PROJECT SUMMARY

Overview:

The overarching goal of this 5-year CAREER proposal is to advance the current knowledge on aging from the perspective of gene networks through integrated research and teaching. Aging is a fundamental question in biology, yet its mechanism remains elusive despite decades of research. Gompertz model describes an exponential increase of mortality rate over time - a ubiquitous feature in biological aging but not in the failures of complex machinery. The negative correlation between the two Gompertz parameters is a conserved demographic characteristic of aging in many organisms including yeast. The stochastic aspect of aging is apparent - Yeast cells from a single colony live to different ages despite their genotypic homogeneity. This coexistence of individual plasticity and a universal demographic characteristic is a puzzling aspect of aging. Hundreds of yeast genes are known to influence lifespan. Paradoxically, not a single gene can be claimed as a direct cause of aging. To complicate things even further, different and even opposite pathways have been observed when yeast aging is measured in dividing or non-dividing cells. This seemingly complicated picture is addressed by the core idea of this proposal - Cellular aging is an emergent property of gene networks.

The defining feature of biological aging, exponential increase of mortality rate, can arise from the PI's network reliability model for cellular aging (NRMCA) with stochastically interacting non-aging components, thereby demonstrating the emergence of cellular aging. The universal negative correlation between the two Gompertz parameters is a prediction of NRMCA. Counter-intuitively, the rate of cellular aging is predicted to be proportional to network robustness and is corroborated by empirical evidences. In this proposal, the PI proposes to advance the current understanding of cellular aging in three directions. Aim 1: Investigate the role of network configuration in cellular aging and evaluate the reliability of the yeast protein interaction network. Aim 2: The PI will first develop a framework to study network modules with limiting effects on aging and apply it to infer network changes in yeast mutants. The PI will then study the role of synthetic lethality in cellular aging, and will finally study the role of renewals/repairs in aging with a focus on mitochondrial renewals. Aim 3: Develop a network framework to study lifespan as a quantitative trait and understand its natural variation.

Intellectual Merit :

Overall, the PI will develop a mathematical framework that can illustrate the basic principles of cellular aging, and can also infer network changes from experimental data of lifespan. Our NRMCA is the first mathematical model that demonstrates the emergent aspect of cellular aging from gene networks, the important role of stochastic heterogeneity in biological aging, and the close-connection between cellular aging and network robustness. NRMCA provides a unifying framework for aging in both dividing and non-dividing cells, and offers insights on aging from the quantitative genetics perspective. By applying NRMCA to evaluate experimental lifespan data and to infer network changes caused by genetic alterations, NRMCA will provide a network approach to interpret experimental data and will shed new light on yeast mutations that affect lifespan.

Broader Impacts :

Our educational goal is to provide cross-disciplinary training to African American students and cultivate their interests in quantitative biology through integrating research with teaching, workshops, a student Chapter for the Society of Industrial and Applied Mathematics, and broad dissemination of educational materials. This project will provide cross-disciplinary training to minority undergraduates in mathematics, computing, and systems biology at a historically black college for women. This project will be integrated into 4 existing courses and 1 proposed new course. The PI will propose a new minor of bioinformatics and systems biology to the college. The PI will build a community and culture of computing on campus by organizing workshops for students and/or faculty. The PI will reach for a broad community through tutorial videos at YouTube, open research projects at GitHub, and an open research blog.

1. Specific aims and career goals

The overarching goal of this 5-year CAREER proposal is to advance current knowledge of aging from network perspective through integrated research and teaching. Aging is a fundamental question in biology, yet mechanistic understanding of aging remains elusive. The PI has proposed a unique mathematical model demonstrating that cellular aging is an emergent property of gene networks http://arxiv.org/pdf/1305.5784.pdf [1]. This model predicts that the rate of aging is proportional to network robustness and heterogeneity plays a key factor in biological aging. In this proposal, the PI aims to develop an innovative and sophisticated network reliability model of cellular aging (NRMCA), and develop computing methods to apply NRMCA to fit experimental data through integrated theoretical, computational, and experimental studies. The major research aims are:

Aim 1 (Years 1-2). Study the role of network configuration in network aging. We will first focus on the power-law configuration and error tolerant features of gene networks. We will then study the aging of the yeast protein interaction network and compare it with random networks.

Aim 2 (Years 2-4). Develop methods to study the limiting interaction modules on network aging and apply them to yeast mutants. We will develop general frameworks and use them to fit experimental data. We will study synthetic lethality because of its predicted special property in network aging. We will also study mitochondrial renewals to understand the role of renewals/repairs in network aging.

Aim 3 (Year 3-5). Study network aging as a quantitative trait. First, we will study one-locus and two-locus models to illustrate the general network principles on heritability of lifespan. Second, we will develop a multi-locus model based on the yeast protein interaction network to understand the yeast natural lifespan variation. We will then use network simulations to evaluate methods for genome-wide-association studies of lifespan variation and develop guidelines.

Our educational aim is to integrate the proposed research into undergraduate education at a historically black college for women, to engage minority students in mathematical, computational and network biology, and to address the urgent needs for training future quantitative biologists.

These aims are based on the PI's experiences, expertise, previous and preliminary findings, and significance of expected results. The PI has a demonstrated record of yeast aging research [2-4], network simulation and analysis [5, 6], computational and experimental genomics [2, 7-9], mathematical modeling [10] and integrating research into teaching [11, 12] (see Chair's letter). Moreover, the PI has developed collaboration with many research groups in the field of yeast aging research (see their letters).

PI Qin's career goals are to become an effective teacher, a nurturing mentor, and an innovative scholar. This project is designed to be both intellectually challenging and reasonably practical to be carried out in a liberal arts college. Qin's long-term research goal is to understand how gene networks influence complex traits. This CAREER proposal will lay the foundation for a life-long endeavor on network biology.

2. Introduction to the network reliability model of cellular aging (NRMCA)

2.1 Brief background on quantitative models of aging

Aging is a fundamental question in biology [13-15]. Aging of an organism or a system can be described by the mortality rate $\mu(t)$, which is the normalized declining rate of viability $S(t)$ over time t:

$$\text{Mortality rate}: \quad \mu(t) = -\frac{1}{S(t)}\frac{dS(t)}{dt} \qquad \text{(Eq. 1)}$$

Aging occurs when $\mu(t)$ is a positive increasing function that indicates increasing chance of dying over age. Mortality rate is also known as the force of mortality, failure rate or hazard rate in various contexts [16-18]. In general, $\mu(t)$ can be a power function for machine aging and an exponential function for biological aging [19, 20].

$$\text{Machine aging}: \quad \mu(t) = C_1 t^{C_2} \qquad \text{(Eq. 2)}$$

$$\text{Biological aging}: \quad \mu(t) = \text{Re}^{Gt} \qquad \text{(Eq. 3)}$$

An organism can be *non-aging* when $\mu(t)$ is a constant in Eq. 1. In this case, drop of viability becomes an exponential decay, which is basically a first-order chemical reaction, just like the exponential decay of radioactive isotopes. Individuals from these populations are as good as new at any time point, and are therefore *non-aging*. Bacterial phages indeed display this kind of non-aging characteristic [21]. It is worth clarifying that non-aging phages are not immortal, they just die with constant mortality rates.

The exponential form of mortality rate (Eq. 3) is known as the Gompertz model of biological aging [22]. The initial mortality rate, R, describes the innate susceptibility to dying. The Gompertz coefficient, G, determines acceleration rate of mortality rate over time and is therefore a parameter for rate of aging. The

exponential increase of mortality rate is a universal characteristic of biological aging, and has been observed in bacteria, yeast, worms, fruit flies, mice, and humans [4, 13, 23-26]. For non-aging phages, their Gompertz coefficient G = 0, corresponding to zero rate of aging. It is therefore perplexing why simple organisms like bacterial phage is non-aging, whereas cellular organisms age.

2.2 Emergent property of cellular aging is shown by NRMCA

Our central hypothesis is that cellular aging is an emergent property of gene networks. Emergent property generally refers to a feature that can be found only at the system level but not at the component level. Classic examples include termite castles, schools of fishes, and flocking of birds. To prove that cellular aging can emerge from gene networks, components in our network model ought to be non-aging. Specifically, we need to demonstrate that Gompertzian aging can arise from gene networks that are made of components with constant mortality rates.

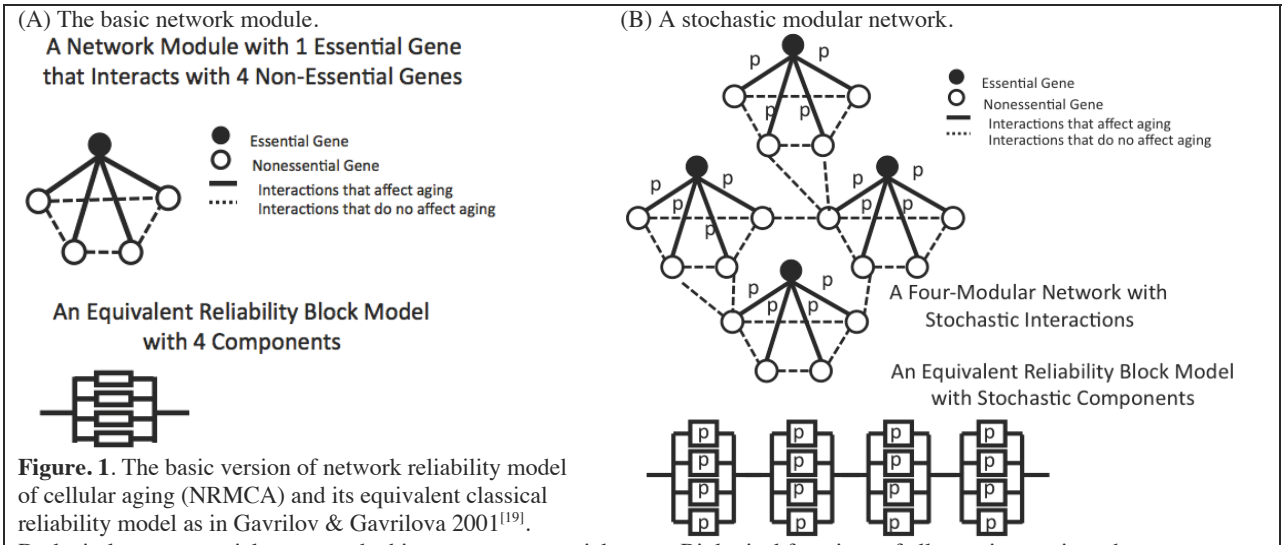


Figure. 1. The basic version of network reliability model of cellular aging (NRMCA) and its equivalent classical reliability model as in Gavrilov & Gavrilova 2001[19]. Dark circles are essential genes, and white ones non-essential genes. Biological functions of all gene interactions decay exponentially, i.e., *non-aging*. When an essential gene loses all of its interactions, it is equivalent to gene deletion and results in cell death. Consequently, only essential genes' interactions influences aging and are represented by solid links. Dashed links are interactions that will not affect aging. Interactions are stochastic, and the chance of an interaction to be initially active is *p*. Independent failures are assumed for both essential modules and gene interactions. Graph and block presentations are mathematically equivalent and will be used interchangeably. Mathematical details are available in our preprint [1].

The basic version of NRMCA contains *m* number of essential modules, and each module contains 1 essential and *n* non-essential genes (Figure 1). Stochastic gene interactions follow a binomial distribution, and the chance of a gene interaction to be active is *p*. The biological function of each interaction is assumed to be *non-aging* and decays exponentially with a constant rate of λ. Death of a cell occurs when an essential gene loses all of its interactions, equivalent to deletion of an essential gene. When failures of modules are independent, analytic approximation for the network mortality rate is

$$\text{Network Mortality Rate}: \quad \mu_{net}(t) = \mathrm{Re}^{Gt} \quad when \quad t << 1/\lambda \qquad \text{(Eq. 4)}$$

$$R = cmnp\lambda(1-p)^{n-1} \; ; \quad G = \frac{n-1}{t_0} = \frac{\lambda p(n-1)}{1-p}; \quad t_0 = \frac{1-p}{p\lambda}$$

where *c* is a normalization constant, t_0 is the 'initial virtual age' proposed by Gavrilov and Gavrilova [19, 20]. Eq. 4 is the Gompertz model of aging. Hence, the defining characteristics of biological aging, the exponential increase of mortality rate over time, can arise from a network model with *non-aging* components. *By definition, we have shown that cellular aging is an emergent property of this model network.* Mathematical details are available in our preprint [1].

In essence, we use network failures to model deteriorating biological functions during aging. The age of each essential module is determined by the maximal age of its interactions. The age of the network is determined by the minimal age of the modules. Aim 3 will explore the influence of this maximum-minimum nature of network aging on lifespan as a quantitative trait. Survivor functions can be

obtained from mortality rate as $S(t) = e^{-\int_0^t \mu(t)dt}$. The cumulative distribution function (CDF) of ages is 1-*S(t)*. Hence, Eq. 4 specifies the CDF for the network stochastic aging process.

NRMCA can be represented by an adjacency matrix **A**, a convenient mathematical tool during numerical simulations. For the basic version of NRMCA in Figure 1, element $a_{i,j}$ in the adjacency matrix follows a binomial distribution and decay with a constant mortality rate. The graph presentations and the block-pathway presentations in Figure 1 are mathematically equivalent, and will be used inter-changeably.

2.3 Important model predictions and properties.

One important prediction is the counter-intuitive positive correlation between the Gompertz coefficient *G* and network robustness – more robust cells have higher rates of aging (Figure 2). Based on Eq. 4, *G* is approximately proportional to the number of active interactions per essential gene (*n*×*p*) and will increase dramatically when chances of these interactions being active become higher (*p*). The number of active interactions per gene can be viewed as a measure of network robustness. Hence, stronger robustness would lead to a faster rate of aging (Figure 2). This prediction is counter-intuitive because aging dynamics is quantified here by two Gompertz parameters, *G* and R, with only G as a measure of rate. In contrast, the colloquial meaning of aging rate actually contains information for both parameters.

Another important prediction is that limiting network modules with redundant functions and non-redundant functions will influence network aging in opposite ways (detailed in Aim 2.1 and Figure 7). Redundant network modules indicate parallel pathways in reliability modeling. Activities of these modules ought to be longer than average if they are the limiting steps of network aging. Non-redundant network modules indicate serial configuration in reliability modeling, and their activities should have shorter than average half-life in order to be limiting steps in network aging.

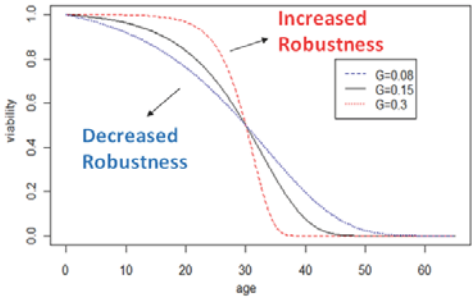


Figure 2. A counter-intuitive prediction of NRMCA: Stronger robustness increases *G*, as shown by the sharper transition of the dying off phase. Supporting evidences are in section 3.2 and 3.3

An important property of the model is the Strehler-Mildvan correlation – the trade-off between *G* and *R* [27]. It can be shown that ln(*R*) ≈ -*BG* + *Intercept*, where *B* and *Intercept* are constants based on n, *λ,* and *p.* This property explains the universality of the Strehler-Mildvan correlation.

More insights and predictions for network aging will be discussed in our proposed study on network configuration, limiting modules, synthetic lethality, renewals/repairs, and heritability of aging.

2.4. Overview and challenges of cellular aging in yeast

Aging likely occurred during evolution in unicellular organisms that predates eukaryotes because aging is observed in bacteria [25, 26, 28, 29]. Tremendous strides have been made toward the mechanistic understanding of aging over the past two decades; yet the very concept of aging is still under debate (For example, see [30]).

As a unicellular organism, the budding yeast *Saccharomyces cerevisiae* has proven to be a good model system for studying mechanisms of cellular aging [31-39]. Many key features of cellular aging were first discovered in yeast before they were established in metazoan cells [28, 32, 40-42]. The lifespan of yeast can be measured in two ways: replicative and chronological lifespan. Replicative lifespan (RLS) is the number of cell cycles that individual mother cells produce before they senesce and cease to divide [31, 35, 43], and is often determined by microdissection. Chronological lifespan (CLS) is how long cells can survive without dividing in stationary phase [33, 44], and is often assessed by quantifying colony-forming units.

Paradoxically, biology of aging becomes evasive once we delve into molecular mechanisms. Although hundreds of yeast genes have been found to influence aging, none of these genes suggests molecular mechanisms that are directly linked to aging. The effect of SIR2 on lifespan is attributed to the "toxic" effect due to accumulation of extrachromosomal rDNA circles [45], a concept that is not only mechanistically obscure but has also been challenged [46]. The effect of TOR pathway on replicative lifespan is attributed to the decreasing ribosome function and translation [40] or to the hyper-activation of cellular functions [30, 47]. The mechanism of TOR on chronological lifespan remains unclear [48]. In fact, it is

speculated that bona fide aging genes do not exist because there are no conserved causes of aging [34, 49, 50]. With no genes as direct causes of aging, it is surprising that calorie restriction (CR) is a universal way of intervention to extend lifespan. In yeast, CR can extend both RLS and CLS [51, 52], despite that there are substantial differences in active pathways between the two different aging processes [53].

A large body of experimental data suggests that complex gene networks are involved in yeast aging. In a large-scale screen, 90 gene deletions were found to extend CLS in BY laboratory strains [54], and 300 gene deletions can shorten CLS [54]. In another screen of RLS, 20% of the gene deletions were found to shorten RLS, whereas 10 out of 564 genes significantly extended RLS [55]. In a quantitative trait study, transgressive segregation of CLS was observed, indicating the involvement of many loci [56]. In collaboration with Jeff Townsend at Yale University, we compared gene expressional profiles of short and long-lived segregants of a wild yeast isolate, and found 15 genes with consistent differential expression levels between the long- and the short-lived progenies[2]. The complex nature of aging has motivated many authors to argue that network is the key to understanding aging [57-68], and their studies have provided inspirations to our model.

It is clear that cellular aging is a stochastic process to a great extent − Genotypically homogenous yeast cells from a single colony will live to different ages even when they are kept in the same laboratory environment. We found that genetic factors contributes ~22% of natural variation in individual RLS [4]. Therefore, it is perplexing that there exists a universal characteristic of aging at the demographic level, known as the Strehler-Mildvan correlation [4] (see section 3.1), despite its likely complex mechanisms and the great plasticity of individual lifespan. We argue that this kind of universality suggests a common principle in the stochastic processes of aging.

2.5 Contribution of the proposed study to the field of cellular aging (Intellectual merit)

To our best knowledge, NRMCA is the first mathematical model to demonstrate the network emergent aspect of cellular aging. It provides a conceptual framework to explain the seemingly inconsistent experimental data, individual plasticity, and universal demographic characteristic of cellular aging in yeast. As an emergent property of networks, cellular aging is a system-level property, which explains the difficulty of pinpointing individual genes as direct causes of aging. The universal Strehler-Mildvan characteristic of aging can be attributed to the common interacting patterns of gene/protein networks shared among most species. The emergent aspect of aging also provides a link between replicative and chronological aging, even though specific pathways differ in the two ways of aging [53].

NRMCA argues that heterogeneity of gene interactions is an important factor between biological aging and non-biological aging. If intrinsic stochastic noises are removed from our model, increase of mortality rate would follow the Weibull model (Eq. 2), which is often the failure model of complex machinery [19, 20]. NRMCA can provide a mechanistic foundation for the antagonistic pleiotropy theory [15] and the disposable soma theory [69] on the evolution of aging. NRMCA is also consistent with the free radical theory that reactive oxygen species (ROS) is a major source of stochastic damages for aging [14].

NRMCA argues that *G* is a measure of robustness. Robustness can reconcile the mutational costs to individuals and the evolutionary benefits to the population [70], because the phenotypic effects are hidden in most conditions. Robustness is related to canalization, and it has been argued that network buffering is a key mechanism of canalization [70-74]. Our model also demonstrates that network biology is a useful way to tackle biological complexity, and can provide insights not easily achieved by reductionist approaches.

On the practical side, NRMCA provides a network model to fit and analyze experimental data and can infer network changes from empirical lifespan data (see Aim 2).

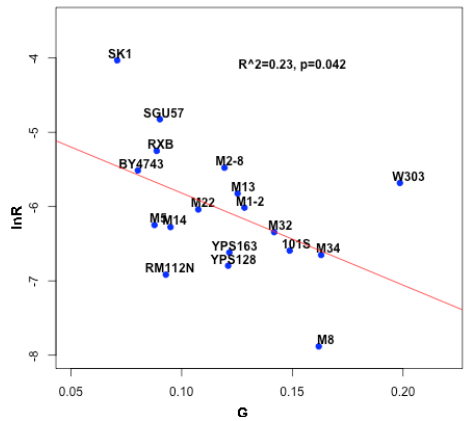
3. Previous and preliminary results that support NRMCA

The PI has a demonstrated record of yeast aging research [2-4], network simulation and analysis [5, 6], computational and experimental genomics[2, 7-9], and mathematical modeling [10]. My group published the first quantitative study on yeast aging using the Gompertz model and reported the Strehler-Mildvan correlation in yeast [4]. My lab has collected more than 150 natural isolates of yeast, whose genomes have been sequenced or partially sequenced. Some highlights are presented below.

3.1 Strehler-Mildvan correlations in yeast populations

We previously reported the Strehler-Mildvan correlation in yeast wild isolates [4]. We have now reaffirmed this observation using many more strains (Figure 3). This kind of universality suggests a common principle and is a key motivation for us to develop NRMCA.

(A) Diploid strains in different backgrounds



(B) Haploid deletion strains in BY4741 background

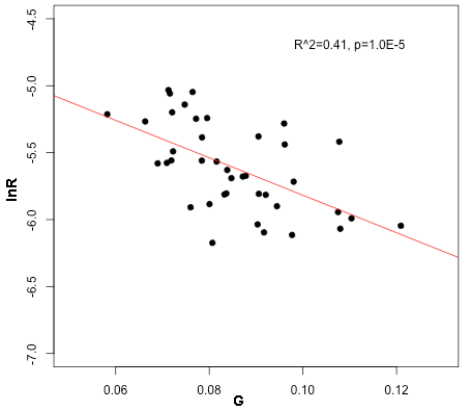


Figure 3. Strehler-Mildvan correlations between the two Gompertz parameters in yeast populations. Negative correlation between ln(R) and G in diploid strains with different genetic backgrounds (A) and in deletion mutants in the BY4741 MAT a background (B). Correlation in the deletion mutants in BY4742 MAT alpha background was also observed but is not shown here. Replicative lifespans of deletion mutants were generously provided by the Kaeberlein group.

3.2. Morphological robustness correlate with RLS.

Cell morphology of yeast deletion mutants is available at the *Saccharomyces cerevisiae* Morphological Database (SCMD, http://scmd.gi.k.u-tokyo.ac.jp/). SCMD provides a list of 501 morphological parameters in four groups: cell shapes, bud sizes, nucleus locations, and actin localizations [75]. Three Spelman students in the PI's lab, Ms. K. Matheson, O. Morrison, and R. Levy, calculated the coefficient of variation (CV) of the normalized 501 parameters for each yeast gene, which describes the unmasked morphological plasticity when the wildtype gene is removed and is therefore a proxy of morphological robustness for the wildtype gene function. This measure of morphological robustness was negatively correlated with RLS (R^2=0.034, p=1.3×10^{-5}), indicating that large morphological variance corresponds to short RLS. This finding is consistent with the known role of morphology in the asymmetric partition of damaged proteins during mitosis [76]. Although active transport is involved[77], mitotic asymmetry can be sufficiently achieved by the slow diffusion of large aggregates, geometry of the mother and daughter cells, and the narrowness of the passage between them [76]. The RLS data used by the students contain RLS for 564 genes measured by the Kaeberlein group [78].

3.3. Diploid cells have larger *G* than haploid cells.

Diploid cells are generally more robust than haploid cells. If we assume the number of molecules to perform a biological function follows a binomial distribution with a probability of p and are drawn from ***n*** number of molecules, coefficient of variation (CV) can be calculated from standard deviation divided by mean: $\sqrt{np(1-p)}/np = \sqrt{(1-p)/np} \propto 1/\sqrt{n}$. Doubling the numbers of molecules should reduce noises by $\sqrt{2}$ and hence increase robustness, as argued by Schrödinger in 1944 [79]. Yeast protein expression noises in G1 phase are much lower in diploids than that of haploid cells [80]. Empirical observations also show that diploid yeast cells are much larger than haploid cells. Using likelihood ratio test, we compared the Gompertz parameters of diploid BY4743 with its haploid counterpart BY4741 and BY4742. We found that G=0.079 in the diploid strain and G=0.065 in haploid counterparts, and the difference is significant at p=3.6x10^{-10}. Hence, diploid cells age faster than haploid cells.

4. Research Plans.

We propose the following plans to achieve three major aims by theoretical studies for basic principles, by developing computing methods to infer network changes from experimental data, and by network simulations based on the yeast protein interaction network - an empirical example of our NRMCA (Figure 4). We will focus on the yeast protein interaction network because it is less biased against essential genes. Although efforts have been made to study genetic interactions for essential genes, only about 30% of the essential genes are covered by a recent study on yeast genetic networks [81]. Our collaborators on theoretical and computational studies include a statistics expert Weibiao Wu at the University of Chicago, and a biophysics expert Yi Jiang at the Georgia State University. Our experimental

collaborators include Matt Kaeberlein at the University of Washington, Fusheng Tang at the University of Arkansas at Little Rock, and John Hartman at University of Alabama at Birmingham. (All support letters are attached.)

Network Configuration (Aim 1, Yr1-2):
Aim 1.1: Power-law configuration.
Aim 1.2: Reliability and aging of yeast protein interaction network.
Limiting Steps in Network Aging (Aim 2, Yr 2-4)
Aim 2.1: Develop a framework for limiting interaction modules and apply it to yeast mutants.
Aim 2.2: Synthetic lethality.
Aim 2.3: Renewals/repairs with focus on mitochondrial renewals.
Network model of Lifespan as a Quantitative Trait (Aim 3, Yr 3-5)
Aim 3.1: One-locus and two-locus models, missing heritability.
Aim 3.2: Multi-locus study and natural lifespan variation.

Figure 4. Overview of our project aims. We will address three main aspects of network aging. Network simulations in aim 1 will be used for aim 2.1, 2.3, and 3.2. Aim 3.1 is connected with aim 2.1. Allele effect on network aging in aim 3.2 will benefit from all previous work, but aim 1.2 and 2.2 are more closely connected.

Aim 1. Study how network configuration influences network aging.

Aim 1.1 Study the impact of power-law and error tolerant network configuration on network aging.

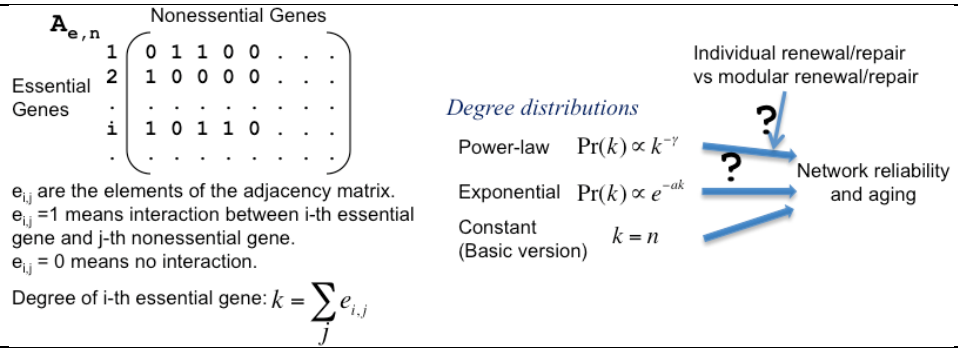


Figure 5, Aim 1.1. Degree distributions on network reliability and aging.

Our NRMCA shows that network configuration and heterogeneity play key roles during the emergence of biological aging. A key source of heterogeneity in gene networks is its power-law feature – the degree distribution of genes follows $\Pr(k) = Z(\gamma)^{-1} k^{-\gamma}$, where *k* is the number of connections per gene, *Z* represents the Ziemman function, and γ is a coefficient [82]. When $\gamma \le 3$, the variance of *Pr(k)* is infinite. For most biological networks, γ is often between 2 and 3 [83, 84] which indicates tremendous amount of heterogeneity in biological networks. In addition, networks with power-law features, such as Internet, are robust to random failures but are fragile to deliberate attacks [85, 86]. In yeast protein networks, highly connected genes, hub-genes, are less likely to directly interact with other hub genes in the protein interaction networks, which contributes to the error tolerance of protein networks [87]. Perturbation of protein concentrations can be mostly buffered locally in yeast protein binding networks, but they can also cascade over more 4 interactions away in certain pathways [88].

We will first adopt an analytic approach to find out the basic principles, in collaboration Dr. Weibiao Wu. With regards to network failures, we only need to focus on the essential genes in NRMCA. By assuming 'perfect error tolerance' in the sense that all hubs are essential genes and essential genes are independent, we found out an approximation for the network mortality rate:

$$\mu_{net}(t) = m\sum_{Klow}^{K\max} \Pr(k) ck (p\lambda)^k t_0^{k-1} (1+t)^{k-1} = m\sum_{Klow}^{K\max} \Pr(k) R_k e^{G_k t} \qquad \text{(Eq. 5)}$$

$$R_k = cmnp\lambda(1-p)^{k-1} \;;\quad G_k = \frac{n-1}{t_0};\quad t_0 = \frac{1-p}{p\lambda},\ when\quad t << 1/\lambda$$

where K_{low} and K_{max} are the lower and upper bounds of connectivity *k* for essential genes, R_k and G_k are the initial mortality rate and Gompertz coefficient for the network module in which an essential gene interacts with *k* non-essential genes. Hence, network mortality rate is basically the weighted mean of modular mortality rates. We will explore closed forms for the network R_{net} and G_{net}, and compare them with those of other network configurations (Figure 5). Intuitively, K_{low} and γ seems to be important factors. The network age is determined by the minimal age of the essential modules. Hence, essential genes with fewer interactions are the limiting steps, if all interactions are assumed to have the same decay rate **λ**.
In addition to power-law distribution, other degree distributions for Pr(k) will be studied for comparisons.

In addition, we plan to use simulation to study network configuration on network aging, because simulation can easily relax some simple assumptions in the analytic approach. There are several ways to simulate power-law gene networks. The preferential attachment model is often used [83]. Alternatively, we can generate the degree distribution based on the power-law, and then pair interacting nodes, in a similar way to a network simulation approach previously used by us[5]. The parameter γ will be ranged from 1 to 3. Control studies will be conducted in networks with fixed numbers of interactions per gene, a Poisson distribution of degrees, and a log-normal distribution of degrees.

To simulate the error-tolerant features, we plan to choose various fractions of nodes with connectivity greater than K_{low} as essential genes. For controls, essential genes will be chosen randomly, or deliberately assigned to nodes with few connections.

We plan to characterize the aging dynamics through *G*, *R*, average and median lifespan, and tail distribution of lifespan. We plan to fit the simulated lifespans to Gompertz, Gompertz-Makeham, and Weibull distribution, and then evaluate them by Akai Information Index.

To simulate the lifespan of an individual cell using NRMCA as in Figure 1, we will first simulate the ages of all gene interactions based on exponential distributions. The age of each essential module is the maximum age of its gene interactions. The age of this individual cell is the minimum age of all of its essential modules. Simulations will be run for a population of cells to generate a survival curve. Three Spelman undergraduates, Ms. E. Dommond, J. Williams and J. Christopher, have written R codes to simulate cellular aging based on the basic model of NRMCA. Another Spelman student, Palpasa Manandhar, wrote some preliminary codes to study network aging in power-law networks with fixed constant failure rate. This preliminary simulation suggests that heterogeneity in degree distributions can dwarf the effect of heterogeneity in failure rates, at least in the parameter space that we tried.

Configuration heterogeneity will likely influence the tail distribution of lifespan, as suggested by a study based on the reliability model [89]. Plateau of mortality rate in late life is an empirical observation and has been attributed to population heterogeneity [90, 91]. It is worth discussing that power-law configuration implies rate-limiting step and is therefore connected to Aim 2. We will also study renewals and repairs in hub and non-hub nodes and compare their effects on network robustness.

Aim 1.2 Study structural reliability and aging process of the yeast protein interaction network, and compare them with permuted networks.

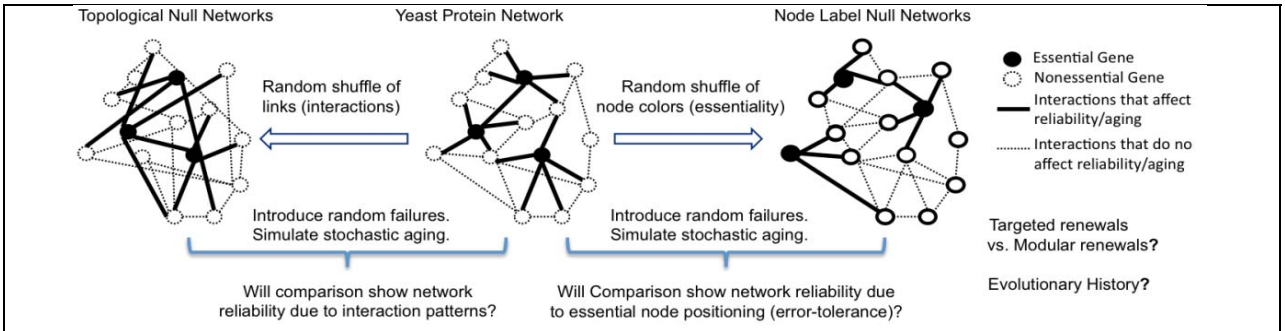


Figure 6, Aim 1.2. Evaluate network reliability and simulate aging process of the yeast protein interaction network.

How reliable and robust is the yeast protein interaction network? This is the first question that we will address here (Figure 6). The structural reliability of a network can be evaluated by a reliability function of the system states [17]. Each functional interaction is represented by 1, and failed one by 0. The combination of all the possible '1's and '0's in all interactions will then constitute the 'state space' for the entire network. How often a network fails in its state space is a measure of how reliable the network is.

Failures in NRMCA are caused by essential nodes, and failure of a single essential node leads to network failure. This indicates that we only need to check the active connectivity of essential nodes, and there is no need to search for traversal paths for the entire network connectivity. The computational complexity of identifying network failure in NRMCA is basically linear O(|V|), where |V| is the number of vertices or nodes. For comparison, the Dijkstra algorithm for a single source-node shortest path runs in O(|E|+|V|log|V|), where |E| is the number of edges. Therefore, it is much faster to calculate network reliability in our models than in other networks (for example, the internet and electric power grids). We will first implement the protein network in adjacency matrix to study basic principles, and then use linked lists to improve efficiency.

We will sample the state space by gradually introducing random interaction failures into networks with increasing failure frequency ε [85]. Plot of chance of network failures ~ ε can then describe network reliability. This kind of curves is often sigmoidal, and the middle point of the transition can be found as a critical transition point. Both the yeast protein network and null network models will be evaluated and compared.

The role of error tolerance can be studied by using null networks with shuffled node labels (Figure 6). The role of network topology can be studied by using null networks with shuffled links (Figure 6).

How does network configuration of yeast protein interaction work influence its aging process? This is the second question that we will address. We plan to simulate the stochastic network aging using the exponential decay function, which is just another way of modeling random failures. The exponential distributed component ages can be conveniently generated using the exponential random number generator. The age of each essential node is the maximum of its interaction ages, and the minimal age of the essential nodes is the network age. Survivor curves of yeast network aging and random networks will be compared. In comparison to the state-space-approach, different failure rates can be assigned to different interactions.

We will start with the yeast mitochondrial protein interaction network, a subset of the yeast protein network. Mitochondrion can be considered as an endosymbiotic prokaryotic cell. The simulation will then be extended to the entire observed yeast protein interaction network [81] (Figure 6).

It is probably expected that biological networks are more reliable than random networks. Alternatively, it is possible that biological networks have less 'pure' structural reliability when renewals/repairs are not considered. If this alternative scenario happened, it would offer us a golden opportunity to compare two different renewals/repair mechanisms: component renewals/repairs versus modular renewals (detailed in Aim 2.3).

Network evolutionary history may influence the reliability evaluation outcomes of the yeast protein interaction network. The PI's postdoctoral research showed that evolution of the yeast protein interaction network mirrors the universal tree of life, and interactions tend to occur between genes with similar evolutionary histories [5]. One option is to model interactions between genes with similar evolutionary histories as more reliable than others. Another option is to model reliability in proportion to their evolutionary history on the universal tree of life, based on our previous work [5].

The PI will partition network simulations into small coding projects for students and integrate network research into courses, parlaying his past experiences (detailed in section 6).

Aim 2. Develop methods to evaluate limiting interaction modules from experimental data, and study renewals/repairs in network aging.

We will first develop a general framework and apply it to yeast mutants (Aim 2.1), then focus on synthetic lethality as a special case (Aim 2.2), and finally study renewals/repairs (Aim 2.3).

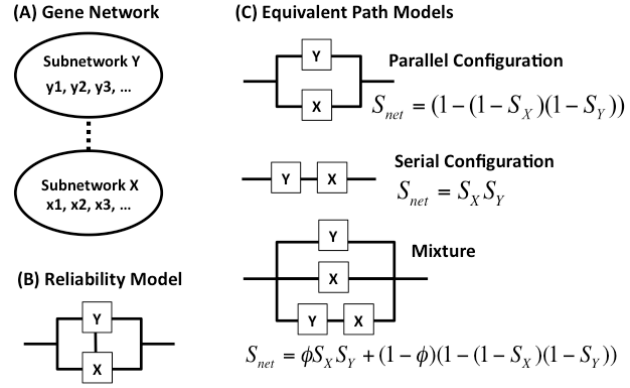


Figure 7, Aim 2.1. A general reliability framework to study network aging and gene interactions. (A) A gene network can be partitioned into two subnetworks, X and Y, each with many interacting genes x_i and y_i. (B) A classical reliability model for X and Y. (C) Network viability $S_{net}(t)$ can be found by the equivalent path models based on the reliability theories [17]. The survival functions for X and Y are S_X and S_Y. The two basic configurations are parallel and serial configurations of X and Y. The mixture of the two basic configuration is a general solution for $S_{net}(t)$ that will be used for developing likelihood methods.

Aim 2.1 Develop a general framework for limiting interaction modules on network aging, and apply it to understand network changes in yeast mutants.

It is important to understand how lifespan can be influenced by genetic alterations. We will address this question by studying how a network module, say module X that contains a list of genes, can influence network aging. In other words, we will study how module X becomes a limiting step in network aging. Figure 7 is our general framework to study a limiting network module X on network aging in the

context of the rest of the network, represented by Y. The general solution of the network survivor function is:

$$S_{net} = \phi S_X S_Y + (1-\phi)(1-(1-S_X)(1-S_Y)) \qquad \text{(Eq. 6)}$$

where ϕ is a mixing parameter. We will use this general solution to develop likelihood methods to analyze experimental data. If we let module Y represents the bulk of the network, S_Y can be modeled by the Gompertz survivor function. There are two special cases at ϕ=0 or 1. The case of ϕ=1 indicates a serial configuration, and ϕ=0 indicates a parallel configuration. It can be seen that limiting network modules in parallel and serial configurations will influence network aging in opposite ways. In serial configuration, the minimal lifespan of the limiting module will determine the network lifespan. In parallel configuration, the maximal lifespan of the limiting module will determine the network lifespan. Parallel configuration indicates redundant functions performed by network modules involved. Serial configuration indicates that network modules perform independent functions. The mixture of the two special cases describes the general case. It is worthy emphasizing that the serial configuration in our NRMCA is equivalent to the limiting path model proposed by Zuk et al. [92].

Our computing approach to apply the general framework in Figure 7 to fit experimental data is outlined in Figure 8. Here, we will ask a basic question here: Does a genetic alternation cause local changes within its interaction module or global network changes? The first scenario is represented by model M1 and second by model M2 in Figure 8. We plan to fit M1 and M2 to experimental lifespan data of yeast mutants and use AIC to evaluate the two models. For M1 model, we plan to approximate the survivor function for the bulk of network S_Y using the wildtype parameters. The network viability S_{net} can be defined by Eq. 6. Fitting of M1 to experimental data can be achieved by minimizing the residual sum of squared (RSS). For M2, we will use the yeast protein interaction network to approximate the bulk of network Y which will reduce free parameters to n_x, λ_x, p_x, λ_y, p_y, and ϕ. We can calculate Akaike information criterion AIC = sample_size x ln(RSS/sample_size)+2x parameter_number, and compare M1 and M2 (Figure 8). We will first use simulated networks with well-designed parameters to develop and evaluate the AIC test statistic, and then apply the method to experimental lifespan measures of yeast mutants based on yeast protein interaction network.

The mixing parameter ϕ describes how much 'unique' function that module X plays with regard to the rest of the network. For example, ϕ should be close to 0 when the deleted gene has a paralog with similar functions (See also Aim 2.2).

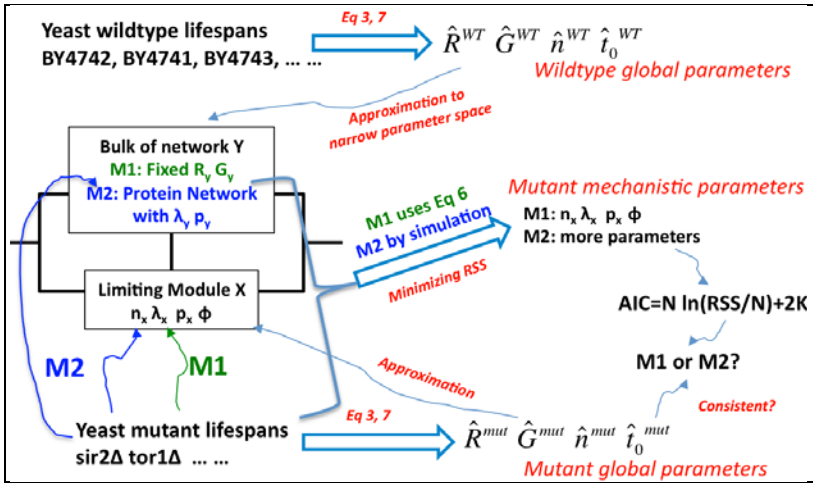


Figure 8, Aim 2.1, continued. Inferring network changes in yeast mutants. Model M1: A genetic alternation cause changes within its own interaction module. Model M2: A genetic alteration causes global network changes. We will first use theoretic network to develop and evaluate the AIC method. We will then apply them to yeast protein network and investigate yeast mutants. For fitting model M1 using the experimental data, the survival function of Y can be approximated by wildtype controls to narrow down parameter space.

In addition, we can estimate the observed $\hat{n}$ and $\hat{t}_0$ from experimental lifespan data using the binomial form of mortality rate and survivor function [1]:

$$\mu_{net}(t) = \alpha(t_0 + t)^{n-1} \quad \text{where} \quad t_0 = \frac{1-p}{p\lambda} \text{ and } \alpha \text{ is a constant} \qquad \text{(Eq 7a)}$$

$$S_{net}(t) = \exp\left(-\int_{t=0}^{t} \mu_{net}(t)\right) = \exp\left(\frac{\alpha({t_0}^n - (t+t_0)^n)}{n}\right) \qquad \text{(Eq 7b)}$$

We can find the probability density function as the product of μ_{net} and S_{net}, and develop maximum likelihood estimations. Although many simple assumptions are used to develop NRMCA, the estimated

$\hat{n}$ and $\hat{t}_0$ can be interpreted as the 'average' number of interactions per essential module and the apparent virtual age.

Our approach here is in the same spirit with population genetics studies. Population genetics parameters such as selection coefficient, mutation rate and effective population size are modeled with many assumptions, but they are nevertheless extremely useful to understand the empirical distributions of nucleotide substations. Likewise, our NRMCA can infer network changes from empirical lifespan distributions and offer mechanistic insights.

Both replicative and chronological lifespans of selected yeast mutants will be measured in the PI's lab. For replicative lifespan assays, a Singer micro-dissection instrument is request to improve the dissection efficacy. For chronological lifespan assays, we will use both flow cytometer method and CFU counting method. A full-time research technician is requested for the experimental effort and lab management. In addition, three yeast labs will share their experimental data with us (see letters of M. Kaeberlein, F. Tang, and J. Hartman).

Aim 2.2 Study how limiting synthetic lethality modules influence network aging.

Synthetical lethality refers to lethality due to double-deletions but not single deletions. Synthetically lethal gene pairs are predicted to have special properties in NRMCA. We will use a synthetically lethal gene pair X and Y as a general example (Figure 9). Because double-deletion of X and Y is lethal, X and Y must perform some essential functions that cannot be replaced by the rest of gene network (represented by Z in Figure 9). Hence, we can model X-Y as an independent network module in serial configuration to Z. The survivor functions of the wildtype S_{WT} and single deletion mutants $S_{x\Delta}$ and $S_{y\Delta}$ are the products of the modular survivor functions (Figure 9). Hence, S_Z will be canceled out in the ratio of S_{WT} / $S_{y\Delta}$, and this ratio is determined only by the modular survivor functions:

$$\frac{S_{WT}}{S_{y\Delta}} = \frac{S_{XY}}{S_X} = \frac{\sum_{k=1}^{n_x+n_y} \Pr(k)(1-(1-e^{\lambda t})^k)}{\sum_{k=1}^{n_x} \Pr(k)(1-(1-e^{\lambda t})^k)} \qquad (Eq\ 8)$$

If both X and Y have the same number of relative stable gene interactions, the above solution can be further simplified as $S_{WT}/S_{y\Delta} \approx (1-(\lambda t)^{n_x})$. Similar properties can be found for $S_{WT}/S_{x\Delta}$, and $S_{x\Delta}/S_{y\Delta}$.

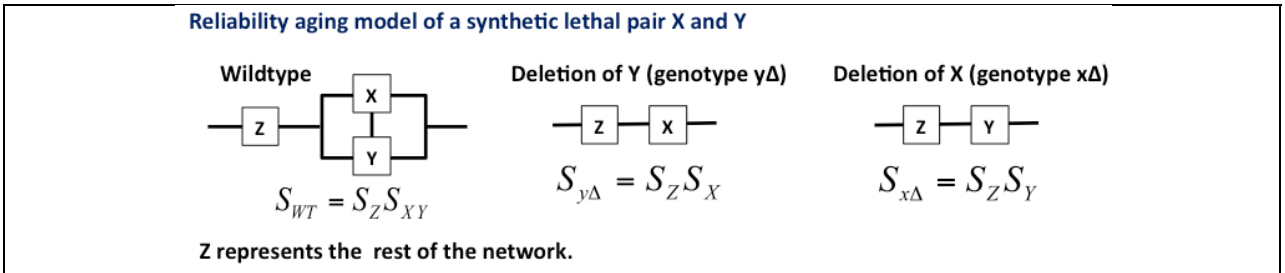


Figure 9. Network survivor functions for a synthetic lethal gene pair X and Y. Following yeast genetic conversion, deletion of gene X and gene Y are represented by lower-cases followed by Δ. Survivor function for Z, the bulk of the network, are canceled out in the ratios of $S_{WT}/S_{y\Delta}$, $S_{WT}/S_{x\Delta}$, and $S_{x\Delta}/S_{y\Delta}$.

The principle illustrated in Figure 9 offers us an opportunity to evaluate local network changes in single deletion mutants of a synthetic lethal gene pairs. Single deletion mutant *xΔ* and *yΔ* can often be obtained from the yeast deletion collection and their replicative lifespans will be measured in parallel to mitigate experimental variations. Technically, survival curves of both wildtype and deletion mutants can be fitted with Gompertz models. Their ratios can then be used to infer changes of the underlying gene interaction modules. If we approximate n_x and n_y using the yeast protein interaction networks, we would be able to estimate λ_x and λ_y and the binomial activation chance p_x, and p_y through numerical fitting.

Synthetic lethality is a special case for the general framework for limiting module in Aim 2.1. Hence, we will apply the general network fitting method as in Figure 8. Comparison of the mixing parameter ϕ in *xΔ*, *yΔ* and WT can tell us how much functional overlap exists between gene X and Y.

Similar to Aim 2.1, lifespan assays will be performed in both the PI's lab and collaborator's labs. One candidate synthetically lethal pair is the HSP90 duplicates in yeast: HSC82 and HSP82. The PI will guide students to pick more synthetic pairs through mini-research proposals (see section 6).

Aim 2.3. Study the role of renewals/repairs in network aging with focus on mitochondrial renewal.

Renewals and repairs play an important role in biological robustness. Self-repairing is also an important factor in several theories of aging [93-95]. Renewals/repairs can be imperfect or perfect. Implicitly, renewals/repairs can be modeled by decreasing the function decay rate λ. Explicitly, renewals/repairs can be modeled by a specified number of stand-by components, and the system fails as the last reserved component fails [17] (Figure 10A). To understand the basic principles of renewals/repairs on aging, we will focus on perfect replacements as in previous studies[96].

Renewals/repairs can target either individual molecules or the entire molecular machinery in a subnetwork (Figure 10B). The first scenario can be seen in the rescue of misfolded proteins by chaperon proteins, and the second can be seen by degradation of mitochondria by lysosome-like vacuole in yeast.

The specific biological question that we will focus on here is the mitochondrial renewals through lysosome-like vacuole [97]. Autophagy of mitochondria is a known lifespan extending strategy [97]. Is there a network reliability advantage of recycling disfunctional mitochondria through lysosome-like vacuole as compared to repairing the individual parts of mitochondria?

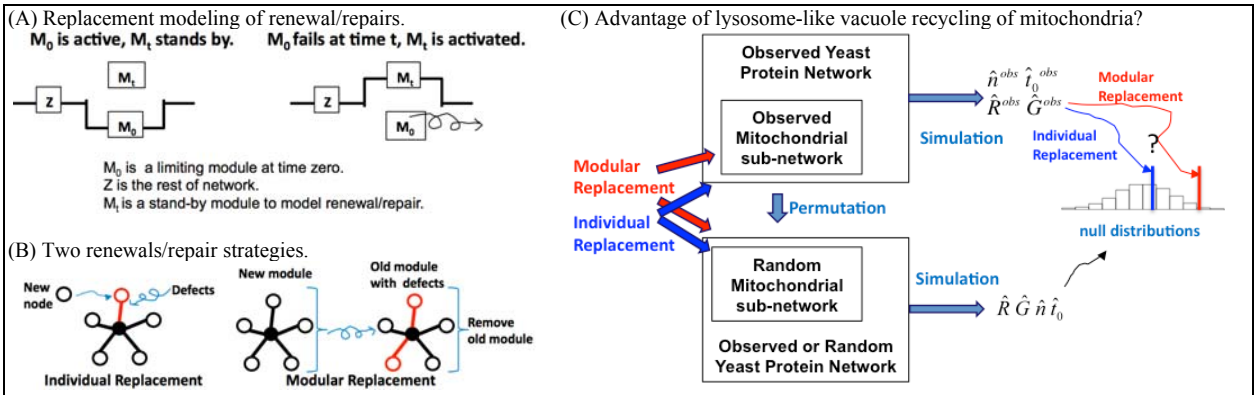


Figure 10, Aim 2.3. Study of renewals/repairs in NRMCA. (A) The initial limiting module M_0 is replaced by M_t at time t. Cost of repair can be introduced as a function of time. Replacement implies the renewal/repair process is instantaneous. (B) Two strategies for renewals/repairs. (C). A specific problem is to focus on mitochondria recycling by lysosome-like vacuole.

We will address this question by simulating the aging process of the yeast protein interaction network (Figure 10C). The mitochondrial subnetwork will be renewed/repaired by individual replacement or modular replacement. Random permutations of the protein network and the mitochondrial subnetwork will be applied to generate the null distributions. One important consideration is how much individual damages in mitochondrial subnetwork should be allowed before mitochondria are recycled by vacuoles. Intuitively, if the threshold is very low, modular replacement would lead to more reliable networks but likely with higher cost. If the threshold is very high, modular replacement would not prevent the network from collapse in time.

We expect that modular repairs make the network more homogenous than individual repairs. Hence, one advantage of recycling mitochondria through vacuole may be to minimize noises and increase network robustness. A drawback of this mitochondria recycling is probably that fewer cells will have extremely long lifespans due to smaller variance. This kind of decreased tail of longevity was argued in a reliability study with perfect replaceable component [89, 96].

Dr. Fusheng Tang has generated many vacuole mutants and measured their replicative lifespans [97, 98]. Dr. Tang has generously shared his data and strains with us (see his letter). Replicative and chronological lifespans of a few interesting mutants, such as atg15Δ and an OSH6 mutant with an altered promoter, will also be measured in PI's lab in order to increase they sample sizes. Heterogeneity of cell populations will be monitored by fluorescent probes, such as propidium iodide or DHE, and quantified by flow cytometer. We will apply the computing approaches in Aim 2.1 to evaluate these mutants and compare them with simulation studies.

Aim 3. Study network impact on aging as a quantitative trait.

Aim 3.1 One-locus and two-locus studies.

Missing heritability is a problem that we will focus on here [92]. Genome-wide association (GWA) studies have revealed many loci for complex human traits, however, these loci can only explain a small proportion of the "total heritability" of the traits[92]. Zuk and colleagues from the Lander group proposed a

limiting pathway (LP) model and found that LPs can inflate the estimation of "total heritability". The pathways in LP models are basically modules in NRMCA. Zuk et al. defined the limiting pathways using the minimal value of a trait [92]. Hence, the LP model is basically a serial configuration of limiting modules in NRMCA (see Figure 7 and Figure 11).

The general gene interaction framework in Figure 7 offers us an opportunity to study lifespan as a quantitative trait. We will start with one-locus and two-locus models in serial configurations to investigate the basic principles (Figure 11). For simplicity, we will assume two alleles X1 and X2 at locus X with similar or redundant roles. Hence, X1 and X2 are in parallel configuration (Figure 11). Network viability can be obtained by grouping the rest network into a super-module Z. Both the mean network lifespan E(t) and variance Var(t) can be calculated using the reliability approach (Figure 11). So are the mean and variance for the modular lifespans of X and Y. Variance of network lifespan is the 'total phenotypic variance' and is the 'total heritability' when error variance is not considered. Variance of X and Y modular lifespans are the 'known heritability'. The 'explained heritability' is the ratio of 'known' versus 'total heritability', which should be 100% for perfectly additive linear models [4, 92].

Locus X and Y may also be in parallel or mixed configurations as in Figure 7. These non-serial network configurations would likely lead to statistical 'epistasis' between X and Y in the linear model.

In addition, we will conduct empirical study of heritability at a single locus underlying lifespan segregation. The PI's lab has found a single locus that underlying 2:2 lifespan segregation in a heterozygous wild isolate M5. We will apply model-fitting approach from Aim 2.1, and use ϕ to figure out whether serial or parallel configuration is the better model for this locus. In collaboration with the Kaeberlein lab, we are trying to identify this locus by sequencing pooled segregants (see his letter). Identification of this single locus may lead to biological insights that can cross-examine the mathematical modeling results.

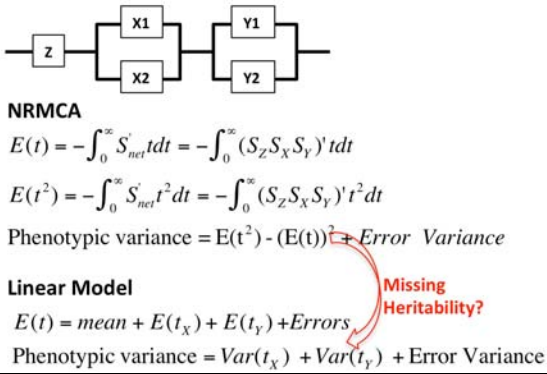


Figure 11, Aim 3.1. A model of two independent loci on lifespan in a diploid background. Probability density function of ages is the derivative $(1-S_{net})' = -S'_{net}$. Phenotypic variance of lifespan can be obtained through the first and second moments. QTL model in haploid background is basically the same as in figure 7. Additive linear model is often used in quantitative genetic studies. Discrepancy between NRMCA and linear model is applicable to address the missing heritability problem. This serial NRMCA model is equivalent to the limiting path model [92], but NRMCA provides an explicit form of variance for lifespan. NRMCA also can describe loci in parallel configuration and generalize them in mixture models.

Aim 3.2 Multi-locus study and implications on natural lifespan variation

How many limiting network modules or loci should we expect to explain yeast natural lifespan variation? The PI and two Spelman students have preliminarily measured chronological lifespan of 33 Sanger yeast strains using propidium iodide and flow cytometer at Princeton during the summer of 2011. A preliminary GWA study showed that linear model has extremely poor statistical power, but Fisher's exact test on genotype-counts for long-, short- and average CLS categories gave 37K 'positive' SNPs out of 200K SNPs in the dataset, using a false discovery rate of 10%. Many of the 'positive' SNPs are located in biological meaningful genes, but a large number of them are likely caused by population structure, similar to another study [99]. From the network perspective, the actual positive SNPs would belong to different limiting network modules.

We aim to use simulation to evaluate the number of potential network modules or loci involved in yeast natural lifespan variation based on the yeast protein interaction network (Figure 12). The candidate limiting loci will be chosen based on a range of empirical evidences: published screens, on-going screens in the Kaeberlein and Hartman labs, known genes on aging, candidate loci from preliminary GWAS, and new hypotheses generated in undergraduate research proposals. In addition, we will use simulated networks to precisely control the number of limiting network modules and evaluate the GWA methods.

We plan to sample K number of limiting genes from the candidates. Each gene can have 2 alleles, which will randomly have fast, slow, or average failure rates for its interactions. The three categories of

failure rates will be based on estimates of the yeast deletion mutants. We plan to randomly pick 30 ~ 500 genotypes ("strains") out of 3^K possible genotypes. For each "strain", we will simulate aging for $N=10^3$ cells to estimate average lifespan. Population structure can be introduced by low levels of mutations in a particular strain. We will run simulations with K in the range of [1,1000]. We will use quantile-quantile plots to compare the simulated and observed lifespan distributions. By sampling the parameter space, our simulations can find the range of K that will most often generate lifespan distributions similar to empirical observations.

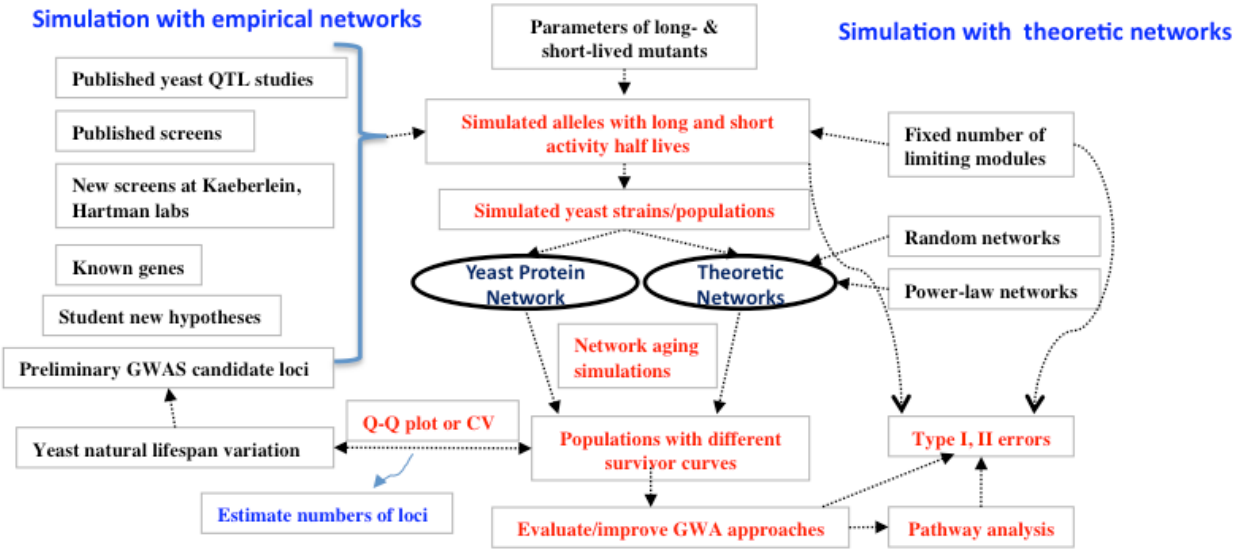


Figure 12. Aim 3.2. Multi-locus model of network aging to study natural lifespan variation. This study can be used to guide future genome-wide association or quantitative trait studies.

The simulated data will be used to compare linear regression method and association test of categorical data, the two methods that we found to work different in GWA study. We will also evaluate standard practices to deal with populations structures, including first principal component of genotypes [100] and the linear mixed model based EMMA [101]. Because of the simulated 'truth', we can compare the type I and type II errors of different methods (see Figure 12).

For both simulated and empirical data, we will identify network modules that are enriched with SNPs statistically associated with lifespan variation, using the yeast protein interaction network as reference. One option is to permute single-gene statistic to generate null distribution, which is analogous to the PI's previous work on network evolution [5]. Alternatively, we can apply 'standard' pathway-association tests, such as gene set enrichment, SNP ratio test, GSA-SNP, ALIGATOR, as reviewed recently [102].

For the preliminary GWA study on chronological lifespan, we plan to use the 150 sequenced yeast strains that we have collected. We will rely on detection methods that are found to be useful by simulation. Alternatively, we plan to choose 'unstructured' strains based on neighbor-joining tree calculated from whole genome genetic variations as in previous work [99, 103, 104]. Association tests on genotype-counts of categorical phenotypic data can be applied using 'unstructured' strains.

Overall, our aim is to estimate the number of limiting network modules involved in yeast natural lifespan variation, evaluate and revise GWA methods, and provide guidelines for QTL and GWA studies.

Long-term directions and career goals: Beyond this five-year plan, the PI plans to use NRMCA to study evolution in age-structured populations and network impact of calorie restriction. The PI will also extend the application of network aging to other organisms.

5. Results from prior NSF support

5.1 NSF CCLI Award # 0837075, $110K, 1/1/2009-12/31/2010, (Qin as PI 1/1/2009-12/31/2009), "Computing in Life Sciences through Hands-on Experience and Case Studies at Tuskegee University". With this educational grant, Qin developed materials for teaching computing in biology, available at http://www.bioinformatics.org/ctls, and had a peer-reviewed teaching publication [12]. Qin has carried out

this project beyond the support of the initial grant. Recently developed materials are at the PI's YouTube [11] and GitHub pages [105].

Contributions to the development of Human Resources in STEM: This grant has enabled Qin to develop and offer teaching materials on computing in biology at Tuskegee University (two course-offers and one three-day faculty workshop with 32 participants), Spelman College (three course offers), Lewis Clark College (one faculty workshop with 7 participants), Delaware State University (one guest lecture with computer lab of ~20 students), and Alabama State University (one workshop for ~15 participants). Four of these institutions are HBCUs.

5.2 NSF RUI Award # 1022294, $293K, 9/1/2010-8/31/2013, PI Qin, "Testing the network hypothesis of cellular aging in *Saccharomyces cerevisiae*". With this RUI, Qin developed the basic NRMCA, and collected and analyzed empirical data to support the network hypothesis (see also section 3). This RUI has led to 1 peer-reviewed research publication [2], 1 manuscript in review, 1 undergraduate honor thesis, and 2 other manuscripts (Underlined authors are current and former Spelman students).

- GUO, Z., A. B. ADOMAS, E. D. JACKSON, H. QIN and J. P. TOWNSEND, 2011 SIR2 and other genes are abundantly expressed in long-lived natural segregants for replicative aging of the budding yeast *Saccharomyces cerevisiae*. FEMS Yeast Res 11: 345-355.
- Qin, H. A network model of cellular aging. In review, see http://arxiv.org/pdf/1305.5784.
- Parnell, L, Undergraduate honor thesis, Study the links between oxidative stress, genomic instability, and cellular aging, May 2012.
- Parnell, L., E. D. Jackson, Parker, M., J. Rodrigues, N. Gupta, B. Mohanty, H. Qin, Hydrogen peroxide induced loss of heterozygosity offers insight on mitotic asymmetry and chronological aging in *Saccharomyces cerevisiae*. In preparation. (Poster, 2012 yeast genetics annual meeting).
- Alexander, A, Montgomery, C. K. Matheson, O. Morrison, A. Story, H. Qin, An analysis of genomic features associated with aging in *S. cerevisiae*. In preparation. (Poster, 2011 SMBE meeting by A. Story in Kyoto Japan).

Contributions to the development of Human Resources in STEM: This RUI has directly supported Qin to mentor over 40 African American female undergraduates to conduct computational and experimental research on cellular aging, with 8 of them now pursuing PhDs in STEM. This project impacted two courses: BIO386 "Genomics, Proteomics, and Bioinformatics' and BIO233 "Microbiology" with a combined total of ~30 enrolled students yearly. In BIO386, students used R to conduct computational genomics projects. In BIO233, students wrote mini research proposals.

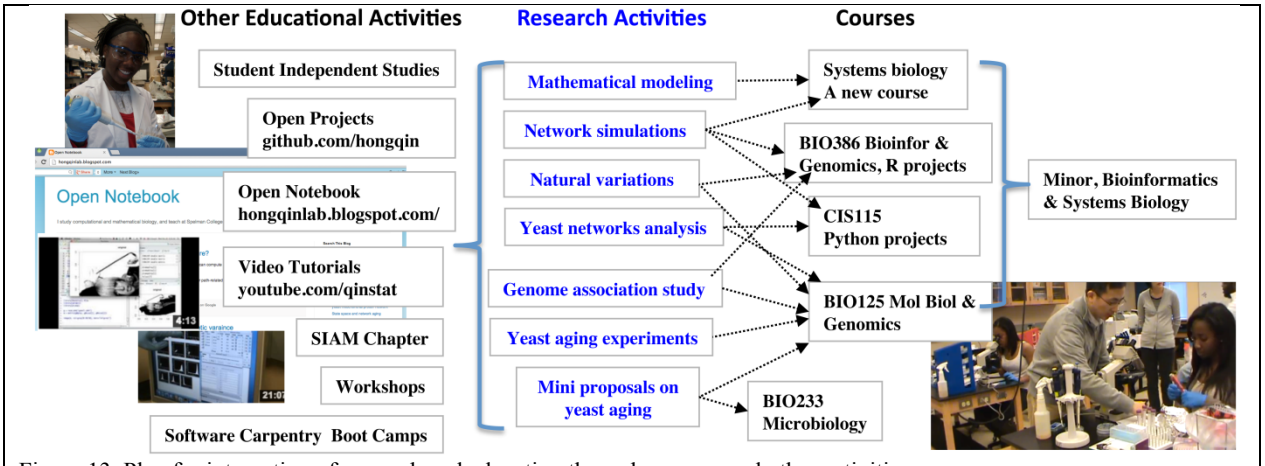


Figure 13. Plan for integration of research and education through courses and other activities.

6. Educational Plan (Broader impact)

Goal: Integrated training of minority students in mathematical, computational, & network biology.

This project will be carried out at Spelman College, a historically black college for women. The overall educational plan (Figure 13) is strongly supported by the Department (see Chair Mark Lee's letter).

6.1. Integrate network biology into existing and proposed new courses.

The PI is experienced in integrating original research into courses (see Chair's letter). As a tenure track faculty at Spelman, the PI teaches 9 credit/contact hours per semester, which provides excellent opportunity for interactions with students. PI will integrate this project into 4 existing and 1 proposed new courses (Figure 13 and Table 1).

Courses	Aims						
	1.1	1.2	2.1	2.2	2.3	3.1	3.2
BIO386	X	X	X	X	X	X	X
CIS115	X	X			X		X
BIO125			X	X			X
BIO233			X	X			X
SysBio	X	X	X	X	X		

Table 1, Mapping between courses and Aims. A course on systems biology will be proposed.

The PI plans to design R and Python coding projects for network simulation and analysis in BIO386 Genomics, Proteomics, and Bioinformatics and CIS115 Introduction to Computing and Informatics.

The PI will lead students to work on mini research proposals on network aging in BIO125 Molecular Biology and Genomics and BIO233 Microbiology. Students will analyze yeast protein networks, read primary research papers, and conduct bioinformatics analysis on yeast genes. Students will be guided to formulate hypothesis on potential mechanisms, students with interesting ideas will be invited to test their hypothesis in the PI's lab.

The PI will propose to the curriculum committee a new introductory course on systems biology that will expose many biology students to the power of mathematical and computational biology. This course will cover basic concepts of dynamic systems, including toggle switch and cell cycles. The PI plans to pilot this course in Year 2 and 3, and proposes to the curriculum committee in Year 4.

6.2. Cross-disciplinary training of minority undergraduates through independent studies.

Students will be trained in mathematical modeling, computational simulation, genomics, next-generation sequence analysis, experimental biology, and flow cytometry. The interdisciplinary nature and genome-scale studies of this project will add a sense of excitement to students. The PI will strive to instill the fundamental values and ethics of research into the students. By taking ownerships of a unique body of data and knowledge through original research, students may discover their love of research and may choose science-related careers. The PI is experienced in training undergraduate researchers for both experimental and computing skills. Data from aging experiments and flow cytometry are routinely analyzed by students using R scripts. The PI has trained over 40 Spelman students in three years, among which four are math majors, two are computer science majors, two have won competitive travel awards to attend international meetings, and 8 of them are pursuing PhDs in STEM. The PI requests funds to support 2 semester students and 2 summer students for this project. The PI will also mentor students from various research programs at Spelman, including HHMI, RISE, and Math RAMP.

6.3. Build a sustainable mathematical and computational biology program through workshops, a new minor, and a student SIAM chapter.

The primary goal here is to build a local community of mathematical modeling and computing, in order to better recruit and retain undergraduates, and go beyond the effort of the PI as a single faculty. The PI was instrumental in establishing a student Chapter of the Society of Industrial and Applied Mathematics (SIAM) in 2012. As the faculty advisor, the PI will work with students and faculty across disciplines. PI will organize faculty and student tutorial workshops to build a critical mass of computing and modeling in research and education. The PI will also organize two coding boot camps with Software Carpentry (see Greg Wilson's letter).

The PI plans to propose an undergraduate minor of bioinformatics and systems biology to Spelman College in Year 4. CIS115, BIO386, and the new course on systems biology will likely be part of the minor (see Chair's letter). The PI will also further develop tutorial videos at youtube.com/qinstat, open research projects at Github.com/hongqin, and share project progress with scientific community through open notebook at hongqinlab.blogspot.com.

6.4. Assessment of educational outcomes.

The PI will frequently discuss with the Department Chair and other senior faculty to evaluate the educational progress, especially about the effectiveness of courses involved. Faculty are evaluated annually on research, education and service at the Spelman Biology Department. Spelman College routinely evaluates the quality of courses through student surveys. Biology Department keeps records of graduated students. We will conduct surveys and solicit feedback from workshop participants. The PI will use LinkedIn and Facebook to track student career paths in order to gauge the long-term impact of PI's educational activities.

References

1. Qin, H. *A network model for cellular aging.* In review; Available from: http://arxiv.org/pdf/1305.5784.pdf.
2. Guo, Z., A.B. Adomas, E.D. Jackson, H. Qin, and J.P. Townsend, *SIR2 and other genes are abundantly expressed in long-lived natural segregants for replicative aging of the budding yeast Saccharomyces cerevisiae.* FEMS Yeast Res, 2011. **11**(4): p. 345-55.
3. Qin, H., M. Lu, and D.S. Goldfarb, *Genomic instability is associated with natural life span variation in Saccharomyces cerevisiae.* PLoS One, 2008. **3**(7): p. e2670.
4. Qin, H. and M. Lu, *Natural variation in replicative and chronological life spans of Saccharomyces cerevisiae.* Exp Gerontol, 2006. **41**(4): p. 448-56.
5. Qin, H., H.H. Lu, W.B. Wu, and W.H. Li, *Evolution of the yeast protein interaction network.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12820-4.
6. Qin, H. and L. Yang, *Detection of changes in transitive associations by shortest-path analysis of protein interaction networks integrated with gene expression profiles.* Proceedings of the IEEE International Conference on Biomedical Engineering and Informatics, 2008. **1**: p. 418-423.
7. Qin, H., W.B. Wu, J.M. Comeron, M. Kreitman, and W.H. Li, *Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes.* Genetics, 2004. **168**(4): p. 2245-60.
8. Kravets, A., H. Qin, A. Ahmad, G. Bethlendy, Q. Gao, and E. Rustchenko, *Widespread occurrence of dosage compensation in Candida albicans.* PLoS ONE, 2010. **5**(6): p. e10856.
9. Qin, H. and A. Driks, *Contrasting evolutionary patterns in the Spore coat of of groups of Bacillus species.* BMC Evolution, Submitted, in revision. .
10. Gilchrist, M.A., H. Qin, and R. Zaretzki, *Modeling SAGE tag formation and its effects on data interpretation within a Bayesian framework.* BMC Bioinformatics, 2007. **8**: p. 403.
11. Qin, H. *Hong Qin's YouTube educational channel*. Available from: http://youtube.com/qinstat/.
12. Qin, H., *Teaching computational thinking through bioinformatics to biology students.* Proceedings of the 40th ACM technical symposium on computer science education, 2009: p. 188-191.
13. Finch, C.E., *Longevity, Senescence, and the Genome*. 1990, Chicago: The University of Chicago Press.
14. Harman, D., *Aging: a theory based on free radical and radiation chemistry.* J Gerontol, 1956. **11**(3): p. 298-300.
15. Williams, G.C., *Pleiotropy, natural selection and the evolution of senescence.* Evolution, 1957. **11**: p. 398-411.
16. Barlow, R.E. and F. Proschan, *Mathematical theory of reliability*. 1996, Philadelphia: Society for Industrial and Applied Mathematics.
17. Leemis, L.M., *Reliability: Probabilistic models and statistical methods. 2nd ed.* . 2009: Lawrence Leemis.
18. Gumbel, E.J., *Statistics of Extremes*. 1958, New York: Columbia University Press.
19. Gavrilov, L.A. and N.S. Gavrilova, *The reliability theory of aging and longevity.* J Theor Biol, 2001. **213**(4): p. 527-45.
20. Gavrilov, L.A. and N.S. Gavrilova, *The biology of life span: A quantitative approach*, ed. V.P. Skulacheve. 1991: Harwood Academic Publishers.

21. De Paepe, M. and F. Taddei, *Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages.* PLoS Biol, 2006. **4**(7): p. e193.
22. Gompertz, B., *On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies.* Philosophical Transactions of the Royal Society of London, 1825. **115**: p. 513-585.
23. Gavrilov, L.A. and N.S. Gavrilova, *Evolutionary theories of aging and longevity.* ScientificWorldJournal, 2002. **2**: p. 339-56.
24. Gavrilov, L.A. and N.S. Gavrilova, *The quest for a general theory of aging and longevity.* Sci Aging Knowledge Environ, 2003. **2003**(28): p. RE5.
25. Stewart, E. and F. Taddei, *Aging in Escherichia coli: signals in the noise.* Bioessays, 2005. **27**(9): p. 983.
26. Stewart, E.J., R. Madden, G. Paul, and F. Taddei, *Aging and death in an organism that reproduces by morphologically symmetric division.* PLoS Biol, 2005. **3**(2): p. e45.
27. Strehler, B.L. and A.S. Mildvan, *General theory of mortality and aging.* Science, 1960. **132**: p. 14-21.
28. Henderson, K.A. and D.E. Gottschling, *A mother's sacrifice: what is she keeping for herself?* Curr Opin Cell Biol, 2008. **20**(6): p. 723-8.
29. Ackermann, M., L. Chao, C.T. Bergstrom, and M. Doebeli, *On the evolutionary origin of aging.* Aging Cell, 2007. **6**(2): p. 235-44.
30. Blagosklonny, M.V., *Aging: ROS or TOR.* Cell Cycle, 2008. **7**(21): p. 3344-54.
31. Mortimer, R.K. and J.R. Johnston, *Life span of individual yeast cells.* Nature, 1959. **183**(4677): p. 1751-2.
32. Kaeberlein, M., *Lessons on longevity from budding yeast.* Nature, 2010. **464**(7288): p. 513-9.
33. Fabrizio, P. and V.D. Longo, *The chronological life span of Saccharomyces cerevisiae.* Methods Mol Biol, 2007. **371**: p. 89-95.
34. Sinclair, D.A., S.J. Lin, and L. Guarente, *Life-span extension in yeast.* Science, 2006. **312**(5771): p. 195-7; author reply 195-7.
35. Park, P.U., M. McVey, and L. Guarente, *Separation of mother and daughter cells.* Methods Enzymol, 2002. **351**: p. 468-77.
36. Laun, P., S. Buttner, M. Rinnerthaler, W.C. Burhans, and M. Breitenbach, *Yeast aging and Apoptosis.* Subcell Biochem, 2012. **57**: p. 207-32.
37. Guarente, L. and C. Kenyon, *Genetic pathways that regulate ageing in model organisms.* Nature, 2000. **408**(6809): p. 255-62.
38. Hekimi, S. and L. Guarente, *Genetics and the specificity of the aging process.* Science, 2003. **299**(5611): p. 1351-4.
39. Koubova, J. and L. Guarente, *How does calorie restriction work?* Genes Dev, 2003. **17**(3): p. 313-21.
40. Kaeberlein, M., C.R. Burtner, and B.K. Kennedy, *Recent developments in yeast aging.* PLoS Genet, 2007. **3**(5): p. e84.
41. Kennedy, B.K., K.K. Steffen, and M. Kaeberlein, *Ruminations on dietary restriction and aging.* Cell Mol Life Sci, 2007. **64**(11): p. 1323-8.
42. Finkel, T., M. Serrano, and M.A. Blasco, *The common biology of cancer and ageing.* Nature, 2007. **448**(7155): p. 767-74.
43. Steffen, K.K., B.K. Kennedy, and M. Kaeberlein, *Measuring replicative life span in the budding yeast.* J Vis Exp, 2009(28).

44. Fabrizio, P. and V.D. Longo, *The chronological life span of Saccharomyces cerevisiae.* Aging Cell, 2003. **2**(2): p. 73-81.
45. Sinclair, D.A. and L. Guarente, *Extrachromosomal rDNA circles--a cause of aging in yeast.* Cell, 1997. **91**(7): p. 1033-42.
46. Ganley, A.R., S. Ide, K. Saka, and T. Kobayashi, *The effect of replication initiation on gene amplification in the rDNA and its relationship to aging.* Mol Cell, 2009. **35**(5): p. 683-93.
47. Blagosklonny, M.V., *Aging-suppressants: cellular senescence (hyperactivation) and its pharmacologic deceleration.* Cell Cycle, 2009. **8**(12): p. 1883-7.
48. Wei, M., P. Fabrizio, F. Madia, J. Hu, H. Ge, L.M. Li, and V.D. Longo, *Tor1/Sch9-regulated carbon source substitution is as effective as calorie restriction in life span extension.* PLoS Genet, 2009. **5**(5): p. e1000467.
49. Steinkraus, K.A., M. Kaeberlein, and B.K. Kennedy, *Replicative aging in yeast: the means to the end.* Annu Rev Cell Dev Biol, 2008. **24**: p. 29-54.
50. Kirkwood, T.B. and C.E. Finch, *Ageing: the old worm turns more slowly.* Nature, 2002. **419**(6909): p. 794-5.
51. Smith, D.L., Jr., J.M. McClure, M. Matecic, and J.S. Smith, *Calorie restriction extends the chronological lifespan of Saccharomyces cerevisiae independently of the Sirtuins.* Aging Cell, 2007. **6**(5): p. 649-62.
52. Lin, S.J., M. Kaeberlein, A.A. Andalis, L.A. Sturtz, P.A. Defossez, V.C. Culotta, G.R. Fink, and L. Guarente, *Calorie restriction extends Saccharomyces cerevisiae lifespan by increasing respiration.* Nature, 2002. **418**(6895): p. 344-8.
53. Laun, P., M. Rinnerthaler, E. Bogengruber, G. Heeren, and M. Breitenbach, *Yeast as a model for chronological and reproductive aging - a comparison.* Exp Gerontol, 2006. **41**(12): p. 1208-12.
54. Powers, R.W., 3rd, M. Kaeberlein, S.D. Caldwell, B.K. Kennedy, and S. Fields, *Extension of chronological life span in yeast by decreased TOR pathway signaling.* Genes Dev, 2006. **20**(2): p. 174-84.
55. Kaeberlein, M., R.W. Powers, 3rd, K.K. Steffen, E.A. Westman, D. Hu, N. Dang, E.O. Kerr, K.T. Kirkland, S. Fields, and B.K. Kennedy, *Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients.* Science, 2005. **310**(5751): p. 1193-6.
56. Kwan, E.X., E. Foss, L. Kruglyak, and A. Bedalov, *Natural polymorphism in BUL2 links cellular amino acid availability with chronological aging and telomere maintenance in yeast.* PLoS Genet, 2011. **7**(8): p. e1002250.
57. Franceschi, C., *Cell proliferation, cell death and aging.* Aging (Milano), 1989. **1**(1): p. 3-15.
58. Kirkwood, T.B. and A. Kowald, *Network theory of aging.* Exp Gerontol, 1997. **32**(4-5): p. 395-9.
59. Kowald, A. and T.B. Kirkwood, *A network theory of ageing: the interactions of defective mitochondria, aberrant proteins, free radicals and scavengers in the ageing process.* Mutat Res, 1996. **316**(5-6): p. 209-36.
60. Soti, C. and P. Csermely, *Aging cellular networks: chaperones as major participants.* Exp Gerontol, 2007. **42**(1-2): p. 113-9.
61. Xue, H., B. Xian, D. Dong, K. Xia, S. Zhu, Z. Zhang, L. Hou, Q. Zhang, Y. Zhang, and J.D. Han, *A modular network model of aging.* Mol Syst Biol, 2007. **3**: p. 147.

62. Budovsky, A., A. Abramovich, R. Cohen, V. Chalifa-Caspi, and V. Fraifeld, *Longevity network: construction and implications.* Mech Ageing Dev, 2007. **128**(1): p. 117-24.
63. Smith, E.D., M. Tsuchiya, L.A. Fox, N. Dang, D. Hu, E.O. Kerr, E.D. Johnston, B.N. Tchao, D.N. Pak, K.L. Welton, D.E. Promislow, J.H. Thomas, M. Kaeberlein, and B.K. Kennedy, *Quantitative evidence for conserved longevity pathways between divergent eukaryotic species.* Genome Res, 2008. **18**(4): p. 564-70.
64. Smith, E.D., B.K. Kennedy, and M. Kaeberlein, *Genome-wide identification of conserved longevity genes in yeast and worms.* Mech Ageing Dev, 2007. **128**(1): p. 106-11.
65. Lorenz, D.R., C.R. Cantor, and J.J. Collins, *A network biology approach to aging in yeast.* Proc Natl Acad Sci U S A, 2009. **106**(4): p. 1145-50.
66. Barea, F. and D. Bonatto, *Aging defined by a chronologic-replicative protein network in Saccharomyces cerevisiae: an interactome analysis.* Mech Ageing Dev, 2009. **130**(7): p. 444-60.
67. Promislow, D.E., *Protein networks, pleiotropy and the evolution of senescence.* Proc Biol Sci, 2004. **271**(1545): p. 1225-34.
68. Rang, C.U., A.Y. Peng, A.F. Poon, and L. Chao, *Ageing in Escherichia coli requires damage by an extrinsic agent.* Microbiology, 2012. **158**(Pt 6): p. 1553-9.
69. Kirkwood, T.B., *Evolution of ageing.* Nature, 1977. **270**(5635): p. 301-4.
70. Wagner, A., *The role of robustness in phenotypic adaptation and innovation.* Proc Biol Sci, 2012. **279**(1732): p. 1249-58.
71. Masel, J. and M.L. Siegal, *Robustness: mechanisms and consequences.* Trends Genet, 2009. **25**(9): p. 395-403.
72. Masel, J. and M.V. Trotter, *Robustness and evolvability.* Trends Genet, 2010. **26**(9): p. 406-14.
73. Levy, S.F. and M.L. Siegal, *Network hubs buffer environmental variation in Saccharomyces cerevisiae.* PLoS Biol, 2008. **6**(11): p. e264.
74. Wagner, A., *Robustness against mutations in genetic networks of yeast.* Nat Genet, 2000. **24**(4): p. 355-61.
75. Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T.L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J.P. Latge, J.M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami, and S. Morishita, *High-dimensional and large-scale phenotyping of yeast mutants.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19015-20.
76. Zhou, C., B.D. Slaughter, J.R. Unruh, A. Eldakak, B. Rubinstein, and R. Li, *Motility and segregation of hsp104-associated protein aggregates in budding yeast.* Cell, 2011. **147**(5): p. 1186-96.
77. Liu, B., L. Larsson, A. Caballero, X. Hao, D. Oling, J. Grantham, and T. Nystrom, *The polarisome is required for segregation and retrograde transport of protein aggregates.* Cell, 2010. **140**(2): p. 257-67.
78. Managbanag, J.R., T.M. Witten, D. Bonchev, L.A. Fox, M. Tsuchiya, B.K. Kennedy, and M. Kaeberlein, *Shortest-path network analysis is a useful approach toward identifying genetic determinants of longevity.* PLoS One, 2008. **3**(11): p. e3802.
79. Schroedinger, E., *What is life?* 1944, Cambridge: Cambridge University Press.
80. Di Talia, S., J.M. Skotheim, J.M. Bean, E.D. Siggia, and F.R. Cross, *The effects of molecular noise and size control on variability in the budding yeast cell cycle.* Nature, 2007. **448**(7156): p. 947-51.

81. Costanzo, M., A. Baryshnikova, J. Bellay, Y. Kim, E.D. Spear, C.S. Sevier, H. Ding, J.L. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R.P. St Onge, B. VanderSluis, T. Makhnevych, F.J. Vizeacoumar, S. Alizadeh, S. Bahr, R.L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.Y. Lin, W. Liang, M. Marback, J. Paw, B.J. San Luis, E. Shuteriqi, A.H. Tong, N. van Dyk, I.M. Wallace, J.A. Whitney, M.T. Weirauch, G. Zhong, H. Zhu, W.A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F.P. Roth, G. Giaever, C. Nislow, O.G. Troyanskaya, H. Bussey, G.D. Bader, A.C. Gingras, Q.D. Morris, P.M. Kim, C.A. Kaiser, C.L. Myers, B.J. Andrews, and C. Boone, *The genetic landscape of a cell.* Science, 2010. **327**(5964): p. 425-31.
82. Aldana, M. and P. Cluzel, *A natural class of robust networks.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8710-4.
83. Barabasi, A.L. and R. Albert, *Emergence of scaling in random networks.* Science, 1999. **286**(5439): p. 509-12.
84. Barabasi, A.L. and Z.N. Oltvai, *Network biology: understanding the cell's functional organization.* Nat Rev Genet, 2004. **5**(2): p. 101-13.
85. Albert, R., H. Jeong, and A.L. Barabasi, *Error and attack tolerance of complex networks.* Nature, 2000. **406**(6794): p. 378-82.
86. Doyle, J.C., D.L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, *The "robust yet fragile" nature of the Internet.* Proc Natl Acad Sci U S A, 2005. **102**(41): p. 14497-502.
87. Maslov, S. and K. Sneppen, *Specificity and stability in topology of protein networks.* Science, 2002. **296**(5569): p. 910-3.
88. Maslov, S. and I. Ispolatov, *Propagation of large concentration changes in reversible protein-binding networks.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13655-60.
89. Finkelstein, M. and J.W. Vaupel, *The relative tail of longevity and the mean remaining lifetime.* DEMOGRAPHIC RESEARCH, 2006. **14**: p. 111-138.
90. Vaupel, J.W., K.G. Manton, and E. Stallard, *The impact of heterogeneity in individual frailty on the dynamics of mortality.* Demography, 1979. **16**(3): p. 439-54.
91. Vaupel, J.W., J.R. Carey, K. Christensen, T.E. Johnson, A.I. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y. Zeng, K.G. Manton, and J.W. Curtsinger, *Biodemographic trajectories of longevity.* Science, 1998. **280**(5365): p. 855-60.
92. Zuk, O., E. Hechter, S.R. Sunyaev, and E.S. Lander, *The mystery of missing heritability: Genetic interactions create phantom heritability.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
93. Kirkwood, T.B., *The nature and causes of ageing.* Ciba Found Symp, 1988. **134**: p. 193-207.
94. Horiuchi, S., *Interspecies differences in the life span distribution: humans versus invertebrates.* Polulation and Development Review, 2003. **29**(Supplement: Life Span: Evolutionary, Ecological, and Demographic Perspectives): p. 127-151.
95. Yashin, A.I., I.A. Iachine, and A.S. Begun, *Mortality modeling: a review.* Mathematical Population Studies, 2000. **8**: p. 305-332.
96. Vaupel, J.W., *Post-Darwinian longevity.* Polulation and Development Review, 2003. **29**(Supplement: Life Span: Evolutionary, Ecological, and Demographic Perspectives): p. 127-151.

97. Tang, F., J.W. Watkins, M. Bermudez, R. Gray, A. Gaban, K. Portie, S. Grace, M. Kleve, and G. Craciun, *A life-span extending form of autophagy employs the vacuole-vacuole fusion machinery.* Autophagy, 2008. **4**(7): p. 874-86.
98. Gebre, S., R. Connor, Y. Xia, S. Jawed, J.M. Bush, M. Bard, H. Elsalloukh, and F. Tang, *Osh6 overexpression extends the lifespan of yeast by increasing vacuole fusion.* Cell Cycle, 2012. **11**(11): p. 2176-88.
99. Connelly, C.F. and J.M. Akey, *On the Prospects of Whole-genome Association Mapping in Saccharomyces cerevisiae.* Genetics, 2012: p. E-publication.
100. Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich, *Principal components analysis corrects for stratification in genome-wide association studies.* Nat Genet, 2006. **38**(8): p. 904-9.
101. Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin, *Efficient control of population structure in model organism association mapping.* Genetics, 2008. **178**(3): p. 1709-23.
102. Wang, K., M. Li, and H. Hakonarson, *Analysing biological pathways in genome-wide association studies.* Nat Rev Genet, 2010. **11**(12): p. 843-54.
103. Schacherer, J., J.A. Shapiro, D.M. Ruderfer, and L. Kruglyak, *Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae.* Nature, 2009. **458**(7236): p. 342-5.
104. Liti, G., D.M. Carter, A.M. Moses, J. Warringer, L. Parts, S.A. James, R.P. Davey, I.N. Roberts, A. Burt, V. Koufopanou, I.J. Tsai, C.M. Bergman, D. Bensasson, M.J. O'Kelly, A. van Oudenaarden, D.B. Barton, E. Bailes, A.N. Nguyen, M. Jones, M.A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E.J. Louis, *Population genomics of domestic and wild yeasts.* Nature, 2009. **458**(7236): p. 337-41.
105. Qin, H. *Hong Qin's GitHub account.* . Available from: https://github.com/hongqin.