Natural genetic and phenotypic variation for aging

Kelly Jin

A dissertation
Submitted in partial fulfillment of the
Requirements for the degree of:

Doctor of Philosophy

University of Washington
2020

Reading Committee:

Daniel Promislow, Chair
Matthew Kaeberlein
Noah Snyder-Mackler

Program Authorized to Offer Degree:

Pathology

University of Washington

Abstract

Natural genetic and phenotypic variation for aging

Kelly Jin

Chair of Supervisory Committee:
Professor Daniel Promislow
Departments of Pathology and Biology

Aging is a complex, highly variable trait influenced by genes, environment, and the interaction between the two. While aging is universal across nearly all species in the animal kingdom, there are many different ways an organism declines in health. As such, in order to fully understand why and how we age, we must explore the variation of aging in populations, both genetically and phenotypically. Historically, two main approaches have been used to try and understand how and why we age, including evolutionary and mechanistic approaches. Evolutionary approaches to studying aging have largely utilized demography, population and quantitative genetics, and mathematical modeling methods to investigate aging from a population perspective. Mechanistic approaches, in contrast, use cellular and molecular methods to drill down to mechanism, usually in the context of highly specific genetic or cellular backgrounds. While these two approaches have each revealed a great deal about both the evolutionary theory underlying *why* we age, as well as specific cellular pathways that may explain *how* we age, there is room for integration of these two complimentary approaches in the current landscape of aging research. An integrated approach to studying aging would incorporate the advantages of population and quantitative genetics, mathematical modeling, and molecular techniques to learn more about aging in individuals within a genetically variable population. Recent advancements in high-throughput genetic and molecular tools, as well as the growing popularity of data sharing, have made such approaches more feasible. In this dissertation, I present three unique but complementary perspectives on how we can use systems biology methods and animal models of genetic variation to learn more about the aging process. In the first study, I investigate the genetic and metabolomic architecture underlying lifespan extension using a collection of 178 genetically variable *Drosophila* lines. In the second study, I introduce a novel model system for

human aging and disease research, the companion dog, and demonstrate that dogs, like humans, show increased levels of comorbidity with age. In my final research chapter, I expand on the versatility of the companion dog as a model for aging by using canine epigenomic profiles to build predictive models of chronological age, and interrogate whether or not those models might also predict the overall health of the animals. Taken together, this collection of highly interdisciplinary studies paves the way for future studies of aging that integrate the advantages of both classical quantitative genetics and cellular approaches to gain a more complete understanding of the biology of aging.

Acknowledgements

For my parents, who overcame so much to provide all the opportunities for my sister and me. For the members of the lab, who helped me in so many different ways. For my friends, who have been unwavering in their support during the hard times. And finally, for Winnie, my constant companion through so many years, and inspiration for much of the work presented here.

Table of Contents

Chapter 1: Introduction

AUTHORS: Kelly Jin[1]

1. Department of Biology, University of Washington, Seattle, WA, USA

Overview

Aging can be defined as the gradual and irreversible decline in health of an organism that is a direct function of how long it has been alive. Aging is widespread across almost all taxa (1), and while aging may be almost universal across the animal kingdom, there are many different ways this gradual decline in health manifests across populations. For example – the lifespan of a fruit fly in your kitchen, the development of cataracts in a dog's eyes, or the number of wrinkles on a person's face – all of these are examples of traits that are associated with age. And as with all complex traits, variation in how these traits age is highly influenced by genetic and environmental components.

The idea that our genes influence how we age is perhaps not a surprising one, and likely evident to each of us in our own lives, as we can see tremendous variation in how people around us age, even those that seemingly live very similar lifestyles. And indeed, estimates of heritability, or the proportion of phenotypic variation that can be explained by genetics, from the lab as well as data from human populations, confirm that there is indeed genetic variation for aging, as measured by traits such as lifespan. In the lab, estimates of heritability of lifespan range from approximately 35% in mice (2) to 41% in fruit flies (3). In human populations, estimates of the heritability of lifespan have historically hovered around 25% (4). These include estimates from twin registries ranging between 20% and 30% (5, 6), although population-based estimates cover a wide range, from 4% to 29%, depending upon ethnic group (7). I use lifespan (age at death) as an example here, but genetic variation likely exists for all age-related traits, many of which, however, are harder to measure and model in a lab. The degree to which age-related traits are influenced by genetics, and how we can use animal models both in and outside of the lab to learn about this variation, is the primary subject of this dissertation.

Evolutionary and mechanistic approaches to studying aging

Over the last century or so, scientists have taken different approaches to discussing and investigating *why* versus *how* we age. The *why* has largely been tackled in the form of evolutionary theory, and supported by foundational concepts from the field of population genetics. Many of these studies were conducted using "classical" quantitative genetics approaches, pre-dating many of the high-throughput sequencing methods that are commonplace in modern research. The question of *how* we age is the one currently receiving the most attention in modern aging research. These studies aim to drill down to specific cellular and molecular mechanisms that underlie aging.

While these approaches have independently contributed a great deal of knowledge to their respective fields, seldom are methods from the two approaches integrated with one another. Now, modern computational and molecular methods make it more feasible to conduct large-scale, systems level studies of genetic variation for aging. In this dissertation, I present three studies that exemplify this approach of merging quantitative genetics and molecular approaches to uncover insights into the aging process at both population and cellular levels.

To begin, I will summarize what we have learned from both evolutionary and mechanistic studies of aging, the advantages of each approach, and the gaps that are still left by both of these two broadly categorized approaches to studying aging.

Mutation accumulation and antagonistic pleiotropy

The fundamental principle underlying the evolutionary theory of aging is the idea that the force of natural selection declines with age, so younger parents have a stronger genetic influence on the next generation as compared to older parents (8). This idea was first alluded to in 1930 in the context of population genetics by Fisher in his work *The Genetical Theory of Natural Selection* (9), and then was explicitly discussed by J. B. S. Haldane, when he proposed it as an explanation for why Huntington's disease, a lethal genetic disorder that has a late age of onset (typically after age 40 (10)), is still so prevalent in the population, despite the dominant lethal nature of the disease-causing mutation (11).

The ideas of Fisher and Haldane led to two related, but distinct theories for the evolution of aging. The first of these is the mutation accumulation (MA) theory, first proposed in 1946, and then further developed in 1952, by Peter Medawar (12, 13). He argued that germ-line mutations will randomly arise in the genome, and most of them will be deleterious or neutral at early ages. MA suggests that mutations that are deleterious at early ages will be eliminated by selection, but the ones that are neutral at early ages with deleterious late-age effects may accumulate due to the decline in the force of selection. Because these mutations are neutral at early ages, natural selection is unable to select against them (13). The second theory was proposed in 1957 by George C. Williams, who argued that the detrimental mutations that cause aging may actually be actively favored by selection if they have a beneficial effect on the organism at early ages (14), resulting in a genetic trade-off between early-life fitness and late-life decrease in health. This theory is known as antagonistic pleiotropy (AP). Both MA and AP theories assume that late-acting deleterious mutations drive the process we know as aging. The theories differ, however, in their assumptions about the early-life effects of these spontaneous germline

mutations. MA assumes that these mutations are neutral in early life, while AP argues that they increase fitness at younger ages.

Studies using animals in the laboratory, as well as natural populations in the wild spanning a wide taxonomic range have tried to find evidence for MA and AP by testing for predictions that each of these theories give rise to (15, 16). The MA model predicts that overall genetic variation for fitness in a population will increase with age, as forces of purifying selection also become less efficient with age (8). Many studies have tested this idea using different fitness traits (reviewed by Promislow and Tatar in 15), including fecundity and mortality, and have found mixed results. Studies that analyze the additive genetic variation for fecundity in *Drosophila* have found both no change in genetic variance with age (17, 18), as well as increases in genetic variation at later age (19). An initial study by Hughes and Charlesworth showed an increase in genetic variation of mortality with age (20), although subsequent work points out possible confounds due to experimental cohort size (19). Historically, testing for evidence of AP has largely revolved around trade-offs between life history traits in natural populations, reviewed nicely by Austad and Hoffman (16). The most common of these observed trade-offs is the negative association between fecundity and longevity, which has been observed in many vertebrate (21, 22), and invertebrate species (23, 24). In the lab, fruit flies are the most commonly used model system for investigating fitness trade-offs. Experimental selection for late-life reproduction results in lines that live longer, which may (25) or may not (26) also coincide with lower early-life reproduction. While measures of reproduction and fecundity vary across studies, making integration of all these findings less straightforward (16), most findings seem to demonstrate the existence of a trade-off between longevity and reproduction. In yeast, worms, and flies, Individual genes have also been demonstrated to extend lifespan but decrease reproduction (16). For example, in flies, hemizygous knockout of *chico*, the fly insulin receptor gene, increases lifespan and decreases fertility, while homozygous deletion of the gene results in sterile, but long-lived females (27). Taken together, studies investigating aging in the context of evolutionary theories, only a small handful of which are discussed here, offer insight into the relationship between genetic variation and aging at a population level.

The hallmarks of aging

While evolutionary approaches primarily have made use of population and quantitative genetics and mathematical theory, modern aging research is dominated by questions revolving around molecular mechanisms of aging. I will refer to these types of studies as mechanistic approaches to aging research. The primary goal of these studies is to identify specific cellular

pathways and molecules that directly influence age-related traits, with the ultimate goal of identifying clinical interventions to help people stay healthier for longer. Many comprehensive reviews now exist that cover cellular mechanisms that may play a role in how we age (28-30). Here, I will highlight one in particular that nicely summarizes these collective findings into nine discrete candidate processes that are now known as the nine “hallmarks of aging” (29). These hallmarks include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intracellular communication. There is an extensive literature on each of these hallmarks (29), and I will not cover them all in this introduction. Instead, I will briefly discuss two hallmarks, deregulated nutrient sensing (with a focus on the effect of dietary restriction (DR) on aging) and epigenetic alterations (with a focus on DNA methylation), as they are most relevant to the chapters included in this dissertation.

The first of these two factors, deregulated nutrient sensing, was shown to play a critical role in aging as far back as 1936 (31), with the first documented evidence of DR-mediated lifespan extension in rodents. Since then, the health and lifespan extending effects of DR have been demonstrated in organisms ranging in complexity from single-celled yeast (32) to non-human primates (33, 34) and humans (35). In *Drosophila*, DR has been shown to significantly extend maximum lifespan by 15-46% (36). As a result, DR is considered by many to be the most robust and efficient form of health and lifespan extension yet known. DR, which can be generally defined as a reduction in nutrient availability without subjecting the organism to malnutrition, is thought to influence longevity through the insulin and IGF-1 signaling (IIS) pathway. The IIS pathway is conserved across almost all organisms in which it has been studied (37), and acts on multiple downstream molecular targets including the FOXO family of transcription factors and the mTOR complex, molecules that are also highly conserved through evolution (37). Despite these highly conserved effects, a handful of studies, including Chapter 1 in this dissertation, are beginning to show that the effects of DR are not universal and are genotype dependent (38-40), highlighting the need to consider genetic variation when investigating potential interventions for aging.

The second major focus of the work described here concerns the wide range of epigenetic alterations documented to occur with age. The epigenome can be defined as the collection of chemical modifications that influence gene expression that are not a result of the actual DNA sequence. These epigenetic alterations include changes in DNA methylation, post-translational histone modifications, and chromatin remodeling (41). Many of these changes ultimately influence how transcriptionally “open” (euchromatic) or “closed” (heterochromatic) the

DNA becomes. General loss of constitutive heterochromatin is associated with age (42). However, most epigenetic changes are highly tissue- and genotype-dependent (41), and therefore, very complex in nature. Of these different modifications, DNA methylation in particular has received a great deal of interest in the last decade, as changes in a handful of DNA methylation sites have been used to build a very accurate predicator of chronological age in both mice and humans (43). Some argue that in addition to being able to predict chronological age, these "clocks" may also inform us about "biological age", or some measure of overall organismal health (43).

Integrating theory and mechanism

While both evolutionary and mechanistic approaches have independently taught us a great deal about how and why we age, integration of these two approaches is currently lacking in the field of aging research. Evolutionary approaches address patterns of aging at the population level, but largely fall short in their ability to point to specific mechanisms of aging, and fail to identify clinical targets for interventions of treatments for age-related disorders. On the other hand, mechanistic studies have identified many specific molecules and pathways that directly modulate age-related traits. While these molecular studies have identified many specific cellular pathways, some of which have led to the development of potential therapeutic interventions for aging, these studies usually ignore genetic variation. They are mostly performed in inbred, isogenic animals that have lived in the lab for many generations. As such, some argue that findings from these experiments may not be relevant to human populations, which are obviously more genetically and environmentally diverse than inbred lab animals. This argument is supported by studies that suggest that the lifespan-extending effects of some aging interventions may be an artifact of lab adaptation of many of the common genetic strains on which these experiments are performed. For example, Harper *et al*. performed DR on grand-offspring of wild- caught mice and found no effect of DR on mean longevity (44). Two possible interpretations of this result include that 1) the DR effect may not apply to animals that have not undergone selection under lab conditions, and/or 2) there is genetic variation for the effect of DR in wild populations, a claim supported by several studies (44). in another study, research found that in the fruit fly gene *Indy*, mutants of which had previously been shown to result in very long-lived flies (45), this lifespan extension was abolished when the mutation was backcrossed into outbred genetic backgrounds (46). Studies such as these suggest that findings from inbred lab-adapted animals may not be broadly applicable to a population of genetically diverse individuals. As such, there is a strong need for studies in the biology of aging that

integrate the perspectives of both of these approaches – leveraging genotypic and phenotypic variation of age-related phenotypes at the population level, along with high-throughput molecular methods – to identify mechanisms underlying aging. Such approaches have the potential not only to point to specific pathways underlying the aging process, but perhaps more critically, they also shed light on how these processes may *vary across a population*, information that is highly relevant in our current "Precision Medicine" era of clinical research.

In order to move forward with such an approach, we must first get a sense of what is known thus far about genetic variation for aging. To this end, in the following sections, I will highlight some large-scale studies of aging that feature genetic variation as an integral component in their study design, what we have learned from them, and what still remains to be discovered. Specifically, I will discuss studies of human and animal models of genetic association of lifespan and lifespan extension. I will then discuss some current challenges in the field, identify potential ways forward to address them, and briefly introduce the remaining chapters in this dissertation and how they address some of these unanswered questions in this rapidly developing field of research.

Genetic variation of aging

Human genome association studies of aging

Various approaches have been used to try and identify genetic loci that regulate human lifespan. These include studies of candidate gene identified from animal models (reviewed thoroughly in (47) and (48)), linkage mapping of long-lived families (49, 50), and genome-wide association studies (GWAS) (51-53). Many candidate genes identified from animal studies have also been found to be associated with longevity in humans, including *CETP*, *HSF2*, sirtuin (*SIRT*) genes, *APOE* genes, and *FOXO3A* (47). Variants found in *APOE,* a gene encoding an apolipoprotein and main cholesterol carrier in the brain, and *FOXO3A,* a gene for a transcription factor in the forkhead family that is linked to many different cell maintenance pathways and is a key regulator of mTOR signaling, have been most consistently replicated, although results vary based on populations (47). Both linkage studies and GWAS test for the association of certain genetic markers (for linkage studies, they are genetic regions, while for GWAS, they are usually single nucleotide variants) with a phenotype of interest. These approaches rely on genetic variation within large populations to identify regions or genes of interest in an unbiased way. The major findings and the challenges faced by both of these approaches are summarized here.

Earliest reported linkage studies of healthy aging identified many putative regions associated with longevity. One region was replicated between at least two studies (50, 54), and subsequently narrowed down to the gene microsomal triglyceride transfer protein (*MTP*) as a candidate gene for longevity (55). The largest linkage study of longevity to date was conducted in the multi-site European Genetics of Healthy Aging (GEHA) Study, which examined 2118 European full sib-pairs over the age of 90 (49). The GEHA study found four loci that were associated with longevity. One of these loci was on the 19q chromosomal region, and was subsequently finely mapped using GWAS in a separate long-lived population to reveal variant alleles at *APOE4* and *APOE2* that accounted for the linkage at that locus. While the GEHA study is frequently discussed among linkage studies of longevity, likely due its identification of *APOE* alleles that replicate findings from other forms of genetic studies of human aging, the majority of linkage studies have been largely unable to replicate findings with one another (56).

Across all GWAS of human lifespan to date, the *APOE* locus is by far the most consistently replicated genetic region to be associated with lifespan that meets genome-wide statistical significance (generally $P \leqslant 5 \times 10^{-8}$) (56). The first of these studies includes one published in 2011 and conducted on 763 long-lived (age of death between 94 – 110 years) and 1085 control (age of death between 45 – 77 years) cases from German biobanks, with replication from a similarly sized German cohort (57). The association of *APOE* with longevity has since been replicated across a multitude of independent GWAS, including people of European (51-53) and Han Chinese (58) descent. The two largest GWAS of longevity to date include a meta-analysis of 14 studies originating from seven European countries covering a total of 7729 long-lived participants ($\geqslant$85 years) and 16121 younger controls (<65 years) (52), and another meta-analysis of 6036 long-lived cases ($\geqslant$90 years) and 3575 controls that died between ages 55 and 80 that were part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, also of European ancestry (53). After *APOE*, the *FOXO3A* locus has also been found to be significantly associated with lifespan and other longevity traits across multiple studies (53, 59, 60), although not to the same extent as *APOE*. While many other GWAS have identified novel candidate genes for longevity, they are largely unable to be replicated between studies, and all show relatively small effect sizes (61). The lack of agreement seen within linkage studies and GWAS may be due to a variety of factors, including power limitations due to small sample size and variable measures of longevity between studies.

Although informative, human studies of healthy aging face many challenges, particularly in study design. Longitudinal sampling of individuals offers greater control over sampling bias, and larger cohort sizes improve statistical power. However, longitudinal samples take time to collect, particularly for studies of extreme old age, and may not be practical for most research timelines. To my knowledge, the largest studies of long-lived individuals are in the low thousands, which is notably smaller than large studies of complex traits such as BMI (62) and height (63), which often have cohort sizes of more than a hundred thousand subjects. To address these challenges in sampling power and study design, some researchers have begun to explore the use of animal models of genetic variation to study aging, both in and outside the lab

Animal models of genetic variation of lifespan

In addition to human studies of genetic variation of lifespan, studies of non-human animal models of genetic variation have also revealed a great deal about genetic variation of age-related traits, particularly lifespan and lifespan extension. In this section, I will briefly discuss strategies used and progress that has been made using these models to study lifespan and lifespan extension, both in and outside of the laboratory environment. Within the lab, I will focus primarily on studies conducted using genetic reference populations from invertebrate models.

Numerous studies have used various animal models to investigate the genetics of natural variation of lifespan. The majority of these studies have been conducted using invertebrate animal models including yeast, nematodes, and flies. Invertebrate models are powerful for aging research for many reasons, including their short-lived nature, large population sizes allowing for large experimental numbers and better phenotypic estimates, and availability of extensive genetic and molecular resources. As in humans, quantitative trait locus (QTL) mapping has been conducted using bi-parental crosses of different strains to try and identify genetic markers of longevity. In one of these studies using *Drosophila*, QTL mapping of a set of 98 recombinant inbred lines (RILs) derived from two fly strains revealed five autosomal QTLs associated with lifespan (64). Follow-up investigations into one of these loci identified the gene Dopa decarboxylase (*Ddc*), an enzyme involved in the synthesis of neurotransmitters dopamine and serotonin, as a candidate gene for lifespan in these strains (65). In 2012, QTL mapping of RILs from a cross between two yeast strains, one lab strain and one clinically derived strain, identified five QTL associated with replicative lifespan, one of which contained the gene *SIR2*, a well-known regulator of replicative lifespan in yeast (66).

While QTL mapping efforts have identified novel and known genes associated with aging, bi-parental crossing schemes only interrogate variation from two genotypes. To address this limited scope of genetic variation, the last decade or so has seen the development and increased usage of experimental genetic reference populations, typically made up of large numbers of synthetic or wild-derived lines of the same species (67-69). Many of these populations have been genetically characterized and therefore serve as extremely powerful tools for mapping quantitative traits within a laboratory setting. Within the invertebrate models, perhaps the most widely used of these reference populations is the *Drosophila* Genetic Reference Panel (DGRP). The DGRP is comprised of approximately 200 *D. melanogaster* inbred lines, each derived from a single inseminated female that was caught from the wild (67, 68). At least three studies have now used GWAS of DGRP lines to investigate lifespan (3, 70, 71). All three estimate the broad sense heritability of lifespan to be between 0.3 and 0.4. The first of these three was conducted in 2014 by Durham *et al*, where they used the DGRP to investigate age-specific fecundity and lifespan (70). They found variants associated with 21 genes using a relatively loose cutoff ($P < 1E-5$) that were all novel for lifespan, but none were validated experimentally (70). While Ivanov *et al.* (3) found enrichment of some genes in the TOR pathway amongst the top ranked variant hits, they were unable to detect any variants that were associated with lifespan that met genome-wide significance. The DGRP has also been used to demonstrate environmental interactions between genetic variation and lifespan. Most recently, in 2020, Huang *et al.* (71) used the DGRP to investigate lifespan in different environmental contexts, specifically three different temperatures. They reveal significant genetic variation for lifespan within each sex and temperature condition, and interactions between genotype and sex and environment. This study is also the first of the three to experimentally validate candidate genes using RNA interference. All of the 15 genes they assess were found to influence lifespan or the interaction between lifespan and sex or environment (71). While these only represent a handful of the many experiments that have incorporated some form of genetic variation into their experimental design to investigate lifespan, collectively they demonstrate a creative approach of interrogating the genetic basis of these very complex trait within the lab.

In addition to lifespan, another age-associated trait that has been studied in the lab in the context of genetic variation is the lifespan extension in response to DR. The first of these studies was conducted in mice in 2010 by Liao *et al.* (39) across a panel of 41 RILs. They found that the majority of the lines did not live longer on DR, and genetic association of these lines found association of fat maintenance with response, where genotypes that maintained fat body levels showed the greatest lifespan extension (72). In 2013, a similar experiment was conducted

in 166 single-gene deletion yeast strains, which also revealed a wide range of response from 79% reduced lifespan to 103% increased lifespan, and that strains with a positive lifespan response to DR were enriched for genes involved in mitochondrial function (38). In flies, there have been at least three studies that have used a collection of genotypes to study DR response (40, 73, 74). The largest and most recent of these was conducted by Wilson *et al.* in 2020 (40), assessing lifespan and climbing ability response to DR in more than 150 DGRP lines. They found that only 50% of strains showed a positive response to DR for both lifespan and climbing, and 35% of strains showed a negative response for both. The results from Wilson *et al.* are expanded upon in Chapter 2 of this dissertation. Collectively, these studies challenge the notion that lifespan extension by DR is universal for all genotypes, further underscoring the value of integrating genetic variation into laboratory studies of aging interventions.

The vast majority of aging studies and, as a result, most what we know about aging come from yeast, worms, flies, and mice in the lab. However, there are limitations to studying aging in laboratory models alone. These limitations include the fact that most laboratory strains are inbred and live in highly controlled and sterilized conditions, which are obviously very far from conditions seem by natural populations. As a result, some have turned to non-laboratory species to study aging. Outside of the lab, the primary approach that has been used to investigate genetic variation of lifespan has been cross-species comparative studies. Across the animal kingdom, large species tend to live longer than small species, while the trend is opposite within species – small individuals within a species live longer than larger individuals on average (75). As such, comparing different species across and within taxonomic clades has the potential to reveal interesting insight into the aging process. Comparative studies of long-lived rodents have been especially valuable in contributing to our understanding of aging mechanisms. The rodent clade is well-suited for comparative aging studies as it contains species that vary widely in body mass and lifespan (75). Disease susceptibility also varies across rodent species, as seen in cancer-prone mice, versus naked mole rats and blind mole rats that almost never get cancer (75). One intriguing finding from these studies suggests that different species of rodents evolved different tumor suppressor mechanisms depending on their body mass and lifespan. Replicative senescence is associated with large body mass, protecting these larger long-lived species from cancer, but leaving them vulnerable to other aging phenotypes associated with cellular senescence (76). In contrast, smaller long-lived species have evolved different mechanisms to protect against cancer, such as high levels of hyaluronan, a polysaccharide that has antitumor and anti-inflammatory properties, in the naked mole rat (77). Recently, the availability of more rodent genomes has also made comparative genomics of rodents more

feasible. In 2018, Sahm *et al.* (78) compared the genomes and transcriptomes of 17 rodent species in hopes of finding positively selected genes associated with longevity. They identified 250 of these genes, some of which were associated with canonical aging pathways including mTOR and IIS pathways, as well as novel candidate genes of aging (78).

There have also been many studies that compare molecular profiles across long-lived and short-lived mammalian species. These include comparative studies of the transcriptome, metabolome, and lipidome, many of which are reviewed nicely by Ma & Galdyshev (79). Collectively, some common longevity-associated pathways have been observed from these cross-species studies, including the role of p53 and associated cell cycle genes showing unique amino acid and gene expression changes in exceptionally long-lived species, including the naked mole rat (77) and African elephants (80). Downregulation of the IIS pathway and upregulation of immune response also seem to trend among long-lived species (79).

Current challenges in genetic variation approaches to study aging

Missing heritability and systems biology

Despite the numerous GWAS on human longevity that now exist, collectively they have failed to account for the full heritability of human lifespan, as the sum of the variance explained by all identified variants is less than most estimates of total heritability (81, 82). This problem of "missing heritability" plagues many genome-wide studies of complex traits, not just lifespan, and may be due to many different factors (81). Some argue that the majority of genetic effects may be non-additive and actually due to complex, epistatic interactions between loci, which cannot be identified in most GWAS (81). Others suggest that genetic effects may actually be epigenetic, or alternatively, due to rare variants that are usually excluded from GWAS of common variants (81). There are also concerns that previous estimates of heritability from twin studies may be overestimating heritability values for most traits (81, 82). While the true explanation of missing heritability may be due to a combination of all of these factors and others, this does not change the fact that candidate genes identified from GWAS of human lifespan and longevity have largely failed to result in actionable insights into direct mechanisms of aging. Furthermore, several studies have also shown that extremely long-lived individuals do not lack risk alleles for common complex diseases (83, 84). Altogether, this suggests that factors that directly influence extreme longevity may lie beyond the realm of common genetic variants.

To address this concern, many researchers have turned to systems biology methods to try and bridge this gap between genotype and phenotype. A relatively new and actively developing field, I define systems biology as the use of biological information from multiple "levels" of the organism (e.g., genome, epigenome, metabolome) paired with computational and mathematical methods to model complex biological phenomena. The rationale behind this approach is rooted in the fact that the biochemical and cellular path from genes to ultimate phenotype is a long one, and that we can learn more by integrating additional molecular measures of cellular signaling with genomic information. By definition, systems biology is a highly interdisciplinary approach that requires a large amount of data and computational resources. The development of high-throughput genetic and molecular tools, as well as the growing popularity of data sharing, have made such approaches more feasible.

Chapters 2 and 4 in this dissertation bring this systems biology approaches to the study of aging. In chapter 2, I integrate two complex levels of cellular information, the genome and the metabolome, using the classic genetic model species *Drosophila melanogaster* to reveal candidate molecular mechanisms that might underlie the lifespan response to DR. My colleagues and I use a combination of molecular, network, and GWAS methods to draw connections between genes, metabolites, and variation in lifespan extension. Not only did this approach identify some known aging-associated genes, but it also revealed novel candidate genes and pathways, lending credence to the application of this interdisciplinary strategy for studying aging or any complex quantitative trait. This work appears in a 2020 issue of PLoS Genetics. Chapter 4, which describes the process of building an age-predictive model using two different epigenome profiles, represents an entirely different way of using a systems approach to studying aging. Here, my collaborators and I use data from an exciting new model species for studying genetic variation of aging—the companion dog (described in further detail below)—to explore the relationship between the epigenome, aging, and the environment. Together, these chapters represent two novel and creative ways to integrate modern day molecular and computational methods with classical population genetic theory to learn more about how genetically variable populations age.

The dog as a model for genetic variation of aging

We have learned a great deal about the genetic basis of aging using both controlled lab animals, as well as comparative, cross-species approaches. However, both of these approaches have their own drawbacks. Lab strains, even those that are derived from wild-caught animals, are usually inbred, and have lived in highly controlled laboratory conditions for

many generations – qualities that obviously do not recapitulate those of natural populations. While cross-species studies address some of these concerns, comparative studies suffer from other kinds of challenges. Aside from differences in sample collection and genetic information, different species are, indeed, very different from each other in many ways. Aside from longevity, they also differ genetically, environmentally, and physiologically, making it very difficult to be sure that novel molecular correlations with longevity are not confounded with other traits.

An exciting, emerging model species addresses many of these challenges: the companion dog, *Canis lupus familiaris*. The domestic dog is a promising and extremely powerful model for aging research, with great translational potential. Our dogs share our environment in a way that cannot be replicated in the lab, allowing us to study a population exposed to the same types and levels of environmental factors to which people are exposed (e.g., second hand smoke, air pollution, ambient noise). In addition, the quality of medical care and documentation for dogs is second only to that of humans, giving researchers an abundance of data sources to mine, including data from veterinary, corporate, and academic sources. Recent studies of canine health and aging include demography of dog populations as functions of sex, sterilization status, and inbreeding (85, 86), demonstrating the feasibility of obtaining and analyzing veterinary data to learn more about canine health and human health through comparative epidemiological approaches. Furthermore, dogs die of similar diseases as humans do (87), suggesting that many things we can learn about canine health will not only benefit dogs, but will invariably benefit us as well. Dogs may also be able to inform us about health and lifespan-extending interventions. In 2017, Urfer *et al.* (88, 89) conducted the first randomized clinical trial testing the effects of short-term, low-dose rapamycin, a pharmacological inhibitor of mTOR, on pet dogs. They found improved measures of cardiac aging in drug-administered animals. This is the first evidence of beneficial health effects of low-dose rapamycin in a non laboratory-raised organism. Studies such as these recognize that, under highly ethically scrutinized clinical settings, companion dogs are an actionable population for potential aging interventions, as many people would support scientific efforts to give their cherished companions healthier and longer lives.

In addition to the environmental and pathological similarities that dogs share with us, companion dogs also serve as an excellent model for aging research due to their unique population structure. Evidence of the first undisputed domesticated dogs date back to approximately 15,000 years BP (90), although some researchers estimate dog domestication may have started as early as 30,000 years ago (91). As such, many thousands of years of artificial selection have resulted in the unique breed-based population structure that we see

today in the modern domestic dog. The dog is the most phenotypically diverse mammal in existence today (92), as evidenced by the extraordinary amount of diversity between dog breeds across morphological and pathological traits. This extreme level of phenotypic heterogeneity between breeds coincides with a high level of genetic homogeneity within breeds (92). This level of genetic homogeneity, although not as extreme as the level of inbreeding seen in laboratory lines, gives researchers some level of confidence about individual breeds and the traits that are associated with them. Two of these traits that are relatively homogenous within breeds, but highly variable across breeds, are lifespan and size. Size, which is inversely correlated with life expectancy across breeds, has been demonstrated to be highly associated with an allele in the *IGF1* gene, a major signaling molecule in the IIS pathway (93). Furthermore, detailed pedigree and breeding information exists for many pet dogs, including mixed breeds and hybrids, allowing for the potential to interrogate the genetic basis for complex traits in a detailed fashion. For all these reasons, the companion dog provides us with a rich landscape for investigating aging.

Chapters 3 and 4 of this dissertation represent studies that use data from companion dogs to learn about aging and human health. In chapter 3, I ask whether or not a population of companion dogs shows patterns of increased multiple-morbidity count (that is, the number of diseases/disorders that an individual has) with age, a phenomenon that has long been observed in human populations, but seldom explored in research settings. We show that, indeed, like people, dogs accumulate more diseases as they age, further demonstrating our ability to model human health in ways that cannot be achieved through other model systems. This work was published in 2016 in Pathobiology of Aging and Age-Related Diseases. In chapter 4, as briefly discussed above, I build epigenetic clocks with data collected from the blood of companion dogs. As previously discussed, dog breeds vary widely in their average life expectancy, making them an interesting species in which to build and test hypotheses about epigenetic age predictors. For example, we may hypothesize that the clock for a larger, short-lived breed may "tick faster" than that of a small, long-lived breed. Chapter 4 lays the conceptual and methodological groundwork for what will no doubt be many future studies to come that interrogate the relationship between canine genetics and aging through the Dog Aging Project.

Closing remarks

In designing and implementing the studies described in this dissertation, my overarching goal has been to integrate multiple levels of biological information in an effort to better understand the mechanisms that underlie genetic and phenotypic variation of aging. As such,

the following chapters represent case studies of highly interdisciplinary approaches to investigating aging. They describe methods and concepts taken from the fields of quantitative genetics, molecular biology, and network modeling. It is my hope that the reader will learn something they did not previously know about the aging process, and be motivated to consider genetic variation when thinking about complex biological phenotypes.

References

1. A. A. Cohen, Aging across the tree of life: The importance of a comparative perspective for the use of animal models in aging. *Biochim Biophys Acta Mol Basis Dis* **1864**, 2680-2689 (2018).
2. B. A. Rikke, C. Y. Liao, M. B. McQueen, J. F. Nelson, T. E. Johnson, Genetic dissection of dietary restriction in mice supports the metabolic efficiency model of life extension. *Exp Gerontol* **45**, 691-701 (2010).
3. D. K. Ivanov *et al.*, Longevity GWAS Using the Drosophila Genetic Reference Panel. *J Gerontol A Biol Sci Med Sci* **70**, 1470-1478 (2015).
4. J. M. Murabito, R. Yuan, K. L. Lunetta, The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. *J Gerontol A Biol Sci Med Sci* **67**, 470-479 (2012).
5. A. M. Herskind *et al.*, The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* **97**, 319-323 (1996).
6. M. McGue, J. W. Vaupel, N. Holm, B. Harvald, Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *J Gerontol* **48**, B237-244 (1993).
7. J. H. Lee *et al.*, Genetic influences on life span and survival among elderly African-Americans, Caribbean Hispanics, and Caucasians. *Am J Med Genet A* **128A**, 159-164 (2004).
8. M. R. Rose, *Evolutionary biology of aging* (Oxford University Press, New York, 1991), pp. ix, 221 p.
9. R. A. Fisher, *The genetical theory of natural selection* (Dover Publications, New York,, ed. 2d rev., 1958), pp. xiv, 291 p.
10. N. S. Wexler *et al.*, Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci U S A* **101**, 3498-3503 (2004).
11. J. B. S. Haldane, *New paths in genetics* (Allen & Unwin, London,, 1941), pp. 206 p.
12. P. B. Medawar (1946) Old age and natural death. (Modern Quarterly), pp 30-56.
13. P. B. Medawar (1952) An Unsolved Problem of Biology (Londonm UK: H. K. Lewis).
14. G. C. Williams (1957) Pleiotropy, natural selection, and the evolution of senescence. (Evolution), p 398.
15. D. E. Promislow, M. Tatar, Mutation and senescence: where genetics and demography meet. *Genetica* **102-103**, 299-314 (1998).
16. S. N. Austad, J. M. Hoffman, Is antagonistic pleiotropy ubiquitous in aging biology? *Evol Med Public Health* **2018**, 287-294 (2018).
17. M. Rose, B. Charlesworth, A test of evolutionary theories of senescence. *Nature* **287**, 141-142 (1980).
18. M. R. Rose, B. Charlesworth, Genetics of life history in Drosophila melanogaster. I. Sib analysis of adult females. *Genetics* **97**, 173-186 (1981).
19. D. E. Promislow, M. Tatar, A. A. Khazaeli, J. W. Curtsinger, Age-specific patterns of genetic variance in Drosophila melanogaster. I. Mortality. *Genetics* **143**, 839-848 (1996).
20. K. A. Hughes, B. Charlesworth, A genetic analysis of senescence in Drosophila. *Nature* **367**, 64-66 (1994).
21. J. F. Lemaître *et al.*, Early-late life trade-offs and the evolution of ageing in the wild. *Proc Biol Sci* **282**, 20150209 (2015).
22. M. R. Robinson, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, L. E. Kruuk, Live fast, die young: trade-offs between fitness components and sexually antagonistic selection on weaponry in Soay sheep. *Evolution* **60**, 2168-2181 (2006).
23. T. Aigaki, S. Ohba, Individual analysis of age-associated changes in reproductive activity and lifespan of Drosophila virilis. *Exp Gerontol* **19**, 13-23 (1984).
24. N. D. White, R. J. Bell, Effects of mating status, sex ratio, and population density on longevity and offspring production of Cryptolestes ferrugineus (Stephens) (Coleoptera: Cucujidae). *Exp Gerontol* **28**, 617-631 (1993).
25. L. S. Luckinbill, M. J. Clare, Selection for life span in Drosophila melanogaster. *Heredity (Edinb)* **55 (Pt 1)**, 9-18 (1985).
26. L. Partridge, K. Fowler, DIRECT AND CORRELATED RESPONSES TO SELECTION ON AGE AT REPRODUCTION IN DROSOPHILA MELANOGASTER. *Evolution* **46**, 76-91 (1992).
27. D. J. Clancy *et al.*, Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* **292**, 104-106 (2001).
28. B. A. Benayoun, E. A. Pollina, A. Brunet, Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol* **16**, 593-610 (2015).
29. C. Lopez-Otin, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* **153**, 1194-1217 (2013).
30. C. Lopez-Otin, L. Galluzzi, J. M. Freije, F. Madeo, G. Kroemer, Metabolic Control of Longevity. *Cell* **166**, 802-821 (2016).
31. C. M. McCay, M. F. Crowell, L. A. Maynard, The effect of retarded growth upon the length of life span and upon the ultimate body size. *Journal of Nutrition* **10**, 63-79 (1935).
32. J. C. Jiang, E. Jaruga, M. V. Repnevskaya, S. M. Jazwinski, An intervention resembling caloric restriction prolongs life span and retards aging in yeast. *FASEB J* **14**, 2135-2137 (2000).

33. R. J. Colman *et al.*, Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201-204 (2009).
34. R. J. Colman *et al.*, Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nat Commun* **5**, 3557 (2014).
35. W. E. Kraus *et al.*, 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory outcomes of a multicentre, phase 2, randomised controlled trial. *Lancet Diabetes Endocrinol* **7**, 673-683 (2019).
36. L. Partridge, M. D. Piper, W. Mair, Dietary restriction in Drosophila. *Mech Ageing Dev* **126**, 938-950 (2005).
37. A. Ruetenik, A. Barrientos, Dietary restriction, mitochondrial function and aging: from yeast to humans. *Biochim Biophys Acta* **1847**, 1434-1447 (2015).
38. J. Schleit *et al.*, Molecular mechanisms underlying genotype-dependent responses to dietary restriction. *Aging Cell* **12**, 1050-1061 (2013).
39. C. Y. Liao, B. A. Rikke, T. E. Johnson, V. Diaz, J. F. Nelson, Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. *Aging Cell* **9**, 92-95 (2010).
40. K. A. Wilson *et al.*, GWAS for Lifespan and Decline in Climbing Ability in Flies upon Dietary Restriction Reveal decima as a Mediator of Insulin-like Peptide Production. *Curr Biol* (2020).
41. P. Sen, P. P. Shah, R. Nativio, S. L. Berger, Epigenetic Mechanisms of Longevity and Aging. *Cell* **166**, 822-839 (2016).
42. B. Villeponteau, The heterochromatin loss model of aging. *Exp Gerontol* **32**, 383-394 (1997).
43. C. G. Bell *et al.*, DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**, 249 (2019).
44. J. M. Harper, C. W. Leathers, S. N. Austad, Does caloric restriction extend life in wild mice? *Aging Cell* **5**, 441-449 (2006).
45. F. Knauf *et al.*, The life-extending gene Indy encodes an exchanger for Krebs-cycle intermediates. *Biochem J* **397**, 25-29 (2006).
46. J. M. Toivonen *et al.*, No influence of Indy on lifespan in Drosophila after correction for genetic and cytoplasmic background effects. *PLoS Genet* **3**, e95 (2007).
47. K. Christensen, T. E. Johnson, J. W. Vaupel, The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* **7**, 436-448 (2006).
48. L. Fontana, L. Partridge, Promoting health and longevity through diet: from model organisms to humans. *Cell* **161**, 106-118 (2015).
49. M. Beekman *et al.*, Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. *Aging Cell* **12**, 184-193 (2013).
50. T. Reed, D. M. Dick, S. K. Uniacke, T. Foroud, W. C. Nichols, Genome-wide scan for a healthy aging phenotype provides support for a locus near D4S1564 promoting healthy aging. *J Gerontol A Biol Sci Med Sci* **59**, 227-232 (2004).
51. J. Deelen *et al.*, Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686-698 (2011).
52. J. Deelen *et al.*, Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* **23**, 4420-4432 (2014).
53. L. Broer *et al.*, GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci* **70**, 110-118 (2015).
54. A. A. Puca *et al.*, A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc Natl Acad Sci U S A* **98**, 10505-10508 (2001).
55. B. J. Geesaman *et al.*, Haplotype-based identification of a microsomal transfer protein marker associated with the human lifespan. *Proc Natl Acad Sci U S A* **100**, 14115-14120 (2003).
56. A. R. Brooks-Wilson, Genetics of healthy aging and longevity. *Hum Genet* **132**, 1323-1338 (2013).
57. A. Nebel *et al.*, A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* **132**, 324-330 (2011).
58. Y. Zeng *et al.*, Novel loci and pathways significantly associated with longevity. *Sci Rep* **6**, 21243 (2016).
59. B. J. Willcox *et al.*, FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* **105**, 13987-13992 (2008).
60. F. Flachsbart *et al.*, Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* **106**, 2700-2705 (2009).
61. P. Eline Slagboom, N. van den Berg, J. Deelen, Phenome and genome based studies into human ageing and longevity: An overview. *Biochim Biophys Acta Mol Basis Dis* **1864**, 2742-2751 (2018).
62. E. K. Speliotes *et al.*, Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-948 (2010).
63. H. Lango Allen *et al.*, Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).
64. S. V. Nuzhdin, E. G. Pasyukova, C. L. Dilda, Z. B. Zeng, T. F. Mackay, Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **94**, 9734-9739 (1997).

65. M. De Luca *et al.*, Dopa decarboxylase (Ddc) affects variation in Drosophila longevity. *Nat Genet* **34**, 429-433 (2003).
66. M. Kaeberlein, M. McVey, L. Guarente, The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. *Genes Dev* **13**, 2570-2580 (1999).
67. T. F. Mackay *et al.*, The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173-178 (2012).
68. W. Huang *et al.*, Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res* **24**, 1193-1208 (2014).
69. G. Liti *et al.*, Population genomics of domestic and wild yeasts. *Nature* **458**, 337-341 (2009).
70. M. F. Durham, M. M. Magwire, E. A. Stone, J. Leips, Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nat Commun* **5**, 4338 (2014).
71. W. Huang *et al.*, Context-dependent genetic architecture of Drosophila life span. *PLoS Biol* **18**, e3000645 (2020).
72. C. Y. Liao *et al.*, Fat maintenance is a predictor of the murine lifespan response to dietary restriction. *Aging Cell* **10**, 629-639 (2011).
73. K. B. Dick, C. R. Ross, L. Y. Yampolsky, Genetic variation of dietary restriction and the effects of nutrient-free water and amino acid supplements on lifespan and fecundity of Drosophila. *Genet Res (Camb)* **93**, 265-273 (2011).
74. P. D. Stanley, E. Ng'oma, S. O'Day, E. G. King, Genetic Dissection of Nutrition-Induced Plasticity in Insulin/Insulin-Like Growth Factor Signaling and Median Life Span in a. *Genetics* **206**, 587-602 (2017).
75. V. Gorbunova, A. Seluanov, Z. Zhang, V. N. Gladyshev, J. Vijg, Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat Rev Genet* **15**, 531-540 (2014).
76. J. Campisi, Cellular senescence as a tumor-suppressor mechanism. *Trends Cell Biol* **11**, S27-31 (2001).
77. R. Buffenstein, The naked mole-rat: a new long-living model for human aging research. *J Gerontol A Biol Sci Med Sci* **60**, 1369-1377 (2005).
78. A. Sahm *et al.*, Long-lived rodents reveal signatures of positive selection in genes associated with lifespan. *PLoS Genet* **14**, e1007272 (2018).
79. S. Ma, V. N. Gladyshev, Molecular signatures of longevity: Insights from cross-species comparative studies. *Semin Cell Dev Biol* **70**, 190-203 (2017).
80. L. M. Abegglen *et al.*, Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA* **314**, 1850-1860 (2015).
81. T. A. Manolio *et al.*, Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).
82. C. A. Reynolds, D. Finkel, A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychol Rev* **25**, 97-112 (2015).
83. S. P. Mooijaart *et al.*, Polymorphisms associated with type 2 diabetes in familial longevity: The Leiden Longevity Study. *Aging (Albany NY)* **3**, 55-62 (2011).
84. M. Beekman *et al.*, Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* **107**, 18046-18049 (2010).
85. S. R. Urfer, M. Kaeberlein, D. E. L. Promislow, K. E. Creevy, Lifespan of companion dogs seen in three independent primary care veterinary clinics in the United States. **7** (2020).
86. J. M. Hoffman, K. E. Creevy, D. E. L. Promislow, Reproductive Capability Is Associated with Lifespan and Cause of Death in Companion Dogs. *Plos One* **8** (2013).
87. J. M. Hoffman, K. E. Creevy, A. Franks, D. G. O'Neill, D. E. L. Promislow, The companion dog as a model for human aging and mortality. *Aging Cell* **17**, e12737 (2018).
88. S. R. Urfer *et al.*, A randomized controlled trial to establish effects of short-term rapamycin treatment in 24 middle-aged companion dogs. *Geroscience* **39**, 117-127 (2017).
89. S. R. Urfer *et al.*, Asymptomatic heart valve dysfunction in healthy middle-aged companion dogs and its implications for cardiac aging. *Geroscience* **39**, 43-50 (2017).
90. G. Larson *et al.*, Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A* **109**, 8878-8883 (2012).
91. M. Germonpré *et al.*, Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *Journal of Archaeological Science* **36**, 473-490 (2009).
92. M. Kaeberlein, K. E. Creevy, D. E. Promislow, The dog aging project: translational geroscience in companion animals. *Mamm Genome* **27**, 279-288 (2016).
93. N. B. Sutter *et al.*, A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**, 112-115 (2007).

Chapter 2: Genetic and metabolomic variation of diet restriction and lifespan in flies

AUTHORS: Kelly Jin[1], Kenneth A. Wilson[2,3], Jennifer N. Beck[2], Christopher S. Nelson[2], George W. Brownridge III[2,4], Benjamin R. Harrison[1], Danijel Djukovic[5], Daniel Raftery[5], Rachel B. Brem[2,3,6], Shiqing Yu[7], Mathias Drton[8], Ali Shojaie[9], Pankaj Kapahi[2,3], Daniel Promislow[1,10]

1. Department of Pathology, University of Washington School of Medicine, Seattle, WA, USA
2. Buck Institute for Research on Aging, Novato, CA, USA
3. Davis School of Gerontology, University of Southern California, University Park, Los Angeles, CA, USA
4. Dominican University of California, San Rafael, CA, USA
5. Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA
6. Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA
7. Department of Statistics, University of Washington, Seattle, WA, USA
8. Department of Mathematics, Technical University of Munich, Munich, Germany
9. Department of Biostatistics, University of Washington, Seattle, WA, USA
10. Department of Biology, University of Washington, Seattle, WA, USA

Abstract

In most organisms, dietary restriction (DR) increases lifespan. However, several studies have found that genotypes within the same species vary widely in how they respond to DR. To explore the mechanisms underlying this variation, we exposed 178 inbred *Drosophila melanogaster* lines to a DR or *ad libitum* (AL) diet, and measured a panel of 105 metabolites under both diets. Twenty four out of 105 metabolites were associated with the magnitude of the lifespan response. These included proteinogenic amino acids and metabolites involved in α-ketoglutarate (α-KG)/glutamine metabolism. We confirm the role of α-KG/glutamine synthesis pathways in the DR response through genetic manipulations. We used covariance network analysis to investigate diet-dependent interactions between metabolites, identifying the essential amino acids threonine and arginine as "hub" metabolites in the DR response. Finally, we employ a novel metabolic and genetic bipartite network analysis to reveal multiple genes that influence DR lifespan response, some of which have not previously been implicated in DR regulation. One of these is *CCHa2R*, a gene that encodes a neuropeptide receptor that influences satiety response and insulin signaling. Across the lines, variation in an intronic single nucleotide variant of *CCHa2R* correlated with variation in levels of five metabolites, all of which in turn were correlated with DR lifespan response. Inhibition of adult *CCHa2R* expression extended DR lifespan of flies, confirming the role of *CCHa2R* in lifespan response. These results provide support for the power of combined genomic and metabolomic analysis to identify key pathways underlying variation in this complex quantitative trait.

Introduction

No two individuals age in exactly the same way (1, 2). Variation in aging, as with all complex traits, is determined by contributions from variation in genes, the environment, and the interaction between the two. Similarly, an individual's response to interventions that slow aging is likely to be equally as complex.

Among these interventions, dietary restriction (DR) has been shown to extend lifespan in almost all model organisms in which it has been tested (3). However, despite this high level of conservation across species (4), several studies suggest that the DR response might be far from universal. For example, Harper *et al*. (5) found no effect of DR on mean longevity in grand-offspring of wild-caught mice. Additionally, results from studies in three different species reveal extensive within-species variation in the DR response (6-10). Liao *et al*. (6) showed that fewer than half of 41 recombinant inbred mouse strains on DR showed a lifespan increase. Several years later, a study of 166 single-gene deletion yeast strains found variation in the DR

response, ranging from a 79% reduction to a 103% increase in lifespan (7). In 2017, Stanley *et al.* (9) measured lifespan for 80 recombinant inbred *Drosophila melanogaster* lines on control and DR diets and found highly significant diet-by-genotype interaction for median life span. Most recently, Wilson *et al.* (10) showed that of 161 naturally-derived inbred lines of *D. melanogaster*, 71% showed a DR-mediated lifespan extension, while the rest showed either a non-significant lifespan response or reduced lifespan under DR ((10), Fig 1A and S1 Fig A). In light of this considerable variation, to fully understand the factors that determine the response to DR, we need to understand not only the mechanisms underlying DR itself, but also the mechanisms that influence *variation* in the DR response within a population.

To better understand why some individuals respond to DR while others do not, we employ a systems biology approach, focusing on the influence of two distinct biological domains—the genome and the metabolome. Given that DR dramatically changes the metabolic state of an organism, metabolomic profiling has become an invaluable tool for investigating the mechanisms underlying DR and aging (11-13). The metabolome, which includes all low-molecular weight molecules within a biological system, plays a crucial role in the biology of aging (14). Examples include the role of specific amino acids like tryptophan or methionine in DR (15, 16), of polyamines in age-related decline in circadian periodicity (17), and of α-ketoglutarate (α-KG) and its derivatives in TOR signaling and epigenetic regulation (18, 19). High-throughput, high-resolution metabolomic profiling methods allow us to capture a snapshot of the circulating products and intermediates of cellular metabolism within a tissue or organism. This ability has proven invaluable for understanding the mechanisms underlying complex traits for many reasons. First, although feedback mechanisms likely exist between all domains within a biological system, the metabolome is generally thought to be downstream of transcriptional and translational regulation. Second, the metabolome is highly influenced by, and therefore integrates information from, both the genome and environment (11, 20). As a result, the metabolome may explain a greater proportion of phenotypic variation within a population than genomic, transcriptomic, or proteomic profiling alone. This is critical, as many genome-wide studies of complex traits explain only a fraction of the phenotypic variation (21). Third, metabolites are involved in all biological processes within the cell, and as such, metabolomic studies can help bridge the gap between genotype and phenotype (20, 22) (23-26). As researchers learn more about the intricacies of the various cellular mechanisms that govern aging, systems biology approaches such as these are essential for revealing mechanisms that underlie these complex phenotypes.

Our group previously showed that DR slows age-related changes in the metabolome and dramatically alters metabolic network structure in a single, wildtype strain of *Drosophila* (11). However, this work did not consider the extent to which the response to DR varies among genotypes. Here, we explore the metabolic signatures from different fly strains that vary widely in their response to DR by profiling the metabolome in 178 inbred fly lines from the *Drosophila* Genetic Reference Panel (DGRP), a set of inbred, fully sequenced strains of *Drosophila* derived from a wild population (27, 28). The DGRP has been successfully used to profile and genetically map complex quantitative traits including longevity (29), oxidative stress resistance (30), microbiota composition (31), and many others (32).

Here, building on a prior analysis of lifespan measurements collected by Wilson *et al.* (10), we extend this with metabolomic profiling of flies from the same experiment. Taken together, this is the largest *Drosophila* study we know of to leverage both the metabolome and genome of a large number of genotypes of the same species to study lifespan response to DR. Specifically, we show that i) DR dramatically remodels the metabolome in consistent ways across 178 different genotypes; ii) several individual metabolites correlate with lifespan response, including amino acids and metabolites involved in α-KG/glutamine metabolism; iii) differential network analysis reveals that metabolite network structure varies with diet and reveals 'hub' metabolites that gain or lose multiple network connections in a diet-specific manner; iv) network analysis of genome-wide association tests reveals candidate gene-to-metabolite-to-phenotype pathways that might underlie the DR response; and v) down-regulation of three of the genes identified from this analysis alters lifespan response, including a neuronal effect of *CCHa2R,* a gene that encodes a neuropeptide receptor thought to be involved in nutrient sensing and satiety response.

Results

Diet specific changes in the metabolome

Targeted metabolomic profiling was performed on 178 DGRP lines for each diet condition. The flies used for metabolomics were collected from the same cohort of flies used for lifespan measurements. The lifespan measurements across the DGRP on AL and DR were first presented and discussed in Wilson *et al*. (10), while the metabolomic data are novel to this study.

After quality filtering (see Methods), 105 metabolites were included in downstream statistical analysis. With this dataset, our goal was to identify biologically meaningful relationships that might exist between diet, lifespan response, the genome, and the metabolome.

To begin, we summarize the effect of DR on the metabolome across the DGRP lines using principal component analysis (PCA). The first principle component (PC) explained 69% of the variation across the entire metabolome, and cleanly separated samples by diet, revealing that DR has an extremely strong effect on the fly metabolome (S2 Fig A).

To determine which metabolites were affected by diet, difference in mean abundance (DR – AL) was calculated for each metabolite across all the lines. Since this calculation was performed on log-normalized and scaled data (see Methods), we interpret the resulting difference as a relative change in abundance (i.e., log(DR) – log(AL) = log(DR/AL)). As determined from paired t-tests, almost all measured metabolites showed a highly significant response to diet (S2 Fig B, Table S1). These results are consistent with previous studies demonstrating a substantial remodeling of the fly metabolome under DR (11).

The effect of DR on lifespan in the DGRP

There is considerable variation in lifespan response as a result of DR across the DGRP (S1 Fig A, (10)). As expected, we also observed a wide range of variation across the DGRP in the difference in lifespan between diets (DR – AL) (Fig 1A). Change in mean lifespan varied among lines from 18 days shorter to 27 days longer in response to DR. Of the 161 DGRP lines that have lifespan measurements, 114 lines (70.8%) lived longer, 30 lines (18.6%) did not show any significant change in lifespan, and 17 (10.5% lines) lived shorter as a result of DR (Fig 1A, (10)).

The primary goal of this study is to investigate the lifespan response to DR across the DGRP. The difference between DR and AL lifespan is one measure of this response. However, this difference value is correlated with AL lifespan (S1 Fig C), which means that any significant associations that we find with the difference in lifespans might actually be due to a relationship with AL lifespan. To remove potentially confounding effects of AL lifespan on the DR response, we derive another lifespan response trait that we call the relative change in lifespan (rLS; see Methods). rLS is highly correlated with the difference in lifespans (Fig 1B), but not with AL lifespan (S1 Fig D). For the remainder of this study, we will use rLS as our primary lifespan response phenotype.

Metabolites are significantly correlated with lifespan response

To determine which metabolites are correlated with lifespan traits (mean lifespan and rLS), we used a simple linear regression to model lifespan phenotypes as a function of each metabolite measured under either diet:

$$\text{Lifespan trait} = \alpha + \beta \cdot \text{metabolite}_{AL} + \varepsilon \quad (1)$$

or

$$\text{Lifespan trait} = \alpha + \beta \cdot \text{metabolite}_{DR} + \varepsilon \quad (2)$$

Looking first for metabolites associated with mean lifespan, we found that 10 metabolites measured under DR were correlated with mean DR lifespan (with a false discovery rate (FDR) = 0.01) (Fig 2A, Table S2). Notably, no metabolites measured under AL were found to be correlated with mean AL lifespan (Fig 2A; Table S3), though 8 metabolites measured under AL were correlated with DR lifespan (Table S3).

We next identified metabolites that were correlated with rLS. Out of a total of 105 measured metabolites, 24 were associated with rLS after FDR correction (Fig 2B, Table S2 and S3). Thirteen metabolites are associated with rLS under both conditions (top right quadrant in Fig 2B), while 10 were only significant when measured under AL (bottom right quadrant in Fig 2B), and only threonine was significant under DR (top left quadrant in Fig 2B).

Next, we asked if the *change* in metabolite abundance was predictive of lifespan response. To do this, we modeled rLS as a function of change in metabolite level ($\Delta mz_i = mz_{DR,i} - mz_{AL,i}$ for metabolite *i*). Of 105 metabolites, the change in abundance of 4 metabolites—2-phosphoenolpyrivic acid (PEP), 2-phosphoglyceric acid, threonine, and arginine—was associated with rLS (Fig 2C, Table S4).

Diet-dependent changes in metabolite networks

Having identified metabolites associated with lifespan and its response to DR across genotypes, we wanted to investigate the relationship between these metabolites and the broader metabolome measured here, and more specifically, to determine how those relationships change between diets. To capture this, we measured covariance among all metabolites and performed differential network analysis to investigate interactions between metabolites across diets (Fig 3A). We grouped interactions into three categories: 1) metabolite-metabolite interactions that are significantly greater (as measured by correlation coefficient ρ;

see Methods) under AL than DR (red edges in Fig 3A; Fig 3B); 2) interactions that are significantly greater under DR than AL (blue edges in Fig 3A; Fig 3C); and 3) interactions that are strongly significant under both conditions (yellow edges in Fig 3A; Fig 3D). We focus solely on positive correlations because there were no significant negative correlations between metabolites in the network.

Some metabolites gain or lose many interactions under one diet in comparison with the other. Of these, the most striking example is arginine, which gains 14 edges under AL (Fig 3E and right asterisk in Fig 3A). Such metabolites may represent 'hubs' in the metabolome network. Hubs are also found under DR conditions, with threonine gaining seven edges (Fig 3F and left asterisk in Fig 3A). Three features of this metabolome network stand out to us. First, the two hub metabolites threonine and arginine are among the top hits for Δmzs associated with rLS (Fig 2C). Second, many of the edges that these hub metabolites gain or lose across diets are shared with other amino acids (Fig 3A). Last, this network includes many of the essential amino acids, most of which are all correlated to one another under both AL and DR (boxed metabolites in Fig 3A). The essential amino acids in *Drosophila* include arginine and the nine essential amino acids in humans (33).

Manipulation of α-KG pathways alters lifespan response to DR

Our analysis of metabolites associated with the lifespan response to DR identified both α-KG and glutamine (Fig 2), which form a sub-network within the overall metabolome network (Fig 3A and D). Given our finding, and previous studies implicating α-KG in DR and TOR signaling (18), we used RNA interference (RNAi) to knock down genes in the α-KG/glutamine pathway to explore its possible role in the effect of DR on lifespan. Specifically, we manipulated the expression of glutamate dehydrogenase (GDH), which catalyzes the reversible conversion between αKG and glutamate, and glutamine synthetase 1/2 (GS1, GS2), which catalyze the conversion of glutamate to glutamine (S3 Fig A). For these experiments we used the inducible GAL4-GeneSwitch (GAL4-gs) drivers to knock down each gene in different parts of the fly upon induction with the compound RU486 (34).

We exposed inducible RNAi fly strains to AL and DR diets and recorded lifespan. To test for a change in lifespan response as a result of RNAi, we tested for a significant RNAi-x-diet interaction term in a Cox Proportional Hazards model (see Methods). Knockdown of *gs2* in the whole fly most strongly affected the lifespan response as compared to knockdown of *gs1* and *gdh*, with knockdown flies living longer on the AL diet (S3 Fig B-D). A greater extension of AL

lifespan was seen with neuron-specific knockdown of GS2, but less so in fat body or gut-specific RNAi (S3 Fig E-G). This suggests that GS2 is part of a diet-specific lifespan pathway active in neurons.

Gene-metabolite-phenotype network analysis for lifespan response

While we found that fully 23% of measured metabolites were associated with rLS, we performed a gene-level genome-wide association study (GWAS) on the rLS trait and only one gene, an uncharacterized coding gene CG6231, was associated with rLS. In light of this, we decided to leverage the metabolite-lifespan associations identified here, creating a bipartite network of genes and metabolites that would allow us to look for genes that might impact lifespan indirectly through their impact on the metabolome. To do this, we first performed multiple individual gene-level GWAS using metabolite levels as quantitative traits. Specifically, we selected metabolites that were associated with rLS. This included 23 AL metabolites and 14 DR metabolites (Fig 2B). The gene-level GWAS resulted in significance scores for each gene-metabolite pair. Briefly, gene scores were assigned by taking the minimum *P* value from all variants associated with a gene after adjustment via at least 10,000 rounds of permutation testing. A detailed explanation of this calculation is included in the Methods. Significant gene-metabolite relationships were used to build a network connecting genes, metabolites, and the lifespan response phenotype. The resulting multi-omic network for AL metabolites is visualized in Fig 4 and the network for DR metabolites in S4 Fig. All genes that appear in the DR network also appear in the AL network, so we display the AL network in the main figures. The purpose of this network is to diagram molecular paths that might regulate lifespan extension, starting from the gene level (teal nodes) to the metabolite level (yellow nodes), and that ultimately influence lifespan response to diet as quantified by rLS (center grey node). The degree, or number of edges, of each gene node is represented by its size in the figure. A complete list of significant AL and DR network edges and scores is provided in Table S5.

Across both the AL and DR network, the gene with the most significant gene-level score was *Vps15* which is associated with AL levels of 2-phosphoglyceric acid and PEP (Table S5). *Vps15* encodes a serine/threonine protein kinase that is part of the PI3-Kinase (PI3K) complex (35). We can also examine gene node degree as an indicator of the gene's role in the signaling network. It is important to note that the degree of a gene node can be interpreted to mean that a gene affects all metabolites its connected to independently, or that a gene more directly affects the level of one metabolite whose level in turn is correlated with other metabolites. In the latter case, some edges may represent indirect gene-metabolite relationships. For example,

CCHa2R, which has a degree of 5 in the AL network (Fig 4), could independently associate with five metabolites. However, it may instead affect the abundance of one metabolite more directly than the others. In both the AL and DR networks, the gene with the largest degree value is *CCHa2R* (Fig 4, S4 Fig). *CCHa2R* encodes the protein CCHamide-2 receptor (CCHa2-R), a neuropeptide receptor thought to be expressed in the brain and gut of fruit flies, and is proposed to be involved in appetite regulation and insulin signaling (36, 37). Thus, our network analysis identified genes that might be involved in regulating lifespan response to DR, including *Ddc*, *SelT*, *jeb*, *Nuak1*, *z*, and several uncharacterized coding genes (Fig 4, S4 Fig, Table S5).

CCHa2R and *SelT* are involved in the lifespan response to DR

To test the ability of our multi-omic network analysis to identify genes involved in the DR response, we used inducible RNAi to measure the effects of gene knockdown on AL and DR lifespans. We focused on the genes *CCHa2R* and *selT*. The single nucleotide polymorphism (SNP) 2R_1939249_SNP identified from metabolite GWAS is a C/T variant in an intron of *CCHa2R*. This SNP was associated with differential abundance of five rLS-associated metabolites under AL (Fig 5A). As an example, the relationship between iso-leucine, rLS, and *CCHa2R* SNP are shown in Figure 5B-D. Knocking down *CCHa2R* using a whole-body driver resulted in an increased mean lifespan under DR, but a slightly decreased mean lifespan under AL, although this effect was not statistically significant under a Cox Proportional Hazards model framework (Fig 5E). Given this trend, we then used a neuronal specific driver to knock down *CCHa2R*, which showed a much greater effect of increasing lifespan under DR, but not AL (Fig 5F).

Manipulation of *SelT* also resulted in changes to diet-specific lifespans (S5 Fig). *SelT* encodes selenoprotein T. Selenoproteins are a family of thioredoxin-disulfide reductases that plays a role in defending against cellular oxidative damage by controlling the redox balance of the cell (38). We found that whole-body *SelT* knockdown increased lifespan on AL while having no effect on lifespan under DR (S5 Fig). Taken together, these results support our network analysis as a map of gene-to-metabolite-to-phenotype pathways that underlie natural variation in the DR response and also highlight the fact that DR response can be influenced through a change in AL lifespan, DR lifespan, or both.

Discussion

Diet restriction is regarded as the most robust form of lifespan extension known, and has been consistently demonstrated to increase longevity in almost all model organisms in which it has been studied (3). However, in almost all of these studies, researchers have used a single strain chosen from a handful of common lab-adapted strains, limiting our ability to gain insight into variation in response to DR in a diverse population. Here, we have leveraged the power of the genetic variation found within the DGRP, together with systems biology methods, to measure and explain genetic variation in the DR response. This work brings three key approaches to bear on the study of DR response, including natural genetic variation, metabolome profiling, and network modeling. By combining all three of these, an approach never used before in the context of DR, we are able to generate novel genetic and biochemical hypotheses about DR. We then validate these findings experimentally using inducible RNAi.

This work adds to a growing body of evidence for genetic variation in the DR response, including similar studies conducted in mice (6), yeast (7), and a different fly genetic reference panel derived from recombinant inbred strains (9). While these previous studies underscore the importance of genetic variation in shaping the DR response, our work focuses in particular not on single-gene knock-out strains or genotypes derived from recombinant inbred strains, but rather on the natural genetic variation derived from a single wild population (28). The translational path from *Drosophila* to humans is a long one, but these results do underscore the possibility that the effect of interventions designed to decrease or delay the onset of age-related decline in human populations could have diverse outcomes from one person to the next, depending on individual genetic makeup and environment history (6-9). The lifespan results we present in this study were also analyzed by Wilson *et al.* (10). We note that Wilson *et al.* found multiple candidate variants that affect diet-dependent longevity, while we identify one that met our significance cutoff (Fig 4 and S4 Fig). This is not surprising, given that these two studies fit two different response variables to different covariates. Wilson *et al.* used a linear model predicting mean lifespan with an interaction term between genotype and diet for their GWAS, while we use relative lifespan as the outcome for our GWAS.

While it is relatively straightforward to measure genetic variation for complex traits, identifying the individual genes that contribute to this variation has proven to be far more challenging (21). To fill this gap, researchers have turned to the metabolome. This has helped not only to define genetic variation, but also to suggest mechanisms that underlie this variation (23-26). Previous studies have found that DR leads to dramatic shifts in metabolism in diverse organisms (39) and that it attenuates metabolic signatures of aging (11, 40, 41). Metabolomic

profiling has become a popular tool for investigating mechanisms underlying DR within single genotypes across a wide array of species, including worms (41, 42), flies (11, 43), mice (40, 44), non-human primates (45), and humans (46, 47). Many of these studies pointed to changes in fatty acid metabolism as a result of DR (40, 44, 47). This is consistent with the finding from Liao *et al.* that across 41 mouse strains, the ones with the least reduction in fat under DR were more likely to show lifespan extension on DR (48). Our study, which is the first that we know of to use metabolome profiling to investigate lifespan response to DR across genotypes, identified 25 metabolites that were correlated with lifespan response either through their baseline abundance or change in abundance across the two diets (Fig 2B and C). We did not profile enough lipid metabolism-related metabolites to be able to test whether or not fat maintenance might be an influential component underlying the DR response. However, one of the genes our bipartite network analysis identified was *jeb* (jelly belly), which encodes a low-density lipoprotein proposed to be involved in neuronal PI3K signaling during nutrient restriction in flies (49). Jelly belly may regulate lifespan response through a mechanism involving the metabolite hypoxanthine (Fig 4). In addition, many of the 25 metabolites associated with lifespan response in our study were related to amino acid metabolism, which we have previously shown to be a significant pathway modified under DR in flies (11).

As we note above, the metabolome can act as a statistical link between genotype and phenotype (25, 26). Here we advance this approach in two important respects. First, we use a novel differential network analysis to contrast AL and DR metabolite networks in a manner that allows straightforward comparison between the two diet conditions (Fig 3), providing the critical and necessary context needed to interpret the interactions that we see under DR. This network captures a snapshot of metabolic relationships that might explain how flies translate nutritional environment to lifespan. For example, others have shown that restricting essential amino acids, methionine in particular, is necessary and sufficient to extend lifespan in flies (50). Our differential network links methionine and other essential amino acids to metabolites that correlate with lifespan response, diagramming metabolic paths that might explain the relationship between methionine, other amino acids, and the DR response (Figs 3, 4, and S4 Fig). Second, while previous studies have typically focused on single molecular domains such as genetics or metabolomics (51-54), here we create a *bipartite network* that links specific genes to metabolites, and specific metabolites to the DR response, allowing us to identify new candidate genes and pathways that we begin to explore here. We use molecular genetic methods to confirm that indeed, many of the genes we identified influence lifespan response under dietary modification, including *CCHa2R* and *SelT*. We also identify other genes that have

been identified by previous independent studies of *Drosophila* longevity. One of these genes, *Ddc,* encodes dopa decarboxylase, which has been suggested to influence natural variation in *Drosophila* longevity via a neuronal mechanism (55), adding further credence to the validity of our network approach. Other genes that will be of particular interest to pursue in future studies include *jeb*, as previously discussed, and *Vps15*, a PI3-Kinase regulator involved in autophagy.

In addition, our RNAi results suggest a role for neuronal perception of nutrient availability in lifespan response to DR. In *Drosophila*, neuronal signaling regulates physiological response to environmental nutrients, including gustatory and metabolic perception of amino acids and sugars (56-58), receipt of signals about nutrient availability from peripheral organs such as the gut (59), and feeding restraint (60). We present evidence for a neuron-specific mechanism involving glutamine/α-KG signaling that influences lifespan under AL. GS2 RNAi, but not GS1, extended the lifespan of flies on AL (S3 Fig). In *Drosophila*, GS1 localizes primarily to the mitochondria while GS2 localizes to the cytosol (61). In light of this, we postulate that cytosolic glutamine and perhaps glutamatergic signaling in neurons play a role in the DR response. This idea is supported by recent findings identifying a set of glutamatergic interneurons that signal *Drosophila* larvae to overcome amino acid limitation and pupate (62). Furthermore, support for the suggestion that regulation of α-KG levels partially explains the lifespan response we observed is provided by work in *Caenorhabditis elegans* showing diet-dependent regulation of mTOR signaling by α-KG (18). This evidence suggests the existence of interactions between glutamine/α-KG signaling and mTOR that ultimately influence lifespan response to DR.

We also found that knocking down expression of neuronal *CCHa2R* resulted in longer lifespan under DR (Fig 5). To our knowledge, this is the first study to implicate *CCHa2R* in an aging/DR-related signaling mechanism. We hypothesize that *CCHa2R* is associated with lifespan due to its influence on metabolism, an idea supported by previous studies in *Drosophila*. CCHa2-R is a G-protein coupled receptor that exclusively binds the neuropeptide CCHa2 and is thought to be involved in nutrient sensing and satiety response (36). CCHa2 is mainly expressed in the fat body and at low levels in the gut and central nervous system (CNS), while its receptor CCHa2-R is highly enriched in the CNS, particularly in *Drosophila* insulin-like peptide (Dilp) producing cells that control the secretion of neuropeptide F and SIFamide in the brain (36). In fly larvae, CCHa2 signaling appears to mediate the secretion of Dilp2 and Dilp5 (36) and signaling between peripheral organs and the brain (37). Given the important role of Dilps both in nutrient signaling and aging (63), this novel connection between CCHa2-R and the effect of DR on aging is not surprising. In addition, the closest human homolog of CCHa2-R is bombesin receptor subtype 3 (64). Mammalian bombesin-like peptides are widely distributed in

the central nervous system and gastrointestinal tract, mirroring the distribution of CCHa2 peptide in *Drosophila* (36), and are thought to mediate signaling between the gut and brain, regulating processes such as smooth-muscle contraction, metabolism, and behavior (65). The striking similarities between *Drosophila* CCHa2 signaling and mammalian bombesin signaling suggest that the processes are conserved to some extent.

Studying the genetic basis of natural trait variation complements existing mutational lab studies, offering critical additional insight into the biology of DR. Like other GWAS of lifespan in the DGRP (29, 66), we sought out to identify natural genetic variation affecting phenotypic variation. Interestingly, the genes and pathways identified by our work along with Durham *et al.* (66) and Ivanov *et al.* (29) largely differ in comparison to the canonical lifespan-associated genes and pathways identified by mutational studies. Although the reasons for this are unclear, these differences further emphasize the importance of systems approaches such as ours for developing a more complete picture of how genetics and environment impact healthy aging.

Limitations of study

Readers should keep two caveats in mind in evaluating the results presented here. First, DGRP metabolite profiles were measured from whole fly bodies. However, different tissues of the fly, including the head, thorax, and abdomen, show different metabolome profiles under AL and DR (11). Given the large number of lines measured here, tissue-specific profiling across the DGRP was outside the scope of the present study. In light of our understanding that neuronal perception of nutrients plays an important role in aging, future studies focused on head or brain specific metabolomics would be of great interest (e.g.(67)).

Second, flies were diet restricted by decreasing the percentage of yeast extract in their food. This diet has been used previously to demonstrate the effects of DR in flies (68-70). There are many different methods of implementing DR that have been published in the *Drosophila* community, (50, 71, 72), and no single diet is considered standard, raising challenges in comparing results across studies. Furthermore, some have argued that different fly strains might have different optimal DR food levels, and that a decrease in fecundity must coincide with increased lifespan in order for the response to truly be considered DR (73). Future work in this area should include genetic variation for survival *and* fecundity, should explore different types of DR, and finally should consider whether genotypes differ in how they respond to DR because they vary in their optimal diet concentration.

Concluding remarks

Our work represents the first effort to understand the role of *naturally occurring variation* in the DR response and the mechanisms that underlie this variation. Studies such as this are critical to the future of health and aging research, as the most common age-related diseases are genetically and phenotypically heterogeneous, and, as a result, likely require treatments tailored to environment and genotype. As such, it is imperative to explore mechanisms of aging and disease not only in single, traditional laboratory strains, but also in the context of genetically heterogeneous animal models.

Methods

Drosophila DR and lifespan

Methods for lifespan experiments were originally described in Wilson *et al.* (10). Briefly, DGRP lines were obtained from the Bloomington *Drosophila* Stock Center. All flies were maintained on standard stock food (1.55% live yeast, 7.5% sugar, 8.5% cornmeal, 0.46% agar, 85% water). Flies were kept on a 12-hour light/dark cycle at 25°C and approximately 65% humidity throughout the experiment. In preparation for lifespan measurements, approximately 15 female flies and 3 male flies were put in each bottle of stock food. Approximately 6 bottles were set up per strain. Five days after setting up stocks, adults were discarded, leaving behind larvae that would be used for lifespan studies. Fourteen days after set-up when the experimental flies were estimated to be 2-3 days old, mated female flies were sorted into approximately 8 vials of either high yeast extract food (AL; 5% yeast extract, 5% sugar, 8.5% cornmeal, 0.46% agar, 85% water) or low yeast extract food (DR; same as previous, but with 0.5% yeast extract), with 25 flies per vial, targeting a total of 200 flies per strain*diet combination. Flies were transferred to fresh vials every other day. Recording of fly deaths commenced 8 days after flies started the new diet. All fly lifespan values presented here represent mean age-at-death measured in days from beginning the experimental diet.

Three flies per line and diet were frozen 5 days after beginning the experimental diet and shipped to the University of Washington in Seattle, WA for metabolomic profiling. We chose to sample at 5 days to try and capture early metabolomic indicators of DR response. Published studies from our group and others have shown metabolomic and other phenotypic differences in response to DR as early as 2-10 days after beginning the experimental diet (11, 43, 74).

RNAi experiments

RNAi experiments were performed utilizing the Gal-UAS system (34). Briefly, temporal knockdown of target genes was carried out using the ubiquitous drug-inducible GeneSwitch driver *Act5C*-GS-Gal4 and tissue specific knockdown was performed using the pan-neuronal driver *elav*-GS-Gal4, the fat body driver *S106*-GS-Gal4, or the gut-specific driver *5966*-GS-Gal4. Following development, flies were sorted onto AL or DR foods that contained 200uM RU486 to allow activation of the GeneSwitch system. Flies were maintained under these conditions throughout life.

Metabolomics sample preparation

Fly samples were prepared following previously described procedure (75-77). Briefly, samples were thawed at room temperature, homogenized in 10:1 PBS:Water, methanol containing known concentrations of 6C13-glucose and 2C13-glutamate was added, samples were vortexed and stored at -20 °C for 20 min. Afterwards, samples were sonicated in an ice bath, centrifuged at 20,600 g and supernatant was recovered and dried. At the end, dried supernatant was reconstituted in a buffer containing known concentrations of 2C13-Tyrosine and 1C13-Lactate. A sample quality control (QC-S) was made by pooling small volumes of randomly chosen 30 prepared and reconstituted fly samples and this QC was used to monitor the data reproducibility.

Liquid Chromatography – Mass Spectrometry (LC-MS)

Each sample was injected twice, 15 µL and 5 µL for analysis in negative and positive ionization modes, respectively. Both chromatographic separations were performed in HILIC mode on two parallel identical amide-based analytical columns. While one column was performing the separation, the other column was getting reconditioned and ready for the next injection. After the chromatographic separation, MS ionization and data acquisition were performed using an AB Sciex QTrap 5500 mass spectrometer (AB Sciex, Toronto, ON, Canada) equipped with an electrospray ionization (ESI) source and operating in multiple-reaction-monitoring (MRM) mode. We monitored 122 and 84 MRM transitions in negative and positive mode, respectively (206 MRM transitions total corresponding to 202 metabolites and 4 stable isotope-labeled internal standards).

Statistical analysis

All analyses described below were carried out using the open source software package R (78). A false discovery rate of $\alpha = 0.01$ using the Benjamini-Hochberg-Yekutieli procedure (79) was used for all multiple comparisons unless otherwise stated.

Calculating relative change in lifespan. To monitor the effect of DR on lifespan, we use the relative change in lifespan (rLS) which was calculated by taking the residuals of a simple linear least squares regression of DR lifespan against AL lifespan (S1 Fig B). The purpose of calculating rLS was to create a phenotype similar to the absolute different between DR and AL lifespan, but removes the potentially confounding effects of AL lifespans driving the lifespan response (Fig 1B and S1 Fig C-D).

Data normalization. All metabolites were log-transformed to approximate a Gaussian distribution. Metabolites with >5% missing measurements were excluded from the analysis. Remaining missing data were imputed using the “impute” R package (80). The final metabolomics dataset was comprised of 356 samples (178 fly lines x 2 diet conditions) with 105 individual metabolite features for each sample. Metabolite data were then mean-centered and unit-scaled within samples to normalize for variation in analyte amount (due to variation in fly weight/size) loaded onto the mass spectrometer. Flies were handled in different experimental batches and these experimental batches were found to have a statistically significant effect on many metabolite levels, but not on lifespan. To correct for the batch effect, we regressed out the effect of batch using a linear model. Depending on the type of analysis, we implemented 2 different types of batch correction: 1) correcting for batch within samples of the same diet, or 2) across all samples regardless of diet. The type of batch correction is indicated in the methods sections below.

Multivariate and univariate analysis. Principle component analysis (PCA) is a form of unsupervised multivariate analysis that partitions out independent variance components across a dataset of multiple, potentially correlated, variables. PCA was performed on all samples using the “prcomp” function in base R to observe how well the metabolome can separate samples by diet.

Differential analysis was performed on individual metabolites to determine which metabolites increase or decrease in response to diet. Batch correction was performed across all samples together regardless of diet for both differential analysis and PCA.

We used a multiple regression model to test for the effects of metabolite abundance on lifespan phenotypes. Genotype-wide mean body mass was used as a covariate in all models. Batch correction was performed on samples within diet for this analysis.

Analysis of RNAi lifespan experiments. We used a multivariate Cox Proportional Hazards regression to model survival as a function of diet (AL or DR), treatment (control or RNAi), and the interaction between diet and treatment. A significant interaction term suggested that the gene targeted by that RNAi treatment influences lifespan response to DR.

One of the assumptions of a Cox Proportional Hazards model is that the ratio of the hazards for individuals in different conditions remains constant over time, which can be evaluated by examining whether or not curves from different treatments cross over one another in a survival plot. While this assumption holds for most of our survival experiments, we acknowledge that for some of our survival experiments, that assumption does not hold (for example, Fig 5E), making some of the test results harder to interpret. However, we decided to apply the Cox model to all of our survival experiments regardless for consistency and interpretability.

Network analysis and genome-wide association. For differential network analysis, we calculated Spearman's ρ correlation matrices for all metabolites within each diet separately and used these to build unweighted differential metabolite networks. Batch correction was performed on samples within diet for this analysis.

To test the difference between the correlation structures under the two diets, we first calculated the difference in the two Spearman's ρ correlation matrices (differential correlation matrix). We then tested whether the differential correlation of each pair of metabolites is significant by using parametric tests for Spearman's ρ correlations, wherein we used the limiting normal distribution of the differential correlation for large sample sizes. More specifically, to improve normal approximations, we considered the scaled Spearman's ρ for a single diet with n samples, i.e., ρ*sqrt((n-2)/(1-ρ^2)). With independent samples, the difference in the scaled correlations between the two diets is approximately normal with mean = 0 and variance = 2. This asymptotic distribution was used to obtain P values for each pair of differential correlations. The resulting P values were adjusted for multiple testing.

Genotype variant calls for the DGRP are publicly available online (http://dgrp2.gnets.ncsu.edu/). Genome-wide association tests were performed in PLINK v.1.07 (81) using an additive linear model fitting phenotype (metabolite levels or lifespan response) as

a function of variant, including *Wolbachia pipientis* infection status and major inversions as covariates. Variants with a minor allele frequency > 0.05 were included in the analysis. To correct for the observation that genes with a greater number of variants are expected by chance to have lower minimum *P* values, we calculated a gene-specific “gene score”. Gene scores for each gene were calculated by initially taking the minimum *P* value from all the variants associated with a particular gene, including variants located within or 1,000 bp upstream or downstream of that gene. Following this, 1,000,000 permutations of this analysis for the top genes were implemented taking the minimum *P* value per gene from of those permutations. Top genes were first identified if they had a preliminary minimum $P \leq 10^{-3}$ from an initial 10,000 permutations of all genes. Permutations were conducted by randomizing the genotype designation among all samples. The final gene score was then calculated by dividing the number of permutations where the observed minimum *P* value was equal to or lower than the empirical minimum *P* value by the total number of permutations. In effect, we are asking if the minimum *P* value within a gene is even smaller than one would expect, given the number of variants found within that gene. A gene was then included in the network if it had a final gene score $\leq 10^{-4.5}$. All networks were visualized using Cytoscape (82).

Acknowledgements

This work was in part facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

Figures

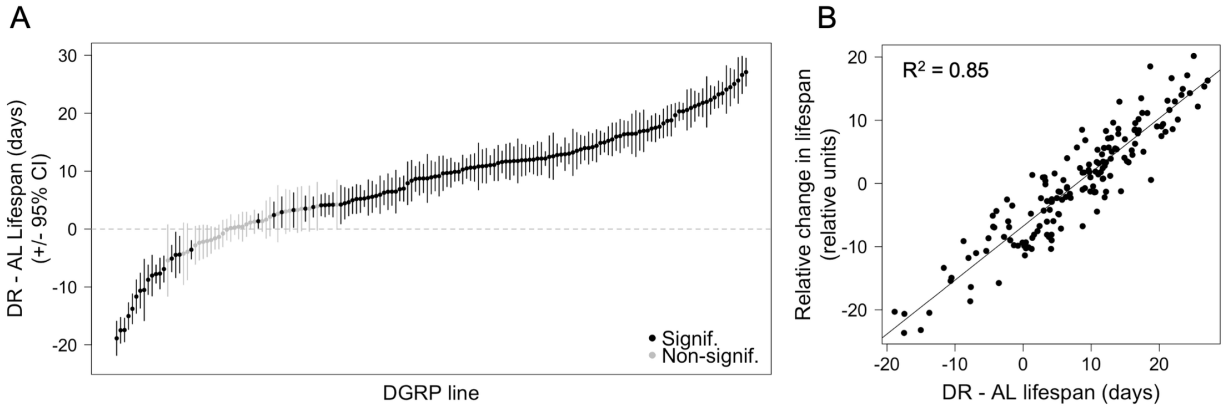


Figure 1. Variation in DR-mediated lifespan extension across the DGRP.

(A) Variation in DR – AL lifespan measured across 161 DGRP lines plotted in ascending order. Each point represents a fly line. Statistical significance was determined using 5% FDR adjusted *P* value from Student's t-tests. Error bars represent 95% confidence interval from t-test. (B) Relationship between change in lifespan and relative change in lifespan (rLS). The two lifespan traits are significantly correlated.

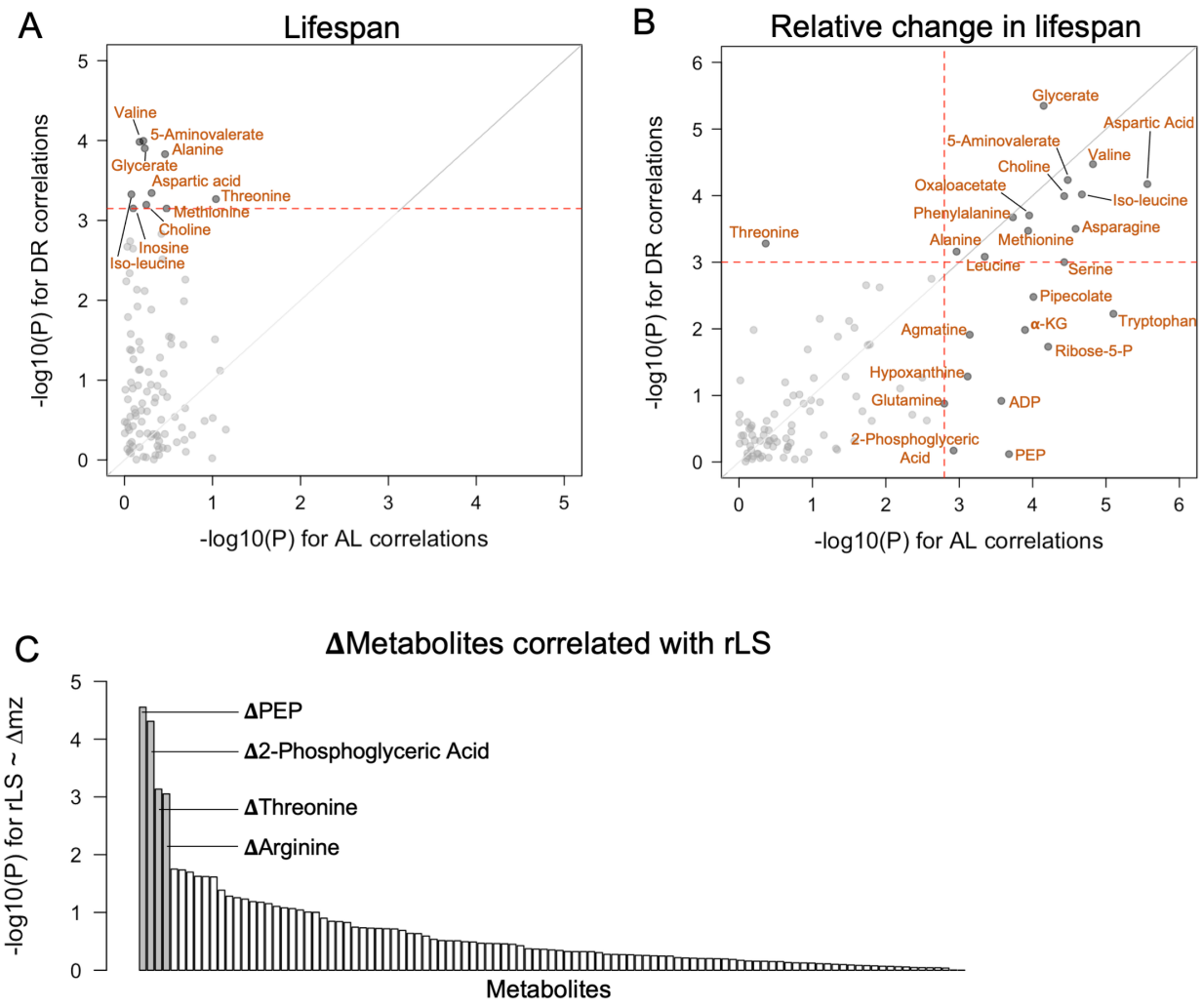


Figure 2. Metabolites significantly correlated with lifespan response.

(A,B) Results of univariate analysis modeling lifespan phenotypes as functions of individual metabolites measured under either AL or DR were performed and -log10(P values) were plotted. Each point represents a single metabolite and the significance of its association with (A) mean lifespan and (B) rLS. Red dotted lines represent FDR cutoff at α = 0.01. (C) Rank of -log10(P values) from linear regression modeling rLS as a function of change in metabolite abundance. Four labeled metabolites passed FDR cutoff of α = 0.05.

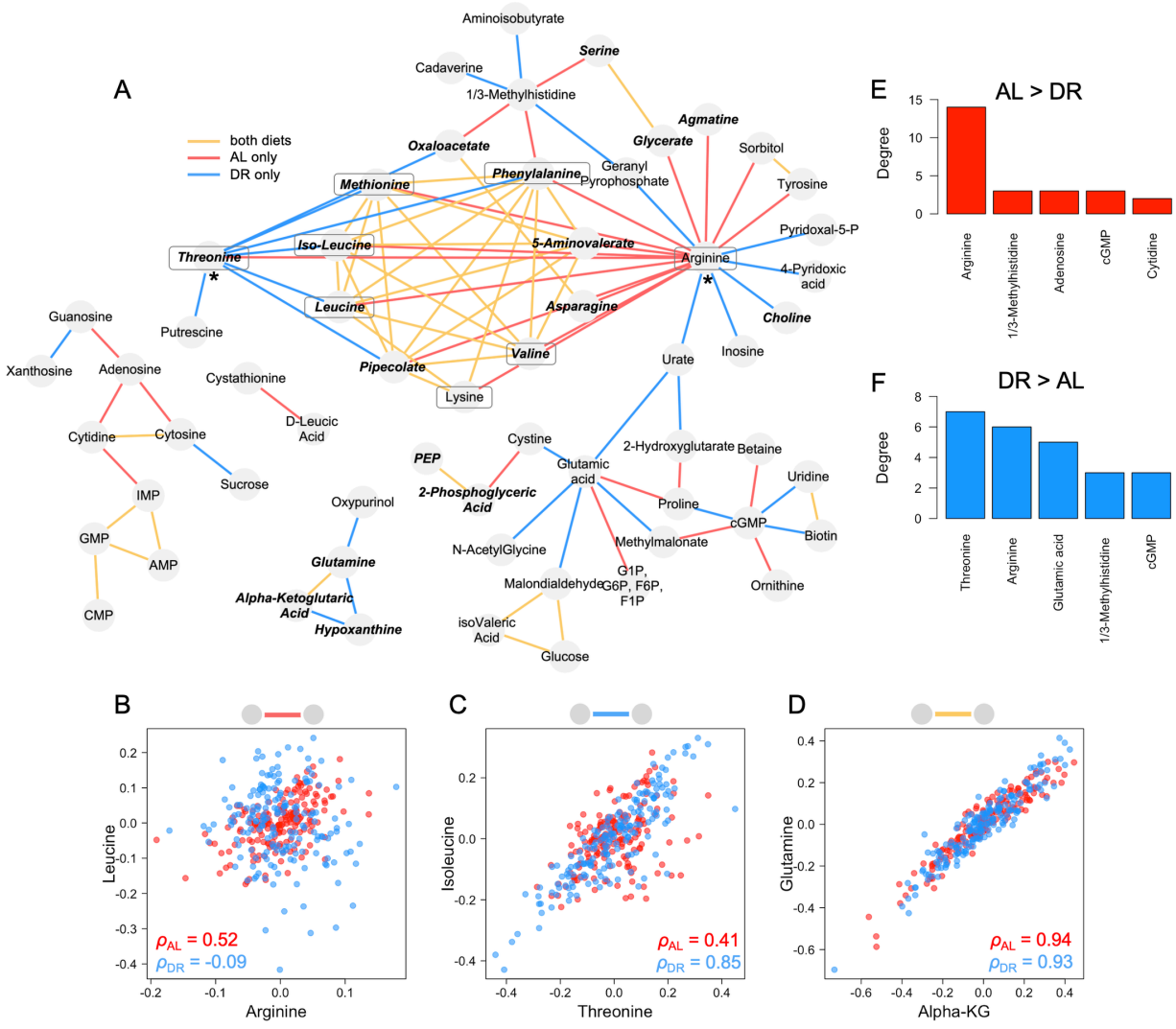


Figure 3. **Diet-dependent correlation network analysis.**

(A) Nodes represent metabolites and edges represent correlation between two metabolites. Edge color denotes correlations that significantly become more positive with AL (red), DR (blue), or have correlation coefficients that exceed abs(0.8) in both diets (yellow). For red and blue edges, only correlation coefficient differences of greater than abs(0.4) are shown. Bolded and italicized metabolites were also found to be significantly associated with rLS. Boxed metabolites are essential amino acids. Asterisks indicate suspected “hub” metabolites. (B-D) Examples of each type of covariance relationship (edge color) are shown. (E-F) Degrees (number of edges) of top 5 most connected nodes under each diet.

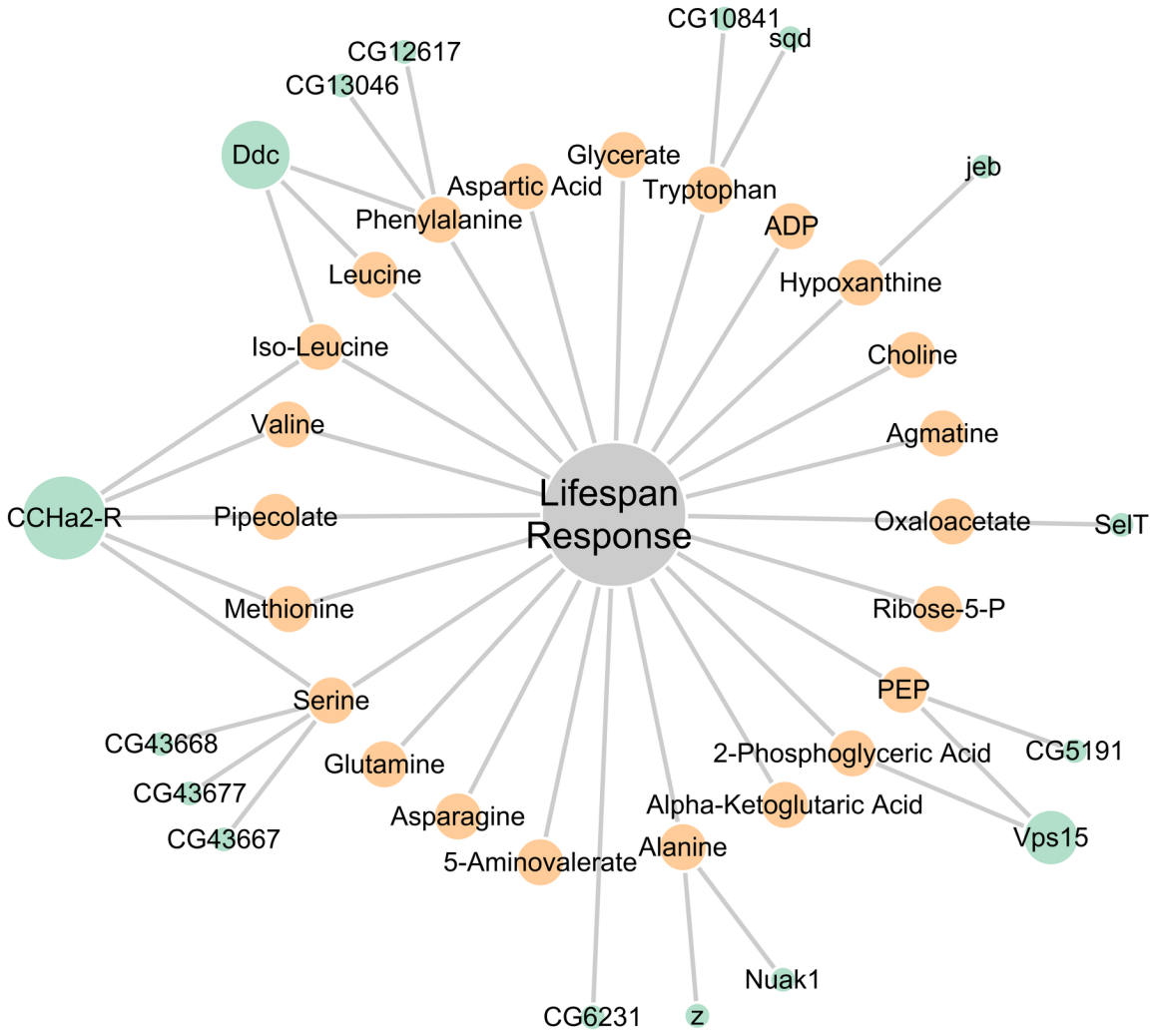


Figure 4. Multi-omic network for lifespan response.

Gene-metabolite-phenotype network was constructed from linear modeling and GWAS results from AL metabolites that were correlated with lifespan response as measured by rLS. Gene nodes are colored in teal, metabolite nodes are colored in yellow. Gene node size is directly proportional to node degree, while metabolite node size is held constant. An edge exists between a metabolite and lifespan response if the metabolite was significantly correlated with rLS at FDR cutoff of $\alpha = 0.01$. An edge exists between a gene and metabolite and/or lifespan response if it the gene had a score of ≤1E-4.5.

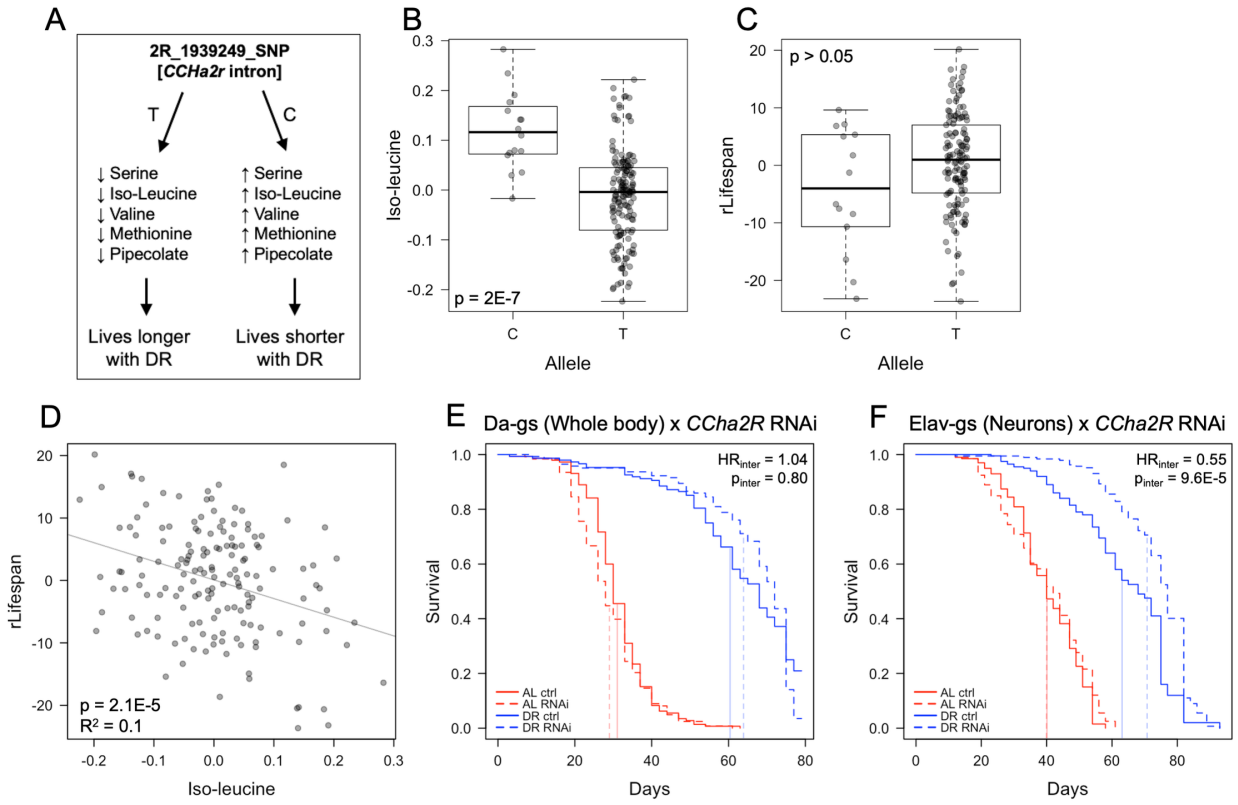


Figure 5. *CCHa2r* is associated with change in metabolite levels and modulates lifespan response to DR.

(A) Diagram of one of the candidate gene pathways identified from metabolite-gene-network analysis, *CCHa2r*, and its relationship with iso-leucine and lifespan response. (B-D) Iso-leucine and its relationship with *CCHa2r* SNP 2R_1939249_SNP and residual lifespan. (E-F) Survival of RNAi (+RU486) versus control (-RU486) flies of *CCHa2r* RNAi in whole-body (D; da-gal4-gs driver) and neurons (E; elav-gal4-gs driver). Vertical lines represent mean lifespan. All lifespan experiments were conducted with 150-200 flies per condition. *P* values from B and C are from plink linear GWAS model. *P* value from D is from a linear regression as summarized in Figure 2. Statistical model in E and F is a Cox Proportional Hazards model fitting survival as a function of diet, RNAi, and the interaction between diet and RNAi. Hazard ratios (HR) and P values are specific to the interaction term.

Supplementary Tables

Table S1. Statistical results for difference in abundance between AL and DR metabolite values.

Metabolite	Relative Change	t-statistic	p-value
Pyruvate	0.35	14.98	3E-33
Lactate	0.05	3.17	2E-03
Oxaloacetate	0.74	60.81	1E-120
PEP	-0.40	-18.00	7E-42
D-GA3P	-1.22	-40.39	3E-91
Glucose	1.73	167.92	4E-197
2-Phosphoglyceric Acid	0.53	24.13	7E-58
G1P, G6P, F6P, F1P	1.23	42.58	6E-95
3-Hydroxybutyric Acid	-0.24	-14.01	2E-30
Succinate	0.45	22.90	8E-55
Alpha-Ketoglutaric Acid	-0.21	-12.31	1E-25
Glycine	-0.42	-70.25	3E-131
Sarcosine	-1.66	-34.73	6E-81
Alanine	1.73	320.08	2E-246
Cadaverine	-0.81	-115.21	2E-168
Serine	0.91	103.57	3E-160
Proline	1.73	141.11	8E-184
5-Aminovalerate	0.45	37.04	3E-85
Valine	-0.01	-0.96	3E-01
Threonine	0.77	58.81	4E-118
Pipecolate	0.74	77.26	3E-138
Leucine	1.44	177.64	2E-201
Iso-Leucine	1.14	111.92	3E-166
Asparagine	0.35	27.00	1E-64
Glutamine	1.38	92.32	1E-151
Lysine	1.40	209.97	4E-214
Glutamic acid	1.29	144.65	1E-185
Methionine	0.45	33.20	6E-78
Histidine	2.06	204.66	3E-212
Phenylalanine	1.39	133.10	2E-179
1/3-Methylhistidine	-0.57	-29.56	2E-70
Arginine	1.89	355.25	2E-254
Tyrosine	0.08	4.04	8E-05
Cystine	-0.72	-23.20	1E-55
isoValeric Acid	-0.50	-49.69	7E-106
Aminoisobutyrate	-1.22	-179.45	4E-202
Betaine	0.64	29.60	2E-70
Taurine	0.98	93.91	6E-153
Aspartic Acid	-0.37	-17.25	9E-40
Carnitine	0.52	40.03	1E-90
Epinephrine	0.23	17.03	4E-39
Glycerate	-0.38	-40.29	4E-91
N-AcetylGlycine	-1.36	-85.20	1E-145
D-Leucic Acid	-0.78	-43.53	2E-96
Anthranilate	-0.97	-88.86	9E-149
2-Hydroxyglutarate	-0.19	-24.01	1E-57
Homogentisate	-0.36	-28.59	3E-68

Xanthurenate	1.66	147.62	3E-187
Pantothenate	0.61	52.37	1E-109
Cystathionine	-1.49	-117.63	6E-170
Tryptophan	0.45	48.89	1E-104
L-Kynurenine	0.14	9.22	9E-17
3-Hydroxykynurenine	0.68	53.57	3E-111
Cytosine	-1.42	-78.84	8E-140
Histamine	1.08	105.10	2E-161
Cytidine	-0.21	-12.50	4E-26
Uridine	-0.74	-61.58	2E-121
Adenosine	-0.32	-13.45	7E-29
Inosine	2.02	255.71	3E-229
1-Methyladenosine	-0.45	-38.49	6E-88
Guanosine	1.13	64.93	2E-125
1-Methylguanosine	-1.25	-162.21	2E-194
GMP	-1.44	-42.75	3E-95
Uracil	-0.85	-86.78	5E-147
Adenine	-1.38	-110.46	3E-165
Hypoxanthine	1.16	77.67	1E-138
Oxypurinol	0.42	28.72	2E-68
Orotate	-0.57	-48.54	3E-104
Allantoin	-0.01	-0.30	8E-01
Urate	1.17	88.79	1E-148
Xanthosine	0.33	24.12	8E-58
CMP	-1.75	-74.90	6E-136
cGMP	-1.47	-50.40	7E-107
AMP	-0.29	-9.41	3E-17
IMP	-0.66	-17.54	1E-40
ADP	-0.95	-51.40	3E-108
Choline	1.59	260.60	1E-230
Methylmalonate	-0.65	-19.39	1E-45
4-Pyridoxic acid	0.39	46.29	8E-101
Biotin	-0.25	-29.16	2E-69
Pyridoxal-5-P	-0.17	-28.79	1E-68
Trimethylamine	-0.85	-67.24	6E-128
Acetylcholine	-1.15	-77.20	3E-138
Glycerol-3-P	0.26	10.77	4E-21
13-HODE	-0.52	-41.53	4E-93
Arachidonate	-1.53	-129.29	4E-177
Acetylcarnitine	1.28	64.81	3E-125
Myristic Acid	-0.93	-93.90	6E-153
Margaric Acid	-0.94	-87.94	5E-148
Linoleic Acid	-0.61	-40.95	3E-92
Linolenic Acid	0.29	30.07	2E-71
Sorbitol	-1.32	-83.50	4E-144
Glyceraldehyde	-0.03	-1.77	8E-02
Inositol	-0.61	-69.71	1E-130
Glucoronate	-1.45	-128.02	2E-176
Lactose	0.39	7.07	3E-11
Sucrose	0.99	56.53	3E-115
Ornithine	-0.81	-37.13	2E-85
Citrulline	-0.73	-39.14	5E-89
Malondialdehyde	-0.02	-2.25	3E-02
Acetoacetate	-0.73	-36.66	1E-84

Putrescine	-1.53	-88.27	3E-148
Agmatine	-1.26	-119.65	3E-171
Ribose-5-P	0.81	56.32	6E-115
Geranyl Pyrophosphate	0.19	18.95	2E-44

Table S2. Statistical results for linear models of lifespan phenotypes as a function of DR metabolite abundance. Bolded values are significant at FDR 1%.

	DR lifespan		AL lifespan		Residual lifespan		Delta lifespan	
Metabolites (DR)	Coef.	P-Value	Coef.	P-Value	Coef.	P-Value	Coef.	P-Value
Pyruvate	0.06	9.8E-01	-0.78	7.9E-01	0.52	8.6E-01	0.84	7.9E-01
Lactate	0.35	9.3E-01	-2.19	5.1E-01	1.64	6.2E-01	2.54	4.8E-01
Oxaloacetate	-18.40	1.5E-03	0.19	9.7E-01	**-18.51**	**2.0E-04**	**-18.59**	**5.8E-04**
PEP	-1.12	7.7E-01	-3.61	2.7E-01	1.01	7.6E-01	2.50	4.9E-01
D-GA3P	-0.49	8.7E-01	-3.90	1.2E-01	1.80	4.8E-01	3.40	2.1E-01
Glucose	18.49	2.3E-03	4.40	4.0E-01	15.90	2.4E-03	14.09	1.3E-02
2-Phosphoglyceric Acid	-0.85	8.3E-01	-3.85	2.5E-01	1.42	6.7E-01	3.00	4.1E-01
G1P, G6P, F6P, F1P	2.68	2.2E-01	2.74	1.5E-01	1.07	5.7E-01	-0.05	9.8E-01
3-Hydroxybutyric Acid	5.36	1.9E-01	1.17	7.4E-01	4.67	1.9E-01	4.19	2.8E-01
Succinate	8.64	4.2E-02	8.82	1.5E-02	3.44	3.5E-01	-0.18	9.6E-01
Alpha-Ketoglutaric Acid	-9.65	2.8E-02	0.13	9.7E-01	-9.73	1.0E-02	-9.78	1.7E-02
Glycine	-18.08	1.3E-01	2.68	7.9E-01	-19.65	5.5E-02	-20.76	6.1E-02
Sarcosine	0.82	6.6E-01	-0.11	9.4E-01	0.88	5.9E-01	0.93	6.0E-01
Alanine	**52.27**	**1.5E-04**	19.84	1.0E-01	**40.59**	**6.9E-04**	32.43	1.3E-02
Cadaverine	12.06	2.6E-01	9.53	3.0E-01	6.46	4.8E-01	2.54	8.0E-01
Serine	-21.78	1.3E-02	5.13	5.0E-01	**-24.80**	**1.0E-03**	-26.92	9.7E-04
Proline	-2.89	5.8E-01	-3.56	4.3E-01	-0.79	8.6E-01	0.68	8.9E-01
5-Aminovalerate	**-19.39**	**1.0E-04**	-3.54	4.2E-01	**-17.30**	**5.8E-05**	**-15.84**	**7.4E-04**
Valine	**-19.28**	**1.0E-04**	-2.60	5.5E-01	**-17.75**	**3.4E-05**	**-16.68**	**3.5E-04**
Threonine	**-18.23**	**5.4E-04**	-4.12	3.7E-01	**-15.80**	**5.2E-04**	-14.11	4.5E-03
Pipecolate	-17.39	7.6E-03	-1.49	7.9E-01	-16.52	3.3E-03	-15.91	9.2E-03
Leucine	-20.31	5.8E-03	1.44	8.2E-01	**-21.16**	**8.3E-04**	-21.75	1.5E-03
Iso-Leucine	**-20.16**	**4.7E-04**	-1.37	7.9E-01	**-19.35**	**9.6E-05**	**-18.79**	**5.0E-04**
Asparagine	-17.66	3.1E-03	1.34	8.0E-01	**-18.45**	**3.2E-04**	**-19.00**	**6.3E-04**
Glutamine	-5.34	2.9E-01	2.03	6.4E-01	-6.53	1.3E-01	-7.37	1.2E-01
Lysine	-16.65	7.3E-02	-2.00	8.0E-01	-15.47	5.4E-02	-14.65	9.2E-02
Glutamic acid	7.38	3.4E-01	5.10	4.5E-01	4.38	5.1E-01	2.28	7.5E-01
Methionine	**-14.39**	**7.1E-04**	-2.11	5.7E-01	**-13.15**	**3.4E-04**	-12.28	2.1E-03
Histidine	15.94	3.6E-02	0.69	9.2E-01	15.54	1.8E-02	15.26	3.2E-02
Phenylalanine	-17.22	2.1E-03	1.05	8.3E-01	**-17.84**	**2.1E-04**	**-18.27**	**4.7E-04**
1/3-Methylhistidine	8.47	5.2E-02	7.95	3.4E-02	3.79	3.2E-01	0.52	9.0E-01
Arginine	35.09	2.9E-02	-0.79	9.5E-01	35.56	1.0E-02	35.88	1.7E-02
Tyrosine	-8.42	1.6E-02	-1.55	6.1E-01	-7.51	1.3E-02	-6.87	3.7E-02
Cystine	4.20	1.6E-01	2.85	2.7E-01	2.53	3.3E-01	1.35	6.3E-01
isoValeric Acid	5.44	3.8E-01	1.54	7.7E-01	4.54	3.9E-01	3.91	5.0E-01
Aminoisobutyrate	7.87	4.7E-01	2.44	7.9E-01	6.44	4.9E-01	5.44	5.9E-01
Betaine	3.37	3.2E-01	0.28	9.2E-01	3.20	2.8E-01	3.09	3.3E-01
Taurine	6.94	1.2E-01	-3.67	3.4E-01	9.10	1.7E-02	10.61	1.0E-02
Aspartic Acid	**-12.10**	**4.5E-04**	-0.46	8.8E-01	**-11.83**	**6.7E-05**	**-11.64**	**3.1E-04**
Carnitine	12.31	5.5E-03	3.36	3.9E-01	10.34	7.1E-03	8.96	3.2E-02
Epinephrine	15.75	7.4E-03	3.73	4.7E-01	13.55	7.6E-03	12.02	3.0E-02
Glycerate	**-28.12**	**1.3E-04**	1.14	8.6E-01	**-28.79**	**4.5E-06**	**-29.26**	**1.8E-05**
N-AcetylGlycine	3.90	3.9E-01	5.11	1.9E-01	0.89	8.2E-01	-1.21	7.8E-01
D-Leucic Acid	-7.53	4.1E-02	-3.77	2.4E-01	-5.31	9.7E-02	-3.76	2.8E-01
Anthranilate	-2.73	7.0E-01	-0.06	9.9E-01	-2.69	6.6E-01	-2.67	6.9E-01
2-Hydroxyglutarate	-11.24	2.8E-01	-1.02	9.1E-01	-10.64	2.4E-01	-10.23	2.9E-01
Homogentisate	-2.16	8.0E-01	-2.28	7.6E-01	-0.82	9.1E-01	0.11	9.9E-01

Xanthurenate	5.99	1.3E-01	1.05	7.6E-01	5.37	1.2E-01	4.94	1.8E-01
Pantothenate	12.81	3.6E-02	5.94	2.6E-01	9.31	7.9E-02	6.87	2.3E-01
Cystathionine	-8.51	1.1E-01	0.19	9.7E-01	-8.62	6.4E-02	-8.70	8.4E-02
Tryptophan	-17.52	4.6E-03	-4.79	3.7E-01	-14.71	6.0E-03	-12.74	2.9E-02
L-Kynurenine	6.76	1.1E-01	-0.44	9.0E-01	7.02	5.2E-02	7.20	6.6E-02
3-Hydroxykynurenine	-3.14	4.2E-01	-6.23	6.0E-02	0.53	8.8E-01	3.09	3.9E-01
Cytosine	-1.83	6.7E-01	-3.79	3.1E-01	0.41	9.1E-01	1.97	6.3E-01
Histamine	16.67	1.0E-02	3.63	5.2E-01	14.53	9.6E-03	13.04	3.3E-02
Cytidine	1.67	6.9E-01	-4.34	2.3E-01	4.23	2.4E-01	6.02	1.2E-01
Uridine	-5.99	3.1E-01	-3.44	5.0E-01	-3.97	4.4E-01	-2.55	6.5E-01
Adenosine	5.03	8.3E-02	3.63	1.5E-01	2.89	2.5E-01	1.40	6.1E-01
Inosine	**30.71**	**7.1E-04**	10.42	1.9E-01	24.58	1.8E-03	20.29	1.8E-02
1-Methyladenosine	1.40	7.2E-01	0.23	9.5E-01	1.26	7.1E-01	1.17	7.5E-01
Guanosine	2.79	6.0E-01	-1.89	6.8E-01	3.90	3.9E-01	4.67	3.5E-01
1-Methylguanosine	10.94	2.4E-01	14.27	7.4E-02	2.54	7.5E-01	-3.33	7.0E-01
GMP	2.60	1.9E-01	1.00	5.6E-01	2.01	2.4E-01	1.60	3.9E-01
Uracil	-4.71	5.1E-01	-2.36	7.0E-01	-3.32	5.9E-01	-2.35	7.3E-01
Adenine	5.56	3.3E-01	-6.38	2.0E-01	9.31	5.9E-02	11.93	2.5E-02
Hypoxanthine	6.79	7.9E-02	0.51	8.8E-01	6.49	5.2E-02	6.28	8.3E-02
Oxypurinol	4.68	2.9E-01	-0.81	8.3E-01	5.16	1.7E-01	5.49	1.8E-01
Orotate	-6.63	3.0E-01	-3.34	5.4E-01	-4.66	4.0E-01	-3.29	5.8E-01
Allantoin	-2.10	4.9E-01	-3.49	1.8E-01	-0.05	9.9E-01	1.39	6.2E-01
Urate	10.55	1.7E-01	-0.59	9.3E-01	10.90	1.0E-01	11.15	1.2E-01
Xanthosine	-1.88	7.3E-01	4.63	3.2E-01	-4.61	3.2E-01	-6.51	2.0E-01
CMP	6.91	3.1E-02	4.52	1.0E-01	4.25	1.3E-01	2.39	4.3E-01
cGMP	-1.32	5.6E-01	-0.30	8.8E-01	-1.14	5.6E-01	-1.01	6.3E-01
AMP	1.58	4.6E-01	1.67	3.7E-01	0.60	7.5E-01	-0.09	9.7E-01
IMP	1.47	3.9E-01	1.91	1.9E-01	0.34	8.2E-01	-0.44	7.8E-01
ADP	-8.69	2.6E-02	-5.80	8.6E-02	-5.28	1.2E-01	-2.89	4.3E-01
Choline	**34.29**	**6.4E-04**	1.24	8.9E-01	**33.56**	**1.0E-04**	**33.05**	**4.3E-04**
Methylmalonate	1.43	5.0E-01	0.16	9.3E-01	1.33	4.7E-01	1.26	5.2E-01
4-Pyridoxic acid	6.82	4.7E-01	5.35	5.1E-01	3.67	6.5E-01	1.47	8.7E-01
Biotin	-2.18	7.9E-01	-1.03	8.8E-01	-1.57	8.2E-01	-1.15	8.8E-01
Pyridoxal-5-P	18.58	1.4E-01	3.34	7.6E-01	16.62	1.3E-01	15.24	2.0E-01
Trimethylamine	7.86	1.4E-01	9.05	4.7E-02	2.54	5.8E-01	-1.19	8.1E-01
Acetylcholine	1.92	6.3E-01	-1.64	6.3E-01	2.89	4.0E-01	3.57	3.4E-01
Glycerol-3-P	0.05	9.9E-01	-4.21	1.5E-01	2.52	3.9E-01	4.25	1.8E-01
13-HODE	4.70	3.3E-01	2.99	4.7E-01	2.94	4.8E-01	1.71	7.1E-01
Arachidonate	-4.95	4.7E-01	-1.98	7.4E-01	-3.78	5.3E-01	-2.97	6.5E-01
Acetylcarnitine	5.91	7.6E-02	4.43	1.2E-01	3.30	2.5E-01	1.48	6.4E-01
Myristic Acid	5.79	4.8E-01	-5.66	4.2E-01	9.12	1.9E-01	11.45	1.3E-01
Margaric Acid	0.42	9.5E-01	-6.31	2.8E-01	4.13	4.8E-01	6.73	2.9E-01
Linoleic Acid	-4.82	1.7E-01	-2.93	3.3E-01	-3.09	3.1E-01	-1.88	5.6E-01
Linolenic Acid	6.65	2.4E-01	4.81	3.3E-01	3.82	4.4E-01	1.84	7.3E-01
Sorbitol	-13.01	1.8E-03	-3.33	3.6E-01	-11.05	2.2E-03	-9.68	1.4E-02
Glyceraldehyde	-8.46	3.3E-02	-6.76	4.8E-02	-4.48	1.9E-01	-1.70	6.5E-01
Inositol	2.62	7.2E-01	-2.76	6.6E-01	4.25	5.0E-01	5.38	4.3E-01
Glucoronate	8.04	2.3E-01	-2.48	6.7E-01	9.50	1.0E-01	10.52	9.3E-02
Lactose	-1.16	4.2E-01	-1.03	4.1E-01	-0.55	6.6E-01	-0.12	9.3E-01
Sucrose	3.04	4.6E-01	3.11	3.8E-01	1.21	7.3E-01	-0.07	9.9E-01
Ornithine	1.92	5.9E-01	-2.43	4.3E-01	3.35	2.8E-01	4.34	2.0E-01
Citrulline	-0.38	9.1E-01	-1.24	6.7E-01	0.36	9.0E-01	0.87	7.8E-01

Malondialdehyde	9.76	1.2E-01	2.85	6.0E-01	8.09	1.3E-01	6.91	2.4E-01
Acetoacetate	0.64	8.7E-01	1.97	5.6E-01	-0.52	8.8E-01	-1.33	7.1E-01
Putrescine	-9.12	7.4E-02	-5.23	2.4E-01	-6.04	1.7E-01	-3.88	4.2E-01
Agmatine	-14.09	4.2E-02	1.53	8.0E-01	-14.99	1.2E-02	-15.62	1.6E-02
Ribose-5-P	8.26	5.5E-02	-0.81	8.3E-01	8.74	1.9E-02	9.07	2.4E-02
Geranyl Pyrophosphate	22.71	1.2E-02	7.76	3.2E-01	18.15	2.0E-02	14.96	7.9E-02

Table S3. Statistical results for linear models of lifespan phenotypes as a function of AL metabolite abundance. Bolded values are significant at FDR 1%.

	DR lifespan		AL lifespan		Residual lifespan		Delta lifespan	
Metabolites (AL)	Coef.	P-Value	Coef.	P-Value	Coef.	P-Value	Coef.	P-Value
Pyruvate	4.24	1.7E-01	1.95	4.6E-01	3.10	2.5E-01	2.29	4.3E-01
Lactate	6.31	1.6E-01	-2.20	5.7E-01	7.60	4.8E-02	8.50	3.9E-02
Oxaloacetate	**-19.99**	**1.7E-04**	-4.06	3.8E-01	**-17.60**	**1.1E-04**	**-15.92**	**1.2E-03**
PEP	10.14	2.1E-03	-0.48	8.7E-01	**10.43**	**2.1E-04**	**10.62**	**4.6E-04**
D-GA3P	0.97	7.4E-01	-2.05	4.0E-01	2.17	3.8E-01	3.02	2.6E-01
Glucose	12.23	2.3E-02	1.17	8.0E-01	11.54	1.2E-02	11.06	2.6E-02
2-Phosphoglyceric Acid	9.52	4.1E-03	0.56	8.5E-01	**9.19**	**1.2E-03**	8.96	3.4E-03
G1P, G6P, F6P, F1P	1.02	7.6E-01	3.59	2.0E-01	-1.09	7.0E-01	-2.57	4.0E-01
3-Hydroxybutyric Acid	5.24	2.8E-01	2.01	6.3E-01	4.05	3.3E-01	3.23	4.7E-01
Succinate	-3.27	3.7E-01	1.39	6.6E-01	-4.08	1.9E-01	-4.65	1.7E-01
Alpha-Ketoglutaric Acid	**-16.40**	**1.4E-04**	-3.79	3.1E-01	**-14.17**	**1.3E-04**	**-12.61**	**1.7E-03**
Glycine	-14.76	1.9E-01	2.36	8.1E-01	-16.15	9.6E-02	-17.12	1.0E-01
Sarcosine	-2.56	2.2E-01	-1.44	4.2E-01	-1.71	3.4E-01	-1.12	5.7E-01
Alanine	49.40	1.0E-03	12.27	3.5E-01	**42.18**	**1.1E-03**	37.13	7.8E-03
Cadaverine	7.12	5.6E-01	-3.27	7.6E-01	9.05	3.9E-01	10.39	3.6E-01
Serine	-22.86	1.6E-03	4.27	5.0E-01	**-25.37**	**3.7E-05**	**-27.13**	**4.1E-05**
Proline	-10.06	2.5E-01	-6.23	4.0E-01	-6.39	4.0E-01	-3.83	6.4E-01
5-Aminovalerate	**-24.97**	**1.4E-04**	-2.92	6.1E-01	**-23.25**	**3.3E-05**	**-22.05**	**2.8E-04**
Valine	**-24.45**	**9.2E-05**	-2.30	6.7E-01	**-23.09**	**1.5E-05**	**-22.14**	**1.3E-04**
Threonine	-11.40	1.3E-01	-10.78	9.1E-02	-5.05	4.3E-01	-0.62	9.3E-01
Pipecolate	-21.97	2.4E-03	3.36	5.9E-01	**-23.95**	**9.7E-05**	**-25.33**	**1.3E-04**
Leucine	-35.45	3.0E-03	0.59	9.5E-01	**-35.80**	**4.5E-04**	**-36.04**	**1.0E-03**
Iso-Leucine	**-30.69**	**1.9E-04**	-1.49	8.4E-01	**-29.82**	**2.1E-05**	**-29.21**	**1.2E-04**
Asparagine	**-21.03**	**4.8E-05**	-4.05	3.7E-01	**-18.64**	**2.6E-05**	**-16.98**	**4.1E-04**
Glutamine	-17.30	1.0E-03	-5.07	2.7E-01	**-14.32**	**1.6E-03**	-12.23	1.3E-02
Lysine	-28.96	2.2E-02	4.70	6.6E-01	-31.73	3.2E-03	-33.67	3.6E-03
Glutamic acid	-10.33	2.9E-01	0.61	9.4E-01	-10.69	2.0E-01	-10.94	2.2E-01
Methionine	-15.40	5.4E-03	4.63	3.3E-01	**-18.12**	**1.2E-04**	**-20.02**	**7.3E-05**
Histidine	20.50	1.0E-02	7.18	3.0E-01	16.27	1.8E-02	13.32	7.2E-02
Phenylalanine	-26.20	1.7E-03	0.63	9.3E-01	**-26.57**	**1.9E-04**	**-26.83**	**4.7E-04**
1/3-Methylhistidine	2.08	4.7E-01	2.07	3.9E-01	0.86	7.2E-01	0.01	1.0E+00
Arginine	2.01	9.1E-01	15.72	2.9E-01	-7.25	6.3E-01	-13.71	4.0E-01
Tyrosine	-6.78	7.6E-02	-0.35	9.1E-01	-6.57	4.5E-02	-6.42	6.9E-02
Cystine	0.58	8.1E-01	-2.00	3.3E-01	1.76	3.9E-01	2.58	2.4E-01
isoValeric Acid	4.11	4.5E-01	1.17	8.0E-01	3.43	4.6E-01	2.95	5.6E-01
Aminoisobutyrate	-0.33	9.8E-01	-4.07	6.7E-01	2.07	8.3E-01	3.74	7.2E-01
Betaine	-9.18	5.8E-02	-6.36	1.2E-01	-5.43	1.9E-01	-2.81	5.3E-01
Taurine	8.46	1.0E-01	-3.53	4.2E-01	10.54	1.6E-02	11.99	1.1E-02
Aspartic Acid	**-12.74**	**1.3E-05**	-1.76	4.9E-01	**-11.70**	**2.7E-06**	**-10.97**	**5.0E-05**
Carnitine	10.48	3.2E-02	5.31	2.0E-01	7.35	8.0E-02	5.17	2.5E-01
Epinephrine	7.34	9.7E-02	-1.34	7.2E-01	8.13	3.2E-02	8.68	3.3E-02
Glycerate	-19.58	1.9E-03	2.96	5.9E-01	**-21.32**	**7.1E-05**	**-22.53**	**9.5E-05**
N-AcetylGlycine	-2.36	5.9E-01	0.43	9.1E-01	-2.61	4.8E-01	-2.78	4.9E-01
D-Leucic Acid	-5.09	1.9E-01	-0.54	8.7E-01	-4.77	1.5E-01	-4.54	2.0E-01
Anthranilate	0.58	9.2E-01	4.53	3.7E-01	-2.09	6.8E-01	-3.95	4.7E-01
2-Hydroxyglutarate	-24.68	6.4E-03	-2.56	7.4E-01	-23.17	2.8E-03	-22.12	8.1E-03
Homogentisate	6.62	2.1E-01	0.66	8.8E-01	6.23	1.7E-01	5.96	2.3E-01

Xanthurenate	5.74	1.7E-01	-0.10	9.8E-01	5.80	1.0E-01	5.84	1.3E-01
Pantothenate	19.16	2.9E-03	6.90	2.1E-01	15.10	6.4E-03	12.26	4.1E-02
Cystathionine	-7.04	1.7E-01	-2.98	5.0E-01	-5.29	2.4E-01	-4.06	4.0E-01
Tryptophan	**-26.92**	**2.0E-04**	0.99	8.8E-01	**-27.50**	**7.9E-06**	**-27.91**	**2.7E-05**
L-Kynurenine	10.14	5.7E-02	0.89	8.5E-01	9.61	3.5E-02	9.25	6.0E-02
3-Hydroxykynurenine	-1.36	7.3E-01	-6.13	7.1E-02	2.25	5.1E-01	4.77	2.0E-01
Cytosine	7.93	1.6E-01	1.15	8.1E-01	7.25	1.3E-01	6.78	1.9E-01
Histamine	15.20	1.1E-02	6.44	2.1E-01	11.41	2.7E-02	8.76	1.2E-01
Cytidine	12.64	2.2E-02	2.00	6.7E-01	11.46	1.6E-02	10.64	3.7E-02
Uridine	-2.47	7.0E-01	4.70	3.8E-01	-5.24	3.3E-01	-7.17	2.2E-01
Adenosine	3.02	4.8E-01	3.31	3.6E-01	1.07	7.7E-01	-0.29	9.4E-01
Inosine	25.01	6.3E-03	2.04	8.0E-01	23.81	2.4E-03	22.97	6.6E-03
1-Methyladenosine	4.01	3.5E-01	-3.03	4.1E-01	5.79	1.1E-01	7.04	7.3E-02
Guanosine	3.57	3.4E-01	3.65	2.5E-01	1.42	6.6E-01	-0.08	9.8E-01
1-Methylguanosine	-2.50	7.8E-01	3.50	6.4E-01	-4.56	5.5E-01	-6.00	4.6E-01
GMP	2.18	2.9E-01	1.30	4.6E-01	1.42	4.2E-01	0.88	6.4E-01
Uracil	9.11	1.6E-01	3.74	5.0E-01	6.90	2.1E-01	5.36	3.7E-01
Adenine	0.25	9.7E-01	-0.04	9.9E-01	0.27	9.6E-01	0.29	9.6E-01
Hypoxanthine	17.24	1.1E-02	-3.64	5.3E-01	**19.38**	**7.7E-04**	**20.88**	**7.5E-04**
Oxypurinol	11.35	5.9E-02	-0.91	8.6E-01	11.89	2.1E-02	12.26	2.7E-02
Orotate	5.60	3.3E-01	8.13	9.9E-02	0.82	8.7E-01	-2.53	6.4E-01
Allantoin	-1.41	6.3E-01	-3.40	1.7E-01	0.59	8.1E-01	1.99	4.6E-01
Urate	10.49	4.9E-02	0.48	9.2E-01	10.20	2.5E-02	10.00	4.2E-02
Xanthosine	3.77	4.9E-01	0.40	9.3E-01	3.54	4.5E-01	3.37	5.0E-01
CMP	6.46	1.8E-02	3.93	9.4E-02	4.14	7.9E-02	2.53	3.2E-01
cGMP	-4.14	1.8E-01	-3.32	2.1E-01	-2.18	4.1E-01	-0.81	7.8E-01
AMP	2.24	4.3E-01	1.45	5.5E-01	1.39	5.7E-01	0.79	7.6E-01
IMP	1.76	3.8E-01	2.08	2.2E-01	0.54	7.6E-01	-0.32	8.6E-01
ADP	-10.47	1.3E-03	-0.53	8.5E-01	**-10.16**	**2.7E-04**	**-9.94**	**9.5E-04**
Choline	34.19	1.3E-03	-5.32	5.6E-01	**37.31**	**3.7E-05**	**39.50**	**5.0E-05**
Methylmalonate	0.09	9.7E-01	-1.70	4.2E-01	1.09	6.1E-01	1.79	4.4E-01
4-Pyridoxic acid	13.02	2.2E-01	-8.44	3.5E-01	17.98	4.6E-02	21.45	2.7E-02
Biotin	8.56	3.0E-01	8.95	2.0E-01	3.29	6.4E-01	-0.39	9.6E-01
Pyridoxal-5-P	8.61	5.5E-01	-11.06	3.7E-01	15.12	2.2E-01	19.68	1.4E-01
Trimethylamine	8.47	1.7E-01	3.14	5.5E-01	6.62	2.2E-01	5.32	3.5E-01
Acetylcholine	-0.01	1.0E+00	-0.47	8.7E-01	0.26	9.3E-01	0.46	8.8E-01
Glycerol-3-P	5.86	1.7E-01	-1.23	7.3E-01	6.59	7.0E-02	7.09	7.0E-02
13-HODE	3.75	4.3E-01	1.90	6.4E-01	2.63	5.2E-01	1.85	6.7E-01
Arachidonate	2.44	7.3E-01	4.39	4.7E-01	-0.14	9.8E-01	-1.94	7.7E-01
Acetylcarnitine	2.27	3.9E-01	3.90	8.2E-02	-0.02	9.9E-01	-1.63	5.0E-01
Myristic Acid	18.61	4.1E-02	-6.09	4.4E-01	22.19	4.4E-03	24.69	3.1E-03
Margaric Acid	5.86	2.9E-01	-7.67	1.0E-01	10.37	2.7E-02	13.52	7.1E-03
Linoleic Acid	-2.37	4.7E-01	-2.12	4.4E-01	-1.12	6.9E-01	-0.25	9.3E-01
Linolenic Acid	8.21	1.8E-01	3.12	5.5E-01	6.37	2.3E-01	5.09	3.7E-01
Sorbitol	-9.17	3.5E-02	-0.64	8.6E-01	-8.80	1.9E-02	-8.53	3.4E-02
Glyceraldehyde	-0.96	8.5E-01	-1.84	6.8E-01	0.12	9.8E-01	0.88	8.6E-01
Inositol	2.56	7.0E-01	-5.53	3.3E-01	5.81	3.0E-01	8.09	1.8E-01
Glucoronate	7.88	2.0E-01	1.49	7.8E-01	7.00	1.8E-01	6.39	2.6E-01
Lactose	-1.91	2.6E-01	-0.21	8.9E-01	-1.79	2.2E-01	-1.70	2.8E-01
Sucrose	0.86	8.0E-01	0.03	9.9E-01	0.84	7.7E-01	0.83	7.9E-01
Ornithine	1.51	6.4E-01	0.59	8.3E-01	1.17	6.7E-01	0.93	7.5E-01
Citrulline	-2.26	5.3E-01	-1.56	6.1E-01	-1.34	6.6E-01	-0.70	8.3E-01

Malondialdehyde	7.72	1.4E-01	1.70	7.0E-01	6.72	1.3E-01	6.02	2.1E-01
Acetoacetate	-0.50	8.8E-01	-2.24	4.4E-01	0.82	7.8E-01	1.74	5.8E-01
Putrescine	-6.87	1.1E-01	-1.50	6.9E-01	-5.98	1.1E-01	-5.37	1.8E-01
Agmatine	-18.11	5.5E-03	1.18	8.3E-01	**-18.81**	**7.2E-04**	**-19.30**	**1.3E-03**
Ribose-5-P	20.72	1.0E-03	-1.44	7.9E-01	**21.57**	**6.1E-05**	**22.17**	**1.3E-04**
Geranyl Pyrophosphate	9.84	2.5E-01	-2.67	7.1E-01	11.40	1.2E-01	12.50	1.1E-01

Table S4. Statistical results for linear models of lifespan phenotypes as a function of delta metabolite (DR - AL metabolite). Bold values are significant at FDR 5%

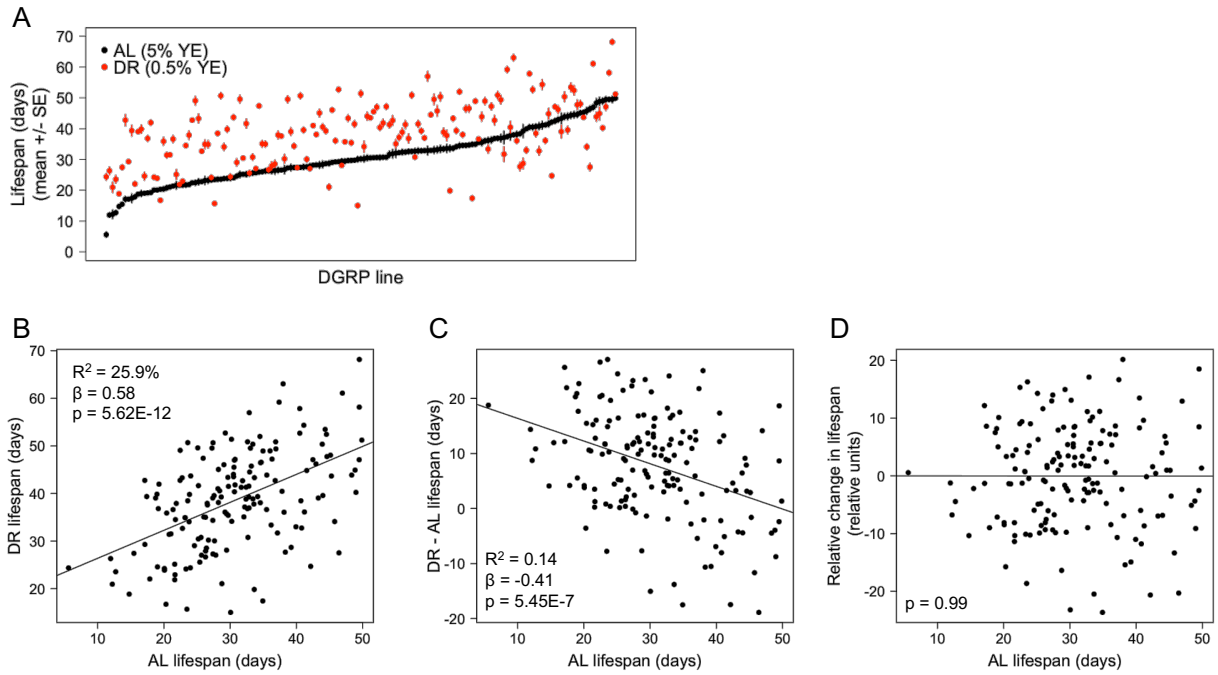
Metabolites (delta)	Residual lifespan	
	Coef.	P-Value
Pyruvate	0.45	8.4E-01
Lactate	-2.70	3.8E-01
Oxaloacetate	0.04	9.9E-01
PEP	**-9.61**	**2.8E-05**
D-GA3P	-2.57	1.4E-01
Glucose	6.97	1.9E-01
2-Phosphoglyceric Acid	**-9.40**	**4.9E-05**
G1P, G6P, F6P, F1P	1.72	3.5E-01
3-Hydroxybutyric Acid	6.87	2.4E-02
Succinate	3.44	1.9E-01
Alpha-Ketoglutaric Acid	0.91	7.7E-01
Glycine	-5.50	5.4E-01
Sarcosine	2.61	2.0E-02
Alanine	22.68	1.8E-02
Cadaverine	4.47	5.5E-01
Serine	3.40	5.7E-01
Proline	2.94	5.0E-01
5-Aminovalerate	-7.47	8.6E-02
Valine	-7.12	9.9E-02
Threonine	**-13.55**	**7.3E-04**
Pipecolate	-2.11	7.0E-01
Leucine	-12.41	5.6E-02
Iso-Leucine	-9.46	6.5E-02
Asparagine	1.75	6.7E-01
Glutamine	1.80	6.1E-01
Lysine	-3.76	6.3E-01
Glutamic acid	7.03	2.3E-01
Methionine	-3.67	3.4E-01
Histidine	2.52	6.3E-01
Phenylalanine	-11.11	2.4E-02
1/3-Methylhistidine	0.85	7.5E-01
Arginine	**31.94**	**8.9E-04**
Tyrosine	-5.57	5.2E-02
Cystine	-1.93	2.6E-01
isoValeric Acid	5.00	3.6E-01
Aminoisobutyrate	11.47	1.4E-01
Betaine	4.14	9.1E-02
Taurine	7.20	1.5E-01
Aspartic Acid	0.30	9.1E-01
Carnitine	3.84	3.5E-01
Epinephrine	1.27	7.4E-01
Glycerate	-3.44	5.4E-01
N-AcetylGlycine	4.16	2.0E-01
D-Leucic Acid	-0.94	7.5E-01
Anthranilate	-0.57	9.1E-01
2-Hydroxyglutarate	11.33	9.9E-02
Homogentisate	-3.32	4.3E-01
Xanthurenate	1.85	6.9E-01
Pantothenate	-0.81	8.6E-01
Cystathionine	-4.40	2.9E-01
Tryptophan	3.40	5.6E-01
L-Kynurenine	3.56	3.1E-01
3-Hydroxykynurenine	-2.98	4.8E-01
Cytosine	-1.71	5.7E-01
Histamine	1.90	7.1E-01
Cytidine	-0.79	8.0E-01

Uridine	3.26	4.8E-01
Adenosine	2.95	1.8E-01
Inosine	8.58	1.9E-01
1-Methyladenosine	-2.27	6.2E-01
Guanosine	2.27	4.6E-01
1-Methylguanosine	12.20	7.8E-02
GMP	0.35	8.2E-01
Uracil	-0.71	9.0E-01
Adenine	7.98	5.9E-02
Hypoxanthine	2.87	4.3E-01
Oxypurinol	3.56	3.4E-01
Orotate	-4.47	3.1E-01
Allantoin	1.46	4.7E-01
Urate	-0.97	8.1E-01
Xanthosine	-3.76	3.2E-01
CMP	-1.77	4.5E-01
cGMP	0.62	7.4E-01
AMP	-0.23	9.0E-01
IMP	-0.53	7.0E-01
ADP	2.91	3.2E-01
Choline	8.60	3.1E-01
Methylmalonate	0.60	7.1E-01
4-Pyridoxic acid	3.06	6.3E-01
Biotin	-0.10	9.9E-01
Pyridoxal-5-P	11.83	1.9E-01
Trimethylamine	-5.78	1.8E-01
Acetylcholine	2.83	4.2E-01
Glycerol-3-P	-5.00	2.4E-02
13-HODE	0.72	8.7E-01
Arachidonate	1.26	7.7E-01
Acetylcarnitine	5.37	4.1E-02
Myristic Acid	-6.44	2.3E-01
Margaric Acid	-8.41	8.3E-02
Linoleic Acid	-2.54	4.8E-01
Linolenic Acid	-2.50	6.5E-01
Sorbitol	-6.19	6.7E-02
Glyceraldehyde	-7.08	1.8E-02
Inositol	-3.06	6.1E-01
Glucoronate	8.25	7.0E-02
Lactose	0.10	9.2E-01
Sucrose	-0.64	8.3E-01
Ornithine	1.41	5.6E-01
Citrulline	0.44	8.8E-01
Malondialdehyde	3.38	5.3E-01
Acetoacetate	0.71	7.9E-01
Putrescine	1.43	6.4E-01
Agmatine	3.10	5.3E-01
Ribose-5-P	-0.68	8.5E-01
Geranyl Pyrophosphate	9.19	1.3E-01

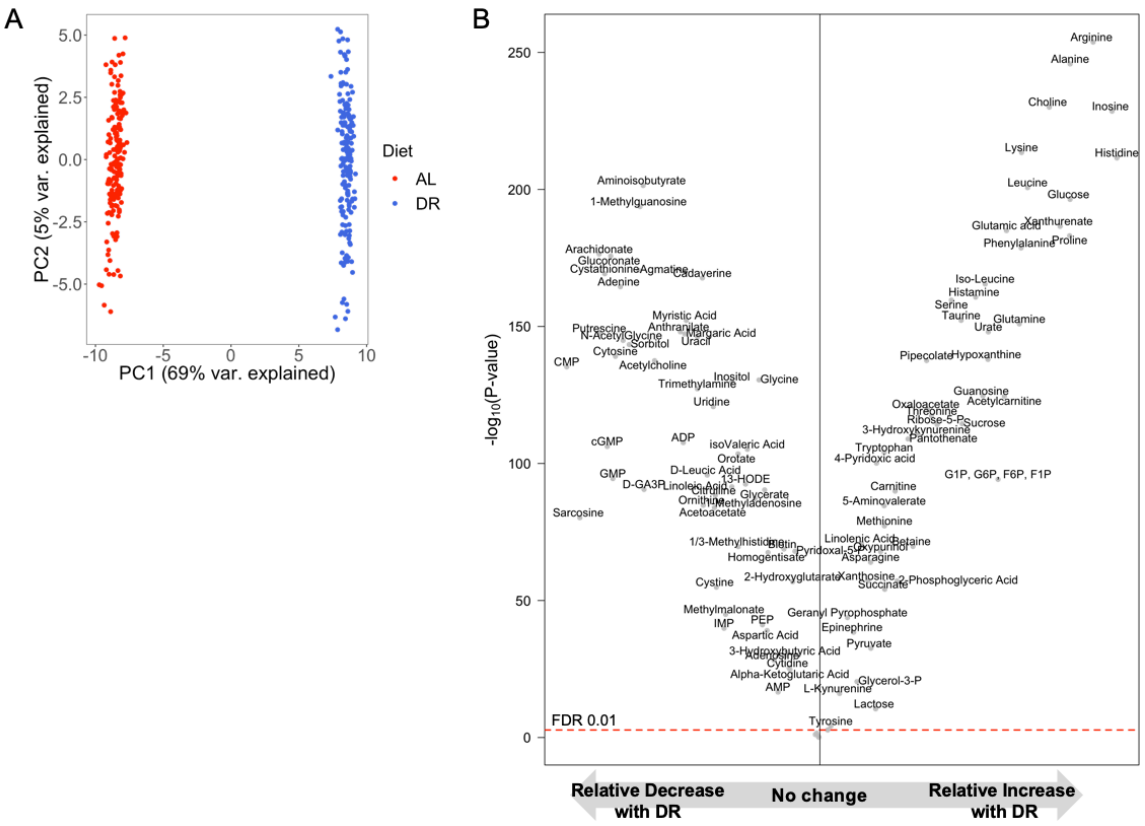
Table S5. Gene scores for AL and DR bipartite networks

AL or DR Network?	GWAS Pheno	Gene	Score
AL	Oxaloacetate	SelT	2.6E-05
AL	Leucine	Ddc	9.0E-06
AL	Iso-Leucine	Ddc	3.0E-05
AL	Iso-Leucine	CCHa2-R	2.5E-05
AL	Methionine	CCHa2-R	1.0E-05
AL	Phenylalanine	Ddc	1.7E-05
AL	Phenylalanine	CG12617	2.4E-05
AL	Phenylalanine	CG13046	2.5E-05
AL	Tryptophan	sqd	1.9E-05
AL	Tryptophan	CG10841	1.1E-05
AL	Hypoxanthine	jeb	8.0E-06
AL	PEP	CG5191	1.3E-05
AL	PEP	Vps15	2.0E-06
AL	2-Phosphoglyceric Acid	Vps15	1.0E-06
AL	Alanine	z	1.2E-05
AL	Alanine	CG43143	4.0E-06
AL	Serine	CCHa2-R	8.0E-06
AL	Serine	CG43667	1.5E-05
AL	Serine	CG43668	1.2E-05
AL	Serine	CG43677	1.1E-05
AL	Valine	CCHa2-R	3.1E-05
AL	Pipecolate	CCHa2-R	2.3E-05
DR	Methionine	CCHa2-R	9.0E-06
DR	Phenylalanine	Ddc	5.0E-06
DR	Phenylalanine	CG13046	2.9E-05
DR	Phenylalanine	CG12617	1.5E-05
DR	Alanine	z	1.0E-05
DR	Alanine	CG43143	3.0E-06
DR	Serine	CG43667	1.3E-05
DR	Serine	CCHa2-R	9.0E-06
DR	Serine	CG43677	1.7E-05
DR	Serine	CG43668	2.1E-05
DR	5-Aminovalerate	CCHa2-R	2.5E-05
DR	Valine	CCHa2-R	2.0E-05
DR	Leucine	Ddc	1.8E-05
AL and DR	rLifespan	CG6231	2.6E-05

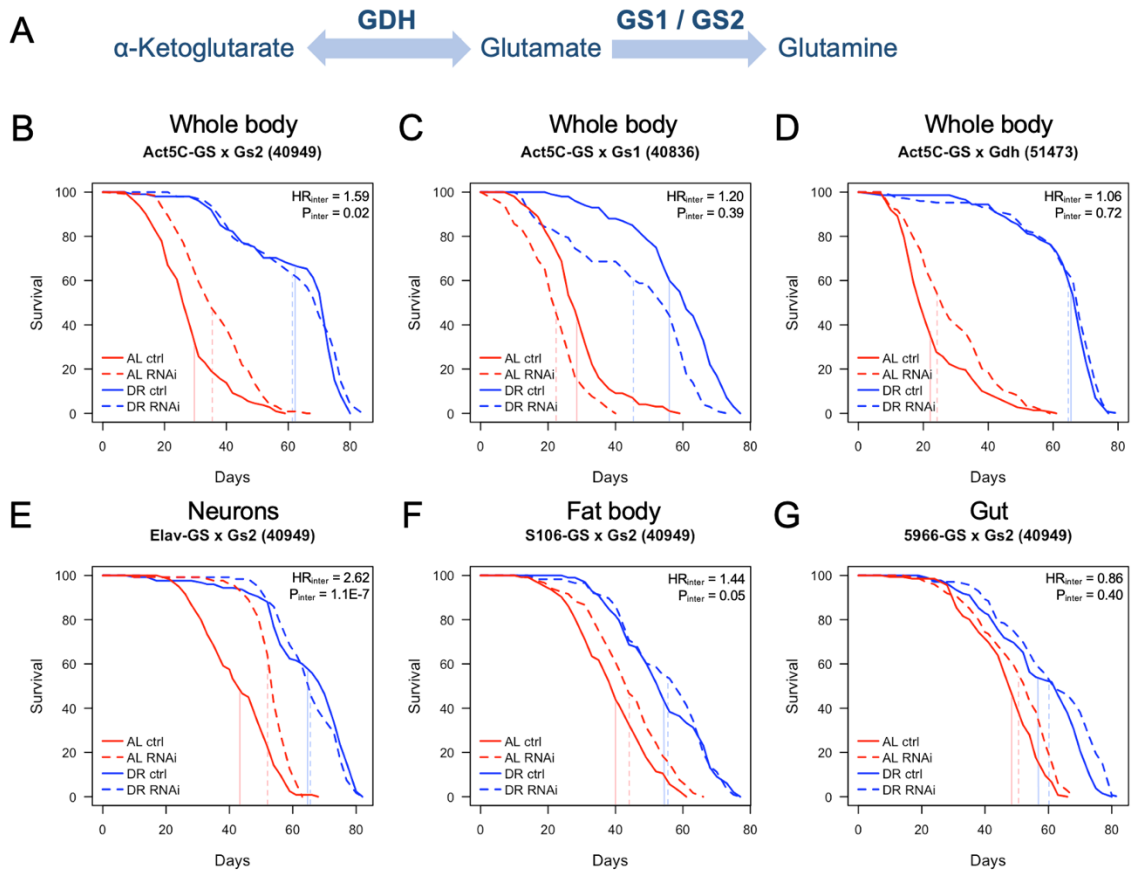
Supplementary Figures



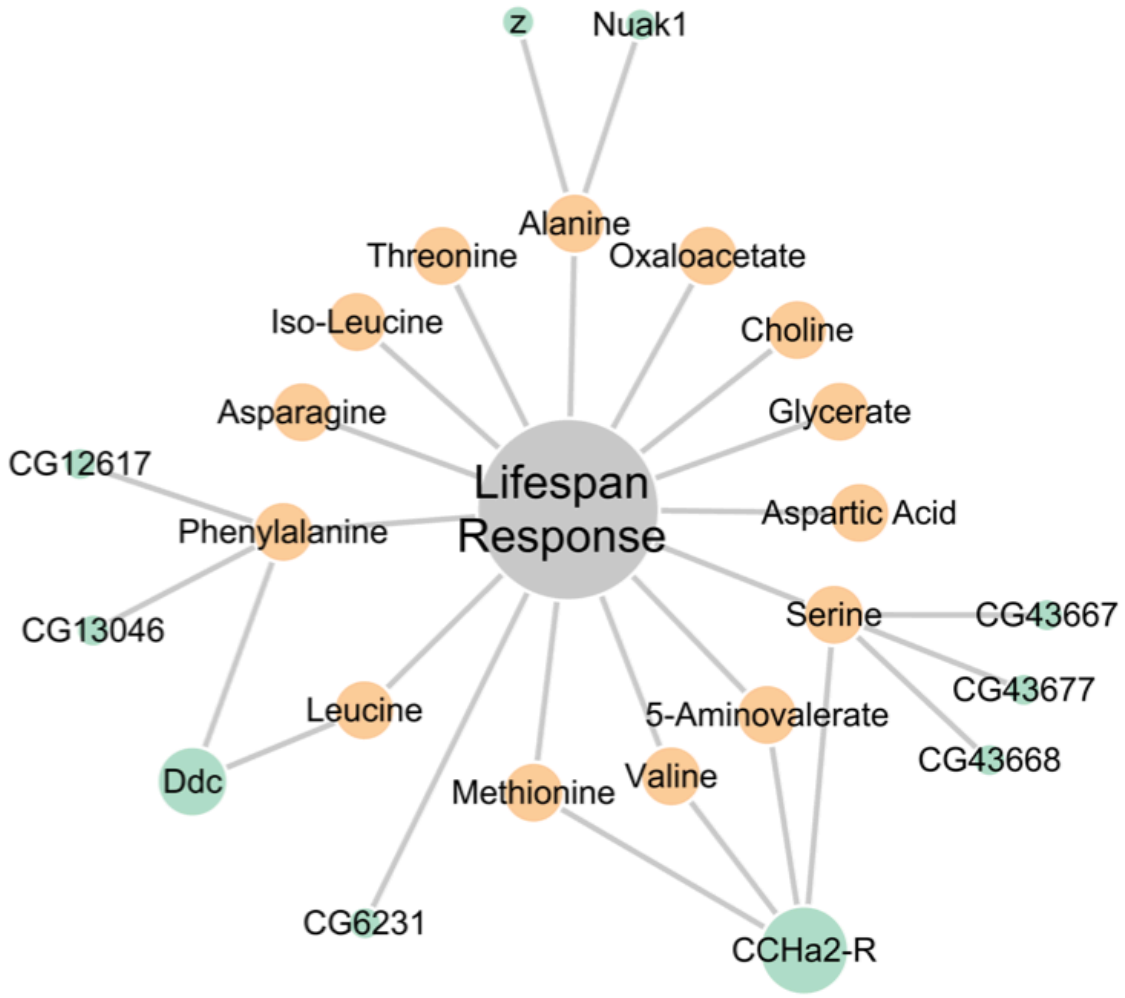
S1 Fig. DGRP lifespan on AL and DR replotted from Wilson *et al*, 2020. (A) Mean lifespan of 161 DGRP lines on AL (5% yeast extract) or DR (0.5% yeast extract) diets. Least squares linear regression of DR lifespan (B) and change in lifespan (DR – AL; C) and rLS (D) as a function of AL lifespan.



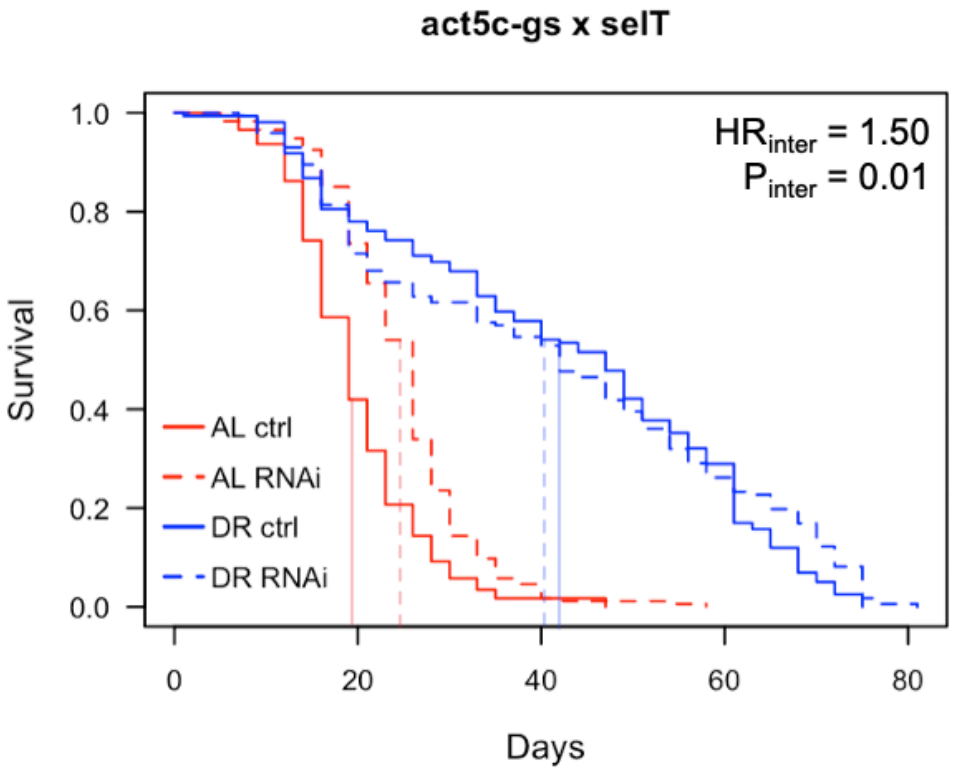
S2 Fig. Diet restriction dramatically remodels the metabolome. (A) PCA of all samples using metabolite profiles colored by diet. Ellipses are drawn with confidence level of 90%. (B) Volcano plot of significance of the difference between metabolite abundance with DR reveals that almost all metabolites are highly significantly changed with DR. Each point on the plot represents a single metabolite result from a pairwise Student's t-test.



S3 Fig. Diet-dependent survival of GDH, GS1, and GS2 RNAi flies. (A) The α-KG/glutamine pathway was manipulated by inhibiting transcript levels of *gs1/2* and *gdh*. (B-G) Survival of RNAi (+RU486) versus control (-RU486) flies of GS2, GS1, and GDH RNAi in whole-body (act5c-gal4-gs driver; B-D), GS2 RNAi in brain (elaV-gal4-gs driver; E), GS2 RNAi in fat body (S106-gal4-gs; F), and GS2 RNAi in gut (5966-gal4-gs; G) on AL and DR. Vertical lines indicate mean survival. Statistical model is a Cox Proportional Hazards model fitting survival as a function of diet, RNAi, and the interaction between diet and RNAi. Hazard ratios (HR) and P values are specific to the interaction term.



S4 Fig. Multi-omic network for lifespan response for DR metabolites. Gene-metabolite-phenotype network was constructed from linear modeling and GWAS results from DR metabolites that were correlated with lifespan response as measured by rLS. Gene nodes are colored in teal, metabolite nodes are colored in yellow. Gene node size is directly proportional to node degree, while metabolite node size is held constant. An edge exists between a metabolite and lifespan response if the metabolite was significantly correlated with rLS at a 1% FDR level of significance. Grey labels are metabolites correlated with rLS with no significant gene associations. An edge exists between a gene and metabolite and/or lifespan response if it the gene had a score of ≤1E-4.5.



S5 Fig. Diet-dependent survival of RNAi fly strains. Survival of inducible RNAi (+RU486) versus control (-RU486) flies of *selT* RNAi. Vertical lines indicate mean survival. Statistical model is a Cox Proportional Hazards model fitting survival as a function of diet, RNAi, and the interaction between diet and RNAi. Hazard ratios (HR) and P values are specific to the interaction term.

References

1. Kirkwood TB, Feder M, Finch CE, Franceschi C, Globerson A, Klingenberg CP, et al. What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment? Mech Ageing Dev. 2005;126(3):439-43. doi: 10.1016/j.mad.2004.09.008. PubMed PMID: 15664632.
2. Rea SL, Wu D, Cypser JR, Vaupel JW, Johnson TE. A stress-sensitive reporter predicts longevity in isogenic populations of Caenorhabditis elegans. Nat Genet. 2005;37(8):894-8. doi: 10.1038/ng1608. PubMed PMID: 16041374; PubMed Central PMCID: PMCPMC1479894.
3. Fontana L, Partridge L. Promoting health and longevity through diet: from model organisms to humans. Cell. 2015;161(1):106-18. doi: 10.1016/j.cell.2015.02.020. PubMed PMID: 25815989; PubMed Central PMCID: PMCPMC4547605.
4. Nakagawa S, Lagisz M, Hector KL, Spencer HG. Comparative and meta-analytic insights into life extension via dietary restriction. Aging Cell. 2012;11(3):401-9. doi: 10.1111/j.1474-9726.2012.00798.x. PubMed PMID: 22268691.
5. Harper JM, Leathers CW, Austad SN. Does caloric restriction extend life in wild mice? Aging Cell. 2006;5(6):441-9. doi: 10.1111/j.1474-9726.2006.00236.x. PubMed PMID: 17054664; PubMed Central PMCID: PMCPMC2923404.
6. Liao CY, Rikke BA, Johnson TE, Diaz V, Nelson JF. Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. Aging Cell. 2010;9(1):92-5. doi: 10.1111/j.1474-9726.2009.00533.x. PubMed PMID: 19878144; PubMed Central PMCID: PMCPMC3476836.
7. Schleit J, Johnson SC, Bennett CF, Simko M, Trongtham N, Castanza A, et al. Molecular mechanisms underlying genotype-dependent responses to dietary restriction. Aging Cell. 2013;12(6):1050-61. doi: 10.1111/acel.12130. PubMed PMID: 23837470; PubMed Central PMCID: PMCPMC3838465.
8. Dick KB, Ross CR, Yampolsky LY. Genetic variation of dietary restriction and the effects of nutrient-free water and amino acid supplements on lifespan and fecundity of Drosophila. Genet Res (Camb). 2011;93(4):265-73. Epub 2011/07/18. doi: 10.1017/S001667231100019X. PubMed PMID: 21767463.
9. Stanley PD, Ng'oma E, O'Day S, King EG. Genetic Dissection of Nutrition-Induced Plasticity in Insulin/Insulin-Like Growth Factor Signaling and Median Life Span in a. Genetics. 2017;206(2):587-602. doi: 10.1534/genetics.116.197780. PubMed PMID: 28592498; PubMed Central PMCID: PMCPMC5499174.
10. Wilson KA, Beck JN, Nelson CS, Hilsabeck TA, Promislow D, Brem RB, Kapahi P. GWAS for lifespan and decline in climbing ability in flies upon dietary restriction reveal decima as a mediator of insulin-like peptide production. Curr Biol. 2020; in press.
11. Laye MJ, Tran V, Jones DP, Kapahi P, Promislow DE. The effects of age and dietary restriction on the tissue-specific metabolome of Drosophila. Aging Cell. 2015;14(5):797-808. doi: 10.1111/acel.12358. PubMed PMID: 26085309; PubMed Central PMCID: PMCPMC4568967.
12. Davies SK, Bundy JG, Leroi AM. Metabolic Youth in Middle Age: Predicting Aging in Caenorhabditis elegans Using Metabolomics. Journal of Proteome Research. 2015;14(11):4603-9. doi: 10.1021/acs.jproteome.5b00442. PubMed PMID: WOS:000364435100015.
13. Tomas-Loba A, de Jesus BB, Mato JM, Blasco MA. A metabolic signature predicts biological age in mice. Aging Cell. 2013;12(1):93-101. doi: 10.1111/acel.12025. PubMed PMID: WOS:000313745100012.
14. Lopez-Otin C, Galluzzi L, Freije JM, Madeo F, Kroemer G. Metabolic Control of Longevity. Cell. 2016;166(4):802-21. doi: 10.1016/j.cell.2016.07.031. PubMed PMID: 27518560.

15. Ables GP, Brown-Borg HM, Buffenstein R, Church CD, Elshorbagy AK, Gladyshev VN, et al. The first international mini-symposium on methionine restriction and lifespan. Front Genet. 2014;5:122. doi: 10.3389/fgene.2014.00122. PubMed PMID: 24847356; PubMed Central PMCID: PMCPMC4023024.
16. Zimmerman JA, Malloy V, Krajcik R, Orentreich N. Nutritional control of aging. Experimental Gerontology. 2003;38(1-2):47-52. doi: 10.1016/s0531-5565(02)00149-3. PubMed PMID: WOS:000181155100007.
17. Zwighaft Z, Aviram R, Shalev M, Rousso-Noori L, Kraut-Cohen J, Golik M, et al. Circadian Clock Control by Polyamine Levels through a Mechanism that Declines with Age. Cell Metab. 2015;22(5):874-85. doi: 10.1016/j.cmet.2015.09.011. PubMed PMID: 26456331.
18. Chin RM, Fu X, Pai MY, Vergnes L, Hwang H, Deng G, et al. The metabolite alpha-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. Nature. 2014;510(7505):397-401. doi: 10.1038/nature13264. PubMed PMID: 24828042; PubMed Central PMCID: PMCPMC4263271.
19. Mishur RJ, Khan M, Munkacsy E, Sharma L, Bokov A, Beam H, et al. Mitochondrial metabolites extend lifespan. Aging Cell. 2016;15(2):336-48. doi: 10.1111/acel.12439. PubMed PMID: 26729005; PubMed Central PMCID: PMCPMC4783347.
20. Hoffman JM, Soltow QA, Li S, Sidik A, Jones DP, Promislow DE. Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster. Aging Cell. 2014;13(4):596-604. doi: 10.1111/acel.12215. PubMed PMID: 24636523; PubMed Central PMCID: PMCPMC4116462.
21. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53. doi: 10.1038/nature08494. PubMed PMID: 19812666; PubMed Central PMCID: PMCPMC2831613.
22. Field KJ, Lake JA. Environmental metabolomics links genotype to phenotype and predicts genotype abundance in wild plant populations. Physiol Plant. 2011;142(4):352-60. doi: 10.1111/j.1399-3054.2011.01480.x. PubMed PMID: 21496032.
23. Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ. The complex genetic architecture of the metabolome. PLoS Genet. 2010;6(11):e1001198. Epub 2010/11/04. doi: 10.1371/journal.pgen.1001198. PubMed PMID: 21079692; PubMed Central PMCID: PMCPMC2973833.
24. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. 2011;477(7362):54-60. Epub 2011/08/31. doi: 10.1038/nature10354. PubMed PMID: 21886157; PubMed Central PMCID: PMCPMC3832838.
25. Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 2008;4(11):e1000282. Epub 2008/11/28. doi: 10.1371/journal.pgen.1000282. PubMed PMID: 19043545; PubMed Central PMCID: PMCPMC2581785.
26. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet. 2010;42(2):137-41. Epub 2009/12/27. doi: 10.1038/ng.507. PubMed PMID: 20037589; PubMed Central PMCID: PMCPMC3773904.
27. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482(7384):173-8. doi: 10.1038/nature10811. PubMed PMID: 22318601; PubMed Central PMCID: PMCPMC3683990.
28. Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014;24(7):1193-208. doi: 10.1101/gr.171546.113. PubMed PMID: 24714809; PubMed Central PMCID: PMCPMC4079974.

29. Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TF, Partridge L, et al. Longevity GWAS Using the Drosophila Genetic Reference Panel. J Gerontol A Biol Sci Med Sci. 2015;70(12):1470-8. doi: 10.1093/gerona/glv047. PubMed PMID: 25922346; PubMed Central PMCID: PMCPMC4631106.
30. Weber AL, Khan GF, Magwire MM, Tabor CL, Mackay TF, Anholt RR. Genome-wide association analysis of oxidative stress resistance in Drosophila melanogaster. PLoS One. 2012;7(4):e34745. doi: 10.1371/journal.pone.0034745. PubMed PMID: 22496853; PubMed Central PMCID: PMCPMC3319608.
31. Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, et al. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of Drosophila melanogaster. Nat Commun. 2015;6:6312. doi: 10.1038/ncomms7312. PubMed PMID: 25692519; PubMed Central PMCID: PMCPMC4333721.
32. Mackay TFC, Huang W. Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. Wiley Interdisciplinary Reviews-Developmental Biology. 2018;7(1). doi: 10.1002/wdev.289. PubMed PMID: WOS:000419120900001.
33. Sang JH, King RC. Nutritional Requirements of Axenically Cultured D*rosophila Melanogaster* Adults. Journal of Experimental Biology1961. p. 793-809.
34. Osterwalder T, Yoon KS, White BH, Keshishian H. A conditional tissue-specific transgene expression system using inducible GAL4. Proc Natl Acad Sci U S A. 2001;98(22):12596-601. doi: 10.1073/pnas.221303298. PubMed PMID: 11675495; PubMed Central PMCID: PMCPMC60099.
35. Stack JH, Herman PK, Schu PV, Emr SD. A membrane-associated complex containing the Vps15 protein kinase and the Vps34 PI 3-kinase is essential for protein sorting to the yeast lysosome-like vacuole. EMBO J. 1993;12(5):2195-204. PubMed PMID: 8387919; PubMed Central PMCID: PMCPMC413440.
36. Sano H, Nakamura A, Texada MJ, Truman JW, Ishimoto H, Kamikouchi A, et al. The Nutrient-Responsive Hormone CCHamide-2 Controls Growth by Regulating Insulin-like Peptides in the Brain of Drosophila melanogaster. PLoS Genet. 2015;11(5):e1005209. doi: 10.1371/journal.pgen.1005209. PubMed PMID: 26020940; PubMed Central PMCID: PMCPMC4447355.
37. Ren GR, Hauser F, Rewitz KF, Kondo S, Engelbrecht AF, Didriksen AK, et al. CCHamide-2 Is an Orexigenic Brain-Gut Peptide in Drosophila. PLoS One. 2015;10(7):e0133017. doi: 10.1371/journal.pone.0133017. PubMed PMID: 26168160; PubMed Central PMCID: PMCPMC4500396.
38. Stadtman TC. Selenocysteine. Annu Rev Biochem. 1996;65:83-100. doi: 10.1146/annurev.bi.65.070196.000503. PubMed PMID: 8811175.
39. Soultoukis GA, Partridge L. Dietary Protein, Metabolism, and Aging. Annual Review of Biochemistry, Vol 85. 2016;85:5-34. doi: 10.1146/annurev-biochem-060815-014422. PubMed PMID: WOS:000379324700003.
40. De Guzman JM, Ku G, Fahey R, Youm YH, Kass I, Ingram DK, et al. Chronic caloric restriction partially protects against age-related alteration in serum metabolome. Age (Dordr). 2013;35(4):1091-104. Epub 2012/06/04. doi: 10.1007/s11357-012-9430-x. PubMed PMID: 22661299; PubMed Central PMCID: PMCPMC3705111.
41. Pontoizeau C, Mouchiroud L, Molin L, Mergoud-Dit-Lamarche A, Dallière N, Toulhoat P, et al. Metabolomics analysis uncovers that dietary restriction buffers metabolic changes associated with aging in Caenorhabditis elegans. J Proteome Res. 2014;13(6):2910-9. Epub 2014/05/22. doi: 10.1021/pr5000686. PubMed PMID: 24819046; PubMed Central PMCID: PMCPMC4059273.
42. Mouchiroud L, Molin L, Kasturi P, Triba MN, Dumas ME, Wilson MC, et al. Pyruvate imbalance mediates metabolic reprogramming and mimics lifespan extension by dietary restriction in Caenorhabditis elegans. Aging Cell. 2011;10(1):39-54. Epub 2010/11/15. doi: 10.1111/j.1474-9726.2010.00640.x. PubMed PMID: 21040400.

43. Avanesov AS, Ma S, Pierce KA, Yim SH, Lee BC, Clish CB, et al. Age- and diet-associated metabolome remodeling characterizes the aging process driven by damage accumulation. Elife. 2014;3:e02077. Epub 2014/04/29. doi: 10.7554/eLife.02077. PubMed PMID: 24843015; PubMed Central PMCID: PMCPMC4003482.
44. Jové M, Naudí A, Ramírez-Núñez O, Portero-Otín M, Selman C, Withers DJ, et al. Caloric restriction reveals a metabolomic and lipidomic signature in liver of male mice. Aging Cell. 2014;13(5):828-37. Epub 2014/07/23. doi: 10.1111/acel.12241. PubMed PMID: 25052291; PubMed Central PMCID: PMCPMC4331741.
45. Rezzi S, Martin FP, Shanmuganayagam D, Colman RJ, Nicholson JK, Weindruch R. Metabolic shifts due to long-term caloric restriction revealed in nonhuman primates. Exp Gerontol. 2009;44(5):356-62. Epub 2009/03/03. doi: 10.1016/j.exger.2009.02.008. PubMed PMID: 19264119; PubMed Central PMCID: PMCPMC2822382.
46. Jablonski KL, Klawitter J, Chonchol M, Bassett CJ, Racine ML, Seals DR. Effect of dietary sodium restriction on human urinary metabolomic profiles. Clin J Am Soc Nephrol. 2015;10(7):1227-34. Epub 2015/04/21. doi: 10.2215/CJN.11531114. PubMed PMID: 25901092; PubMed Central PMCID: PMCPMC4491302.
47. Collet TH, Sonoyama T, Henning E, Keogh JM, Ingram B, Kelway S, et al. A Metabolomic Signature of Acute Caloric Restriction. J Clin Endocrinol Metab. 2017;102(12):4486-95. doi: 10.1210/jc.2017-01020. PubMed PMID: 29029202; PubMed Central PMCID: PMCPMC5718701.
48. Liao CY, Rikke BA, Johnson TE, Gelfond JA, Diaz V, Nelson JF. Fat maintenance is a predictor of the murine lifespan response to dietary restriction. Aging Cell. 2011;10(4):629-39. doi: 10.1111/j.1474-9726.2011.00702.x. PubMed PMID: 21388497; PubMed Central PMCID: PMCPMC3685291.
49. Cheng LY, Bailey AP, Leevers SJ, Ragan TJ, Driscoll PC, Gould AP. Anaplastic lymphoma kinase spares organ growth during nutrient restriction in Drosophila. Cell. 2011;146(3):435-47. doi: 10.1016/j.cell.2011.06.040. PubMed PMID: 21816278.
50. Grandison RC, Piper MD, Partridge L. Amino-acid imbalance explains extension of lifespan by dietary restriction in Drosophila. Nature. 2009;462(7276):1061-4. doi: 10.1038/nature08619. PubMed PMID: 19956092; PubMed Central PMCID: PMCPMC2798000.
51. Lorenz DR, Cantor CR, Collins JJ. A network biology approach to aging in yeast. Proc Natl Acad Sci U S A. 2009;106(4):1145-50. Epub 2009/01/21. doi: 10.1073/pnas.0812551106. PubMed PMID: 19164565; PubMed Central PMCID: PMCPMC2629491.
52. Wuttke D, Connor R, Vora C, Craig T, Li Y, Wood S, et al. Dissecting the gene network of dietary restriction to identify evolutionarily conserved pathways and new functional genes. PLoS Genet. 2012;8(8):e1002834. Epub 2012/08/09. doi: 10.1371/journal.pgen.1002834. PubMed PMID: 22912585; PubMed Central PMCID: PMCPMC3415404.
53. Ghosh S, Wanders D, Stone KP, Van NT, Cortez CC, Gettys TW. A systems biology analysis of the unique and overlapping transcriptional responses to caloric restriction and dietary methionine restriction in rats. FASEB J. 2014;28(6):2577-90. Epub 2014/02/26. doi: 10.1096/fj.14-249458. PubMed PMID: 24571921; PubMed Central PMCID: PMCPMC4021438.
54. Hou L, Wang D, Chen D, Liu Y, Zhang Y, Cheng H, et al. 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; PubMed Central PMCID: PMCPMC5110149.
55. De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, Mackay TF. Dopa decarboxylase (Ddc) affects variation in Drosophila longevity. Nat Genet. 2003;34(4):429-33. doi: 10.1038/ng1218. PubMed PMID: 12881721.

56. Bjordal M, Arquier N, Kniazeff J, Pin JP, Leopold P. Sensing of amino acids in a dopaminergic circuitry promotes rejection of an incomplete diet in Drosophila. Cell. 2014;156(3):510-21. doi: 10.1016/j.cell.2013.12.024. PubMed PMID: 24485457.
57. Linford NJ, Ro J, Chung BY, Pletcher SD. Gustatory and metabolic perception of nutrient stress in Drosophila. Proc Natl Acad Sci U S A. 2015;112(8):2587-92. doi: 10.1073/pnas.1401501112. PubMed PMID: 25675472; PubMed Central PMCID: PMCPMC4345594.
58. Ro J, Pak G, Malec PA, Lyu Y, Allison DB, Kennedy RT, et al. Serotonin signaling mediates protein valuation and aging. Elife. 2016;5. doi: 10.7554/eLife.16843. PubMed PMID: 27572262; PubMed Central PMCID: PMCPMC5005037.
59. Dus M, Lai JS, Gunapala KM, Min S, Tayler TD, Hergarden AC, et al. Nutrient Sensor in the Brain Directs the Action of the Brain-Gut Axis in Drosophila. Neuron. 2015;87(1):139-51. doi: 10.1016/j.neuron.2015.05.032. PubMed PMID: 26074004; PubMed Central PMCID: PMCPMC4697866.
60. Pool AH, Kvello P, Mann K, Cheung SK, Gordon MD, Wang L, et al. Four GABAergic interneurons impose feeding restraint in Drosophila. Neuron. 2014;83(1):164-77. doi: 10.1016/j.neuron.2014.05.006. PubMed PMID: 24991960; PubMed Central PMCID: PMCPMC4092013.
61. Caggese C, Barsanti P, Viggiano L, Bozzetti MP, Caizzi R. Genetic, Molecular and Developmental Analysis of the Glutamine-Synthetase Isozymes of Drosophila-Melanogaster. Genetica. 1994;94(2-3):275-81. doi: Doi 10.1007/Bf01443441. PubMed PMID: WOS:A1994QH20800020.
62. Jayakumar S, Richhariya S, Reddy OV, Texada MJ, Hasan G. Drosophila larval to pupal switch under nutrient stress requires IP3R/Ca(2+) signalling in glutamatergic interneurons. Elife. 2016;5. doi: 10.7554/eLife.17495. PubMed PMID: 27494275; PubMed Central PMCID: PMCPMC4993588.
63. Post S, Karashchuk G, Wade JD, Sajid W, De Meyts P, Tatar M. Insulin-Like Peptides DILP2 and DILP5 Differentially Stimulate Cell Signaling and Glycogen Phosphorylase to Regulate Longevity. Front Endocrinol (Lausanne). 2018;9:245. Epub 2018/05/28. doi: 10.3389/fendo.2018.00245. PubMed PMID: 29892262; PubMed Central PMCID: PMCPMC5985746.
64. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics. 2011;12:357. Epub 2011/08/31. doi: 10.1186/1471-2105-12-357. PubMed PMID: 21880147; PubMed Central PMCID: PMCPMC3179972.
65. Ramos-Alvarez I, Martin-Duce A, Moreno-Villegas Z, Sanz R, Aparicio C, Portal-Nunez S, et al. Bombesin receptor subtype-3 (BRS-3), a novel candidate as therapeutic molecular target in obesity and diabetes. Molecular and Cellular Endocrinology. 2013;367(1-2):109-15. doi: 10.1016/j.mce.2012.12.025. PubMed PMID: WOS:000315611600012.
66. Durham MF, Magwire MM, Stone EA, Leips J. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. Nat Commun. 2014;5:4338. Epub 2014/07/08. doi: 10.1038/ncomms5338. PubMed PMID: 25000897.
67. Harvanek ZM, Lyu Y, Gendron CM, Johnson JC, Kondo S, Promislow DEL, et al. Perceptive costs of reproduction drive ageing and physiology in male Drosophila. Nat Ecol Evol. 2017;1(6):152. doi: 10.1038/s41559-017-0152. PubMed PMID: 28812624; PubMed Central PMCID: PMCPMC5657004.
68. Katewa SD, Akagi K, Bose N, Rakshit K, Camarella T, Zheng X, et al. Peripheral Circadian Clocks Mediate Dietary Restriction-Dependent Changes in Lifespan and Fat Metabolism in Drosophila. Cell Metab. 2016;23(1):143-54. doi: 10.1016/j.cmet.2015.10.014. PubMed PMID: 26626459; PubMed Central PMCID: PMCPMC4715572.

69. Nelson CS, Beck JN, Wilson KA, Pilcher ER, Kapahi P, Brem RB. Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. BMC Genomics. 2016;17(1):867. doi: 10.1186/s12864-016-3137-9. PubMed PMID: 27809764; PubMed Central PMCID: PMCPMC5095962.
70. Zid BM, Rogers AN, Katewa SD, Vargas MA, Kolipinski MC, Lu TA, et al. 4E-BP extends lifespan upon dietary restriction by enhancing mitochondrial activity in Drosophila. Cell. 2009;139(1):149-60. doi: 10.1016/j.cell.2009.07.034. PubMed PMID: 19804760; PubMed Central PMCID: PMCPMC2759400.
71. Chapman T, Partridge L. Female fitness in Drosophila melanogaster: An interaction between the effect of nutrition and of encounter rate with males. Proceedings of the Royal Society B-Biological Sciences. 1996;263(1371):755-9. doi: 10.1098/rspb.1996.0113. PubMed PMID: WOS:A1996UW13400013.
72. Kapahi P, Zid BM, Harper T, Koslover D, Sapin V, Benzer S. Regulation of lifespan in Drosophila by modulation of genes in the TOR signaling pathway. Current Biology. 2004;14(10):885-90. doi: 10.1016/j.cub.2004.03.059. PubMed PMID: WOS:000221681600025.
73. Partridge L, Piper MD, Mair W. Dietary restriction in Drosophila. Mech Ageing Dev. 2005;126(9):938-50. doi: 10.1016/j.mad.2005.03.023. PubMed PMID: 15935441.
74. Lang S, Hilsabeck TA, Wilson KA, Sharma A, Bose N, Brackman DJ, et al. A conserved role of the insulin-like signaling pathway in diet-dependent uric acid pathologies in Drosophila melanogaster. PLoS Genet. 2019;15(8):e1008318. Epub 2019/08/15. doi: 10.1371/journal.pgen.1008318. PubMed PMID: 31415568; PubMed Central PMCID: PMCPMC6695094.
75. Sperber H, Mathieu J, Wang Y, Ferreccio A, Hesson J, Xu Z, et al. The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. Nat Cell Biol. 2015;17(12):1523-35. doi: 10.1038/ncb3264. PubMed PMID: 26571212; PubMed Central PMCID: PMCPMC4662931.
76. Du J, Rountree A, Cleghorn WM, Contreras L, Lindsay KJ, Sadilek M, et al. Phototransduction Influences Metabolic Flux and Nucleotide Metabolism in Mouse Retina. J Biol Chem. 2016;291(9):4698-710. doi: 10.1074/jbc.M115.698985. PubMed PMID: 26677218; PubMed Central PMCID: PMCPMC4813492.
77. Chiao YA, Kolwicz SC, Basisty N, Gagnidze A, Zhang J, Gu H, et al. Rapamycin transiently induces mitochondrial remodeling to reprogram energy metabolism in old hearts. Aging (Albany NY). 2016;8(2):314-27. doi: 10.18632/aging.100881. PubMed PMID: 26872208; PubMed Central PMCID: PMCPMC4789585.
78. R Core Team. R: A language and environment for statistical computing. In: Computing RFfS, editor. Vienna, Austria2014.
79. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological. 1995;57(1):289-300. PubMed PMID: WOS:A1995QE45300017.
80. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: impute: Imputation for microarray data.: R package version 1.56.0.; 2018.
81. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMCPMC1950838.
82. Gao J, Tarcea VG, Karnovsky A, Mirel BR, Weymouth TE, Beecher CW, et al. Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. Bioinformatics. 2010;26(7):971-3. doi: 10.1093/bioinformatics/btq048. PubMed PMID: 20139469; PubMed Central PMCID: PMCPMC2844990.

Chapter 3: Multiple morbidities in companion dogs - A novel model for investigating age-related disease

AUTHORS: Kelly Jin[1], Jessica M. Hoffman[2], Kate E. Creevy[3], Dan G. O'Neill[4], Daniel Promislow[1,5]

1. Department of Pathology, University of Washington School of Medicine, Seattle, WA, USA
2. Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA
3. College of Veterinary Medicine, Texas A&M University, College Station, TX, USA
4. Veterinary Epidemiology, Economics and Public Health, The Royal Veterinary College, Hatfield, UK
5. Department of Biology, University of Washington, Seattle, WA, USA

Abstract

The proportion of men and women surviving over 65 years has been steadily increasing over the last century. In their later years, many of these individuals are afflicted with multiple chronic conditions, placing increasing pressure on healthcare systems. The accumulation of multiple health problems with advanced age is well documented, yet the causes are poorly understood. Animal models have long been employed in attempts to elucidate these complex mechanisms with limited success. Recently, the domestic dog has been proposed as a promising model of human aging for several reasons. Mean lifespan shows two-fold variation across dog breeds. In addition, dogs closely share the environments of their owners and substantial veterinary resources are dedicated to comprehensive diagnosis of conditions in dogs. However, while dogs are therefore useful for studying multimorbidity, little is known about how aging influences the accumulation of multiple concurrent disease conditions across dog breeds. The current study examines how age, body weight, and breed contribute to variation in multimorbidity in over 2000 companion dogs visiting private veterinary clinics in England. In common with humans, we find that the number of diagnoses increases significantly with age in dogs. However, we find no significant weight or breed effects on morbidity number. This surprising result reveals that while breeds may vary in their average longevity and causes of death, their age-related trajectories of morbidities differ little, suggesting that age of onset of disease may be the source of variation in lifespan across breeds. Future studies with increased sample sizes and longitudinal monitoring may help us discern more breed-specific patterns in morbidity. Overall, the large increase in multimorbidity seen with age in dogs mirrors that seen in humans and lends even more credence to the value of companion dogs as models for human morbidity and mortality.

Introduction

Global human populations are aging rapidly, with 17% of the population – a predicted 1.6 billion people – expected to be over the age of 65 by the year 2050 (1). Among these older individuals, a large percentage will be afflicted with multiple morbidities. Incidence of many of these morbidities, including but not limited to diabetes mellitus, arthritis, hypertension, osteoporosis, neurodegenerative disease and various forms of cancer, increases with age. Furthermore, these diseases often present with each other as multiple (two or more) morbidities (1, 2). However, it is currently unknown if accumulation of multiple morbidities is a consequence or a cause of biological aging and shortened longevity. The significant increase in morbidity count with age suggests that multimorbidity is a consequence of aging; however, increases in

multiple chronic conditions are associated with shorter longevity (3), indicating multimorbidity as a potential cause of biological aging. Close examination of why certain diseases present together and why others do not, as well as the trajectories of disease accumulation throughout age, can offer critical insights into the biology of aging. Furthermore, multiple morbidities are often complex to manage in the clinic, and undeniably result in a heavy burden on social and healthcare infrastructures currently in place (4, 5). For these reasons, the phenomenon of age-related multimorbidities has long been recognized as a key area for research in the field of aging (6).

Experimental laboratory animal models are commonly used to study singular, age-related diseases. However, studies exploring how different age-related conditions present and interact with one another are difficult to perform in experimental models. As such, the genetic and physiological mechanisms underlying different multimorbidity patterns and trends have yet to be adequately addressed in current animal models. One reason for this is that most of the popular aging models (flies, yeast, worms) are too evolutionarily distant from humans to model all aspects of individual chronic diseases, let alone multiple ones. Furthermore, although many murine models do often present with multiple morbid conditions (e.g. 7), aging out large cohorts of genetically diverse mouse populations is costly, and there are no established systems known to us for diagnosing, treating, and preventing chronic conditions in mice. And finally, the environmental factors that influence aging in laboratory animals are likely to be wildly different from the environmental effects that humans experience.

In the current study, we discuss the domestic dog, Canis lupus familiaris, as a potentially powerful model for studying the age-related accumulation of morbid conditions. Dogs offer a host of advantages for studying disease, perhaps the most pronounced of which is their unique breed-based population structure. The domestic dog, which is the most phenotypically diverse mammal species on earth (8), exhibits high phenotypic heterogeneity among breeds, in addition to phenotypic and genetic homogeneity within breeds (9). In addition, breeds of dogs tend to present with distinct patterns of diseases and causes of death (10). Dogs and humans also share many diseases that present in old age, and the quality of medical care for dogs is second only to that for humans. Furthermore, companion dogs share our environment and its associated disease risk factors in a way that can never be replicated in a laboratory. These facts make the companion dog an exciting model in which to study complex diseases.

Here, we present multimorbidity patterns across an extensive database of electronic veterinary records in dogs with the aims of 1) understanding how multimorbidity trends vary with age, body weight, and breed within a population, and 2) highlighting the translational potential of the dog

model in future studies of multimorbidity and aging. We are interested in body weight because of its large influence on lifespan – small breeds tend to live longer than larger breeds (11, 12). Thus, we hypothesize that both dog body weight and age will have significant effects on multimorbidity patterns, specifically that the number of morbidities will increase with increasing observed age and body weight of dogs presenting in an extensive database of electronic medical records from veterinary clinics in the UK.

Results

Distribution of multiple morbidities

The original database consisted of 3225 adult dogs (aged over one year at their final record), of which 2586 (80.2%) had both body weight and age data, and were thus included in the analysis. The distribution of morbidity scores across all dogs in the dataset more closely follow a negative-binomial distribution rather than a Poisson distribution (AICnegbinom < AICpois; Figure 1a). Morbidity scores across dogs separated by weight class follow very similar distributions (Figure 1c-e; Mann-Whitney test results: $P > 0.05$ for all pairwise comparisons), suggesting there are very few differences between differently sized dogs in terms of the number of morbidities accumulated over the lifetime.

The distributions of the ages of the dogs at the times of their visits are shown in Table 1 and Figure 1b, f-h. In contrast to the size class-specific morbidity score distributions, the shape of age-distributions varies noticeably between weight classes. Amongst small dogs, the greatest number of dogs seen by the veterinarian were one-year-old, while in medium and large dogs, that age increases (3, and 5 years-old, respectively).

Age, but not body weight or breed, affects multiple morbidities

In order to determine whether age and/or body weight are significantly associated with multimorbidity score, we ran a GLM of morbidity score as a function of age, weight, and their interaction. As expected, there was a highly significant positive effect of age on number of diagnoses ($P = 2.34E-12$, Table 2, Figure 2). This trend represents a highly significant effect of age on morbidity score across all dogs. However, failing to support the initial study hypothesis, no significant association was detected between body weight ($P = 0.777$) nor age-by-weight interactions ($P = 0.278$) on multimorbidity score.

We then broke down the weight classes into individual breeds (Figure 3). Our breed-based analysis surveyed 12 breeds comprising 1278 dogs. In medium and large breeds, age was significantly associated with morbidity score ($P<7.31E-04$, Table 3), consistent with trends observed in the entire dataset (Table 2). Specific breeds did not vary with respect to morbidity score in any weight class. Interestingly, no significant effects of breed, age, or their interaction were found on morbidity score in small dogs.

Methods

Data processing

Multimorbidity data on a large cohort of dogs were acquired from the VetCompass database of the Royal Veterinary College in the United Kingdom (13). Data were included from dogs attending primary practice veterinary clinics in England over the course of 3.5 years, from 2009-2013. Some dogs visited their veterinarians several times over the period while others had a single visit. All defining symptoms/diagnoses recorded for each dog during the study period were extracted, and we considered each defining symptom/diagnosis to be a morbidity for that dog. Full details of the methods used for developing the data set used for this study have been previously published (13). We removed descriptors or procedures that were not related to the health of the animal or were determined by one of the authors (KEC) to be too vague (eg. “puppy vet check”, “ID chip insertion”, “nail clip”). The total number of diagnoses for each dog was summed and treated as the morbidity score for each dog. Thus, all included diagnoses were weighted equally. While some dogs had more than one included veterinary record across the time period of data collection, only a single age was recorded for each dog. This age was assigned by subtracting the dog’s birthdate from the last day it visited the veterinary clinic. Dogs under one year of age were removed from the analysis due to their constantly changing body weights. In addition, only those dogs with both body weight and age data recorded were included in the analysis. Dogs were separated into small (<10 kg), medium (10-20 kg), or large (20+ kg) body weight classes. Weight class cutoffs were based on a modified version of cutoffs previously assigned by a veterinarian in another study (14). We then chose those breeds with the largest sample sizes in the database (i.e., breeds represented by at least 50 individuals), and grouped them by breed standard into their respective weight classes: small (Jack Russell terrier, West Highland white terrier, Shih-Tzu, and Cavalier King Charles spaniel), medium

(border collie, Staffordshire bull terrier, Cocker spaniel, and English Springer spaniel), and large (Labrador retriever, golden retriever, and Rottweiler). Breeds were grouped into weight classes based on American Kennel Club average weights (15). Mixed breeds, which represented the largest sized sample population, were excluded from this analysis, allowing us to determine the effects of genetic homogeneity on multimorbidity number. Ethical approval for the study was granted by the Royal Veterinary College Ethics and Welfare Committee (reference number 2015/1369).

Statistical Models

Due to the discrete and over-dispersed (i.e. non-Gaussian) nature of the morbidity data, we first used a generalized linear model (GLM) with a negative binomial distribution to discover the effects of age, body weight, and their interaction on the morbidity scores, treating all factors as fixed effects. Next, we looked amongst individual breeds within each different weight class to see if there was significant breed variation for age-related changes in multimorbidity within similar sized dogs. Similar to our weight analysis, we ran a GLM with a negative binomial distribution for each individual weight class for the effects of breed, age, and their interaction on morbidity score. All statistical analyses were completed in the program R (16). We used a model selection approach to identify the best distribution of the data based on Akaike's Information Criterion (AIC) (17). For each candidate model, an AIC was computed using the R package fitdistrplus (18), and the model with the lowest AIC was selected as the best supported distribution. Regressions were performed using the R package MASS (19).

Discussion

We began our investigation by exploring the distribution of morbidity scores across all dogs in the VetCompass dataset. We found that all morbidity scores follow a non-Poisson distribution (Figure 1a, c-e). The Poisson distribution is used to describe events that are randomly scattered over space or time. Departures from Poisson, therefore, are often interpreted as departures from randomness and neutrality (20). Thus, much like how departures from Hardy Weinberg equilibrium suggest the presence of forces acting upon genetic variation in a population (21), the deviations from Poisson in our data imply that accumulation of multi-morbid conditions is not a random process and that other forces may be significantly influencing multimorbidity rates in dogs. At present, we are unable to identify what these external forces

may be, but the question is an exciting one, that further underscores the research potential of the canine model for understanding multimorbidity.

We also found that the shapes of the morbidity distributions did not differ substantially between body weight classes of dog. This observation is intriguing, as it is well documented that large dogs have shorter lifespans than small dogs (10). Therefore, one might expect that larger breeds may have a different distribution of disease accumulation than smaller breeds. However, this does not appear to be the case, as all size classes of dogs have similar morbidity score distributions.

Of many potential factors that might influence this unexpected observation, one is the distribution of ages of presenting dogs. Despite similar distributions of morbidity scores, we discovered differences among body weight groups of dogs in the most frequent age of presentation to the veterinary practices (Figure 1 f-h). This interesting observation indicates that there may be external factors influencing the age at which dogs are most likely to present to the clinic, and those factors may vary based on size and/or breed of the animal. We need more data on the timing of disease manifestation between different size classes, and by definition, different breeds of dogs. In addition to the breed-specific differences in the age at onset of different diseases, extrinsic forces may also affect these patterns. For example, recent increases in popularity of certain small dog breeds (e.g. rising ownership of Pugs in England (22)) would result in an observed distribution skewed towards younger small dogs. In a growing population, younger dogs would make up a larger proportion of the sample population. Given the recent growth in this small, relatively long-lived breed, this demographic phenomenon could skew morbidity counts towards lower levels.

We found that across all dogs, multimorbidities increase significantly with age, but not with size (Table 2, Figure 2). Overall this trend holds within breeds of different size classes as well. We know that frequencies of specific diagnoses differ among size classes of dogs (10, 13). However, within our dataset, the total number of morbidities for each dog is dependent almost exclusively on age with no detectable effect of weight or breed. Different breeds are known to experience different morbidities, as well as to have different causes of death (10, 23). In this light, we would have expected to find that certain breeds were more prone to multiple morbidities, but our study failed to identify this association. This suggests that there is little effect of breed on the overall number of multiple morbidities, only the types of morbidities. However, replicate and controlled studies are needed to test this hypothesis more directly.
Additionally, the choice to weight each diagnosis equally may have allowed the morbidity scores to appear the same across breeds, while the actual burden of disease comprising each score

differed. Existing morbidity indices for humans often weight specific measures by their potential effect on overall function (24, 25). Future clinical studies on aging in dogs could benefit greatly from developing such weighted measures.

Our breed-based analysis showed similar patterns as seen in the dataset overall, with one exception. Small dog breeds failed to show a significant effect of age on morbidity score, suggesting that compared to larger dogs, small dogs may differ physiologically in a way that attenuates the effect of age on multimorbidity. This hypothesis is consistent with the observation that smaller dog breeds tend to live longer than larger breeds, as well as the idea that smaller dogs are may be aging more slowly (11, 12).

Overall, our results highlight a novel and valuable application of the companion dog as a model for age-related multimorbidities in humans. We believe the key to its value, in part, lies in the dog's ability to model non-biological variables in addition to biological ones. Biological variables, including sex, neutering, body weight and age, are important variables that can contribute to the development of chronic disease, but they are not the only important variables. Non-biological factors include, but are not limited to, environmental, behavioral, social, and economic components and have large and complex effects on both canine and human health. For example, several previous papers employing the VetCompass database have highlighted dramatic financial and social effects from pet insurance on disease diagnosis and survival rates. Compared with non-insured dogs, dogs that are insured are four times more likely to be diagnosed with hyperadrenocorticism (26), four times more likely to be diagnosed with cranial cruciate disease (27), and four times more likely to be diagnosed with mast cell tumor (28). Non-insured diabetic dogs have 1.7 times the hazard of death compared with insured dogs (29). Consequently, we believe that with the increase in available data, and improvements in data science methods, the scientific and medical research community can now incorporate these non-biological factors when investigating complex biological processes such as aging. Given that dogs and humans share many environmental risk factors such as the same air and water pollutants, similar levels of exercise by exercising together, and similar potential economic constraints, such as access to medical care, the potential of the dog both as a sentinel (30) and translational model is high.

Caveats

While this is the largest multimorbidity analysis in the dog to date, our analysis was not without its limitations. Firstly, our sample size of just over 2500 dogs is still relatively small, especially once divided by breed. To better understand age-related changes in multimorbidities,

we need studies with many more dogs at various ages in each breed, and a better understanding of the distribution of the complete at-risk population in terms of age, size and breed. Furthermore, the relatively small sample size of this dataset did not allow us to look at the effects of sex and sterilization status, the latter of which significantly effects disease risk and mortality (14, 31).

Second, as described above, we treated every diagnosis the dog had been given with equal weight, other than those removed for vagueness and/or irrelevance as described in the Methods section. As such, some morbidities may only have minor effects on the physiology and health of the dog. Many studies in humans have used similarly unweighted disease counts as measures of multimorbidity (25), though inclusion criteria for specific diseases can vary. Our current unweighted approach may have acted to limit our ability to detect the true differences of accumulated morbidity between small and large dogs. That is to say, while the distributions of multiple morbidities we present are similar in small and large dogs (Figure 1c, e), conditions in large dogs could be more detrimental to overall health than those in small dogs.

Third, owners may choose to euthanize dogs before they develop their “maximum” morbidity levels. The original study from which these data were derived reported that 88.9% of the recorded deaths were by euthanasia (13). This could have a strong confounding effect on certain breeds; since we know that diverse breeds differ in disease profile, some of the more debilitating and painful diseases that lead to earlier euthanasia might be over-represented among certain breeds (10, 32).

Fourth, the nature of this dataset does not allow us to know if these diagnoses were chronic within each dog or were only present at one specific time point. Many human studies focus specifically on multiple chronic conditions (e.g. 2), which we were not able to ascertain from this population. An ideal canine study of chronic multimorbidities would follow dogs throughout life to discover the degree to which different conditions persist throughout the lifespan of the dog.

Finally, other unexamined factors likely affect the observed distribution (which might differ from the true distribution) of morbidities. For example, the probability that an owner brings a dog to clinic might be a function of the animal’s multimorbidity score, leading to ascertainment bias. Higher multimorbid individuals may be more likely to be brought to the clinic; additionally, these animals may be shorter-lived overall, leading to a right truncation of the data due to high-morbidity individuals existing in that state for a shorter time period compared to those individuals with fewer morbidities. Furthermore, individual morbidity scores can move backwards in the real world with the cure of some morbidities due to pharmacological interventions or lifestyle

changes, which would not be reflected in our cross-sectional data such. Hidden variables such as these can only be revealed and accounted for with long term longitudinal studies, an approach that will be key to future advancement in our understanding of the causes and consequences of aging and age-related disease.

Conclusions

Here, we have presented the largest multimorbidity analysis in the companion dog to date. We find little variation in morbidity scores across dogs of varying body weight groups despite previously published dramatic variation in causes and ages of death across weight classes and individual breeds (10, 33). Furthermore, most of the variation in morbidity score patterns can be explained by age rather than the body size or breed of the dog, suggesting that age is a major, if not the most influential, risk factor for accumulation of disease in dogs, as it is in humans. Although much more investigation is needed to identify mechanisms of multimorbidity, this early study has revealed interesting insights into the architecture of morbidity across dog breeds. This study further shows that the companion dog can be an excellent model for studying disease variation and age-related decline in health.

Acknowledgements

This study was supported in part by NIH R24 AG044284 (DEP) and NIH T32 GM095421 (KJ). We are grateful to The Kennel Club, The Kennel Club Charitable Trust and Dogs Trust for supporting VetCompass in the UK.

Tables

Table 1. Summary statistics of distributions of ages (in years) of dogs in the VetCompass data

	Mean	Minimum	1st Quartile	Median	3rd Quartile	Max	Count
Small dogs	6.52	1.01	2.87	5.45	9.6	18.48	766
Medium dogs	6.95	1.01	3.36	6.11	10.16	18.74	692
Large dogs	7.07	1	3.84	6.53	9.95	17.52	1128
All dogs	6.87	1	3.39	6.08	9.92	18.74	2586

Table 2. Results of linear model on age, weight, and age-by-weight interaction

Variable	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	0.463	6.12E-02	7.575	3.59E-14
Age	5.15E-02	7.35E-03	7.013	2.34E-12
Weight	7.87E-04	2.78E-03	0.283	0.777
Age*Weight	3.67E-04	3.39E-04	1.084	0.278

Table 3. Results of analysis of variance for breed and age within different weight classes.

Small Breeds				
Variable	Estimate	Std. Error	t-value	p-value
Intercept	0.762	0.217	3.504	4.58E-04
Age	8.96E-05	6.54E-05	1.369	0.171
Cavalier King Charles Spaniel	0.182	0.297	0.613	0.540
Jack Russell	-0.447	0.257	-1.738	0.082
West Highland White Terrier	-0.510	0.319	-1.599	0.110
Yorkshire Terrier	-0.321	0.285	-1.125	0.261
Age*Cavalier King Charles Spaniel	4.80E-05	9.64E-05	0.498	0.618
Age*Jack Russell	7.41E-05	7.65E-05	0.969	0.333
Age*West Highland White Terrier	1.45E-04	9.33E-05	1.563	0.118
Age*Yorkshire terrier	4.96E-05	8.63E-05	0.574	0.566

Medium Breeds				
Variable	Estimate	Std. Error	t-value	p-value
Intercept	0.302	0.219	1.379	0.168
Age	2.02E-04	6.00E-05	3.378	7.31E-04
Cocker Spaniel	0.279	0.280	0.996	0.319
English Springer Spaniel	0.303	0.283	1.071	0.284
Staffordshire Bull Terrier	0.184	0.252	0.732	0.464
Age*Cocker Spaniel	-4.80E-05	8.62E-05	-0.557	0.578
Age*English Springer Spaniel	4.82E-07	8.16E-05	0.006	0.995
Age*Staffordshire Bull Terrier	-9.93E-05	7.59E-05	-1.307	0.191

Large Breeds				
Variable	Estimate	Std. Error	t-value	p-value
Intercept	2.53E-02	0.280	0.090	0.928
Age	3.02E-04	8.14E-05	3.716	2.02E-04
Labrador Retriever	0.507	0.303	1.670	0.095
German Shepherd	-4.52E-02	0.344	-0.131	0.895
Age*Labrador Retriever	-1.65E-04	9.00E-05	-1.841	0.066
Age*German Shepherd	-5.32E-06	1.06E-04	-0.050	0.960

Figures

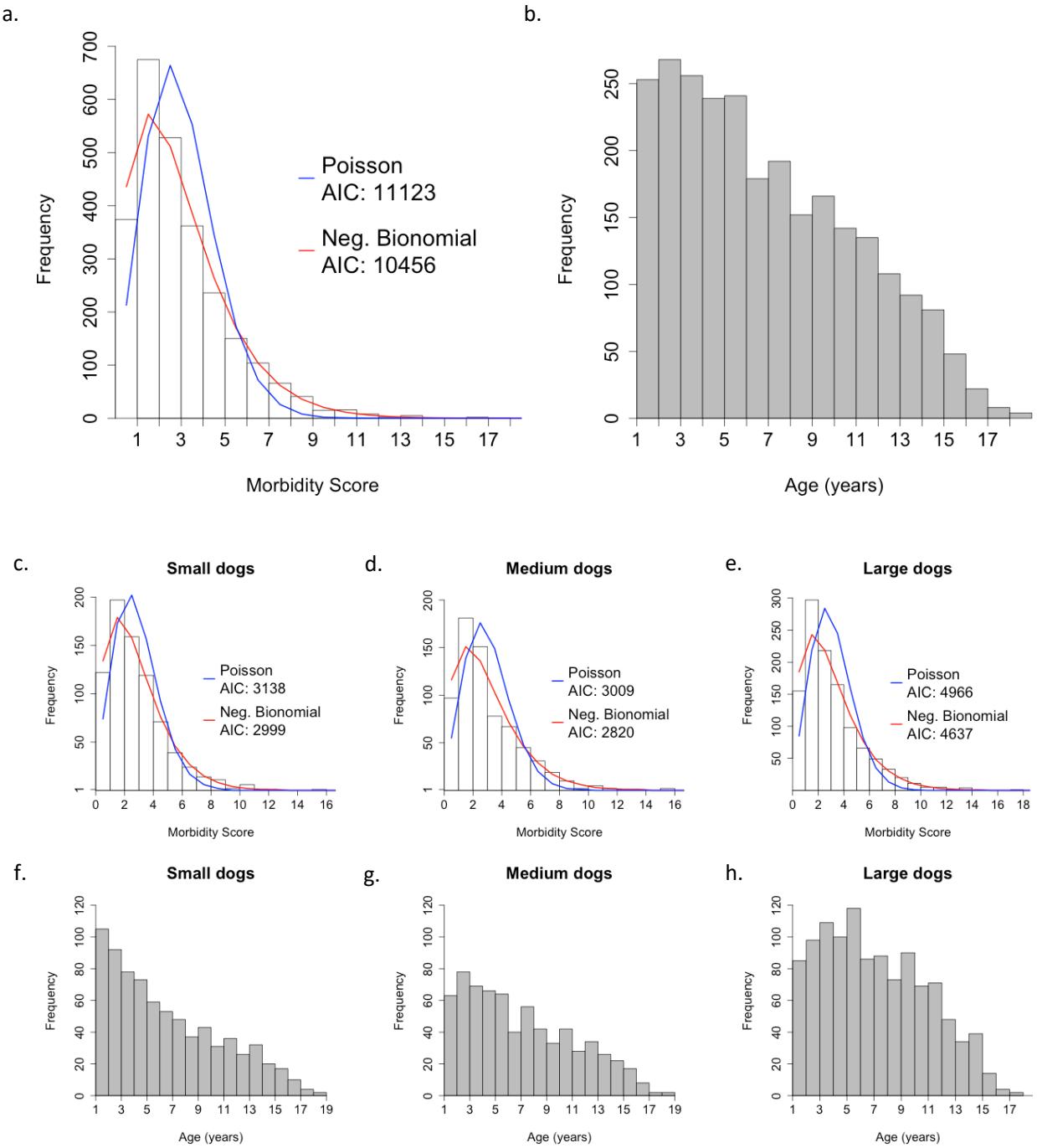


Figure 1. Morbidity scores and ages of dogs in the VetCompass dataset. Shown are distributions of morbidity scores (a) and age of dog at veterinary visit (b) across all dogs in the datasets. Morbidity score (c-e) and age (f-h) distributions of dogs are also visualized by body weight class as labeled. Total number of animals in each weight class are as follows: n_{small} = 766, n_{medium} = 692, n_{large} = 1128, n_{total} = 2586.

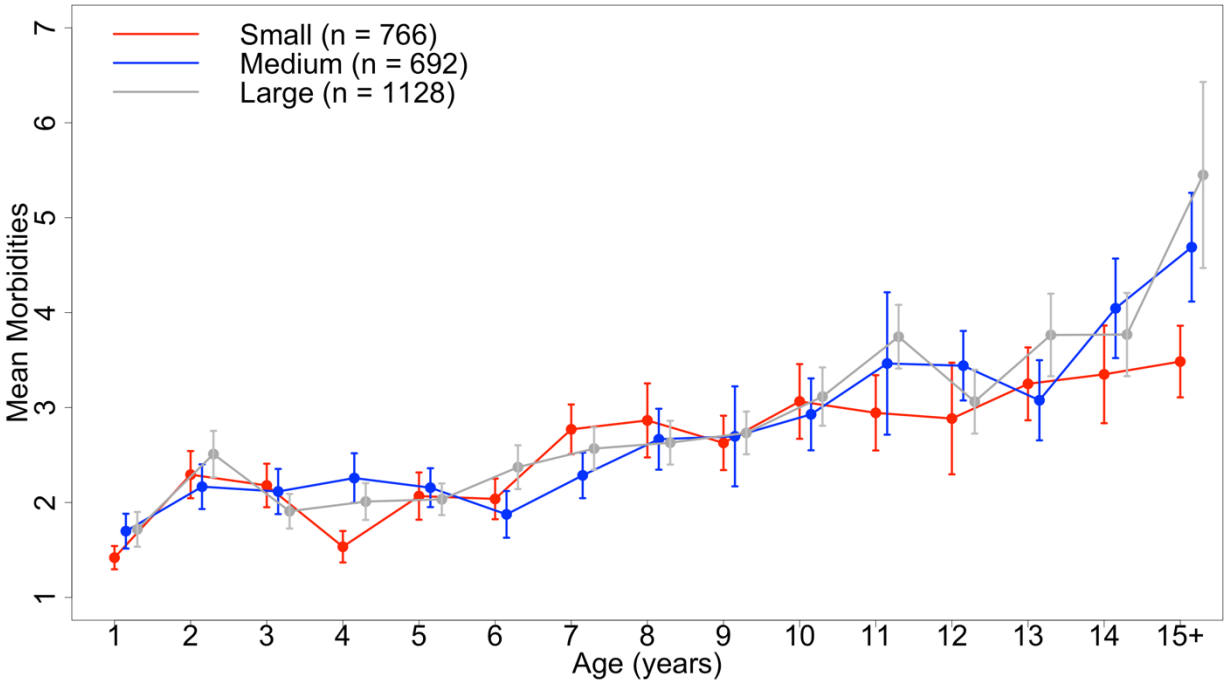


Figure 2. Age-related changes in morbidity scores for dogs by body weight class. Error bars indicate +/- one standard error. Due to small sample size, all dogs within a weight class with age ≥ 15 years were grouped together for visualization purposes.

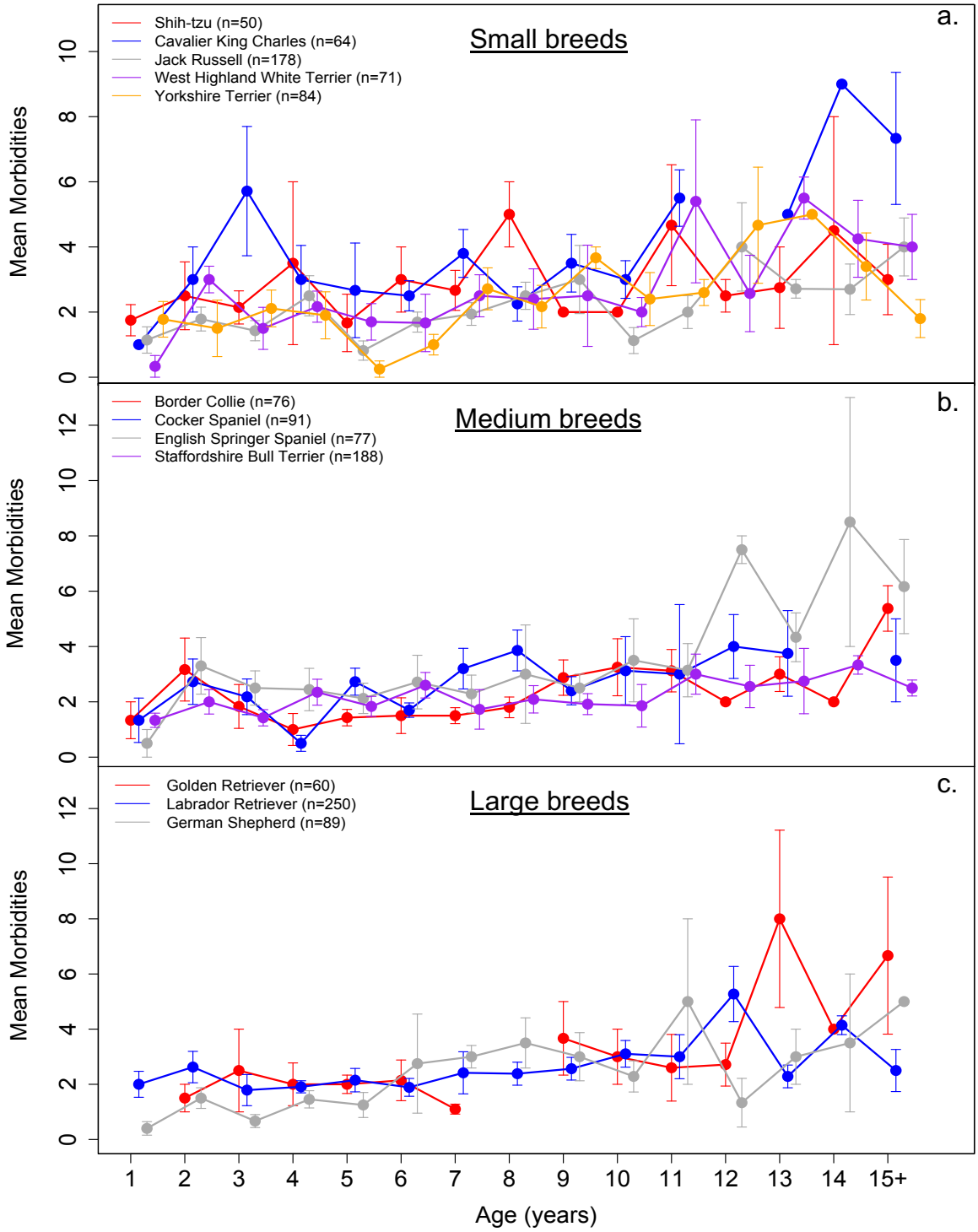


Figure 3. Age-related changes in morbidity scores for dogs by breed. Breeds are grouped by small (a), medium (b), and large (c) weight class. Error bars indicate +/- one standard error. Due to small sample size, all dogs within a breed with age ≥ 15 years were grouped together for visualization purposes.

References

1. He W, Goodkind D, Kowal P. An Aging World: 2015. In: Bureau USC, editor. Washington, DC: U.S. Government Publishing Office; 2016. p. P95.
2. Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. Health Qual Life Outcomes. 2009;7:82. doi: 10.1186/1477-7525-7-82.
3. DuGoff EH, Canudas-Romo V, Buttorff C, Leff B, Anderson GF. Multiple Chronic Conditions and Life Expectancy A Life Table Analysis. Med Care. 2014;52(8):688-94.
4. Bock JO, Konig HH, Brenner H, Haefeli WE, Quinzler R, Matschinger H, et al. Associations of frailty with health care costs - results of the ESTHER cohort study. Bmc Health Serv Res. 2016;16. doi: ARTN 128; 10.1186/s12913-016-1360-3.
5. Picco L, Achilla E, Abdin E, Chong SA, Vaingankar JA, McCrone P, et al. Economic burden of multimorbidity among older adults: impact on healthcare and societal costs. Bmc Health Serv Res. 2016;16. doi: ARTN 173; 10.1186/s12913-016-1421-7.
6. Yancik R, Ershler W, Satariano W, Hazzard W, Cohen HJ, Ferrucci L. Report of the national institute on aging task force on comorbidity. J Gerontol A Biol Sci Med Sci. 2007;62(3):275-80.
7. Treuting PM, Linford NJ, Knoblaugh SE, Emond MJ, Morton JF, Martin GM, et al. Reduction of age-associated pathology in old mice by overexpression of catalase in mitochondria. J Gerontol A Biol Sci Med Sci. 2008;63(8):813-22.
8. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, et al. A single IGF1 allele is a major determinant of small size in dogs. Science. 2007;316(5821):112-5. doi: 10.1126/science.1137045.
9. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, et al. Genetic structure of the purebred domestic dog. Science. 2004;304(5674):1160-4. doi: DOI 10.1126/science.1097406.
10. Fleming JM, Creevy KE, Promislow DE. Mortality in north american dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. J Vet Intern Med. 2011;25(2):187-98. doi: 10.1111/j.1939-1676.2011.0695.x.
11. Kraus C, Pavard S, Promislow DE. The size-life span trade-off decomposed: why large dogs die young. Am Nat. 2013;181(4):492-505. doi: 10.1086/669665.
12. Galis F, Van Der Sluijs I, Van Dooren TJM, Metz JAJ, Nussbaumer M. Do large dogs die young? J Exp Zool Part B. 2007;308b(2):119-26. doi: 10.1002/jez.b.21116.
13. O'Neill DG, Church DB, McGreevy PD, Thomson PC, Brodbelt DC. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. PLoS One. 2014;9(3):e90501. doi: 10.1371/journal.pone.0090501.
14. Hoffman JM, Creevy KE, Promislow DEL. Reproductive Capability Is Associated with Lifespan and Cause of Death in Companion Dogs. Plos One. 2013;8(4). doi: ARTN e61082; 10.1371/journal.pone.0061082.
15. AKC Dog Breeds: American Kennel Club; (cited 2016). Available from: http://www.akc.org/dog-breeds/.
16. R Core Team. R: A language and environment for statistical computing. In: Computing RFfS, editor. Vienna, Austria2014.
17. Burnham KP, Anderson DR, Burnham KP. Model selection and multimodel inference : a practical information-theoretic approach. 2nd ed. New York: Springer; 2002. xxvi, 488 p. p.
18. Delignette-Muller ML, Dutang C. fitdistrplus: An R Package for Fitting Distributions. J Stat Softw. 2015;64(4):1-34.
19. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth ed. New York: Springer; 2002.
20. Frank SA. The common patterns of nature. J Evolution Biol. 2009;22(8):1563-85. doi: 10.1111/j.1420-9101.2009.01775.x.
21. Rodriguez S, Gaunt TR, Day INM. Hardy-Weinberg Equilibrium Testing of Biological Ascertainment for Mendelian Randomization Studies. Am J Epidemiol. 2009;169(4):505-14. doi: 10.1093/aje/kwn359.
22. O'Neill DG, Darwent EC, Church DB, Brodbelt DC. Demography and health of Pugs under primary veterinary care in England. Canine Genet Epidemiol. 2016;3:5. doi: 10.1186/s40575-016-0035-z.
23. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. 2016;7. doi: ARTN 10460; 10.1038/ncomms10460.
24. Charlson ME, Pompei P, Ales KL, Mackenzie CR. A New Method of Classifying Prognostic Co-Morbidity in Longitudinal-Studies - Development and Validation. J Chron Dis. 1987;40(5):373-83. doi: Doi 10.1016/0021-9681(87)90171-8.
25. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of Multimorbidity and Morbidity Burden for Use in Primary Care and Community Settings: A Systematic Review and Guide. Ann Fam Med. 2012;10(2):134-41. doi: 10.1370/afm.1363.
26. O'Neill DG, Scudder C, Faire JM, Church DB, McGreevy PD, Thomson PC, et al. Epidemiology of hyperadrenocorticism among 210,824 dogs attending primary-care veterinary practices in the UK from 2009 to 2014. J Small Anim Pract. 2016;57(7):365-73. doi: 10.1111/jsap.12523.
27. Taylor-Brown FE, Meeson RL, Brodbelt DC, Church DB, McGreevy PD, Thomson PC, et al. Epidemiology of Cranial Cruciate Ligament Disease Diagnosis in Dogs Attending Primary-Care Veterinary Practices in England. Vet Surg. 2015;44(6):777-83. doi: 10.1111/vsu.12349.

28. Mattin M, O'Neill D, Church D, McGreevy PD, Thomson PC, Brodbelt D. An epidemiological study of diabetes mellitus in dogs attending first opinion practice in the UK. Vet Rec. 2014;174(14):349. doi: 10.1136/vr.101950.
29. Shoop SJ, Marlow S, Church DB, English K, McGreevy PD, Stell AJ, et al. Prevalence and risk factors for mast cell tumours in dogs in England. Canine Genet Epidemiol. 2015;2:1. doi: 10.1186/2052-6687-2-1.
30. Scotch M, Odofin L, Rabinowitz P. Linkages between animal and human health sentinel data. BMC Vet Res. 2009;5:15. doi: 10.1186/1746-6148-5-15.
31. Hart BL, Hart LA, Thigpen AP, Willits NH. Long-term health effects of neutering dogs: comparison of Labrador Retrievers with Golden Retrievers. PLoS One. 2014;9(7):e102241. doi: 10.1371/journal.pone.0102241.
32. O'Neill DG, Church DB, McGreevy PD, Thomson PC, Brodbelt DC. Longevity and mortality of owned dogs in England. Vet J. 2013;198(3):638-43. doi: 10.1016/j.tvjl.2013.09.020.
33. Chase K, Jones P, Martin A, Ostrander EA, Lark KG. Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic. J Hered. 2009;100 Suppl 1:S37-41. doi: 10.1093/jhered/esp011.

Chapter 4: Age and environment modifies different components of the canine epigenome

AUTHORS: Kelly Jin[1], Elisabeth Goldman[2], Viktoria Usova[1], Alex Chitsazan[3], Anneke Kakebeen[3], Unity Jeffery[4], Kate E Creevy[4], Andrea Wills[3], Noah Snyder-Mackler[5,6], Daniel Promislow[1,7]

2. Department of Pathology, University of Washington, Seattle, WA, USA
2. Department of Anthropology, University of Oregon, Eugen, OR, USA
3. Department of Biochemistry, University of Washington, Seattle WA, USA
4. College of Veterinary Medicine, Texas A&M University, College Station, TX, USA
5. Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA
6. School of Life Sciences, Arizona State University, Tempe, AZ, USA
7. Department of Biology, University of Washington, Seattle, WA, USA

Introduction

As we age, physiological function steadily declines, while the risk of morbidity and mortality steadily rises (1). Although age is a reliable predictor of overall health across a population, when we observe measures of health at the individual level, there are those who seem to age faster or slower than others (2). Developing tools to measure this variation, and revealing the genetic and environmental factors that influence it, are active areas of research that will help us understand how and why we age.

One promising area of study includes the aging epigenome. The epigenome consists of the collection of structural and biochemical changes in the cell that alter gene expression levels without changing the actual DNA sequence, including DNA methylation, histone modifications, and changes to chromatin accessibility (3). The epigenome integrates information from both genes and environment, and is rich with changes that correlate with and potentially directly influence organismal aging (3, 4). Changes in diverse epigenetic elements, including global loss of constitutive heterochromatin (5, 6), general histone loss (7, 8), and global and local changes in DNA methylation (9-11) have all been associated with aging in vertebrate systems. In recent decades, significant resources have been invested into using epigenetic markers to develop predictive models of age in hopes not only of predicting chronological age, but also of estimating intrinsic measures of overall health compared to the population mean. Researchers have argued that the amount by which predicted age departs from chronological age can be taken as a measure of underlying health, referred to as biological age or age acceleration.

Many different age predictors now exist in addition to epigenetic measures of DNA methylation (12-15), including a wide range of other molecular and clinical measurements, such as telomere length (16), gene expression transcripts (17, 18), protein glycosylation and abundance (19, 20), metabolite levels (21), and composite clinical biomarkers such as systolic blood pressure and cholesterol levels (22-24). Across these different age predictors, the most accurate and well-characterized predictors of age are the DNA methylation (DNAm) clocks, which are also the most thoroughly validated by independent studies. Two such clocks include Horvath's 353 CpG multi-tissue DNAm age estimator (13), and Hannum's 71 CpG single-tissue DNAm age estimator (14). Both Horvath's and Hannum's DNAm clocks have been shown to be highly predictive of chronological age (13, 14), predictive of all-cause mortality and lifespan (25, 26), and claim to measure biological age or age acceleration in an organism or tissue (13). Indeed, age acceleration as measured by Horvath's and Hannum's clocks are predictive of mortality and other age-related phenotypes and diseases (27-29). However, there are still many areas of uncertainty that surround estimates of epigenetic age, such as disagreement between

different DNAm clocks (30), a possible upper limit or plateau to DNAm clock estimates when surveyed across longitudinal samples (31), and the sensitivity of the clock to cell-type proportions or training set size (32). As the level of clinical and public interest in these predictors grows, it becomes imperative for the field to carefully and thoughtfully probe the biological and clinical relevance of the age predictions that come from these clocks.

Applications of the DNAm clock now span many diverse areas of clinical and biological research, including human sub-population specific clocks (33), development of other non-human mammalian species clocks including in chimpanzees, mice, dogs, and humpback whales (34-37), and pairing of clocks with studies of lifespan-extending interventions (38). However, there are still many aspects of DNAm clocks that remain poorly understood. For example, it is unclear if age-associated DNA methylation changes are simply a consequence of other underlying cellular processes, or if the genetic loci that show these changes play a direct, causal role in age-related physiological decline. If the latter is true, measurements of other components of the epigenome may also be as or more predictive of chronological and/or biological age. In this study, we address this very question by using two different features of the epigenome – methylation status and chromatin accessibility – to build independent age predictors. We define chromatin accessibility as regions of open chromatin that are accessible specifically to the modified transposase used in the assay for transposes-accessible chromatin using sequencing (ATAC seq, (39)). Chromatin accessibility as measured by ATAC seq has been used in a wide variety of basic and clinical research fields, including embryonic development (40), tumor development (41), and aging/age-related disorders (42, 43).

We build and compare these two age predictors across a population of individuals from an exciting and relatively new model system of aging – *Canus lupus familiaris*, or the companion dog. The dog is an attractive model for aging research for a multitude of reasons. First, as the most phenotypically variable mammal on earth, dogs demonstrate considerable variation in lifespan and disease susceptibility across breeds. Larger breeds tend to have shorter lifespans than smaller breeds (44, 45). This suggests that larger breeds may be aging faster than smaller breeds (46), leading us to hypothesize that for a given age, individuals of larger breeds should have greater values of biological age or age acceleration than individuals of smaller breeds as measured by an epigenetic clock. Second, the unique breed-based population structure results in high levels of genetic homogeneity within breeds, coupled with high levels of genetic *heterogeneity* between breeds, affording researchers some level of control over genetics as well as increased statistical power when estimating mean phenotypic measurements across breeds. And lastly, dogs share our environment in a way that can never be replicated in laboratory

settings. They are exposed to the same kinds of environmental factors as people are, such as second-hand smoke, air pollution, or ambient noise. This affords researchers the opportunity to learn about the effect of these factors on human health from dog data.

Here we measured DNA methylation and chromatin accessibility profiles of peripheral blood mononuclear cells (PBMCs) from 70 companion dogs. We used reduced representation bisulfite sequencing (RRBS seq, (47)) to measure DNA methylation and ATAC seq (39) to profile global chromatin accessibility. With these data we developed a DNAm clock and, to our knowledge, the first ever ATAC seq clock, for canine age. We also carried out univariate modeling to estimate the effects of age and other biological and environmental factors on each feature. We found that 1) the DNAm clock predicts age more accurately than the ATAC seq clock, 2) residual age measures from the two clocks are not correlated with one another, and 3) a large number of ATAC seq features correlate with rural or suburban home location of the animals, suggesting that chromatin accessibility might be highly sensitive to environmental factors such as home location.

Results

Cohort of companion dogs

We measured chromatin accessibility and methylation status of PBMCs in 70 healthy companion dogs using ATAC and RRBS-seq, respectively. The distributions of age and breed size of the cohort are shown in Figures 1A and 1B. There was no correlation between age and breed size of profiled dogs (Fig 1C). The most highly represented breeds included Dachshunds, Border Collies, Labrador Retrievers, and Australian Shepherds (Fig 1D). However, the cohort was mostly composed of breeds represented by only one individual animal.

The canine epigenetic clock

To evaluate the ability of our methylation and ATAC data to predict age and assess biological age in our cohort of dogs, we built predictors of chronological age using both datasets, which from this point onward will be referred to as DNAm and ATAC clocks. We built all of our clocks using an elastic net model. Due to our relatively limited sample size, we compared two methods of feature selection. These methods include, 1) allowing the elastic net model to select from all features, or 2) implementing a feature pre-selection method and using only the top 50 most often-selected features to build the model (see Methods for details).

To obtain a preliminary assessment of model performance across both datasets using the different feature selection methods, we built 250 different models by selecting 80% of our samples at random for a training set, and then using the remaining "untouched" 20% of samples to test the model. This process was repeated 250 times to generate a distribution of performance measures for each independently built model. The methylation data resulted in more accurate age predictions compared to the ATAC data as assessed by R^2 and RMSE values of predicted versus observed age values (Fig 2A), and both models performed significantly better using 50 features (Fig 2A).

The final methylation and ATAC clocks were built using LOOCV and feature pre-selection method using 50 features (see Methods). Both models were able to predict chronological age. However, the methylation clock predicted age more accurately than the ATAC clock (R^2_{meth} = 0.91, R^2_{ATAC} = 0.55; Fig 2B). For both models, the optimal mixing parameter was α = 0.1, indicating both data types performed better with models more closely resembling Ridge regression than Lasso regression. The optimal regularization parameter was λ = 0.33 for the DNAm clock and λ = 0.71 for the ATAC clock.

To assess whether or not there was any concordance between the location of features selected by these two clocks, we mapped the top 50 features of each clock to the nearest single gene and looked for overlap between the two lists. There were no genes that overlapped between the top 50 features selected by each clock (Fig 2C). We then extended the search to the top 100, 200, and 500 most often-selected features, and found an overlap of 2 genes when we look at the top 200 features, and 15 genes when we look at the top 500 features from both clocks (P = 0.015 from random sampling of data, Fig 2D).

Residual age does not correlate with breed weight

Next, we evaluated the ability of our clocks to assess biological health, or biological age. We generated an estimate of biological age by taking the residuals of a linear regression of predicted on observed age. From this point onward, we refer to this measure as "residual age". Our method of measuring residual age is comparable to the measures of "age acceleration" from Horvath's clocks (36). Our general hypothesis is that for a given age, larger breeds, which are shorter lived and age at a more rapid rate (44, 46), will show a higher biological age than smaller dogs. Furthermore, we assume that a higher residual age value suggests an older biological age, or in other words that individual's epigenome may look older than their chronological age, while a lower residual age value suggests a younger biological age.

To test the hypothesis that dogs from shorter-lived large breeds have higher residual age values than individuals of smaller breeds, we examined the relationship between breed size and residual age as measured by both methylation or ATAC clocks. We found no relationship between breed size and residual age as estimated by either clock (Fig 3A and B).

To assess whether or not the two measures of residual age were measuring the same or different aspects of aging, we examined the relationship between the two residual age measures. We found no relationship between residual age as measured by the methylation clock and residual age as measured by the ATAC clock (Fig 3C). Collectively, we find no evidence that residual age as estimated from either clock is a measure of biological age or health.

Age, sex, weight, and home location explain many epigenetic features

The epigenome is influenced by both biological and environmental factors (3). Accordingly, we assessed the effect of age, sex, weight, and home location on individual features from both datasets. Home location was reported by owners as either “rural”, “suburban with yard”, “suburban without yard”, or “urban”. There was only one owner that reported an “urban” home location, so it was grouped with “suburban without yard”.

To test for an effect of environment on the epigenome, we modeled each individual methylome or ATAC-seq feature as a function of age, sex, breed weight, home location, and cell-type using OLS linear regression (see Equation 1 in Methods,). In total, 14,334 methylation and 15,417 ATAC features were fitted using this model. A significance value was generated for each feature and each term in the model.

To compare the impact of the four main independent predictors (age, sex, weight, and home location) on our two epigenetic datasets, we used a genomic window analysis to compare the sites with the lowest P-value across each 10kb genomic window. Generally, we find that more methylation sites than ATAC sites are associated with age, sex, and breed weight (Fig 4A, B, and D.; Table 1). However, the reverse was true for home location, which had a substantially stronger association with ATAC sites than methylation sites (Fig 4B). Over half of the ATAC features (n = 8173) were associated with home location at an FDR cutoff of 10% (Table 1). The relationship between home location and the ATAC feature with the strongest home location association is shown in Fig 4E.

Genes near ATAC peaks that are associated with home location are enriched for neurological phenotypes

To assess whether or not individual epigenetic features associated with age, sex, or home location are related to biologically relevant genes or functional domains, we performed gene enrichment on the genes that fell closest to the features identified from our previous univariate analysis. We used the Human Phenotype Ontology (HPO) database to search against our gene lists, which tests for enrichment of orthologous canine genes in a database of genes associated with different human pathological and physiological phenotypes (48). We performed gene enrichment on three groups of genes, including those closest to 1) age-associated methylation sites (n = 41 genes), 2) sex-associated methylation sites (n = 16 genes), and 3) home location-associated ATAC sites (n = 1310 genes). No significant HPO terms were associated with genes near age-associated methylation sites. However, three HPO terms were associated with genes near sex-associated methylation sites (Table 2), all related to X-linked or sex-specific phenotypes, suggesting our univariate analysis is identifying biologically relevant epigenetic features. Finally, 48 HPO terms were associated with genes that fall near home location-associated ATAC sites (Table 3), the majority of which are related to cognitive and neuro-physiological phenotypes. Taken together, these results suggest that home location may play a large role in canine cognitive and neurological function.

Discussion

DNA methylation is highly predictive of chronological age, and many have suggested that it could be even more informative than true chronological age in predicting the biological health of the organism. While DNAm clocks have delivered promising results in their estimates of biological age, there are still many unanswered questions with regards to these clocks, including whether or not other components of the epigenome can also be used to inform us about chronological and/or biological age. Here, we explore this question by examining two types of epigenetic profiles and their relationship with age and environment in the companion dog.

This is the second study to successfully build a DNAm clock for dogs, and the first that we know of to build an ATAC clock in any organism. This study is also novel in that all of our datasets (methylation, ATAC, and flow cytometry to measure cell types for control) were generated in the same set of animals. We show that the DNAm clock is a much more accurate

predictor of chronological age than the ATAC clock (Fig 2B). This result is not unexpected given how well DNAm clocks have performed in other studies (29).

There are two interesting findings from this study that we would like to highlight here. The first is the lack of concordance between the DNAm and ATAC clocks. Since we know that DNA methylation influences global chromatin accessibility, we originally hypothesized that the features selected by each clock would come from similar regions of the genome. However, genes that fell near the top 200 ATAC features most often selected by the model did not show any significant overlap with the top 200 DNAm features. It was only when we examined the top 500 features that we see a statistically significant overlap (Fig 2 C and D). Since our models were only built with the top 50 features, this indicates the genetic signal driving the two epigenetic age clocks are not coming from the same genetic loci. Furthermore, neither of the residual age measures derived from either of these clocks correlate with breed, or with one another (Fig 3). While this may be due to the small sample size of our present study, an alternative interpretation may be that DNAm and chromatin accessibility are influenced in very different ways by age and other intrinsic or extrinsic factors, and therefore each tells us very different stories about the health of an animal, despite the fact that DNA methylation contributes to global chromatin accessibility.

Secondly, when we analyzed each DNAm and ATAC feature separately using linear regression, we found striking differences in the impact of environment on the two datasets. We found that age was associated with 106 DNAm sites, but no ATAC sites (Fig 4A; Table 1). Since the DNAm clock is a more accurate predictor age than the ATAC clock (Fig 2), this result was not surprising. In contrast, we found over half of the ATAC sites (n = 8,173; Fig 4C; Table 1) were associated with home location at a 10% significance level, while no DNAm sites were associated with home location. While we are suspicious of this very high number of correlated sites, when we visualize the individual ATAC features and their relationship with home location, we indeed see a correlation (eg. Fig 4E), suggesting that home location of these animals may have a strong influence on chromatin accessibility that is measurable by ATAC seq, but less so through RRBS seq. One possible explanation may be that the effect of home location is acting through epigenetic routes other than methylation, as suggested by studies associating air pollution components with histone modifications (49, 50). The association of these particular sites with genes that are involved in human neurological development and function is yet more intriguing, as this might suggest some link between home location and neurological/cognitive health in dogs, possibility acting through chromatin remodeling. Taken together, this suggests that while ATAC seq might not perform as well as DNAm for the purposes of chronological age

prediction, ATAC-derived chromatin information could be highly predictive of an entirely different component of organismal health. The biology that underlies the possible connection between home location and epigenetic reprogramming is an intriguing one, as researchers have already found a link between adverse environmental exposures, particularly in early life, and epigenetic variation (51). These environmental exposures include a diversity of factors including air pollution (52), heavy metals (53, 54), malnutrition (55, 56), and maternal stress (57). The vast majority of studies about the effect of these environmental factors on epigenetic modifications are in the context of DNA methylation, likely due to fact that DNA methylation profiling methods have been around longer than assays that profile chromatin accessibility. However, we hope studies like ours and others, for example those that have found effects of lead exposure (58) or a Western diet (59) on chromatin accessibility, will inspire further investigation of the effects of environment on epigenetic remodeling outside of DNA methylation.

As previously discussed, ours is the second study to build an epigenetic clock in dogs that we know of. The first study to do so was described by Thompson *et al.* in 2017 (36). While the two studies both build DNAm clocks in companion dogs, they also differ in important ways. One of the primary features of the Thompson study was to compare the DNAm clock in dogs to ones built from wolves and humans. Here, our primary objective was to compare two different types of epigenetic clocks using a single population of dogs. Like our residual age values, the age acceleration estimates from the Thompson clock also show no significant correlation with breed size. Future efforts should compare the features and performance of both of these DNAm clocks.

There are at least three caveats to keep in mind when evaluating the results of this study. First, the sample size of this study (n = 70 dogs) is small, and while we are still able to build a highly predictive age model with this group of animals, the lack of correlation of our residual age measures with breed might be due to lack of statistical power due to a small sample size. In future, as part of the Dog Aging Project (60), we intend to build annual epigenetic clocks in a set of 1500 dogs followed longitudinally over the course of their lives. Our results establish the feasibility, and provide us with a lower bound on efficacy, for such measures. These future studies will include efforts to build not only a global biological clock for all dogs, but breed-specific ones as well.

Second, the metadata for the dogs in this study, including age, breed, and the environmental survey questions, was reported by the owners and not verified through objective measures. While we have no reason to believe any of the self-reported responses are

inaccurate, we acknowledge that information about pets, particularly age and breed, are not always well documented and might be subject to error.

Lastly, we also acknowledge that our interpretations of the biological differences between RRBS and ATAC seq data highlighted in this study are influenced by the inherent differences between the two data types. RRBS seq provides us with read counts of methylated and unmethylated CpG sites, which we converted to percentages prior to statistical analysis, bounded by 0 and 1. This is contrasted with ATAC seq data, which are presented as read count data from sequencing peaks with no theoretical upper bound. While both of these data were transformed to normality prior to model-building, it is possible that the inherent differences between the data types influenced how the models performed. To address this, future efforts will include in-depth analysis examining how different simulated data types perform using similar methodology.

Notwithstanding these caveats, our results point to the exciting new landscape of future studies of health and aging offered by the companion dog. The unique breed structure and highly variable longevity patterns of the domestic dog offer straightforward hypotheses to generate and test in the context of aging research. Dogs suffer from very similar diseases that humans do (61), and are exposed to the same environmental factors that we are. Furthermore, canine health itself, independent from modeling human health, is an important area of study, as many of us care a great deal about our canine companions. Canine biological and chronological age clocks have many applications, as they have the potentially not only to inform us about the health of pets, but also to generate very accurate estimates of chronological age, as the majority of adopted or rescued animals have no veterinary records with which to inform owners about age. We hope that studies such as ours will generate more enthusiasm and excitement about using companion dogs to learn about human health.

Methods

Study cohort

All dogs were recruited at Texas A&M University and comprised of pets of staff and student volunteers. All animals were declared to be healthy by the owner, although no formal veterinary exams were performed. Age, breed, and environmental survey information were reported by each owner. Sixty-eight out of 70 animals were sterilized, so we chose not to include sterilization status as a factor in this study. All procedures for this study were reviewed

and approved by the TAMU Institutional Animal Care and Use Committee (IACUC 2016-0224 CA). Because dog owners provided information about their dogs in the home environment, the study was also reviewed and approved by the TAMU Institutional Review Board (IRB2016-0532D). Informed consent was obtained from all owners at the time of enrollment.

Whole blood was drawn and PBMCs were isolated in Texas, frozen, and then shipped to Seattle, Washington where the remaining epigenetic profiling and analyses were performed.

Sample collection and PBMC isolation

Using a needle and syringe, blood (5 mL) was collected from a peripheral vein by routine venipuncture and immediately transferred to K_2EDTA vacutainers. Blood was mixed with an equal volume of 2% fetal bovine serum (HyClone) in phosphate buffered saline (HyClone), and transferred to a barrier tube (SepMate-15, StemCell technologies) prefilled with 4.5 mL of density gradient medium (Lymphoprep 1.077, StemCell technologies). After centrifugation at 1200 g for 15 minutes at room temperature, the supernatant was collected and washed 3 times with 10 mL of 2% fetal bovine serum in phosphate buffered saline by centrifugation at 300 g for 10 minutes at room temperature. Based on a hemocytometer count, cells were resuspended at a concentration of 1 x 10^6 per mL in fetal bovine serum with 10% DMSO. After 25 minutes incubation at room temperature, cells were transferred to a – 80ºC freezer within a Styrofoam container. Samples were held at -80ºC for a maximum of 4 days before shipping on dry ice. Once arriving in Seattle, samples were rapidly thawed at 37ºC for 60 seconds, a small volume was stained with Trypan Blue, and then counted using a hemocytometer to obtain cell concentration and viability estimates. Samples were then immediately distributed into aliquots for downstream analyses, including ATAC seq, RRBS seq, and flow cytometry analysis.

ATAC seq library preparation

ATAC seq was performed on canine PBMCs largely following the original protocol from Buenrostro *et al.* (39), with some modifications. Briefly, 250,000 cells were washed 3x in 1x PBS by spinning for 2 min at 2000 g. In contrast to the original published methods, we skipped the cell lysis step and moved immediately to the transposition reaction by adding the transposition buffer and transposase directly to the washed cell pellet (62). Transposition was carried out at 37ºC for 1 hour. DNA from the transposed sample was then purified using a Qiagen Minelute kit

as per manufacturer's instructions. PCR amplification of purified DNA was then conducted using Nextera PCR primers and NEB Next High-Fidelity 2x PCR Master Mix (cat no. M0541s) using the recipe and cycling program as previously described (39). Amplification was monitored in parallel using qPCR in order to reduce GC and size bias. The amplified reaction was then purified using a Qiagen PCR Cleanup kit. The final library was eluted in Qiagen Elution Buffer (10 mM Tris Buffer, pH 8) and stored at -20ºC until ready for sequencing.

Samples were prepared as described above in batch sizes ranging from 6-12 samples. After all the samples were processed, all libraries were pooled for sequencing using a 75 bp paired end Illumina Nextseq platform at the Brotman Baty Institute at the University of Washington.

ATAC seq data analysis

Software and parameters used for adaptor trimming, read alignment, and peak calling parameters were followed as described in Kakabeen *et al*. (62). A consensus peak set was used to determine feature signal for all samples. The consensus peaks were called on a merged BAM file composed of equally-subsampled reads from all samples in the experiment. Peaks with summits that were closer than 500 bp to one another were merged and considered as a single feature. Peaks were filtered to include peaks with a median coverage of >20 reads across all samples. Peaks that mapped to mitochondrial or DNA scaffolds were also removed. After filtering, 15,417 features remained in the dataset. Count values were then converted to reads per kilobases mapped (RPKM) by dividing the number of reads at each peak region by the peak width (estimated from Macs2 peak-calling software) and total reads mapped for each sample. These values were then log transformed, centered, and scaled prior to model building.

RRBS seq library preparation

RRBS libraries were generated from ~300 ng of DNA extracted from canine PBMCs following a modified version of Boyle *et al*. (63).

RRBS seq data analysis

Samples were sequenced on the Illumina HiSeq platform at the Northwest Genomics Center. Sequenced reads were trimmed with software Trim Galore!, and trimmed reads were

mapped to the dog genome (CanFam 3.1). Total methylated and unmethylated CpG sites were counted from mapped reads. CpG sites were filtered to include sites with a mean depth of 5X and median methylation level between 0.1 and 0.9 to exclude constitutively hyper- or hypo-methylated sites. Sites that mapped to mitochondrial or scaffold DNA were also removed. After filtering, 14,334 sites remained in the dataset. These values were then arcsine square root transformed prior to model building.

Flow cytometry

Cryopreserved canine PBMC samples were thawed in a 37°C water bath, half of the amount of each sample was used for flow cytometrical staining, and half was re-frozen for future analysis. Samples for flow cytometry were transferred into 50 ml conical tubes and diluted in RPMI-1640 culture media. Cells were washed twice with RPMI-1640 by spinning at 300g for 8 minutes, the resulting cellular pellets were resuspended in 50 uL of FACS staining buffer (2% fetal bovine serum in PBS) and stained with 18 uL of antibody cocktail, which includes FITC-conjugated anti-canine CD3 clone CA17.2A12 (Bio-Rad MCA1774F), PE-Cyanine 7-conjugated anti-canine CD4, clone YKIX302.9 (eBioscience 25-5040-42), Pacific Blue-conjugated anti-canine CD8, clone YCATE55.9 (Bio-Rad MCA1039PB), APC-AlexaFluor 750-conjugated anti-human CD11b clone Bear1 (Beckman Coulter A97052), Brilliant Violet 605-conjugated anti-human CD14, clone M5E2 (Becton Dickenson 564054), Alexa Fluor 647-conjugated anti-canine CD21, clone CA2.1D6 (Bio-Rad MCA1781A647), Brilliant Violet 785-conjugated anti-mouse/human CD44, clone IM7 (Biolegend 103059), PE-conjugated anti-human CD62L clone FMC46 (Bio-Rad MCA1076PE), Brilliant UltraViolet 395-conjugated anti-human CD94, clone HP-3D9 (BD OptiBuild 743954). Cells were stained for 20 minutes at 4°C and washed twice with FACS staining buffer. After the last wash, stained cells were resuspended in FACS buffer containing 7-AAD (1:500 dilution) and immediately run on an LSR Fortessa flow cytometer (BD Biosciences). Data were analyzed using FlowJo 10. Doublets were excluded based on FSC-A/FSC-H and SSC-A/SSC-H gating. Lymphocytes, monocytes and granulocytes were gated based on FSC-A and SSC-A parameters, confirmed by lineage-restricted expression of CD11b and CD14. T cells were defined as CD3+/CD21- lymphocytes, B cells were defined as CD3-/CD21+ lymphocytes, NK cells were defined as CD3-/CD21-/CD94+ lymphocytes; within T cells we identified the following populations: CD94+ T cells defined as CD3+/CD21-/CD94+ lymphocytes and conventional CD94- T cells defined as CD3+/CD21-/CD94- lymphocytes. CD4 and CD8 T cells were defined within CD94- T cells as CD4+/CD8-/CD3+/CD21-/CD94-

lymphocytes and CD4-/CD8+/CD3+/CD21-/CD94- lymphocytes, respectively. Double-positive and double-negative T cells were defined within CD94- T cells as CD4+/CD8+/CD3+/CD21-/CD94- lymphocytes and CD4-/CD8-/CD3+/CD21-/CD94- lymphocytes, respectively. Within CD4 and CD8 T cells we defined CD62L- and CD62L+ subsets as well as CD44Low and CD44High subsets.

Statistical analysis

All data analysis and visualization was performed using statistical analysis software R (64). P-values were adjusted for multiple comparisons using the Benjamini-Hochberg-Yekutieli procedure (65).

Principle component analysis (PCA): PCA was conducted using the prcomp() function in R on cell-type proportions from the flow cytometry data. PC scores were included as covariates in the univariate modeling described below.

Epigenetic clock: We use the R packages glmnet and caret to build epigenetic clocks using either ATAC or RRBS data. We used an elastic net model, which is a penalized regression model that allows for tuning of two parameters, the mixing parameter (α) and regularization parameter (λ). Briefly, α can range from 0 to 1, where an $\alpha = 0$ is equivalent to Ridge regression, and an $\alpha = 1$ is equivalent to Lasso regression. The role of the regularization parameter is to minimize mean-squared error. The greater the value of λ, the greater the penalty. We trained our models by optimizing both α and λ.

We use two sampling methods to evaluate the ability of either ATAC seq or RRBS seq data to build a predictive clock. We call the first method the random-sampling method. Using the random-sampling method, we randomly select 80% of the samples to train the model, and then use the remaining “unseen” 20% of samples to evaluate the model. This sampling process was repeated 500 times, allowing us to generate a distribution of 500 RMSE and R^2 terms (one per model) to evaluate performance. The second sampling method we used to build the model was a leave-one-out-cross validation (LOOCV) approach, where all but one of the samples were used to build a model, and the model was used to predict the age of the left-out sample. This method gives one predicted age per sample, and thus a single RMSE and R^2 term. We refer to the model generated from this sampling method as the final model.

We also tried building models with and without feature pre-selection. To pre-select features, we applied the random-sampling method to both datasets and observed the features that were most often selected across all 500 iterations. We then reran the model building processes using only the top 50 most frequently selected features. We call this process feature pre-selection. We show substantial improvement in predictive ability using the pre-selection method (Fig 2A), so we used this method to build our final model. We acknowledge there could be pitfalls to using this pre-selection method, which we discuss in the Discussion.

Univariate modeling: We modeled individual features using an ordinary least square (OLS) linear model. We used the following model:

$$\text{Feature} \sim \text{age} + \text{weight} + \text{sex} + \text{home\_location} + \text{cell\_PC1} + \text{cell\_PC2} \qquad \text{(Eqn 1)}$$

where cell_PC1 and cell_PC2 are the PC scores from the PCA of all cell-type proportions obtained from flow cytometry. We used the Tracy-Widom test (66) to determine that the first two components were significant, and therefore included them in our model to account for correlation between blood cell types and epigenetic signal.

Gene enrichment analysis: We used the R package gprofiler2 and web-based version of the same tool, g:Profiler (67), to perform gene enrichment analysis using the Human Phenotype Ontology database (48).

Tables

Table 1. Number of methylation or ATAC features associated with different model terms at FDR 10%

	Age	Sex	Weight	Home Location	Cell-type PC1	Cell-type PC2
Methylation (n = 14,334)	106	130	6	0	1	0
ATAC (n = 15,417)	0	14	0	8173	0	0

Table 2. Human Phenotype ontology terms associated with sex-correlated methylation sites

Term name	Adjusted P-value	Term size	Query size	Intersection size
X-linked inheritance	2.96E-11	229	16	13
Gonosomal inheritance	3.32E-11	231	16	13
X-linked recessive inheritance	1.20E-07	170	16	10

Table 3. Human Phenotype ontology terms associated with home-location-correlated ATAC sites

Term name	Adjusted P-value	Term size	Query size	Intersection size
Abnormality of the cerebrum	1.91E-13	1495	1310	604
Aplasia/Hypoplasia involving the central nervous system	5.49E-13	1362	1310	557
Abnormality of forebrain morphology	1.13E-12	1516	1310	608
Hypertonia	3.28E-11	819	1310	359
Decreased head circumference	3.85E-10	978	1310	412
Abnormality of brain morphology	5.30E-10	1973	1310	747
Abnormality of skull size	6.90E-10	1204	1310	490
Global developmental delay	7.60E-10	1427	1310	566
Abnormality of nervous system physiology	9.56E-10	2883	1310	1028
Abnormal muscle tone	1.05E-09	1653	1310	641
Microcephaly	1.92E-09	969	1310	406
Abnormality of nervous system morphology	2.60E-09	2364	1310	868
Morphological abnormality of the central nervous system	3.08E-09	2162	1310	804
Aplasia/Hypoplasia of the cerebrum	4.58E-09	1168	1310	474
Abnormal muscle physiology	1.72E-08	1995	1310	747
Neurodevelopmental abnormality	1.78E-07	2170	1310	798
Abnormality of the nervous system	4.64E-07	3084	1310	1076
Neurodevelopmental delay	5.26E-07	1717	1310	649
Spasticity	5.36E-07	679	1310	292
Abnormality of higher mental function	6.40E-07	2085	1310	768
Generalized hypotonia	4.19E-06	914	1310	372
Abnormality of central motor function	6.42E-06	1523	1310	579
Brain atrophy	7.48E-06	574	1310	249
Abnormality of the musculature	8.36E-06	2264	1310	819
Abnormality of movement	1.47E-05	1683	1310	630
Atrophy/Degeneration affecting the cerebrum	2.44E-05	528	1310	230
Upper motor neuron dysfunction	2.86E-05	1032	1310	409
Atrophy/Degeneration affecting the central nervous system	3.36E-05	653	1310	275
Cerebral atrophy	6.72E-05	522	1310	226
Muscular hypotonia	7.33E-05	1403	1310	533
Intrauterine growth retardation	0.000173561	432	1310	191
Neurological speech impairment	0.00019775	935	1310	371
Slanting of the palpebral fissure	0.000233279	520	1310	223
Abnormality of the eye	0.000350596	2411	1310	856
Abnormal skull morphology	0.000461777	1768	1310	649
Abnormality of the forehead	0.000497074	761	1310	308
Sloping forehead	0.001073436	139	1310	74
Ventriculomegaly	0.001240387	515	1310	218
Abnormal axial skeleton morphology	0.001267397	2048	1310	737
Abnormality of the palpebral fissures	0.001763654	600	1310	248
Abnormal cortical gyration	0.002030469	271	1310	126
Abnormality of the cerebral white matter	0.002745752	718	1310	289
Abnormality of skeletal morphology	0.003847062	2416	1310	851
Abnormality of the cerebral subcortex	0.005130708	802	1310	317
Abnormality of the skeletal system	0.007196294	2514	1310	880
Feeding difficulties	0.008488871	754	1310	299
Seizures	0.00897571	1493	1310	550
Abnormality of the cerebral cortex	0.009731357	309	1310	138

Figures

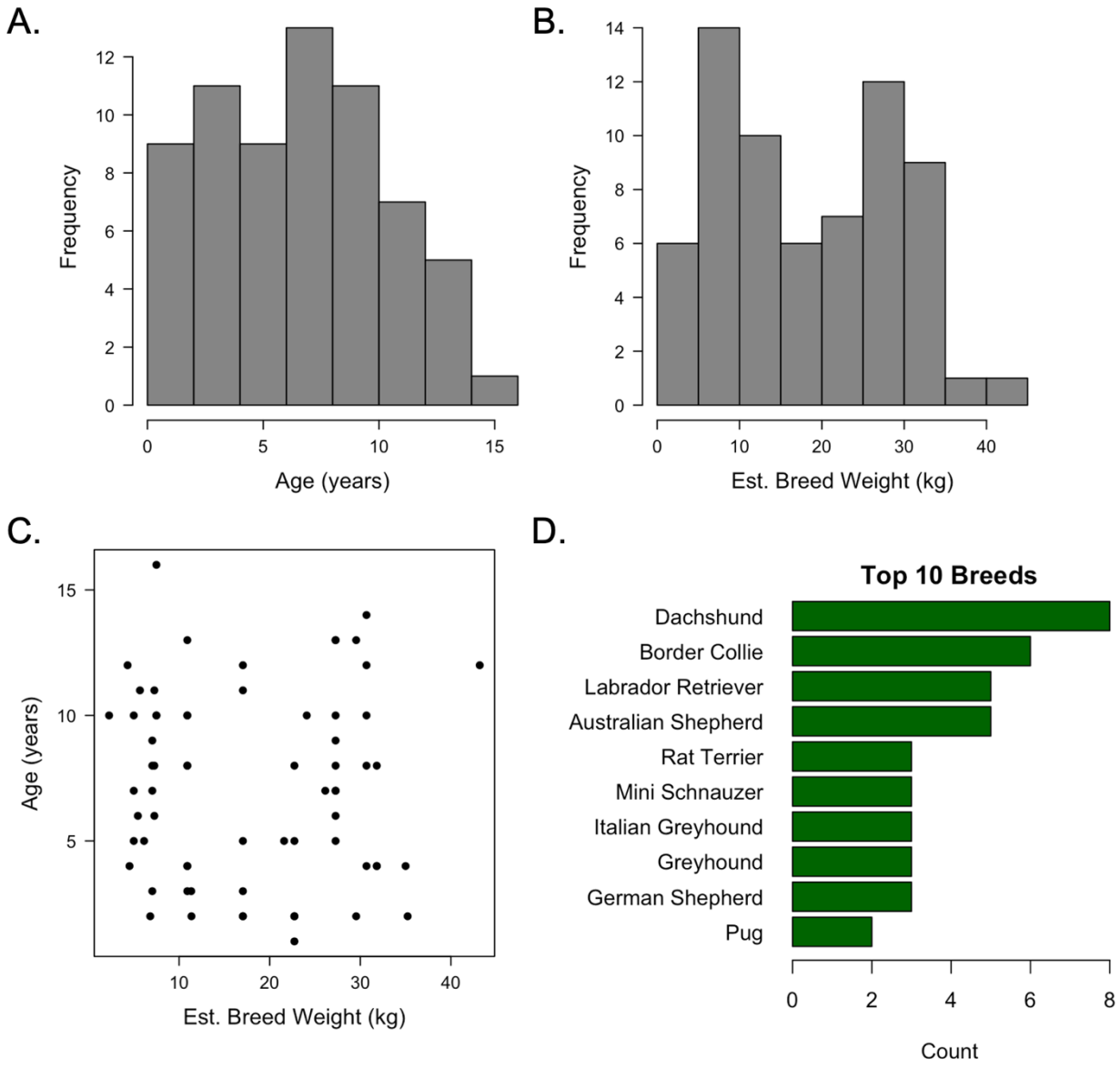
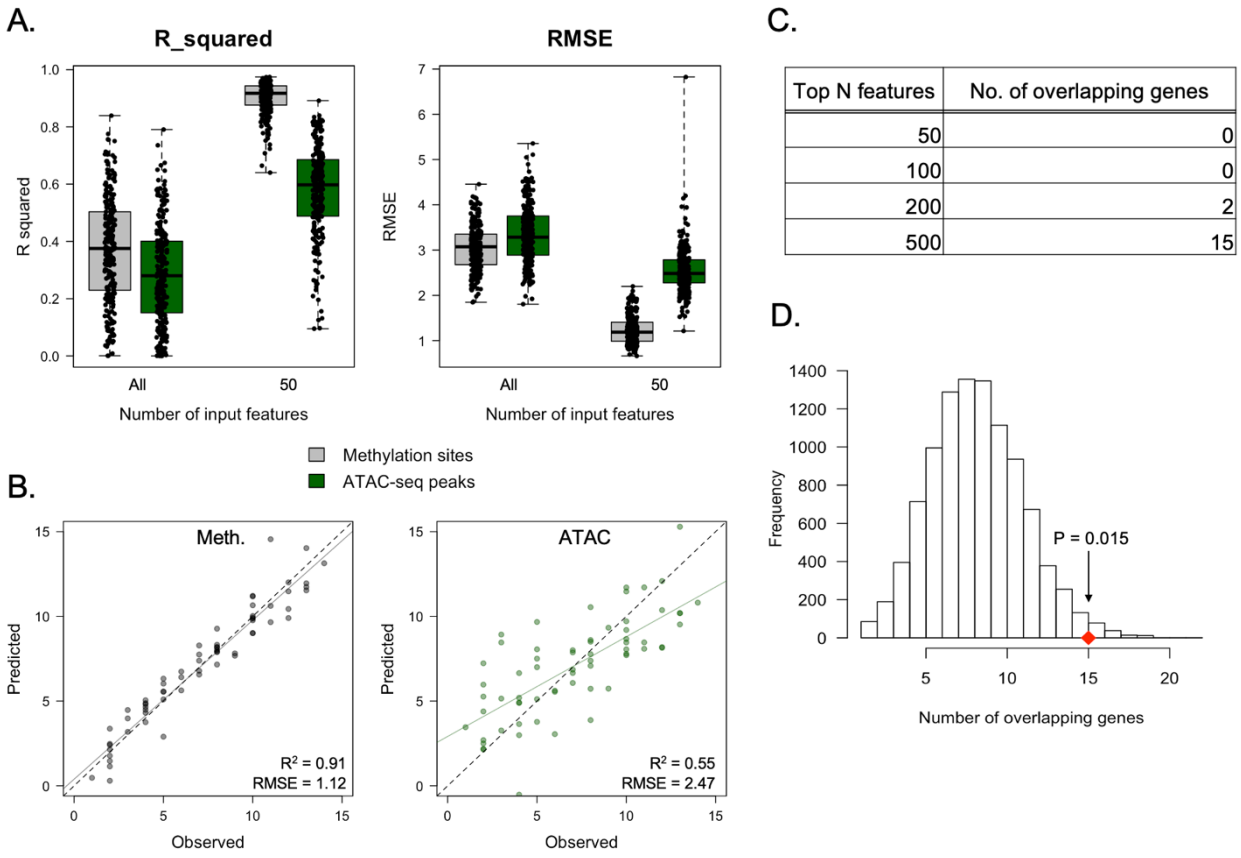


Figure 1. Sample cohort information. A-B. Age and estimated breed weight distribution of 70 dogs in the cohort. C. Correlation between age and estimated breed weight. D. Top 10 most highly represented breeds in the cohort.



Top N features	No. of overlapping genes
50	0
100	0
200	2
500	15

Figure 2. The canine epigenetic clock. A. Distribution of R^2 and RMSE between predicted and observed age values for different epigenetic clock models built from 250 randomly sampled training/testing sets. B. Final model performance of different clock models using top 50 features. C. The number of overlapping genes from top N ATAC clock features and top N DNAm clock features. D. Distribution of 10,000 permutations of the number of overlapping genes from 500 randomly chosen ATAC features and 500 randomly chosen DNAm features. Red diamond indicates the real overlap of genes between the true top 500 ATAC and top 500 DNAm features.

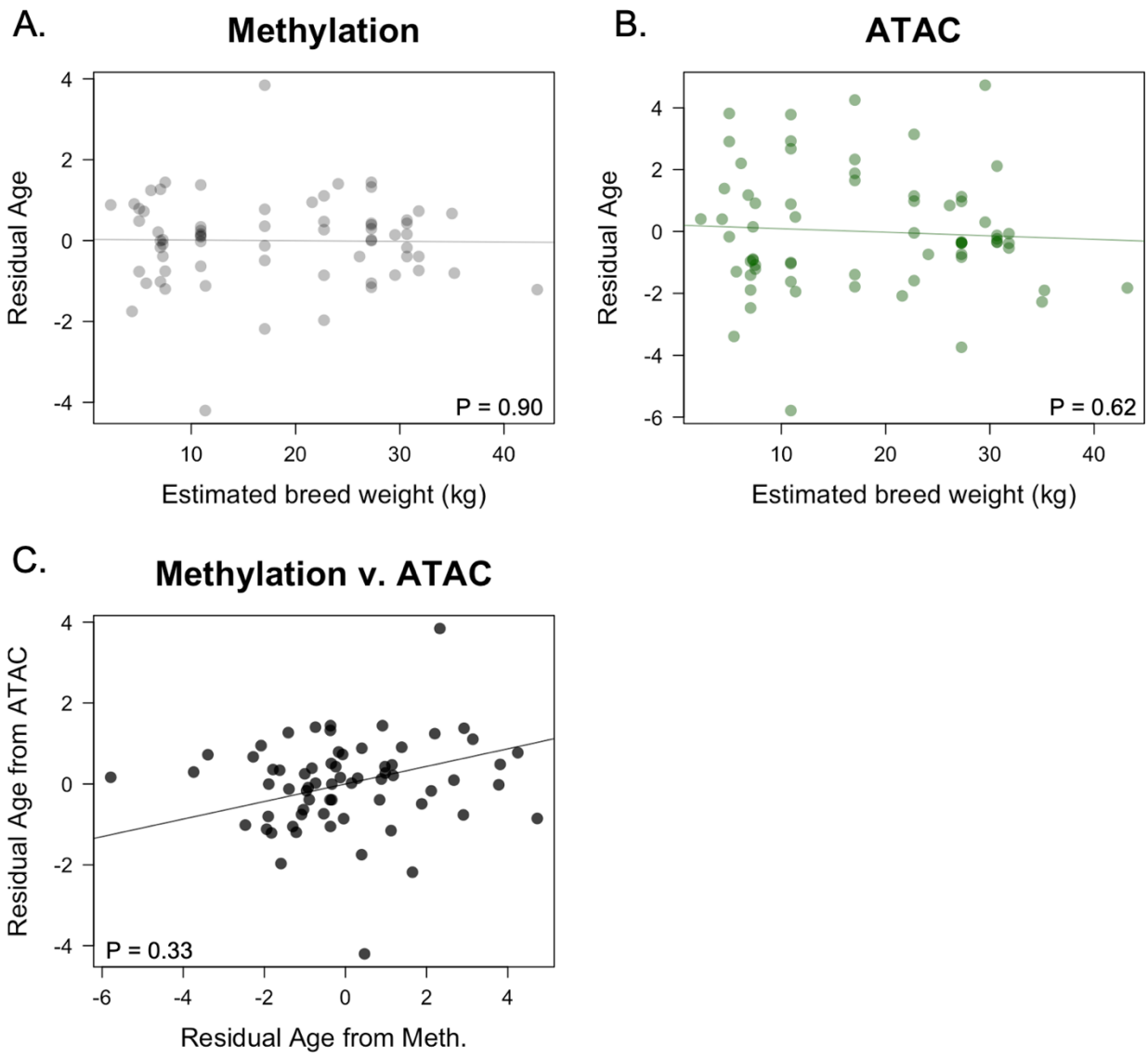


Figure 3. Residual age estimates. A-B. Relationship between estimated breed weight and residual age prediction from 50-feature methylation (A) and ATAC (B) epigenetic age clocks. C. There is no correlation between residual age predicted from methylation and ATAC epigenetic clocks.

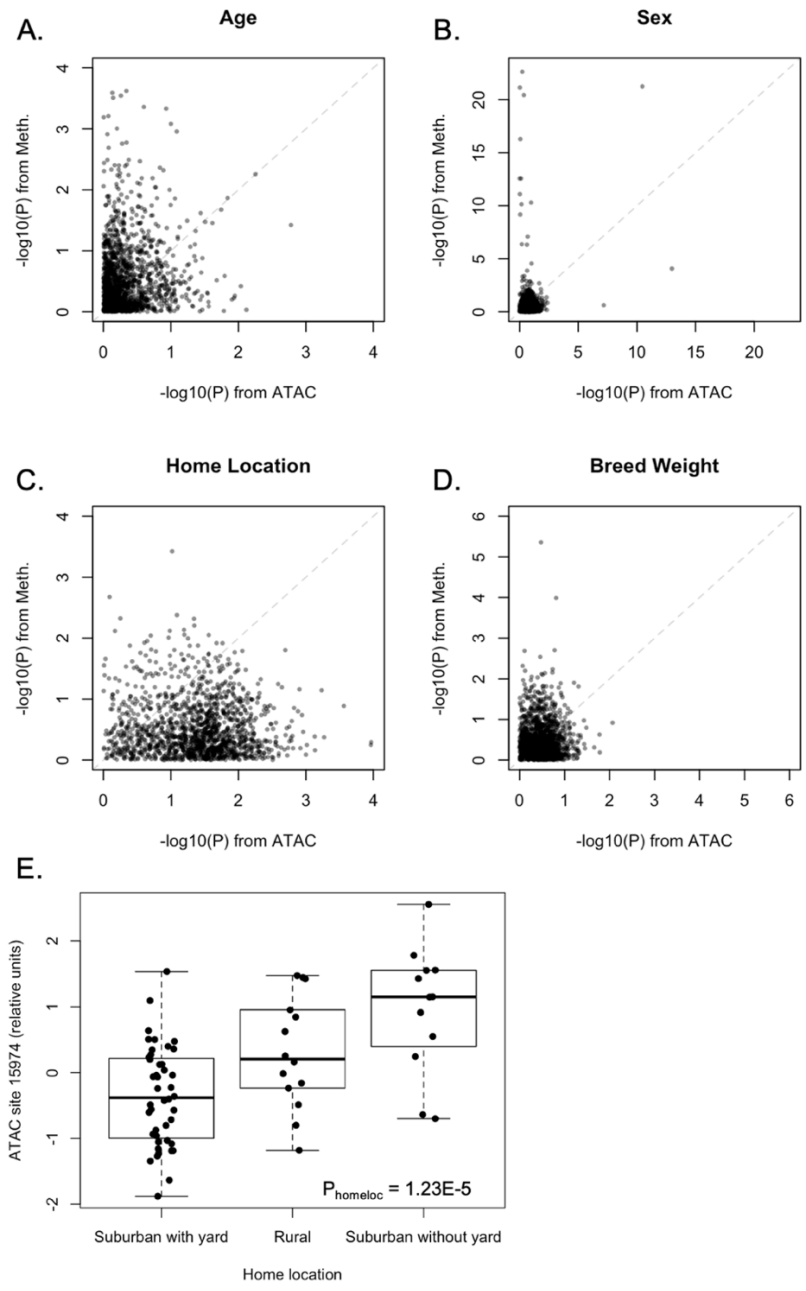


Figure 4. The effect of age, sex, home location, and breed weight on epigenetic features. Multiple linear regression was performed on each methylation and ATAC feature, including age, sex, weight, home location, and cell-type PCs as predictors. Shown here are correlations of minimum P-values from age (A), sex (B), home location (C), and weight (D) terms across 10 kb genomic windows between methylation and ATAC features. E. Relationship between ATAC peak 15974 and home location. P-value is from home-location term in ANOVA model.

References

1. M. Kaeberlein, P. S. Rabinovitch, G. M. Martin, Healthy aging: The ultimate preventative medicine. *Science* **350**, 1191-1193 (2015).
2. K. Christensen *et al.*, "Looking old for your age": genetics and mortality. *Epidemiology* **15**, 251-252 (2004).
3. P. Sen, P. P. Shah, R. Nativio, S. L. Berger, Epigenetic Mechanisms of Longevity and Aging. *Cell* **166**, 822-839 (2016).
4. P. Oberdoerffer, D. A. Sinclair, The role of nuclear architecture in genomic instability and ageing. *Nat Rev Mol Cell Biol* **8**, 692-702 (2007).
5. P. Trojer, D. Reinberg, Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**, 1-13 (2007).
6. R. C. Allshire, H. D. Madhani, Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229-244 (2018).
7. W. Dang *et al.*, Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802-807 (2009).
8. R. J. O'Sullivan, S. Kubicek, S. L. Schreiber, J. Karlseder, Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat Struct Mol Biol* **17**, 1218-1225 (2010).
9. D. G. Hernandez *et al.*, Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* **20**, 1164-1172 (2011).
10. V. K. Rakyan *et al.*, Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* **20**, 434-439 (2010).
11. A. E. Teschendorff *et al.*, Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-446 (2010).
12. S. Bocklandt *et al.*, Epigenetic predictor of age. *PLoS One* **6**, e14821 (2011).
13. S. Horvath, DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).
14. G. Hannum *et al.*, Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367 (2013).
15. M. E. Levine *et al.*, An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573-591 (2018).
16. E. H. Blackburn, C. W. Greider, J. W. Szostak, Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med* **12**, 1133-1138 (2006).
17. A. C. Holly *et al.*, Towards a gene expression biomarker set for human biological age. *Aging Cell* **12**, 324-326 (2013).
18. M. J. Peters *et al.*, The transcriptional landscape of age in human peripheral blood. *Nat Commun* **6**, 8570 (2015).
19. J. Krištić *et al.*, Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789 (2014).
20. C. Menni *et al.*, Circulating Proteomic Signatures of Chronological Age. *J Gerontol A Biol Sci Med Sci* **70**, 809-816 (2015).
21. J. Hertel *et al.*, Measuring Biological Age via Metabonomics: The Metabolic Age Score. *J Proteome Res* **15**, 400-410 (2016).
22. D. W. Belsky *et al.*, Quantification of biological aging in young adults. *Proc Natl Acad Sci U S A* **112**, E4104-4110 (2015).
23. M. E. Levine, Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J Gerontol A Biol Sci Med Sci* **68**, 667-674 (2013).
24. Q. Li *et al.*, Homeostatic dysregulation proceeds in parallel in multiple physiological systems. *Aging Cell* **14**, 1103-1112 (2015).
25. R. E. Marioni *et al.*, DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**, 25 (2015).
26. L. Perna *et al.*, Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* **8**, 64 (2016).
27. S. Horvath *et al.*, Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* **111**, 15538-15543 (2014).
28. A. Quach *et al.*, Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)* **9**, 419-446 (2017).
29. C. G. Bell *et al.*, DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**, 249 (2019).
30. D. W. Belsky *et al.*, Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? *Am J Epidemiol* **187**, 1220-1230 (2018).
31. R. E. Marioni *et al.*, Tracking the Epigenetic Clock Across the Human Life Course: A Meta-analysis of Longitudinal Cohort Data. *J Gerontol A Biol Sci Med Sci* **74**, 57-61 (2019).

32. Q. Zhang *et al.*, Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med* **11**, 54 (2019).
33. S. Horvath *et al.*, An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* **17**, 171 (2016).
34. H. Ito, T. Udono, S. Hirata, M. Inoue-Murayama, Estimation of chimpanzee age based on DNA methylation. *Sci Rep* **8**, 9998 (2018).
35. S. Maegawa *et al.*, Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* **20**, 332-340 (2010).
36. M. J. Thompson, B. vonHoldt, S. Horvath, M. Pellegrini, An epigenetic aging clock for dogs and wolves. *Aging (Albany NY)* **9**, 1055-1068 (2017).
37. A. M. Polanowski, J. Robbins, D. Chandler, S. N. Jarman, Epigenetic estimation of age in humpback whales. *Mol Ecol Resour* **14**, 976-987 (2014).
38. A. Sziráki, A. Tyshkovskiy, V. N. Gladyshev, Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell* **17**, e12738 (2018).
39. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
40. J. Wu *et al.*, The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657 (2016).
41. K. Davie *et al.*, Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet* **11**, e1004994 (2015).
42. D. M. Moskowitz *et al.*, Epigenomics of human CD8 T cell differentiation and aging. *Sci Immunol* **2** (2017).
43. J. Wang *et al.*, ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat Commun* **9**, 1364 (2018).
44. G. J. Patronek, D. J. Waters, L. T. Glickman, Comparative longevity of pet dogs and humans: implications for gerontology research. *J Gerontol A Biol Sci Med Sci* **52**, B171-178 (1997).
45. F. Galis, I. Van der Sluijs, T. J. Van Dooren, J. A. Metz, M. Nussbaumer, Do large dogs die young? *J Exp Zool B Mol Dev Evol* **308**, 119-126 (2007).
46. C. Kraus, S. Pavard, D. E. Promislow, The size-life span trade-off decomposed: why large dogs die young. *Am Nat* **181**, 492-505 (2013).
47. A. Meissner *et al.*, Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-5877 (2005).
48. P. N. Robinson *et al.*, The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* **83**, 610-615 (2008).
49. C. Liu *et al.*, Characterization of genome-wide H3K27ac profiles reveals a distinct PM2.5-associated histone modification signature. *Environ Health* **14**, 65 (2015).
50. L. Cantone *et al.*, Inhalable metal-rich air particles and histone H3K4 dimethylation and H3K9 acetylation in a cross-sectional study of steel workers. *Environ Health Perspect* **119**, 964-969 (2011).
51. A. Vaiserman, Epidemiologic evidence for association between adverse environmental exposures in early life and epigenetic variation: a potential link to disease susceptibility? *Clin Epigenetics* **7**, 96 (2015).
52. F. Perera *et al.*, Relation of DNA methylation of 5'-CpG island of ACSL3 to transplacental exposure to airborne polycyclic aromatic hydrocarbons and childhood asthma. *PLoS One* **4**, e4488 (2009).
53. A. Domingo-Relloso *et al.*, Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. *Environ Health Perspect* **128**, 67005 (2020).
54. A. P. Sanders *et al.*, Cadmium exposure and the epigenome: Exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. *Epigenetics* **9**, 212-221 (2014).
55. B. T. Heijmans *et al.*, Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* **105**, 17046-17049 (2008).
56. E. W. Tobi *et al.*, DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet* **18**, 4046-4053 (2009).
57. T. F. Oberlander *et al.*, Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics* **3**, 97-106 (2008).
58. M. Khalid, M. Abdollahi, Epigenetic modifications associated with pathophysiological effects of lead exposure. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* **37**, 235-287 (2019).
59. A. Christ *et al.*, Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming. *Cell* **172**, 162-175.e114 (2018).
60. M. Kaeberlein, K. E. Creevy, D. E. Promislow, The dog aging project: translational geroscience in companion animals. *Mamm Genome* **27**, 279-288 (2016).
61. J. M. Hoffman, K. E. Creevy, A. Franks, D. G. O'Neill, D. E. L. Promislow, The companion dog as a model for human aging and mortality. *Aging Cell* **17**, e12737 (2018).

62. A. D. Kakebeen, A. D. Chitsazan, M. C. Williams, L. M. Saunders, A. E. Wills, Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of regeneration in neural progenitors. *Elife* **9** (2020).
63. P. Boyle *et al.*, Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol* **13**, R92 (2012).
64. R Core Team (2014) R: A language and environment for statistical computing. ed R. F. f. S. Computing (Vienna, Austria).
65. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
66. N. E. Karoui, Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Annals of Probability* **35**, 663-714 (2007).
67. U. Raudvere *et al.*, g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019).