Project Description

Research on the biology of aging has demonstrated that, while chronological aging is unavoidable, the rate of physiological aging is strongly influenced by environmental conditions and conserved genetic pathways, and organisms alter the rate at which they age in response to external cues (1). The most potent example of this is dietary restriction (DR), reduced food intake without malnutrition, which slows aging in every organism tested thus far and protects against multiple chronic diseases, including cancer, cardiovascular disease, and neurodegeneration (2)**.** Elucidating how DR promotes healthy aging will allow the development of novel therapeutics to prevent and treat age-related diseases.

Advances in genome editing and genomic measurement technologies have made it possible to rapidly conduct precise interventional experiments and measure genome-wide responses. Further, automated parallel experimentation systems can perform experiments on thousands of model organisms simultaneously. An exciting opportunity exists to significantly accelerate progress in uncovering the molecular mechanisms of diseases of aging, if we can close the loop between experimentation and analysis by using optimal experimental design – that is, adaptively choosing what experiments to perform next to gain the most useful information. However, novel statistical methods are needed for designing optimal experiments in this setting.

In particular, recently developed automated parallel experimentation systems for *C. elegans* (3, 4) have the potential to enable faster and more accurate inference of complex biological networks such as the gene regulatory network mediating the effect of DR on longevity. For instance, the Fontana lab at Harvard Medical School has developed the "Lifespan Machine", an apparatus that performs experiments on tens of thousands of *C. elegans* worms simultaneously, applying individual treatments and measuring results via imaging, all in a completely automated fashion (3). To fully tap the potential of systems like this, **statistical methods are needed that can adaptively choose experiments to maximize the gain of useful information towards solving a given scientific problem**. This would make it possible to build systems for **I**nference, **D**esign of experiments, and **E**xperimentation in an **A**utomated **L**oop (IDEAL, Fig. 1), enabling a massive speed-up in our ability to characterize complex biological systems.

We propose to develop statistical methods and perform laboratory experiments that would provide a proof of concept of IDEAL, focusing on the causal functional network mediating the longevity-inducing effect of DR, using *C. elegans* as a model organism. We propose two aims that each start at different stages of the IDEAL cycle. Aim 1 will leverage existing data sets from the literature and from the Mair lab to build new statistical models and methods, providing analysis and experimental design that will inform future experimental stages. Meanwhile, Aim 2 will start at the experimental stage of the loop, developing automated platforms and novel data sets that will allow us to infer biomarkers of physiological age and predictors of the response to dietary restriction, informing future statistical analysis and design stages. This project will help lay the foundation for future work continuing the cycle between basic science and statistical modeling as each discipline informs the other, supported by federal funding secured using the results generated by this pilot collaboration.

Aim 1. Modeling and Experimental Design Methodology for Inferring Mediators of DR on Healthspan.

Methods to infer a model of the causal network mediating DR via iterative experimentation & analysis.

When studying the molecular mechanisms of a particular disease or therapeutic intervention, such as the healthful effects of dietary restriction, we would ideally like to obtain an accurate mechanistic model of the network of causal relationships among the key transcripts, proteins, genetic variants, epigenetic regulators, and metabolites involved (5). Many methods for inferring gene regulatory networks have been proposed, including ordinary differential equation (ODE) models, Boolean networks, Bayesian networks, and correlation or information-theoretic approaches (6). Often these methods are used to attempt to infer large-scale networks from a fixed set of data, a daunting task. In contrast, we propose to focus on the smaller-scale network mediating the DR response, and crucially, we propose to adaptively conduct new experiments in an iterative cycle between analysis and experimentation.

As an initial step, we will model aggregate steady-state expression levels using linear ODEs. This has the benefit of being biologically plausible, at least approximately (7), and capable of representing causal feedback relationships, which are crucial for homeostasis. We will exploit existing knowledge of functional effects to constrain the model space, by using the Ensembl Variant Effect Predictor (VEP) to predict whether a genetic variant has a functional effect. We will construct an informative prior for which genes are likely to interact, by

using the functional annotation of genomic features and putative gene function available at WormBase (www.wormbase.org), including Gene Ontology (GO) terms describing the associated biological process, cellular component, and molecular function. To further inform the prior for which genes are likely to interact, we will use an extensive deleteome data set for *S. cerevisiae* (Baker's yeast) with genome-wide RNA expression measurements on 1,484 single-gene deletion mutants, covering a quarter of all yeast genes (8).

Network estimation and uncertainty quantification will be performed by modeling aggregate/bulk gene expression values as distributed around the steady-state of the ODE, enforcing a sparsity constraint on the entries of the interaction matrix (which is biologically-motivated, since each gene can only be regulated by a few other genes), and using regularization or Bayes to favor parameter values consistent with the prior information and constraints. *C. elegans* expression and survival data from the Mair lab along with selected high-quality publicly available data sets will be used.

To perform optimal experimental design, it is first necessary to define an objective function that specifies the quantity to be optimized. Since we are interested in understanding and predicting the effect on healthy aging, we define the objective as minimizing the survival prediction error, with mean squared error as the loss function. Given a prediction algorithm taking the inferred mediating variables as input, performing optimal experimental design is straightforward in principle: choose an experiment that maximizes the expected decrease in prediction error. However, in practice this is not easy due to the large number of potential experiments as well as the complexity of the network model and the prediction algorithm. We intend to develop novel statistical methodology to address this problem.

Aim 2. Experiments to Define Biomarkers of Physiological Age and Predictors of Life Expectancy.

Automated Survival Analysis of Variable Responses to DR and Related Anti-Aging Therapeutics.

Finding biological variation that predicts individual responses to a healthspan modifier is not possible using human populations, and economically unfeasible in mammalian systems. Here we use *C. elegans*, which age rapidly in three weeks, as a model organism to assess aging in real time as we apply different interventions. Previous aging studies in *C. elegans* have used one wild type isogenic strain (N2). We have available 40 wild isolate populations that have been extensively sequenced, with around 95,000 single nucleotide variations and 30,000 indels per isogenic strain, on average. We will use automated survival analysis via "Lifespan Machines" to begin to measure the response to genetic, dietary, and pharmacological interventions known to increase lifespan of N2 animals. This proposal will allow us to install one incubator containing 8 flatbed scanners for the lifespan machine and related software (3). We have consulted extensively with the Fontana lab about installation and setup, along with other labs using this setup. Critically, this pilot will allow us to run non-human-scored unbiased experiments to measure *C. elegans* longevity with a high degree of resolution and throughput. This will increase the robustness and resolution of our aging data, and moving forward will provide invaluable tools for IDEAL. For this pilot study we will begin by analyzing how different genetic backgrounds (wild type strains) respond to the lifespan-extending effects of Metformin, which robustly increases lifespan of *C. elegans* and mice (9, 10), and has been suggested to promote healthy aging in humans (11). We will identify input genomic variation between strains that is predictive of the output variation in response to a given treatment. Using CRISPR/Cas9 genome editing, we will then validate predicted effects of genetic variation on response by recreating that allelic change in our reference N2 control line. We expect strains to be 'non-responders' or 'super responders' to metformin, and we expect to be able to CRISPR-edit the N2 baseline strain to mimic these effects after our analysis. Future work will extend these studies to other interventions to determine causal mechanisms that predict response to treatment.

Impact: Our vision is to build an interdisciplinary **Harvard Aging Research Program (HARP)** for (A) tightly integrated laboratory experimentation and statistical analysis to characterize molecular mechanisms of aging in model systems such as nematode worms, mice, and human cell lines, (B) translating findings from model systems to humans using extensive observational data from epidemiological cohorts such as the Rotterdam study, and (C) collaborating to implement healthspan-increasing therapeutics in health policy and human clinical trials. Harvard is uniquely positioned to achieve this vision, having laboratory scientists working on the basic biology of aging, statisticians, epidemiologists, health policy experts, and ready access to hospitals for administering clinical trials. Seed funding for this project would enable us to demonstrate a proof of concept for part (A), focused on DR in the nematode *C. elegans*.

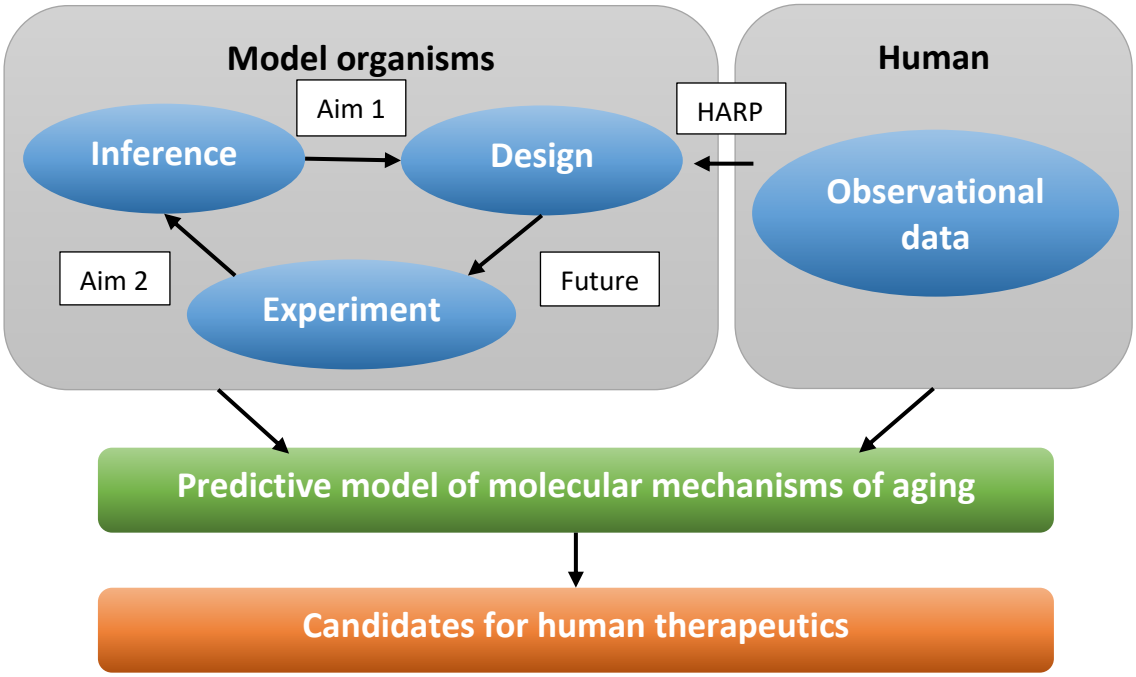

Figure 1. **I**nference, **D**esign of experiments, and **E**xperimentation in an **A**utomated **L**oop (**IDEAL**) Strategy for Harvard Aging Research Program (**HARP**).

1. Fontana L, Partridge L, Longo VD. Extending healthy life span--from yeast to humans. Science. 2010;328(5976):321-6. Epub 2010/04/17. doi: 10.1126/science.1172539. PubMed PMID: 20395504.
2. Mair W, Dillin A. Aging and survival: the genetics of life span extension by dietary restriction. Annual review of biochemistry. 2008;77:727-54. Epub 2008/04/01. doi: 10.1146/annurev.biochem.77.061206.171059. PubMed PMID: 18373439.
3. Stroustrup N, Ulmschneider BE, Nash ZM, Lopez-Moyado IF, Apfeld J, Fontana W. The Caenorhabditis elegans Lifespan Machine. Nat Methods. 2013;10(7):665-70. doi: 10.1038/nmeth.2475. PubMed PMID: 23666410.
4. Kinser HE, Pincus Z. High-throughput screening in the C. elegans nervous system. Mol Cell Neurosci. 2016. doi: 10.1016/j.mcn.2016.06.001. PubMed PMID: 27265309; PMCID: PMC5136351.
5. Hou L, Wang D, Chen D, Liu Y, Zhang Y, Cheng H, Xu C, Sun N, McDermott J, Mair WB, Han JD. A Systems Approach to Reverse Engineer Lifespan Extension by Dietary Restriction. Cell Metab. 2016;23(3):529-40. doi: 10.1016/j.cmet.2016.02.002. PubMed PMID: 26959186; PMCID: 5110149.
6. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models-a review. Biosystems. 2009;96(1):86-103. doi: 10.1016/j.biosystems.2008.12.004. PubMed PMID: 19150482.
7. Chen T, He HL, Church GM. Modeling gene expression with differential equations. Pac Symp Biocomput. 1999:29-40. PubMed PMID: 10380183.
8. Kemmeren P, Sameith K, van de Pasch LA, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW, van Heesch S, Kashani MM, Ampatziadis-Michailidis G, Brok MO, Brabers NA, Miles AJ, Bouwmeester D, van Hooff SR, van Bakel H, Sluiters E, Bakker LV, Snel B, Lijnzaad P, van Leenen D, Groot Koerkamp MJ, Holstege FC. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell. 2014;157(3):740-52. doi: 10.1016/j.cell.2014.02.054. PubMed PMID: 24766815.
9. Cabreiro F, Au C, Leung KY, Vergara-Irigaray N, Cocheme HM, Noori T, Weinkove D, Schuster E, Greene ND, Gems D. Metformin retards aging in C. elegans by altering microbial folate and methionine metabolism. Cell. 2013;153(1):228-39. Epub 2013/04/02. doi: 10.1016/j.cell.2013.02.035. PubMed PMID: 23540700.
10. Martin-Montalvo A, Mercken EM, Mitchell SJ, Palacios HH, Mote PL, Scheibye-Knudsen M, Gomes AP, Ward TM, Minor RK, Blouin MJ, Schwab M, Pollak M, Zhang Y, Yu Y, Becker KG, Bohr VA, Ingram DK, Sinclair DA, Wolf NS, Spindler SR, Bernier M, de Cabo R. Metformin improves healthspan and lifespan in mice. Nat Commun. 2013;4:2192. doi: 10.1038/ncomms3192. PubMed PMID: 23900241; PMCID: 3736576.
11. Barzilai N, Crandall JP, Kritchevsky SB, Espeland MA. Metformin as a Tool to Target Aging. Cell Metab. 2016;23(6):1060-5. doi: 10.1016/j.cmet.2016.05.011. PubMed PMID: 27304507.