

### **PCAViewer Documentation**

 

 Description:
 Principal Components Analysis Viewer

 Author:
 Marc-Danie Nazaire, Joshua Gould (Broad Institute), gp-help@broad.mit.edu

 The Institute for Genomics Research
 5.0

The PCAViewer displays the results from principal component analysis (the PCA module). It is adapted from TIGR's MultiExperiment Viewer (<u>http://www.tigr.org/software/tm4/mev.html</u>).

#### Summary

As discussed in Raychaudhuri, et al (2000: "A PCA analysis of DNA microarray data can consider the genes as variables or the experiments as variables or both. When genes are variables, the analysis creates a set of 'principal gene components' that indicate the features of genes that best explain the experimental responses they produce. When experiments are the variables, the analysis creates a set of 'principal experiment components' that indicate the features of the experimental conditions that best explain the gene behaviors they elicit.

The viewer provides two tabs, Components and Projection. Use the Components tab to plot the eigenvectors of selected principal components, as shown in <u>Figure 1</u>. Use the Projection tab to display a 2D plot that projects experiments/genes onto two principal components, as shown in <u>Figure 2</u>, or a 3D plot that projects them onto three principal components, as shown in <u>Figure 3</u>. (The viewer projects experiments or genes onto principal components depending on whether the principal component analysis considered the experiments or genes as variables, respectively.)

The viewer provides the following menu commands:

File>Save As	Save the plot to an image file.
File>Export Plot Table	Exports the plot table to a tab delimited text file.
Help>Help Contents	Display information about the viewer and how to use it.

#### **Components tab**

In the Components tab, the left pane lists each principal component. Select one or more principal components to plot their eigenvectors. The % Var field at the bottom of the list sums the % Variation of the selected principal components.

The right pane plots the eigenvectors of the selected principal components. The table under the plot lists the individual values that comprise each vector.

Right-click on the plot to display the following menu:

Properties	Set display options for the plot.
Save As	Save the plot to an image file.
Print	Print the plot.
Zoom in	Zoom in on either axis or both axes. Alternatively, click and drag to zoom in on an area of the plot.

Zoom out

Zoom out on either axis or both axes.

Auto range

Reset the default zoom level for either axis or both axes.



Figure 1. Components tab plotting eigenvectors for four principal components

### **Projection tab**

In the Projection tab, the left pane lists each principal component. To project the genes or experiments (whichever the principal component analysis considered as variables) onto principal components: select two or three principal components and click Plot. The % Var field at the bottom of the list sums the % Variation of the selected principal components.

The right pane shows a <u>2D plot</u> of two principal components or a <u>3D plot</u> of three principal components. If a cls, sample info file, or featureset file is provided then the points in the projection are colored according to the groups specified. **Note**: Java 3D version 1.3.1 or greater must be installed (<u>https://java3d.dev.java.net/binary-builds.html</u>) in order to display a 3D plot. The table under the plot lists the values that are being plotted in the projection of each gene/experiment on the selected principal components.

To manipulate a 2D plot:

- Hover over a point to display its details.
- Click and drag or click Zoom In to zoom in on an area of the 2D plot.
- Click and drag or click Zoom Out to zoom out on an area of the 2D plot.
- Click Reset Plot to reset the default zoom level.
- Click Clear Plot to clear the right pane.
- Select the Display tick marks checkbox to show/hide tick marks along the axes.

• Right-click on the plot to display the following menu:

Properties	Set display options for the plot.
Save As	Save the plot to an image file.
Print	Print the plot.
Zoom in	Zoom in on either axis or both axes.
Zoom out	Zoom out on either axis or both axes.
Auto range	Reset the default zoom level for either axis or both axes.

To manipulate a 3D plot:

- Hover over a point to display its details.
- Alt+click and drag down on the vertical axis or click *Zoom In* to zoom in.
- Alt+click and drag up on the vertical axis or click *Zoom Out* to zoom out.
- Click and drag an axis to rotate the plot.
- Right-click and drag an axis to move the entire plot.
- Click Reset Plot to reset the plot to its default aspect.
- Click Clear Plot to clear the right pane.
- Select the *Display tick marks* checkbox to show/hide tick marks along the axes.
- Select the *Display grid lines* checkbox to show/hide grid lines on the plot.
- Right-click on the plot to display the following menu:

Save As Save the plot to an image file.

How to assign groups in a plot:

- 1. Select a point by clicking on a point in the plot or clicking on a row in the projection table.
- 2. Click on Set Group and provide a name and color for the new group.
- 3. To assign additional points to a recently created group, repeat step 1 and click on the colored icon to the left of *Set Group*.
- 4. To assign additional points to a previous group that is not the recent one, repeat step 1 and step2 but provide the name of the previous group.

The new group assignments for the point(s) should now appear under the Group column in the plot table.

How to remove a new group from the plot:

1. Click on *Remove Group* and check the box next to the group(s) to remove. Then click OK.

The group assignments for the points in the groups that were removed should now be reset to their original groups.

🚣 PCA¥iewer										_ 8 ×
File Help										
Components Projection										
Principal Component	Eigenvalue	% Variation				2D Pr	ojection			
1	1.9870408E8	88.917%				2011	ojeenon			
2	3846020.5	1.721%	7,000,000							
3	3364154.5	1.505%	6.500.000							
4	2949850.75	1.32%	6 000 000							
5	1867015.25	0.835%	5 500 000							
6	1514342.875	0.678%	5,000,000							
7	1192336.125	0.534%	1,000,000							
8	919274.062	0.411%	4,500,000							
9	806856.25	0.361%	4,000,000	•						
10	725710.625	0.325%	3,500,000							
11	615190.625	0.275%	3,000,000	• • • • • •						
12	570486.688	0.255%	2,500,000	•						
13	526281.188	0.236%	2,000,000							
14	448194.125	0.201%	1,500,000							
15	435792.594	0.195%	∾ 1.000.000							
16	403967.938	0.181%	8 500,000							
17	388824.594	0.174%	- C							
18	348798.219	0.156%	- 500,000	• • • • • • • • • • • • • • • • • • •						
19	331435.969	0.148%	.E -500,000							
20	323352.219	0.145%	-1,000,000							
21	278686.062	0.125%	-1,500,000	1 <b></b> /						
22	268118.75	0.12%	-2,000,000							
23	252465.219	0.113%	-2,500,000							
24	237581.297	0.106%	-3,000,000	• •						
25	223852.344	0.1%	-3,500,000							
26	218734.562	0.098%	-4,000,000							
27	203898.062	0.091%	-4.500.000							
28	201017.0	0.09%	-5.000.000	•						
29	187341.688	0.084%	5,500,000							
30	179344.719	0.08%	0,000,000							
31	158329.172	0.071%	-6,000,000							
32	152882.031	0.068%	-6,500,000							
33	123478.648	0.055%	-7,000,000							
34	114508.852	0.051%		AML 14 : [-1.8605262E7.	-6862103.5100.000	-5 000 000	ò	5 000 000	10.000.000	15 000 000
35	107957.656	0.048%					Prin Comn 1			
36	102335.156	0.046%								
37	93493.43	0.042%				AL	L 🗢 AML			
38	86575.281	0.039%						-	-	
				Sample	Pnn.	Comp. 1	Prin. Com	p. 2	Grou	10 dt
			ALL_14402_T-cell		-1.7682194E7		1269544.5		ALL	<u>^</u>
			ALL_17638_T-cell		-1.7617346E7		186219.78125		ALL	
			ALL_22474_T-cell		-1.7005492E7		3868334.25		ALL	
			AML_12		-1.7634174E7		-1662190.375		AML	
			AML 13		-1.6770848E7		2419068.25		AML	
			AML 14		-1 8605262E7		-6862103 5		AMI	-
			have been as a second s							M
% Var 90	.638	Plot Reset Plot	Clear Plot			Zoom In Zoom Out	Set Group	Remove Group	<ul> <li>Display tick marks</li> </ul>	<ul> <li>Display grid lines</li> </ul>

Figure 2. Projection tab projecting genes onto two principal components



Figure 3. Projection tab projecting genes onto three principal components

### References

• Raychaudhuri S, Stuart JM, and Altman RB. (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac. Symp. Biocomput. 455-466.

### Parameters

Name	Description
dataset file	The dataset analyzed by the PCA moduleres, .gct
s matrix file	The s matrix created by the PCA moduleodf
t matrix file	The t matrix created by the PCA moduleodf
u matrix file	The u matrix created by the PCA moduleodf
cls.or.sample.info.class	A cls or sample info file – .cls, .txt
sample.info.class	The name of the class column to use in the sample info file (required if sample info file provided)
featureset file	The feature set filegmt, .gmx, .grp

### Platform dependencies

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