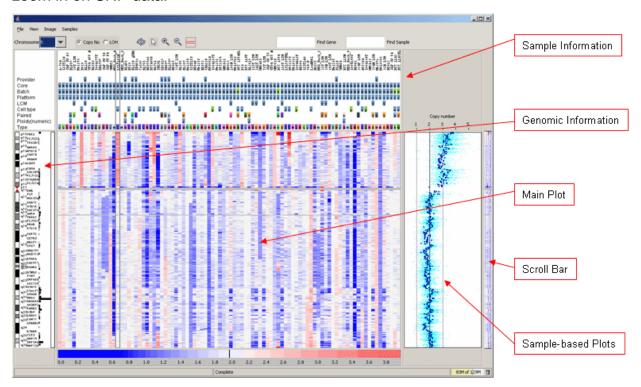


#### **SNPViewer Documentation**

Module name: SNPViewer

**Description:** Displays SNP data plotting copy numbers and LOH values **Author:** Jim Robinson (Broad Institute), <u>gp-help@broad.mit.edu</u>

**Summary:** The SNPViewer displays SNP data. Sample data displays on the y-axis and SNP data on the x-axis. The main plot shows either copy number or LOH data. The area to the right of the main plot displays data for a selected sample (sample-based plots). At the far right, a scroll bar shows where you are on the chromosome as you zoom in on SNP data.



This document describes how to start the SNPViewer and then discusses each area of the viewer:

- Starting the SNPViewer
- Sample Information
- Genomic Information
- Main Plot: Copy Number or LOH
- Scroll Bar
- Sample-based Plots
- Menu Bar
- Tool Bar



## **Starting the SNPViewer**

To start the SNPViewer:

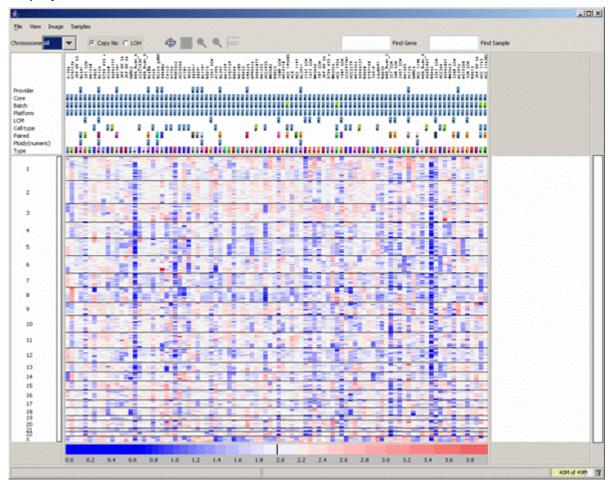
- 1. In GenePattern, select the SNPViewer module.
- 2. Enter the SNPViewer parameters:

dataset filename	Optional. Full path and file name for a copy number file (cn or xcn) sorted by chromosome and physical location. The xcn file format includes genotype information, which allows you to view LOH data in the SNPViewer. Typically, you use the SNPFileCreator and CopyNumberDivideByNormals modules to create .cn or .xcn files.
	If specified, GenePattern automatically loads this data after starting the SNPViewer. If omitted, omit the other filename parameters as well and load your data after starting the SNPViewer. See the instructions below.
sample info filename	Optional. Full path and file name of the sample information file. The sample information file must include the following columns:
	<ul> <li>Array: identifier for the SNP array</li> </ul>
	<ul> <li>Sample: identifier for the biological sample used to generate the SNP array data</li> </ul>
	Any other columns in the sample information file appear as sample attributes in the <u>Sample Information</u> area.
sample list filename	Optional. Full path and file name of a sample list file, which contains one sample identifier per line. If specified, the viewer loads only these samples and loads them in this order.
max memory	Required. Maximum amount of memory available to the SNPViewer. The default value of 700m is appropriate for users with at least 1GB of memory. Users with .5 GB should lower this to perhaps 300m.
	The SNPViewer requires approximately 1m of memory per sample for a 250K SNP chip array. Most users find 700m or 300m sufficient.

Note: For descriptions of .cn, .xcn, and sample information files, see File Formats.

# GenePattern

3. Click *OK*. After a few moments, the SNPViewer starts, loads the dataset, and displays it:



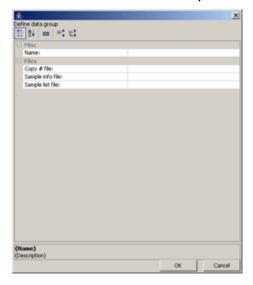
 Select a chromosome from the drop-down list in the tool bar to display a more detailed view of an individual chromosome (as shown at the beginning of this document).

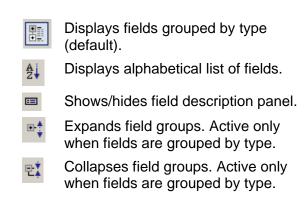
After starting the SNPViewer, you can use the File menu to load .cn or .xcn files into the viewer. This is particularly useful for viewing .cn or .xcn files that are on your local drive; for example, files that you have downloaded from GenePattern. (When you specify a local file as a parameter to the SNPViewer module, GenePattern copies the specified file to the GenePattern server. For large cn or .xcn file, this can be time consuming.)



To load SNP data into the SNPViewer:

1. Select File>New Data Group. The Define Data Group window appears:





## 2. Enter the requested information:

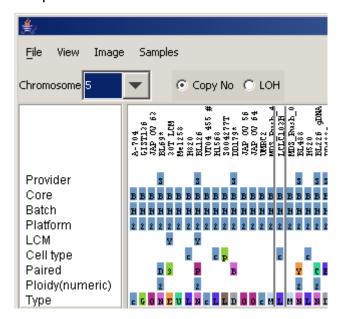
Name	Required. Name for this SNP data group.	
Copy # file	Required. Full path and file name for a copy number file (cn or xcn) sorted by chromosome and physical location. The xcn file format includes genotype information, which allows you to view LOH data in the SNPViewer. Typically, you use the SNPFileCreator and CopyNumberDivideByNormals modules to create .cn or .xcn files.	
Sample info file	Optional. Full path and file name of the sample information file. The sample information file must include the following columns:	
	<ul><li>Array: identifier for the SNP array</li></ul>	
	<ul> <li>Sample: identifier for the biological sample used to generate the SNP array data</li> </ul>	
	Any other columns in the sample information file appear as sample attributes in the <u>Sample Information</u> area.	
Sample list file	Optional. Full path and file name of a sample list file, which contains one sample identifier per line. If specified, the viewer loads only these samples and loads them in this order.	

3. Click OK to display the SNP data in the viewer.



## **Sample Information**

The y-axis of the main plot represents samples. The sample information area at the top of the viewer window lists sample names and attributes. The attributes come from the sample information file.



The colored blocks represent the attribute values for each sample. For each attribute, each unique value is assigned a color; therefore, a quick scan shows you the distribution of attribute values. To display an attribute value, hover over a colored block.

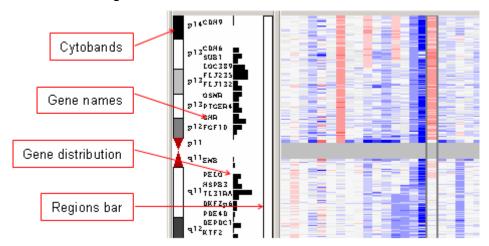
Use the Samples menu to hide, filter, and sort attributes:

- To hide or show attributes, select Samples>Show Sample Attributes. In the Sample Attributes window, select the attributes that you want to appear in the sample information area. By default, all attributes are selected.
- To filter samples by attribute value, select Samples>Filter by Attribute. The Sample Attributes window lists each attribute and all of its value. Select the attribute values to use for the filter: a logical OR combines values for an attribute and a logical AND combines attributes. For example, to display only samples that have a tissue Type of Breast Cancer OR Lung AND a Cell Type of cell\_line, under Type select both Breast Cancer and Lung and under Cell Type select cell\_line.
- To sort samples by attribute value, select Samples>Sort by Attribute. In the Sample Attributes window, select the attribute to use for sorting the samples. For example, select Type to sort samples based on the value of the Type attribute.



#### **Genomic Information**

Genomic information is displayed in the left panel of the viewer. Use the View menu to show/hide the genomic information.



By default, from left to right, the SNP information area displays:

- 1. Cytobands, both graphically and by name.
- 2. Gene names. Due to space constraints, only selected genes are shown. More gene names are displayed as you zoom in on the chromosome.
- 3. A bar graph that shows gene distribution. The longer the bar the more genes there are in this area. Hover over a bar to display the names of all genes in this area.
- 4. A regions bar that highlights user-defined regions of interest. Initially, the bar is empty (white). Regions of interest are described below.

SNP IDs are not shown by default, but can be displayed before gene names. Due to space constraints, only selected SNP IDs are shown. More SNP IDs are displayed as you zoom in on the chromosome.

**Note:** The SNPViewer displays cytoband and gene information from the Human Genome assembly of May 2004 (hg17).



### Regions of Interest

To define a region of interest:

1. In the tool bar, select the Define Region tool:



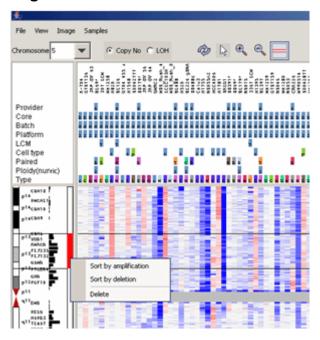
2. In the plot, click the start of a chromosomal region and the end of that region. The viewer annotates this region of interest by adding a red section to the region bar, as shown in the figure below.

Right-click the red section in the region bar to display the following menu:

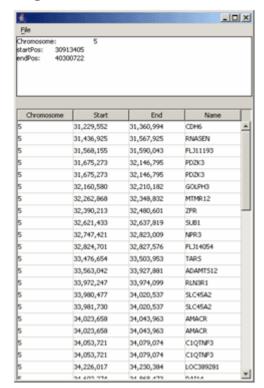
Sort by amplification	Sorts all samples based on SNP amplification in this region. The SNP amplification for a sample across this region is the number of SNPs with a copy number greater than 2.5.
Sort by deletion	Sorts all samples based on SNP deletions found in this region. The SNP deletion for a sample in this region is the number of SNPs with a copy number less than 1.5.
Delete	Deletes the annotation that marks this as a region of interest.

To see more detail about a region of interest, double-click the red section in the region bar. A window displays the start and end position of each gene in the region.

#### **Region of Interest**



#### **Region Detail**

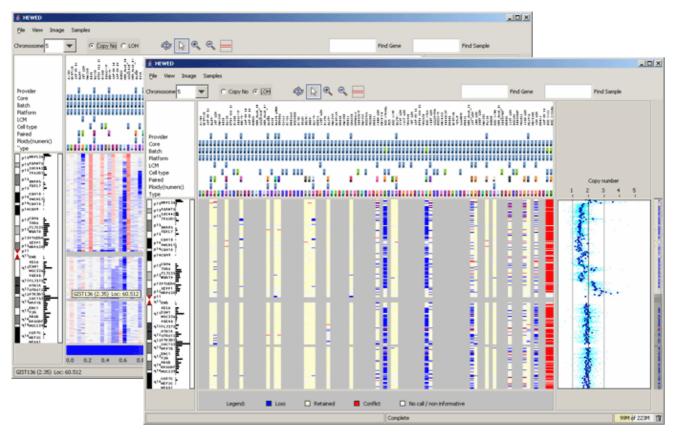




## Main Plot: Copy Number or LOH

The main plot of the SNPViewer displays either copy number or loss of heterozygosity (LOH) data. To switch between plots, select the *Copy No* or *LOH* radio button in the tool bar.

**Note**: To view LOH data, your SNP data file must include genotype information. A copy number file in the .xcn file format includes this information.

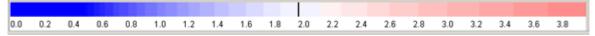


In the plot, rows represent chromosomal positions. Horizontal gray bars in the plot indicate regions where SNP data is unavailable.

Columns represent samples. In the LOH view, samples with no LOH data available are shown in gray. The data shown here, for example, includes unpaired samples.

Hover over a data point in the plot to display the sample name, copy number, and chromosomal position for that data point.

The legend below the plot defines the colors used to display the data. In the Copy Number view, the legend is a heat map:



To modify the thresholds for the heat map, drag the black bar that separates blue from red. For example, if you move the bars to the position shown below, copy numbers less than 1 display in blue, copy numbers greater than 3 in red, and all others in white:





Double-click on the legend to restore the default heat map thresholds.

For more information about copy number data, see the CopyNumberDivideByNormals module documentation. For more information about LOH data, see the LOHPaired module documentation.

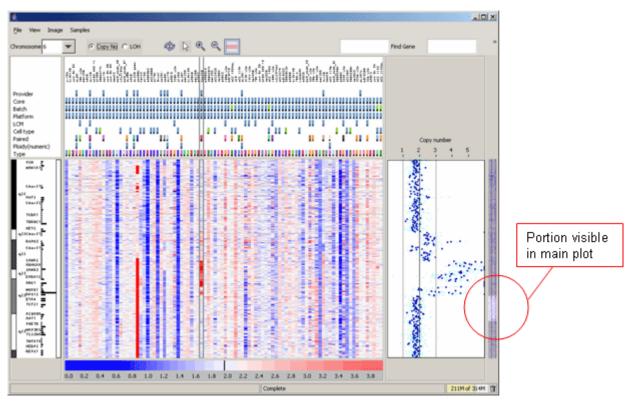
### **Scroll Bar**

When you select a chromosome from the drop-down list in the tool bar, the main plot displays SNP data for the chromosome and the scroll bar displays a narrow view of the chromosome plot.

To zoom in on a portion of the chromosome, click the Zoom In icon on the tool bar and then click a chromosomal location in the main plot. The viewer zooms in, centering on the selected chromosomal position. The scroll bar displays the chromosome plot, highlighting the portion that is visible in the viewer, as shown below.

To scroll up or down the chromosome, drag the highlighted portion of the scrolling region.

**Note**: Zooming and scrolling effect only the chromosomal (vertical) aspect of the viewer. The sample (horizontal) aspect remains static.



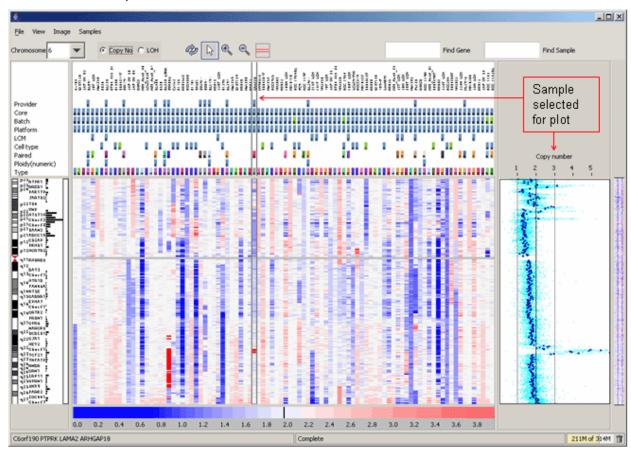


## **Sample-based Plots**

The area at the right of the viewer displays information for a selected sample. The SNPViewer currently has one sample-based plot, Copy # Plot, which plots the median copy number in dark blue and all copy numbers in light blue. The main plot uses median copy numbers to generate the copy number heat map.

To display the sample-based plot:

- 1. Select View>Show Copy # Plot menu to show (or hide) the sample-based plot.
- 2. In the tool bar, select the Select tool (the arrow). This tool is selected by default.
- 3. Click a sample in the main plot. The sample-based plot displays information for the selected sample.





## Menu Bar

File	New Data Group	Loads SNP information into the viewer; see Starting the SNPViewer.
	Exit	Closes the SNPViewer.
View	Show Regions Show Cytobands Show SNP IDs Show Genes Show Gene Distribution	Shows/hides SNP information on the x-axis of the viewer; see SNP Information.
	Show Copy # Plot	Shows/hides sample-based based plots in the far right panel of the viewer; see <a href="Sample-based">Sample-based</a> <a href="Plots">Plots</a> .
Image	Export screeenshot	Exports the SNPViewer window to an encapsulated postscript file (.eps format).
Samples	Show Sample Attributes Filter by Attribute Sort by Attribute	Shows/hides sample information on the y-axis of the viewer; see <u>Sample Information</u> .

## **Tool Bar**

Chromosome 5	Displays all chromosomes (default) or zooms in to a single chromosome.
© Copy No C LOH	Displays copy numbers or LOH calls in the main plot; see Main Plot: Copy Number or LOH.
<b>©</b>	Refreshes the SNPViewer window.
Q	Select mode (default) allows you to select samples in the main plot; see <u>Sample-based Plots</u> . Available only when viewing individual chromosomes.
<b>€</b>	Zoom in to display a smaller portion of the genome in greater detail.  Available only when viewing individual chromosomes.
	Zoom out to display a larger portion of the genome in less detail.  Available only when viewing individual chromosomes.
	Define a region of interest; see Regions of Interest. Available only when viewing individual chromosomes.
Find Gene	Displays the chromosome that contains the specified gene. Gene names are not case-sensitive; however, many names contain



	spaces, which must be included in the search string.
Find Sample	Selects the specified sample. Sample names are case-sensitive and many names contain spaces, which must be included in the search string.

## **Parameters**

For descriptions of the parameters, see Starting the SNPViewer.

## References

None.

# Platform dependencies

Task type: Visualizer

CPU type: any OS: any Language: Java JVM level: 1.5