# 1. Results from Prior NSF Support

## 1.1 NSF CCLI Award # 0837075, 2009, “Computing in Life Sciences through Hands-on Experience and Case Studies at Tuskegee University”, PI Qin.

With this support, Dr. Qin developed teaching materials – lecture notes, computer lab exercises, modules on bioinformatics and computational biology, available at http://www.bioinformatics.org/ctls. This support resulted in a peer-reviewed teaching publication {Qin, 2009 #1388}:

* Qin, H., Teaching computational thinking through bioinformatics to biology students. Proceedings of 40th ACM Technical Symposium on Computer Science Education, 2009: p. 188-191.

*Contributions to the development of Human Resources in STEM*: This grant has enabled Qin to offer portions of the developed modules and materials at Tuskegee University (two course-offers in one year, many guest lectures, and one three-day faculty workshop with 32 participants), Spelman College (four course offers and four guest lectures in two years), Lewis Clark College (one faculty workshop with 7 participants), Delaware State University (one guest lecture with computer lab to ~20 students), and Alabama State University (one workshop for ~15 faculty and students). Four of these institutions are HBCUs. Although Qin was PI on this grant for only one year, Qin has been carrying out the educational activities beyond the support of this grant.

## 1.2 NSF RUI Award # 1022294, 2010-2012, “Testing the network hypothesis of cellular aging in *Saccharomyces cerevisiae*”. PI Qin.

With this RUI support, Qin laid the groundwork for this career proposal. In less than 2 years, this RUI support has led to one peer-reviewed publication {Guo, 2011 #1089}, one undergraduate honor thesis, and three manuscripts in preparation (Underlined authors are current and formal 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.
* 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., Domminque, P. J. Rodrigues, N. Gupta, B. Mohanty, H. Qin, Hydrogen peroxide induced loss of heterozygosity offers insights on mitotic asymmetry and chronological aging in *Saccharomyces cerevisiae*. In preparation. (Poster presentation at ABCRM 2012, Abstract submitted to the 2012 yeast genetics annual meeting).
* Montgomery, C. K. Matheson, O. Morrison, A. Story, H. Qin, An analysis of genomic features associated with aging in *S. cerevisiae*. In preparation. (Poster presentation at the 2011 SMBE meeting by A. Story in Kyoto Japan)
* Qin, H. A network model of cellular aging. (Poster presentation in computational cell biology meeting, CSHL, March 2011. Oral presentation, annual meeting of Society of Mathematical Biology, Knoxville, July, 2012).

*Contributions to the development of Human Resources in STEM*: This grant has enabled Qin to mentor over 25 African American female undergraduates to conduct computational and experimental research on cellular aging in less than two years. Two of them have won travel awards to attend the 2010 Yeast Genetics & Molecular Biology Meeting at Vancouver, Canada and the 2011 Society for Molecular Biology and Evolution at Kyoto, Japan, respectively. Part of this project has been developed into investigation-based learning modules that have been integrated with two courses: BIO320 “Genomics, Proteomics, and Bioinformatics’ and BIO233 “Microbiology” with a combined total of ~30 enrolled students yearly. In BIO320, students learned to use R to conduct research computational genomics. In BIO233, learn to write mini research proposals on a Wiki- website (<http://sunrays.spelman.edu/bgd/>).

## Table 1. Overview of planned project activities and time line

|  |  |
| --- | --- |
| Years | Project Activities |
| 1-2  2-3  3-4  3-5 | **Theoretical Component –** *The foundation of the whole project.*  Aim 1 Develop a theoretical framework for studying cellular aging based on gene networks. Aim 1.1. Study the impact of power-law and error tolerant network configurations on cellular aging.Aim 1.2. Introduce complex and dependable gene interactions and limiting modules into the network aging model.Aim 1.3. Introduce renewal/repair mechanisms into the network aging model. Aim 1.4. Introduce ploidity and alleles into the network reliability model and study how gene interactions influence lifespan and aging process as quantitative traits. |
| 1-2  1-2  2-3  1-5  1-5  2-5 | **Empirical Componen**t – *Examine biological implications and provide feedback to theoretical work.*  Aim 2. Empirically study the connections between robustness and cellular aging through experimental and computational genomics. Aim 2.1. Use sequenced genomes to compute genes’ mutation variance as mutational capacitances.Aim 2.2. Develop and/or compute a comprehensive set of measures for genes’ phenotypic capacitances.Aim 2.3. Use integrated analysis of capacitance measures and interaction data to identify candidate capacitor genes and limiting network modules on cellular robustness, and study their association with aging-related traits.Aim 2.4. Experimentally characterize how robustness affects chronological aging dynamics, tolerances to oxidative stress and genomic instability by loss of heterozygosity, and expressional robustness in ~100 deletion mutants of candidate capacitors, synthetic lethal pairs, and random singleton genes as controls.Aim 2.5 Experimentally test whether capacitors and robustness limiting network modules tend to be associated with natural variation in lifespan and/or aging-related traits.Aim 2.6. Develop an ODE model for glucose, intracellular H2O2 and superoxide changes in chronological aging. |
| 1-5  1-5  3-5  1-5 | **Educational Component** – *Operational plan of the research components through undergraduate research*  Aim 3. Integrated training on modeling, computing and genome biology to minority students. Aim 3.1. Engage minority undergraduates in interdisciplinary research.Aim 3.2. Integrate original research in computational and genome biology into courses – FYE, CIS115, BIO125, BIO233 and BIO386Aim 3.3. Develop a new course on systems biology for undergraduates Aim 3.4 Build a sustainable undergraduate program on computing and modeling through faculty workshops, an undergraduate minor, a student club, and outreach. |

# 2. Career Goals, Institutional Environment, and Specific Aims

PI Qin’s career goals are to become an effective teacher, a nurturing mentor, and an innovative scholar. Qin’s long-term research goal is to understand how gene networks influence complex traits, especially those with emergent aspects, such as aging.

Spelman College provides an excellent academic environment for Qin to pursue his career goals as a teacher-scholar. As a liberal arts college for women with African American descents, Spelman is forward-looking college aiming to prepare students as future leaders. Spelman has a 39% acceptance rate, consistently ranked by US News among the 100 best liberal arts colleges and the best HBCU. More than one third of the Spelman’s roughly 2000 students pursue degrees in Science and Engineering. The student to faculty ratio is 11:1, giving excellent opportunity for close interactions. Spelman College has a long history of emphasize research in student learning experience. The Spelman college strategic plan for 2015 aims to promote a campus culture of research, expand student research and interdisciplinary training. The interdisciplinary nature of this proposed research and its close integration with education are perfectly aligned with Spelman College’s strategic plan.

The Biology Department at Spelman is constantly revising its curriculum to provide students with the best available learning experience. Over 65% of 2012 biology graduates have had at least one semester of mentored research experience. Since 2010, there are about 40~50 biology students present their research accomplishments on the Spelman Research Day annually. The department’s never-ending drive of curriculum innovation has led to more than 20 years of continuous support from the HHMI educational grant, through competitive renewal applications, to the Spelman College. In May 2012, Spelman was given a HHMI Capstone Award as one of the 11 schools for ‘sustanied excellence and important contributions to undergraduate science education” {HHMI, 2012 #2395}. The biology department strongly supports the proposed research and educational activities by PI (See Chair’s support letter).

For this proposal, Qin focuses on aging – a fundamental question in biology. Although many hypotheses and theories have been formulated to explain the causes of aging, mechanistic understanding of aging are far from clear. PI Qin approaches this question by focusing on cellular aging using *Saccharomyces cerevisiae* as a model system and chooses to gain mechanistic insights of cellular aging based on gene networks. One unique aspect of this proposal is Qin’s network model of cellular aging, which demonstrates theoretically that cellular aging can be an emergent property of model gene networks (see section XX). This mathematical model also predicts that network robustness is proportional to the rate of aging. It also shows that population heterogeneity due to stochastic variation is a key factor in shaping the dynamics characteristics of cellular aging.

The overall plan is composed by three major components – theoretic, empirical, and educational components (Table 1). The integrated natures of these components allows many projects to proceed in parallel – The theoretical work provides directions for computational and experimental efforts, experimental results need to be interpreted by computational analysis under theoretic frameworks, and computational analysis and experimental results will in turn help to improve and revise the theoretic models. These aims are based on the P.I.’s experience and expertise, our preliminary findings, our collection of yeast natural isolates, and the significance of expected results. This project will also be partitioned into modules and small projects and will be integrated into student learning and research.

# 3. Background, Previous and Preliminary Findings, Significance and Intellectual Merit

## 3.1. Brief overview of yeast aging and current challenges

Aging is a fundamental question in biology {Finch, 1990 #401;Harman, 1956 #1036;Williams, 1957 #273}. Although tremendous strides have been made toward the mechanistic understanding of aging over the past two decades, somehow the very concept of aging is still under debate (For example, see {Blagosklonny, 2008 #506}).

The aging of cells that undergo asymmetric divisions likely arose early in the evolution of both prokaryotes and eukaryotes {Stewart, 2005 #642;Stewart, 2005 #643;Henderson, 2008 #474}. As a unicellular organism, the budding yeast *Saccharomyces cerevisiae* has proven to be a good model system for studying mechanisms of cellular aging {Mortimer, 1959 #303}.[add kennedy, longo papers] Many key features of cellular aging were first discovered in yeast before they were established in metazoan cells {Kaeberlein, 2010 #787;Kaeberlein, 2007 #494;Kennedy, 2007 #496;Henderson, 2008 #474;Finkel, 2007 #472}. The life span of yeast can be measured in two ways: replicative and chronological life spans. Replicative life span (RLS) is the number of cell cycles that individual mother cells produce before they senesce and cease to divide {Mortimer, 1959 #303;Park, 2002 #88;Steffen, 2009 #473}. The actual number of daughters produced by a cohort of mother cells is determined by microdissection. Chronological life span (CLS) is how long cells can survive without dividing in stationary phase {Fabrizio, 2003 #282;Fabrizio, 2007 #483}. The number of surviving cells in a population is assessed over time by quantifying colony-forming units. Both replicative aging and chronological aging are defined based on the concept of cell cycle: RLS is a measurement of cell cycles that a single mother cell can accomplish, and CLS is the capability of cells to reenter cell cycle from a non-dividing state.

Paradoxically, the biology of aging becomes evasive once we delve into the molecular mechanisms. Although hundreds of genes in yeast have been found to influence aging, none of these genes suggests molecular mechanisms that are directly linked to aging. The effect of SIR2 on life span is attributed to the “toxic” effect due to accumulation of extrachromosomal rDNA circles {Sinclair, 1997 #250}, a concept that is not only mechanistically obscure and but has also been challenged {Ganley, 2009 #644}. The effect of TOR pathway on replicative life span is attributed to the decreasing ribosome function and translation {Kaeberlein, 2007 #494} or to the hyper-activation of cellular functions {Blagosklonny, 2008 #506;Blagosklonny, 2009 #503}. The mechanism of TOR on chronological life span remains unclear {Wei, 2009 #499}. In fact, it is speculated that bona fide aging genes do not exist because there are no conserved causes of aging {Sinclair, #477;Steinkraus, 2008 #475;Kirkwood, 2002 #626}. With no genes as direct causes of aging, it is surprising that calorie restriction (CR) is a universal way of intervention to extend life span in many organisms including yeast. In yeast, CR can extend both RLS and CLS [REF], despite that substantially different network pathways are activated in the two different aging processes [Lau 05 ref] .

It is clear that cellular aging is largely a stochastic process - Genotypically homogenous yeast cells from a single colony will live to different ages. Genetic factors only contributes about 22% of the natural variation in replicative lifespan {Qin, 2006 #461}. Therefore, it is perplexing that there exists a universal characteristic of aging at the demographic level, known as the Strehler-Mildvan correlation, despite the diverse genotypic and environmental factors that can influence the aging process and great plasticity of individual lifespan. This kind of universality, PI Qin argues, suggests a common principle in the stochastic processes of aging.

A large body of experimental data suggests complex gene networks are involved in yeast aging. In a large scale screen, deletions of 90 genes were found to extend CLS in BY laboratory strains, and only 16 of them are TOR related {Powers, 2006 #564}. The remaining 74 genes are associated with iron homeostasis, cell wall organization and biogenesis, transport, and many have unknown functions {Powers, 2006 #564}. Deletion of 300 genes can shorten CLS {Powers, 2006 #564}. In another screen of RLS, 20% of the gene deletions were found to shorten RLS, whereas 10 out of 564 genes significantly extend RLS {Kaeberlein, 2005 #486}. Six of the 10 genes are implicated in the TOR pathway, four others are a ubiquitin protease, an isocitrate dehyodrogenase, and two proteins have unknown functions. In a quantitative trait study, transgressive segregation of CLS was observed, indicating the involvement of many loci {Kwan, 2011 #1518}. 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, including genes involved in gene silencing, stress response, and mitochondrial function {Guo, 2011 #1089}.

PI Qin proposes that cellular aging is an emergent property of gene/protein regulatory networks, which will provide 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-Mildavn characteristic of aging can be attributed to the common interacting patterns of gene/protein networks shared among most species. The conserved effect of CR may also be understood from its pleiotropic effect on the network as a whole.

## 3.2 Gompertz model – a quantitative definition of biological aging

The dynamics of biological aging can be defined by the two-parameter Gompertz aging model {Gompertz, 1825 #1151}:

|  |  |
| --- | --- |
|  | (Eq 1) |

where, *m* is mortality rate; *s* is the survival fraction of a population; *t* is time. Mortality rate *m* is basically the normalized declining rate of *s*. The initial mortality rate, *m0*, describes the innate susceptibility to dying. The Gompertz coefficient, *G*, determines acceleration rate of the mortality rate over time and is therefore a parameter for rate of aging. The analytic form of *s* can be solved from Eq 1 by using initial condition of s=1 at t=0.

The exponential increase of mortality rate is a universal characteristic of aging in eukaryotic species, including yeast, worm, fruit fly, mouse, and human {Finch, 1990 #401;Gavrilov, 2002 #272;Gavrilov, 2003 #415;Qin, 2006 #461}. Both replicative and chronological aging display the sigmoid shape of survival curves, which implies the exponential increase of mortality rate during the dying-off phase. Qin and Lu 2006 found a negative linear correlation between G and ln(*R0*) in yeast {Qin, 2006 #461}. This negative linear correlation was first reported in human {Strehler, 1960 #380}, known as the Strehler-Mildvan correlation, and has been extensively studied. This kind of universality is one of the motivations for mathematical modeling of cellular aging by PI Qin because it suggests a common underlying principle.

Given the Gompertz definition of biological aging, an organism could have a constant mortality rate when G=0. This means that the drop of viability follows a simple exponential decay, just like a first-order chemical reaction such as the exponential decay of radioactive isotopes. Individuals from these populations will then be as good as new at any time point, and are therefore ‘non-aging’. Bacterial phages indeed display this kind of non-aging characteristics [Depape 08?]. It is worthy clarifying that non-aging bacterial phages are not ‘immortal’, they just die with constant mortality rates.

## 3.3. A gene network model of cellular aging – A unique mathematical approach proposed by PI Qin

### 3.3.1 Rationale:

Emergent property refers to a feature form at system levels but cannot be found at the component levels. Classical examples include termite castles, school of fishes, and flocking of birds. To prove that cellular aging is an emergent property of model gene network, components of this model network ought to be non-aging. Specifically, PI Qin needs to demonstrate that Gompertzian aging of a system can occur in gene networks that are made of non-aging components, i.e., components with constant mortality rates.

Although intracellular molecular networks can be partitioned into gene regulatory networks, protein networks, and metabolic networks etc, Qin’s network model is an abstractive network in order to demonstrate the common principle. The proposed term ‘gene network’ should be viewed as a generic term that goes beyond the cis- and trans-gene expression regulations.

### 3.3.2 The basic network model of cellular aging

To gain mechanistic insights on cellular aging, PI Qin developed a basic mathematical model of cellular aging based on gene interaction network. This model network is made of only nonaging components - the function of gene interactions decrease with a constant mortality rate (Figure XXX). Death of a cell occurs when an essential gene loses all of its interactions, equivalent to the deletion of an essential gene. Gene interactions are modeled as intrinsically stochastic, following a Poisson distribution in the prototype. Qin was able to show that the characteristics of biological aging, the exponential increase of mortality rate over time, can arise from this gene network model. Hence, by definition, cellular aging is an emergent property of this model network.

The basic network contains *K* essential modules, and each module contains one essential and *n* non-essential genes. The biological functionality of each interaction is assumed to be non-aging and decline exponentially with a constant rate of µ. Death of a cell occurs when an essential gene lose all of its interactions to non-essential genes, equivalent to the experimental deletion of an essential gene. Initially, the probability of an interaction in each module being functional is assumed to be p. The mortality rate of the this system is

, when t 1/µ. Eq 1

where *m0* = *CKnpµ(1-p)n-1* , and *G = µp(n-1)/(1-p)*, *µ* is the constant mortality rate of each interaction, *C* is a normalization constant (Qin, preparation). [Figure XXX to show m0 and R] This network model has its roots in the reliability model of aging, and Eq 1 can be found using the ‘initial virtual age’ method developed by Gavrilov and Gavirolva {Gavrilov, 1991 #1785;Gavrilov, 2001 #397} .

The rate of aging, *G*, is approximately proportional to the number of interactions per essential gene (*n*) and will increases dramatically at chances of these interactions being active become higher (*p*). The number of interactions per gene can be viewed a measure of network robustness. In real gene networks with much more complex interactions, the loss of one gene’s function could be remediated by repair mechanisms, other genes with overlapping functions {Wagner, 2000 #567;Gu, 2003 #537}, or alternative pathways through network buffering mechanisms {Wagner, 2000 #567}. The ability of biological systems to maintain homeostasis is often termed ‘robustness’. Formally, robustness can be defined as persistence of phenotypes in the presence of genetic, environmental, or stochastic changes {Balouri, 2008 #533;Masel, 2009 #532}. In the present proposal, Qin aims to study how complex gene interactions influence *G* and *m0* in order to extend the basic network model of aging.

### 3.3.3 Model properties, predictions, implications and limitations.

The most important property of the model is the Strehler-Mildvan correlation – the trade-off between *G* and *m0*. Using the same method as in the work of Gavrilov and Gavriolova {Gavrilov, 1991 #1785;Gavrilov, 2001 #397}, it can shown that ln(*m0*) ≈ -*BG + Intercept*, where *B* and *Intercept* are constants based on K, *µ,* and *p.* This property can explain the ‘conserved’ demographic characteristic of aging.

The most important prediction is the counter-intuitive positive correlation between the Gompertz coefficient *G* and network robustness – more robust cells have higher rate of aging. It is counter-intuitive because aging dynamics is quantified here by two Gompterz parameters, *G* and m0, with only G as a measure of rate. In contrast, the colloquial meaning of aging rate actually contains information for both parameters.

The model indicates that the emergent principle of aging characteristics remain the same between replicative and chronological aging, even though specific pathways likely differ in the two ways of aging.

The model argues that heterogeneity of gene interactions is an important factor in shaping the dynamics of the biological aging. If intrinsic stochastic noises are removed from the model, increase of mortality rate of the model follows the Weibull model, which is often the failure model of complex machinery. This kind of population heterogeneity can be attributed to stochastic variation, genetic heterogeneity and/or environmental perturbations.

We like to mention that the current basic model is over-simplified because we aimed to derive analytic solutions. The theoretic component of this proposal aims to develop more sophisticate models

## 3.4. Comparison of our model with other network related work on aging

The role of networks in aging has also been discussed by others, explicitly or implicitly. Franschi argued that aging is associated with cellular networks {Franceschi, 1989 #638}. Kowald and Kirkwood proposed that the effect of defective mitochondria, aberrant proteins, and free radicals should be integrated as a “network theory of aging” {Kirkwood, 1997 #57;Kowald, 1996 #48}. Soti and Csermely proposed a “weak link theory of aging”- age-related damage affects the low affinity, transient interactions (weak links) and lead to increased noise, destabilization and diversity {Soti, 2007 #634}. Eventually, these noises cannot be distinguished from normal signals. Hence, aging is “deterioration of emergent network properties”, which is similar to the second scenario of yeast aging that we are discussing {Soti, 2007 #634}. Rattan suggested that many genes influencing life span were not “selected” as “genes for aging” and their effect on aging is due to a number of “functionally coupled genes” {Rattan, 1995 #622}. Rattan argued that most hypothesized aging models imply “directly or indirectly that progressive failure of homeostatic mechanism is crucial for the process of aging” {Rattan, 1995 #622}.

The importance of network in aging is also recognized by many others. Xue and colleagues discovered in *C. elegans* that “aging genes” are more enriched at the interfaces between network modules and argues that “aging is linked to the dynamic network stability” {Xue, 2007 #518}. Bell and colleagues studied the human protein interaction network and found genes involved in life span regulation tend to be inter-connected hubs. They experimentally verified 18 genes in worm and found a third of them can extend worm’s life span {Bell, 2009 #529}. Budovsky and colleagues explored the human orthologs of genes associated with aging in model organisms. They found that hub genes are often involved in age-related diseases {Budovsky, 2007 #525}. They further found that tumor suppressors tend to extend life span, whereas oncogenes tend to shorten life span, which are argued as evidence for the evolutionary and molecular links between aging and cancer. Smith and colleagues compared genes associated with aging in yeast and worm and found that yeast orthologs of worm genes associated with aging also tend to alter life span in yeast, indicating a conserved mechanism of life span regulation between yeast and worm {Smith, 2008 #522;Smith, 2007 #498}. Lorenz and colleagues focused on a 10-gene network from the Snf1 signaling pathway in yeast and engineered perturbation to infer causal network interactions and found new interaction associated with aging {Lorenz, 2009 #478}. Barea and Bonatto argued that protein interaction network can synthesize different theories of aging and used yeast protein network to explore the connection between replicative and chronological aging {Barea, 2009 #524}. Promislow compared yeast protein associated with aging to five traits and found that protein associated with aging are more connected than expected by chance, which is argued as evidence of antagonistic pleiotropy theory for the evolution of senescence {Promislow, 2004 #530}. Promislow also argued that the network characteristics of proteins associated with aging may be useful to predict new genes associated with aging {Promislow, 2004 #530}.

In comparison, our model is built on the mechanistic insight that cellular aging is an emergent property of gene networks, and offers unique insights and predictions,

## 3.5 Intellectual merit, significance and broad implications

Our model only demonstrates that cellular aging is an emergent property of gene network. It can be further extrapolated that aging is an emergent property of complex living systems in general. Viewing aging as an emergent property may provide mechanistic foundation for some evolutionary theories on aging, such as the antagonistic pleiotropy theory {Williams, 1957 #273} and the disposable soma theory {Kirkwood, 1977 #56}. Not only can we better understand the concept of pleiotropy from the network perspective, we will probably also look to networks to understand the connection between advantages at early life and detrimental effects at late life {Williams, 1957 #273}. This kind of trade-off is also the central argument of the disposable soma theory {Kirkwood, 1977 #56}.

Our theoretic model also argues that aging is one of the best measures of robustness - the Gompertz parameter is a measure of robustness (Gottschling, personal communication). Robustness can reconcile the mutational costs to individual and the evolutionary benefit to the population {Wagner, 2012 #2242}, because the phenotypics effect are hidden in most conditions. Robustness is related to canalization, and it has been argued that network buffering is a key mechanism of canalization {Levy, 2008 #606;Masel, 2009 #532;Wagner, 2000 #567}. Our model suggests a mechanistic link of robustness, gene networks, and aging.

Our model also demonstrates that network biology is a useful way to tackle biological complexity, and network approaches can provide insights that cannot be easily achieved by reductionist approaches.

## 3.6. Other previous and preliminary results in support of the proposed network model

### 3.6.1. Student robustness project, morphology robustness is correlated with RLS

### 3.6.2. H2O2-LOH (need to fit CLS with Gompertz model)

Over-expression of SOD2 can extend CLS {Fabrizio, 2003 #2022}, in consistent with our Cb/Cv ~ CLS correlation.

Table 1. Replicative and chronological life spans in yeast strains (ADD Tg/Tc, Cb/Cv)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strain | ARLS | *m0* | *G* | CLS |
| 101S | 31.3±0.8 | 0.0012±0.0007 | 0.14±0.02 | 3.4±0.2 |
| M1-2 | 27.9±1.1 | 0.0024±0.0011 | 0.13±0.02 | 10.4±3.2 |
| M13 | 26.7±1.2 | 0.0030±0.0011 | 0.13±0.01 | 16.3±3.7 |
| M14 | 36.6±1.5 | 0.0019±0.0007 | 0.09±0.01 | 7.2 ±0.2 |
| M22 | 31.8±1.3 | 0.0024±0.0009 | 0.11±0.01 | 5.2±2.1 |
| M2-8 | 24.8±0.8 | 0.0042±0.0010 | 0.12±0.01 | 4.1±0.7 |
| M32 | 28.1±0.8 | 0.0018±0.0005 | 0.14±0.01 | 6.4±0.8 |
| M34 | 26.8±1.0 | 0.0013±0.0007 | 0.16±0.02 | 5.2±0.4 |
| M5 | 36.7±1.0 | 0.0040±0.0008 | 0.07±0.01 | 4.9±0.5 |
| M8 | 35.2±0.9 | 0.0004±0.0002 | 0.16±0.02 | 10.5±0.1 |
| SGU57 | 23.6±1.5 | 0.0080±0.0022 | 0.09±0.01 | 9.3±0.7 |
| RM11 | 44.6±1.5 | 0.0010±0.0004 | 0.09±0.01 | 12.7±2.9 |
| YPS128 | 35.0±1.2 | 0.0011±0.0005 | 0.12±0.01 | 3.3 ±0.1 |
| YPS163 | 34.4±0.8 | 0.0008±0.0003 | 0.14±0.01 | 4.2±1.1 |
| BY4743 | 33.2±0.9 | 0.0040±0.0013 | 0.08±0.01 | 9.7±1.7 |
| SK1 | 22.0±1.3 | 0.011 ±0.003 | 0.08±0.01 | 5.0±0.9 |
| W303 | 18.7±0.6 | 0.0034±0.0011 | 0.20±0.02 | 17.2±3.9 |

ARLS (Average replicative life span), CLS (Chronological life span). Standard deviations are estimated by bootstrapping. The numbers of bootstraps equal the sample sizes. These strains are a unique and important resource of this project.

### *3.6.3.* Flow-cytometer based measure of CLS and ROS *.*

|  |
| --- |
| C:\Documents and Settings\hqin\My Documents\Dropbox\MRI.dropbox\sandboxMRI\figures\A3.071410\SangerA3-PI-viability\Slide1.TIF |
| Figure 1. Propidium iodide (PrI) staining can be used to measure chronological lifespan in *S. cerevisiae*. PrI-negative cells are considered as viable cells. Small inserts are histograms of log-transformed PrI signals. When cells become older, the PrI peak shift to the right, indicating more cell become PrI-positive. This experiment was done using a BD FACSCanto II at the FHCRC. |

PI Qin has developed a protocol to measure CLS of the budding yeast using conventional flow cytometer at the Fred Hutchinson Cancer Research Center in the summer of 2010 and at the Princeton University in the summer of 2011. Briefly, aliquots of yeast cells in either depleted media or water are taken periodically, sonicated for 2 seconds, and are stained by propidium iodide (PrI). PrI is a fluorescent dye that enters cells with damaged membranes, and can be used as a marker for certain types of cell death {Fontaine, 2008 #1043;Lopez-Amoros, 1995 #1049;Yu, 2011 #1038;Hagiwara, 2011 #1040}. PrI negative cells are considered viable cells, and PrI positive cells are considered dead (Figure 1). The fractions of live and dead cells are modeled by t-mixture models with Box-Cox transformation or log-transformation, and are estimated by expectation maximization procedure using the R package flowClust {Lo, 2009 #1247;Lo, 2008 #1249}.

There are at least two major shortcomings of the PrI-based viability assay that can be greatly improved by the requested image flow cytometer. (1) PrI can stain nucleic DNA, mitochondrial DNA, and cytosolic RNA. These background signals vary substantially in ~70 yeast natural isolates that we measured. ImageStreamX can separate the more condensed nucleic DNA signals from the more diffused signals of mitochondria DNA or cytosolic RNA. (2) Some dead yeast cells are PrI-negative, presumably due to apoptotic degradation of nucleic DNAs. Identification of these dead cells from live cells can be improved by morphometric features.

### 3.6.4. Radicicol on hsp90 (Need to revise this).

Hsp90 is a phenotypic capacitor that can buffer mutations in its substrate proteins {Rutherford, 2003 #641;Rutherford, 1998 #640}. As a result, Hsp90 plays an important role in gene network robustness. We found that radicicol, an Hsp90 inhibitor, can extend CLS in some natural isolates at 2.5nM (? Check this, need figure). The survival curve displayed increased initial mortality rate and decreased Gompertz coefficient as compared to those of the untreated samples. This CLS-extension effect of radicol is only observed in two strains of the four natural isolates that we measured. At 5nM however, ridicicol did not extend CLS, indicating its lifespan extension effect is hormetic.

Two copies of the Hsp90 gene exist in yeast genome: HSC82 and HSP82, which makes molecular manipulation a complicated process {Nathan, 1995 #609}. Using engineered haploid lab strains (W303), Harris et al found that reduced Hsp90 activities do not change the mean RLS {Harris, 2001 #608}. However, neither Gompertz modeling nor sophisticated statistical tests such as LRT was used. In fact, based on visual examination of their data (Figure 1 in their report), the drop of viability is sharper in the wildtype but less so in the mutants with reduced Hsp90 activities {Harris, 2001 #608}. This indicates that *G* is larger in the wildtype but is smaller in the mutants, which is consistent with our hypothesized role of Hsp90 in cellular aging. Notice that the rate of aging can change significantly without changing the mean life span (See Eq3).

We found many single-nucleotide polymorphisms in the promoters and coding regions of HSC82 and HSP82 (http://www.sanger.ac.uk/gbrowse/gbrowse/cere\_dmc/#search). It is also known that certain fungal Hsp90 is resistant to radicicol inhibition due to a leucine-to-isoleucine substitution {Prodromou, 2009 #610}. Hence, reasonable yeast natural variation should exist in responses to Hsp90 inhibitors.

One way to assay the tolerance to these inhibitors is to grow yeast strains in different concentrations of these inhibitors using a high throughput growth curve assay instrument – such as the Bioscreen C (http://www.growthcurvesusa.com/) . From the lag, mid-point, and plateau of the growth curves, we can estimate each strain’s dose-dependent responses, such as the lethal dose at 50% (LD50), (for similar applications, see {Fraud, 2005 #651;Powers, 2006 #564}).

We will use regression to test whether tolerances to Hsp90 inhibitors are associated with their rates of aging in these natural isolates of yeast. For one or two strains, we will perform replicative life span assay under several inhibitor doses, for example, LD10, LD25, and LD50. We will then use regression to test whether aging rates are associated with doses of Hsp90 inhibitors.

Alternatively, flow cytometry can be used to detect the fraction of live and dead cells. The LIVE/DEAD® Cell Viability kit from Invitrogen will be used to stain yeast cells. Finally, plating assays and colony-forming units will be used to verify the high-throughput measures.

Because of the complexities of aging, we will look for partial correlation between tolerance to Hsp90 inhibitors and aging rates by controlling for other factors, such as responses to oxidative stress (see below), tolerance to TOR inhibitors (see below), genomic instability and mitotic asymmetry (see our previous publication {Qin, 2008 #516}).

Flatt review: use Hsp90 and Heat Schock proteins are examples of canalization.

Heat shock protein 90 (Hsp90) is a promising cancer drug target, as multiple oncogenic proteins are destabilized simultaneously when it loses its activity in tumor cells. Highly selective Hsp90 inhibitors, including the natural antibiotics geldanamycin (GdA) and radicicol (RAD), inactivate this essential molecular chaperone by occupying its nucleotide binding site. Often cancer drug therapy is compromised by the development of resistance, but a resistance to these Hsp90 inhibitors should not arise readily by mutation of those amino acids within Hsp90 that facilitate inhibitor binding, as these are required for the essential ATP binding/ATPase steps of the chaperone cycle and are tightly conserved. Despite this, the Hsp90 of a RAD-producing fungus is shown to possess an unusually low binding affinity for RAD but not GdA. Within its nucleotide binding site a normally conserved leucine is replaced by isoleucine, though the chaperone ATPase activity is not severely affected. Inserted into the Hsp90 of yeast, this conservative leucine to isoleucine substitution recreated this lowered affinity for RAD in vitro. It also generated a substantially enhanced resistance to RAD in vivo. Co-crystal structures reveal that the change to isoleucine is associated with a localized increase in the hydration of an Hsp90-bound RAD but not GdA. To the best of our knowledge, this is the first demonstration that it is possible for Hsp90 inhibitor resistance to arise by subtle alteration to the structure of Hsp90 itself.{Prodromou, 2009 #610}

### I am here, 2012 June 11.

### 3.6.5. Ploidity

For a given set of measured life spans of N cells (), the likelihood function based on the Gompertz model is:

.

Hence, the log-likelihood is:

.

Maximization of the above log-likelihood function will yield the maximal likelihood estimations of the Gompertz parameters for a data set of life span. We have implemented the maximization procedure in the R-statistics language and environment. This likelihood approach enables us to design nested models to statistically compare *I* and *G* values of different strains (see section 4.3).

Protein expression noise in G1 phases are much lower in diploids (lower CV) than that of haploid cells {Di Talia, 2007 #2408}.

If stochastic variation in the number of key molecular causes gene expression noises, then doubling their numbers ought to reduce the noises by sqrt(2) {Schroedinger, 1944 #2414}.

CV in diploid is sqrt(2) lower than haploid cells (Shoediz book, cite in Di Talia 07 nature paper.

stochastic model is necessary for aging.

### 3.6.6 CR effect on yeast aging??(Need nested model here)

### GWA study (move to research plan?)

# 4. Research Plan, Expected Results, and Alternative Approaches. (need to strengthen connection between aim 1 and aim 2. Aim 1 -> prediction, Aim 2-> robustness modules -> simulation study in Aim 1. )

## Aim 1, Theoretical Component: Develop a theoretical framework for studying cellular aging based on gene networks.

The goal here is to gain better insights on the emergent aspect of cellular aging from the network perspective. We partition this aim into a list of sub-aims. It is important to emphasize that these sub-aims are intertwined and inter-dependent.

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

(Doyle’s tradeoff of robustness and fragility)

Bet-heding evaluation (decrased arithmetic mean versus increased genometric mean fitness ref 87 in Levy 2012).

Our network model predicts that heterogeneity plays a key role during the emergence of biological aging. One of the key sources of heterogeneity in gene networks is its power-law feature – the degree distribution of genes follows

P(k) = Z(gamma)-1k-gamma, Eq. 1 (powerlaw)

where k is the number of connections per gene, Z represents the Ziemman function, and gamma is a coeffient {Aldana, 2003 #1784}. When gamma <=3, the variance of P(k)= Z(gamma)-1k-gamma is infinite. For most biological networks, gamma is often between 2 and 3 {Barabasi, 1999 #1520}(cite a review on this), which indicates tremendous amount of heterogeneity in biological networks. Networks with power-law features, such as Internet, are robust to random failures but are fragile to deliberate attacks {Albert, 2000 #1528;Doyle, 2005 #1521}.

In addition, 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 protein networks {Maslov, 2002 #1651}. 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 {Maslov, 2007 #1771}.

Here, we will use simulation to study how power-law degree distribution and error tolerance features of gene networks influence the aging process.

There are several ways to simulate power-law gene networks. The preferential attachment model (BA99 model) is often used. Alternatively, we can generate the degree distribution based on P(k)= Z(gamma)-1k-gamma (Eq1), and then pairs interacting nodes in a similar way to the network modeling approach implemented previously by us {Qin, 2003 #566}.

In our network model of aging, we need to categorize the nodes in the network as ‘essential’ or ‘non-essential’ genes. For simplicity, we will choose 5%, 10%, and 15% of nodes with top-ranked degrees as essential genes.

For comparison and controls, we will simulate the aging processes of networks with fixed numbers of interactions per gene, a Poisson distribution of degrees, and a log-normal distribution of degrees.

To character the aging process, we will calculate the mean, median, minimal, maximal life spans, m0, G, Makeham constant, Weibull distribution, Akai Information Index for model comparison, the mean and median life span of 10% longest-lived individuals (the long tail).

I need a table to summarize network features and parameters to be simulated, and parameters to characterize the aging process: degree distribution, node partition, interaction functional decay, gamma parameter of power-law, aging: mean, median, maximal, m0, R, M, tail, last 10%

In addition, we will introduce noises into exponential decay rates using either log-normal model or failty model as in {Finkelstein, 2006 #1514}.

Misses Ericka Dommond, Jessika Williams and Jessica Christopher, three Spelman undergraduates mentored by PI Qin, have prototyped some R codes to simulate cellular aging based on some simple network models. They found that simple network models with fixed degree per node but with log-normal distributed rates of exponential decay can give rise to Gompertz characteristics of aging. They further found that in networks with Poisson degree distribution, the effect of heterogeneous exponential decay rates can be dwarfed by the heterogeneity of degree distribution. (Describe the simulation process: maximum of component in modules, and minimum of module age for system. )

Network evolution simulation {Presser, 2008 #1762}

{Deeds, 2006 #1764}

It is important for us to compare our work here with the traditional reliability approach by Finkelstein and Vaupel {Finkelstein, 2006 #1514} that studied the heterogenous decay rates on the tail distribution of life span. Finkelstein and Vauple used a frailty model to introduce noises, and proposed a summary statistics to measure the tail distribution (What is their main findings?).

Plateau of mortality rate in population indicate heterogenous distribution (James Vaupel’s work on gamma distribution)

network configuration on cellular aging by comparing the aging characteristics of network with power-law connection configuration and Poisson configuration. Study how biological noises influence aging.

Related to Aim 1.2, we plan to investigate whether power-law configuration play a role in CR effect (modeled by moderate negative cooperation of Hill equation for dependable interactions).

Doyle argues that systems evolved to be robust against general perturbations are fragile against certain types of rare perturbations {Doyle, 2005 #1521}. Doyle futher argues there are trade-off between robustness and fragility {Doyle, 2011 #2125;Doyle, 2007 #2126;Tanaka, 2005 #2127;Csete, 2004 #2130;Csete, 2002 #2131}

### Aim 1.2. Introduce complex gene interactions and dependable gene interactions into the network reliability model.

Our prototype network reliability model of aging does not consider interactions between essential genes. Interactions between hub genes (and likely essential genes) indeed occur below random expectations in real gene network {Maslov, 2002 #1651}, they do exist nevertheless. When interactions are allowed between essential genes, it means that the failures of two functional modules are not independent, which will make analytic studies very challenging. Here, we will use simulation to explore how failures of dependent functional modules affect the aging process.

Another interaction type of great interest is the synthetic lethality, which occurs when more than two gene deletions lead to cell death. Many single-gene deletions do not lead to cell death in yeast, but double deletion of two genes can result in a lethal phenotype. When one of the synthetic lethal genes is deleted, we expect that it should decrease the redundancy/robustness -> smaller G and larger m0, and perhaps larger M. The effect of the synthetic lethal pairs on cellular aging can be simulated straightforwardly, and its prediction can be tested experimentally (see Aim 3.3).

From the computing and coding perspective, interaction between essential genes and synthetic lethal genes can quickly increase computational complexity of the program and require increasing book-keeping. To simply the modeling and coding process, we plan to merge dependent essential modules into super-modules (I need a diagram to illustrate my plan for interacting essential modules and synthetic lethality. Box two interacting essential modules. ). We will simulate the aging of these super-modules by additional bookkeeping, and the rest of the independent modules can still be simulated using the original fast algorithm.

Our prototype network reliability model of aging assumes that the functional of each interaction is independent. In reality, many gene interactions may influence each other, and the decrease of their function may be associated and intertwined.

In the traditional reliability model, if we assumed component mortality rate is proportional to the number of failures (i.e., linear assumption), Gomertz model of aging can be approximately obtained in a special case {Gavrilov, 1991 #1785;Gavrilov, 2011 #1786;Le Bras, 1976 #1787}. This is has been termed an avalanche–like model {Gavrilov, 2011 #1786;Gavrilov, 1991 #1785}.

To implement a similar linear increase model of component failure rate in our network reliability model, we need to decide how changes are distributed among modules. For simplicity, we will start with dependent changes within modules. It is also reasonable to assume that dependent changes largely occur within each essential module. We can then object-oriented programming techniques to implement the aging of each module.

In previous studies based on traditional reliability models, the slope of the linear model was chosen to be one, presumably because the authors focused on analytic solutions. We argue that the slopes of the linear model describe how much gene interactions affect each other’s failure rate - in other words, how much deleterious effects are shielded by other gene interactions. Hence, the slope of the linear models is related to network robustness. From this perspective, it seems more reasonable to model these dependable changes using sigmoid functions such as the Hill function. Using Hill functions will also generalize our model on gene interactions because linear models can be viewed as special cases of sigmoid functions. The non-linearity of Hill function can be adjusted by the Hill coefficient. The Hill coefficient can also describe positive and negative changes. We plan to simulate aging process using a series of Hill coefficient, including both positive and negative ones. For comparison, we will also simulate aging process using a various slopes of linear models for dependable failure rates. We will focus on how these parameters will influence the parameters of the aging process as summarized in Table 2 (??). The network configuration can be fixed number of interactions per module, random configuration, or power-law configuration. One question is particular interest to us: Can moderate negative Hill cooperation regenerate some features of calorie restriction on cellular aging in random networks and/or power-law networks?

### Aim 1.3. Introduce renewal/repair mechanisms into the network reliability model.

Intracellular gene networks are dynamic networks – Gene products are continuously synthesized and damaged products are often repaired. From the mathematical perspective, renewal and repair processes can be both modeled in the same way, as in repairable engineering systems {Vaupel, 2003 #1791;Finkelstein, 2006 #1514}. The self-repairing property plays an important in formatting theories of aging {Kirkwood, 1988 #51;Horiuchi, 2003 #1789;Yashin, 2000 #1790}.

Repairs can be either imperfect or perfect. To explore the theoretic impact of repairs on cellular aging, we will focus on perfect repairs as in previous studies {Vaupel, 2003 #1791}. Perfect repair is equivalent to replacement of machinery components in the reliability engineering {Barlow, 1996 #1792;Leemis, 2009 #1793}. In reliability engineering, repairs are modeled by a specified number of stand-by components, and the systems fails as the last reserved component fails {Leemis, 2009 #1793}. This replacement approach was also used in previous modeling on aging {Vaupel, 2003 #1791;Finkelstein, 2006 #1514}. Even with the assumption of perfect repairs, the aging process of complex systems is challenging to be described. Because of the complexity of studying repairs in large systems, previous work addressed systems with only one replaceable component {Finkelstein, 2006 #1514}.

We plan to introduce perfect repairs into our network aging model by studying its effect first in single-module network and then in multiple-module networks. We will assume that each defect gene interaction will be repaired by replacing the essential gene product involved. This assumption means that aging clocks of all interactions of the replaced essential gene will be re-set.

Intuitively, repairs should increase network robustness, and should lead to large G and smaller m0. This expectation is consistent with previous results that repairs could decrease the relative tail of longevity and presumably sharpens the transition of the dying off phase of the survival curves {Vaupel, 2003 #1791}; Finkelstein, 2006 #1514}.

We will also compare repairs of the highly connected genes and less-connect genes in power-law networks. As control, repair in random networks will also be studied. It was argued verbally that weak-linked genes in networks are key to aging [REF from a europena lab? ]. Our simulation study here can address this question quantitatively.

Vaupel 2003 proposed to a relative tail of longevity by comparing two quantiles, say 50% and 90%. This non-parametric method will be as an alternative approach to the Gompertz model to characterize the aging process.

Stochastic ordering, shaked and Shantikhumar, 1993

### Aim 1.4. Introduce ploidity and alleles into the network reliability model and study how gene interactions influence lifespan and aging process as quantitative traits.

Ploidity and alleles are two related genetic factors that contribute to both robustness and plasticity in eukaryotic organisms, including *S. cerevisiae*. The importance of different allelic interactions on aging can be seen by the transgressive segregation of CLS reported by Kwan and Bedalov {Kwan, 2011 #1518}. In this study, the CLS of two haploid parental strains are similar, but their haploid progeny display a wide range of CLS. This kind of transgressive segregation can be sufficiently generated by a single antagonistic QTL based on additive model, and has been argued to play important roles in adaptation and speciation {Rieseberg, 2003 #1795;Rieseberg, 1999 #1797}.

In our network aging model, we plan to introduce two alleles only for the essential genes (for diploid organisms such as *S. cerevisiae*). When an essential gene has two alleles that each has a different failure rate, we will be facing the question how to model lethal phenotype. A natural option is to treat the two alleles as redundant components which is equivalent to parallel components in reliability engineering models. Mathematically, this two-allele model for one gene is serendipitously analogous to the synthetic lethality model (two single-allele genes) in Aim 1.2, because non-essential genes do not contribute to lethality in our model.

Zuk et al proposed a limiting pathway (LP) model for disease that is analogous to traditional reliability models {Zuk, 2012 #1798}. Zuk and colleagues show that when there are two LP in the model, … … ??(how could additive model leads to epistasis?).

Cryptic genetic variation refers to those variation that are capable of affect some traits but were buffered genetic robustness, such as the genetic capacitance of Hsp90 {Gibson, 2004 #2157;Rutherford, 1998 #640}. The slope of regression line between phenotypic reponses and generations (lines) were termed ‘realized heritability h2’ {Rutherford, 1998 #640}. It was argued that capacitors can ‘conditionally’ release genetic variations and may provide adaptive advantages {Rutherford, 1998 #640}. Cryptic variation was found to promote rapid adaptation in a ribozyme {Hayden, 2011 #2155}.

Our network model does not need the additive assumption for QTLs.

#### Implication: aging as a quantitative trait by simulating interacting and non-interacting genes. Dominance using diplod alleles

Hidden variation (see Hermisson and Wagner, 2004 Genetics, hidden variation and genetic robustness)

Mutation and aging, LOH modeling

Evolutionary and ecological implication of age-structure yeast population.

Emergent property modeling and selection on genes. Canalization. Pleiotropy.

Threshold trait?

How would statistic epistasis arise from allelic effect on aging in networks

Dependable interaction -> robustness

### Expected outcomes , significance, and alternative approaches

Aging is one of the best measurements of robustness (Quote from Dan Gottschling, personal communication).

Biological heterogeneity and noises is an important factor in biological aging.

Improve our understanding of lifespan as a quantitative trait, in reference to Zuk work from the Lander lab.

m0 , G, and tail distribution in large population

Gompertz versus Weibull distribution

Synthetic lethal interactions

Birth/repair effect (more reduncy -> large G, small m0)

Provide guidance for experimental studies

## Aim 2, Empirical Component: Study how robustness influences cellular aging through experimental and computational approaches.

To empirically study how network robustness affect aging, we will focus on capacitor genes and synthetic lethal pairs.

### Aim 2.1. Use sequenced yeast genomes to compute genes’ mutation variance and identify candidate genetic capacitors

(A gene’s cryptic variation could be either due to its intrinsic robustness or due to its interacting genes, such as capacitors. Compare with the 300 morphological stabilizers by Levy and Siegel {Levy, 2008 #606}.

Gibson and Dorkins argued that a proportion of nucleotide polymorphism are conditional non-neutral {Gibson, 2004 #2157}.

Gibson and Dorkins argued that CGV may be detected by the increase of genetic variance as compared to environmental variance and statistical interaction between alleles {Gibson, 2004 #2157}. Alleles are maintained in populations mainly for their contribution to survival under stressful conditions, as first demonstrated by observations in E. coli that many metabolic loci only show growth defects in nutrient limiting conditions {Dykhuizen, 1980 #2226}.

Network data: protein interactions complex and unified? {Wang, 2009 #2319;Collins, 2007 #2320}. Predicted yeast functional network {Lee, 2004 #1737}. Genetic network {Costanzo, 2010 #2074}. Synthetic lethal {Tong, 2004 #1648}. (What the difference?)

A human currated reference network {Myers, 2006 #2324}. Co-expression network from 3 expression data sets {Gasch, 2000 #611;Spellman, 1998 #2327}.

#### 2.1.1

Polymorphism in ‘wildtype’ individuals suggests that these sequence changes do not substantially affect fitness. Both synonymous and nonsynonymous nucleotide diversity can contain ‘cryptic genetic variation’ that can affect phenotypes once the robustness of an organism is compromised {Masel, 2009 #532}.

Polymorphism in each gene may indicate the extent of cryptic genetic variation that an organism can tolerate, and hence is a proxy of mutational robustness with respect to that gene.

Genetic capacitors (also called phenotypic capacitors in some papers) can act as switches of the degree of robustness {Masel, 2009 #532}. One of the best studied genetic capacitors is Hsp90, first reported in the fruit fly [Cite Rutherford, Lindquist papers]. The yeast Hsp90 is encoded by two copies HSP82 and 88? in yeast.

Biological robustness from the systems biology perspective {Kitano, 2007 #1488;Kitano, 2004 #1495;Kitano, 2007 #1488;Kitano, 2004 #1495}. It argues that robustness is the maintenance of specific function against perturbations and often requires the systems to change its mode of operation in sufficiently flexible ways to deal with the perturbations {Kitano, 2007 #1488;Kitano, 2004 #1495;Kitano, 2007 #1488;Kitano, 2004 #1495}.

Compare to Segal’s phenotypic robustness screen based on morphology {Levy, 2008 #606;Levy, 2008 #606}.

Random and hub gene deletion on robustness in E. coli used variance in growth rate to gauge robustness in deletion mutants. {Cooper, 2006 #535;Cooper, 2006 #535}

Wagner argued that biological systems are mutationally robust for two reasons: There are a large number of equivalent solutions for organisms to solve a given problem, and natural selection may favor a solution that are robust than others {Wagner, 2005 #1872;Wagner, 2005 #1872}. Wagner further argued that mutational robustness can emerge as a by-product of selection for robustness {Wagner, 2005 #1872;Wagner, 2005 #1872}. Wagner’s reasoning suggest ??

Method: The most accepted metric for mutations is the synonymous nucleotide diversity, as it has been applied to estimate mutational rates in E. coli genes {Martincorena, 2012 #1873;Martincorena, 2012 #1873}. We will use a similar approach to estimate the synonymous and nonsynonymous nucleotide diversity using sequenced strains of *S. cerevisiae*. Codon bias, recombination

Low mutation, low mutational variance indicate mutational robustness or genetic canalization {Hermisson, 2004 #1515;de Visser, 2003 #1516 ;Hermisson, 2004 #1515;de Visser, 2003 #1516 }.

*Genome resources*: As we are preparing this proposal, we have over 70 strains with sequenced genomes: 36 strains with sequenced and aligned genomes by the Saccharomyces Genome Resequencing Project (SGRP) at the Wellcome Trust Sanger Institute, at least 25 strains with contigs available at the Saccharomyces crevisiae Strain Proejct at Genome Institute at Washington University (NCBI SRA: at least over 20 entries with clearly identified strain background information Y2209, YB210, sigma, etc. At least 10 of them have assembled contigs and supercontigs). There are 30 genomes are available at Saccharomyces Genome Database (SGD), many of which are sequenced by Wustl. Furthermore, in collaboration with Dr. David Botstein, we are also sequencing the 11 natural isolates that we have phenotyped lifespans, genomic instability, and tolerance to oxidative stress.

Platform bias (454), Sequencing errors.

Sanger genome contain some ambiguous nucleotides {Ramazzotti, 2012 #1874;Ramazzotti, 2012 #1874}

Because Wustl did not impute sequence, so there are less number of CDS in these genomes {Ramazzotti, 2012 #1874;Ramazzotti, 2012 #1874}.

In addition, 35 strains of S. paradoxus are also available from *SGRP*.

*Expected outcomes, caveats, alternative approaches*: The Sanger sequenced strains is based on low-coverage sequences {Liti, 2009 #679;Liti, 2009 #679}, and some SNPs and indels are imputed based on ancestral recombination graphs using closely related genomes. This approach likely will remove rare SNPs and indels from the samples. Because we are interested in the overall mutational pattern of each gene, rare SNPs would not affect our estimations in substantial ways.

As one alternative to nucleotide diversity in the coding regions, we can also estimate the nucleotide diversity in the 5’ UTR regions. As antoher alternative, we can use copy number variations to estimate the ‘dosage robustness’ of each gene, which can be compared to the dosage genetic screen results from Boone lab. CNVs can be inferred from the alignment quality information, with the caveat that it may also caused by mapping errors. The short-reads sequences from the Wustl strains can be used to infer CNVs. One of the question of CNV with respect to yeast aging is to see whether copy number of the rDNA locus is realted with yeast life span variation. (Need more reference here. How Genomic DNA are prepared? ). In addition, CNVs can be used see whether the number of rDNA unit at the rNDA locus is associated with natural life span variation. Spontaneous rDNA copy number variation modulate Sir2 levels and may affect aging {Michel, 2005 #423;Michel, 2005 #423}. There are conflicting evidence on whether extrachromosomal rDNA circles or high recombination rate at this locus is associated with RLS {Lindstrom, 2011 #673;Lindstrom, 2011 #673}{Kobayashi, 2006 #2029;Kobayashi, 2004 #2025;Kobayashi, 2011 #691;Ganley, 2009 #644} [How about telomerase?]. rDNA and telomere, genome instability {Kobayashi, 2011 #691}. Gotschling show that Met15LOH is a readout of rDNA locus instability. LOH is a sign for genome wide elevation of hyper recombination.

PI Qin has extensive experiences in sequence analysis and bioinformatics analysis. Qin will also collaborate with Dr. Titus Brown (Michigan State University) on the next-generation sequence analysis.

#### 2.1.2. Compare genetic capacitors with phenotypic capacitors

Phenotypic capacitors can be measured by morphology, growth fitness in diverse conditions. See morphology database. Chemical genomic database.

### Aim 2.2. Develop a comprehensive set of quantitative measures for genes’ phenotypic capacitances.

#### Crowd-sourcing approach, 0, 1, 2, approach to label qualitative data. (Bioinformatics skill to parse, merge, and annotate the data in CIS 115, BIO125, Bio233). Ideal undergraduate bioinformatics project.

#### Acetic acid, survival data {Mira, 2010 #2344}, expression data {Li, 2010 #2352}.

#### Botstein lab, survival data, quiesnce data.

#### Morphological capacitor

#### Growth capacitor from chemicalgenome dataset {Hillenmeyer, 2010 #2042}

#### mRNA expression level capacitors?

#### Caveat, strain background, ploidity.

### Aim 2.3. Integrated analysis and comparison of various capacitance measures based on gene interactions patterns, and aging-related traits.

#### Current work focused on Physical Interaction data, expand to protein complex data, genetic interaction data.

#### Weka and BioWeka {Gewehr, 2007 #2332;Frank, 2004 #2333}.

#### Aim 2.3.1 Study the connection of gene interactions, robustness, and aging

#### the statistical associations of multiple proxies of robustness with aging measurements. robustness ~ RLS and CLS on using yeast gene deletion mutants

genome-wide fitness data {Hillenmeyer, 2010 #2042;Hillenmeyer, 2010 #2042}, data in 2008 science paper {Hillenmeyer, 2008 #541;Hillenmeyer, 2008 #541}.

Robustness ~ GFP CV, H2O2-LOH, radicicol effect, CR, rapamycin effect.

Duplicate essential genes (Jianzhi Zhang’s paper)

Duplicate synthetic lethal genes

Does CR improve robustness?

Synthetic lethal gene deletion mutants

Tolerance to genomic instability and cellular aging, H2O2-induced LOH

Radicicol on life span (down stream signals)

rapamycin, CR effect on DHE and DHR signals

GFP CV on robustness measure of lifespan influencing deletion mutants.

Resources

CLS screen by Kaeberlein {Burtner, 2011 #779;Burtner, 2011 #779}

Growth data in NaCl. Prophecy, protein complexes {Warringer, 2003 #261;Warringer, 2003 #1969;Warringer, 2003 #261;Warringer, 2003 #1969}

Review SGD

Caveat: neighboring-gene effect {Ben-Shitrit, 2012 #1970;Ben-Shitrit, 2012 #1970}

Dosage effect {Magtanong, 2011 #2020;Magtanong, 2011 #2020}

##### Robustness

###### Flowcytometer on GFP mutant collections, expression variance ~ noise ~ ~ morphology robustness ~ robustness

###### This is similar to Rutherford and Yoshi’s yeast work.

###### Ref: Silander 2011 Plos genetic, use flow cytometer to measure GFP ~ phenotypic noise in E coli. Plot of CV ~ mean {Silander, 2012 #1467;Silander, 2012 #1467}

###### Nature Review, biological robustness, Hiroaki Kitano 2004 {Kitano, 2004 #1495;Kitano, 2007 #1488;Kitano, 2004 #1495;Kitano, 2007 #1488}

### Aim 2.4. Experimentally characterize how robustness affects chronological aging dynamics, tolerances to both oxidative stress and genomic instability by loss of heterozygosity, and expressional robustness in ~100 deletion mutants of candidate capacitors, synthetical lethal pairs, and random singleton genes as controls.

Many duplicates has lost backup capacity. {Li, 2010 #2364}

PI Qin's lab has studying genomic instability using loss of heterozygosity (LOH) at the MET15 locus in yeast strains [Qin 08 REF]. Recently, Misses Erin Jackson and Lindsay Parnell, one formal and one current Spelman students, developed a protocol to measure H2O2-dose dependent LOH (see preliminary results).

We currently know the replicative lifespan of 564 deletion mutant {Managbanag, 2008 #563;Kaeberlein, 2005 #486;Kaeberlein, 2005 #306} and CLS measured in synthetic medium for 550 deletion mutants {Burtner, 2011 #779}.

Caveat: It was argued that the release of hidden genetic variance dueto a major mutation or environmental stress does not demonstrate canalization of the wild-type genotype {Hermisson, 2004 #1515}.

The focus is connection between robustness and CR lifespan extension effect. DR extends CLS independent of sirtuins {Smith, 2007 #2024}.

### 2012 May 15, I am here ####

{Andersen, 2008 #906}

Yeast Fitness Database, fitdb.stanford.edu rapamycin effect

Antioxidanat screen {Wu, 2011 #1879}

286 H2O2-sensitive Saccharomyces cerevisiae deletion mutants were screened to identify genes involved in cellular adaptation to H2O2 {Ng, 2008 #1893}

#### CR, Radicicol, rapamycin on DHE, DHR signals

#### Measure RLS (?) and CLS

#### Oxidative stress and LOH

#### Aim 2.4.1. Estimate expressional robustness using GFP for genes with known effect on aging and/or representative levels of mutational robustness.

GFP variance coefficient of variation (CV)

Uri Alon’s GFP paper in E coli {Silander, 2012 #1467}

Noise (as measured in CV) in transporters are argued to be advantages and positively selected. {Zhang, 2009 #1966}

#### Aim 2.4.2. Experimentally test some interaction patterns, especially synthetic lethal pairs and duplicates, on cellular aging in S. cerevisae.

Pick synthetic lethal pairs with high or low interacting degrees in genetic network (I need present preliminary analysis on this).

##### I need to emphasize the my GWS do not rely on rare alleles, in contrast to GWS in human disease studies. In fact, I am testing whether essential genes and genes with strong deletion fitness effect tend to influence aging more than non-essential genes.

###### Use unstructured isolates.

### Aim 2.5. Experimentally test whether robustness-influencing genes tend to be associated with natural variation in lifespan and aging-related traits. {Gat-Viks, 2010 #1972}

Use multiple phenotypes to improve statistis power? For Firsher’s exact test, this seems reasonable. Genotype type rows is the same, but more phenotypes rows (still have sample size problem?). (This can be addressed by simulation study) (model test approach).

GWAS in heterogous SNPs?

PI-based permeability life span

Glucose on DHE, DHR in CLS

Calorie restriction effect on DHE, DHR-signals

L0 variation, LOH -> mutational genetic robustness, Introduce LOH to selected sequenced strains

H2O2 and paraquat tolerance

CLS by CFU and micro-colony analysis

(Bedalov’s microplating assay on lifespan)

#### CR, Radicicol, rapamycin on DHE, DHR signals

#### Aim 2.5.1

#### Aim 2.5.2. Compare the results of GWS with phenotypic measures of deletion mutants and machine learning predictions.

##### GWA-aging (probably not a good idea in NSF proposal, more suited for NIH proposal)

###### Human longevity -GWA study, {Bergman, 2007 #1503;Barzilai, 2010 #1502} A gene interaction sub-network was constructed for longevity-associated genes.

###### {Atzmon, 2008 #1504;Barzilai, 2006 #1505;Atzmon, 2006 #1506;Barzilai, 2003 #1507}

###### {Boger, 2011 #1508;Boes, 2009 #1509;Heid, 2008 #1510;Boes, 2008 #1511}

###### power analysis in gwa {Klein, 2007 #1512}

###### design of gwa {Amos, 2007 #1513}

###### stratification may bring additional bias in yeast gwa.

##### Candidate gene approach

### Aim 2.6. Develop a prototype ODE based model for cellular senescence, proliferation, and ROS hormesis. Brain and Cousens model allows for hormesis {Belz, 2012 #2154}[Brain, Cousen 1989, Weed Res]

H2DCF-DA, DHE, and TO-PRO3 were used to monitor H2O2, O2·-, and viability simultaneously {Cossarizza, 2009 #1288}. H2DCF-DA and PrI were used simultaneously to monitor apoptotic features {Laun, 2005 #1251;Mason, 2005 #1257}. Double stain of live yeast cells with FM4-64, to monitor vacuole, and with Hoechst, to monitor DNA content, can be done by established methods {Kvam, 2006 #1261}. Hoechst and PrI can be used to monitor both live and dead cells {Poot, 1997 #1273}. Invitrogen provides a free iPad app to simulate the multiplexing of fluorophores. We have been using this app to guide experimental designs of flow cytometry.

Rowe thesis page 67 argues that DHR detects more multiple species, DHE is for superoxide, HPF for \*OH, Amplex Red for h2O2.

Preliminary data on H2O2, O2 change during CLS, and CR effect. I need to use SCD because its CLS is shorter than that in YPD.

DHR also detect hydrogen peroxide, nitric oxide, and peroxynitrite

H2O2 can be detected by dihydrorhodamine 123 {Qin, 2008 #2396;Weinberger, 2010 #864;Mesquita, 2010 #851}

{Jamieson, 1998 #875;Carmel-Harel, 2000 #2372}

CR and rapamycin effect, cite Xie’s paper

#### An ODE approach in PNAS {Lorenz, 2009 #478}

(Brief background) There are at least 3 cell cycle checkpoints, G1/S checkpoint, S checkpoint, and G2/M checkpoint. In replicative aging, majority of the cells die with round shape or small buds during replicative aging, indicating they die or are arrested in G1 or S phase {Hartwell, 1974 #1017} based on our empirical observations. In chronological aging, dead cells indicate that they cannot restart the cell cycle to form colonies. So, we will use G1/S and S checkpoints to model cell death in both replicative and chronological aging. G1/S checkpoint has been argued to key control of cell size and the ‘START’ of cell cycle {Hartwell, 1977 #1012;Rupes, 2002 #1868}. To model cell cycle arrest, one option is to use a ‘permanent’ G1 phase. In yeast, the cyclin-dependent kinas activator Cln3 promotes entry into S phase.

It is important to distinguish cellular senescence from cellular quiescence. Quiescence refers to arrest of cell cycle in health cells in unfavorable growth conditions, such as limited nutrients (? Broach and Washburne review on this. Do they have the same definition?). Senescence, we argue, can be modeled as cell arrest in good growth conditions.

(Need Jianhua Xin and Tyson collaboration letter for short visits)

#### Aim 2.6.1. Quantify the glucose dose-dependent changes of H2O2, superoxide, CLN3-GFP (?), cell cycle distribution, and aging-related changes in selection yeast mutants.

ROS change in replicative aging {Lam, 2011 #1880}

Swi6 is a redox sensor {Chiu, 2011 #1881}

SGD only show CLN3 in G1/S transition (CLN3, CLN2 is in Tyson’s model). SGD shows Cln3 regulates Cln1 &2. DNA content and cell cycle distribution by DNA content analysis. Con-focal fluorescence microscope on ROS levels, bud scars, glucose levels.

Chronologled aged cells, Enrich old cells by magnetic beads, and then flow cytometer? Mother-cell enrichment cassette. Imaging Quantification?

Time course of DHE, DHR staining will be carried out.

Growth signal induce superoxide and inhibit quiescence {Weinberger, 2010 #864}.

Mitochondrial hormesis promote ROS , and retrograde ROS signaling {Ristow, 2011 #1034}

CR increase H2O2 and SOD2-> decrease superoxide -> long Chronological lifespan {Mesquita, 2010 #851}. It argues for hormesis effect of H2O2.

H2O2, acetic acid {Ludovico, 2002 #1848;Ludovico, 2001 #1849}, replicative and chronological aging can increase ROS -> Yca1 (yeast caspase1) apoptosis {Madeo, 2004 #1831}.

ROS hormesis implies that the system is also hysteresis: For a cell system already in extreme high superoxide state, bring its H2O2 down to the moderate level should not extend its lifespan.

DNA damage and replication defects in S-phase can lead to apoptosis-like cell death in yeast {Burhans, 2003 #873}

Replicativelly aged cells show apoptosis features {Laun, 2001 #1841}

Chronological aging leads to apoptosis {Herker, 2004 #274}.

Apoptosis pathways are conserved in yeast {Madeo, 2004 #1831}. Expression of mammalian Bax can trigger apoptotic changes in yeast {Ligr, 1998 #1860}.

In other model systems, cell cycle progress and antiapoptosis was shown to be regulated by Pim-1 and c-Myc {Shirogane, 1999 #1850}.

Mitochondria and Bcl-2 can induce growth arrest and mortality in S. cerevisiae {Greenhalf, 1996 #1851}.

Response to rapamycin, a life span extension antibiotic, is also mediated by UTH1, a gene required for Bax-induced cell death in yeast {Camougrand, 2004 #1862;Camougrand, 2003 #1865;Priault, 2002 #1866}.

Sugar (glucose) is likely a limiting nutrient for yeast cells in the wild environment {Granot, 2003 #1853}. Glucose (?) can induce cell death with rapid production of ROS and apoptotic features including RNA and DNA degradation, membrane damage, nucleus fragmentation and cell shrinkage.

The role of cell morphology in mitotic asymmetry and yeast aging is a newly discovered phenomenon {Zhou, 2011 #1157}. As a single cell organism, it is important for mother yeast cells to prevent aging factors, such as damaged proteins, from passing on to new-born daughter cells. Active transport is observed too {Liu, 2010 #958}[Add Nystrom’s active retrograde transport here] Surprisingly, this 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 {Zhou, 2011 #1157}. This observation is consistent with the morphological changes of yeast cells during replicative aging. During microdissection analysis of replicative aging, it can be seen that young yeast cells are generally smaller in size with elliptical shapes. [add pnas paper, microfluidic study, increase size, delayed timing, etc] Old yeast cells become larger, round in shape, and often have membrane blebbing (irregular bulges) {Steffen, 2009 #473;Kennedy, 1994 #19}. Mitotic asymmetry between mother and daughter cells breaks down around the time when most cells lose viability in the population, and leads to much higher levels of genomic instability in the daughter cells {McMurray, 2004 #911;McMurray, 2003 #244}. The morphological changes coincide with the slow-down of cell division during replicative aging. Old mother cells become larger and take longer to divide. Daughter cells from young mother cells are usually much smaller than their mother cells, but daughter cells from old mother cells can often be similar in size to their mothers {Kennedy, 1994 #19}. We observed similar break-downs of mitotic asymmetry during chronological aging of yeast cells {Qin, 2008 #516}. Collectively, these results suggest that the role of cell morphology in aging warrants detailed studies.

#### Aim 2.6.2 Develop an ODE model

#### To read: evolving reliability of yeast cell cycle model {Braunewell, 2009 #1876;Braunewell, 2007 #1878}

#### Cell cycle expression data {Spellman, 1998 #2327}.

#### Cell cycle model (need Tyson support letter)

##### Many cells are arrested but they are not senescent, so the model has to be arrest in the presence of growth signals.

##### Ref {Blagosklonny, 2011 #676}

## Future studies and long-term career goals.

There are a number of directions for the long-term research on network and emergent properties.

A new framework to study gene interactions in quantitative genetics

Statistical genetics

Fitness, robustness, recombination, and the reason for sex in yeast

Robustness, evolvability, fragility

Hormesis of ROS

Simulate aging based on ‘real’ yeast gene works

Hopefully, more experimental data would be available to quantify the gene interaction strength, model, in 5 years.

Recombination and segregation of alleles (sex) into the network model

Mitotic asymmetry

Apoptotic pathway and cellular aging

Age-dependent mutation and age-induced hyper-recombination in age-structured cell populations

Calorie restriction effect on network robustness and gene interactions

Network robustness and drug persistence

Drug persistence of single cell organisms such as bacteria and yeast

Network models of speciation

Evolution of age-structured population

Stable population structure through Leslie matrix

Fertility <-(-)-robustness (more redundancy) –(+)->fitness

Optimal life for a species (using mutational effect, and tradeoff, see Lin Chao’s work).

Age-structured population -> Game theory

Trade-off

Expansion into E. coli and other yeast species, crab-tree negative species Kluyveromyces lactis {Rizzetto, 2012 #2397}

# 5. Educational Plan

The educational plan is essentially the operational plan to carry out the proposed research plans through undergraduate based research and teaching activities. The focus here is to effectively engage undergraduates in cutting-edging research and integrated research into student learning.

## Aim 3. Integrated training on modeling, computing and genome biology to minority students.

### Aim 3.1. Engage minority undergraduates in interdisciplinary research.

This project will mainly be carried out by students at Spelman College, a historically black college for women. Students will be trained in mathematical modeling, computational simulation, genomics, next-generation sequence analysis, experimental genetics, and flow cytometry. The interdisciplinary nature and genome-scale studies of this project will add a sense of excitement to students. This kind of interdisciplinary training and the combination of experimental and computing experience will greatly enhance student preparation for graduate schools. The participated students will gain ownerships to a unique body of data and knowledge through original research in computational and systems biology.

It is important like to emphasize that this project is both practically assessable by undergraduates and intellectually challenging. Students will learn to deal with the frustrations and rewards that will inevitably arise from thinking through the challenging questions and problems posted in this project. Dr. Qin will also strive to instill the fundamental values and ethics of research into the students. By taking ownership of the projects, students may discover their love of research and may choose science-related careers. Dr. Qin especially encourages his students to learn more about genomics and computational biology.

Qin is experienced in training undergraduate researchers for both experimental and computing skills. Qin only accepts student researchers who can accept and sign a learning contract – a document that stipulate expectation of research activities and responsible conducts of both mentees and mentor. Qin requires all students in his lab to learn basic R programming. Data from aging experiments and flow cytometry are routinely analyzed by students using R scripts in the lab. Important experimental and computational methods are also provided to students on Qin’s Youtube educational channel (www.youtube.com/qinstat) to speed up the learning process. All student researchers are required to use GoogleDoc and Dropbox to share experimental protocols, data, and computational analysis. Qin also encourages students to present their findings at various national and international meetings. These meetings will enable our students to present their accomplishments on statistical imaging and cell population studies, and to learn more about graduate schools and career opportunities in science.

Qin has demonstrated records of effectively engaging undergraduates in research. Qin has trained over 35 undergraduate researchers at Spelman in three years, among which four are math majors, two are computer science majors, two won competitive travel awards to attend international meetings, several are pursuing graduate trainings in STEM, and at least two are pursuing Ph.D. training in computing and/or genomics related fields.

### Aim 3.2. Integrate original research in computational and genome biology into courses - FYE, CIS115, BIO125, BIO233, and BIO386.

As a tenure track faculty in the Spelman Biology department, Qin teaches 9 credit/contact hours per semester, which provides excellent opportunity to integrate original research into curriculum. Qin currently teaches First Yeast Experiences (FYE), BIO233 Microbiology, and BIO320(BIO386) Genomics, Proteomics and Bioinformatics. Qin plans to teach CIS115 Introduction to Computing and Informatics and BIO125 Molecular Biology and Genomics in next academic year. The biology core course sequence at Spelman include four 100 level courses. Consequently, BIO125 targets sophomores, BIO233 targets juniors, and BIO386 targets juniors and seniors.

In FYE, Qin plans to cover topics on hypothesis-driven paradigm of scientific research, scientific inquiry as a way of life, responsible research conduction, and controversies in research using SIR2 as an example. FYE targets freshmen and is a recruitment venue for PI.

In CIS115, Qin plans to teach basic bioinformatics programming in Python, regular expression, and computational thinking. Student group projects can include parsing published genomics data from various sources, pipe-line design for large scale sequence analysis, and network/graph analysis. CIS 115 targets freshmen and sophomores and aims to train student with basic programming skills.

In BIO125, Qin plans to introduce original genomics research on cellular aging by taking advantage of the yeast deletion collection and the GFP collection. BIO125 targets sophomores at Spelman and aims to prepare students to carry out independent research projects. The current BIO125 uses yeast as a model to study the impact mutations in MSH2 on DNA repair and focuses on molecular biology. Qin plans to switch the focus to genomics and gene networks. Qin plans to lead students to study LOH, ROS and protein expression changes in response to H2O2 and paraquat using the diploid deletion mutants and GFP fusion strains (aim 2.X??), gene interaction analysis using Cytoscape, protein domain analysis, protein 3D structure visualization using Deep Viewer. Students will also learn to take research note properly and to write research reports. Research integrity and ethics will also be emphasized. With departmental support (See Chair’s letter from Dr. Mark Lee), Qin plans to first developed detailed experimental course plans in a pilot course in the first two project years, and then expand to all four sessions in BIO125. Qin will also parlay his past experience of teaching a HHMI phage genomics course and bioinformatics courses into BIO125.

In BIO233, Qin plans to lead to students work on mini research proposals. In these semester-long writing assignments, students will study genes are known to influence life span but without clear molecular connections to aging. Students will be guided to formulate hypothesis and propose experiments to test them. Students will learn to use bioinformatics tools to identify homolog/orthologs, perform domain analysis, gene/protein interaction analysis, and generate their hypothesis. Students are required to write their mini proposals on wiki based website developed by Qin. Qin has piloted this crowd-sourcing open science approach in Fall of 2011 (http://sunrays.spelman.edu/bgd/wiki/index.php/Main\_Page). One student, Yamisha Rutherford, carried out her proposed research project in Spring 2012 and won a first place for Poster presentation on Spelman Research Day. Qin plans to scales up this crowd-sourcing open science approach and encourage more students to carry out their proposed experiments.

In BIO386, Qin teaches R programming, data analysis, and computational methods. BIO386 targets upper-class students and honor students. BIO386 is based on BIO320 - a project-based computing course that Qin has taught for three years at Spelman. Past students reflected that BIO320 is like a rotation experience in graduate schools and made them prepared for graduate training. Qin revised BIO320 to BIO386 because the new course emphasizes writing – students are required to write manuscript-styled project reports, and because the new course will meet the rigorous requirements of the Spelman Honor Program. Qin plans to lead BIO386 students in groups to carry out the proposed computational modeling, simulation, and analysis in aims 1 and 2.

Qin is experienced in teaching computing to undergraduates and integrate original research into courses. He has developed many hands-on laboratory modules for bioinformatics (see www.bioinformatics.org/ctls). For this project, he plans to develop Youtube short videos to teach basic concepts and skills (see www.youtube.com/qinstat) to facilitate student training and learning.

Overall, Qin fully embraces the notion that learning by investigation is the most effective way of learning, and has extensive experiences of integrating research into curriculum and teaching computing and bioinformatics at undergraduate level. Qin not only has experiences, skills, energy and determination to lead student to carry out the proposed research projects using a crow-sourcing approach, he also has the departmental and institutional support to provide the best available investigation-based learning experiences to Spelman students.

### Aim 3.3. Develop a new course of systems biology for undergraduates.

The core mission of this proposal is to understand cellular aging from the systems perspective. PI Qin believes that an introductory course of systems biology will not only better prepare Spelman students to participate in the proposed research, but also provide a venue to synthesize and disseminate the research findings of this project. In addition, this course can expose many biology students to the power of mathematical modeling and computational analyses.

The proposed course will cover basic concepts of dynamics systems, ordinary differential equations, phase diagrams, and bifurcation analysis. Classical example of toggle switch, cell cycle model will be discussed. Students will learn to use Xppaut and R to study system behaviors. Qin also plans to guide students to develop ODE based models for glucose dependent changes of intracellular H2O2 and superoxide, explore the parameter space to study the dynamic behaviors, and compare the simulations to experimental observations (see aim 2.6??). Qin plans to develop this course by simplifying the CSHL course materials of Computational Cell Biology (Qin attended in 2011). Qin will also learn from the Cellular Biophysics and Modeling course offered at the College of William and Mary. Instructor of this course, Dr. Gregory Smith, has generously shared his entire course materials with Qin. Qin plans to first pilot this new course in Year 3 and 4, and then propose to the college curriculum committee for formal offering in Year 5.

### Aim 3.4. Build a sustainable undergraduate program on computing and modeling through faculty workshops, an undergraduate minor, a student club, and outreach.

The primary goal here is to build a local community of computing and modeling in undergraduate research and teaching, in order to better recruit and retain undergraduates to participate in research with strong aspects of computing and modeling and take courses on computing and modeling. This aim will also generate broad impact that reaches beyond the effort of PI as a single faculty.

For faculty workshops, Qin plans to organize tutorial workshops for faculty to adopt R and other computing and modeling tools in research and education in Year 1 and 3. One software, V-cell, is extremely promising for modeling exercises in classroom and undergraduate research. Dr. Raquell Holmes from University of Connecticut Health Center (V-cell developer) has graciously agreed to visit Spelman to organize a workshop on V-cell and modeling (see her letter of support). PI Qin organized one workshop on computing and gave two workshops on bioinformatics in the past.

To establish a long-lasting presence of computing at Spelman, Qin plans to propose an undergraduate minor of bioinformatics and systems biology. This idea is enthusiastically supported by both biology and computer science departments. CIS115 is a direct result of Qin’s initiative and collaboration with Dr. Alfred Watkins in the computer science department. Tentatively, the required courses of this new minor will include CIS115, the proposed new course on systems biology, BIO386, a bioinformatics course being developed by Dr. Maloney, biostatistics, and core biology requirements. We plan to propose this new minor to Spelman curriculum committee in project year 3.

To cultivate student interest in computing and modeling and establish a presence of computing in Spelman student community, Qin plans to guide a group of students to form a student club of computing, develop a social group on Facebook that provides a forum for students and alumni from the Spelman computing program to share their career experiences. Currently, this Facebook group has four faculty and 12 current and past Spelman students. T-shirts and pens with “Computing @ Spelman” will be designed and distributed for recruitment events held on campus.

To recruit promising high school students to Spelman, PI Qin will also devote significant amount of effort on outreach. Qin has cultivated a partnership with two Science teachers at the Cedar Grove High School with a predominately African American student enrollment. Qin visited Cedar Grove multiple times, and one teacher also visited his lab. In fall of 2011, Qin and two Spelman students designed yeast genetic lab to illustrate the advantage of diploid versus haploid against recessive mutations. The two Spelman students designed a classroom exercise using colored twizzlers. For this project, Qin plans to strength his collaboration with Cedar Grove STEM program and AP course (**see letter of support, Hairston**).

Culture of research

Moving beyond one faculty; a drop of water, to build a wave, and hopefully contribute to sea changes in biology. To build a minor in computational biology with more faculty participation.

More broad base of student pool that will become more interested and more prepared to do research in computational biology, and pursue career in computational biology.

Cellular aging as an emergent property of gene network using math and experimental approaches.

Systems biology is the one of the best examples of quantitative biology, in which {Robeva, 2010 #2384}. Generally systems biology uses several modeling approaches: Difference or differential equations and Boolean networks.

To lay the stepping-stones of career path for computational and systems biologists not just for PI Qin, but for many others as wells.

Earl Achibad, GCAT,

Morehosue minor Bioinformatics core requirements:

3 courses beyond one faculty;

bioinformatics consortium, a step-stone, first step;

## How to assess outreaching effort (need suggestions)

# Synergistic projects:

## Genes under recent selection and their associations with diseases in human populations using the HapMap Phase III project data (with data from African ancestry in Southwest USA).

# Letters

Lee’s support letter, eligibility, two pages

Mohanty letter on genomic instability collaboration

Titus Brown on NGS (budget of visit)

Yuri chernoff letter of collaboration on mentoring, postdoc training, yeast group meeting

Tyson letter of collaboration on cell cycle model (budget of visit and short stay)

Kaeberlein letter on RLS (budget of visit and short stay)

GAS support letter on R workshop

Botstein letter on computing account, yeast expertise, and yeast sequencing

more hosuse support letter, CAU support letter,

# To do:

## Tech quote to HR

## RUI supplment is not allowed.

## Metropolis fitting of Kaeberlein’s data, ask for collaboration letter

## Microscope, pH meter, mg balance, dissection scope for micro-colony counting.

# Key references

Dahmi 2011 paper {Dhami, 2011 #2354} (use complex data from {Krogan, 2006 #2357})

Kaeberlein’s 564 genes on lifespan regulation {Kaeberlein, 2005 #486}

Hsp90

Gavrilet’s realibity model

Flatt’s robustness review, Jonna Masel’s review on robustness

Pleiotropic theory of aging

Laun review paper on ROS

Landry paper. Yeast as a ecology (statistics paper)

Oxidative stress in wine yeast {Orozco, 2012 #2373}

Ty1 element is associated with increased LOH in chronological aging {Maxwell, 2011 #2377}.

Seringhaus, Gerstein, Bayesian prediction of essential genes, 2006 {Seringhaus, 2006 #2379}

Computing has changed biology – biology education must catch up mathematical biology educaiton {Pevzner, 2009 #2382}

Essential mathematical biology, by Nicholas F Britton Physics today, 2010

Mathematical biology education {Robeva, 2010 #2384;Robeva, 2009 #2385}

Why is math biology so hard? Math biology is often perfect for undergraduate research projects{Reed, 2004 #2387}

Education for a biocomplex future {Gross, 2000 #2388}, finding a blance, {Gross, 2004 #2391}

## 1. Qin, H. (2009). Teaching computational thinking through bioinformatics to biology students. *Proceedings of the 40th ACM technical symposium on computer science education*, 188-191.

## 2. Guo, Z., Adomas, A. B., Jackson, E. D., Qin, H. & Townsend, J. P. (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, 345-55.

## 3. Finkelstein, M. & Vaupel, J. W. (2006). The relative tail of longevity and the mean remaining lifetime. *DEMOGRAPHIC RESEARCH* 14, 111-138.

## 4. Kwan, E. X., Foss, E., Kruglyak, L. & Bedalov, A. (2011). Natural polymorphism in BUL2 links cellular amino acid availability with chronological aging and telomere maintenance in yeast. *PLoS Genet* 7, e1002250.

## 5. Proulx, S. R., Nuzhdin, S. & Promislow, D. E. (2007). Direct selection on genetic robustness revealed in the yeast transcriptome. *PLoS One* 2, e911.

## 6. Hermisson, J. & Wagner, G. P. (2004). The population genetic theory of hidden variation and genetic robustness. *Genetics* 168, 2271-84.

## 7. de Visser, J. A., Hermisson, J., Wagner, G. P., Ancel Meyers, L., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., Gibson, G., Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., von Dassow, G., Wagner, A. & Whitlock, M. C. (2003). Perspective: Evolution and detection of genetic robustness. *Evolution* 57, 1959-72.

## 8. Blagosklonny, M. V. (2011). Cell cycle arrest is not senescence. *Aging (Albany NY)* 3, 94-101.

## 9. Silander, O. K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U. & Ackermann, M. (2012). A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli. *PLoS Genet* 8, e1002443.

## 10. Kitano, H. (2004). Biological robustness. *Nat Rev Genet* 5, 826-37.

## 11. Kitano, H. (2007). Towards a theory of biological robustness. *Mol Syst Biol* 3, 137.

## 12. Bergman, A., Atzmon, G., Ye, K., MacCarthy, T. & Barzilai, N. (2007). Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Comput Biol* 3, e170.

## 13. Barzilai, N., Gabriely, I., Atzmon, G., Suh, Y., Rothenberg, D. & Bergman, A. (2010). Genetic studies reveal the role of the endocrine and metabolic systems in aging. *J Clin Endocrinol Metab* 95, 4493-500.

## 14. Atzmon, G., Pollin, T. I., Crandall, J., Tanner, K., Schechter, C. B., Scherer, P. E., Rincon, M., Siegel, G., Katz, M., Lipton, R. B., Shuldiner, A. R. & Barzilai, N. (2008). Adiponectin levels and genotype: a potential regulator of life span in humans. *J Gerontol A Biol Sci Med Sci* 63, 447-53.

## 15. Barzilai, N., Atzmon, G., Derby, C. A., Bauman, J. M. & Lipton, R. B. (2006). A genotype of exceptional longevity is associated with preservation of cognitive function. *Neurology* 67, 2170-5.

## 16. Atzmon, G., Rincon, M., Schechter, C. B., Shuldiner, A. R., Lipton, R. B., Bergman, A. & Barzilai, N. (2006). Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol* 4, e113.

## 17. Barzilai, N., Atzmon, G., Schechter, C., Schaefer, E. J., Cupples, A. L., Lipton, R., Cheng, S. & Shuldiner, A. R. (2003). Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 290, 2030-40.

## 18. Boger, C. A., Gorski, M., Li, M., Hoffmann, M. M., Huang, C., Yang, Q., Teumer, A., Krane, V., O'Seaghdha, C. M., Kutalik, Z., Wichmann, H. E., Haak, T., Boes, E., Coassin, S., Coresh, J., Kollerits, B., Haun, M., Paulweber, B., Kottgen, A., Li, G., Shlipak, M. G., Powe, N., Hwang, S. J., Dehghan, A., Rivadeneira, F., Uitterlinden, A., Hofman, A., Beckmann, J. S., Kramer, B. K., Witteman, J., Bochud, M., Siscovick, D., Rettig, R., Kronenberg, F., Wanner, C., Thadhani, R. I., Heid, I. M., Fox, C. S. & Kao, W. H. (2011). Association of eGFR-Related Loci Identified by GWAS with Incident CKD and ESRD. *PLoS Genet* 7, e1002292.

## 19. Boes, E., Coassin, S., Kollerits, B., Heid, I. M. & Kronenberg, F. (2009). Genetic-epidemiological evidence on genes associated with HDL cholesterol levels: a systematic in-depth review. *Exp Gerontol* 44, 136-60.

## 20. Heid, I. M., Boes, E., Muller, M., Kollerits, B., Lamina, C., Coassin, S., Gieger, C., Doring, A., Klopp, N., Frikke-Schmidt, R., Tybjaerg-Hansen, A., Brandstatter, A., Luchner, A., Meitinger, T., Wichmann, H. E. & Kronenberg, F. (2008). Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* 1, 10-20.

## 21. Boes, E., Kollerits, B., Heid, I. M., Hunt, S. C., Pichler, M., Paulweber, B., Coassin, S., Adams, T. D., Hopkins, P. N., Lingenhel, A., Wagner, S. A. & Kronenberg, F. (2008). INSIG2 polymorphism is neither associated with BMI nor with phenotypes of lipoprotein metabolism. *Obesity (Silver Spring)* 16, 827-33.

## 22. Klein, R. J. (2007). Power analysis for genome-wide association studies. *BMC Genet* 8, 58.

## 23. Amos, C. I. (2007). Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 16 Spec No. 2, R220-5.

References

## Laplacian matrix( googlel ranking?) number of spanning trees (robustness)

### Kirchhoff's theorem or Kirchhoff's matrix tree theorem

Magnetic beads sorting of old cells, {Caballero, 2011 #953}

Asymmetric partition in yeast, septin-dependent diffusion is lost in bud6Delta mutant, {Shcheprova, 2008 #2365}. Bud6Delta can be used as control for DHE, DHR signals?

Active retrograde transport is also involved. {Liu, 2010 #958}

Kirkwood’s model on asymmetry uses fitness, can be used on my modeling approach. It still used a fixed threshold D\* to simulate death. It is ODE model and can be used to study theoretical implications {Rashidi, 2012 #2367}

For ROS modeling: Active and inactive mitochondria are partitioned differently between mother and daughter cells {Klinger, 2010 #1059}. Paraquat and H2O2 –depedent aconitase actitive were reported at 3 concentration. Sigmoidal response is clear in figure 3. Aconitase activity assay, DHE, DHR measures were used.

Heterogeneity of dead cells is quantified by microfluic and continuous monitoring. {Lee, 2012 #2368}. It also show exponential increased of division time and gradual increased of cell size (measured by area under microscope).

Yeast C-limiting array data, 6. Fazio A, Jewett MC, Daran-Lapujade P, Mustacchi R, Usaite R, Pronk JT,

Workman CT, Nielsen J: Transcription factor control of growth rate

dependent genes in *Saccharomyces cerevisiae*: A three factor design.

*BMC Genomics* 2008, 9:341.