Analysis of Robustness, Network Features, and Cellular Aging in *S. cerevisiae*

Dr. Hong Qin
Spelman College
Jessica Coates, Ella Harley, Teneisha McDonald

December 9, 2012

Abstract

Cellular aging is a fundamental biological process. However, the mechanisms associated with how aging occurs are unknown. Current hypotheses on cellular aging suggest that process is due to underlying gene and protein networks. Within the model organism S. cerevisiae, lifespan can be measured as the length of replicative lifespan (RLS) or chronological lifespan. Through computational methods, we measured the coefficient of variation (CV) for expression datasets from NCBIGEO to determine if there existed a correlation between the RLS, fitness, and interaction degree of protein and genetic networks. Based on computational observations, we were able to determine there exists a strong positive correlation between CV and fitness within all of the data sets. There also existed a relationship between permutated CV values and protein interaction networks and positive gene interaction networks for all of the data sets. These findings suggest that as the fitness of the organism increases the lifespan of the organism also increases and that the degree of protein and positive gene networks could affect lifespan..

Introduction

Saccharomyces cerevisiae is a single celled organism and is commonly known as yeast. While being one of the simplest organisms, it is a eukaryotic cell like the human cell with many similarities. Some of the similarities include: their cells contain a nucleus with chromosomes, they undergo cell division, and also share some of the same proteins. The entire yeast genome

contains approximately 6000 genes and statistically, 31% of the genes can be compared to a robust homolog within the human genome, though this is only an estimate (1). The possibility that the protein and genetic interactions within yeast cells might also exist in human cells makes it an ideal model organism.

Aging, by definition, is the progressive loss of function on a cellular and molecular level over time, resulting in decrease in fertility and increase in mortality (2). Cellular aging is a fundamental biological process that, although is widely studied, has yet to be fully understood. This fact has rendered cellular mechanisms that result in aging and, at times, age-related disease an important topic among biological research. Yeast *S. cerevisiae* has a lifespan measured in two ways. These life spans are replicative life span and chronological life span. Replicative life span is the average number of daughter cells of a single mother cell and includes only dividing cells. Yeast grow by budding out new cells, unlike humans which undergo mitosis. Chronological life span is the duration period of cells in the stationary phase, non-dividing cells. Aging is described quantitatively using Gompertz model of aging. Gompertz model of aging describes aging as an exponential increase in mortality rate over time (3).

$$m = -\frac{1}{s}\frac{ds}{dt} = R_0 e^{Gt}$$

$$s = e^{\left(R_0/G\right)\left(1-e^{Gt}\right)}$$

Figure 1. Gompertz model of aging. *e* is Euler's Number (*e* = 2.71828...). R0 is initial mortality rate, innate susceptibility to dying, lifespan potential at birth. G is the coefficient for mortality rate acceleration.

Within the human genome and yeast, fitness is a measure of RLS in response to genetic mutations in gene pairs. It can be said that yeast with a p value of less than .05 is correlated to the RLS. Inversely, those p values that are found to be greater than .05 are found to have none to

significantly low correlation.

Previous research has been found that cellular aging is proposed to occur as an emergent property at the cellular level through gene and protein networks. This hypothesized mechanism causes reason to suggest there exists a correlation between the cells life span and the robustness that can be experimentally observed and calculated through computational methods.Through using R Studio, we aim to statistically evaluate and find a correlation between the robustness and the rate at which a cell ages in S.cerevisiae .

Methods and Materials

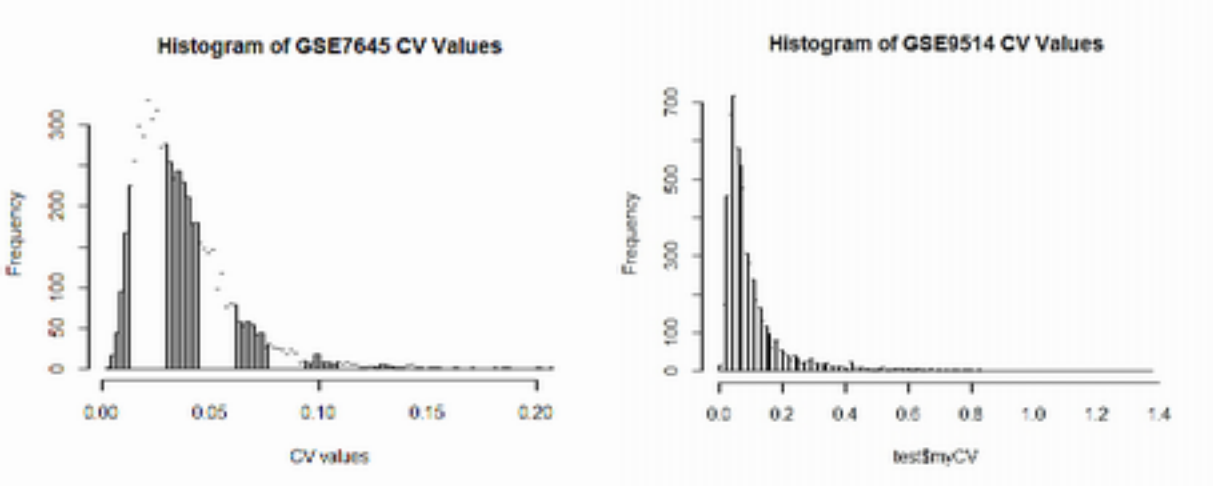
In order to address our aim, three Saccharomyces cerevisiae datasets were obtained from the NCBIGEO database and used for data analysis. The three sets included were based on expressions data for oxidative stress (GSE7645), stress throughout fermentation (GSE8536), and response to heme deficiency and hypoxia (GSE9514). Through the use of R studio programming software, specific opening reading frames and probes were found within each dataset and used to create an expression matrix. Values for signal were then normalized for use in calculating the coefficient of variation (cv). The coefficient of variation was calculated as the dividend of the standard deviation of these values by the mean. A .csv file of the calculated CV values were then used in further data analysis.

To determine if there existed a correlation between the CV values, standard deviation, or mean values from the NCBIGEO data sets and other S. cerevisiae data sets, data sets containing experimental data on fitness, replicative lifespan, and gene/protein networks were obtained (Fraser et. al). Linear regression analysis was then performed to determine a potential correlation

using R studio programming software. Calculated values for permutation were also used to determine the difference of expression for the coefficient of variation or standard deviation in protein networks and/or genetic networks.

Results

When calculating the CV values for the NCBIGEO datasets, the values for the GSE7645 data set ranged from 3.2E-03 to 2.08E-01 with a median value of 3.13E-01 (Figure 2A). In data set GSE9514, the CV values ranged from 0.006247 to 0.068826 (Figure 2B). In the data set GSE8536, the data shows no correlation.. The data set for GSE8536 ranged in value from 1.2E-03 to 2.01E-01 with a median value of 1.02E-01 (Figure 3C).



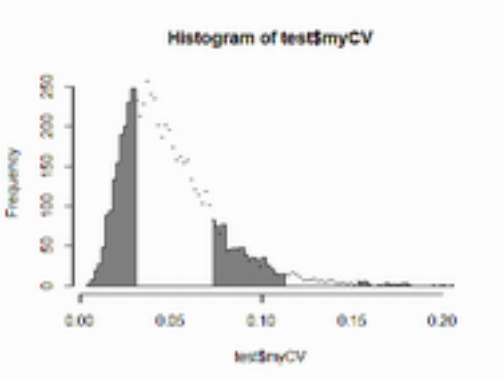


Figure 2. Histograms of Associated Coefficient of Variation Values Derived from NCBIGEO data sets. CV values from NCBIGEO data sets. CV values were calculated based on standard deviation and mean values obtained from data sets. A) Frequency of CV values within GSE7645 data set B)Frequency of CV values within GSE9514 data set.C)Frequency of CV values within GSE8536.

All assays for the correlation between RLS and mean,standard deviation, and CV values showed there was no relationship when compared with the GSE7645 data set. The same trend was observed in the data set GSE9514, which revealed there was no correlation, yielding p-values ($p>0.5$).

Assays for the correlation between CV values from the GSE7645 data set and the fitness data obtained from Fraser revealed a strong positive correlation between the values ($p=2.2E-16$) (Figure 3A). For the GSE9514 data set, assays for the correlation between these two factors syehowed a strong positive correlation, p value 2.2E-16 (Figure 3B). There was also a strong positive correlation between the CV values and fitness data for Figure 3C (GSE8536). The observed p value ws 2.2E-16.

Correlation assay to determine a relationship between CV values from the GSE7645 data set and protein/gene network interactions revealed no significant correlations ($p>0.5$). GSE8536 also showed no significant correlation with a p-value calculated at 0.16 For the data set GSE9514, there was no correlation between the GSE9514 CV and protein interaction network and positive gene networks. Their respective p values were 0.2687 and 0.07592. There, however, was a weak correlation between the GSE9514 CV and negative gene networks, p- value 0.04129.

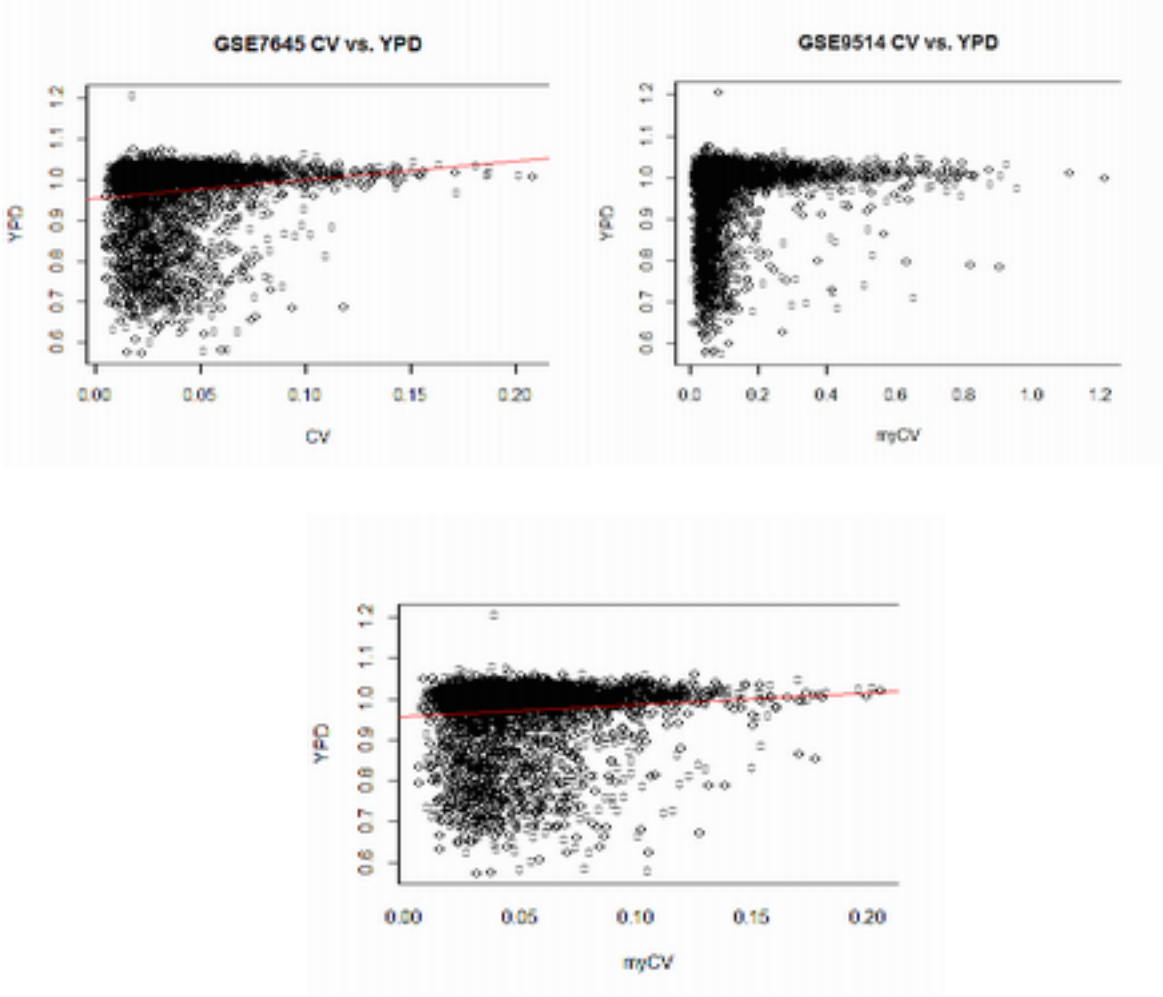
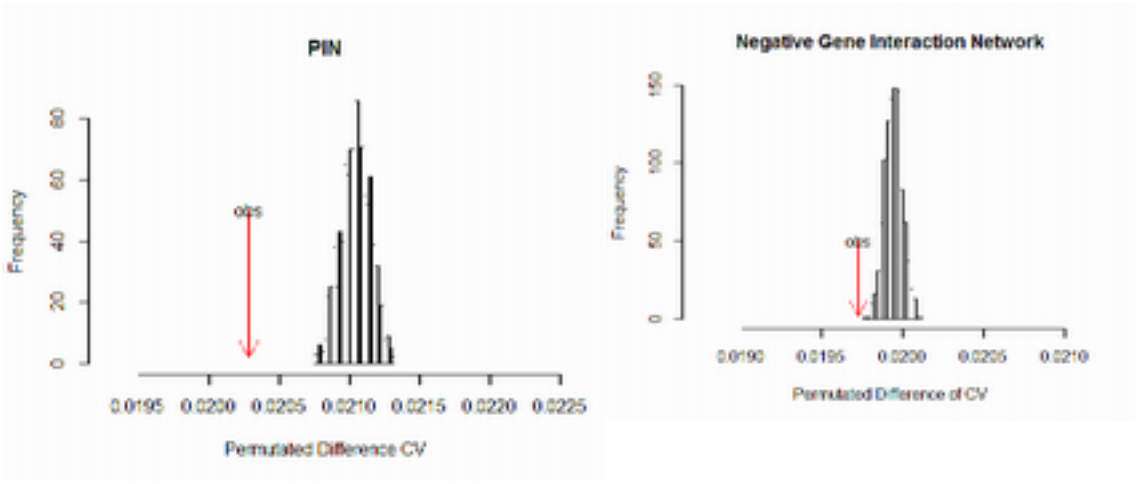


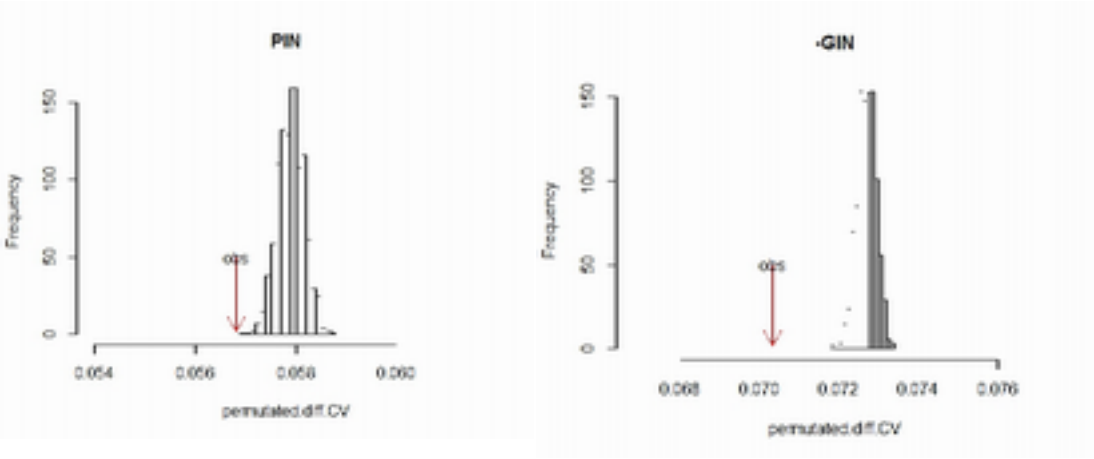
Figure 3. Linear Regression Models of CV Values versus Measurements of Cellular Aging in S. cerevisiae. Linear regression analysis of correlated values for cellular aging vs coefficient of variation values. ($p<0.05$). A) GSE7645 CV vs. Fitness. B) GSE9514 CV vs. Fitness. C) GSE8536 CV vs. Fitness.

When using permutation to investigate the correlation between protein interaction network and GSE7645 CV values, there was an associated p value of 0.001 (Figure 4A). For There existed no relationship between positive gene networks and CV values for GSE7645 ($p=0.998$). However, there did exist a correlation between CV values for GSE7645 and negative gene networks ($p=0$) (Figure 4B). For GSE9514 CV, there was an associated p value of 0.001 when investigating the relationship between CV and protein interaction networks (Figure 4C). No existing relationship was shown when GSE9514 CV was compared with positive gene

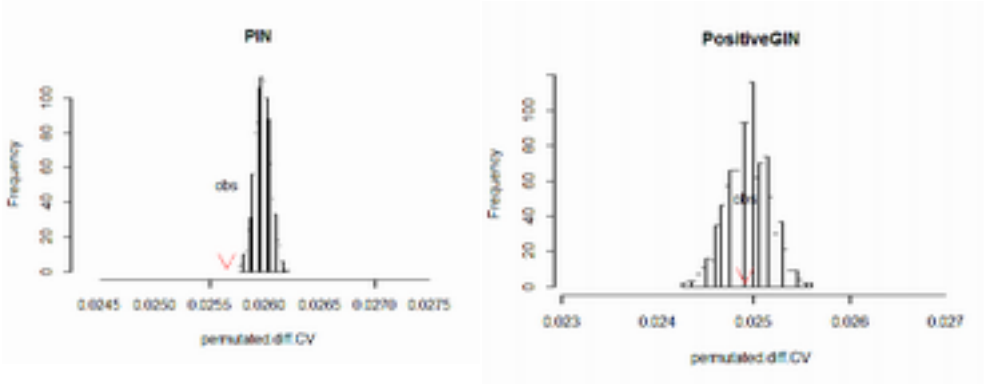
networks. A relationship was shown between GSE9514 CV and negative gene networks, with a given p-value of 0 (Figure 4D). For the data sets of GSE8536 (Figure E) recorded observation showed that there was no correlation between the CV and protein interaction networks (Figure E). However, there seemed to be a significant correlation between the CV and positive gene network (Figure F).



A.



C.



E.

Figure 4. Calculated Difference of Expression CV in Protein Networks. A) Permutated Difference of CV frequency (GSE7645) in protein interaction networks. B) Permutated Difference of CV frequency (GSE7645). C) Permutated Difference of CV frequency (GSE9514) in protein interaction networks. D) Permutated Difference of CV frequency (GSE9514). E) Permutated Difference of CV frequency (GSE8536) in protein interaction networks. F) Permutated Difference of CV frequency (GSE8536).

Discussion

The observed computational data suggests that fitness, protein interactions, and negative gene interactions are correlated to cellular aging. However, when performing experimental observational analysis, there is room for errors to occur. A potential source of error could have occurred within the encoding or naming files that were saved from Rstudio. This could result in incorrect data being processed by Rstudio and an alteration of interpretation of results.Another possible source of error could result from the improper interpretation of graph and results. This could not only come from reading the graphs incorrectly but also from misinterpretation of R-values which could change the correlation between all three results. Although, research and observation has been conducted that does not mean the work stops there. There is still future studies and research that could be performed to that could not only lead to future discoveries but also to have a more accurate results (if the same results were reached during multiple trials). These future experiments could include the correlation between other genomes of this nature. Also, it would be interesting to see if the results differ in the presence of different environmental

factors.

References:

1) Twyman, Richard. *Model Organism: Yeast.* The Human Genome. Available: http://genome.wellcome.ac.uk/doc_WTD020808.html. Accessed 09 December 2012.
2) Esra Borklu Yucel*, Kutlu O. Ulgen. *A Network-Based Approach on Elucidating the MultiFaceted Nature of Chronological Aging in S. cerevisiae.* Available from: http://spelelearn.spelman.edu/file.php/2622/papers/PLoS_ONE_2011_Borklu_Yucel-1.pdf. Accessed 09 December 2012.
3) Qin, H. and M. Lu, Natural variation in replicative and chronological life spans of Saccharomyces cerevisiae. Exp Gerontol, 2006. 41(4): p. 448-56.