Array Track

Tutorial 4:
Data Exploring Tools

Data Exploring Tool

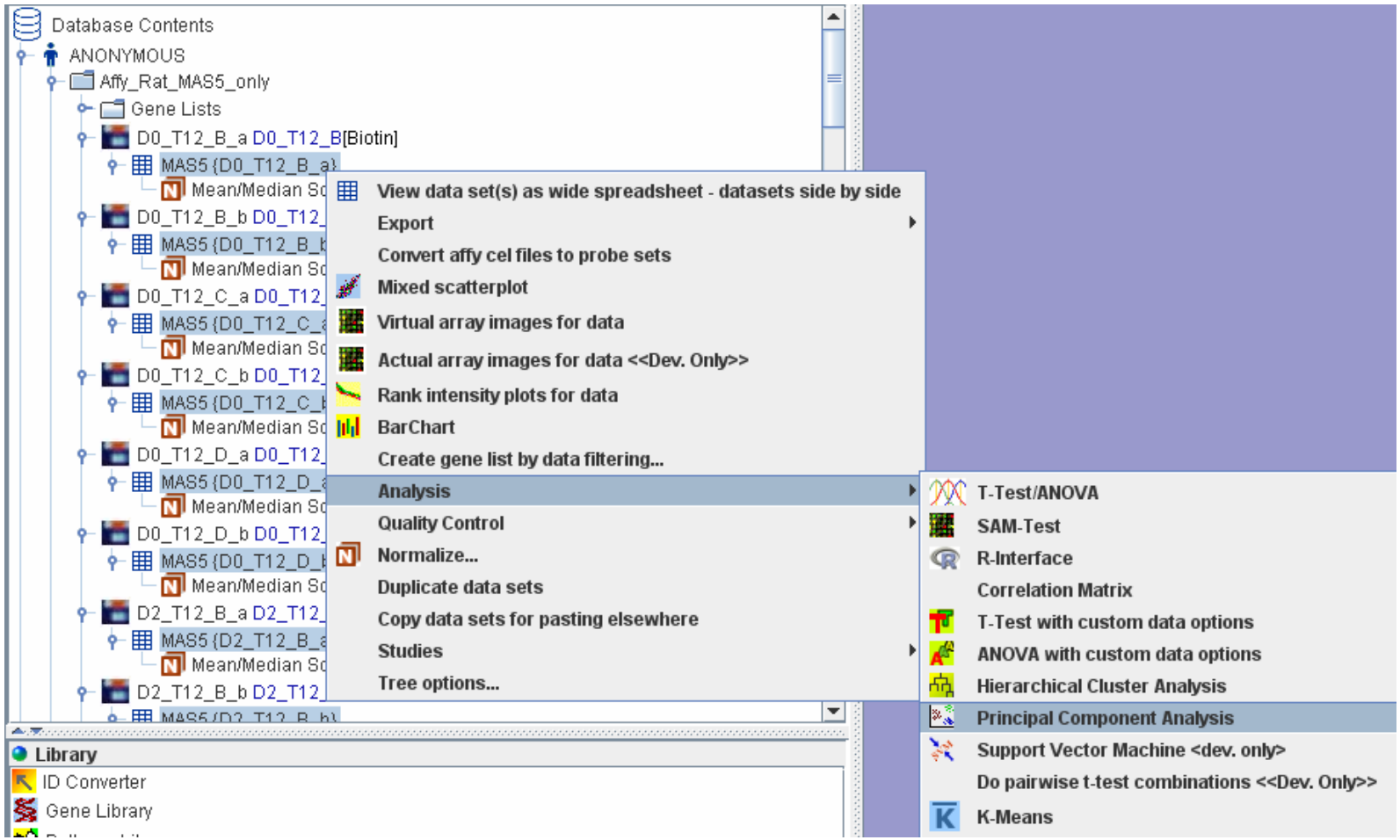
ArrayTrack provides some tools for data analysis. These tools can be accessed from the Tool panel, pull-down menu or database panel.

This tutorial will cover Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Correlation matrix, Scatter plot and K-Means.

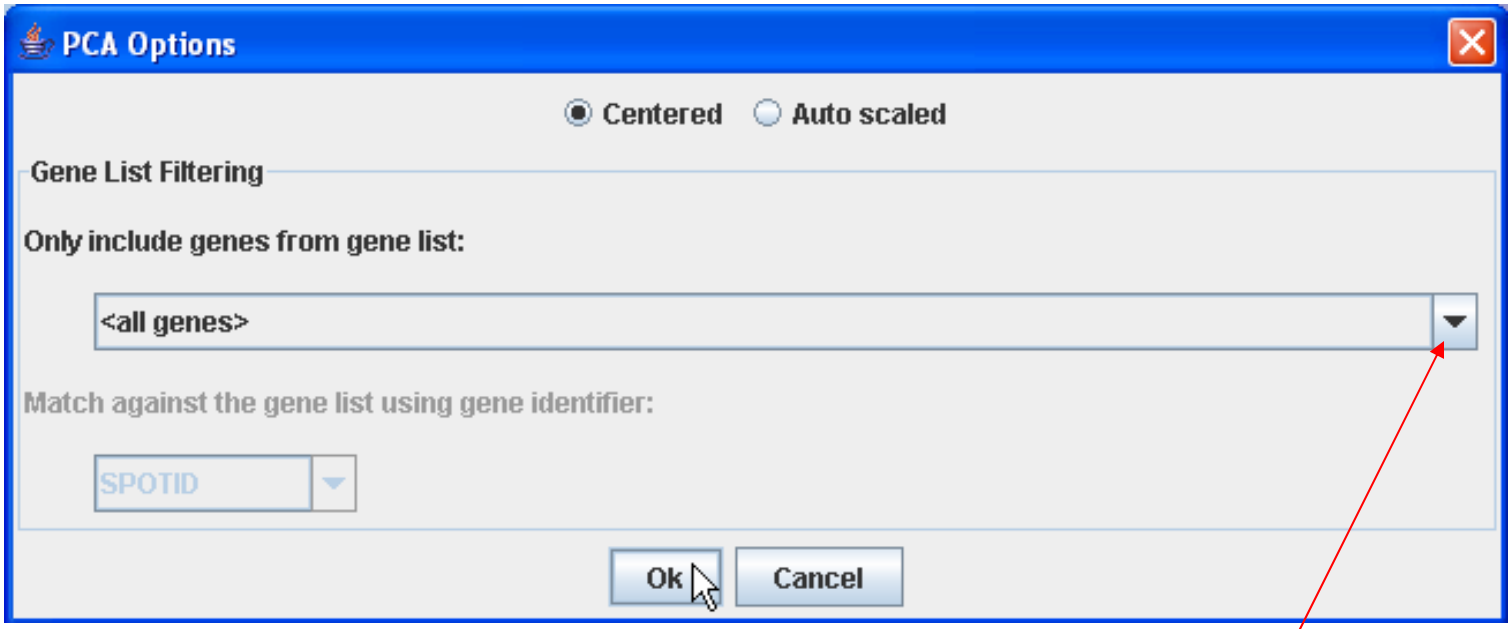
Principal Component Analysis (PCA)

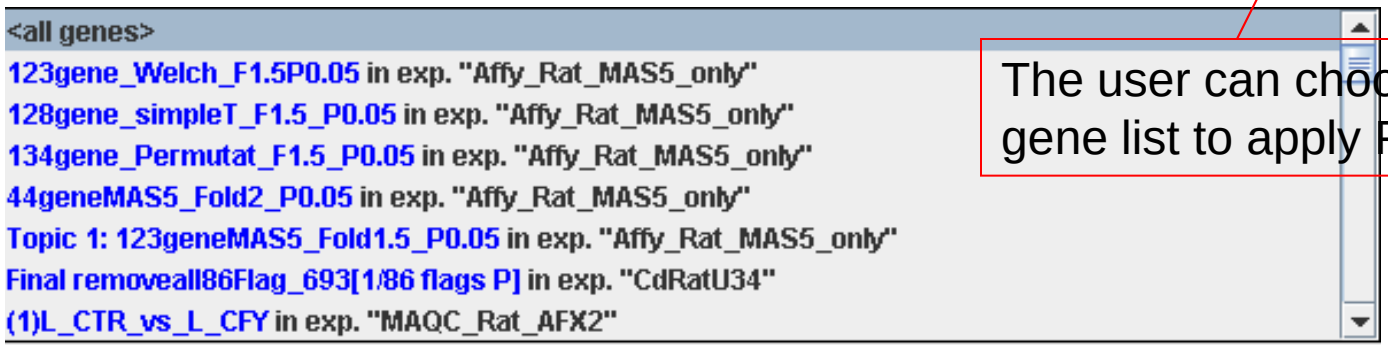
Select dataset, right-click, choose Analysis ->Principal Component Analysis

Right-click the selected data, choose “Analysis “ ->PCA



PCA - continued

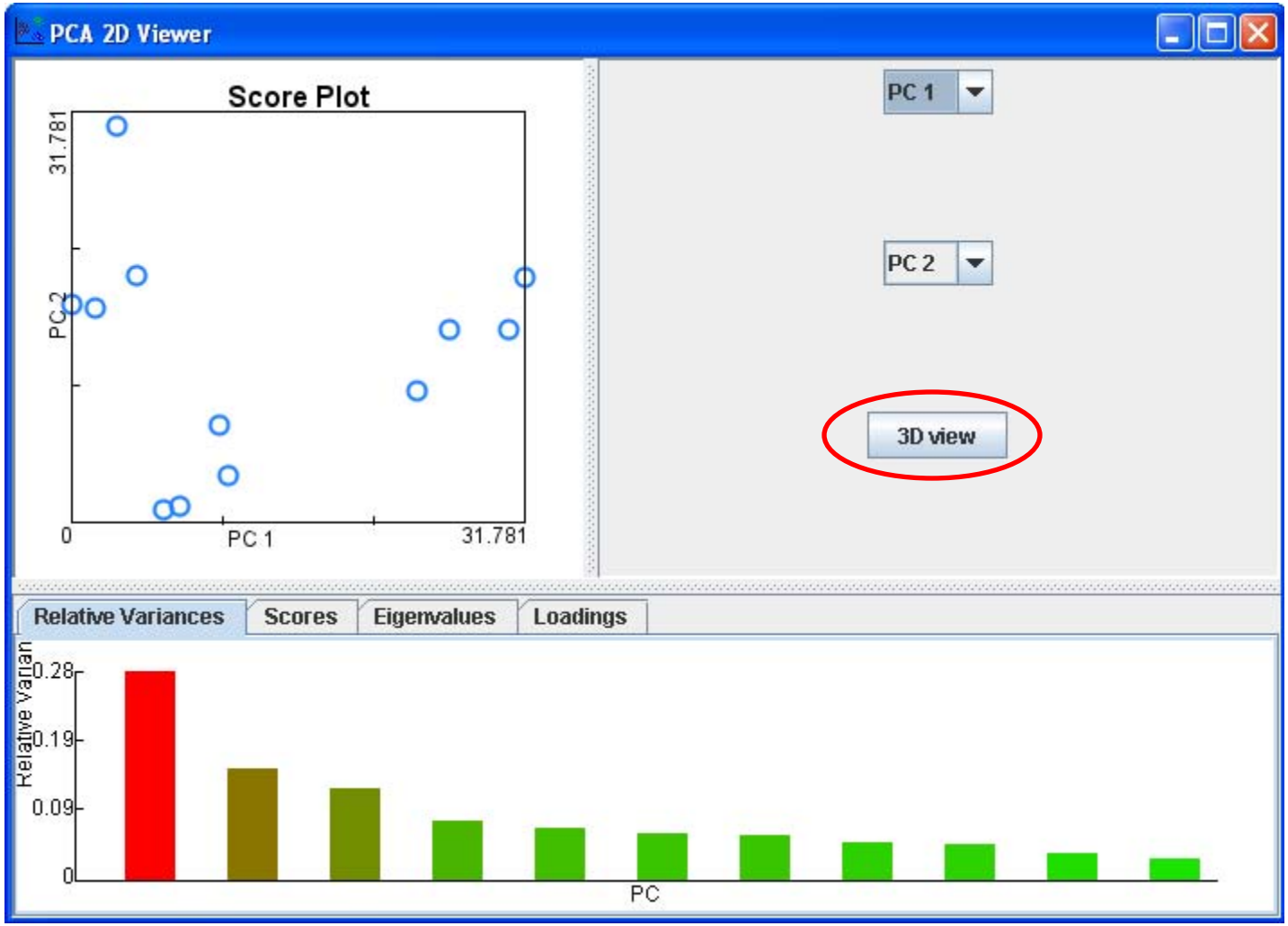




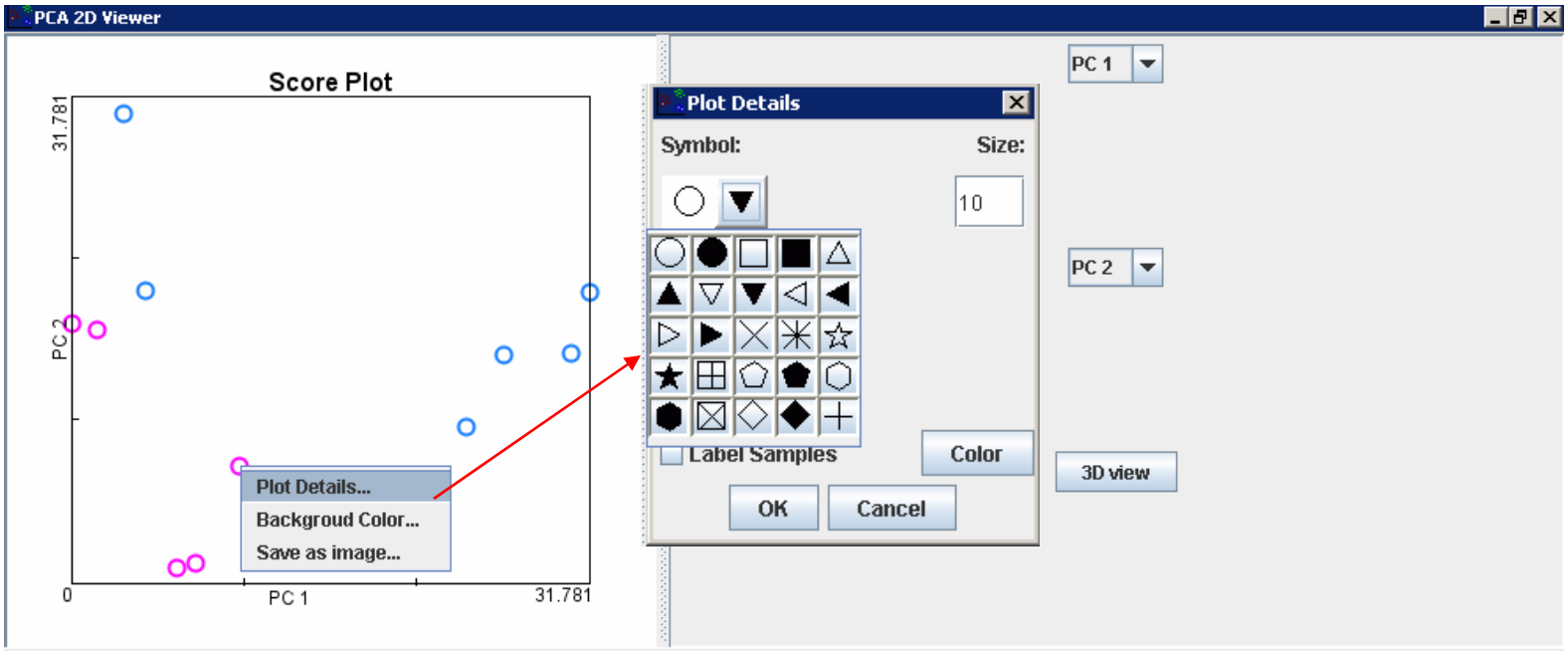
The user can choose a specific gene list to apply PCA

PCA - continued

This is the PCA 2D view. Click 3D view button will bring out the 3D image, see slide #7.



PCA - continued



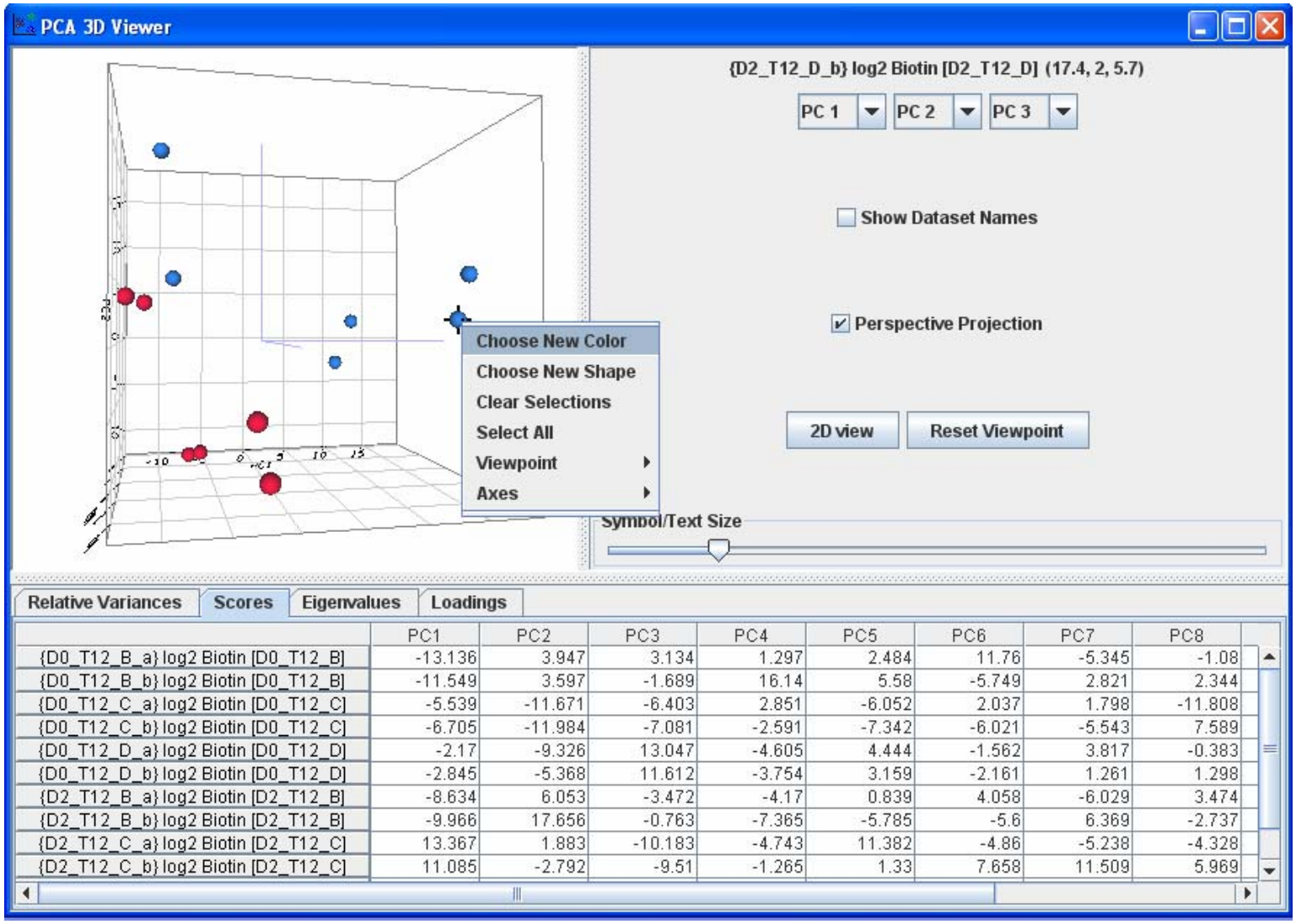
Relative Variances | Scores | Eigenvalues | Loadings

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
{D0_T12_B_a} log2 ...	-13.136	3.947	3.134	1.297	2.484	11.76	-5.345	-1.08	-8.292	0.949	-0.407
{D0_T12_B_b} log2 ...	-11.549	3.597	-1.689	16.14	5.58	-5.749	2.821	2.344	0.857	-0.197	0.441
{D0_T12_C_a} log2 ...	-5.539	-11.671	-6.403	2.851	-6.052	2.037	1.798	-11.808	3.318	0.724	-0.322
{D0_T12_C_b} log2 ...	-6.705	-11.984	-7.081	-2.591	-7.342	-6.021	-5.543	7.589	-4.763	-0.571	-0.048
{D0_T12_D_a} log2 ...	-2.17	-9.326	13.047	-4.605	4.444	-1.562	3.817	-0.383	-0.308	-2.452	7.788
{D0_T12_D_b} log2 ...	-2.845	-5.368	11.612	-3.754	3.159	-2.161	1.261	1.298	2.968	5.285	-8.243
{D2_T12_B_a} log2 ...	-8.634	6.053	-3.472	-4.17	0.839	4.058	-6.029	3.474	11.513	-3.398	1.007
{D2_T12_B_b} log2 ...	-9.966	17.656	-0.763	-7.365	-5.785	-5.6	6.369	-2.737	-2.731	-0.009	0.071
{D2_T12_C_a} log2 ...	13.367	1.883	-10.183	-4.743	11.382	-4.86	-5.238	-4.328	-2.11	2.197	0.584
{D2_T12_C_b} log2 ...	11.085	-2.792	-9.51	-1.265	1.33	7.658	11.509	5.969	-0.231	0.466	-0.541
{D2_T12_D_a} log2 ...	18.645	6.006	5.644	5.048	-7.659	0.922	-3.804	1.281	1.659	6.975	3.686
{D2_T12_D_b} log2 ...	17.447	2	5.664	3.157	-2.381	-0.482	-1.616	-1.619	-1.88	-9.969	-4.015

Under Score tab, there is the table displaying the scores for all the hybridizations. Select some Rows, the corresponding spots in the PCA plot will be highlighted. Right-click those spots Choose Plot Details, the user can define the shape of the spots and color of the spot

PCA - continued

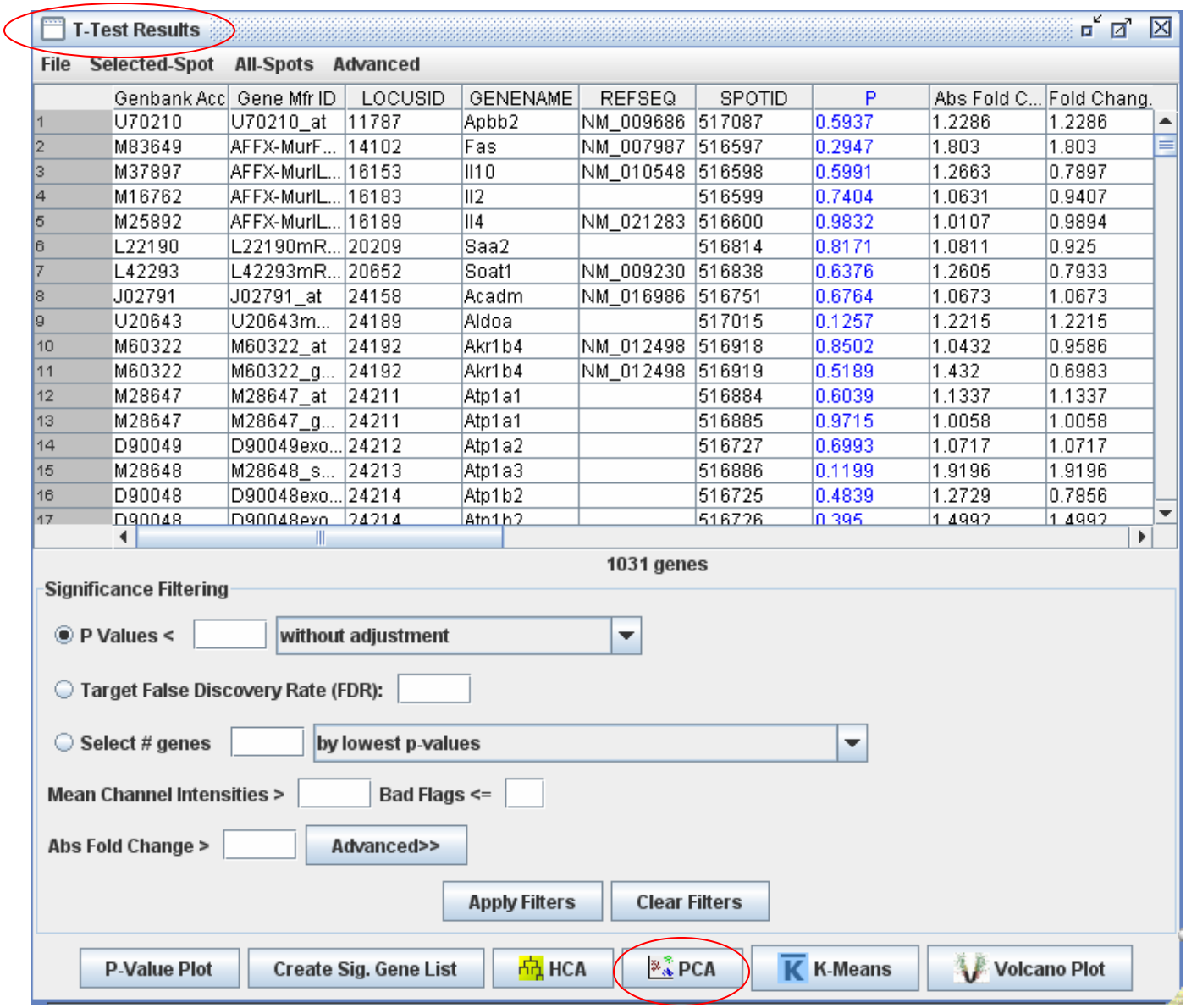
PCA 3D view



	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
{D0_T12_B_a} log2 Biotin [D0_T12_B]	-13.136	3.947	3.134	1.297	2.484	11.76	-5.345	-1.08
{D0_T12_B_b} log2 Biotin [D0_T12_B]	-11.549	3.597	-1.689	16.14	5.58	-5.749	2.821	2.344
{D0_T12_C_a} log2 Biotin [D0_T12_C]	-5.539	-11.671	-6.403	2.851	-6.052	2.037	1.798	-11.808
{D0_T12_C_b} log2 Biotin [D0_T12_C]	-6.705	-11.984	-7.081	-2.591	-7.342	-6.021	-5.543	7.589
{D0_T12_D_a} log2 Biotin [D0_T12_D]	-2.17	-9.326	13.047	-4.605	4.444	-1.562	3.817	-0.383
{D0_T12_D_b} log2 Biotin [D0_T12_D]	-2.845	-5.368	11.612	-3.754	3.159	-2.161	1.261	1.298
{D2_T12_B_a} log2 Biotin [D2_T12_B]	-8.634	6.053	-3.472	-4.17	0.839	4.058	-6.029	3.474
{D2_T12_B_b} log2 Biotin [D2_T12_B]	-9.966	17.656	-0.763	-7.365	-5.785	-5.6	6.369	-2.737
{D2_T12_C_a} log2 Biotin [D2_T12_C]	13.367	1.883	-10.183	-4.743	11.382	-4.86	-5.238	-4.328
{D2_T12_C_b} log2 Biotin [D2_T12_C]	11.085	-2.792	-9.51	-1.265	1.33	7.658	11.509	5.969

PCA - continued

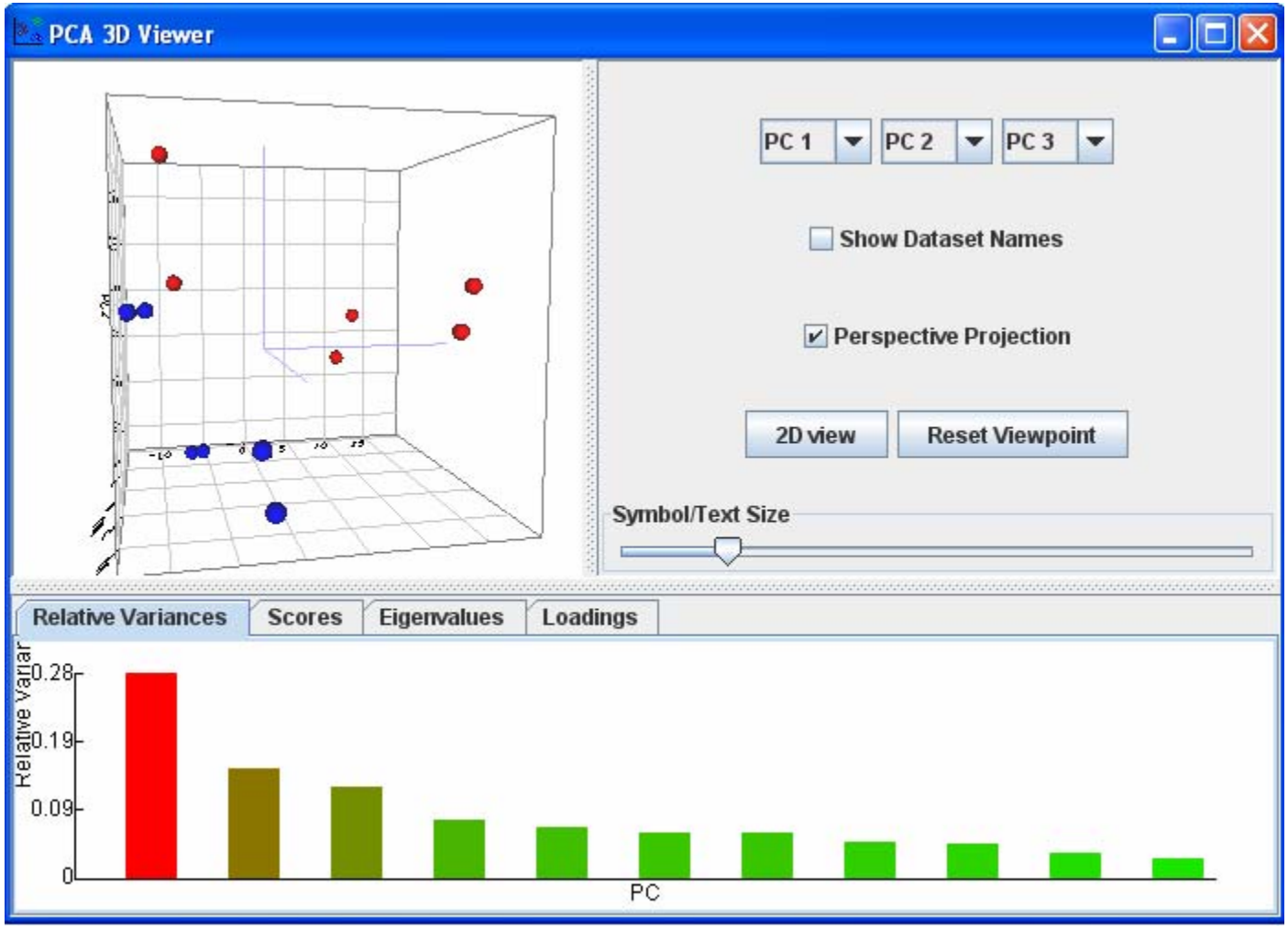
PCA can also be accessed from T-test result



	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang.
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.5937	1.2286	1.2286
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.2947	1.803	1.803
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.5991	1.2663	0.7897
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.7404	1.0631	0.9407
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.9832	1.0107	0.9894
6	L22190	L22190mR...	20209	Saa2		516814	0.8171	1.0811	0.925
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.6376	1.2605	0.7933
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.6764	1.0673	1.0673
9	U20643	U20643m...	24189	Aldoa		517015	0.1257	1.2215	1.2215
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.8502	1.0432	0.9586
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.5189	1.432	0.6983
12	M28647	M28647_at	24211	Atp1a1		516884	0.6039	1.1337	1.1337
13	M28647	M28647_g...	24211	Atp1a1		516885	0.9715	1.0058	1.0058
14	D90049	D90049exo...	24212	Atp1a2		516727	0.6993	1.0717	1.0717
15	M28648	M28648_s...	24213	Atp1a3		516886	0.1199	1.9196	1.9196
16	D90048	D90048exo...	24214	Atp1b2		516725	0.4839	1.2729	0.7856
17	D90048	D90048exo...	24214	Atp1b2		516726	0.395	1.4992	1.4992

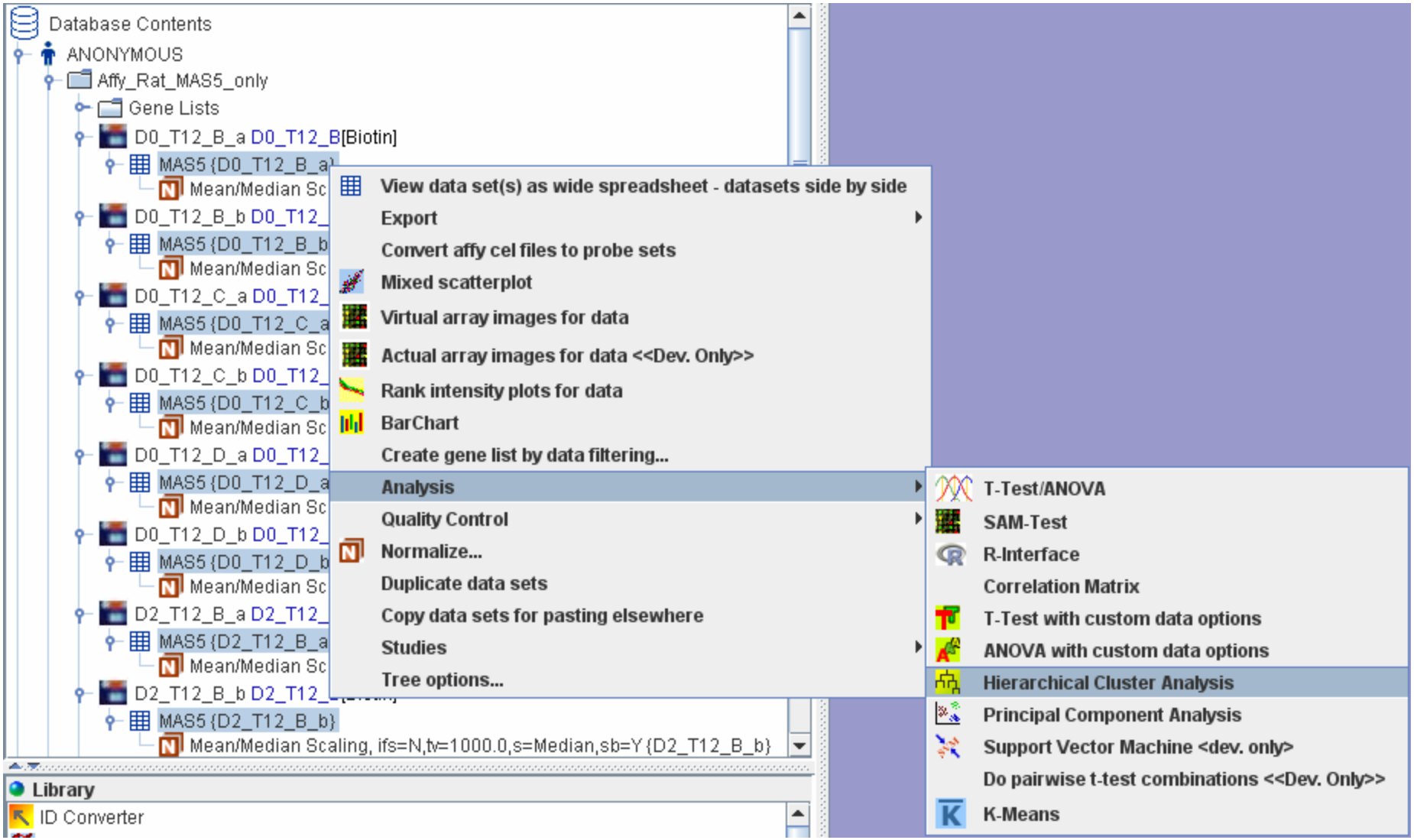
PCA - continued

PCA from T-test result. Since data are assigned in two groups in T-test, here The PCA 3D view automatically shows the two groups in two colors.

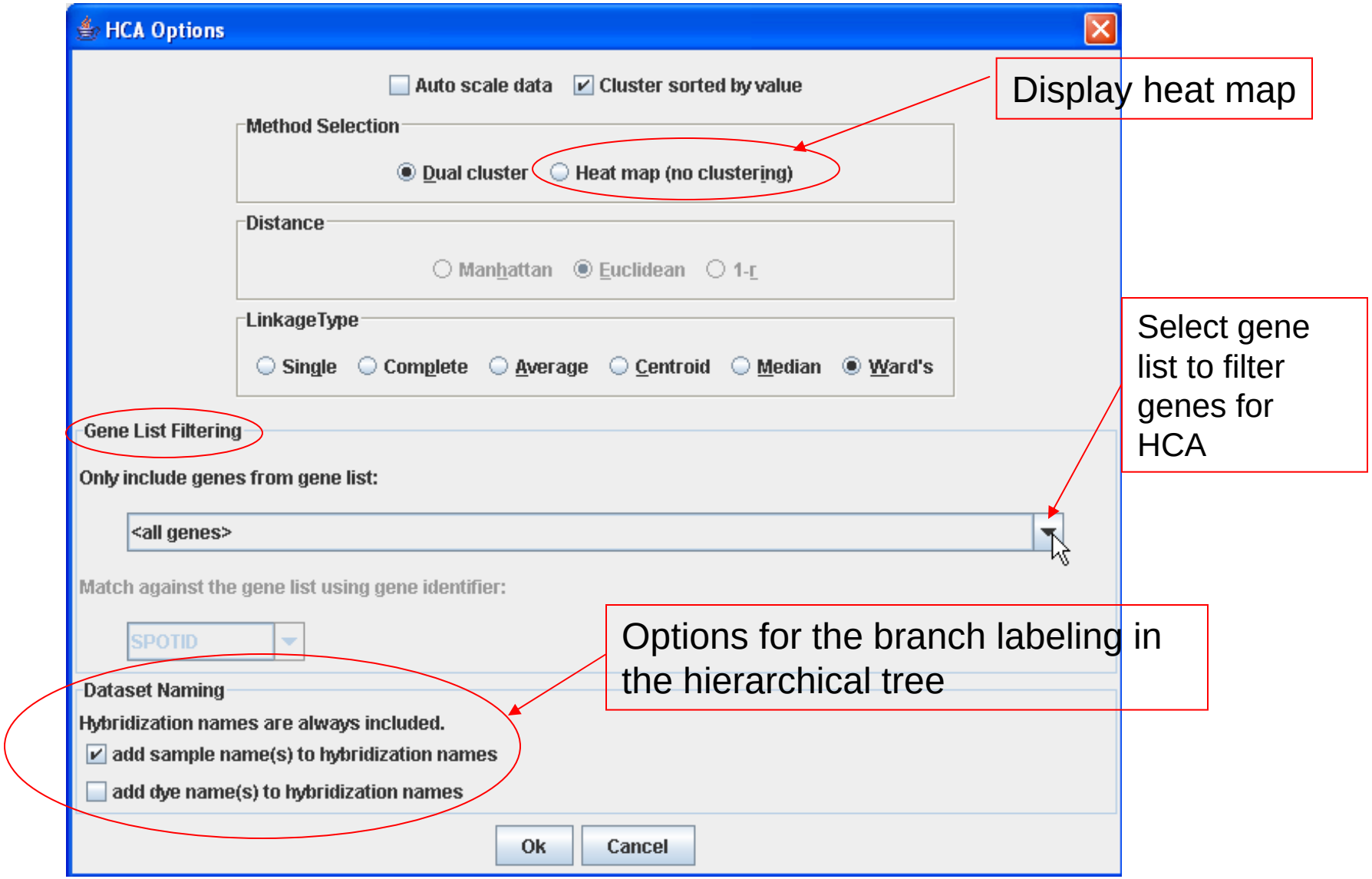


Hierarchical Cluster Analysis (HCA)

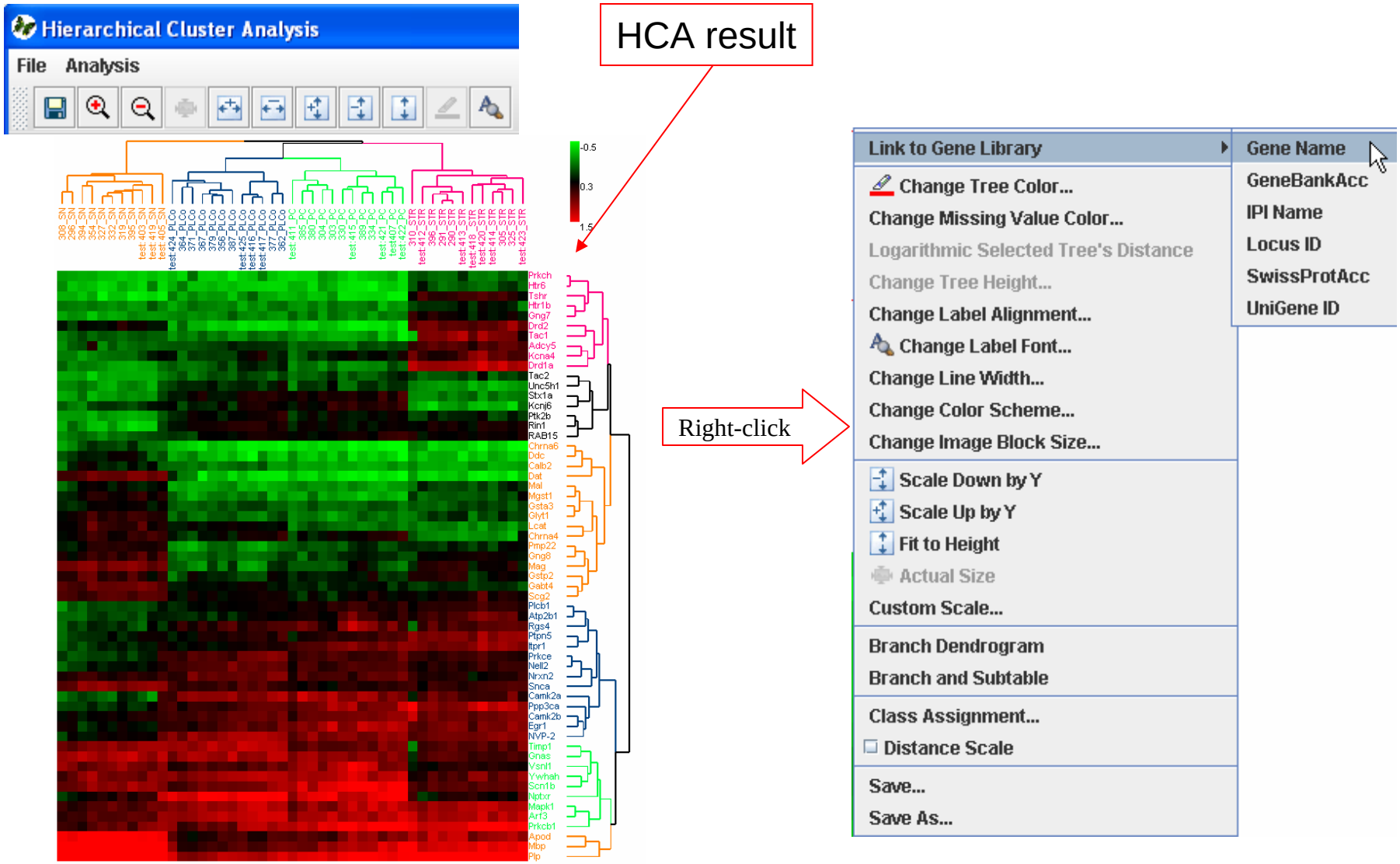
Select dataset, right-click, choose Analysis ->Hierarchical Cluster Analysis



HCA - continued

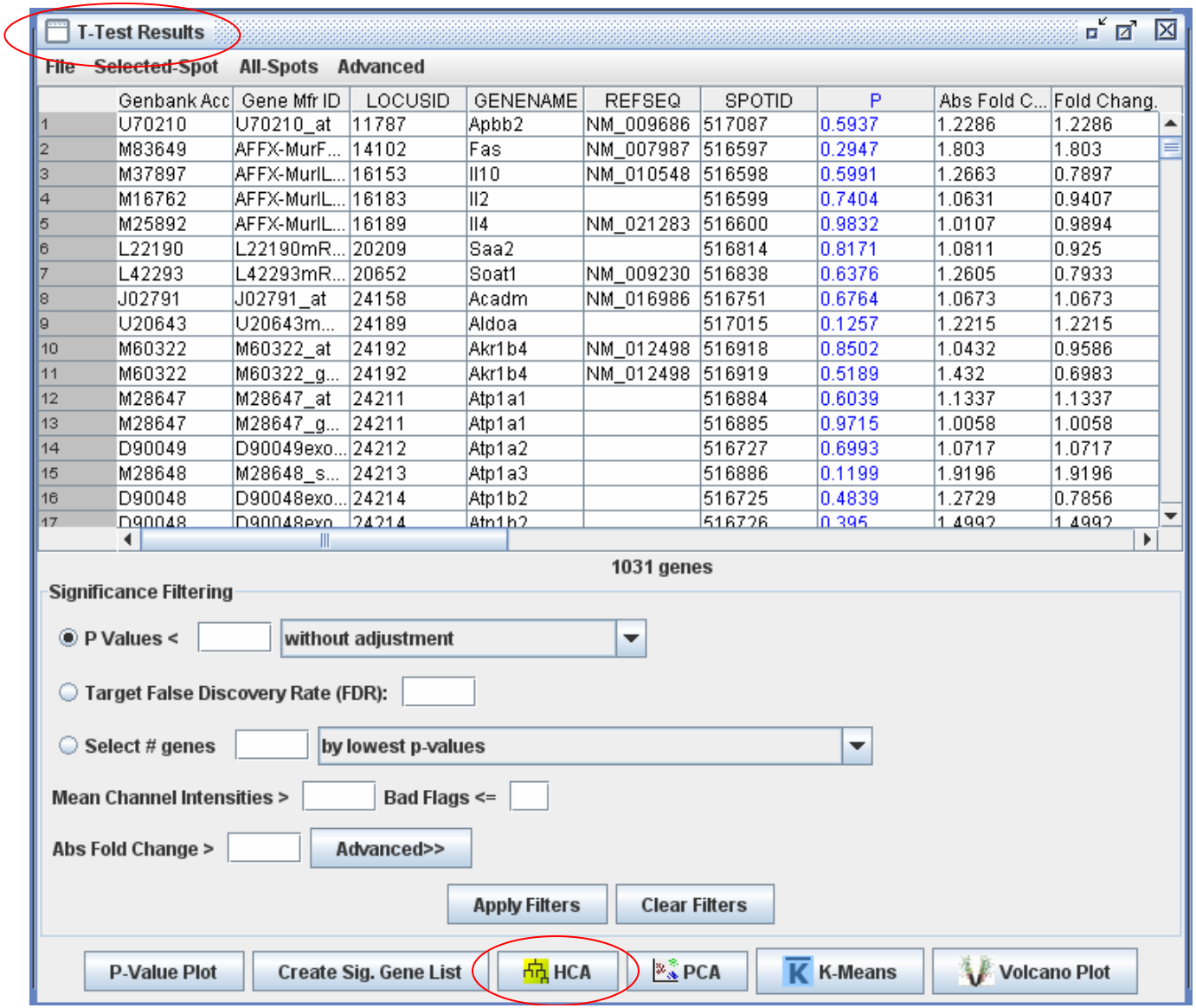


HCA - continued



HCA - continued

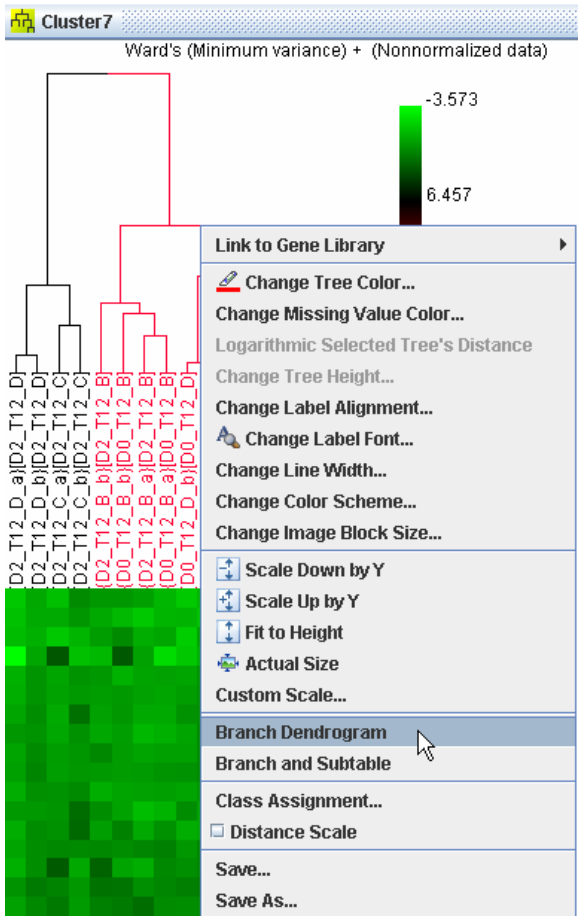
HCA can also be accessed from T-test result.

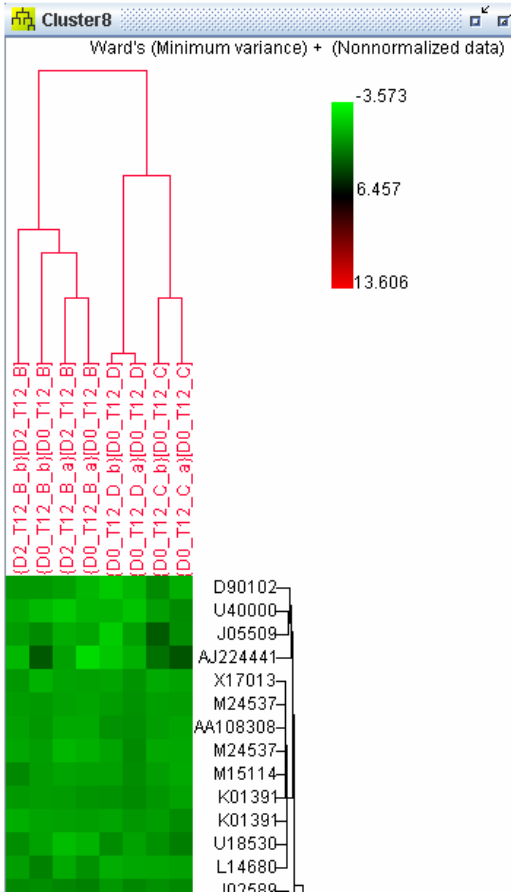


	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	Abs Fold C...	Fold Chang.
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.5937	1.2286	1.2286
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.2947	1.803	1.803
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.5991	1.2663	0.7897
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.7404	1.0631	0.9407
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.9832	1.0107	0.9894
6	L22190	L22190mR...	20209	Saa2		516814	0.8171	1.0811	0.925
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.6376	1.2605	0.7933
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.6764	1.0673	1.0673
9	U20643	U20643m...	24189	Aldoa		517015	0.1257	1.2215	1.2215
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.8502	1.0432	0.9586
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.5189	1.432	0.6983
12	M28647	M28647_at	24211	Atp1a1		516884	0.6039	1.1337	1.1337
13	M28647	M28647_g...	24211	Atp1a1		516885	0.9715	1.0058	1.0058
14	D90049	D90049exo...	24212	Atp1a2		516727	0.6993	1.0717	1.0717
15	M28648	M28648_s...	24213	Atp1a3		516886	0.1199	1.9196	1.9196
16	D90048	D90048exo...	24214	Atp1b2		516725	0.4839	1.2729	0.7856
17	D90048	D90048exo...	24214	Atp1b2		516726	0.395	1.4992	1.4992

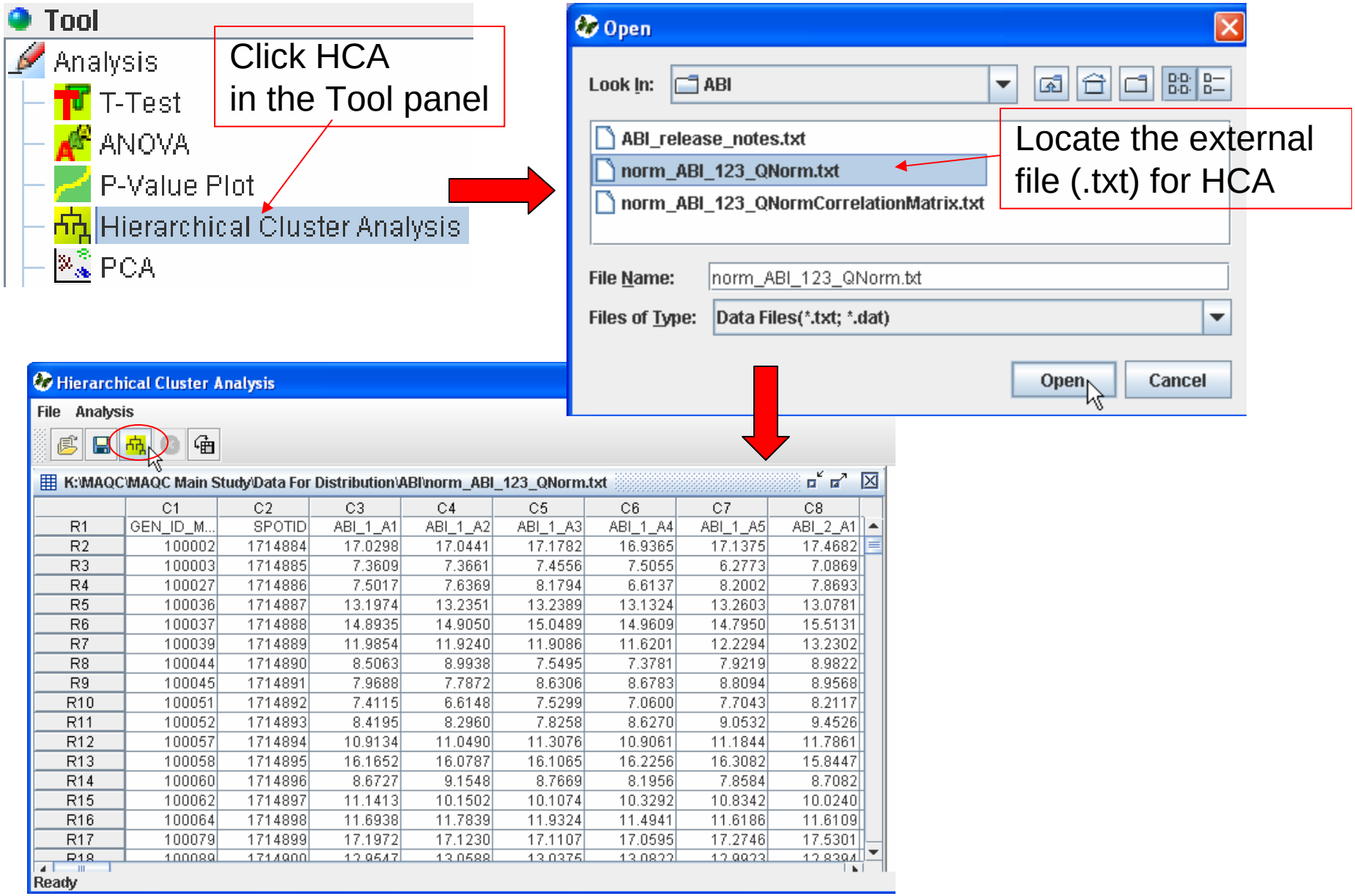
Extensive Features in HCA

- Zoom in and zoom out but clicking or
- Change the font and the color of the label for each branch of the tree.
- From the HCA plot there is a link to Gene Library according to the available IDs



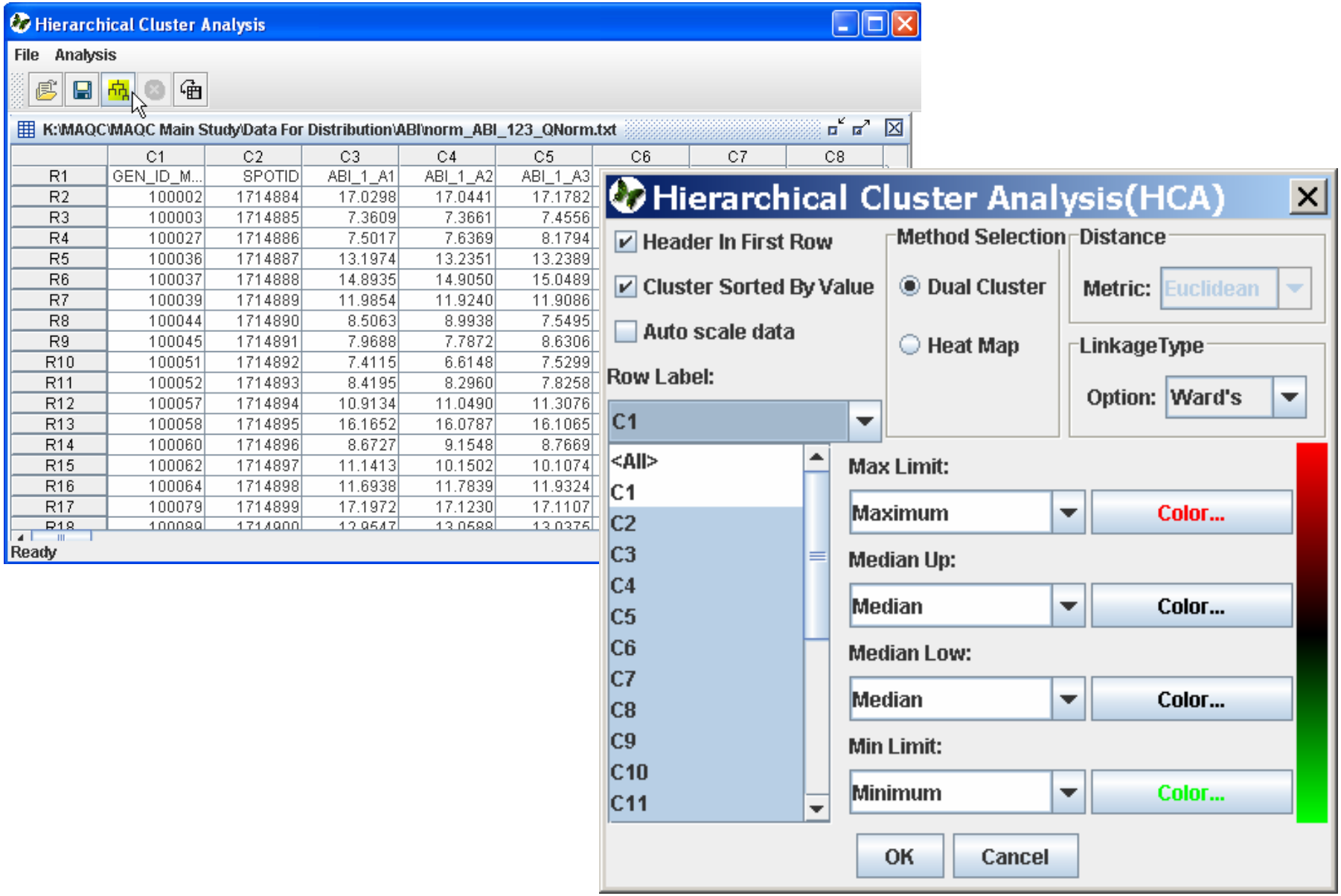


Apply HCA and PCA to the external files



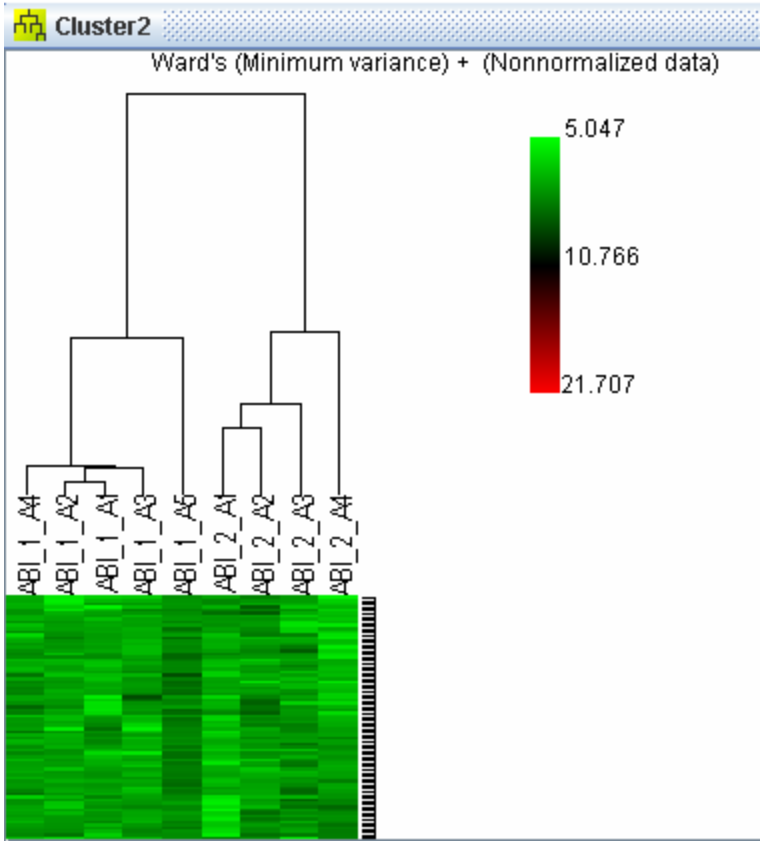
	C1	C2	C3	C4	C5	C6	C7	C8
R1	GEN_ID_M...	SPOTID	ABI_1_A1	ABI_1_A2	ABI_1_A3	ABI_1_A4	ABI_1_A5	ABI_2_A1
R2	100002	1714884	17.0298	17.0441	17.1782	16.9365	17.1375	17.4682
R3	100003	1714885	7.3609	7.3661	7.4556	7.5055	6.2773	7.0869
R4	100027	1714886	7.5017	7.6369	8.1794	6.6137	8.2002	7.8693
R5	100036	1714887	13.1974	13.2351	13.2389	13.1324	13.2603	13.0781
R6	100037	1714888	14.8935	14.9050	15.0489	14.9609	14.7950	15.5131
R7	100039	1714889	11.9854	11.9240	11.9086	11.6201	12.2294	13.2302
R8	100044	1714890	8.5063	8.9938	7.5495	7.3781	7.9219	8.9822
R9	100045	1714891	7.9688	7.7872	8.6306	8.6783	8.8094	8.9568
R10	100051	1714892	7.4115	6.6148	7.5299	7.0600	7.7043	8.2117
R11	100052	1714893	8.4195	8.2960	7.8258	8.6270	9.0532	9.4526
R12	100057	1714894	10.9134	11.0490	11.3076	10.9061	11.1844	11.7861
R13	100058	1714895	16.1652	16.0787	16.1065	16.2256	16.3082	15.8447
R14	100060	1714896	8.6727	9.1548	8.7669	8.1956	7.8584	8.7082
R15	100062	1714897	11.1413	10.1502	10.1074	10.3292	10.8342	10.0240
R16	100064	1714898	11.6938	11.7839	11.9324	11.4941	11.6186	11.6109
R17	100079	1714899	17.1972	17.1230	17.1107	17.0595	17.2746	17.5301
R18	100089	1714900	12.9547	13.0588	13.0375	13.0822	12.9923	12.8394

Apply HCA and PCA to the external files – continued



	C1	C2	C3	C4	C5	C6	C7	C8
R1	GEN_ID_M...	SPOTID	ABI_1_A1	ABI_1_A2	ABI_1_A3			
R2	100002	1714884	17.0298	17.0441	17.1782			
R3	100003	1714885	7.3609	7.3661	7.4556			
R4	100027	1714886	7.5017	7.6369	8.1794			
R5	100036	1714887	13.1974	13.2351	13.2389			
R6	100037	1714888	14.8935	14.9050	15.0489			
R7	100039	1714889	11.9854	11.9240	11.9086			
R8	100044	1714890	8.5063	8.9938	7.5495			
R9	100045	1714891	7.9688	7.7872	8.6306			
R10	100051	1714892	7.4115	6.6148	7.5299			
R11	100052	1714893	8.4195	8.2960	7.8258			
R12	100057	1714894	10.9134	11.0490	11.3076			
R13	100058	1714895	16.1652	16.0787	16.1065			
R14	100060	1714896	8.6727	9.1548	8.7669			
R15	100062	1714897	11.1413	10.1502	10.1074			
R16	100064	1714898	11.6938	11.7839	11.9324			
R17	100079	1714899	17.1972	17.1230	17.1107			
R18	100089	1714900	12.9547	13.0588	13.0375			

Apply HCA and PCA to the external files – continued

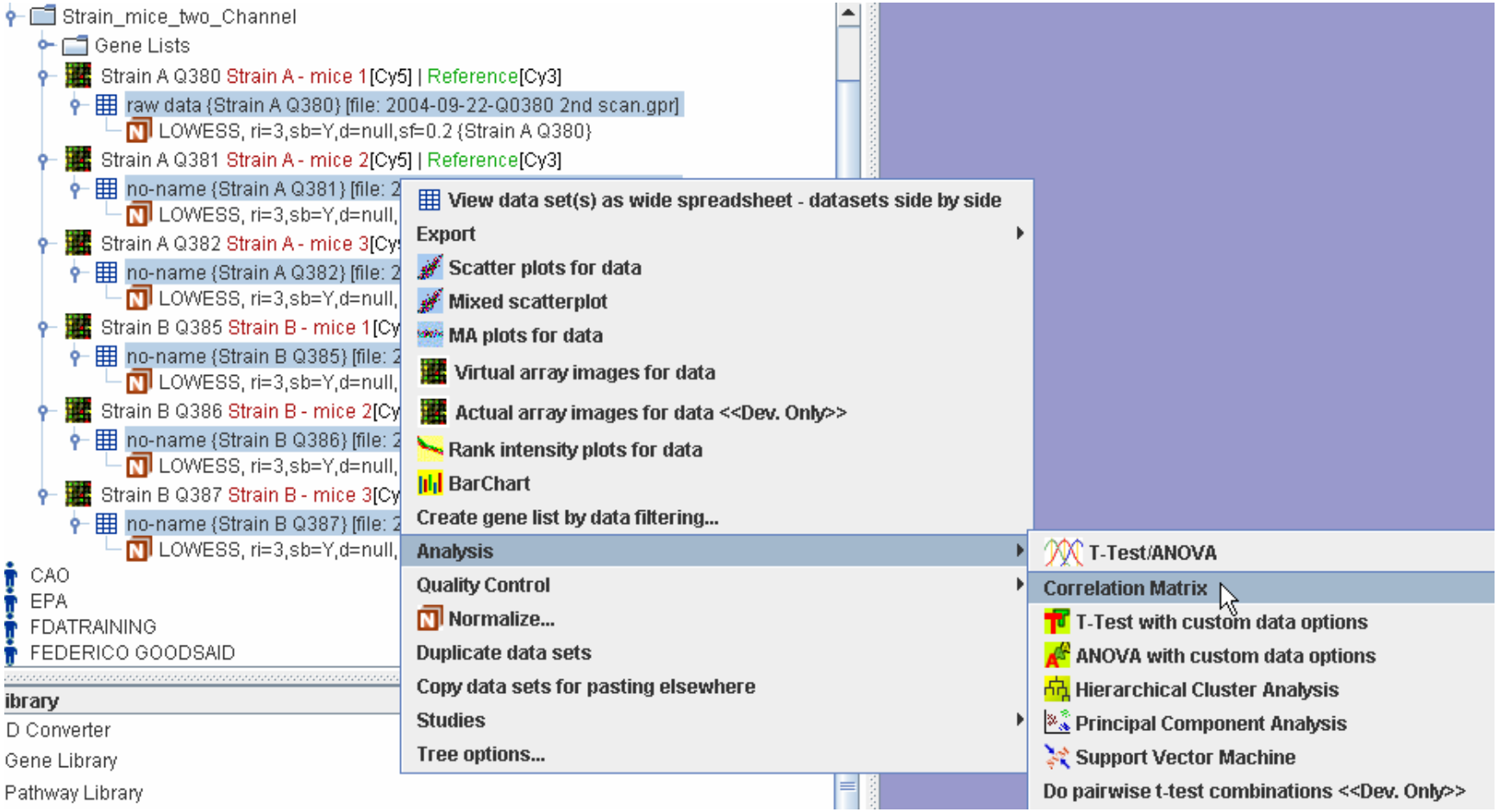


The procedure for applying PCA to the external files is same as applying HCA.

Correlation Matrix

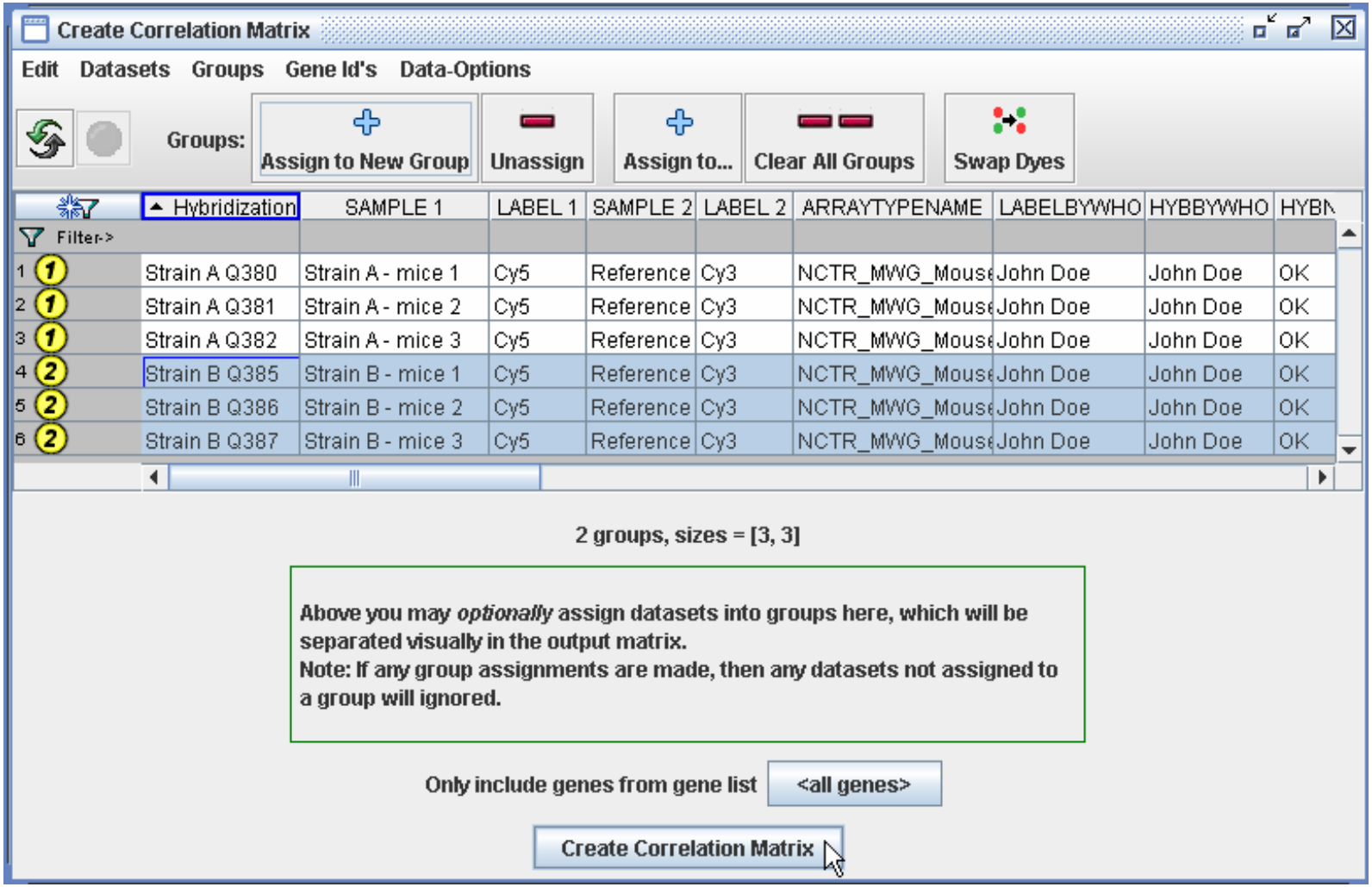
The correlation matrix function shows the correlation between column i and column j of the original matrix. It is used to find out the correlation between two groups of data in ArrayTrack.

Select data set, right-click, choose Analysis -> Correlation Matrix



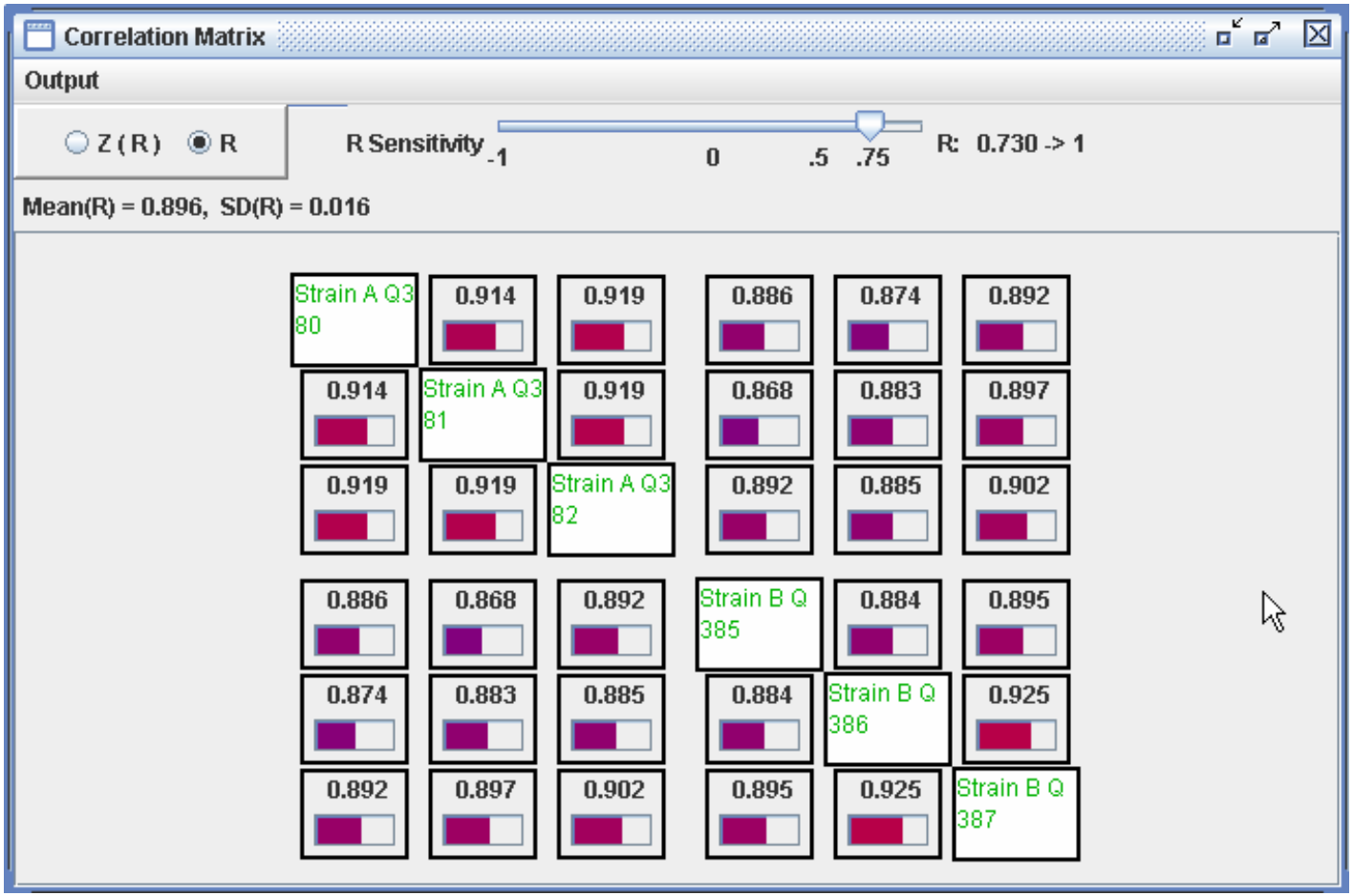
Correlation Matrix - continued

Assign dataset to groups, click “Create Correlation Matrix” button.



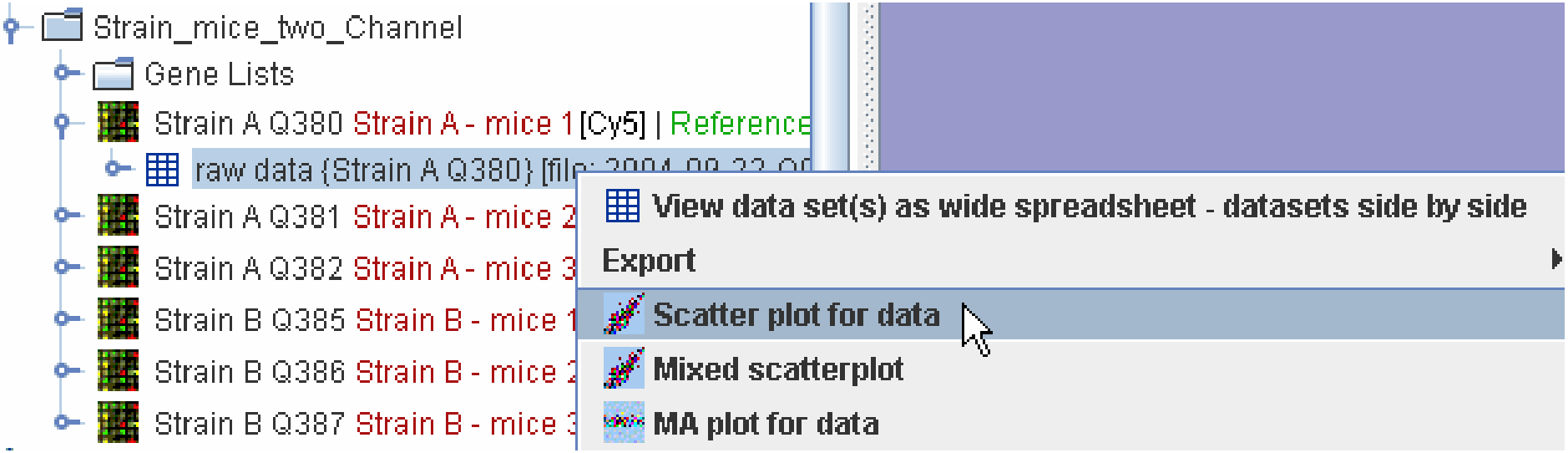
Correlation Matrix - continued

The diagonal elements of the correlation matrix will be 1 since they are the correlation of a column with itself. The correlation matrix is also symmetric since the correlation of column i with column j is the same as the correlation of column j with column i.

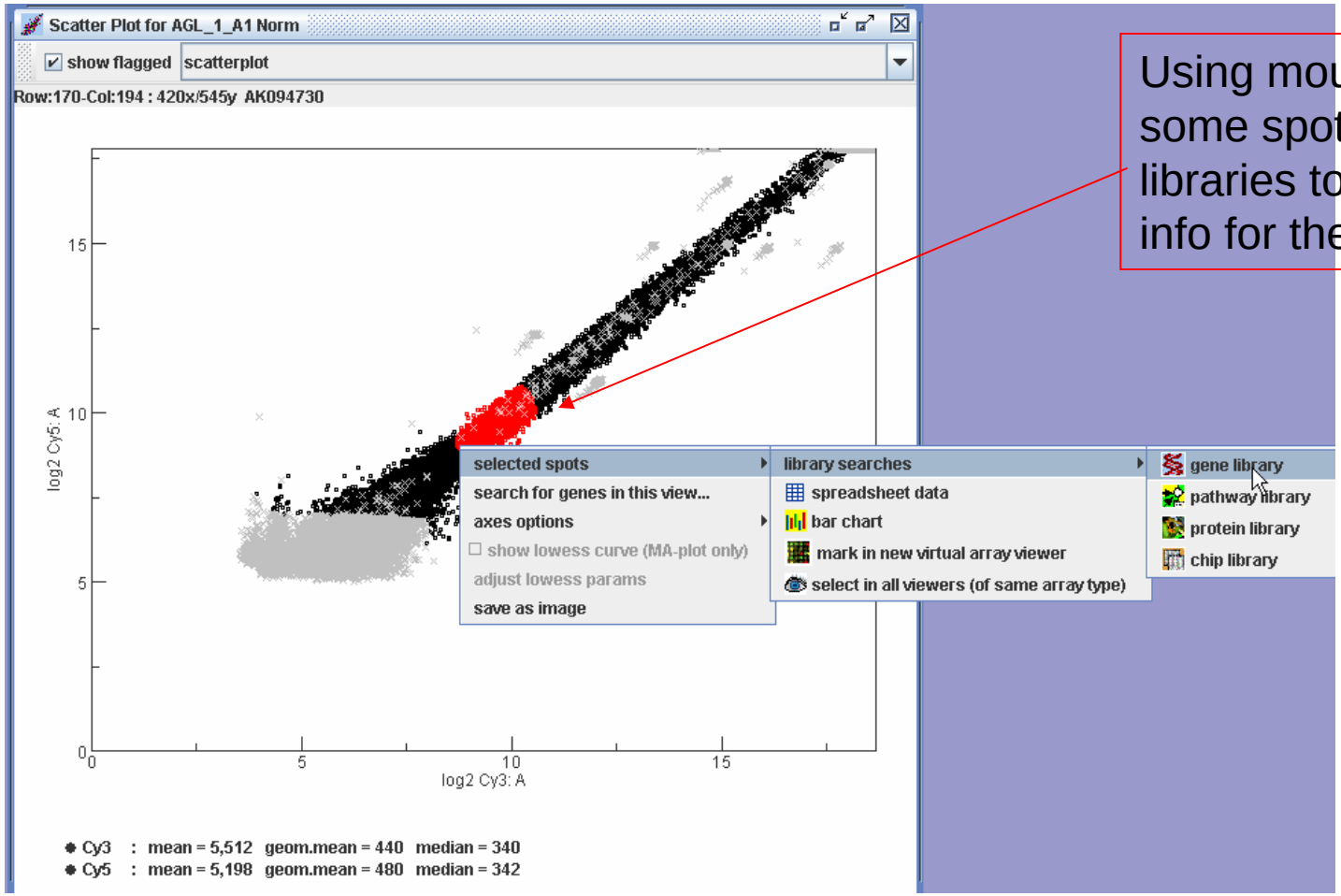


Scatter Plot

Scatter plot shows the fluorescence intensity data of the Cy3 channel versus those of the Cy5 channel for the same array.
Select data, right-click, choose “Scatter plot for data”.



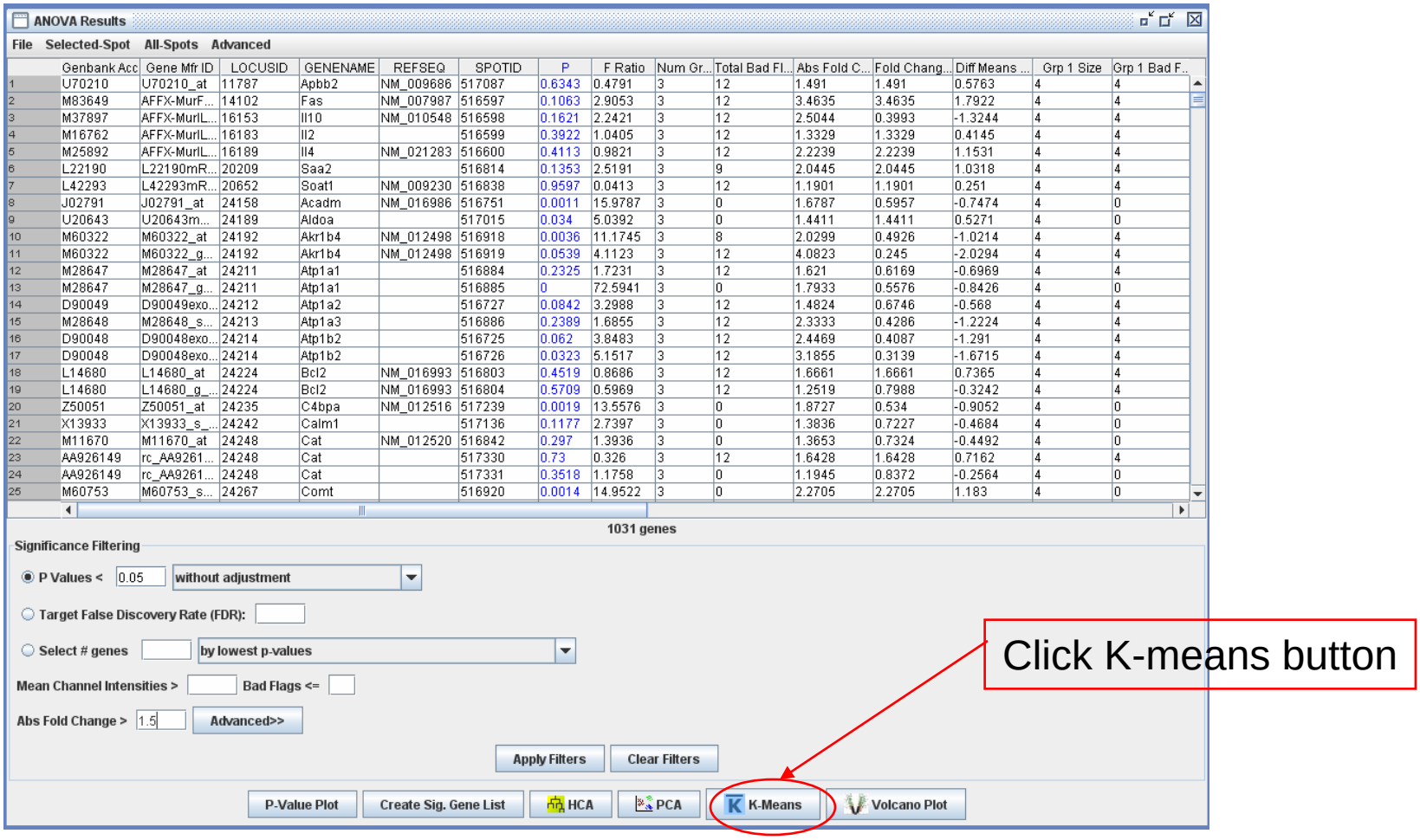
Scatter Plot - continued



Using mouse to circle some spots and approach libraries to get genome info for these spots

K-means

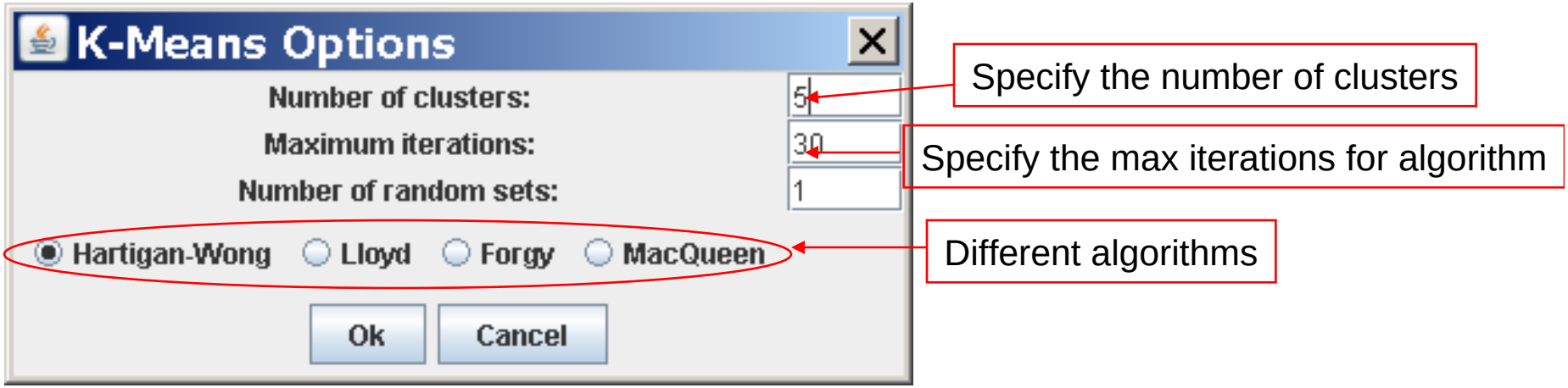
K-means is an algorithm to cluster objects into k clusters based on gene attributes, and calculate the centroid or mean point for each clusters. Users can access K-means from T-test/ANOVA result.



	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	P	F Ratio	Num Gr...	Total Bad Fl...	Abs Fold C...	Fold Chang...	Diff Means ...	Grp 1 Size	Grp 1 Bad F..
1	U70210	U70210_at	11787	Apbb2	NM_009686	517087	0.6343	0.4791	3	12	1.491	1.491	0.5763	4	4
2	M83649	AFFX-MurF...	14102	Fas	NM_007987	516597	0.1063	2.9053	3	12	3.4635	3.4635	1.7922	4	4
3	M37897	AFFX-MurIL...	16153	Il10	NM_010548	516598	0.1621	2.2421	3	12	2.5044	0.3993	-1.3244	4	4
4	M16762	AFFX-MurIL...	16183	Il2		516599	0.3922	1.0405	3	12	1.3329	1.3329	0.4145	4	4
5	M25892	AFFX-MurIL...	16189	Il4	NM_021283	516600	0.4113	0.9821	3	12	2.2239	2.2239	1.1531	4	4
6	L22190	L22190mR...	20209	Saa2		516814	0.1353	2.5191	3	9	2.0445	2.0445	1.0318	4	4
7	L42293	L42293mR...	20652	Soat1	NM_009230	516838	0.9597	0.0413	3	12	1.1901	1.1901	0.251	4	4
8	J02791	J02791_at	24158	Acadm	NM_016986	516751	0.0011	15.9787	3	0	1.6787	0.5957	-0.7474	4	0
9	U20643	U20643m...	24189	Aldoa		517015	0.034	5.0392	3	0	1.4411	1.4411	0.5271	4	0
10	M60322	M60322_at	24192	Akr1b4	NM_012498	516918	0.0036	11.1745	3	8	2.0299	0.4926	-1.0214	4	4
11	M60322	M60322_g...	24192	Akr1b4	NM_012498	516919	0.0539	4.1123	3	12	4.0823	0.245	-2.0294	4	4
12	M28647	M28647_at	24211	Atp1a1		516884	0.2325	1.7231	3	12	1.621	0.6169	-0.6969	4	4
13	M28647	M28647_g...	24211	Atp1a1		516885	0	72.5941	3	0	1.7933	0.5576	-0.8426	4	0
14	D90049	D90049exo...	24212	Atp1a2		516727	0.0842	3.2988	3	12	1.4824	0.6746	-0.568	4	4
15	M28648	M28648_s...	24213	Atp1a3		516886	0.2389	1.6855	3	12	2.3333	0.4286	-1.2224	4	4
16	D90048	D90048exo...	24214	Atp1b2		516725	0.062	3.8483	3	12	2.4469	0.4087	-1.291	4	4
17	D90048	D90048exo...	24214	Atp1b2		516726	0.0323	5.1517	3	12	3.1855	0.3139	-1.6715	4	4
18	L14680	L14680_at	24224	Bcl2	NM_016993	516803	0.4519	0.8686	3	12	1.6661	1.6661	0.7365	4	4
19	L14680	L14680_g_...	24224	Bcl2	NM_016993	516804	0.5709	0.5969	3	12	1.2519	0.7988	-0.3242	4	4
20	Z50051	Z50051_at	24235	C4bpa	NM_012516	517239	0.0019	13.5576	3	0	1.8727	0.534	-0.9052	4	0
21	X13933	X13933_s_...	24242	Calm1		517136	0.1177	2.7397	3	0	1.3836	0.7227	-0.4684	4	0
22	M11670	M11670_at	24248	Cat	NM_012520	516842	0.297	1.3936	3	0	1.3653	0.7324	-0.4492	4	0
23	AA926149	rc_AA9261...	24248	Cat		517330	0.73	0.326	3	12	1.6428	1.6428	0.7162	4	4
24	AA926149	rc_AA9261...	24248	Cat		517331	0.3518	1.1758	3	0	1.1945	0.8372	-0.2564	4	0
25	M60753	M60753_s...	24267	Comt		516920	0.0014	14.9522	3	0	2.2705	2.2705	1.183	4	0

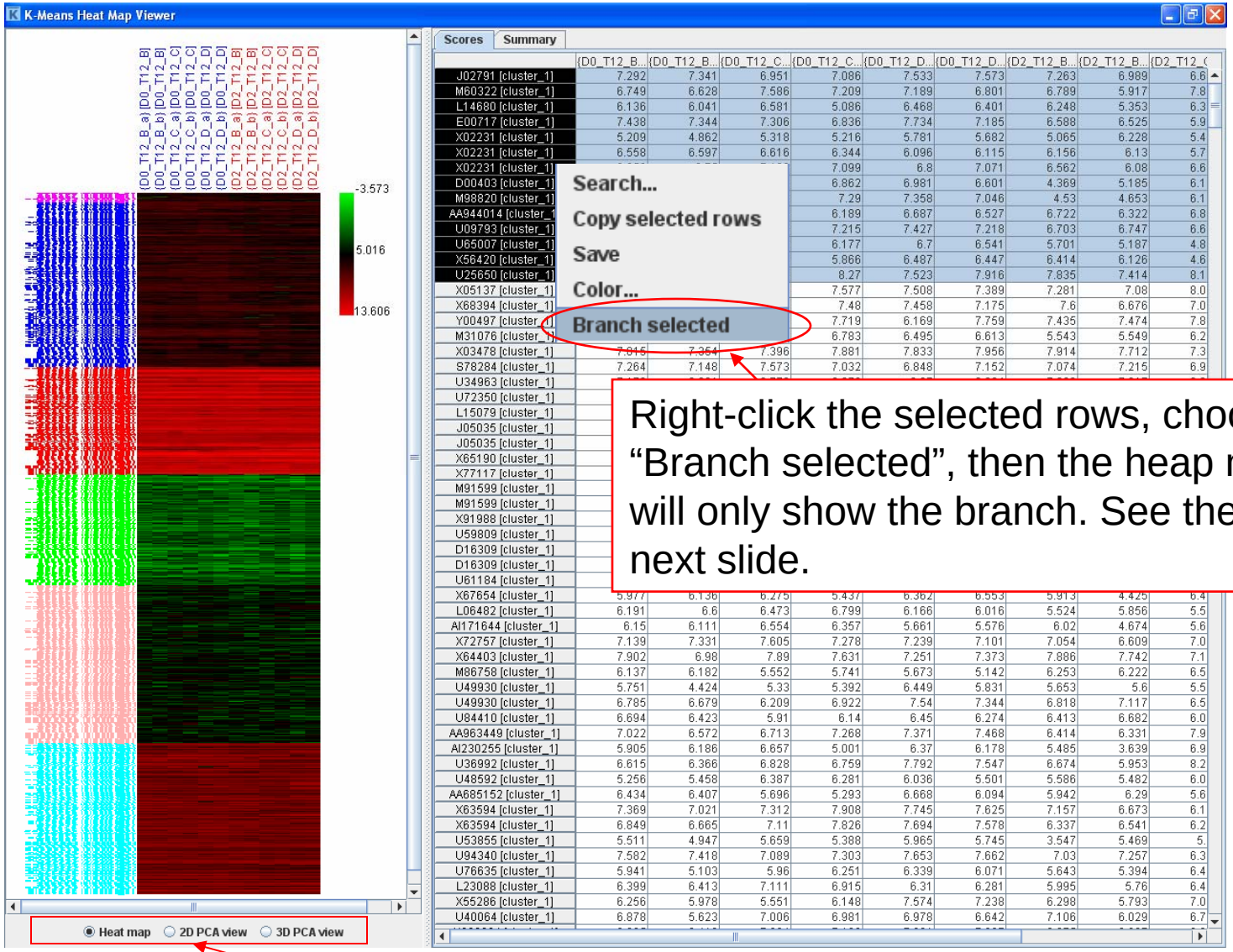
K-means (continued)

The following window will pop up with some default values



Click OK button

K-means (continued)



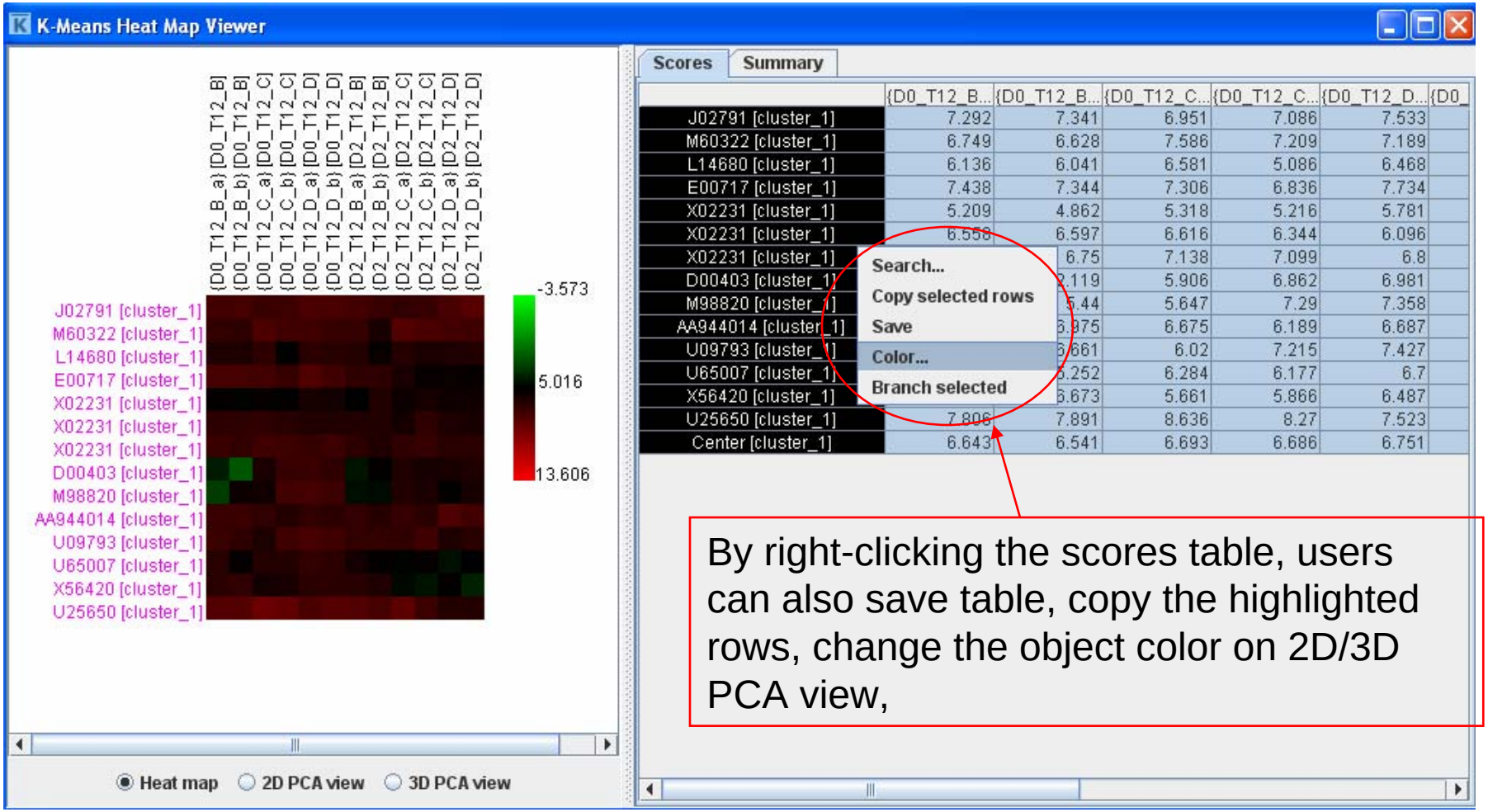
Clusters are highlighted in different colors

Right-click the selected rows, choose “Branch selected”, then the heap map will only show the branch. See the next slide.

Options for heat map, 2D PCA view or 3D PCA view

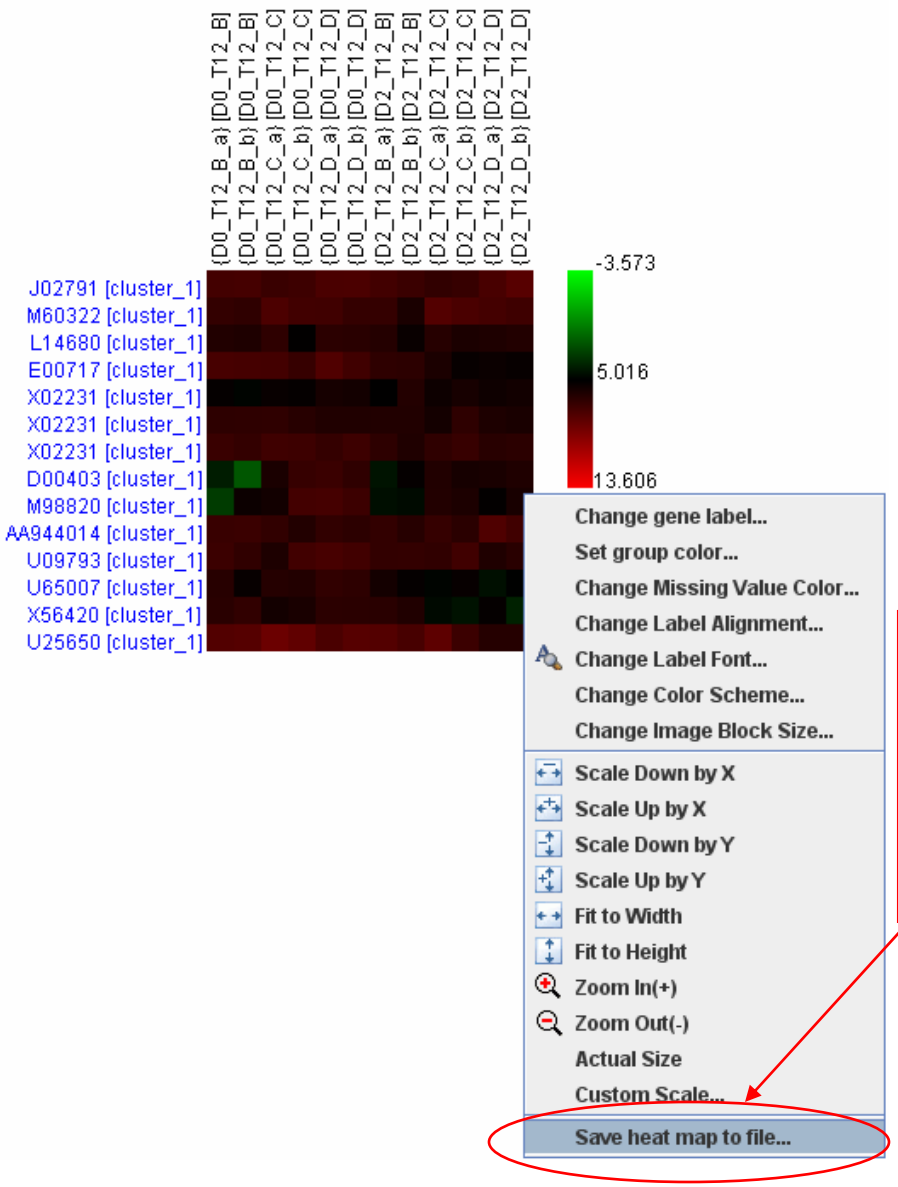
K-means (continued)

A branch of the previous heat map



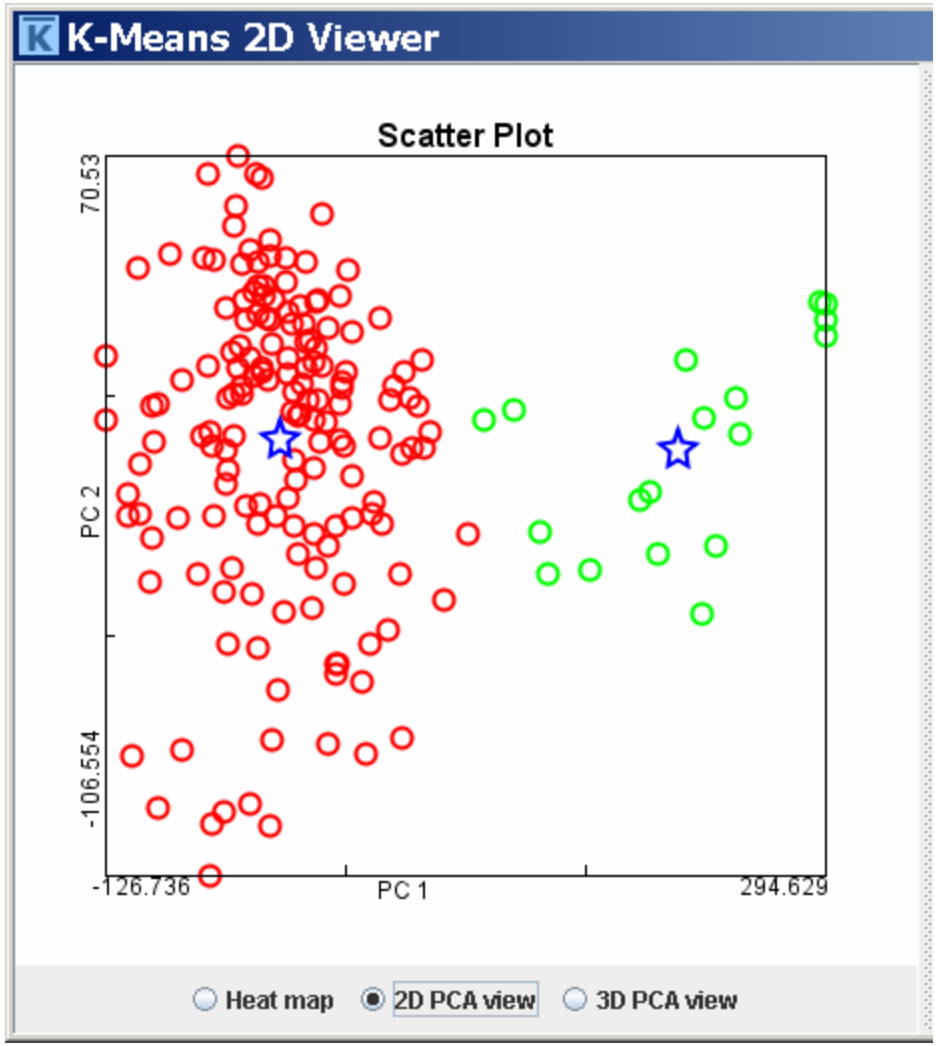
By right-clicking the scores table, users can also save table, copy the highlighted rows, change the object color on 2D/3D PCA view,

K-means (continued)



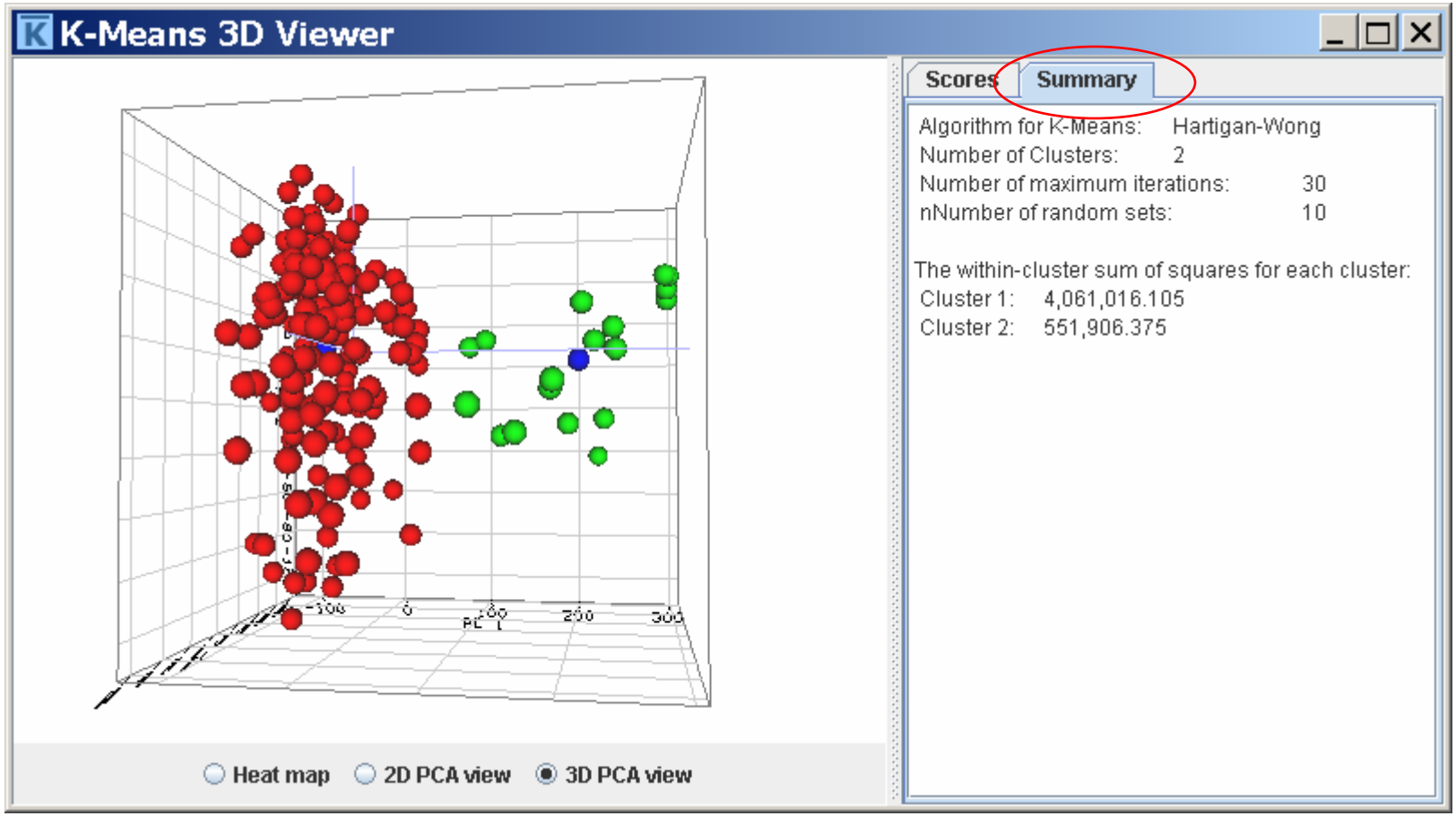
Right-click the heat map, user can choose “Save heat map to file…” to save the heat map. There are also many other options for users to change the looks of the heat map.

K-means (continued)



2D view of K-means. Each cluster is marked in different colors. The blue stars are centroids.

K-means (continued)

3D PCA view of K-means

Summary of K-means