I-SPY
USER'S GUIDE

Version 1.0

NATIONAL® CANCER INSTITUTE

Center for Bioinformatics

July 30, 2007

Credits and Resources

I-SPY *Development and Management Teams*		
Development	***Documentation***	***Product and Program Management***
Subha Madhavan	Laura Jackel	Subha Madhavan
Alex Jiang	Ying Long	
Kevin Rosso	Huaitian Liu	
Ryan Landy		
Himanso Sahni		
David Bauer		
Huaitian Liu		
Michael Harris		
Ram Bhattaru		
Ye Wu		
Ying Long		
Don Swan		
Dana Zhang		
Vesselina Bakalov		
Nick Xiao		
Gregg Silk		

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

TABLE OF CONTENTS

About This Guide

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

Purpose

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

Audience

This guide is designed to assist researchers and investigators using the I-SPY Analysis Portal application.

Topics Covered

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

- *Chapter 1* provides an overview of the I-SPY program.
- *Chapter 2* provides instructions to start using I-SPY.
- *Chapter 3* describes how to search on patient or sample identiers. The results show the patients that fulfill the criteria.
- *Chapter 4* describes how to perform a clinical and an *IHC (Immunohistochemistry)* Level of Expression and Loss of Expression query.
- *Chapter 5* extends the basic knowledge of the previous chapters and shows you how to work with five types of high order analyses.
- *Chapter 6* describes how to view all the results generated from searches and high order analyses.
- *Chapter 7* describes how to manage user-defined and study-defined patient and gene identifier lists.

- *Appendix A* provides a data dictionary listing the variables for I-SPY.
- *Glossary* describes terms used in this guide.

Text Conventions Used

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

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

Table Documentation conventions

CHAPTER 1
ABOUT I-SPY

This chapter introduces you to I-SPY and provides an overview of I-SPY functions.

Topics in this chapter include:

About I-SPY

The NCI Center for Bioinformatics (NCICB), in collaboration with physicians, researchers, and cooperative groups, has designed I-SPY. Clinical trials are critical to identifying markers and mechanisms of resistance in therapy, and I-SPY is a multi-center clinical trial for women undergoing neoadjuvant chemotherapy from breast cancer. I-SPY is a web-based system which supports correlative data analysis and centralized reporting of results to catalyze the transition from uniform to tailored care.

I-SPY facilitates collaboration, provides an infrastructure for data management, analysis and communication, and develops a commitment to sharing information and developing data standards.

About I-SPY Functions

Users can perform a variety of tasks in I-SPY. *Table 1.1* describes each I-SPY task.

Task	Description
Perform a Patient or Sample Identifier Lookup	Search the database for patient or sample identifiers. Display, download, and save the data associated with the search criteria. See *Conducting an Identifier Lookup*.
Perform a Search	Perform one of the following types of queries: • Clinical query • *IHC* Level of Expression query • IHC Loss of Expression query See *Conducting Searches*.
Perform a High Order Analysis	Run the following types of *higher order analyses*: • Class comparisons • Hierarchical clustering • Principal component analyses • Correlation scatter plot • Categorical plot analysis See *High Order Analysis*.
View Results	View Search and High Order Analysis results. See *Viewing Results*.
Manage Lists	Manage user-defined or study-defined patient or gene identifier lists. You can use them to filter queries or perform analysis. See *Managing Lists*.

Table 1.1 I-SPY user tasks

CHAPTER

2

GETTING STARTED WITH I-SPY

This chapter introduces general I-SPY procedures and how to obtain help to use I-SPY.

Topics in this chapter include:

Launching I-SPY

To launch I-SPY, follow these steps:

1. Go to the I-SPY portal on the NCICB website:
   http://ispy-analysis-stage.nci.nih.gov

The I-SPY login page appears *(Figure 2.1).*

Figure 2.1 I-SPY login page

Creating a User Account

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

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

OR

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

Logging In

To log into I-SPY, you need the username and password assigned to you by the I-SPY Administrator.

1. On the login page, enter your **username** and **password**.

Figure 2.2 I-SPY login

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

Accepting I-SPY Provisions

Once you log in, the Legal Rules of the Road page appears. After reading the provisions, click the **CLICKING HERE** link *(Figure 2.3)* in the lower right-hand corner.

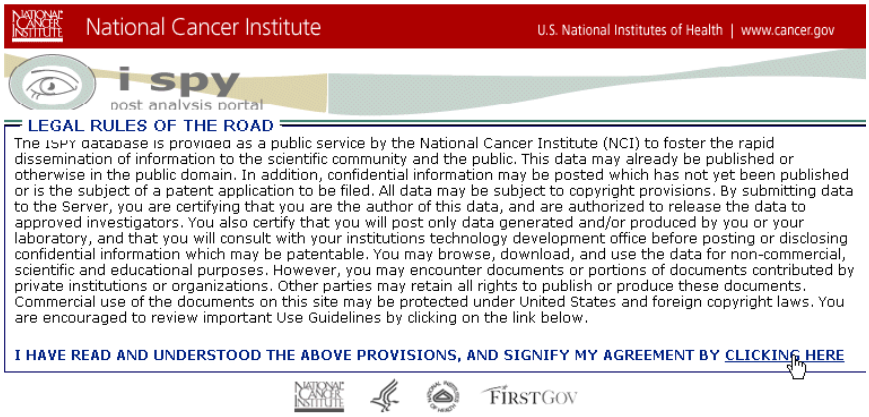


Figure 2.3 Legal Rules of the Road page

The I-SPY workspace appears (*Figure 2.4*).

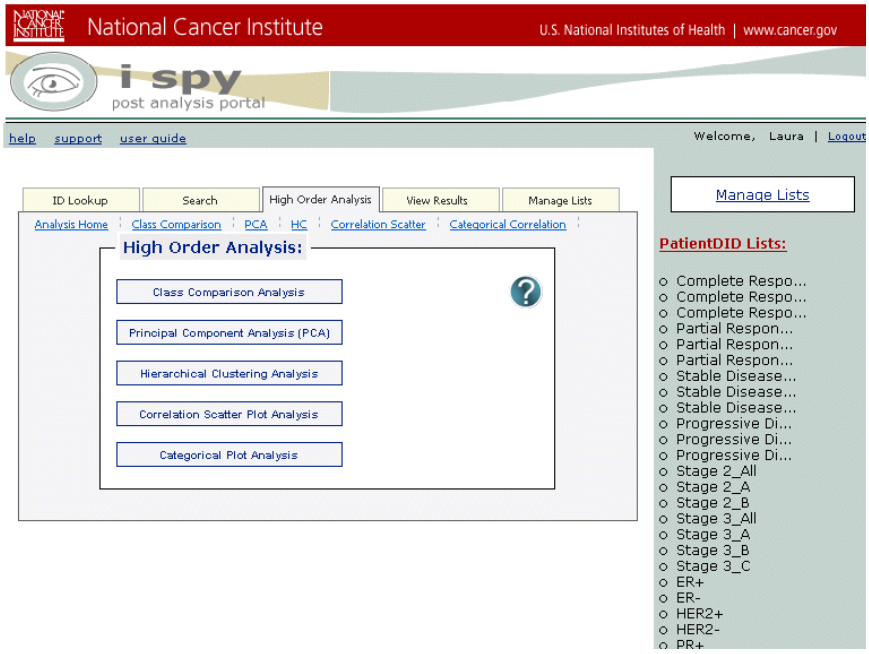


Figure 2.4 The I-SPY workspace

Welcome to I-SPY

The I-SPY workspace comprises a set of five tabs, a side, help links, and a logout link. The five tabs enable you to perform the following functions:

1. Perform an ID lookup.
2. Perform complex searches.
3. Perform higher order analyses.
4. View results of the searches and analyses.
5. Manage lists.

The side displays pre-defined patient identifier lists and the default gene identifier list. As you add your own lists, the new lists will appear in red.

Getting Help

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

Figure 2.5 I-SPY's menu

Table 2.1 describes each item on the I-SPY menu.

Help	***How to Access***
Complete online help	To access the complete version of online I-SPY help, click the **help** link located in the I-SPY menu. For complete page-level help, click on any I-SPY page. For brief field help, click [?].
Application support	To obtain support for I-SPY, click the **support** link located under the I-SPY menu.
User's Guide	To access a pdf version of the *I-SPY User's Guide*, click the **user guide** link located under the I-SPY menu.
Download Integrate Data File	To download the Integrated Data File, click the **integrated data file** link. The file is an Excel spreadsheet comprising all of the clinical data for I-SPY integrated with the MRI data. The file also containes all of the patient to sample ID mappings for each data type.

Table 2.1 Getting help with I-SPY

Application Support

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

Click the **support** link in the I-SPY menu. The NCICB Applications Support Group page appears.

Logging Out

To log out of I-SPY, follow these steps.

1. On the I-SPY workspace, click the **logout** link (*Figure 2.6*) in the upper right-hand corner.

Figure 2.6 Logout link

The I-SPY login page appears.

CHAPTER

3

CONDUCTING AN IDENTIFIER LOOKUP

This chapter describes how to use I-SPY to look up patient identifiers or sample identifiers.

Topics in this chapter include:

ID Lookup Overview

The ID Lookup function enables you to find information about samples for a given patient by entering either sample or patient identifiers. Once you perform the lookup, you can also perform the following tasks:

- *Displaying/Hiding the Patient Sample Information*
- *Downloading Patient Sample Information to an Excel File*
- *Creating a PatientDID List with the ID Lookup*

Note: You can use a *PatientDID list* to filter I-SPY queries and perform data analysis.

Looking Up a Patient or Sample Identifier

When you search for a patient identifier, I-SPY displays the patient along with all the samples associated with the patient. If you search for a sample identifier, I-SPY displays the patient associated with the sample identifier. To perform an ID lookup, follow these steps:

1. From the ID Lookup page (*Figure 3.1*), enter a valid patient identifier, such as `1001`, or enter a valid sample identifier, such as `209512`.

   Note: To enter multiple identifiers, separate the identifiers with commas. For example, enter `1001,1002`.

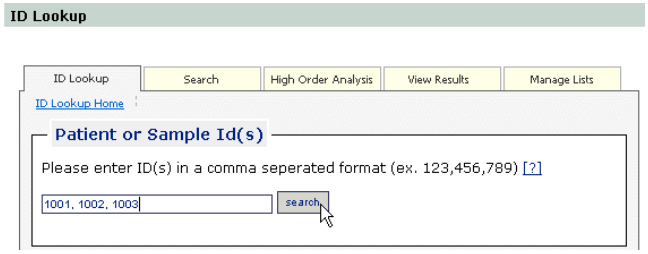


Figure 3.1 Entering identifiers

2. Click the **Search** button.

   The patients associated with the identifier(s) appear below the **Search** button (*Figure 3.2*).

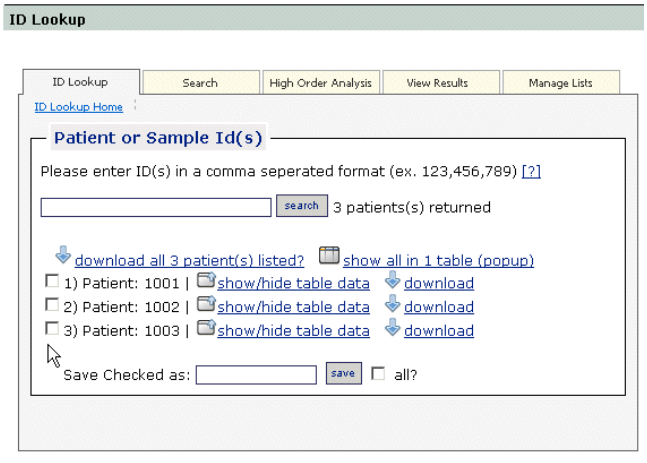


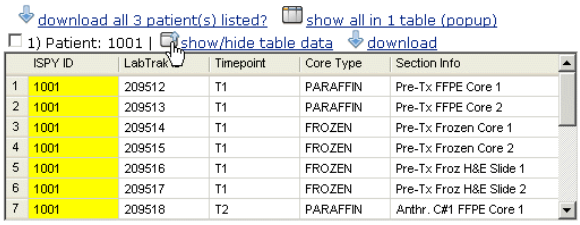
Figure 3.2 Found patients

Displaying/Hiding the Patient Sample Information

Once you perform an ID Lookup, you can display sample information for either an individual patient or multiple patients. Follow these steps:

1. To display all the samples collected for an individual patient click next to the patient's row.

   The table highlights the lookup criteria (*Figure 3.3*).



	ISPY ID	LabTrak	Timepoint	Core Type	Section Info
1	1001	209512	T1	PARAFFIN	Pre-Tx FFPE Core 1
2	1001	209513	T1	PARAFFIN	Pre-Tx FFPE Core 2
3	1001	209514	T1	FROZEN	Pre-Tx Frozen Core 1
4	1001	209515	T1	FROZEN	Pre-Tx Frozen Core 2
5	1001	209516	T1	FROZEN	Pre-Tx Froz H&E Slide 1
6	1001	209517	T1	FROZEN	Pre-Tx Froz H&E Slide 2
7	1001	209518	T2	PARAFFIN	Anthr. C#1 FFPE Core 1

Figure 3.3 Patient table data

Table 3.1 describes the sample information associated with the patient.

Item	*Special Instructions*
ISPY ID	The identifier for the patient.
LabTrak ID	The identifier for the sample collected.
Timepoint	The timepoint when the sample was collected. • **T1** • **T2** (for samples): 24 to 96 hours after the first cycle. • **T2** (for MRI): Any time between day 14 of cycle 1 and day 1 of cycle 2. • **T3**: Inter-regimen • **T4**: Prior to surgery for response evaluation forma and sample post-surgery
Core Type	The type of substance used to store the sample.
Section Info	

Table 3.1 Understanding the patient table data page

Note: To hide the table, data click .

2. To display data for multiple patients in one table click above the list of patients.

   All the patients' data are shown in one table listed in descending order by patient identifier.

   Note: To hide the table data, close the window.

Downloading Patient Sample Information to an Excel File

From the ID Lookup page, you can download one patient's sample data to a file or download all the listed patients' data to the same file. Follow these steps.

To download an individual patient's data to an Excel file, follow these steps:

1. Click next to the patient for which you want to download data.
2. Name the file and select a location.

   The individual patient's data is saved to the Excel file.

If you searched for *multiple* patients, to save all the patients' data to the same Excel file, follow these steps.

1. Click above the list of patients to download all the patient sample information to one file.
2. Name the file and select a location.

   All the patients' data is saved to the same Excel file listed in descending order by patient identifier.

Creating a PatientDID List with the ID Lookup

From the ID Lookup page, you can save multiple patients' data to a *PatientDID list*. You can use PatientDID lists to further filter a query or analyze data. To create an I-SPY PatientDID list, follow these steps.

1. Select the box next to each patient to be saved to the PatientDID list (*Figure 3.4*) or select the **All** box to select all of the patients.

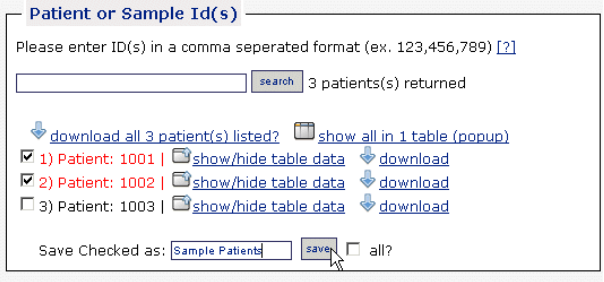


Figure 3.4 Saving to a PatientDID list

2. Name the list.
3. Click the **save** button.

   The list name appears in red in the side bar at the bottom of the PatientDID Lists.

   Note: To further modify the new PatientDID list, see *Managing Lists*.

CHAPTER

4

CONDUCTING SEARCHES

This chapter describes how to perform queries to generate results, such as reports and plots.

Topics in this chapter include:

Search Overview

The Search function enables you to perform queries from the following categories:

- *Performing a Clinical Query*
- *Performing an IHC Level of Expression Query*
- *Performing an IHC Loss of Expression Query*

Report results are listed on the View Results page.

Performing a Clinical Query

A *clinical query* enables you to generate clinical reports using customized search criteria. The search criteria filter the report based on clinical, MR, or pathology parameters. For example, you can create a clinical query that finds patients between the ages of 31 and 50 and had a complete response within timepoints T1 and T2.

To define a clinical query, follow these steps:

1. On the Clinical Query Form page, you are required to fill in at least one search criteria (*Figure 4.1*).

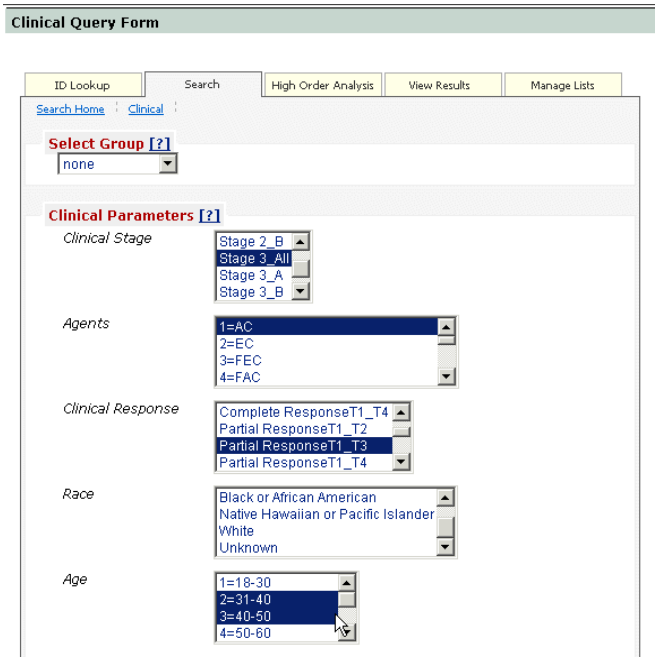


Figure 4.1 Clinical Query Form page (top portion)

2. *Table 4.1* lists the available search criteria:

Note: To select more than one option in a list box, SHIFT-click or CTRL-click.

Criteria	*Item Name*	*Special Instructions*
Select Group	Select Group	Select a group to filter the query to a collection of patients. Lists that you created appear in red.
Clinical Parameters	Clinical Stage	Select a clinical stage to further filter the query: • **Stage 2_All** • **Stage 2_A** • **Stage 2_B** • **Stage 3_All** • **Stage 3_A** • **Stage 3_B** • **Inflammatory**
	Agents	Select an agent to further filter the query.

Table 4.1 Clinical Query search criteria instructions

Criteria	*Item Name*	*Special Instructions*
	Clinical Response	Select one or more clinical responses and the appropriate timepoint range. ● **Complete Response** ● **Partial Reponse** ● **Stable Disease** ● **Progressive Disease**
	Race	Select one or more races.
	Age	Select one or more age ranges.
MR Parameters	Morphology	Select an MRI parameter to further filter the query based on the radiologist measurement.
	Percent LD	This group of options enables you to specify the percentage of LD (Longest Diameter) change in the size of the tumor between two timepoints. ● Select the timepoint range in which to analyze the percentage of LD change. For example, **PERCENT_LD_CHANGE_T1_T2** queries the percentage of change in LD between T1 Pre-treatment and T2 Early Treatment. ● Select the greater than/equal to (**>=**) or the less than/equal to (**<=**) option. ● Enter the percentage of LD change to search for in the selected timepoint range.
Pathology	Pathology Tumor Size	Specify the tumor size and associated biomarkers to filter the query. Select the greater than/equal to (**>=**) or less than/equal to (**<=**) indicator and enter a value in centimeters of the tumor size.
	Status	Specify the Pathology Status: **ER+**: Estrogen receptor positive **ER-**: Estrogen receptor negative **PR+**: Progesterone receptor positive **PR-**: Progesterone receptor positive **HER2+**: HER2 positive **HER2-**: HER2 negative

Table 4.1 Clinical Query search criteria instructions

3. Once you fill in at least one search criteria, you are required to enter a name for the clinical query. The name must be unique among all the queries in the current session.

   To clear all the entries on the page, click the **Clear** button.

4. To submit the query and generate the Clinical report, click the **Submit** button.

The Clinical report will be listed on the View Results page. From the report, you can save all or some of the patients' data to an I-SPY PatientDID list. You can use PatientDID lists to further filter a query or perform analysis. See *Clinical Reports*.

Performing an IHC Level of Expression Query

An *IHC Level of Expression* query enables you to filter a search with one or more timepoints, biomarkers, and stain characteristics. The report results list records that satisfy the specified search criteria.

To perform a *IHC* Level of Expression query, follow these steps:

1. On the IHC Level of Expression Query Form page, the following search criteria are available to filter the query (*Figure 4.2*).

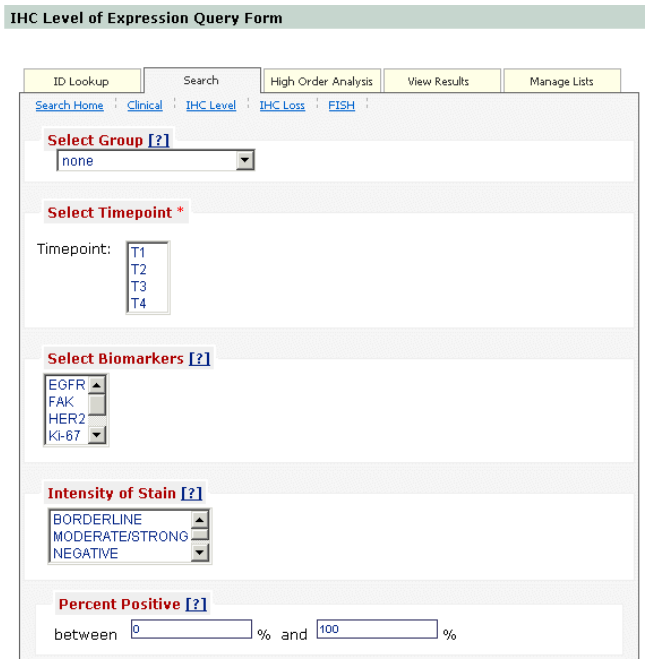


Figure 4.2 IHC Level of Expression Query Form page (top portion)

Table 4.2 lists the available search criteria:

Note: To select more than one option in a list box, SHIFT-click or CTRL-click.

Criteria	*Special Instructions*
Select Group	Select a *group* to filter the query to a collection of patients. Lists that you created appear in red.
Select Timepoint	You are required to specify at least one timepoint during which the selected criteria are fulfilled. **T1**, **T2**, **T3**, **T4**
Select Biomarkers	Select one or more biomarkers to filter the query: • **P27** • **Ki-67** • **EGFR** • **CCND1** • **P53** • **HER2** • **BCL2** • **FAK**
Intensity of Stain	Select an option that best describes the intensity of stain: • **Negative** • **Borderline** • **Weak** • **Moderate_Strong** • **Unevaluable**
Percent Positive	Enter the percent positive range to filter the query.
Localization of Stain	Select an option that best describes the localization of stain: • **None** • **Membrane** • **Nucleus** • **Cytoplasm** • **Membrane_and_Cytoplasm** • **Nuclear_and_Cytoplasm** • **NA** or Not Applicable
Distribution of Stain	Select an option that best describes the distribution of stain: • **None** • **Homogenous** • **Heterogenous**

Table 4.2 IHC Level of Expression Query search criteria instructions

2. Once you fill in the search criteria, you are required to enter a name for the IHC query. The name must be unique among all the queries in the current session.

To clear all the entries on the page, click the **Clear** button.

3. To submit the query and generate the IHC Level of Expression report, click the **Submit** button.

   The IHC Level of Expression report will be listed on the View Results page. From the report, you can save all or some of the patients' data to an I-SPY PatientDID list. You can use a PatientDID list to further filter a query or perform analysis. For more information, see *IHC Level of Expression Search Results*.

Performing an IHC Loss of Expression Query

An *IHC Loss of Expression* query enables you to filter a search with one or more timepoints, the P27 biomarker only, and invasive and benign range characteristics. The report results list records that satisfy the specified search criteria.

To perform a *IHC* Loss of Expression query, follow these steps:

1. On the IHC Loss of Expression Query Form page, the following search criteria are available to filter the query (*Figure 4.2*).

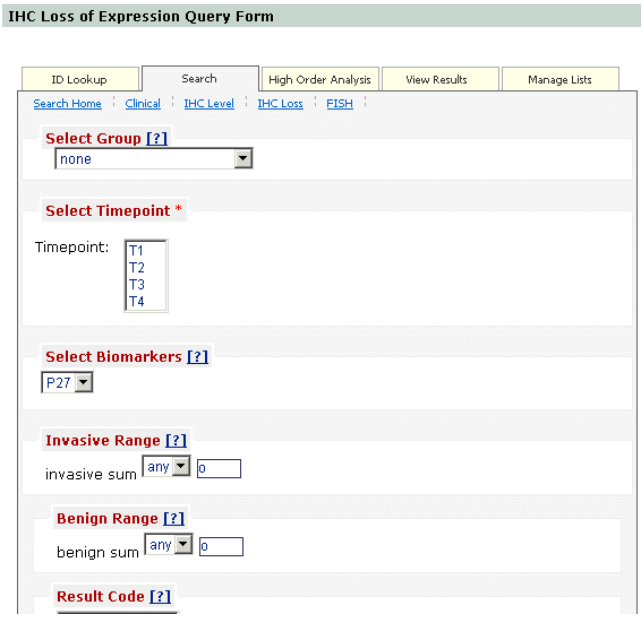


Figure 4.3 IHC Loss of Expression Query Form page (top portion)

Table 4.2 lists the available search criteria:

Note: To select more than one option in a list box, SHIFT-click or CTRL-click.

Criteria	*Special Instructions*
Select Group	Select a *group* to filter the query to a collection of patients. Lists that you created appear in red.
Select Timepoint	You are required to specify at least one timepoint during which the selected criteria are fulfilled. **T1**, **T2**, **T3**, **T4**
Select Biomarkers	IHC Loss of Expression data is available only for biomarker **P27**.
Invasive Range	Specify equal to, greater than, or less than to define the range invasive sum: • = • >= • <= Specify the value for the invasive sum.
Benign Range	Specify equal to, greater than, or less than to define the range benign sum: • = • >= • <= Specify the value for the benign sum.

Table 4.2 IHC Loss of Expression Query search criteria instructions

2. Once you fill in the search criteria, you are required to enter a name for the IHC query. The name must be unique among all the queries in the current session.

   To clear all the entries on the page, click the **Clear** button.

3. To submit the query and generate the IHC Loss of Expression report, click the **Submit** button.

   The IHC Loss of Expression report will be listed on the View Results page. From the report, you can save all or some of the patients' data to an I-SPY PatientDID list. You can use a PatientDID list to further filter a query or perform analysis. For more information, see *IHC Loss of Expression Search Results*.

CHAPTER

5

HIGH ORDER ANALYSIS

This chapter describes how to use I-SPY to run higher order analyses.

Topics in this chapter include:

High Order Analysis Overview

The *High Order Analysis* function enables you to perform the following analyses:

- *Performing a Class Comparison*
- *Performing a Principal Component Analysis*
- *Performing Hierarchical Clustering Analysis*
- *Performing Correlation Scatter Plot Analysis*
- *Performing Categorical Plot Analysis*

Report results are listed on the View Results page.

Performing a Class Comparison

A *Class Comparison analysis* alows you to identify genes and reporters that are differentially expressed between two groups. To perform a *Class Comparison*, follow these steps:

1. The Class Comparison Analysis Form page *(Figure 5.1)* enables you to define the criteria to perform a class comparison.

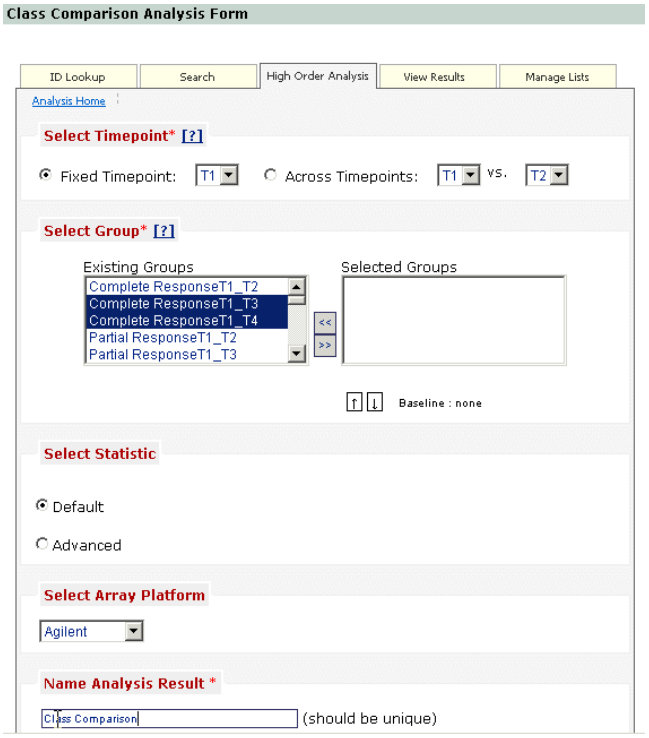


Figure 5.1 Class Comparison Analysis Form page

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

Criteria	*Item Name*	*Special Instructions*
Select Timepoint	Fixed Timepoint	Select a timepoint in which to perform the analysis. This option compares *two groups* at the same timepoint.
	Across Timepoints	Select a range of timepoints in which to perform the analysis. This option analyzes *one group* at different timepoints.

Table 5.1 Class Comparison criteria instructions

Criteria	*Item Name*	*Special Instructions*
Select Group	Existing Groups Selected Groups	Select a *group* to filter the query to a collection of patients. Lists that you created appear in red. For a *Fixed Timepoint* analysis, select two groups (compares two groups at the same timepoint). For an *Across Timepoints* analysis, select one group (analyzes one group at different timepoints).
	Baseline	For a *Fixed Timepoint* analysis, the baseline is determined by the second group in the **Selected Groups** box. For an *Across Timepoints* analysis, the baseline is determined by the first timepoint in the chosen range. The **(baseline)** appears in red next to your selection.
Select Statistic	Default	Select to perform a default statistical analysis.
	Advanced	Select to define additional statistical analysis options.
	• Statistical Method	Select the appropriate statistical method: • **T-test: Two Sample Test** to identify genes showing statistically significant differences between two samples. • **Wilcoxon Test: Man-Whitney Test** is the non-parametric test analog to the independent two-sample t-test. This test is used in place of a two-sample t-test when the populations being compared are not normal.
	• Multiple Comparison Adjustment	*Family-wise Error Rate (FWER)*: Bonferroni *False Discover Rate (FDR)*: Benjamini-Hochberg
	• Fold Change	The default is >=2. Specify the threshold for the differential regulation. This returns differential expression ratios between tumor and non-tumor samples for a particular reporter.
	• p-value	The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples. The default is <=0.05.
Select Array Platform	Select Array Platform	Select the array platform.

Table 5.1 Class Comparison criteria instructions

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

Performing a Principal Component Analysis

A *Principal Component analysis* is a dimensionality reduction algorithm, which identifies clusters of samples that may have similar gene expression profiles. To perform a *Principal Component analysis*, follow these steps:

1. The Principal Component Analysis (PCA) Form page *(Figure 5.2)* enables you to define criteria to perform a PCA.

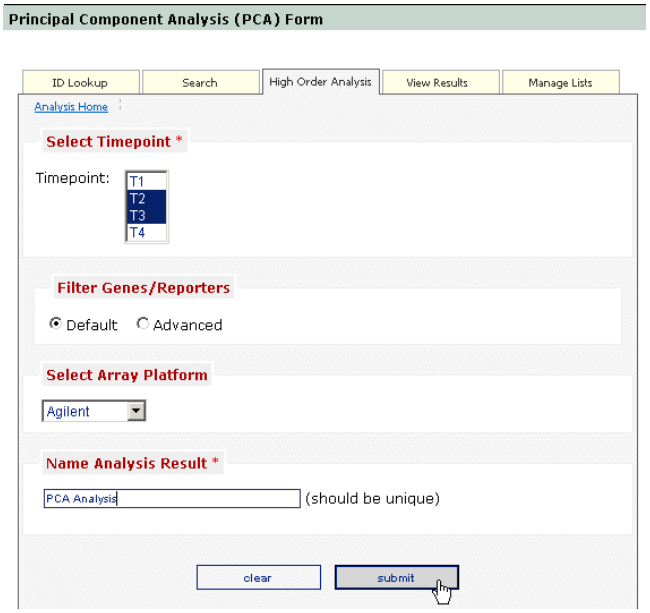


Figure 5.2 Selecting Principal Component Analysis criteria

2. You are required to complete at least one criteria for the Principal Component analysis. *Table 5.2* lists the available criteria:

Criteria	*Item Name*	*Special Instructions*
Select Timepoint	Timepoint	Select one or more timepoints in which to perform the analysis.
Filter Genes/Reporters	Default	Select to perform a default statistical analysis.
	Advanced	Select to define additional gene/reporter filters.
	• Constrain reporters by variance (Gene Vector) percentile: %	Enter a percentage which selects the reporters whose variances of the log ratio (or Log2 signals) across all experiments were among the top percentile of variance of all reports identified. For example, 70% chooses reporters with the top 30 (100 - 70) percentile of variance.
	• Constrain by GeneList	Select a *Gene list* to filter the query. Lists that you created appear in red. The default gene list is **defaultGene1**.
Select Array Platform	Select Array Platform	Select an array platform.

Table 5.2 Principal Comparison Analysis criteria instructions

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

Performing Hierarchical Clustering Analysis

Hierarchical Clustering analysis creates a dendrogram of the samples in the analysis. To perform a *Hierarchical Clustering*, follow these steps:

1. The Hierarchical Clustering Analysis Form *(Figure 5.3)* enables you to fill in criteria for a hierarchical clustering.

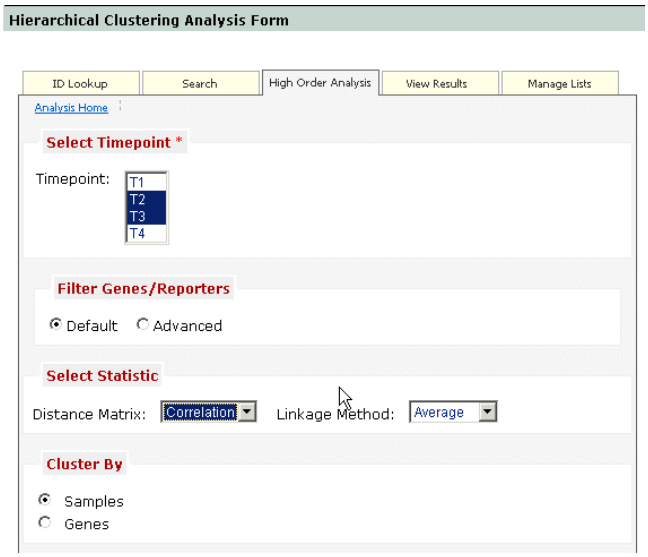


Figure 5.3 Selecting Hierarchical Clustering criteria

2. You are required to enter at least one step for the hierarchical clustering. *Table 5.3* lists the available criteria:

Criteria	*Item Name*	*Special Instructions*
Filter Genes/Reporters	Default	Select to perform a default statistical analysis.
	Advanced	Click to define additional gene/reporter filters.
	• Constrain reporters by variance (Gene Vector) percentile: %	Enter a percentage which selects the reporters whose variances of the log ratio (or Log2 signals) across all experiments were among the top percentile of variance of all reports identified. For example, 70% chooses reporters with the top 30 (100 - 70) percentile of variance.
	• Constrain by GeneList	Select a *Gene list* to filter the query. Lists that you created appear in red. The default gene list is **defaultGene1**.

Table 5.3 Hierarchical Clustering criteria instructions

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

Table 5.3 Hierarchical Clustering criteria instructions

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

Performing Correlation Scatter Plot Analysis

A *Correlation Scatter Plot analysis* enables you to select two continuous variables and plot them against each other. The variables can be gene expression values or a clinical parameter like MRI percent longest diameter change.

The following are examples of how you can use a Correlation Scatter Plot Analysis.

- **Cross platform validation**: Select the same gene on two different platforms and display the correlation between the expression values.
- **Interreporter validation**: Select the same gene (but different reporters) on the same platform.
- **Gene expression correlation**: Investigate the relationships between gene expression values for two different genes.
- **Clinical parameter and gene expression relationship**: Investigate the relationship between a clinical parameter and the gene expression values for a given gene.

To perform a Correlation Scatter Plot analysis, follow these steps:

1. The Correlation Scatter Analysis Form *(Figure 5.3)* enables you to fill in criteria for generate a correlation scatter plot.

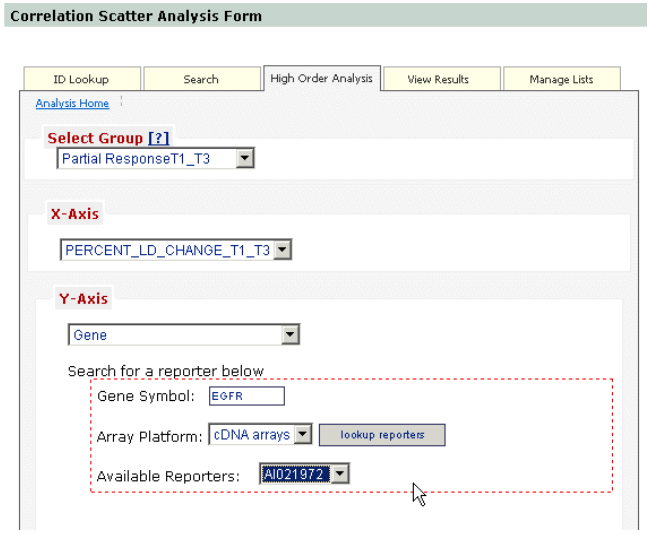


Figure 5.4 Selecting Correlation Scatter Plot criteria

2. You are required to enter at least one step for the correlation scatter plot. *Table 5.4* lists the available criteria:

Criteria	***Special Instructions***
Select Group	Select a *group* to filter the query to a collection of patients. Lists that you created appear in red.
X-Axis	Enter gene information or select a timepoint range in which to analyze the percentage of LD (Longest Diameter) change.
• Gene	Select **Gene**. Note that to generate the plot, you must select a gene symbol for one axis. • Enter a gene symbol. • Select an array. • Click the **Lookup Properties** button. • Select a reporter.

Table 5.4 Correlation Scatter Plot criteria instructions

Criteria	*Special Instructions*
• PERCENT LD_CHANGE	Select an option to analyze the percentage of LD (Longest Diameter) change in the size of the tumor. • **PERCENT_LD_CHANGE**: Include all samples and compare against all Percent_LD_Change values. • **PERCENT_LD_CHANGE_T1_T2**: Display the percentage of LD change between timepoints T1 Pre-treatment and T2 Early Treatment. • **PERCENT_LD_CHANGE_T1_T3**: Display the percentage of LD change between timepoints T1 Pre-treatment and T3 Between Treatment Regimes. • **PERCENT_LD_CHANGE_T1_T4**: Display the percentage of LD change between timepoints T1 Pre-treatment and T4 Pre-surgery.
Y-Axis	Same as the X-axis options.
Correlation	Select a distance matrix option: **Pearson** correlation: Pearson's Correlation Coefficient measures the strength of the linear relationship between two variables. Assumptions are the following: linear relationship between two variables; continuous random variables; both variables must be normally distributed; and two variables must be independent of each other. **Spearman** correlation: Spearman's Rank Correlation Coefficient is a non-parametric measure of correlation. It assesses how well an arbitrary monotonic function could describe the relationship between two variables without making any assumptions about the frequency distribution of the variables. It does not require the assumption that the relationship between the variables is linear. It can be used for variables measured at the ordinal level.

Table 5.4 Correlation Scatter Plot criteria instructions

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

Performing Categorical Plot Analysis

The *Categorical Plot analysis* enables you to select one or more groups of patients as defined in the I-SPY Manage Lists function, and view a box-and-whisker plot of a continuous variable for the patients in the selected groups. These groups can be the pre-defined groups defined in I-SPY or groups that you create with the Manage Lists function. The following example describes how creating lists in the I-SPY Manage Lists function (see *Combining Existing Lists to Create a New List* on page 49) can generate categorical plots for specific needs.

- Using the I-SPY Manage Lists function, create two lists
  - A Triple Negative list combining the ER-, HER2-, PR- groups
  - A Triple Positive list combining the ER+, HER2+, PR+ groups
- Specify Categorical Plot criteria to compare the Percent Longest Diameter Change for timepoints T1 to T4 with values for patients in the Triple Positive group versus patients in the Triple Negative group.

Other general uses of box-and-whisker plots include the following:

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

To perform a Categorical Plot Analysis, follow these steps:

1. The Categorical Plot Analysis Form *(Figure 5.3)* enables you to fill in the criteria to generate a categorical plot.

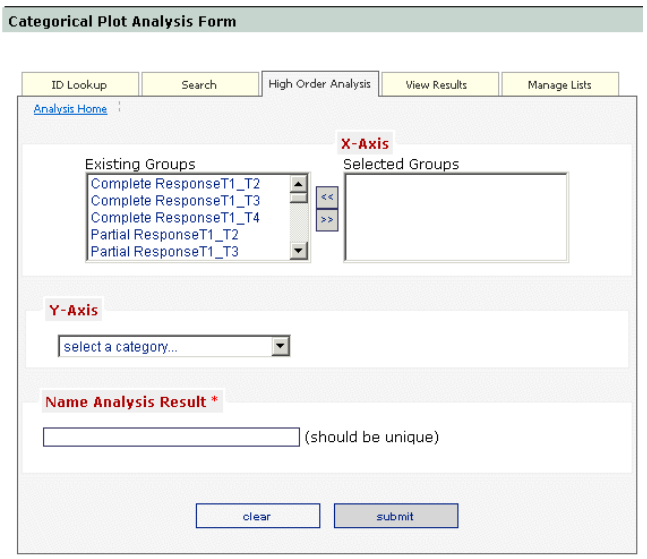


Figure 5.5 Selecting Categorical Plot Analysis criteria

2. You are required to complete an entry for the categorical plot. *Table 5.5* lists the available criteria:

Criteria	*Special Instructions*
Select Group (X-axis)	Select the *groups* for the X-axis to filter the query to a collection of patients. Lists that you created appear in red.
Y-axis	Enter gene information or select a timepoint range in which to analyze the percentage of LD (Longest Diameter) change.
• Gene	Select **Gene**. Note that to generate the plot, you must select a gene symbol for one axis. • Enter a gene symbol. • Select an array. • Click the **Lookup Properties** button. • Select a reporter.
• PERCENT LD_CHANGE	Select an option to analyze the percentage of LD (Longest Diameter) change in the size of the tumor. • **PERCENT_LD_CHANGE**: Include all samples and compare against all Percent_LD_Change values. • **PERCENT_LD_CHANGE_T1_T2**: Display the percentage of LD change between timepoints T1 Pre-treatment and T2 Early Treatment. • **PERCENT_LD_CHANGE_T1_T3**: Display the percentage of LD change between timepoints T1 Pre-treatment and T3 Between Treatment Regimes. • **PERCENT_LD_CHANGE_T1_T4**: Display the percentage of LD change between timepoints T1 Pre-treatment and T4 Pre-surgery.

Table 5.5 Correlation Plot Analysis criteria instructions

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

CHAPTER

6

VIEWING RESULTS

This chapter describes reports and search results that I-SPY returns after queries and and high order analyses.

Topics in this chapter include the following:

Results Overview

The View Results page shows a collection of reports previously viewed in a particular user session. This allows you to compare reports by opening them in separate windows. You can view results generated with the Search function and the High Order Analysis function.

Using I-SPY Report Icons

Table 6.1 describes icons that appear at the top of most I-SPY report results:

Icon	Special Instructions
	Show or hide the report tools that appear below the report title and above the report data.
	Displays the query details at the bottom of the report used to generate the data.
	Prints the report.
	Saves the report to a spreadsheet file.
	Displays online help for the current report along with a complete outline of the I-SPY online help.
	Closes the report window.

Table 6.1 I-SPY Report Icons

Search Results

The following results are generated from the Search function:

- *Clinical Reports*
- *IHC Level of Expression Search Results*
- *IHC Loss of Expression Search Results*

View Results (*Figure 6.1*) displays the query name and lists the output generated for the query.

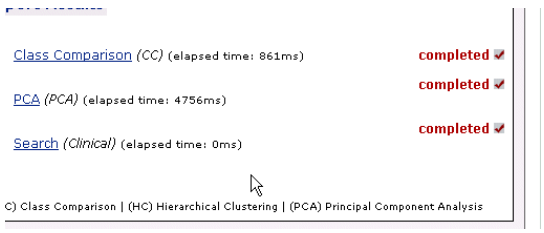


Figure 6.1 Search Results

Clinical Reports

A *Clinical report* displays the demographic, clinical, MR, and pathology data for a given set of patients (*Figure 6.2*). From a clinical report, you can create a *PatientDID list* to further filter queries. If there is an icon in the NCIA Image column of the report, you can also access images associated with a patient.

ISPY_ID	NCIA_IMAGE	INST_ID	AGE	AGECAT	RACE_ID
1199	-	939=U Penn	37.68	2= >30-40	White
1210	-	939=U Penn	40.58	3= >40-50	White
1075	-	508=U Ca at San Francisco	33.47	2= >30-40	White
1155	-	939=U Penn	42.65	3= >40-50	White
1125	-	2527=U Texas SW Med Ctr	44.85	3= >40-50	American Indian or Alaska Native
1056	-	508=U Ca at San Francisco	35.37	2= >30-40	White
1211	-	443=Chicago	48.11	3= >40-50	White
1085	-	540=Georgetown	43.93	3= >40-50	White
1206	-	530=Mem Sloan-Kettering Cancer Ctr	43.03	3= >40-50	White

Figure 6.2 Clinical Report

Table 6.2 refers you to the appropriate section of the data dictionary for descriptions of all the columns in the Clinical report:

Column Names	Cross Reference for Descriptions
ISPY_ID through SURVDTD	See *Patient Demographic Data Dictionary*
CHEMOCAT through HERCEPTIN	See *Chemotherapy Summary Data Dictionary*
DOSEDENSEANTHRA through HER2COMMUNITYMETHOD	See *On-Study Data Dictionary*
SURGERYLUMPECTOMY through REASON_NO_SURG	See *Post-Surgery Summary Data Dictionary*
RTTHERAPY through RTOTHER	See *Follow-Up Data Dictionary*
TSizeClinical through CLINRESPT1_T4	See *Response Evaluation Data Dictionary*
Morphological pattern at T1 through MR % change T3_T4	See *MR Data Dictionary*

Table 6.2 I-SPY Clinical report data items cross-references

Creating a PatientDID List from a Clinical Report

On any Clinical page, you can select and save patients to a *PatientDID list*.

1. There are two ways to select patients on the Clinical window:
   - To select an individual, select the box in the **I-SPY ID** column (*Figure 6.3*).

Figure 6.3 Checking the I-SPY ID column

   - To select all of the patients, select the **All** box (*Figure 6.4*).

Figure 6.4 Selecting all of the samples on the Clinical window

   To clear all of the patients, uncheck the **All** box.

2. To save the patients to a PatientDID list, enter a name for the list (*Figure 6.5*).

Figure 6.5 Saving Selected Samples on the Clinical page

3. Click the **save checked** button. Sample List Saved appears.
4. Click the **OK** button. The new PatientDID list is now displayed in red in the side bar at the bottom of the PatientDID names (*Figure 6.6*). Mouse over the name and the data items appear.

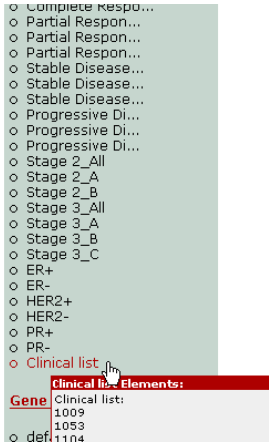


Figure 6.6 Saved PatientDID list

Accessing an NCIA Image from a Clinical Report

To access an NCIA image from a clinical report, follow these steps:

Click the NCIA icon  in the **NCIA Image** column (*Figure 6.7*).

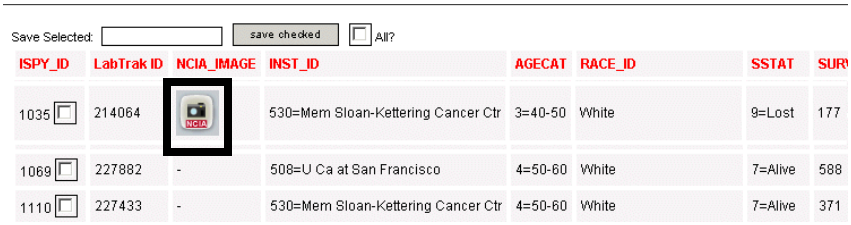


Figure 6.7 Clinical Report with NCIA Image icon

The *NCIA* web site appears. As a first time user, register with NCIA to obtain a username and password. Then you can log in and have access to these images.

IHC Level of Expression Search Results

An *IHC Level of Expression report* displays the patients selected from the IHC Level of Expression search (*Figure 6.3*). On the IHC Level of Expression page, you can select and save patients to a *PatientDID list* to further filter queries. You can also sort the patients using the red triangles in the column name.

Figure 6.8 IHC Level of Expression Report

IHC Loss of Expression Search Results

An *IHC Loss of Expression report* displays the patients selected from the IHC Loss of Expression search (*Figure 6.3*). On the IHC Level of Expression page, you can select and save patients to a *PatientDID list* to further filter queries. You can also sort the patients using the red triangles in the column name.

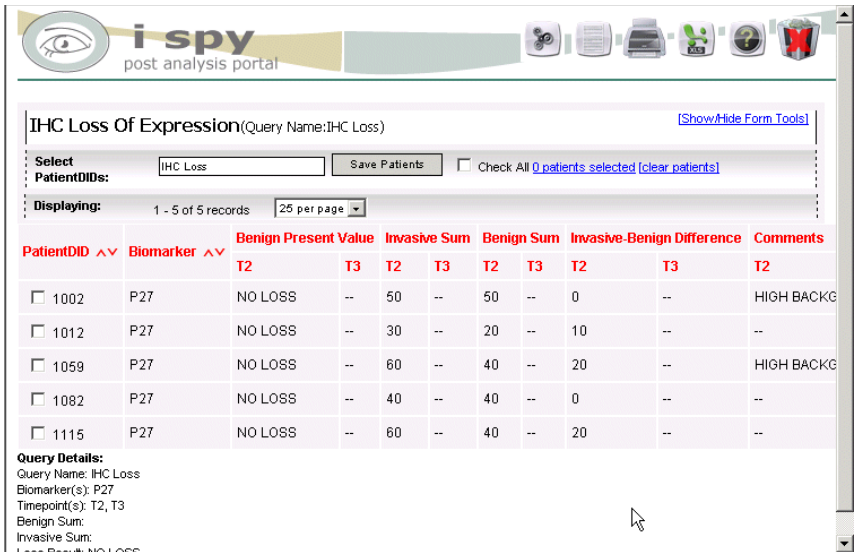


Figure 6.9 IHC Loss of Expression Report

High Order Analysis Results

The following reports are generated from the High Order Analysis function:

- *Class Comparison Report*
- *Principal Component Analysis Plot*
- *Hierarchical Clustering Report*
- *Correlation Scatter Plot*
- *Categorical Plot Analysis*

View Results (*Figure 6.10*) displays the query name and lists the output generated for the query.

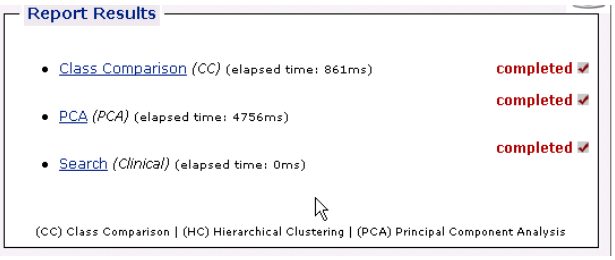


Figure 6.10 HOA Results

Class Comparison Report

The *Class Comparison report* (*Figure 6.11*) displays group average, fold change, and *p*-value based on the search parameters that you selected. For a **T-test** or **Wilcoxon** Statistical Method analysis (*Figure 6.11*), the Class Comparison report is as follows.

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

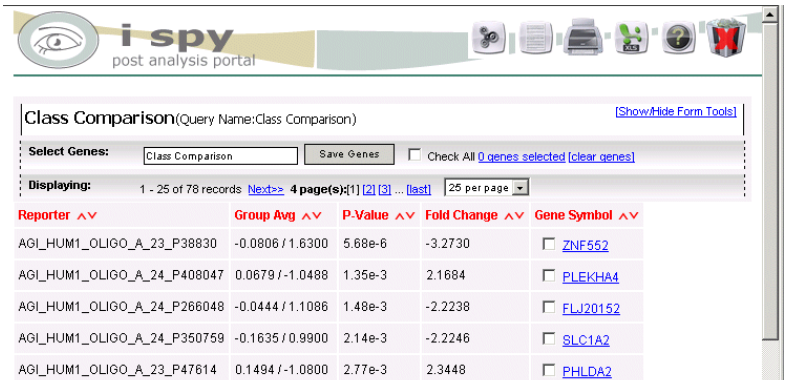


Figure 6.11 Class Comparison Report

Creating a Gene List (Select Genes toolbar)

On the Class Comparison page, you can select and save the genes to a *Gene list*. To create a Gene list, follow these steps (*Figure 6.12*):

1. To select all of the genes in result list, click the **Check All** box.
2. To select some of the genes, check the box in the Gene Symbol column.

   Note: To clear the selected genes, click the **clear genes** link.
3. To save the selected genes, enter a unique name for the file next to **Select Genes**, or maintain the current name, which varies based on the type of Statistical Method selected for the analysis.

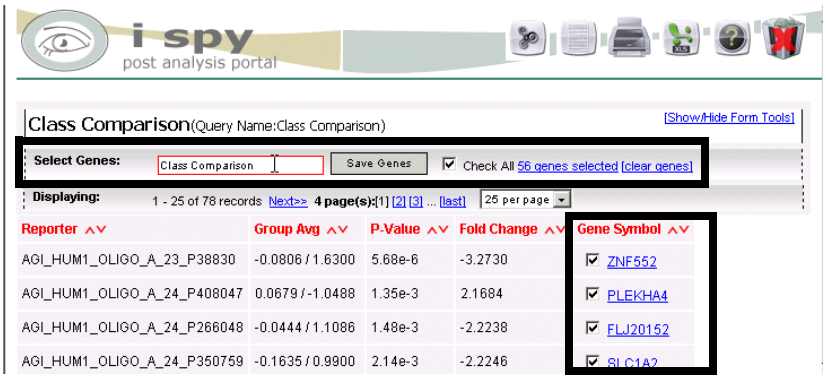


Figure 6.12 Selecting Genes instructions

4. Click the **Save Genes** button.

   The results are saved.

5. Click the **OK** button.

   The new Gene list appears in red in the side bar at the bottom of the Gene Lists names. Mouse over the name and the data items appear.

Resorting Column Results

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

1. If a report column has red triangles pointing up and down next to the name, you can sort a column of numeric or alphabetical values (*Figure 6.13*).

Figure 6.13 Sorting column results

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

Showing Additional Information

When results are listed in a report, row or column items may appear as links. Click the link to display additional information about the item.

For example, to display more information about a gene, click the name link (*Figure 6.14*).

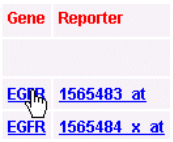


Figure 6.14 The Gene column

The Cancer Genome Anatomy Project (CGAP) browser opens.

Principal Component Analysis Plot

The *Principal Component Analysis plot* (*Figure 6.15*) is a two-dimensional graph which plots the various principal components from the analyses.

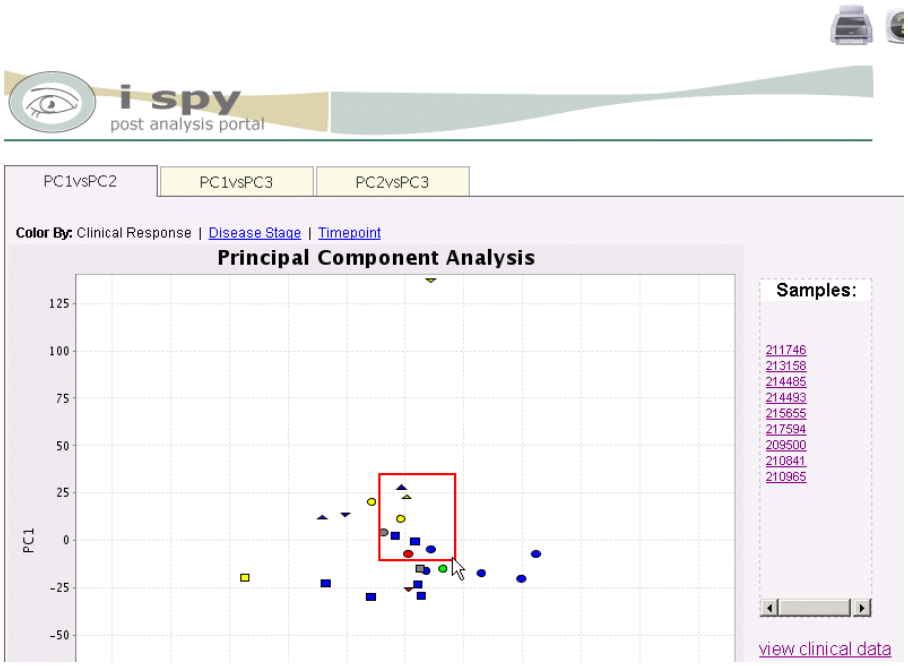


Figure 6.15 Principal Component Analysis Plot

Table 6.3 describes other areas in the plot:

Area	Description
Tabs	You can click on the three tabs at the top of the graph to display the following: • PC1vsPC2 • PC1vsPC3 • PC2vsPC3
Color By	Each point on the graph represents a sample, and by default, the samples are colored by **Clinical Response**. To color by **Disease stage** or **Timepoint**, click the appropriate link.
Legend	At the bottom of the graph, a legend defines how the different shapes in the graph indicate different survival ranges for patients.
Samples	The **Samples** area enables you to select, review, and display clinical data for samples in the plot (see *Selecting Samples of Interest in a Plot*).

Table 6.3 Areas of the Principal Component Analysis Plot

Selecting Samples of Interest in a Plot

To select the samples of interest in an I-SPY plot, follow these steps:

1. Click and drag a rectangle around the samples.

   A red rectangle appears around the samples, and the list of the samples appears on the right-hand side (*Figure 6.16*).

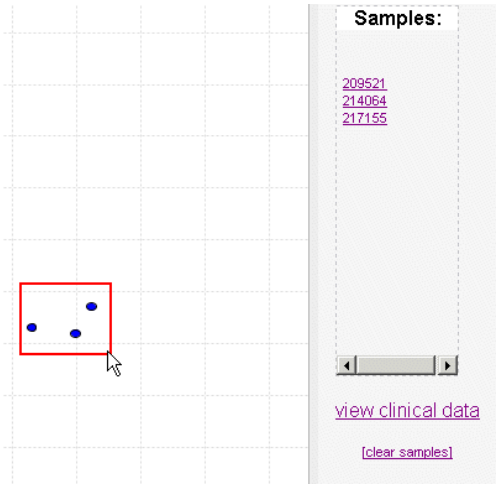


Figure 6.16 Sampling from a clinical plot

2. To help lasso the points on the plot and identify the location of these points, mouse over a sample name in the list.

A yellow circle appears on the plot where the sample is located (*Figure 6.17*).

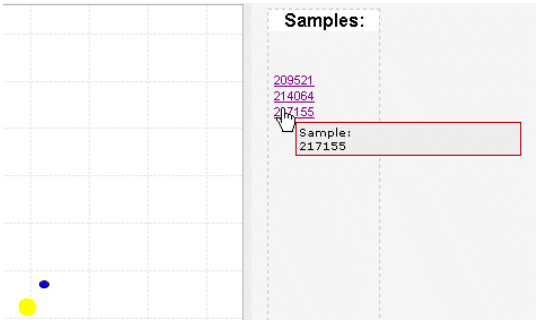


Figure 6.17 Lasso the points

3. To generate clinical data for the selected samples and save the samples, click the **view clinical data** link. See *Clinical Reports*.

   To select another group of samples, click the **clear samples** link and start again.

Hierarchical Clustering Report

The *Hierarchical Clustering report* (*Figure 6.18*) displays the dendrogram from the hierarchical clustering analysis and a clinical report. The dendrogram is organized based on the gene expression profiles of the samples. Samples with similar profiles are placed closer together on the tree. To adjust the size of the graph, move the box on the **Image Control** bar in the top lefthand corner.

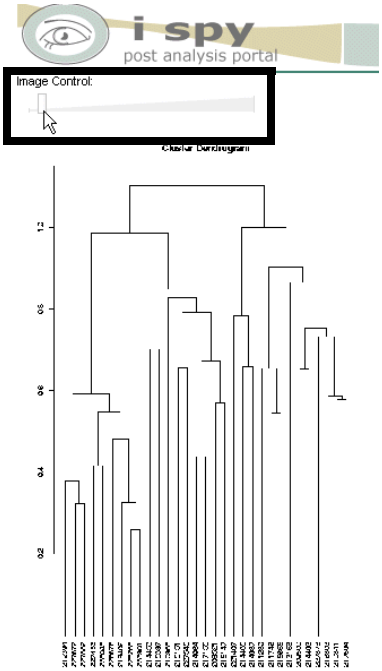


Figure 6.18 Hierarchical Clustering Dendrogram

A clinical report appears beneath the dendrogram where you can create a *PatientID list* or display an image associated with a patient. For more information, see *Clinical Reports*.

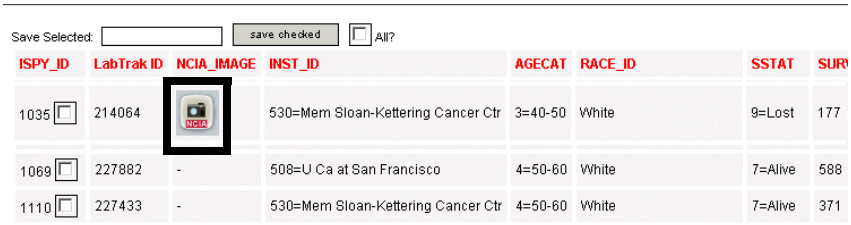


Figure 6.19 Hierarchical Clustering Clinical Report

Correlation Scatter Plot

The *Correlation Scatter plot* (*Figure 6.20*) is a visualization used to compare two continuous variables. The X-axis represents the values for one of the variables and the Y-axis represents the values for the other variable.

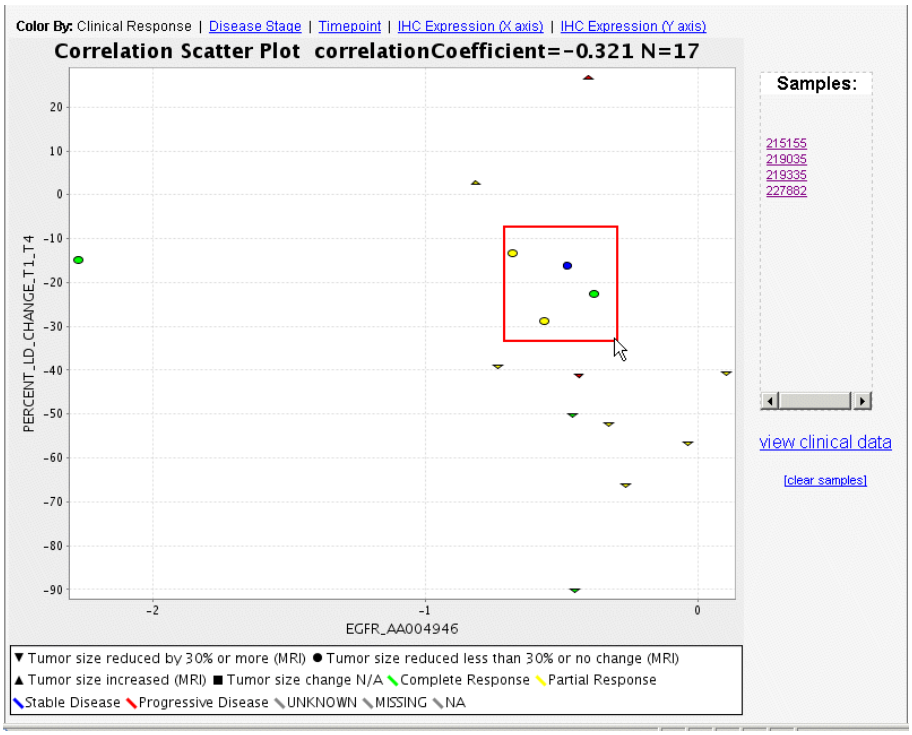


Figure 6.20 Correlation Scatter Plot

Table 6.4 describes other areas in the plot:

Area	Description
Color By	Each point on the graph represents a sample, and by default, the samples are colored by **Clinical Response**. To color by **Disease stage** or **Timepoint**, click the appropriate link. To plot IHC expression on the X- or Y-axis, click the **IHC Expression X axis** or **IHC Expression Y axis** link. For example, selecting color by **IHC Expression (X Axis)**, colors the points on the plot based on the IHC Expression of the gene on the X-axis. Selecting color by **IHC Expression (Y Axis)** colors the points on the plot based on the IHC Expression of the gene on the Y-axis.
Correlation Coefficient	Computed and displayed in the title. Correlation coeffients with values close to 1 are highly correlated. Values close to -1 indicate an inverse relationship. Values close to 0 indicate no correlation between the parameters.
N=	The number of samples in the plot.
Legend	At the bottom of the graph, a legend defines how the different shapes on the graph indicate different survival ranges for patients.
Samples	The **Samples** area enables you to select, review, and display clinical data for samples in the plot (see *Selecting Samples of Interest in a Plot*).

Table 6.4 Areas of the Correlation Scatter Plot

Categorical Plot Analysis

The *Categorical Plot analysis* (*Figure 6.21*) displays a box-and-whisker plot of a continuous variable for patients in selected groups.

The following items in the graph indicate the following:

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

Example uses of box-and-whisker plots include the following:

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

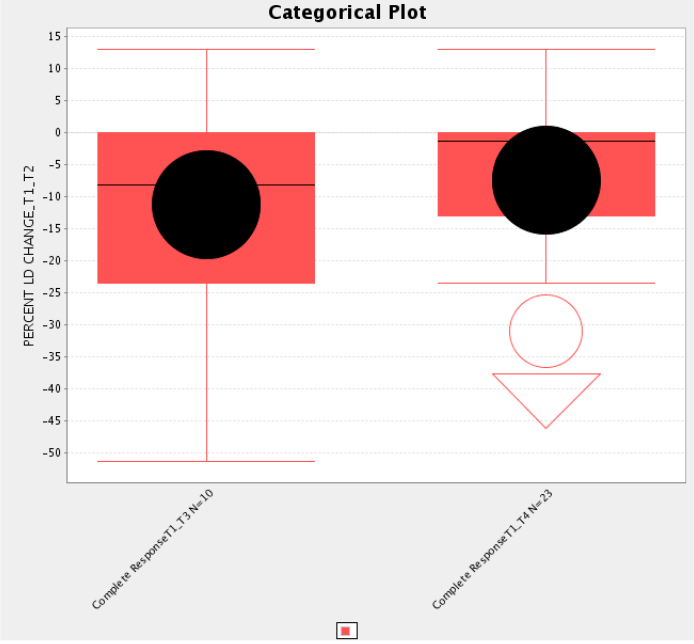


Figure 6.21 Categorical Plot

CHAPTER 7
MANAGING LISTS

This chapter describes how to manage patient and gene lists.

Topics in this chapter include:

Managing Lists Overview

The I-SPY Manage Lists function centralizes all activities pertaining to the creation and management of user-defined, as well as study-defined **PatientDID Lists** and **Gene Lists**. With these lists, you can further refine queries or facilitate analysis. Using the Manage List function, you can perform the following functions:

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

Note: On the Manage Lists page, you can minimize the number of lists displayed by clicking the PatientDID Lists or Gene Lists headings.

PatientDID Lists

An I-SPY *PatientDID list* is a list of patients with certain characteristics that you can use to filter a query or perform analysis. These lists are pre-defined in the Manage Lists function, or you can create PatientDID lists from the ID Lookup function, IHC Search results, and Hierarchical Clustering results. The side bar displays all PatientDID lists. Mouse over a list name in the side bar to display the list's data items. Note that user-defined lists appear in red type in the side bar.

Gene Lists

An I-SPY *Gene list* is a list of genes with certain characteristics that you can use to filter a query or perform analysis. The pre-defined gene list is **defaultGene1**, or you can create your own list with the Class Comparison results. The side bar displays the names of the Gene lists. Mouse over a list name in the side bar to display the list's data items. Note that user-defined lists appear in red type in the side bar.

Viewing the Data Items in a List

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

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List** or **Gene List**).

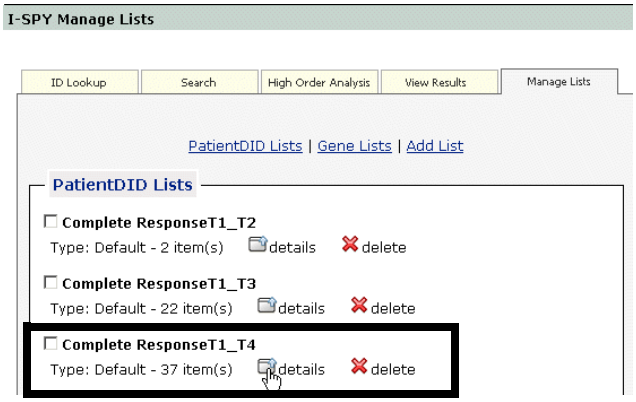


Figure 7.1 List types and Details

2. Find a list to be viewed.

   Note: If you mouse-over the list name on the Manage Lists page, the author, creation date, and comments appear.

3. Click the **details** icon to display all of the items in the list.

   Note: You can also mouse-over the list name in the side bar and the list's data item names appear in a popup window.

4. To export the list, click the **export list** link at the bottom of the data item list.

Removing Data Items to Create a New List

You may delete items from an existing list, then view the new list or save the list. To delete data items, follow these steps:

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List** or **Gene List**).
2. Find the list you want to change, and click on the box next to the list name.
3. Click the **details** icon to display all the items in the selected list.

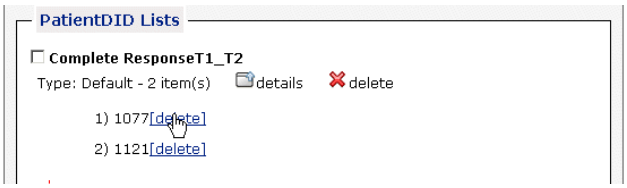


Figure 7.2 Deleting data items

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

   Once you remove the items, you can view the new list or save the list to your computer.
5. Click the **export link** at the bottom of the items list.
6. Click **Open** or **Save**.

Combining Existing Lists to Create a New List

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

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List** or **Gene List**).
2. Click the box next to the list(s) to be used for creating a new list.

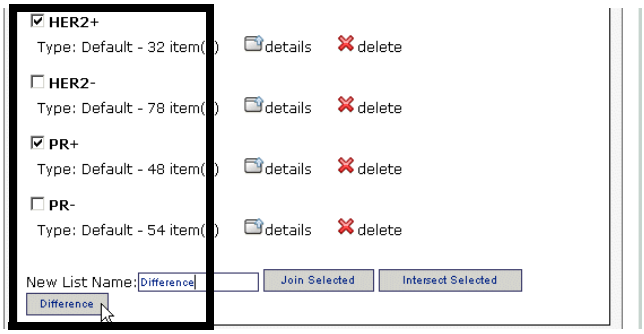


Figure 7.3 Combining existing lists

3. In the **New List Name** box, enter a unique name for the new list you are creating, and then click the appropriate button:

   Note: You cannot select more than two lists when using the **Difference** option.

   - **Join** combines two or more categories into a on list.
   - **Intersect** creates a new list from only the items that appear on more than one selected list category.
   - **Difference** creates up to two lists each comprising items that appeared in one of the selected lists. For example, if you select **HER2+** and **PR+**, the new lists are "HER2+ - PR+" comprising the items that appeared in the HER2 list only and "PR+ - HER2+" comprising the items appearing in the PR+ list only (*Figure 7.4*).

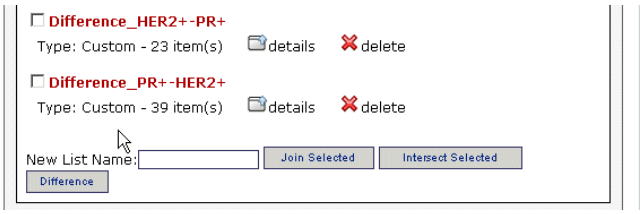


Figure 7.4 New Difference lists

The new list(s) appear on the Manage Lists page and in the side bar in red (*Figure 7.4*).

Deleting an Entire List

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

1. At the top of the Manage List page, click on the type of lists you would like to view (**PatientDID List** or **Gene List**).
2. Find the list you want to delete, and click the box next to the list name. Click more than one box to select multiple lists for deletion.

Figure 7.5 Deleting an entire list

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

Adding a New "Custom" List

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

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

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

2. To upload a list, follow these steps:

   1. Click **Upload List** at the top of the box.

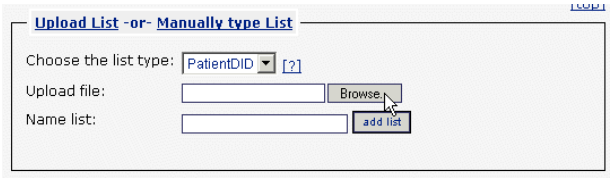


Figure 7.6 Uploading a list

   2. From the **Choose the list type** drop-down list box, select the list to be uploaded.
   3. Click the **Browse** button beside the **Upload file** box. Navigate to and select the file to be uploaded.

   Note: The list must have one ID on each line, and the file must end with a Return.

   4. Enter a unique name for the list, and then click the **Add List** button. The new list appears on the side bar in red.

3. To create and add a list manually, follow these steps:

   1. Click **Manually Type List** at the top of the box.

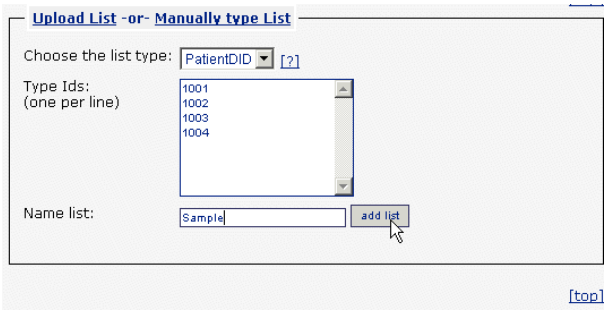


Figure 7.7 Manually typing a list

   2. From the **Choose the list type** drop-down list box, select the list to be uploaded.
   3. In the **Type Ids** box, enter items into the text block by typing them one to a line. End the list with a Return.

4. Enter a unique name for the list, and then click the **Add List** button. The new list appears on the side bar in red.
5. To open and view the newly created list or save it to your computer, click on the list name in the side bar. Click **Open** or **Save**.

6. To open and view the newly created list or save it to your computer, click on the list name in the side bar. Click **Open** or **Save**.

APPENDIX A
DATA DICTIONARY

This appendix lists the contents of the data dictionary.

Topics in this appendix include:

I-SPY Data Dictionary

Table A.1 lists the variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
ISPY_ID (or De_ID)	I-SPY identifier uniquely corresponds 1 to 1 to the CALGB patient identifier	Integer 1001-1239
DataExtractDt	Date data was downloaded from the CALGB database, for this transfer it is April 16, 2007	Date 10 format mm/dd/yyyy
Inst_ID	Registering Institution 372=U North Carolina at Chapel Hill 443=Chicago 508=U Ca at San Francisco 530=Mem Sloan-Kettering Cancer Ctr 540=Georgetown 2527=U Texas SW Med Ctr 939=U Penn 2051=UWash 2527=UTexas 2647=U Alabama (Birmingham) 2790=ECOG	Num

Table A.1 I-SPY variables

Patient Demographic Data Dictionary

Table A.2 lists Patient Demographic variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
ISPY_ID (or De_ID)	I-SPY identifier uniquely corresponds 1 to 1 to the CALGB patient identifier	Integer 1001-1239
DataExtractDt	Date data was downloaded from the CALGB database, for this transfer it is April 16, 2007	Date 10 format mm/dd/yyyy

Table A.2 Patient demographics variables

Variable Name	Variable Description	Format
Inst_ID	Registering Institution 372=U North Carolina at Chapel Hill 443=Chicago 508=U Ca at San Francisco 530=Mem Sloan-Kettering Cancer Ctr 540=Georgetown 2527=U Texas SW Med Ctr 939=U Penn 2051=UWash 2527=UTexas 2647=U Alabama (Birmingham) 2790=ECOG	Num
AgeCat	Patient Age Category 1= 18-30 2= >30-40 3= >40-50 4= >50-60 5= >60-70 6= >70-80 7= >80-<89 >=89 not available	Integer
Age	Patient Age	Integer
Race_id	Patient Race (1 case is 136 for white, black and American Indian or Alaska Native)	Character
Sstat	Survival Status (7=Alive, 8= Dead, 9=Lost)	Integer
SurvDtD	Survival date (time from study entry to death or last follow-up)	Integer

Table A.2 Patient demographics variables

Chemotherapy Summary Data Dictionary

Table A.3 lists Chemotherapy Summary data variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
Chemo A=Adria C=Cytoxan E=Epirubicin F=5-Fu Td=Docetaxel Tp=Paclitaxel H=Herceptin TpTd =Crossover from Tp to Td or Td to Tp	Neo-Adjuvant Chemotherapy Regimen 1=AC 2=EC 3=FEC 4=FAC 5=A 6=AC → Td 7=AC → Tp 8=A → Td 9=A → Tp 12=AC → TdH 13=AC → TpH 14=FEC → Tp 15=EC → Tp 16=AC → TpTd 17= A → Tp → C 19=AC → Td → Xeloda 20=EC → Tp → Carboplatin 21=FEC → Tp → Abraxane 22=AC → Td → Navelbine → Xeloda 23=AC → Tp → Vinorelbine → Tarceva 24= C → Tp	Integer
ChemoCat	Chemotherapy Group Category (Condensed from prior variable "Chemo") 1=Anthracycline Only Regimen 2=Anthracycline Plus Taxane 4=Anthracycline + Taxane + Other	Integer
DoseDenseAnthra	Dose Dense Anthracycline Therapy? 0= Standard Therapy, q3 wks 1= Dose Dense Therapy, q2 wks	Integer
DoseDenseTaxane	Dose Dense Taxane Therapy? 0= Standard Therapy, q3 wks 1= Dose Dense Therapy, q2 wks	Integer
Tam	Tamoxifen received 1=Yes	Integer

Table A.3 Chemotherapy summary data variables

Variable Name	Variable Description	Format
Herceptin	Herceptin received 1=Yes	Integer

Table A.3 Chemotherapy summary data variables

On-Study Data Dictionary

Table A.4 lists On-Study data variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
MenoStatus	Menopausal Status 1=Pre (<6 mo since LMP AND no prior bilateral ovariectomy AND not on estrogen replacement) 2= Post (prior bilateral ovariectomy OR > 12 mo since LMP with no prior hysterectomy) 5=Indeterminate, not 1 or 2 above	Integer
SentinelNodeSample	Was Sentinel node sampling performed pre-treatment? 0=no 1=yes	Integer
SentinelNodeResult	Sentinel Node Biopsy Results pre-treatment 0=Negative 1=Positive	Integer
HistTypeInv	Histologic Type of Invasive Tumor (On-Study) 1=Necrosis 2=Ductal Carcinoma 3=Lobular 4=Mixed Ductal/Lobular carcinoma 5=Other 6=No invasive tumor present	Integer
HistologicGradeOS	Combined Histologic Grade - On-study (According to SBR/Elston Classification) 1=Grade I (low) 2= Grade II (intermediate) 3= Grade III (high) 4= Indeterminate	Integer
ER_TS	Estrogen Receptor Status – Total Score Total Score = ER_PS+ ER_IS Considered Allred Score; **3 is positive**	Integer

Table A.4 On-Study summary data variables

Variable Name	Variable Description	Format
PgR_TS	Progesterone Receptor Status – Total Score Total Score = PgR_PgS+ PgR_IS Considered Allred Score**; 3 is positive**	Integer
The following Her2 measures were performed in the **community** as opposed to centrally If FISH is 3+.. consider positive; if IHC is 2+ look at FISH result		
Her2CommunityPos (Based on Algorithm of the Her2 community measured variables)	Her2 Summary as measured in the Community 0=negative 1=positive* 2= indeterminate** 3=not done *Any 3+ or FISH ratio 2 ** 2+	Integer
Her2CommunityMethod	Her2 Summary method as measured in the Community 1= IHC 2=FISH 3= Other 4=Unknown	Integer

Table A.4 On-Study summary data variables

Post-Surgery Summary Data Dictionary

Table A.5 lists Post-Surgery Summary data variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
SurgeryLumpectomy	Surgery Procedure Performed was Partial mastectomy/ lumpectomy/ excisional biopsy 0=no 1=yes	Integer
SurgeryMastectomy	Surgery Procedure Performed was Mastectomy, NOS 0=no 1=yes	Integer
InitLump_FupMast	Initial Lumpectomy Surgery followed by Mastectomy Surgery at a later date 1=Yes	Integer

Table A.5 Post-surgery summary data variables

Variable Name	Variable Description	Format
Surgery	Did patient have extensive Primary Surgery ***immediately*** following chemotherapy? 1=Yes, Lumpectomy/Mastectomy or Both 0=Delayed surgery (>8 weeks); following RT 2= Delayed surgery (>8 weeks); other than primary RT 3=No, Disease Progression 4=No, Adverse Event 5=No, Patient Refusal 6=No, Other Reason, Not specified	Integer
DCISonly	DCIS only thing left following surgery 1=Yes	Integer
PTumor1Szcm_Micro	Primary Tumor **Pathological** Tumor Size – Microscopic, measured in cm	Number with decimal
HistologicTypePS	Histologic Type of Primary Tumor (Post-Surgery) 1=Necrosis 2=Ductal Carcinoma 3=Lobular 4=Mixed Ductal/Lobular carcinoma 5=Other 6=No residual invasive breast cancer	Integer
HistologicGradePS	Combined Histologic Grade -Post-Surgery (According to SBR/Elston Classification) 1=Grade I (low) 2= Grade II (intermediate) 3= Grade III (high) 4= Indeterminate 5=No residual invasive breast cancer	Integer
NumPosNodes	Total Number positive Axillary + Sentinel (post) Nodes, post-chemotherapy	Number
NodesExamined	Total Number of Axillary + Sentinel (post) nodes Examined, post-chemotherapy	Number

Table A.5 Post-surgery summary data variables

Variable Name	Variable Description	Format
PathologyStage	Pathology Assessment Staging 1= Stage 0 (DCIS only) 2= Stage I 3= Stage IIA 4=Stage IIB 5= Stage IIIA 6= Stage IIIB 7=Stage IIIC 8= Stage IV	Integer
ReasonNoSurg	Principal Reason Breast Conserving Surgery Not Performed 1=Multicentric Disease 2=Inflammatory Disease 3=Diffuse microcalcifications 4=Patient Choice/ Family history 5=Institutional Norm 6=Specific anatomy of primary 7=Other (Note this data comes from both the C-911 and C-931 forms)	

Table A.5 Post-surgery summary data variables

Follow-Up Data Dictionary

Table A.6 lists Follow-Up Summary data variables, description, and format for the data dictionary:

Variable Name	Variable Description	Format
RtTherapy	Has patient received adjuvant radiation therapy (prior to treatment failure or second primary cancer)? 1=No , 2= Yes	Integer
RtBreast	Radiation to Breast 0= No, 1 = Yes	Integer
RtBoost	Radiation to Boost 0= No, 1 = Yes	Integer
RtAxilla	Radiation to Axilla 0= No, 1= Yes	Integer
RtSNode	Radiation to Supraclavicular node 0= No, 1= Yes	Integer
RtIMamNode	Radiation to Internal Mammary node 0=No, 1 = Yes	Integer

Table A.6 Follow-up summary data variables

Variable Name	Variable Description	Format
RTChestW	Radiation to Chest Wall 0=No, 1= Yes	Integer
RtOther	Radiation to Other Site 0= No, 1= Yes	Integer

Table A.6 Follow-up summary data variables

Response Evaluation Data Dictionary

Table A.7 lists Response Evaluation data variables, description, and format for the data dictionary:

Response evaluation form coincides with MRI schedule.

- T2 = Early Treatment Day1, cycle 2
- T3= Inter-Regimen
- T4= Prior to Surgery for Response evaluation form and sample Post-surgery

Variable Name	Variable Description	Format
TSize**Clinical**	Size of Primary Tumor (cm) – Clinical Assessment at Baseline	Number with decimal
NSize**Clinical**	Size of Largest Palpable Node (cm) – Clinical Assessment at Baseline	Number with decimal
StageTe	Disease Stage T (metastasis) Baseline	Char
StageNe	Disease Stage N (metastasis) Baseline	Char
StageMe	Disease Stage M (metastasis) Baseline	Char
ClincalStage (Based on algorithm)	Clinical Staging at Baseline 1= Stage 0 2= Stage I 3= Stage IIA 4=Stage IIB 5= Stage IIIA 6= Stage IIIB 7=Stage IIIC 8= Stage IV 9=Inflammatory	Integer
ClinRespT1_T2 (If missing from form used an algorithm to determine) RECIST criteria	Clinical Response Baseline to Early Treatment 1= CR 2=PR 3=Stable Disease 4=Progressive Disease	Integer

Table A.7 Response Evaluation summary data variables

Variable Name	Variable Description	Format
ClinRespT1_T3 (If missing from form used an algorithm to determine) RECIST criteria	Clinical Response Baseline to Inter-Regimen 1= CR 2=PR 3=Stable Disease 4=Progressive Disease 5=Inter-Regimen Chemo not given	Integer
ClinRespT1_T4 (If missing from form used an algorithm to determine) RECIST criteria	Clinical Response Baseline to Pre-Surgery 1= CR 2=PR 3=Stable Disease 4=Progressive Disease 6=No Surgery Performed	Integer

Table A.7 Response Evaluation summary data variables

MR Data Dictionary

Table A.7 lists MR data variables, description, and format for the data dictionary:

Variable Name	Variable Description
Morphological Pattern at T1	Morphological pattern at T1 Pre-Treatment - Baseline. 1=Single uni-centric mass with well-defined margin 2=Multi-lobulated mass with well-defined margins 3=Area enhancement with irregular margins (WITH nodularity) 4=Area enhancement with irregular margins (WITHOUT nodularity) 5=Septal spread; streaming
LES_T1	Lesion type at T1 Pre-Treatment - Baseline. 0=None corresponding to index 1=Mass 2=Regional
LES_T2	Lesion type at the T2 Early Treatment timepoint (as indicated on the M4 form). 0=None corresponding to index 1=Mass 2=Regional
LES_T3	Lesion type at the T3 Between Treatment Regimes timepoint (as indicated on the M4 form). 0=None corresponding to index 1=Mass 2=Regional
LES_T4	Lesion type at the T4 Pre-Surgery timepoint (as indicated on the M4 form). 0=None corresponding to index 1=Mass 2=Regional
LD_T1	Longest Diameter (LD) in the cancer mass at T1 Pre-Treatment - Baseline.
LD_T2	Longest Diameter (LD) in the cancer mass at the T2 Early Treatment timepoint.
LD_T3	Longest Diameter (LD) in the cancer mass at the T3 Between Treatment Regimes timepoint.
LD_T4	Longest Diameter (LD) in the cancer mass at the T4 Pre-Surgery timepoint.

Table A.2 I-SPY Clinical report MR data items Icons

Variable Name	Variable Description
MRI % change T#_T#	The percentage of Longest Dimension (LD) change between the two timepoints. • T1 Pre-treatment - Baseline M3 form • T2 Early Treatment • T3 Between Treatment Regimes • T4 Pre-Surgery

Table A.2 I-SPY Clinical report MR data items Icons

GLOSSARY

Acronyms and other terms referred to in the chapters of this User's Guide are described in this glossary.

Term	*Definition*
CGAP	Cancer Genome Anatomy Project
Class Comparison	Differential gene expression across the tumor types will be evaluated by calculating the typical *t*-statistic for each reporter. Both parametric and non-parametric *p*-value will be computed.
False Discovery Rate (FDR)	The expected proportion of Type I errors among rejected hypotheses in simultaneous testing of multiple null hypotheses.
Family-wise Error Rate (FWER)	Denotes the probability of having at least one false significant test result within the set of tested hypotheses.
Gene List	A pre-defined or user-defined list in I-SPY comprising genes with a set of characteristics. Used to filter a query. These are shown in the Manage Lists function.
Group	A pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics.
Hierarchical Clustering	A statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left.
High Order Analysis	After data preprocessing (filtering and normalization), further statistical analysis of gene expression data are performed.

Table B.1 Glossary of I-SPY terms

Term	Definition
IHC (Immunohistochemistry)	Method of analyzing and identifying cell types based on the binding of antibodies to specific components of the cell. It is sometimes referred to as immunocytochemistry.
NCIA	National Cancer Imaging Archive
NCI	National Cancer Institute
NCICB	National Cancer Institute Center for Bioinformatics
PatientDID List	A pre-defined or user-defined list in I-SPY comprising patients with a set of characteristics. Used to filter a query. These are shown in the Manage Lists function.
Principal Component Analysis	Commonly used in microarray research as a tool. It is designed to capture the variance in a dataset in terms of principle components. In effect, one is trying to reduce the dimensionality of the data to summarize the most important, for example defining parts, while simultaneously filtering out noise.

Table B.1 Glossary of I-SPY terms

INDEX

J

L

M

N

O

P

R

S

T

U

V

W

Y