CSCI 5465

Lab#8

Quincy Gu

12/09/2019

Part I. Understanding the gene expression study

**(1). What was the purpose of the study and what experiments did the authors do to answer the question they were interested in?**

The purpose of this study is to investigate the responses from the genomic expression patterns in the yeast Saccharomyces cerevisiae to the diverse environmental transitions.

The authors apply the DNA microarrays experiments to do the analysis in order to answer the question they were interested in.

**(2). What are some examples of different environments or stresses the authors subjected yeast cells to before measuring their gene expression?**

The examples of different environments or stresses the authors subjected yeast cells to before measuring their gene expression are listed in the following:

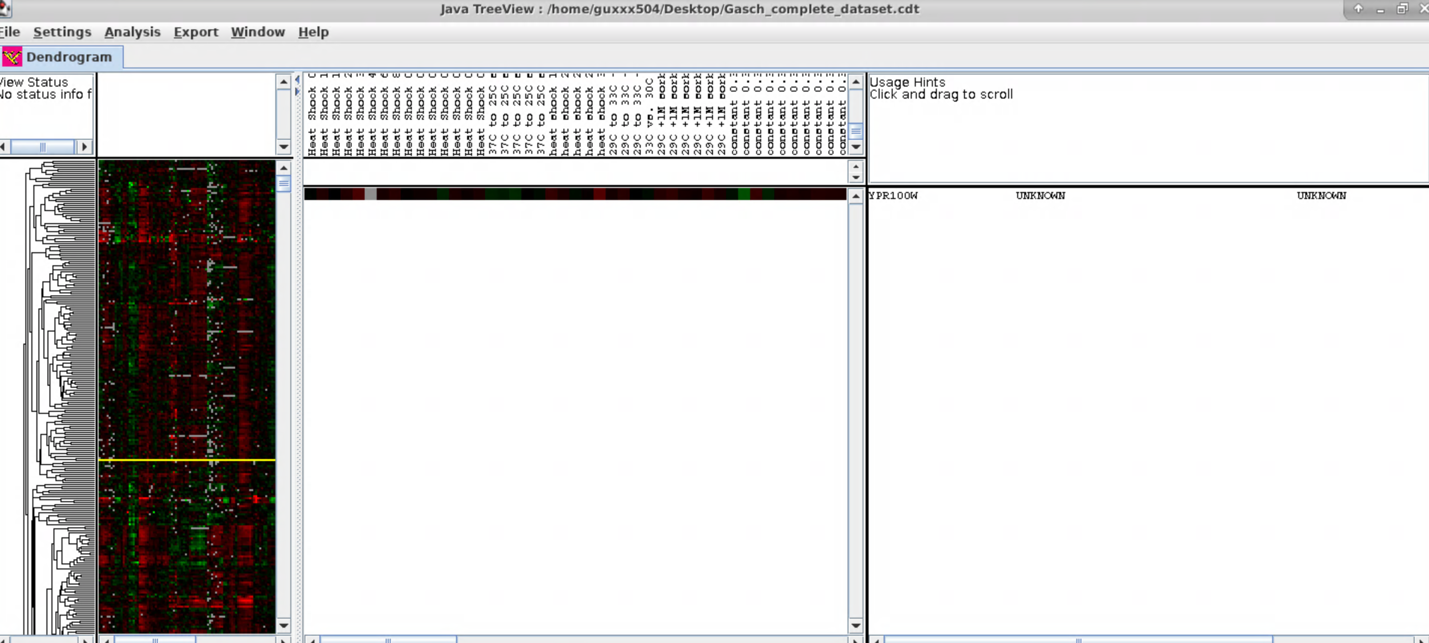
* Temperature Shocks
* Hydrogen Peroxide
* The Superoxide-generating Drug Menadione
* The Sulfhydryl-oxidizing Agent Diamide
* The Disulfide-reducing Agent Dithiothreitol
* Hyper- & Hypo-osmotic Shock
* Amino Acid Starvation
* Nitrogen Source Depletion
* Progression into Stationary Phase

**(3). Describe one of the significant conclusions of the study.**

Defined a term as the “environmental stress response (ESR)” to describe the global expression programs in response to a diverse set of stresses, including their specific features and a common response to all of the stressful conditions. By studying how initiation of the ESR contributes to cellular resistance to various stresses, it helps to understand the role of this program in the yeast life cycle (Gasch et al, 4242).

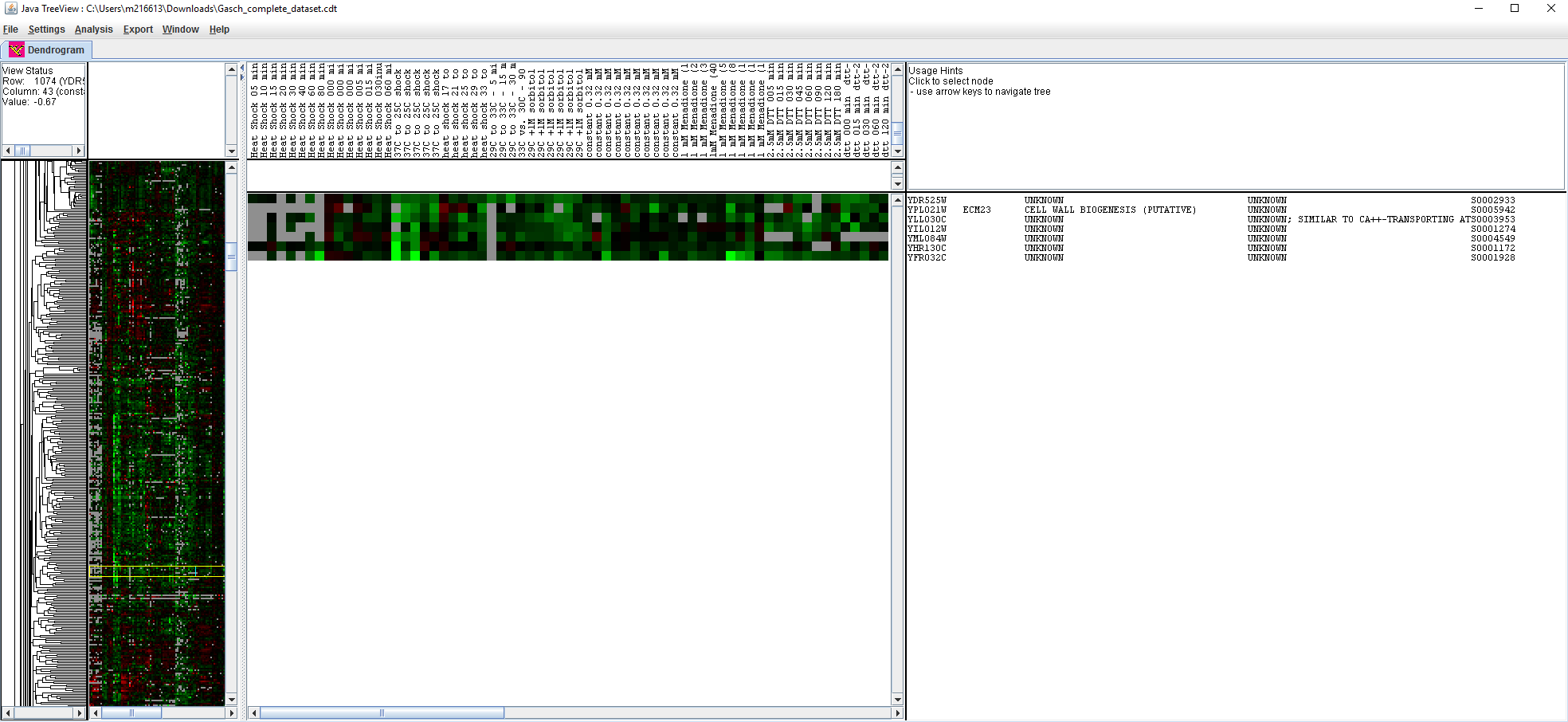
Part II. Exploring the gene expression data with public software tools

(1) + (2).



(3).

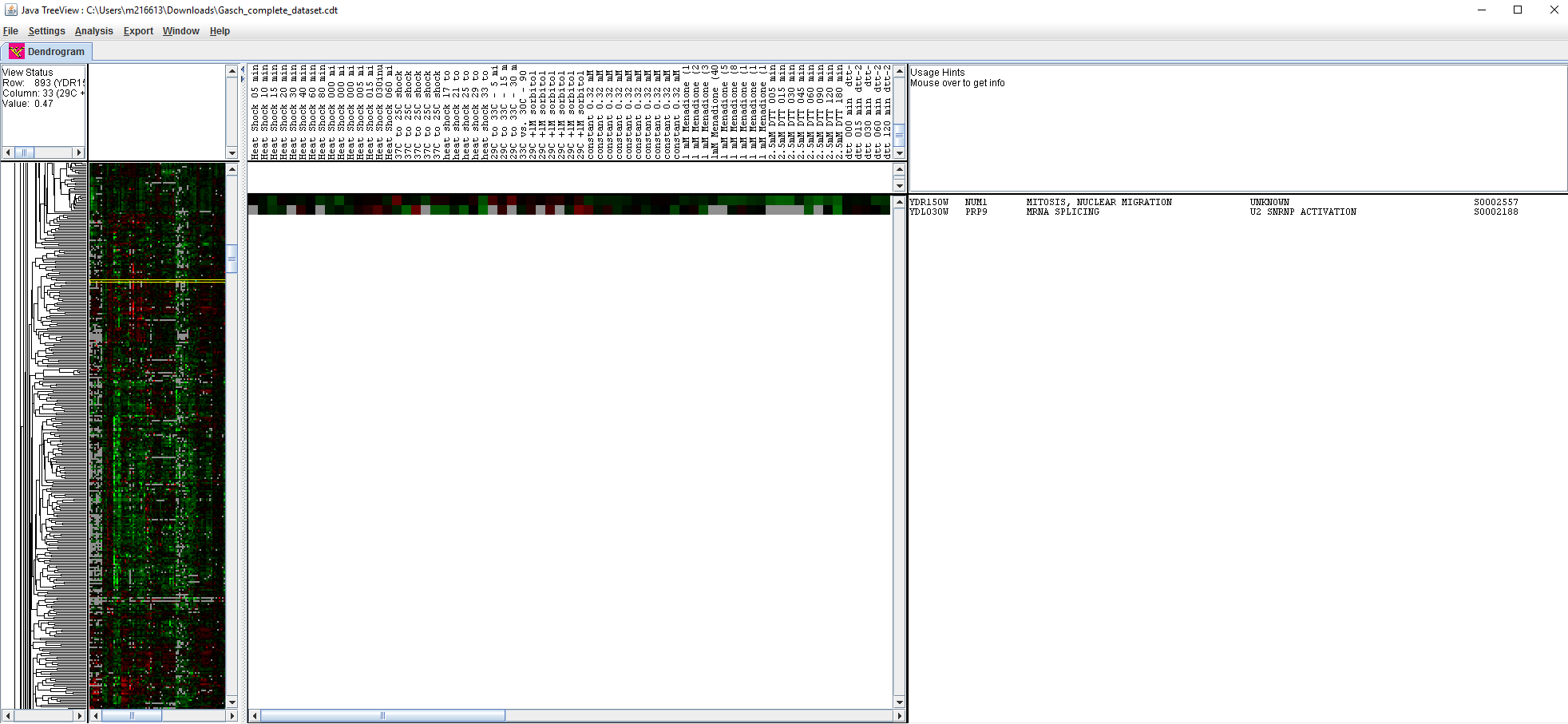
I attached the screenshot of the region I randomly picked up in the following:



The set of locus tags of the genes which are clustered in this region includes YDR525W, YPL021W (ECM23), YLL030C, YIL012W, YML084W, YHR130C, and YFR032C. All of these genes are protein coding genes, and the lineages of these genes are the same, which includes Eukaryota, Fungi, Ascomycota, Saccharomycotina, Saccharomycetes, Saccharomycetales, Saccharomycetaceae, and Saccharomyces.

(4).

I attached the screenshot of the cluster I’ve chosen for this question in the following:



The set of genes I’ve chosen are listed in the following:

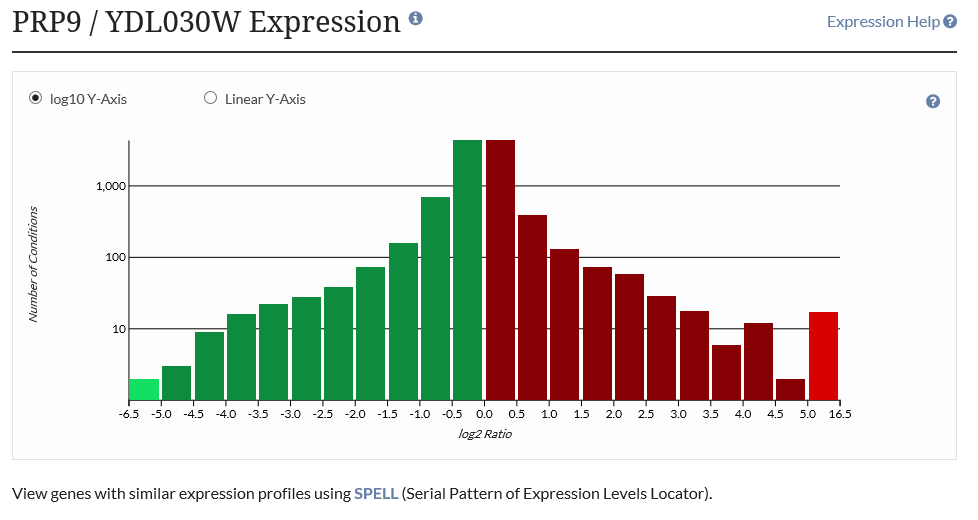
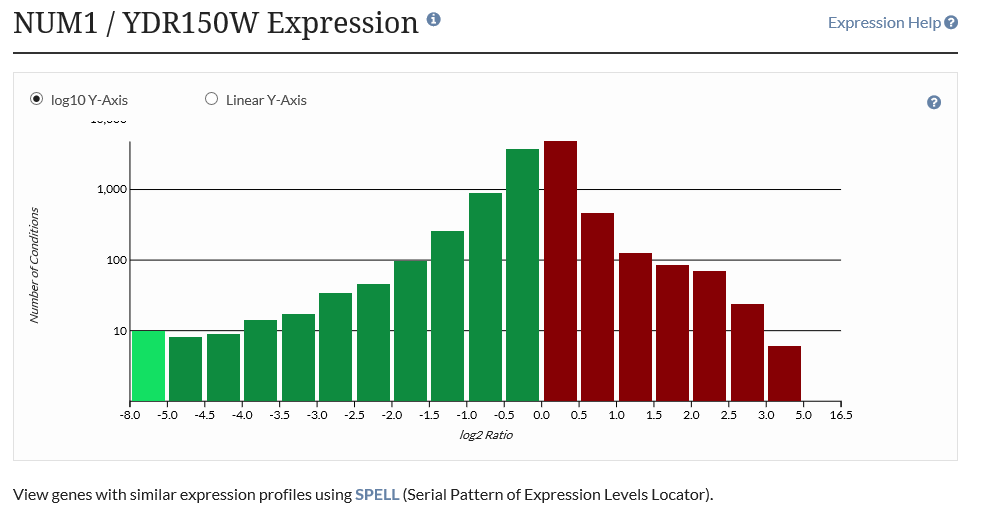
* NUM1 (YDR150W)
* PRP9 (YDL030W)

a). It does make sense that they cluster together.

Because they both located in chromosome IV and have the similar function.

b). Based on the reference webpage provided in the lab 8 assignment sheet, the website provided the figure to view genes with similar expression using SPELL (Serial Pattern of Expression Levels Locator).

I attached the figures for NUM1 and PREP9 in the following:

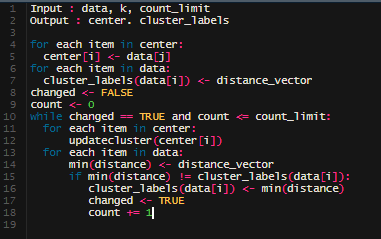


It is obvious that the distributions of the 2 genes with its similar expressions are pretty similar. And this does make sense to give their similar functions.

c). ARG5,6 is the gene that clustered with ARG1 gene. A TATA binding protein regulatory network could be the pattern of expression that drive their similarity. Since the number of conditions associate with this is more than the others, so I think these genes are more active under these conditions.

Part III. Clustering gene expression profiles using the k-means algorithm

(1). I attached the screenshot of my pseudo-code in the following:



And the stopping criterion for the algorithm is



(4).

a. There are 38 iterations for k-means to converge.

Note this would be different each time you run the code

e. There are some clear patterns in terms of what functions are represented.

Part IV. Interpreting your clusters using the Gene Ontology and Public tools

(4). These functions do make sense given the conditions the genes responded to in part III above.