

I-SPY USER'S GUIDE

Version 1.5



NATIONAL[®]
CANCER
INSTITUTE

Center for Biomedical Informatics
and Information Technology

January 6, 2009

CREDITS AND RESOURCES

<i>I-SPY Development and Management Teams</i>		
<i>Development</i>	<i>Documentation</i>	<i>Product and Program Management</i>
Subha Madhavan	Dana Zhang	Anand Basu
Alex Jiang	Ying Long	Shine Jacobs
Kevin Rosso	Huaitian Liu	Subha Madhavan
Ryan Landy	Laura Jackel	
Himanso Sahni	Jill Hadfield	
David Bauer		
Huaitian Liu		
Michael Harris		
Ram Bhattaru		
Ye Wu		
Ying Long		
Don Swan		
Dana Zhang		
Vesselina Bakalov		
Nick Xiao		
Gregg Silk		
Shuang Cai		

<i>Contacts and Support</i>	
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	1
Purpose	1
Audience	1
Topics Covered	1
Text Conventions Used	2
 Chapter 1	
Getting Started with I-SPY	3
About I-SPY	3
How to Cite I-SPY Data	4
Launching I-SPY	4
Creating a User Account	5
Logging In	5
Accepting I-SPY Provisions	6
I-SPY Workspace	7
I-SPY Menu	7
I-SPY Tabs	8
I-SPY Side Bar	9
Application Support	9
Logging Out	10
 Chapter 2	
Conducting an Identifier Lookup	11
ID Lookup Overview	11
Looking Up a Patient or Sample Identifier	12
Displaying/Hiding the Patient Sample Information	13
Downloading Patient Sample Information to an Excel File	14
Creating a PatientDID List with the ID Lookup	14
 Chapter 3	
Conducting Searches	15
Search Overview	15

Performing a Clinical Query	15
Performing an IHC Level of Expression Query	18
Performing an IHC Loss of Expression Query	20
Performing a p53 Query	21

Chapter 4

High Order Analysis25

High Order Analysis Overview	25
Performing a Class Comparison	26
Performing a Principal Component Analysis	28
Performing Hierarchical Clustering Analysis	29
Performing Correlation Scatter Plot Analysis	31
Performing Categorical Plot Analysis	33
Performing a GenePattern Analysis	35
Launching GenePattern	37

Chapter 5

Viewing Results39

Results Overview	39
Using I-SPY Report Icons	40
Search Results	40
Clinical Reports	41
IHC Level of Expression Search Results	43
IHC Loss of Expression Search Results	44
p53 Search Results	44
High Order Analysis Results	45
Class Comparison Report	45
Principal Component Analysis Plot	47
Hierarchical Clustering Report	49
Correlation Scatter Plot	50
Categorical Plot Analysis	52
GenePattern Analysis Reports	53

Chapter 6

Managing Lists57

Managing Lists Overview	57
Adding New Lists	58
Combining Existing Lists to Create a New List	58
Removing Data Items to Create a New List	59
Uploading a List	60
Manually Entering a List	60
Viewing the Data Items in a List	61

Deleting an Entire List	61
Exporting a List	62
Appendix A	
Data Dictionary	59
I-SPY Data Dictionary	60
Patient Demographic Data Dictionary	60
Chemotherapy Summary Data Dictionary	62
On-Study Data Dictionary	63
Post-Surgery Summary Data Dictionary	66
Follow-Up Data Dictionary	68
Response Evaluation Data Dictionary	70
MR Data Dictionary	73
Pathology Data Dictionary	74
Appendix B	
Glossary	71
Index	75

ABOUT THIS GUIDE

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

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

Purpose

This guide provides an overview of 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. It also provides instructions for getting started using I-SPY.
- [Chapter 2](#) describes how to search using patient or sample identifiers. The results show the patients that fulfill the criteria.
- [Chapter 3](#) describes how to perform a clinical and other queries on I-SPY data.
- [Chapter 4](#) extends the basic knowledge of the previous chapters and shows you how to work with high order analyses.
- [Chapter 5](#) describes how to view all the results generated from searches and high order analyses.
- [Chapter 6](#) 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.
- [Appendix B](#) 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.
<i>Italics</i>	Highlights references to other documents, sections, figures, and tables.	See <i>Figure 4.5</i> .
Italic boldface monospace type	Represents text that you type.	In the New Subset text box, enter <i>Proprietary Proteins</i> .
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

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:

- *About I-SPY* on page 3
- *Launching I-SPY* on page 4
- *Creating a User Account* on page 5
- *Logging In* on page 5
- *I-SPY Workspace* on page 7
- *Application Support* on page 9
- *Logging Out* on page 10

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.

How to Cite I-SPY Data

When referencing the I-SPY data set, please cite National Cancer Institute as the source, including year of first production release (2007), the I-SPY website (<http://ispy.nci.nih.gov>) and the accessed date.

For Example:

National Cancer Institute. 2007. I-SPY home page.
<<http://ispy.nci.nih.gov>>. Accessed 2007 September 24

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 1.1*).

National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov

i spy
post analysis portal

Progress in finding better therapies for breast cancer treatment is hampered by the lack of opportunity to integrate and rapidly test novel therapeutics in the clinical setting. In order to catalyze the transition from uniform to tailored care, clinical trials to identify markers and mechanisms of resistance to therapy are critical. A collaboration of physicians, researcher and cooperative groups are conducting one such study, the I-SPY Trial, a multi-center clinical trial of women undergoing neoadjuvant chemotherapy from breast cancer. In order to assist in the conduct of the trial and the analysis of results, a great deal of attention has been paid to facilitating collaboration, providing infrastructure for data management, analysis, and communication, and developing a commitment to sharing information and developing data standards.

The NCI Center for Bioinformatics (NCICB) is designing a web-based system to support correlative data analysis and centralized reporting of results.

username:
password:

[request username/password](#)

powered by
colintegrator

Figure 1.1 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 email 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**.

request username/password'." data-bbox="272 467 461 532"/>

username:

password:

• [request username/password](#)

Figure 1.2 I-SPY login

2. Click the **Submit** button. If your login is successful, the Legal Rules of the Road page appears ([Figure 1.3](#)).

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 1.3*) in the lower right-hand corner.

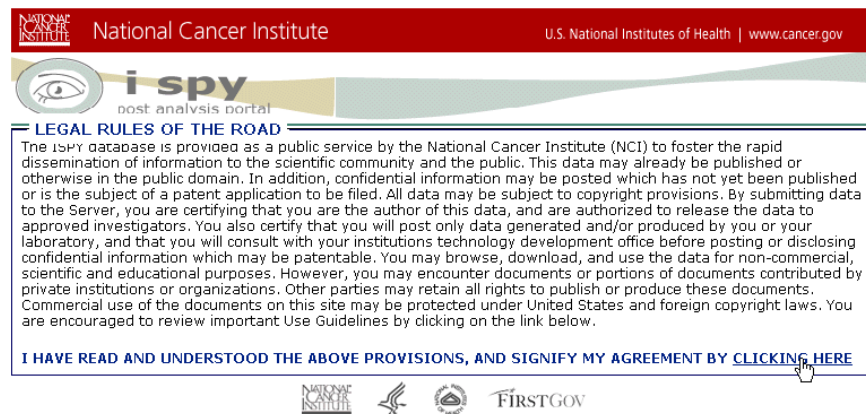


Figure 1.3 Legal Rules of the Road page

The I-SPY workspace appears (*Figure 1.4*).

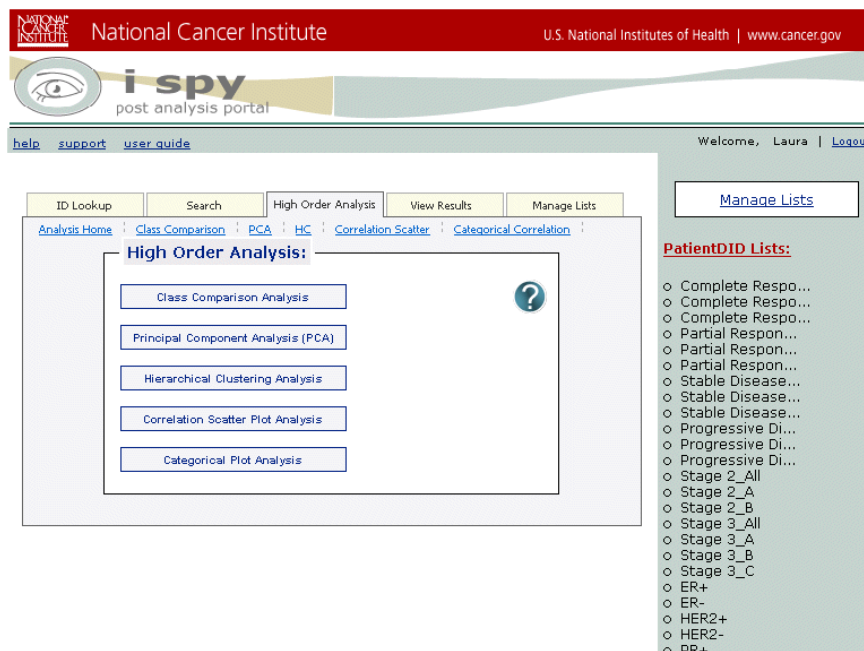


Figure 1.4 The I-SPY workspace

I-SPY Workspace

The I-SPY workspace enables quick access to all I-SPY functions and information. To access I-SPY functions use the tabs.

- *I-SPY Menu* on page 7
- *I-SPY Tabs* on page 8
- *I-SPY Side Bar* on page 9

I-SPY Menu

The I-SPY menu (*Figure 1.5*) provides access to information on how to use I-SPY, how to obtain application support, and how to log out.



Figure 1.5 I-SPY's menu

Table 1.1 describes each item on the I-SPY menu.



Menu Option	Function
help	Click to access the complete version of online I-SPY help. For complete page-level help, click  on any I-SPY page. For brief field help, click  .
support	Click to obtain support for I-SPY.
user guide	Click to access a pdf version of the <i>I-SPY 1.5 User's Guide</i> .
integrated data file	Click to download the Integrated Data File, which is the Excel spreadsheet file comprising all of the clinical data for I-SPY integrated with the MRI data. The file also contains all of the patient to sample ID mappings for each data type
Logout	Click to log out of I-SPY.

Table 1.1 I-SPY menu

I-SPY Tabs

When you log into I-SPY, you can access all I-SPY functions from the I-SPY tabs shown below.



Figure 1.6 I-SPY tabs

Table 1.2 describes each I-SPY tab on the workspace.

Tab Name	Function
ID Lookup	Search the database for patient or sample identifiers. Display, download, and save the data associated with the search criteria. For more information, see Conducting an Identifier Lookup .
Search	Perform one of the following types of queries: <ul style="list-style-type: none"> • Clinical query • IHC Level of Expression query • IHC Loss of Expression query • p53 query For more information, see Conducting Searches .
Analysis	Run higher order analyses including: <ul style="list-style-type: none"> • Class comparisons • Hierarchical clustering • Principal component analyses • Correlation scatter plot • Categorical plot analysis • GenePattern analysis For more information, see High Order Analysis .
Manage Lists	Manage user-defined or study-defined patient or gene identifier lists. For more information, see Managing Lists .
View Results	View results generated for searches and high order analyses. For more information, see Viewing Results .

Table 1.2 I-SPY tabs

I-SPY Side Bar

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

Information Displayed	Function
PatientDID List	Displays the default PatientDID lists provided with I-SPY, and displays in red any PatientDID lists added to I-SPY. See Managing Lists Overview on page 57.
Gene List	Displays the default Gene list provided with I-SPY, and displays in red any Gene lists added to I-SPY with the Class Comparison function. See Managing Lists Overview on page 57.

Table 1.3 Blue side bar

Note: To display a list's data items, hover over the name in the side bar and a popup displays the data items. To export a list to a spreadsheet file, double-click the list name.

Application Support

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

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

Logging Out

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

1. On the I-SPY workspace, click the **logout** link ([Figure 1.7](#)) in the upper right-hand corner.



Figure 1.7 Logout link

The I-SPY login page appears.

CHAPTER 2

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* on page 11
- *Looking Up a Patient or Sample Identifier* on page 12

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*

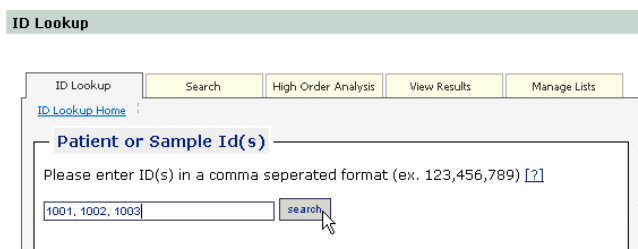
Note: You can use a *PatientDID list* to filter I-SPY queries and perform data analysis.

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 2.1](#)), 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.

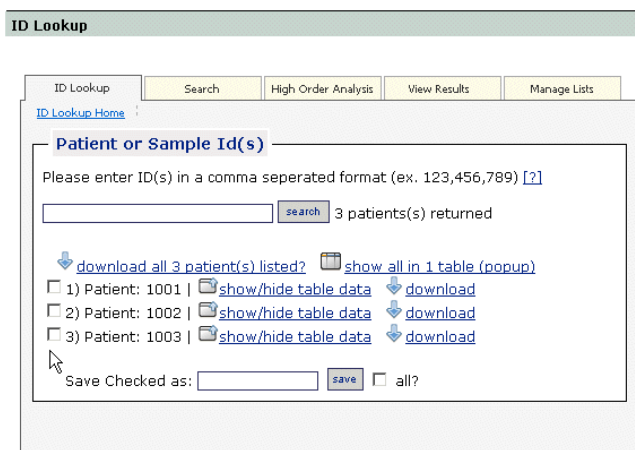


The screenshot shows the 'ID Lookup' page with a navigation bar containing 'ID Lookup', 'Search', 'High Order Analysis', 'View Results', and 'Manage Lists'. Below the navigation bar is a link to 'ID Lookup Home'. The main section is titled 'Patient or Sample Id(s)' and contains the instruction 'Please enter ID(s) in a comma separated format (ex. 123,456,789) [?]'. A text input field contains the text '1001, 1002, 1003' and a 'search' button is located to its right.

Figure 2.1 Entering identifiers

2. Click the **Search** button.

The patients associated with the identifier(s) appear below the **Search** button ([Figure 2.2](#)).




The screenshot shows the 'ID Lookup' page with the same navigation bar. The 'Patient or Sample Id(s)' section now displays the search results. It shows a 'search' button and the text '3 patient(s) returned'. Below this, there are three links: 'download all 3 patient(s) listed?', 'show all in 1 table (popup)', and a list of three patients. Each patient entry includes a checkbox, the patient ID, a 'show/hide table data' link, and a 'download' link. The patients are: 1) Patient: 1001, 2) Patient: 1002, and 3) Patient: 1003. At the bottom, there is a 'Save Checked as:' field, a 'save' button, and a checkbox labeled 'all?'.

Figure 2.2 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 2.3*).

[download all 3 patient\(s\) listed?](#)
[show all in 1 table \(popup\)](#)
☐ 1) Patient: 1001 | [show/hide table data](#) [download](#)

	ISPY ID	LabTrak ID	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&I FFPE Core 1


Figure 2.3 Patient table data

Table 2.1 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. <ul style="list-style-type: none"> • 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 2.1 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.

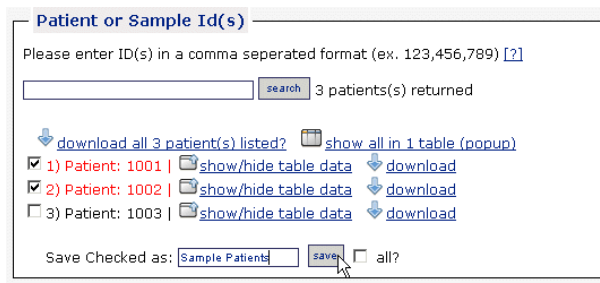
1. Click  above the list of patients to download all the patient sample information to one file.
2. Name the file and select a location.

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 2.4](#)) or select the **All** box to select all of the patients.



Patient or Sample Id(s)

Please enter ID(s) in a comma separated format (ex. 123,456,789) [\[?\]](#)

3 patients(s) returned

[download all 3 patient\(s\) listed?](#) [show all in 1 table \(popup\)](#)

<input checked="" type="checkbox"/>	1) Patient: 1001	show/hide table data	download
<input checked="" type="checkbox"/>	2) Patient: 1002	show/hide table data	download
<input type="checkbox"/>	3) Patient: 1003	show/hide table data	download

Save Checked as: ☐ all?

Figure 2.4 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 3

CONDUCTING SEARCHES

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

Topics in this chapter include:

- [Search Overview](#) on page 15
- [Performing a Clinical Query](#) on page 15
- [Performing an IHC Level of Expression Query](#) on page 18
- [Performing an IHC Loss of Expression Query](#) on page 20
- [Performing a p53 Query](#) on page 21

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 a p53 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 who had a complete response within timepoints T1 and T2.

To define a clinical query, follow these steps:

1. On the **Search** tab, click the **Clinical Query** button. The Clinical Query Form page opens ([Figure 3.1](#)).

Figure 3.1 Clinical Query Form page (top portion)

2. Select the appropriate search criteria.
- 3.

Note: You must enter a name for the search (at the bottom of the page). The name must be unique among all the queries in the current session. Selecting other criteria is optional.

[Table 3.1](#) describes the available search criteria. 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: <ul style="list-style-type: none"> • 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 3.1 Clinical Query search criteria instructions

Criteria	Item Name	Special Instructions
	Clinical Response	Select one or more clinical responses and the appropriate timepoint range. <ul style="list-style-type: none"> • 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. <ul style="list-style-type: none"> • 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 3.1 Clinical Query search criteria instructions

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

- or -

5. 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 Search tab, click the IHC Level of Expression Query button. The IHC Level of Expression Query Form opens (*Figure 3.2*).

IHC Level of Expression Query Form

ID Lookup Search High Order Analysis View Results Manage Lists

[Search Home](#) [Clinical](#) [IHC Level](#) [IHC Loss](#) [FISH](#)

Select Group [?]

none

Select Timepoint *

Timepoint: T1 T2 T3 T4

Select Biomarkers [?]

EGFR
FAK
HER2
Ki-67

Intensity of Stain [?]

BORDERLINE
MODERATE/STRONG
NEGATIVE

Percent Positive [?]

between 0 % and 100 %

Figure 3.2 IHC Level of Expression Query Form page (top portion)

2. Select the appropriate search criteria.

Note: You must enter a name for the search (at the bottom of the page). The name must be unique among all the queries in the current session. Selecting other criteria is optional.

Table 3.2 lists the available search criteria. To select more than one option in a list box, SHIFT-click or CTRL-click.

Criteria	Special Instructions
Select Group	Select a <i>group</i> 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: <ul style="list-style-type: none"> • P27 • Ki-67 • EGFR • CCND1 • P53 • HER2 • BCL2 • FAK
Intensity of Stain	Select an option that best describes the intensity of stain: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • None • Homogenous • Heterogenous

Table 3.2 IHC Level of Expression Query search criteria instructions

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

- or -

4. 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 Search tab, click the IHC Loss of Expression Query button. The IHC Loss of Expression Query Form opens ([Figure 3.3](#)).

Figure 3.3 IHC Loss of Expression Query Form page (top portion)

2. Select the appropriate search criteria.
- 3.

Note: You must enter a name for the search (at the bottom of the page). The name must be unique among all the queries in the current session. Selecting other criteria is optional.

[Table 3.2](#) lists the available search criteria. 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: <ul style="list-style-type: none"> • = • >= • <= Specify the value for the invasive sum.
Benign Range	Specify equal to, greater than, or less than to define the range benign sum: <ul style="list-style-type: none"> • = • >= • <= Specify the value for the benign sum.

Table 3.3 IHC Loss of Expression Query search criteria instructions

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

- or -

5. 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](#).

Performing a p53 Query

A [p53](#) query enables you to generate p53 mutation reports using customized search criteria. The search criteria filter the report based on patient groups, timepoints, p53 mutation status, and mutation types.

To define a p53 query, follow these steps:

1. On the Search tab, click the p53 Query button. The p53 Query Form opens ([Figure 3.1](#)).

P53 Query Form

Figure 3.4 P53 Query Form

2. Select the appropriate search criteria.
- 3.

Note: You must enter a name for the search (at the bottom of the page). The name must be unique among all the queries in the current session. Selecting other criteria is optional.

Note:

[Table 3.4](#) lists the available search criteria. 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
P53 Mutation Status	Select the mutation status by which to further filter the query: <ul style="list-style-type: none"> • WILD TYPE • MUTANT

Table 3.4 P53 search criteria

Criteria	Special Instructions
Mutation Type	Select the mutation type by which to further filter the query: <ul style="list-style-type: none"> • missense • nissense & frameshift • silent • in frame • splice site • nonsense • frameshift

Table 3.4 P53 search criteria (Continued)

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

- or -

5. To submit the query and generate the p53 search report, click the **Submit** button.

See [p53 Search Results](#) on page 44 for information about the search report.

CHAPTER 4

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](#) on page 25
- [Performing a Class Comparison](#) on page 26
- [Performing a Principal Component Analysis](#) on page 28
- [Performing Hierarchical Clustering Analysis](#) on page 29
- [Performing Correlation Scatter Plot Analysis](#) on page 31
- [Performing Categorical Plot Analysis](#) on page 33
- [Performing a GenePattern Analysis](#) on page 35
- [Launching GenePattern](#) on page 37

High Order Analysis Overview

Click the High Order Analysis tab to display the high order analysis options, shown in the list below. Follow these link for more information about each function:

- [Performing a Class Comparison](#) on page 26
- [Performing a Principal Component Analysis](#) on page 28
- [Performing Hierarchical Clustering Analysis](#) on page 29
- [Performing Correlation Scatter Plot Analysis](#) on page 31
- [Performing Categorical Plot Analysis](#) on page 33
- [Performing a GenePattern Analysis](#) on page 35

Report results are listed on the View Results page.

Performing a Class Comparison

A *Class Comparison analysis* allows 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 4.1*) enables you to define the criteria to perform a class comparison.

Figure 4.1 Class Comparison Analysis Form page

2. You are required to complete at least one criteria for the class comparison. *Table 4.1* 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 <i>two groups</i> at the same timepoint.
	Across Timepoints	Select a range of timepoints in which to perform the analysis. This option analyzes <i>one group</i> at different timepoints.

Table 4.1 Class Comparison criteria instructions

Criteria	Item Name	Special Instructions
Select Group	Existing Groups Selected Groups	<p>Select a group to filter the query to a collection of patients. Lists that you created appear in red.</p> <p>For a <i>Fixed Timepoint</i> analysis, select two groups (compares two groups at the same timepoint).</p> <p>For an <i>Across Timepoints</i> analysis, select one group (analyzes one group at different timepoints).</p>
	Baseline	<p>For a <i>Fixed Timepoint</i> analysis, the baseline is determined by the second group in the Selected Groups box.</p> <p>For an <i>Across Timepoints</i> analysis, the baseline is determined by the first timepoint in the chosen range.</p> <p>The (baseline) appears in red next to your selection.</p>
Select Statistic	Default	Select to perform a default statistical analysis.
	Advanced	Select to define additional statistical analysis options.
	<ul style="list-style-type: none"> Statistical Method 	<p>Select the appropriate statistical method:</p> <ul style="list-style-type: none"> 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.
	<ul style="list-style-type: none"> Multiple Comparison Adjustment 	<p>Family-wise Error Rate (FWER): Bonferroni</p> <p>False Discover Rate (FDR): Benjamini-Hochberg</p>
	<ul style="list-style-type: none"> 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.
	<ul style="list-style-type: none"> p-value 	<p>The probability for obtaining the differences in expression values between tumor (or a subtype of tumor) and non-tumor samples.</p> <p>The default is ≤ 0.05.</p>
Select Array Platform	Select Array Platform	Select the array platform.

Table 4.1 Class Comparison criteria instructions

- You must enter a title/name for this analysis in the **Name Analysis Result** text box. This name must be unique among all your queries in this session.

- To submit your criteria and create a 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:

- The Principal Component Analysis (PCA) Form page ([Figure 4.2](#)) enables you to define criteria to perform a PCA.

Figure 4.2 Selecting Principal Component Analysis criteria

- You are required to complete at least one criteria for the Principal Component analysis. [Table 4.2](#) 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.
	<ul style="list-style-type: none"> 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.

Table 4.2 Principal Comparison Analysis criteria instructions

Criteria	Item Name	Special Instructions
	<ul style="list-style-type: none"> 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 4.2 Principal Comparison Analysis criteria instructions

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

Performing Hierarchical Clustering Analysis

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 4.3](#)) enables you to fill in criteria for a hierarchical clustering.

Figure 4.3 Selecting Hierarchical Clustering criteria

2. You are required to enter at least one step for the hierarchical clustering. [Table 4.3](#) lists the available criteria:

Criteria	Item Name	Special Instructions
Filter Genes/Reporters	Default	Select to perform a default statistical analysis.
	Advanced	Click to define additional gene/reporter filters.

Table 4.3 Hierarchical Clustering criteria instructions

Criteria	Item Name	Special Instructions
	<ul style="list-style-type: none"> 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.
	<ul style="list-style-type: none"> 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 Statistic	Distance Matrix	Select a distance matrix option: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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 4.3 Hierarchical Clustering criteria instructions

- You must enter a title/name for this analysis in the **Name Analysis Result** text box. This name must be unique among all your queries in this session.
- To submit your criteria and create a 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 4.3](#)) enables you to fill in criteria for generate a correlation scatter plot.

Figure 4.4 Selecting Correlation Scatter Plot criteria

2. You are required to enter at least one step for the correlation scatter plot. [Table 4.4](#) 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.

Table 4.4 Correlation Scatter Plot criteria instructions

Criteria	Special Instructions
X-Axis	Enter gene information or select a timepoint range in which to analyze the percentage of LD (Longest Diameter) change.
<ul style="list-style-type: none"> Gene 	<p>Select Gene. Note that to generate the plot, you must select a gene symbol for one axis.</p> <ul style="list-style-type: none"> Enter a gene symbol. Select an array. Click the Lookup Properties button. Select a reporter.
<ul style="list-style-type: none"> PERCENT LD_CHANGE 	<p>Select an option to analyze the percentage of LD (Longest Diameter) change in the size of the tumor.</p> <ul style="list-style-type: none"> 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	<p>Select a distance matrix option:</p> <p>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.</p> <p>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.</p>

Table 4.4 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 [Deleting an Entire List](#) on page 61) 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 4.3](#)) enables you to fill in the criteria to generate a categorical plot.

Figure 4.5 Selecting Categorical Plot Analysis criteria

2. You are required to complete an entry for the categorical plot. [Table 4.5](#) 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.
<ul style="list-style-type: none"> • Gene 	Select Gene . Note that to generate the plot, you must select a gene symbol for one axis. <ul style="list-style-type: none"> • Enter a gene symbol. • Select an array. • Click the Lookup Properties button. • Select a reporter.

Table 4.5 Correlation Plot Analysis criteria instructions

Criteria	Special Instructions
<ul style="list-style-type: none"> PERCENT_LD_CHANGE 	<p>Select an option to analyze the percentage of LD (Longest Diameter) change in the size of the tumor.</p> <ul style="list-style-type: none"> 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 4.5 Correlation Plot Analysis criteria instructions

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

Performing a GenePattern Analysis

I-SPY provides two means to use GenePattern to analyze gene expression data.

- This section describes sending data from I-SPY to [GenePattern](#) for analyses whose results can then be viewed in I-SPY.
- The following section describes how to launch GenePattern itself so you can work within that application.

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

- On the High Order Analysis page, click the **Send the Data to GenePattern Analysis** button.
- The GenePattern Integration Form page ([Figure 4.3](#)) that opens enables you to define the criteria for perform the GenePattern data analyses.

Gene Pattern Integration Form

ID Lookup
Search
High Order Analysis
View Results
Manage Lists

[Analysis Home](#)
[Class Comparison](#)
[PCA](#)
[HC](#)
[Correlation Scatter](#)
[Categorical Plot](#)
[GP Analysis](#)

Select Group and Time Points*

Select 2 or More Groups
Complete ResponseT1_T2

Select 1 or More Time Points
T1
T2
T3
T4

Selected Groups
Complete ResponseT1_T2 --- T2-T3

Filter Genes/Reporters

Select a gene or reporter list: defaultGene1

Select Array Platform

Agilent

Name Analysis Result *

Test (should be unique)

Figure 4.6 Selecting GenePattern criteria

Fill in the criteria to be used in the GenePattern analysis. [Table 4.6](#) lists the available criteria:

Criteria	Item Name	Special Instructions
Select Group and Timepoints	<ul style="list-style-type: none"> Select 2 or More Groups 	<p>Note: You must select more than one group. If you have chosen to compare groups with ONE fixed timepoint, move two groups into the Selected Groups box. Your baseline group is determined by the second group in the box.</p> <p>To select two or more groups, click a group and CTRL + click additional groups or click the first group and SHIFT + click the last group.</p>
	<ul style="list-style-type: none"> Select 1 or More Time Points 	<p>If you have chosen to compare groups ACROSS timepoints, move one group into the Selected Groups box. Your baseline group is now determined by the first chosen timepoint.</p> <ul style="list-style-type: none"> T1 Pre-treatment T2 Early Treatment T3 Between Treatment Regimes T4 Pre-surgery
	<ul style="list-style-type: none"> << button >> button 	<p>To move groups to the Selected Groups box, click >>.</p> <p>To remove a group from the Selected Groups list, select the name and click <<.</p>
	Selected Groups	These are the groups to be included in the GenePattern analysis. Note: You must select at least one patient group.

Table 4.6 GenePattern Integration criteria instructions

Criteria	Item Name	Special Instructions
Filter Genes/Reports	Select a gene or reporter list	Drop down the list box and select a saved list of differentially expressed genes or reporters. If you have not added a Gene List or a Reporter List reporter list, none appears.
Select Array	Select an Array Platform	Select Agilent or cDNAarrays option.
	Select a Chromosome	Drop down the list box and select the chromosome number for analysis.

Table 4.6 *GenePattern Integration 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, click the **Submit** button.

To view the results generated by the analysis, see [GenePattern Analysis Reports](#).

Launching GenePattern

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

CHAPTER 5 VIEWING RESULTS

This 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*](#) on page 39
- [*Using I-SPY Report Icons*](#) on page 40
- [*Search Results*](#) on page 40
- [*High Order Analysis Results*](#) on page 45

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 5.1 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 5.1 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](#)
- [p53 Search Results](#)

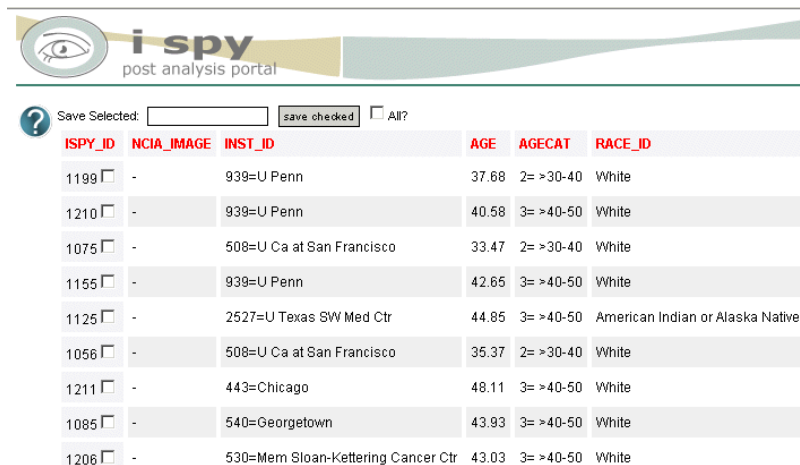
View Results ([Figure 5.1](#)) displays the query name and lists the output generated for the query.

Class Comparison (CC) (elapsed time: 861ms)	completed ✓
PCA (PCA) (elapsed time: 4756ms)	completed ✓
Search (Clinical) (elapsed time: 0ms)	completed ✓
C) Class Comparison (HC) Hierarchical Clustering (PCA) Principal Component Analysis	

Figure 5.1 Search Results

Clinical Reports

A *Clinical report* displays the demographic, clinical, MR, and pathology data for a given set of patients ([Figure 5.2](#)). 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	AGE_CAT	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 5.2 Clinical Report

[Table 5.2](#) 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
DOSESENSEANTHRA 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 5.2 I-SPY Clinical report data items cross-references

Creating a PatientDID List from a Clinical or p53 Report

On any Clinical (or the p53) page, you can select and save patients to a [PatientDID list](#).

- 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 5.3](#)).

ISPY_ID	NCIA_IMAGE	INST_ID	AGECAT	RACE_ID	SSTAT	SURVDTD
1104	<input checked="" type="checkbox"/>	-	2527=U Texas SW Med Ctr	4=50-60	Black or African American	7=Alive 516
1128	<input checked="" type="checkbox"/>	-	508=U Ca at San Francisco	2=31-40	White	7=Alive 416
1053	<input checked="" type="checkbox"/>	-	372=U North Carolina at Chapel Hill	5=60-70	White	7=Alive 723
1009	<input checked="" type="checkbox"/>	-	508=U Ca at San Francisco	3=40-50	Asian	7=Alive 1021

Figure 5.3 Checking the I-SPY ID column

- To select all of the patients, select the Check **All** box ([Figure 5.4](#)).

[<<back](#)
 Save Selected: ☒ All?

ISPY_ID	NCIA_IMAGE	INST_ID
---------	------------	---------

Figure 5.4 Selecting all of the samples on the Clinical window

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

- To save the patients to a PatientDID list, enter a name for the list (or leave the default name by which you named the query) ([Figure 5.5](#)).

[<<back](#)
 Save Selected: ☒ All?

ISPY_ID	NCIA_IMAGE	INST_ID
---------	------------	---------

Figure 5.5 Saving Selected Samples on the Clinical page

- Click the **Save Patients** button. Sample List Saved appears.
- Click the **OK** button. The new PatientDID list is now displayed in red in the right side bar at the bottom of the PatientDID names ([Figure 5.6](#)). Mouse over the name and the data items appear.

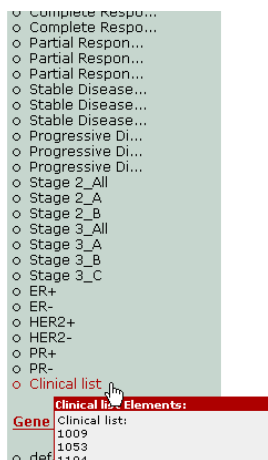


Figure 5.6 Saved PatientDID list

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 5.7*).

Save Selected: ☐ All?




ISPY_ID	LabTrak ID	NCIA_IMAGE	INST_ID	AGECAT	RACE_ID	SSTAT	SURV
1035 <input type="checkbox"/>	214064		530=Mem Sloan-Kettering Cancer Ctr	3=40-50	White	9=Lost	177
1069 <input type="checkbox"/>	227882	-	508=U Ca at San Francisco	4=50-60	White	7=Alive	588
1110 <input type="checkbox"/>	227433	-	530=Mem Sloan-Kettering Cancer Ctr	4=50-60	White	7=Alive	371

Figure 5.7 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 5.3*). 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.

IHC Level Of Expression(Query Name:IHC LOE) [\[Show/Hide Form Tools\]](#)

Select PatientDIDs: ☐ Check All [0 patients selected](#) [\[clear patients\]](#)

Displaying: 1 - 25 of 84 records [Next>>](#) [4 page\(s\)](#) [\[1\]](#) [\[2\]](#) [\[3\]](#) ... [\[last\]](#) [\[25 per page\]](#)

PatientDID ^v	Biomarker ^v	% Positive		Intensity		Localization		Result Code		Sum
		T2	T3	T2	T3	T2	T3	T2	T3	
<input type="checkbox"/> 1001	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO
<input type="checkbox"/> 1002	HER2	51-75	--	BORDERLINE	--	MEMBRANE AND CYTOPLASM	--	--	--	NO
<input type="checkbox"/> 1004	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO
<input type="checkbox"/> 1005	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO
<input type="checkbox"/> 1008	HER2	10-25	--	WEAK	--	MEMBRANE	--	--	--	NO
<input type="checkbox"/> 1009	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO
<input type="checkbox"/> 1010	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO
<input type="checkbox"/> 1012	HER2	1-9	--	NEGATIVE	--	NONE	--	--	--	NO

Figure 5.8 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 5.3](#)). 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.

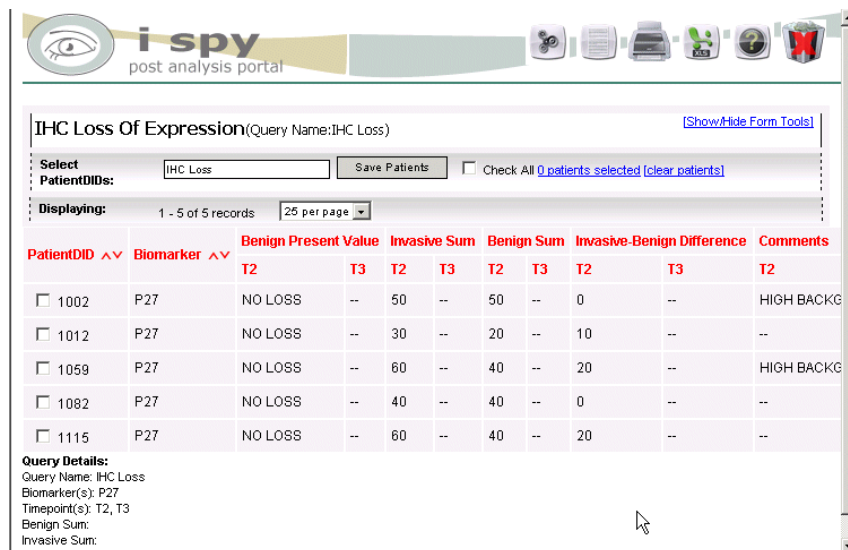


Figure 5.9 IHC Loss of Expression Report

p53 Search Results

The [p53](#) search results list records that satisfy the search criteria you specified ([Performing a p53 Query](#) on page 21).

To open and view search results, follow these steps:

1. On the View Results tab, under Report Results, click the hypertext link showing the name of your query. This opens the I-Spy Post Analysis Portal, showing the report details ([Figure 5.10](#)).

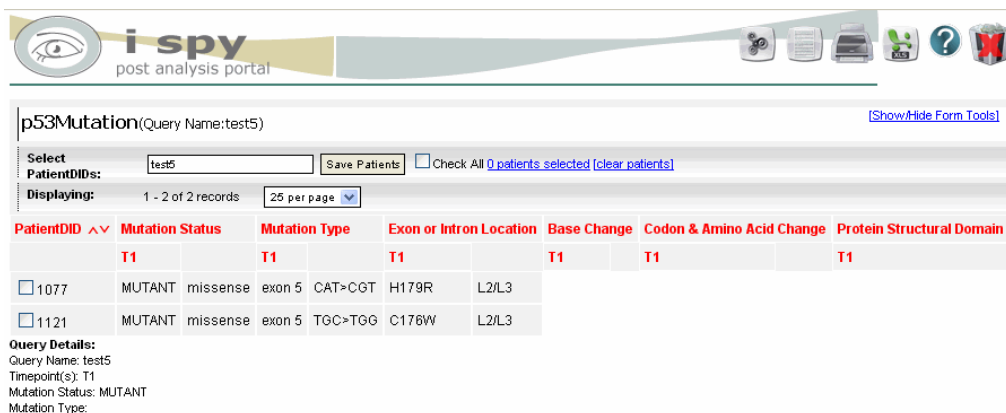


Figure 5.10 p53 query results

The report includes the following fields: PatientDIDs, Mutation Status, Mutation Type, Exon or Intron Location, Base Change, Codon & Amino Acid Change, and Protein Structure Domain.

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 [Creating a PatientDID List from a Clinical or p53 Report](#) on page 42.

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](#)
- [GenePattern Analysis Reports](#)

View Results ([Figure 5.11](#)) displays the query name and lists the output generated for the query.

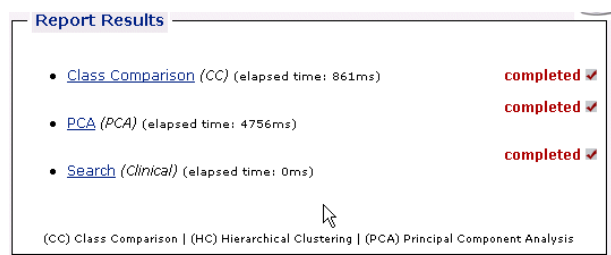



Figure 5.11 HOA Results

Class Comparison Report

The *Class Comparison report* ([Figure 5.12](#)) 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 5.12](#)), 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.



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

Select Genes: Save Genes ☐ Check All 0 genes selected [\[clear genes\]](#)

Displaying: 1 - 25 of 78 records [Next>>](#) 4 page(s) [\[1\]](#) [\[2\]](#) [\[3\]](#) ... [\[last\]](#) 25 per page

Reporter ^v	Group Avg ^v	P-Value ^v	Fold Change ^v	Gene Symbol ^v
AGI_HUM1_OLIGO_A_23_P38830	-0.0806 / 1.6300	5.68e-6	-3.2730	<input type="checkbox"/> ZNF552
AGI_HUM1_OLIGO_A_24_P408047	0.0679 / -1.0488	1.35e-3	2.1684	<input type="checkbox"/> PLEKHA4
AGI_HUM1_OLIGO_A_24_P266048	-0.0444 / 1.1086	1.48e-3	-2.2238	<input type="checkbox"/> FLJ20152
AGI_HUM1_OLIGO_A_24_P350759	-0.1635 / 0.9900	2.14e-3	-2.2246	<input type="checkbox"/> SLC1A2
AGI_HUM1_OLIGO_A_23_P47614	0.1494 / -1.0800	2.77e-3	2.3448	<input type="checkbox"/> PHLDA2

Figure 5.12 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 5.13](#)):

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.



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

Select Genes: Save Genes ☒ Check All 56 genes selected [\[clear genes\]](#)

Displaying: 1 - 25 of 78 records [Next>>](#) 4 page(s) [\[1\]](#) [\[2\]](#) [\[3\]](#) ... [\[last\]](#) 25 per page

Reporter ^v	Group Avg ^v	P-Value ^v	Fold Change ^v	Gene Symbol ^v
AGI_HUM1_OLIGO_A_23_P38830	-0.0806 / 1.6300	5.68e-6	-3.2730	<input checked="" type="checkbox"/> ZNF552
AGI_HUM1_OLIGO_A_24_P408047	0.0679 / -1.0488	1.35e-3	2.1684	<input checked="" type="checkbox"/> PLEKHA4
AGI_HUM1_OLIGO_A_24_P266048	-0.0444 / 1.1086	1.48e-3	-2.2238	<input checked="" type="checkbox"/> FLJ20152
AGI_HUM1_OLIGO_A_24_P350759	-0.1635 / 0.9900	2.14e-3	-2.2246	<input checked="" type="checkbox"/> SLC1A2

Figure 5.13 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 5.14*).



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

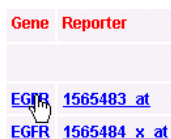
Figure 5.14 Sorting column results

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 5.15*).



Gene	Reporter
EGFR	1565483_at
EGFR	1565484_x_at

Figure 5.15 The Gene column

The Cancer Genome Anatomy Project (CGAP) browser opens.

Principal Component Analysis Plot

The *Principal Component Analysis plot* (*Figure 5.16*) is a two-dimensional graph which plots the various principal components from the analyses.

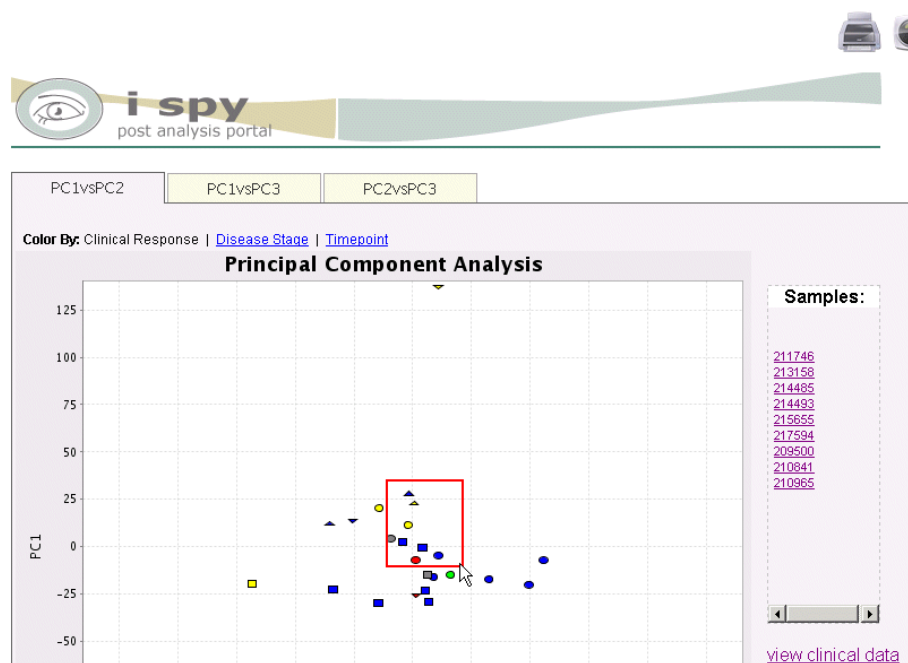


Figure 5.16 Principal Component Analysis Plot

Table 5.3 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: <ul style="list-style-type: none"> • 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 5.3 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 5.17).

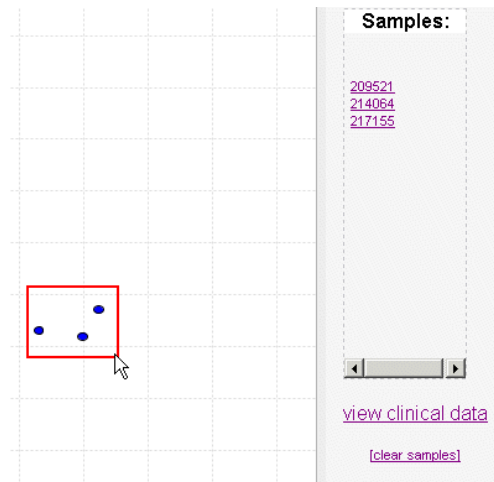


Figure 5.17 Sampling from a clinical plot

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 5.18*).

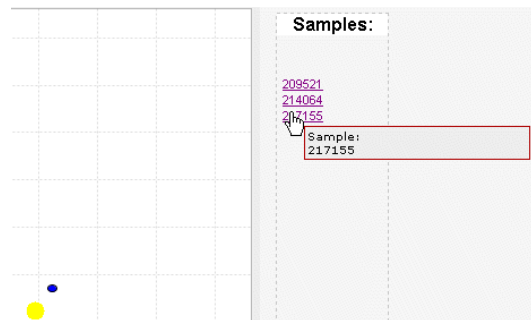


Figure 5.18 Lasso the points

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 5.19*) 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.

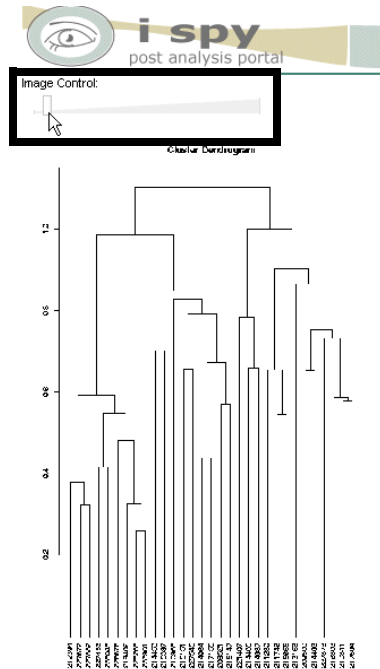


Figure 5.19 Hierarchical Clustering Dendrogram

A clinical report appears beneath the dendrogram where you can create a [PatientDID](#) [list](#) or display an image associated with a patient. For more information, see [Clinical Reports](#).

Save Selected: ☐ All?


ISPY_ID	LabTrak ID	NCIA_IMAGE	INST_ID	AGECAT	RACE_ID	SSTAT	SUR
1035 <input type="checkbox"/>	214064		530=Mem Sloan-Kettering Cancer Ctr	3=40-50	White	9=Lost	177
1069 <input type="checkbox"/>	227882	-	508=U Ca at San Francisco	4=50-60	White	7=Alive	588
1110 <input type="checkbox"/>	227433	-	530=Mem Sloan-Kettering Cancer Ctr	4=50-60	White	7=Alive	371

Figure 5.20 Hierarchical Clustering Clinical Report

Correlation Scatter Plot

The *Correlation Scatter plot* ([Figure 5.21](#)) 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.

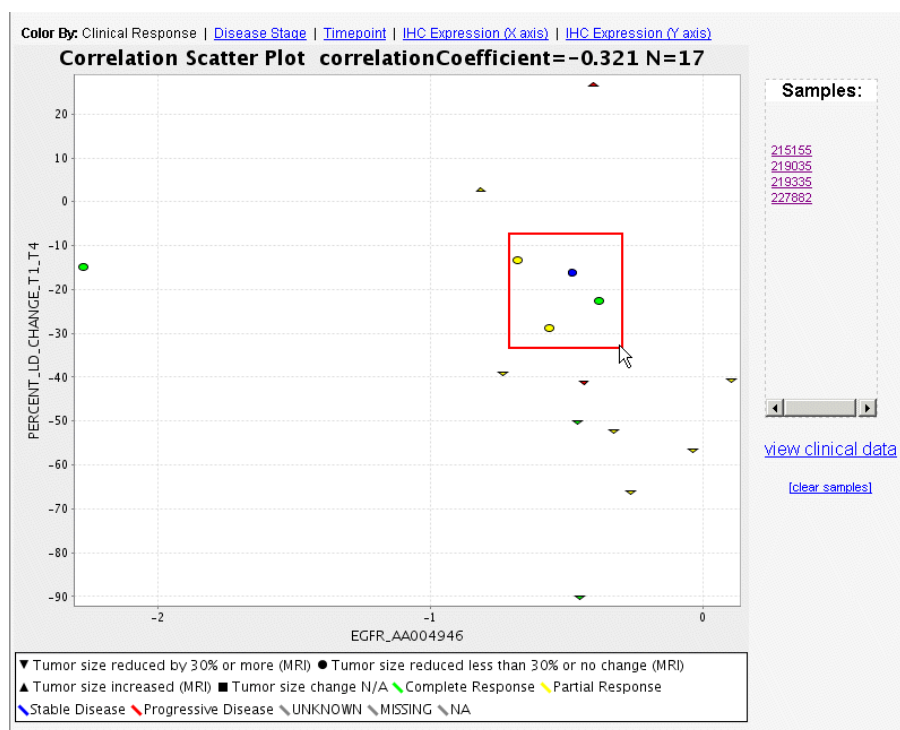


Figure 5.21 Correlation Scatter Plot

Table 5.4 describes other areas in the plot:

Area	Description
Color By	<p>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.</p> <p>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.</p>
Correlation Coefficient	Computed and displayed in the title. Correlation coefficients 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.

Table 5.4 Areas of the Correlation Scatter Plot

Area	Description
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 5.4 Areas of the Correlation Scatter Plot

Categorical Plot Analysis

The *Categorical Plot analysis* ([Figure 5.22](#)) 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.

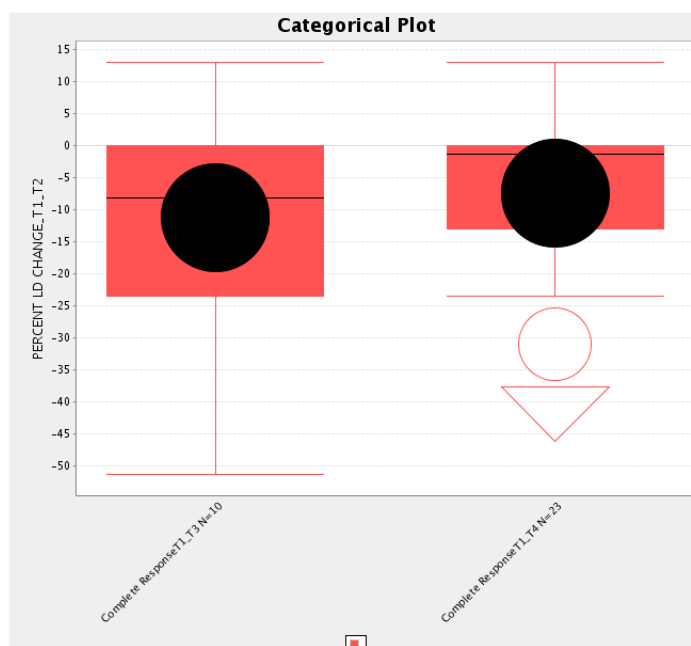


Figure 5.22 Categorical Plot

GenePattern Analysis Reports

If you have launched a [GenePattern](#) Analysis from the High Order Analysis tab, I-SPY directs you to the View Results tab where you can monitor the status of the task you have initiated. ([Figure 5.23](#)). From this tab, you can also launch GenePattern itself to perform tasks within that application. (See [Performing a GenePattern Analysis](#) on page 35.).

I-SPY View Gene Pattern Job

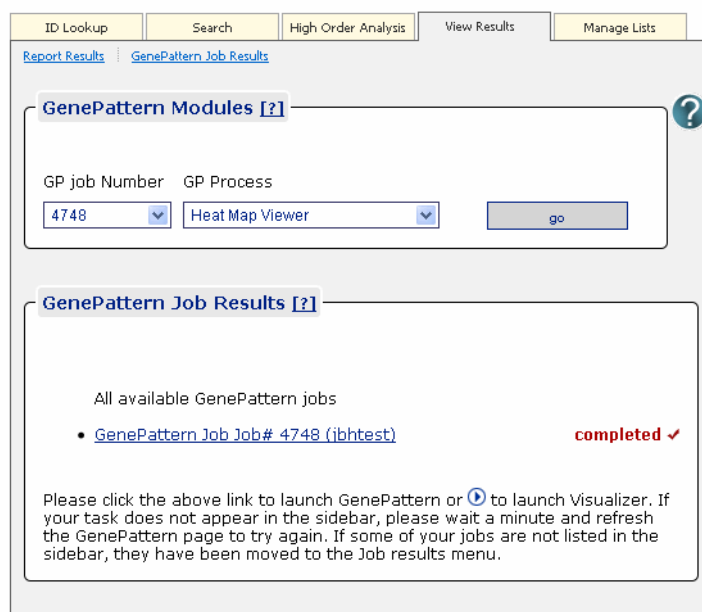


Figure 5.23 GenePattern Job results page

The GenePattern Job Results (lower panel) in the View Results tab ([Figure 5.23](#)) indicates that your query has been sent to GenePattern. It displays the number assigned to the job and its current status. When the status shows **Completed**, the job number appears in the Gene Pattern Modules section in the upper portion of the page.

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

1. Select a completed job by number. If you have more than one job in the queue, select from the drop down list the data to be analyzed.
2. Select the analysis (described in [Table 5.5](#)) that you want GenePattern to perform on your data.

Analysis Method	Description
HeatMap Viewer	Displays values in a heat map format where the largest values are displayed as the reddest (hot), the smallest values are displayed as the bluest (cool), and intermediate values are a lighter color of either blue or red.

Table 5.5 Gene expression data analyses performed by GenePattern from I-SPY

<i>Analysis Method</i>	<i>Description</i>
Hierarchical Clustering	Genes or other expression data are clustered according to predetermined algorithms.
K-Nearest Neighbors	Classifies a sample by assigning it the label most frequently represented among the k nearest samples.
Comparative Marker Selection	Finds the genes in a dataset that are most closely correlated with the two phenotypes (ALL and AML).

Table 5.5 Gene expression data analyses performed by GenePattern from I-SPY

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

The Gene Pattern Modules section of the page displays a message telling you to be patient. Your results should be finished shortly ([Figure 5.24](#))

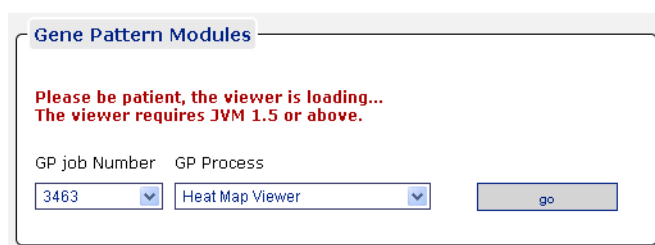


Figure 5.24 GenePattern Modules analysis results monitor

The GenePattern Job Results section of the page lists the status of the analysis. Once the job shows **completed**, click the hypertext link to view the result. Your results display based on your analysis selection, as described in the following bullets.

- If HeatMap viewer was selected as the analysis method, then the GenePattern HeatMap Viewer Applet opens up and display the values in heat map format ([Figure 5.25](#)).

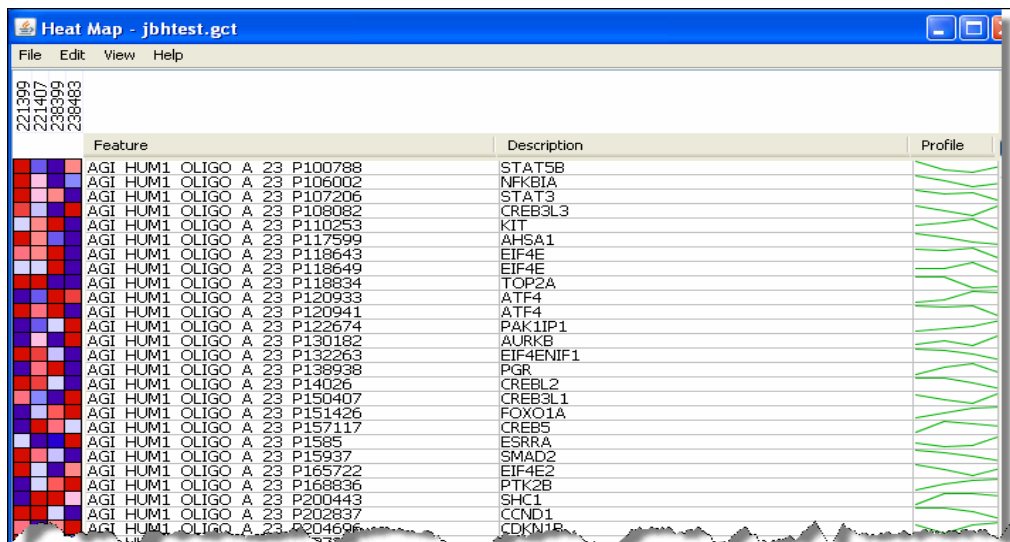


Figure 5.25 Example of HeatMap Viewer analysis results

- If Hierarchical Clustering, K-Nearest Neighbors or Comparative Marker Selection was selected, a GenePattern job is launched from I-SPY. The status of the job is displayed as GenePattern Analysis Results.

The Welcome to GenePattern page appears with general instructions on how to use the application ([Figure 5.26](#)). Refer to **Recent Jobs** on the right sidebar for the tasks that you can perform with the I-SPY data.

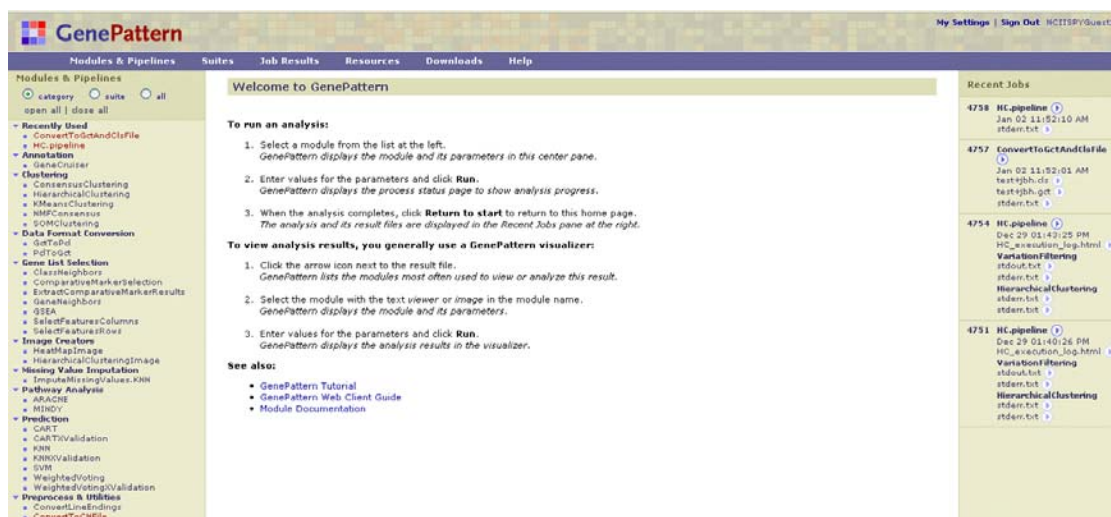


Figure 5.26 GenePattern window

If your job is not there, click **JobResults** tab.

For complete documentation on how to use GenePattern, click **Help** on the blue bar at the top of the window.

For information about performing another GenePattern query, see [Performing a GenePattern Analysis](#) on page 35.

CHAPTER 6 MANAGING LISTS

This chapter describes how to manage patient and gene lists.

Topics in this chapter include:

- *Managing Lists Overview* on page 57
- *Adding New Lists* on page 58
- *Viewing the Data Items in a List* on page 61
- *Deleting an Entire List* on page 61
- *Exporting a List* on page 62

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.

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

Adding New Lists

Table 6.1 lists how to add each list type to I-SPY.

List Type	How to Add a List
PatientDID List	<p>I-SPY provides a collection of PatientDID lists, or you can create new PatientDID lists. See the following:</p> <ul style="list-style-type: none"> • Combining Existing Lists to Create a New List • Removing Data Items to Create a New List • Uploading a List • Manually Entering a List • Creating a PatientDID List with the ID Lookup on page 14 • Creating a PatientDID List from a Clinical or p53 Report on page 42
Gene List	<p>I-SPY provides a default Gene list, or you can create a new Gene list. See the following:</p> <ul style="list-style-type: none"> • Combining Existing Lists to Create a New List on page 58 • Removing Data Items to Create a New List on page 59 • Uploading a List on page 60 • Manually Entering a List on page 60 • Creating a Gene List (Select Genes toolbar) on page 46
Reporter List	<p>To add a Reporter List, see the following:</p> <ul style="list-style-type: none"> • Uploading a List • Manually Entering a List

Table 6.1 Adding I-SPY lists

Combining Existing Lists to Create a New List

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

1. At the top of the Manage List page, click on the type of 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 6.1 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 Selected** combines two or more categories into a on list.
- **Intersect Selected** 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 6.2*).

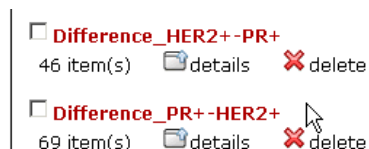


Figure 6.2 New Difference lists

The new list names appear on the Manage Lists page and in the side bar in red. When you hover over the name in the side bar, the list's data items appear popup window. When you hover over the new list name on the Manage List page, the creation date is updated, but the Author and Notes are **null**.

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.

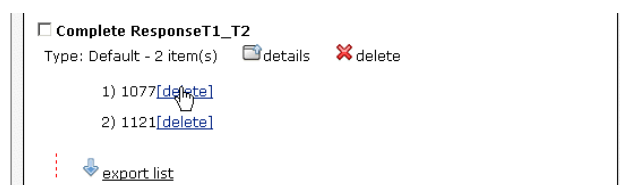


Figure 6.3 Deleting data items

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 export the list to your computer. See *Exporting a List* on page 62.

Uploading a List

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

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

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

2. Click **Upload List** at the top of the box (*Figure 6.4*).

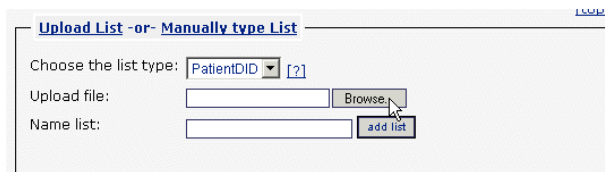


Figure 6.4 Uploading a list

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

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

Manually Entering a List

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

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

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

1. Click **Manually Type List** at the top of the box (*Figure 6.5*).

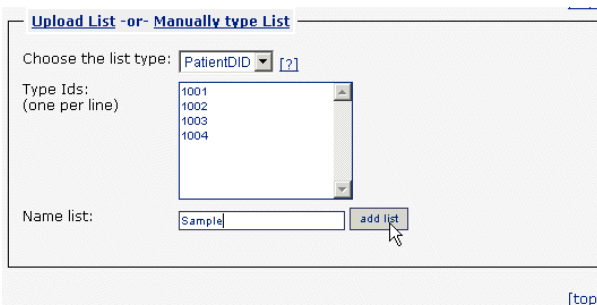


Figure 6.5 Manually typing a list

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

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

The new list name appears under the appropriate category on the Manage Lists page and in the side bar in red.

5. To display the values in the list, click **Details**.

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

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

The data items appear (*Figure 6.6*).

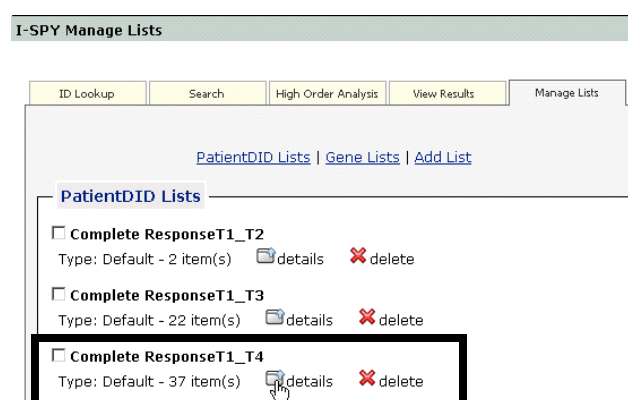


Figure 6.6 List types and Details

Note: If you mouse-over a list name in the side bar, the list's data items appear in a popup window.

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.
3. To delete the selected lists, click an **x delete** icon.



Figure 6.7 Deleting an entire list

4. The lists are removed.

Exporting a List

To export a list, follow these steps:

1. At the top of the Manage List page, click **PatientDID Lists** or **Gene Lists**.
2. Find a list to be exported, and click the **details** icon to display all of the items in the list.
Note: To export a list from the side bar, double-click the list name. Open the file in Notepad or save the patient identifiers to a spreadsheet file.
3. Scroll to the bottom of the list of data items, and click the **export list** link.
4. Open the list in Notepad or save the list to spreadsheet file on your computer.

APPENDIX A DATA DICTIONARY

This appendix lists the contents of the data dictionary.

Topics in this appendix include:

- *I-SPY Data Dictionary* on page 60
- *Patient Demographic Data Dictionary* on page 60
- *Chemotherapy Summary Data Dictionary* on page 62
- *On-Study Data Dictionary* on page 63
- *Post-Surgery Summary Data Dictionary* on page 66
- *Follow-Up Data Dictionary* on page 68
- *Response Evaluation Data Dictionary* on page 70
- *MR Data Dictionary* on page 73
- *Pathology Data Dictionary* on page 74

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 September 24, 2007.	Date 10 format mm/dd/yyyy
Inst_ID	Registering institution: <ul style="list-style-type: none"> • 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 U Wash • 2527 U Texas • 2647 U Alabama (Birmingham) • 2790 ECOG 	Number
Height	Height in centimeters (cm).	Number
Weight	Weight in kilograms (kg)	Number
BSA	Body surface area in m ²	Number

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
AgeCat	Patient age category: <ul style="list-style-type: none"> • 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

Table A.2 Patient demographics variables

Variable Name	Variable Description	Format
Age	Patient age	Integer
Race_id	Patient race: <ul style="list-style-type: none"> • 1 Caucasian • 3 African American • 4 Asian • 5 Native Hawaiian/Pacific Islander • 6 American Indian/Alaskan Native • 136 multi-racial: Caucasian, African American and American Indian or Alaska Native) 	Character
Sstat	Survival Status: <ul style="list-style-type: none"> • 7 Alive • 8 Dead • 9 Lost 	Integer
SurvDtD	Survival date. Time from study entry to death or last follow-up (time unit in days).	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: <ul style="list-style-type: none"> • 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"): <ul style="list-style-type: none"> • 1 Anthracycline Only Regimen • 2 Anthracycline Plus Taxane • 4 Anthracycline + Taxane + Other 	Integer
DoseDenseAnthra	Dose Dense Anthracycline therapy: <ul style="list-style-type: none"> • 0 Standard Therapy, q3 wks • 1 Dose Dense Therapy, q2 wks • 2 Other 	Integer

Table A.3 Chemotherapy summary data variables

Variable Name	Variable Description	Format
DoseDenseTaxane	Dose Dense Taxane Therapy: <ul style="list-style-type: none"> 0 Standard Therapy, q3 wks 1 Dose Dense Therapy, q2 wks 2 Other 	Integer
Tam	Indicates whether Tamoxifen was received: <ul style="list-style-type: none"> 0 No 1 Yes 	Integer
Herceptin	Indicates whether Herceptin was received: <ul style="list-style-type: none"> 0 No 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: <ul style="list-style-type: none"> 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	Indicates whether the sentinel node sampling was performed pre-treatment: <ul style="list-style-type: none"> 0 no 1 yes 	Integer
SentinelNodeResult	Sentinel Node Biopsy Results pre-treatment: <ul style="list-style-type: none"> 0 Negative 1 Positive 	Integer

Table A.4 On-Study summary data variables

Variable Name	Variable Description	Format
HistTypeInv	Histologic type of invasive tumor on-Study): <ul style="list-style-type: none"> • 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): <ul style="list-style-type: none"> • 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
PgR_TS	Progesterone Receptor Status – Total Score. Total Score PgR_PgS+ PgR_IS Considered Allred Score; > 3 is positive	Integer
ERpos	Estrogen Receptor Status (Allred Score or Community determined): <ul style="list-style-type: none"> • 0 Negative • 1 Positive • 2 Indeterminate 	Integer
PgRpos	Progesterone Receptor Status (Allred Score or Community determined): <ul style="list-style-type: none"> • 0 Negative • 1 Positive • 2 Indeterminate 	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		

Table A.4 On-Study summary data variables

Variable Name	Variable Description	Format
Her2CommunityPos (Based on Algorithm of the Her2 community measured variables)	Her2 summary as measured in the Community: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 1 IHC • 2 FISH • 3 Other • 4 Unknown 	Integer
FineNeedle	Indicates whether the surgical procedure involved a Fine Needle Aspiration: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
CoreNeedle	Indicates whether the surgical procedure involved a Core Needle: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
Incisional	Indicates whether the surgical procedure involved an Incisional <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
BilateralCa	Indicates whether the patient has bilateral breast cancer: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
Laterality	Tumor Laterality: <ul style="list-style-type: none"> • 1 Left • 2 Right 	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	Indicates whether the surgery procedure performed was a partial mastectomy/ lumpectomy/ excisional biopsy: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
SurgeryMastectomy	Indicates whether the surgery procedure performed was a mastectomy, NOS <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
InitLump_FupMast	Indicates whether the initial lumpectomy surgery was followed by mastectomy surgery at a later date <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
Surgery	Indicates whether the patient had an extensive Primary Surgery immediately following chemotherapy: <ul style="list-style-type: none"> • 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	Indicates whether the DCIS was the only thing left following surgery: <ul style="list-style-type: none"> • 1 Yes • 2 No • 3 No Surgery 	Integer
PTumor1Szcm_Micro	The microscopic size of the primary tumor Pathological tumor measured in centimeters.	Number with decimal

Table A.5 Post-surgery summary data variables

Variable Name	Variable Description	Format
HistologicTypePS	<p>The histologic type of primary tumor (Post-Surgery):</p> <ul style="list-style-type: none"> • 1 Necrosis • 2 Ductal Carcinoma • 3 Lobular • 4 Mixed Ductal/Lobular carcinoma • 5 Other • 6 No residual invasive breast cancer 	Integer
HistologicGradePS	<p>The combined histologic grade post-surgery according to SBR/Elston Classification:.</p> <ul style="list-style-type: none"> • I Grade I (low) • II Grade II (intermediate) • III Grade III (high) • No Surgery • Unavailable 	Integer
NumPosNodes	<p>The total number positive axillary + sentinel (post) nodes, post-chemotherapy:</p> <ul style="list-style-type: none"> • Numerical value • No Surgery 	Number
NodesExamined	<p>The total number of axillary + sentinel (post) nodes examined, post-chemotherapy:</p> <ul style="list-style-type: none"> • Numerical value • No Surgery 	Number
PathologyStage	<p>The pathology assessment staging:</p> <ul style="list-style-type: none"> • 0 • 1 Stage 0 (DCIS only) • I Stage I • IIA Stage IIA • IIB Stage IIB • IIIA Stage IIIA • IIIB Stage IIIB • IIIC Stage IIIC • IV Stage IV 	Integer

Table A.5 Post-surgery summary data variables

Variable Name	Variable Description	Format
ReasonNoSurg	<p>The principal reason why the breast conserving surgery was not performed:</p> <ul style="list-style-type: none"> • 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	<p>Indicates whether the patient received adjuvant radiation therapy (prior to treatment failure or second primary cancer):</p> <ul style="list-style-type: none"> • 1 No • 2 Yes 	Integer
RtBr	<p>Indicates radiation to the breast:</p> <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtBrTD	Total radiation boost to the breast.	Integer
RtBo	<p>Indicates whether there was a radiation boost:</p> <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtBoTD	Total radiation boost – total dose.	Integer
RtAx	<p>Indicates radiation to axilla:</p> <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtAxTD	Indicates the total dose of radiation to axilla:	Integer

Table A.6 Follow-up summary data variables

Variable Name	Variable Description	Format
RtSN	Indicates radiation to supraclavicular node: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtSNTD	The total dose of radiation to supraclavicular node.	Integer
RtIM	Indicates radiation to internal mammary node: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtIMTD	The total dose of radiation to the internal mammary node.	Integer
RtCW	Indicates radiation to the chest wall: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtCWTD	The total dose of radiation to the chest wall.	Integer
RtOt	Indicates radiation to another site: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
RtOtTD	The total dose of radiation to another site.	Integer
LocalProgress	Indicates local progression: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
DistProgress	Indicates distant progression: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
AromataseI	Indicates aromatase inhibitor: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
OvarianSup	Indicates ovarian suppression: <ul style="list-style-type: none"> • 0 No • 1 Yes 	Integer
OvarianAbl	Indicates ovarian ablation: <ul style="list-style-type: none"> • 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

Note: All Dates are relative to Study Entry/Registration Date. For example positive numbers imply dates are after study entry and negative numbers signify that an event occurred prior to study entry.

Variable Name	Variable Description	Format
TSizeClinical	Size of primary tumor (cm) – Clinical Assessment at Baseline	Number with decimal
NSizeClinical	Size of largest palpable node (cm) – Clinical Assessment at Baseline 0 No palpable nodes	Number with decimal
StageTe	Disease Stage T (metastasis) Baseline {e=cleaned data}	Char
StageNe	Disease Stage N (metastasis) Baseline {e=cleaned data}	Char
StageMe	Disease Stage M (metastasis) Baseline {e=cleaned data}	Char
ClinicalStage (Based on algorithm)	The clinical staging at baseline <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 1 CR • 2 PR • 3 Stable Disease • 4 Progressive Disease • 6 No Surgery Performed 	Integer
T4Baseline T4Early T4Int T4PreS	T4 tumor status at: <ul style="list-style-type: none"> • Baseline • Early Treatment • Inter-Regimen • Pre-Surgery • 1 Skin Only • 2 Chest Wall Only • 3 Skin and Chest Wall • 4 Inflammatory • 5 No skin involvement 	Integer
BaseAxillary EarlyAxillary IntAxillary PreSAxillary	Axillary lymph node involvement at: <ul style="list-style-type: none"> • Baseline • Early Treatment • Inter-regimen • Pre-surgery • 0 No • 1 Yes 	Integer
BaseInternalM EarlyInternalM IntInternalM PreSInternalM	Internal Mammary lymph node involvement at Baseline: <ul style="list-style-type: none"> • Early Treatment • Inter-regimen • Pre-surgery • 0 No • 1 Yes 	Integer

Table A.7 Response Evaluation summary data variables

Variable Name	Variable Description	Format
BaseSupra EarlySupra IntSupra PreSSupra	Supraclavicular lymph node involvement at Baseline <ul style="list-style-type: none"> • Early Treatment • Inter-regimen • Pre-surgery • 0 No • 1 Yes 	Integer
BaseInfra EarlyInfra IntInfra PreSInfra	Infraclavicular lymph node involvement at <ul style="list-style-type: none"> • Baseline • Early Treatment • Inter-regimen • Pre-surgery • 0 No • 1 Yes 	Integer

Table A.7 Response Evaluation summary data variables

MR Data Dictionary

Table A 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. <ul style="list-style-type: none"> • 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. <ul style="list-style-type: none"> • 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). <ul style="list-style-type: none"> • 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). <ul style="list-style-type: none"> • 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). <ul style="list-style-type: none"> • 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.8 I-SPY Clinical report MR data items

Variable Name	Variable Description
MRI % change T#_T#	<p>The percentage of Longest Dimension (LD) change between the two timepoints.</p> <ul style="list-style-type: none"> • T1 Pre-treatment - Baseline M3 form • T2 Early Treatment • T3 Between Treatment Regimes • T4 Pre-Surgery

Table A.8 I-SPY Clinical report MR data items

Pathology Data Dictionary

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

Variable Name	Variable Description
InSituDz	<p>Indicates whether In Site disease is present in the surgical specimen:</p> <ul style="list-style-type: none"> • Present • Absent • No Surgery
InSituHisto	<p>Histology of In Situ disease if present in surgical specimen.</p> <ul style="list-style-type: none"> • DCIS • LCIS • N/A • No Surgery
InSituSpan	<p>Longest diameter span (in cm) of In Situ disease if present in surgical specimen.</p> <ul style="list-style-type: none"> • numerical • N/A • No Surgery • Indeterminate
%InSitu	<p>Percentage of the overall tumor, identified within the surgical specimen, that is made up of by in situ disease (numerical up to the hundreths decimal point).</p> <ul style="list-style-type: none"> • Numerical • N/A • No Surgery • Indeterminate

Table A.1 I-SPY Pathology data items

Variable Name	Variable Description
InSituGrade	<p>Nuclear grade of In Situ disease in the surgical specimen.</p> <ul style="list-style-type: none"> • Low • Intermediate • High • N/A • Indeterminate • No Surgery
InvDz	<p>Indicates whether there was any residual invasive disease present within the breast at time of surgery. (Note: This solely represents residual disease in the breast, independent of lymph node status.)</p> <ul style="list-style-type: none"> • Present • Absent • No Surgery
InvDzHisto	<p>Histologic type of Invasive disease present in the surgical specimen.</p> <ul style="list-style-type: none"> • Ductal • Lobular • Mixed • Mucinous • N/A • No Surgery
LVI	<p>Indicates whether there is a Lymphovascular invasion (LVI). LVI is defined as tumor cells within spaces lined by endothelium).</p> <ul style="list-style-type: none"> • Yes • No • No Surgery • Unavailable
InvDzMultiFoc	<p>Indicates whether the invasive component in the surgical specimen is multifocal (more than one major site of invasive disease:</p> <ul style="list-style-type: none"> • Yes • No • N/A • No Surgery • Unavailable
InvDzCellularity	<p>Percentage of the span of the tumor in the surgical specimen that is composed of cellular (as opposed to stromal) tissue. Note: In Situ disease is counted as cellular tissue and was included in the measurement of cellularity.</p> <ul style="list-style-type: none"> • Numerical value • Not applicable • No Surgery • Indeterminate

Table A.1 I-SPY Pathology data items

Variable Name	Variable Description
SurgMargins	<p>Surgical margins. Note: If there are separately submitted margins, the TRUE margin status is reflected here.</p> <ul style="list-style-type: none"> • 0 (Negative for invasive and in situ disease) • 1 (Positive for invasive disease, completion mastectomy was performed, additional invasive disease WAS found) • 2 (Positive for invasive disease, re-excision was performed, additional invasive disease WAS found) • 3 (Positive for invasive disease, re-excision was performed, no additional invasive or in situ disease was found) • 4 (Positive for invasive disease, re-excision was NOT performed, patient had radiation instead) • 5 (Positive for in situ disease, re-excision was performed, no additional in situ or invasive disease was found) • No surgery
MetSzLN	<p>Size in centimeters (to the hundredths decimal point) of the largest lymph node metastasis (from either the sentinel or axillary lymph node dissection). This is the diameter of largest positive LN metastasis (ANL or SLN) (CM).</p> <ul style="list-style-type: none"> • Numerical value • No Surgery • Unavailable
yT	<p>Pathologic Tumor stage. Y signifies that this stage is being assigned following neoadjuvant chemotherapy.</p> <ul style="list-style-type: none"> • T0 • Tis • T1 • T2 • T3 • T4 – Inflammatory • T4 – Skin • T4 – Inflammatory, skin, chest wall <p>Note: If patient is T4, the pathologist is prompted to specify why; answer choices for this are: Skin, Chest wall, Skin and Chest wall, Inflammatory.</p> <ul style="list-style-type: none"> • No Surgery
yN	<p>Pathologic Nodal Stage. Y signifies that this stage is being assigned following neoadjuvant chemotherapy.</p> <ul style="list-style-type: none"> • Nx • N0 • N1 • N2 • N3 • No Surgery

Table A.1 I-SPY Pathology data items

Variable Name	Variable Description
yM	Pathologic Metastasis Stage. Y signifies that this stage is being assigned to the following neoadjuvant chemotherapy.- Mx <ul style="list-style-type: none"> • M0 • M1 • No Surgery
PCR	Indicates whether there is a Pathologic Complete Response (PCR). PCR is defined as no residual invasive disease. Only patients who are Stage 0 are considered PCR. Those patients who had In Situ disease only were considered disease free. <ul style="list-style-type: none"> • Yes • No • N/A
RCB Index	Residual Cancer Burden Index (defined by Fraser Symmans et al, J Clin Oncol. 2007 Oct 1;25(28):4414-22. Epub 2007 Sep 4.) is based on a calculation that takes into account the diameter in two directions of the tumor, the % of the tumor made up of by In Situ disease, the cellularity of the tumor, the number of lymph nodes positive for metastasis, and the diameter of the largest lymph node metastasis. <ul style="list-style-type: none"> • Numerical value • Unavailable
RCBclass	Residual Cancer Burden Class is a categorization as defined by Fraser Symmans et al (J Clin Oncol. 2007 Oct 1;25(28):4414-22. Epub 2007 Sep 4.) which is based on a patient's residual cancer burden index. <ul style="list-style-type: none"> • 0 (RCB Index 0) • I (RCB Index less than or equal to 1.36) • II (RCB Index greater than 1.36 and less than or equal to 3.28) • III (RCB Index greater than 3.28) • Unavailable
RCB_PATHSZ_1	The longest diameter span (in centimeters) over which the cellularity of the residual breast disease was measured; this measurement being one of the parameters used in the calculation of Residual Cancer Burden Index. <ul style="list-style-type: none"> • Numerical value • No Surgery
RCB_PATHSZ_2	A second measurement of the diameter (in cm) over which the cellularity of the residual breast disease was measured, taken in a plan perpendicular to the longest diameter measured for RCB_PATHSZ_1. <ul style="list-style-type: none"> • Numerical value • No Surgery

Table A.1 I-SPY Pathology data items

Variable Name	Variable Description
PTUMOR1SZCM_MICRO_1	Primary Tumor Pathological Tumor Size, longest diameter – Microscopic, measured in centimeters. <ul style="list-style-type: none"> • Numerical value • No Surgery
PTUMOR1SZCM_MICRO_2	A second measurement of the Primary Tumor Pathological Tumor Size; this is the diameter measurement taken in a plane perpendicular to the longest diameter measured for PTUMOR1SZCM_MICRO_1 – Microscopic, measured in centimeters. <ul style="list-style-type: none"> • Numerical value • No Surgery

Table A.1 I-SPY Pathology data items

APPENDIX B 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.
Comparative Marker Selection	An analysis which finds the genes in a dataset that are most closely correlated with the two phenotypes (ALL and AML)
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.
GenePattern	An application developed at the Broad Institute. Genepattern enables researchers to access various methods to analyze genomic data.

Table B.1 Glossary of I-SPY terms

Term	Definition
Group	A pre-defined or user-defined PatientDID list comprising patient identifiers with certain characteristics.
HeatMap Viewer	An analysis which displays values in a heat map format where the largest values are displayed as the reddest (hot), the smallest values are displayed as the bluest (cool), and intermediate values are a lighter color of either blue or red.
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.
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.
K-Nearest Neighbors	An analysis which classifies a sample by assigning it the label most frequently represented among the k nearest samples.
NCIA	National Cancer Imaging Archive
NCI	National Cancer Institute
NCICB	National Cancer Institute Center for Bioinformatics
p53	p53 is a tumor suppressor gene whose mutations are found in most tumor types, and so contribute to the complex network of molecular events leading to tumor formation.
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

Term	Definition
Reporter List	A pre-defined or user-defined list in I-SPY comprising reporters with a set of characteristics. Used to filter a query.

Table B.1 Glossary of I-SPY terms

INDEX

Symbols

46

A

Across Timepoints, definition 26

Adding

list, typing 60

list, uploading 60

adding

new list (combining) 58

new list (removing data items) 59

user account 5

analyses

Categorical Plot 33

Class Comparison 26

Correlation Scatter Plot 31

GenePattern 35

Hierarchical Clustering 29

p53 21

Principal Component 28

Application Support 9

audience

of the User's Guide 1

Average linkage, definition 30

B

Baseline, assigning 27

black dot in box

on Categorical Plot 52

Box and Whisker plot

Categorical Plot 33

C

Categorical Plot

Analysis, description 33

Analysis, report, description 52

Analysis Form,description 33

black dot in box defined 52

circles defined 52

High Order Analysis 33

line in box defined 52

triangles defined 52

uses for 52

uses for Box and Whisker plot 33

using Managing Lists 33

CGAP 71

Chemotherapy Summary data dictionary 62

circles

on Categorical Plot 52

citing

I-SPY data 4

Class Comparison 71

Analysis, definition 26

Analysis Form,description 26

example (View Results) 45

High Order Analysis 26

report description (View Results) 45

Clinical plot

selecting samples of interest 48

Clinical Query

definition 15

filling in form 16

Form page, description 16

Clinical report

creating PatientDID list 42

description 41

saving genes 46

saving samples 42

selecting samples 42

closing

a report 40

combining lists 58

Comparative Marker Selection 71

GenePattern results 55

Complete linkage, definition 30

Constrain by GeneList, definition 29, 30

Constrain reporters by variance (Gene Vector)

percentile, definition 28, 30

Correlation, definition 32

Correlation Coefficient, definition 51

Correlation Scatter Plot

- Analysis, definition 31
- Analysis Form, description 31
- examples of 31
- High Order Analysis 31
- report, description 50

Creating

- list, typing 60
- list, uploading 60

creating

- new list (combining) 58
- new list (removing data items) 59
- user account 5

D

data

- citing I-SPY datasets 4

- data analyses, GenePattern 35

- data dictionary 60

- Chemotherapy Summary 62
- Follow-Up Summary 68
- MR 73
- Pathology 74
- Patient Demographic 60
- Post-Surgery Summary 66
- Reponse Evaluation 70

data items

- removing from a list 59
- viewing in a list 61

- defaultGene1, description 30

deleting

- list 61

- Difference option, description 59

- Distance Matrix, definition 30

downloading

- ID Lookup data 14

E

e-mailing

- NCICB Application Support 5

- Euclidean distance, definition 30

Excel file

- downloading ID Lookup data 14

Exporting

- list 62

F

- False Discovery Rate (FDR) 71

- Family-wise Error Rate (FWER) 71

- Fixed Timepoint, definition 26

- Fold Change, definition 27

Follow-Up Summary

- data dictionary 68

G

Gene list 71

- creating with Manage Lists 58
- saving on Class Comparison report 46
- selecting and saving 46

GenePattern

- Comparative Marker Selection report 55
- HeatMap Viewer report 54
- hierarchical clustering report 55
- initiating analyses from I-SPY 35
- K-nearest neighbors report 55
- launching 37

GenePattern Home

- description 71

Group

- selecting 16, 22

group 72

H

- HeatMap Viewer 72

- HeatMap Viewer, GenePattern report 54

help

- documentation 7

help link 7

Hierarchical Clustering 72

- Analysis, definition 29
- Analysis Form, description 29
- dendrogram description 50
- GenePattern results 55
- High Order Analysis 29
- NCIA icon 43
- page description (View Results) 49
- report, description 49

High Order Analysis 72

High Order Analysis function

- Categorical Plot 33
- Class Comparison 26
- Correlation Scatter Plot 31
- Hierarchical Clustering 29
- overview 25
- Principal Component 28

horizontal line in box

- on Categorical Plot 52

I

ID Lookup function

- creating PatientDID list 14
- displaying patient sample information 13
- downloading data 14
- overview 11
- steps 12

- IHC (Immunohistochemistry) 72

- IHC Expression X/Y axis, Color By, definition 51

IHC Level of Expression

- defined 18
- filling in Query form 18
- report, description 43
- report, example 43

IHC Loss of Expression

- defined 20
- filling in Query form 20
- report, description 44
- report, example 44

Image Control bar 49**integrated data file link 7****Intersect option, description 59****I-SPY**

- citing data 4
- creating user account 5
- description 3
- functions 8
- getting help 7
- launching 4
- logging in 5
- logging out 10
- menu 7
- side bar 9
- tabs 8
- workspace 7
- workspace screen 6

ISPY ID

- definition 13

J**Join option, description 59****K****K-Nearest Neighbors 72****K-nearest neighbors GenePattern results 55****L****LabTrak ID**

- definition 13

launching

- GenePattern 37
- I-SPY 4

LD

- Longest Diameter 17, 32, 35, 73

Legal Rules of the Road page 6**LES**

- lesion type 73

line in box

- on Categorical Plot 52

Linkage Method, definition 30**Lists**

- exporting 62

typing a list 60

uploading a list 60

lists

- combining lists 58
- creating a new list 58, 59
- deleting 61
- removing data items 59
- viewing data items 61

logging in

- I-SPY 5

logging out

- I-SPY 10

Logout link 7**logout link 10****M****Manage Lists page**

- description 57

Managing lists

- overview 57

Managing Lists function

- creating PatientDID list 42
- for Categorical Plot analysis 33

Manually Type List options, description 60**mappings**

- file with Patient ID to Sample IDs 7

Microsoft Excel

- downloading ID Lookup data 14

Morphology, definition 17**MR**

- data dictionary 73

N**NCI 72****NCIA 72****NCIA icon 43****NCICB 72****NCICB Application Support 9****O****On-Study data dictionary**

- data dictionary
- On-Study 63

Overview

- managing lists 57

overview

- High Order Analysis function 25
- ID Lookup function 11
- Search function 15
- View Results function 39

P

- p53
 - definition 72
- p53 query
 - description 21
 - filling in form 21
 - report, description 44
- Pathology
 - data dictionary 74
- Patient Demographic data dictionary 60
- PatientDID list 72
 - creating with ID Lookup 14
 - creating with Manage Lists 58
 - in Clinical report 42
 - saving on Clinical report 42
 - user list side bar 42
- Patient ID
 - displaying sample information 13
 - mappings to sample ID 7
 - searching for 12
- Pearson correlation, definition 30, 32
- Percent LD, definition 17
- PERCENT LD_CHANGE, definition 32, 35
- Post-Surgery Summary
 - data dictionary 66
- Principal Component
 - Analysis, defined 28
 - Analysis Form, description 28
 - Analysis report, description 47
 - High Order Analysis 28
 - report example 47
- Principal Component Analysis 72
- printing
 - a report 40
- purpose
 - of the User's Guide 1
- p-value, definition 27

R

- red list name
 - user created list 42
- removing
 - data items in a list 59
- Reporter list, defined 73
- reports
 - Categorical Plot 52
 - Class Comparison 45
 - Correlation Scatter Plot report example 50
 - GenePattern Comparative Marker Selection 55
 - GenePattern HeapMap Viewer 54
 - GenePattern hierarchical clustering 55
 - GenePattern K-nearest neighbors 55
 - Hierarchical Clustering 49

- p53 44
- Principal Component Analysis 47
- using icons 40
- viewing 39

- Response Evaluation
 - data dictionary 70

S

- sample
 - selecting on Clinical report 42
 - selecting on plot 48
- Sample ID
 - displaying patient information 13
 - mappings to Patient ID 7
 - searching for 12
- Samples area
 - selecting samples 48
- samples of interest 48
- saving
 - a report 40
 - genes on Class Comparison 46
 - samples on Clinical report 42
- search
 - p53 results 44
- Search function
 - Clinical Query 16
 - IHC Level of Expression Query 18
 - IHC Loss of Expression Query 20
 - overview 15
- search function
 - p53 query 21
- searching
 - patient IDs and samples 12
- Select Genes toolbar 46
- Select Group
 - assigning 16
- side bar
 - export PatientDID list 62
 - red Gene list (Class Comparison) 46
 - red name user created 42
 - red PatientDID list (Clinical report) 42
 - red PatientDID list (Search) 14
- side bar, description 9
- Single linkage, definition 30
- sorting
 - report columns 46
- Spearman, definition 32
- spreadsheet
 - saving report to 40
- Statistical Method, definition 27
- support link 7

T

T1

Pre-treatment - Baseline timepoint [74](#)

T2

Early Treatment timepoint [74](#)

T3

Between Treatment Regimes timepoint [74](#)

T4

Pre-surgery timepoint [74](#)tabs, shown [8](#)Timepoint, definition [13](#)

timepoints

T1 through T4 define [74](#)

triangles

on Categorical Plot [52](#)top of columns to sort [46](#)

T-test Two Sample Test

definition [27](#)output [45](#)

Typing

list [60](#)**U**

Uploading

list [60](#)

User's Guide

audience [1](#)organization of [1](#)purpose [1](#)text conventions [2](#)

user account

creating [5](#)user guide link [7](#)**V**

viewing

data items in a list [61](#)

View Results function

overview [39](#)**W**

Wilcoxon Test Man-Whitney Test

definition [27](#)output [45](#)**Y**

yellow circle

meaning on plot [49](#)

