

## SNP Cluster Plots

**Author:** Greta Linse Peterson, Golden Helix, Inc.

### Overview

This function creates scatter plots based on A and B allele intensities that can be split on SNP genotypes to create tri-colored cluster plots. The function will work for up to 100 SNPs at a time, although a separate scatterplot is created for each SNP.

You will need genotypes for up to 100 SNPs, and you must also have the Affymetrix CEL files available for import.

This document provides instructions on how to import the intensity data from the CEL files, how to use the function, and how to split the scatter plots based on genotypes.

### Recommended Directory Location

Save the script to the following directory:

\*..\Application Data\Golden Helix SVS\UserScripts\SVS\Tools\

**Note:** The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between operating systems. The easiest way to locate this directory on your computer is to open SVS and go to **Tools >Open Folder > UserScripts Folder** and save the script in the **\SVS\Tools** folder. If you save the script to the proper folder, it will be accessible from the project navigator **Tools** menu.

### Obtaining the Required Datasets

In order to create cluster plots you need the following items:

1. Genotypes for all samples for at most 100 SNPs.
2. Affymetrix CEL files for the Allele A and B intensities

Assuming that genotype analysis has been performed to determine significant SNPs and that cluster analysis now needs to be performed, the genotypes should already be available in the project. See **Figure 1** for a sample project. The relevant nodes are number 25, the Association Test results where the top 25 markers are selected (see **Figure 2**, and number 26, where all other SNPs (and phenotypes) are inactivated except for these markers (see **Figure 3**).

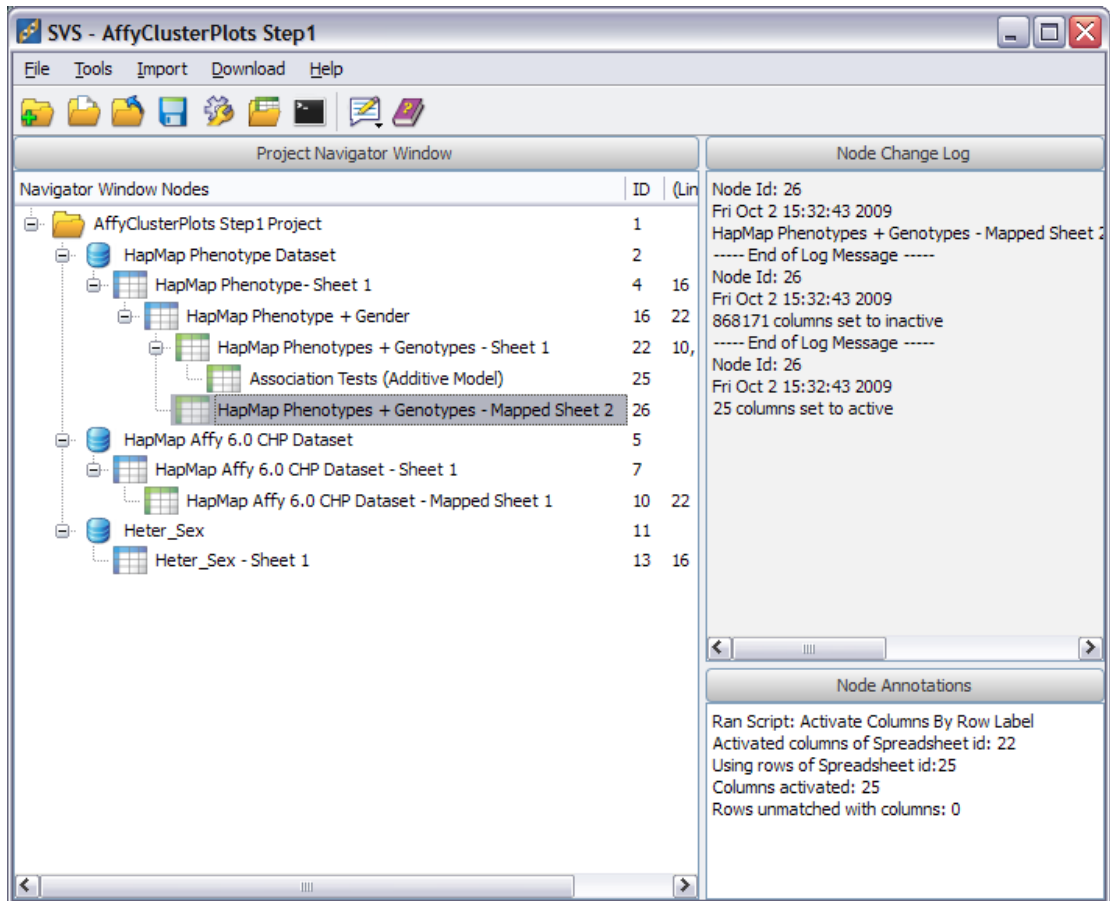


Figure 1: HapMap project with association results on gender

Map	Marker	R	1	R	2	R	3
		Corr/Trend	P	Corr/Trend	-log10 P	Corr/Trend	R
1	SNP_A-8631489	5.733930256	1275e-056	55.24154759	38386	-0.97748	1152385415
2	SNP_A-8401046	9.740143648	08917e-054	53.0114346	38079	-0.95714660	3234602
3	SNP_A-8547953	4.894364926	34959e-052	51.31030365	25585	-0.9599848	41489156
4	SNP_A-8325240	1.05829468	427999e-051	50.97539338	52905	-0.93640	1025876785
5	SNP_A-4304555	5.361777129	33609e-006	5.27069124	10829	0.277425	128574724
6	SNP_A-2092035	5.750584656	15366e-006	5.24028799	78283	-0.276525660	183003
7	SNP_A-8680119	5.937889836	995091e-006	5.22636786	393905	0.276112938	232312
8	SNP_A-2206294	6.98720328	716988e-006	5.15569662	110636	0.274008768	420886
9	SNP_A-8372209	9.014187280	62345e-006	5.04507342	317913	-0.2706849	11867888
10	SNP_A-8391943	1.52318114	779921e-005	4.81724844	12861	-0.26371875	1244533
11	SNP_A-2164086	1.52318114	779921e-005	4.81724844	12861	-0.26371875	1244533
12	SNP_A-2250877	1.536569669	50338e-005	4.81344774	36439	0.26360	1110234547
13	SNP_A-4208034	1.610660029	58149e-005	4.7929961	11871052	0.262967256	672239
14	SNP_A-2300919	1.72122872	236835e-005	4.764161415	40031	-0.262071209	18048
15	SNP_A-8505180	1.936352389	20692e-005	4.71301560	427888	0.26096042	5316611
16	SNP_A-1815694	2.092368574	40098e-005	4.67936181	1132991	0.25941967	5387037
17	SNP_A-2305874	2.447722466	26159e-005	4.61123782	598892	0.25727152	5080428
18	SNP_A-2090743	2.901323650	97878e-005	4.53740382	169926	-0.25492472	6337964
19	SNP_A-1915166	2.90714352	2042e-005	4.53653352	12457	0.2548969	47044924
20	SNP_A-8638676	3.06242271	435297e-005	4.51393486	267755	0.2541746	44857863
21	SNP_A-8607724	3.166440003	15045e-005	4.49942873	645319	0.2537100	1313334
22	SNP_A-8454246	3.238251899	01636e-005	4.48968937	11763	-0.2538699	4330706
23	SNP_A-8331724	3.567448889	78842e-005	4.4476422	4022925	0.2520449	47359688
24	SNP_A-8288234	3.898150995	11941e-005	4.40914134	246902	-0.2508005	48555345
25	SNP_A-2296771	4.189098516	47735e-005	4.3778794	2606528	0.249785989	402922
26	SNP_A-8573561	5.26812874	119799e-005	4.2783436	2040185	-0.2474521	54886071
27	SNP_A-8460199	5.275104690	29514e-005	4.2777689	1688973	0.24651163	32851483
28	SNP_A-4287427	5.63185524	172114e-005	4.2493485	1660204	0.24557488	2558495
29	SNP_A-8435839	5.698321707	22038e-005	4.2442530	35664	0.24540659	1476520

Figure 2: Association Test results, sorted ascending on p-value, top markers selected

Map	CELFILE	SNP_A-8479801	SNP_A-8631489	SNP_A-8475540
1	NAO6985_GW6_C	B_B	B_B	A_B
2	NAO6991_GW6_C	B_B	B_B	B_B
3	NAO6993_GW6_C	B_B	A_B	B_B
4	NAO6994_GW6_C	B_B	A_B	B_B
5	NAO7000_GW6_C	B_B	B_B	B_B
6	NAO7019_GW6_C	B_B	B_B	B_B
7	NAO7022_GW6_C	B_B	A_B	B_B
8	NAO7029_GW6_C	B_B	A_B	B_B
9	NAO7034_GW6_C	B_B	A_B	B_B
10	NAO7048_GW6_C	B_B	A_B	B_B
11	NAO7055_GW6_C	B_B	B_B	B_B
12	NAO7056_GW6_C	B_B	B_B	B_B
13	NAO7345_GW6_C	B_B	B_B	B_B

Figure 3: Only top markers are active in the Phenotype + Genotype spreadsheet

Next, the allele intensities need to be imported.

- 1) Open the Affymetrix CEL import window by going to **Import > Affymetrix > CEL**.
- 2) Select the CEL files using the File and Directory choosers. See **Figure 4** for an example of importing some HapMap data from Affymetrix 6.0 CEL files.

**Import Options**

Platform: GenomeWideSNP\_6

Import Options

Reference Set

All samples

Defined subset from spreadsheet

0=Ref 1=Non-Ref Column:  Select Sheet

Don't include reference samples in output spreadsheet

Precomputed populations All (270 HapMap Samples)

Marker map: Affy SNP6 Marker Map.dsm (Local) Select Marker Map

CDF File: GenomeWideSNP\_6.Full.cdf (Local) View Library Folder

Specify temp directory  Browse

Datasets to Import

A/B intensities  Quantile normalized A/B intensities

Unmerged LogRs in separate channels

LogR ratios - samples as columns  LogR ratios - samples as rows

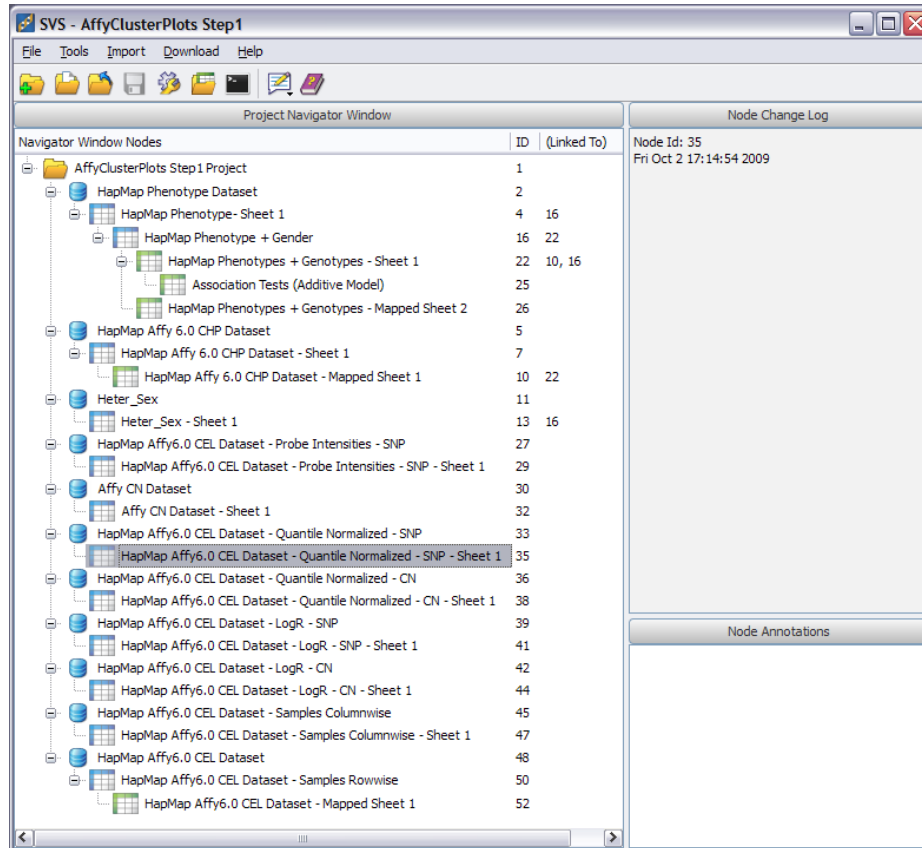
Dataset name base: Affy CEL Dataset

< Back Finish Cancel Help

Figure 4: Affymetrix CEL file import window, settings for Affymetrix 6.0 CEL files

- 3) **IMPORTANT:** Notice that the option “**Quantile normalized A/B intensities**” is checked. This is a crucial step to import the allele A and B intensities. Otherwise only the log ratio data will be imported.

- 4) After the CEL files are imported, the project should contain the CEL dataset and the normalized intensity data. See **Figure 5**. In the case of Affymetrix 500k datasets, there will be two Quantile Normalized SNP datasets, one for NSP and one for STY arrays. These two sheets can either be joined together or treated separately.



**Figure 5: Project after importing CEL files including intermediate steps**

### Using the Function

- 1) To use the function, go to **Tools > SNP Cluster Plots**.
- 2) Select the genotype spreadsheet that has up to 100 markers active.
- 3) Next, select the Quantile Normalized SNP spreadsheet.
- 4) The Quantile Normalized SNP spreadsheet will be transformed from two columns per sample (first column for A allele intensity, second column for B allele intensity) to two columns per marker, (first column for A allele intensity, second column for B allele intensity) column headers labeled accordingly (see Error! Reference source not found.). XY scatter plots will also be generated for each marker selected. These plots can be split on genotypes to create up to three colored genotype plots.

- 5) To finish creating a genotype plot, follow the following steps.
  - a. Select and open an XY scatter plot (**Figure 6**).

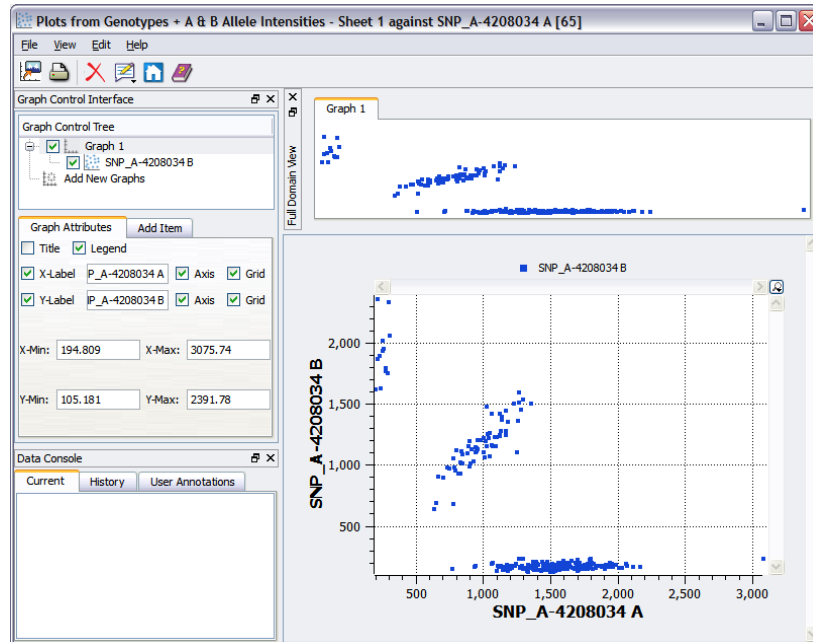


Figure 6: XY scatter plot of allele A & B intensities for SNP\_A-4208034

- b. Select the graph item labeled SNP\_A-4208034 B, and click on the **Filter** tab, **Figure 7**.

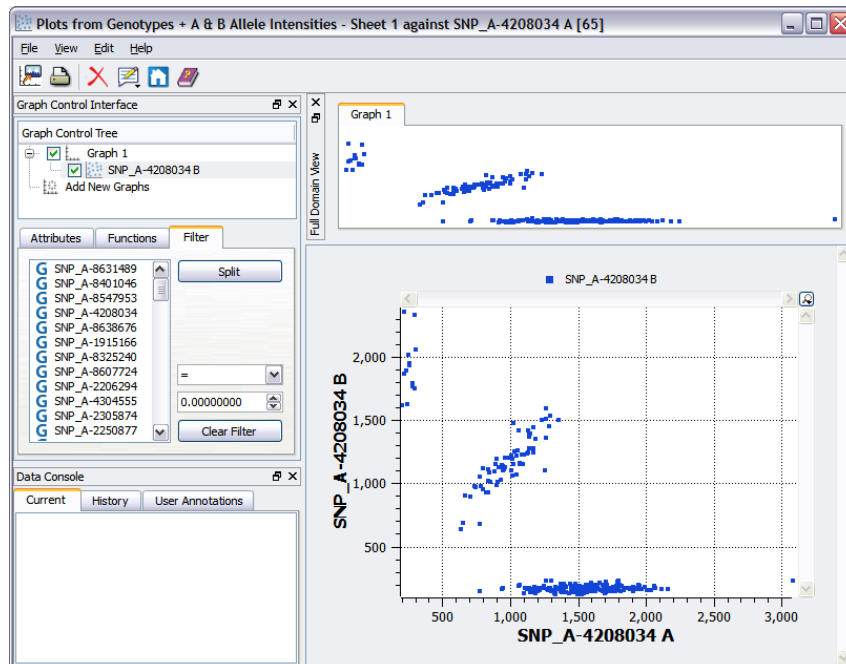
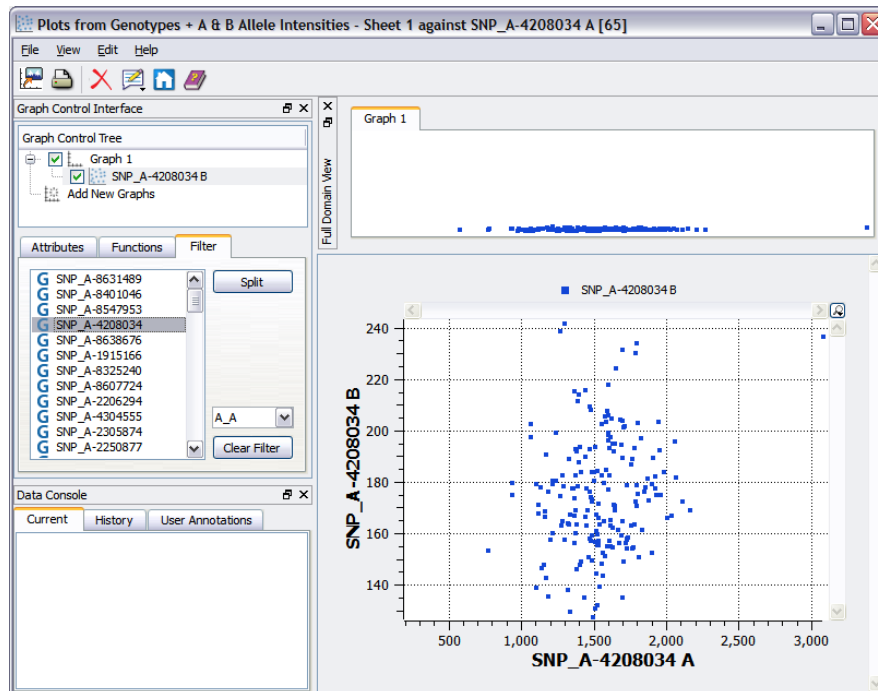


Figure 7: Graph item with Filter tab selected

- c. Next, select SNP\_A-4208034 in the filter list, make sure that there is a blue G next to the item selected, and click **Split**, **Figure 8**.



**Figure 8: A & B Allele intensity scatter plot before clicking "Split" on the same marker genotypes**

The resultant plot will have a different color for each genotype found for all samples for the selected marker. The plot can be edited to change the X and Y axis labels, the legend labels, and to add a title to the graph. The plot can be printed or saved by going to **File > Save as Image** or **Print**.

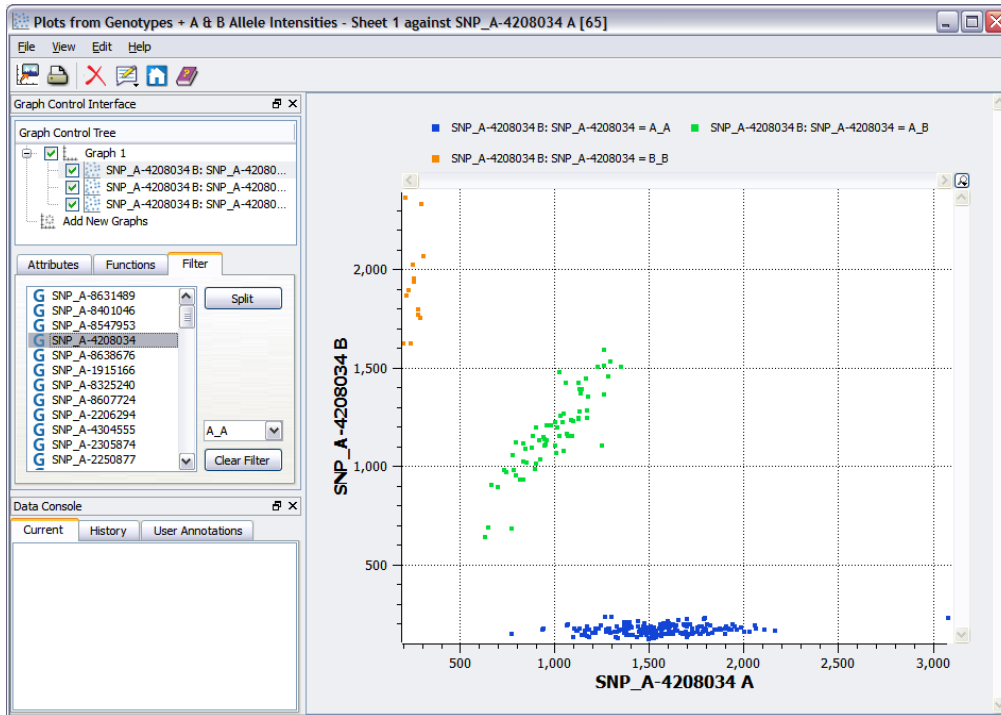


Figure 9: Genotype Cluster Plot for SNP\_A-4208034