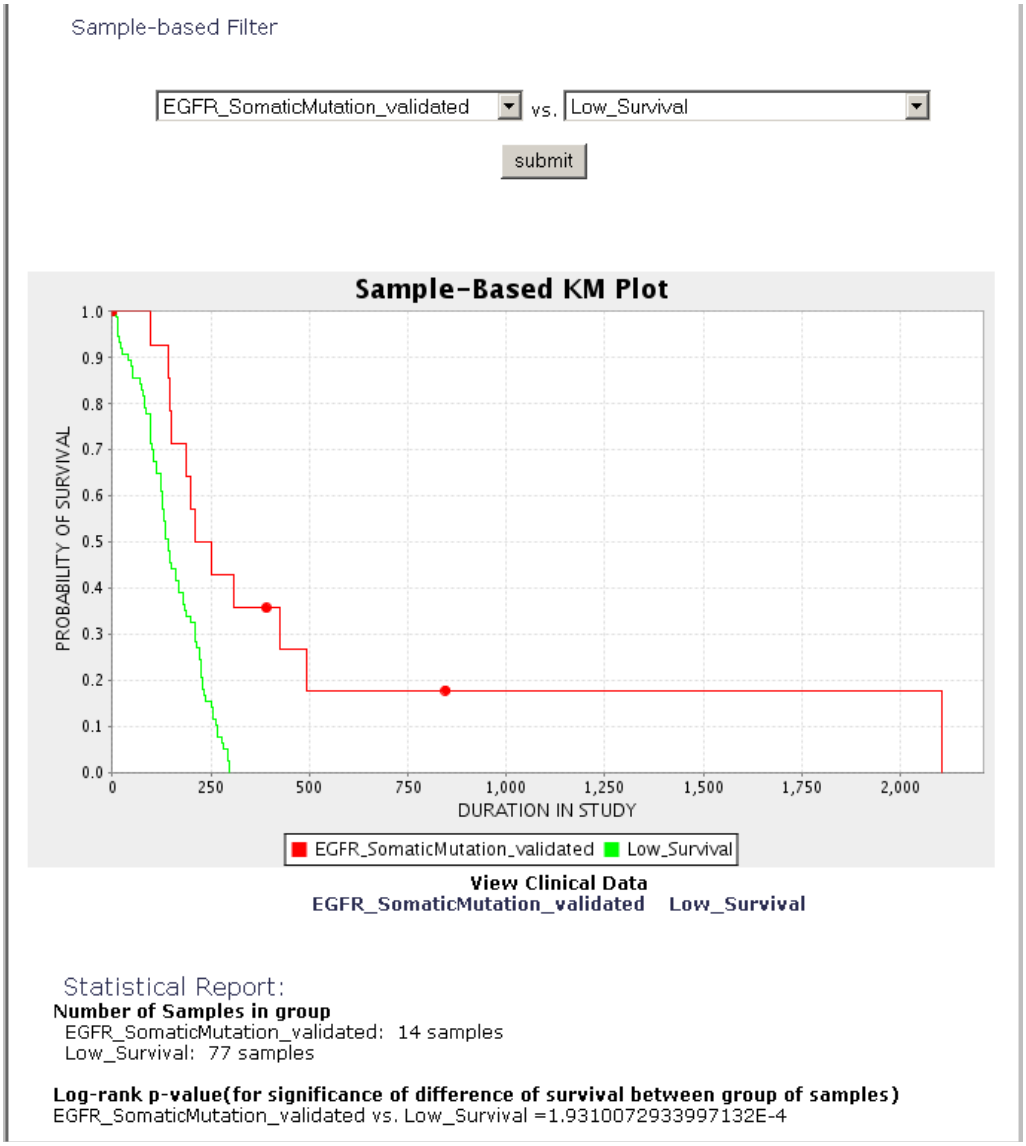


Figure 5.7 Sample-Based KM PLot

4. To compare other *sample* groups, select the two groups of interest from the Sample-based Filter drop-down lists at the top of the page, and then click **Submit**.

   A new Sample-Based Kaplan-Meier survival plot appears.

5. To view clinical data associated with the sample groups you selected, directly below the plot, click a sample group link, for example, EGFR_Somatic_validated.

   The Clinical Report appears in a new browser window. See *Navigating to the Query Results Page* on page 59.

Related Topics:

- *Understanding Sample-Based KM Plots*
- *Creating Gene Expression-Based Kaplan-Meier Plots*
- *Creating Sample-Based KM Plots*
- *Creating Gene Expression Plots*

Understanding Sample-Based KM Plots

A Sample-Based *Kaplan-Meier* plot *(Figure 5.8)* shows the survival rate at each time point for *samples* with certain genome characterization characteristics (e.g., EGFR mutation levels in tumor samples are greater than those in the non-tumor samples by 3 fold or higher). Kaplan-Meier estimates are calculated based on the last follow-up time and the censor status (0=alive, 1=dead) from the samples of interest. The Kaplan-Meier estimates are then plotted against the survival time. The points that correspond to the events with censor status of 0 are indicated on the graph.

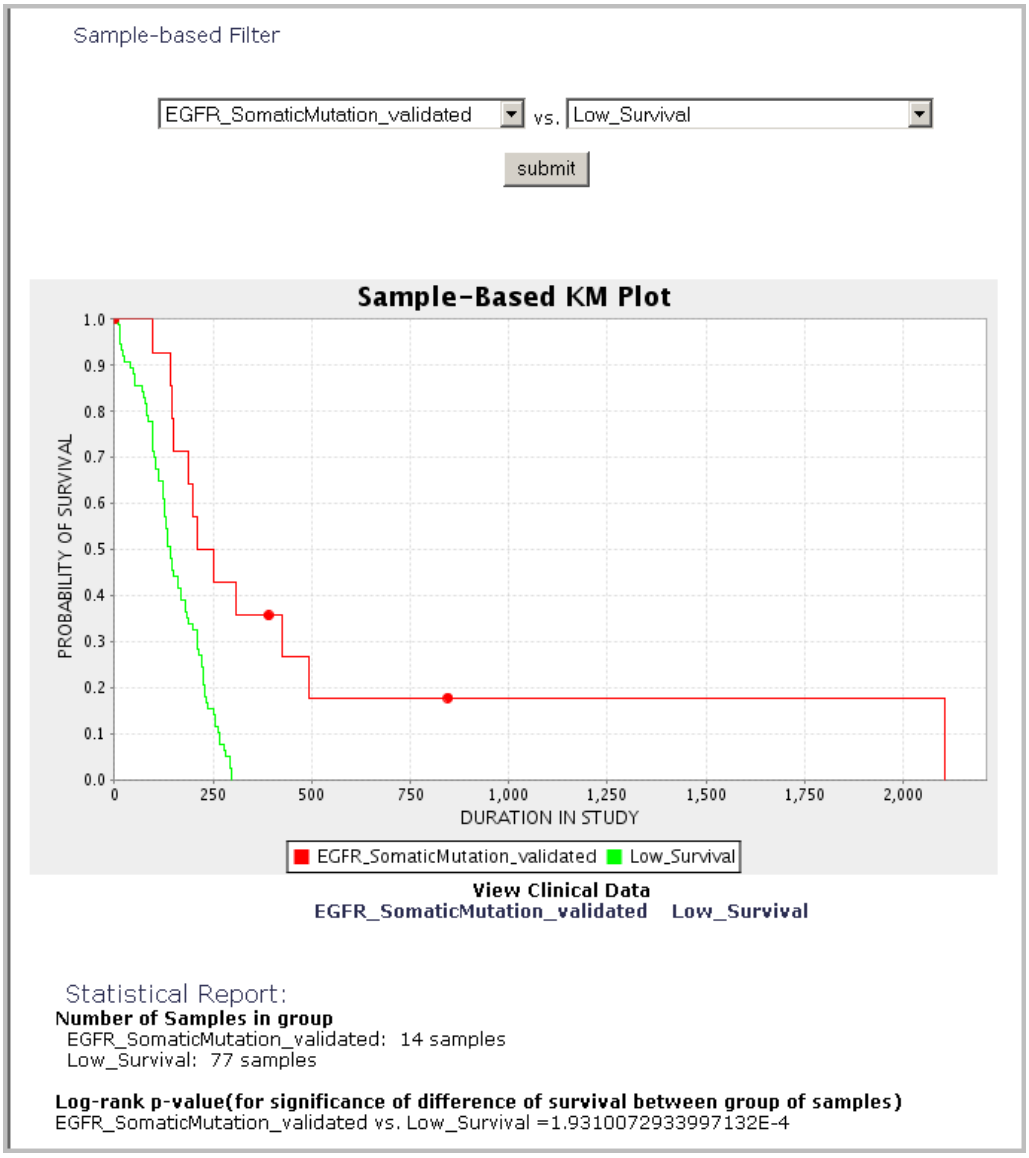


Figure 5.8 Sample-Based KM Plot

Table 5.2 describes areas on the Sample-based Kaplan-Meier Survival Plot.

Item	***Special Instructions***
Sample-based Filter	To filter the plot, select new criteria from the drop-down lists at the top of the page, and then click **Submit**.
View Clinical Data	To display clinical data for the selected *sample* groups, click the group link. For more information, see *Working With Clinical Reports* on page 60
Statistical Report	• **Number of Samples in the group** specifies the number of samples, per group in the plot. • **Log-rank *p*-value** indicates the significance of the difference in survival between any two groups of samples segregated based on gene expression of the gene of interest. The log rank *p*-value is calculated using the Mantel-Haenszel procedure. The *p*-values are recalculated every time a new threshold is selected.

Table 5.2 Sample-Based KM Plot page description

Chapter 6
Analysis Tools

This chapter describes how to use the Cancer Molecular Analysis Portal to perform principal component and gene pattern analyses, and to access associated analysis applications.

Topics in this chapter include:

Overview

The Analysis Tools page provides access to the following tools and applications:

- Principal Component Analysis (PCA)
- Gene Pattern Analysis
- Gene Pattern Home
- Integrated Heatmap Viewer
- Cancer Genome Workbench (CGWB)

Principal Component Analysis

Principal component analysis (PCA) is method of identifying and highlighting patterns in data for the purpose of finding similarities and differences. PCA algorithms compress

the number of dimensions of data to make the visual display more meaningful in terms of pattern recognition.

Related Topics:

- *Selecting Search Criteria for Principal Component Analysis*
- *Working With PCA Plots*
- *Changing the PCA Display*
- *Selecting and Saving Samples in PCA Plots*

Selecting Search Criteria for Principal Component Analysis

Principal Component Analysis (PCA) enables you to find the main dimensions of variation in a multi-dimensional data set.

How to Select Criteria For Analysis

1. On the Analysis Home page, click **Principal Component Analysis**.

   The Principal Component Analysis page appears (*Figure 6.1*).

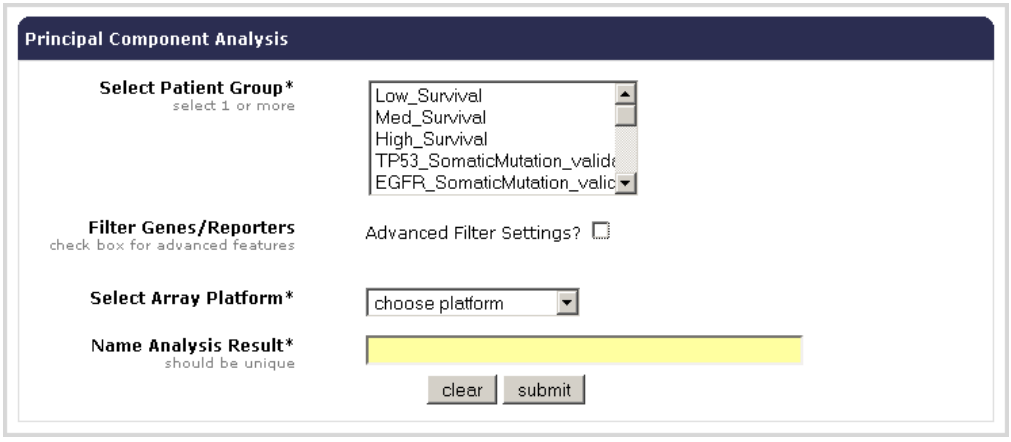


Figure 6.1 Analysis Tools – Principal Component Analysis

2. Select the *patient groups* to query from the patient group list by doing any of the following:
   - To select a single group, click the group name.
   - To select several contiguous groups, click the first group, press and hold the SHIFT key, and select the last group in the series. Or, drag your cursor from the first group to the last group in the series.
   - To select several discontinuous groups, click one group and CTRL + click additional groups.

   Note: You must select a minimum of two lists.
3. To view results with the default filter, select the **Default** option.
4. To constrain your query by variance (Gene Vector) percentile, click the **Advanced** filter option.

The Advanced options are displayed (*Figure 6.2*).

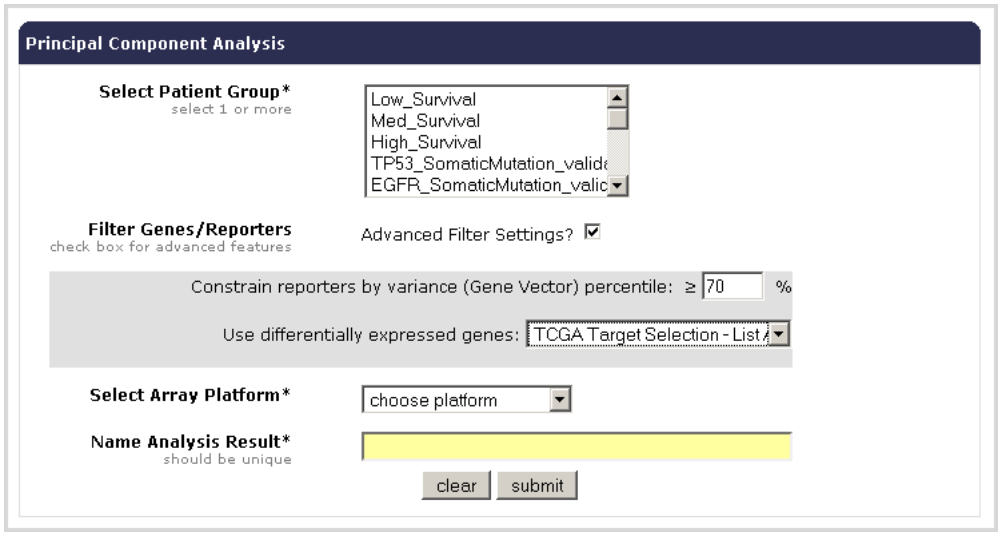


Figure 6.2 Advanced Filter Settings

5. Type a percentage in the text box provided to select the reporter genes whose variances of the log ratio (or log2 signals) across all experiments were among the top percentile of variance of all reporter genes identified. For example, 70% selects reporter genes with the top 30 (100 - 70) percentile of variance.
6. To filter by differentially expressed genes, select TCGA from the drop-down list.

   Note: Additional targets will be available as data is uploaded to the system.
7. Select an array from the **Select Array Platform** drop-down list.

   Note: In the REMBRANDT context, Affy HT Human Genome U133 Plus 2.0 is the only array available.
8. Type a unique name for the query.
9. Click **Submit**.

   The system processes your query and presents a link to the results (*Figure 6.3*).

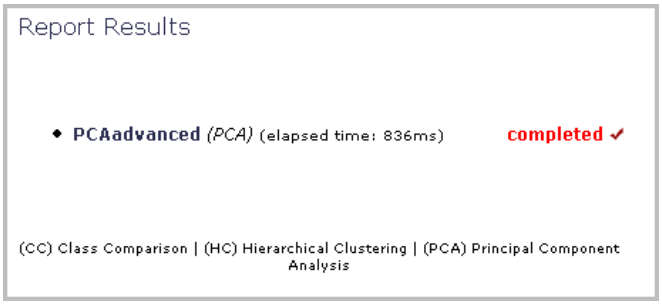


Figure 6.3 Query Process Completed

10. When the **completed** message is displayed, click the link to access the PCA plot.

The PCA plot appears. (See *Figure 6.4*.)

Working With PCA Plots

The principal component analysis (PCA) plot displays the results of your query. PCA enables you to find the main dimensions of variation in a multi-dimensional data set.

Each point on the graph represents a *sample*, and each sample group is color-coded (*Figure 6.4*).

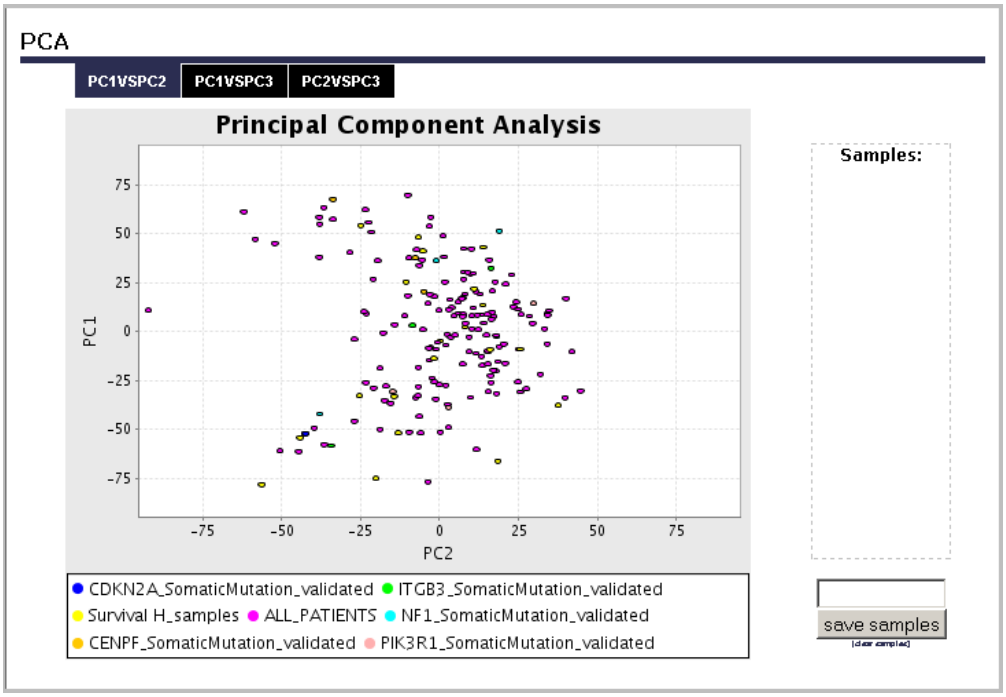


Figure 6.4 PCA Plot

Changing the PCA Display

To display **PC1 versus PC2**, **PC1 versus PC3**, or **PC2 versus PC3**, click the appropriate tab above the plot.

PC1 is the first principal component, which accounts for most of the variation in the multidimensional data set. After correcting for the variation due to PC1, PC2—the second principal component—is the dimension that accounts for most of the remaining variation in the dataset. Subsequent analyses correct for remaining variation.

Selecting and Saving Samples in PCA Plots

The Samples area enables you to select, review, and save *samples* in the plot.

How to How to Select and Save Samples

1. On the PCA plot, drag your cursor over a sample of interest, or to select a group of samples, drag your cursor across two or more samples. A red outline appears around your sample(s), and the sample name(s) is/are added to the **Samples** list to the right of the plot (*Figure 6.5*).

   Tip: Hover your cursor over a sample name to view the entire ID.

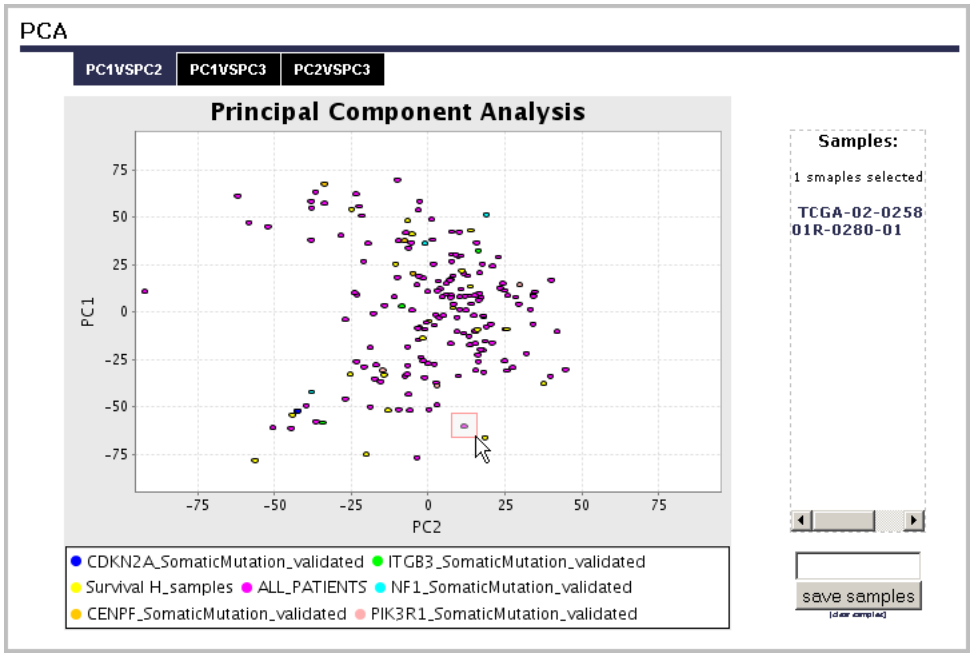


Figure 6.5 PCA Plot – Selected Sample

2. To save the sample list, in the **Save Samples** field, type a unique name for the list, and then click **Save Samples**.

   The sample list appears on the List Management page. See *Viewing and Managing Lists* on page 79.

Gene Pattern Analysis

The Gene Pattern Analysis tool enables you to select groups of patients, genes, and array platform data to use as criteria for analysis in the GenePattern application. GenePattern provides access to a broad array of computational methods used to analyze genomic data.

Related Topics:

- *Selecting Criteria for Gene Pattern Analysis*
- *Principal Component Analysis*

Selecting Criteria for Gene Pattern Analysis

You can use the Cancer Molecular Analysis Portal's gene pattern analysis tool to select samples for computation in the GenePattern application.

How to How to Select Criteria For Gene Pattern Analysis

1. On the **Analysis Tools** page, click **Gene Pattern Analysis**.

   The Analysis Module page appears (*Figure 6.6*)

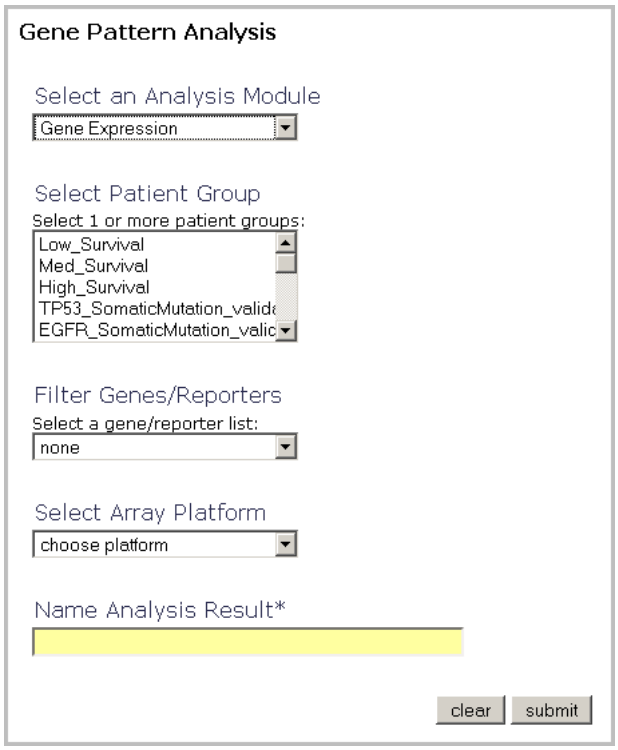


Figure 6.6 Analysis Tools – Gene Pattern Analysis, TCGA Context

Note: Limited query criteria are available for REMBRANDT data as shown in *Figure 6.7*.

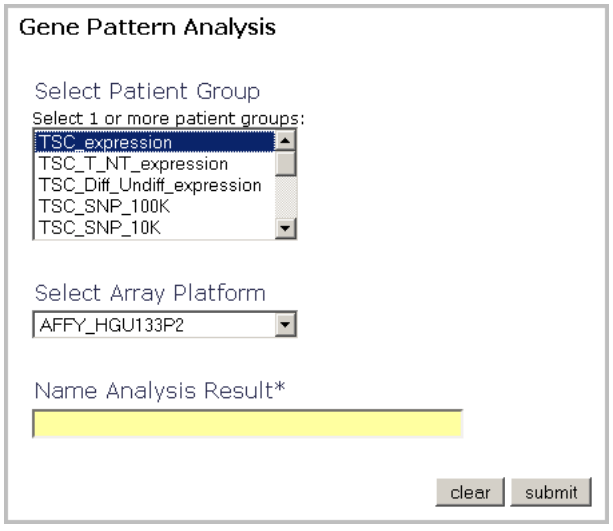


Figure 6.7 Analysis Tools – Gene Pattern Analysis, REMBRANDT Context

2. Select either **Gene Expression** or **Copy Number** from the module drop-down list (*Figure 6.8*).

   Note: The copy number feature is not available for REMBRANDT data.

   If you selected **Copy Number**, the **Gene Pattern Analysis** page expands to display chromosome number filter criteria.

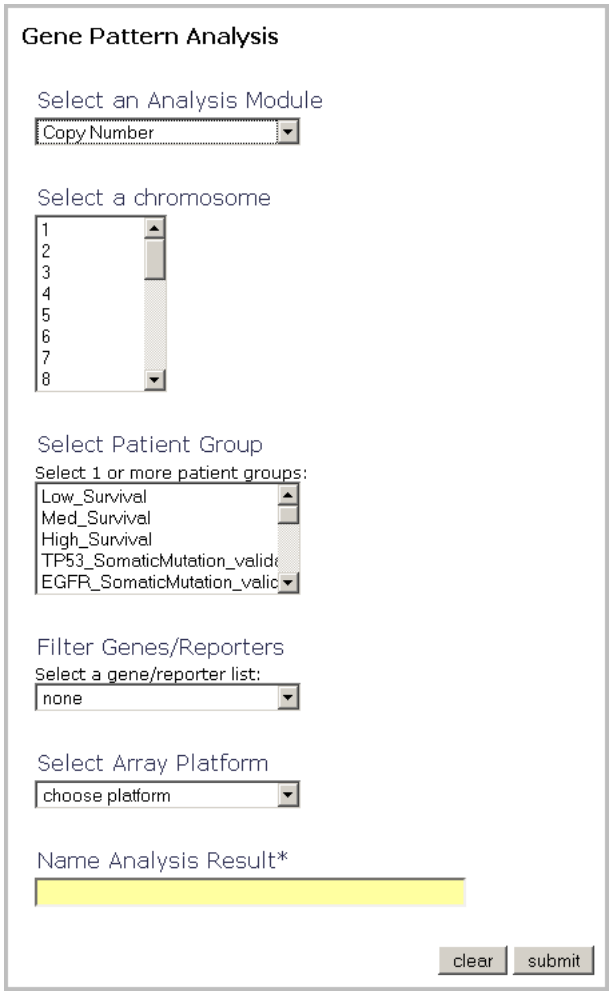


Figure 6.8 Gene Pattern Analysis, Copy Number Module, TCGA Context

3. If you selected the **Copy Number** module, select the chromosome number (1-22, X, Y) of interest from the **Select a Chromosome** drop-down list.
4. Select one or more *patient groups* to query from the **Select Patient Group** list. For instructions on selecting multiple groups, see step 2 on page 70.
5. To filter by differentially expressed genes, select TCGA from the **Filter Genes/ Reporters** drop-down list.

   Note: Additional targets will be available as data is uploaded to the system.
6. Select an array from the **Select Array Platform** drop-down list.

   Note: In the REMBRANDT context, Affy HT Human Genome U133 Plus 2.0 is the only array available.
7. Type a unique name for the query in the **Name Analysis Result** field.
8. Click **Submit**.

   The system processes your query and presents a link to the results (*Figure 6.9*).

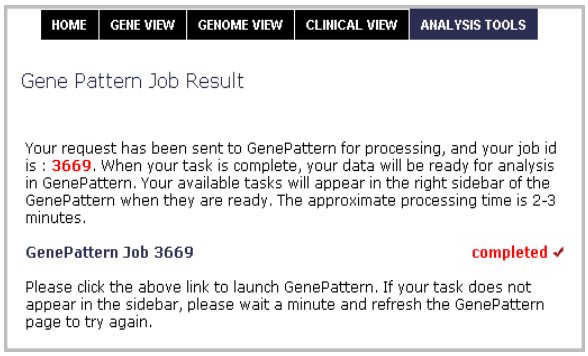


Figure 6.9 Query Process Completed

9. When the **completed** message is displayed, click the link to access the GenePattern application.
10. Follow the instructions provided on the GenePattern website to conduct your analysis.

GenePattern Home

The GenePattern Home tool provides direct access to the GenePattern application. Follow the instructions provided on the GenePattern website to conduct your study.

Integrated Heatmap Viewer

The Heatmap Viewer tool provides direct access the CMA Portal Genome View. For details, see *Chapter 4, Working With Genome Views,* on page 47.

Cancer Genome Workbench

The Cancer Genome Workbench tool provides direct access to the Cancer Genome Workbench application. Follow the instructions provided in the application help on the Cancer Genome Workbench website to conduct your study.

CHAPTER 7
MANAGING LISTS

This chapter describes how to manage patient and/or gene lists by editing application-defined lists and creating new custom lists.

Topics in this chapter include:

List Management Overview

The Manage Lists function centralizes all activities pertaining to the creation and management of user-defined, as well as study-defined, *PatientDID* lists, gene lists, and *reporter gene lists*. With these lists, you can further refine queries to facilitate analysis.

Related Topics:

- *Creating Custom Lists*
- *Viewing and Managing Lists*
- *Deleting Items From a List*
- *Deleting Lists*

List Types

The CMA Portal sidebar contains a number of lists designed to facilitate analysis by providing sets of predefined data elements (patient ID, gene names, etc.).

These lists include:

- *PatientDID* Lists

- Gene Lists
- *Reporter Gene* Lists

Related Topics:

- *About PatientDID Lists*
- *About Gene Lists*
- *About Reporter Gene Lists*

About PatientDID Lists

DID information is data that has been disassociated from a patient's personally identifiable information (PII). The CMA Portal system provides DIDs rather than complete IDs in order to protect patient privacy.

The PatientDID list section on the Manage Lists page contains somatic mutations (data type)/somatic variants compiled from genomic sequencing centers. Each list in the section comprises a set of patients with a given set of unique characteristics. You can use default or custom lists to filter your queries.

Biorepositories that contain high quality cancer collections were chosen to provide clinical data and biospecimens (tumor and normal tissues) for analysis. Clinical datapoints collected include, but are not limited to:

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgery

Caution: CMA Portal saves your custom *PatientDID* lists for your current session only. Once you log out of your session, you can not retrieve them.

Table 7.1 describes the system-defined patient survival groups.

TCGA PatientDID List	***Description***
ALL_Patients	List of all patients available for investigation.
Low_Survival	List of patients who survived up to, and including, 300 days.
Med_Survival	List of patients who survived longer than 300 days but fewer than 900 days.
High_Survival	List of patients who survived longer than 900 days.

Table 7.1 PatientDID List descriptions

Related Topics:

- *List Types*

About Gene Lists

Each list in the Genes Lists section on the Manage Lists page contains a set of genes of interest. Currently you can use either default TCGA or REMBRANDT genes, or custom lists to filter your queries.

Note: Currently the available data displayed is provided by TCGA Centers as of a given date. When working in the REMBRANDT context, the data displayed is provided by REMBRANDT collaborators. Other gene targets will be available in subsequent releases of this portal.

You can download TCGA gene lists at:
http://gforge.nci.nih.gov/docman/view.php/259/7051/TCGA%20target%20lists.xls

About Reporter Gene Lists

You can create and use custom *reporter gene* lists for your queries.

Note: No predefined reporter gene lists are available currently. However, you can define custom lists.

Viewing and Managing Lists

Figure 7.1 illustrates the PatientDID, Gene, and Reporter lists as they appear with their respective content when managing lists. Default CMA Portal lists are displayed in black text; custom lists you may have created in your current session are displayed in red text. In addition to the lists displayed in the figure, registered CMA Portal users can create custom clinical data lists. See *Creating Custom Lists* on page 83.

Note: You may have to scroll down the page to see all lists.

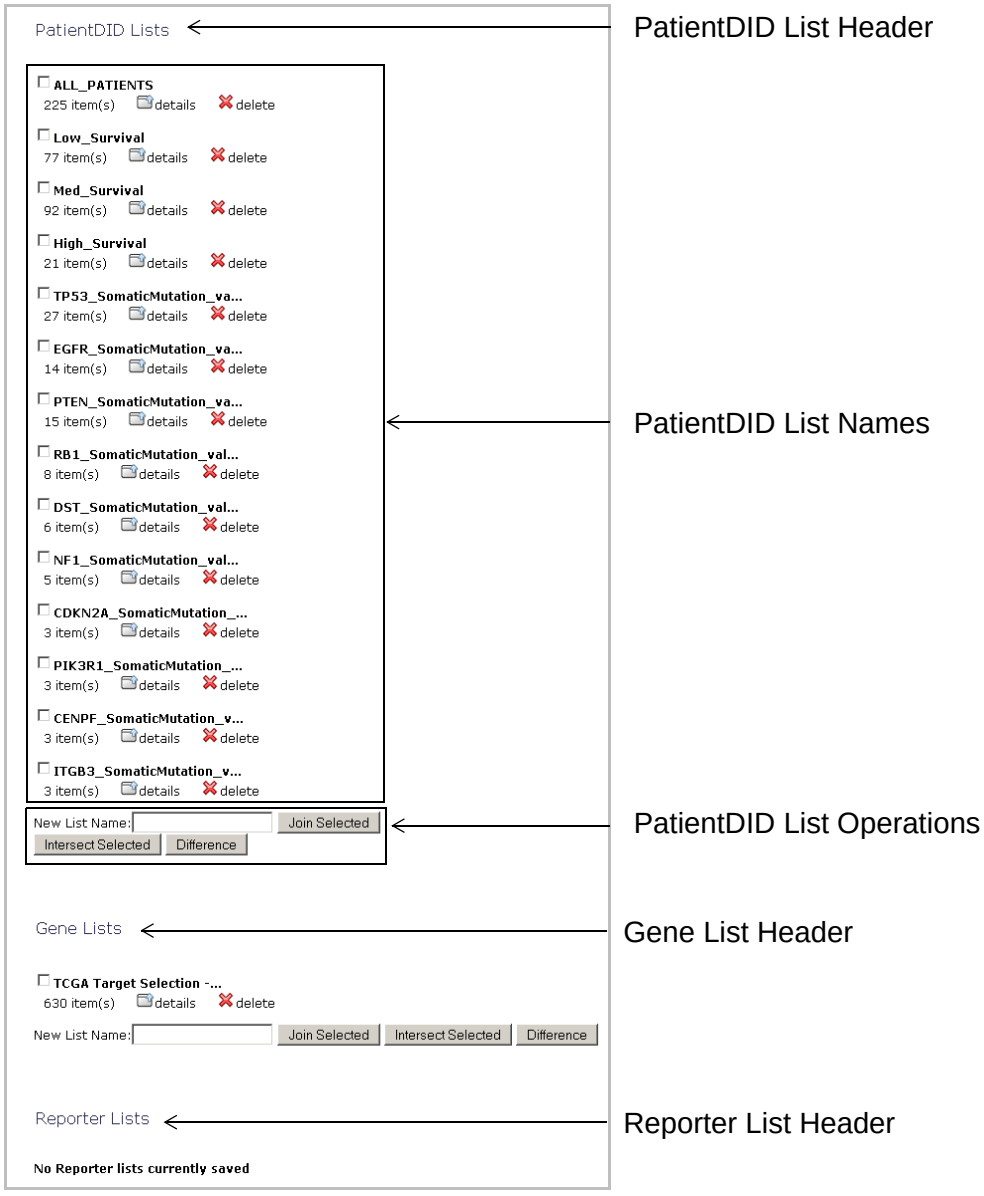


Figure 7.1 List Management Page – TCGA Context

The CMA Portal enables you to manage the content of predefined and custom lists as follows:

- View details of a list's content. See *Viewing List Details* on page 81.
  - View a list's creation date and notes. See *Viewing List Creation Data and Notes* on page 81.
- Delete a list. See *Deleting Lists* on page 82.
  - Delete an item from a list. See *Deleting Items From a List* on page 82.

Related Topics:

- *Viewing List Details*

Viewing List Details

How to How to View the Individual Data Items In a List

1. At the top of the **List Management** page, click the type of list you want to view (**PatientDID Lists**, **Gene Lists**, or **Reporter Lists**).
2. To view the items in a list, next to the list name, click **Details**. Click **Details** again to hide the details.

The list details appear below the list name (*Figure 7.2*).

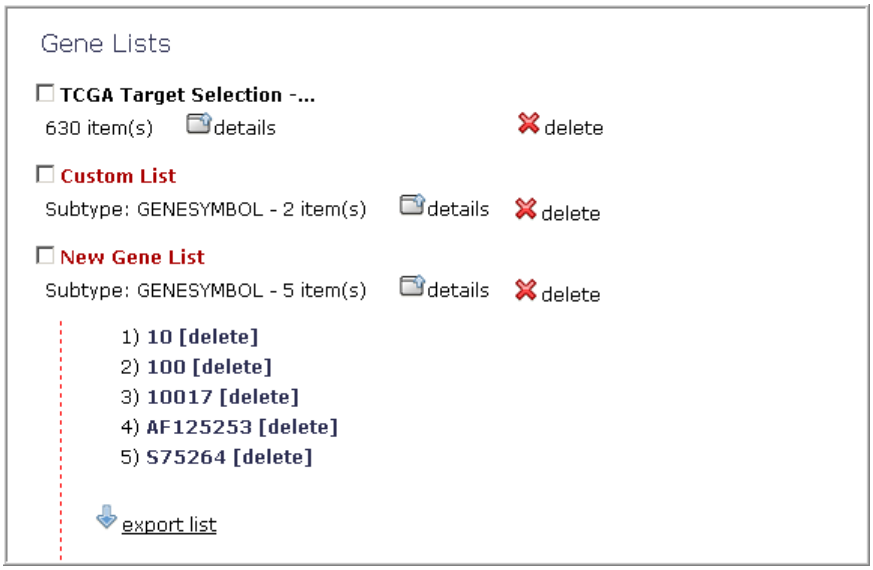


Figure 7.2 List Details

Viewing List Creation Data and Notes

Each list has metadata associated with it, including its author, creation date, and notes.

How to How to View the List Metadata

- Hover your cursor over the name of a list on the **List Management** page. The list's creation date and notes appear in a popup window.

For example, *Figure 7.3* displays metadata for the EFGR Somatic Mutation (validated) list.

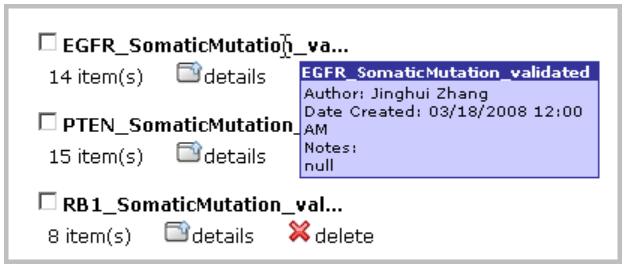


Figure 7.3 PatientDID Lists Section – EGFR List Metadata

Deleting Lists

You can delete any list on the List Management page, but you can not retrieve a list that you have deleted.

How to How to Delete a List

1. At the top of the **List Management** page, click the type of lists you want to delete (**PatientDID Lists**, **Gene Lists**, or **Reporter Lists**) (*Figure 7.4*).

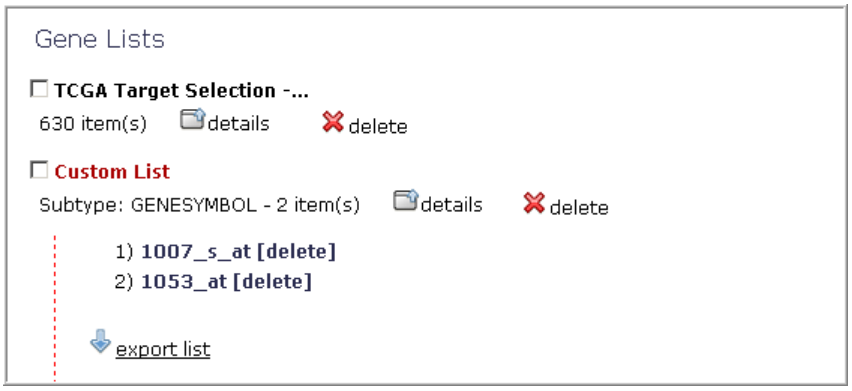


Figure 7.4 Gene Lists – Custom List Details

Caution: You can not retrieve a list once you have deleted it.

2. Click the **delete** icon (**x**) next to the name of the list you want to delete.

The list is deleted.

Deleting Items From a List

You can delete any item on the List Management page, but you can not retrieve an item that you have deleted.

How to How to Delete One or More List Items

1. Click the **Details** icon next to a list to display its content.

Caution: You can not retrieve a list once you have deleted it.

2. To the right of the item you want to delete, click **[delete]**.

The item is deleted from the list.

Creating Custom Lists

You can create new lists from two or more existing lists by joining or intersecting two or more lists; or by subtracting one list from another. *Figure 7.5* illustrates the end result of each of these operations.

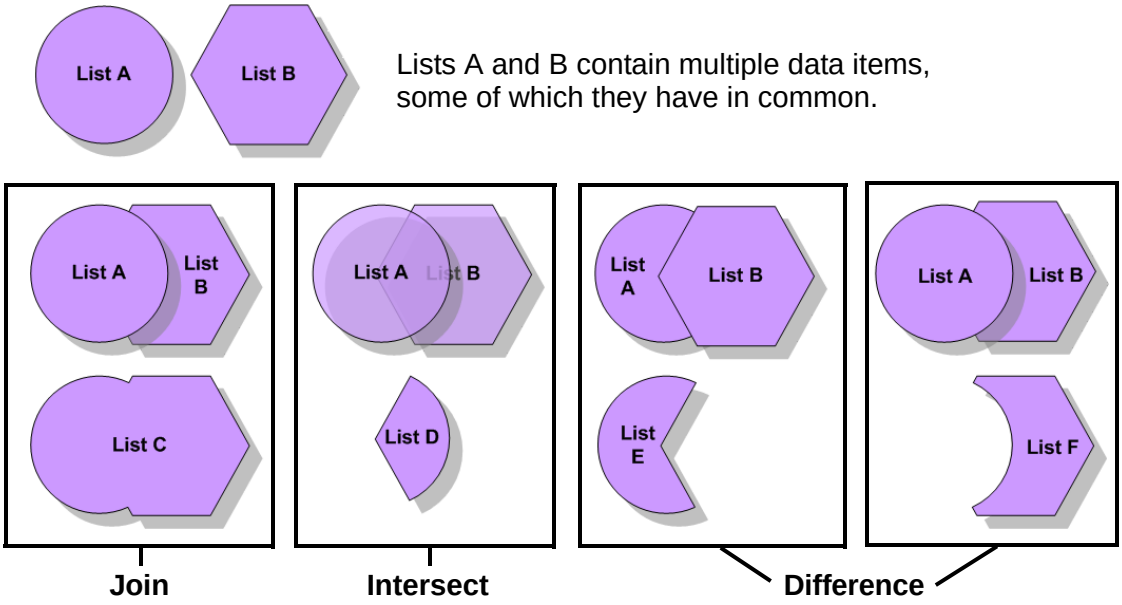


Figure 7.5 List Functions

In the illustration, List A is represented by a circle and List B by a hexagon. Each list contains multiple data items, for example, gene symbols, some of which they have in common. As a result of commingling the lists, you can create new lists as follows:

- List C (Join) – Data elements of Lists A and B combined.
- List D (Intersect) – Data elements that are common to both List A and List B
- List E (Difference) – Data elements that are *not* common to both List A and List B, where common data elements are subtracted from List A
- List F (Difference) – Data elements that are *not* common to both List A and List B, where common data elements are subtracted from List B

How to How to Create a Custom List From Existing Lists

1. At the top of the sidebar, click **List Management**.
2. To navigate to the group of lists from which you want to create a custom list, click the appropriate link at the top of the List Management page (*Figure 7.6*). For example, to add to the list of genes, click Gene List.

Figure 7.6 List Management Menu <<Tech writer: create new screen shot when "Add List" functionality is operational>>

3. Select the check box next to the list name(s) that you want to include in a given operation to create a new list (*Figure 7.1*).

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

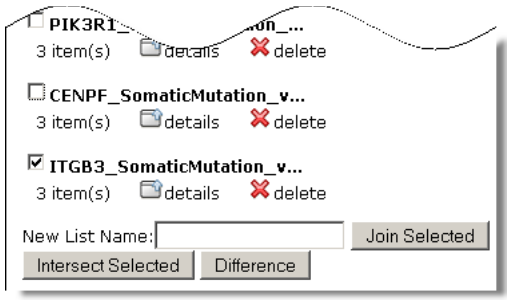


Figure 7.7 Patient DID List

4. In the **New List Name** box, type a unique name for the new list and then do one of the following operations:

   - To create a new list that contains all data items from two or more lists, click **Join**.
   - To create a new list that contains only data items that the selected lists have in common, click **Intersect**. If the selected lists have no items in common, the list is created but contains no data items.
   - To create new lists that contain only data items that are unique to each of two lists, click **Difference**. *Figure 7.8* provides examples of the two lists that are created as a result of the Difference operation.

     For example, if you select **All Patients** and those with **TP53** mutations, two new lists are created as follows:

     — **Difference_High_Survival** – Contains items unique to the High Survival group. (High Survival DIDs minus TP53 DIDs)

     — **Difference_TP53** – Contains items unique to the TP53 group. (TP53 DIDs minus High Survival DIDs.)

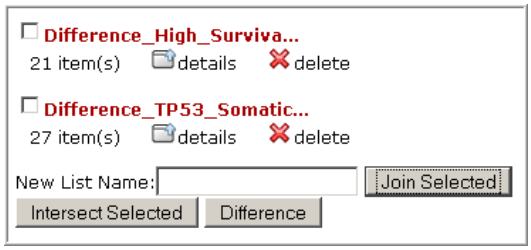


Figure 7.8 New Difference Lists

   New lists appear on the Manage Lists page in red text.

5. To view the items in new lists, hover your cursor over the name of the list in the sidebar.

   The list's data items appear in a popup window. See *Figure 7.3* on page 82.

6. To view the date and time that the list was created, hover your cursor over the name of the list in the **List Management** page.

   The list's creation date and time appear in a popup window.

Related Topics:

- *Viewing and Managing Lists*
- *Deleting Items From a List*
- *Deleting Lists*

GLOSSARY

This glossary defines acronyms, abbreviations, and terminology used in this guide.

Term	***Definition***
Affy_SNP6	DNA analysis array that covers 1.8 million genetic markers, including more than 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation.
Affy HT Human Genome U133	High-throughput expression profile of approximately 40,0000 transcripts and variants.
Affy Human Exon 1.0	Contains approximately one million predicted and confirmed exons.
Agilent 8 x 15K Human miRNA-specific Microarray	Contains probes for 723 human and 76 human viral microRNAs from the Sanger database v.10.1.
Agilent Whole Human Genome	High-density profiling analysis tool that covers over 41,000 unique human genes and transcripts
amplified	Condition when the number of copies of a gene is equal to or greater than 2.5 times the original.
ANOVA	Analysis of Variance which simplifies the F-test, where F-test is the mean square for each main effect and the interaction effect divided by the *within* variance. A one-way ANOVA or single factor ANOVA tests differences between the groups classified only on one independent variable.
BCM	Baylor College of Medicine
caBIG®	Cancer Biomedical Informatics Grid
CBIIT	(National Cancer Institute) Center for Bioinformatics and Information Technology
controlled data	Non-public, protected data. Ensures patient privacy.
copy number	Differences in the number of copies of a particular gene in a single *sample*. Also known as copy number variation (CNV).
deleted	Condition when one or more genes have been deleted.
DOD	Date of death
DODFU	Date of death or last follow-up
DOP	Date of procedure

Term	Definition
gene expression	Process by which proteins are made from the instructions encoded in DNA.
intensity	Geometric mean value calculated for each comparison group.
Kaplan-Meier survival plot	Survival probability for the user-defined set of criteria as a function of time and survival differences as analyzed by the log-rank test.
LOH	Loss of heterozygosity
methylation	Enzymatic addition of methyl (CH3) group to DNA which causes inactivation of that region.
mismatch (MM)	Condition in which DNA bases from one strand are not complementary to the bases from the other strand.
mutated	Alteration in DNA sequence that is either induced by a mutagen or is spontaneous
NCI	National Cancer Institute
NCICB	National Cancer Institute Center for Bioinformatics. Now known as NCI-CBIIT (National Cancer Institute Center for Bioinformatics and Information Technology)
normalization	Used to designing relational database tables and minimizing duplicated data.
open data	Data that is available to the general public.
overexpressed	TheCancer Molecular Analysis Portal when the fold ratio is twice the control value or average.
patient group	Pre-defined or user-defined patientDID list comprising patient identifiers with certain characteristics
patientDID	Data from tissue *sample* analysis from which personally identifiable information has been removed such that a researcher can no longer trace data back to an individual patient.
perfect match (PM)	Set of oligonucleotides whose sequence exactly matches the gene of interest
probe	Labeled segment of DNA that is used to bind to and identify a gene or mRNA transcript.
probe set	Multiple probe pairs. Each probe pair consists of two groups of probes—a *perfect match* (PM) and a *mismatch* (MM).
reporter gene	Gene that codes for a product that can readily be measured, such as a fluorescing protein. Often used for expression studies of heterologous promoters.
sample	Biological tissue sample
sample threshold	A final percentage threshold applied to the *samples* used to determine whether the gene is an anomaly.
SNP	Single-nucleotide polymorphism
standard deviation	Statistical measure of spread or variability.
SVG plugin	Integrates with your Web browser as a plug-in and enables you to display SVG images like the pathway diagram.

Term	***Definition***
TCGA	The Cancer Genome Atlas
tumor mutation samples	The subset of tumor *samples* where a mutation has been found in that particular gene.
underexpressed	TheCancer Molecular Analysis Portal when the fold ratio is less than twice the control value or average.
value threshold	The initial threshold applied to data to determine an anomaly.

INDEX

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

W