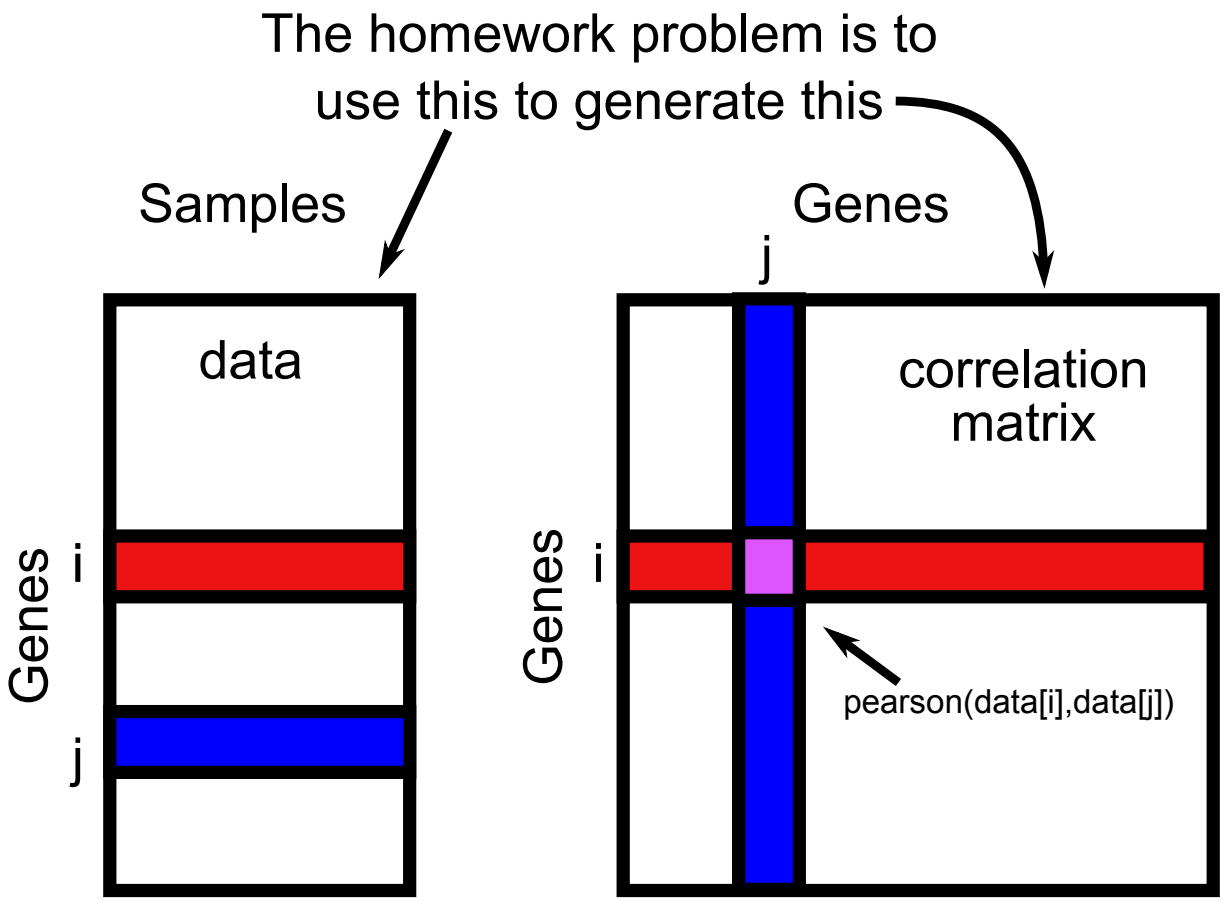
Practical Bioinformatics

Mark Voorhies

5/16/2019

What was Mark even asking for yesterday?



List tricks

Adding data to a list:

```
mylist = []
mylist.append(3)
mylist += [4,5,6]
```

List tricks

Adding data to a list:

```
mylist = []
mylist.append(3)
mylist += [4,5,6]
```

Lists of lists:

```
matrix = [[ 1, 2, 3, 4],
          [ 5, 6, 7, 8],
          [ 9,10,11,12]]
```

Pearson distances

Pearson similarity

$$s(x,y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2}\sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

Pearson distances

Pearson similarity

$$s(x,y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2}\sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

Pearson distance

$$d(x,y) = 1 - s(x,y)$$

Pearson distances

Pearson similarity

$$s(x,y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2}\sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

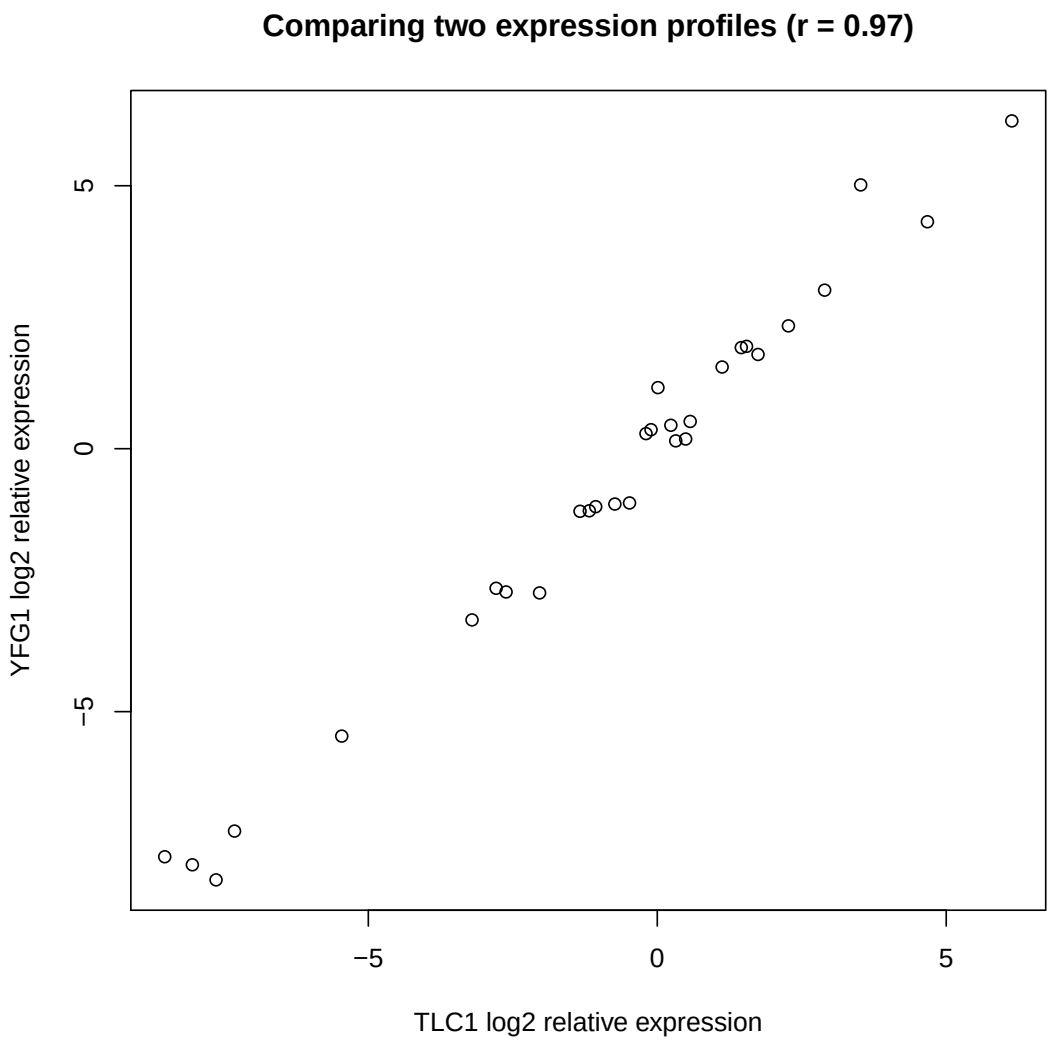
Pearson distance

$$d(x,y) = 1 - s(x,y)$$

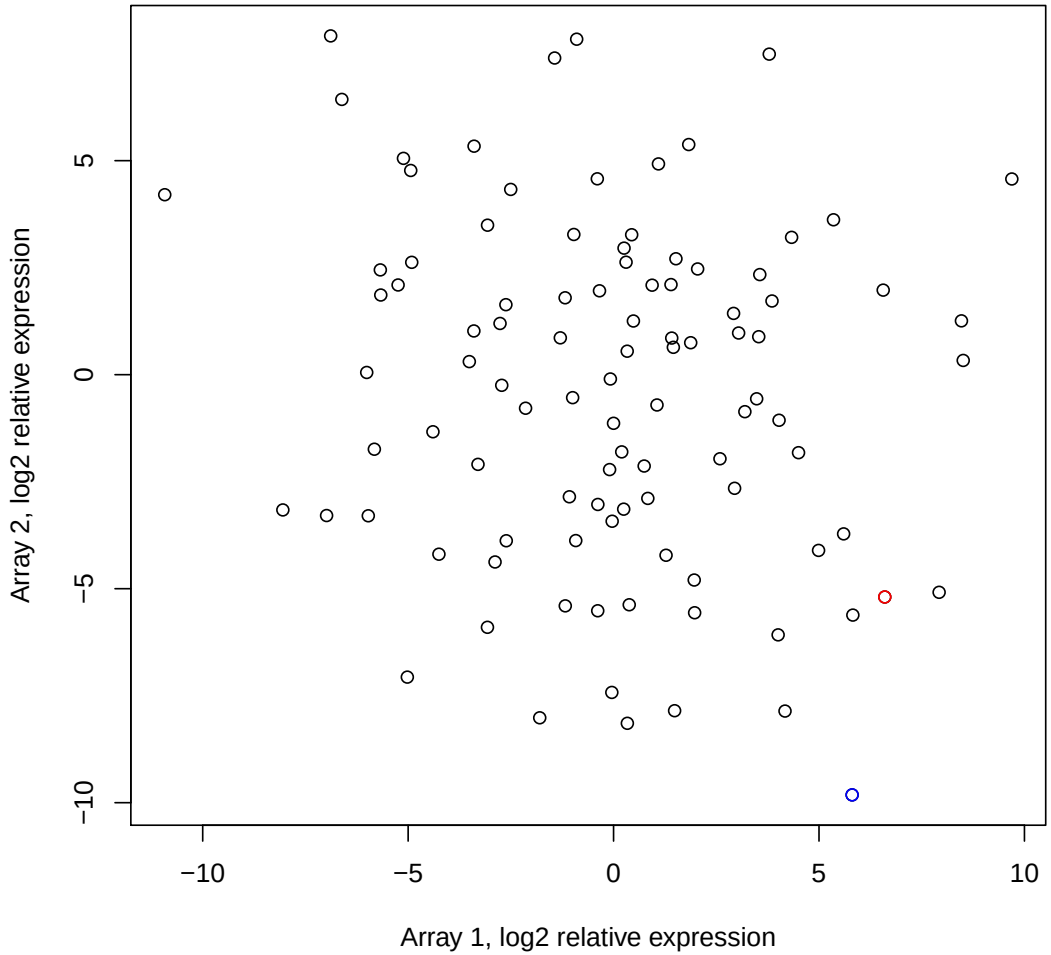
Euclidean distance

$$\frac{\sum_i^N (x_i - y_i)^2}{N}$$

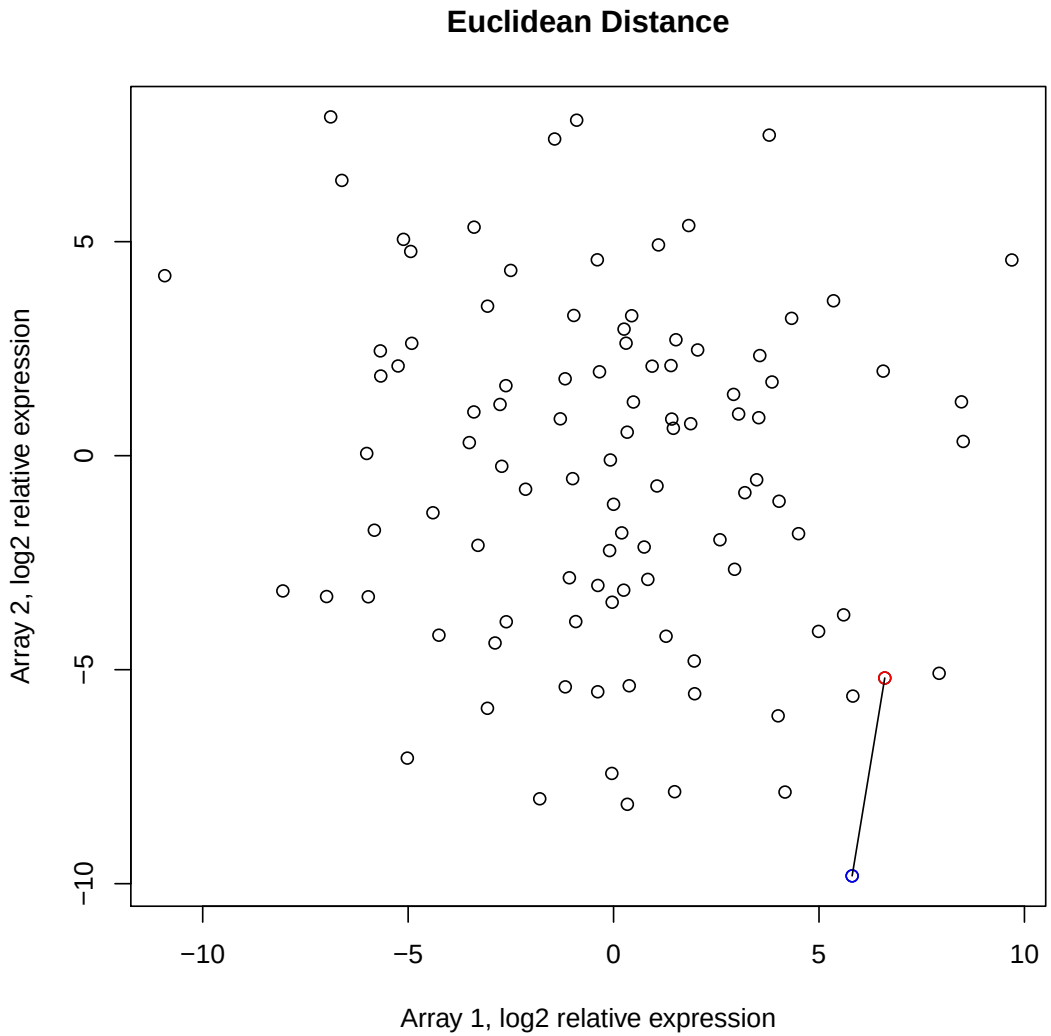
Comparing all measurements for two genes



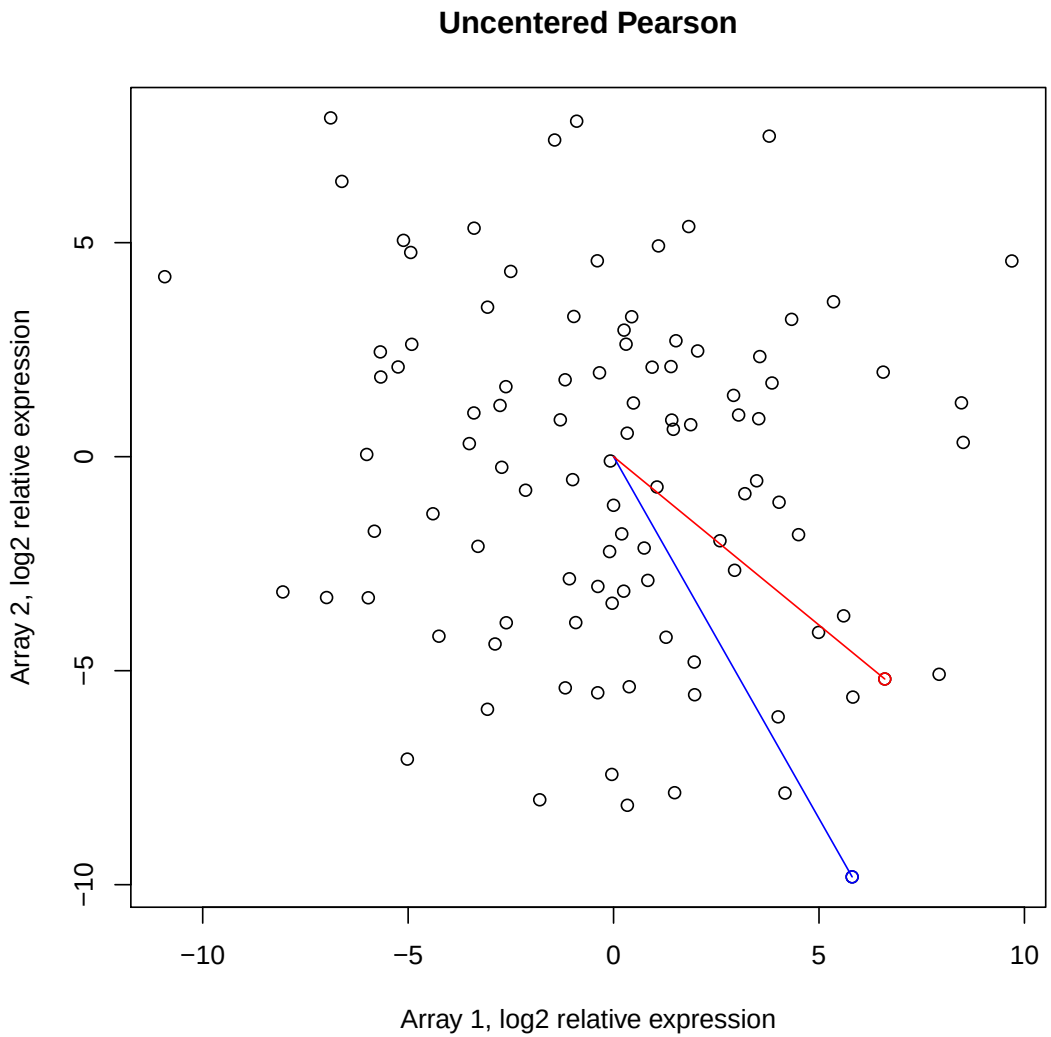
Comparing all genes for two measurements



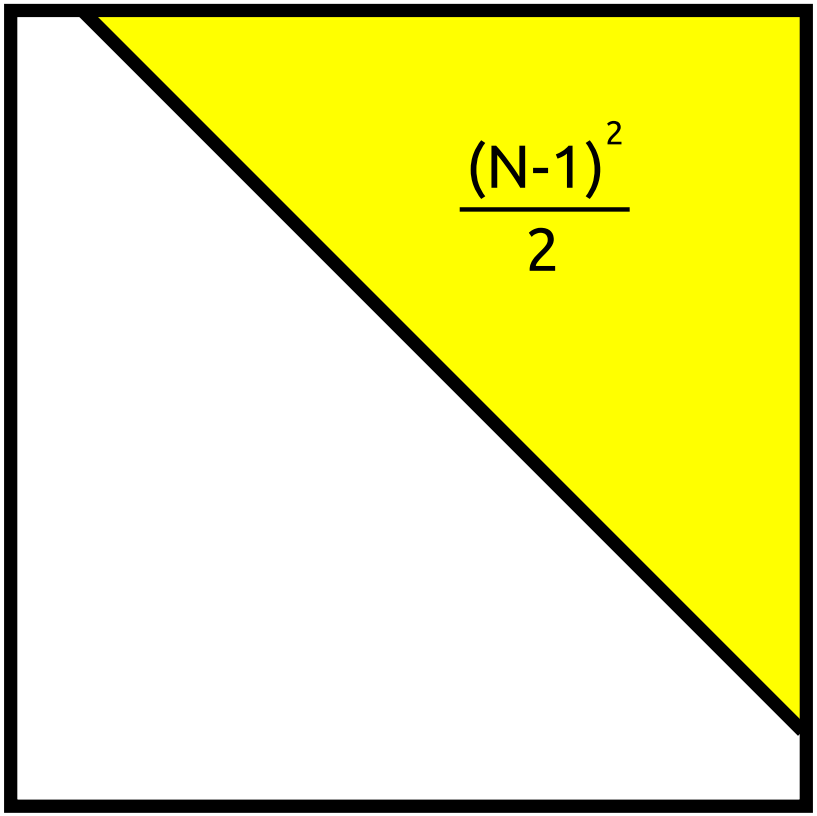
Comparing all genes for two measurements



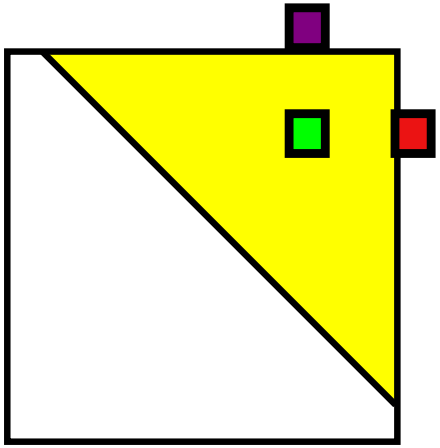
Comparing all genes for two measurements

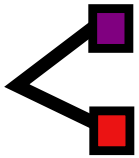


Measure all pairwise distances under distance metric

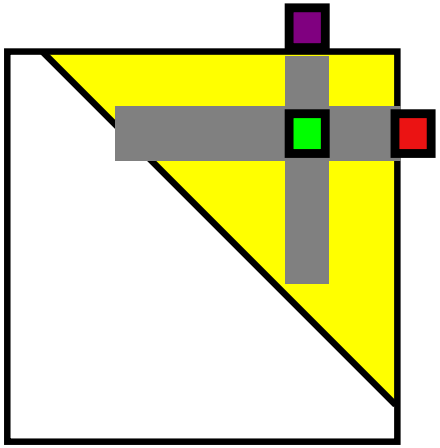


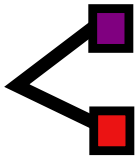
Hierarchical Clustering



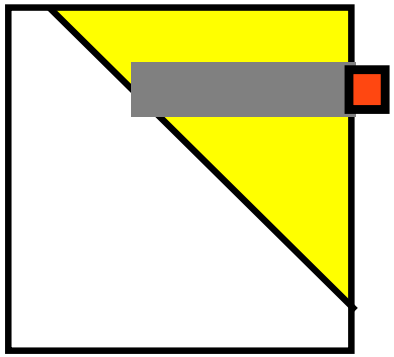


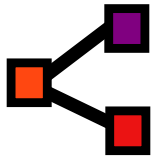
Hierarchical Clustering



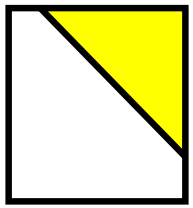


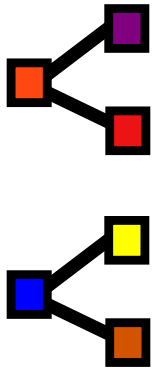
Hierarchical Clustering



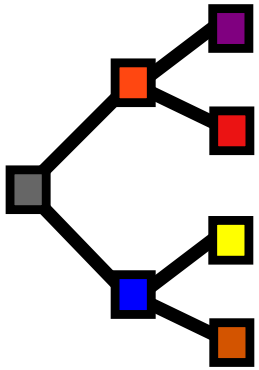


Hierarchical Clustering





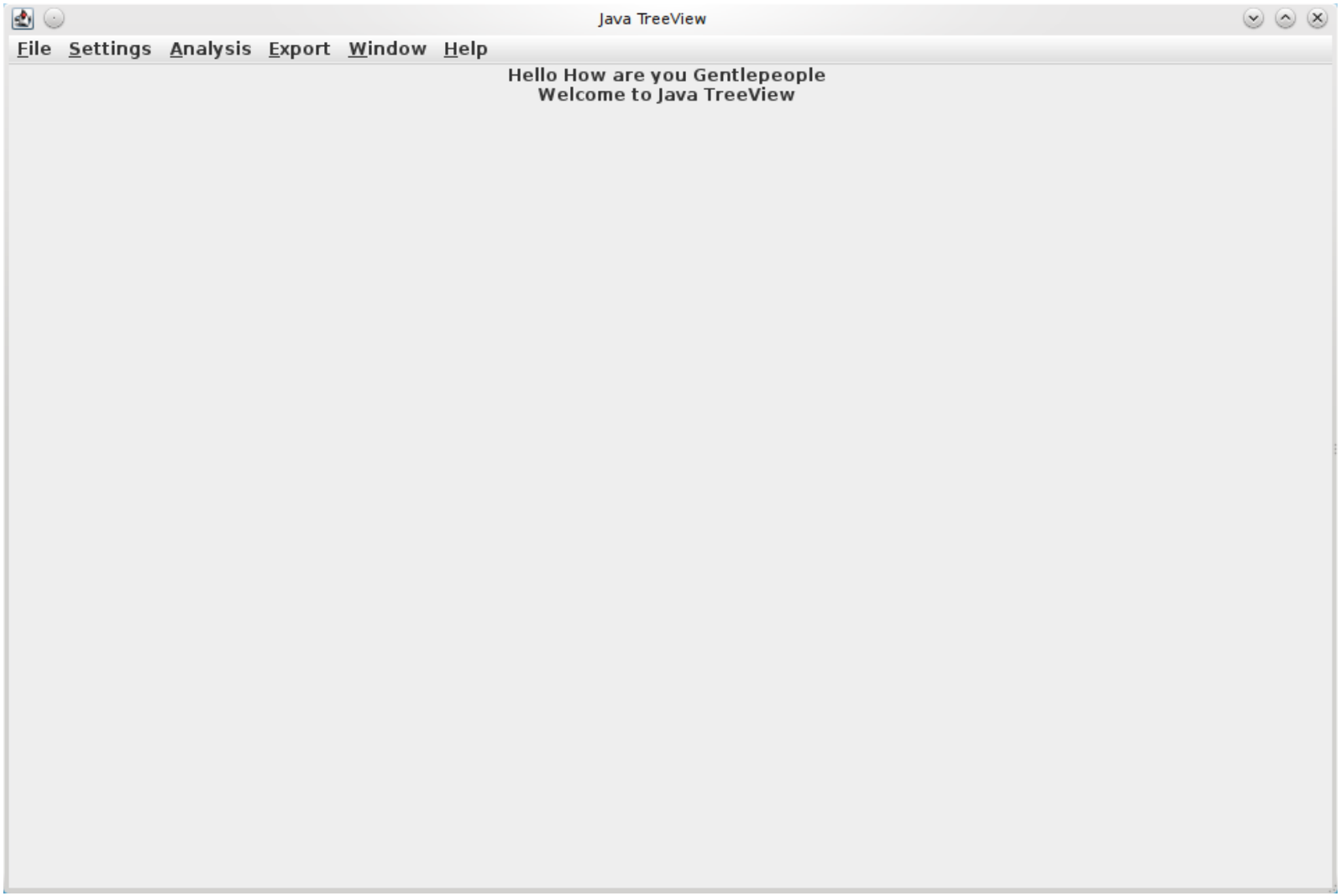
Hierarchical Clustering



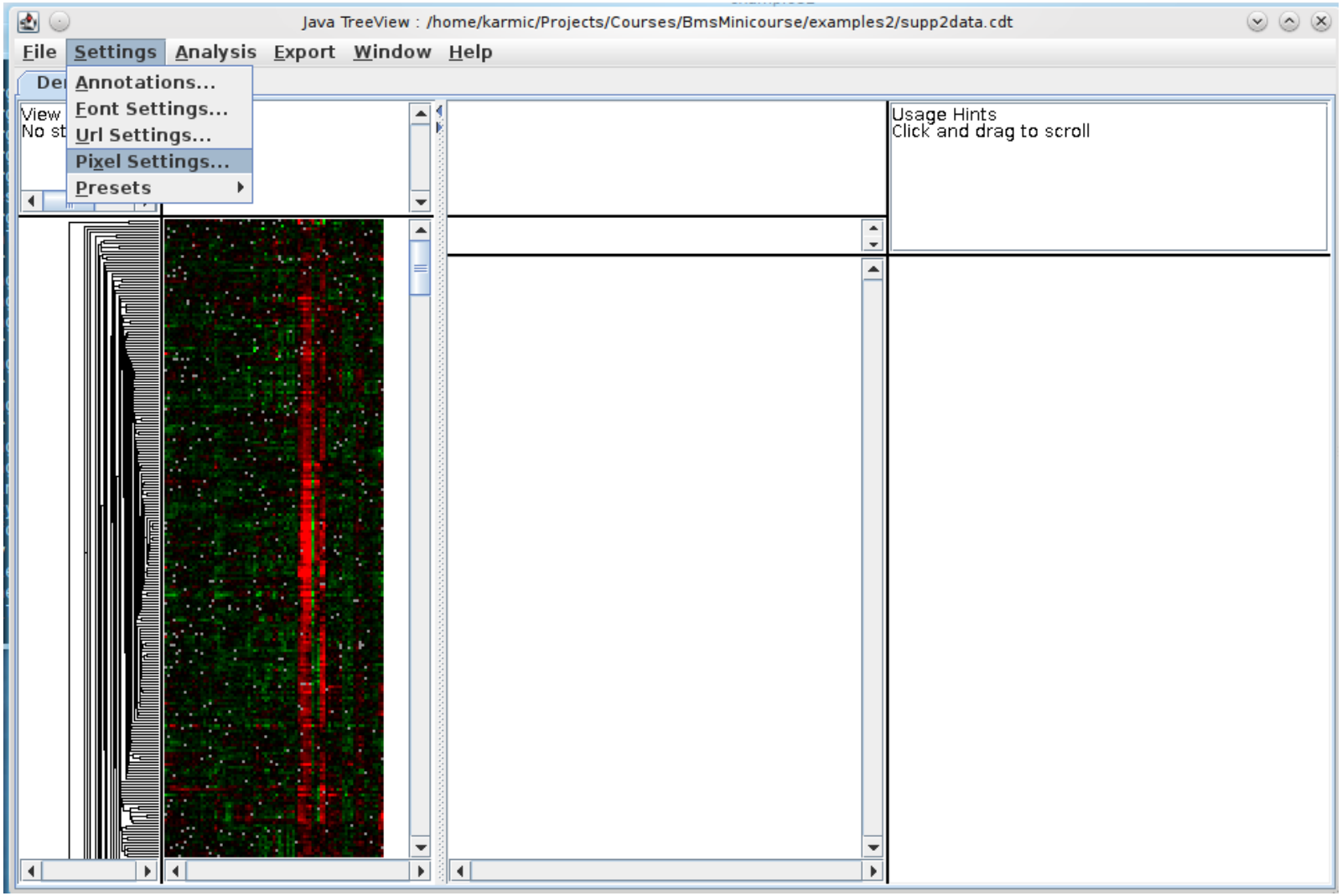
It's hard work at times, but you have to be realistic. If you have a large database with many variables and your goal is to get a good understanding of the interrelationships, then, unless you get lucky, this complex structure is bound to require some hard work to understand.

Bill Cleveland and Rick Becker
http://stat.bell-labs.com/project/trellis/interview.html

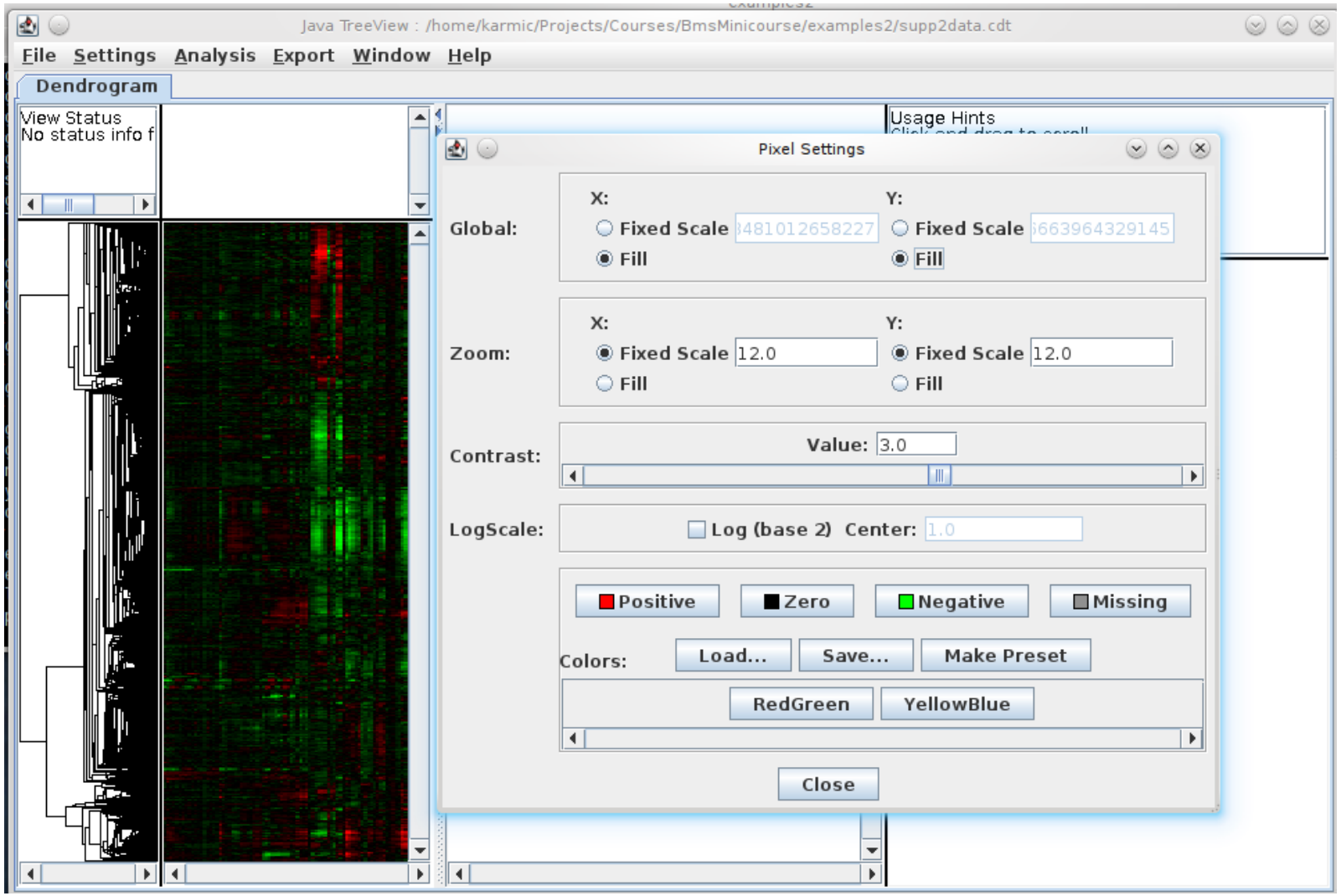
Using JavaTreeView



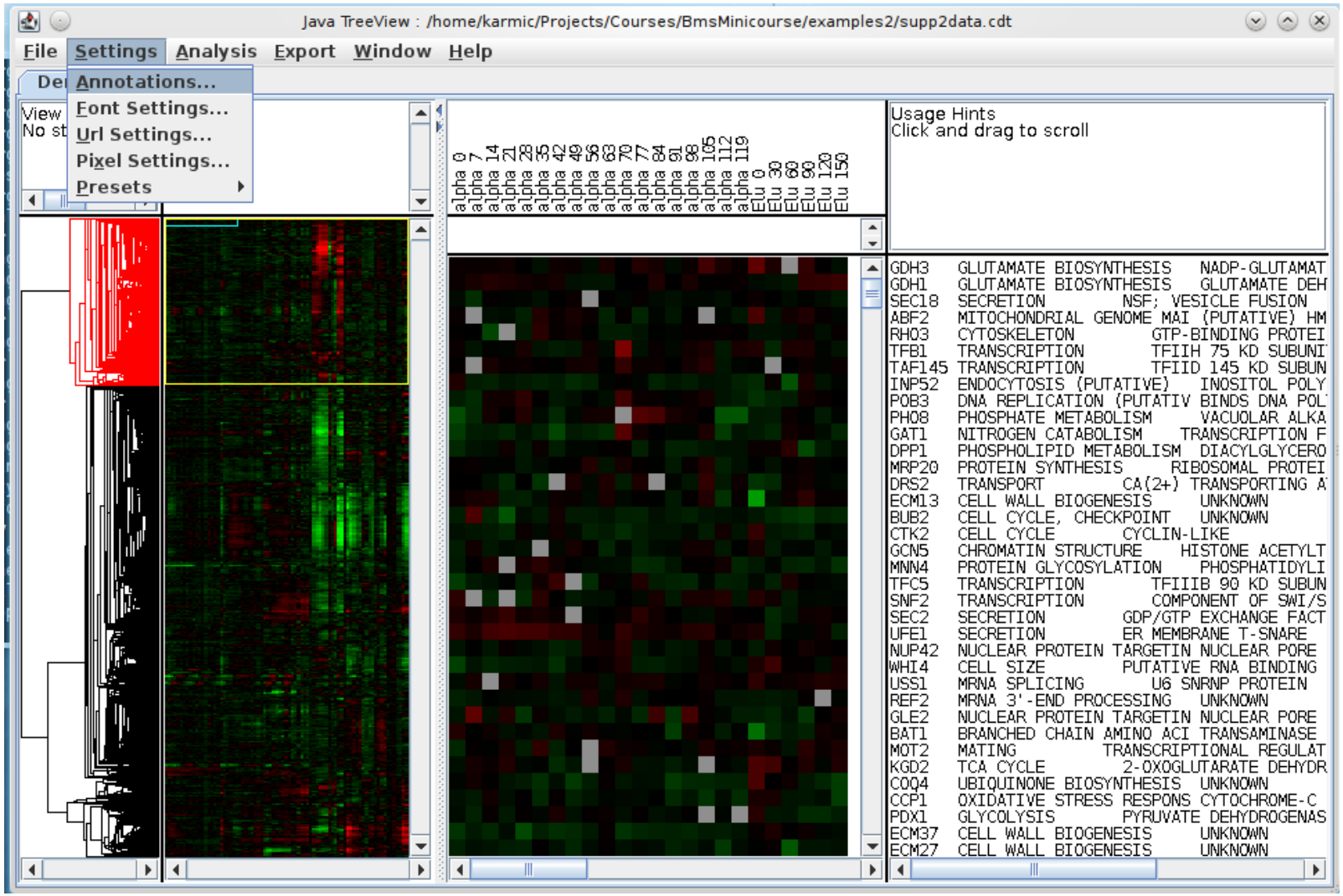
Adjust pixel settings for global view



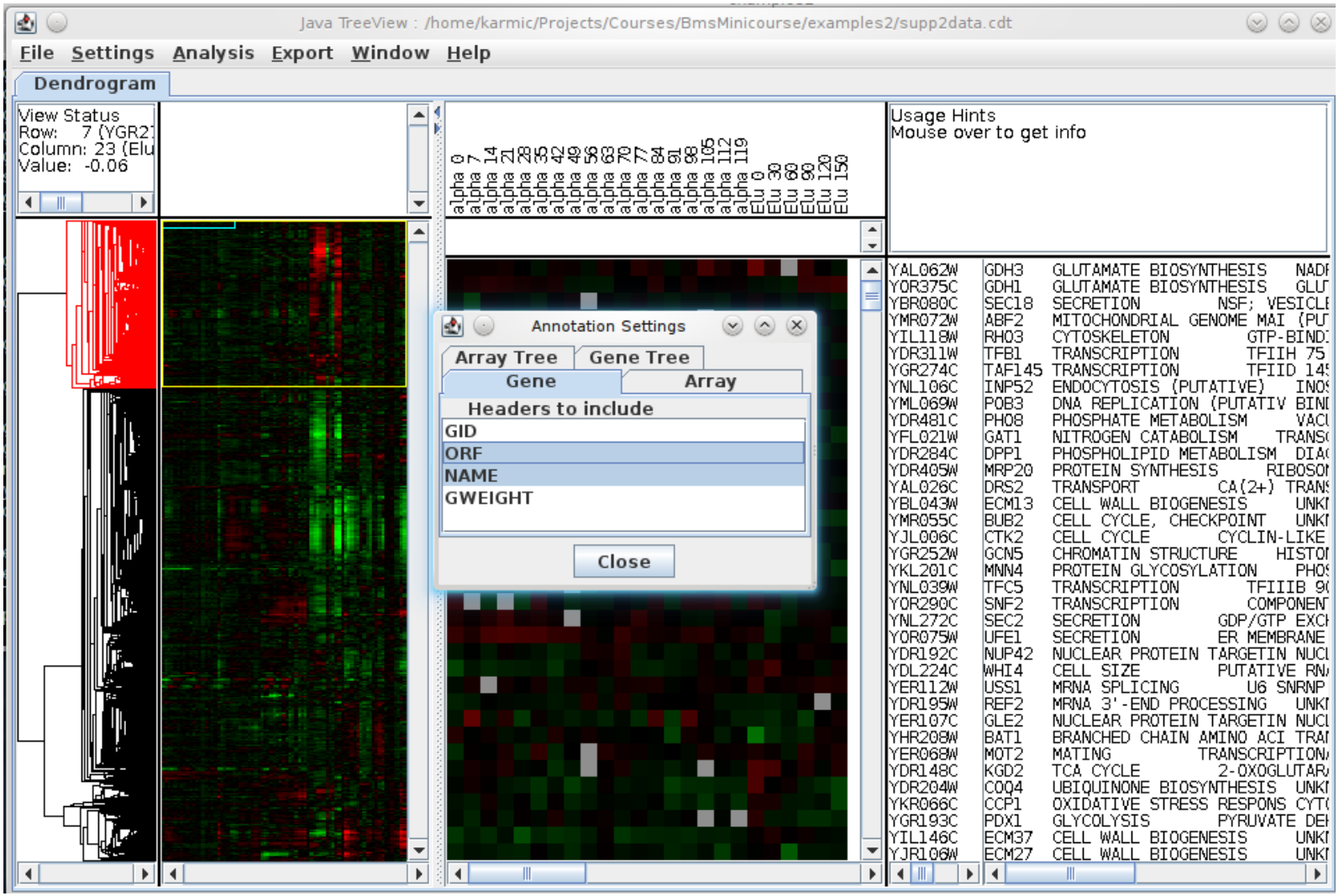
Adjust pixel settings for global view



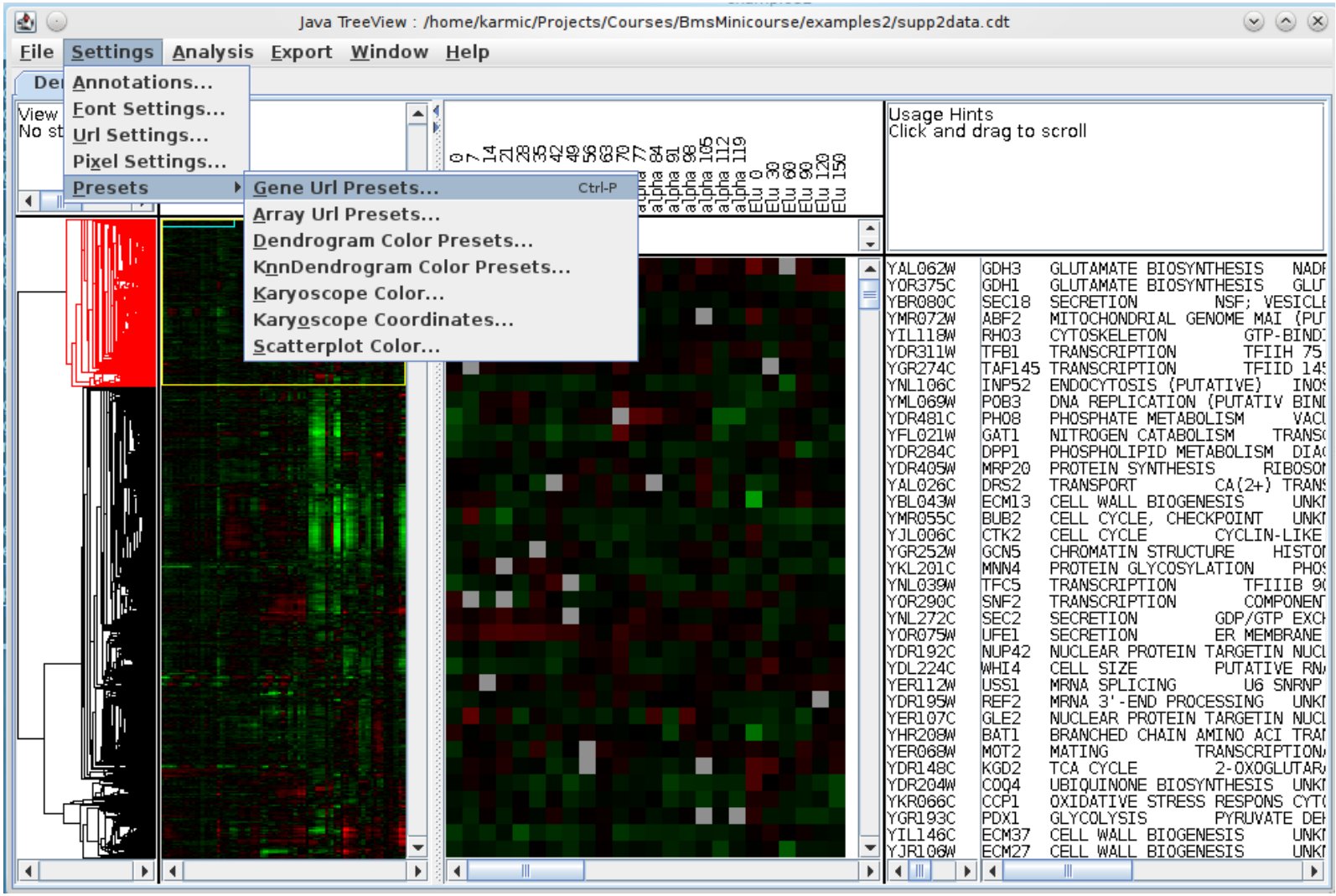
Select annotation columns



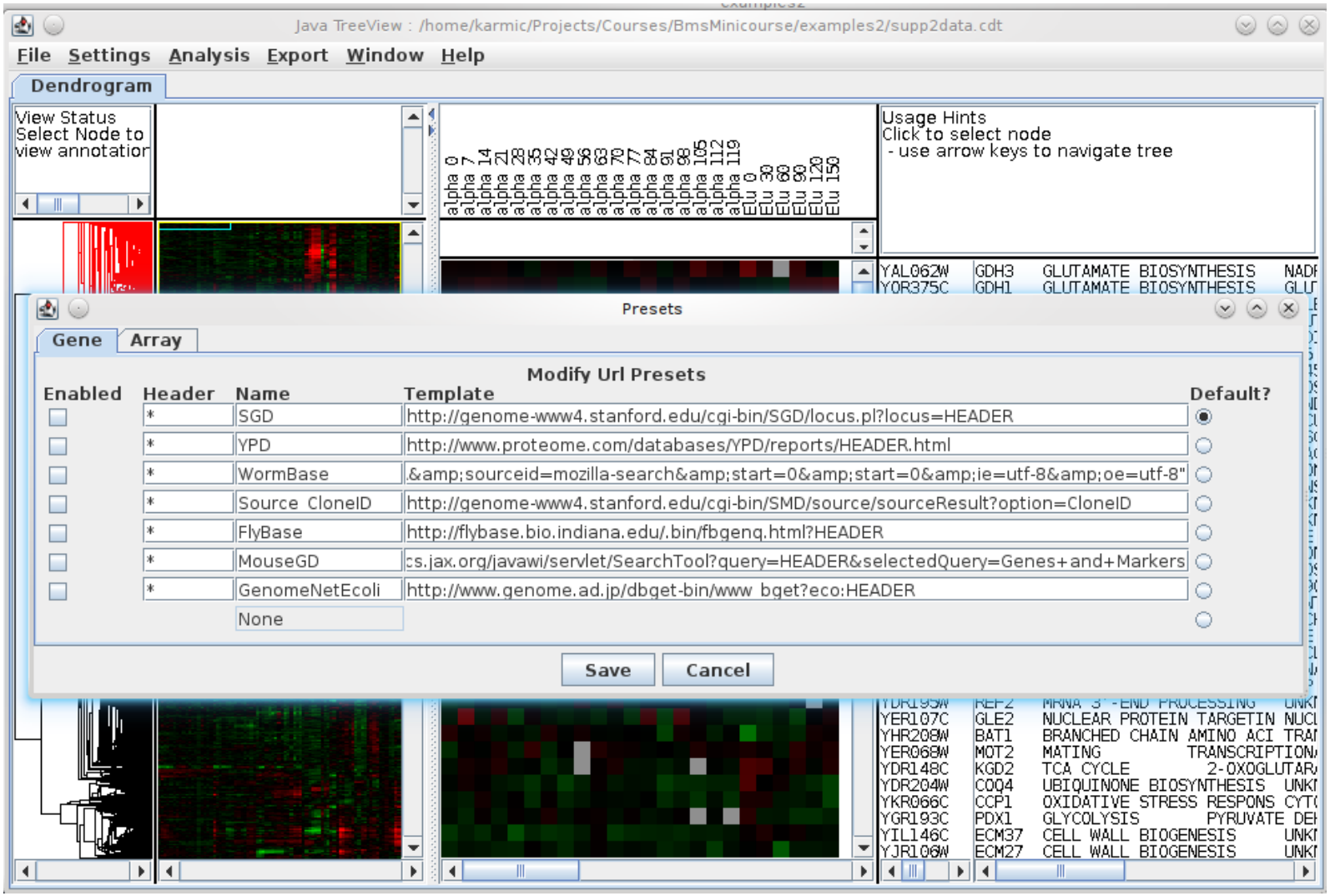
Select annotation columns



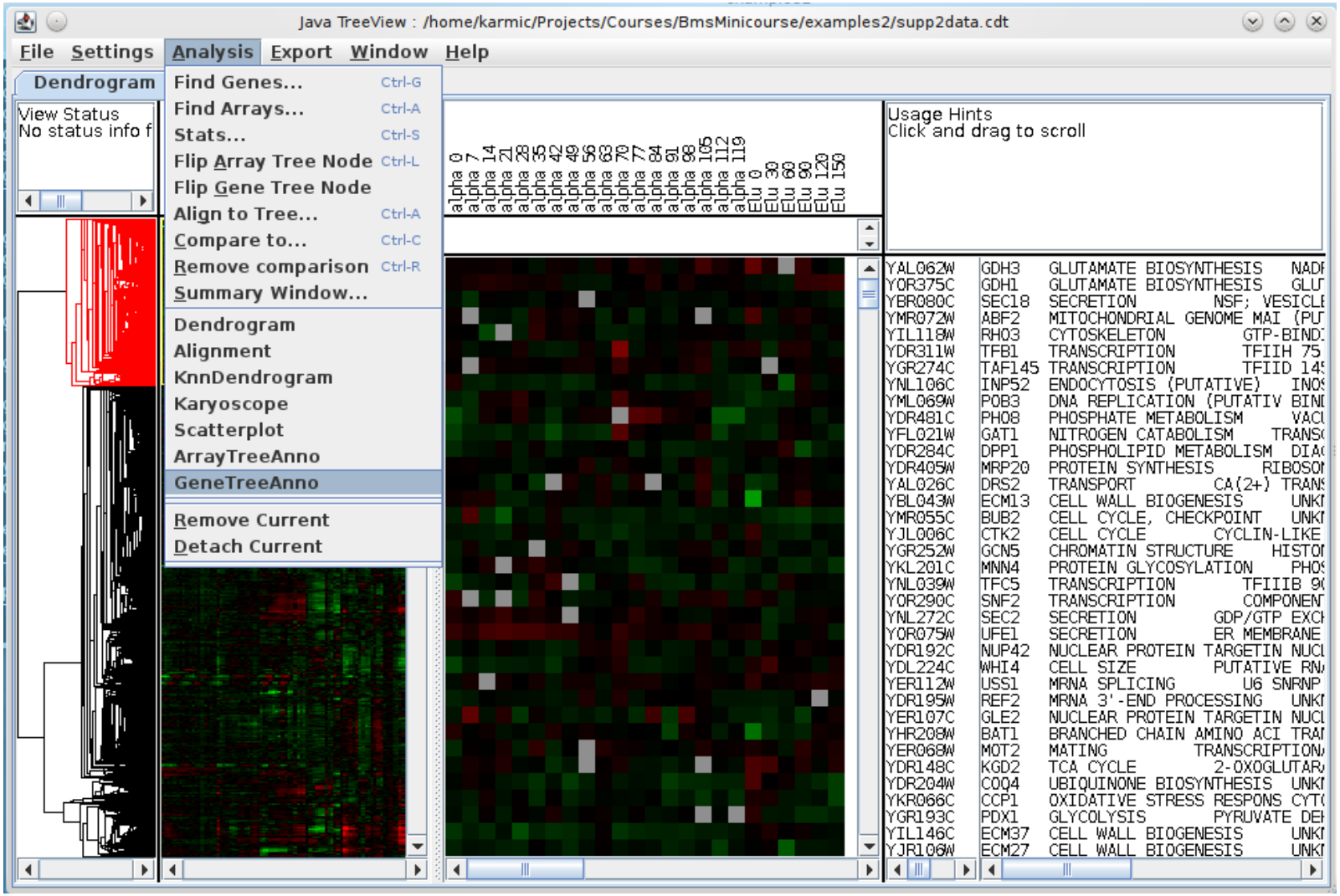
Select URL for gene annotations



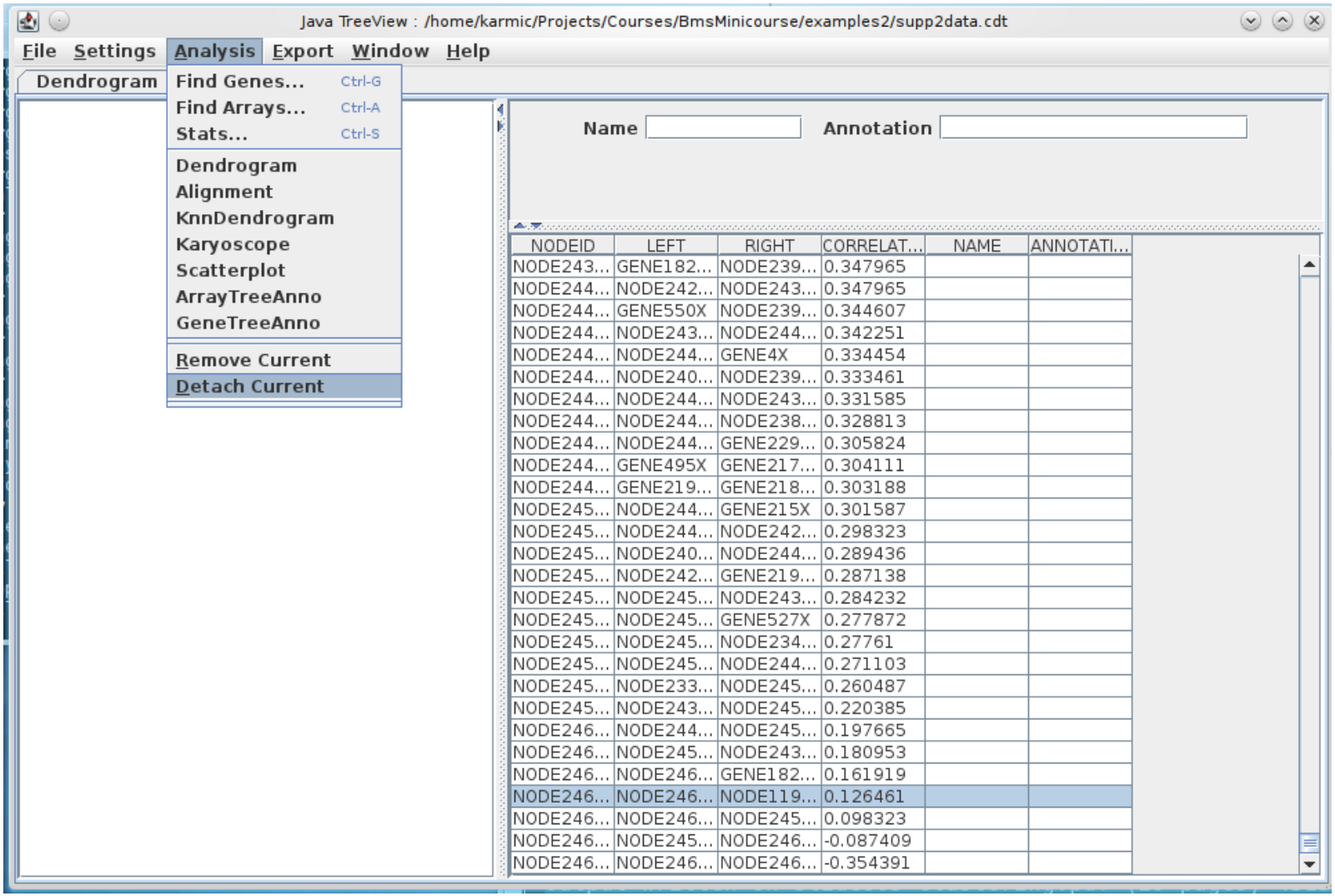
Select URL for gene annotations



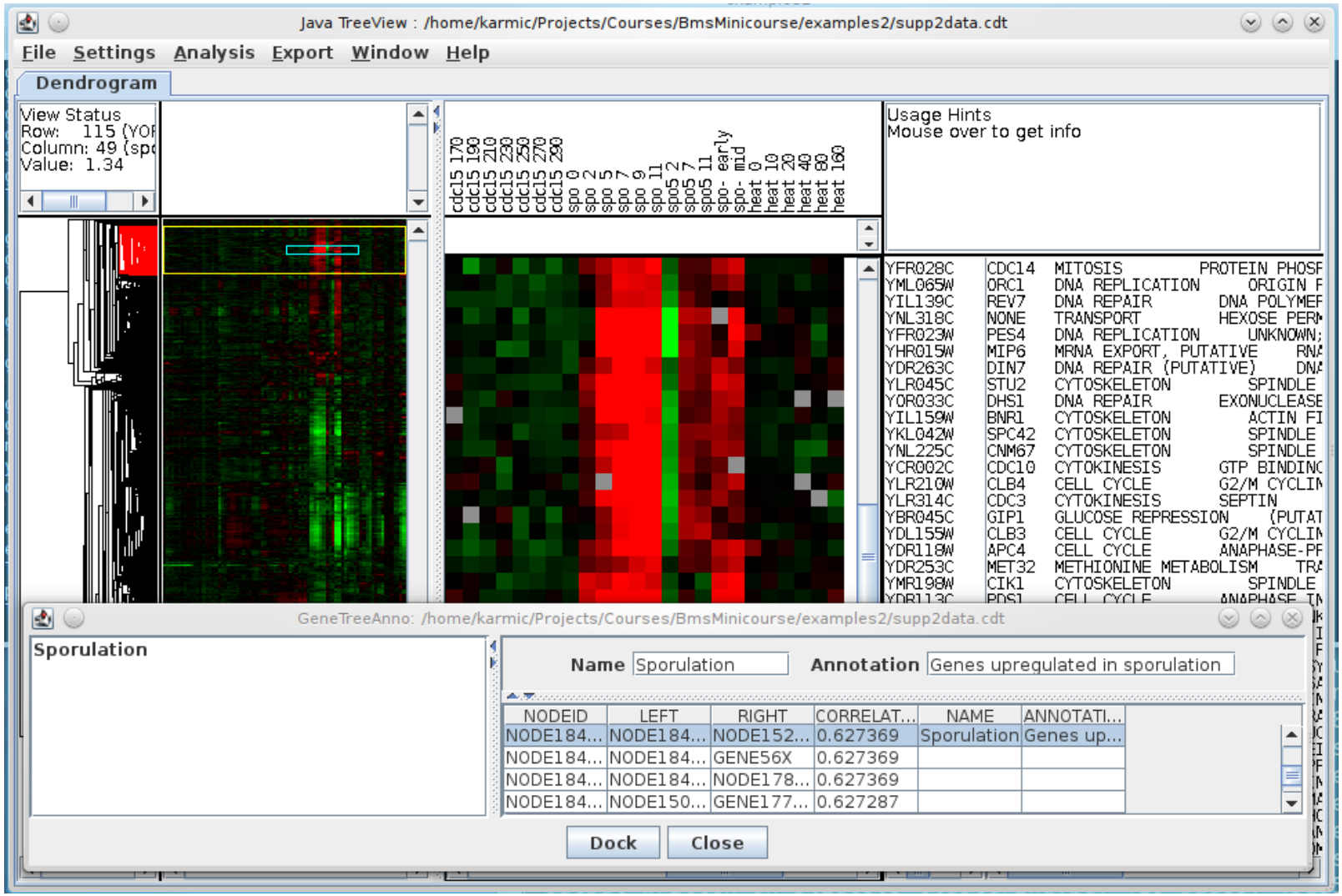
Activate and detach annotation window



Activate and detach annotation window



Activate and detach annotation window



Homework

1. Write functions to reproduce the shuffling controls in figure 3 of the Eisen paper (removing correlations among genes and/or samples).