



tranSMART Foundation

tranSMART

User's Guide

Version 1.2
Edition: 1





This version 1.2 of the tranSMART documentation was created by tranSMART Foundation and is made available under a [Creative Commons Attribution 3.0 Unported License](#). Earlier versions of this tranSMART documentation were created by Johnson & Johnson and Recombinant Data Corporation (ConvergeHEALTH by Deloitte) and used under a [Creative Commons Attribution 3.0 Unported License](#).

Some company, product, and service names referenced in the documentation may be trademarks or service marks of others.

This document is licensed under the Creative Commons Attribution 3.0 Unported License.

You are free:

- to share** — to copy, distribute and transmit the work
- to remix** — to adapt the work
- to make commercial use of the work

Under the following conditions:

Attribution — You must attribute your use of the work in the following manner:

“This version 1.2 of the tranSMART documentation was created by tranSMART Foundation and is made available under a [Creative Commons Attribution 3.0 Unported License](#). Earlier versions of this tranSMART documentation were created by Johnson & Johnson and Recombinant Data Corporation (ConvergeHEALTH by Deloitte) and used under a [Creative Commons Attribution 3.0 Unported License](#).

Some company, product, and service names referenced in the documentation may be trademarks or service marks of others.”

Any blank pages in this document are intentionally inserted to allow correct double-sided printing.

Contents

Chapter 1: Getting Started with tranSMART	1
Tools	1
Logging In.....	2
Chapter 2: Browse.....	3
Overview of the Browse UI	3
Defining Search Filters	4
Keyword Search.....	5
Using the Filter Browser	6
Managing Active Filters	7
Viewing Studies in the Program Explorer Tree.....	8
Viewing and Exporting Files in Browse Folders	9
Opening a Study in Analyze View	10
Chapter 3: Analyze	13
Overview of the UI	13
Using Analyze — Basics.....	15
The Navigation Tree	16
Visual Cues in the Navigation Tree	16
Selecting Studies for Analysis	17
Public and Private Studies	18
Viewing a Study Description	19
Serial Numeric Data	20
Defining the Cohorts.....	21
Selecting Criteria.....	21
Specifying a Numeric Value	22
Joining Multiple Criteria for a Subset Definition	25
Modifying or Deleting Criteria.....	26
Saving Subset Definitions.....	26
Retrieving Saved Subset Definitions	28
Exporting Cohort Data	30
The Export Jobs List	31
Exporting Data Directly into Galaxy	32
Viewing Sample Data.....	34

Chapter 4: Summary Statistics for Analysis.....	35
Generating Summary Statistics	35
Significance Tests.....	38
View Subset Breakdown by Concept	39
Defining Points of Comparison	41
Results of a Comparison.....	41
Printing the Contents of Summary Statistics View	42
Copying Individual Charts in Summary Statistics View	42
Viewing Analysis Data in Grid View	43
Grid View Display Options	44
Chapter 5: Advanced Workflow Analyses	45
Running the Analyses	46
aCGH Survival Analysis	46
Box Plot with ANOVA	48
Correlation Analysis.....	50
Forest Plot.....	51
Frequency Plot for aCGH	55
Group Test for aCGH	56
Group Test for RNASeq	57
Heatmaps	59
IC50 Dose Response Curve.....	66
Line Graph	68
Logistic Regression.....	69
PCA	71
Scatter Plot with Linear Regression.....	73
Survival Analysis.....	77
Table with Fisher Test	78
Waterfall Plot.....	81
High Dimensional Data.....	83
Data Binning.....	84
Data Binning Using Box Plot with ANOVA.....	84
Data Binning Using Forest Plot	85
Data Binning Using Survival Analysis	87
Data Binning Using Table with Fisher Test	88
Running Across-Trial Analyses	90
Viewing Recent Analysis Jobs.....	90
Viewing a Logged Job	91

Chapter 6: Third-Party Analysis Tools	93
Dalliance Genome Browser	93
Quick Tour	93
Genomic Variants	94
MetaCore Enrichment Analysis	95
MetaCore Enrichment Analysis Based on Marker Selection Data	97
Configuration.....	97
Chapter 7: Sample Explorer.....	99
Select a Primary Search Filter	99
View and Refine Sample Search Results	100
Select and Remove Search Filters.....	101
Manage the Search Result List	102
Chapter 8: Gene Signatures and Gene Lists.....	105
Creating a Gene Signature	105
Step 1. Adding the Genes to a Text File	105
Step 2. Creating the Gene Signature	107
Performing Actions on Your Gene Signatures	111
Performing Actions on Other Users' Signatures	112
Viewing a Gene Signature Definition	113
Chapter 9: GWAS.....	115
Searching for a Study	115
GWAS Filter Browsers	116
Viewing Search Results	120
Analysis View.....	121
Table View	125
Manhattan Plot.....	126
Chapter 10: Data Upload	131
Uploading Analysis Data.....	131
File Templates	133
Chapter 11: Administration	135
Administrator Privileges	135
The Administrator's Console	135
Tasks in the Administrator's Console.....	136

Managing tranSMART Users	137
Understanding User Roles and Access Rights	138
Managing User Accounts	140
Managing Galaxy Users	143
View or Delete Galaxy Users	143
Managing tranSMART Roles	143
Understanding Role / URL Mappings	144
Managing User Roles	146
Accessing the Administrator's Console	149
Managing Security for Analyze Studies.....	149
Securing a Study.....	150
Managing Groups	150
Managing Access Privileges.....	153
Viewing the tranSMART Access Log	156
Displaying the Access Log	156
Exporting the Access Log to a Spreadsheet.....	156
Specifying the Timeframe for the Access Log	156
Browse Tool Administration	157
Creating Program Explorer Objects.....	157
Editing and Deleting Objects.....	159
Common Features for Creating and Editing Objects	160
Uploading Files to Folders.....	161
Appendix A: Download Analysis Data as R Data	163
Downloading Raw R Data	163
Reviewing R Version Information.....	164
Export R Code and Data	165
Appendix B: Glossary	167

Chapter 1

Getting Started with tranSMART

The tranSMART application reflects the efforts of various informatics groups to integrate data from internal and external data sources within a single data warehouse, and to provide scientific end users the tools to search for, view, and analyze the data in the warehouse.

The core internal data is a historical base of biomarker data from gene expression, RBM, and SNP experiments involving GWAS, Metabolic GWAS, and eQTL data types, and including both raw and analyzed data.

The tranSMART application includes a search tool that lets researchers query this vast ocean of disparate data for studies of interest and related resources.

Another major tranSMART feature, called Analyze, allows authorized users to create and study cohorts of patients that have been involved in completed clinical research efforts. Analyze includes an Across Trials feature that allows users to define cohorts made up of patients from multiple studies.



There may be some minor differences between the UI objects illustrated in this guide and the ones you see on your screen.

Tools

The tranSMART menu bar is shown below:



The following tools appear on the menu bar:

- **Browse** — Search across studies and analyses for research data related to search filters that you specify.
- **Analyze** — View study data for subjects that you select, based on criteria that you specify. Also, compare data generated for subjects in two different cohorts, based on criteria and points of comparison that you specify.
- **Sample Explorer** — Search for datasets of tested tissue and blood samples, within categories such as tissue type, pathology, and test type (such as gene expression or SNP).
- **Gene Signature/Lists** — View definitions of existing gene signatures and add new gene signature definitions.
- **GWAS** — View genetic variants in individuals to find those that may be associated with a trait of interest, such as a major disease.

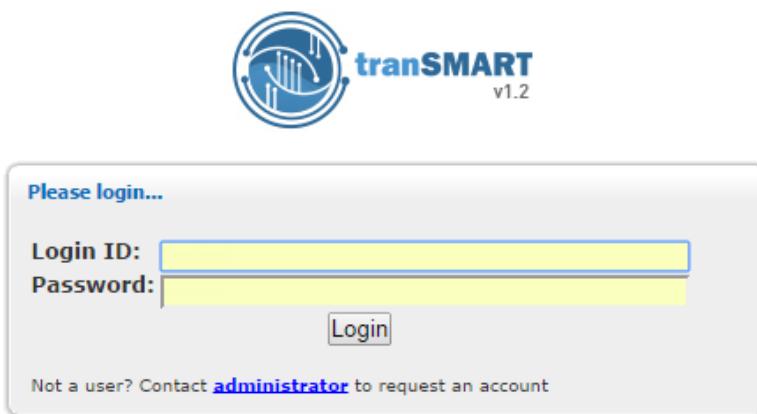
- **Upload Data** — Upload analysis data for a study.
- **Admin** — Perform administrative tasks such as creating tranSMART user accounts.
- **Utilities** — Contains submenus providing supplementary information or actions.

Logging In

To log into tranSMART:

1. Type the address of the tranSMART software into your browser's URL field.

The login screen appears:



2. Type your tranSMART login credentials and click **Login**.

Chapter 2

Browse

The tranSMART Browse feature provides a fine-grained search capability from a single user interface into studies and other data sources of interest.

You define a search query by typing search keywords into the text box, by selecting pre-defined search filters from one or more filter browsers, or by any combination of these methods. tranSMART conducts the search across multiple data sources.

The Browse feature supports filters based on one or more of the following kinds of information:

- Keywords that you specify, such as part of a study or analysis name
- Reference SNP (RS) identifiers
- Individual genes and all genes in a gene signature
- Chromosomes and a specific position within a chromosome
- Diseases and observations



Administrators only: For information on creating and editing Browse objects, see [Browse Tool Administration](#) on page 157.

Overview of the Browse UI

The figure below shows the Browse interface. It is divided into two panes:

Left pane

Use this pane to define search filters to retrieve the studies of interest.

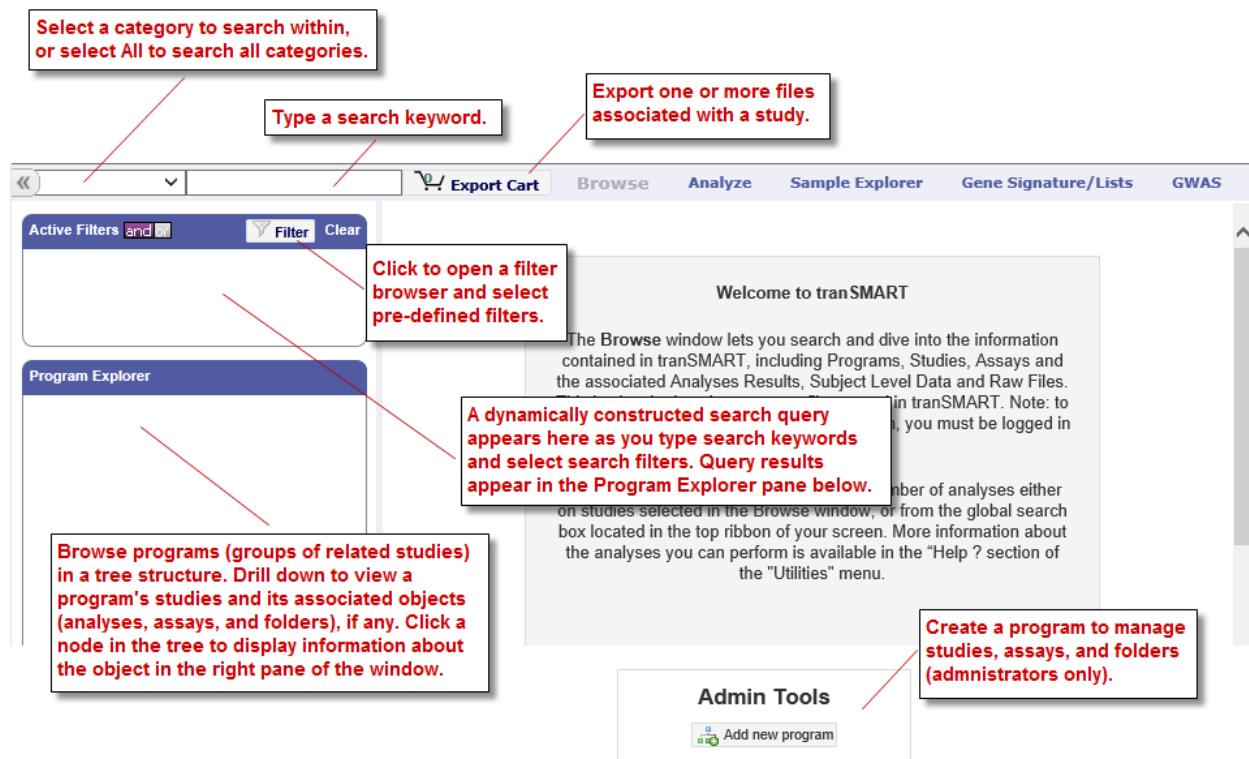
The Active Filters area displays any pre-defined filters that you have selected and free-text search keywords that you have typed; for example, the title of a study. The Program Explorer area displays the matching studies.

Right pane

Use this pane to view information about the studies and any associated objects (analyses, assays, folders) listed in the Program Explorer.

Some studies that you display in this pane can be opened in Analyze view, where you define cohorts within the study and perform a variety of analyses of the cohort data.

Features of the Browse page are shown below:



Defining Search Filters

You define search filters to retrieve just the resources that interest you.

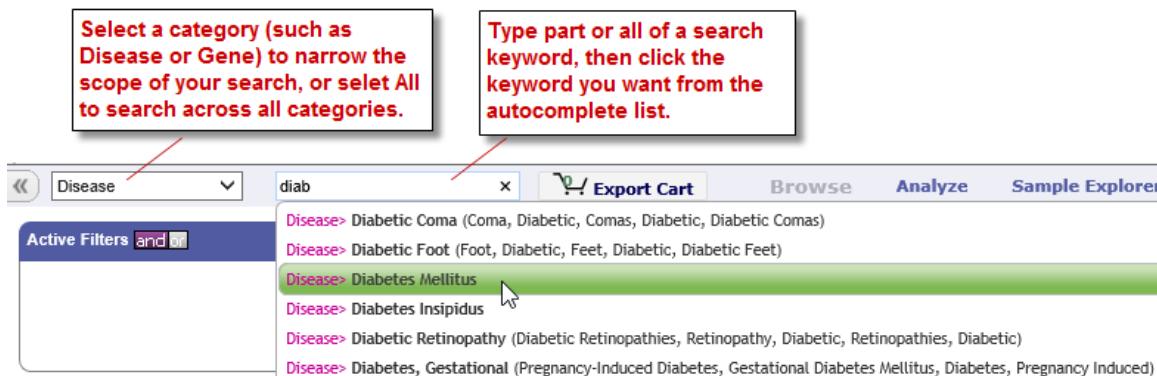
Search filters are identical in the Browse and Analyze windows. When you define a filter, it will be applied in Browse and in Analyze.

This section describes how to use keywords as search filters, using the boxes above the Active Filters area, and how to select pre-defined search filters from the [Filter Browser](#) (page 6).

The search filters you type and that you select from the Filter Browser are displayed in the [Active Filters](#) area (page 7).

Keyword Search

The following figure shows the controls for defining a keyword search:



There are two types of keyword searches:

- Keyword searches based on dictionaries. These searches apply to all categories of pre-defined metadata, such as Disease.
- Keyword searches based on free-text fields, such as names and descriptions, as well as file names and content (for indexed files). In Analyze, free-text searches apply to all tree nodes in the Navigate Terms pane.

To define a keyword search (any category except Free Text):

1. Select one of the categories in the category dropdown control, or select **All** to search across all categories.



Gene searches return all matches of the gene, not just results that are statistically significant.

2. Specify part or all of a search keyword in the text field to the right of the category dropdown.

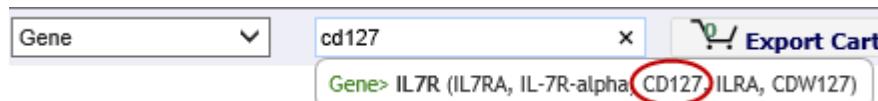
When you type at least two characters in the field, transSMART begins to search within the specified category and lists keywords that begin with those characters. The search text is not case sensitive.

The following figure shows a list of the keywords displayed when the characters **il7** are typed and the category **Gene** is selected:



Up to 15 keywords can be displayed at one time. If you don't see the one you want, type more characters into the field.

Note that the search looks for matches based on the characters at the beginning of a keyword (in bold) or, as shown below, at the beginning of a keyword synonym (in parentheses).



3. Click the keyword you want, but do not press Enter or Return.

When you click the keyword, the following actions occur:

- The search begins immediately. The contents of the Program Explorer are updated, and a result is displayed in the right pane.
- The search filter appears in the Active Filters area:



You can add more filters by repeating the steps above, by selecting filters from the Filter Browser, or by a combination of these actions.



Search filters for SNPs, genes, and gene signatures do not filter out studies and analyses that omit the specified SNP or gene. However, the only records returned for an analysis are those that contain the specified SNP or gene. If an analysis contains no references to the SNP or gene, no records are returned for that analysis.

To define a Free Text keyword search:

1. Select **Free Text** in the category dropdown control.
2. Type the entire keyword (consisting of one or more words) and press **Enter**.

For example, after you type the following Free Text keyword and press Enter, transSMART searches for data sources containing "primary breast tumor" but not those containing just "breast tumor":



Using the Filter Browser

The Filter Browser lets you select one or more search filters to include in your search query.

transSMART adds all of your search filters, including those you type into the [keyword search](#) field (page 5), into the [Active Filters](#) area (page 7).

To work with the Filter Browser:

1. Click the filter icon in the **Active Filters** box:



A list of all filters from which you can select appears.

2. Click the down arrow next to a filter type to see all available entries for that type.
3. Select the filter(s) you want to use from the list. You can select more than one filter from the same filter type or filters from different types.

All selections appear in the **Active Filters** pane, along with any search keywords you might have specified.

Managing Active Filters

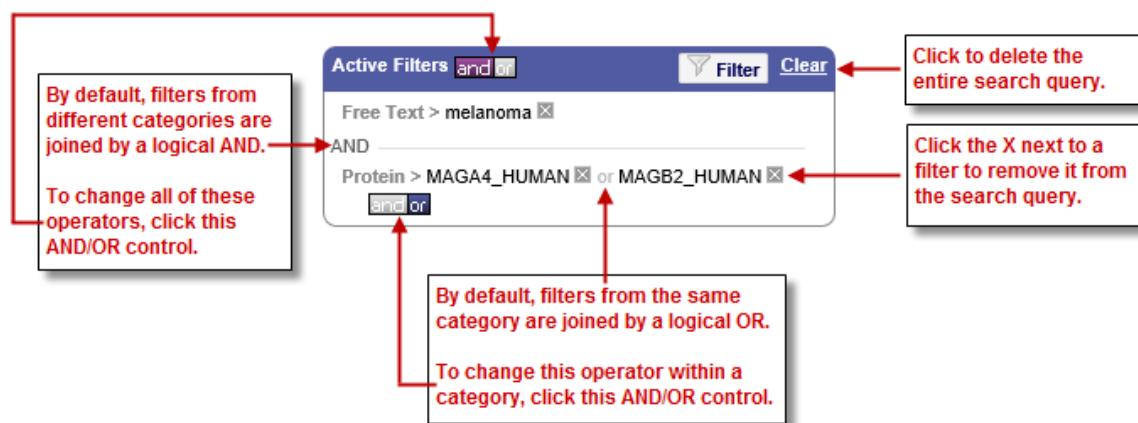
The Active Filters area displays the entire search query that you build using the [keyword search feature](#) (page 5) and/or [filter browser](#) feature (page 6).

Each filter that you define is added to the search query. Each time you add a filter to the search query, the result set in the right side of the Browse page is modified to satisfy the entire search query.

The following search query in Active Filters will return data sources involving melanoma and the human protein MAGA4 or MAGB2:



Note the following controls in the Active Filters pane:



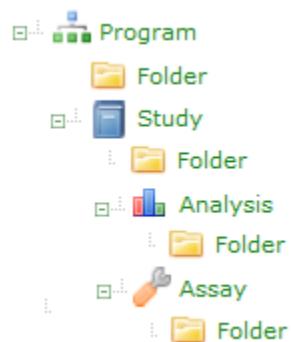
Viewing Studies in the Program Explorer Tree

The Program Explorer displays the results of the search query in the Active Filters box. As the search query changes, the contents of the Program Explorer changes along with it.



Administrators only: For information on creating and editing the objects in the Program Explorer, see [Browse Tool Administration](#) on page 157.

The following illustration shows the hierarchy of objects in the Program Explorer tree. Note that each node in the tree is associated with an icon that represents the type of object at that node:



Program is the top-level component of the hierarchy whose purpose is to group related studies together. Most of the time a program is defined by a molecular target, but it may also be a disease or a pathway.

Study is a collection of subjects on which one or several assays were performed. It can be a clinical trial, a preclinical study, or a discovery experiment.



Icons for studies that can be opened in Analyze view are designated by a yellow star () on the icon.

Assay is an investigative procedure for qualitatively or quantitatively assessing the amount or functional activity of an entity. An assay is defined by a unique experimental protocol.

Analysis is a result obtained by analyzing data from a study. In most cases, an analysis is a signature; that is, a list of molecular entities affected by a particular experimental condition or phenotype.

Folders contain one or several files with information about the associated program, study, analysis, or assay.

Click an object name to view information about the object in the right pane of the Browse window.

Viewing and Exporting Files in Browse Folders

Folders are used to store any type of files. You can view and download files.

To view a file in a folder:

1. Click the folder name in the Program Explorer tree.

Information about the folder appears in the right panel, including a list of files in the folder.

2. Click the name of the file to view.

Files of common formats (such as .doc and .pdf) will be displayed in the corresponding application.

To export files:

1. Click the **Add to Export** button at the far right of a file to export, or click **Export all** to export all files in the folder:

Associated Files			
File Name	Created on	Updated on	
GSE4382-GPL180_series_matrix.txt.gz	2014-09-09	2014-09-09	Add to export
GSE4382-GPL2776_series_matrix.txt.gz	2014-09-09	2014-09-09	Add to export
GSE4382-GPL2777_series_matrix.txt.gz	2014-09-09	2014-09-09	Add to export
			Export all

The files are not exported immediately. Instead, they are added to the Export Cart.

2. Optionally, as you continue to work in the Browse window, add files from other folders to the Export Cart.
 3. When finished adding files to the Export Cart and ready to export the files, click the **Export Cart** button at the top of the Browse window:



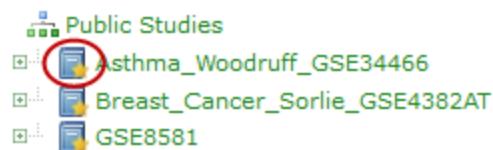
4. In the Export Files dialog box, click **Export Selected Files**.

Note the following:

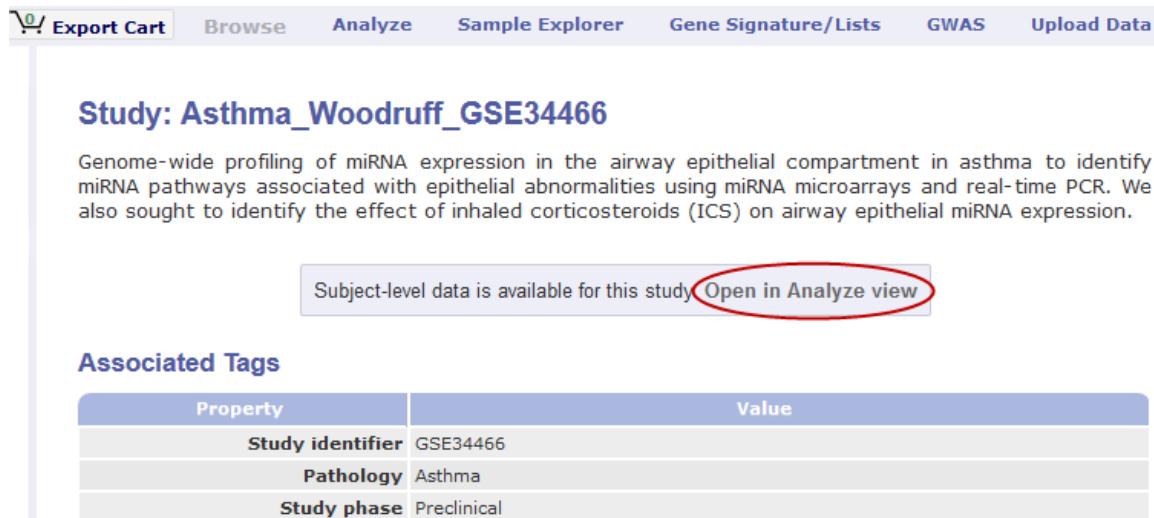
- transSMART formats the file to export as a zip file, assigns it the name `export.zip`, and downloads the file to the `Downloads` directory on your computer.
 - If a file named `export.zip` already exists in the directory, transSMART changes the name to `export-1.zip` (or `export-2.zip`, `export-3.zip`, and so forth, depending on how many files have been exported previously).
 - If multiple files are selected for export on the Export Files dialog box, all are downloaded in one zip file.

Opening a Study in Analyze View

Some studies that have been selected in the Program Explorer tree can be opened in Analyze view. The icon for these studies is displayed with a yellow star (), as shown below:



To open these studies in Analyze View, first click the study in the Program Explorer tree, then click the **Open in Analyze view** button as shown below:



The screenshot shows the transSMART interface with the following elements:

- Top navigation bar: Export Cart, Browse, Analyze, Sample Explorer, Gene Signature/Lists, GWAS, Upload Data.
- Main title: Study: Asthma_Woodruff_GSE34466.
- Description: Genome-wide profiling of miRNA expression in the airway epithelial compartment in asthma to identify miRNA pathways associated with epithelial abnormalities using miRNA microarrays and real-time PCR. We also sought to identify the effect of inhaled corticosteroids (ICS) on airway epithelial miRNA expression.
- Text: Subject-level data is available for this study. **Open in Analyze view** (this text is circled in red).
- Section: Associated Tags.
- Table: Shows study properties.

Property	Value
Study identifier	GSE34466
Pathology	Asthma
Study phase	Preclinical

transSMART displays the Comparison tab of the Analyze window and opens the study you were just viewing in the Browse window.

In both the Analyze and Browse windows, note that the study has been added to the Active Filters pane, and that the results of the search query are now restricted to that single study.

Opening a Study in Analyze View

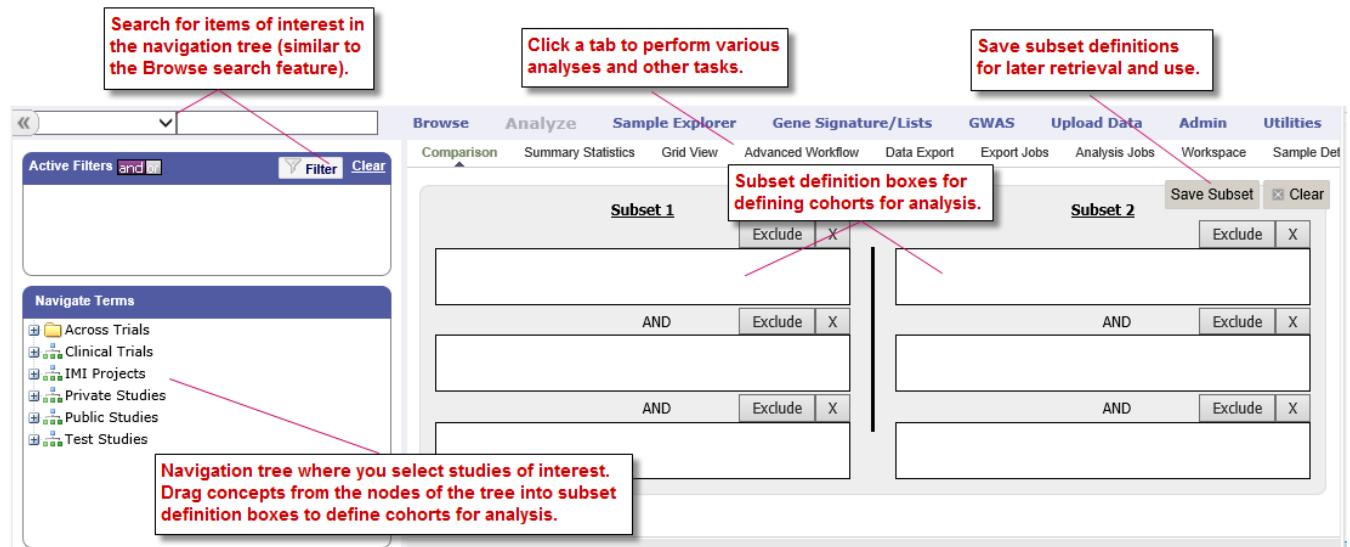
Chapter 3

Analyze

Analyze lets you compare data generated for test subjects in two different cohorts, based on criteria and points of comparison that you specify. Analyze is useful to help you test a hypothesis that involves the criteria and points of comparison that you select.

Overview of the UI

The following figure shows key areas of the Analyze interface:



The page is divided into two panes:

Left pane

The Left pane provides a Microsoft Windows Explorer-like navigation tree where you select the criteria for membership in the cohorts and the points of comparison between the cohorts.

In this pane, select the study of interest and the concepts for defining cohorts for analysis.

Right pane

The Right pane lets you define the criteria that test subjects must satisfy to become members of one of the two cohorts being compared. Each of these cohorts is called a *subset* because it typically contains only some of the subjects involved in the study.

You define the criteria for the subsets in the subset definition boxes. Subjects who do not satisfy the criteria you define are excluded from the subsets and therefore, from the analysis.

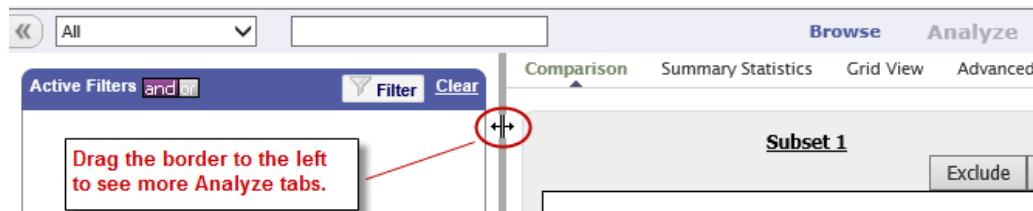
The following table describes the tabs and buttons in the right pane of Analyze:

Tab or Button	Use this Tab or Button To...
Comparison tab	<p>Display the subset definition boxes.</p> <p>Click this tab from any other view to remove the currently displayed view (for example, Summary Statistics view or Grid view) and redisplay the subset definition boxes. This allows you to further refine the subsets for the comparison.</p> <p>See Selecting Criteria on page 21.</p>
Summary Statistics tab	<p>Generate the following:</p> <ul style="list-style-type: none"> ▪ Tables and charts that provide information about the subjects in the subsets. ▪ Analyses of criteria included in the subset definitions. <p>The tables and charts are displayed in Summary Statistics view.</p> <p>See Generating Summary Statistics on page 35.</p>
Grid View tab	<p>Display the comparison and analysis data in grid format.</p> <p>See Viewing Analysis Data in Grid View on page 43.</p>
Advanced Workflow tab	<p>Display advanced analyses and visualizations of the data — for example, heatmaps, scatter plots, survival analyses, and many others.</p> <p>See Chapter 5: Advanced Workflow Analyses.</p>
Data Export tab	<p>Select data to export for further analysis in an external tool.</p> <p>See Exporting Cohort Data on page 30.</p>
Export Jobs tab	<p>Display previously exported jobs.</p> <p>See The Export Jobs List on page 31.</p>
Analysis Jobs tab	<p>Review analyses you have run previously and view the status of analyses you have chosen to run in the background.</p> <p>See Viewing Recent Analysis Jobs on page 90.</p>
Workspace tab	<p>Perform actions related to a saved cohort definition for an analysis, including restoring the saved cohort definition back into the subset definition boxes.</p> <p>See Retrieving Saved Subset Definitions on page 28.</p>
Sample Details tab	<p>View information about the selected data in the Sample Explorer.</p> <p>See Viewing Sample Data on page 34.</p>

Tab or Button	Use this Tab or Button To...
Galaxy Export tab	Export data directly to the Galaxy platform. See Exporting Data Directly into Galaxy on page 32.
Genome Browser tab	View data in the Dalliance Genome Browser. See Dalliance Genome Browser on page 93.
MetaCore Enrichment Analysis tab	Provide enrichment of a gene list to evaluate the significance of the genes to the studied phenotype and/or patient cohort. See MetaCore Enrichment Analysis on page 95 .
Save Subset button	Save the subset definition. This allows you to regenerate the comparison at a later time without having to reconstruct the criteria used in the comparison. See Saving Subset Definitions on page 26.
Clear button	Clear all data that has been specified by the user, including the data in the subset definition boxes and in the Advanced Workflow variable input boxes.



If the tabs at the rightmost end of the tab bar are not visible, try shrinking the left pane by dragging the vertical border to the left:



Using Analyze — Basics

Three basic tasks are involved in using Analyze:

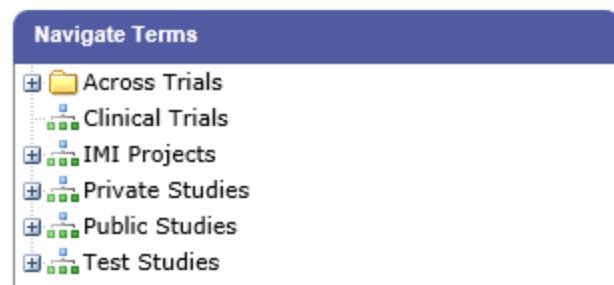
- Identify the study to include in the comparison. Using the Across Trials folder, multiple studies can be included in the comparison.
- Specify the criteria for membership in the two cohorts. Note that some analyses in Advanced Workflow only allow for the specification of one cohort at this time.
- Select the function to perform, such as Summary Statistics, Grid View, or Advanced Workflow, from the tab bar.



You may see the notations **NA** and **Unknown** in the study data. **NA** indicates not applicable and **Unknown** indicates not available.

The Navigation Tree

The Analyze navigation tree is located in the Navigate Terms pane:



The navigation tree looks and works much like Microsoft Windows Explorer. Windows Explorer is a hierarchy of folders, sub-folders, and files. The navigation tree is a hierarchy of folders and sub-folders (the branches) and values (the leaves) that reflect aspects of the trial, such as research metrics, compounds used, and patient demographics.

In Analyze, all levels of the tree, including branches and leaves, are referred to as nodes.

The following figure shows typical top-level nodes of a study. Some studies may not require all of these nodes, and others may require additional nodes:

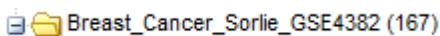


The nodes you see on your screen may differ from those listed here. Only those domains present in your data will appear in your navigation tree.

Visual Cues in the Navigation Tree

Each concept node in the navigation tree displays the following information about the concept:

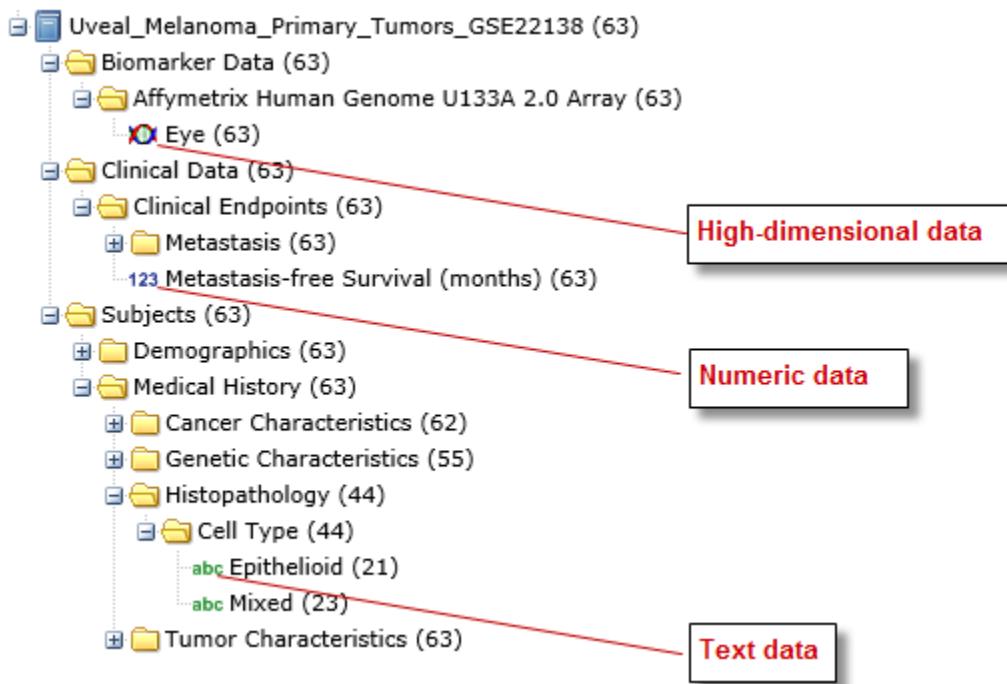
- The numbers in parentheses at nodes of the tree indicate the number of subjects to whom that node applies. For example, in the figure below, there are a total of 167 subjects in the study:



Nodes within the Across Trials folder do not indicate the number of subjects associated with the node.

- In tranSMART, data values are represented in one of three ways: by number, by text, or by high dimensional data (SNP, gene expression, etc.) stored as arrays.

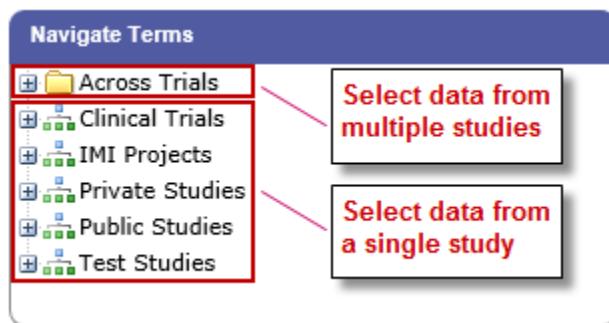
The three types of data values and their associated icons are illustrated below:



Selecting Studies for Analysis

Select studies for analysis in the navigation tree, located in the Navigate Terms section of the left pane of the Analyze window.

You can select data from a single study, or you can select data from multiple studies located in the Across Trials folder.



To select a study, click the + icon (⊕) next the study name:



You can then drill down into the study to find the data to use to define the cohorts for the analysis.

Searching for a Study

You define search filters with the Analyze tool as you do with the Browse tool. For information, see [Defining Search Filters](#) on page 4.

Selecting Data from Multiple Studies in the Across Trials Folder

The Across Trials folder is a special folder that contains data from multiple studies. You defined cohorts from this folder in the same way that you define cohorts from a single-study folder.

Common categories of data from the multiple studies are loaded into the same nodes of the tree; for example, the Female node contains female subjects across all the studies that are loaded into the Across Trials folder.

Use the Across Trials folder to include data from multiple studies in your analysis. For example, you may want to determine whether age at diagnosis is correlated with survival in breast cancer patients, regardless of which study the subjects participated in.

Structure of the Across Trials Tree

The data in the Across Trials folder needs to be curated so that each study has the same hierarchical folder structure and naming conventions. Only those folders and values with the same names will be displayed in the Across Trials folder.

Public and Private Studies

Analyze studies can be either public or private. Public studies can be found in both the **Public Studies** folder of the Analyze navigation tree and in the research-specific folders.

You can perform all the operations described in this chapter on public studies. No special privileges are required.

To perform operations described in this chapter on a private study, a transSMART Administrator must assign you access rights to the study. Access rights are based on the following access levels:

Access Level	Privileges
VIEW	Define the criteria for the cohorts to be compared, generate summary statistics for the cohorts, and specify points of comparison for the cohorts.
EXPORT	All privileges of the VIEW access level, plus the ability to export comparison data or expression data to a Microsoft Excel spreadsheet.
OWN	All VIEW and EXPORT privileges. This access level can only be assigned to the owner of the study.

If you do not have access rights to the study you want (that is, if the study is grayed out), contact a transSMART Administrator. The administrator will contact the study owner to find out if you should be granted VIEW access, EXPORT access, or no access.



Even if you have no access rights to a private study, you can read a description of the study. For information, see [Viewing a Study Description](#) on page 19.

Viewing a Study Description

You can view a description of any Analyze study, whether or not you have access rights to the study.

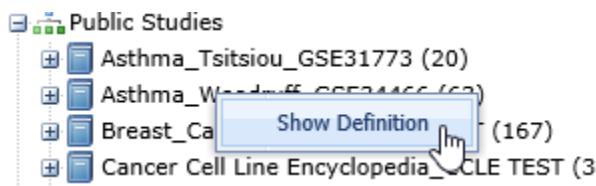
To view a description of a study:

1. In Analyze, open the top level node for the list of studies you are interested in; for example, click the + icon (⊕) next to Public Studies to open the list of public studies:



2. Right-click the particular study you are interested in.

3. Click the **Show Definition** popup:

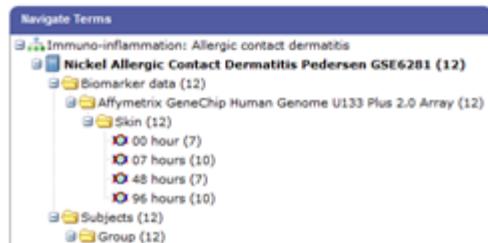


The Show Concept Definition dialog box appears, showing the title, description, and other information about the study.

Serial Numeric Data

transSMART supports serial numeric data (high or low dimensional); that is, a numeric variable that has been measured in a series of conditions for each subject (for example, several timepoints). The conditions cannot be specific to each subject but are shared by all subjects; for example, a measurement performed at 0, 7, 48, and 96 hours for the various subjects.

In the Analyze navigation tree, serial data is represented by several leaves of the same type in a folder, with each leaf representing a condition with a label; for example:



In the transSMART database, each condition can be described by a numeric value (such as for time series or dose response) or by a categorical value (such as in the case of a series of tissues derived from each subject).

When the value characterizing each sample is numeric, it is also associated with a unit. In the case of time series, for example, the value associated with each sample will be time duration, and the unit can be hours (a single unit is used for the complete series).

In Analyze, serial data specificities can be best exploited using Line Graph and Heatmap.

Defining the Cohorts

You define the cohorts for an analysis by selecting criteria that members of each cohort must satisfy. For example, cohort members might be required to satisfy a weight or age requirement. Analyze lets you build a set of criteria for each cohort that can be as simple or as complex as you need.

The cohorts you define are called *subsets*. Typically, after your criteria are applied, the members of a resulting cohort are a subset of all the subjects that participated in the study.

Selecting Criteria

To define a cohort, select criteria (called *concepts*) from a study in the navigation tree and drag them into the subset definition boxes. With studies in the Across Trials folder, concepts include data from multiple studies.

Linked event data, non-linked event data, and NGS data can all be used to populate the cohorts.

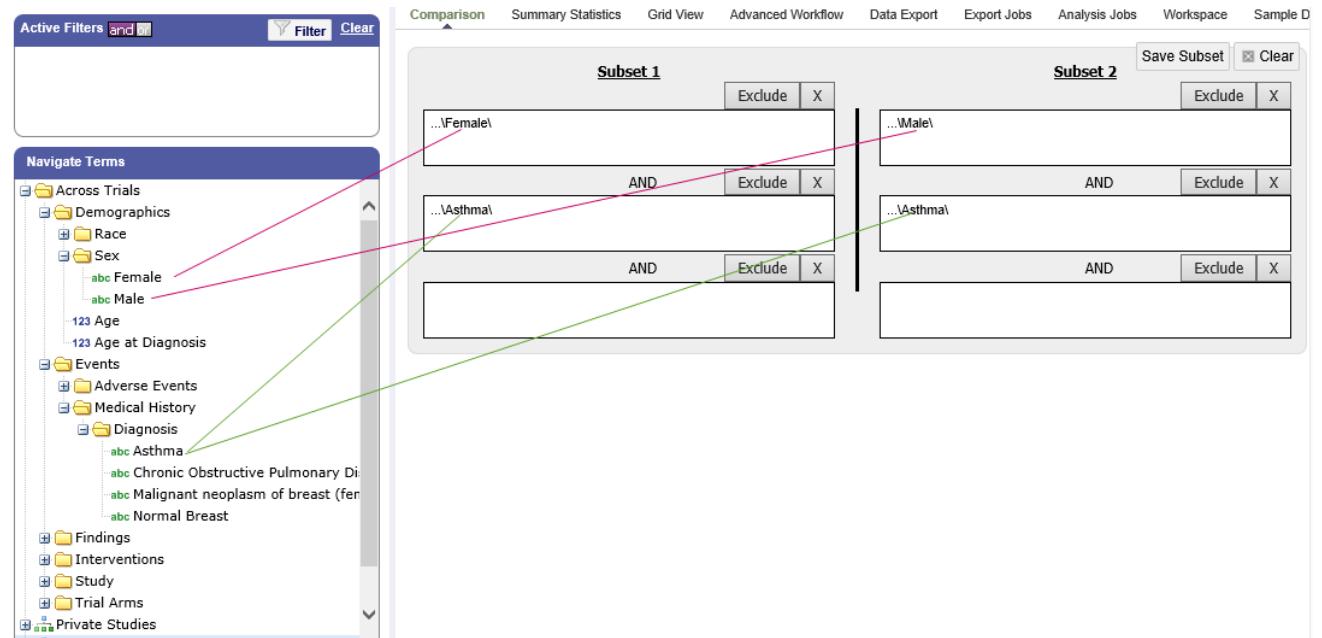
Single Study Example

In the following example from a single asthma study, female patients have been dragged into Subset 1 and male patients into Subset 2:

The screenshot shows the Analyze software interface for defining cohorts. On the left, there is a navigation tree titled "Asthma_Tsitsiou_GSE31773 (20)". Under "Subjects", there are "Demographics (20)" and "Gender (20)". Under "Gender (20)", there are "Female (13)" and "Male (7)". A red arrow points from the "Female (13)" node in the navigation tree to the "Subset 1" box in the main panel. Another red arrow points from the "Male (7)" node in the navigation tree to the "Subset 2" box in the main panel. The main panel has tabs for "Comparison", "Summary Statistics", "Grid View", "Advanced Workflow", "Data Export", "Export Jobs", "Analysis Jobs", "Workspace", and "Sample ID". The "Subset 1" panel contains the text "...Female" and an "Exclude" button. The "Subset 2" panel contains the text "...Male" and an "Exclude" button. There are also "Save Subset" and "Clear" buttons at the top of each panel.

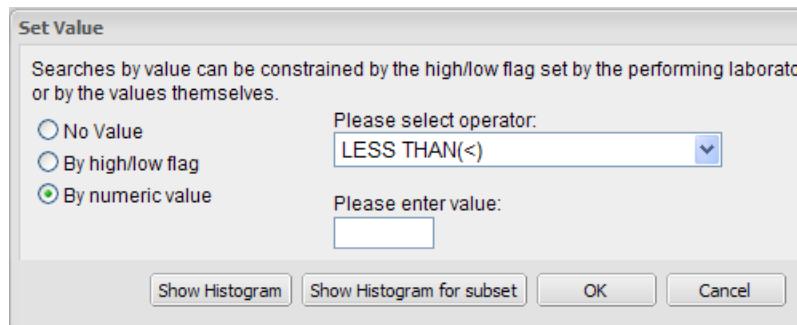
Across Trials Example

In the following example, males and females from the studies loaded into the Across Trials folder have been dragged into Subsets 1 and 2. However, because the concept Asthma has also been dragged into both Subset 1 and Subset 2, the cohorts include only males and females from the asthma studies in the Across Trials folder, not males and females from any of the other studies in the Across Trials folder.

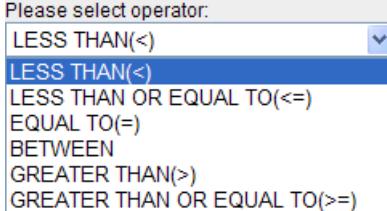


Specifying a Numeric Value

When you drag a numeric concept into a subset definition box, the Set Value dialog box appears:

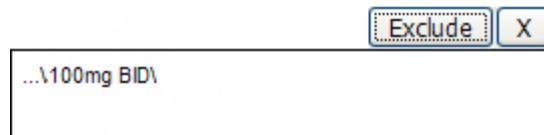


Use the Set Value dialog to specify how you want to constrain the numeric values to use in the subset definition. To do so, first select one of the following choices:

Selection	Description
No Value	<p>Values are not constrained. All the numeric data associated with the concept are factored into the subset definition.</p> <p>If you select No Value, no other information is required. Click OK to add the concept with all its associated numeric data to the subset.</p>
By high/low flag	<p>If the data was grouped into high/low/normal ranges during curation and loading, it is possible to select the range to factor into the subset definition.</p> <p>When you select By high/low flag, the Please select range field appears. Select the range you want and click OK.</p>
By numeric value	<p>Values are constrained by an exact value or a range of values.</p> <p>After you select By numeric value:</p> <ul style="list-style-type: none"> ▪ Select one of the following numeric operators in the Please select operator dropdown:  <ul style="list-style-type: none"> ▪ In Please enter value, type the numeric value that the operator applies to. <p>For example, to constrain the ages of subjects to 50 years or younger, select LESS THAN OR EQUAL TO(≤) in the dropdown, then type 50 in the Please enter value field.</p> <ul style="list-style-type: none"> ▪ Click OK. <p>See the next section for information on viewing the numeric values associated with the concept and that you can select from.</p>



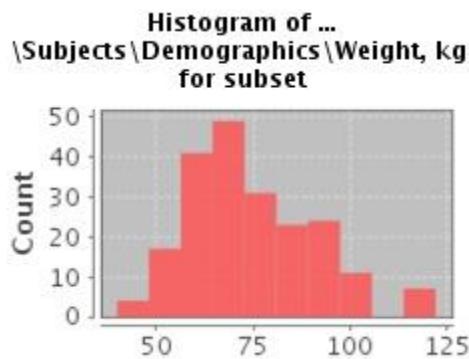
When finished defining the numeric constraint on the Set Value dialog, be sure to click **OK** and not press the **Enter** key. Pressing **Enter** will activate the subset button that has focus — the **Exclude** button in the example below:



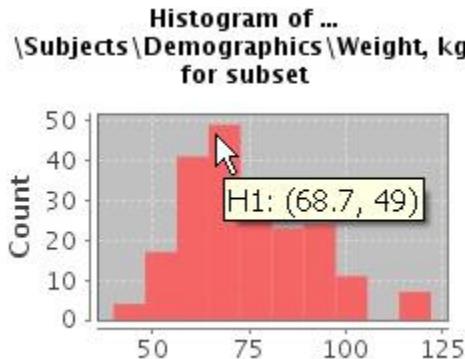
Viewing the Numeric Values Associated with a Concept

Note the buttons **Show Histogram** and **Show Histogram for subset** in the Set Value dialog. The histograms show how the numeric values associated with the concept that you placed in the subset box are distributed among the subjects across both subsets, or in the particular subset you are currently defining, respectively.

A histogram may be helpful in determining the number to set as the constraining factor for a concept. For example, suppose you drag a Weight concept into a subset box, then click **Show Histogram for subset**. In the following histogram of the weights of test subjects, the weights range from about 25 kg to just under 125 kg. The largest bin represents fewer than 50 subjects. You may want to use these weight parameters to help you determine the value to set for the weight concept.



You can get more specific information about the number of subjects represented by a particular bin and the average of the values in the bin by hovering the mouse cursor over the bin you are interested in. For example, in the following figure, the largest bin represents 49 subjects with an average weight of 68.7 kg:



Joining Multiple Criteria for a Subset Definition

Multiple criteria for a subset definition are joined by one of the following logical operators: AND, OR, or AND NOT.

The rules for joining multiple criteria are as follows:

- Criteria in separate subset definition boxes are joined by an AND operator.

For example, the following definition boxes select only male subjects, AND males whose weights are between 65 kg and 90 kg:

The image shows a 'Subset 1' dialog box. It contains two separate definition boxes. The top box has the text '...!Male' and an 'Exclude' button. The bottom box has the text '...!Weight between 65 and 90' and an 'Exclude' button. Between the two boxes is the word 'AND'.

- Criteria within the same subset definition box are joined by an OR operator.

For example, to use the extreme ends of the weight scale for your weight criterion, you might add the following to a definition box:

The image shows a 'Subset 1' dialog box containing a single definition box with the text '...!Weight <=50' and '...!Weight >=100'.

These criteria select subjects whose weight is either 50 kg or less, OR 100 kg or greater.

- To join a definition box with an AND NOT operator, click the **Exclude** button above the definition box.

The figure below selects only male subjects, but not those who weigh between 50 kg and 100 kg:

The image shows a 'Subset 1' dialog box containing a single definition box with the text '...!Male' (with an 'Exclude' button) and '...!Weight between 50 and 100' (with an 'Include' button).

Note that when you click the **Exclude** button, the button label changes to **Include**, allowing you to join the criteria in the box with an AND operator later if you choose.

Modifying or Deleting Criteria

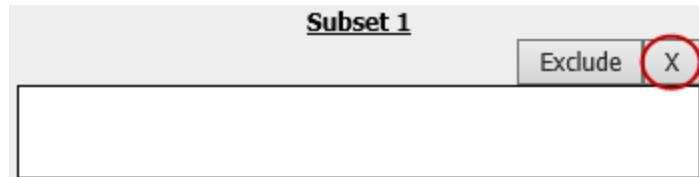
To delete or modify a criterion in a subset definition box, right-click the criterion and select either **Delete** or **Set Value**.



Set value displays only when the criterion is a numeric value.

Show Definition displays for any type of criterion. Use this option to review the node before modifying or deleting it.

To remove the entire contents of a subset definition box from the subset definition, click the **X** icon (above the box:

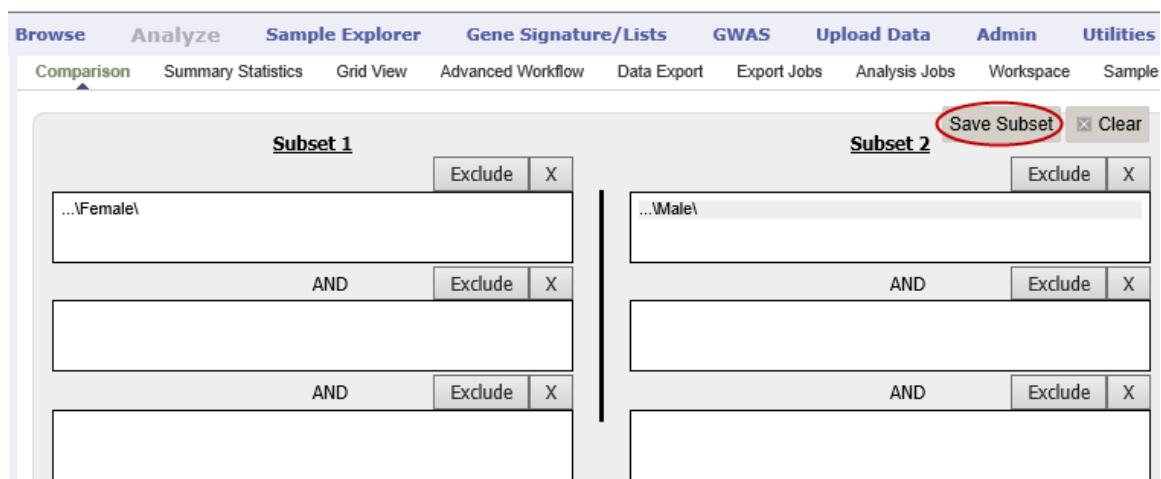


Saving Subset Definitions

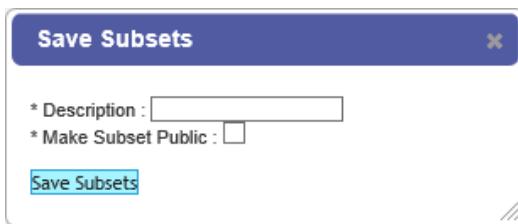
You can save your subset criteria in order to regenerate the subsets at a later time without having to define the criteria again.

To save a subset definition:

1. In **Analyze**, select a study of interest.
2. Define the cohorts whose data points will be represented.
3. Click the **Save Subset** button to save the criteria:



The Save Subsets dialog box appears:



4. Enter a description of the subsets in the **Description** field.
5. Optionally, clear **Make Subset Public** to make this subset available only to yourself:
 - If the subset is public**, all others are able to view it.
 - If the subset is not public**, only the user who created it can view it.
6. Click **Save Subsets**.

The subset information displays immediately in the Workspace tab in the **Subset Manager** portion of the Workspace page:

Subset Manager											Search: <input type="text"/>	Show <input type="button" value="10"/> entries
Description	Study	Query	Use	Email	Link	Created by	Delete	Public	Create Date			
4382						guest			01-19-2015			
Cell line						admin			08-20-2014			
gse34466 gender						admin			01-26-2015			
wsc						admin			01-22-2015			

For information about the Workspace tab, including retrieving saved subsets, see [Retrieving Saved Subset Definitions](#) on page 28.

Retrieving Saved Subset Definitions

The **Workspace** tab of the Analyze window is where a saved subset definition can be retrieved.

To retrieve a saved subset definition, click the corresponding radio button in the **Use** column:

Subset Manager										
Description	Study	Query	Use	Email	Link	Created by	Delete	Public	Create Date	
4382			<input type="radio"/>			guest			01-19-2015	
Cell line			<input type="radio"/>			admin			08-20-2014	
gse34466 gender			<input checked="" type="radio"/>			admin			01-26-2015	
wsc			<input type="radio"/>			admin			01-22-2015	

The retrieved subset definition remains in the Subset Manager until you explicitly delete it.

For information on saving a subset definition, see [Saving Subset Definitions](#) on page 26.

Subset Manager Overview

The following table describes the features of the Subset Manager:

Column	Description
Search	In this field, type one or more characters of a subset definition description. As you type, transSMART refines the list to include only the studies that match what you type.
Show <i>n</i> entries	Specify the maximum number of studies to include in a single page of the list.
Description	The description provided for the subset when saved. Also: <ul style="list-style-type: none"> ■ Click the pencil icon to edit the subset definition description. Only the user who created the subset definition can edit the description. ■ Click the arrow icon next to Description to sort the list alphabetically by the descriptions.
Study	The study ID. Click the arrow icon next to Study to sort the list by study IDs.
Query	Hover the mouse pointer over to review a saved subset definition without returning to the Comparison tab.

Column	Description
Use	<p>Click the Use radio button to populate the subset definition boxes on the Comparison tab with the saved criteria, then click OK to acknowledge the message that any existing criteria in the subset definition boxes will be overridden.</p> <p>After you click OK, the Comparison tab appears with the subset boxes populated with the saved criteria.</p>
Email	<p>Click the Email icon to email the saved subset definition to yourself and colleagues, as appropriate.</p>
Link	<p>Click the Link icon to see the URL of a subset definition.</p>
Created by	<p>The username of the person who created the subset definition.</p> <p>Click the arrow icon next to Created by to sort the list by usernames.</p>
Delete	<p>Click the Delete icon to delete this subset definition from the Subset Manager list and transSMART.</p> <p>Note: Only the user who created the subset definition can delete it.</p>
Public	<p>Indicates whether the subset definition will be accessible by others or only by the person who created and saved the subset definition or by an administrator. The Public setting is the default when the subset definition is saved.</p> <ul style="list-style-type: none"> ■ Public (): Accessible by the user who saved the subset definition and others. ■ Private (): Accessible only by the user who saved the subset definition. <p>Note: If a subset is based on a study that a user does not have sufficient privileges to see, the user will not be able to restore the subset definition to the subset definition boxes. Seeing a saved subset definition does not grant new privileges to users for the associated study.</p>
Create Date	<p>The date the subset definition was created and saved.</p> <p>Click the arrow next to Create Date to sort the list by date.</p>
First/Previous/ Next/Last	<p>Navigate through the pages of a multi-page list.</p>

Exporting Cohort Data

You can export data for one or both cohorts by defining the cohort(s) and clicking the **Data Export** tab. You can either download the data immediately after the export, or you can run the export in the background and download the data at a later time from the **Export Jobs** tab.

Downloaded data is saved to a location you specify in tab-separated format. Export metadata (information about the cohort definition and filters that selected the data to export) is downloaded in a separate file from the data itself.

To export data to your local machine or a network location:

1. Define one or both cohorts as described in [Defining the Cohorts](#) on page 21.
 2. Click the **Data Export** tab.
- The Data Export page appears with your selected cohorts.
3. Optionally, drag additional nodes from the study into the export criteria to filter the data to export:

Subset 1	
Selected Cohort	(Public Studies\GSE4698\Subjects\Demographics\Sex\Female\) AND (Public Studies\GSE4698\Samples and Timepoints\Sample Source\Bon
Clinical & Low Dimensional Biomarker Data (Drag and drop low dimensional nodes here to filter the exported data.)	Data is available for 18 patients Export (.TXT) <input checked="" type="checkbox"/>
Messenger RNA data (Microarray) (Drag and drop high dimensional nodes here to filter the exported data.)	Tab separated file is available for Affymetrix Human Genome U133A Array: 18 patients Export (.TSV) <input type="checkbox"/>

Because some studies have hundreds of concepts associated with each patient, adding one or more filters allows you to limit the exported data to only you need to work with.

4. Select the checkbox for the type of data to export:

Subset 1	
Selected Cohort	(Public Studies\GSE4698\Subjects\Demographics\Sex\Female\) AND (Public Studies\GSE4698\Samples and Timepoints\Sample Source\Bon
Clinical & Low Dimensional Biomarker Data (Drag and drop low dimensional nodes here to filter the exported data.)	Data is available for 18 patients Export (.TXT) <input checked="" type="checkbox"/>
Messenger RNA data (Microarray) (Drag and drop high dimensional nodes here to filter the exported data.)	Tab separated file is available for Affymetrix Human Genome U133A Array: 18 patients Export (.TSV) <input type="checkbox"/>

Above, only clinical and low dimensional data is being exported.

5. Click the **Export Data** button at the bottom of the page.

6. Do one of the following:

- When the export completes, download the data to your PC or a network location.
- With a large data set, click the **Run in Background** button on the Job Status dialog box. You can download the data at a later time from the **Export Jobs** tab.
- Optionally, click the **Cancel** button to cancel the export.

Both exported jobs and canceled jobs appear listed on the Export Jobs tab. Jobs remain listed on this tab for seven days. See [The Export Jobs List](#) on page 31 for information about this list.

The Export Jobs List

A list of all exported jobs over the last seven days is displayed when you click the **Export Jobs** tab. The list includes all jobs: successes, errors, and pending jobs.

Browse	Analyze	Sample Explorer	Gene Signature/Lists	GWAS	Upload Data	Admin	Util
Comparison	Summary Statistics	Grid View	Advanced Workflow	Data Export	Export Jobs	Analysis Jobs	Workspace
<hr/>							
Name	Query Summary	Status	Started On				
admin-DataExport-100202		Completed	2015-02-05 21:35:36.991				
admin-DataExport-100183		Completed	2015-02-05 20:03:27.631				
admin-DataExport-100182		Completed	2015-02-05 19:34:00.402				
admin-DataExport-100149		Completed	2015-02-05 21:36:01.767				

The list contains the following columns:

Column	Description
Name	The name of the export job. Jobs use the naming convention: <i>User - Type of Job Run - Job ID:</i> 
Query Summary	Displays the query that was run to generate the subset.

Column	Description
Status	The status of the export job: <ul style="list-style-type: none"> ▪ Completed — The job has finished and the data is available for download. ▪ Started — The job has been started and is still processing. ▪ Error — The job did not complete due to an error. ▪ Cancelled — The job was cancelled and will not complete.
Started On	The date and time that the export was started.

Exporting Data Directly into Galaxy

If you have the Galaxy data analysis tool installed, you can export cohort data from tranSMART into Galaxy in either of these ways:

- Export the data and download the data files to your local PC or a network location, using the tranSMART **Data Export** and **Export Jobs** tabs, and then open Galaxy and import the data.
- Export the data directly into Galaxy using the **Galaxy Export** tab.

For information about the Galaxy software, see <http://galaxyproject.org/>.



Exporting data into Galaxy using the **Galaxy Export** tab requires both of the following:

- That a tranSMART administrator has associated your tranSMART user ID with a Galaxy key.
- That Galaxy be configured to support exports from tranSMART. See the Galaxy documentation for configuration instructions.

To export data using the Galaxy Export tab:

1. Define one or both cohorts as described in [Defining the Cohorts](#) on page 21.
2. Click the **Data Export** tab and define the data to export, as described in steps 2 through 4 in section [Exporting Cohort Data](#) on page 30.
3. Click the **Export Data** button at the bottom of the page, but do not download the data when prompted to do so.

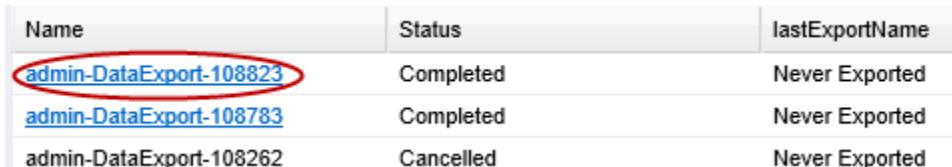
Note that data exports are listed on both the **Export Jobs** tab and the **Galaxy Export** tab.

4. Click the **Galaxy Export** tab:



Name	Status	lastExportName	lastExportTime	exportStatus
admin-DataExport-108823	Completed	Never Exported		
admin-DataExport-108783	Completed	Never Exported		
admin-DataExport-108262	Cancelled	Never Exported		

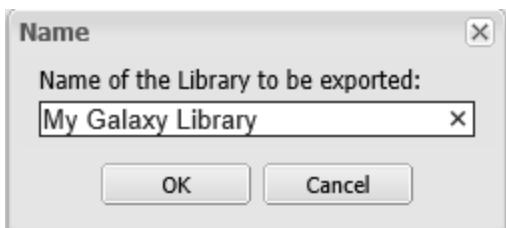
5. When the Status column for the exported data shows **Completed**, click the name of the job to export to Galaxy:



Name	Status	lastExportName
admin-DataExport-108823	Completed	Never Exported
admin-DataExport-108783	Completed	Never Exported
admin-DataExport-108262	Cancelled	Never Exported

The Name dialog box appears.

6. Type the name of the Galaxy data library where the data will be exported, then click **OK**.



7. Click the **Refresh** button at the bottom of the page.

The status of the export is updated as shown below:



Name	Status	lastExportName	lastExportTime	exportStatus
admin-DataExport-108823	Completed	My Galaxy Library	2015-02-11 00:00:00.0	Started
admin-DataExport-108783	Completed	Never Exported		
admin-DataExport-108262	Cancelled	Never Exported		

When the export to Galaxy is complete, the completion status is reflected in the **exportStatus** column.

Viewing Sample Data

If the cohort data includes data that has been loaded into the Sample Explorer, you can view information about the sample data without having to explicitly open the Sample Explorer and searching for the data.

To view sample data for the cohort(s) defined in Analysis:

1. Define one or both cohorts as described in [Defining the Cohorts](#) on page 21.
2. Click the **Sample Details** tab:

The screenshot shows the 'Sample Details' tab selected in the top navigation bar. Below the tabs, there are two sections labeled 'Subset 1' and 'Subset 2'. Each section has a text input field containing a search term ('...breast cancer' in Subset 1), an 'Exclude' button, and a close ('X') button. To the right of the subsets are 'Save Subset' and 'Clear' buttons.

The Sample Explorer opens, displaying any cohort data that has been loaded in the Sample Explorer:

Dataset Explorer Patient Selection							Sample Contact Information	Collapse All	Expand All	
BioBank	Subject Treatment	Source Organism	Data Type	Tissue	Pathology	Aliquot Count				
160	-	human	-	blood	MS	1				
929	-	human	-	brain	MS	1				
292	-	human	-	brain	MS	1				
105	-	human	-	brain	MS	1				
494	-	human	-	brain	MS	1				

For information about this page of the Sample Explorer, see [View and Refine Sample Search Results](#) on page 100.

Chapter 4

Summary Statistics for Analysis

This chapter explains how to review study information and populate the cohorts you use for analyses.

Generating Summary Statistics

When you finish defining criteria for the cohorts to compare — the subsets — click the **Summary Statistics** button.



As an alternative to generating summary statistics, you can view a breakdown of a particular subset by a selected concept (see [View Subset Breakdown by Concept](#) on page 39).

tranSMART displays tables and charts of information that describe the subsets. The information is displayed in Summary Statistics view in the following sections:

- A summary of the criteria used to define subsets to compare.

For example:

Query Summary for Subset 1	Query Summary for Subset 2
(\Public Studies\Public Studies\Lymphoma_Lenz_GSE10846\Subjects\Medical History\Cancer Stage\Stage 4) AND (\Public Studies\Public Studies\Lymphoma_Lenz_GSE10846\Subjects\Medical History\Chemotherapy\CHOP-Like Regimen)	(\Public Studies\Public Studies\Lymphoma_Lenz_GSE10846\Subjects\Medical History\Cancer Stage\Stage 4) AND (\Public Studies\Public Studies\Lymphoma_Lenz_GSE10846\Subjects\Medical History\Chemotherapy\R-CHOP-Like Regimen)

- A table showing the number of subjects in each subset that match the subset criteria.

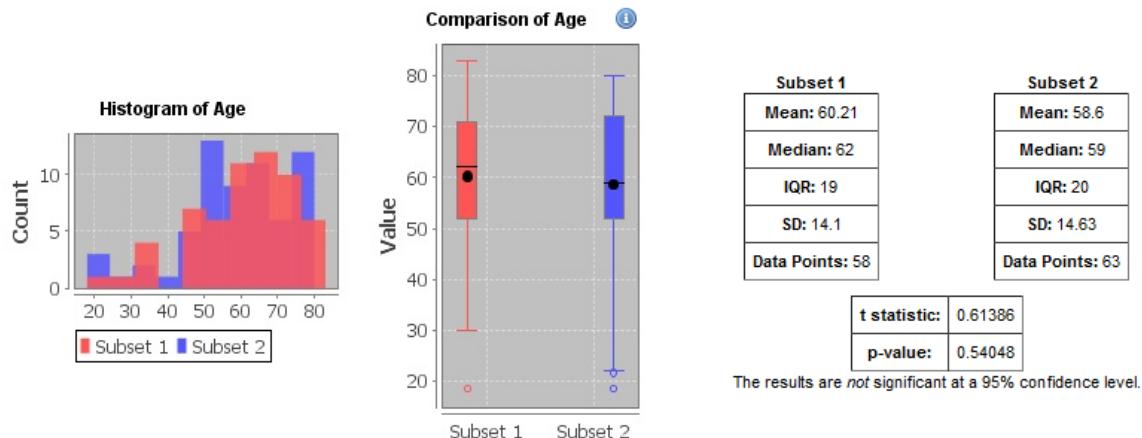
For example:

Subject Totals		
Subset 1	Both	Subset 2
58	0	63

In this example, 58 subjects matched the criteria for Subset 1 and 63 matched the criteria for Subset 2. No (0) subjects matched the criteria for both subsets.

- Tables and charts that show how the subjects who match the criteria fit into age, sex, and race demographics.

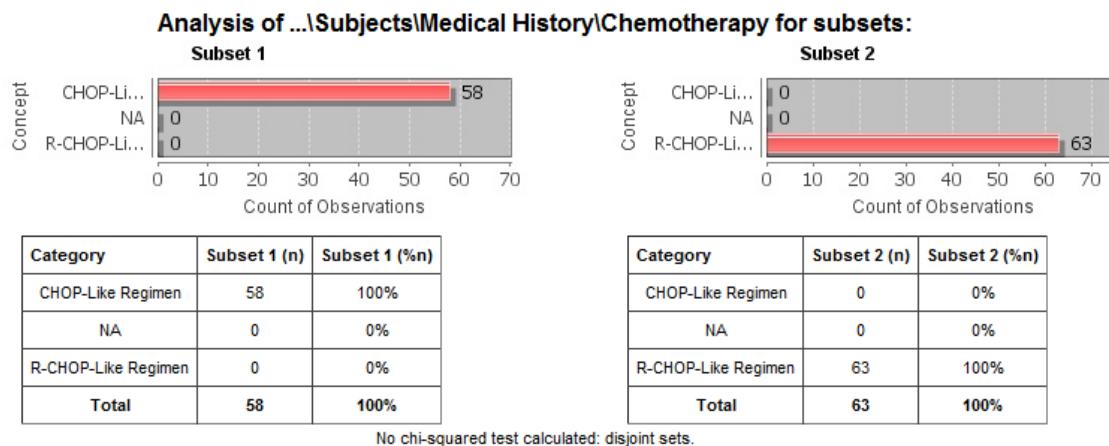
This example shows the age portion of the demographics data only:



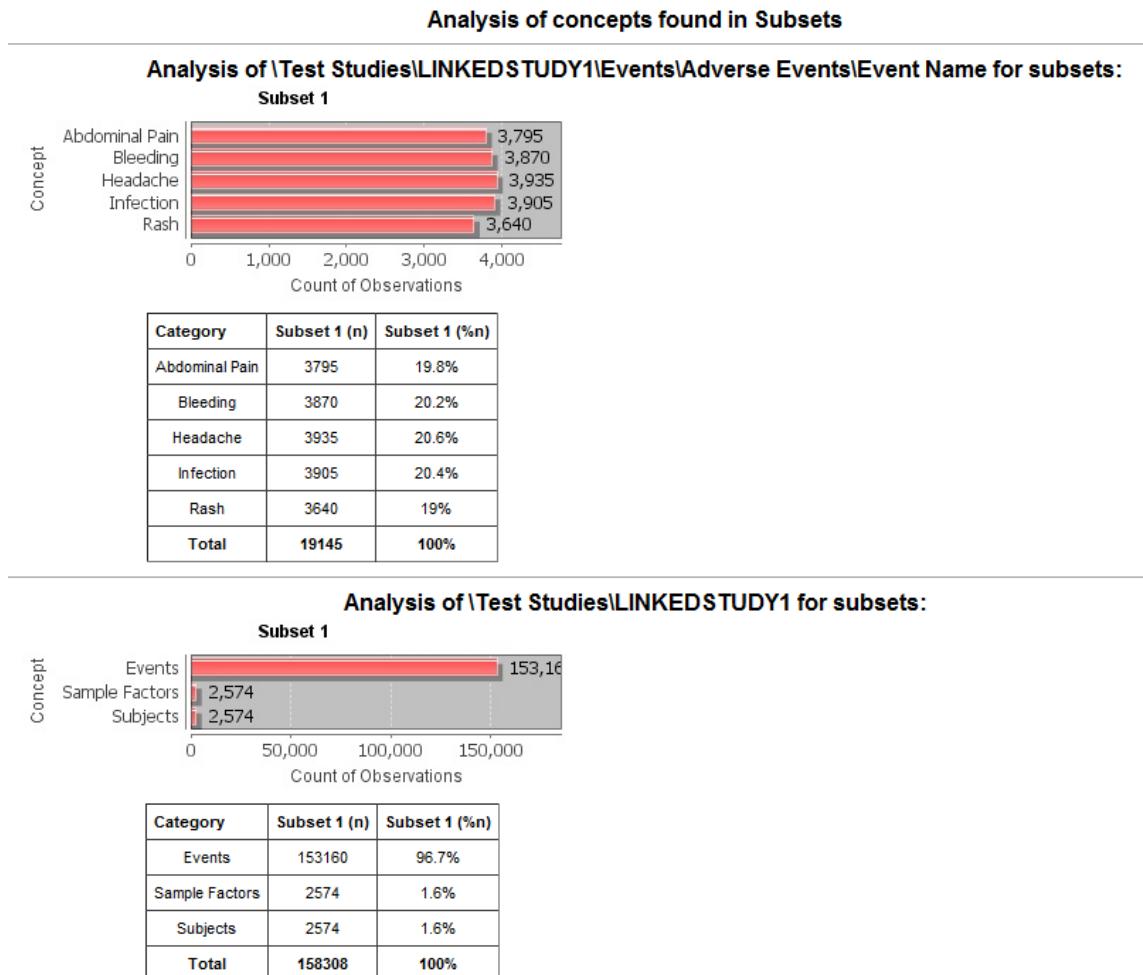
- Analyses of the concepts you added to the subsets from the navigation tree. The data displayed reflects the data used to generate the summary statistics.

The next examples show analysis of concepts for a non-linked event, a linked event, and NGS data.

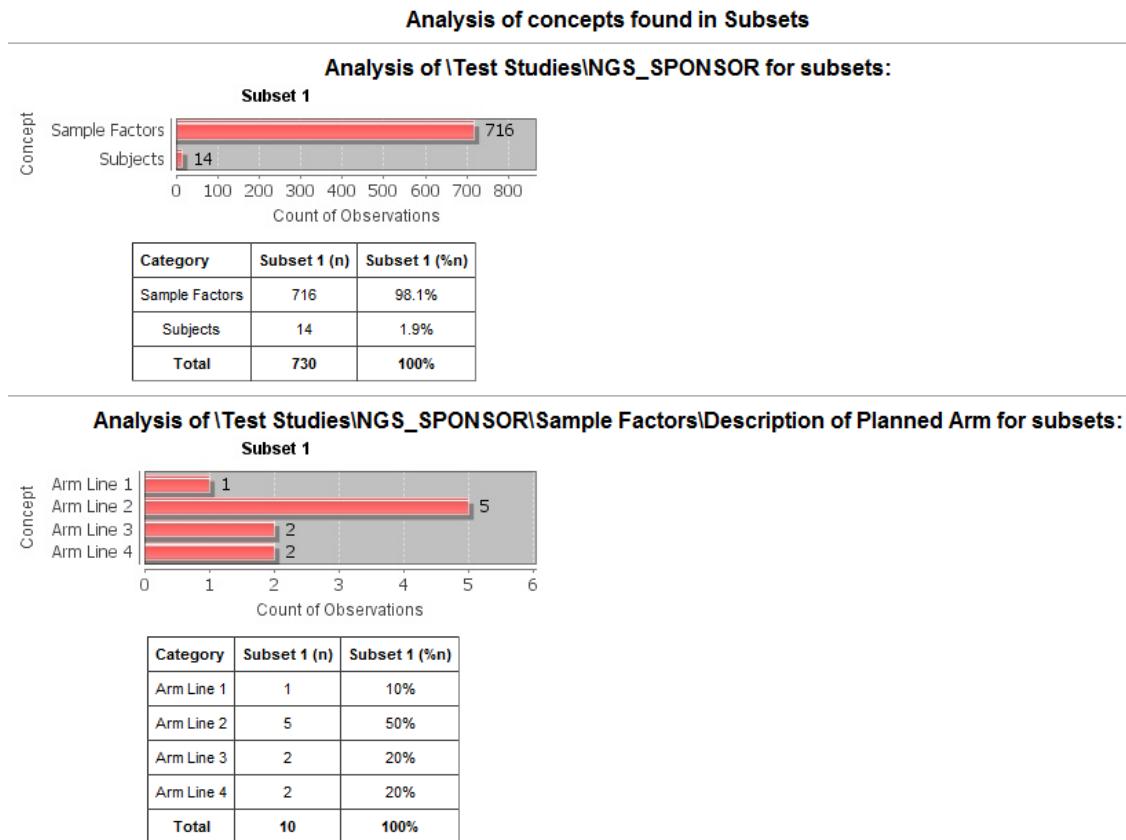
Example 1: Non-linked event. This example shows the analysis of the chemotherapy concept:



Example 2: Linked event. This example shows the analysis of concepts for adverse events:



Example 3: NGS data. This example shows the analysis of concepts for description of planned arm:



Significance Tests

The analyses include the results of significance testing that Analyze performs:

t statistic:	0.61386
p-value:	0.54048

The results are *not* significant at a 95% confidence level.

Significance testing is designed to indicate whether the reliability of the statistics is 95% or greater, based on p-value.

Analyze calculates the significance result using either t-test or chi-squared statistics to determine the p-value:

- For continuous variables (for example, subject weight or age), a t-test compares the observed values in the two subsets.

tranSMART uses the following Java method to calculate the t-test statistic:

[http://commons.apache.org/math/apidocs/org/apache/commons/math3/stat/inference/TTest.html#tTest\(double\[\],%20double\[\]\)](http://commons.apache.org/math/apidocs/org/apache/commons/math3/stat/inference/TTest.html#tTest(double[],%20double[]))

- For categorical values (for example, diagnoses), a chi-squared test compares the counts in the two subsets.

transSMART uses the following Java method to calculate the chi-squared statistic:

[http://commons.apache.org/math/apidocs/org/apache/commons/math3/stat/inference/ChiSquareTest.html#chiSquareTest\(long\[\]\[\]\)](http://commons.apache.org/math/apidocs/org/apache/commons/math3/stat/inference/ChiSquareTest.html#chiSquareTest(long[][]))

If there is not enough data to calculate a test, Analyze displays a message indicating the insufficient quantity of data. In addition, significance test results are not displayed in the following circumstances:

- If two identical subsets are defined. In this case, the significance test results are not meaningful.
- If all subjects in the first subset have one set of values for the categorical value and all subjects in the second subset have other categorical values. For example, suppose you set Subset 1 to contain only males and Subset 2 to contain only females. If you then try to show statistics by gender, tables similar to the following would result:

Category	Subset 1 (n)	Subset 1 (%n)
Female	54	100%
Male	0	0%
Unknown	0	0%
Total	54	100%

Category	Subset 2 (n)	Subset 2 (%n)
Female	0	0%
Male	64	100%
Unknown	0	0%
Total	64	100%

No chi-squared test calculated: disjoint sets.

In this case, the chi-squared function doesn't return meaningful results.

View Subset Breakdown by Concept

Generating summary statistics provides data for all subsets defined by study cohorts. You can view data for a particular subset, however, as follows:

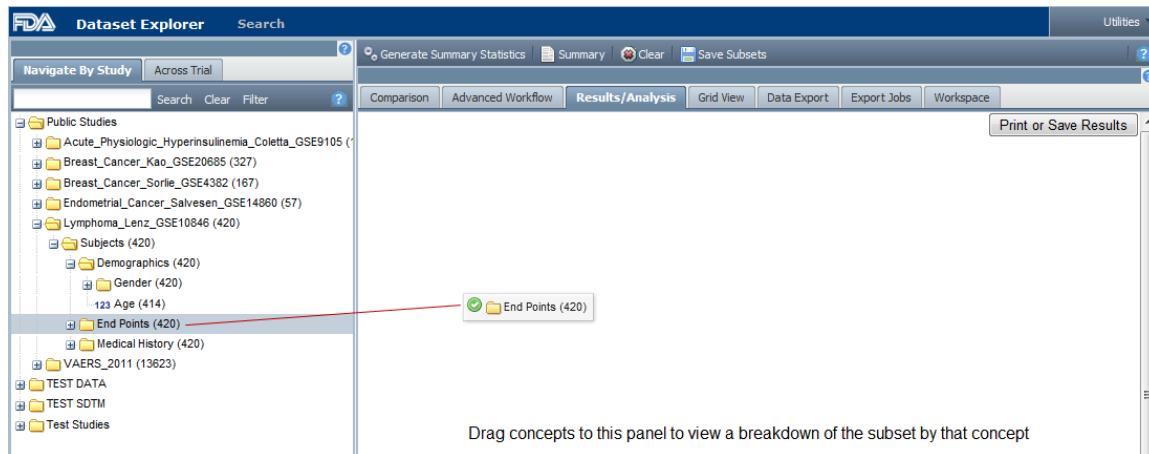
- Select a cohort from the navigation tree and drag it into a subset; for example:

The screenshot shows the transSMART interface with a navigation tree on the left. The tree includes categories like 'Public Studies' and 'Subjects (420)', with further subdivisions such as 'Demographics (420)' and 'Gender (420)'. Under 'Gender (420)', the 'Female' node is selected and highlighted in grey. A red arrow points from this node to the 'Subset 1' panel on the right. The 'Subset 1' panel contains a single condition: '...!Female'. Below this, there are two empty boxes separated by an 'AND' operator, each with an 'Exclude' button. The 'Subset 2' panel is also empty and has its own 'Exclude' button. The overall interface is clean with a white background and light blue header elements.

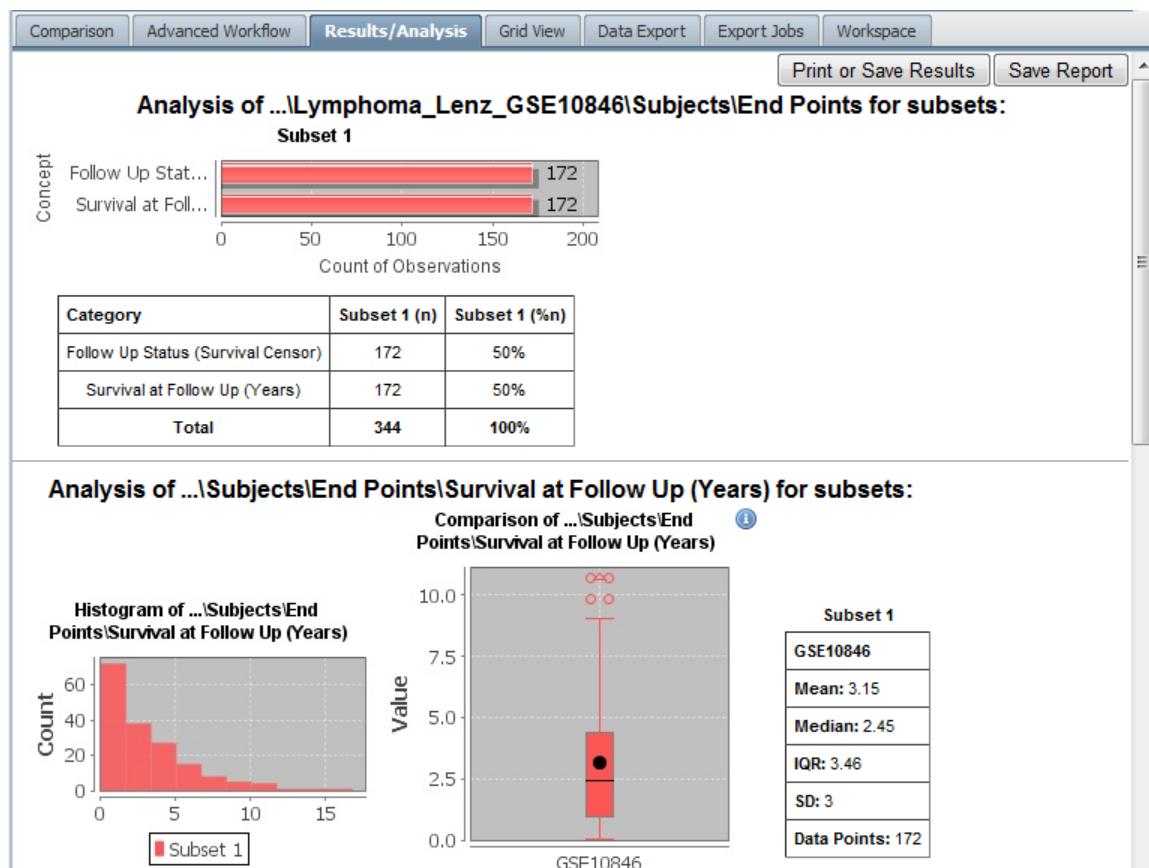
- Click the **Summary Statistics** tab.

Generating Summary Statistics

3. Drag and drop a folder from the navigation tree into the empty page; for example:



transSMART calculates the results and displays the data for the given subset and concept:



Defining Points of Comparison

Once you establish the subsets of subjects that you want to compare, you can apply one or more points of comparison to the subsets. A *point of comparison* is a concept in the navigation tree.

To apply a point of comparison to the subsets:

1. You must already have defined the subsets and have generated summary statistics for the subsets, as described in the previous section.
2. Drag the concept that you want to introduce as the point of comparison from the navigation tree and drop it anywhere inside the Summary Statistics view.

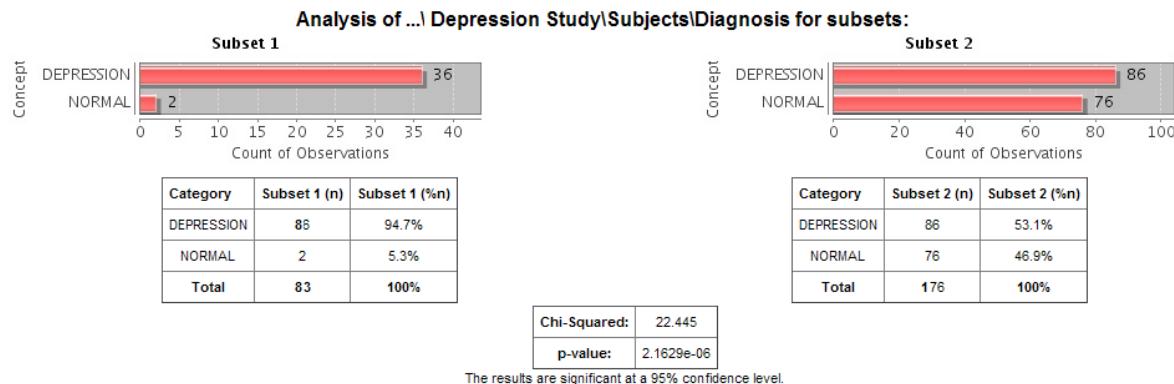
As soon as you drop the point of comparison into the Summary Statistics view, tranSMART begins to compare the subsets based on that point of comparison. When finished, tranSMART displays a side-by-side summary of how the subjects in each subset match or respond to the point of comparison.

Results of a Comparison

In a comparison of subjects in a psychological study, suppose Subset 1 contains subjects with a substance abuse problem and Subset 2 contains subjects with no substance abuse assessment.

After the subsets are defined and summary statistics are generated, a diagnosis of depression is dropped into the Summary Statistics view as a point of comparison. tranSMART displays a side-by-side comparison of the subjects in each subset, indicating that almost all the subjects with a substance abuse problem have been diagnosed with depression, while that diagnosis for those with no substance abuse problem is more evenly split.

The comparison is placed at the top of the Summary Statistics view, above the demographic definitions plus any other earlier comparisons:



To keep the size of the preceding figure within production limits, the demographics (age, sex, and race) portions of the figure are excluded.



Query details accessed through the **Summary** button do not reflect points of comparison.

Printing the Contents of Summary Statistics View

You can print the contents of Summary Statistics view as shown below.

1. In Summary Statistics view, click the **Print** button:

The screenshot shows the 'Summary Statistics' view within a software application. At the top, there is a navigation bar with tabs: 'Browse', 'Analyze', 'Sample Explorer', 'Gene Signature/Lists', 'GWAS', 'Upload Data', and 'Admin'. Below the navigation bar, there is a sub-navigation menu with items: 'Comparison', 'Summary Statistics' (which is currently selected), 'Grid View', 'Advanced Workflow', 'Data Export', 'Export Jobs', 'Analysis Jobs', and 'Worksp...'. On the far right of this menu, there is a 'Print' button, which is circled in red. The main content area is titled 'Summary Statistics' and contains two sections: 'Query Summary for Subset 1' and 'Query Summary for Subset 2'. Each section displays a list of study names.

Query Summary for Subset 1	Query Summary for Subset 2
(\\Public Studies\Public Studies\GSE13168\Treatment Group\Cell Culture Pretreatment\Fluticasone\)	(\\Public Studies\Public Studies\GSE13168\Treatment Group\Cell Culture Pretreatment\No treatment\)

The entire contents of Summary Statistics view appear in a separate browser window.

2. Click **Print this page**.

Copying Individual Charts in Summary Statistics View

If you are interested in a particular chart in the Summary Statistics view, you can copy the chart to a file, as follows:

1. With the Summary Statistics view displayed, click **Print**.

The entire contents of the Summary Statistics view appear in a separate browser window.

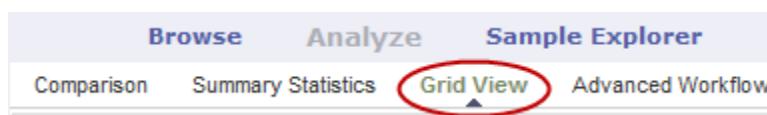
2. Right-click the chart you want to copy.
3. In the Internet Explorer popup menu, click **Save Image As**.
4. In the Save Image dialog, specify the name, location, and the file type for the chart.
5. Click **Save**.

Viewing Analysis Data in Grid View

If you are displaying analysis data in the various tables and charts of Summary Statistics view, and want to view the data in a single table, use the **Grid View** option.

Access Grid View as follows:

1. Click the **Analyze** tool and define your cohorts as described earlier in this chapter.
2. Click **Summary Statistics**.
3. Click **Grid View**.



4. Optionally, you can drag and drop additional points of comparison into the grid, and new columns will appear for that data.

You can drag a node from any level of the tree into the grid.

Sample of Grid View for a public study:

Comparison	Summary Statistics	Grid View	Advanced Workflow	Data Export	Export Jobs	Analysis Jobs	Workspace	Sample Details	Gala
Grid View									
Subject	Patient	Samples	Subset	Trial	Sex	Age	Race	male	female
1000384597	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384598	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384604	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384605	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384607	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384608	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384611	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384613	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female
1000384614	GSE8581GS...		subset1	GSE8581	female	NULL	NULL	NULL	female



The ID assigned in the **Subject** column is the internal tranSMART ID that is assigned at the time of data loading. The ID in the **Patient** field contains the original subject ID that was provided in the data.

Grid View Display Options

- **Sort the grid by a specific column.** Click the down-arrow icon (▼) next to the column heading you want to sort by, then select **Sort Ascending** or **Sort Descending**.
 - **Hide or redisplay columns.** Click the down-arrow icon next to any column heading, click **Columns** as shown below, then select or deselect columns to hide or redisplay:

Samples ▲	▼	Subset	Trial	Sex
		A Z Sort Ascending	8581	fema
		Z A Sort Descending	8581	fema
		Columns		
		subset1	GSE	<input checked="" type="checkbox"/> Subject
		subset1	GSE	<input checked="" type="checkbox"/> Patient
		subset1	GSE	<input checked="" type="checkbox"/> Samples
		subset1	GSE	<input checked="" type="checkbox"/> Subset
		subset1	GSE	<input checked="" type="checkbox"/> Trial
		subset1	GSE	<input checked="" type="checkbox"/> Sex
		subset1	GSE	<input checked="" type="checkbox"/> Age
		subset1	GSE	<input checked="" type="checkbox"/> Race
		subset1	GSE	<input checked="" type="checkbox"/> male
		subset1	GSE	<input checked="" type="checkbox"/> female
		subset1	GSE	

If a column name does not appear in the menu, you have not included the associated concept in the analysis. For example, Diagnosis has not been included in the analysis above.

Chapter 5

Advanced Workflow Analyses

tranSMART provides the ability to generate the following analyses and visualizations:

- [aCGH Survival Analysis](#) (page 46)
- [Box Plot with ANOVA](#) (page 48)
- [Correlation Analysis](#) (page 50)
- [Forest Plot](#) (page 51)
- [Frequency Plot for aCGH](#) (page 55)
- [Group Test for aCGH](#) (page 56)
- [Group Test for RNASeq](#) (page 57)
- [Heatmaps](#) (page 59):
 - [Standard Heatmap](#) (page 60)
 - [Hierarchical Clustering](#) (page 61)
 - [K-Means Clustering](#) (page 63)
 - [Marker Selection](#) (page 64)
- [IC50 Dose Response Curve](#) (page 66)
- [Line Graph](#) (page 68)
- [Logistic Regression](#) (page 69)
- [PCA](#) (page 71)
- [Scatter Plot with Linear Regression](#) (page 73)
- [Survival Analysis](#) (page 77)
- [Table with Fisher Test](#) (page 78)
- [Waterfall Plot](#) (page 81)

Advanced Workflows use the R software environment for statistical computing and to generate analyses and visualizations. For more information, visit <http://www.r-project.org>.

Running the Analyses

To begin to run any analysis:

1. In **Analyze**, open the study of interest, or open the Advanced Trials folder to run an analysis of data from multiple studies.
2. Define the cohort(s) you want to analyze by dragging one or more concepts into empty subset definition boxes. For more information, see [Defining the Cohorts](#) on page 21.

The following sections describe how to run specific analyses after you perform the above steps.

Optionally, after you run an analysis, you can download the associated R data by clicking **Download raw R data** below the visualization.



Visualization may take a few seconds to a few minutes to appear.

aCGH Survival Analysis

This is a statistical test (logrank) for survival data and called copy number data. The testing is recommended to be performed on high-dimensional data nodes containing chromosomal region information.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform an aCGH Survival Analysis:

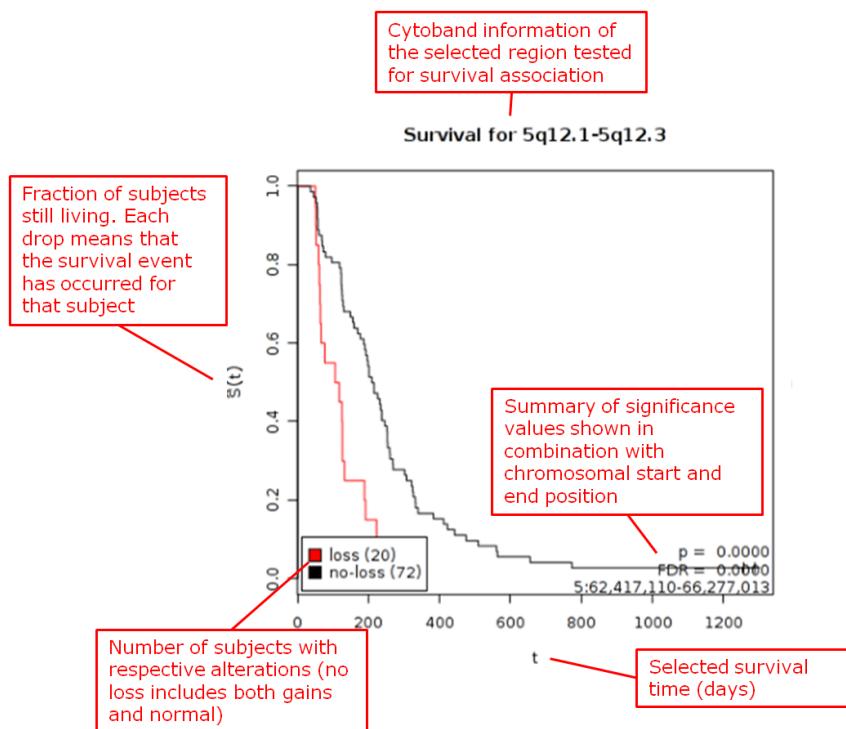
1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **aCGH Survival Analysis**.
The Variable Selection section appears.
3. Define the following variables:
 - Region**: A high-dimensional data node containing the chromosomal regions.
 - Survival Time**: A numerical data node containing survival time of interest (for example: Overall Survival [days]).
 - Censoring Variable**: A categorical data node indicating status subjects in which the selected event in survival time did NOT happen (for example, for Overall Survival Time, the Censoring Variable to select is Alive).
 - Alteration type**: The type of chromosomal alteration used to test association to survival (gains, losses, both).

- **Permutations:** The significance of the p-values is evaluated through permutations, and a false discovery rate is calculated. At least 10,000 permutations are recommended for final calculations. This will require a significant amount of time. (Permutations can be lowered for exploratory purposes.)
4. Click **Run Analysis**. As this may take a while, users are advised to select the option **Run Job in Background** in the popup window. The analysis can be retrieved at a later time in the **Analysis Jobs** tab.

Results appear in two sections:

1. The chromosomal regions present in the high-dimensional data node are shown in a table, appended with p-values and false discovery rates.
2. You can sort through the chromosomal regions, click on a region of interest and press the button **Show Survival Plot**. This will plot the survival curve for that particular region (see example below).

You can also opt to click **Download Result**, in which case both the table and all survival plots are obtained.



Reference

Wiel et al. (2005) "CGHMultiArray: exact p-values for multi-array comparative genomic hybridization data." *Bioinformatics* 21: 3193-3194.

Box Plot with ANOVA

A box plot with ANOVA analysis displays a box and whisker plot with corresponding analysis of variance in the sample(s).

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a box plot with ANOVA analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Box Plot with ANOVA**.

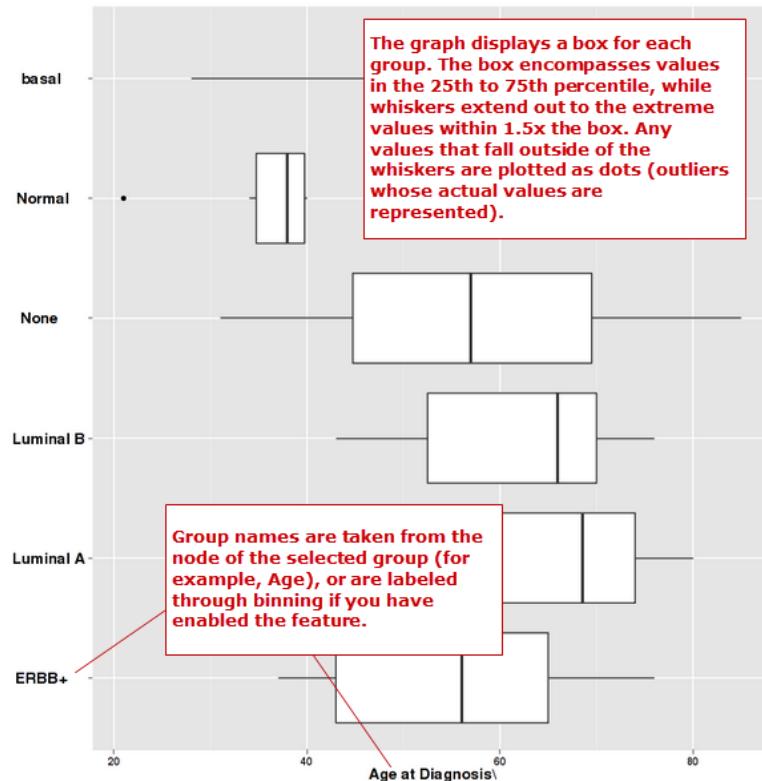
The Variable Selection section appears.

3. Define an independent variable and a dependent variable, following the instructions above the boxes. You must use one categorical variable and one continuous variable. The boxes are plotted based on the categorical variable:
 - If the *independent variable* is categorical, the boxes are plotted horizontally.
 - If the *dependent variable* is categorical, the boxes are plotted vertically.
 - If you select two continuous variables, you must bin one to create a categorical value.
4. Optionally, enable binning by selecting **Enable binning**.

Data binning refers to a pre-processing technique used to reduce minor observation errors. Clusters of data are replaced by a value representative of that cluster (the central value). For information on binning, see [Data Binning Using Box Plot with ANOVA](#) on page 84.

5. Click **Run**.

Your analysis appears below:



ANOVA Result

p-value	0.00181
F value	4.11

F value and p-value produced by ANOVA calculations on all groups.

Group	Mean	n
ERBB+	54.5	11
Luminal A	64.3	28
Luminal B	62.1	11
None	57.5	44
Normal	37.3	6

The mean and population size of each group.

Pairwise t-Test p-Values

	ERBB+	Luminal A	Luminal B	None	Normal
Luminal A	5.19e-02	NA	NA	NA	NA
Luminal B	2.07e-01	6.59e-01	NA	NA	NA
None	5.34e-01	4.56e-02	3.28e-01	NA	NA
Normal	1.65e-02	3.65e-05	6.65e-04	1.20e-03	NA

Correlation Analysis

In a correlation analysis, you are using statistical correlation to assess the relationship between variables.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a correlation analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Correlation Analysis**.
The Variable Selection section appears.
3. Define two or more continuous (or numerical) variables (for example, Age).
4. Indicate how you want to run the correlation in the **Run Correlation** dropdown menu.
5. Select the analysis you want to perform from the **Correlation Type** dropdown menu:

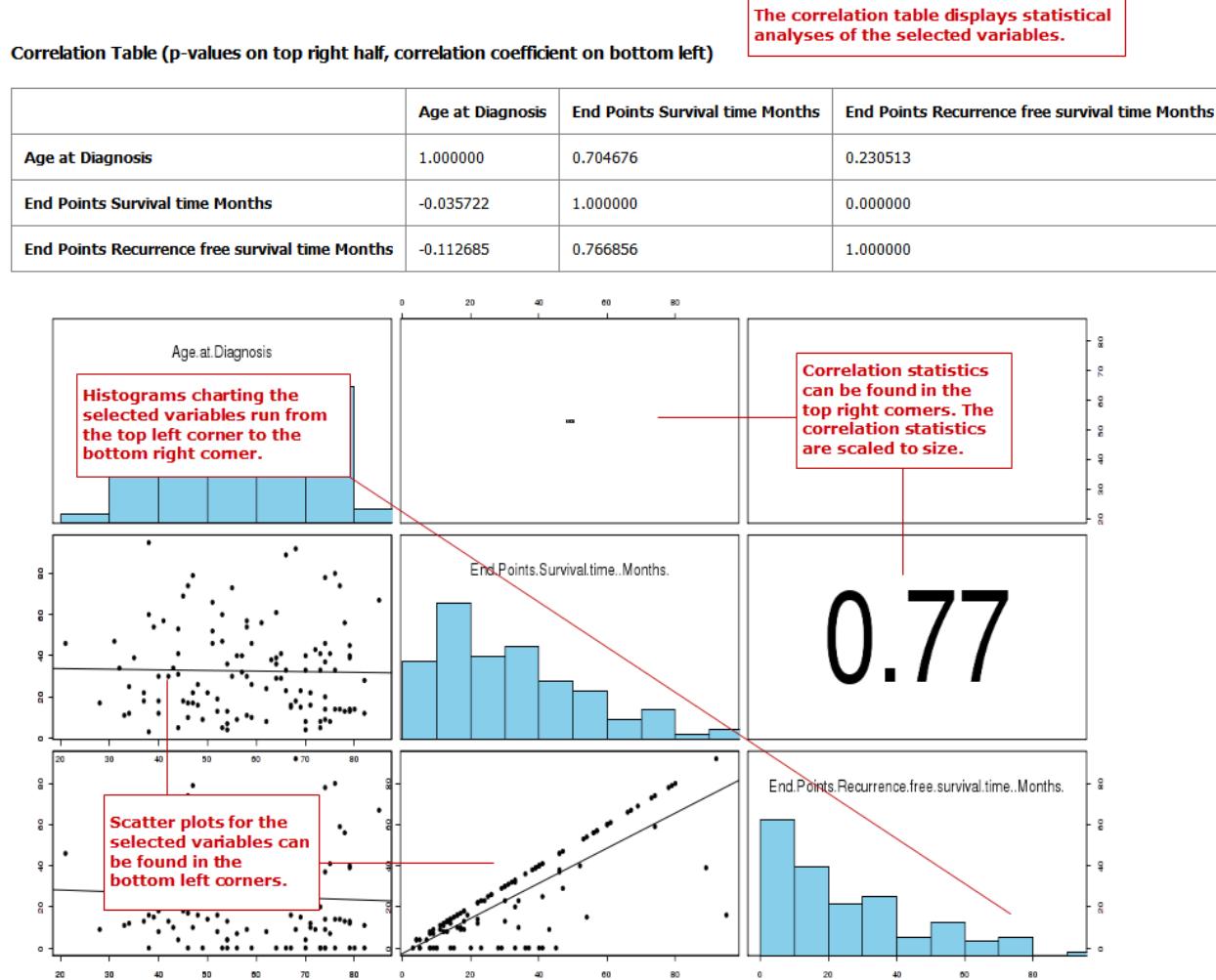

Note: At this time, correlations are run by variable only.



The analyses listed under **Correlation Type** refer to different regression algorithms.

6. Click **Run**.

Your analysis appears below:



Forest Plot

A forest plot graphically displays the relative strength of treatment effects among various cohorts (for example, people who took the same drug). Relative strength can be calculated in two ways:

- As relative risk given exposure to a treatment or an environmental factor — that is, the probability of an event occurring in a group of exposed subjects measured against the probability of the event occurring in a group of non-exposed subjects.
- As an odds ratio — that is, the odds of an event occurring in one group measured against the odds of an event occurring in a different group.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a forest plot analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.

2. Select **Forest Plot**.

The Variable Selection section appears.

3. Define the following variables:

- **Independent variable:** Specifies the experimental or treatment variable being measured in the analysis. If this variable is continuous, it requires binning.
- **Control or Reference variable:** Indicates the control or reference variable for the analysis; for example, no treatment or placebo. If this variable is continuous, it requires binning.
- **Dependent Variable:** Indicates the event outcome. Variables entered must be mutually exclusive; for example, Alive and Dead.

If there is only one node in the concept you want to use for Dependent variable, use the checkbox below the box to create the second node. For example, the only node in Gender is Female. tranSMART presumes that each subject for whom Female does not apply is Other.

If this variable is continuous, it requires binning.

- **Stratification Variable:** Stratifies the relationship between the dependent and independent variables by the variable specified here. For example, if you add the stratification variable Cancer Stage, data is plotted and displayed for each stage. Without stratification, data displays as a single summary value in the graph.

If this variable is continuous, it requires binning.

4. Optionally, enable binning by clicking the **Enable** button.

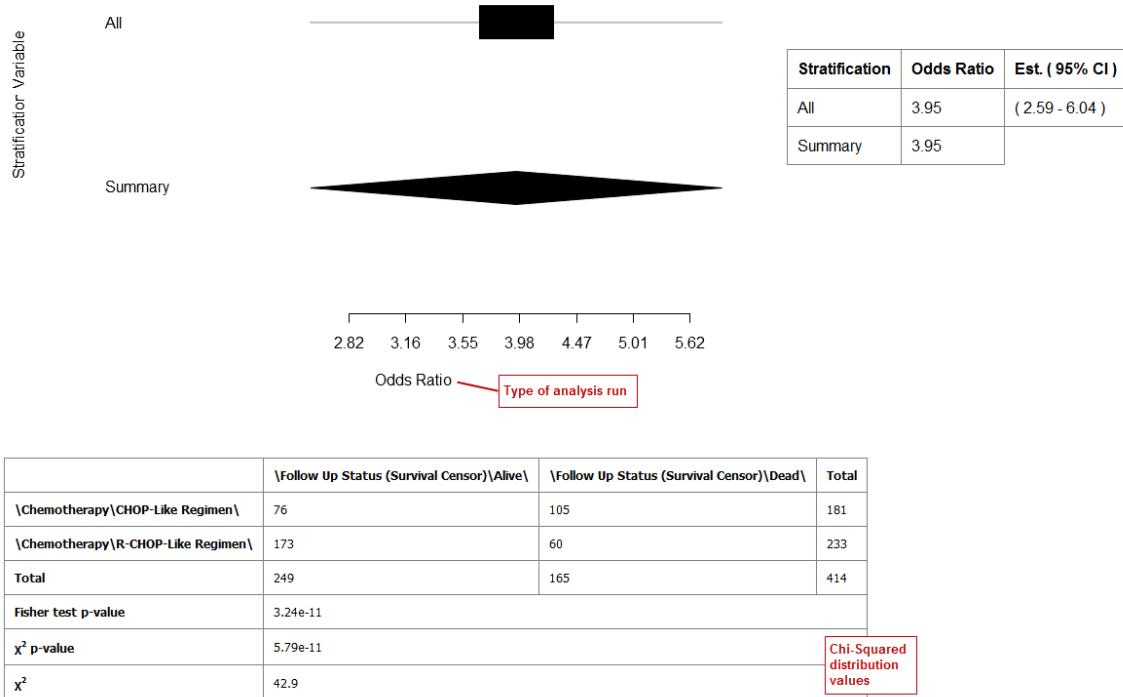
For information, see [Data Binning Using Forest Plot](#) on page 85.

5. In **Statistic Type**, click **Odds Ratio or Relative Risk**.

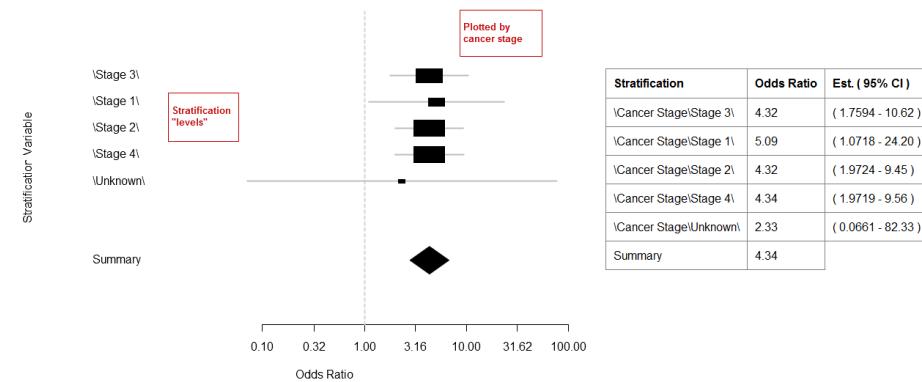
6. Click **Run**.

Your analysis appears below:

Example 1: Odds Ratio analysis run without stratification:



Example 2: Odds ratio analysis with stratification:



Five stratified tables are shown, each corresponding to a cancer stage: Stage 3, Stage 1, Stage 2, Stage 4, and Unknown. Each table includes columns for Follow Up Status (Alive/Dead), Total, and Fisher test p-value. Red boxes highlight the Chi-squared p-value for each stage.

\Cancer Stage\Stage 3\			
	\Follow Up Status (Survival Censor)\Alive\	\Follow Up Status (Survival Censor)\Dead\	Total
\Chemotherapy\CHOP-Like Regimen\	18	21	39
\Chemotherapy\R-CHOP-Like Regimen\	46	12	58
Total	64	33	97
Fisher test p-value	0.00102		
χ^2 p-value	0.00157		
χ^2	9.99		

\Cancer Stage\Stage 1\			
	\Follow Up Status (Survival Censor)\Alive\	\Follow Up Status (Survival Censor)\Dead\	Total
\Chemotherapy\CHOP-Like Regimen\	21	7	28
\Chemotherapy\R-CHOP-Like Regimen\	36	2	38
Total	57	9	66
Fisher test p-value	0.0303		
χ^2 p-value	0.00516		
χ^2	3.79		

\Cancer Stage\Stage 2\			
	\Follow Up Status (Survival Censor)\Alive\	\Follow Up Status (Survival Censor)\Dead\	Total
\Chemotherapy\CHOP-Like Regimen\	23	32	55
\Chemotherapy\R-CHOP-Like Regimen\	51	16	67
Total	74	48	122
Fisher test p-value	0.000171		
χ^2 p-value	0.00024		
χ^2	13.5		

\Cancer Stage\Stage 4\			
	\Follow Up Status (Survival Censor)\Alive\	\Follow Up Status (Survival Censor)\Dead\	Total
\Chemotherapy\CHOP-Like Regimen\	14	44	58
\Chemotherapy\R-CHOP-Like Regimen\	37	26	63
Total	51	70	121
Fisher test p-value	0.000198		
χ^2 p-value	0.000247		
χ^2	13.4		

\Cancer Stage\Unknown\			
	\Follow Up Status (Survival Censor)\Alive\	\Follow Up Status (Survival Censor)\Dead\	Total
\Chemotherapy\CHOP-Like Regimen\	0	1	1
\Chemotherapy\R-CHOP-Like Regimen\	3	4	7
Total	3	5	8
Fisher test p-value	1		
χ^2 p-value	1		
χ^2	0		

Frequency Plot for aCGH

This analysis plots the copy number alteration frequencies for different groups. This analysis is performed on high-dimensional data nodes containing chromosomal region information.



This analysis represents a quick way to investigate alteration frequencies of the selected groups and is very similar to the advanced workflow analysis [Group Test for aCGH](#) (page 56), in which statistical testing is performed. It is advisable to use the Frequency Plot for aCGH analysis for exploratory purposes before performing statistical testing (which requires a significant amount of time).

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a Frequency Plot for aCGH analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.

2. Select **Frequency Plot for aCGH**.

The Variable Selection section appears.

3. Define the following variables:

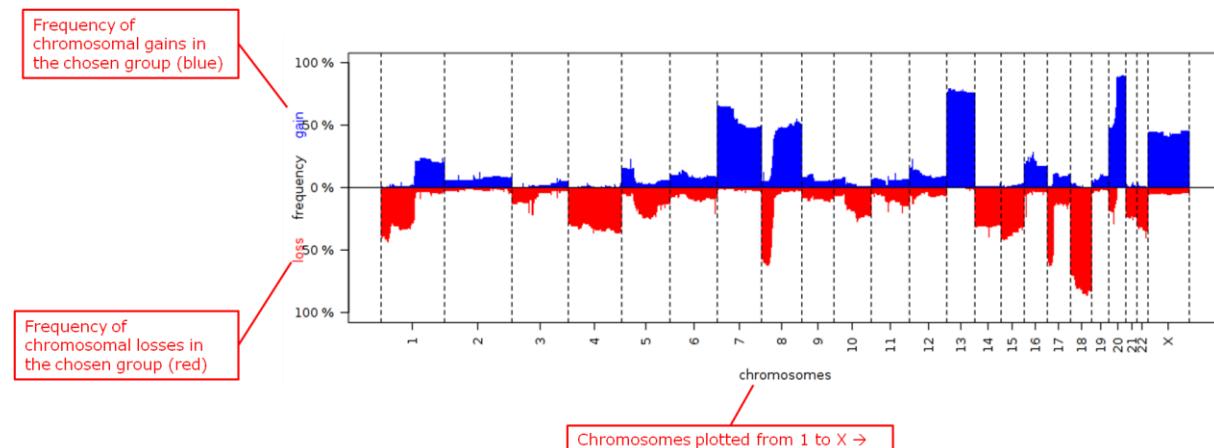
- ArrayCGH**: A high-dimensional data node containing the chromosomal regions.
- Group**: Categorical data nodes separating the samples into two or more groups (though only one group may be plotted as well).

4. Click **Run Analysis**.

Result

Frequency plots of copy number alterations in each defined group are shown. Frequencies of chromosomal gains are in blue and chromosomal losses are in red.

Example of a plot of one group:



Reference

Mark A. van de Wiel, Kyung In Kim, Sjoerd J. Vosse, Wessel N. van Wieringen, Saskia M. Wilting and Bauke Ylstra. " CGHcall: calling aberrations for array CGH tumor profiles." *Bioinformatics*, 23, 892-894.

Group Test for aCGH

Three different statistical tests are available to determine potential differences in status of copy number alterations between various groups. The testing is recommended to be performed on high-dimensional data nodes containing chromosomal region information.

This analysis plots the copy number aberration frequencies for different groups and indicates significant different regions between these groups.

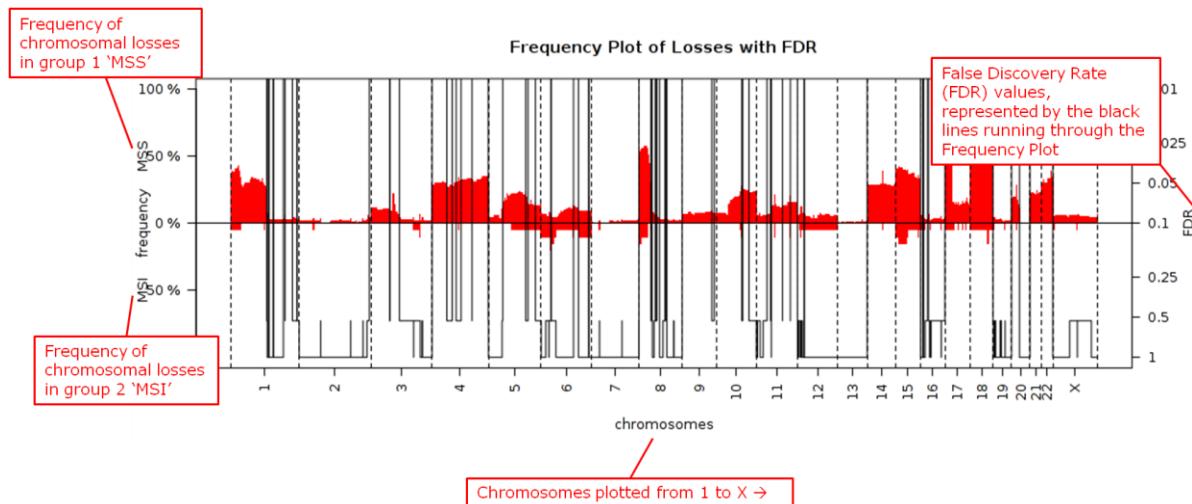
To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a Group test for aCGH analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Group Test for aCGH**.
The Variable Selection section appears.
3. Define the following variables:
 - Region**: A high-dimensional data node containing the chromosomal regions.
 - Group**: Categorical data nodes separating the samples into two or more groups.
 - Statistical Test**: Select the test to perform:
 - **Chi-square**: Test for the association between alteration pattern and group label. Supports multiple comparisons.
 - **Wilcoxon**: Rank-sum test for two groups.
 - **Kruskal-Wallis**: Generalization for Wilcoxon for more than two groups.
 - Alteration type**: The type of chromosomal alteration used to test the association (gains, losses, both).
 - Permutations**: The significance of the p-values is evaluated through permutations, and a false discovery rate is calculated. At least 10,000 permutations are recommended for final calculations. This will require a significant amount of time. (Permutations can be lowered for exploratory purposes in lieu of generating a Frequency Plot for aCGH.)
4. Click **Run Analysis**. As this may take a while, consider selecting the option **Run Job in Background** in the popup window. The analysis can be retrieved at a later time in the **Analysis Jobs** tab.

Results appear in two sections:

1. The chromosomal regions present in the high-dimensional data node are shown in a table, appended with p-values and false discovery rates.
2. Frequency plots of copy number alterations in each defined group are shown. In particular, "Mirror frequency plots" are shown; for example:



Reference

Wiel et al. (2005) "CGHMultiArray: exact p-values for multi-array comparative genomic hybridization data." *Bioinformatics* 21: 3193-3194.

Group Test for RNASeq

For microarrays, the abundance of a particular transcript is measured as a fluorescence intensity, effectively a continuous response, whereas for digital gene expression (DGE) data the abundance is observed as a count. One of the fundamental data analysis tasks, especially for gene expression studies, involves determining whether there is evidence that counts for a transcript or exon are significantly different across experimental conditions. The software package edgeR (empirical analysis of DGE in R), which forms part of the Bioconductor project, is designed to examine differential expression of count-based expression data between two or more groups.

The Group Test for RNASeq analysis is recommended to be performed on high-dimensional data nodes containing RNASeq-based read count observations. The results of the analysis comprise an ordered table of the differentially expressed genes (or tags, or exons, etc.) and plots visualizing the level of (dis)similarity of individual samples (MDS plot) as well as the DGE data (MA plot).

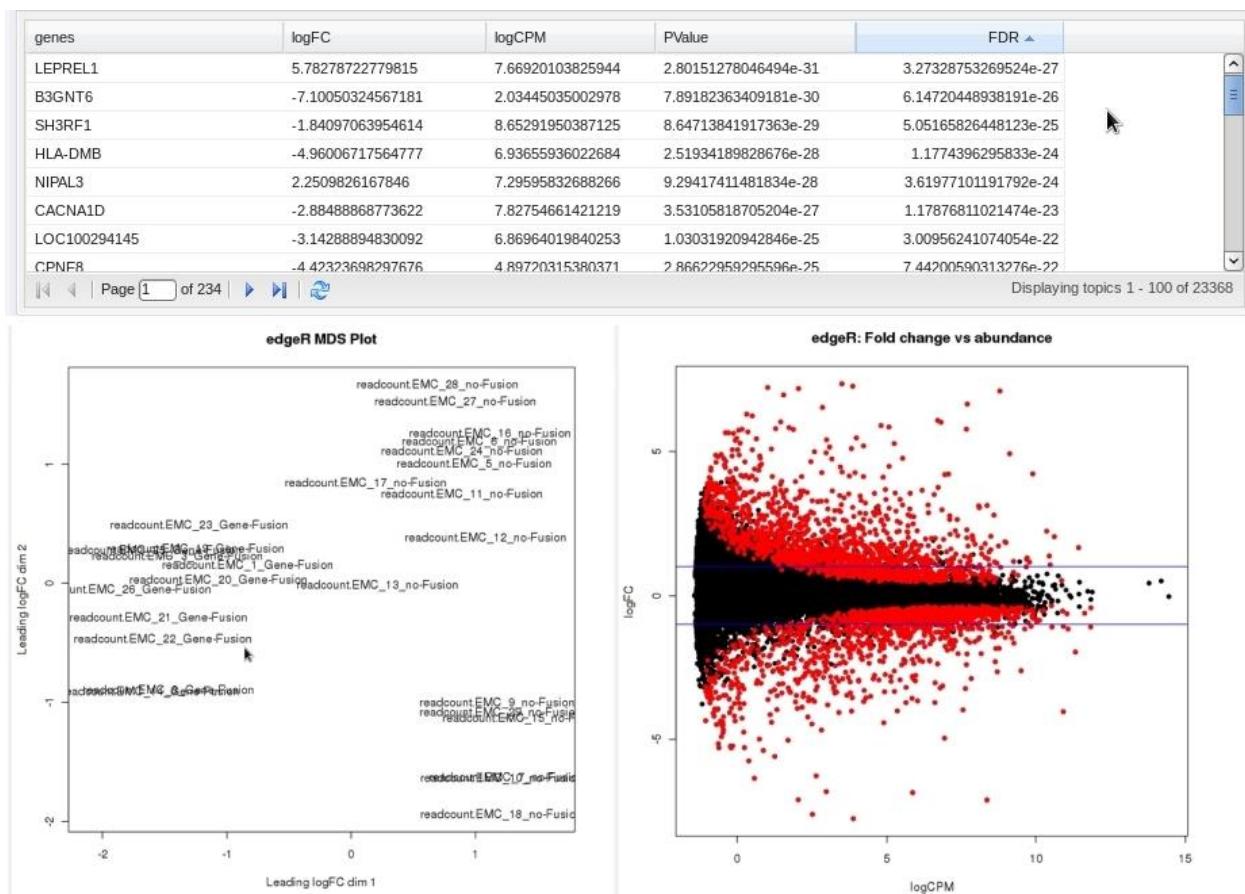
To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a Group Test for RNASeq analysis:

1. Click the **Advanced Workflow** tab, and then open the **Analysis** menu.
2. Select **Group Test for RNASeq**.
The Variable Selection section appears.
3. Define the following variables:
 - RNASeq**: A high-dimensional data node containing RNASeq-based read count data.
 - Group**: Categorical data nodes separating the samples into two or more groups.
 - Analysis Type**: Select the type of analysis to perform:
 - two group unpaired
 - multi-group
4. Click **Run Analysis**. As this may take a while, consider selecting the option **Run Job in Background** in the popup window. The analysis can be retrieved at a later time in the **Analysis Jobs** tab.

Results appear in two sections:

- An ordered table of the differentially expressed genes (or tags or exons, etc.) including fault changes, abundances, p-values, and false discovery rates.
- An MDS plot visualizing the level of (dis)similarity of individual samples, and an MA plot (fold change versus abundance) visualizing the RNASeq data.



Reference

Mark D. Robinson, Davis J. McCarthy and Gordon K. Smyth (2009) "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* (2010) 26 (1): 139-140.

Heatmaps

In Analyze, a heatmap is a matrix of data points for a particular set of biomarkers, such as genes, at a particular point in time and/or for a particular tissue sample in the study, as measured for each subject in the study.

In an Analyze heatmap:

- The color red indicates higher-than-normal expression
- The color green indicates lower-than-normal expression
- Biomarkers appear in the y-axis, and subjects appear in the x-axis.



A heatmap can display data points for up to 1000 samples.

Analyze uses the R software environment for statistical computing and to generate analyses and visualizations. For more information, visit <http://www.r-project.org>.

You can generate the following types of heatmaps:

- [Standard Heatmap](#) (page 60)
- [Hierarchical Clustering](#) (page 61)
- [K-Means Clustering](#) (page 63)
- [Marker Selection](#) (page 64)

Standard Heatmap

A standard heatmap is a visualization of biomarker data points with no indication of patterns, groupings, or differentiation among the data points.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a standard heatmap analysis:

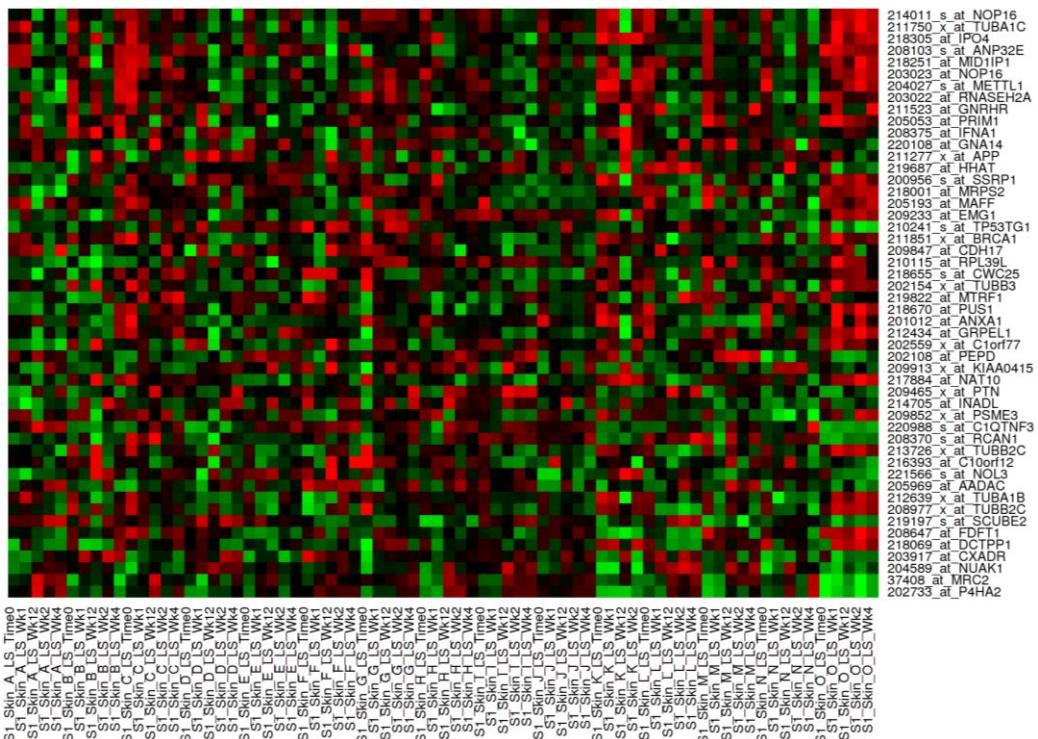
1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Heatmap**.

The Variable Selection section appears.

3. Drag a high-dimensional data node (●), or several high-dimensional nodes in the case of serial data, into the Variable Selection box.
4. Click the **High Dimensional Data** button.
The Compare Subsets-Pathway Selection dialog box appears.
5. Specify the platform and other filters for the analysis.
For information, see [High Dimensional Data](#) on page 83.
6. Click **Apply Selections**.
7. In **Max rows to display**, type the maximum number of rows in the heatmap.
8. Optionally, select either or both of the following:
 - Group by subject (instead of node) for multiple nodes
 - Calculate z-score on the fly

9. Click **Run**.

Your analysis appears below:



With serial data, the heatmap will display the various conditions ordered by increasing associated value, such as in chronological order for a time series.

Hierarchical Clustering

Hierarchical clustering is a visualization of patterns of related data points in gene expression data.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a hierarchical clustering heatmap analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Hierarchical Clustering**.
3. Drag a high-dimensional data node (●) into the Variable Selection box.
4. Click the **High Dimensional Data** button.

The Compare Subsets-Pathway Selection dialog box appears.

5. Specify the platform and other filters for the analysis.

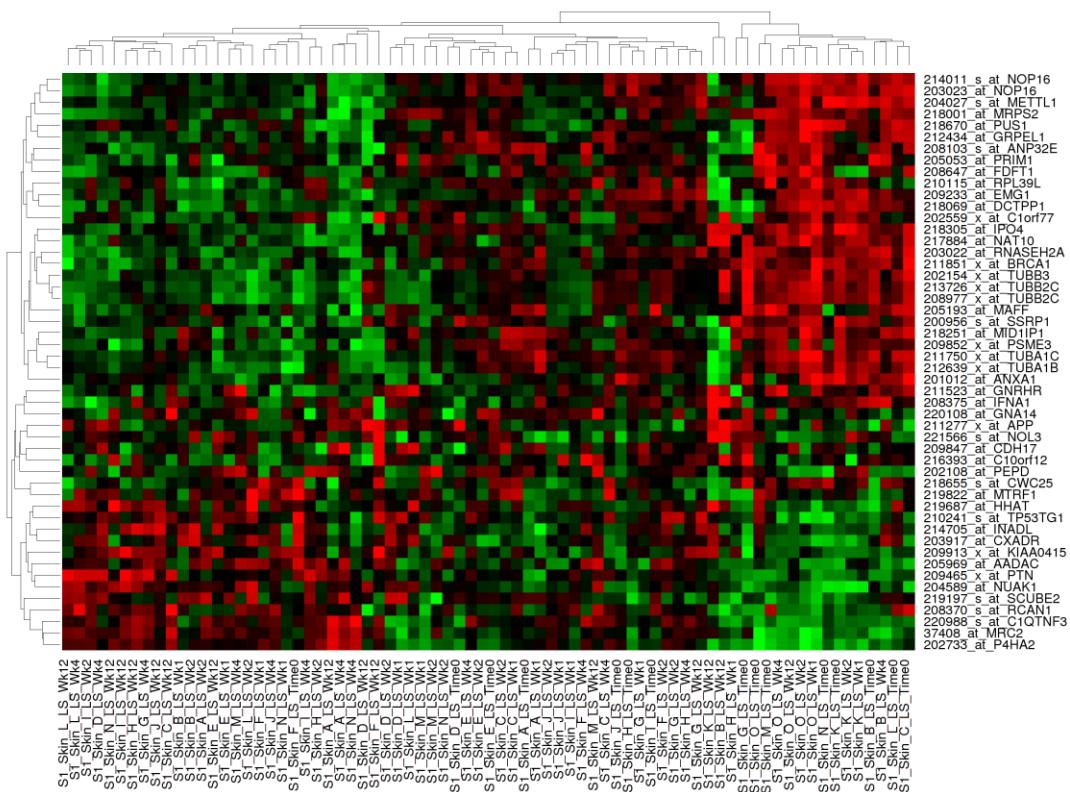
For information, see [High Dimensional Data](#) on page 83.

6. Click **Apply Selections**.
 7. In **Max rows to display**, type the maximum number of rows in the heatmap.
 8. Optionally, select one or more of the following:

- Apply clustering for rows
 - Apply clustering for columns
 - Calculate z-score on the fly

- ### **9. Click **Run**.**

Your analysis appears below:



To read more about Hierarchical Clustering, visit:
<http://www.ics.uci.edu/~eppstein/280/cluster.html>

K-Means Clustering

K-Means clustering is a visualization of groupings of the most closely related data points, based on the number of groupings you specify.



The K-Means analysis clusters columns only. Rows are not clustered.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a k-means clustering heatmap analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **K-Means Clustering**.

The Variable Selection section appears.

3. Drag a high-dimensional data node (H) into the Variable Selection box.
4. Click the **High Dimensional Data** button.

The Compare Subsets-Pathway Selection dialog box appears.

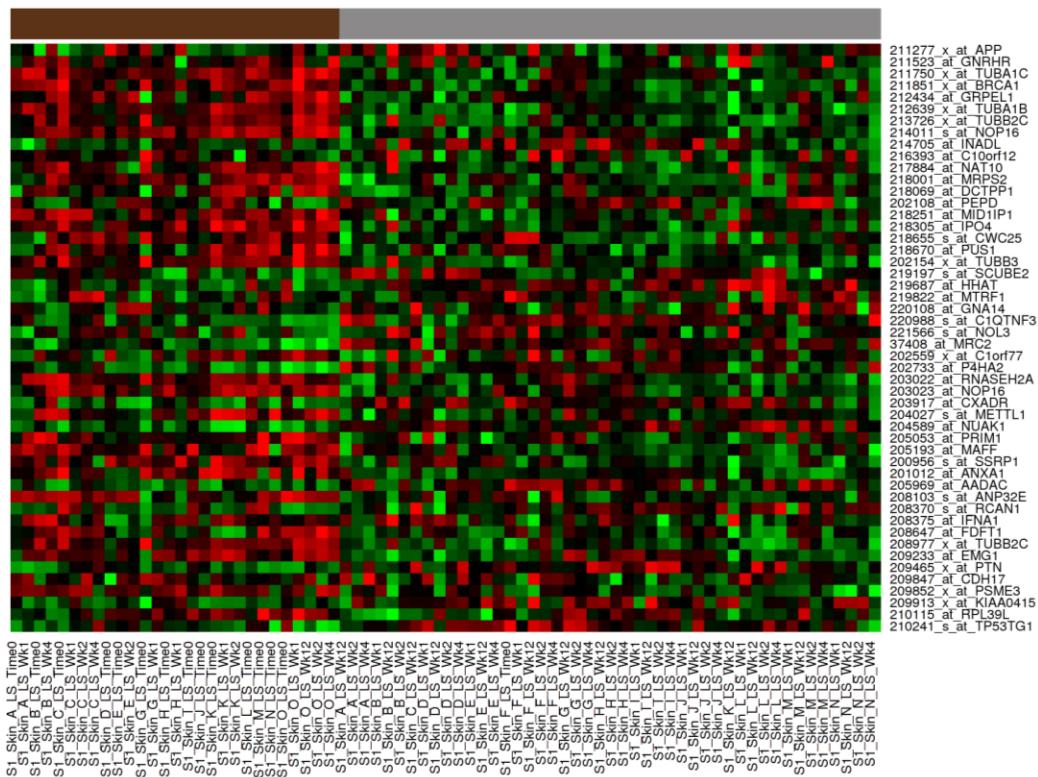
5. Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

6. Click **Apply Selections**.
7. In **Number of clusters**, type the number of clusters to include in the heatmap.
8. In **Max rows to display**, type the maximum number of rows in the heatmap.
9. Optionally, select **Calculate z-score on the fly**.

10. Click **Run**.

Your analysis appears below. Clusters are represented by the colored bars at the top of the heatmap:



To read more about K-Means Clustering, visit:
<http://www.ics.uci.edu/~eppstein/280/cluster.html>

Marker Selection

A marker selection heatmap is a visualization of differentially expressed genes in distinct phenotypes. Specifically, the algorithm determines the set of genes which is most differently expressed between the two subsets. This list of differentially expressed genes is subsequently presented in a table, along with a variety of accompanying statistics.

Optionally, you can run a MetaCore Enrichment Analysis from a generated Marker Selection heatmap.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.



Two subsets must be specified when using a Marker Selection heatmap.

To perform a marker selection heatmap analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
 2. Select **Marker Selection**.
- The Variable Selection section appears.
3. Drag a high-dimensional data node (●) into the Variable Selection box.
 4. Click the **High Dimensional Data** button.

The Compare Subsets-Pathway Selection dialog box appears.

5. Specify the platform and other filters for the analysis.

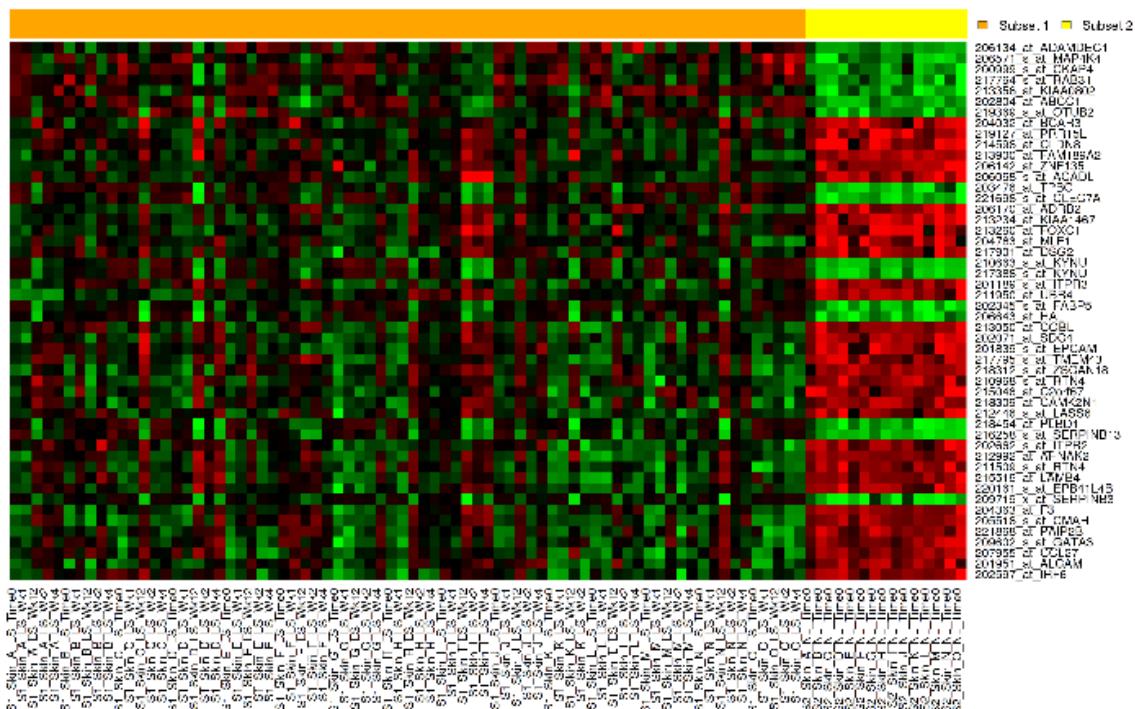
For information, see [High Dimensional Data](#) on page 83.

6. Click **Apply Selections**.
7. In the **Number of Markers** field, type a numeric value. This will determine the number of differentially expressed genes that are returned.
8. Optionally, select either or both of the following:

- Group by subject (instead of node) for multiple nodes
- Calculate z-score on the fly

9. Click **Run**.

Your analysis appears below. The subsets are represented by the colored bars at the top of the heatmap:



A table of the top markers appears below the heatmap. You can sort the table by clicking any of the column headings. Optionally, you can view MetaCore settings and run a MetaCore Enrichment Analysis by clicking the buttons above the table.

For information about MetaCore Enrichment Analysis, see [MetaCore Enrichment Analysis](#) on page 95.

The following table represents a portion of the data from the Marker Selection heatmap illustrated above:

Table of top Markers METACORE SETTINGS [Run MetaCore Enrichment Analysis](#)

Gene Symbol	Probe ID	Log2(fold change) S2 vs S1	t	P-value	Adjusted P-value	B
KIAA1467	213234_at	-1.83250044462678	-7.60654308806882	2.62360440632487e-13	5.34795522185261e-09	19.6420805129124
EPB41L4B	220161_s_at	-1.79522586242916	-7.36314327740463	1.29610451288309e-12	8.6858746864399e-09	18.1204028206329
CLDN8	214598_at	-1.7644895433702	-7.33291504107679	1.57652720878327e-12	8.6858746864399e-09	17.9339084641723
SDC4	202071_at	-1.79083400396042	-7.32084988381699	1.70444950675822e-12	8.6858746864399e-09	17.859627197573
CLEC7A	221698_s_at	1.74909978868116	7.21985359855876	3.26343028517116e-12	1.33043525865858e-08	17.241320459041
UBR4	211950_at	-1.68831401895957	-7.04400775010408	9.95845511986347e-12	3.38321915272162e-08	16.1798600734978
EPCAM	201839_s_at	-1.71321763310572	-7.00158366136835	1.29959522007723e-11	3.78442128086489e-08	15.9266818596976
F3	204363_at	-1.71621899571343	-6.9724120176461	1.55961445459661e-11	3.83507933112402e-08	15.753253805065



For more information on the analyses used in Marker Selection, visit: <http://mathworld.wolfram.com/bonferronicorrection.html>.

IC50 Dose Response Curve

IC50 dose response curve analyses measure the effectiveness of a compound in inhibiting certain biological processes.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

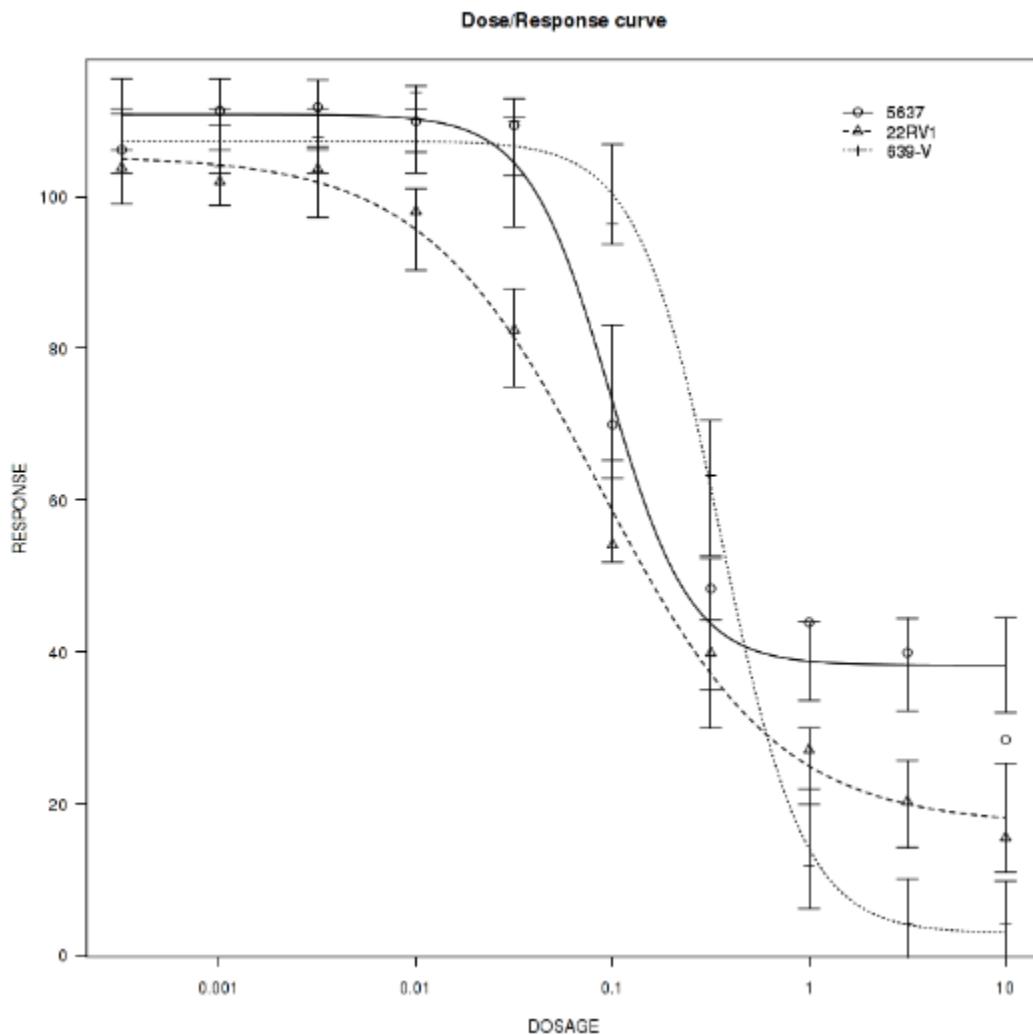
To perform an IC50 dose response curve analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **IC50**.
3. Define the following variables:

Variable	Description
Cell Lines	The categorical value that represents the cell lines to plot.
Concentration Variable	The continuous variable that represents the dosage of a compound at a given concentration level.

4. Click **Run**.

Your analysis appears below:

All Cell Lines Dosage/Response Curve

Line Graph

A line graph is designed to plot serial numeric data (high or low dimensional); that is, a numeric variable that has been measured in a series of conditions for each subject (for example, several timepoints). For more information on serial data, see [Serial Numeric Data](#) on page 20.

In a line graph, the various conditions are plotted along the x-axis, at scale (unless you check the **Plot evenly spaced** option) when the conditions are associated with a numeric value. For example, time series data will be plotted on scale with time.

For categorical conditions, data points are evenly spaced along the x-axis.

The measurement of interest can be plotted for one or several groups (for example, treatment groups) of the defined subsets.



Each group will be plotted as a distinct line on the graph, unless you select **Plot individuals** as the graph type. In that case, each individual is plotted as a distinct line, using different colors for each group.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a line graph analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Line Graph**.

The Variable Selection section appears.

3. Drag and drop several nodes of serial data into the **Time/Measurement Concepts** selection box. To define the groups, drag and drop nodes into the **Group Concepts** selection box.

If no group concept is defined, the defined subsets are used as one group.

Note that the order of the data points along the x-axis is controlled by the value defining each condition, even with the **Plot evenly spaced** option selected; for example, in chronological order for time series.

4. If you included high dimensional data in either concept box, click the **High Dimensional Data** button for that box.

The Compare Subsets-Pathway Selection dialog box appears.

- a. Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

- b. Click **Apply Selections**.

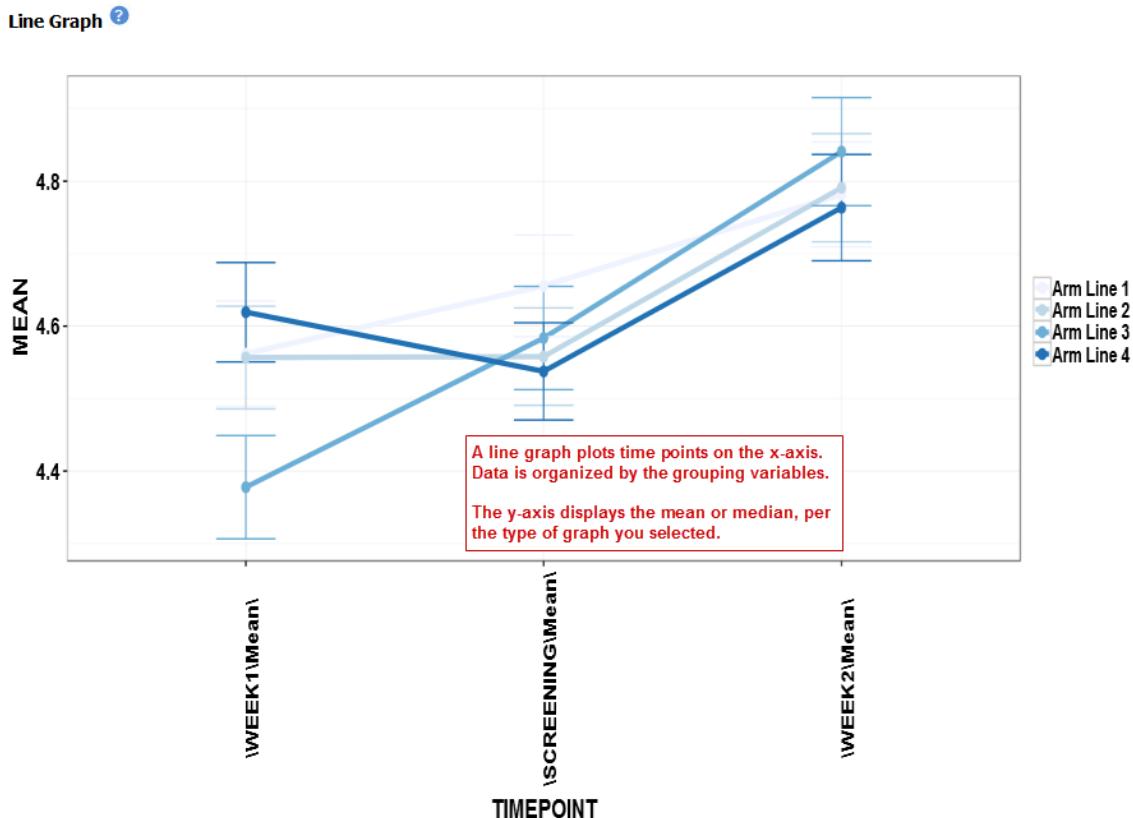
5. Optionally, select one or both of the following:

- Enable binning
- Plot evenly spaced

6. In **Graph Type**, select the type of line graph you want to display.

7. Click **Run**.

Your analysis appears below:



Logistic Regression

Logistic regression is a type of regression analysis used to predict the outcome of a variable that can take on a limited number of categories based on one or more predictors. A logistic regression analysis displays a categorical value predictive of a numerical value.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a logistic regression analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.

2. Select **Logistic Regression**.

The Variable Selection section appears.

3. Define the **Independent Variable** and the **Outcome** variables, following the instructions above the entry boxes.



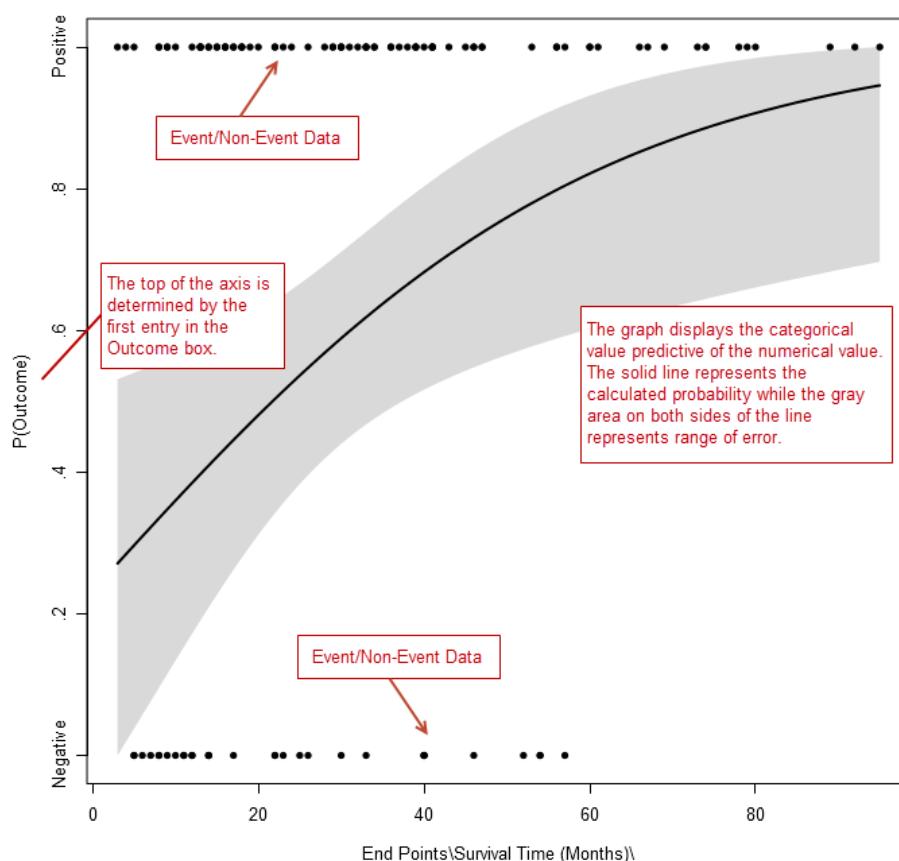
The categorical Outcome variable must use two — and *only two* — nodes.

The top of the logistic regression plot is determined by the first entry in the **Outcome** variable box.

4. Optionally, select **Enable binning**.

5. Click **Run**.

Your analysis appears below. Note that raw data (Event/Non-Event data) is plotted along the top and bottom of the analysis.



Logistic Regression

Logistic Regression Result

Model	binomial generalized linear model glm(Outcome~Independent)			
Coefficients				
	p-Value	Estimate	Z Value	Standard Error
Intercept	0.957	0.0211	0.0543	0.369
Y	0.00865	0.0327	2.63	0.0125
Deviance Residuals				
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-2.01	-1.27	0.675	0.883	1.13
Null deviance: 133 on 112 degrees of freedom				
Residual deviance: 124 on 111 degrees of freedom				
Overall Model p-Value: 0.183				

For every digit increase in the numeric concept, the odds of having the top categorical concept be true go up by $E(Y \ Estimate)$

The null deviance expresses how well the response is predicted by a model with nothing but an intercept

PCA

In a principal component analysis (PCA), the total number of variables in the dataset is reduced to a smaller number of variables – the principle components of the dataset.

Principal component variables are calculated from correlated variables in the total dataset. In other words, the principal component analysis is a workflow used to identify variance in a dataset. The analysis can be run on an entire microarray chip, or on a pathway.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.



Only one subset may be specified in this analysis. Information in Subset 2 will be ignored.

To perform a PCA analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **PCA**.

The Variable Selection section appears.

3. Drag a high-dimensional data node () into the Variable Selection box.
4. Click the **High Dimensional Data** button.

The Compare Subsets-Pathway Selection dialog appears.

Running the Analyses

5. Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

6. Click **Apply Selections**.

7. Optionally, select either or both of the following:

- Use experiment/node as variable instead of probe (multiple nodes only)
- Calculate z-score on the fly

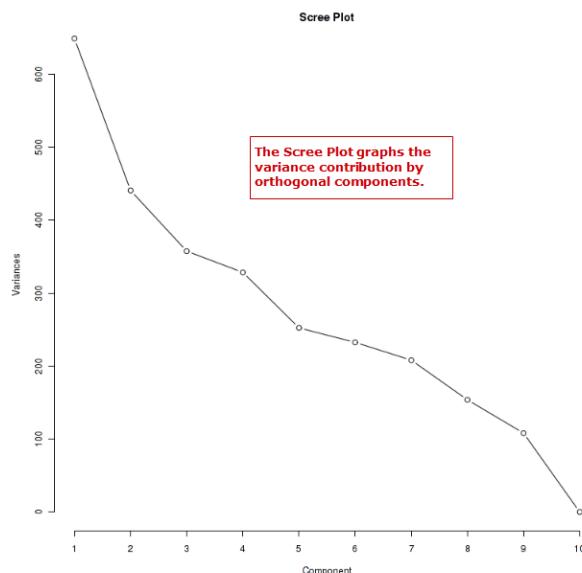
Click **Run**. Your analysis appears below:

Component Summary		
Primary Component	Eigen Value	Percent Variance
PC1	649.09441	23.77064
PC2	440.686	16.13847
PC3	357.5004	13.09211
PC4	328.47809	12.02927
PC5	252.33128	9.24068
PC6	232.52903	8.5155
PC7	208.12087	7.62164
PC8	153.7603	5.63089
PC9	108.15558	3.96079
PC10	0	0

A Principal Component Analysis (PCA) is commonly used as a tool in exploratory data analysis.

Data is split into orthogonal components, and the genes/probes that contribute the most to the components are displayed.

The Component Summary table displays the orthogonal components that your data has been broken into, and how much of the overall variance they are contributing to the variance in the total data set (percent variance).



The Scree Plot graphs the variance contribution by orthogonal components.

Gene list by proximity to Component

Component 1	Component 2	Component 3	Component 4
X5404_OGN -0.061	X4728_ANXAB1 0.06	X185_MKI67 -0.067	X569_YME1L1 -0.068
X1228_OPA1 -0.055	X2045_SORBS2 0.058	X4398_TGFA 0.067	X4278_PPP1R14C 0.067
X2129_DENND4B -0.054	X1094_DST	X4967_CRABP1 0.059	X3571_CNTN3 0.062
X738_LILRA2 -0.053	X1130_MGST1	X2787_KRT17 0.056	X2054_MS4A6A 0.061
X2573_FABP7 0.052	X4340_MMP7 0.055	X2481_SORL1 0.059	X4996_CSM1 -0.06
X4195_CHI3L2 0.052		X4967_CRABP1 0.059	X4689_ACSS3 -0.056

The Gene List table lists genes that make up each orthogonal component.



For more information regarding PCAs, see: <http://psb.stanford.edu/psb-online/proceedings/psb00/raychaudhuri.pdf>.

Scatter Plot with Linear Regression

A scatter plot displays values for two variables within a dataset, with a line that best fits the slope of the data.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a scatter plot with linear regression analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Scatter Plot with Linear Regression**.

The Variable Selection section appears.

3. Define an independent variable and a dependent variable. Both variables should be continuous (for example, Age) and can be high dimensional data.
4. If you included high dimensional data in either variable box, click the **High Dimensional Data** button for that box.

The Compare Subsets-Pathway Selection dialog box appears.

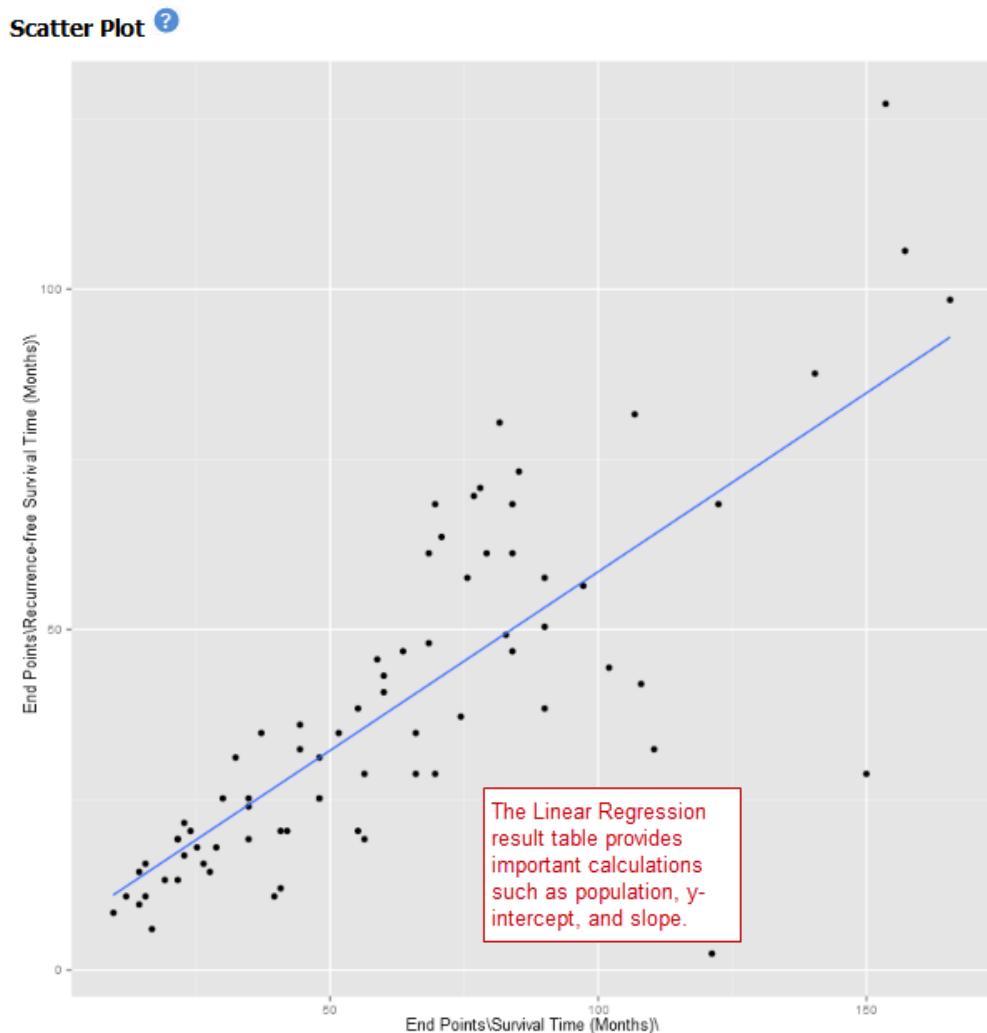
- a. Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

- b. Click **Apply Selections**.

5. Click **Run**.

Your analysis appears below:



Linear Regression Result

Number of Subjects	75
Intercept	5.96
Slope	0.525
r-squared	0.603
adjusted r-squared	0.597
p-value	2.78e-16

Log₁₀ Transformation

Often there will be a large spread between values in the x-axis of a scatter plot analysis. You can use the **log₁₀** option to transform the values in the x-axis, making the graph easier to analyze.

To use the log₁₀ transformation:

1. Select the study you want to use and drag it into a Subset Definition box.
2. Select the **Scatter Plot with Linear Regression** analysis.
3. Enter the independent and dependent variables.
4. Check the box next to **Perform log₁₀ transformation on independent variable** (below the **Independent Variable** box):

Independent Variable

Select a continuous variable from the Data Set Explorer Tree and drag it into the box.

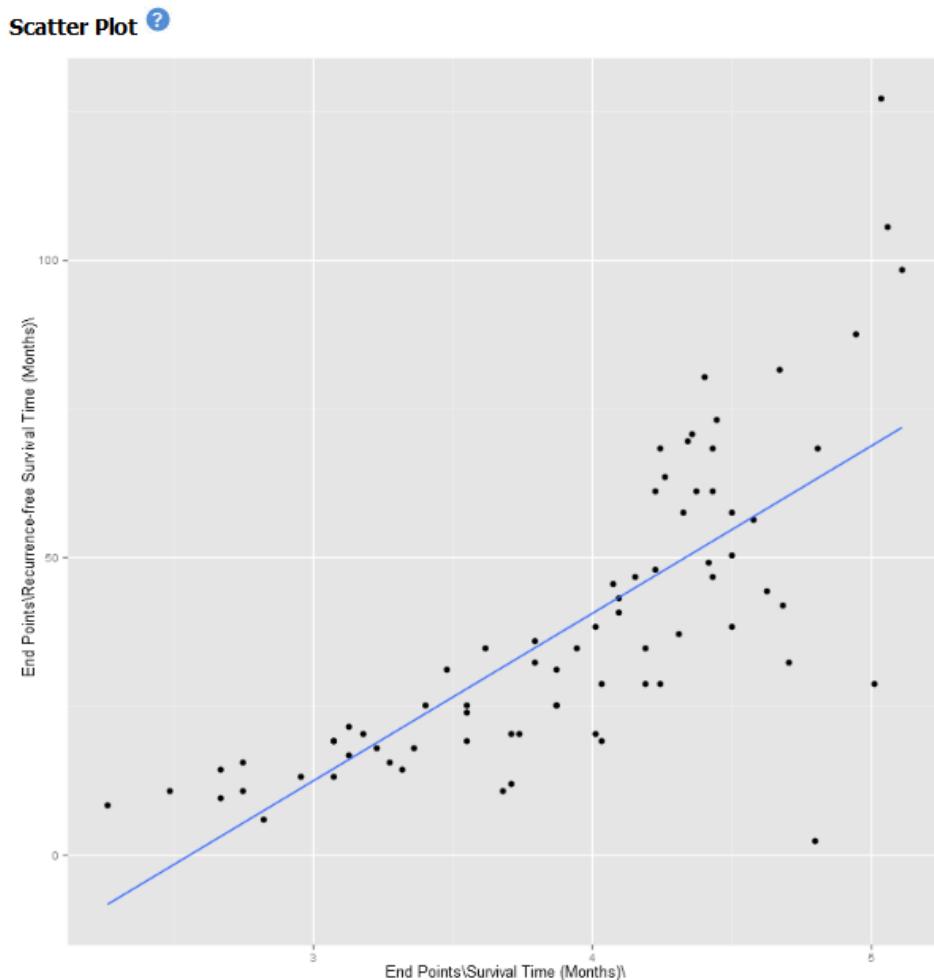


High Dimensional Data

Clear

Perform log₁₀ transformation on independent variable

5. Click **Run**. Your analysis appears below:



Linear Regression Result

Number of Subjects	75
Intercept	-71.9
Slope	28.2
r-squared	0.563
adjusted r-squared	0.557
p-value	9.21e-15

Note the difference between the x-axis on the scatter plot shown previously (no \log_{10} transformation) and the graph shown immediately above. On the first graph, the x-axis values are plotted by multiple of 50 — 50, 100, 150. When the \log_{10} transformation is applied, the x-axis values are plotted per much lower values — 3, 4, and 5. The Linear Regression Result values reflect the recalculated data.

Survival Analysis

A survival analysis displays time-to-event data.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a survival analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select Survival Analysis.

The Variable Selection section appears.

3. Define the following variables:

- Time:** A numerical measure of duration; for example, Overall Survival Time (Years).
- Category:** The groups into which the data will be split in order to compare the time measured; for example, Cancer Stage. This variable is optional. If you do use it, you must enter two nodes for the comparison.

If this variable is continuous, it requires binning.

- Censoring Variable:** Specifies which patients had the event whose time is being measured. For example, if the Time variable selected is **Overall Survival Time (Years)**, an appropriate event variable is **Patient Death**.

4. Optionally, select **Enable binning**.

For details, see [Data Binning Using Survival Analysis](#) on page 87.

Running the Analyses

5. Click **Run**.

Your analysis appears below:

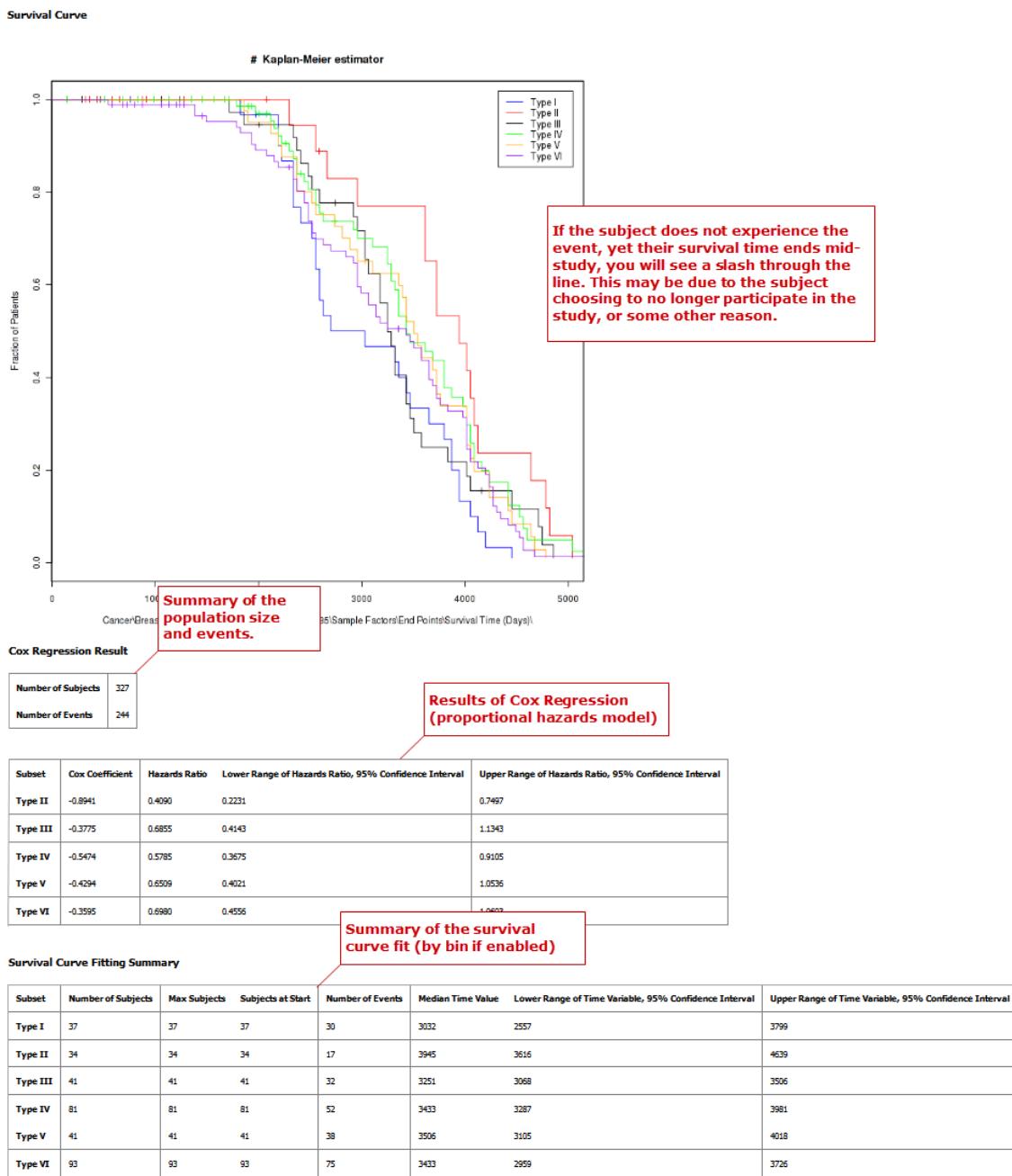


Table with Fisher Test

A Fisher Test analysis examines the significance of associated categorical variables.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To perform a table with fisher test analysis:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.

2. Select **Table with Fisher Test**.

The Variable Selection section appears.

3. Define independent and dependent variables, following the instructions over the **Independent Variable** and **Dependent Variable** boxes.

4. If you included high dimensional data in either variable box, click the **High Dimensional Data** button for that box.

The Compare Subsets-Pathway Selection dialog box appears.

a. Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

b. Click **Apply Selections**.

5. Optionally, select **Enable binning**.

If you select this option, the first, or top, variable in the Dependent Variable box will be used as the conditional variable to calculate the binary outcome. Multiple variables can be categorized into two distinct groups by enabling the Data Binning option. The variable selected in Bin 1 will be used as the conditional variable to calculate the binary outcome.

For information on binning with this type of analysis, see [Data Binning Using Table with Fisher Test](#) on page 88.

6. Click **Run**.

Your analysis appears below:

	Alive	Dead	NA	Total
Lymphoma	249	165	3	417
NA	0	0	3	3
Total	249	165	6	420

Population of each category's members

Fisher test p-value	5e-04
χ^2	208
χ^2 p-value	2.95e-42

Chi-squared distribution values

Table with Fisher Test with Linked Events

If you run the Table with Fisher test analysis using linked events data, the analysis contains two levels for each portion of the analysis: subject-level and event-level.

Using a linked event study, define your variables as described above the **Independent Variable** and **Dependent Variable** boxes. Then click **Run** to create the analysis.

Note that there are now two sets of results for each type of data presented.

Subject Level			
	No	Yes	Total
Bleeding	747	748	1495
Headache	761	768	1529
Total	1508	1516	3024

Subject Level:
Population of each category's members

Event Level			
	No	Yes	Total
Bleeding	1904	1966	3870
Headache	1946	1989	3935
Total	3850	3955	7805

Event Level:
Population of each category's members

Subject Level	
Fisher test p-value	0.942
χ^2	0.00506
χ^2 p-value	0.943

Subject Level:
Chi-squared distribution values

Event Level	
Fisher test p-value	0.839
χ^2	0.0409
χ^2 p-value	0.84

Event Level:
Chi-squared distribution values

Waterfall Plot

A waterfall plot displays a bar chart where a single bar represents each sample in a cohort. Bars are sorted by selected variables and displayed in ascending order. You can further refine the display by specifying ranges that will shade bars accordingly.

To begin the analysis, see [Running the Analyses](#) on page 46, then perform the following steps.

To generate a waterfall plot:

1. Click the **Advanced Workflow** tab, then open the **Analysis** menu.
2. Select **Waterfall**.

The Variable Selection section appears.

3. Define the required variable by selecting a continuous data node from the Dataset Explorer tree and dragging it into the Data Node definition box:

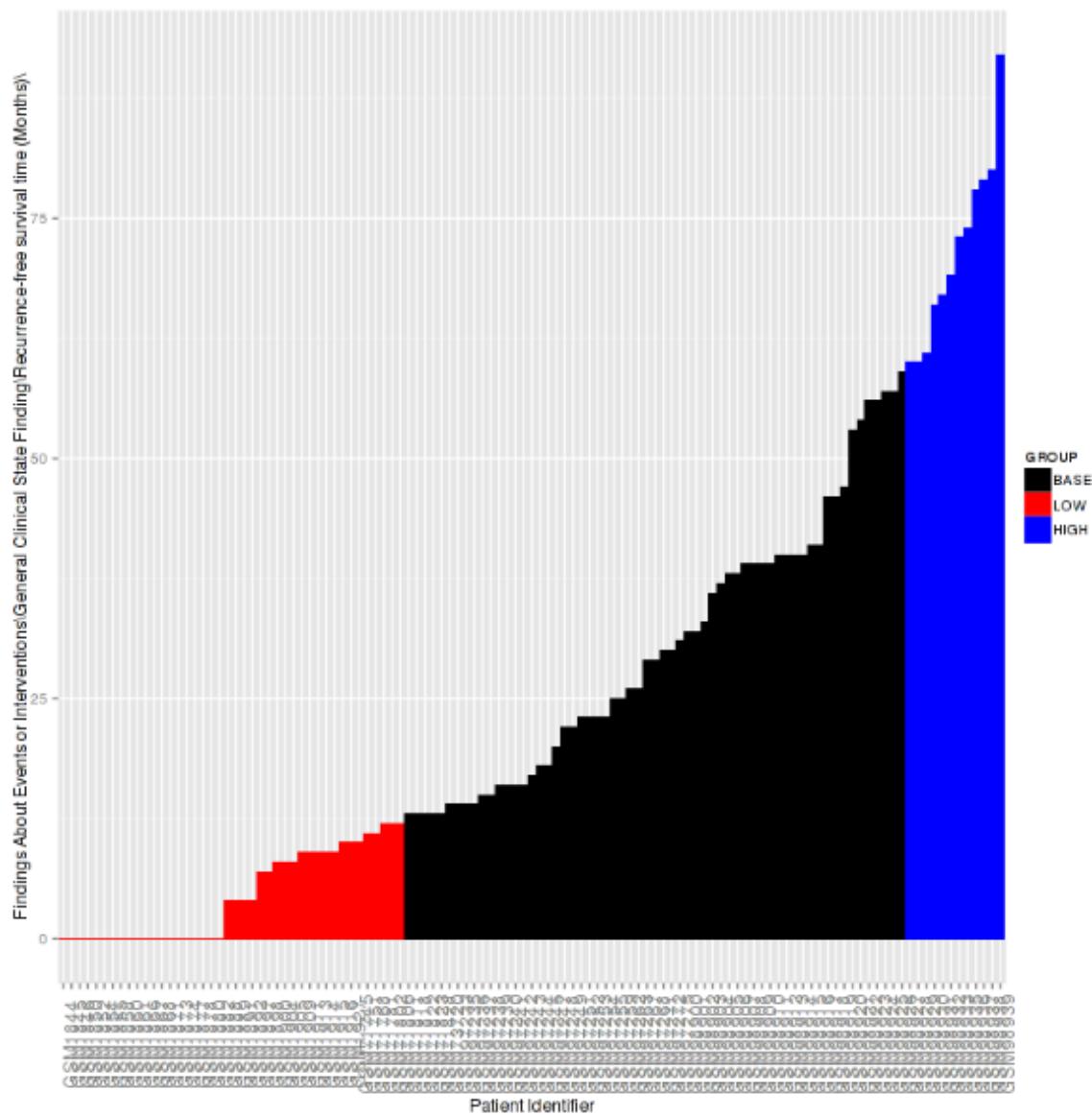


Continuous data nodes are indicated by the **(123)** icon to the left of study data.

4. In **Low Range**, select the appropriate operator from the dropdown menu, then type the value of the low range.
5. In **High Range**, select the appropriate operator from the dropdown menu, then type the value of the low range.
6. Optionally, if you would like the variable, as well as the specified ranges, to appear within separate subsets in the **Comparison** tab, click **Select inputs as Cohort**.

7. Click **Run**.

Your analysis appears below:



High Dimensional Data

The High Dimensional Data button available within the Advanced Workflow section of Analyze allows you to specify additional inputs for selected variables. These inputs help filter specific information of value (such as platforms, samples, and genes or pathways).



The High Dimensional Data feature must be used when you perform an analysis using high dimensional data (such as SNP, gene expression, RBM, etc.) symbolized by the DNA icon (DNA). Additionally, the High Dimensional Data feature cannot be used without high dimensional data.

When you click the **High Dimensional Data** button while setting up an analysis, the Compare Subsets-Pathway Selection dialog box appears. transSMART will attempt to pre-populate default values in the associated fields of the dialog box based on the underlying data in the variable selection box.

The dialog box has the following filters:

Filter	Description
Marker Type	The platform type (for example, Gene Expression, SNP, mRNA, etc.) used to collect biomarker data in the study.
GPL Platform	The specific name of the platform used in the study.
Sample	The type of sample tested in the study.
Tissue	The type of tissue tested in the study.
Select a Gene/Pathway/mirID/UniProtID	The gene or other item of interest. Separate multiple entries with a comma. If you would like to run the analysis on the entire chip, leave this field blank.
Aggregate Probes?	<p>The checkbox can be selected if the variable chosen is either gene expression data or SNP copy number data.</p> <p>If the checkbox is selected, the algorithm WGCNA (weighted correlation network analysis) is employed. For genes that are comprised of multiple probes, WGCNA selects the probe that best represents the overall expression level or copy number.</p> <p>This checkbox does not apply to all advanced workflows.</p> <p>Note: WGCNA was developed by the Department of Human Genetics at UCLA. For more information, see http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/.</p>

When finished defining the filters, click **Apply Selections**, then continue setting up the analysis in the Variable Selection section.

Data Binning

Data binning refers to a pre-processing technique used to reduce observation errors and to allow continuous variables to become categorical. Clusters of data are replaced by a value representative of that cluster (the central value).



The data displayed after binning represents the data available in the study. If, for example, you have selected to bin based on date range (0-10 years of age), yet there is only data available for subjects eight years old and up, the bin will display the age range as 8-10.

Data Binning Using Box Plot with ANOVA

When conducting a Box Plot with ANOVA analysis, at least one of the variables selected should be a continuous variable (for example, age), and the other should be a categorical value (for example, tumor stage).

A continuous variable can be viewed as a categorical value using the binning feature, described below. Alternatively, binning can be used to regroup categorical data to consider it as a single variable. For example, if histological grade with values such as *Well Defined*, *Moderately Well Defined*, and *Poorly Defined* are selected, you can group *Moderately Well Defined* with *Poorly Defined* and treat them as one group for the purposes of this analysis.

To use the data binning feature with a box plot analysis:

1. Begin to set up a Box Plot with ANOVA analysis by following the instructions in section [Box Plot with ANOVA](#) on page 48.
2. Enable binning by selecting **Enable binning**.
3. Define the following and then click **Run**.

Field	Description	Comments
Variable	Select which variable should define the groups (Independent or Dependent) from the dropdown menu.	If the <i>independent variable</i> defines the groups, boxes will be plotted horizontally. If the <i>dependent variable</i> defines the groups, boxes will be plotted vertically
Variable Type	Select whether the variable you have defined above is continuous or categorical from the dropdown menu.	A continuous variable can be turned into a categorical variable when you use the binning feature.
Number of Bins	Type the number of bins you would like data to be organized in.	This step may require trial and error based on how you want to display data.

Field	Description	Comments
Bin Assignments	Select how you would like data to be binned from the dropdown menu. Note: This feature can only be used when the variable type selected above is continuous.	<ul style="list-style-type: none"> ■ Evenly Distribute Population: Assigns bins based on the underlying data. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-40), (40-80), (81-85), (86-90), (90-92)]. ■ Evenly Spaced Bins: Creates bins based on the overall range of the variable. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-20), (21-40), (41-60), (61-80), (81-100)].
Manual Binning	Select the checkbox if you want to bin manually. Note: This is the only binning method available if you are trying to bin a categorical variable type.	<p>Complete the binning form that populates as a result of checking the Manual Binning box.</p> <ul style="list-style-type: none"> ■ For continuous data: <ul style="list-style-type: none"> ■ For categorical data:

Data Binning Using Forest Plot

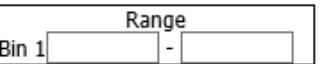
Data binning is used in forest plot analyses if the variable you want to use is continuous (for example, age) but needs to be viewed as categorical data. As an alternative, binning can be used to regroup categorical data to consider it as a single variable. For example, if histological grade with values such as *Well Defined*, *Moderately Well Defined*, and *Poorly Defined* are selected, you can group *Moderately Well Defined* with *Poorly Defined* and treat them as one group for the purposes of this analysis.

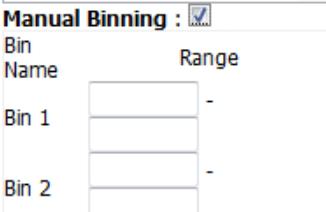
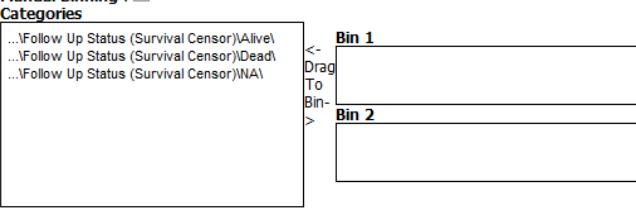
To use the data binning feature with a forest plot analysis:

1. Begin to set up a Forest Plot analysis by following the instructions in section [Forest Plot](#) on page 51.

2. Enable binning by clicking the **Enable** button.

3. Define the following and then click **Run**.

Field	Description	Comments
Variable	Select the variable(s) you want to bin by checking the Bin the [variableType] Variable box next to the appropriate variables. You can bin from none to all four variables.	Example for binning an independent variable: Independent Variable Bin the Independent Variable <input checked="" type="checkbox"/>  Variable Type Continuous 
Variable Type	Select whether the variable you have defined above is continuous or categorical from the dropdown menu.	A continuous variable can be turned into a categorical variable when you use the binning feature.
Number of Bins	Used with the Dependent and Stratification Variables only.	Enter the number of bins into which you would like data to be organized. This step may require trial and error based on how you want to display data.
Bin Assignments	Select how you would like data to be binned from the dropdown menu. Note: This is only an option when binning a continuous variable in the Dependent or Stratification input boxes.	<ul style="list-style-type: none"> ▪ Evenly Distribute Population: Assigns bins based on the underlying data. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-40), (40-80), (81-85), (86-90), (90-92)]. ▪ Evenly Spaced Bins: Creates bins based on the overall range of the variable. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-20), (21-40), (41-60), (61-80), (81-100)].

Field	Description	Comments
Manual Binning	<p>For Dependent and Stratification variables: Select the Manual Binning checkbox if you want to bin manually.</p> <p>Note: This is the only binning method available if you want to bin a categorical variable.</p>	<p>Complete the binning form that populates as a result of checking the Manual Binning box.</p> <ul style="list-style-type: none"> ■ For continuous data:  <ul style="list-style-type: none"> ■ For categorical data: 

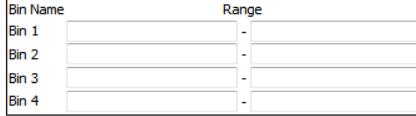
Data Binning Using Survival Analysis

Data binning is used in survival analyses if the variable you want to use is continuous (for example, age) but needs to be viewed as categorical data. Alternatively, binning can be used to regroup categorical data to consider it as a single variable. For example, if histological grade with values such as *Well Defined*, *Moderately Well Defined*, and *Poorly Defined* are selected, you can group *Moderately Well Defined* with *Poorly Defined* and treat them as one group for the purposes of this analysis.

To use the data binning feature with a survival analysis:

1. Begin to set up a Survival Analysis by following the instructions in section [Survival Analysis](#) on page 77.
2. Enable binning by selecting **Enable binning**.
3. Define the following and then click **Run**.

Field	Description	Comments
Variable Type	Select whether the variable you have defined above is continuous or categorical.	A continuous variable can be treated as a categorical variable when you use the binning feature.

Field	Description	Comments
Number of Bins	Type the number of bins you would like data to be organized in.	This step may require trial and error based on how you want to display data.
Bin Assignments	Select how you would like data to be binned. Note: This feature can only be used when the variable type selected above is continuous.	<ul style="list-style-type: none"> Evenly Distribute Population: Assigns bins based on the underlying data. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-40), (40-80), (81-85), (86-90), (90-92)]. Evenly Spaced Bins: Creates bins based on the overall range of the variable. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-20), (21-40), (41-60), (61-80), (81-100)].
Manual Binning	Select the checkbox if you want to bin manually. Note: This is the only binning method available if you are trying to bin a categorical variable type.	<p>Complete the binning form that populates as a result of checking the Manual Binning box.</p> <ul style="list-style-type: none"> For continuous data:  <ul style="list-style-type: none"> For categorical data: 

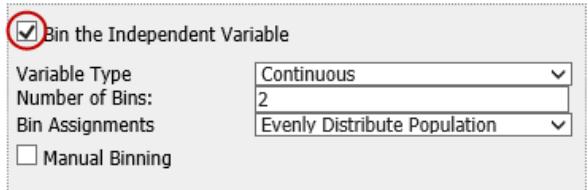
Data Binning Using Table with Fisher Test

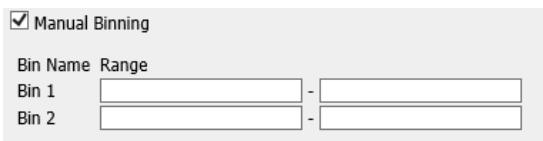
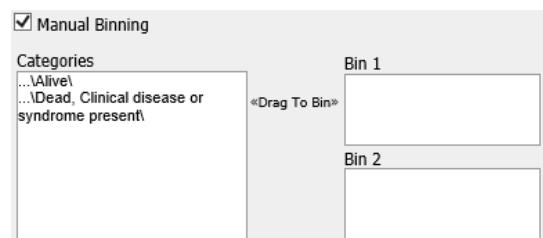
Data binning is used in Fisher Test analyses if the variable you want to use is continuous (for example, age) but needs to be viewed as categorical data. Alternatively, binning can be used to regroup categorical data to consider it as a single variable. For example, if histological grade with values such as *Well Defined*, *Moderately Well Defined*, and *Poorly Defined* are selected, you can group *Moderately Well Defined* with *Poorly Defined* and treat them as one group for the purposes of the analysis.

To use the data binning feature with a Fisher Test analysis:

- Begin to set up a Table with Fisher Test analysis by following the instructions in section [Table with Fisher Test](#) on page 78.
- Enable binning by selecting **Enable binning**.

3. Define the following and then click **Run**.

Field	Description	Comments
Variable	Select the variable(s) you want to bin by checking the Bin the [Variable Type] Variable box next to the appropriate variables. You can bin from none to both variables.	Example for binning an independent variable: 
Variable Type	Select whether the variable you have defined above is continuous or categorical.	A continuous variable can be treated as a categorical variable when you use the binning feature.
Number of Bins	Type the number of bins you would like data to be organized in.	This step may require trial and error based on how you want to display data.
Bin Assignments	Select how you would like data to be binned. Note: This feature can only be used when the variable type selected above is continuous.	<ul style="list-style-type: none"> ▪ Evenly Distribute Population: Assigns bins based on the underlying data. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-40), (40-80), (81-85), (86-90), (90-92)]. ▪ Evenly Spaced Bins: Creates bins based on the overall range of the variable. For example, if the majority of the subjects in the study were elderly, bins based on age could look like: [(1-20), (21-40), (41-60), (61-80), (81-100)].

Field	Description	Comments
Manual Binning	Select the checkbox if you want to bin manually. Note: This is the only binning method available if you are trying to bin a categorical variable type.	Complete the binning form that populates as a result of checking the Manual Binning box. ■ For continuous data:  ■ For categorical data: 

Running Across-Trial Analyses

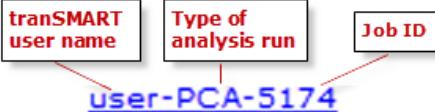
You run analyses based on cohorts defined from the Across Trials folder just as you do analyses based on cohorts defined from single-study folders.

Viewing Recent Analysis Jobs

The **Analysis Jobs** tab allows you to review analyses you have run previously, and also to see the status of analyses you have chosen to run in the background.

Each advanced workflow that you have run in the past seven days is logged in the Jobs tab in a spreadsheet format.

The columns of information in the Analysis Jobs tab are described below:

Column	Description
Name	The name of the analysis run. The format of the name is as follows: 

Column	Description
Status	The status of the analysis. Statuses are explained below: <ul style="list-style-type: none"> ▪ Completed — The job has finished and a visualization is available. ▪ Started — The job has been started and is still processing. ▪ Uploading File — You have selected to load additional data into your visualization, and the data is still in the process of uploading to tranSMART. ▪ Error — The job did not complete due to an error. ▪ Cancelled — The job was cancelled and will not complete.
Run Time	The time the analysis took to process.
Started On	The date and time that the analysis was first started.



Click the **Refresh** button to view any changes that have been made since the Analysis Jobs tab initially populated:



Viewing a Logged Job

Each advanced analysis that you have run in the previous seven days will be logged in the **Analysis Jobs** tab. You may view the visualization again by selecting it from the list.

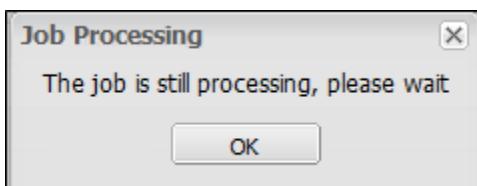
To run a logged advanced workflow:

1. In Analyze, click the **Analysis Jobs** tab:
2. Click the hyperlink of the analysis you are interested in viewing:

Name	Status	Run Time	Started On
user-Compare_5221	Completed	8.707 seconds	2011-01-01 00:00:00.000
user-Select-5207	Started		2011-01-01 00:00:00.000
user-PCA-5179	Uploading file		2011-01-01 00:00:00.000
user-Select-5207	Error		2011-01-01 00:00:00.000
user-PCA-5174 68	Started		2011-01-01 00:00:00.000
user-PCA-5179	Error		2011-01-01 00:00:00.000



If you click on a job that has not been completed, the following dialog box appears:



Chapter 6

Third-Party Analysis Tools

The following third-party tools are embedded into tranSMART Analyze:

- [Dalliance Genome Browser](#) (page 93)
- [MetaCore Enrichment Analysis](#) (page 95)

Dalliance Genome Browser

The Dalliance Genome Browser allows you to visualize genomic data, compare genomic variants between different patient cohorts, compare different types of genomic information, and compare the data with public genomic information, such as COSMIC variations.

To view tranSMART data in the Genome Browser:

1. In **Analyze**, open the study of interest, or open the Advanced Trials folder to run an analysis of data from multiple studies.
2. Define your cohort(s) as described in [Defining the Cohorts](#) on page 21.
3. Click the **Genome Browser** tab to display the data in the Genome Browser:



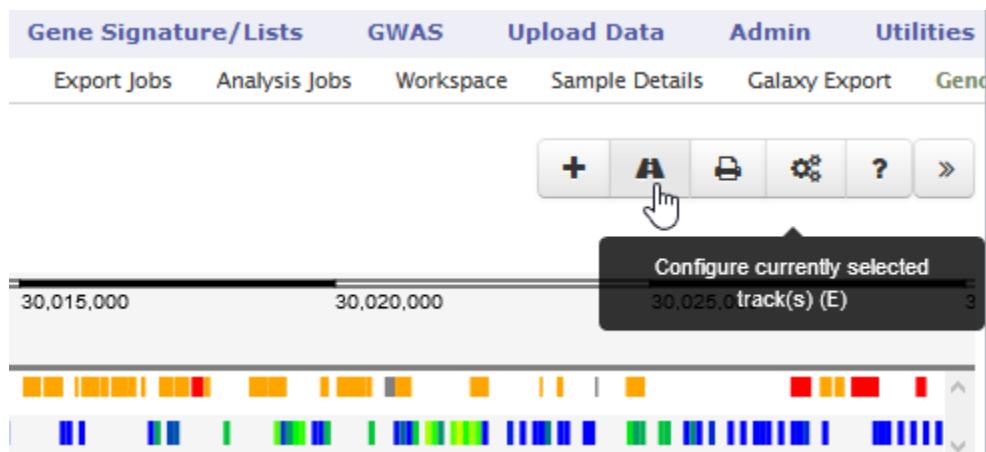
4. Optionally, to add additional data from the study data, drag the concepts of interest from the study into the Genome Browser.

Quick Tour

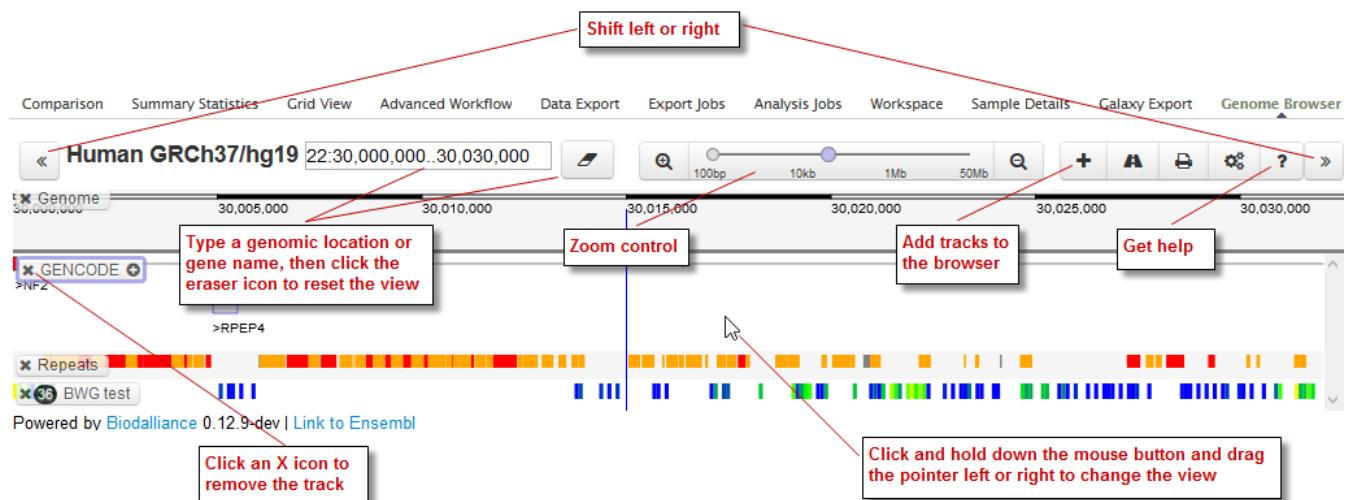
Information about the Genome Browser is located here:

- Getting Started: <http://www.biodalliance.org/started.html>
- Adding Data: <http://www.biodalliance.org/adding.html>

To see a description of a UI control, hover the mouse pointer over the control:

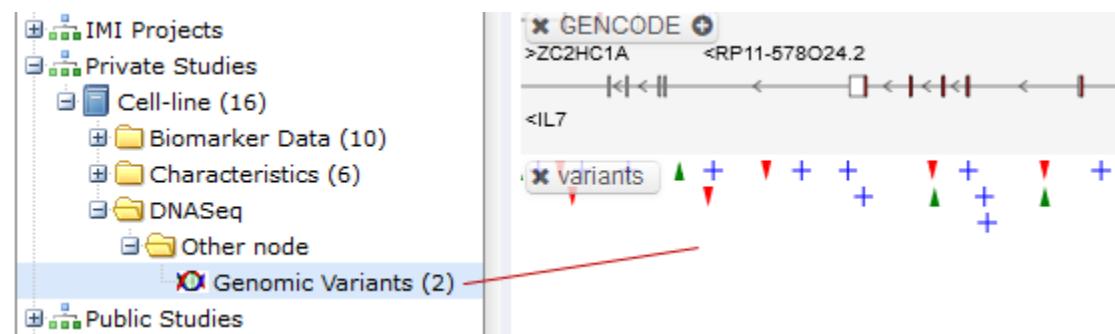


Some highlights of the UI are shown below:



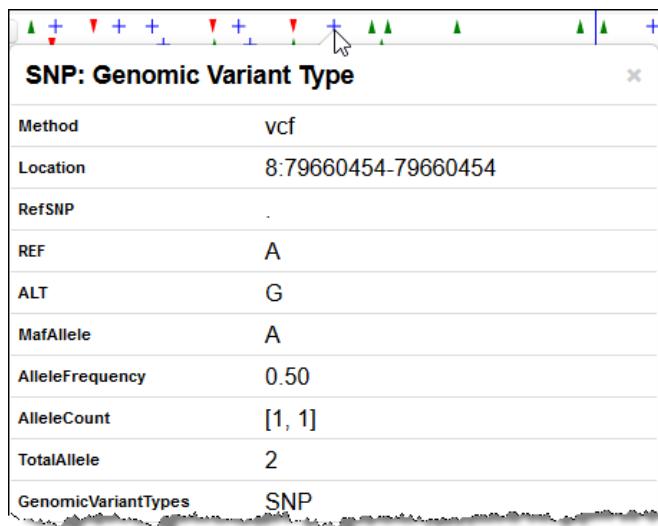
Genomic Variants

If the study has a genomic variants concept, drag it into the browser for a visualization of variants such as SNPs, unidentified mutations, insertions, and deletions:



Note that:

- A green upward-pointing arrow represents an insertion.
- A red downward-pointing arrow represents a deletion.
- Clicking an icon displays details about the variant. Below, a blue plus-sign icon is clicked to display details about a SNP:



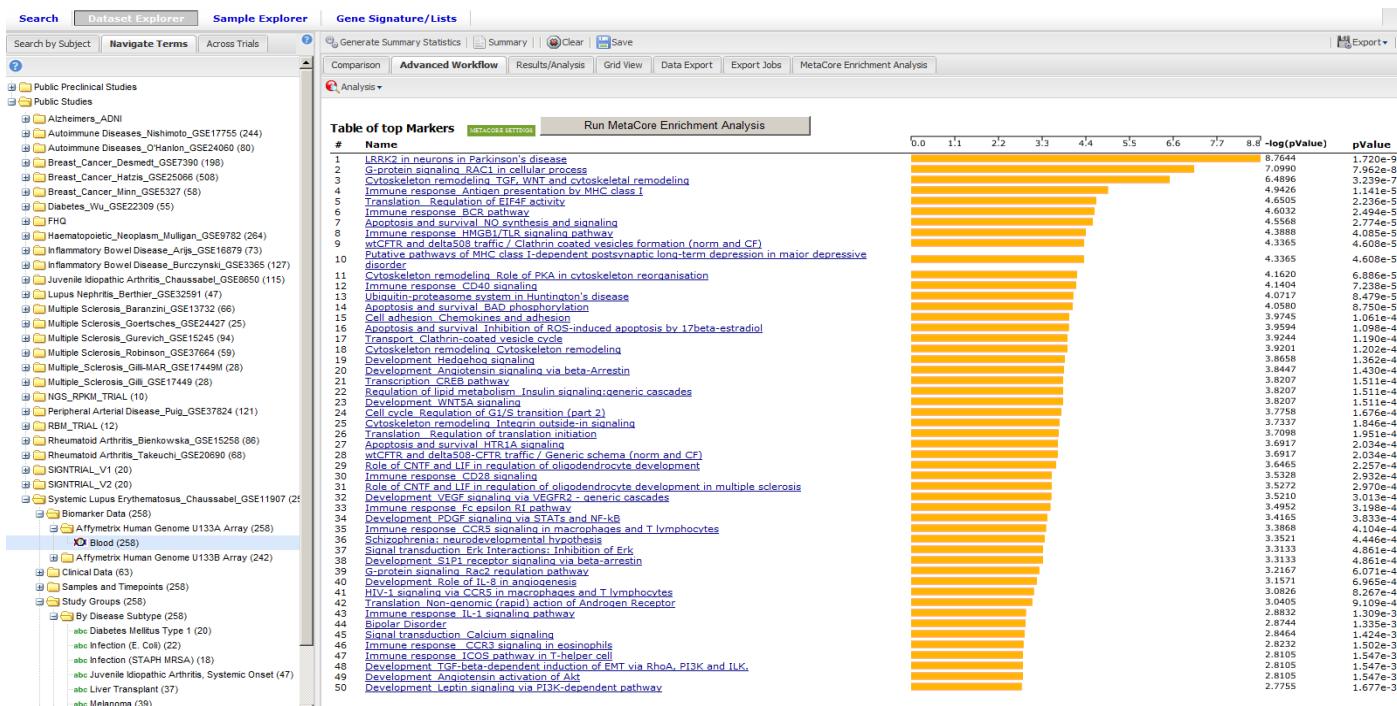
MetaCore Enrichment Analysis

Enrichment analysis is one of the main approaches to understanding the biology behind genes or an expression profile, finding the most significant pathways and processes related to a studied phenotype, validating the relevance of a gene signature, and other use cases.

The enrichment analysis plug-in offered by Thomson Reuters for tranSMART includes options for publically available or the entire portfolio of MetaCore pathway maps, which can be further enhanced by Specialty Module pathway maps created for different diseases.

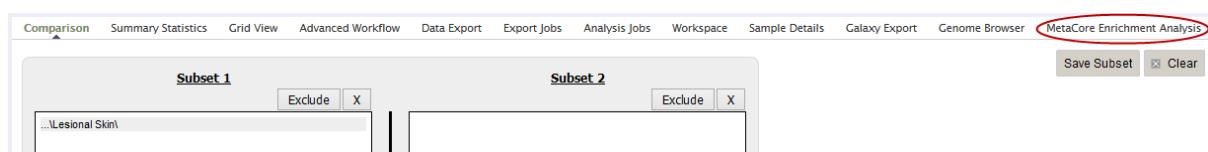
In the following figure, a histogram represents enrichment analysis results across MetaCore and disease pathway maps in the tranSMART interface.

MetaCore Enrichment Analysis



To perform a MetaCore Enrichment Analysis:

- Define a cohort as described in [Defining the Cohorts](#) on page 21.
- Click the **MetaCore Enrichment Analysis** tab:



- Drag a high-dimensional data node (●) into the Variable Selection box.
- Click the **High Dimensional Data** button.

The Compare Subsets-Pathway Selection dialog appears.

- Specify the platform and other filters for the analysis.

For information, see [High Dimensional Data](#) on page 83.

- Either click **Run Workflow** to run the analysis now, or click **Apply Selections** to define more parameters for the analysis and continue with the steps below.
- Optionally, specify the z-score threshold for the data.
- Optionally, click **MetaCore Settings** to view your settings.
- Click **Run** to run the analysis.

MetaCore Enrichment Analysis Based on Marker Selection Data

The enrichment analysis feature complements the Marker Selection advanced workflow (see [Marker Selection](#) on page 64) by providing enrichment of a gene list generated by the workflow to evaluate the significance of the genes to the studied phenotype and/or patient cohort.

Configuration

MetaCore Enrichment Analysis is an additional grails plugin. It is attached to a project in `BuildConfig.groovy`:

```
plugins {
    ...
    if (!dm) {
        ...
        runtime ':transmart-metacore-plugin:1.2.2-SNAPSHOT'
        ...
    } else {
        ...
    }
}
```

For both free and MetaCore enrichments, you need to specify the following line in your `~/.grails/transmartApp/Config.groovy`:

```
com.thomsonreuters.transmart.metacoreAnalyticsEnable = true
```

You don't need any extra settings in order for free enrichment to work.

To use MetaCore's account, create a special table that will store user preferences. Execute `search_user_settings.sql` under `searchapp` or system Oracle user or the appropriate script for PostgreSQL, otherwise you will not be able to use full enrichment functionality. You can find scripts for creating this table here:

https://github.com/transmart/transmart-data/tree/master/ddl/postgres/searchapp/search_user_settings.sql

https://github.com/transmart/transmart-data/blob/master/ddl/oracle/searchapp/search_user_settings.sql

If you want all users to use their personal MetaCore account, you don't need to do anything else. If you want an ability to use a common account for enrichments (users will have a choice), specify the default MetaCore credentials in `~/.grails/transmartApp/Config.groovy`:

```
com.thomsonreuters.transmart.metacoreURL = 'https://portal.genego.com'
com.thomsonreuters.transmart.metacoreDefaultLogin = 'metacore_login'
com.thomsonreuters.transmart.metacoreDefaultPassword = 'metacore_password'
```


Chapter 7

Sample Explorer

Sample Explorer lets you search for tissue and blood samples of interest so that you can learn more about the samples.

The Sample Explorer window has two panes:

■ Right pane

Lets you initiate a search for samples using one or more pre-defined filters. For information, see [Select a Primary Search Filter](#), next.

After you initiate a search, the pre-defined filters are replaced by search results. For information, see [View and Refine Sample Search Results](#) on page 100.

■ Left pane

Reflects the currently selected filters and the number of a filter's samples that appear in the search results.

Select a Primary Search Filter

This pane of the Sample Explorer window lets you initiate a search for samples by selecting the primary search filter. The primary search filters appear when you first open Sample Explorer, or, after you have begun a search, when you click **Clear Search** in the upper-left corner of the pane.

Click a filter to initiate a sample search.

Select a primary search filter

Browse for filter

By BioBank 105 (1) 160 (1) 249 (1) 292 (1) 326 (1) 484 (1) 494 (1) 695 (1) 843 (1) 929 (1)	By Subject Treatment - (10)	By Source Organism human (10)
By Data Type - (10)	By Sample Treatment - (10)	By Tissue blood (6) brain (4)
By Pathology MS (10)	By Data Set EXP:GSE4382 (10)	

Note that the number of samples that are associated with a filter appear in parentheses after the filter name.

View and Refine Sample Search Results

After you select a [primary search filter](#):

- The right pane displays the search results.
- The left pane contains all selected and de-selected search filters, allowing you to narrow the search results.

The following figure illustrates the sections of the Sample Explorer after a primary search filter has been selected:

The screenshot shows the Sample Explorer window with various search filters selected on the left and a results grid on the right. Red callouts provide instructions for different UI elements:

- Selected search filters:** Points to the sidebar where filters like "By BioBank" and "By Tissue" are listed.
- Clear the search results and re-display the primary search filters.** Points to the top right corner of the main pane.
- Click to the right of any column heading to sort and group results and to select the columns to display in the grid.** Points to the "Tissue" column header in the results grid.
- Click for contact information about the selected sample.** Points to a "Contact" button in the grid header.
- Check or clear a check box to select or de-select a filter.** Points to a checkbox for "blood (0)" in the "By Tissue" section.
- The number of samples associated with this filter that appear in the search results, whether or not the filter is selected.** Points to the "(0)" value next to "blood" in the "By Tissue" section.

BioBank	Subject Treatment	Source Organism	Data Type	Tissue	Pathology	Aliquot Count
929	-	human	-	brain	MS	1
292	-	human	-	brain	MS	1
105	-	human	-	brain	MS	1
494	-	human	-	brain	MS	1

Tasks you can perform in the Sample Explorer window include:

- [Select and remove search filters](#)
- [Sort and group the search results, and add/remove search result columns](#)

Select and Remove Search Filters

You can refine a sample search result by adding and removing search filters, including the primary filter you initially selected. Search filters are listed in the left pane of the Sample Explorer window.

To select or remove a search filter, check or clear the check box next to the filter name.



Clicking a filter name rather than the check box next to the name will select that filter and deselect all currently selected filters.

The filters you select are joined together in a search string by the logical operators **AND** and **OR**, as follows:

- Filters within a filter category (such as **DataType**) are joined by **OR**.
- Filters in different filter categories are joined by **AND**.

For example, the search string for the filter selections illustrated below is:

(RBM OR Gene Expression) AND (Colorectal Cancer OR Gastric Cancer)

By DataType

- RBM (90)
- Gene Expression (691)
- SNP (40)

By Pathology

- Liver, Cancer of (236)
- Colorectal Cancer (194)
- Gastric Cancer (186)
- Rheumatoid Arthritis (90)
- Oesophageal Cancer (36)

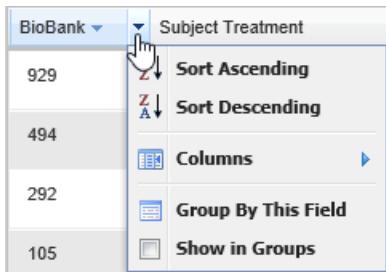
Manage the Search Result List

You can make the following adjustments to the search result list:

Sort by Column

To sort the result list by the contents of a column:

1. Click the right side of the column heading to pull down the menu:

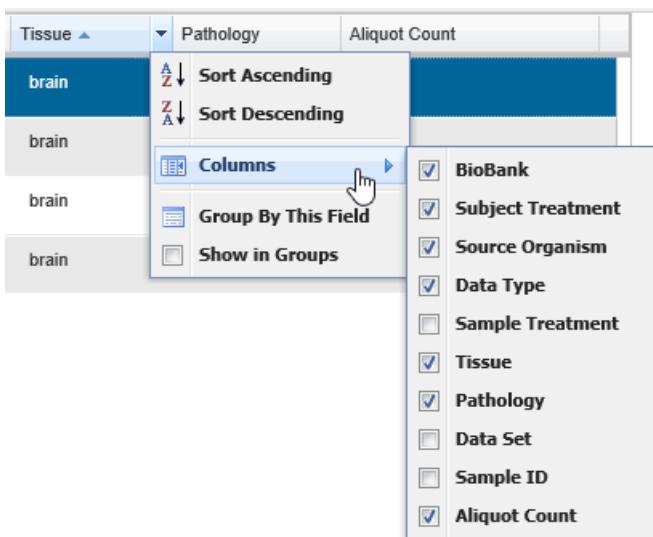


2. Click **Sort Ascending** or **Sort Descending**.

Add and Remove Columns

To add and remove columns:

1. Click the right side of the column header to pull down the menu.
2. Hover the mouse pointer over **Columns** to display the submenu of column headings:



3. Check or clear the check boxes to add or remove columns from the search result.



If the columns bunch together, slightly drag the right edge of the frame around one of the column headings.

Group and Ungroup Search Results

To group search results by the contents of a particular column:

1. Click the right side of the column header to pull down the menu.
2. Click **Group by This Field**.

Tissue	Pathology	Aliquots
brain		
blood		
blood		
brain		
blood	MS	1

To view the samples in all groups:

- Click **Expand All** in the upper-right corner of the search result.

Sample Contact Information Collapse All **Expand All**

BioBank ▲	Subject Treatment	Source Organism	Data Type	Tissue
⊕ Tissue: blood (6 Items)				
⊕ Tissue: brain (4 Items)				

To view the samples in a particular group:

- Click the plus-sign icon next to the group name:

Sample Contact Information Collapse All Expand All

BioBank ▲	Subject Treatment	Source Organism	Data Type	Tissue
⊕ Tissue: blood (6 Items)				
⊕ Tissue: brain (4 Items)				

To ungroup the search result:

- Pull down the menu and click the Show in Groups menu item:

The screenshot shows a list of samples under the heading 'BioBank'. A context menu is open over the first item, 'Tissue: blood (6)'. The menu options are: Sort Ascending, Sort Descending, Columns, Group By This Field, and Show in Groups. The 'Show in Groups' option has a checked checkbox and is highlighted with a blue border. The menu is displayed over several sample entries, with the first entry being '160' and the last visible entry being '843'. Below the main list, there is a section titled 'Tissue: brain (4 Items)'.

Subject	Treatment	Source
160	-	hun
249	-	hun
326	-	hun
484	-	hun
695	-	hun
843	-	hun

Tissue: brain (4 Items)

Chapter 8

Gene Signatures and Gene Lists

The tranSMART gene signature wizard guides you through the process of creating a gene signature or gene list. You specify whether the gene signature or list is publicly available to other tranSMART users or is reserved for your private use.

Once you create the gene signature or list, it can be used in tranSMART searches to find clinical studies and experiments where the differentially regulated genes overlap with the genes contained in the gene signature or list. This will generate a set of hypotheses about diseases or treatments that may have similar genes deregulated, and that can help you develop a further set of experiments.



This chapter uses the term "gene signature" to refer to both gene signatures and gene lists.

Creating a Gene Signature

There are two basic tasks involved in creating a gene signature:

1. Add the list of genes for the gene signature to a text file.
Genes can be indicated by gene symbol or by their associated probe set ID.
2. Use the gene signature wizard to define the information on which the gene signature is based, such as species, source of data, and test type, and also to import into the gene signature definition the text file containing the genes.

Step 1. Adding the Genes to a Text File

The gene signature wizard expects to import the genes for the gene signature from a tab-separated text file. The file must contain one, and possibly two, columns of information:

- First column — A list of gene symbols or probe set IDs.
- Optional second column — The fold change ratios associated with the gene symbols or probe set IDs.

The fold change ratios can be either **actual values** (for example, 12.8 or -12.8) or one of the following **composite values**:

- 1.** All down-regulated gene expressions.
- 1.** All up-regulated gene expressions.
- 0.** No change.

The following table shows the different ways you can specify the genes for your gene signature:

Contents of File	Format	Examples
Gene symbols only	<i>GeneSymbol</i>	TCN1 IL1RN KIAA1199 GOS2
Gene symbols, actual fold change	<i>GeneSymbol</i> <tab> <i>ActualFC</i>	CXCL5 -19.19385797 IL8RB -18.21493625 FPR1 -17.6056338 FCGR3A -15.69858713
Gene symbols, composite fold change	<i>GeneSymbol</i> <tab> <i>CompositeFC</i>	CXCL5 -1 IL8RB -1 MMP3 0 SOD2 1
Probe set IDs only	<i>ProbesetID</i>	224301_x_at 1398191_at Dr.2473.1.A1_at A_24_P93251
Probe set IDs, actual fold change	<i>ProbesetID</i> <tab> <i>ActualFC</i>	224301_x_at - 19.19385797 1398191_at - 18.21493625 Dr.2473.1.A1_at - 17.6056338 A_24_P93251 - 15.69858713
Probe set IDs, composite fold change	<i>ProbesetID</i> <tab> <i>CompositeFC</i>	224301_x_at -1 1398191_at 0 Dr.2473.1.A1_at 1 A_24_P93251 -1

Step 2. Creating the Gene Signature

1. In tranSMART, click the **Gene Signature/Lists** menu.
2. Click the **New Signature** button.

The first page of the gene signature wizard appears:

Gene Signature Create

Instructions ▾

Page 1: Definition:

Signature/List Name*

Description

Note, the creator of this signature will be 'Anthony Ioven' at the current system time

Meta-Data Cancel



Required fields on gene signature wizard pages are marked with a red asterisk (*).

You can find additional information about the gene signature wizard by clicking **Information** on any wizard page.

3. Specify a name (required) and an optional description for your gene signature, then click Meta-Data to proceed to the next gene wizard page.

Creating a Gene Signature

The second page of the gene signature wizard appears:

Gene Signature Create

Instructions ▾

Page 2: Meta-Data:

Source of list	<input type="button" value="select source ▾"/>
Owner of data	<input type="button" value="select owner of the data ▾"/>
Stimulus	i.e. LPS, polyIC, etc: Dose, units, and time: <input type="text"/>
Treatment	Drug treatment used in assay: Dose, units, and time: <i>OR Enter:</i> J&J Compound: <input type="button" value="select compound ▾"/> Protocol Number: <input type="text"/>
PMIDs (comma separated)	<input type="text"/>
Species*	<input type="button" value="select relevant species ▾"/>
Technology Platform*	<input type="button" value="select tech platform ▾"/>
Tissue Type	<input type="button" value="select relevant tissue ▾"/>
Experiment Type	<input type="button" value="select experiment type ▾"/> If applicable, ATCC designation: <input type="text"/>
<input type="button" value="Definition"/> <input type="button" value="Next"/> <input type="button" value="Cancel"/>	

4. Specify values in the required fields **Species** and **Technology Platform**, and also in any other relevant fields, then click **Next** to proceed to the final gene signature wizard page:

5. Specify values in the required field **P-value Cutoff**.
6. In the section **File Upload Information**, describe the text file you created in the section [Step 1. Adding the Genes to a Text File](#) on page 105, using the required fields **File Information** and **Upload File**:

- In the **File schema** section of **File Information**, select **Gene Symbol <tab> Metric Indicator** or **Probe Set Symbol <tab> Metric Indicator**, depending on the method you chose to specify the genes.
- In the **Fold change metric** section of **File Information**, select one of the following choices from the dropdown:

Fold Change Metric Indicator	Description
Actual fold change	The text file contains actual fold change values for each gene symbol or probe set ID.
Not used	The text file contains gene symbols or probe set ID only. There are no associated fold change values.
-1 (down), 1 (up), 0 (optional for unchanged)	The fold change values are not actual values. They simply represent whether the gene expression was down-regulated (-1), up-regulated (1), or unchanged (0).

- In **Upload File**, specify the path and name of the file that contains the genes to import. Use the **Browse** button to select the file from the navigation tree.

- Specify values in any other relevant fields on this gene wizard page, then click **Save** to save the gene signature.

The new gene signature appears in the **Gene Signature List** at the top of the Gene Signature/List view:

Gene Signature List											
My Signatures (1) ▲											
Name	Author	Date Created	Species	Tech Platform	Tissue Type	Public List	Gene	# Genes	# Up-Regulated	# Down-Regulated	
Trainee9 Gene Signature	Training Account	2009-08-08	Human	GPL8300	Lung	No	No	18	7	11	-- Select Action --

Making a New Gene Signature Public

By default, a newly created gene signature is private.

To make a gene signature public:

- In the **Gene Signature List**, click the **Select Action** dropdown to the right of the gene signature you just created.
- Click **Make Public** in the dropdown list:

Gene Signature List											
My Signatures (1) ▲											
Name	Author	Date Created	Species	Tech Platform	Tissue Type	Public List	Gene	# Genes	# Up-Regulated	# Down-Regulated	
Trainee9 Gene Signature	Training Account	2009-08-08	Human	GPL8300	Lung	No	No	18	7	11	-- Select Action --

Public Signatures (11) ▼											
Gene Signature List											
My Signatures (1) ▲											

- Select Action --
- Select Action --
- Clone
- Delete
- Edit
- Edit Items
- Excel Download
- Make Public**



After you click **Make Public**, the value in the **Public** column for the gene signature changes from **No** to **Yes**:

Gene Signature List											
My Signatures (1) ▲											
Name	Author	Date Created	Species	Tech Platform	Tissue Type	Public List	Gene	# Genes	# Up-Regulated	# Down-Regulated	
Trainee9 Gene Signature	Training Account	2009-08-08	Human	GPL8300	Lung	Yes	No	18	7	11	-- Select Action --



tranSMART users assigned the role `ROLE_ADMIN` have access to both public and private gene signatures.

Performing Actions on Your Gene Signatures

To edit or perform other actions on a gene signature in your gene signature list:

1. In tranSMART, click the **Gene Signature/Lists** menu.

The **Gene Signature List** appears, containing all the genes you have created:

Gene Signature List										
My Signatures (1) ▲										
Name	Author	Date Created	Species	Tech Platform	Tissue Type	Public List	Gene # Genes	# Up-Regulated	# Down-Regulated	- Select Action -
Trainee9 Gene Signature	Training Account	2009-08-08	Human	GPL8300	Lung	Yes	No	18	7	11

2. Click the **Select Action** dropdown for the gene signature you are acting on. The dropdown contains all the actions you can perform on the gene signature:

Action	Description
Clone	Create an exact duplicate of the gene signature definition (<i>except</i> for the text file containing the gene symbols and fold change values), and display the definition in the gene signature wizard. Cloning a gene signature helps you create a new gene signature with a similar definition to an existing one. However, it is expected you will import a different set of genes into the gene signature.
Delete	Delete the gene signature.
Edit	Open the gene signature in the gene signature wizard for editing. The gene signature wizard displays all the information in the gene signature, including the reference to the text file containing the list of genes and fold change values. If you want to choose a different text file, click the following label: Upload New File Only to Override Existing Items ▾ To save any changes you make during editing, you must click the Save button on the third page of the wizard.
Edit Items	Add, delete, or modify one or more genes in the text file containing the gene symbols and fold change values.
Excel Download	Generate the entire contents of the gene signature, including the information in the text file containing the gene symbols and fold change values, to a Microsoft Excel spreadsheet. The gene signature definition and gene symbols/fold change values are written to separate spreadsheets.

Action	Description
Make Public	<p>Make a private gene signature public.</p> <p>Note: To make a public gene signature private, edit the gene signature and set the Public? field to No on the first page of the gene signature wizard:</p> <div style="text-align: center;"> <input type="radio"/> Yes <input checked="" type="radio"/> No </div>

Performing Actions on Other Users' Signatures

You can perform actions on gene signatures that other transSMART users have created. The gene signatures you can access and the actions you can perform on them depend on the role assigned to your transSMART user ID, as follows:

Role	Authorized Actions
ROLE_ADMIN	All actions on all gene signatures, both public and private.
ROLE_SPECTATOR ROLE_STUDY_OWNER ROLE_DATASET_EXPLORER_ADMIN	Only Clone and Excel Download , and only on public gene signatures.

To edit or perform actions on a gene signature other than your own:

1. In transSMART, click the **Gene Signature/Lists** menu.
2. Click **Public Signatures** to open the list of public gene signatures:

Gene Signature List

My Signatures (1)										
Name	Author	Date Created	Species	Tech Platform	Tissue Type	Public List	Gene # Genes	# Up-Regulated	# Down-Regulated	Select Action
Trainee9 Gene Signature	Training Account	2009-08-08	Human	GPL8300	Lung	No	No	18	7	11

Public Signatures (11)



transSMART users assigned the role `ROLE_ADMIN` will see **Other Signatures** instead of **Public Signatures**.

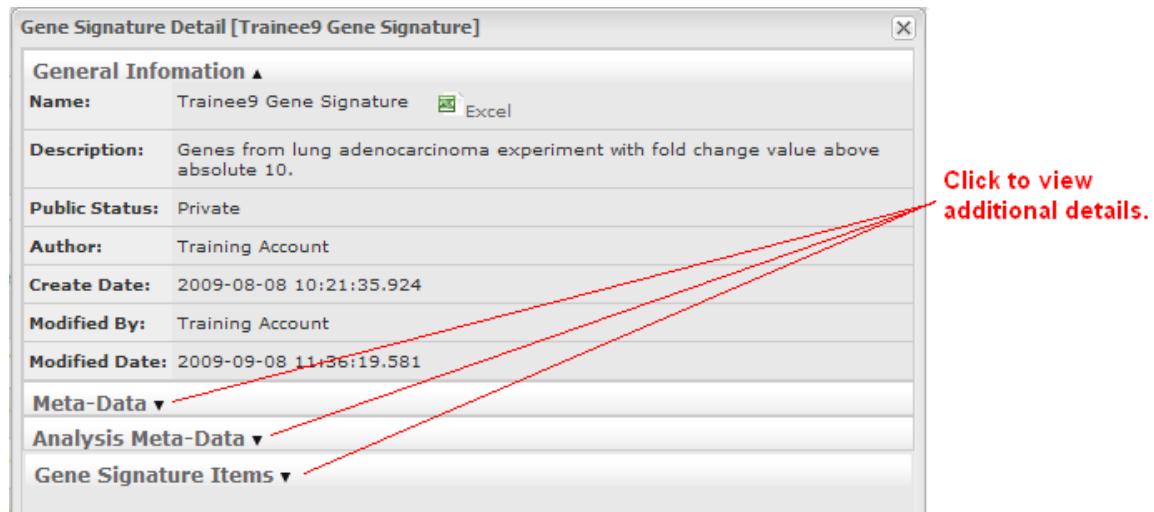
3. Click the **Select Action** dropdown for the gene signature you want to act on.
4. Select the action you want to perform on the gene signature.

Viewing a Gene Signature Definition

You can view the definition of a gene signature, including its list of genes and fold change values, for any gene signature you are authorized to access.

To view a gene signature definition, click the **Detail** icon (grid) next to the gene signature name.

The Gene Signature Detail dialog box appears, containing the gene signature definition:



Chapter 9

GWAS

The genome-wide association study (GWAS) tool is used to find relationships between genetic variants in different individuals and a trait of interest. Though the tool is called "GWAS," the feature supports genome-wide association studies (GWAS), expressional quantitative trait loci (eQTL), and metabolic associations (mQTL).

This application is comprised of a GWAS page to query across GWAS, a data loading panel to import GWAS files, and an interactive Manhattan Plot application called the "Genome-Wide Association Visual Analyzer" (GWAVA).

Searching for a Study

The GWAS search feature is similar to the search feature of the Browse tool. You can build a search query using keywords, pre-defined filters, or any combination.

Features of the GWAS page are shown below:

The screenshot shows the GWAS search interface with several callout boxes highlighting features:

- Select a category to search within, or select All to search all categories.** (points to the 'All' dropdown in the top left)
- Type a search keyword.** (points to the search input field in the top center)
- Tasks you can perform on the items that appear in Analysis View or Table View.** (points to the 'Analysis View' tab in the top center)
- Query results appear below in either Analysis View or Table View.** (points to the bottom section of the interface)
- A dynamically constructed search query appears here as you type search keywords and select search filters.** (points to the search input field)
- Click to open a filter browser in one of these categories and select pre-defined filters.** (points to the 'Filter Browser' section on the left)

Keyword searches and the Active Filters pane work as they do with the Browse tool. For information, see [Defining Search Filters](#) (page 4) and [Managing Active Filters](#) (page 7).

GWAS Filter Browsers

The Filter Browser pane contains the following GWAS filter categories:

Filter Browser	(?)
Analyses	»
Data Type	»
Study	»
Region of Interest	»

There are two types of GWAS filter browsers:

- **Resource** browsers that let you select a filter for a particular tranSMART resource, such as studies, analyses, and data types (for example, GWAS, metabolic GWAS, GWAS Fail, eQTL).
- **Human genome region of interest** browser that lets you specify the region of the human genome to investigate.

Resource Browsers

The resource browsers (studies, analyses, data types) all operate the same way.

To select a filter from a resource browser:

1. Click the name of the filter browser category to open; for example, Study as shown below:

Filter Browser	(?)
Analyses	»
Data Type	»
Study	»
Region of Interest	»

2. In the browser popup window, filter options are listed on the left. Select a filter by clicking the plus sign (+) to the right of the filter name. The selected filter is added to the right part of the browser.

The screenshot shows a browser window titled "Study". At the top, there is a search bar containing the letter "k". Below the search bar, a list of filter options is displayed. One item in the list, "Chronic Kidney Disease Gen", has its plus sign (+) circled in red. The status bar at the top right indicates "0 items selected". At the bottom right of the browser window, there is a large blue button labeled "Select".

With filters that have long names, note that:

- Even if the plus sign (+) is partially obscured by the name, you can still click it.
- Hovering the mouse pointer over the name displays the full name.

3. Optionally:

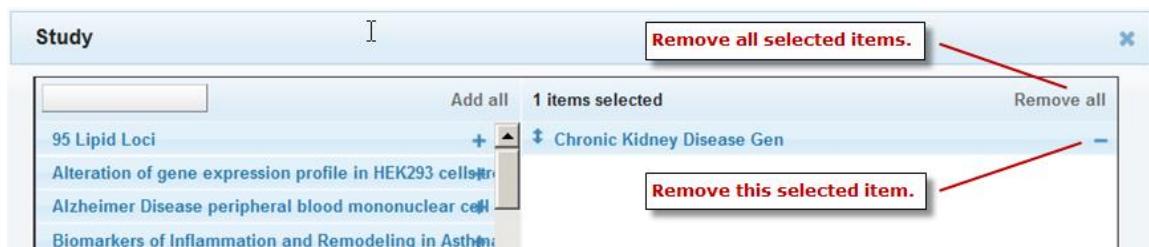
- Add additional filters within this browser by repeating the previous step.
- Narrow the list of filter options by typing characters into the text box at the top left of the browser. transSMART lists all those filters whose names include the characters, in a contiguous string, anywhere in the name (not case-sensitive):

The screenshot shows a browser window titled "Study". At the top, there is a search bar containing the letters "kid". Below the search bar, a list of filter options is displayed. The first item in the list, "Chronic Kidney Disease Gen", has its plus sign (+) circled in red. The status bar at the top right indicates "0 items selected".

- Select all the filters in the list by clicking **Add all**:

The screenshot shows a browser window titled "Study". At the top, there is a search bar containing the letter "k". Below the search bar, a list of filter options is displayed. The status bar at the top right indicates "1 items selected". The selected item in the list is "Chronic Kidney Disease Gen", and its plus sign (+) is circled in red. At the bottom right of the browser window, there is a large blue button labeled "Select".

- Remove a selected filter by clicking the minus sign (-) to the right of the selected filter name, or remove all selected filters by clicking **Remove all**:



4. When finished selecting filters from this browser, click **Select** at the bottom right of the browser. Your selections will be added to the Active Filters area.
5. Optionally, select filters from a different browser by repeating the above steps.

All of the selected filters will become part of the same search query and be included in the Active Filters area.

Region of Interest Browser

The Region of Interest browser lets you specify a particular area of the human genome as a search filter.



Search filters for regions of interest do not filter out studies and analyses that omit the region of interest. However, the only records returned for an analysis are those that contain the specified region of interest. If an analysis does not reference the region of interest, no records are returned for that analysis.

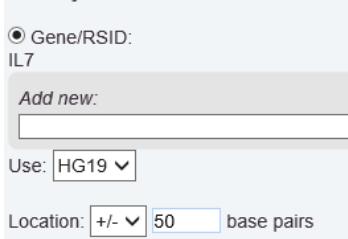
To specify a region of interest:

1. In the Filter Browser pane, click **Region of Interest**.

The Region of Interest browser appears:

The screenshot shows the 'Region of Interest' browser window. It has sections for 'Filter by:' (Gene/RSID and Chromosome), 'Add new:' (text input field), 'Use:' (dropdown set to HG19), 'Location:' (dropdown set to +/-, input field, and base pairs unit), and 'Select' (button). A red box highlights the 'Select' button.

2. Define the filter as described in the following table:

Filter by	Description
Gene	<p>1. Select the Gene/RSID radio button.</p> <p>2. Type one or more characters in the Add new (not case sensitive). transSMART begins to search for gene names or synonyms that begin with the characters you typed.</p> <p>Up to 15 keywords are displayed. If you don't see the one you want, type more characters into the field.</p> <p>3. Click the gene of interest once it has been auto-suggested.</p> <ul style="list-style-type: none"> ▪ To select another gene, repeat the above steps. ▪ To remove a selected gene, click the gene name. <p>4. Optionally, in the Use field, select the Human Genome version to use as the basis of this search. The default is the current version.</p> <p>5. Optionally, in Location, specify the number of base pairs above, below, or both above and below the specified genes to include in the region of interest.</p> <p>If you do not specify a location, the region of interest will be the specified genes only.</p> <p>For example, the following selects a region that spans 50 base pairs above and below the gene IL7, based on Human Genome version 19:</p>  <p>6. Optionally, specify a p-value cutoff in the p-value field. Only those results with a p-value at or below the cutoff are returned. If you do not specify a p-value cutoff, all matches within the region of interest are returned.</p> <p>7. When finished defining the region of interest, click Select. The filter is added to the search query in the Active Filters area.</p>
RS Identifier	Define the region of interest based on an RS identifier the same way you would define one for a gene. In step 2, type the RSID, which consists of the letters rs followed by at least one numeric character.

Filter by	Description
Chromosome	<p>1. Select the Chromosome radio button.</p> <p>2. Select the number of the chromosome of interest from the dropdown list.</p> <p>3. Optionally, in the Use field, select the Human Genome version to use as the basis of this search. The default is the current version.</p> <p>4. Optionally, in the Position text box, type the <i>exact</i> position number of interest.</p> <p>If you do not specify a position, the region of interest will be the entire chromosome.</p> <p>5. Optionally, in the two fields after the Position text box, specify the number of base pairs above, below, or both above and below the specified chromosomal position to include in the region of interest.</p> <p>If you specify a position but not a range of base pairs, the region of interest will be the specified position within the chromosome.</p> <p>For example, the following selects a region of interest that spans the base pair at position 57694854 and the 500 base pairs above it within chromosome 12, based on Human Genome version 19:</p>  <p>6. Optionally, specify a p-value cutoff in the p-value field. Only those results with a p-value at or below the cutoff are returned. If you do not specify a p-value cutoff, all matches within the region of interest are returned.</p> <p>7. When finished defining the region of interest, click Select. The filter is added to the search query in the Active Filters area.</p>

3. Optionally, repeat the above steps to add an additional region of interest to the search query.

Viewing Search Results

Search results appear in the right pane of the GWAS page.

You can view search results in the following forms:

- [Analysis View](#) (page 121)
- [Table View](#) (page 125)
- [Manhattan Plot](#) (page 126)

You can also export Analysis View and Table View data and visualizations.

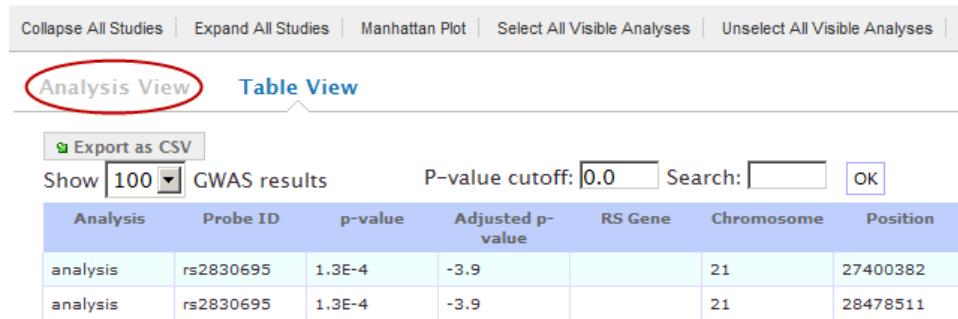
Tabs on the GWAS Page

The following tabs are displayed on the GWAS page:

Tab	Description
Collapse All Studies	Hides the analysis names that appear under the names of listed studies.
Expand All Studies	Lists the names of each study's analyses under the study's name. Only the names of analyses that satisfy the search query in Active Filters are listed.
Manhattan Plot	<p>Launches the external GWAVA application, which will display data from all selected analyses.</p> <p>A selected analysis is one whose check box next to its name is checked:</p>  <p>The screenshot shows a list of analyses under the heading 'Analyses'. One analysis, 'lysine - M01301 - meta analysis', has a red circle around its checkbox, indicating it is selected.</p>
Select All Visible Analyses	Selects the check boxes for all analyses in all listed studies.
Unselect All Visible Analyses	De-selects the check boxes for all visible studies. A de-selected analysis will not appear in a Manhattan Plot.
Add Selected to Filter	Adds all selected analyses to the Active Filters pane.

Analysis View

Analysis View is the default view on the GWAS page. To display this view, click the **Analysis View** button:



The screenshot shows the GWAS Analysis View interface. At the top, there is a toolbar with buttons for 'Collapse All Studies', 'Expand All Studies', 'Manhattan Plot', 'Select All Visible Analyses', and 'Unselect All Visible Analyses'. Below the toolbar, there are two tabs: 'Analysis View' (which is circled in red) and 'Table View'. Underneath the tabs, there are several input fields: 'Export as CSV', 'Show 100 GWAS results' (with a dropdown menu), 'P-value cutoff: 0.0', 'Search: []', and an 'OK' button. Below these inputs is a table with the following columns: Analysis, Probe ID, p-value, Adjusted p-value, RS Gene, Chromosome, and Position. Two rows of data are shown:

Analysis	Probe ID	p-value	Adjusted p-value	RS Gene	Chromosome	Position
analysis	rs2830695	1.3E-4	-3.9		21	27400382
analysis	rs2830695	1.3E-4	-3.9		21	28478511

Tasks

You can perform the following tasks in Analysis View:

- Browse the list of studies, view information about a study, and expand the list of the analyses of a study.
See [Browse the Study List](#) on page 122.
- View metadata for a particular analysis.
See [View Metadata for an Analysis](#) on page 123.
- View the data in a particular analysis, filter the data, export the data to a comma-separated text file, and display the data in a QQ Plot (GWAS data only).
See [View, Filter, and Export Analysis Data](#) on page 123.

Browse the Study List

Before a search query is defined in Active Filters, the GWAS page is displayed in Analysis View with all studies listed. You can view the entire list of studies using the scroll bar on the page.

As you add search filters to the Active Filters area, the studies that appear in the list narrows, based on the search filters you have defined.

You can perform the following tasks for a study:

- View metadata for the study.

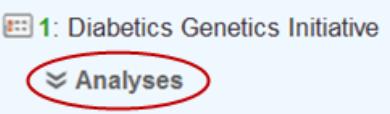
To do so, click the icon to the left of the study name:



Information about the study is displayed, such as the description of the study, the institution that conducted the study, and data availability.

- Expand the study by pulling down a list of the study's analyses that satisfy the current search query in Active Filters.

To do so, click the **Analyses** button under the study name:



Optionally, pull down the analyses for all listed studies by clicking the **Expand All Studies** tab at the top of the Faceted Browser page.

View Metadata for an Analysis

You can view a variety of information about an analysis, such as description of the analysis, type of data collected, data sample size, tissue type, cell type, and analysis platform.

To do so, click the icon to the left of the analysis name:

6: Type 2 Diabetes Meta Analysis

Analyses

- Meta analysis of the DGI and WTCCC pooled controls Type 2 Diabetes - META_T2D - T2D meta
- Meta analysis of the DGI and WTCCC pooled controls Type 2 Diabetes - META_T2D - Meta-wtccc-pooled-controls

View, Filter, and Export Analysis Data

This section describes how to:

- Display the data for a particular analysis of a study.
- Filter the data according to p-value and/or search keyword.
- Export the data to a comma-separated text file.
- Display the data as a QQ Plot.

Typically, before you view analysis data, you will define a search query to narrow the lists of studies and analyses that appear in Analysis View.



To upload analysis data for a study, see Chapter 10: [Data Upload](#).

To view analysis data, and optionally filter and export the data:

1. In Analysis View, navigate to the study that contains the analysis.
2. Click the **Analyses** button under the study name to expand the list of analyses for the study.

A list appears containing the study's analyses:

The screenshot shows a list of analyses under the heading "1: Diabetics Genetics Initiative". The "Analyses" section is highlighted with a red oval. Below it, five items are listed, each with a checkbox and a small icon:

- Broad Body Mass Index (diabetic) - BROAD_BMI_DIAB - Additive
- Broad Body Mass Index (non-diabetic) - BROAD_BMI_ND - Additive
- Broad Insulin - BROAD_INS - Additive
- Broad Fasting Plasma Glucose - BROAD_GLC - Additive
- Broad High-density lipoprotein - BROAD_HDL - Additive

- Click the name of the analysis of interest.

The rows of analysis data appear below the analysis name:

The screenshot shows the "Broad Insulin - BROAD_INS - Additive" analysis results. Several annotations are overlaid on the interface:

- Display the data in a QQ Plot.**: A callout box points to the "QQ Plot" tab at the top.
- Filter rows by search keyword.**: A callout box points to the search bar labeled "Search: [] OK".
- Export the data to a comma-separated text file.**: A callout box points to the "Export as CSV" button.
- Change the number of rows displayed.**: A callout box points to the "Show 10 entries" dropdown.
- P-value cutoff: 0.0**: A callout box points to the "P-value cutoff: 0.0" input field.
- Filter rows by p-value equal to this value or below (0.0 disables this filter).**: A callout box points to the "Search: [] OK" button.
- Display the previous or next set of rows.**: A callout box points to the "Previous" and "Next" navigation buttons.
- Scroll to view more columns of data.**: A callout box points to the scroll bar on the right side of the table.

Probe ID	p-value	Adjusted p-value	RS Gene	Chromosome	Position
rs16954006	2.7030000000000303E-6	5.56815395430127		18	49299819
rs9925729	7.78400000000082E-6	5.1087971727397	CACNG3	16	24330389
rs1568209	9.54500000000098E-6	5.020224067270309	SLCO3A1	15	92581290
rs10490210	1.160000000000038E-5	4.93554201772020			65170482
rs7590631	1.395000000000107E-5	4.85542579			65174408
rs2168144	1.416000000000045E-5	849243560139689		1	110657532
rs11102058	1.530000000000035E-5	770574152079299		1	110666410
rs4641310	2.009000000000047E-5	4.697020063251749	UBL4B	1	110654498

Showing 1 to 10 of 384756 entries [Previous] [Next]

4. Optionally, filter the data results through one or both of the following methods and then click **OK** (do not press Enter or Return):
 - Specify a p-value in the **P-value cutoff** field.
Only those rows whose **p-value** column contains a p-value at or below the specified p-value are returned.
Setting **P-value-cutoff** to **0.0** disables the p-value filter.
 - Specify a search keyword in the **Search** field. All data columns are searchable.
5. Optionally, click **Export as CSV** to export the filtered data to a comma-separated text file.
6. Optionally, click **QQ Plot** to display the filtered data in a QQ Plot. To export the image, click **Export as PNG**.

Table View

Table View lets you perform the following tasks:

- View analysis data from multiple analyses in a single table.
- Filter the rows of analysis data by p-value and/or a search keyword.
- Export the analysis data to a comma-separated text file.



The contents of Table View are determined by the filters in the Active Filters area. Selecting an individual analysis by checking the check box next to the analysis name in Analysis View will not cause the analysis to be included in Table View.

To view analysis data in Table View:

1. Define search filters that will retrieve the records you want to view.



Be sure to filter your search as narrowly as possible. Not only will this result in a table that contains only the most pertinent data, but it will reduce the time required to retrieve and display the data.

2. Click the **Table View** button:

The screenshot shows the GWASome software interface with the following details:

- Top navigation bar: Collapse All Studies, Expand All Studies, Manhattan Plot, Select All Visible Analyses, Unselect All Visible Analyses.
- Tab bar: Analysis View (selected), Table View (circled in red).
- Search results: 1 study with 2 analyses in 0.023 seconds.
- Study list: 6: Type 2 Diabetes Meta Analysis.
- Analysis list: Analyses (with a dropdown arrow).

3. Optionally, filter the data results through one or both of the following methods and then click **OK** (do not press Enter or Return):

- Specify a p-value in the **P-value cutoff** field.

Only those rows whose **p-value** column contains a p-value at or below the specified p-value are returned.

Setting **P-value-cutoff** to **0.0** disables the p-value filter.

- Specify a search keyword in the **Search** field. All data columns are searchable.



Setting a p-value or search keyword in Analysis View for a particular analysis will not filter the data that appears in Table View. To filter Table View records by these parameters, you must define the filters in Table View itself.

4. Optionally, click **Export as CSV** to export the filtered data to a comma-separated text file.

The screenshot shows the GWAS Table View interface. At the top, there are buttons for 'Collapse All Studies', 'Expand All Studies', 'Manhattan Plot', 'Select All Visible Analyses', and 'Unselect All Visible Analyses'. Below these are tabs for 'Analysis View' and 'Table View', with 'Table View' selected. A red box highlights the 'Export as CSV' button. Another red box highlights the message 'These results have been filtered according to gene/chromosome region criteria.' A third red box highlights the 'Filter rows by search keyword.' button. A fourth red box highlights the 'Show [100] GWAS results' dropdown and the 'P-value cutoff: 0.0' input field. A fifth red box highlights the 'Search:' input field and the 'OK' button. A sixth red box highlights the 'Change the number of rows displayed.' dropdown. A seventh red box highlights the 'Export the data to a comma-separated text file.' button. A eighth red box highlights the 'Filter rows by p-value equal to this value or below (0.0 disables this filter.)' message. A ninth red box highlights the 'Display the previous or next set of rows.' button. A tenth red box highlights the 'Scroll to view more columns of data.' message. The main area displays a table with columns: Analysis, Probe ID, p-value, Adjusted p-value, RS Gene, Chromosome, Position, BETA, and DIRECTION_OF_EFFECT. The first row shows data for TCF7L2 on chromosome 10 at position 114788815. The second row shows data for TCF7L2 on chromosome 10 at position 114788815. The bottom of the table shows pagination: 'Showing 1 to 61 of 61 entries' with 'Previous' and 'Next' buttons, and a scroll bar.

Manhattan Plot

You can view GWAS data from selected analyses in a Manhattan Plot. Manhattan Plots are generated by the Genome-Wide Association Visual Analyzer (GWAVA) application.

The GWAVA application lists all GWAS analyses for selection. GWAVA can run and display multiple analyses at the same time on the same Manhattan Plot.

Only standard GWAS data can be viewed in GWAVA. eQTL and mQTL data are not supported.

Alternatively, GWAVA can be launched and used as a separate application independent of the transSMART user interface

To display analysis data in a Manhattan Plot:

1. Optionally, define search filters using the keyword search and Filter Browser features.

Doing so will reduce the number of studies and analyses that you will need to browse through in Analysis View when selecting the analyses to include in the Manhattan Plot.



If you define any gene or gene signature filters, those genes will appear in the GWAVA Gene Model Selection window.

2. In Analysis View, do one of the following:

- Select the check box next to each analysis whose data will be included in the Manhattan Plot:

23: Metabolite

Analyses

lysine - M01301 - meta analysis

At least one analysis must be selected.

- Click the **Select All Visible Analyses** tab to select all analyses for all listed studies.

3. Click the **Manhattan Plot** tab.

4. In the Manhattan Plot Options dialog box, select the human genome version to use as the basis for the selected data, and optionally, specify a p-value cutoff:

Manhattan Plot Options

SNP Annotation Source: Human Genome 19

P-value cutoff: 0.05

Plot

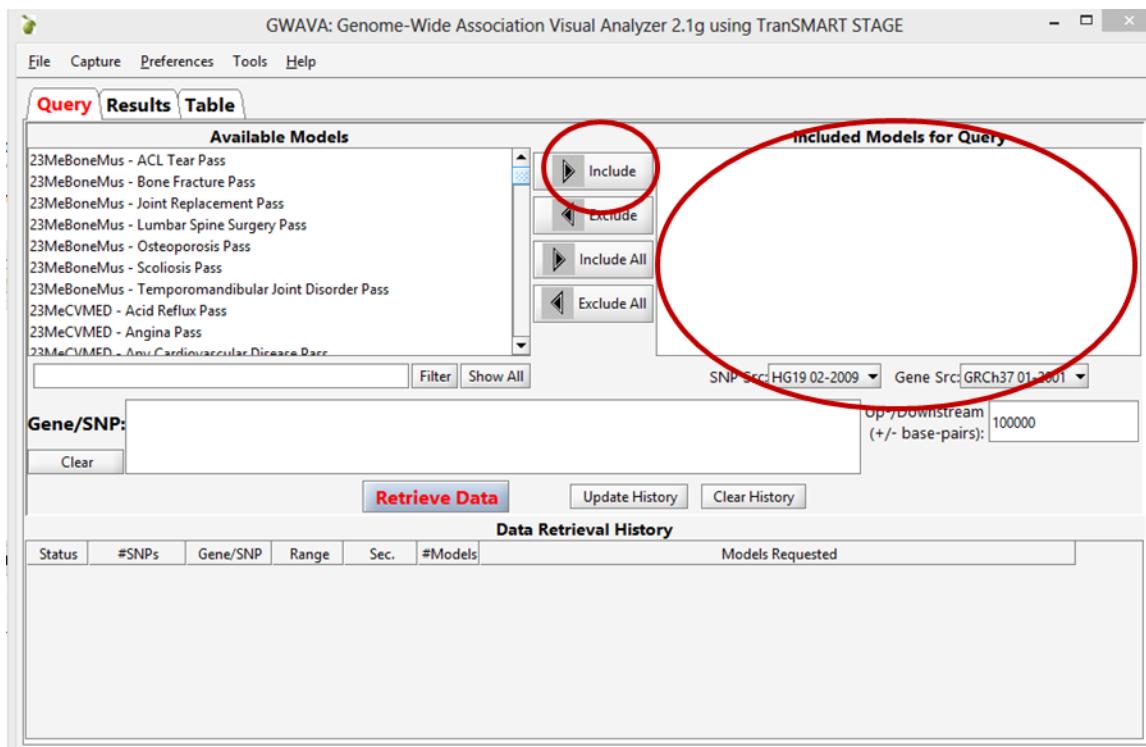
If you specify a p-value cutoff, the only data included in the Manhattan Plot will be from records containing the specified p-value or below.

Viewing Search Results

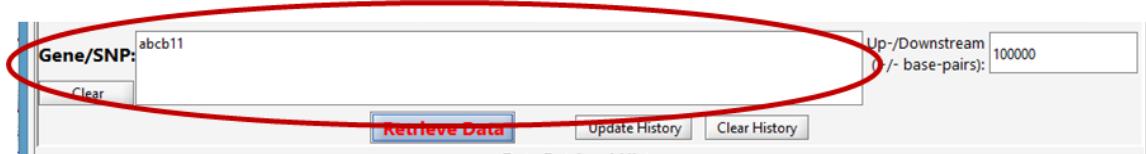
5. Click **Plot**.

The GWAVA application opens.

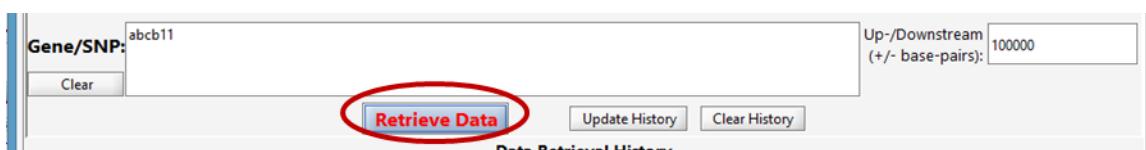
6. In the GWAVA window, select the analysis or analyses of interest from the left and click the **Include** button. The selected items are moved into the panel on the right:



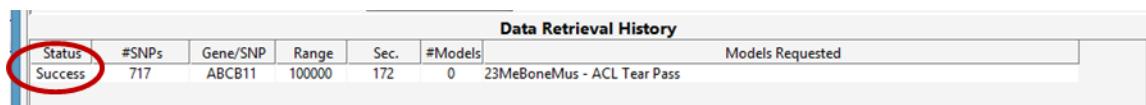
7. Enter a list of genes or RSIDs (one or more):



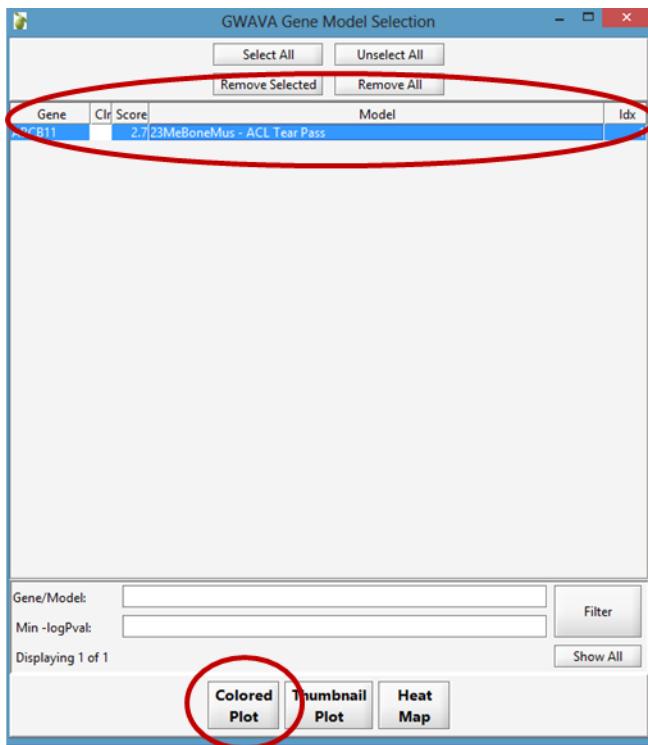
8. Click the **Retrieve Data** button at the bottom to begin running the analysis/analyses.



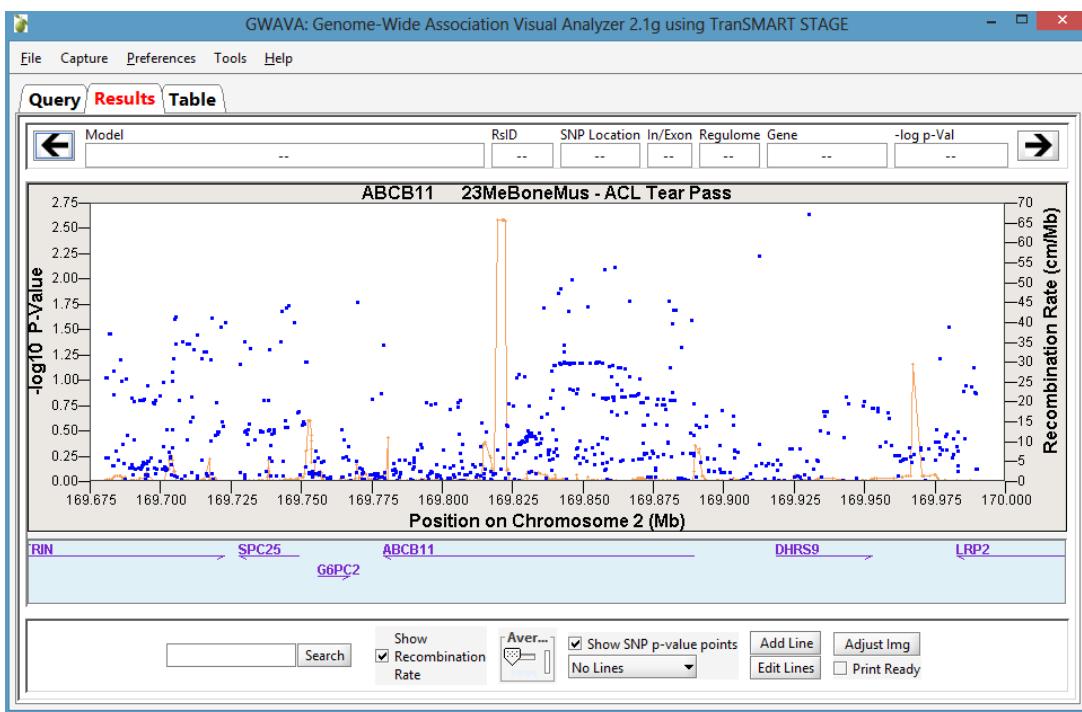
9. Once the job status moves from Working to Success, the job is complete.



10. From the GWAVA Gene Model Selection window, select the completed analysis/analyses and click the **Colored Plot** button.

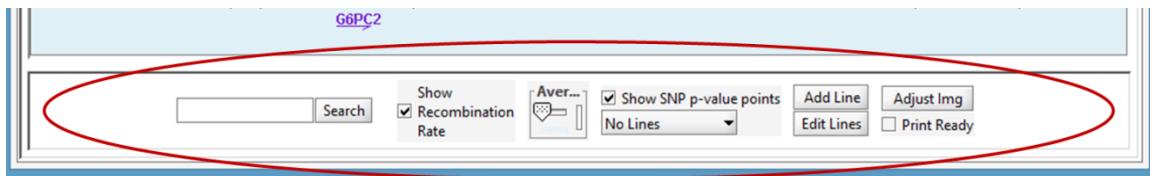


11. The Manhattan Plot is displayed in the Results tab of the GWAVA window:



Viewing Search Results

12. Optionally, manipulate the image using the features on the bottom of the window or by directly highlighting the plot itself. For example, the recombination rate can optionally be displayed, the plot can be zoomed in and out on the chromosomal range, and horizontal lines can be added to the plot as needed.



13. To export, click the **Capture > Save plot to file** menu option.

Chapter 10

Data Upload

The **Upload Data** menu at the top of the tranSMART window lets you upload analysis data for a study. It also lets you define information about the analysis (the analysis metadata), such as analysis name, description, data type, sample size, tissue, phenotype, and so on.

Data that you upload through the two-page data upload form is fully integrated with the studies and analyses that have already been loaded in the tranSMART data warehouse. The data is immediately searchable via tranSMART Browse.

Analysis data is uploaded from a tab-separated text file. To ensure that this file is in the proper format for uploading, use one of the provided templates. For information, see [File Templates](#) on page 133.

Uploading Analysis Data

Before you begin to upload analysis data, you must have a properly formatted file containing the data in a directory on your local computer or on a network server that is accessible from your computer.

To ensure that the data is in a properly formatted file, use one of the file templates provided. For information, see [File Templates](#) on page 133.

To upload analysis data:

1. At the top of the tranSMART window, click the **Upload Data** menu.

The first Upload Data page appears. On this page, the fields **Study**, **Analysis Type to Upload**, and **Analysis Name** are required.

2. In **Study**, select the name of the study associated with the analysis.

To do so, do either of the following:

- Type part of the study name in the text box. tranSMART lists all study names that contain those characters in a contiguous string anywhere in the name (not case sensitive). When you see the name you want, click it. If you do not see the name, type more characters.
- Click the **Browse** button to browse the list of study names. Select the one you want and click **Select**.

The study metadata appears under the study name. If the selected study is not the one you want, click **Change** to select a different name.



If you do not see the study you want, click the **Email administrator** button at the upper right corner of the page to inquire about the study.

- In **Analysis Type to Upload**, select the data type of the analysis data.

- In **Analysis Name**, type a name for the analysis.

This is the name that will appear in the list of analyses for the study.

- In **Analysis Description**, type a description of the analysis.

The description should contain enough information to help a researcher who is scanning the analyses find the ones that are of interest.

- Click **Enter metadata**.

The second page of the data upload form appears. The page name will reference the data type you selected on the previous page.

- In **File**, click the **Browse** button to navigate to the tab-delimited text file that contains the analysis data.

Optionally, click **Download Template** to open or save the template for the data type that you are uploading.

- Complete the upload data form by providing information for the remaining fields on the form, as described in the table below:

Analysis Metadata	Data Type	Description
Disease or Phenotype	All	<p>Type one or more characters in the name of a disease or observation that is relevant to the analysis. transSMART lists the names that contain the characters in a contiguous string anywhere in the name.</p> <p>Click the name of the relevant disease or observation. If you do not see the name you want, type more characters.</p> <p>To add another disease or observation, click Add new.</p>
Population	All	Specify the population of individuals to whom the analysis applies.
Sample Size	All	Specify the number of subjects who were included in the study.
Tissue	All	The type of tissue on which testing was performed.
Cell Type	All	The type of cell on which testing was performed.
(Genotype) Platform: Vendor	All	<p>The genotype platform vendor.</p> <p>To add another vendor and platform, click Add new.</p>

Analysis Metadata	Data Type	Description
(Genotype) Platform	All	The specific genotype platform involved in the study. Platform names will vary with the vendor you select.
Genome Version	All	The human genome version on which the analysis data is based.
Expression Platform: Vendor	eQTL	The gene expression platform vendor. To add another vendor and platform, click Add new .
Expression Platform	eQTL	The specific gene expression platform involved in the study. Platform names will vary with the vendor you select.
Model Name	All	The name of the model that was used to perform the analysis.
Model Description	All	The description of the model.
Statistical Test	All	The statistical test that was used to analyze the data.
P-value cutoff <=	All	The p-value threshold applied to the data being uploaded. The uploaded data contains records with a p-value that is equal to or less than the specified p-value cutoff.
Research Unit	All	The research unit that performed the analysis or to whom the analysis is relevant..

9. Click **Upload** to upload the analysis data and metadata.

File Templates

Analysis data must be contained in a properly formatted, tab-separated text file.

tranSMART includes templates for the supported analysis data types (GWAS, Metabolic GWAS, eQTL). Be sure to use these templates for the analysis data to upload.

The data must contain values in the rsid and p-value columns.

You can download the templates from the data upload form in transSMART. To do so:

1. At the top of the transSMART window, click the **Upload Data** tab.

The first page of the upload data form appears. The templates are available from the second page, so you must provide information on the required fields of this page to proceed to the next.

2. In **Study**, select any study name.
3. In **Analysis Type to Upload**, select the data type for the analysis data that you will upload.
4. In **Analysis Name**, type any name.
5. Click **Enter metadata**.

The second page of the data upload form appears.

6. Click **Download Template** to the right of the File field:

The screenshot shows a web-based form titled "Upload GWAS Data". At the top right is a link "Email administrator". Below the title is a note: "If you are unable to locate the relevant autocomplete fields, email the administrator by clicking the button above". The main area has a "File:" label with a browse button "Browse..." and a "Download Template" button. A red circle highlights the "Download Template" button. Below the file input field is a note: "Upload should be a tab-delimited plain text file".

You will be prompted to open or save the template for the data type you specified in step 3.

7. Click **Save**, specify a location for the file to be saved, and click **Save**.
8. Close the Download dialog box.
9. Click the **Cancel** button on the upload data form.

Chapter 11

Administration

As a tranSMART User Administrator, your responsibilities include the following tasks:

- Adding new users to the tranSMART access list
- Granting permissions to users through role assignments and access rights to private Analyze studies
- Creating user groups, and assigning these groups access rights to private Analyze studies
- Creating and mapping user roles
- Setting up security for private Analyze studies



This chapter is intended for tranSMART User Administrators only. Other users will not be able to see the controls or perform the tasks described in this chapter.

Administrator Privileges

In addition to performing administration tasks, you, as a tranSMART User Administrator, are a tranSMART super-user with access privileges to all tranSMART data sources, including all Analyze studies and both public and private gene signatures. You also have access to the tranSMART access log.

tranSMART user administrators are assigned the role `ROLE_ADMIN`.

The Administrator's Console

To access the console where you perform administrator tasks, click the **Admin** menu:



On initialization, the administrator's console displays the transSMART access log:

Access Time	User	Event	Event Message
2015-02-12 16:49:37.959	admin	UploadData-Index	Upload Data index page
2015-02-12 16:45:01.07	admin	UploadData-Index	Upload Data index page
2015-02-12 16:38:30.809	admin	Login	Mozilla/5.0 (Windows NT 6.3; WOW64; rv:35.0) Gecko/20100101 Firefox/35.0
2015-02-12 16:02:32.358	admin	Login	Mozilla/5.0 (Windows NT 6.3; WOW64; rv:35.0) Gecko/20100101 Firefox/35.0

Tasks in the Administrator's Console

The tasks you can perform as administrator are listed vertically along the left edge of the administrator's console. The following table summarizes these tasks.



The **Secure Object Paths** and **Add Study** tasks are no longer used. While these tasks remain on the Administrator's console at this time, they are not available for use nor are they addressed in any way.

Category	Task	Description
Access Log	View Access Log	Display the transSMART access log.
Groups	Group List	List all user groups and edit or delete groups.
	Create Group	Create a user group.
	Group Membership	Add users to a group or remove users from a group.
Users	User List	List all transSMART users or edit or delete users.
	Create User	Create a transSMART user.
Galaxy Users	User List	List of transSMART users who can export data to Galaxy via the Galaxy Export tab.
	Create User	Associate a transSMART user ID with a Galaxy key.
Access Control	Access Control by Group	Grant users and groups access privileges to private Analyze studies, or remove access privileges for users and groups.

Category	Task	Description
	Access Control by Study	Grant users and groups access privileges to private Analyze studies, or remove access privileges for users and groups.
Study	Study List	List the Analyze studies that are protected by access control.
	Add Study	Not used
Secure Object Paths	SecureObjectPath List	Not used
	Add SecureObjectPath	Not used
Roles	Role List	List all transSMART roles and edit or delete roles.
	Create Role	Create a transSMART role.
RequestMap Setup	Requestmap List	Display mappings between transSMART roles and the transSMART URLs that each role grants access to, and edit or delete mappings.
	Requestmap Create	Create a mapping between a role and a transSMART URL.
Package	Build Information	View details about the current transSMART build.

Managing transSMART Users

Managing users involves the following tasks:

- Creating user accounts
- Editing and deleting user accounts
- Assigning users roles
- Assigning users and groups access rights to private Analyze studies

Understanding User Roles and Access Rights

Users are granted permissions to access private Analyze studies in two ways:

- Through roles
- Through the access level assigned to the user or group for a private study

For information about access levels, see [Access Levels](#) on page 153.

User Roles

When you create or edit a user account, you can assign the user one or more of the roles in the following table.

For information on creating or editing a user account, see [Managing User Accounts](#) on page 140.

Role	Permissions
ROLE_SPECTATOR	<p>tranSMART Search</p> <ul style="list-style-type: none">■ All functions <p>Analyze</p> <ul style="list-style-type: none">■ Access to a private study if the user is assigned a VIEW or EXPORT access level for the study.■ Export ability for a private study if the user is assigned an EXPORT access level for the study.■ Access to all studies in the Public Studies folder. No access level is required. <p>Notes:</p> <ul style="list-style-type: none">■ Users with this role cannot be assigned the OWN access level for a study.■ Assign this role to the user.

Role	Permissions
ROLE_STUDY_OWNER	<p>transSMART Search</p> <ul style="list-style-type: none"> ■ All functions <p>Analyze</p> <ul style="list-style-type: none"> ■ Access to a private study if the user is assigned a VIEW, EXPORT, or OWN access level for the study. ■ Export ability for a private study if the user is assigned an EXPORT or OWN access level for the study. ■ Access to all studies in the Public Studies folder. No access level is required. <p>Note: Private studies are categorized by Centers (CBER, CDER, and CDRH). The role STUDY_OWNER applies at the Center level. For more information, see Access Levels on page 153.</p>
ROLE_DATASET_EXPLORER_ADMIN	<p>transSMART Search</p> <ul style="list-style-type: none"> ■ All functions <p>Analyze</p> <ul style="list-style-type: none"> ■ Access to all studies ■ Export ability for all studies <p>Note: The Analyze administrator has no user administration permissions.</p>
ROLE_ADMIN	<p>transSMART Search</p> <ul style="list-style-type: none"> ■ All functions <p>Analyze</p> <ul style="list-style-type: none"> ■ Access to all studies ■ Export ability for all studies <p>User Administration</p> <ul style="list-style-type: none"> ■ Full user administration functions
ROLE_PUBLIC_USER	<p>transSMART Search</p> <ul style="list-style-type: none"> ■ Search functions against public data only. <p>Analyze</p> <ul style="list-style-type: none"> ■ Access to studies in the Public Studies folder only. ■ Export ability for all public studies. <p>Note: This is a limited access role used for trainee accounts.</p>



For information about creating new roles that you can assign to users, see [User Roles](#) on page 138.

Access Rights to Analyze Studies

Analyze studies can be either public or private. Public studies are in the **Public Studies** folder of the Analyze navigation tree. All other studies are private.

Access rights to public and private studies are as follows:

- **Public studies:** All tranSMART users have full access to the studies in the Public Studies folder. No access level is required for these studies.
- **Private studies:** By default, tranSMART users cannot access private studies. To allow a user to make comparisons between cohorts in a private study, you must grant the user access rights to that particular study.

If a user does not have access rights to a particular private study, the study is grayed out when the user displays the list of studies in the Analyze navigation tree.



Even if the user does not have access rights to a private study, he/she can see a description of the study by right-clicking the study name in the navigation tree and then clicking **Show Definition**.

Managing User Accounts

Creating a User Account

To create a user account:

1. Click the **Admin** menu to display the administrator's console.
2. Under **Users**, click **Create User**.

The Create User window appears:

Create User

WWID:	<input type="text"/>
Login Name:	<input type="text"/>
Full Name:	<input type="text"/>
Password:	<input type="text"/>
Email:	<input type="text"/>
Enabled:	<input type="checkbox"/>
Description:	<input type="text"/>
Show Email:	<input type="checkbox"/>
Assign Roles:	
ROLE_ADMIN	<input type="checkbox"/>
ROLE_STUDY_OWNER	<input type="checkbox"/>
ROLE_SPECTATOR	<input type="checkbox"/>
ROLE_DATASET_EXPLORER_ADMIN	<input type="checkbox"/>
ROLE_PUBLIC_USER	<input type="checkbox"/>

 **Create**

- Provide values for the fields in the Create User window, as follows:

Field	Description	Required
WWID	Unique database identifier.	Yes
Login Name	The user's login ID.	Yes
Full Name	The name to display in the tranSMART window for this user.	Yes
Password	The user's password.	Yes
Email	The user's email address.	No
Enabled	Check this box to enable the user to log into tranSMART. If you leave the box blank, the user's account is disabled, and the user will not be able to log into tranSMART.	No
Description	An optional description of the user. The description appears in the user list (displayed with the User List task in the administrator's console).	No

Field	Description	Required
Show Email	<p>Check this box to display the user's email address.</p> <p>Note: The email display functionality is reserved for future use. Currently, the user's email address is displayed only when you or another User Administrator view or edit a user's account.</p>	No
Assign Roles	<p>Assign one or more roles to the user by checking the boxes next to the names of the roles to assign.</p> <p>If you do not check any of the boxes, the user will not be able to log into tranSMART.</p> <p>Note: For information about the roles you can assign to the user, see the section User Roles on page 138.</p>	Yes

- When you are finished defining the user account, click **Create**.

Editing or Deleting a User Account

To edit or delete a user account:

- Click the **Admin** menu to display the administrator's console.
 - Under **Users**, click **User List**.
- The AuthUser List window appears.
- Click the column heading **Full Name** to sort the list of user names alphabetically.
- Sorting the list may help you find the name in the list of users.



You can sort any of the columns in the AuthUser List by clicking the column heading.

- Locate the name of the user whose account you want to edit or delete.
 - Click **Show** for the account to edit or delete.
- The User window appears.
- Take one of the following actions:

- To delete the account, click **Delete**. Then click **OK** to confirm the deletion.



Deleting a user account does not delete the user's records in the access log.

Records of Analyze studies are independent of any associated user account. For example, if a user is the owner of a particular study, the study remains in Analyze after the user is deleted, even if no other user has access privileges for the study.

- To edit the account, click **Edit**. After making the edits, click **Update**.

Managing Galaxy Users

tranSMART users who have the Dalliance Galaxy data analysis tool installed and configured for use with tranSMART can export data from tranSMART directly into Galaxy. To enable this feature for Galaxy users, you must associate their tranSMART user ID with their Galaxy key.

1. Click the **Admin** menu to display the administrator's console.

2. Under **Galaxy Users**, click **Create User**.

The Create User window appears.

3. Provide values for all the fields in the Create User window, as shown below:

Field	Description
Username of User	The user's tranSMART user ID.
Galaxy Key	The user's Galaxy key.
Email	The user's email address.

4. Click **Create**.

View or Delete Galaxy Users

To view the list of Galaxy users or to delete a Galaxy user:

1. Click the **Admin** menu to display the administrator's console.

2. Under **Galaxy Users**, click **User List**.

The Galaxy User List window appears.

3. To delete a user as a Galaxy user, click **Delete** to the right of the user's row.

Managing tranSMART Roles

A role is mapped to one or more tranSMART URLs. Each URL provides access to a tranSMART resource.

If a user is assigned a particular role, the user is able to access the URL mapped to the role, and therefore, to the resource available through the URL.

For example, the role `ROLE_ADMIN` is mapped to the URL pattern `/authUser/**` on the tranSMART site. At this location, users assigned `ROLE_ADMIN` (that is, administrators like yourself) can view, create, edit, and delete tranSMART user accounts.

A URL pattern can be mapped to one or more roles. Since `/authUser/**` is mapped to no other role than `ROLE_ADMIN`, only users assigned this role can perform tasks on user accounts.

Understanding Role / URL Mappings

Roles are mapped to URLs on the Requestmap List window of the administrator's console:

Requestmap List

ID	URL Pattern	Roles	
1	<code>/requestmap/**</code>	<code>ROLE_ADMIN</code>	Show
2	<code>/role/**</code>	<code>ROLE_ADMIN</code>	Show
3	<code>/authUser/**</code>	<code>ROLE_ADMIN</code>	Show
5	<code>/**</code>	<code>IS_AUTHENTICATED_REMEMBERED</code>	Show
6	<code>/login/**</code>	<code>IS_AUTHENTICATED_ANONYMOUSLY</code>	Show
7	<code>/css/**</code>	<code>IS_AUTHENTICATED_ANONYMOUSLY</code>	Show
8	<code>/js/**</code>	<code>IS_AUTHENTICATED_ANONYMOUSLY</code>	Show
9	<code>/images/**</code>	<code>IS_AUTHENTICATED_ANONYMOUSLY</code>	Show
10	<code>/search/loadAJAX**</code>	<code>IS_AUTHENTICATED_ANONYMOUSLY</code>	Show
1753751	<code>/accessLog/**</code>	<code>ROLE_ADMIN</code>	Show
1753752	<code>/authUserSecureAccess/**</code>	<code>ROLE_ADMIN</code>	Show
1753753	<code>/secureObject/**</code>	<code>ROLE_ADMIN</code>	Show
1753754	<code>/secureObjectPath/**</code>	<code>ROLE_ADMIN</code>	Show

URLs in this window are expressed as fragments of URLs called URL patterns. transSMART determines the full URL to associate with a role by adding the URL pattern to the root URL for the transSMART site. For example, if the transSMART root URL is `https://transmart.mysite.com/transmart` and the URL pattern is `/authUser/**`, the complete URL mapped to the role `ROLE_ADMIN` is the following:

`https://transmart.mysite.com/transmart/authUser/**`

The request map supports the `**` pattern-matching characters. For example, in the above URL, the URL pattern `/authUser/**` matches both of the following URLs:

URL	Purpose
<code>https://transmart.mysite.com/transmart/authUser/list</code>	View, edit, and delete transSMART users.
<code>https://transmart.mysite.com/transmart/authUser/create</code>	Create transSMART users.

Default Role / URL Mappings

The following table describes the pre-defined mappings between tranSMART roles and URL patterns:

URL Pattern	Mapped Role	Purpose
/accessLog/**	ROLE_ADMIN	View the tranSMART access log. When you click the Admin menu to access the administrator's console, the log is displayed by default.
/authUser/**	ROLE_ADMIN	Create, view, edit, and delete tranSMART users. Currently, only tranSMART administrators can perform these tasks.
/role/**	ROLE_ADMIN	Create, view, edit, and delete tranSMART roles. Currently, only tranSMART administrators can perform these tasks.
/requestmap/**	ROLE_ADMIN	Create, view, edit, and delete mappings between roles and URLs. Currently, only tranSMART administrators can perform these tasks.
/authUserSecureAccess/**	ROLE_ADMIN	Create, view, edit, and delete a user's access rights to specific clinical trials.
/secureObject/**	ROLE_ADMIN	Create, view, edit, and delete IDs and other attributes of a clinical trial.
/secureObjectPath/**	ROLE_ADMIN	No longer used.

URL Pattern	Mapped Role	Purpose
/ **	IS_AUTHENTICATED_REMEMBERED	<p>Attempt to access any transSMART URL.</p> <p>Note that:</p> <ul style="list-style-type: none"> ▪ If the user has not yet logged into transSMART, the transSMART login screen appears. ▪ If the user successfully logs in, or if the user is already logged in, access to the specified URL depends upon the user's role.
/login/**	IS_AUTHENTICATED_ANONYMOUSLY	These URLs can be accessed by anyone.
/css/**	IS_AUTHENTICATED_ANONYMOUSLY	
/js/**	IS_AUTHENTICATED_ANONYMOUSLY	
/images/**	IS_AUTHENTICATED_ANONYMOUSLY	
/search/loadAJAX**	IS_AUTHENTICATED_ANONYMOUSLY	



The roles IS_AUTHENTICATED_REMEMBERED and IS_AUTHENTICATED_ANONYMOUSLY cannot be edited, deleted, or explicitly assigned to users.

Managing User Roles



In some cases, application development may be required to support role-based functionality.

Creating a Role

To create a transSMART user role:

1. Click the **Admin** menu to display the administrator's console.
 2. Click **Create Role**.
- The Create Role window appears.
3. In **Role Name**, type a name for the role.

Role names must be upper case and must be prefixed with `ROLE_` — for example:

Create Role

Role Name:	<input type="text" value="ROLE_VIEW_LOG"/>
Description:	<input type="text"/>
Create	

In this example, a user assigned the role `ROLE_VIEW_LOG` can view the access log on the administrator's console but cannot perform any of the other tasks on the console.

4. In **Description**, type a description for the role.

A description is required.

5. Click **Create**.

You must now map the role to a URL. Choose one of the following actions:

- Adding a Role to an Existing Request Map (page 147)
- Creating a New Request Map (page 148)

Adding a Role to an Existing Request Map

1. If the administrator's console isn't already displayed, click the **Admin** menu to display it.
2. Click **Requestmap List**.
3. Click **Show** for the mapping to which you want to add a new role:

9	/images/**	IS_AUTHENTICATED_ANONYMOUSLY	Show
10	/search/loadAJAX**	IS_AUTHENTICATED_ANONYMOUSLY	Show
1753751	/accessLog/**	ROLE_ADMIN	Show
1753752	/authUserSecureAccess/**	ROLE_ADMIN	Show
1753753	/secureObject/**	ROLE_ADMIN	Show

4. Click **Edit**.
5. In **Roles (comma-delimited)**, type a comma and a space character after the rightmost role in the field, then type the name of the role to add to the map.

Edit Requestmap

ID:	1753751
URL Pattern:	<input type="text" value="/accessLog/**"/>
Roles (comma-delimited):	<input type="text" value="ADMIN, ROLE_VIEW_LOG"/>
Update Delete	

6. Click **Update**.

Creating a New Request Map

1. If the administrator's console isn't already displayed, click the **Admin** menu to display it.
2. Click **Requestmap Create**.
3. In **URL Pattern**, type the URL pattern to map to a role.
 Double-check your entry to ensure that the URL exists. tranSMART does not validate the entry.
4. In **role (comma-delimited)**, type the role name in upper case.
If you are mapping multiple roles to the URL, separate the role names with a comma.
5. Click **Create**.

Assigning a Role to a User

You assign a role to a user when you create or edit the user's account. For instructions, see [Managing User Accounts](#) on page 140.

Editing or Deleting a Role

To edit or delete a role:

1. If the administrator's console isn't already displayed, click the **Admin** menu to display it.
2. Click **Role List**.
3. Click **Show** for the role to edit or delete.
4. Take one of the following actions:
 - To delete the role, click **Delete**. Then click **OK** to confirm the deletion.
 - To edit the role, click **Edit**. After making the edits, click **Update**.

Editing or Deleting a Request Map

To edit or delete a mapping between a role and a URL:

1. If the administrator's console isn't already displayed, click the **Admin** menu to display it.
2. Click **Requestmap List**.
3. Click **Show** for the map to edit or delete.

4. Take one of the following actions:

- To delete the map, click **Delete** and then click **OK** to confirm the deletion.
- To edit the map, click **Edit**. After making the edits, click **Update**.

Accessing the Administrator's Console

There are two ways for a user to attempt to access the administrator's console:

- Click the **Admin** menu on the transSMART window (see [The Administrator's Console](#) on page 135).
The **Admin** menu is displayed only for users who are assigned the role `ROLE_ADMIN`.
- Enter the complete URL for the administrator's console:

`https://transmart.mysite.com/transmart/accessLog/list`

Partial Administrator Rights

If a user is assigned a role that is mapped to one of the tasks on the administrator's console, that user can access the console and click on all of the links to administrator tasks. However, the only task the user will be allowed to perform is the one authorized through a role.

For example, suppose you create the role `ROLE_VIEW_LOG` to allow a user to view the transSMART access log. A user with this role can view the log by entering the full URL for this administrator task — for example:

`https://transmart.mysite.com/transmart/accessLog/list`

However, if the user clicks on any of the other links on the administrator's console, the access-denied message is displayed.

Managing Security for Analyze Studies

Users are able to perform operations with private Analyze studies only if you or another administrator grant the user (or a group that the user belongs to) access rights to do so.

Before you can assign a user or a user group access rights to a protected study, the following tasks must be performed:

1. The study must be loaded into a database server.
2. You must protect the study by defining it as a secure object, using the transSMART administrator's console.

If tranSMART is deployed on multiple servers, this step must be performed on each server separately, after the study has been loaded to the corresponding database server.

Securing a Study

When a study is loaded, the data loader indicates whether the study is to be secured. Depending on its status, the study is created (secured) in or removed (not secured) from BIOMART.BIO_EXPERIMENT as well as these security concepts:

- SEARCHAPP.SEARCH_SECURE_OBJECT
- I2B2DEMODATA.PATIENT_TRIAL
- I2B2DEMODATA.OBSERVATION_FACT

If the **Add Study** option doesn't perform this step, the application should be changed to do so or the **Add Study** option should be removed.

You can also run the stored procedure I2B2_SECURE_STUDY, after a study is loaded, to add or remove security.

Managing Groups

Access privileges for a study can be assigned to users individually or to a group of users. Assigning access privileges to a group of users can be more convenient than assigning privileges individually.

Creating a Group

To create a group:

1. Click the **Admin** menu to display the administrator's console.
2. Click **Create Group**.

The following window appears:

Create User Group

Name:	<input type="text"/>
Description:	<input type="text"/>
Enabled:	<input type="checkbox"/>
Unique Id:	<input type="text"/>

 **Create**

3. In **Name**, assign a name to the group.
4. Optionally, in **Description**, type an optional description of the group.
5. To enable the group's privileges, select **Enabled**.
6. Leave **Unique ID** blank. A unique ID will be assigned to the group.
7. Click **Create**.

In the following figure, the group Test Group has been created. Note that it currently has no members or privileges to access any studies.

User Group

info UserGroup 5113 created

Id:	5113
Enabled:	true
Description:	Group definition for test purposes.
Group Category:	USER_GROUP
Name:	Test Group
Type:	GROUP
Members:	
Access to Studies:	

  **Edit** **Delete**

Managing a Group's Users

To add users to a group, or remove users from a group:

1. Click the **Admin** menu to display the administrator's console.
2. Click **Group Membership**.

The following window appears:

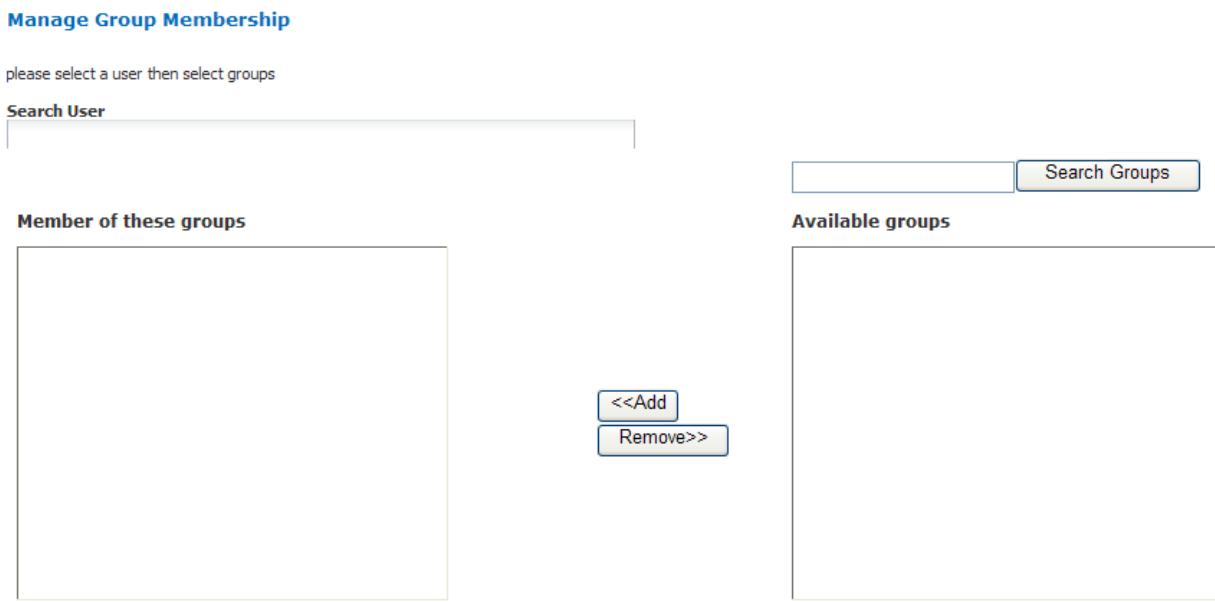
Manage Group Membership

please select a user then select groups

Search User

Member of these groups

Available groups



3. In **Search User**, type part or all of a user name, then select the name from the autotype dropdown.

Next you will specify the group that the user is being added to or removed from.

4. Click **Search Groups**.

The list of the available groups appears in the **Available groups** box.

5. Click the group name, then click **Add** to add the user to the group, or **Remove** to remove the user from the group.

In the figure below, the specified user has been added to the group Test Group:

Manage Group Membership

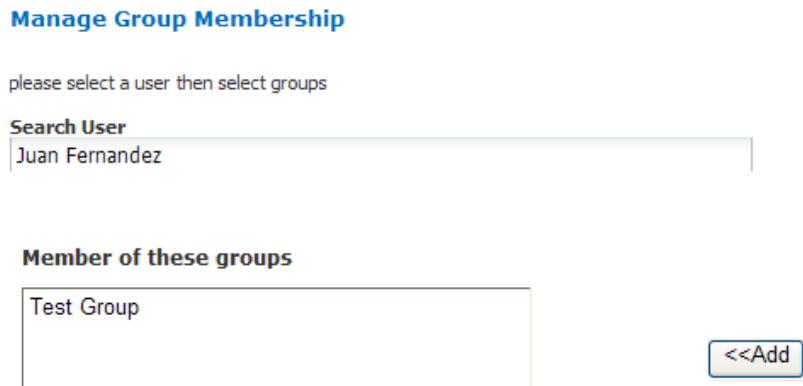
please select a user then select groups

Search User

Juan Fernandez

Member of these groups

Test Group



6. Click another administrative task, or leave the administrator's console. No Save action is required.

Editing or Deleting a Group

1. Click the **Admin** menu to display the administrator's console.
2. Click **Group List**.
3. Click the ID of the group to edit or delete.
4. In the User Group window, click **Edit** or **Delete**:
 - If editing, make the changes and click **Update**.
You may need to scroll down to the bottom of the window to see the edit fields.
 - If deleting, click **Delete**, then click **OK** to confirm the deletion.

Managing Access Privileges

You assign a user or group access privileges to a study by assigning the user or group a particular access level for the study. Access levels determine the kinds of operations that the user can perform when accessing the study.

Access Levels

Individual users and groups of users can be assigned the following access levels for a study:

Access Level	Description
OWN	User is the owner of the study with full access privileges.
EXPORT	User is not the owner of the study, but the user can define cohorts and points of comparison from the study. The user can also export all generated summary statistics and comparison data to a Microsoft Excel spreadsheet.
VIEW	User is not the owner of the study, but the user can define cohorts and points of comparison from the study. However, the user cannot export any data.

Managing Access Privileges for a User or Group

In the Manage Study Access for User/Group window, you can perform the following tasks:

- Assign or remove access privileges to one or more studies for a user or group.
- Assign the access level for the access privileges.

To assign a user or group access privileges for a study:

1. Click the **Admin** menu to display the administrator's console.
2. Click **Access Control by Group**.

The following window appears:

Manage Study Access for User/Group

Search User/Group

Access Level **VIEW**

Has Access for these studies

Available studies:

<<Add
Remove>>

3. In **Search User/Group**, type part or all of a user or group name, then select the name from the autotype dropdown.
4. In the **Available studies** box, select one or more studies that the members of the group can access, then click **Add**.
5. In **Access Level**, select the access level (VIEW, EXPORT, OWN), to give to the members of the group for the selected studies.

For descriptions of these access levels, see [Access Levels](#) on page 153.

6. Click another administrative task, or leave the administrator's console. No Save action is required.

If you now click **Groups > Group List**, and then click the ID of the new group you created in [Creating a Group](#) on page 150, you will see the members of the groups the studies to which the members have access privileges, and the access level for each study.

Managing Access Privileges for a Study

In the Manage Study Access window, you can perform the following tasks:

- Assign or remove access privileges to one or more users or groups for a secure object (such as a study or an entire study category).
- Assign the access level for the access privileges.

To grant access privileges to a study:

1. Click the **Admin** menu to display the administrator's console.
2. Click **Access Control by Study**.

The following window appears:

Manage Study Access

Secure Object:

Access Level:

User/Group Assigned Access

User/Group Without Access

3. In **Secure Object**, select the study or study category to which access is being granted.
4. In the **User/Group Without Access** box, select the users and/or groups who can access the secure object, then click **Add**.
5. In **Access Level**, select the access level (VIEW, EXPORT, OWN) for accessing this secure object by the selected users/ groups.

For descriptions of these access levels, see [Access Levels](#) on page 153.

6. Click another administrative task, or leave the administrator's console. No Save action is required.

Viewing the tranSMART Access Log

The Access Log lets you view tranSMART events such as logins, logouts, searches, and Analyze analyses. For each event, the log notes the time and date of the event and the user who performed the operation.

The access log displays events beginning with the most recent.

Displaying the Access Log

When you open the administrator's console, the log is displayed by default.

If you are in a different window of the administrator's console and want to display the access log, click **View Access Log**.

Exporting the Access Log to a Spreadsheet

To export the access log to a Microsoft Excel spreadsheet:

1. With the access log displayed, click **Export to Excel**.
2. Specify whether you want to display the access log within a spreadsheet, or immediately save the spreadsheet to a file.

Specifying the Timeframe for the Access Log

By default, the log shows all events, starting with the most recent event and extending back to show one week before the end date.

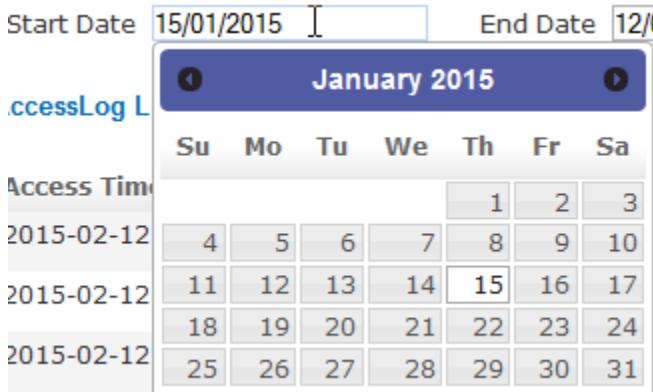
You can specify a particular timeframe for the events you want to display or export.

To specify a timeframe:

1. With the access log displayed, type the date of the earliest events to display in the **Start Date** text box.

Date format is dd/mm/yyyy.

Alternatively, select the start date from the calendar that appears when you place the mouse pointer inside the Start Date or End Date text box.



2. Repeat Step 1 for the **End Date** field.
3. Click **Filter**.

All events within the specified timeframe display.



If the **End Date** is before the **Start Date**, the event list contains no entries.

Browse Tool Administration

This section describes how to create and modify the following objects in the Browse Program Explorer:

- Programs
- Studies
- Analyses
- Assays
- Folders

For descriptions of these objects, see [Viewing Studies in the Program Explorer Tree](#) on page 8.

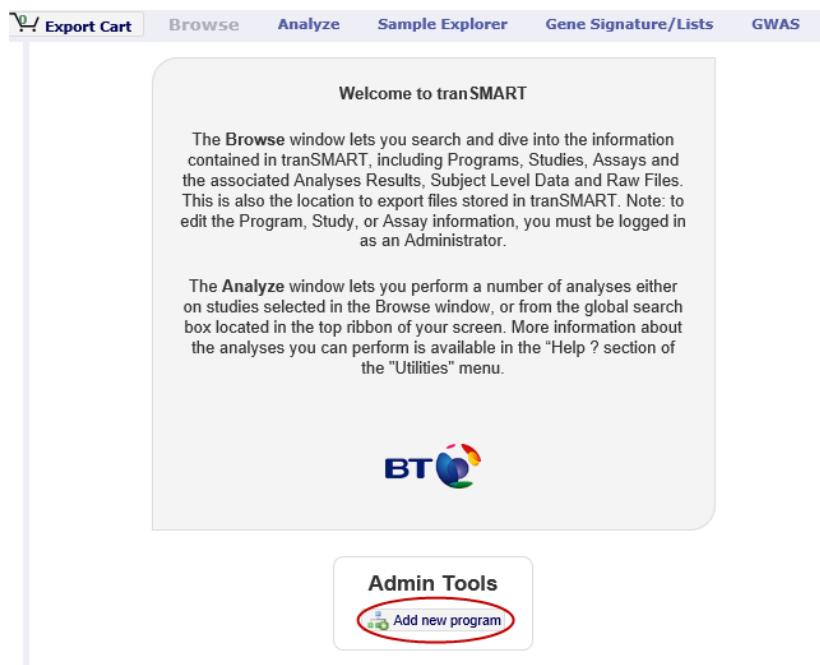
Creating Program Explorer Objects

A program is the highest-level object in the Program Explorer tree. The procedure for creating a program is different than for creating all other objects.

Creating a Program

To create a program in the Program Explorer:

1. Click **Browse** in the tranSMART menu bar.
2. Click **Add new program** under the Welcome to tranSMART box:



3. Define the fields in the Create Program dialog box, then click **Save**.

Creating Other Program Explorer Objects

Studies, analyses, assays, and folders are child objects of some other object. For example, you can create a study under a program, an analysis under a study, or a folder under an analysis or another folder.

To create a child object:

1. Select its parent object in the Program Explorer.

The child objects that can be created under the parent appear as buttons in the upper-right corner of the Browse window; for example:

Program: Public Studies Add new study Add new folder

This program gathers public studies.

2. Click the appropriate button to open the Create... dialog box.
3. Define the fields in the Create... dialog box, then click **Save**.

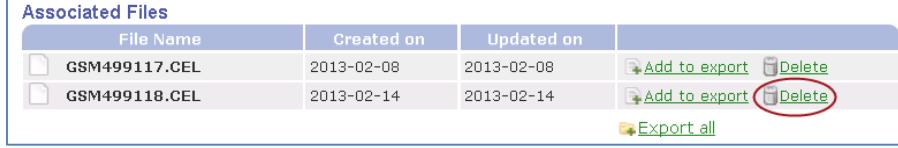
The following table shows the objects you can create for a selected object in the Program Explorer:

Selected Object in Program Explorer	Child Objects You Can Create
Program	<ul style="list-style-type: none"> ▪ Study ▪ Folder
Study	<ul style="list-style-type: none"> ▪ Analysis ▪ Assay ▪ Folder
Analysis	<ul style="list-style-type: none"> ▪ Folder
Assay	<ul style="list-style-type: none"> ▪ Folder
Folder	<ul style="list-style-type: none"> ▪ Sub-Folder

Editing and Deleting Objects

The following table describes how to edit and delete Program Explorer objects:

Task	Description						
Editing an object	<p>To edit an object, click the object in the Program Explorer, then click the pencil icon that appears in the Browse window:</p> <p>Folder: Processed data</p> <p>Files with gene expression intensity values.</p> <p>Associated Tags</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="background-color: #d9e1f2;">Property</th> <th style="background-color: #d9e1f2;">Value</th> <th style="background-color: #d9e1f2; width: 40px;"></th> </tr> </thead> <tbody> <tr> <td>File Type</td> <td>Processed data</td> <td></td> </tr> </tbody> </table> <p>Define the fields in the Edit... dialog box, then click Save.</p>	Property	Value		File Type	Processed data	
Property	Value						
File Type	Processed data						
Deleting analyses, assays, and folders	<p>To delete an analysis, assay, or folder, click the object in the Program Explorer, then click the Delete this... button in the upper right corner of the window; for example:</p> <p>Folder: Processed data</p> <p>Files with gene expression intensity values.</p> <p>Note: Only analyses, assays, and folders can be deleted from within the Browse window. Programs and studies must be deleted from the database directly.</p>						

Task	Description
Deleting files	<p>To delete a file from a folder, click the folder in the Program Explorer, then click the Delete button at the right:</p> 

Common Features for Creating and Editing Objects

The following table shows the features that apply to all Program Explorer objects when you are creating or editing an object in a Create... or Edit... dialog box:

Feature	Description
Required fields	<p>Fields whose names are followed by a red asterisk are required:</p> 
Autocomplete fields	<p>Shaded fields are autocomplete fields. Type one or more characters at the beginning of the value that you want to assign to the field, and transSMART will display a list of text strings that begin with those characters. Select the value to assign from the displayed list.</p>  <p>Alternatively, insert the cursor in the field and press the Down arrow key to select from an alphabetical list of suggested field values.</p>
Multiple-value fields	<p>Some autocomplete fields allow multiple values to be assigned. These fields contain the label Add new next to the field.</p> 

Feature	Description
Removing a value from a multi-value field	<p>To remove a value from a multi-value field, click the blue x icon next to the value:</p> 
Close vs. Cancel buttons	<p>Both buttons close the Create... or Edit... dialog box, and any changes you made in the dialog box are abandoned. However, with Cancel, a warning message appears before the dialog box is closed. With Close, the dialog box is closed immediately with no warning message.</p>

Uploading Files to Folders

Folders allow you to attach files to an object. For example, you might add a folder to contain files pertaining to the analysis of a study, or a gene list for an analysis.

You can upload any type of file to a folder. However, the free-text search feature will only search files in a format that can be text-indexed, such as Microsoft Word documents, text files, and electronically generated PDFs.

Files can be uploaded to a folder via FTP and can be stored on the application server.

Appendix A

Download Analysis Data as R Data

Analyses run through the Advanced Workflow tool within Analyze use R for computation. tranSMART allows you to:

- Download data files that were exported from the tranSMART database for use in an analysis. These files can then be used in other external tools.
- Download the R scripts that tranSMART used in the generation of an analysis.
- Review the R version information used for a particular analysis.
- Export R code and its data from tranSMART so you execute the R code in the R program independently of tranSMART.



For more information on The r Project for Statistical Computing, visit the following site: www.r-project.org.

Downloading Raw R Data

To download advanced workflow analyses as raw R data:

1. Click the tranSMART **Analyze** tool to display the Analyze window.
2. Select the study you want to use and drag it into a subset definition box in Subset 1.
3. Click the **Advanced Workflow** tab and select the analysis you want to run.
4. Define the variables accordingly.
5. Click **Run**.

Your analysis appears below the variable selection boxes.

6. Click **Download raw R data** at the bottom of the page.

A dialog box similar to the following appears:



7. Decide whether you want to open the file or save it to your hard drive, then click **OK**.

Reviewing R Version Information

To review the R version information for an analysis:

1. Click the tranSMART **Analyze** tool to display the Analyze window.
2. Select the study you want to use and drag it into a subset definition box in Subset 1.
3. Click the **Advanced Workflow** tab and select the analysis you want to run.
4. Define the variables accordingly.
5. Click **Run**.

Your analysis appears below the variable selection boxes.

6. Click **R Version Information** at the bottom of the page.

The following R version data displays:

- Information about the R installation:

R Version Information

R version 2.15.1 (2012-06-22)

Platform: x86_64-pc-mingw32/x64 (64-bit)

locale:

```
[1] LC_COLLATE=English_United States.1252
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252
```

attached base packages:

```
[1] stats graphics grDevices datasets methods base
```

other attached packages:

```
[1] MASS_7.3-18 stringr_0.6.1 Cairo_1.5-1 ggplot2_0.9.2.1
[5] plyr_1.7.1
```

loaded via a namespace (and not attached):

```
[1] colorspace_1.1-1 dichromat_1.2-4 digest_0.5.2 grid_2.15.1
[5] gtable_0.1.1 labeling_0.1 memoise_0.1 munsell_0.4
[9] proto_0.3-9.2 RColorBrewer_1.0-5 reshape2_1.2.1 scales_0.2.2
[13] tools_2.15.1
```

Version information

The base "attached" packages that came with R and are loaded into the environment to run the analysis

Other attached packages (add-ons) that were downloaded to run the analysis

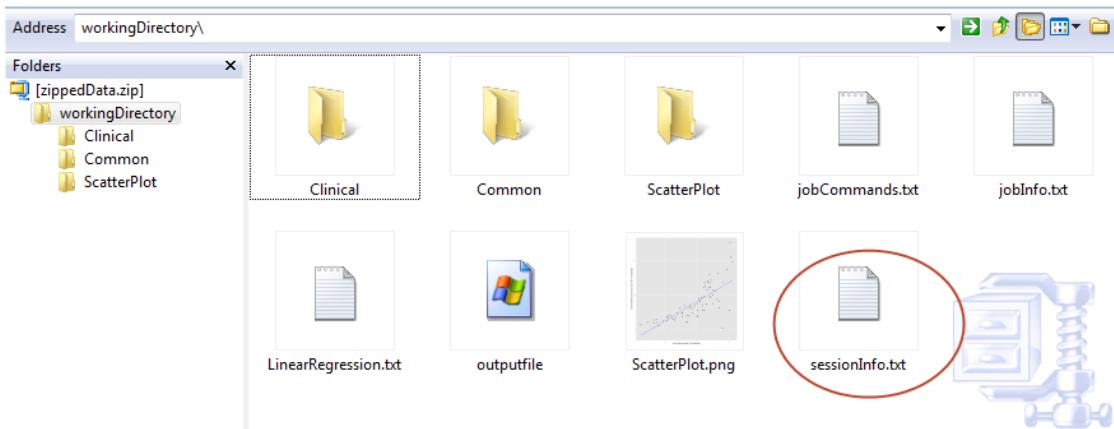
Not attached but available to R through namespaces

- A table of the packages that are included in the installation. Although not all packages may be used, all are listed. The version number is especially useful — if you want to re-create an analysis, the version number indicates the specific version used to create the original analysis.

This sample shows a portion of the table:

"Package"	"LibPath"	"Version"	"Priority"	"Depends"
"base"	"C:/Program Files/R/R-2.15.1/library"	"2.15.1"	"base"	NA
"Biobase"	"C:/Program Files/R/R-2.15.1/library"	"2.16.0"	NA	"R (>= 2.10), utils, BiocGenerics (>= 0.1.0)"
"BiocGenerics"	"C:/Program Files/R/R-2.15.1/library"	"0.2.0"	NA	"methods, graphics, stats"
"BioInstaller"	"C:/Program Files/R/R-2.15.1/library"	"1.4.7"	NA	"R (>= 2.15.0)"
"bitops"	"C:/Program Files/R/R-2.15.1/library"	"1.0-4.1"	NA	NA
"boot"	"C:/Program Files/R/R-2.15.1/library"	"1.3-4"	"recommended"	"R (>= 2.14.0), graphics, stats"
"Cairo"	"C:/Program Files/R/R-2.15.1/library"	"1.5-1"	NA	"R (>= 2.4.0)"
"caTools"	"C:/Program Files/R/R-2.15.1/library"	"1.13"	NA	"R (>= 2.2.0), bitops"
"class"	"C:/Program Files/R/R-2.15.1/library"	"7.3-3"	"recommended"	"R (>= 2.5.0), stats, utils"

You can view the full table in grid format when you select the **Download raw R data** option. Just click **sessionInfo.txt** to open the file:



Export R Code and Data

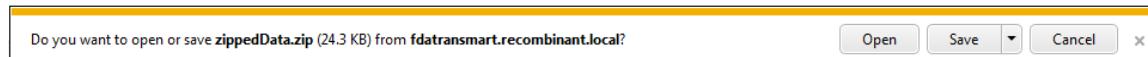
To export R code and data so it is executable in R:

1. Click the tranSMART **Analyze** tool to display the Analyze window.
2. Select the study you want to use and drag it into a subset definition box in Subset 1.
3. Click the **Advanced Workflow** tab and select the analysis you want to run.
4. Define the variables accordingly.
5. Click **Run**.

Your analysis appears below the variable selection boxes.

6. Click **Download raw R Data** at the bottom of the page.

A dialog box similar to the following appears:



7. Open the file and copy the R commands from the file jobcommands.txt.

Be sure you delete the analysis image from the Advanced Workflow page.

8. Paste the R commands from the downloaded file into R.

9. Run R.

R creates the appropriate analysis image.

Prerequisites for using R

1. When opening R, change the working directory to your downloaded raw R data folder `setwd("C:\\users\\username\\Desktop\\workingDirectory")`.
2. Install the following packages:
 - a. `install.packages("plyr")`
 - b. `install.packages("ggplot2")`
 - c. `install.packages("Cairo")`
 - d. `install.packages("rmeta")`
 - e. `install.packages("visreg")`

Appendix B

Glossary

AGGREGATE PROBES

Used in Analyze, the Aggregate Probes checkbox allows you to group probes used in high-dimensional data samples to form a total quantity that analyses will be performed on.

ANALYSIS OF VARIANCE (ANOVA)

Analysis of Variance (ANOVA) is a statistical method used in Analyze to make concurrent comparisons between two or more means in a box plot.

ANALYSIS VIEW

Used in the Search tool, the Analysis View option displays the statistically significant analyses from your search filter(s).

ANALYZE

Analyze lets you compare data generated for test subjects in two different cohorts, based on criteria and points of comparison that you specify. Analyze is useful to help you test a hypothesis that involves the criteria and points of comparison that you select.

ANTI-REGULATION

An analysis of a statistically significant experiment returned from a search against a gene signature or list is designated as *co-regulated* or *anti-regulated*.

ARRAY DATA

See: [Microarray](#)

ARRAYEXPRESS

Database of gene expression and other microarray data at the European Bioinformatics Institute (EBI).

See <http://www.ebi.ac.uk/arrayexpress> for details.

BINOMIAL DISTRIBUTION

Graph that displays the discrete probability distribution of obtaining n successes out of N Bernoulli trials.

See <http://mathworld.wolfram.com/BinomialDistribution.html> for details.

BIOMARKER

Short for Biological Marker, a biomarker is a key molecular or cellular event that links a specific environmental exposure to a health outcome.

BOX PLOT

Also known as a Box and Whisker Plot, a box plot is a histogram-like method of displaying data. Box plots are useful when conveying location and variation information in datasets.

CATEGORICAL VARIABLE

Also known as a nominal value, a categorical variable is one that has two or more categories, but with no intrinsic ordering to the categories. An example of a categorical value is hair color — there is no way to order these variables from highest to lowest.

CENSORING VALUE

Used in Survival Analyses. The Censoring Value specifies which patients had the event whose time is being measured. For example, if the Time variable selected is Overall Survival Time (Years), an appropriate censoring variable is Patient Death.

CHI SQUARED

Let the probabilities of various classes in a distribution be p_1, p_2, \dots, p_k , with observed frequencies m_1, m_2, \dots, m_k . The quantity

$$\chi_s^2 = \sum_{i=1}^k \frac{(m_i - N p_i)^2}{N p_i}$$

is therefore a measure of the deviation of a sample from expectation, where N is the sample size.

COHORT

A group of subjects who share specific events or characteristics. Cohorts are defined in the subset definition boxes of the Analyze tool.

CONTINUOUS VARIABLE

Continuous variables have an infinite number of values between two points. For example, age or temperature.

CO-REGULATION

An analysis of a statistically significant experiment returned from a search against a gene signature or list is designated as *co-regulated* or *anti-regulated*.

CORRELATION ANALYSIS

A type of Regression Analysis, correlation analysis measures the correlation coefficient — the linear association between two variables. Values of the correlation coefficient are always between -1 and +1. A correlation coefficient of +1 indicates that two variables are perfectly related in a positive linear sense, while a correlation coefficient of -1 indicates that two variables are perfectly related in a negative linear sense.

COX COEFFICIENT

The Cox coefficient refers to the coefficients in a Cox regression model (also known as the proportional hazards model for survival-time). The analysis investigates the effects of one or more variables upon the time a specified event takes to happen. The cox coefficient relates to a hazard; a positive coefficient indicates a worse prognosis, while a negative coefficient indicates a protective effect of the variable.

DATA BINNING

Defers to a data pre-processing technique used to reduce observation errors and to allow continuous variables to become categorical. Clusters of data are replaced by a value representative of that cluster (often but not necessarily, the central value).

DATA WAREHOUSE

A database used for reporting and analysis.

DATASET

Collection of data, most commonly presented in a tabular form where each column represents a specific variable, and each row represents a value for that variable.

DEPENDENT VARIABLE

In an experiment, the dependent variable is the response that is measured.

DIFFERENTIAL MODULATION

DOWN-REGULATION

An analysis of a statistically significant experiment returned from a search against a pathway is designated as *up-regulated* or *down-regulated*.

ENTREZ GENE

Reference sequences for a wide range of species. For details, see <http://www.ncbi.nlm.nih.gov/gene/>.

ENTREZ GLOBAL

Federated search engine that allows users to search various health sciences databases at the National Center for Biotechnology Information (NCBI) website.

See www.ncbi.nlm.nih.gov/Entrez/ for details.

FOLD CHANGE RATIO

A number describing how much a quantity changes going from an initial to a final value. An initial value of 50 and a final value of 100 corresponds to a fold change of 2 (a two-fold increase).

GENE

Stretches of DNA and RNA that code for a polypeptide or for an RNA chain — contains hereditary molecular information.

GENE CHIP

See: [Microarray](#)

GENE EXPRESSION

The flow of genetic information from gene to protein; the process, or the regulation of the process, by which the effects of a gene are manifested; the manifestation of a heritable trait in an individual carrying the gene or genes that determine it.

GENE EXPRESSION OMNIBUS

GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community. For more information, see <http://www.ncbi.nlm.nih.gov/geo>.

GENE SET ENRICHMENT ANALYSIS (GSEA)

Computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (for example, phenotypes).

See <http://www.broadinstitute.org/gsea/index.jsp> for details.

GENE SIGNATURE

A group of genes whose combined expression pattern is uniquely characteristic of a medical condition or other clinical outcome of interest.

GENE SYMBOL

A unique abbreviation of a gene name consisting of italicized uppercase Latin letters and Arabic numbers. we use Entrez as the full list of genes (related to but not identical to HUGO)

See <http://www.genenames.org/> for details.

GENECARDS

Database that offers information about human genes (and mouse homologues).

See <http://www.genecards.org> for details.

GOOGLE SCHOLAR

Google application that provides a search of scholarly literature across multiple disciplines and sources.

See <http://scholar.google.com> for details.

GPL PLATFORM

A Platform record is composed of a summary description of the array or sequencer and, for array-based Platforms, a data table defining the array template. Each Platform record is assigned a unique and stable GEO accession number (GPLxxx). A Platform may reference many Samples that have been submitted by multiple submitters.

HEATMAP

Display of differential expression. Individual values contained in the matrix are represented by colors.

HIERARCHICAL CLUSTERING

Hierarchical clustering is a type of clustering analysis whose goal is to organize data so that the objects in the same cluster are more similar to each other than to those in other clusters.

HIGH DIMENSIONAL DATA

Datasets where the intersection of a subject and measurement is comprised of hundreds or thousands of points. For example, in a low dimensional data measurement such as height, the intersection of subject and measurement is one number (ex. 180 cm), whereas in a high dimensional data measurement such as gene expression in a lymph node, the measurement is 50,000 individual probe expression values.

HISTOGRAM

A visual representation of the distribution of data values within a dataset.

HOMOLOGY

The basis for comparative biology — where organs/structures from one organism are compared to a similar organ/structure in a different organism.

IN VITRO STUDY

Those that are conducted using components of an organism that have been isolated from their usual biological surroundings.

IN VIVO STUDIES

Experimentation using a whole, living organism.

INDEPENDENT VARIABLE

In an experiment, the independent variable is the variable that is manipulated.

JOB

In Valhalla, a job refers to a command you have given Analyze to process or export data. Jobs and job-related events can be found within the **Jobs** tab in Analyze.

KENDALL CORRELATION

Kendall's rank correlation provides a distribution-free test of independence and a measure of the strength of dependence between two variables.

K-MEANS CLUSTERING

The K-Means clustering heatmap clusters genes and/or samples into a specified number of clusters. The result is k clusters, each centered around a randomly-selected data point.

LINE GRAPH

Line graphs illustrate the temporal relationship between two major variables.

MARKER SELECTION

Marker Selection is a display of the top differentially expressed genes between two specified cohorts. .

MESH ONTOLOGY

MeSH is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity.

MICROARRAY

A two-dimensional array on a chip or solid surface that assays large amounts of DNA material.

MRNA ANALYSIS

Assays that quantify the expression levels of all mRNA molecules in an experiment.

NAVIGATION TREE

The Window's Explorer-like, hierarchical representation of study data that has been loaded into Analyze.

NCBI

The National Center for Biotechnology Information.

See <http://www.ncbi.nlm.nih.gov/> for details.

NUMERIC-NODE

Used in Analyze, numeric-nodes are indicated by the (123) symbol, numeric nodes indicate that the data values associated with the concept are only numeric (for example, age values, date values, etc.). For more information, see [Continuous Variable](#).

ONTOLOGY

A hierarchical description of the concepts and relationships that can exist for an agent or a community of agents.

ORTHOGONAL COMPONENT

When performing statistical analysis, independent variables that affect a particular dependent variable are said to be orthogonal if they are uncorrelated, since the covariance forms an inner product.

PATHOLOGY

The study of diagnosis and disease.

PATHWAY

A group of genes interacting to form an aggregate biological function.

PEARSON CORRELATION

Obtained by dividing the covariance of the two variables by the product of their standard deviations

PRINCIPAL COMPONENT ANALYSIS

A Principal Component Analysis (PCA) is commonly used as a tool in exploratory data analysis. Data is split into orthogonal components, and the genes/probes that contribute the most variance to the components are displayed.

PROBE SET

A probe set is a collection of probes designed to interrogate a given sequence.

PROBE SET ID

A probe set ID is used to refer to a probe set, which looks like the following:

12345_at or 12345_a_at or 12345_s_at or 12345_x_at

The last three characters (_at) identify the probe set strand.

P-VALUE

The number corresponding probability that the occurrences of your experiment and analysis did not happen by chance. P-value cutoffs are often 0.05 or 0.01 — when the value is under the threshold, the result is said to be statistically significant.

R

R is a language and environment for statistical computing and graphics.

See <http://www.r-project.org> for details.

RBM DATA

Rules Based Medicine. They provide an array measurement of metabolites

REGRESSION ALGORITHMS

Algorithms that are particularly suited for mining data sets that have high dimensionality (many attributes), including transactional and unstructured data.

RHO-VALUE

Also known as Spearman's rho, the rho-value is a non-parametric measure of statistical dependence between two variables. See: [Spearman Correlation](#).

R-VALUE

The value assigned to a correlation coefficient.

SCATTER PLOT

Type of graph that uses Cartesian coordinates to display values for two variables for a set of data.

SEARCH FILTER

A biomedical concept used to define search criteria in the Search tool.

SEARCH STRING

A sequence of biomedical concepts used to define search criteria in the Search tool.

SLOPE

The steepness of the line of best fit in a graph ($\Delta y/\Delta x$).

SNP DATA

Single Nucleotide Polymorphism. DNA sequence data marking variation occurring when a single nucleotide — A, T, C or G — in the genome.

SPEARMAN CORRELATION

The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (, also signified by rho-value) measures the strength of association between two ranked variables.

STATISTICAL SIGNIFICANCE

Results of analyses on data that are statistically significant indicate a confidence level that the results did not happen by chance.

SUBSET

A smaller grouping of participants in a study. See [cohort](#).

SURVIVAL ANALYSIS

Assessment of the amount of time that a person or population lives after a particular intervention or condition.

T STATISTIC

Ratio of the departure of an estimated parameter from its notional value and its standard error.

TABLE WITH FISHER TEST

Examines the significance of associated categorical variables.

TEA ANALYSES

Target Enrichment Analysis (TEA) measures the enrichment of a gene signature, gene list, or pathway in a microarray expression experiment.

TEA P-VALUE

These normalized p-values are intermediate values in the TEA calculation. To be considered a statistically significant analysis, an analysis must have at least one matching biomarker with a TEA p-Value of less than 0.05.

TEA SCORE**TEXT-NODE**

Indicated by the (abc) symbol, text nodes indicate that the data values associated with the concept are only textual (for example, race or gender). For more information, see [Categorical Variable](#).

TISSUE TYPE

The specific type of tissue that has been used in the experiment (for example, breast tissue, lung tissue, etc.)

UP-REGULATION

An analysis of a statistically significant experiment returned from a search against a pathway is designated as *up-regulated* or *down-regulated*.

X-AXIS

The horizontal axis of a two-dimensional Cartesian coordinate system.

Y-AXIS

The vertical axis of a two-dimensional Cartesian coordinate system.

Glossary