# Molecular Patterns and Signatures of Longevity

# Molecular Patterns and Signatures of Longevity

A dissertation presented

by

Siming Ma

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor: Professor Vadim N. Gladyshev Siming Ma

**Molecular Patterns and Signatures of Longevity**

Abstract

Since their divergence from a common ancestor some 200 million years ago, mammals have undergone significant diversification in physiology, morphology, habitat, size, and longevity. The maximum lifespan of mammalian species ranges from under 3 to over 200 years, but the molecular basis of such variation is poorly understood. While many genes, pathways, dietary interventions, and pharmacological compounds have been shown to influence the lifespan of model organisms, it is not known whether the same mechanisms are responsible for the longevity variation across different species. This thesis presents the analyses of gene expression and the levels of metabolites, chemical elements, and/or proteins, across multiple organs and tissues of up to 42 species of mammals, as well as the analyses of 5 long-lived mouse models, 22 natural isolates of yeast, and 16 species of fruit flies, to identify the molecular patterns and signatures associated with species longevity. The results show that longer-lived mammals up-regulate ribosomal proteins and genes involved in DNA repair, and down-regulate ubiquitin-mediated proteolysis and apoptotic functions. Some of the metabolic changes in long-lived mammals, such as higher levels of sphingomyelins and glycerophospholipids but lower levels of polyunsaturated triacylglycerols, were also observed in long-lived mouse models. Yeast strains of varying replicative lifespan differed in their aerobic respiration capacity, attributable to different protein composition in mitochondria. Long-lived fruit flies overexpressed the genes involved in lipid metabolism but suppressed the genes involved in neuronal development. Many genes previously implicated in lifespan control in model organisms also showed the expected correlation with the longevity traits across species. This thesis presents the snapshots of the complex changes associated with species natural lifespan variation and offers new insights into the mechanisms of longevity control and potential lifespan extension strategies.

# Table of Contents


**Chapter 1 Introduction** **1**
Aging Within Single Species 2
Longevity Across Different Species 7
Scope of Dissertation 13
References 15

**Chapter 2 Methods and Data Quality** **21**
Phylogenetic Regression 22
*Ordinary Least Squares approach* *25*
*Generalized Least Squares approach* *27*
*Maximum Likelihood approach* *29*
*Different trait evolution models* *31*
*Two-step verification procedure for robustness* *36*
Species Without Reference Genomes 37
*RNA sequencing, read alignment and read counting* *37*
*Identification of gene orthologs across species* *39*
*Ortholog quality assessment* *42*
Data Variability 45
*Within-species variation vs. between-species variation* *46*
*Samples from different datasets* *47*
*Metabolic effects of drug treatment and gene knockout* *48*
References 49

**Chapter 3 Mammalian Metabolome** **51**
Abstract 52
Introduction 53
Results 55
*Metabolite conservation and divergence among organs* *55*
*Metabolite profiles reflect organ functions* *63*
*Metabolites with lineage-specific changes* *66*
*Metabolome remodeling and longevity variation* *70*
*Body mass and longevity signatures* *71*
*Insights from the analysis of long-lived mouse models* *76*
Discussion 84
Experimental Procedures 86
References 91

**Chapter 4 Mammalian Ionome** **97**
Abstract 98
Introduction 99
Results 100
*Conservation of the ionomes of mammalian organs* *100*
*Interactions among elements are indicative of biological functions* *104*
*Ionomes of mammalian organs* *106*
*Lineage-specific distribution of elements* *112*
*Nature alters elemental composition to adjust lifespan and body mass* *117*
Discussion 123
Experimental Procedures 124
References 126

**Chapter 5 Yeast Transcriptome And Proteome** **129**
Abstract 130

Introduction 131
Results 133
*Variation in replicative lifespan and growth rate across natural yeast isolates* *133*
*Phenotypic variation across strains* *136*
*Correlation between phenotype and lifespan* *140*
*Networks and pathways represented by top hits* *144*
*Mitochondrial abundance and composition of the strains* *146*
*Comparison of related long-lived and short-lived strains* *149*
Discussion 150
Experimental Procedures 154
References 156

**Chapter 6 Fly Transcriptome** **159**
Abstract 160
Introduction 161
Result 163
*Life history traits in Drosophila* *163*
*Gene expression variation reflects evolutionary relationships* *167*
*Expression variation is best described by stabilizing selection* *170*
*Gene expression and longevity* *173*
*Changes in gene expression drive changes in species lifespan* *177*
Discussion 180
Experimental Procedures 181
References 185

**Chapter 7 Rodent Fibroblasts** **191**
Abstract 192
Introduction 193
Results 195
*Longevity trait variation among the species* *195*
*Gene expression by RNA sequencing* *198*
*Gene expression patterns in fibroblasts follow phylogeny* *203*
*Expression of many genes correlates with longevity traits* *206*
*Genes showing positive correlation with lifespan* *209*
*Expression of genes showing negative correlation with lifespan* *212*
*Metabolites correlating with longevity traits* *214*
Discussion 219
Experimental Procedures 222
References 225

**Chapter 8 Mammalian Transcriptome** **231**
Preliminary Results and Discussion 231
*Biological samples and RNA sequencing* *231*
*Segregation of samples by organ origin* *236*
*Organ-specific expression patterns* *239*
*Genes associated with longevity* *240*
*Gene expression and metabolic pathways* *245*
References 247

**Chapter 9 Conclusions** **249**
Summary of Main Findings 250
Within-species vs. Cross-species Lifespan Variations 254
Potential Pitfalls 257
Future Directions 259
References 261

To my parents, my grandma,

Uncle Michael, Aunt Pauline, Aunt Mingjia, and Uncle Pak Kin,

for their care, support, and encouragement;

To Prof. Huay Cheem Tan and Mr. Richard Yap,

for their trust and confidence in me over the past 10 years;

And to Prof. Vadim N. Gladyshev and all members of the lab,

for a truly memorable and enriching learning journey.

# Chapter 1 Introduction

**The great ships full of boys and girls sent in search of the immortal medicine by the Chinese Emperor Shih Huang Ti**. By Utagawa Kuniyoshi (1797-1861). Source: http://www.kuniyoshiproject.com/Warrior%20triptychs%201839-1841,%20Part%20I%20%28T47-T62%29.htm

According to Benjamin Franklin, "The only things certain in life are death and taxes." Although the deadline for filing IRS tax returns comes every April, most of us are more concerned about the one that comes once a lifetime. Legend has it that more than 2000 years ago the First Emperor of China sent hundreds of young men and women out to the sea in search of the elixir of immortality. The First Emperor expired before the return of this expedition, but the quest for eternal youth has never ceased. Today, there is an active research enterprise focused on the biology of aging and are many health and supplement products claiming anti-aging effects; yet the secrets of longevity remain unsolved.

## AGING WITHIN SINGLE SPECIES

The research aging and longevity can be largely divided into two areas. One is the aging process within a single species (Figure 1.1). As an organism grows old, a large number of changes occur on the cellular level; Lopez-Otin and colleagues summarize 9 typical hallmarks of aging: increased genome instability, telomere attrition, changes in epigenetic markers, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, induction of cellular senescence, exhaustion of stem cell population, and altered intercellular communication (Lopez-Otin et al., 2013). On the physiological level, the aging phenotypes include reduction in hair re-growth, dermal thickness and subcutaneous adipose in mice (Tyner et al., 2002); decreases in pharyngeal pumping, body movement and chemotaxis in *C. elegans* (Collins et al., 2008); and reduction in locomotion, reproduction and climbing activities in flies (Iliadi and Boulianne, 2010). In addition, in the absence of extrinsic causes of death, the mortality rate (i.e. number of deaths per unit population per unit time) of most species increases with age (Jones et al., 2014), producing a concave, downward sloping survival curve. Much effort has been directed to identify the genes, pathways, and treatments that can either delay these age-related changes, or increase the lifespan of the species (i.e. shifting the survival curve to the right), or both (Figure 1.1B).

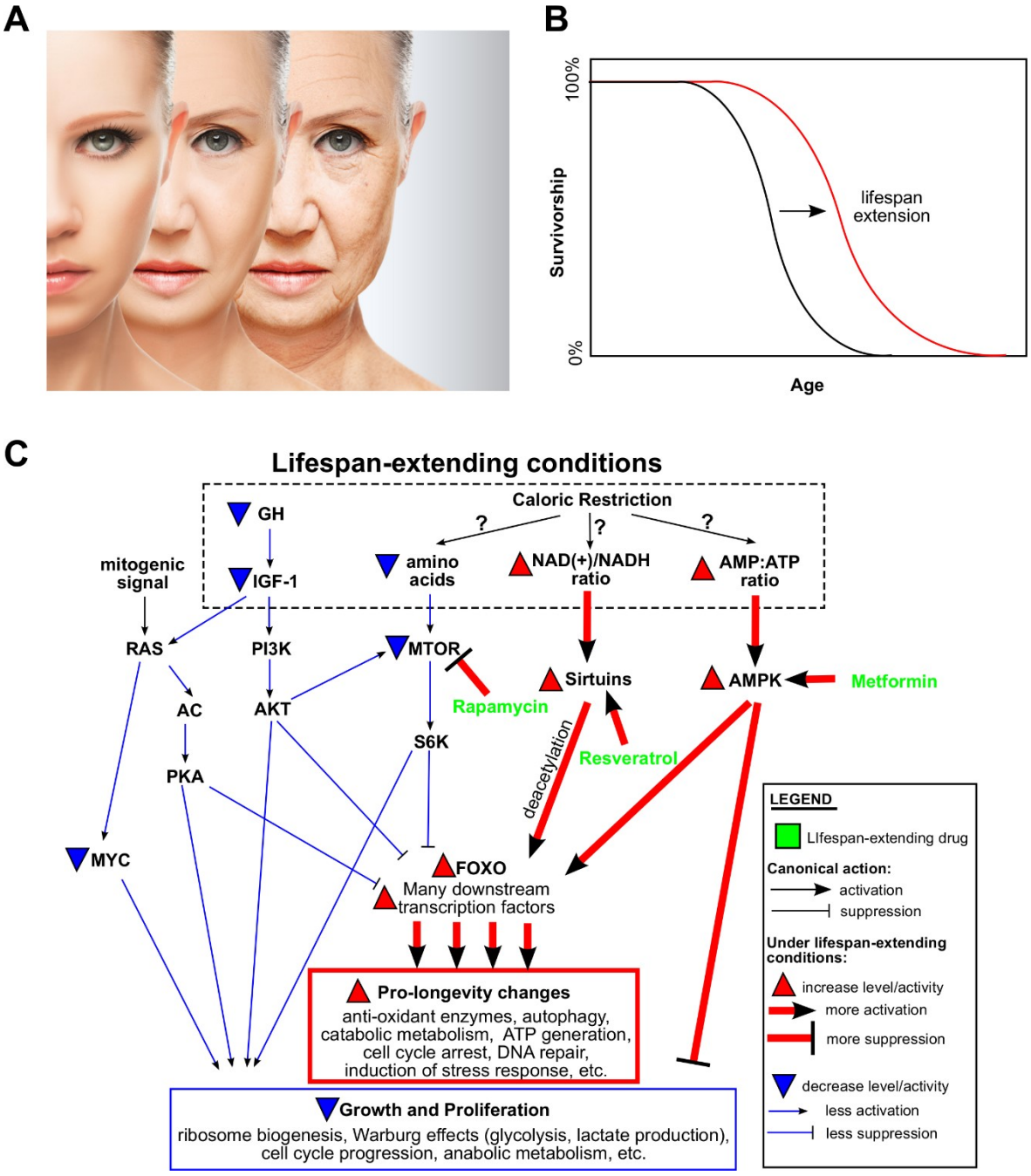



**Figure 1.1. Aging within a single species.**

**(A) Physiological changes with age.** Photo credit: JenkoAtaman, Adobe Stock.

**(B) A typical survival curve.** A successful lifespan extension strategy shifts or scales the black curve to the red.

**(C) Major pathways implicated in lifespan extension.** Only selected components of the pathways are shown. Cross-talks among the pathways are omitted.

Research across various model organisms has identified several pathways involved in lifespan control (Fontana et al., 2010; Haigis and Sinclair, 2010; Kenyon, 2010) (Figure 1.1C). One is the GH/IGF-1/PI3K/AKT/FOXO axis (GH: growth hormone; IGF-1: insulin-like growth factor 1; PI3K: phosphoinositide 3-kinase; AKT: protein kinase B; FOXO: forkhead box O). FOXO proteins are a group of transcription factors controlling a number of downstream processes to promote antioxidant enzymes synthesis, cell-cycle arrest, and energy homeostasis, with the overall effects of increasing stress resistance and enhancing survival (Carter and Brunet, 2007). In the presence of growth signal IGF-1, AKT phosphorylates FOXO to keep them in the cytoplasm, and the cell favors a program of growth and proliferation. Ablation of IGF-1 signaling pathway has been found to extend lifespan, first reported in the nematode *C. elegans* (Friedman and Johnson, 1988; Kenyon et al., 1993), and later confirmed in flies and mammals (Fontana et al., 2010; Taguchi and White, 2008). Modulation of GH signaling, which in turn affects IGF-1 secretion, can also affect longevity (Swindell, 2007). Mutant strains such as Snell dwarf (defective in anterior pituitary development) (Flurkey et al., 2001) and GH receptor knockout (GHRKO) (Coschigano et al., 2003) are long-lived, due to altered signaling in GH itself or through GH-stimulated production of IGF-1. Human individuals with mutation in the gene coding for GH receptor ("Laron syndrome") exhibit dwarfism but also have strikingly low rates of cancer and diabetes (Wade, 2011).

Three other pathways, TOR/S6K (TOR: target of Rapamycin; S6K: ribosomal S6 kinase), sirtuins, and AMPK (AMP-activated protein kinase), are often termed "nutrient sensing", given their abilities to monitor and respond to the metabolic and energy status of the cells (Figure 1.1C). TOR becomes activated by abundance of amino acids; it in turns actives S6K to promote growth and proliferation (Jewell et al., 2013). On the other hand, during amino acid deprivation, TOR activities decrease and a transcription program of stress resistance is initiated (Gallinetti et al., 2013). Rapamycin, an inhibitor of mTORC1, produces similar effects and leads to 23-26% increase in median lifespan of mice (Harrison et al., 2009; Miller et al., 2014). Sirtuins are a family of enzymes with deacetylase activities that require the splitting of $NAD^+$ (nicotinamide adenine dinucleotide) into nicotinamide and Ac-ADP-ribose during the deacetylation process (Michan and Sinclair, 2007). SIRT1 interacts with and deacetylates a number of proteins to promote cell survival and DNA repair

and reduce inflammation (Haigis and Sinclair, 2010). During aging, the $NAD^+$/NADH ratio in the cell decreases and there is a loss of mitochondrial oxidative phosphorylation system components, but supplement with $NAD^+$ precursor can reverse these aging effects in a SIRT1-dependent manner (Gomes et al., 2013). The anti-aging effects of sirtuins activation have been shown in yeast, flies, worms, and mammals (Kaeberlein et al., 1999; Michan and Sinclair, 2007; Rogina and Helfand, 2004; Tissenbaum and Guarente, 2001). Resveratrol and other sirtuins activators are promising candidates to pharmacologically induce these beneficiary effects (Baur et al., 2006; Hubbard et al., 2013; Wood et al., 2004). AMPK is also an important sensor of energy level: it remains inactive when cellular ATP is abundant, but becomes activated as ATP is consumed and turned into AMP (Burkewitz et al., 2014; Hardie et al., 2012). Activated AMPK up-regulates a number of downstream processes, including mitochondrial biogenesis, beta oxidation, glucose uptake, and autophagy, while inhibiting protein synthesis (Burkewitz et al., 2014). Metformin is a chemical activator of AMPK and has been shown to extend lifespan in *C. elegans* (Onken and Driscoll, 2010).

Signaling through these nutrient sensing pathways likely explains the lifespan extension by dietary restriction (DR), in which an animal's dietary calorie intake is reduced while maintaining normal balance of nutrients. DR is the first method proven effective in extending mammalian longevity, and it has been validated in yeast, worms, flies, and mice (Fontana et al., 2010; McCay et al., 1935; Sinclair, 2005). Animals under DR experience substantial metabolic remodeling, with significant changes in endocrine levels, fat oxidation, reactive oxygen species production and protein turnover, mimicking the responses to mild biological stress (Sinclair, 2005). However, the exact mechanisms of DR are not fully clear: in *Drosophila* it seems the reduction of nutrients, rather than calories, is responsible for the lifespan extension (Mair et al., 2005); reduction of methionine alone is sufficient to produce the longevity effect (Lee et al., 2014); and the benefit of DR can be achieved even when initiated late in life (Mair et al., 2003). The effectiveness of DR in primates is still debated (Colman et al., 2014; Mattison et al., 2012). Nevertheless, numerous lines of evidence have linked DR to each of these nutrient sensing pathways, and there are also cross-talks among the signaling cascade components. It is possible that DR may produce the condition of low amino acids, high $NAD^+$/NADH ratio and high AMP/ATP ratio, thereby impacting all of these pathways (Figure 1.1C). It was recently

demonstrated that mice with only one copy of *Myc* exhibited increased lifespan, reduced serum IGF-1, increased AMPK activity and reduced AKT, TOR and S6K activities (Hofmann et al., 2015), suggesting yet another potential intervention point for lifespan extension.

While many of the lifespan experiments were conducted in metazoa, particualrly fruit flies, worms, and mice, the unicellular budding yeast *Saccharomyces cerevisiae* was particularly useful for early steps in our understanding of aging and longevity. The lifespan of yeast can be defined as either "replicative" (RLS; the number of daughter cells produced by a mother cell before senescence) or "chronological" (CLS; the length of time a yeast cell can survive in a non-dividing state) (Kaeberlein et al., 2007), and several known regulators of aging were discovered using these assays. It was found that reducing glucose or amino acids in culture media can extend both RLS and CLS (Jiang et al., 2000; Reverter-Branchat et al., 2004). Genome-wide screens of single-gene deletion strains for extended lifespan have identified mutations that decrease TOR activity (Kaeberlein et al., 2005; Powers et al., 2006). The role of sirtuins in lifespan regulation was also first demonstrated in yeast, as overexpression of *SIR2* increased RLS by suppressing homologous recombination at rDNA repeats and preventing accumulation of extrachromosomal rDNA circles, whereas the deletion of *SIR2* produced the opposite effects (Kaeberlein et al., 1999; Kennedy et al., 1995; Sinclair and Guarente, 1997). Sir2 orthologs were subsequently found to mediate lifespan extension in both worms and flies (Longo and Kennedy, 2006; Rogina and Helfand, 2004; Tissenbaum and Guarente, 2001).

## LONGEVITY ACROSS DIFFERENT SPECIES

The other area of aging research looks at the lifespan differences across species, which is the subject of this dissertation. Let's consider a simple example: a boy, a cat, and a mouse, all living under the same roof, can be considered under the influence of a similar set of environmental factors. In the absence of diseases and accidents, the boy will almost certainly be around to celebrate his 60th birthday. If he (or the cat) is lucky, he may keep the cat as a pet for 10-15 years. However, it is almost impossible for the mouse to survive beyond 4 years of age. What causes such dramatic difference in lifespan?

If we focus on mammals only, the differences in longevity are already remarkable (Figure 1.2A). All modern mammals descend from a common ancestor that lived ~210 million years ago, yet they exhibit more than 100-fold differences in lifespan and 50 million-fold variation in body weight (Tacutu et al., 2013) (Figure 1.2B). On the one extreme are small and short-lived species: Etruscan shrews (*Suncus etruscus*) weigh ~ 2g and live up to 3.2 years. The other extreme are the large and long-lived beasts: African elephants is the largest land mammal, weighing up to 6 tons and living to 70 years; in the ocean, bowhead whales (*Balaena mysticetus*) can weigh > 100 tons and are estimated to live > 200 years (Tacutu et al., 2013). In general, the longer-lived species also tend to be bigger, produce fewer offspring, grow more slowly, and have lower mass-specific metabolic rates (Peters, 1986; Sacher, 1959; Western, 1979), suggesting these life history traits may be modulated by the same underlying evolutionary forces. In addition, certain lineages have evolved to live longer as a whole: despite their small body sizes (~10-20 g; similar to shrews and small rodents), most bats can live for 10-20 years (compared to < 4 years in shrews and small rodents) (Seim et al., 2013) (Figure 1.2B). In other instances, exceptionally long-lived species have emerged sporadically among short-lived taxonomic relatives: the naked mole rat (*Heterocephalus glaber*) lives ten times longer than other rodents of comparable size (Buffenstein, 2008; Fang et al., 2014; Kim et al., 2011). Humans (*Homo sapiens*) are also exceptionally long-lived: Jeanne Calment of France held the longevity record of 122 years and 164 days (Whitney, 1997), while gorillas, chimpanzees and orangutans can live to ~ 60 years (Tacutu et al., 2013).

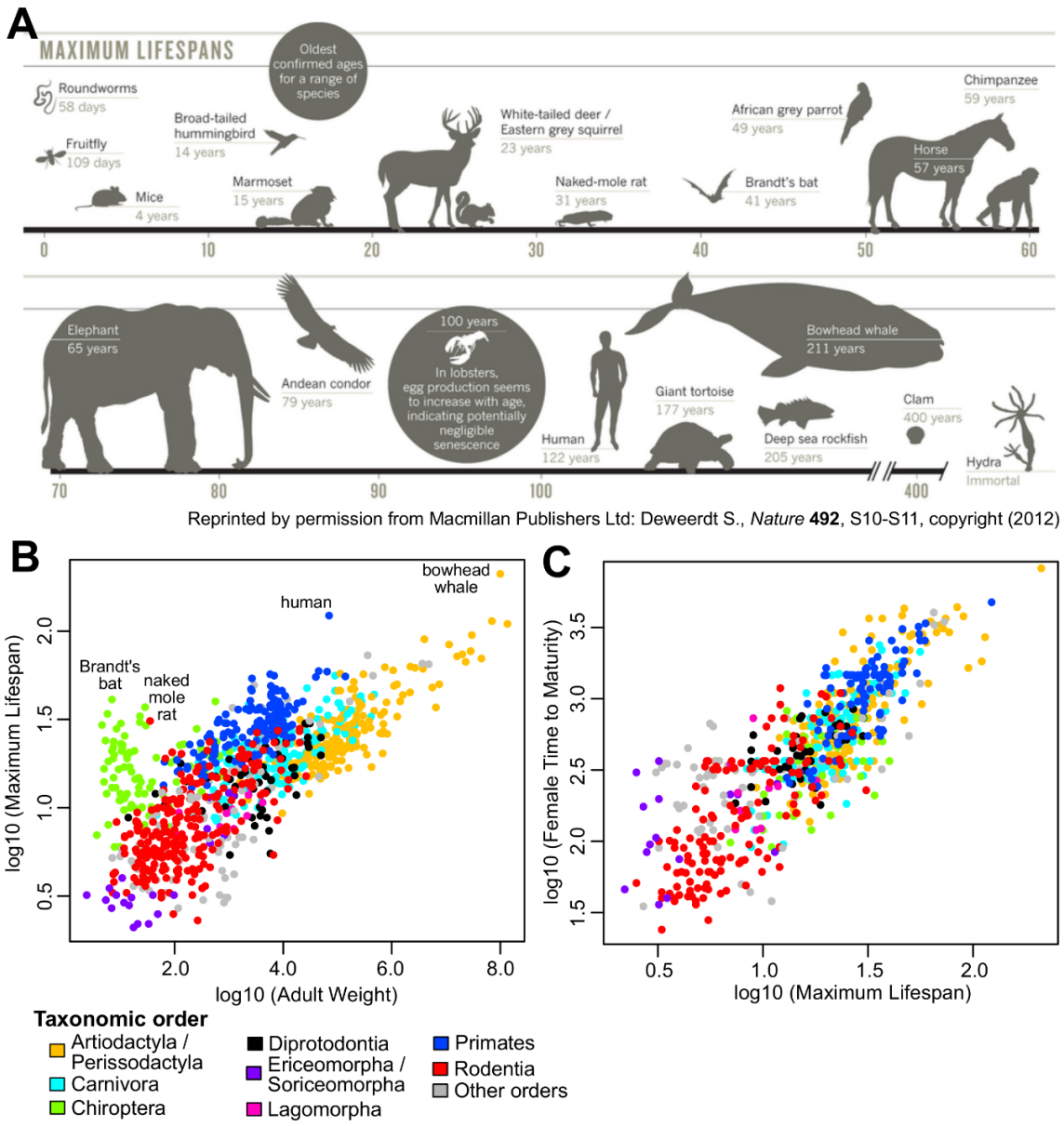



**Figure 1.2. Longevity across different species.**

**(A) Variation in maximum lifespan.** Reprinted by permission from Macmillan Publishers Ltd (Deweerdt, 2012).

**(B) Maximum lifespan correlates positively with adult weight.** The points are colored by taxonomic orders. Selected species were labeled.

**(C) Female Time to Maturity correlates positively with maximum lifespan.** Colored as in panel B.

Life expectancy can be impacted by unnatural causes of death, such as diseases, accidents, homicide, and suicide. Over the last century, the life expectancy at birth in developed countries has increased by 27 years, largely achieved by the improvement in public health care (Hayflick, 2007). Nevertheless, heathcare alone cannot change the biological limit of longevity: even with the best provision of nutrients and care in a laboratory, no mice can survive beyond 5 years. Therefore, when talking about longevity of a species here, we refer to its maximum lifespan or maximum lifespan potential, i.e. the longest time a member of the species has been observed to live, or the longest time it has the potential to live (assuming free from predation and other external causes of death). For most animals this refers to their lifespans in captivity (Holliday, 2006); but for the exceptionally long-lived species, one has to rely on unconventional methods. For example, a male Brandt's bat (*Myotis brandtii*) was recaptured in the wild in Siberia of Russia, 41 years after it was originally banded (Podlutsky et al., 2005). Harpoon points made of ivory and stone were found in three bowhead whales captured in 2007; such weapons were last manufactured in New England in about 1880 (Gardner, 2007). Since maximum lifespan may suffer from reporting bias, as an alternative one can track female time to maturity (i.e. time taken to reach puberty), which may be measured more easily and correlate significantly with maximum lifespan (Pearson correlation 0.83; Figure 1.2C).

One useful framework to view the variation in longevity is the *r*/*K* selection theory, which highlights the links among lifespan, reproduction, and selection pressure (Austad, 1997; MacArthur and Wilson, 1967; Pianka, 1970). Briefly, the *r*-selected species tend to be small in size, have short generation time, and produce many offspring, although each of the offspring has relatively low probability of reaching adulthood. Examples include small rodents, which often occupy the bottom of the food chain and compete by their sheer numbers. On the other hand, *K*-selected species are large in size, mature slowly, produce fewer offspring, but each has high chance of survival and relatively long lifespan. Large animals like humans and whales are *K*-selected species and are usually on the top of the food chain; they win by quality, not quantity. Although the *K*-selected species have high chance of survival, over-reproduction and over-population may cause them to quickly run out of food. Therefore, the optimal strategy choice depends on the evolutionary force and selection pressure, and some experiments suggest that changes in these factors can indeed influence species lifespan. A study

of two Virginia opossum populations separated for ~5000 years reveals that the population on an island with less exposure to predators had greater survivorship, reduced litter sizes, slower acceleration of age-specific mortality, and fewer signs of physiological aging, when compared to the control population living on mainland (Austad, 1993). Experiments on *Drosophila* also showed that in populations selected for late reproduction, the lifespan of flies increased significantly, together with delayed senescence and reduced early fecundity (Luckinbill et al., 1984).

Yet, on the molecular level, the mechanism of lifespan variation across species is still poorly understood. For example, are those pathways involved in the aging regulation within a single species also applicable across different species? Do the naturally long-lived species have a transcriptomic and metabolic state similar to those laboratory animals under DR? Although telomeres tend to shorten with age and stress (Epel et al., 2004) and there have been reports linking telomere length with longevity in human individuals and birds (Cawthon et al., 2003; Heidinger et al., 2012), the lengths of telomere across different species in fact show negative correlation with maximum lifespan (Gomes et al., 2011), e.g. humans have the shortest telomeres, but the longest lifespan among primates (Kakuo et al., 1999). Other authors have quantified a number of biochemical and enzymatic parameters using samples across various species and revealed that the longer-lived ones probably have greater maintenance capacity, suffer less damage, and are more efficient at repair than the shorter-lived ones (Holliday, 1997, 2006) (Table 1.1). However, until recently there have been only very few cross-species comparative studies analyzing the entire mammalian transcriptome or metabolome (Brawand et al., 2011; Fushan et al., 2015; Merkin et al., 2012). Besides the difficulty in obtaining reliable samples, the large-scale quantification of gene expression and metabolites only recently became cost effective, due to advance in high-throughput sequencing and mass spectrometry. In contrast to the studies comparing treatment and control samples of a single species, cross-species analyses bring about special statistical considerations due to the phylogenetic relationship and statistical non-independence. Furthermore, to date there are only < 50 mammalian species with publicly available complete genomes, posing significant challenges on the read alignment of RNA sequencing data without reference genomes.

Another valuable approach has been to compare exceptionally long-lived species with closely related species characterized by more common lifespan, identifying features associated with exceptional longevity. Examples include amino acid changes in Uncoupling Protein 1 (UCP1) and production of high-molecular-mass hyaluronan in the naked mole rat (Kim et al., 2011; Tian et al., 2013); unique sequence changes in IGF1 and GH receptors in Brandt's bat (Seim et al., 2013); gene gain and loss associated with DNA repair, cell-cycle regulation, and cancer, as well as alteration in insulin signaling in the bowhead whale (Keane et al., 2015; Seim et al., 2014); and duplication of the p53 gene in elephants (Abegglen et al., 2015). Again, it is important to ascertain whether these mechanisms are unique characteristics of certain exceptionally long-lived species, or they can also be extended to account for the general lifespan variation (Partridge and Gems, 2002).

**Table 1.1. Comparison of maintenance and repair efficiencies across different mammalian species.** Table is based on (Holliday, 1997).

| Parameters measured | Longevity Correl. | Species | Reference |
|---|---|---|---|
| Longevity of fibroblasts *in vitro* | positive | Mouse, rat, kangaroo rat, mink, rabbit, bat horse, human | (Rohme, 1981) |
| Longevity of erythrocytes *in vitro* | positive | | |
| Poly (ADP-ribose) polymerase activity | positive | Elephant, pig, rabbit, horse, rat, guinea pig, marmoset, sheep, chimpanzee, gorilla, donkey, cattle | (Grube and Bürkle, 1992) |
| Rate of DNA repair | Positive | Mouse, rat, shrew, hamster, rabbit, dog, cat, cow, horse, elephant, corolla, human | (Cortopassi and Wang, 1996) |
| Cross-linking of collagen | negative | Human, bovine, rat | (Yamauchi et al., 1988) |
| Mitochondrial membrane peroxidizability index | negative | Mouse, rat, guinea pig, sheep, dog, pig, cow, horse | (Pamplona et al., 1998) |
| Liver fatty acid peroxidizability index | negative | Mouse, rat, guinea pig, dog, pig, cow, horse | (Pamplona et al., 2000) |
| Rate of telomere shortening | negative | Finch, tree swallow, penguin, tern, Leach's storm-petrel | (Haussmann et al., 2003) |
| Mitochondrial free radicals production | negative | Combining data from a number of studies | (Perez-Campo et al., 1998) |
| Lung glutathione reductase activity | negative | | |
| Brain glutathione peroxidase activity | negative | | |
| Liver catalase activity | negative | | |
| Oxidative damage to DNA | negative | Mouse, rat, monkey, human | (Adelman et al., 1988) |
| Oxidative damage to mitochondrial DNA | negative | Mouse, rat, guinea pigs, rabbit, sheep, pig, cow, horse | (Barja and Herrero, 2000) |
| Susceptibility to protein oxidation | negative | Mouse, rat, rabbit, pig, pigeon | (Agarwal and Sohal, 1996) |
| Capacity to convert benzo(a)pyrene to water-soluble metabolites | negative | Hamster, mouse, rat, guinea pig, rabbit, cow, elephant, human | (Moore and Schwartz, 1978) |

## SCOPE OF DISSERTATION

This dissertation presents the analyses of transcriptomes, metabolomes, proteomes and/or ionomes, across different species of mammals and *Drosophila*, and natural isolates of budding yeast *Saccharomyces cerevisiae*, to identify the molecular patterns and signatures associated with species longevity.

Chapter 2 focuses on the key methods employed in the analyses. It explains the rationale and implementation of phylogenetic regression to account for the phylogenetic relationship of the cross-species data, and the two-step verification procedure to assess data robustness. It also presents the method for identifying ortholog sets and aligning RNA sequencing reads, especially for the species without publicly available genomes. Finally, given that the samples were collected from different species with potentially significant variation in diet, gender, and other parameters, it addresses the issue of data variability.

Chapter 3 presents the metabolome analysis of brain, heart, kidney, and liver across 26 mammalian species. The metabolites significantly enriched or depleted in a particular organ (e.g. low glutamine in kidney) or among a particular phylogenetic lineage of species (e.g. low methionine sulfoxide in liver of bats; the bile acid conjugation strategies among carnivores, herbivores and omnivores) are discussed, in light of the known organ- and species-specific physiologies. Using phylogenetic regression, the metabolites with significant positive or negative correlation to species longevity were identified (e.g. long-lived species had high urate:allantoin ratio but low tryptophan degradation products). The results are compared to the metabolic changes in 5 long-lived mouse models, and certain overlapping metabolic changes (e.g. up-regulation of sphingomyelin and down-regulation of polyunsatureated triacylglycerols) are discussed.

Chapter 4 presents the analysis of 18 metal and non-metal elements among the mammalian species and organs discussed in Chapter 3. Here, the groups of elements with similar distribution patterns (e.g. copper and zinc) were identified. The kidney and liver levels of selenium were also found to correlate with the number of selenocysteine residuals of selenoprotein P sequences of the species. In terms of longevity, liver selenium levels negatively correlated with species lifespan,

whereas kidney and liver cadmium levels showed strong positive correlation, but it raised the interesting question of correlation and causality.

Chapter 5 presents the analysis of transcriptome, proteome, and metabolome of 22 strains of *S. cerevisiae*. The natural isolates with high replicative lifespan up-regulated oxidative phosphorylation and mitochondrial respiratory chains and down-regulated protein targeting and nitrogen compound biosynthesis. Interestingly, the strains differed not by mitochondrial numbers, but by the composition of mitochondria, with the longer-lived strains having higher levels of proteins of the pyruvate dehydrogenase complex, Complex III, and Complex IV, and the shorter-lived strains with higher expression of outer membrane translocases and mitochondrial chaperones.

Chapter 6 examines the gene expression variation across 14 species of *Drosophila*, with mean lifespan ranging from 8 days to ~60 days. It was observed that the longer-lived flies up-regulate the genes in lipid metabolism and down-regulated those in neuronal development. Many genes previously shown to affect lifespan in model organisms also showed similar directions of change with respect to species lifespan.

Chapter 7 presents the analysis of gene expression and metabolites in primary skin fibroblasts of 16 species of mammals (most of which were rodents). Regression against the longevity traits revealed positive correlation with the genes involved in DNA repair and negative correlation with the genes involved in proteolysis.

Similar conclusions were reached in Chapter 8, which profiled the gene expression in brain, kidney, and liver across 42 mammalian species. Chapter 8 also reports up-regulation of ribosomal proteins and down-regulation of tricarboxylic acid cycle; such patterns were also observed in hibernating animals.

Chapter 9 summarizes the main findings across these datasets and discusses a number of lessons learned from these analyses.

# REFERENCES


Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S.*, et al.* (2015). Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA *314*, 1850-1860.

Adelman, R., Saul, R.L., and Ames, B.N. (1988). Oxidative damage to DNA: relation to species metabolic rate and life span. Proc. Natl. Acad. Sci. USA *85*, 2706-2708.

Agarwal, S., and Sohal, R.S. (1996). Relationship between susceptibility to protein oxidation, aging, and maximum life span potential of different species. Exp. Gerontol. *31*, 365-372.

Austad, S.N. (1993). Retarded senescence in an insular population of Virginia opossums (Didelphis virginiana). J. Zool. *229*, 695-708.

Austad, S.N. (1997). Comparative aging and life histories in mammals. Exp. Gerontol. *32*, 23-38.

Barja, G., and Herrero, A. (2000). Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. FASEB J. *14*, 312-318.

Baur, J.A., Pearson, K.J., Price, N.L., Jamieson, H.A., Lerin, C., Kalra, A., Prabhu, V.V., Allard, J.S., Lopez-Lluch, G., Lewis, K.*, et al.* (2006). Resveratrol improves health and survival of mice on a high-calorie diet. Nature *444*, 337-342.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Buffenstein, R. (2008). Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. J. Comp. Physiol. B *178*, 439-445.

Burkewitz, K., Zhang, Y., and Mair, W.B. (2014). AMPK at the nexus of energetics and aging. Cell Metab. *20*, 10-25.

Carter, M.E., and Brunet, A. (2007). FOXO transcription factors. Curr. Biol. *17*, R113-114.

Cawthon, R.M., Smith, K.R., O'Brien, E., Sivatchenko, A., and Kerber, R.A. (2003). Association between telomere length in blood and mortality in people aged 60 years or older. Lancet *361*, 393-395.

Collins, J.J., Huang, C., Hughes, S., and Kornfeld, K. (2008). The measurement and analysis of age-related changes in Caenorhabditis elegans. WormBook, 1-21.

Colman, R.J., Beasley, T.M., Kemnitz, J.W., Johnson, S.C., Weindruch, R., and Anderson, R.M. (2014). Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. Nat. Commun. *5*, 3557.

Cortopassi, G.A., and Wang, E. (1996). There is substantial agreement among interspecies estimates of DNA repair activity. Mech. Ageing Dev. *91*, 211-218.

Coschigano, K.T., Holland, A.N., Riders, M.E., List, E.O., Flyvbjerg, A., and Kopchick, J.J. (2003). Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. Endocrinology *144*, 3799-3810.

Deweerdt, S. (2012). Comparative biology: Looking for a master switch. Nature *492*, S10-11.

Epel, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., and Cawthon, R.M. (2004). Accelerated telomere shortening in response to life stress. Proc. Natl. Acad. Sci. USA *101*, 17312-17315.

Fang, X., Seim, I., Huang, Z., Gerashchenko, M.V., Xiong, Z., Turanov, A.A., Zhu, Y., Lobanov, A.V., Fan, D., Yim, S.H.*, et al.* (2014). Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. Cell Rep. *8*, 1354-1364.

Flurkey, K., Papaconstantinou, J., Miller, R.A., and Harrison, D.E. (2001). Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proc. Natl. Acad. Sci. USA *98*, 6736-6741.

Fontana, L., Partridge, L., and Longo, V.D. (2010). Extending healthy life span--from yeast to humans. Science *328*, 321-326.

Friedman, D.B., and Johnson, T.E. (1988). A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics *118*, 75-86.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Gallinetti, J., Harputlugil, E., and Mitchell, J.R. (2013). Amino acid sensing in dietary-restriction-mediated longevity: roles of signal-transducing kinases GCN2 and TOR. Biochem. J. *449*, 1-10.

Gardner, D. (2007). Whale survives harpoon attack 130 years ago to become 'world's oldest mammal'. In Daily Mail.

Gomes, A.P., Price, N.L., Ling, A.J., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P.*, et al.* (2013). Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. Cell *155*, 1624-1638.

Gomes, N.M., Ryder, O.A., Houck, M.L., Charter, S.J., Walker, W., Forsyth, N.R., Austad, S.N., Venditti, C., Pagel, M., Shay, J.W.*, et al.* (2011). Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. Aging Cell *10*, 761-768.

Grube, K., and Bürkle, A. (1992). Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span. Proc. Natl. Acad. Sci. USA *89*, 11759-11763.

Haigis, M.C., and Sinclair, D.A. (2010). Mammalian sirtuins: biological insights and disease relevance. Annu. Rev. Pathol. *5*, 253-295.

Hardie, D.G., Ross, F.A., and Hawley, S.A. (2012). AMPK: a nutrient and energy sensor that maintains energy homeostasis. Nat. Rev. Mol. Cell Biol. *13*, 251-262.

Harrison, D.E., Strong, R., Sharp, Z.D., Nelson, J.F., Astle, C.M., Flurkey, K., Nadon, N.L., Wilkinson, J.E., Frenkel, K., Carter, C.S.*, et al.* (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature *460*, 392-395.

Haussmann, M.F., Winkler, D.W., O'Reilly, K.M., Huntington, C.E., Nisbet, I.C., and Vleck, C.M. (2003). Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. Proc. Biol. Sci. *270*, 1387-1392.

Hayflick, L. (2007). Biological aging is no longer an unsolved problem. Ann. N. Y. Acad. Sci. *1100*, 1-13.

Heidinger, B.J., Blount, J.D., Boner, W., Griffiths, K., Metcalfe, N.B., and Monaghan, P. (2012). Telomere length in early life predicts lifespan. Proc. Natl. Acad. Sci. USA *109*, 1743-1748.

Hofmann, J.W., Zhao, X., De Cecco, M., Peterson, A.L., Pagliaroli, L., Manivannan, J., Hubbard, G.B., Ikeno, Y., Zhang, Y., Feng, B.*, et al.* (2015). Reduced expression of MYC increases longevity and enhances healthspan. Cell *160*, 477-488.

Holliday, R. (1997). Understanding ageing. Philos. Trans. R. Soc. Lond. B Biol. Sci. *352*, 1793-1797.

Holliday, R. (2006). Aging is no longer an unsolved problem in biology. Ann. N. Y. Acad. Sci. *1067*, 1-9.

Hubbard, B.P., Gomes, A.P., Dai, H., Li, J., Case, A.W., Considine, T., Riera, T.V., Lee, J.E., E, S.Y., Lamming, D.W.*, et al.* (2013). Evidence for a common mechanism of SIRT1 regulation by allosteric activators. Science *339*, 1216-1219.

Iliadi, K.G., and Boulianne, G.L. (2010). Age-related behavioral changes in Drosophila. Ann. N. Y. Acad. Sci. *1197*, 9-18.

Jewell, J.L., Russell, R.C., and Guan, K.L. (2013). Amino acid signalling upstream of mTOR. Nat. Rev. Mol. Cell Biol. *14*, 133-139.

Jiang, J.C., Jaruga, E., Repnevskaya, M.V., and Jazwinski, S.M. (2000). An intervention resembling caloric restriction prolongs life span and retards aging in yeast. FASEB J. *14*, 2135-2137.

Jones, O.R., Scheuerlein, A., Salguero-Gomez, R., Camarda, C.G., Schaible, R., Casper, B.B., Dahlgren, J.P., Ehrlen, J., Garcia, M.B., Menges, E.S.*, et al.* (2014). Diversity of ageing across the tree of life. Nature *505*, 169-173.

Kaeberlein, M., Burtner, C.R., and Kennedy, B.K. (2007). Recent developments in yeast aging. PLoS Genet. *3*, e84.

Kaeberlein, M., McVey, M., and Guarente, L. (1999). The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev. *13*, 2570-2580.

Kaeberlein, M., Powers, R.W., 3rd, Steffen, K.K., Westman, E.A., Hu, D., Dang, N., Kerr, E.O., Kirkland, K.T., Fields, S., and Kennedy, B.K. (2005). Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science *310*, 1193-1196.

Kakuo, S., Asaoka, K., and Ide, T. (1999). Human is a unique species among primates in terms of telomere length. Biochem. Biophys. Res. Commun. *263*, 308-314.

Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I.*, et al.* (2015). Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. *10*, 112-122.

Kennedy, B.K., Austriaco, N.R., Jr., Zhang, J., and Guarente, L. (1995). Mutation in the silencing gene SIR4 can delay aging in S. cerevisiae. Cell *80*, 485-496.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. Nature *366*, 461-464.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P.*, et al.* (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223-227.

Lee, B.C., Kaya, A., Ma, S., Kim, G., Gerashchenko, M.V., Yim, S.H., Hu, Z., Harshman, L.G., and Gladyshev, V.N. (2014). Methionine restriction extends lifespan of Drosophila melanogaster under conditions of low amino-acid status. Nat. Commun. *5*, 3592.

Longo, V.D., and Kennedy, B.K. (2006). Sirtuins in aging and age-related disease. Cell *126*, 257-268.

Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194-1217.

Luckinbill, L.S., Arking, R., Clare, M.J., Cirocco, W.C., and Buck, S.A. (1984). Selection for Delayed Senescence in Drosophila melanogaster. Evolution *38*, 996-1003.

MacArthur, R.H., and Wilson, E.O. (1967). The Theory of Island Biogeography (Princeton, Princeton University Press).

Mair, W., Goymer, P., Pletcher, S.D., and Partridge, L. (2003). Demography of dietary restriction and death in Drosophila. Science *301*, 1731-1733.

Mair, W., Piper, M.D., and Partridge, L. (2005). Calories do not explain extension of life span by dietary restriction in Drosophila. PLoS Biol. *3*, e223.

Mattison, J.A., Roth, G.S., Beasley, T.M., Tilmont, E.M., Handy, A.M., Herbert, R.L., Longo, D.L., Allison, D.B., Young, J.E., Bryant, M.*, et al.* (2012). Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. Nature *489*, 318-321.

McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The Effect of Retarded Growth Upon the Length of Life Span and Upon the Ultimate Body Size: One Figure. J. Nutr. *10*, 63-79.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593-1599.

Michan, S., and Sinclair, D. (2007). Sirtuins in mammals: insights into their biological function. Biochem. J. *404*, 1-13.

Miller, R.A., Harrison, D.E., Astle, C.M., Fernandez, E., Flurkey, K., Han, M., Javors, M.A., Li, X., Nadon, N.L., Nelson, J.F.*, et al.* (2014). Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. Aging Cell *13*, 468-477.

Moore, C.J., and Schwartz, A.G. (1978). Inverse correlation between species lifespan and capacity of cultured fibroblasts to convert benzo(a)pyrene to water-soluble metabolites. Exp. Cell Res. *116*, 359-364.

Onken, B., and Driscoll, M. (2010). Metformin induces a dietary restriction-like state and the oxidative stress response to extend C. elegans Healthspan via AMPK, LKB1, and SKN-1. PLoS One *5*, e8758.

Pamplona, R., Portero-Otin, M., Riba, D., Requena, J.R., Thorpe, S.R., Lopez-Torres, M., and Barja, G. (2000). Low fatty acid unsaturation: a mechanism for lowered lipoperoxidative modification of tissue proteins in mammalian species with long life spans. J. Gerontol. A Biol. Sci. Med. Sci. *55*, B286-291.

Pamplona, R., Portero-Otin, M., Riba, D., Ruiz, C., Prat, J., Bellmunt, M.J., and Barja, G. (1998). Mitochondrial membrane peroxidizability index is inversely related to maximum life span in mammals. J. Lipid Res. *39*, 1989-1994.

Partridge, L., and Gems, D. (2002). Mechanisms of ageing: public or private? Nat. Rev. Genet. *3*, 165-175.

Perez-Campo, R., Lopez-Torres, M., Cadenas, S., Rojas, C., and Barja, G. (1998). The rate of free radical production as a determinant of the rate of aging: evidence from the comparative approach. J. Comp. Physiol. B *168*, 149-158.

Peters, R.H. (1986). The ecological implications of body size, Vol 2 (Cambridge University Press).

Pianka, E.R. (1970). On r- and K-Selection. Am. Nat. *104*, 592-597.

Podlutsky, A.J., Khritankov, A.M., Ovodov, N.D., and Austad, S.N. (2005). A new field record for bat longevity. J. Gerontol. A Biol. Sci. Med. Sci. *60*, 1366-1368.

Powers, R.W., 3rd, Kaeberlein, M., Caldwell, S.D., Kennedy, B.K., and Fields, S. (2006). Extension of chronological life span in yeast by decreased TOR pathway signaling. Genes Dev. *20*, 174-184.

Reverter-Branchat, G., Cabiscol, E., Tamarit, J., and Ros, J. (2004). Oxidative damage to specific proteins in replicative and chronological-aged Saccharomyces cerevisiae: common targets and prevention by calorie restriction. J. Biol. Chem. *279*, 31983-31989.

Rogina, B., and Helfand, S.L. (2004). Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proc. Natl. Acad. Sci. USA *101*, 15998-16003.

Rohme, D. (1981). Evidence for a relationship between longevity of mammalian species and life spans of normal fibroblasts in vitro and erythrocytes in vivo. Proc. Natl. Acad. Sci. USA *78*, 5009-5013.

Sacher, G.A. (1959). Relation of Lifespan to Brain Weight and Body Weight in Mammals. In Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing), G.E.W. Wolstenholme, and M. O'Conner, eds. (Chichester, John Wiley & Sons, Ltd), pp. 115-141.

Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L*., et al.* (2013). Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat. Commun. *4*, 2212.

Seim, I., Ma, S., Zhou, X., Gerashchenko, M.V., Lee, S.G., Suydam, R., George, J.C., Bickham, J.W., and Gladyshev, V.N. (2014). The transcriptome of the bowhead whale Balaena mysticetus reveals adaptations of the longest-lived mammal. Aging (Albany NY) *6*, 879-899.

Sinclair, D.A. (2005). Toward a unified theory of caloric restriction and longevity regulation. Mech. Ageing Dev. *126*, 987-1002.

Sinclair, D.A., and Guarente, L. (1997). Extrachromosomal rDNA circles--a cause of aging in yeast. Cell *91*, 1033-1042.

Swindell, W.R. (2007). Gene expression profiling of long-lived dwarf mice: longevity-associated genes and relationships with diet, gender and aging. BMC Genomics *8*, 353.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-1033.

Taguchi, A., and White, M.F. (2008). Insulin-like signaling, nutrient homeostasis, and life span. Annu. Rev. Physiol. *70*, 191-212.

Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., Mao, Z., Nevo, E., Gorbunova, V., and Seluanov, A. (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. Nature *499*, 346-349.

Tissenbaum, H.A., and Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature *410*, 227-230.

Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C*., et al.* (2002). p53 mutant mice that display early ageing-associated phenotypes. Nature *415*, 45-53.

Wade, N. (2011). Ecuadorean Villagers May Hold Secret to Longevity. In The New York Times.

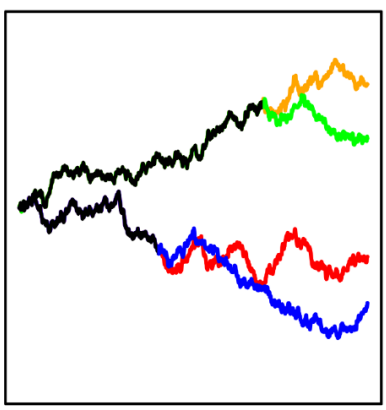
Western, D. (1979). Size, life history and ecology in mammals. Afr. J. Ecol. *17*, 185-204.

Whitney, C.R. (1997). Jeanne Calment, World's Elder, Dies at 122. In The New York Times.

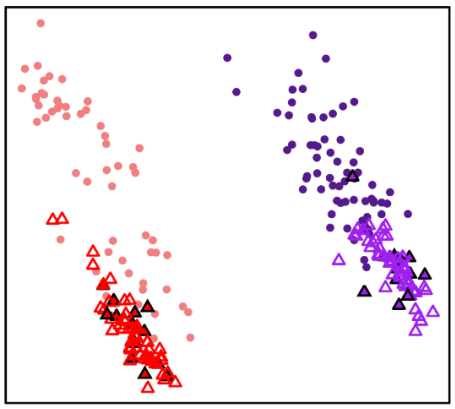
Wood, J.G., Rogina, B., Lavu, S., Howitz, K., Helfand, S.L., Tatar, M., and Sinclair, D. (2004). Sirtuin activators mimic caloric restriction and delay ageing in metazoans. Nature *430*, 686-689.

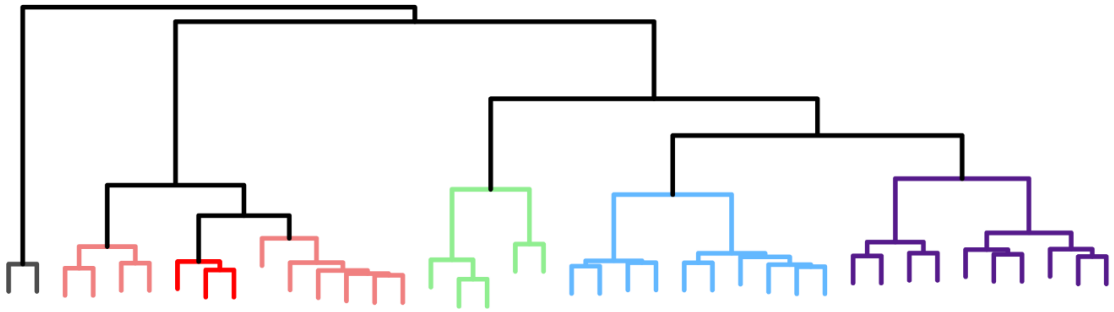
Yamauchi, M., Woodley, D.T., and Mechanic, G.L. (1988). Aging and cross-linking of skin collagen. Biochem. Biophys. Res. Commun. *152*, 898-903.

# Chapter 2 Methods and Data Quality



| | Sp. A | Sp. B | Sp. C | Sp. D |
|---|---|---|---|---|
| Sp. A | 1 | 0.70 | 0 | 0 |
| Sp. B | 0.70 | 1 | 0 | 0 |
| Sp. C | 0 | 0 | 1 | 0.40 |
| Sp. D | 0 | 0 | 0.40 | 1 |





## PHYLOGENETIC REGRESSION

Given two sets of variables, *x* and *y*, their statistical relationship can be estimated using regression analysis. With respect to longevity and aging research, one can perform regression between any molecular measurement (e.g., gene expression or metabolite level) and longevity trait (e.g., maximum lifespan or time to maturity), to identity the genes or metabolites showing statistically significant correlations.

Simple linear regression by ordinary least squares approach is perhaps the simplest and most commonly encountered form of regression in biology research (e.g. in the form of two-sample t-test). For the analysis to be valid, it requires the input data to meet a number of statistical assumptions, some of which concern the regression residual (the difference between the predicted value based on the regression model, and the actual observed value). Specifically, the variances of the residuals are assumed to be equal ("homoscedasticity") and uncorrelated ("independence") (Logan, 2010). While these assumptions are likely true for the random samples from a single species, the picture becomes much more complicated for the samples collected across many different species.

Take as an example a study comparing brain glucose levels among different species of rodents and carnivores. It is likely that the readings from all the rodent species will be similar to one another, and the readings among the carnivores will be also similar. To determine whether there is significant difference between the rodents and carnivores, one needs to consider not only the difference in mean values, but also the variance (or standard deviation). Since all the rodent species are phylogenetically related due to their shared evolutionary history (the same is also true for the carnivores), the data variance is likely to be smaller than if these species were completely unrelated. If one then applies simple t-test (which assumes the data points are independent) without regards to the phylogenetic relationship, the variance will be underestimated and the significance will be overstated.

The solution to this problem has been provided by a number of authors in comparative and evolutionary biology (Butler and King, 2004; Felsenstein, 1985; Freckleton et al., 2002; Garland et al., 1993; Grafen, 1989; Martins and Hansen, 1997; Pagel, 1999). Known as "phylogenetic regression", "phylogenetically independent contrast", or "phylogenetic generalized least squares", this

approach basically incorporates the phylogenetic relationship of the species, modifies the assumption of independence and homoscedasticity, and performs regression analysis using generalized least squares approach. The overall effect is to increase the stringency for statistical significance (i.e. the p values under phylogenetic regression are usually much larger (less significant) than those under simple linear regression), such that many of the candidates considered significant under simple linear regression will no longer meet the cut-off. In other words, the approach requires that a relationship across the species must persist, above and beyond what can be expected based on phylogeny.

The following sections will elaborate on: ordinary least squares (OLS) approach and its assumptions; generalized least squares (GLS) approach and relaxation of OLS assumptions; maximum likelihood (ML) approach and its relation to GLS; different trait evolution models (i.e. different ways to describe the phylogenetic relationship); and a two-step verification procedure to test robustness of results. The methods will be illustrated using the following numerical example of longevity trait and metabolite level across 12 different species (Table 2.1 and Figure 2.1):

**Table 2.1. Hypothetical dataset showing metabolite and trait variations across different species.**

| Species | Metabolite | Trait |
|---|---|---|
| Guinea Pig | 7.525 | -0.682 |
| Porcupine | 8.946 | 0.293 |
| Chinchilla | 8.139 | -0.121 |
| Chipmunk | 8.191 | -0.044 |
| Red Squirrel | 8.744 | -0.024 |
| Fox Squirrel | 9.322 | 0.025 |
| Beaver | 8.803 | -0.022 |
| African Grass Rat | 7.863 | -0.682 |
| Gerbil | 7.294 | -0.652 |
| Meadow Vole | 8.368 | -0.625 |
| Cotton Rat | 7.160 | -0.783 |
| White-footed Mouse | 8.357 | -0.322 |

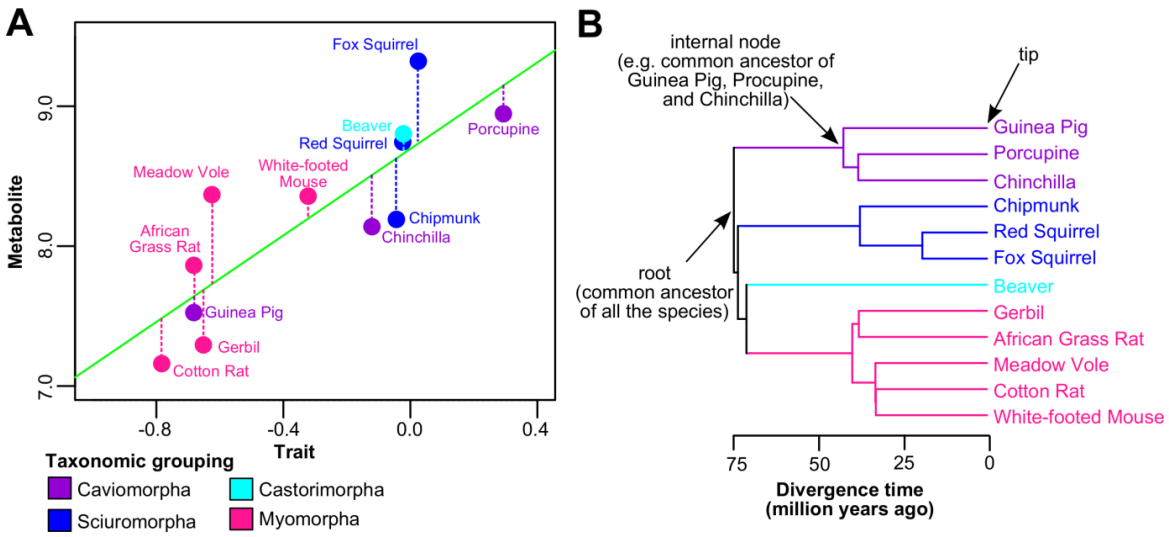



**Figure 2.1. An example for phylogenetic regression.**

**(A) Plot of metabolite against trait.** The data are based on Table 2.1. The species are colored by taxonomic grouping. The green line indicates the OLS solution. The vertical dotted lines indicate the regression residuals.

**(B) Phylogenetic tree of the species.** The root, tip, and internal node of the tree are indicated.

## Ordinary Least Squares approach

For the dependent variable $\boldsymbol{y}$ and independent variable $\boldsymbol{X}$, the general form of the regression model is:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{y}$ is an n×1 matrix (for n species)
- $\boldsymbol{X}$ is an n×2 matrix, containing 1s in the 1st column and the predictor in the 2nd column
- $\boldsymbol{\beta}$ is an 2×1 matrix containing the vertical intercept and slope
- $\boldsymbol{\varepsilon}$ is an n×1 matrix containing the regression residuals, with the assumptions that:
    - $\boldsymbol{\varepsilon}$ follows multivariate normal distribution;
    - $E(\boldsymbol{\varepsilon}) = 0$ (i.e. mean value is 0), and
    - $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$, where $\boldsymbol{I}$ is an identity matrix (i.e. independence and homoscedasticity)

For an estimator $\widehat{\boldsymbol{\beta}}$, the regression residual becomes:

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\widehat{\beta}}$$

And the Residual Sum of Squares (RSS) is:

$$\boldsymbol{\varepsilon^T\varepsilon} = \left(\boldsymbol{y} - \boldsymbol{X\widehat{\beta}}\right)^T\left(\boldsymbol{y} - \boldsymbol{X\widehat{\beta}}\right)$$

The OLS solution $\widehat{\boldsymbol{\beta}}$ is one that minimizes RSS:

$$\frac{\partial\left(\boldsymbol{\varepsilon^T\varepsilon}\right)}{\partial\widehat{\boldsymbol{\beta}}} = 0$$

$$\frac{\partial}{\partial\widehat{\boldsymbol{\beta}}}\left(\boldsymbol{y^Ty} - \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^Ty} - \boldsymbol{y^TX\widehat{\beta}} + \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^TX\widehat{\beta}}\right) = 0$$

$$-2\boldsymbol{X^Ty} + 2\boldsymbol{X^TX}\,\widehat{\boldsymbol{\beta}} = 0$$

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X^TX}\right)^{-1}\boldsymbol{X^Ty}$$

Hence $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X^TX}\right)^{-1}\boldsymbol{X^Ty}$ is the OLS solution. The statistics of $\widehat{\boldsymbol{\beta}}$ is given in Table 2.2.

**Table 2.2. Regression by ordinary least squares (OLS) approach.**

<table>
<tr>
<td>General form:<br>$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$<br>$\boldsymbol{\varepsilon}$ is an n×1 matrix containing residuals in the model and assuming:<br>$\boldsymbol{\varepsilon}$ follows multivariate normal distribution;<br>$E(\boldsymbol{\varepsilon}) = 0$, and<br>$Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$, where $\boldsymbol{I}$ is an identity matrix</td>
<td>Numerical example: $\boldsymbol{y}$ is metabolite level; $\boldsymbol{X}$ is trait<br>$$\boldsymbol{y} = \begin{bmatrix} 7.525 \\ 8.946 \\ 8.139 \\ 8.191 \\ 8.744 \\ 9.322 \\ 8.803 \\ 7.863 \\ 7.294 \\ 8.368 \\ 7.160 \\ 8.357 \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & -0.682 \\ 1 & 0.293 \\ 1 & -0.121 \\ 1 & -0.044 \\ 1 & -0.024 \\ 1 & 0.025 \\ 1 & -0.022 \\ 1 & -0.682 \\ 1 & -0.652 \\ 1 & -0.625 \\ 1 & -0.783 \\ 1 & -0.322 \end{bmatrix}, \quad Var(\boldsymbol{\varepsilon}) = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$</td>
</tr>
<tr>
<td>Formula to find predictor:<br>$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y}$$</td>
<td>$$(\boldsymbol{X^T X})^{-1} = \begin{bmatrix} 0.146 & 0.207 \\ 0.207 & 0.683 \end{bmatrix}$$<br>$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 8.695 \\ 1.546 \end{bmatrix}$$</td>
</tr>
<tr>
<td>Variance:<br>$$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X^T X})^{-1} = \frac{RSS}{d.f.}(\boldsymbol{X^T X})^{-1} = \frac{\boldsymbol{\varepsilon^T \varepsilon}}{n-2}(\boldsymbol{X^T X})^{-1}$$<br>where<br>• RSS: residual sum of squares<br>• d.f.: degree of freedom<br>• $\sigma^2$ is estimated by RSS/d.f.<br>Standard error (S.E.):<br>$$S.E.(\widehat{\boldsymbol{\beta}}) = \sqrt{Var(\widehat{\boldsymbol{\beta}})}$$<br>Statistics:<br>$$t\ value = \frac{\widehat{\boldsymbol{\beta}}}{S.E.(\widehat{\boldsymbol{\beta}})}$$</td>
<td>$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 7.525 \\ 8.946 \\ 8.139 \\ 8.191 \\ 8.744 \\ 9.322 \\ 8.803 \\ 7.863 \\ 7.294 \\ 8.368 \\ 7.160 \\ 8.357 \end{bmatrix} - \begin{bmatrix} 1 & -0.682 \\ 1 & 0.293 \\ 1 & -0.121 \\ 1 & -0.044 \\ 1 & -0.024 \\ 1 & 0.025 \\ 1 & -0.022 \\ 1 & -0.682 \\ 1 & -0.652 \\ 1 & -0.625 \\ 1 & -0.783 \\ 1 & -0.322 \end{bmatrix} \begin{bmatrix} 8.695 \\ 1.546 \end{bmatrix} = \begin{bmatrix} -0.115 \\ -0.202 \\ -0.369 \\ -0.436 \\ 0.086 \\ 0.589 \\ 0.142 \\ 0.223 \\ -0.393 \\ 0.639 \\ -0.324 \\ 0.160 \end{bmatrix}$$<br>$$Var(\widehat{\boldsymbol{\beta}}) = \frac{1.497}{12-2}\begin{bmatrix} 0.146 & 0.207 \\ 0.207 & 0.683 \end{bmatrix} = \begin{bmatrix} 0.0219 & 0.0310 \\ 0.0310 & 0.1023 \end{bmatrix}$$<br>S.E. of intercept: $\sqrt{0.0219} = 0.148$<br>S.E. of slope: $\sqrt{0.1023} = 0.320$<br>Intercept: t value = 8.695/0.148 = 58.75, p value = $4.9\times10^{-14}$<br>Slope: t value = 1.546/0.320 = 4.831, p value = $6.9\times10^{-4}$</td>
</tr>
</table>

## Generalized Least Squares approach

Let's still consider the regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ and keep the assumptions that $\boldsymbol{\varepsilon}$ follows multivariate normal distribution and $E(\boldsymbol{\varepsilon}) = 0$, but now let $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$, such that:

- The diagonal elements of $\boldsymbol{\Sigma}$ may be different (heteroscedasticity)
- The off-diagonal elements of $\boldsymbol{\Sigma}$ may be non-zero (autocorrelation)

If we can find a matrix $\boldsymbol{W}$, such that:

$$\boldsymbol{W^T W = \Sigma^{-1}}$$

Then, the regression model can be transformed to:

$$(\boldsymbol{Wy}) = (\boldsymbol{WX})\boldsymbol{\beta} + (\boldsymbol{W\varepsilon})$$

Making use of the results $Var(\boldsymbol{X}) = E\left[\big(\boldsymbol{X} - E(\boldsymbol{X})\big)\big(\boldsymbol{X} - E(\boldsymbol{X})\big)^T\right]$; $E(\boldsymbol{\varepsilon}) = 0$; and $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$:

$$\begin{aligned}
&Var(\boldsymbol{W\varepsilon}) \\
&= E\left[\big(\boldsymbol{W\varepsilon} - E(\boldsymbol{W\varepsilon})\big)\big(\boldsymbol{W\varepsilon} - E(\boldsymbol{W\varepsilon})\big)^T\right] \\
&= E[(\boldsymbol{W\varepsilon})(\boldsymbol{W\varepsilon})^T] \\
&= E[\boldsymbol{W\varepsilon\varepsilon^T W^T}] \\
&= \boldsymbol{W}E[\boldsymbol{\varepsilon\varepsilon^T}]\boldsymbol{W^T} \\
&= \boldsymbol{W}Var(\boldsymbol{\varepsilon})\boldsymbol{W^T} \\
&= \boldsymbol{W}\sigma^2\boldsymbol{\Sigma W^T} \\
&= \sigma^2\boldsymbol{I}
\end{aligned}$$

In other words, $(\boldsymbol{Wy}) = (\boldsymbol{WX})\boldsymbol{\beta} + (\boldsymbol{W\varepsilon})$ satisfies the assumption of OLS and can be solved using the formula:

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}} \\
&= \big((\boldsymbol{WX})^T(\boldsymbol{WX})\big)^{-1}(\boldsymbol{WX})^T(\boldsymbol{Wy}) \\
&= \big(\boldsymbol{X^T W^T WX}\big)^{-1}\boldsymbol{X^T W^T Wy} \\
&= \big(\boldsymbol{X^T}\big(\boldsymbol{W^T W}\big)\boldsymbol{X}\big)^{-1}\boldsymbol{X}^T(\boldsymbol{W}^T\boldsymbol{W})\boldsymbol{y} \\
&= (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}
\end{aligned}$$

By analogy, the RSS is given by:

$$(\boldsymbol{W\varepsilon})^T(\boldsymbol{W\varepsilon}) = \boldsymbol{\varepsilon}^T\boldsymbol{W}^T\boldsymbol{W\varepsilon} = \boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$$

The GLS solution is presented in Table 2.3, assuming the matrix $\boldsymbol{\Sigma}$ is already known (based on Brownian Motion model, Figure 2.1B).

**Table 2.3. Regression by generalized least squares (GLS) approach.**

<table>
<tr>
<td>General form:<br>$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$<br><br>$\boldsymbol{\varepsilon}$ is an n×1 matrix containing residuals in the model and assuming:<br><br>$\boldsymbol{\varepsilon}$ follows multivariate normal distribution;<br><br>$E(\boldsymbol{\varepsilon}) = 0$, and<br><br>$Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$,</td>
<td>Numerical example: Y is metabolite level; X is trait<br>$\boldsymbol{y} = \begin{bmatrix} 7.525 \\ 8.946 \\ 8.139 \\ 8.191 \\ 8.744 \\ 9.322 \\ 8.803 \\ 7.863 \\ 7.294 \\ 8.368 \\ 7.160 \\ 8.357 \end{bmatrix}$, $\boldsymbol{X} = \begin{bmatrix} 1 & -0.682 \\ 1 & 0.293 \\ 1 & -0.121 \\ 1 & -0.044 \\ 1 & -0.024 \\ 1 & 0.025 \\ 1 & -0.022 \\ 1 & -0.682 \\ 1 & -0.652 \\ 1 & -0.625 \\ 1 & -0.783 \\ 1 & -0.322 \end{bmatrix}$,<br>$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.43 & 0.43 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.43 & 1 & 0.49 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.43 & 0.49 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 & 0.5 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0.5 & 1 & 0.74 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0.5 & 0.74 & 1 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 0.05 & 1 & 0.49 & 0.47 & 0.47 & 0.47 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 0.05 & 0.49 & 1 & 0.47 & 0.47 & 0.47 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 0.05 & 0.47 & 0.47 & 1 & 0.56 & 0.56 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 0.05 & 0.47 & 0.47 & 0.56 & 1 & 0.56 \\ 0 & 0 & 0 & 0.02 & 0.02 & 0.02 & 0.05 & 0.47 & 0.47 & 0.56 & 0.56 & 1 \end{bmatrix}$</td>
</tr>
<tr>
<td>Formula to find predictor:<br>$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$</td>
<td>$(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} = \begin{bmatrix} 0.225 & 0.163 \\ 0.163 & 0.670 \end{bmatrix}$<br>$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 8.700 \\ 1.676 \end{bmatrix}$</td>
</tr>
<tr>
<td>Variance:<br>$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X^T\Sigma^{-1}X})^{-1} = \frac{RSS}{d.f.}(\boldsymbol{X^T\Sigma^{-1}X})^{-1} = \frac{\boldsymbol{\varepsilon^T\Sigma^{-1}\varepsilon}}{n-2}(\boldsymbol{X^T\Sigma^{-1}X})^{-1}$<br>where<br>• RSS: residual sum of squares<br>• d.f.: degree of freedom<br>• $\sigma^2$ is estimated by RSS/d.f.<br><br>Standard error (S.E.):<br>$S.E.(\widehat{\boldsymbol{\beta}}) = \sqrt{Var(\widehat{\boldsymbol{\beta}})}$<br><br>Statistics:<br>$t\ value = \frac{\widehat{\boldsymbol{\beta}}}{S.E.(\widehat{\boldsymbol{\beta}})}$</td>
<td>$\boldsymbol{\varepsilon} = \begin{bmatrix} 7.525 \\ 8.946 \\ 8.139 \\ 8.191 \\ 8.744 \\ 9.322 \\ 8.803 \\ 7.863 \\ 7.294 \\ 8.368 \\ 7.160 \\ 8.357 \end{bmatrix} - \begin{bmatrix} 1 & -0.682 \\ 1 & 0.293 \\ 1 & -0.121 \\ 1 & -0.044 \\ 1 & -0.024 \\ 1 & 0.025 \\ 1 & -0.022 \\ 1 & -0.682 \\ 1 & -0.652 \\ 1 & -0.625 \\ 1 & -0.783 \\ 1 & -0.322 \end{bmatrix} \begin{bmatrix} 8.700 \\ 1.676 \end{bmatrix} = \begin{bmatrix} -0.032 \\ -0.245 \\ -0.358 \\ -0.435 \\ 0.084 \\ 0.580 \\ 0.140 \\ 0.306 \\ -0.313 \\ 0.716 \\ -0.227 \\ 0.197 \end{bmatrix}$<br>$Var(\widehat{\boldsymbol{\beta}}) = \frac{2.852}{12-2}\begin{bmatrix} 0.225 & 0.163 \\ 0.163 & 0.670 \end{bmatrix} = \begin{bmatrix} 0.0642 & 0.0464 \\ 0.0464 & 0.1911 \end{bmatrix}$<br><br>S.E. of intercept: $\sqrt{0.0642} = 0.253$<br>S.E. of slope: $\sqrt{0.1911} = 0.437$<br><br>Intercept: t value = 8.700/0.253 = 34.35, p value = $1.0\times10^{-11}$<br>Slope: t value = 1.676/0.437 = 3.83, p value = $3.3\times10^{-3}$</td>
</tr>
</table>

## Maximum Likelihood approach

In addition to the OLS and GLS approaches, the regression parameter can also be estimated using Maximum Likelihood (ML). For example, suppose *x* is a univariate normal distribution with known mean $\mu$ and variance $\sigma^2$, i.e. $x \sim N(\mu, \sigma^2)$, so the probability density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In other words, if we take *n* random samples from the distribution *x* with known mean and variance, then the frequency distribution of these *n* values is described by the above equation. The ML approach can be considered the reverse process: given *n* values taken randomly from the distribution *x*, we can estimate the mean and variance of such distribution. Mathematically, the likelihood function for samples ($x_1, x_2, x_3, \ldots, x_n$) is defined as:

$$L(\mu, \sigma^2 | x) = \prod_{r=1}^{n} f(x_r | \mu, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{r=1}^{n}(x_r-\mu)^2}$$

which is often presented in the log form:

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{r=1}^{n}(x_r - \mu)^2$$

The parameters $\mu$ and $\sigma^2$ can be found iteratively, such that the likelihood *L* (or log-likelihood *ln(L)*) is at maximum.

When this result is extended to k-dimensional, multivariate normal distribution $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and variance $\sigma^2\boldsymbol{\Sigma}$, i.e. $\boldsymbol{x} \sim N_k(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$, the probability density function is given by:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k \sigma^2 \boldsymbol{\Sigma}}} e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})}{2\sigma^2\boldsymbol{\Sigma}}} = (2\pi)^{-\frac{k}{2}} |\sigma^2\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

For *n* random samples, the likelihood function and log-likelihood function are:

$$L(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma} | \boldsymbol{x}) = \prod_{r=1}^{n} f(\boldsymbol{x_r} | \boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}) = (2\pi)^{-\frac{nk}{2}} |\sigma^2\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{r=1}^{n}(\boldsymbol{x_r}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_r}-\boldsymbol{\mu})}$$

$$\ln(L) = -\frac{nk}{2}\ln(2\pi) - \frac{n}{2}\ln(|\sigma^2\boldsymbol{\Sigma}|) - \frac{1}{2\sigma^2}\sum_{r=1}^{n}(\boldsymbol{x_r} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_r} - \boldsymbol{\mu})$$

Recall that for the regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, the residual $\boldsymbol{\varepsilon}$ is assumed to follow multivariate normal distribution with mean $E(\boldsymbol{\varepsilon}) = 0$ and variance $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$. Given $\boldsymbol{\varepsilon}$ is an n×1 matrix, the log-likelihood is:

$$\begin{aligned}
&\ln(L) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(|\sigma^2\boldsymbol{\Sigma}|) - \frac{1}{2\sigma^2}\,\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(|\sigma^2\boldsymbol{\Sigma}|) - \frac{1}{2}(RSS)
\end{aligned}$$

When $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ (the GLS solution), RSS is at minimum, i.e. *ln(L)* is at maximum. In other words, the GLS (or OLS, in case of homoscedasticity and independence) approach and the ML approach is equivalent.

## Different trait evolution models

So far, we have assumed that the matrix $\boldsymbol{\Sigma}$ in the variance $Var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$ is already known. The matrix $\boldsymbol{\Sigma}$ is sometimes called the variance-covariance (VCV) matrix, in which the diagonal elements describe the variance of each sample, and the off-diagonal elements describe the covariance between two selected samples. In the case of OLS, the matrix $\boldsymbol{\Sigma}$ is an identity matrix, such that the diagonal elements are all 1s (and scaled by $\sigma^2$ to give equal variance for the residuals, i.e. homoscedasticity) and the off-diagonal elements are all 0s (i.e. independence among the residuals). In the case of GLS, the diagonal elements can be unequal and the off-diagonal elements can be non-zeros. This section will focus first on the situation of non-zero off-diagonals.

The evolutionary relationship of any given number of species can be represented using a phylogenetic tree (Figure 2.1B). The tips of the tree showing the species names denote the present states of the species, whereas the root of the tree denotes the state of the common ancestor of all the species in the distant past. The internal nodes represent speciation events when two branches start to diverge, and the branch lengths are proportional the amount of changes (Baum, 2008). Commonly, the tree can be constructed based on nucleotide or protein sequence variations across the species, and calibrated using fossils to represent the estimated time of divergence.

### The Basics: Brownian Motion model

The evolution of a trait among species can be modeled by a Brownian Motion (BM) process (Felsenstein, 1985; Revell, 2010; Revell et al., 2008). Consider a simple example of 4 species A, B, C, and D (Figure 2.2). Setting the relative time as 0.0 at the root and 1.0 at the tip, the phylogenetic tree illustrates three speciation events: at relative time 0.0, the ancestor of Species A and B splits from the ancestor of Species C and D; at relative time 0.4, Species C and D split apart; and at relative time 0.7, Species A and B split apart (Figure 2.2A). In other words, between relative time 0.0 and 0.4 Species C and D evolve along the identical path, and between relative time 0.0 and 0.7 Species A and B evolve along the identical path (Figure 2.2B, the black line). Only after the respective speciation events then they start to evolve independently.

Under the BM model, the amount of changes in a trait is proportional to time. If the rate of change (per unit time) in the trait is $\sigma^2$ and is constant over the entire tree, then the variance between Species A and the root of the tree (i.e. the common ancestor of all 3 species) is $Var(A) = \sigma^2 \times 1.0$, where 1.0 is the relative time between the root and tip. Similarly, $Var(B) = \sigma^2 \times 1.0$, $Var(C) = \sigma^2 \times 1.0$ and $Var(D) = \sigma^2 \times 1.0$ for Species B, C and D, respectively. In addition, since Species A and B evolve together between relative time 0.0 and 0.7, their covariance is $Cov(A, B) = \sigma^2 \times 0.7$. Similarly, the covariance between Species C and D is $Cov(C, D) = \sigma^2 \times 0.4$. All the information can be summarized using a VCV matrix $\boldsymbol{\Sigma}$, where the diagonal is the relative time between the root and tip of the tree, and the off-diagonal is the relative time between the root and the last common ancestor of the two species (Figure 2.2C). Hence, variance in trait $\boldsymbol{\varepsilon}$ is $Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma}$ and can be solved by GLS.

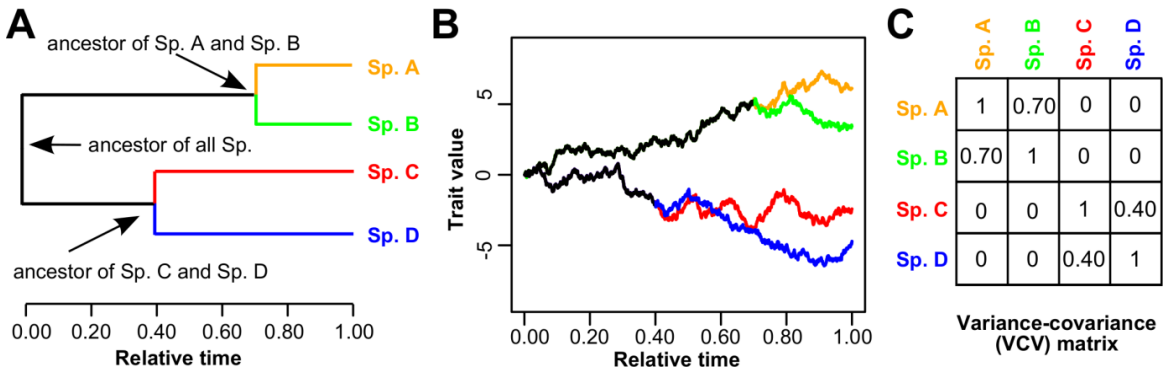



**Figure 2.2. Brownian Motion (BM) model of trait evolution.**

**(A) Phylogenetic tree showing the relative time of evolution.**

**(B) Simulation of trait evolution.** The common ancestor of all 4 species is assumed to have a trait value 0 (at relative time 0.0). The variance $\boldsymbol{\sigma^2}$ is assumed to be 0.02. The trait was simulated over 1000 time points.

**(C) The variance-covariance (VCV) matrix for the species.**

**Variation: Null model and Lambda model**

In BM model, we assume that the phylogenetic tree fully and accurately reflects the evolution of trait. To relax this assumption, a number of ways have been proposed to transform the phylogenetic tree and VCV, one of the most common being "Pagel's lambda" (Pagel, 1999). In the Lambda model, the off-diagonal elements of the VCV are scaled by a factor lambda that ranges between 0 and 1 (Figure 2.3). There are two interesting situations. When lambda = 1, it is the same as BM model (Figure 2.2C). When lambda = 0, all the off-diagonals are 0 and the diagonals are 1, i.e. it becomes an identity matrix and variance in trait $\boldsymbol{\varepsilon}$ is $Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$, satisfying the OLS assumptions; this is also called the Null model. In other words, the Null model assumes that the species are effectively independent of one another.

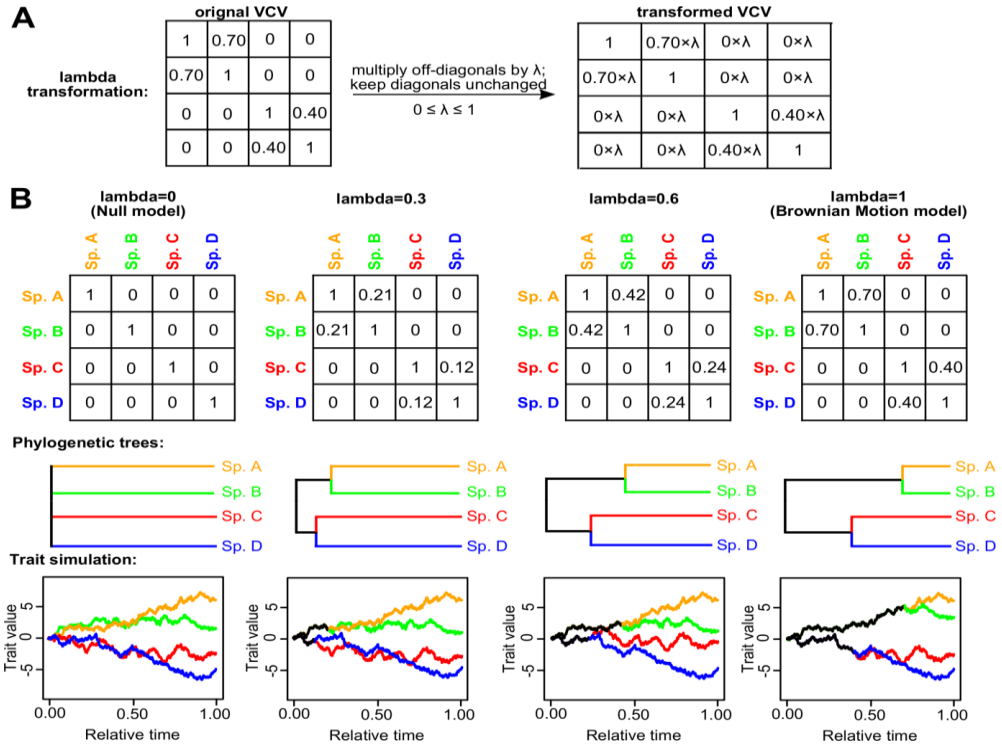



**Figure 2.3. Lambda model of trait evolution.**

**(A) Lambda transformation of variance-covariance (VCV) matrix.** The original VCV is based on Brownian Motion model in Figure 2.2.

**(B) The effects of different values of lambda.** From left to right: the lambda value is 0, 0.3, 0.6 and 1, and the corresponding phylogenetic trees and trait simulation are shown below. When lambda=0, it becomes the "Null" model.

## Variation: Ornstein-Uhlenbeck model

Under the BM model, the differences in a trait among the species simply increase with divergence time. On the other hand, for certain traits (e.g. body temperature) there may be physical limits and boundaries, or there may be an optimal value. The Ornstein-Uhlenbeck (OU) model (Butler and King, 2004; Martins and Hansen, 1997) builds on the BM model, by adding the selection strength parameter α and the optimal trait value θ, such that the trait values of the species will be pulled towards θ. The effects of OU transformation on the VCV and trait evolution are illustrated in Figure 2.4. Note that when α=0, it becomes the BM model.

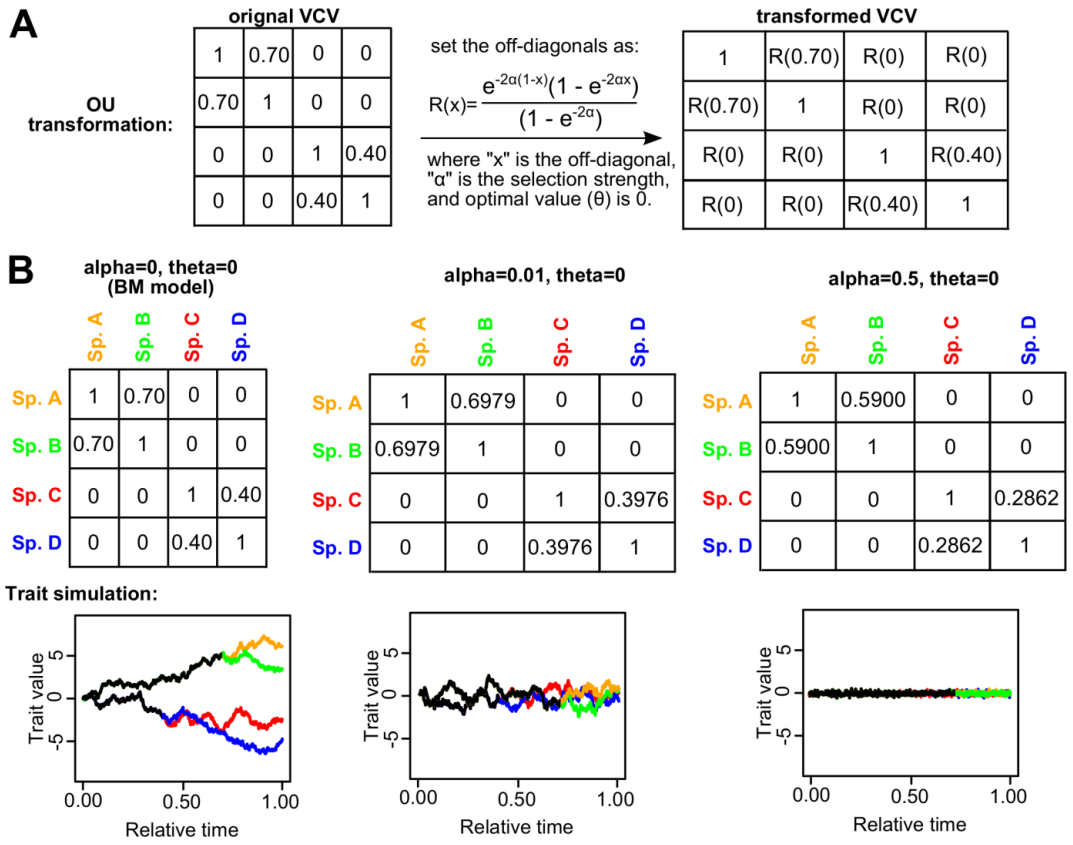



**Figure 2.4. Ornstein-Uhlenbeck (OU) model of trait evolution.**

**(A) OU transformation of variance-covariance (VCV) matrix.** The original VCV is based on Brownian Motion model in Figure 2.2.

**(B) The effects of different values of alpha.** The optimal trait value (theta) is set as 0 in all cases. When alpha=0, it becomes the BM model.

**Selecting the best model**

Recall that for the hypothetical dataset (Figure 2.1), the regression of metabolite level against longevity trait across 12 different species can be solved using GLS (Table 2.3), or if the species are assumed to be independent of one another, using OLS (Table 2.2). In addition, the maximum likelihood (ML) approach can also obtain the same results.

Both GLS and ML require 3 inputs: the independent variable (e.g. longevity trait), the dependent variable (e.g. metabolite level), and the VCV. The form of the VCV depends on the assumed model of trait evolution, and here four models (Null, BM, Lambda, and OU) are presented. Different VCV can be generated by changing the model parameters (e.g. lambda or alpha), and then be used as input for GLS or ML to obtain the regression slope and p value. The ML approach has an additional advantage, that it calculates explicitly the likelihood under each set of parameters (i.e. the goodness of fit of data), and the parameters are estimated at the same time as the regression statistics. The best model will then be the one with the largest likelihood (Lavin et al., 2008). This approach also means that one needs not make any *a priori* assumption on the mode of trait evolution: if the data do suggest that the input data are independent and no phylogenetic correction is needed, then the ML approach should simply return the Null model as the best model.

Table 2.4 illustrates the model selection using the hypothetical dataset. Since the OU model produces the largest likelihood (i.e. least negative log likelihood), it is considered the best model and its regression slope and p value will be reported. Note that in this case, the BM model is considered inferior to the Null model based on maximum likelihood, and the Lambda model returns an estimated lambda = 0 (i.e. effectively the Null model).

**Table 2.4. The best model for the hypothetical dataset.** See Table 2.1 for the values of the longevity trait and metabolite level. See Table 2.3 for the variance-covariance (VCV) matrix of the phylogenetic tree.

| Model | Log likelihood | Slope coefficient | Slope p value | Estimated Parameter(s) |
|---|---|---|---|---|
| Null | -4.540 | 1.546 | $6.887\times10^{-4}$ | -- |
| BM | -6.697 | 1.676 | $3.299\times10^{-3}$ | -- |
| Lambda | -4.540 | 1.546 | $6.887\times10^{-4}$ | Lambda = 0 |
| OU | -4.534 | 1.543 | $7.110\times10^{-4}$ | Alpha = 11.2 |

## Two-step verification procedure for robustness

Since the regression may be influenced by outlier values, a two-step verification procedure is developed to assess the robustness of the results. In the first step, the point with the largest residual is identified and removed, and regression is performed using the remaining points (Figure 2.5A). This ensures the regression line will not be skewed by a single outlier, and the resulting p value is reported as "p value.robust". In the second step, regression is repeated by removing each of the remaining points (already excluding the outlier point in the first step), one at a time, and the largest (i.e. least significant) p value is reported as "p value.max" (Figure 2.5B). This ensures the relationship is generalizable and does not depend on any single species.

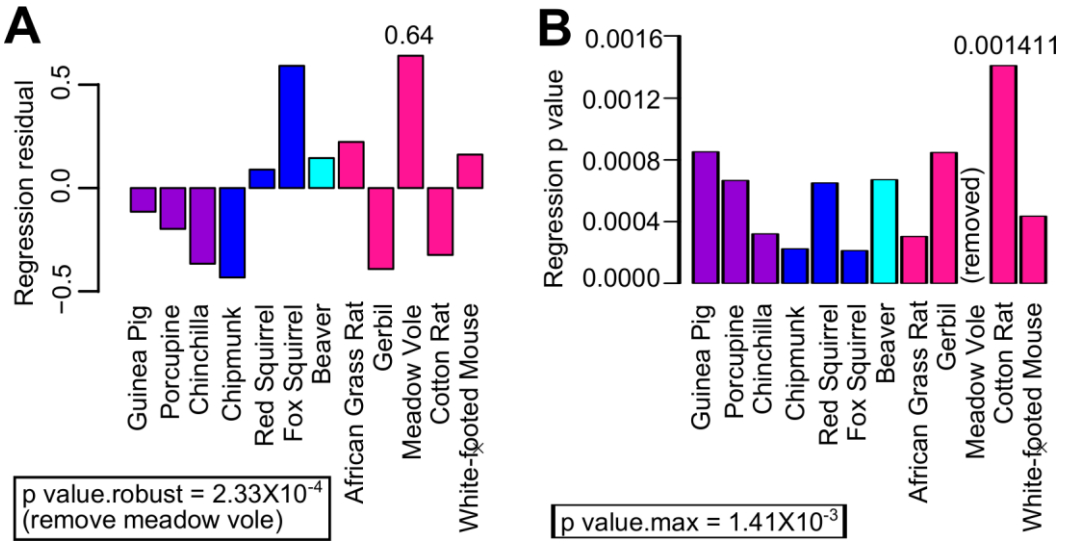



**Figure 2.5. Two-step verification procedure.**

**(A) First step: remove the point with largest residual.** Meadow vole has the largest residual (0.64) and is removed. Regression using the remaining species gives p value = $2.33\times10^{-4}$ (i.e. p value.robust).

**(B) Second step: remove each species, one at a time.** Removal of cotton rat gives the largest (least significant) p value = $1.41\times10^{-3}$ (i.e. p value.max). Meadow vole is already removed after the first step.

# SPECIES WITHOUT REFERENCE GENOMES

## RNA sequencing, read alignment and read counting

The general workflow for quantifying gene expression by RNA sequencing is illustrated in Figure 2.6. RNA molecules are isolated from biological samples and fragmented into short sequences. Using reverse transcription, these fragments are converted into a sequence library and the nucleotide sequences can be read by a sequencing machine. If the reference genome for the species is publicly available, the short reads can be aligned to the reference genome (e.g. using alignment software STAR (Dobin et al., 2013)) and the number of aligned reads can then be counted (e.g. using featureCounts function of Subread (Liao et al., 2014)) to represent the gene expression values.

If the species has no complete genome, then the transcriptome may be *de novo* assembled from the short reads (e.g. using software Trinity (Grabherr et al., 2011)). By relying on the overlapping sequences among the short reads, these reads may be "stitched" back together to derive the original RNA sequences. These assembled sequences can then be used for read alignment and read counting. However, for those genes with low expression, there may not be sufficient reads to regenerate the complete sequences.

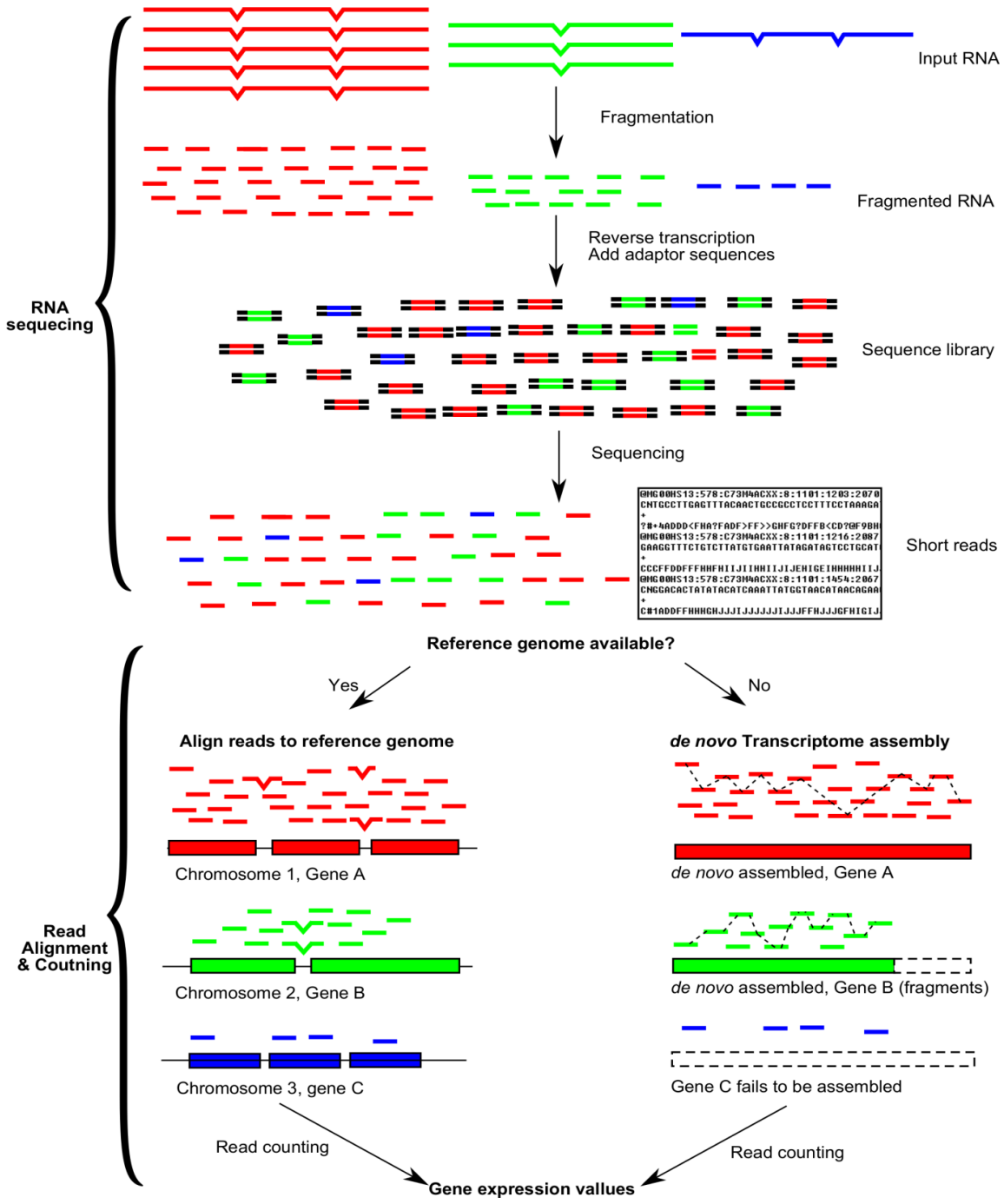



**Figure 2.6. Schematic workflow of gene expression analysis by RNA sequencing.** When the reference genome is available, the reads can be aligned to the reference genome directly. When the reference is not available, the transcriptomes can be assembled *de novo* and then used for read alignment. The assembly of transcriptomes replies on "stitching" together the short reads with overlapping sequences. For modestly expressed genes, a part of or the entire transcript sequences may fail to be assembled.

## Identification of gene orthologs across species

For studies involving samples from the same species, it is relatively straightforward to identify the differentially expressed genes. On the other hand, for gene expression comparison across different species, several complications can arise. First is the adjustment for the phylogenetic relationship of the species, as discussed above. Second is the identification of gene orthologs (i.e. genes in different species that evolved from a common ancestral gene by speciation). In order to make meaningful comparison of the read counts of a particular gene, one needs to decide on the sequence of this gene in each species to be used for read alignment and read counting. For the commonly studied species, the ortholog sets can be downloaded from a number of online databases (e.g. HomoloGene of NCBI; BioMart of Ensembl; and Multiz Alignment of UCSC Genome Browser). However, for the less common species (especially those without complete genomes), there is little information on gene orthology relationship. Aligning reads to the genome of a related species is often far from ideal: for example, only 13% of the reads of African grass rat fibroblasts could be uniquely mapped to the mouse genome (even though both belong to the same Family Muridae), and the alignment rate was even lower for red squirrel (about 5%). Given that a few of the studies presented here involved multiple species without complete genomes, a pipeline was developed to identify ortholog sequences across the species (Figure 2.7).

**Step 1: generate mouse reference**

Based on the *Mus musculus* Ensembl genome and annotation (release 78), the longest transcript was extracted for each protein-coding gene locus, after confirming the presence of start codon and stop codon and the proper reading frame. Those transcripts containing highly repetitive or highly similar sequences (e.g. genes coding for histones and olfactory receptors) were identified and removed using BLAST (at e value cut-off $10^{-6}$) (Camacho et al., 2009), ensuring that the read alignment would be unique and unambiguous. This generated the Mouse Reference, representing the coding sequences of 16,816 unique protein-coding genes.

**Step 2: identify species-specific ortholog sets**

For each species, the transcriptome was assembled *de novo* using Trinity (Grabherr et al., 2011). BLAST (with "dc-megablast" option) was performed between Mouse Reference and the assembled transcriptome (and the published genome, if available) to identify the reciprocal best hits (Tatusov et al., 1997). The sequences were trimmed down to open reading frame (i.e. flanked by start and stop codons) using Exonerate (Slater and Birney, 2005). Within each ortholog sets, multiple sequence alignment was performed using MUSCLE (Edgar, 2004) and the percentage of sequence identity was assessed by MView (Brown et al., 1998).The sequence fragments or missing sequences due to poor coverage were filled up using the consensus of related species.

**Step 3: read mapping, counting, filtering and normalization**

The RNA sequencing reads were aligned to the species-specific ortholog sets using STAR (Dobin et al., 2013) and read counting was performed by featureCounts (Liao et al., 2014). Those ortholog sets with too high counts (i.e. read counts contributing to >5% of the total counts) or too low counts (i.e. less than 10 counts in > 30% of the samples) were discarded. The library sizes were scaled by trimmed mean of M-values (TMM) method, log10-transformed, and quantile-normalized (Robinson and Oshlack, 2010).

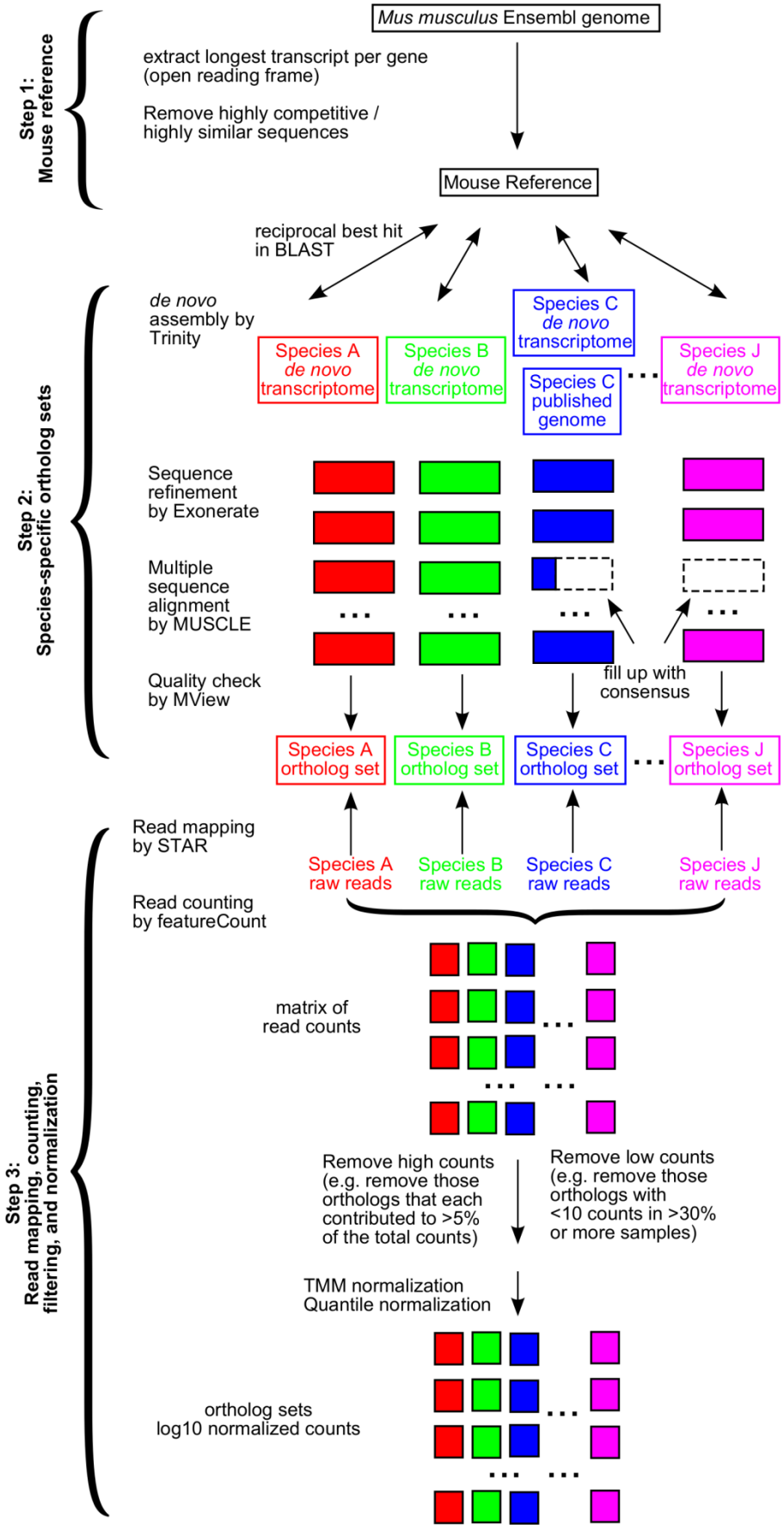



**Figure 2.7. The pipeline to obtain the species-specific ortholog sets and expression values.**

## Ortholog quality assessment

Before proceeding with data analysis, we assessed the quality of the orthologs in terms of the following areas (Figure 2.8; illustrated using data from Chapter 8):

**Filling up with consensus**

The sequence fragments and missing sequences were filled up using the consensus of related species, for two main purposes. First, the ortholog sequences would be of similar lengths, so that read count variations across the species would not be influenced by length differences. Second, if a gene was expressed at low level in one species (and its sequence failed to be assembled *de novo*), the consensus sequence of the related species could be used for read alignment and counting.

In the mammalian transcriptome dataset (Chapter 8), 80% of the orthologs did not require filling up or were filled up <10% of sequence length, whereas only 4% were filled up 90~100% of sequence length (Figure 2.8A). In terms of standardized expression values, there was no significant bias against those filled up using consensus (Figure 2.8A), perhaps except for a slight decline in the 90-100% category. Therefore, the filling up procedure did not negatively impact the overall results.

**Aligned to genomes vs. aligned to ortholog sets**

Unlike a complete genome, the *de novo* assembled ortholog sequences did not contain introns and non-coding regions. The 5' and 3' untranslated regions were also removed to facilitate sequence alignment. For those species with publically available genomes, ~82% of the reads could be aligned to the complete genomes and ~46% of the reads could be aligned to the ortholog sets. For those without genomes, ~42% of the reads could be aligned to the ortholog sets (Table 2.5, Figure 2.8B). The Spearman correlation coefficients between the read counts based on ortholog set alignment and those based on genome alignment were > 0.95 for most species (Table 2.5, Figure 2.8B).

**Our orthology definition vs. Ensembl orthology definition**

For those species with annotated genomes in Ensembl, we compared our orthology definition with the orthology definition in Ensembl. Orthology information for ~10,000 to ~15,000 of our orthologs could be found in Ensembl, and 90-99% of them matched our orthology definition (Table 2.5), suggesting the results of our pipeline were consistent with other databases.

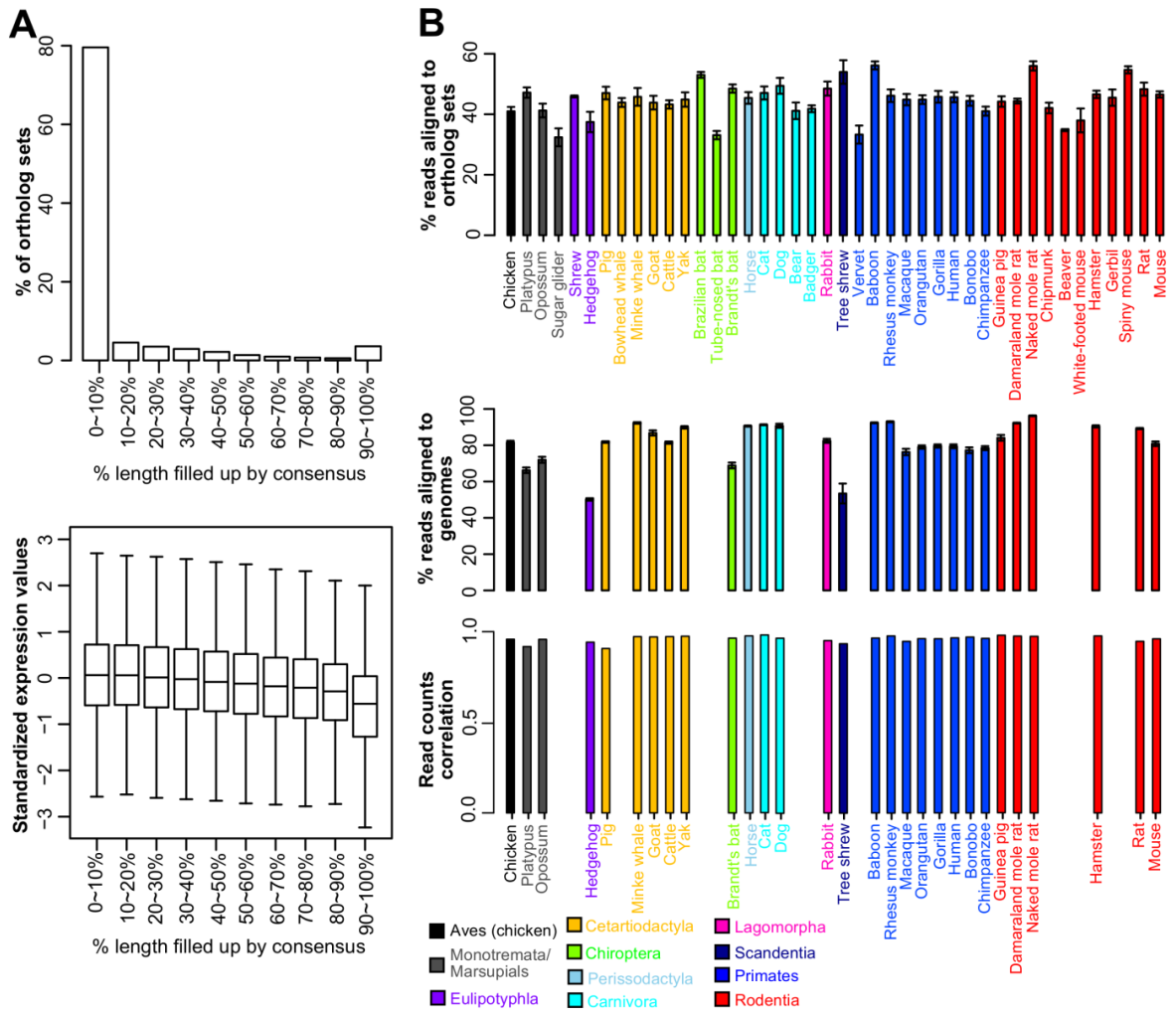



**Figure 2.8. Assessment of ortholog sets quality.**

**(A) The effect of filling up with consensus.** The ortholog sets were categorized by their lengths that required filling up (top) and the corresponding relative expression values were shown (bottom).

**(B) Comparing reads alignment to ortholog sets and to genome.** The percentages of reads in each species aligned to the ortholog sets were shown (top). For those species with complete genomes, the reads were also aligned to the genomes (middle), and the average Spearman correlation coefficient between the read counts of ortholog set alignment and the read counts of genome alignment was calculated for these species (bottom). See Table 2.5 for more details.

**Table 2.5. Assessment of ortholog sets quality.** For each species, the reads were aligned to *de novo* assembled ortholog sets, and if available, to the complete genome. The average Spearman correlation coefficient between the read counts of ortholog set alignment and of genome alignment was calculated (see Figure 2.8B). For those species with Ensembl annotations, the ortholog definition from Ensembl were compared with our ortholog definition.

| Species | Alignment Rates and Counts | | | | Ortholog definition | | |
|---|---|---|---|---|---|---|---|
| | Aligned to ortholog sets | Source of genome | Aligned to complete genome | Read counts correlation (mean) | No. with Ensembl definition | No. of matching definition | % of matching definition |
| Chicken | 41.0% | Ensembl | 81.6% | 0.958 | 11,249 | 11,182 | 99.4% |
| Platypus | 47.2% | Ensembl | 66.2% | 0.918 | 11,363 | 10,667 | 93.9% |
| Opossum | 41.2% | Ensembl | 72.0% | 0.958 | 13,569 | 13,454 | 99.2% |
| Sugar glider | 32.4% | N/A | | | | | |
| Shrew | 45.9% | N/A | | | | | |
| Hedgehog | 37.4% | Ensembl | 50.1% | 0.942 | 11,510 | 11,440 | 99.4% |
| Pig | 47.0% | Ensembl | 81.8% | 0.907 | 13,880 | 13,598 | 98.0% |
| Bowhead whale | 43.9% | N/A | | | | | |
| Minke whale | 45.8% | NCBI | 92.3% | 0.974 | | | |
| Goat | 43.9% | NCBI | 86.8% | 0.971 | | | |
| Cattle | 43.3% | Ensembl | 81.4% | 0.973 | 15,095 | 15,005 | 99.4% |
| Yak | 44.9% | NCBI | 89.9% | 0.974 | | | |
| Brazilian bat | 53.0% | N/A | | | | | |
| Tube-nosed bat | 33.1% | N/A | | | | | |
| Brandt's bat | 48.5% | NCBI | 68.9% | 0.964 | | | |
| Horse | 45.4% | Ensembl | 90.6% | 0.977 | 14,866 | 14,808 | 99.6% |
| Cat | 47.0% | Ensembl | 91.3% | 0.981 | 14,827 | 14,743 | 99.4% |
| Dog | 49.4% | Ensembl | 90.7% | 0.966 | 14,900 | 14,819 | 99.5% |
| Bear | 41.1% | N/A | | | | | |
| Badger | 41.8% | N/A | | | | | |
| Rabbit | 48.5% | Ensembl | 82.4% | 0.951 | 14,039 | 13,942 | 99.3% |
| Tree shrew | 54.0% | Ensembl | 53.4% | 0.935 | 11,841 | 11,776 | 99.5% |
| Vervet | 33.3% | N/A | | | | | |
| Baboon | 56.2% | NCBI | 92.4% | 0.965 | | | |
| Rhesus monkey | 46.1% | NCBI | 92.9% | 0.977 | | | |
| Macaque | 44.8% | Ensembl | 76.3% | 0.947 | 14,771 | 14,615 | 98.9% |
| Orangutan | 44.9% | NCBI | 79.0% | 0.962 | | | |
| Gorilla | 45.8% | Ensembl | 79.6% | 0.961 | 14,760 | 14,657 | 99.3% |
| Human | 45.6% | Ensembl | 79.4% | 0.966 | 15,409 | 14,934 | 96.9% |
| Bonobo | 44.4% | NCBI | 77.3% | 0.970 | | | |
| Chimpanzee | 41.0% | Ensembl | 78.4% | 0.963 | 14,576 | 14,527 | 99.7% |
| Guinea pig | 44.2% | Ensembl | 84.1% | 0.979 | 14,592 | 14,521 | 99.5% |
| Damaraland mole Rat | 44.4% | NCBI | 92.2% | 0.976 | | | |
| Naked mole rat | 56.0% | NCBI | 96.2% | 0.974 | | | |
| Chipmunk | 42.1% | N/A | | | | | |
| Beaver | 34.8% | N/A | | | | | |
| White-footed mouse | 38.0% | N/A | | | | | |
| Hamster | 46.6% | NCBI | 90.4% | 0.976 | | | |
| Gerbil | 45.5% | N/A | | | | | |
| Spiny mouse | 54.7% | N/A | | | | | |
| Rat | 48.3% | Ensembl | 89.2% | 0.947 | 15,791 | 14,493 | 91.8% |
| Mouse | 46.5% | Ensembl | 80.8% | 0.961 | 16,817 | 16,817 | 100.0% |

## DATA VARIABILITY

Each of the studies presented in this dissertation involves biological samples from different species or strains, often obtained from different sources. While care has been taken to match the samples by biological age and sex (e.g. for the mammalian species in Chapter 3, Chapter 4, and Chapter 8, the samples were obtained from young, male adults (except for female for vervet and horse)), other factors such as diet, circadian cycle, and even measurement inaccuracy, will introduce additional data variation. Furthermore, we rely on the public database for the longevity trait information; such data will not be entirely accurate either. Here the issue of data variability will be discussed in more detail.

## Within-species variation vs. between-species variation

Chapter 3 presents the metabolite data of brain, heart, kidney, and liver of 26 mammalian species. Chapter 8 presents the gene expression data of brain, kidney, and liver of chicken and 41 mammalian species. Analysis of Variance (ANOVA) revealed that organs, species and their interaction together accounted for 91% of the total variation in the metabolite dataset, and 87% of the total variation in the gene expression dataset, whereas the differences between replicates accounted for only 9% and 13%, respectively (Figure 2.9). This suggests the within-species variation is likely to be much smaller than the between-species variation.

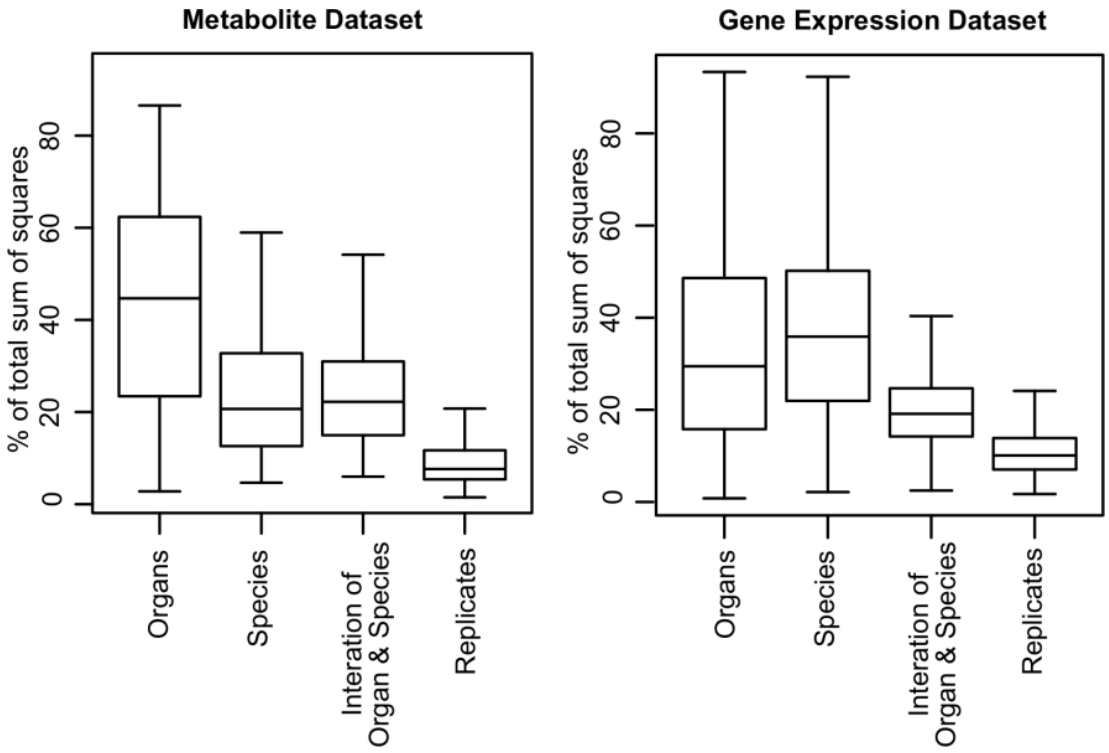



**Figure 2.9. Sources of data variation.** The box plots showed the distribution of the ANOVA results for individual metabolites (left) or genes (right).

## Samples from different datasets

In Chapter 8, besides the samples collected and sequenced by our laboratory (Fushan et al., 2015), other published RNA sequencing samples were also incorporated (Brawand et al., 2011; Merkin et al., 2012). For the species that were common across different datasets, we analyzed their variations by hierarchical clustering. For example, Fushan *et al.* (Fushan et al., 2015) reported expression data in brain, kidney, and liver for 3 mice and 2 cattle; Brawand *et al.* (Brawand et al., 2011) reported data in brain, cerebellum, heart, kidney, and liver for 3 mice (as well as testis for 2 of the mice); and Merkin *et al.* (Merkin et al., 2012) reported data in brain, heart, kidney, and liver for 2 cattle (Figure 2.10). The result showed that the samples segregated first by organ, then by species, and there was no obvious segregation by the sources of sample. The variations between the replicates were smaller than the variations between species and between organs.

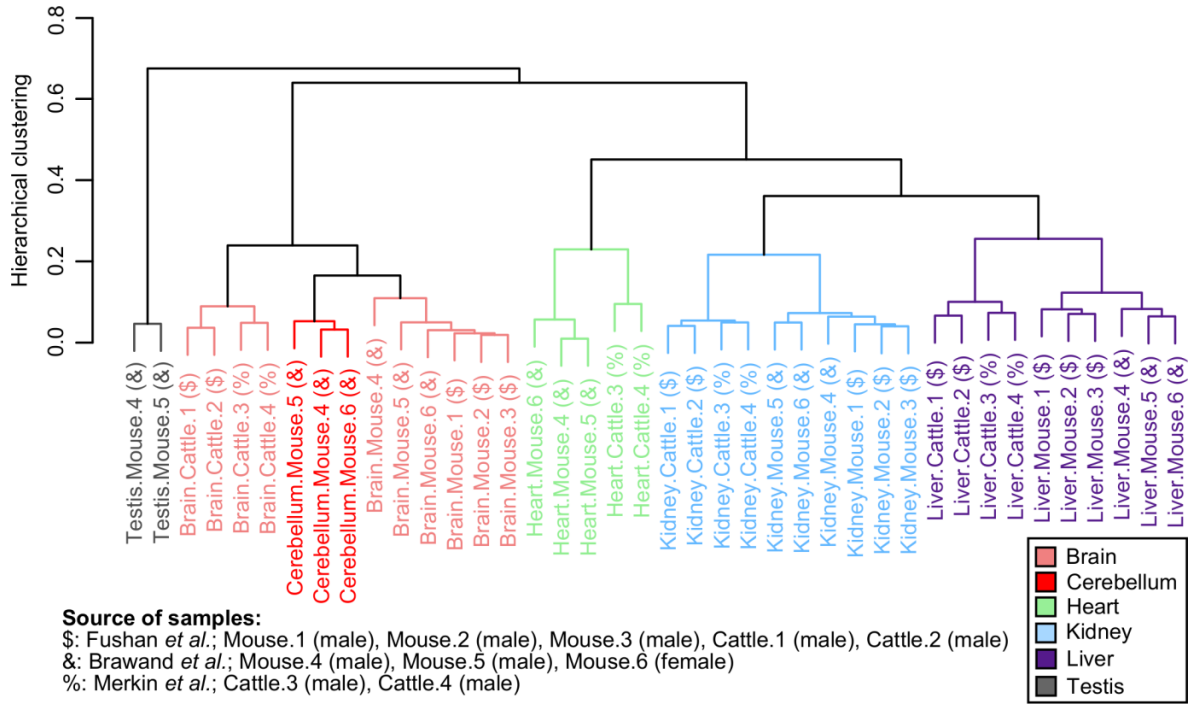



**Figure 2.10. Clustering of the samples from different sources.** The symbols in the parenthesis indicate the source of sample. Hierarchical clustering was performed on a distance matrix of 1 minus Pearson correlation coefficient of expression values, using complete linkage.

## Metabolic effects of drug treatment and gene knockout

In Chapter 3, metabolite profiling was performed on brain, heart, kidney and liver of 26 mammalian species, as well as on brain and liver of 5 long-lived mouse models (i.e. caloric restriction (CR), rapamycin treatment (RAP), acarbose treatment (ACA), growth hormone receptor knock-out (GHRKO), and Snell dwarf mutant mice). Compared to wild-type, these long-lived mice exhibited certain metabolic changes (e.g. lower level of polyunsaturated triacylglycerols and higher levels of sphingomyelin) that were also found to correlate with longevity across various mammalian species (Chapter 3). To visualize the extent of metabolic changes, we performed Principal Component Analysis (PCA) on brain and liver data of the long-lived mice, wild-type mice, and the other mammalian species (Figure 2.11). The result showed that while these treatments and gene manipulation changed the metabolite levels in these mice, the long-lived mouse samples still clustered more closely to the wild-type mice than to the other species, suggesting that the inherent metabolic signatures of the species are relatively robust, and the variations across biological replicates are minor compared to the variations across different species and organs.

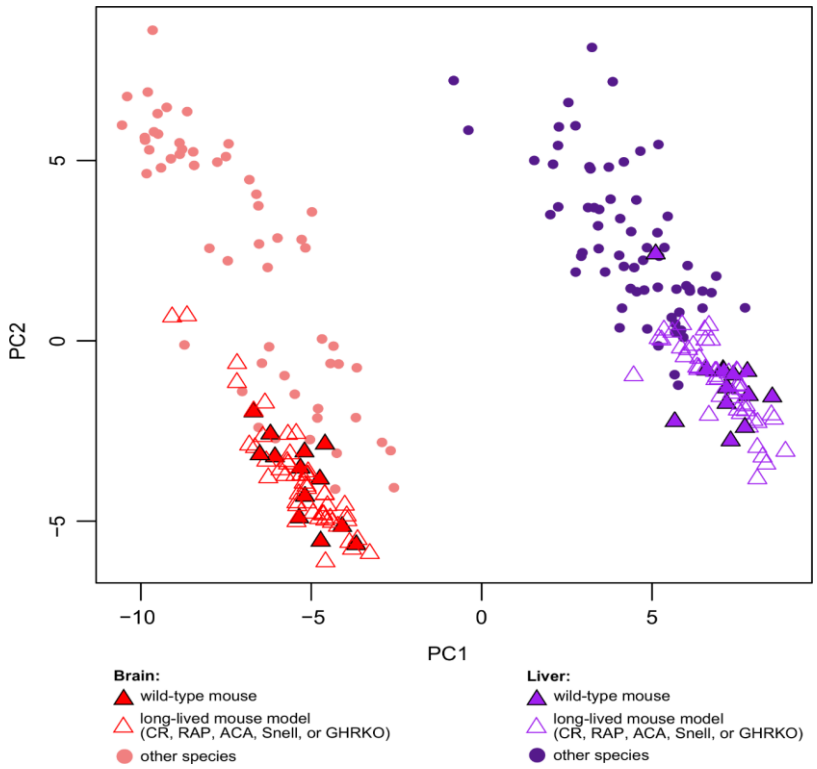



**Figure 2.11. Sources of data variation.** The box plots showed the distribution of the ANOVA results for individual metabolites (left) or genes (right).

# REFERENCES


Baum, D. (2008). Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. Nature Education *1*.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Brown, N.P., Leroy, C., and Sander, C. (1998). MView: a web-compatible database search or multiple alignment viewer. Bioinformatics *14*, 380-381.

Butler, M.A., and King, A.A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am. Nat. *164*, 683-695.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792-1797.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. Am. Nat. *125*, 1-15.

Freckleton, R.P., Harvey, P.H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. *160*, 712-726.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Garland, T., Dickerman, A.W., Janis, C.M., and Jones, J.A. (1993). Phylogenetic Analysis of Covariance by Computer Simulation. Syst. Biol. *42*, 265-292.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644-652.

Grafen, A. (1989). The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci. *326*, 119-157.

Lavin, S.R., Karasov, W.H., Ives, A.R., Middleton, K.M., and Garland, T., Jr. (2008). Morphometrics of the avian small intestine compared with that of nonflying mammals: a phylogenetic approach. Physiol. Biochem. Zool. *81*, 526-550.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Logan, M. (2010). Biostatistical design and analysis using R : a practical guide (Chichester, UK ; Hoboken, NJ, Wiley-Blackwell).

Martins, E.P., and Hansen, T.F. (1997). Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. Am. Nat. *149*, 646-667.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593-1599.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877-884.

Revell, L.J. (2010). Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. *1*, 319-329.

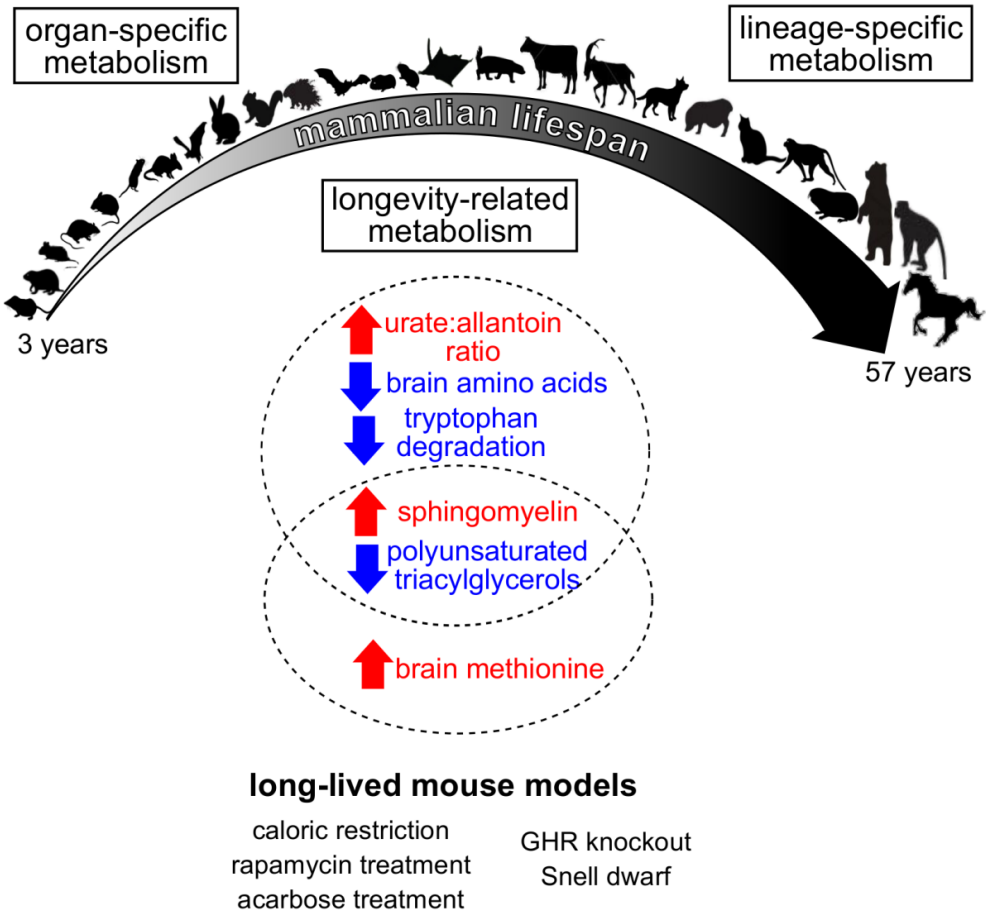
Revell, L.J., Harmon, L.J., and Collar, D.C. (2008). Phylogenetic Signal, Evolutionary Process, and Rate. Syst. Biol. *57*, 591-601.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. Science *278*, 631-637.

# Chapter 3 Mammalian Metabolome




This chapter is based on the following publication:

Siming Ma*, Sun Hee Yim*, Sang-Goo Lee, Eun Bae Kim, Sang-Rae Lee, Kyu-Tae Chang, Rochelle Buffenstein, Kaitlyn N. Lewis, Thomas J. Park, Richard A. Miller, Clary B. Clish, Vadim N. Gladyshev. **Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity.** *Cell Metab.,* 22 (2015), pp. 332-43

* equal contribution.

**Author Contributions**

S.M. carried out data analyses. S.H.Y. carried out preliminary analyses. C.B.C. supervised metabolite profiling. S.H.Y., S.G.L., E.B.K., S.R.L., K.T.C., R.B., K.N.L., T.J.P., and R.A.M. provided, collected, and prepared samples. All authors contributed to data interpretation. V.N.G. coordinated the study. S.M. and V.N.G. wrote the paper with input from all authors.

# ABSTRACT

Biological diversity among mammals is remarkable. Mammalian body weights range seven orders of magnitude and lifespans differ more than 100-fold among species. While genetic, dietary, and pharmacological interventions can be used to modulate these traits in model organisms, it is unknown how they are determined by natural selection. By profiling metabolites in brain, heart, kidney, and liver tissues of 26 mammalian species representing ten taxonomical orders, we report metabolite patterns characteristic of organs, lineages, and species longevity. Our data suggest different rates of metabolite divergence across organs and reveal patterns representing organ-specific functions and lineage-specific physiologies. We identified metabolites that correlated with species lifespan, some of which were previously implicated in longevity control. We also compared the results with metabolite changes in five long-lived mouse models and observed some similar patterns. Overall, this study describes adjustments of the mammalian metabolome according to lifespan, phylogeny, and organ and lineage specialization.

# INTRODUCTION

All modern mammals descend from a common ancestor that lived ~210 million years ago and have since undergone remarkable diversification in morphology, life history, and other characteristics. Their body parts, such as tongues, ears, fingers, and feet, have been modified for numerous functions including nectar-feeding, echolocating, swimming, flying, and digging; their body weights range from under 2 g (Etruscan shrew, *Suncus etruscus*) to over 150 tons (blue whale, *Balaenoptera musculus*); and their maximum lifespans differ by more than 100-fold (Tacutu et al., 2013). Many of the traits affecting development, body weight, and lifespan (i.e., the life history traits) are often correlated. Longer-lived species tend to be bigger, produce fewer offspring, grow more slowly, and have lower mass-specific metabolic rates (Peters, 1986; Sacher, 1959; Western, 1979), indicative of modulation by the same underlying evolutionary forces. Certain lineages (e.g., bats (Seim et al., 2013) and primates) have evolved to live longer as a whole, whereas other instances of exceptional longevity have emerged sporadically among short-lived taxonomic relatives (e.g. the naked mole rat (*Heterocephalus glaber*) lives ten times longer than other rodents of comparable size (Buffenstein, 2008; Fang et al., 2014; Kim et al., 2011)).

Longevity is elastic and can vary along a continuum, but the underlying factors are only starting to be characterized. Research in model organisms revealed several important molecular players, such as insulin-like growth factor 1 (IGF-1) (Friedman and Johnson, 1988; Holzenberger et al., 2003; Tatar et al., 2001), mechanistic target of rapamycin (mTOR) (Kenyon, 2010; Vellai et al., 2003), and sirtuins (Lin et al., 2000; Tissenbaum and Guarente, 2001). Dietary and pharmacological interventions can also extend lifespan in diverse organisms (Harrison et al., 2009; McCay et al., 1935; Weindruch et al., 1986). In particular, lifespan of laboratory mice can be increased by restriction of food or methionine (Flurkey et al., 2010; Sun et al., 2009), administration of rapamycin (Harrison et al., 2009; Miller et al., 2014) or acarbose (Harrison et al., 2014), or certain genetic mutations (Ladiges et al., 2009). Rapamycin, an inhibitor of mTORC1, leads to 23-26% increase in median lifespan of mice (Miller et al., 2014). Acarbose inhibits glycoside hydrolases (the enzymes that digest complex carbohydrates to absorbable sugars in the gastrointestinal tract) and is used clinically to blunt post-

prandial glucose surges in diabetic patients; it seems plausible that limiting peak glucose concentrations may explain its longevity benefits. Mutant strains such as Snell dwarf (defective in anterior pituitary development) (Flurkey et al., 2001) and growth hormone receptor knockout (GHRKO) (Coschigano et al., 2003) are also long-lived, due to altered signaling in growth hormone (GH) itself or through GH-stimulated production of IGF-1.

How longevity is modulated during evolution to produce both long-lived and fit animals, however, is still unclear. Lifespan is an inherent characteristic of a species and remains relatively stable through generations, but it can also change in either direction over time. In order to vary lifespan on an evolutionary time scale, a number of biological pathways may need to be altered, rewired or reprogrammed. Omics-scale comparative studies across multiple species are instrumental in understanding the evolution of mammalian genomes and gene expression (Brawand et al., 2011; Lindblad-Toh et al., 2011). To gain insights into the metabolic basis of mammalian diversity and longevity, we quantified metabolite levels in brain, heart, kidney, and liver tissues of 26 species of mammals and identified metabolites with organ-, lineage-, and trait-specific patterns. We described the metabolite divergence and distribution in different organs; linked the lineage-specific metabolic patterns to lineage-specific physiologies; and identified metabolites with positive or negative correlation to longevity traits. In addition, we profiled the metabolites in brain and liver of five long-lived mouse models (caloric restriction, rapamycin treatment, acarbose treatment, GHRKO, and Snell dwarf) and compared the observed changes with the mammalian longevity signatures. Our study provides the first glimpse into how metabolism may have been altered to modulate mammalian lifespan.

# RESULTS

## Metabolite conservation and divergence among organs

We applied targeted metabolite profiling to quantify the metabolite levels in brain, heart, kidney, and liver of 26 species of mammals, representing 10 taxonomical orders and covering a wide range of longevity-associated traits (Figure 3.1, Table 3.1). The species were matched by biological age (all young adults) and sex (all were males, except for horse and vervet). Biological replicates (i.e. samples from multiple individuals of a species) were collected for most of the species (Table 3.1). Those metabolites with more than 20% missing values in a particular organ were excluded from analysis in that organ (Figure 3.2A). In total, 162 water soluble metabolites and 100 lipids were reliably detected across 235 samples. Across the biological replicates, over 90% of the measurements had coefficient of variation (i.e. standard deviation divided by mean) < 0.06 (Figure 3.2B).

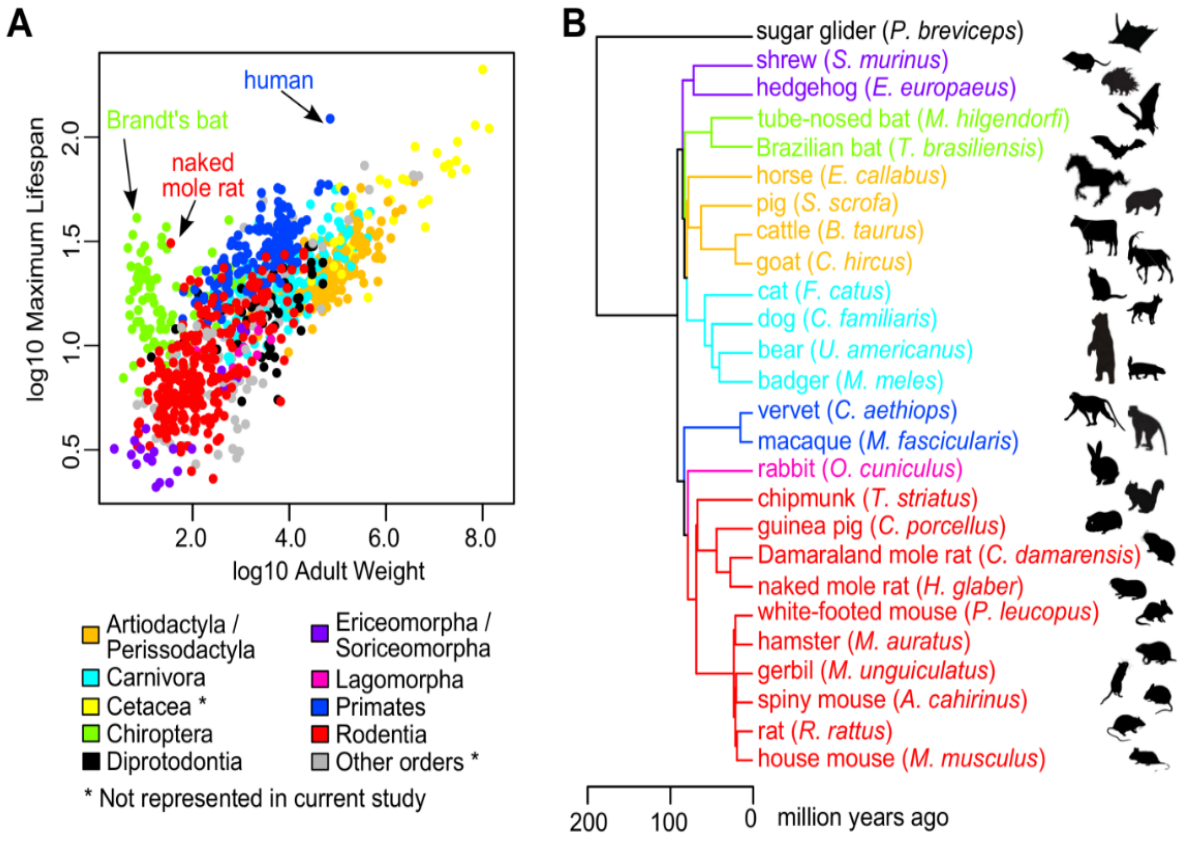



**Figure 3.1. Diversification of mammals.**

**(A) Maximum lifespan correlates positively with body mass.** Maximum lifespan (years) was plotted against adult weight (grams) on log10-scale for 995 mammalian species from AnAge database (Tacutu et al., 2013), color coded by taxonomical orders. To simplify the color scheme, Artiodactyla and Perissodactyla were grouped together, and Ericomorpha and Soriceomorpha were grouped together.

**(B) Phylogeny of the mammals examined in the current study.** Branches and tips are colored according to taxonomical orders (same color scheme as in (A)). Divergence times were based on previous estimates (Fushan et al., 2015; Meredith et al., 2011). Animal silhouettes are for illustration only.

**Table 3.1. 26 mammalian species samples and life history traits.** Adult Weight (AW), Maximum Lifespan (ML) and Female Time to Maturity (FTM) were obtained from AnAge database (Tacutu et al., 2013). Average Lifespan (AL) were obtained from Longevity Records of Max Planck Institute for Demographic Research (Carey and Judge, 2000). Maximum Lifespan Residuals (MLres), Female Time to Maturity Residuals (FTMres), and Average Lifespan Residual (ALres) were computed using the following allometric equations: ALres = AL/(2.16×$AW^{0.205}$); MLres =

| Common Name | Scientific Name | Sex | No. of biological replicates | | | | AW (gram) | AL (year) | ML (year) | FTM (day) | ALres | MLres | FTMres |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Br | Ht | Kd | Lv | | | | | | | |
| sugar glider | *Petaurus breviceps* | M | 2 | 2 | 2 | 2 | 129.3 | 14 | 17.8 | 236 | 2.3911 | 1.7335 | 1.0521 |
| shrew | *Suncus murinus* | M | 0 | 0 | 0 | 3 | 39.7 | 2.5 | 3.2 | 36 | 0.5439 | 0.3733 | 0.2074 |
| hedgehog | *Erinaceus europaeus* | M | 2 | 2 | 2 | 2 | 1213.5 | 8 | 11.7 | 253 | 0.8634 | 0.8089 | 0.6938 |
| tube-nosed bat | *Murina hilgendorfi* | M | 0 | 3 | 3 | 3 | 12.8 | 6 | 9 | 180 | 1.6477 | 1.2493 | 1.3266 |
| Brazilian bat | *Tadarida brasiliensis* | M | 0 | 0 | 0 | 1 | 13.3 | 8 | 12 | 273 | 2.1779 | 1.6551 | 1.9936 |
| horse | *Equus caballus* | F | 2 | 0 | 3 | 3 | 300000 | 30 | 57 | 914 | 1.0463 | 1.6961 | 0.7582 |
| pig | *Sus scrofa* | M | 3 | 3 | 3 | 3 | 135000 | 18 | 27 | 334 | 0.7394 | 0.9078 | 0.3295 |
| cattle | *Bos taurus* | M | 3 | 0 | 3 | 3 | 548500 | 20 | 20 | 548 | 0.6164 | 0.5426 | 0.3988 |
| goat | *Capra hircus* | M | 3 | 3 | 3 | 3 | 61000 | 15 | 20.8 | 406 | 0.7252 | 0.7897 | 0.4758 |
| cat | *Felis catus* | M | 3 | 3 | 3 | 3 | 3900 | 14 | 30 | 289 | 1.1893 | 1.7349 | 0.6152 |
| dog | *Canis familiaris* | M | 3 | 3 | 3 | 3 | 10148.9 | 11 | 21.8 | 274 | 0.7681 | 1.0891 | 0.4739 |
| bear | *Ursus americanus* | M | 3 | 0 | 3 | 3 | 154250 | 30 | 34 | 1278 | 1.1992 | 1.1201 | 1.2247 |
| badger | *Meles meles* | M | 2 | 2 | 2 | 2 | 11050 | 16 | 18.6 | 365 | 1.0979 | 0.9172 | 0.6198 |
| vervet | *Chlorocebus aethiops* | F | 1 | 1 | 1 | 1 | 5620 | 30 | 30.8 | 1034 | 2.3647 | 1.6843 | 2.0332 |
| macaque | *Macaca fascicularis* | M | 2 | 2 | 2 | 2 | 6362.5 | 38 | 39 | 1238 | 2.92 | 2.0926 | 2.3696 |
| rabbit | *Oryctolagus cuniculus* | M | 3 | 3 | 3 | 3 | 2167.9 | 9 | 9 | 730 | 0.8624 | 0.5694 | 1.765 |
| chipmunk | *Tamias striatus* | M | 2 | 2 | 2 | 2 | 89.6 | 8 | 9.5 | 187 | 1.473 | 0.9786 | 0.9027 |
| guinea pig | *Cavia porcellus* | M | 3 | 3 | 3 | 3 | 639.1 | 8 | 12 | 66 | 0.9847 | 0.9152 | 0.208 |
| Damaraland mole rat | *Cryptomys damarensis* | M | 1 | 1 | 1 | 1 | 131.3 | 12 | 15.5 | 511 | 2.0431 | 1.506 | 2.2704 |
| naked mole rat | *Heterocephalus glaber* | M | 2 | 0 | 1 | 4 | 35.3 | 30 | 31 | 228 | 6.6861 | 3.6824 | 1.3472 |
| white-footed mouse | *Peromyscus leucopus* | M | 4 | 2 | 2 | 4 | 22.3 | 1 | 7.9 | 73 | 0.2449 | 1.0067 | 0.4765 |
| hamster | *Mesocricetus auratus* | M | 3 | 3 | 3 | 3 | 108.2 | 2 | 3.9 | 48 | 0.3543 | 0.3903 | 0.2224 |
| gerbil | *Meriones unguiculatus* | M | 3 | 3 | 3 | 3 | 64.8 | 2 | 6.3 | 43 | 0.3936 | 0.6819 | 0.2227 |
| spiny mouse | *Acomys cahirinus* | M | 3 | 0 | 3 | 3 | 42 | 4 | 5.9 | 58 | 0.8603 | 0.6825 | 0.33 |
| rat | *Rattus rattus* | M | 3 | 3 | 3 | 3 | 117 | 2 | 4.2 | 90 | 0.3487 | 0.4153 | 0.41 |
| house mouse | *Mus musculus* | M | 3 | 3 | 3 | 3 | 18 | 2 | 4 | 42 | 0.5117 | 0.5267 | 0.2872 |

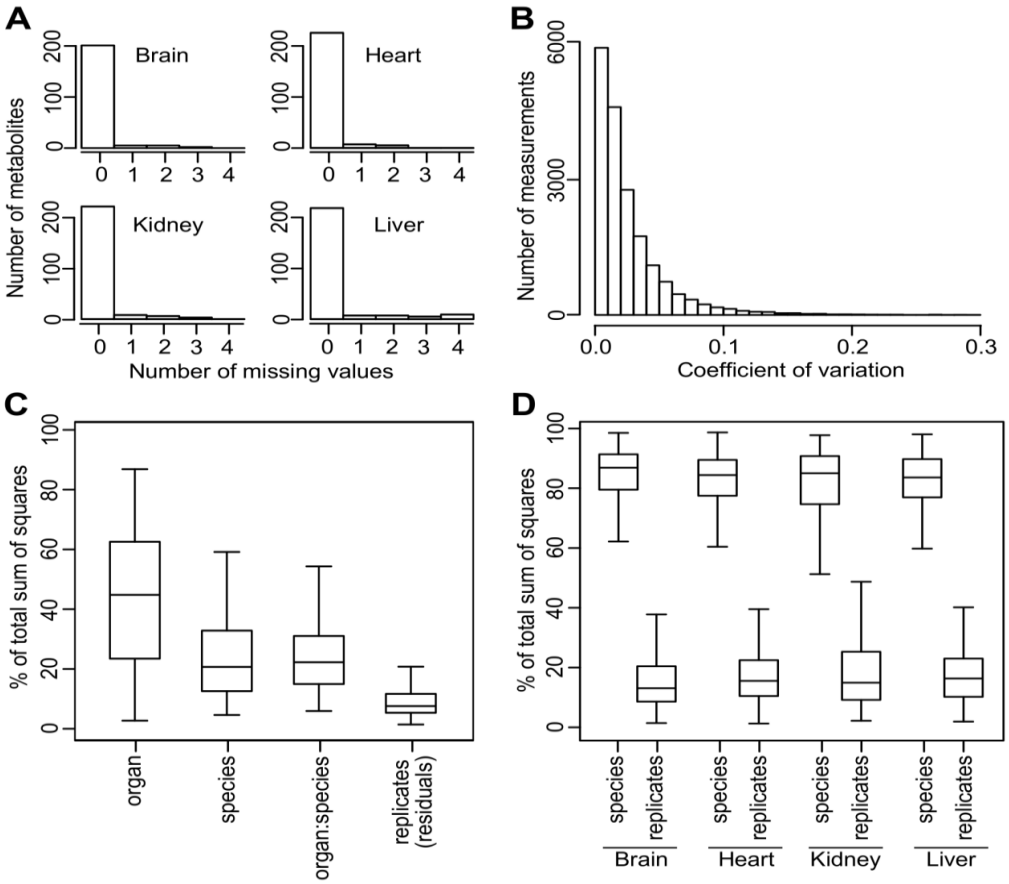



**Figure 3.2. Mammalian species dataset quality assessment.**

**(A) Number of metabolites with missing values in each organ.** Those metabolites with more than 20% missing values were excluded from analysis.

**(B) Coefficient of variation among biological replicates.** Coefficient of variation was computed as standard deviation divided by mean, using only those samples with biological replicates. The 90th percentile was 0.06 and 95th percentile was 0.08.

**(C) Percentage of total variation in metabolite levels attributed to organ, species, and biological replicates.** The plot indicates the percentage of total sum of squares in analysis of variance (ANOVA) attributed to the respective factors. The model "Metabolite Level ~ Organ + Species + Organ:Species" was fitted to each metabolite across the four organs ("Organ:Species" denotes the interaction term).

**(D) Percentage of total variation in metabolite levels in each organ attributed to species and biological replicates.** The model "Metabolite Level ~ Species" was fitted to each metabolite within the indicated organ.

Principal Component Analysis revealed the samples segregated predominantly by organ origin, with most of the replicates clustering together and the first three Principal Components accounting for about 50% of the total variance (Figure 3.3A). Variations due to organ, species, and biological replicates were assessed by analysis of variance (ANOVA). In terms of the overall data, organ and species origins accounted for over 80% of total variation (Figure 3.2C). In terms of the data within each organ, species origins accounted for over 80% of variation (Figures 3.2D). This indicated the between-species and between-organ variations were much greater than the within-species variations. The clustering pattern also agreed with those based on mammalian gene expression profiles (Brawand et al., 2011; Fushan et al., 2015), suggesting that metabolite levels and organ-specific metabolism were generally well conserved across the mammals.

The phylogenetic relationship of many mammals has been established based on fossil and molecular evidence (Brawand et al., 2011; Fushan et al., 2015; Meredith et al., 2011). To determine if their metabolite levels recapitulate this relationship, we constructed phylograms using the metabolite levels in each organ and found them largely consistent with the reference phylogeny (Figure 3.4A). The brain phylogram had the shortest tip-to-root branch lengths (Figures 3.4B and 3.4C), and we found by data simulation that increasing branch lengths might be due to deviation from phylogeny and presence of random noise (Figure 3.4D). The brain samples also showed highest Spearman correlation coefficients (Figure 3.3B) and had the largest proportion of metabolites with high phylogenetic signals (i.e. Pagel's lambda > 0.9 (Pagel, 1999) and Blomberg's K > 1 (Blomberg et al., 2003); Figures 3.3C and 3.3D), suggesting that brain metabolites are most conserved among the four organs and have evolved largely according to the phylogeny (Brawand et al., 2011). In contrast, the metabolites in the other examined organs diverge to much greater extent, possibly due to stronger environmental influence or other selection pressures.

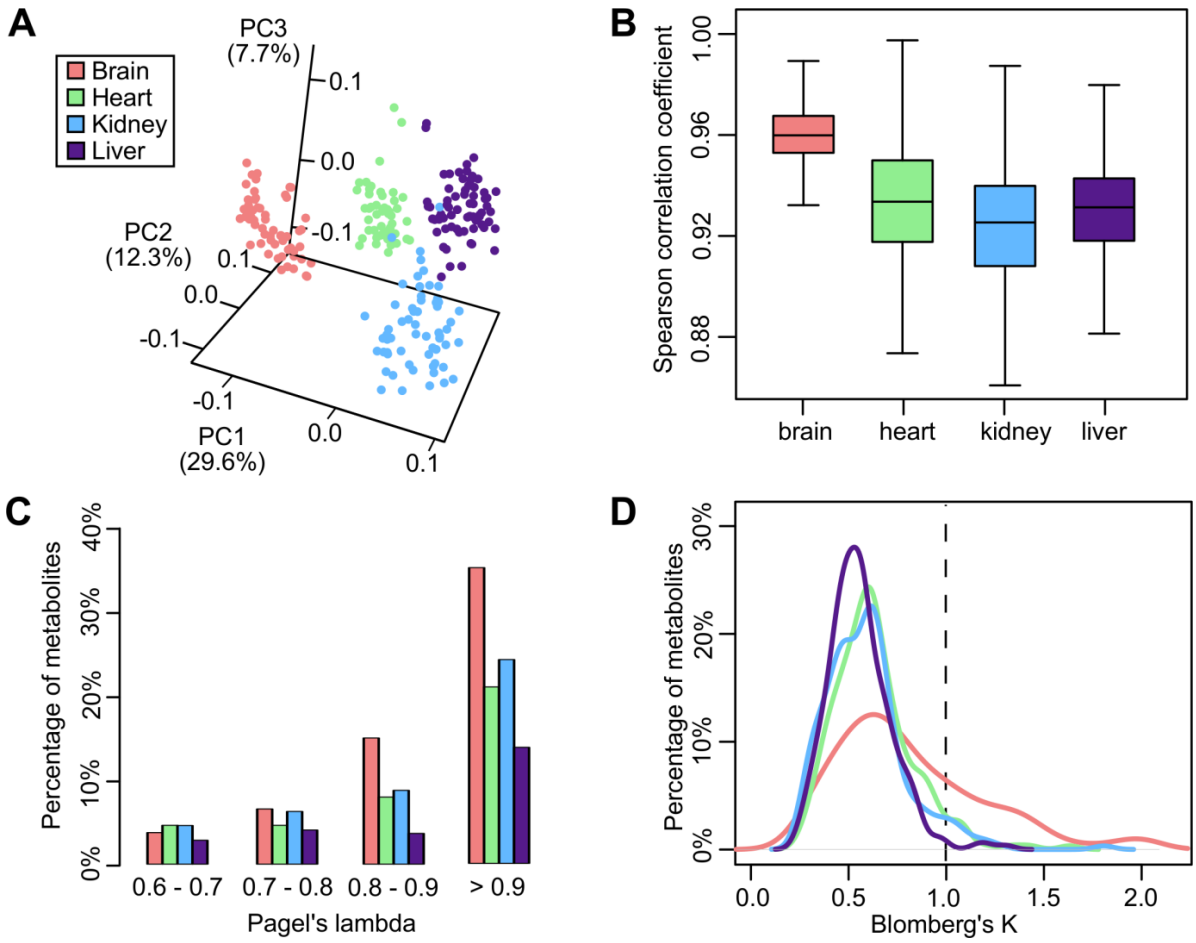



**Figure 3.3. Metabolite divergence and correlation.**

**(A) Samples segregate predominantly according to organ origin.** Values in parenthesis indicate the percentage of variance explained by each of the first three Principal Components (PCs). Biological replicates were treated as individual points.

**(B) Brain samples show highest Spearman correlation coefficients.** The box plot shows the pair-wise correlation among the samples in each organ. Wilcoxon rank sum test p value < $2\times10^{-16}$ for brain against each of the other organs.

**(C) and (D) Brain has the largest percentage of metabolites with high phylogenetic signals.** In (C), only Pagel's lambda > 0.6 are shown. In (D), the dotted line indicates Blomberg's K=1.0.

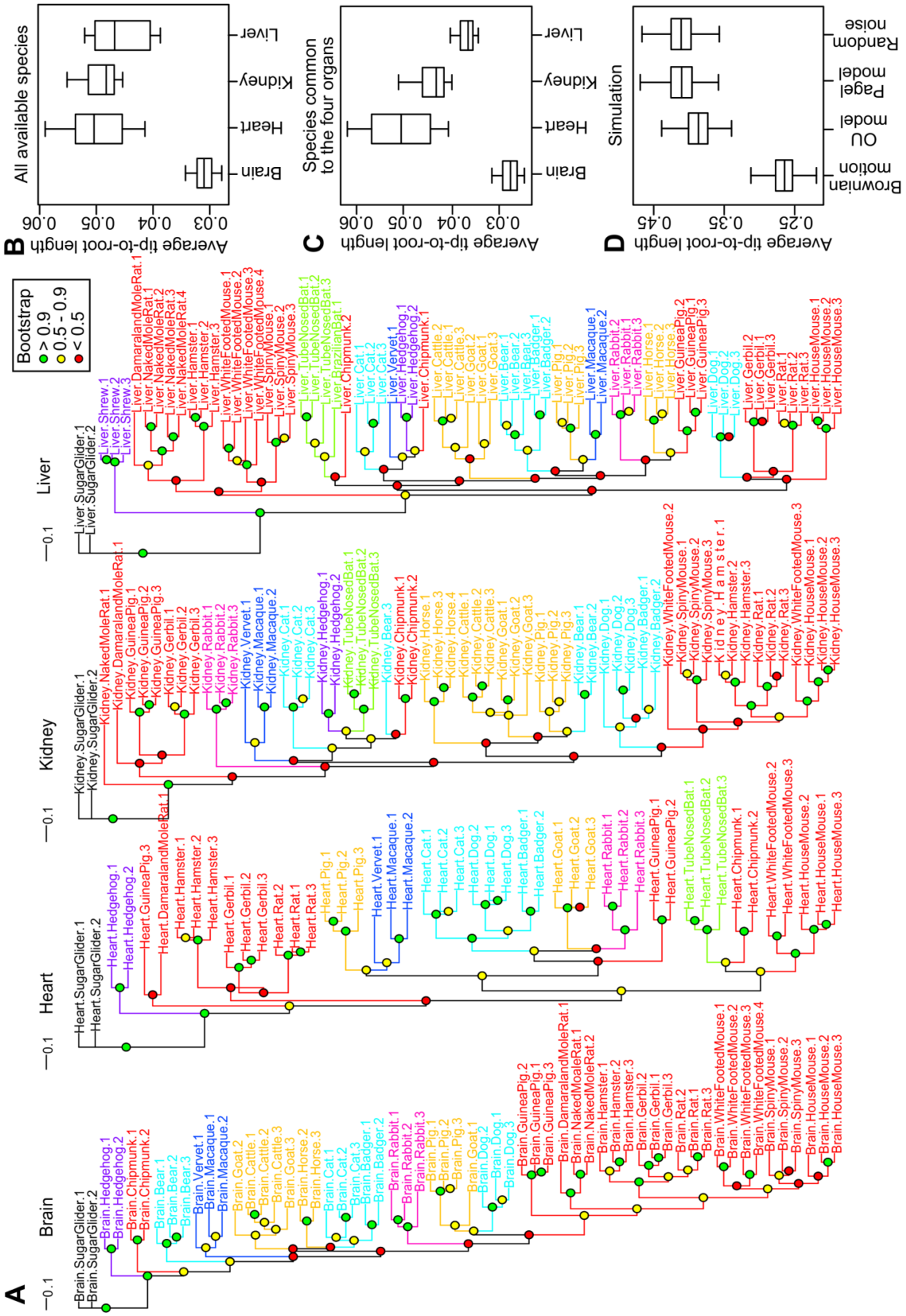
**Figure 3.4. Metabolite divergence in mammalian organs.**

**(A) Phylograms based on metabolite levels in each organ.** The trees were constructed by neighbor joining method using a distance matrix of 1 minus Spearman correlation coefficients. Biological replicates were treated as individual tips and the branches were colored according to taxonomical orders. Colors of nodes indicate 1000-time bootstrap values.

**(B) and (C) Metabolites diverge least in brain.** The average tip-to-root branch lengths excluded the branch leading to sugar glider (the out-group). The box plot represents the results of 1000 trees generated by bootstrap in each organ, using (B) all the species available or (C) only those species common to all the four organs. The central bands indicate median values and the whiskers indicate 5th and 95th percentiles. Wilcoxon rank sum test p values $< 2\times10^{-16}$ (Bonferroni-adjusted) for brain versus each of the other organs.

**(D) Simulation of tip-to-root branch lengths.** The box plot represents the results of 1000 phylograms based on simulated data. For "Brownian motion", the reference phylogenetic tree was used directly (i.e. trait evolution follows phylogeny). For "OU model", the tree was transformed with a restraining force ($\alpha$=1) to mimic the Ornstein-Uhlenbeck (OU) process. For "Pagel model", the tree was transformed by Pagel's lambda ($\lambda$=0.5). For "Random noise", random normal variables with mean 0 and standard deviation 1 were added to the simulated data from "Brownian motion" to mimic the effect of non-phylogenetic variation.

**Figure 3.4 (Continued)**



## Metabolite profiles reflect organ functions

The metabolite profile of an organ is expected to reflect its biological functions. We visualized the distribution of metabolites on a heat map (Figure 3.5A) and used the Wilcoxon rank sum test to identify metabolites selectively enriched or depleted in a particular organ (in comparison with at least two other organs).

18 out of the 19 proteinogenic amino acids measured (cysteine was not quantified) were found at moderate to high levels in kidney relative to the other organs, likely due to reabsorption at the renal proximal tubule. One exception was glutamine (Figure 3.5B), which is routinely metabolized by kidney for nitrogen disposal and acid-base balance. Glutamine is broken down to ammonia and glutamate, helping to remove excessive protons and generate bicarbonate ions (van de Poll et al., 2004).

The metabolite profile of the heart largely reflected its energy demand. Heart tissue was depleted of amino acids and many other metabolites, but enriched with acylcarnitines (especially short-chain acylcarnitines, Figure 3.5C) and triacylglycerols (TAGs). Acylcarnitines help transport fatty acids across mitochondrial inner membranes (Vaz and Wanders, 2002), whereas carnitine acts as an acetyl group acceptor, buffering the cellular pool of coenzyme A (CoA) and preventing inhibition of pyruvate dehydrogenase, especially in tissues dependent on beta-oxidation (Hoppel, 2003).

In contrast, the brain normally relies on glucose for fuel and contains relatively few TAGs. However, it had high concentrations of glycerophospholipids and a number of sphinogmyelins (SMs) (Figure 3.5D), which are both key constituents of animal cell membranes. In particular, SMs are mainly found in the myelin sheaths surrounding nerve cell axons. Neurotransmitters gamma-aminobutyric acid (GABA) and glutamate were also present at high levels.

Liver was enriched with a wide range of metabolites including amino acids, glycerophospholipids, carbohydrates, and steroids. Some of them were significantly higher than in the other organs, likely indicative of liver-specific pathways. For example, sucrose and lactose were found at very high concentrations in liver (Figure 3.5E), as these sugars are not routinely utilized by the other organs. Bile acid components such as glycocholate, taurocholate, taurodeoxycholate and

taurochenodeoxycholate were restricted mostly to liver (Figure 3.5E), since primary bile acids are synthesized by liver cells from cholesterol.

For the lipids, we grouped them according to LIPID MAPS Classification System (Experimental Procedures) (Fahy et al., 2007). Within each category, we compared the relative percentage abundance of individual lipids in our study with those previously reported in human plasma (Quehenberger et al., 2010). Significant correlations were observed for TAG, lysophosphatidylcholine (LPC), phosphatidylcholine (PC), and cholesteryl ester (CE) (Table 3.2), suggesting the overall lipid composition was conserved across mammals.

**Table 3.2. Correlation between the abundance of lipids measured in our study and previously reported lipids in human plasma.** Human plasma data are based on (Quehenberger et al., 2010). For each class of lipid molecules, the relative percentage abundance of individual lipid molecules in a mammalian species in brain (Br), heart (Ht), kidney (Kd), or liver (Lv) was computed and compared with that reported in human plasma. Pearson ($r$) and Spearman ($\rho$) correlation coefficients between the observed and the reported abundance were calculated. The 25th, 50th and 75th percentiles of the correlation coefficients (across all the mammalian species) are shown below. Coefficients > 0.60 are highlighted in bold. TAG: triacylglycerol; LPC: lysophosphatidylcholine; PC: phosphatidylcholine; LPE: lysophosphatidylethanolamine; SM: sphingomyelin; CE: cholesteryl ester.

| **Lipid class** | | **TAG** | | **LPC** | | **PC** | | **LPE** | | **SM** | | **CE** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Correlation** | | ***r*** | ***ρ*** | ***r*** | ***ρ*** | ***r*** | ***ρ*** | ***r*** | ***ρ*** | ***r*** | ***ρ*** | ***r*** | ***ρ*** |
| **Percentile** | | | | | | | | | | | | | |
| Br | 25th | 0.30 | 0.36 | **0.86** | **0.64** | 0.36 | **0.65** | 0.19 | 0.15 | -0.03 | 0.36 | **0.63** | **0.50** |
| | 50th | 0.50 | 0.53 | **0.87** | **0.71** | 0.39 | **0.68** | 0.23 | 0.50 | 0.04 | 0.48 | **0.80** | **0.80** |
| | 75th | **0.61** | **0.69** | **0.88** | **0.71** | 0.43 | **0.70** | 0.36 | 0.50 | 0.10 | **0.64** | **0.95** | **0.80** |
| Ht | 25th | **0.74** | **0.72** | **0.76** | **0.66** | 0.48 | **0.66** | 0.11 | -0.17 | 0.06 | 0.39 | **0.93** | **0.71** |
| | 50th | **0.80** | **0.82** | **0.82** | **0.83** | 0.56 | **0.75** | 0.25 | -0.03 | 0.09 | 0.43 | **0.98** | **0.82** |
| | 75th | **0.85** | **0.86** | **0.86** | **0.94** | **0.74** | **0.79** | 0.37 | 0.09 | 0.23 | 0.50 | **0.99** | **0.89** |
| Kd | 25th | 0.58 | **0.65** | **0.81** | **0.82** | 0.60 | **0.75** | -0.24 | -0.33 | 0.21 | 0.36 | **0.84** | **0.70** |
| | 50th | **0.68** | **0.73** | **0.86** | **0.86** | **0.66** | **0.77** | -0.09 | -0.23 | 0.35 | 0.43 | **0.94** | **0.75** |
| | 75th | **0.83** | **0.82** | **0.88** | **0.93** | **0.72** | **0.83** | 0.16 | -0.07 | 0.54 | 0.48 | **0.98** | **0.83** |
| Lv | 25th | **0.69** | **0.74** | **0.69** | **0.71** | **0.69** | **0.77** | -0.27 | -0.36 | -0.06 | 0.29 | **0.69** | **0.61** |
| | 50th | **0.79** | **0.81** | **0.77** | **0.86** | **0.80** | **0.81** | -0.11 | -0.23 | -0.02 | 0.35 | **0.91** | **0.68** |
| | 75th | **0.86** | **0.85** | **0.86** | **0.92** | **0.84** | **0.85** | 0.08 | -0.03 | 0.06 | 0.38 | **0.97** | **0.81** |

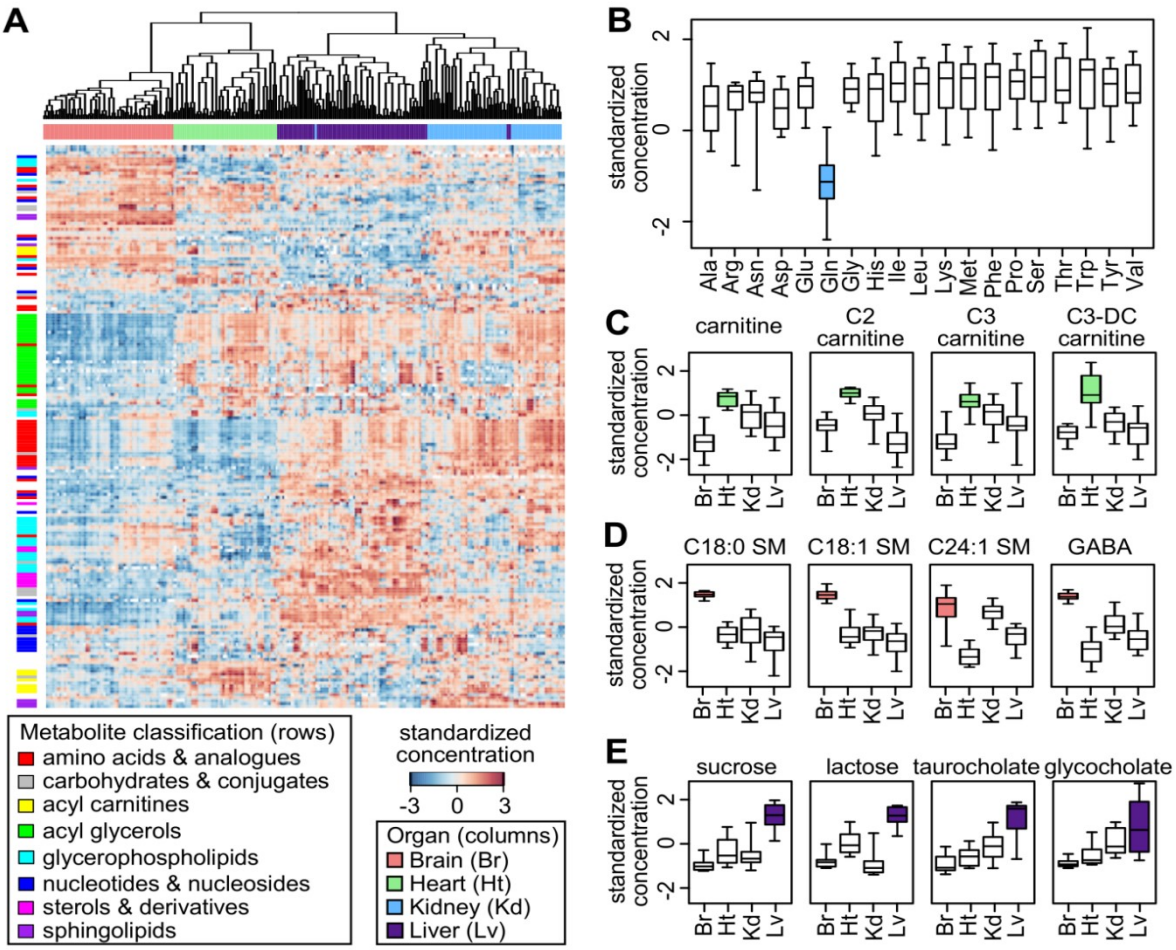



**Figure 3.5. Distribution of metabolites across the organs.**

**(A) The overall pattern visualized on a heat map.** Hierarchical clustering was performed on standardized concentrations using average linkage. Each row represents one metabolite and each column represents one biological sample. Selected classes of metabolites are highlighted.

**(B) Kidney is depleted of glutamine.** Each box represents the range of standardized concentrations for a particular amino acid in kidney across the mammals.

**(C) Heart is enriched with carnitine and short-chain acylcarnitines.** The alternative names are acetylcarnitine (C2 carnitine), propionylcarnitine (C3 carnitine), and malonylcarnitine (C3-DC carnitine).

**(D) Brain is enriched with sphingomyelins (SM) and the neurotransmitter gamma-aminobutyric acid (GABA).**

**(E) Liver has high levels of sucrose, lactose, and bile acid components.**

## Metabolites with lineage-specific changes

Since a particular lineage of mammals often exhibits biological and physiological features distinctive from the others, we grouped the species by taxonomic orders or families and applied phylogenetic ANOVA (Garland et al., 1993) to determine if the concentration of a metabolite in one group was significantly different from the other groups.

Bats (Order Chiroptera) showed significantly reduced levels of methionine sulfoxide in both kidney and liver (phylogenetic ANOVA p value = 0.003 in kidney and 0.002 in liver), while their methionine levels were typical of other mammals (Figure 3.6A). Methionine sulfoxide is produced by oxidation of methionine by reactive oxygen species (ROS) and in most species its level increases during aging or oxidative stress (Berlett and Stadtman, 1997). Bats are the longest-lived mammalian order after controlling for the effect of body size and there is evidence that they produce less ROS and are more resistant to oxidative stress. For example, cave Myotis bats and Mexican free-tailed bats (both with maximum lifespan potential of 12 years) show lower protein carbonylation and ubiquitination in liver than mice and their cells are more resistant to protein oxidation (Salmon et al., 2009; Shi et al., 2010). Bat mitochondria from heart also produce less hydrogen peroxide than those from shrew and white-footed mouse (Brunet-Rossinni, 2004), although the differences are less than divergence in their maximum lifespans (Buffenstein et al., 2008). Hence, low methionine sulfoxide levels are consistent with reduced oxidative stress generally observed in bats.

Several genetic and physiological features of African mole rats (Order Bathyergidae) are distinct from those of other rodents (Fang et al., 2014; Kim et al., 2011), so we compared Bathyergidae against the other examined species, as well as against the other rodents. Several metabolites were detected in both comparisons across multiple organs, including enrichment of acetylglycine (in heart and liver), enrichment of trimethylamine N-oxide (in brain and heart), and depletion of allantoin (in brain, heart, kidney, and liver) (Figure 3.6B). This depletion of allantoin in the Bathyergidae is particularly striking, since other African rodents (in particular the Cricetidae) excrete high levels of allantoin (Buffenstein et al., 1985). Using the gene expression data for some of these species (Fushan et al., 2015), we confirmed the positive correlation between uricase expression

and allantoin level in liver (Figure 3.6C), with particularly low expression in the naked mole rat. In mammals, degradation of purine produces urate, which is then converted to allantoin by the enzyme uricase and excreted in urine (Buffenstein et al., 1985; Ngo and Assimos, 2007). In humans and other higher primates, the gene coding for uricase is a pseudogene and urate is excreted instead. However, these genetic changes were not found in the African mole rat enzymes, so the low uricase expression appears to be achieved by a different mechanism.

Since the mammalian species used in this study include carnivores, insectivores, omnivores, and herbivores, we wondered if the dietary preferences would also be reflected in the metabolic profiles, especially in terms of different bile acid conjugates. Bile acids can be conjugated with either taurine or glycine, depending on their concentrations in liver and affinities for the enzyme bile acid CoA:amino acid N-acyltransferase. Most animals conjugate exclusively with taurine, whereas glycine conjugation is limited to certain placental mammals and herbivores (Huxtable, 2002; Vessey, 1978). Indeed, in liver the taurocholate:glycocholate ratio correlated positively with taurine:glycine ratio (Pearson correlation coefficient = 0.74; Figure 3.6D). Rabbit and guinea pig are known to employ glycine-conjugation only and had low values for both ratios, while cat, being an obligate carnivore, was high in both (Figure 3.6D). Based on the clustering pattern, hedgehog and shrew (predominantly insectivores), as well as animals belonging to Order Carnivora probably use taurine-conjugation only, whereas most rodents of the Family Muridae and animals of Orders Artiodactyla and Perissodactyla use both taurine- and glycine-conjugation (Figure 3.6D).

**Figure 3.6. Metabolite patterns reflect species physiology.**

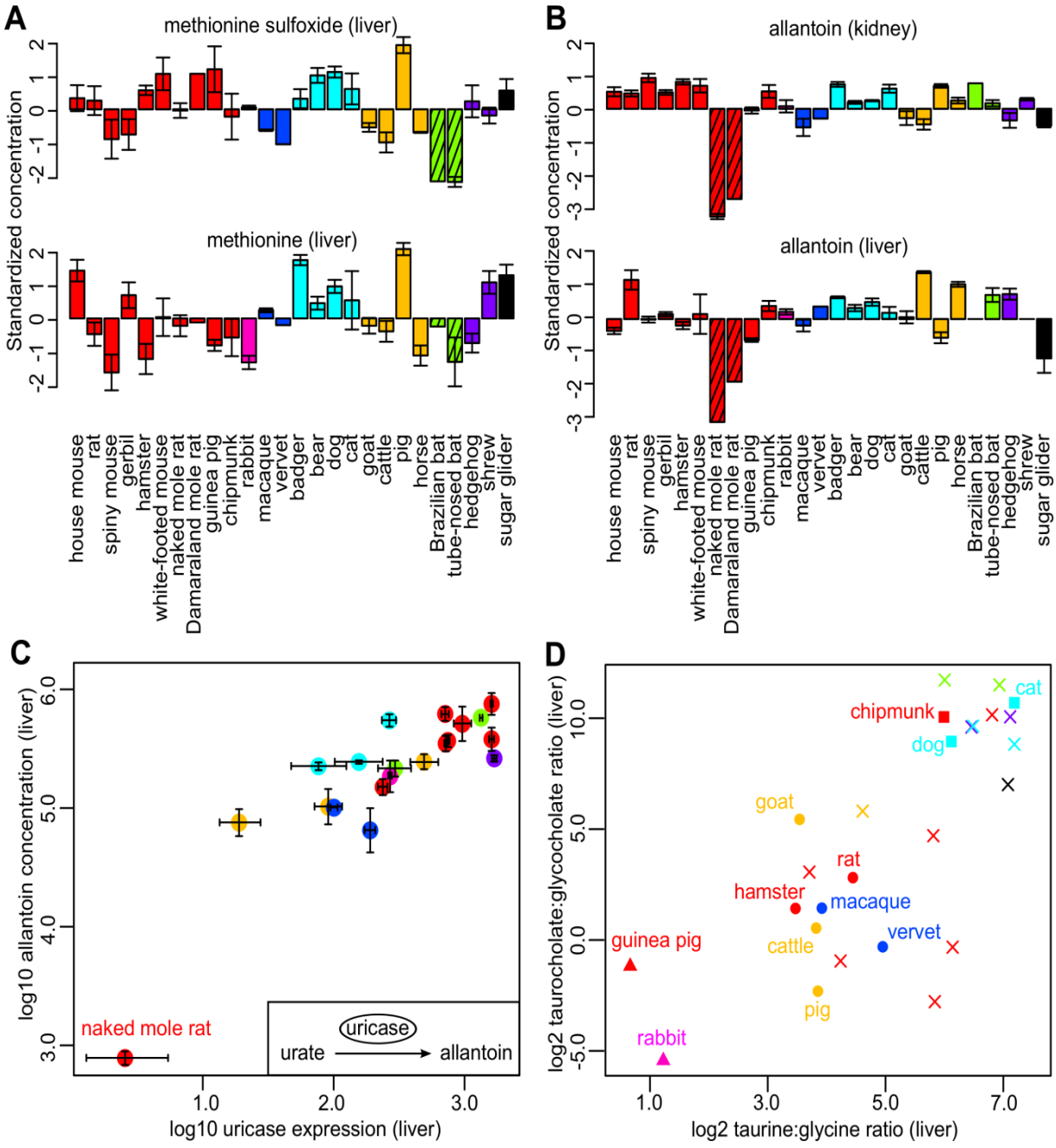
**(A) Bats have low levels of methionine sulfoxide in liver.** The error bars indicate standard errors (only for those with biological replicates). The species are colored according to taxonomical orders (same color scheme as in Figure 3.1A). The bars representing the bats are shaded.

**(B) African mole rats have low levels of allantoin in kidney and liver.** The bars representing the naked mole rat and Damaraland mole rat are shaded.

**(C) Liver allantoin levels correlate positively with uricase expression.** The error bars indicate standard errors in gene expression measurements (horizontal direction) or in metabolite measurements (vertical direction). The correlation relationship is robust (correlation coefficients using all points: Pearson = 0.86, Spearman = 0.78; excluding naked mole rat: Pearson = 0.76, Spearman = 0.74). Gene expression data were not available for Damaraland mole rat.

**(D) Use of taurine and glycine for bile acid conjugation among the mammals.** The species with known conjugated strategy are highlighted. Square (■): taurine-conjugation only; triangle (▲): glycine-conjugation only; circle (●): both taurine- and glycine-conjugation. Cross (×): conjugation strategy unknown.

**Figure 3.6 (Continued)**



## Metabolome remodeling and longevity variation

Next we examined the general trend in longevity and body mass across these species. We obtained the traits data from public databases (Carey and Judge, 2000; Tacutu et al., 2013) and focused primarily on Adult Weight (AW) and the longevity metrics Average Lifespan (AL), Maximum Lifespan (ML), Female Time to Maturity (FTM), as well as their body mass adjusted residuals (i.e. ALres, MLres, and FTMres, respectively) (Table 3.1). While AL and ML are most closely related to the concept of longevity, FTM can be measured more easily and may be less prone to reporting bias. They also correlated strongly with one another (Pearson correlation coefficient = 0.91 between AL and ML; 0.87 between AL and FTM; 0.84 between ML and FTM). Since adult weight correlates positively with lifespan, the longevity residuals were computed to remove the body mass influence. To account for the evolutionary relationship of the species, we performed regression by phylogenetic generalized least squares (Chapter 2, "Phylogenetic Regression") (Table 3.3). Different models of trait evolution were tested and within-species variations were incorporated in the calculation and we applied a two-step verification procedure to assess the robustness of the results (Chapter 2, "Phylogenetic Regression"). The results were also adjusted for False Discovery Rate. p value.robust < 0.01 was chosen as the cut-off and the top hits were tabulated across the organs and traits. (Table 3.3).

When the results were visualized on a heat map (Figure 3.7A), a few observations became apparent. Within each organ the correlations with the longevity metrics were largely consistent. Although the reported lifespan data were obtained from different databases and might not be entirely accurate, they did not significantly affect the calculated correlation, suggesting the results were robust to sample variation or measurement inaccuracy. On the other hand, the patterns were rather distinct across the four organs, suggesting the metabolites in different organs may follow different trajectories during evolution. By pooling the top hits (p value.robust < 0.01) of the two sets of longevity metrics (i.e. combining AL, ML, and FTM as one set; ALres, MLres, and FTMres as the other set), a number of positively and negatively correlating pathways were found to be enriched in each organ (Figure 3.7B).

## Body mass and longevity signatures

With respect to adult weight, creatinine (Crn) showed significant positive correlation in all four organs (p value.robust < $10^{-8}$; Figure 3.7C, Table 3.3). A related metabolite, creatine (Cr), also emerged as a top hit in heart and liver. It is well known that urinary and serum Crn levels increase with body mass (especially lean body mass) (Forbes and Bruining, 1976), as most Crn is derived from Cr in skeletal muscles and larger animals tend to have greater muscle mass. On the other hand, several glycerophospholipids (e.g. C16:0 LPE, C22:6 LPE, C18:0 LPC, C22:6 LPC) negatively correlated with body mass, especially in brain and heart (Figure 3.7B). A number of triacylglycerols showed significant but opposite trends in heart (positive correlation) and kidney (negative correlation) (Figure 3.7B).

In terms of the longevity traits, negative correlation was observed for amino acids, LPC, LPE, and metabolites involved in thiamine metabolism, whereas positive correlation was observed mainly for SM (Figure 3.7B, Table 3.3). LPC and LPE are generated by phospholipase-dependent hydrolysis of PC and PE, respectively. Phospholipase A2 ($PLA_2$) activity releases fatty acids such as arachidonic acid from sn-2 position of glycerol backbone of phospholipids and is commonly associated with inflammatory signaling in mammalian tissues. For example, elevated circulating lipoprotein-associated $PLA_2$ activity is linked to coronary artery disease in humans (Rosenson and Stafforini, 2012), supporting a potential inverse relationship between phospholipase activities (and hence LPC and LPE levels) and longevity.

Similar to the situation with body mass, TAG as a whole showed opposing trends to longevity in heart (positive) and kidney (negative). Closer examination revealed that the negative correlations in kidney were largely attributed to TAG with polyunsaturated fatty acid (PUFA) side chains (i.e. multiple double bonds, Figure 3.7D), whereas the positive correlations in heart were due to TAG with saturated or monounsaturated fatty acid (MUFA) side chains. A recent study on human plasma lipidomes of middle-aged offspring of nonagenarians revealed a signature of 19 lipid species associating with female familial longevity, including high levels of SM and low levels of PUFA TAG (Gonzalez-Covarrubias et al., 2013). Analysis of phospholipids in heart of a number of mammals also

revealed a negative correlation between double bond content and maximum lifespan (Pamplona et al., 2000). Naked mole rat tissues contain much lower levels of docosahexaenoic acid -containing (with 6 double bonds) phospholipids compared to mouse (Mitchell et al., 2007). Since PUFA are particularly sensitive to peroxidation damage, reduced level of polyunsaturated TAG in long-lived species may reflect their enhanced resistance to oxidative stress.

Allantoin correlated negatively with longevity in brain, kidney and liver, whereas urate showed some moderate positive correlation (Table 3.3). Furthermore, urate:allantoin ratio showed significant positive correlation with ML, ALres and MLres in kidney (p value.robust < $10^{-3}$; Figure 3.7E), indicating that long-lived mammals had higher urate and lower allantoin levels. The ranges of p values in kidney remain significant even when each species was left out one at a time (p value.max = $1.21\times10^{-2}$ for ML, $1.89\times10^{-3}$ for ALres, and $1.60\times10^{-4}$ for MLres), so the observation is generally applicable across the examined mammals and does not depend on any particular species. A previous study in primate and non-primate mammals also found significant positive correlation between maximum lifespan potential and urate concentration in serum and brain per specific metabolic rate (Cutler, 1984). Interestingly, humans have the highest serum urate level and are the longest-lived primates (Cutler, 1984). The naked mole rat, the longest-living rodent, also had very low levels of uricase expression in liver (Figure 3.6C). Allantoin can also be produced from urate by free radical oxidation (Kaur and Halliwell, 1990), and studies on human samples suggest high allantoin level may be a marker of oxidative stress (Kand'ar and Zakova, 2008; Yardim-Akaydin et al., 2006).

The liver concentrations of two tryptophan degradation products, anthranilic acid and kynurenine, showed robust negative correlation with longevity (Table 3.3). Over 95% of free tryptophan is degraded via the kynurenine pathway, with the first rate-limiting step catalyzed by indoleamine 2,3-dioxygenase (IDO) or tryptophan 2,3-dioxygenase (TDO). Anthranilic acid is produced from enzymatic hydrolysis of kynurenine. Several studies have linked tryptophan metabolism to aging and longevity. Knockdown of *tdo-2* gene in *C. elegans* can suppress the toxicity of aggregation-prone proteins and extend lifespan (van der Goot et al., 2012). Fruit flies with TDO deficiency live significantly longer than wild-type controls (Oxenkrug, 2010). In mammals, reducing dietary tryptophan can extend lifespan and delay age-related changes in rats and mice (De Marte and

Enesco, 1986; Segall and Timiras, 1976), and the kynurenine:tryptophan ratio in humans increases with aging (Capuron et al., 2011; Frick et al., 2004). In agreement, we also observed significant negative correlation with longevity for kynurenine:tryptophan ratio and anthranilic acid:tryptophan ratio in liver (Figure 3.7F, Table 3.3).

Reducing dietary amino acids levels has proved effective in lifespan extension (Grandison et al., 2009; Lee et al., 2014; Min and Tatar, 2006). Here the amino acids in brain showed negative correlation predominantly with FTM and FTMres (Figure 3.7B, Table 3.3), implying that mammals that mature more slowly (which are usually also longer-lived) tend to have lower levels of brain amino acids. The levels of branched chain amino acids such as leucine and isoleucine are also low in long-lived Ames dwarf mice (Wijeyesekera et al., 2012), which are defective in adenohypophyseal development and have stunted growth. 4-pyridoxate (catabolite of vitamin B6) in brain and thiamine (vitamin B1) in kidney and liver also negatively correlated with lifespan. They are required, respectively, for the synthesis of pyridoxal phosphate (PLP) and thiamine pyrophosphate (TPP), which are the essential cofactors for many enzymes involved in amino acid metabolism (Eliot and Kirsch, 2004; Lonsdale, 2006). Overall, the result is consistent with reduced rate of mass-specific metabolism in longer-lived mammals.

**Table 3.3. (see attached Excel file) Metabolites with significant correlation to body mass and longevity traits.** Phylogenetic regression was performed on each metabolite in each organ against Adult Weight (AW); Average Lifespan (AL); Maximum Lifespan (ML); Female Time to Maturity (FTM); Average Lifespan Residual (ALres); Maximum Lifespan Residual (MLres); and Female Time to Maturity Residual (FTMres). The values are "p value.robust": the regression slope p value after the point with the largest residual error was removed. Only those metabolites with p value.robust < 0.01 are shown. For those with positive correlation, the p values are rendered positive. For those with negative correlation, the p values are rendered negative.

**Figure 3.7. Metabolites correlating with body mass and longevity.**

**(A) Overview of correlation with body mass and longevity.** The grids represent the robust regression p value ("p value.robust") between metabolite levels in each organ and the indicated traits ("AW": Adult Weight; "AL": Average Lifespan; "ML": Maximum Lifespan; "FTM": Female Time to Maturity; "ALres": Average Lifespan Residual; "MLres": Maximum Lifespan Residual; "FTMres": Female Time to Maturity Residual). Only p value.robust < 0.01 are shown in color, with positive correlation in red and negative correlation in blue. Selected classes of metabolites are highlighted by rows (same color scheme as in Figure 3.5A). See Table 3.3 for more details.

**(B) Top pathways correlating with body mass and longevity.** The grids represent the pathway enrichment analysis p values (only p values < 0.01 are shown in color), with positive correlation in red and negative correlation in blue. For the purpose of enrichment analysis, the top hits (p value.robust < 0.01) in AL, ML, and FTM were pooled together, and the top hits in ALres, MLres and FTMres were also pooled together. "Monounsaturated TAG" refers to TAG with at most 2 double bonds in total. "Polyunsaturated TAG" refers to TAG with 3 or more double bonds.

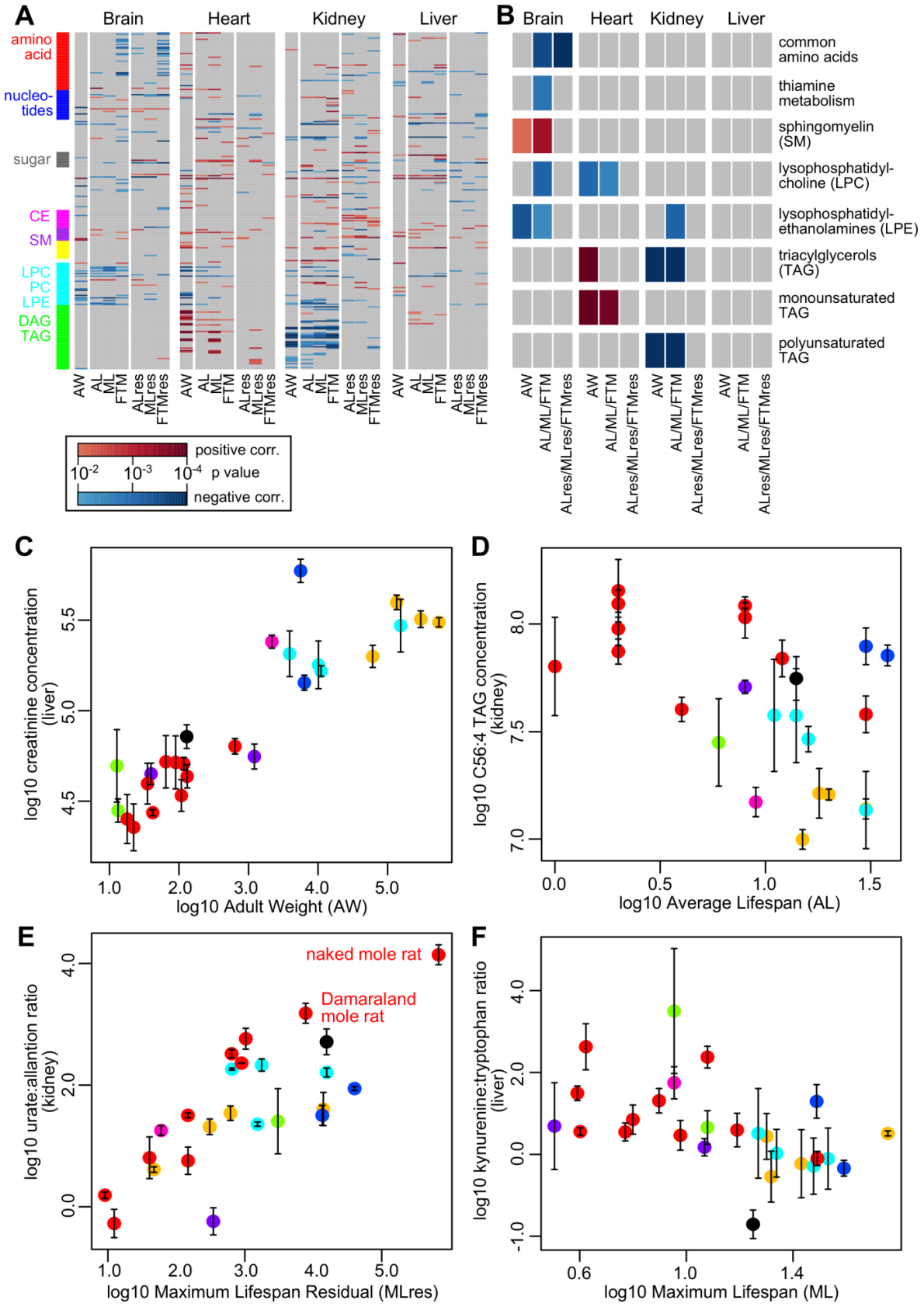
**(C) Liver creatinine level correlates positively with Adult Weight.** The vertical error bars indicate standard error. The points are colored according to taxonomical orders (same color scheme as in Figure 3.1A). Regression p value.robust = $1.01\times10^{-10}$; p value.max = $4.20\times10^{-10}$.

**(D) Kidney C56:4 TAG level correlates negatively with Average Lifespan.** Regression p value.robust = $9.75\times10^{-3}$; p value.max = $3.70\times10^{-2}$.

**(E) Kidney urate:allantoin ratio correlates positively with Maximum Lifespan Residual.** The points representing the naked mole rat and the Damaraland mole rat are indicated. Regression p value.robust = $8.41\times10^{-6}$; p value.max = $1.60\times10^{-4}$.

**(F) Liver kynurenine:tryptophan ratio correlates negatively with Maximum Lifespan.** Regression p value.robust = $7.23\times10^{-3}$; p value.max = $1.89\times10^{-2}$.

**Figure 3.7 (Continued)**



## Insights from the analysis of long-lived mouse models

To compare our results with established long-lived animal models, we performed metabolite profiling on brain and liver tissues of mice under caloric restriction (CR), rapamycin treatment (RAP), acarbose treatment (ACA), as well as GHRKO and Snell dwarf mice (Snell), against their respective wild type controls under control diets (Table 3.4). Five age-matched (approximately one-year old when sacrificed) biological replicates were collected for each condition, with both males and females for CR, RAP and ACA, and males only for GHRKO and Snell (Table 3.4). In total, 358 metabolites were reliably quantified across the 120 samples, and 241 of these metabolites overlapped with the mammalian dataset (Figure 3.8). Across the biological replicates, over 90% of the measurements had coefficient of variation < 0.06 (Figure 3.8A). Segregation of the samples in each organ was examined by hierarchical clustering (Figures 3.8D and 3.8E).

**Table 3.4. Five long-lived mouse models.** Sex, age of sacrifice and weight are indicated.

| Treatment | Sex | Age (days) | Weight (gram) | Sample labels | |
|---|---|---|---|---|---|
| | | | | Brain | Liver |
| Control | F | 367 | 43.6 | Brain.Control.F.1 | Liver.Control.F.1 |
| Control | F | 367 | 49.8 | Brain.Control.F.2 | Liver.Control.F.2 |
| Control | F | 374 | 28.3 | Brain.Control.F.3 | Liver.Control.F.3 |
| Control | F | 374 | 39.9 | Brain.Control.F.4 | Liver.Control.F.4 |
| Control | F | 392 | 30.1 | Brain.Control.F.5 | Liver.Control.F.5 |
| Caloric restriction | F | 369 | 26.5 | Brain.CR.F.1 | Liver.CR.F.1 |
| Caloric restriction | F | 369 | 21.5 | Brain.CR.F.2 | Liver.CR.F.2 |
| Caloric restriction | F | 376 | 22.9 | Brain.CR.F.3 | Liver.CR.F.3 |
| Caloric restriction | F | 376 | 25.5 | Brain.CR.F.4 | Liver.CR.F.4 |
| Caloric restriction | F | 373 | 26.7 | Brain.CR.F.5 | Liver.CR.F.5 |
| Rapamycin | F | 369 | 29.5 | Brain.Rapamycin.F.1 | Liver.Rapamycin.F.1 |
| Rapamycin | F | 369 | 26.3 | Brain.Rapamycin.F.2 | Liver.Rapamycin.F.2 |
| Rapamycin | F | 376 | 35.6 | Brain.Rapamycin.F.3 | Liver.Rapamycin.F.3 |
| Rapamycin | F | 376 | 30.6 | Brain.Rapamycin.F.4 | Liver.Rapamycin.F.4 |
| Rapamycin | F | 378 | 33.9 | Brain.Rapamycin.F.5 | Liver.Rapamycin.F.5 |
| Acarbose | F | 375 | 29.5 | Brain.Acarbose.F.1 | Liver.Acarbose.F.1 |
| Acarbose | F | 376 | 29.3 | Brain.Acarbose.F.2 | Liver.Acarbose.F.2 |
| Acarbose | F | 376 | 30.4 | Brain.Acarbose.F.3 | Liver.Acarbose.F.3 |
| Acarbose | F | 376 | 28.4 | Brain.Acarbose.F.4 | Liver.Acarbose.F.4 |
| Acarbose | F | 376 | 28.7 | Brain.Acarbose.F.5 | Liver.Acarbose.F.5 |

**Table 3.4 (Continued)**

| Treatment | Sex | Age (days) | Weight (gram) | Sample labels | |
|---|---|---|---|---|---|
| | | | | Brain | Liver |
| Control | M | 364 | 40.6 | Brain.Control.M.1 | Liver.Control.M.1 |
| Control | M | 362 | 30.7 | Brain.Control.M.2 | Liver.Control.M.2 |
| Control | M | 362 | 40.7 | Brain.Control.M.3 | Liver.Control.M.3 |
| Control | M | 381 | 45.8 | -- | Liver.Control.M.4 |
| Control | M | 376 | 43.2 | Brain.Control.M.5 | Liver.Control.M.5 |
| Control | M | 371 | 47.1 | Brain.Control.M.6 | -- |
| Caloric restriction | M | 371 | 37.1 | Brain.CR.M.1 | Liver.CR.M.1 |
| Caloric restriction | M | 371 | 31.7 | Brain.CR.M.2 | Liver.CR.M.2 |
| Caloric restriction | M | 373 | 33.3 | Brain.CR.M.3 | Liver.CR.M.3 |
| Caloric restriction | M | 372 | 32.5 | Brain.CR.M.4 | Liver.CR.M.4 |
| Caloric restriction | M | 377 | 36.1 | Brain.CR.M.5 | Liver.CR.M.5 |
| Rapamycin | M | 364 | 37.9 | Brain.Rapamycin.M.1 | Liver.Rapamycin.M.1 |
| Rapamycin | M | 362 | 33 | Brain.Rapamycin.M.2 | Liver.Rapamycin.M.2 |
| Rapamycin | M | 362 | 40.5 | Brain.Rapamycin.M.3 | Liver.Rapamycin.M.3 |
| Rapamycin | M | 372 | 41.5 | Brain.Rapamycin.M.4 | Liver.Rapamycin.M.4 |
| Rapamycin | M | 377 | 28.5 | Brain.Rapamycin.M.5 | Liver.Rapamycin.M.5 |
| Acarbose | M | 369 | 36.3 | Brain.Acarbose.M.1 | Liver.Acarbose.M.1 |
| Acarbose | M | 369 | 41.7 | Brain.Acarbose.M.2 | Liver.Acarbose.M.2 |
| Acarbose | M | 381 | 34.2 | Brain.Acarbose.M.3 | Liver.Acarbose.M.3 |
| Acarbose | M | 376 | 34.1 | Brain.Acarbose.M.4 | Liver.Acarbose.M.4 |
| Acarbose | M | 376 | 27.8 | Brain.Acarbose.M.5 | Liver.Acarbose.M.5 |
| Control | M | -- | -- | Brain.GHRWT.1 | Liver.GHRWT.1 |
| Control | M | -- | -- | Brain.GHRWT.2 | Liver.GHRWT.2 |
| Control | M | -- | -- | Brain.GHRWT.3 | Liver.GHRWT.3 |
| Control | M | -- | -- | Brain.GHRWT.4 | Liver.GHRWT.4 |
| Control | M | -- | -- | Brain.GHRWT.5 | Liver.GHRWT.5 |
| GHR knockout | M | -- | -- | Brain.GHRKO.1 | Liver.GHRKO.1 |
| GHR knockout | M | -- | -- | Brain.GHRKO.2 | Liver.GHRKO.2 |
| GHR knockout | M | -- | -- | Brain.GHRKO.3 | Liver.GHRKO.3 |
| GHR knockout | M | -- | -- | Brain.GHRKO.4 | Liver.GHRKO.4 |
| GHR knockout | M | -- | -- | Brain.GHRKO.5 | Liver.GHRKO.5 |
| Control | M | -- | -- | Brain.SnellWT.1 | Liver.SnellWT.1 |
| Control | M | -- | -- | Brain.SnellWT.2 | Liver.SnellWT.2 |
| Control | M | -- | -- | Brain.SnellWT.3 | Liver.SnellWT.3 |
| Control | M | -- | -- | Brain.SnellWT.4 | Liver.SnellWT.4 |
| Control | M | -- | -- | Brain.SnellWT.5 | Liver.SnellWT.5 |
| Snell mutant | M | -- | -- | Brain.SnellMut.1 | Liver.SnellMut.1 |
| Snell mutant | M | -- | -- | Brain.SnellMut.2 | Liver.SnellMut.2 |
| Snell mutant | M | -- | -- | Brain.SnellMut.3 | Liver.SnellMut.3 |
| Snell mutant | M | -- | -- | Brain.SnellMut.4 | Liver.SnellMut.4 |
| Snell mutant | M | -- | -- | Brain.SnellMut.5 | Liver.SnellMut.5 |

**Figure 3.8. Long-lived mouse model dataset quality assessment.**

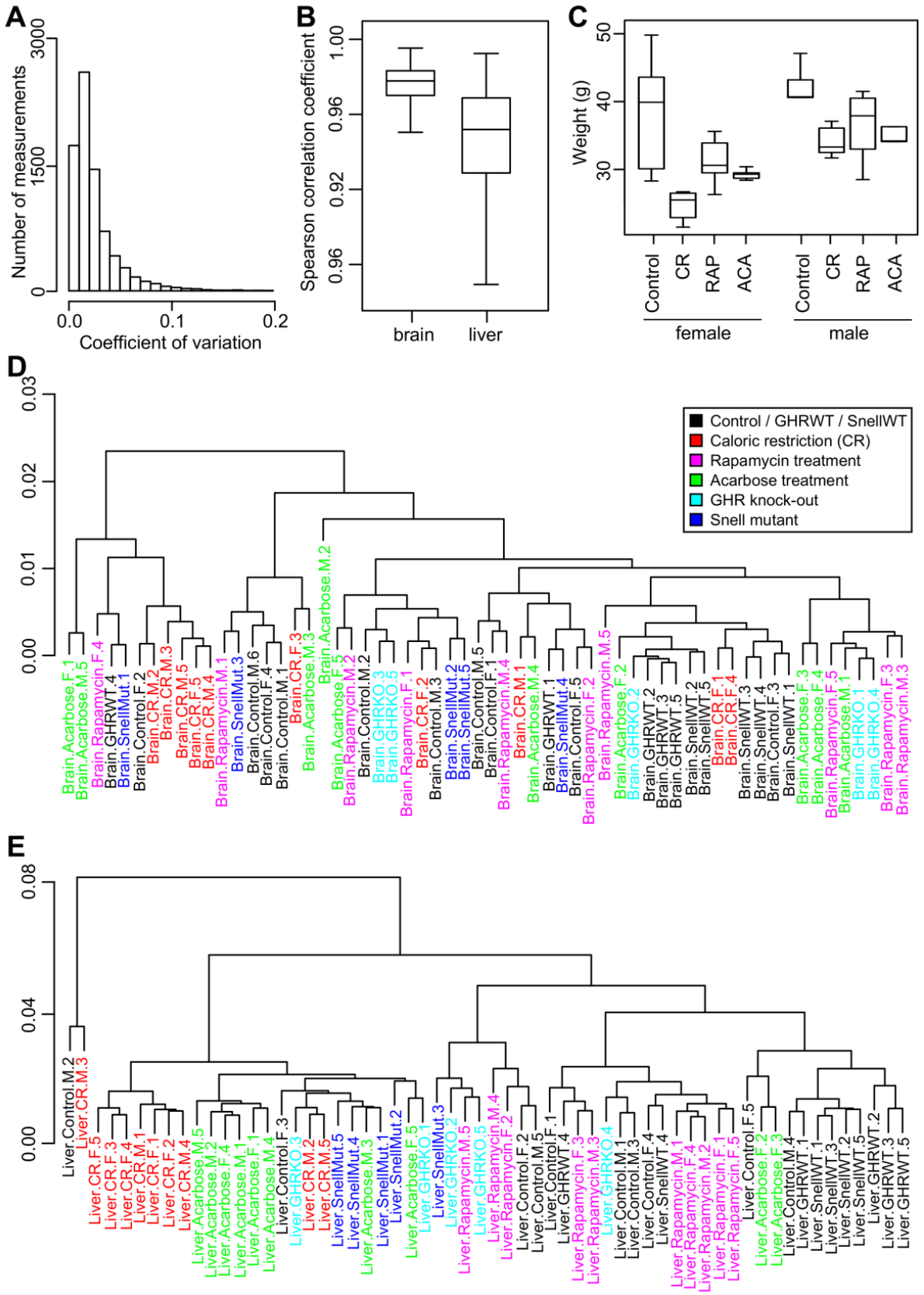
**(A) Coefficient of variation among biological replicates.** $90^{th}$ percentile: 0.05; $95^{th}$ percentile: 0.07.

**(B) Spearman correlation coefficients among brain samples and among liver samples.**

**(C) Weights of the animals of the long-lived models.** In both male and female, there is no significant difference in weights among mice under caloric restriction (CR), rapamycin treatment (RAP) and acarbose treatment (ACA) (pairwise t-test p value > 0.05).

**Clustering of the samples in (D) brain and (E) liver.** The samples are colored by treatment types. The hierarchical clustering was based on 1 minus Spearman correlation coefficient and used complete linkage.

**Figure 3.8 (Continued)**



We identified the metabolites differentially distributed between the long-lived mouse models and the corresponding controls and performed pathway enrichment analysis (Table 3.5, Figure 3.9 ). The long-lived mouse model dataset clustered with the mouse data in mammalian dataset (Figure 3.9C), indicating the overall metabolic signatures inherent in the species were well preserved. Interestingly, while a significant number of top hits were found in liver, the brain metabolite levels did not change much between the treatment and control (Figure 3.9A) and they were more conserved than those in liver (Figure 3.8B). The blood-brain barrier may help keep the brain metabolism in tight homeostasis and refractory to external modulations. The only exception was the Snell mice, which are defective in anterior pituitary development. Compared to control, the Snell mice brain likely exhibits a shift from oxidative phosphorylation towards glycolysis (Figure 3.9B).

In liver, CR, ACA and Snell mice produced very similar metabolic shifts, and these patterns were observed in both males and females (Figures 3.9A and 3.9B). Remarkably, there was extensive reduction in PUFA TAG levels across all these three models (Figures 3.9B and 3.9D, Table 3.5), which was consistent with the longevity signature we identified across the mammalian species and may indicate reduced susceptibility to peroxidation damage and oxidative stress in the long-lived mice. While the low PUFA TAG levels might be partly explained by the lower weights of these long-lived mice, this signature was not observed in GHRKO dwarf mice or in RAP mice. There were no significant differences in body weight among CR, RAP, and ACA in either gender (Figure 3.8C). In addition, the long-lived mouse models exhibited elevated levels of SM (in particular C14:0 SM, C16:0 SM, C18:0 SM and C18:1 SM), which also showed positive correlation in longevity in the mammalian species dataset. Previously, SM levels were reported to be low in old mice but at normal level in those under chronic CR (De Guzman et al., 2013), and were found to be high in the serum of centenarians (Montoliu et al., 2014). High SM levels are also associated with human female familial longevity (Gonzalez-Covarrubias et al., 2013). Sphingosine-linked fatty acids like ceramides are often regarded as “damage-associated molecular patterns” and may cause inflammatory damage by activating Nlrp3 inflammasome (De Guzman et al., 2013; Vandanmagsar et al., 2011). Elevated SM levels may also reflect reduced turnover to ceramides.

Other similarities as well as differences exist between our two datasets and those in the literature. For example, methionine is found at high levels in long-lived Ames dwarf mice, which may represent an increased methionine flux to transsulfuration and improved oxidative stress resistance (Wijeyesekera et al., 2012). Methionine level is also high in brain of male CR mice and liver of female ACA mice (Table 3.5). LPC levels were previously found to decrease with age but maintained in CR mice (De Guzman et al., 2013); in both our datasets they were low in long-lived animals. Furthermore, the mammalian dataset signatures of high urate:allantoin ratio and low kynurenine:tryptophan ratio were either insignificant or showed the opposite trends in the mouse models (Table 3.5).

To quantify the similarity between the longevity signatures from our two datasets, we counted the number of top hits in both datasets that had the same direction of correlation to longevity and compared that with the probability of getting similar results by chance (Experimental Procedures). The liver signatures of Snell, CR, and ACA mice matched very well to those based on AL, ML, and FTM in kidney of the mammalian dataset (Figure 3.9E). In addition, these liver signatures also clustered together (Figure 3.9F), suggesting lifespan extension by CR, acarbose treatment and in Snell mutants may affect certain common pathways, where rapamycin treatment and growth hormone receptor knockout may achieve lifespan extension via different mechanisms.

**Table 3.5. (see attached Excel file) Metabolites differentially distributed in long-lived mouse models.** Metabolites differentially distributed in brain samples and liver samples were identified with respect to the matching controls. Only those metabolites with p value < 0.01 are shown. For those with positive correlation, the p values are rendered positive. For those with negative correlation, the p values are rendered negative. CR: caloric restriction; RAP: rapamycin treatment; ACA: acarbose treatment; GHRKO: growth hormone receptor knockout; Snell: Snell dwarf mouse; F: female; M: male.

**Figure 3.9. Metabolites differentially distributed in long-lived mouse models.**

**(A) Overview of metabolite differential distribution.** The grids represent the linear model p values for differential distribution in the indicated long-lived mouse models with respect to their corresponding controls in brain and liver (CR: caloric restriction; RAP: rapamycin treatment; ACA: acarbose treatment; GHRKO: growth hormone receptor knockout; Snell: Snell dwarf mouse; F: female; M: male). Only p value < 0.01 are shown in color, with positive correlation in red and negative correlation in blue. Selected classes of metabolites are highlighted by rows (same color scheme as in Figure 3.5A). See Table 3.5 for more details.

**(B) Top enriched pathways.** The grids represent the pathway enrichment analysis p values (only p values < 0.01 are shown in color), with positive correlation in red and negative correlation in blue. For brain, only Snell is shown.

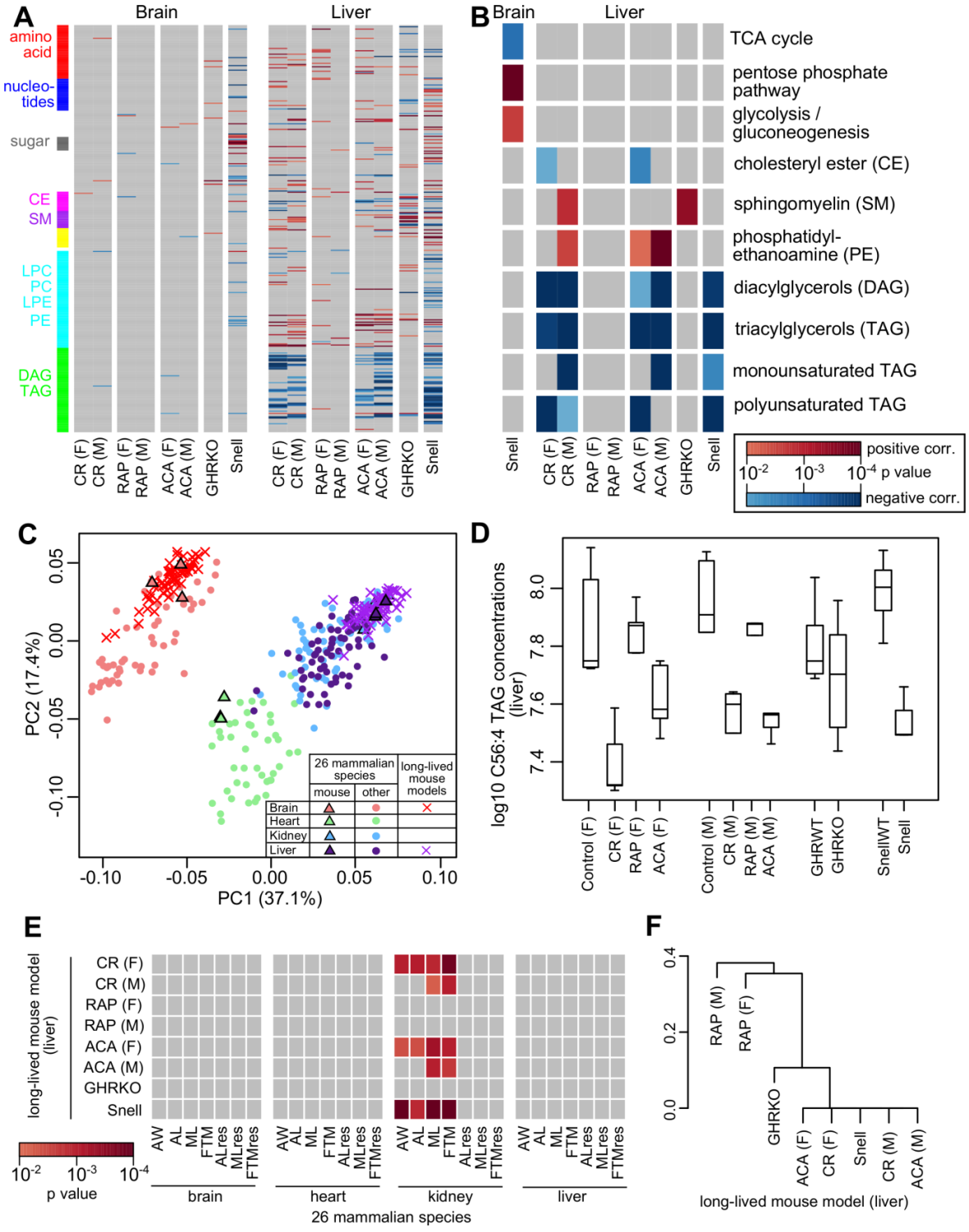
**(C) Long-lived mouse models data cluster well with mammalian species data.** Values in parenthesis indicate the percentage of variance explained by each Principal Component (PC). Biological replicates were treated as individual points.

**(D) Liver C56:4 TAG level across the long-lived mouse models.** C56:4 TAG levels were significant lower in CR(F) (p value = $2.21\times10^{-5}$), CR(M) (p value = $8.17\times10^{-3}$), ACA(F) (p value = $7.22\times10^{-3}$), and Snell (p value = $9.07\times10^{-5}$), compared to their respective controls.

**(E) Overlap among longevity signatures.** Between each pair of comparison, the numbers of metabolites with matching and opposite direction of correlation to longevity were calculated. The p value was based on binomial statistics, assuming equal probability of getting a match or a mismatch by chance. Only p value < 0.01 are shown in color.

**(F) Hierarchical clustering of the long-lived mouse models.** The distance matrix is based on the similarity among the longevity signatures (i.e. pairwise binomial p values). Only the liver data are shown. The mouse models are as shown above.

**Figure 3.9 (Continued)**



# DISCUSSION

Mammals have diversified dramatically over the tens of millions of years of evolution with remarkably different longevity profiles. How are their lifespans modulated by evolution while preserving competitiveness within their ecological niches? Which metabolites are involved and, more generally, how is metabolism adjusted in order to increase lifespan? While most of research on the control of lifespan was performed on single model organisms, our study addressed these questions by analyzing metabolite levels in several organs across the class of Mammalia. We found that metabolites in brain diverged less than in the other examined organs and the organ-differential distribution of metabolites represented their respective biological functions. The lineage-specific metabolite features we identified reflect known physiology of animals (e.g., low oxidative stress in bats) and also offer some new insights (e.g., bile acid conjugation strategies among mammals and diminished conversion of urate to allantoin in African mole rats). With regard to the longevity traits, we identified metabolites previously implicated in lifespan control as well as several new candidates. In particular, long-lived mammals were associated with low polyunsaturated triacylglycerols, low tryptophan degradation products, and low brain amino acids; as well as high sphingomyelin levels and high urate:allantoin ratio. Comparison of our signatures with the metabolite changes in long-lived mouse models indicated some overlap with mice under CR, mice treated with acarbose and Snell dwarf mice, especially for decrease in polyunsaturated triacylglycerols and increase in sphingomyelin. Similar changes were also previously reported in studies on human centenarians and other long-lived animal models. Furthermore, these three mouse models produced metabolite signatures distinct from those observed in rapamycin treatment and growth hormone receptor knockout mice, so the lifespan extension effects may have been achieved via different mechanisms.

Our study also reveals some unexpected complexities in analyzing metabolites and longevity. While some metabolites show consistent correlation with longevity traits across multiple organs, many patterns seem to be organ-specific. In the long-lived mouse models, many liver metabolites change significantly compared to the controls, but the brain exhibits very little perturbation. Furthermore, the longevity signatures in liver of the long-lived mouse models matched with the

kidney (but not liver) signature across the mammalian species, suggesting certain aspects of the mammalian longevity signatures may be distinct from the long-lived mouse models. In addition, the molecular mechanisms underlying the lifespan extension in these mouse models are not yet well-delineated, and differences among various long-lived mouse models have been previously reported. For example, in heart, kidney, and liver tissues, Snell and GHRKO mice showed different levels of chaperone mRNAs (Swindell et al., 2009). A low calorie diet is beneficial to Ames dwarf mice (Bartke et al., 2001) but not to GHRKO mice (Bonkowski et al., 2009). Expression of genes related to xenobiotic detoxification in liver are distinctly different among rapamycin-treated mice, CR mice (Miller et al., 2014), and GHR deletion mice (Li et al., 2013). CR mice also differ from rapamycin-treated mice in terms of leptin, FGF-21, and glucose tolerance (Lamming et al., 2013; Miller et al., 2014).

Compared with research that focuses on a single species, the current study benefited from the large effects of trait differences. While various factors such as feeding status, circadian cycle, gender, and body weight differences can introduce additional noise, ANOVA suggests that the variation between different species is generally much greater than variation among replicates of the same species. Even with the ablation of GHR or anterior pituitary, the brain and liver profiles of the long-lived mice still clustered well with the mouse data in the mammalian dataset and very similar longevity signatures were also obtained from both males and females of the same long-lived model. However, our study also suffers from a number of limitations. The current study does not prove causality between the metabolites and longevity traits, as the metabolite levels may influence and also be influenced by longevity. The number of metabolites quantified here only represents a fraction of the entire metabolome space and potentially important candidates may have been missed by our targeted approach. Many metabolites correlated strongly among one another and can inflate the signals observed. The metabolic fluxes through pathways and the metabolic changes during aging would not be reflected in our data either. While the biological implications of many metabolites identified here are far from fully understood, our study provides the first report of metabolite signatures of longevity across the mammalian spectrum, from which future studies should benefit.

## EXPERIMENTAL PROCEDURES

### Animal samples

Descriptions of the 26 mammalian species are provided in Table 3.1. The mammalian organ samples were obtained from various sources (Fushan et al., 2015). The animals were young adults and all were males, except for horse and vervet. Immediately after sacrificing, whole liver, kidney, heart, or frontal parts of brain were frozen in liquid nitrogen and stored at -80ºC until further use. To ensure comparability of data derived from homologous organs between species, each organ was ground in liquid nitrogen-cooled mortar and used for metabolite extraction. Most tissue samples were prepared in biological duplicates or triplicates (i.e. samples from different animals). Tissue samples were homogenized in water and normalized to protein concentration prior to metabolite analyses.

Descriptions of the five long-lived mouse models are provided in Table 3.4. All these models as well as genotype and diet matched controls were from the colonies at University of Michigan Medical School. Liver and brain cortex samples were taken at 12 months of age from male and female mice treated from 4 months of age with rapamycin (14.7 ppm, as in (Miller et al., 2014)), or acarbose (1000 ppm, as in (Harrison et al., 2014)), or from mice subjected to 40% dietary restriction, or from untreated littermate control mice of the genetically heterogeneous stock UM-HET3, in which each mouse was genetically unique but shared the same set of inbred grandparents (C57BL/6J, BALB/cByJ , C3H/HeJ, and DBA/2J). Liver and brain cortex samples from Snell dwarf (Flurkey et al., 2001) and GHRKO (Coschigano et al., 2003) males, and their corresponding littermate controls, were taken from young adults aged 4 to 6 months.

### Mass spectrometry quantification and normalization

To measure polar metabolites and lipids in tissue homogenates, we used three LC-MS methods as previously described (Townsend et al., 2013). Briefly, two targeted polar metabolite profiling methods were developed using reference standards of each metabolite to determine chromatographic retention times and mass-spec multiple reaction monitoring transitions, declustering potentials and collision energies. Negative ionization mode data were acquired using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Tissue

homogenates (30 μL) were extracted using 120 μL of 80% methanol (VWR) containing 0.05 ng/μL inosine-$^{15}N_4$, 0.05 ng/μL thymine-$d_4$, and 0.1 ng/μL glycocholate-$d_4$ as internal standards (Cambridge Isotope Laboratories). Positive ionization mode data were acquired using a 4000 QTRAP triple quadrupole mass spectrometer (AB SCIEX) coupled to an 1100 Series pump (Agilent) and an HTS PAL autosampler (Leap Technologies). Tissue homogenates (10 μL) were extracted using nine volumes of 74.9:24.9:0.2 (v/v/v) acetonitrile/methanol/formic acid containing stable isotope-labeled internal standards (0.2 ng/μL valine-d8, Isotec; and 0.2 ng/μL phenylalanine-d8 (Cambridge Isotope Laboratories)). Tissue homogenates (10 μL) were extracted for lipid analyses with 190 μL of isopropanol containing 1-dodecanoyl-2-tridecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids). MS analyses were carried out using electrospray ionization and Q1 scans in the positive ion mode. For each method, internal standard peak areas were monitored for quality control and MultiQuant 1.2 software (AB SCIEX) was used for automated peak integration. Metabolite peaks were manually reviewed for quality of integration and compared against a known standard to confirm identity.

**Data processing and quality assessment**

For the 26 mammalian species dataset, raw data were log10-transformed to conform to normal distribution; Shapiro–Wilk test confirmed assumption of normalcy was valid for over 75% of the measurements. Mean and standard error were computed across the biological replicates. Standardized concentrations (i.e. scaled to mean = 0 and standard deviation = 1) were used in cross-metabolite analysis. For the 5 long-lived mouse models dataset, raw data were log10-transformed and those metabolites missing in any one of the models in a particular organ were excluded from analysis in that organ. To render the two datasets comparable, the mean metabolite values in house mouse brain and liver of the mammalian species dataset were used as baselines to scale the long-lived mouse model dataset and R package “sva” was used to removed potential batch effects (Leek et al., 2014).

**Organ-specific phylograms**

The phylograms were constructed using the neighbor-joining (NJ) method (Saitou and Nei, 1987; Studier and Keppler, 1988) using sugar glider as the out-group. The distance matrix was based on 1 minus Spearman correlation coefficient. Reproducibility of the bifurcation pattern was assessed

using a 1000-time bootstrap procedure, by random sampling of a subset of the metabolite to build phylogram and repeating the procedure 1000 times. The degree of metabolite divergence was estimated using the average tip-to-root branch length of organ-specific phylogram. For the bootstrap procedure, one replicate per organ per species was randomly selected to assemble a pseudo-dataset for building phylogram. The procedure was repeated 1000 times to calculate the average tip-to-root branch length (excluding the branch leading to the out-group sugar glider). Similar results were produced by using only those species for which data were available for all four organs.

**Tip-to-root branch length simulation**

To investigate how various parameters might affect the tip-to-root branch length of NJ-phylogram, we simulated four scenarios ("Brownian motion", "Random noise", "OU model", and "Pagel's model) using R packages "phytools" (Revell, 2012) and "geiger" (Harmon et al., 2008) (Figure 3.4D). In each scenario, 300 simulations were run according to its parameter settings to generate a (300×26) dataset, mimicking the number of metabolites and species in the current study. A phylogram was constructed from each dataset using NJ method and the average tip-to-root branch length was calculated. The procedure was repeated 1000 times for each scenario.

**Phylogenetic signals**

More closely related species tend to resemble each other more than if they were drawn randomly from a phylogenetic tree, so their traits may be statistically non-independent. This phylogenetic relatedness, or "phylogenetic signal", can be detected using a number of metrics (Munkemuller et al., 2012). Pagel's lambda and Blomberg's K were computed using R package "phytools" (Revell, 2012). Those metabolites with Pagel's lambda > 0.9 and Blomberg's K > 1 were considered to have high phylogenetic signal.

**Pathway enrichment analysis**

Pathway information was obtained from ConsensusPathDB (Kamburov et al., 2009) and Human Metabolome Database (HMDB) (Wishart et al., 2013). For ConsensusPathDB, only pathways with known KEGG IDs were incorporated. For the lipids, customised pathways were created for sphingomyelin (SM); cholesterol ester (CE); monoacyl glycerophosphocholines (i.e. lysophosphatidylcholine (LPC)); diacyl glycerophosphocholines (i.e. phosphatidylcholine (PC));

monoacyl glycerophosphoenthanoamines (i.e. lysophosphatidylethanolamine (LPE)); diacyl glycerophosphoenthanoamines (i.e. phosphatidylethanolamine (PE)); monoacyl glycerols (MAG); diacyl glycerol (DAG); and triacyl glycerol (TAG). Acylcarnitines were further grouped into "short-chain" (up to 8 carbons), "medium-chain" (9 to 12 carbons), and "long-chain" (more than 12 carbons). Triacylglycerols were further grouped into monounsaturated TAG (MUFA-TAG, those with 2 or less double bounds in total) and polyunsaturated TAG (PUFA-TAG, those with more than 2 double bonds in total). Analysis was performed on pathways with at least 5 but less than 100 metabolites. Enrichment statistics was based on a hypergeometric distribution (Tavazoie et al., 1999). Odd ratios and expected counts were calculated as previously described (Gentleman et al., 2013).

**Organ-differential distribution of metabolites and lipid composition**

Paired Wilcoxon rank sum test was used to identify metabolites with organ-differential distribution for all combinations of organ pairs. To qualify as a top hit, a metabolite must show differential distribution (Bonferroni adjusted p value < 0.05) in at least 2 organ pairs. For lipid composition, the relative percentage abundance of individual lipid molecules within their own categories (i.e. TAG, LPC, LPE, PC, SM, or CE) were computed and compared with those previously reported in human plasma. Those lipid molecules with more than 10% relative abundance were considered the major species.

**Phylogenetic ANOVA**

To determine lineage-specific changes in metabolite levels, the species were grouped by taxonomical orders or families, and phylogenetic ANOVA was applied to determine if the concentration of a metabolite in one group was significantly different from that in other groups. A standard ANOVA assumes independence of observations, but this was not true in the current study as the animals were related phylogenetically. In phylogenetic ANOVA, the F-value of standard ANOVA is compared to a null distribution generated by stimulating trait evolution on a reference phylogeny, thus accounting for the non-independence of species. Phylogenetic ANOVA was performed using R package "phytools" (Revell, 2012).

**Regression by generalized least square and test for robustness**

See Chapter 2. The cut-off for top hits was p value.robust < 0.01.

**Differentially distributed metabolites in long-lived mouse models**

R package "limma" (Smyth, 2005) was used to identify differentially distributed metabolites between treatment and control groups in the long-lived mouse models. Pathway enrichment analysis was performed on the top hits (p value < 0.01).

**Longevity signature similarities**

Binomial statistics and 5000-time bootstrap procedure were used to assess the degree of similarity among the longevity signatures. Given any two signatures, the number of metabolites with matching directions of correlation to longevity and the number of metabolites with opposite directions of correlation to longevity were calculated. For binomial statistics, p values were computed by assuming equal probability of obtaining a match or a mismatch by chance. For the bootstrap procedure, metabolites were assigned matching or opposite directions randomly. P values were computed as the percentage of trials yielding greater number of matches (by chance) than the observed results.

# REFERENCES


Bartke, A., Wright, J.C., Mattison, J.A., Ingram, D.K., Miller, R.A., and Roth, G.S. (2001). Extending the lifespan of long-lived mice. Nature *414*, 412.

Berlett, B.S., and Stadtman, E.R. (1997). Protein oxidation in aging, disease, and oxidative stress. J. Biol. Chem. *272*, 20313-20316.

Blomberg, S.P., Garland, T., Jr., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution *57*, 717-745.

Bonkowski, M.S., Dominici, F.P., Arum, O., Rocha, J.S., Al Regaiey, K.A., Westbrook, R., Spong, A., Panici, J., Masternak, M.M., Kopchick, J.J.*, et al.* (2009). Disruption of growth hormone receptor prevents calorie restriction from improving insulin action and longevity. PLoS One *4*, e4567.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Brunet-Rossinni, A.K. (2004). Reduced free-radical production and extreme longevity in the little brown bat (Myotis lucifugus) versus two non-flying mammals. Mech. Ageing Dev. *125*, 11-20.

Buffenstein, R. (2008). Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. J. Comp. Physiol. B *178*, 439-445.

Buffenstein, R., Campbell, W.E., and Jarvis, J.U. (1985). Identification of crystalline allantoin in the urine of African Cricetidae (Rodentia) and its role in their water economy. J. Comp. Physiol. B *155*, 493-499.

Buffenstein, R., Edrey, Y.H., Yang, T., and Mele, J. (2008). The oxidative stress theory of aging: embattled or invincible? Insights from non-traditional model organisms. Age (Dordr) *30*, 99-109.

Capuron, L., Schroecksnadel, S., Feart, C., Aubert, A., Higueret, D., Barberger-Gateau, P., Laye, S., and Fuchs, D. (2011). Chronic low-grade inflammation in elderly persons is associated with altered tryptophan and tyrosine metabolism: role in neuropsychiatric symptoms. Biol. Psychiatry *70*, 175-182.

Carey, J.R., and Judge, D.S. (2000). Longevity records : life spans of mammals, birds, amphibians, reptiles, and fish (Odense, Odense University Press).

Coschigano, K.T., Holland, A.N., Riders, M.E., List, E.O., Flyvbjerg, A., and Kopchick, J.J. (2003). Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. Endocrinology *144*, 3799-3810.

Cutler, R.G. (1984). Urate and ascorbate: their possible roles as antioxidants in determining longevity of mammalian species. Arch. Gerontol. Geriatr. *3*, 321-348.

De Guzman, J.M., Ku, G., Fahey, R., Youm, Y.H., Kass, I., Ingram, D.K., Dixit, V.D., and Kheterpal, I. (2013). Chronic caloric restriction partially protects against age-related alteration in serum metabolome. Age (Dordr) *35*, 1091-1104.

De Marte, M.L., and Enesco, H.E. (1986). Influence of low tryptophan diet on survival and organ growth in mice. Mech. Ageing Dev. *36*, 161-171.

Eliot, A.C., and Kirsch, J.F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu. Rev. Biochem. *73*, 383-415.

Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007). LIPID MAPS online tools for lipid research. Nucleic Acids Res. *35*, W606-612.

Fang, X., Seim, I., Huang, Z., Gerashchenko, M.V., Xiong, Z., Turanov, A.A., Zhu, Y., Lobanov, A.V., Fan, D., Yim, S.H.*, et al.* (2014). Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. Cell Rep. *8*, 1354-1364.

Flurkey, K., Astle, C.M., and Harrison, D.E. (2010). Life extension by diet restriction and N-acetyl-L-cysteine in genetically heterogeneous mice. J. Gerontol. A Biol. Sci. Med. Sci. *65*, 1275-1284.

Flurkey, K., Papaconstantinou, J., Miller, R.A., and Harrison, D.E. (2001). Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proc. Natl. Acad. Sci. USA *98*, 6736-6741.

Forbes, G.B., and Bruining, G.J. (1976). Urinary creatinine excretion and lean body mass. Am. J. Clin. Nutr. *29*, 1359-1366.

Frick, B., Schroecksnadel, K., Neurauter, G., Leblhuber, F., and Fuchs, D. (2004). Increasing production of homocysteine and neopterin and degradation of tryptophan with older age. Clin. Biochem. *37*, 684-687.

Friedman, D.B., and Johnson, T.E. (1988). A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics *118*, 75-86.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Garland, T., Dickerman, A.W., Janis, C.M., and Jones, J.A. (1993). Phylogenetic Analysis of Covariance by Computer Simulation. Syst. Biol. *42*, 265-292.

Gentleman, R., Falcon, S., and Sarkar, D. (2013). Category: Category Analysis. R package.

Gonzalez-Covarrubias, V., Beekman, M., Uh, H.W., Dane, A., Troost, J., Paliukhovich, I., van der Kloet, F.M., Houwing-Duistermaat, J., Vreeken, R.J., Hankemeier, T.*, et al.* (2013). Lipidomics of familial longevity. Aging Cell *12*, 426-434.

Grandison, R.C., Piper, M.D., and Partridge, L. (2009). Amino-acid imbalance explains extension of lifespan by dietary restriction in Drosophila. Nature *462*, 1061-1064.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. Bioinformatics *24*, 129-131.

Harrison, D.E., Strong, R., Allison, D.B., Ames, B.N., Astle, C.M., Atamna, H., Fernandez, E., Flurkey, K., Javors, M.A., Nadon, N.L.*, et al.* (2014). Acarbose, 17-alpha-estradiol, and nordihydroguaiaretic acid extend mouse lifespan preferentially in males. Aging Cell *13*, 273-282.

Harrison, D.E., Strong, R., Sharp, Z.D., Nelson, J.F., Astle, C.M., Flurkey, K., Nadon, N.L., Wilkinson, J.E., Frenkel, K., Carter, C.S.*, et al.* (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature *460*, 392-395.

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloen, A., Even, P.C., Cervera, P., and Le Bouc, Y. (2003). IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature *421*, 182-187.

Hoppel, C. (2003). The role of carnitine in normal and altered fatty acid metabolism. Am. J. Kidney Dis. *41*, S4-12.

Huxtable, R.J. (2002). Expanding the Circle 1975–1999: Sulfur Biochemistry and Insights on the Biological Functions of Taurins. In Taurine 4 : taurine and excitable tissues, L. Corte, R.J. Huxtable, G. Sgaragli, and K.F. Tipton, eds. (Springer US), pp. 1-25.

Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2009). ConsensusPathDB--a database for integrating human functional interaction networks. Nucleic Acids Res. *37*, D623-628.

Kand'ar, R., and Zakova, P. (2008). Allantoin as a marker of oxidative stress in human erythrocytes. Clin. Chem. Lab. Med. *46*, 1270-1274.

Kaur, H., and Halliwell, B. (1990). Action of biologically-relevant oxidizing species upon uric acid. Identification of uric acid oxidation products. Chem. Biol. Interact. *73*, 235-247.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P.*, et al.* (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223-227.

Ladiges, W., Van Remmen, H., Strong, R., Ikeno, Y., Treuting, P., Rabinovitch, P., and Richardson, A. (2009). Lifespan extension in genetically modified mice. Aging Cell *8*, 346-352.

Lamming, D.W., Ye, L., Astle, C.M., Baur, J.A., Sabatini, D.M., and Harrison, D.E. (2013). Young and old genetically heterogeneous HET3 mice on a rapamycin diet are glucose intolerant but insulin sensitive. Aging Cell *12*, 712-718.

Lee, B.C., Kaya, A., Ma, S., Kim, G., Gerashchenko, M.V., Yim, S.H., Hu, Z., Harshman, L.G., and Gladyshev, V.N. (2014). Methionine restriction extends lifespan of Drosophila melanogaster under conditions of low amino-acid status. Nat. Commun. *5*, 3592.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2014). sva: Surrogate Variable Analysis. R package version 3.10.0.

Li, X., Bartke, A., Berryman, D.E., Funk, K., Kopchick, J.J., List, E.O., Sun, L.Y., and Miller, R.A. (2013). Direct and Indirect Effects of Growth Hormone Receptor Ablation on Liver Expression of Xenobiotic Metabolizing Genes. Am. J. Physiol. Endocrinol. Metab. *305*, E942-950.

Lin, S.J., Defossez, P.A., and Guarente, L. (2000). Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae. Science *289*, 2126-2128.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E.*, et al.* (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476-482.

Lonsdale, D. (2006). A review of the biochemistry, metabolism and clinical benefits of thiamin(e) and its derivatives. Evid. Based Complement. Alternat. Med. *3*, 49-59.

McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The Effect of Retarded Growth Upon the Length of Life Span and Upon the Ultimate Body Size: One Figure. J. Nutr. *10*, 63-79.

Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simao, T.L., Stadler, T.*, et al.* (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science *334*, 521-524.

Miller, R.A., Harrison, D.E., Astle, C.M., Fernandez, E., Flurkey, K., Han, M., Javors, M.A., Li, X., Nadon, N.L., Nelson, J.F.*, et al.* (2014). Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. Aging Cell *13*, 468-477.

Min, K.J., and Tatar, M. (2006). Restriction of amino acids extends lifespan in Drosophila melanogaster. Mech. Ageing Dev. *127*, 643-646.

Mitchell, T.W., Buffenstein, R., and Hulbert, A.J. (2007). Membrane phospholipid composition may contribute to exceptional longevity of the naked mole-rat (Heterocephalus glaber): a comparative study using shotgun lipidomics. Exp. Gerontol. *42*, 1053-1062.

Montoliu, I., Scherer, M., Beguelin, F., DaSilva, L., Mari, D., Salvioli, S., Martin, F.P., Capri, M., Bucci, L., Ostan, R.*, et al.* (2014). Serum profiling of healthy aging identifies phospho- and sphingolipid species as markers of human longevity. Aging (Albany NY) *6*, 9-25.

Munkemuller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. Methods Ecol. Evol. *3*, 743-756.

Ngo, T.C., and Assimos, D.G. (2007). Uric Acid nephrolithiasis: recent progress and future directions. Rev. Urol. *9*, 17-27.

Oxenkrug, G.F. (2010). The extended life span of Drosophila melanogaster eye-color (white and vermilion) mutants with impaired formation of kynurenine. J. Neural Transm. *117*, 23-26.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877-884.

Pamplona, R., Portero-Otin, M., Ruiz, C., Gredilla, R., Herrero, A., and Barja, G. (2000). Double bond content of phospholipids and lipid peroxidation negatively correlate with maximum longevity in the heart of mammals. Mech. Ageing Dev. *112*, 169-183.

Peters, R.H. (1986). The ecological implications of body size, Vol 2 (Cambridge University Press).

Quehenberger, O., Armando, A.M., Brown, A.H., Milne, S.B., Myers, D.S., Merrill, A.H., Bandyopadhyay, S., Jones, K.N., Kelly, S., Shaner, R.L.*, et al.* (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. J. Lipid Res. *51*, 3299-3305.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217-223.

Rosenson, R.S., and Stafforini, D.M. (2012). Modulation of oxidative stress, inflammation, and atherosclerosis by lipoprotein-associated phospholipase A2. J. Lipid Res. *53*, 1767-1782.

Sacher, G.A. (1959). Relation of Lifespan to Brain Weight and Body Weight in Mammals. In Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing), G.E.W. Wolstenholme, and M. O'Conner, eds. (Chichester, John Wiley & Sons, Ltd), pp. 115-141.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Salmon, A.B., Leonard, S., Masamsetti, V., Pierce, A., Podlutsky, A.J., Podlutskaya, N., Richardson, A., Austad, S.N., and Chaudhuri, A.R. (2009). The long lifespan of two bat species is correlated with resistance to protein oxidation and enhanced protein homeostasis. FASEB J. *23*, 2317-2326.

Segall, P.E., and Timiras, P.S. (1976). Patho-physiologic findings after chronic tryptophan deficiency in rats: a model for delayed growth and aging. Mech. Ageing Dev. *5*, 109-124.

Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L*., et al.* (2013). Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat. Commun. *4*, 2212.

Shi, Y., Buffenstein, R., Pulliam, D.A., and Van Remmen, H. (2010). Comparative studies of oxidative stress and mitochondrial function in aging. Integr. Comp. Biol. *50*, 869-879.

Smyth, G.K. (2005). Limma: linear models for microarray data (New York, Springer).

Studier, J.A., and Keppler, K.J. (1988). A note on the neighbor-joining algorithm of Saitou and Nei. Mol. Biol. Evol. *5*, 729-731.

Sun, L., Sadighi Akha, A.A., Miller, R.A., and Harper, J.M. (2009). Life-span extension in mice by preweaning food restriction and by methionine restriction in middle age. J. Gerontol. A Biol. Sci. Med. Sci. *64*, 711-722.

Swindell, W.R., Masternak, M.M., Kopchick, J.J., Conover, C.A., Bartke, A., and Miller, R.A. (2009). Endocrine regulation of heat shock protein mRNA levels in long-lived dwarf mice. Mech. Ageing Dev. *130*, 393-400.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-1033.

Tatar, M., Kopelman, A., Epstein, D., Tu, M.P., Yin, C.M., and Garofalo, R.S. (2001). A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science *292*, 107-110.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nat. Genet. *22*, 281-285.

Tissenbaum, H.A., and Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature *410*, 227-230.

Townsend, M.K., Clish, C.B., Kraft, P., Wu, C., Souza, A.L., Deik, A.A., Tworoger, S.S., and Wolpin, B.M. (2013). Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin. Chem. *59*, 1657-1667.

van de Poll, M.C., Soeters, P.B., Deutz, N.E., Fearon, K.C., and Dejong, C.H. (2004). Renal metabolism of amino acids: its role in interorgan amino acid exchange. Am. J. Clin. Nutr. *79*, 185-197.

van der Goot, A.T., Zhu, W., Vazquez-Manrique, R.P., Seinstra, R.I., Dettmer, K., Michels, H., Farina, F., Krijnen, J., Melki, R., Buijsman, R.C*., et al.* (2012). Delaying aging and the aging-associated decline in protein homeostasis by inhibition of tryptophan degradation. Proc. Natl. Acad. Sci. USA *109*, 14912-14917.

Vandanmagsar, B., Youm, Y.H., Ravussin, A., Galgani, J.E., Stadler, K., Mynatt, R.L., Ravussin, E., Stephens, J.M., and Dixit, V.D. (2011). The NLRP3 inflammasome instigates obesity-induced inflammation and insulin resistance. Nat. Med. *17*, 179-188.

Vaz, F.M., and Wanders, R.J. (2002). Carnitine biosynthesis in mammals. Biochem. J. *361*, 417-429.

Vellai, T., Takacs-Vellai, K., Zhang, Y., Kovacs, A.L., Orosz, L., and Muller, F. (2003). Genetics: influence of TOR kinase on lifespan in C. elegans. Nature *426*, 620.

Vessey, D.A. (1978). The biochemical basis for the conjugation of bile acids with either glycine or taurine. Biochem. J. *174*, 621-626.

Weindruch, R., Walford, R.L., Fligiel, S., and Guthrie, D. (1986). The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. J. Nutr. *116*, 641-654.

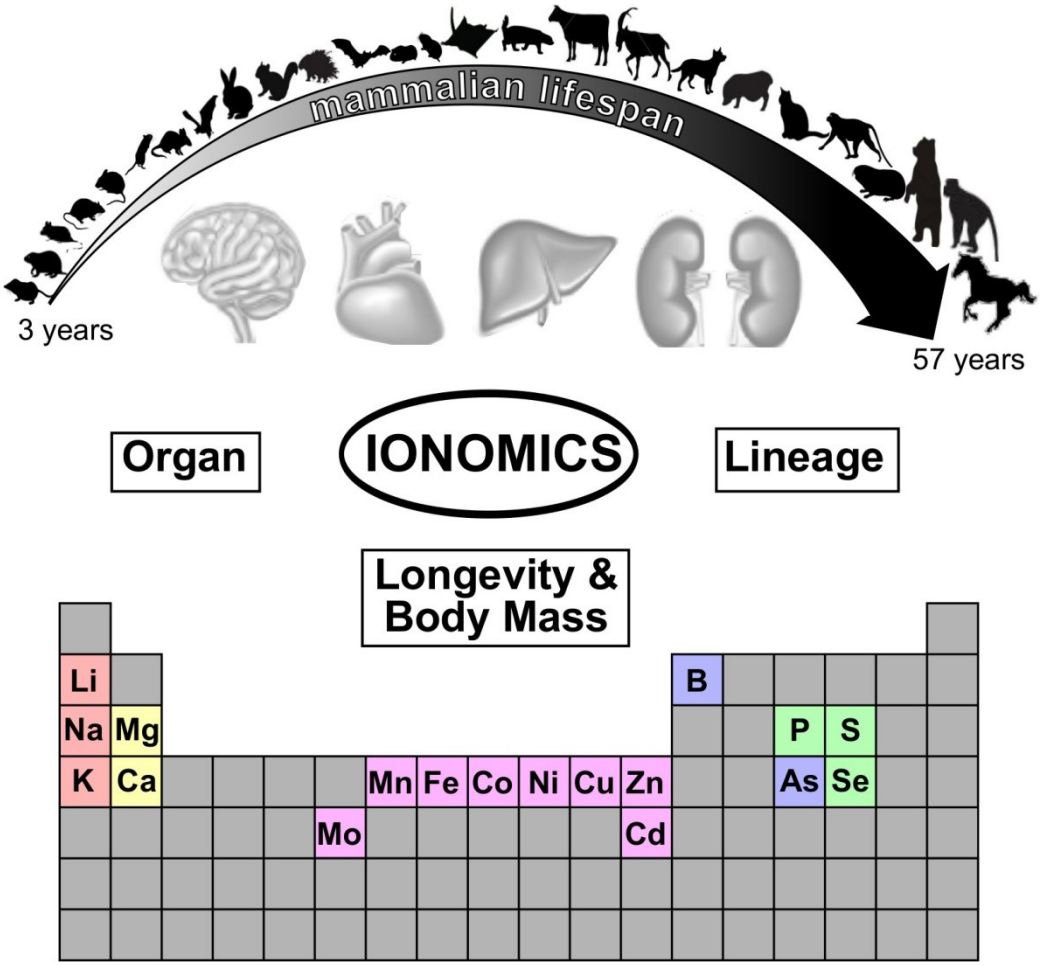
Western, D. (1979). Size, life history and ecology in mammals. Afr. J. Ecol. *17*, 185-204.

Wijeyesekera, A., Selman, C., Barton, R.H., Holmes, E., Nicholson, J.K., and Withers, D.J. (2012). Metabotyping of long-lived mice using 1H NMR spectroscopy. J. Proteome Res. *11*, 2224-2235.

Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E*., et al.* (2013). HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. *41*, D801-807.

Yardim-Akaydin, S., Sepici, A., Ozkan, Y., Simsek, B., and Sepici, V. (2006). Evaluation of allantoin levels as a new marker of oxidative stress in Behcet's disease. Scand. J. Rheumatol. *35*, 61-64.

# Chapter 4 Mammalian Ionome




This chapter is based on the following publication:

Siming Ma, Sang-Goo Lee, Eun Bae Kim, Thomas J. Park, Andrei Seluanov, Vera Gorbunova, Rochelle Buffenstein, Javier Seravalli, Vadim N. Gladyshev. **Organization of the Mammalian Ionome According to Organ Origin, Lineage Specialization, and Longevity.** *Cell Rep.,* 13 (2015), pp. 1319-26

**Author contributions**

S.M. carried out data analyses. S.G.L., E.B.K., T.J.P., A.S. V.G. and R.B. provided, collected and prepared samples. J.S. carried out ICP-MS measurements. All authors contributed to data interpretation. V.N.G. coordinated the study. S.M. and V.N.G. wrote the paper with input from all authors. We thank Anton Turanov and Sun Hee Yim for help with sample collection or preparation.

# ABSTRACT

Trace elements are essential to all mammals, but their distribution and utilization across species and organs remains unclear. Here, we examined 18 elements in the brain, heart, kidney, and liver of 26 mammalian species and report the elemental composition of these organs, the patterns of utilization across the species, and their correlation with body mass and longevity. Across the organs, we observed distinct distribution patterns for abundant elements, transition metals, and toxic elements. Some elements showed lineage-specific patterns, including reduced selenium utilization in African mole rats, and positive correlation between the number of selenocysteine residues in selenoprotein P and the selenium levels in liver and kidney across mammals. Body mass was linked positively to zinc levels, whereas species lifespan correlated positively with cadmium and negatively with selenium. This study provides insights into the variation of mammalian ionome by organ physiology, lineage specialization, body mass, and longevity.

# INTRODUCTION

The full set of elements used by organisms, or the ionome, supports diverse cellular functions (Eide et al., 2005; Salt et al., 2008). Transition metals alone are estimated to be required by more than one third of enzymes (Holm et al., 1996; Waldron and Robinson, 2009). Selenium (Se) and iodine (I) are used as components of proteins or hormones. Together with manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), copper (Cu), zinc (Zn), and molybdenum (Mo), these trace elements are needed only in minute quantities, but often act as important protein cofactors and active site components. Their deficiency or overload can result in severe pathological conditions (Fraga, 2005; Goldhaber, 2003).

In contrast, the metals sodium (Na), magnesium (Mg), potassium (K), calcium (Ca), as well as nonmetals phosphorus (P) and sulfur (S), are required in much larger quantities and are often called macronutrients. Some exist as free ions for establishing the electrochemical gradient across biological membranes (e.g. $Na^+$ and $K^+$); others reside in specific subcellular compartments as signaling molecules (e.g. $Ca^{2+}$). Many are constituents of macromolecules like proteins (e.g. sulfur) and nucleic acids (e.g. phosphate groups), or key structural components in bones, shells and exoskeletons (e.g. calcium phosphate minerals). Yet another group of elements, including lithium (Li), arsenic (As), and cadmium (Cd), are present in the environment and can be readily taken up by plants and animals, but have no apparent biological functions. Depending on the quantity, these elements elicit different biological responses, features that underlie both their use in medical treatments when applied in moderate concentrations, and their toxicity when absorbed in excess.

While a number of large-scale cross-species ionomics studies have been performed in plants (Ozaki et al., 2000; Watanabe et al., 2007; White et al., 2012), similar studies are lacking in mammals. In particular, the variation of element levels across organs, species, and lineages is not well understood. Since the use of these elements is likely shaped by evolution and environmental constraints, one may also be able to identify the links between the ionome and life-history traits (e.g., body mass, time to maturity, and longevity). Crucially, the nature of these questions means that one may need to look across a spectrum of organisms and organs to identify the common trends.

Recent advances in sequencing technology have enabled comparative genomics analyses to reveal the evolution of element utilization (Zhang and Gladyshev, 2009). In this study, we characterized the mammalian ionome by directly quantifying 18 elements in brain, heart, kidney and liver of 26 mammalian species, providing insights into the organization, distribution, and evolution of utilization of elements in mammals.

# RESULTS

## Conservation of the ionomes of mammalian organs

We analyzed 233 freshly frozen samples from the brain, heart, kidney, and liver of 26 mammalian species representing 10 taxonomic orders (Figure 4.1A and Table 4.1). All animals were young adults, and at least two biological replicates (i.e. different individual animals) were obtained for most species. The tissue concentrations of Li, B, Na, Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Mo, and Cd were quantified by four independent runs of inductively-coupled plasma mass spectrometry (ICP-MS) (Malinouski et al., 2014). After filtering and normalization, the final data quality was assessed graphically (Figure 4.2). The batch effect was removed using R package "sva" (Leek et al.) (Figures 4.2A and 4.2B). Over 90% of the measurements had coefficient of variation < 0.14 (Figure 4.2C).

Both principal component analysis (PCA) and heat map showed the samples generally clustered according to their organ origin (Figures 4.1B and 4.1C). The first three principal components (PCs) accounted for ~ 65% of the total variance (Figure 4.1B), suggesting the elemental composition of each organ was generally conserved. The heat map also revealed a few clusters of elements with similar distribution patterns (Figure 4.1C), such as the transition metals Mo, Mn, Co, and Fe; the various isotopes of Cu and Zn; as well as the macronutrients P, K, S and Mg.

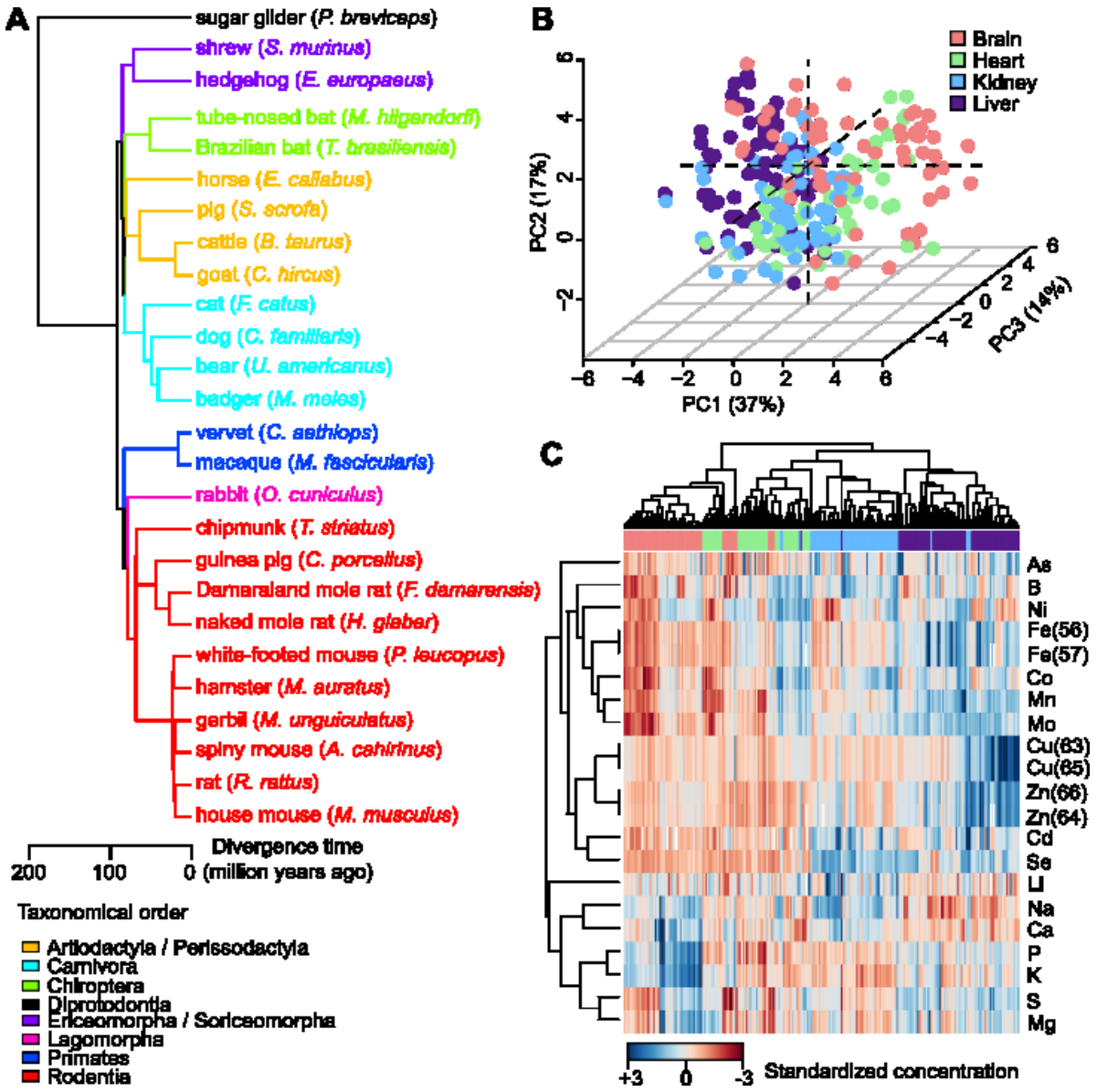



**Figure 4.1. Basic features of the mammalian ionome.**

**(A) Mammals examined in this study.** The species and their common names are indicated and the branches are colored according to taxonomical orders.

**(B) Principal Component Analysis reveals segregation of samples by organ origin.** Biological replicates are presented as individual points and colored by organ. The percentage variation explained by each principal component (PC) is indicated in parentheses.

**(C) Overview of the mammalian ionome using heat map.** Each row represents one element or isotope. Each column represents one biological sample (same color scheme as in (B)). Hierarchical clustering was performed using 1 minus Pearson correlation coefficient with average linkage.

**Table 4.1. 26 mammalian species samples and life history traits.** Adult Weight (AW), Maximum Lifespan (ML) and Female Time to Maturity (FTM) were obtained from AnAge database (Tacutu et al., 2013). Maximum Lifespan Residuals (MLres) and Female Time to Maturity Residuals (FTMres) were computed using the following allometric equations: $MLres = ML/(4.88 \times AW^{0.153})$; $FTMres = FTM/(78.1 \times AW^{0.217})$. Br: Brain; Ht: Heart; Kd: Kidney; Lv: Liver.

| Common Name | Scientific Name | Sex | No. of biological replicates | | | | AW (gram) | ML (year) | FTM (day) | MLres | FTMres |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Br | Ht | Kd | Lv | | | | | |
| sugar glider | *Petaurus breviceps* | M | 2 | 2 | 2 | 2 | 129.3 | 17.8 | 236 | 1.7335 | 1.0521 |
| shrew | *Suncus murinus* | M | 0 | 0 | 0 | 3 | 39.7 | 3.2 | 36 | 0.3733 | 0.2074 |
| hedgehog | *Erinaceus europaeus* | M | 2 | 2 | 2 | 2 | 1213.5 | 11.7 | 253 | 0.8089 | 0.6938 |
| tube-nosed bat | *Murina hilgendorfi* | M | 0 | 3 | 2 | 3 | 12.8 | 9 | 180 | 1.2493 | 1.3266 |
| Brazilian bat | *Tadarida brasiliensis* | M | 0 | 0 | 0 | 1 | 13.3 | 12 | 273 | 1.6551 | 1.9936 |
| horse | *Equus caballus* | F | 2 | 0 | 3 | 3 | 300000 | 57 | 914 | 1.6961 | 0.7582 |
| pig | *Sus scrofa* | M | 3 | 3 | 3 | 3 | 135000 | 27 | 334 | 0.9078 | 0.3295 |
| cattle | *Bos taurus* | M | 3 | 0 | 3 | 3 | 548500 | 20 | 548 | 0.5426 | 0.3988 |
| goat | *Capra hircus* | M | 3 | 3 | 3 | 3 | 61000 | 20.8 | 406 | 0.7897 | 0.4758 |
| cat | *Felis catus* | M | 3 | 3 | 3 | 3 | 3900 | 30 | 289 | 1.7349 | 0.6152 |
| dog | *Canis familiaris* | M | 3 | 3 | 3 | 3 | 10148.9 | 21.8 | 274 | 1.0891 | 0.4739 |
| bear | *Ursus americanus* | M | 3 | 0 | 3 | 3 | 154250 | 34 | 1278 | 1.1201 | 1.2247 |
| badger | *Meles meles* | M | 2 | 2 | 2 | 2 | 11050 | 18.6 | 365 | 0.9172 | 0.6198 |
| vervet | *Chlorocebus aethiops* | F | 1 | 1 | 1 | 1 | 5620 | 30.8 | 1034 | 1.6843 | 2.0332 |
| macaque | *Macaca fascicularis* | M | 2 | 2 | 2 | 2 | 6362.5 | 39 | 1238 | 2.0926 | 2.3696 |
| rabbit | *Oryctolagus cuniculus* | M | 3 | 3 | 3 | 3 | 2167.9 | 9 | 730 | 0.5694 | 1.765 |
| chipmunk | *Tamias striatus* | M | 2 | 2 | 2 | 2 | 89.6 | 9.5 | 187 | 0.9786 | 0.9027 |
| guinea pig | *Cavia porcellus* | M | 3 | 3 | 3 | 3 | 639.1 | 12 | 66 | 0.9152 | 0.208 |
| Damaraland mole rat | *Cryptomys damarensis* | M | 1 | 1 | 1 | 1 | 131.3 | 15.5 | 511 | 1.506 | 2.2704 |
| naked mole rat | *Heterocephalus glaber* | M | 2 | 0 | 1 | 4 | 35.3 | 31 | 228 | 3.6824 | 1.3472 |
| white-footed mouse | *Peromyscus leucopus* | M | 4 | 2 | 2 | 4 | 22.3 | 7.9 | 73 | 1.0067 | 0.4765 |
| hamster | *Mesocricetus auratus* | M | 3 | 3 | 3 | 3 | 108.2 | 3.9 | 48 | 0.3903 | 0.2224 |
| gerbil | *Meriones unguiculatus* | M | 3 | 3 | 3 | 3 | 64.8 | 6.3 | 43 | 0.6819 | 0.2227 |
| spiny mouse | *Acomys cahirinus* | M | 3 | 0 | 3 | 3 | 42 | 5.9 | 58 | 0.6825 | 0.33 |
| rat | *Rattus rattus* | M | 3 | 3 | 3 | 3 | 117 | 4.2 | 90 | 0.4153 | 0.41 |
| house mouse | *Mus musculus* | M | 3 | 3 | 3 | 3 | 18 | 4 | 42 | 0.5267 | 0.2872 |

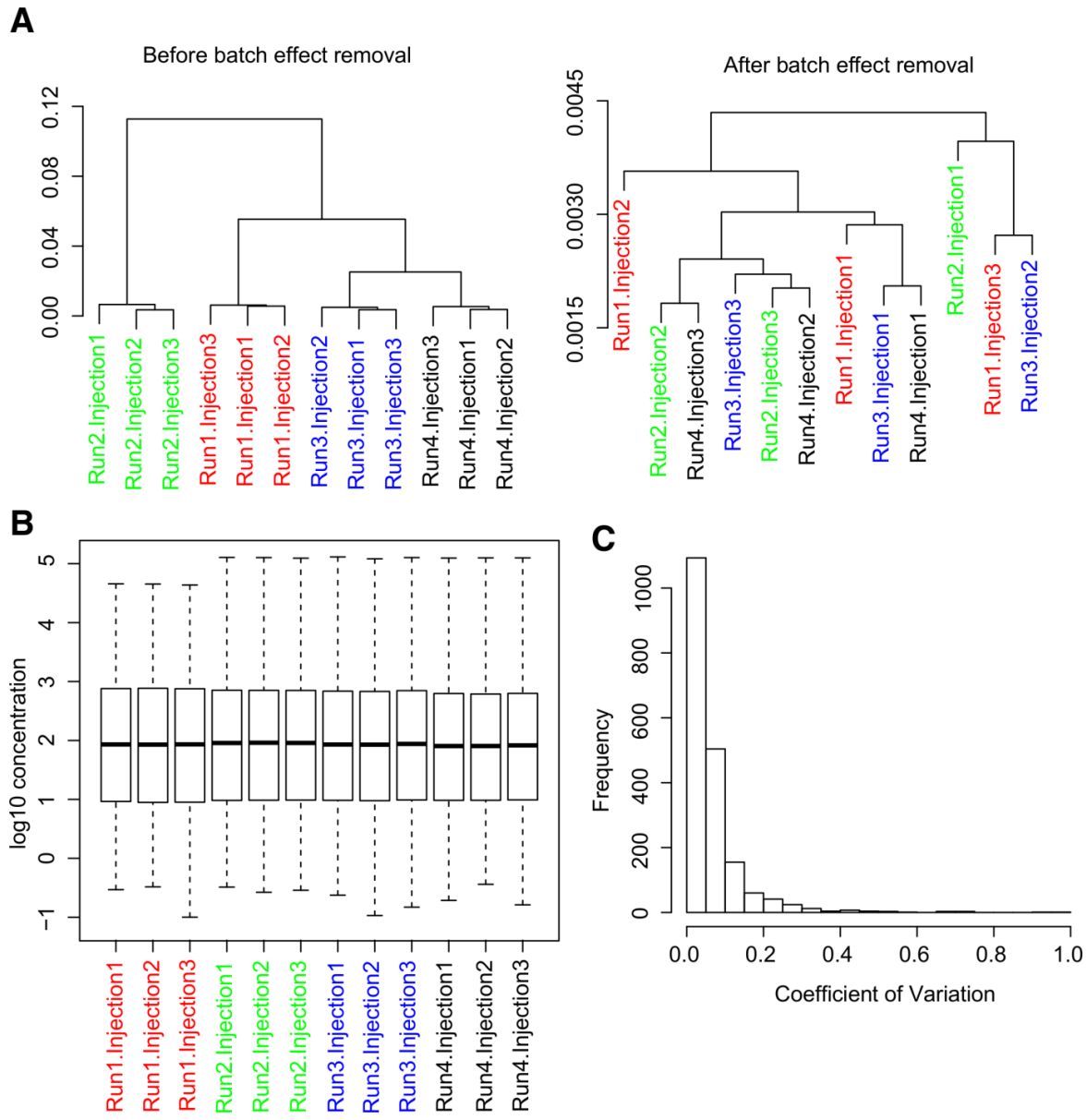



**Figure 4.2. Processing and normalization of data.**

**(A) Removal of batch effect.**

**(B) Results from all runs followed similar distribution after batch effect removal.**

**(C) Most samples had small coefficients of variation.** Coefficient of variation was calculated as standard deviation divided by absolute mean across the biological replicates ($90^{th}$ percentile = 0.14).

## Interactions among elements are indicative of biological functions

To explore the relationship among the elements, we computed Spearman correlation coefficients for all possible element pairs, using the full data (Figure 4.3A) and individually within each organ (Figure 4.4). As expected, the isotopes of the same element showed very strong correlations (all coefficients > 0.99), suggesting that mammals lack the ability to distinguish different isotopic forms. In addition, there were several clusters of elements with high correlations (Figures 4.3A and 4.3B), suggesting potential commonality in uptake mechanism or biological functions.

One such cluster consisting of Fe, Mo, Mn, Co and Ni was consistently observed in brain, heart, and kidney (all pair-wise correlation coefficients > 0.4; Figure 4.4). These transition metals often exist as divalent ions, some of which are known to interact. Divalent metal transporter 1 (DMT1) is a key metal transporter with a very broad substrate range that includes $Fe^{2+}$, $Mn^{2+}$, $Co^{2+}$, $Ni^{2+}$, and $Cd^{2+}$ ions (Gunshin et al., 1997; Mackenzie et al., 2007). $Mn^{2+}$ uptake can be coupled with $Fe^{2+}$ uptake via both transferrin-dependent and transferrin-independent pathways (Roth, 2006). Fe-S clusters and heme groups are required for biosynthesis of the prosthetic group Moco (molybdenum cofactor) and for functioning of most Mo enzymes (Hamza et al., 1998; Mendel and Bittner, 2006). A trace element study of 96 fern and fern ally species also reported correlation coefficient of 0.62 between Fe and Co concentrations (Ozaki et al., 2000). Although in vertebrates Ni has no known biological function, in bacteria and archaea Ni and Co utilize similar or common transport systems (Zhang et al., 2009).

Cu and Zn, two of the most abundant trace metals in the body, also correlated positively in multiple organs (Figure 4.4). Interaction between Cu and Zn at the physiological level is well documented, as Zn administration is used to antagonize Cu overload in patients with Wilson's disease (Hill and Link, 2009). Cellular transport of Cu and Zn depends on their respective transporters: ZnT and Zip transporter families for Zn; high affinity transporter CTR1, chaperone COX17, CSS, ATOX1, and P-type ATPase ATP7A, ATP7B for Cu (Amaravadi et al., 1997; Eide, 2006; Hamza et al., 1999; Liuzzi and Cousins, 2004). Zn is required for the catalytic activities of more than 200 enzymes and also serves a structural role in transcription factors, whereas Cu is found in several metabolic

enzymes, as well as cytochrome c oxidase (Complex IV) of the electron transport chain. Some enzymes (e.g. Cu/Zn superoxide dismutase) contain both elements in the active sites. Therefore, the observed correlation may reflect the physiological requirement for proper balance of Cu and Zn.

Interestingly, both Se and Cd were high in kidney and liver, and low in brain and heart (Figure 4.3C). While Se is mainly found in the form of selenocysteine (Sec) in selenoproteins, Cd has no known biological functions in higher organisms. Both elements are toxic at high concentrations and their levels in plants and animals often depend on the environmental concentrations in soil, water, or food (Clemens et al., 2013; Hurst et al., 2013). Another toxic element, As, also clustered very closely with Cd and Se (Figure 4.3B), suggesting that all three elements are treated by mammals as toxic species and are handled in a similar manner, although Se is primarily recognized as an essential element.

## Ionomes of mammalian organs

Next we analyzed the elemental distribution across the four organs (Figures 4.3C and 4.5A). Liver had the highest or the second highest levels of 15 out of the 21 elements and/or isotopes (Figure 4.3C), likely reflecting its central role in catabolism, anabolism, and detoxification. As electrolytes become concentrated during renal ultrafiltration and reabsorption, kidney had relatively high levels of Li, Na, and Ca. Brain and heart are less diverse in their metabolic functions and contained low levels of many elements. Nevertheless, P and K were found at the highest concentrations in brain, with Na and Ca also present at high levels. Our data also agreed well with a previous report of metal levels in humans (Katoh et al., 2002) (Figure 4.5B).

We also examined the gene expression patterns of some enzymes known to utilize trace elements (Figure 4.3D) (Fushan et al., 2015). Sulfite oxidase (SUOX), xanthine oxidase/ dehydrogenase (XDH), and aldehyde oxidase (AOX1) require the molybdenum cofactor (Moco) for their activities, whereas molybdenum cofactor biosynthesis protein 1 (MOCS1) catalyzes the biosynthesis of Moco. They were all highly expressed in liver and showed significant correlation with Mo concentration (Spearman correlation coefficient SUOX: 0.58; XDH: 0.49; AOX1: 0.58; MOCS1: 0.68). Arginase 1 (ARG1) binds Mn at the active site and showed strong correlation with Mn concentration (Spearman correlation coefficient 0.58). The selenoproteins glutathione peroxidase 1 (GPX1), thioredoxin reductases TXNRD1 and TXNRD2, and selenoprotein P (SEPP1) also showed significant correlation to Se levels (Spearman correlation coefficient GPX1: 0.64; TXNRD1: 0.27; TXNRD2: 0.40; SEPP1: 0.61).

**Figure 4.3. Correlation and distribution of elements.**

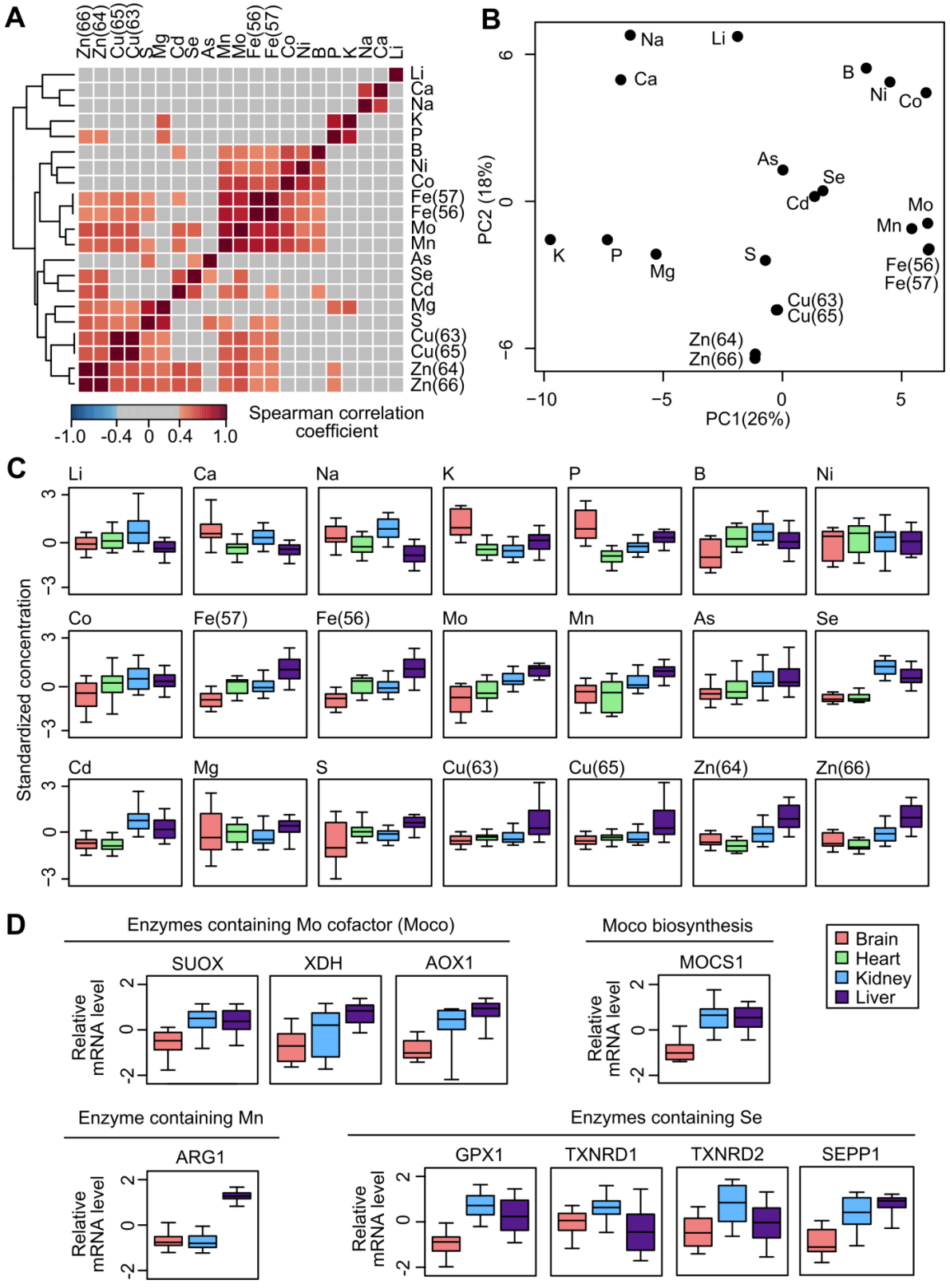
**(A) Elements form clusters of significant correlation.** Pairwise correlation coefficients among the elements are shown. Coefficients > 0.4 or < -0.4 (approximately equal to p value < 0.05) are highlighted in color. The correlations shown are based on data for all four organs.

**(B) Principal Component Analysis reveals distinct clusters of elements.** The elements and isotopes are projected on the first two principal components (PCs).

**(C) Distribution of elements across organs.** The box plots represent standardized concentrations in the brain, heart, kidney, and liver, with the central bands indicating median values and the whiskers indicating $5^{th}$ and $95^{th}$ percentiles.

**(D) Gene expression patterns of enzymes that utilize Mo, Mn and Se are consistent with the organ distribution of these elements.** Gene expression data are based on a previous study (Fushan et al., 2015); data for heart are not available. SUOX: sulfite oxide; XDH: xanthine oxidase/dehydrogenase; AOX1: aldehyde oxidase; MOCS1: Molybdenum cofactor biosynthesis protein 1; ARG1: arginase 1; GPX1: glutathione peroxidase 1; TXNRD1: thioredoxin reductase 1; TXNRD2: thioredoxin reductase 2; SEPP1: selenoprotein P.

**Figure 4.3 (Continued)**



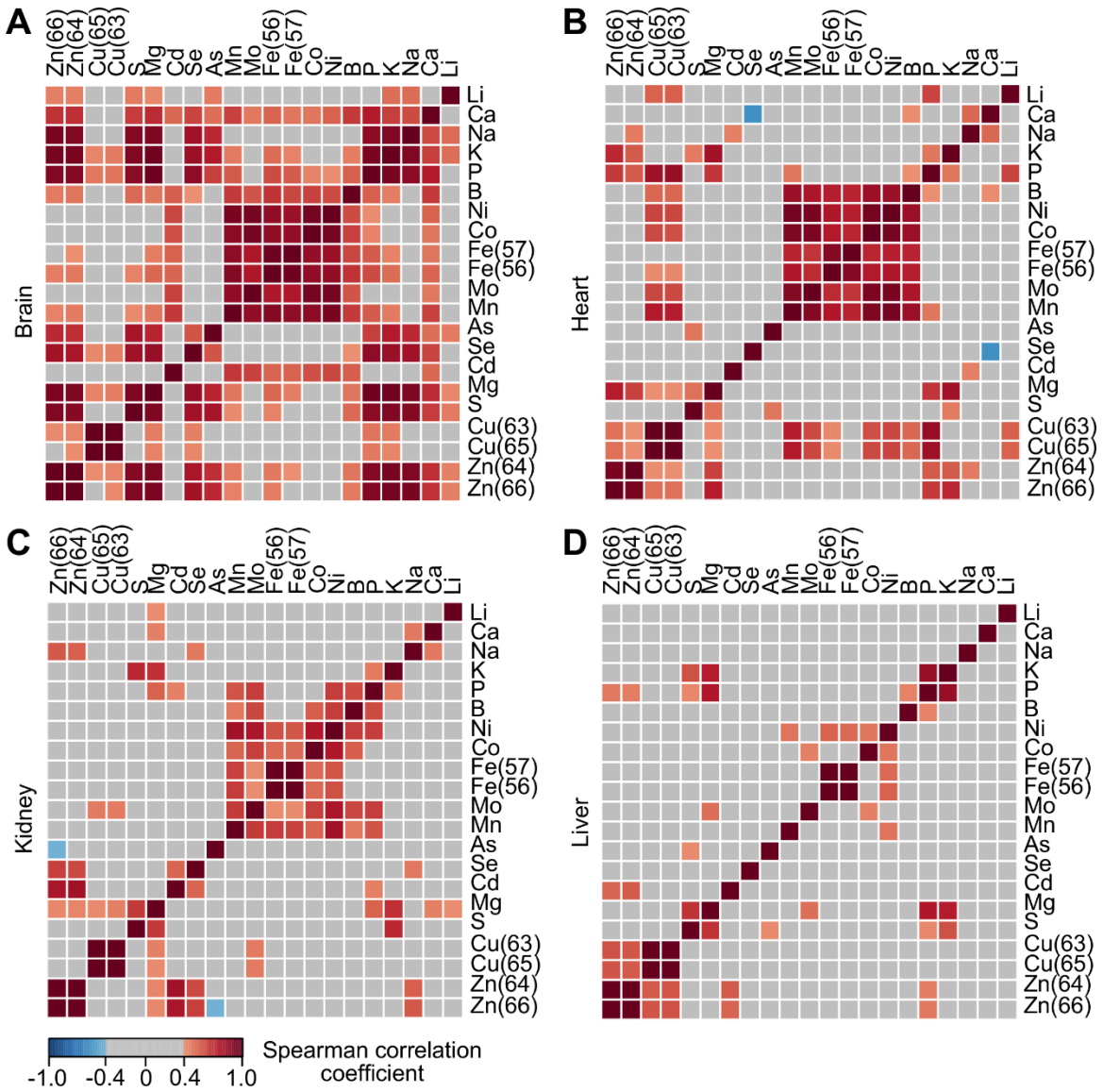



**Figure 4.4. Correlations among elements in (A) brain; (B) heart; (C) kidney; and (D) liver.**

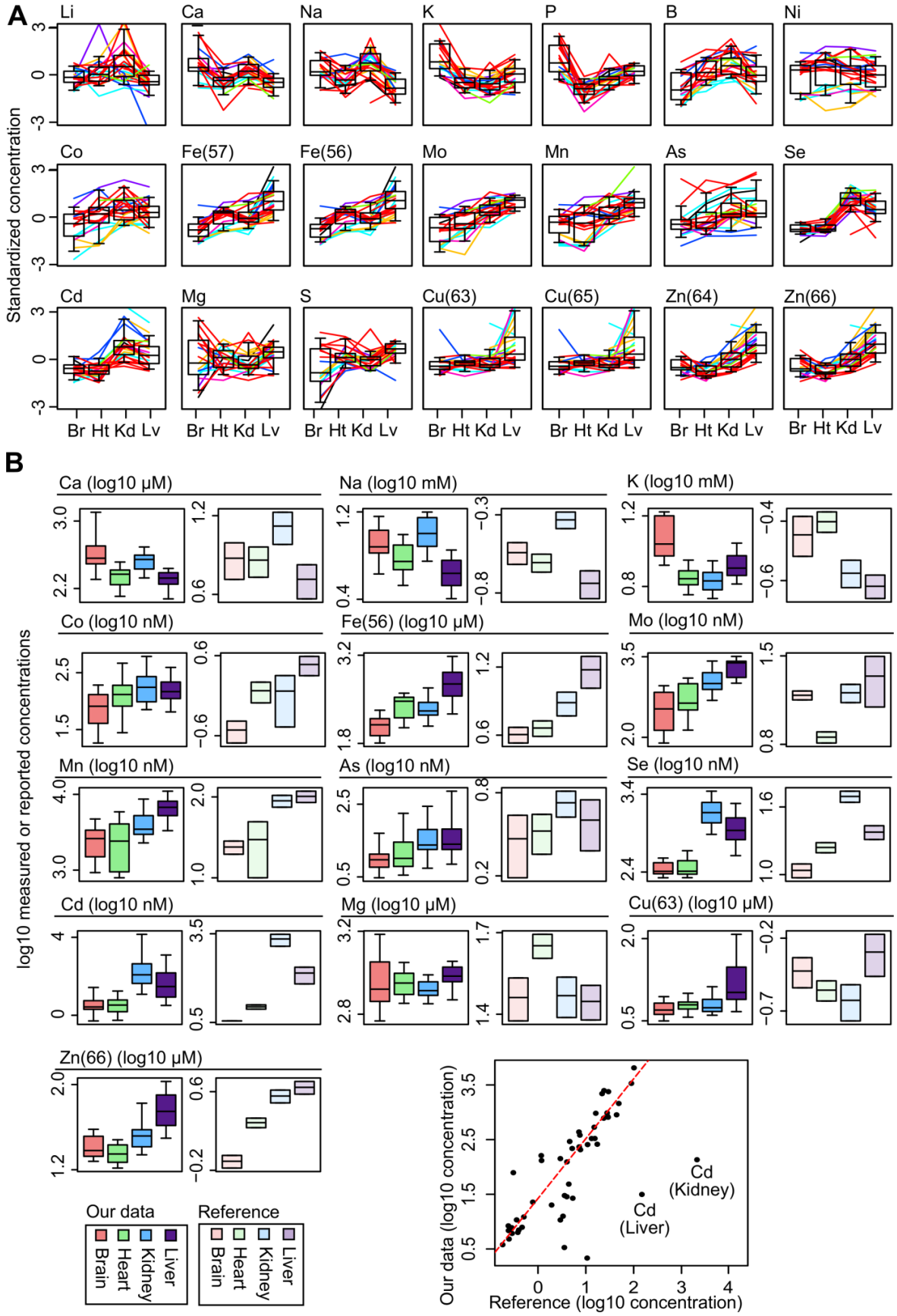
Only Spearman correlation coefficients > 0.4 or < -0.4 are highlighted in color. The correlations are based on the data in the respective organs.

**Figure 4.5. Organ differential distribution of elements.**

**(A) Trajectories of element distribution across the four organs. "Br", "Ht", "Kd", and "Lv" refer to brain, heart, kidney, and liver, respectively.** Each line represents one species and is colored according to taxonomical order.

**(B) The distribution patterns of elements across different organs.** For each element, the box plot on the left represents the log10 concentrations measured in our studies, and the box plot on the right represents the log10 concentrations (converted to the appropriate molar concentrations) reported in (Katoh et al., 2002) ("reference"). For our data, the central band represents the median value and the box represents first and third quartiles. For the reference data, the central band represents the mean value and the box represents ±0.5 standard deviation. The insert at the bottom shows a plot of the median values (across the elements and organs) in our study versus those in the reference (Spearman correlation coefficient = 0.81). Cd levels are the main outliers. The differences in absolute values are likely due to different amounts of tissues used.

**Figure 4.5 (Continued)**



## Lineage-specific distribution of elements

To identify lineage-specific patterns, i.e. whether some elements were significantly higher or lower in a particular group of related species, we quantified the phylogenetic signal using Pagel's lambda and Blomberg's K (Revell, 2012), and performed phylogenetic ANOVA (Harmon et al., 2008) to explicitly test for differential distribution in a particular taxonomical order or family.

Mg, S, P, and K in brain exhibited significant phylogenetic signals (Figure 4.6A). They were also enriched in the brain of rodents compared to non-rodents (phylogenetic ANOVA p value < 0.05 for Mg, S, K and 0.07 for P). The distribution patterns were very similar between Mg and S as well as between K and P, supporting our observation above that these elements clustered with strong correlation (Figure 4.3).

Among the animals examined were two species of desert-dwelling African mole rats (the naked mole rat and the Damaraland mole rat), whose genetic and physiological features are quite distinct from other rodents (Fang et al., 2014; Kim et al., 2011; LaVinka and Park, 2012; Maina et al., 2001). In particular, naked mole rats face significant oxidative stress, even though they are the longest lived rodents (Andziak et al., 2006; Buffenstein, 2005). In terms of element distribution, both had significantly lower levels of Se in kidney and liver than all other species (Figure 4.6B; phylogenetic ANOVA p value=0.007 for kidney and 0.005 for liver), and to a lesser extent, than other rodents (p value=0.13 for kidney and 0.16 for liver). Se levels and glutathione peroxidase activity (Andziak et al., 2005; Malinouski et al., 2012) are lower in the naked mole rat liver and kidney than in mouse tissues, in part due to an early stop codon in the sequence coding for GPX1 (Kasaikina et al., 2011). The same early stop codon is also found in Damaraland mole rat (Fang et al., 2014).

Mammalian selenoproteomes consist of 24-25 selenoproteins (Kryukov et al., 2003). Selenoprotein mRNAs have a structure called SECIS element that supports co-translational insertion of Sec at UGA codon. While most selenoproteins contain only single Sec residue, the number of Sec residues in plasma selenoprotein P (SelP, coded by SEPP1) vary greatly across different organisms (Lobanov et al., 2008). All but one of these Sec residues are found in the C-terminal region and SelP is involved in the transport of Se throughout the body. We observed the number of Sec residues in

SelP correlated remarkably with the measured Se levels in kidney and liver (Figures 4.6C; Pearson correlation coefficient = 0.62 for kidney and 0.42 for liver). For example, both naked mole rat and Damaraland mole rat had only 7 Sec residues and contained very little Se in kidney and liver, whereas pig, the carnivores (dog, bear, and badger), and the bats had 14-16 Sec residues in SelP and relatively high Se levels in both organs (Figure 4.7). Thus, the use of Se in mammals is tuned by altering the number of Sec residues in the C-terminal region of SelP.

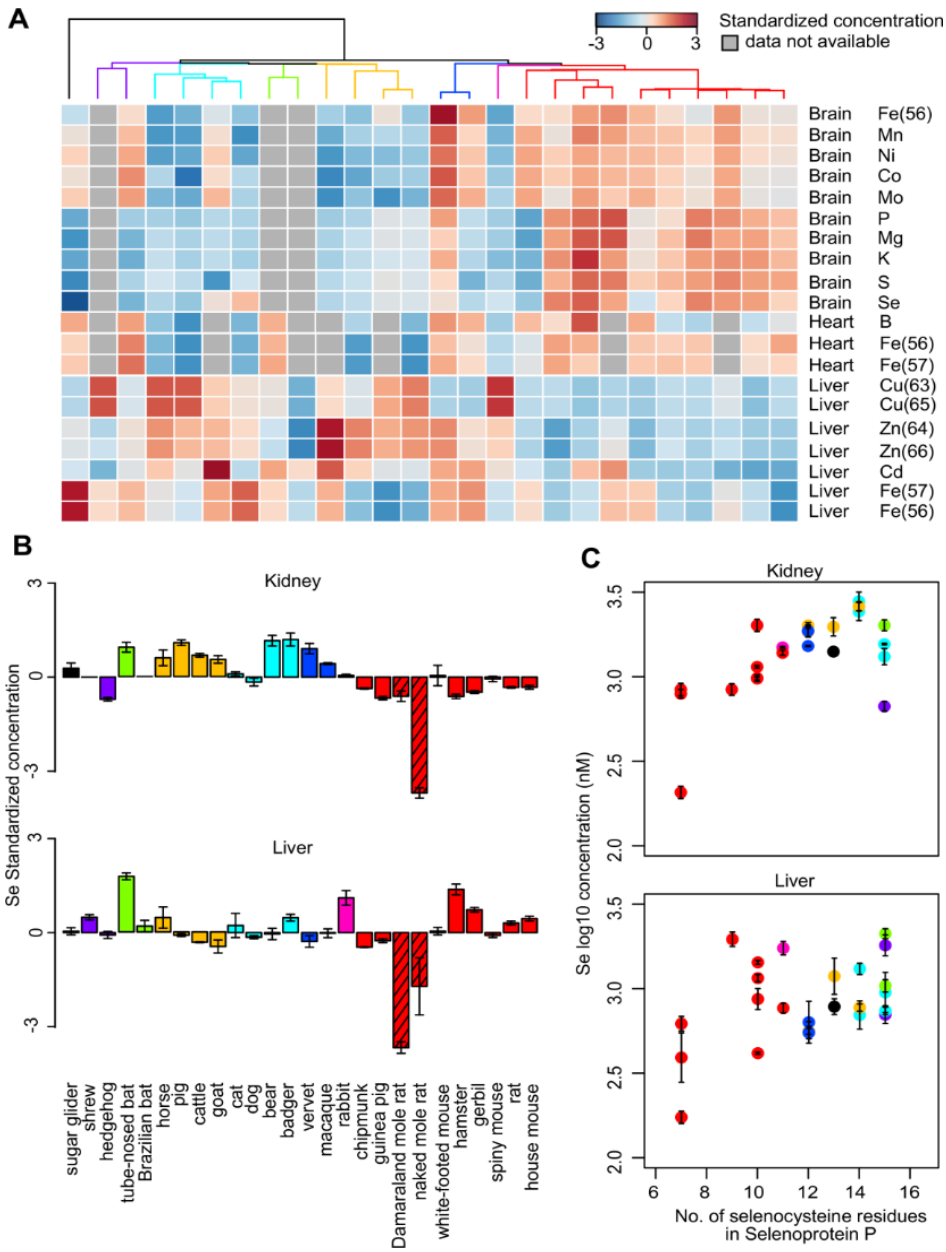



**Figure 4.6. Lineage specific distribution of elements.**

**(A) Elements with significant phylogenetic signals.** The grids are colored based on the standardized concentration and the columns are arranged by phylogenetic relationship. Only those elements with p values < 0.05 for both Pagel's λ and Blomberg's K are shown.

**(B) Low kidney and liver Se in African mole rats.** The error bars represent standard error. The bars corresponding to the naked mole rat and Damaraland mole rat are shaded.

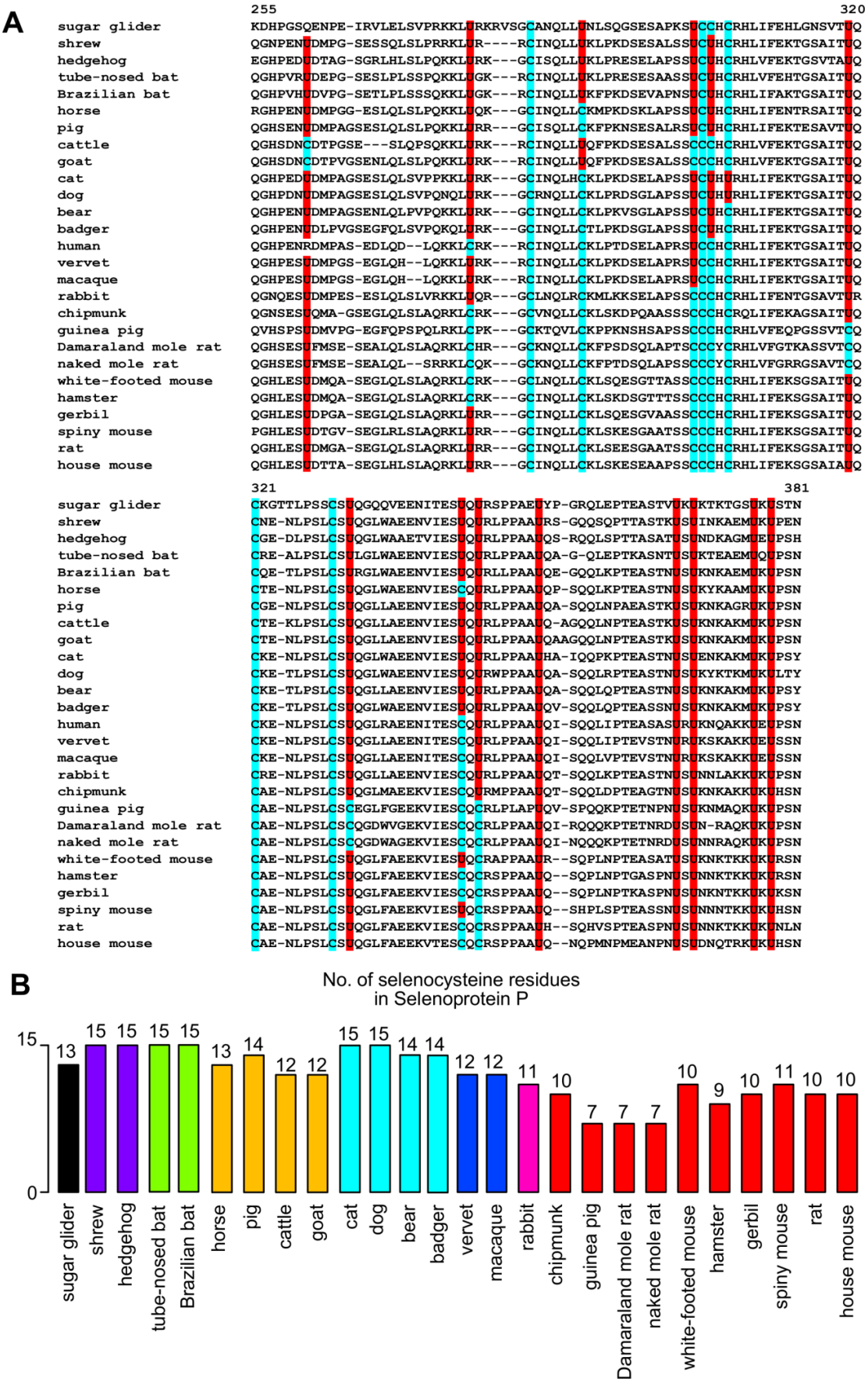
**(C) Kidney and liver Se levels correlate with the number of Sec residues in selenoprotein P.** The error bars represent standard error. The points are colored by taxonomical order. Pearson correlation coefficient: 0.62 for kidney and 0.42 for liver.

**Figure 4.7. Mammalian Selenoprotein P.**

**(A) Multiple sequence alignment of the C-terminal region of Selenoprotein P.** The positions of Sec ("U") are highlighted in red; cysteine (C) are highlighted in cyan. All these species have one additional Sec residue in the N-terminal region (not shown). Position numbering is based on human sequence (included here for reference). The sequences of sugar glider, shrew, hedgehog, tube-nosed bat, Brazilian bat, horse, pig, goat, cat, dog, bear, badger, vervet, rabbit, chipmunk, guinea pig, white-footed mouse, hamster, gerbil, spiny mouse, rat, and house mouse were based on assembled transcriptomes in (Fushan et al., 2015). We confirmed they were consistent with the records in NCBI. The sequences of human, macaque, and naked mole rat were extracted from UCSC Multiz 100-way Alignment. The sequence of Damaraland mole rat (XM_010633881.1) was obtained from NCBI.

**(B) Number of Sec residues in Selenoprotein P.** The number includes the additional Sec residue in the N-terminal region (i.e. one more than the number of Sec shown in (A)).

**Figure 4.7 (Continued)**



## Nature alters elemental composition to adjust lifespan and body mass

To understand the relationship between element levels and longevity, we obtained the lifespan data from AnAge database (Tacutu et al., 2013) and performed regression analysis using generalized least squares, to account for the phylogenetic relatedness of species and incorporate the measurement errors (Felsenstein, 1985; Grafen, 1989; Ives et al., 2007). We focused on adult weight as well as 4 longevity traits: Maximum Lifespan (ML), Maximum Lifespan Residual (MLres), Female Time to Maturity (FTM), and Female Time to Maturity Residual (FTMres) (Table 4.1). FTM was used as an alternative measure for longevity, since it is easier to quantify and less prone to reporting bias than ML (Pearson correlation coefficient between ML and FTM = 0.84, p value < $10^{-7}$). The residuals (MLres and FTMres) describe the portion of the trait that has not been accounted for by body mass. We evaluated different evolutionary models, selected the best-fit model by maximum likelihood, and reported the slope coefficient and p value (Figure 4.8A, Table 4.2; Chapter 2, "Phylogenetic Regression"). A two-step verification procedure was applied to remove potential outliers and ensure the results were generalizable (Chapter 2).

**Figure 4.8. Elements correlating with life history traits.**

**(A) Correlation of elements in each organ with Adult Weight, Maximum Lifespan (ML), Maximum Lifespan Residual (MLres), Female Time to Maturity (FTM), and Female Time to Maturity Residual (FTMres).** Only those with regression p value.robust < 0.02 are indicated in colors: red – positive correlation; blue – negative correlation (see Table 4.2 for more details). Green asterisks indicate those with q value.robust < 0.05.

**Adult Weight correlates (B) positively with liver Zn level and (C) negatively with heart Fe level.** The error bars represent standard error (already incorporated in regression calculation). The points are colored by taxonomical order. (B) p value.all = $4.41\times10^{-7}$; q value.all = $9.25\times10^{-5}$; p value.robust = $4.15\times10^{-9}$; q value.robust = $1.70\times10^{-8}$; p value.max = $2.48\times10^{-7}$; q value.max = $1.04\times10^{-4}$. (C) p value.all = $4.91\times10^{-4}$; q value.all = 0.0187; p value.robust = $6.00\times10^{-4}$; q value.robust = 0.0194; p value.max = $1.21\times10^{-3}$; q value.max = 0.036.

**Maximum Lifespan correlates (D) positively with liver Cd level and (E) negatively with liver Se level.** Selected species are indicated. (D) p value.all = $4.28\times10^{-6}$; q value.all = $4.49\times10^{-4}$; p value.robust = $5.63\times10^{-6}$; q value.robust = $4.73\times10^{-4}$; p value.max = $2.50\times10^{-6}$; q value.max = $2.10\times10^{-4}$. (E) p value.all = 0.0331; q value.all = 0.266; p value.robust = 0.0126; q value.robust = 0.120; p value.max = 0.0335; q value.max = 0.285.

**Figure 4.8 (Continued)**

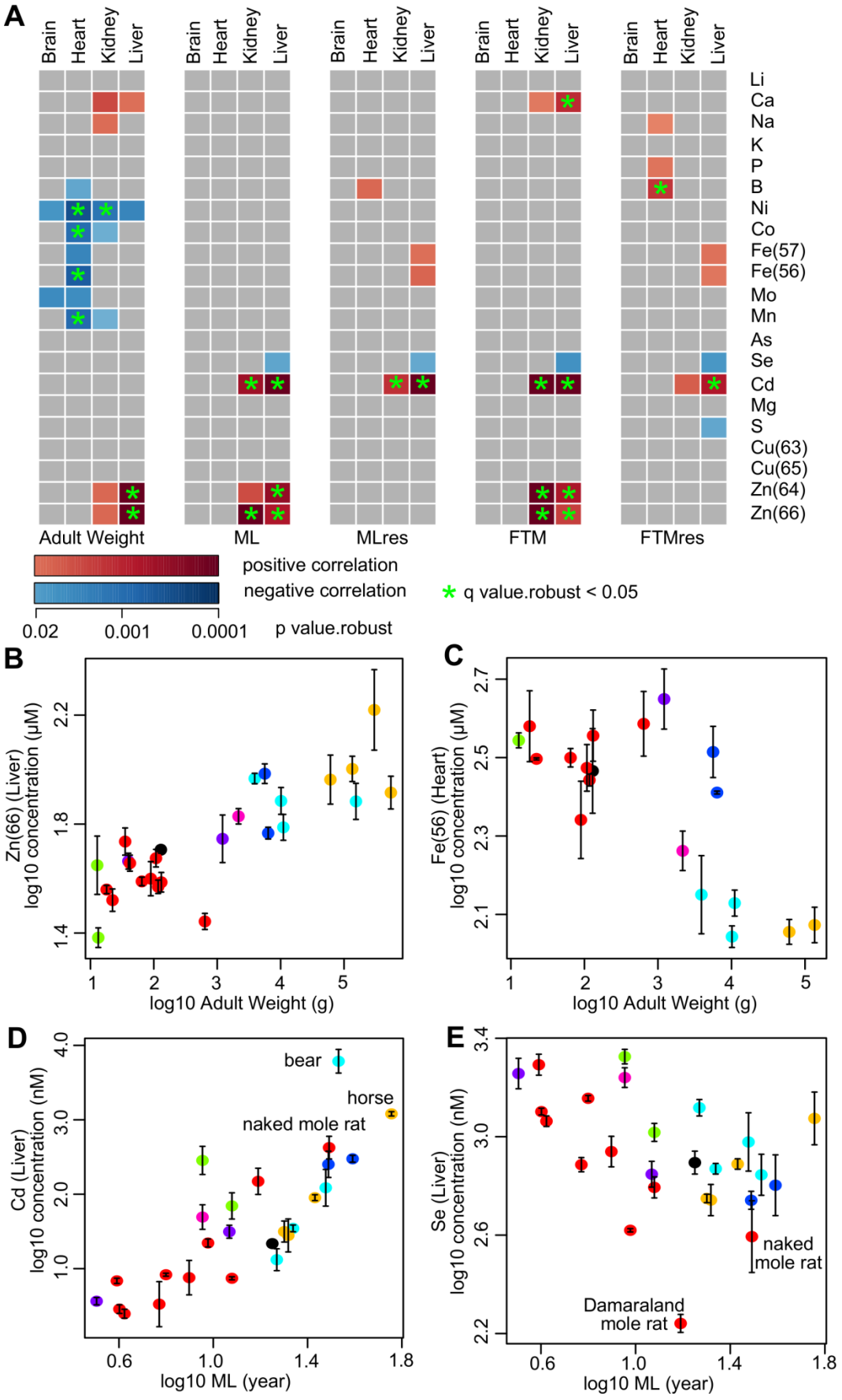


**Table 4.2. Phylogenetic regression of element level in each organ with Adult Weight, Maximum Lifespan (ML), Maximum Lifespan Residual (MLres), Female Time to Maturity (FTM), and Female Time to Maturity Residual (FTMres).** "Slp.all", "p value.all", and "q value.all" refer to the slope coefficient, regression p value, and False Discovery Rate adjustment q value, respectively, using all the data points. "p value.robust" and "q value.robust" refer to the regression parameters after removing the point with the largest residual error. "p value.max" and "q value.max" refer to the maximal p value and FDR adjustment q value, when each one of the species was left out, one at a time. Those with p value.robust < 0.02 are listed. "E" refers to exponent with base 10 (e.g. "4.3E-03" denotes $4.3\times10^{-3}$).

**Table 4.2 (Continued)**

| Trait | Organ | Element | Slp. all | p value. all | q value. all | p value. robust | q value. robust | p value. max | q value. max |
|---|---|---|---|---|---|---|---|---|---|
| Adult Weight | Br | Ni60 | -0.24 | 4.3E-03 | 7.5E-02 | 6.9E-03 | 8.8E-02 | 7.8E-03 | 1.1E-01 |
| | Br | Mo95 | -0.18 | 7.8E-03 | 1.1E-01 | 4.4E-03 | 6.6E-02 | 6.5E-03 | 1.0E-01 |
| | Ht | B11 | -0.07 | 1.4E-02 | 1.5E-01 | 1.4E-02 | 1.3E-01 | 3.2E-02 | 2.8E-01 |
| | Ht | Mn55 | -0.21 | 2.6E-03 | 5.2E-02 | 9.9E-04 | 2.4E-02 | 5.8E-03 | 9.4E-02 |
| | Ht | Fe56 | -0.12 | 4.9E-04 | 1.9E-02 | 6.0E-04 | 1.9E-02 | 1.2E-03 | 3.6E-02 |
| | Ht | Fe57 | -0.11 | 2.3E-03 | 5.1E-02 | 3.4E-03 | 5.4E-02 | 7.2E-03 | 1.0E-01 |
| | Ht | Co59 | -0.21 | 6.6E-03 | 1.0E-01 | 1.1E-03 | 2.5E-02 | 6.9E-03 | 1.0E-01 |
| | Ht | Ni60 | -0.37 | 9.6E-04 | 2.5E-02 | 3.6E-04 | 1.4E-02 | 1.5E-03 | 4.2E-02 |
| | Ht | Mo95 | -0.20 | 1.0E-02 | 1.3E-01 | 4.6E-03 | 6.6E-02 | 3.2E-02 | 2.8E-01 |
| | Kd | Na23 | 0.06 | 1.4E-02 | 1.5E-01 | 9.0E-03 | 9.9E-02 | 5.4E-03 | 9.2E-02 |
| | Kd | Ca40 | 0.04 | 1.4E-02 | 1.5E-01 | 3.4E-03 | 5.4E-02 | 3.0E-02 | 2.8E-01 |
| | Kd | Mn55 | -0.05 | 5.8E-02 | 3.6E-01 | 2.0E-02 | 1.6E-01 | 4.5E-02 | 3.2E-01 |
| | Kd | Co59 | -0.10 | 3.4E-02 | 2.7E-01 | 1.8E-02 | 1.5E-01 | 3.6E-02 | 2.8E-01 |
| | Kd | Ni60 | -0.24 | 4.2E-03 | 7.5E-02 | 1.6E-03 | 3.3E-02 | 1.1E-03 | 3.6E-02 |
| | Kd | Zn64 | 0.07 | 7.2E-04 | 2.2E-02 | 7.5E-03 | 8.8E-02 | 2.9E-03 | 6.1E-02 |
| | Kd | Zn66 | 0.07 | 7.3E-04 | 2.2E-02 | 7.6E-03 | 8.8E-02 | 3.0E-03 | 6.1E-02 |
| | Lv | Ca40 | 0.02 | 1.8E-01 | 5.4E-01 | 9.9E-03 | 1.0E-01 | 1.1E-01 | 4.5E-01 |
| | Lv | Ni60 | -0.15 | 9.6E-03 | 1.3E-01 | 3.4E-03 | 5.4E-02 | 2.5E-03 | 5.8E-02 |
| | Lv | Zn64 | 0.12 | 1.2E-07 | 5.0E-05 | 4.2E-09 | 1.7E-06 | 8.6E-07 | 1.8E-04 |
| | Lv | Zn66 | 0.11 | 4.4E-07 | 9.3E-05 | 1.7E-08 | 3.6E-06 | 2.5E-07 | 1.0E-04 |
| ML | Kd | Zn64 | 0.35 | 3.2E-05 | 1.7E-03 | 3.5E-03 | 5.4E-02 | 5.3E-05 | 2.8E-03 |
| | Kd | Zn66 | 0.33 | 4.0E-05 | 1.9E-03 | 1.9E-05 | 1.1E-03 | 7.4E-05 | 3.5E-03 |
| | Kd | Cd111 | 2.48 | 1.5E-04 | 6.2E-03 | 4.8E-04 | 1.7E-02 | 4.1E-04 | 1.4E-02 |
| | Lv | Zn64 | 0.36 | 2.5E-03 | 5.2E-02 | 3.5E-04 | 1.4E-02 | 5.5E-03 | 9.2E-02 |
| | Lv | Zn66 | 0.33 | 4.0E-03 | 7.5E-02 | 7.4E-04 | 2.1E-02 | 9.4E-03 | 1.2E-01 |
| | Lv | Se78 | -0.36 | 3.3E-02 | 2.7E-01 | 1.3E-02 | 1.2E-01 | 3.4E-02 | 2.8E-01 |
| | Lv | Cd111 | 1.94 | 4.3E-06 | 4.5E-04 | 5.6E-06 | 4.7E-04 | 2.5E-06 | 2.1E-04 |
| MLres | Ht | B11 | 0.29 | 4.4E-02 | 3.2E-01 | 7.9E-03 | 9.0E-02 | 6.1E-02 | 3.4E-01 |
| | Kd | Cd111 | 2.33 | 1.4E-02 | 1.5E-01 | 1.6E-03 | 3.3E-02 | 2.3E-03 | 5.7E-02 |
| | Lv | Fe56 | 0.48 | 4.9E-02 | 3.3E-01 | 7.1E-03 | 8.8E-02 | 4.5E-02 | 3.2E-01 |
| | Lv | Fe57 | 0.46 | 6.0E-02 | 3.6E-01 | 9.6E-03 | 1.0E-01 | 7.2E-02 | 3.6E-01 |
| | Lv | Se78 | -0.44 | 3.7E-02 | 2.8E-01 | 1.5E-02 | 1.3E-01 | 7.8E-02 | 3.7E-01 |
| | Lv | Cd111 | 2.07 | 8.2E-04 | 2.3E-02 | 3.7E-06 | 3.9E-04 | 3.2E-04 | 1.2E-02 |
| FTM | Kd | Ca40 | 0.11 | 4.1E-02 | 3.1E-01 | 1.3E-02 | 1.2E-01 | 4.5E-02 | 3.2E-01 |
| | Kd | Zn64 | 0.26 | 7.9E-06 | 6.6E-04 | 8.7E-07 | 1.2E-04 | 1.7E-06 | 2.1E-04 |
| | Kd | Zn66 | 0.24 | 1.2E-05 | 8.5E-04 | 6.4E-05 | 3.4E-03 | 3.3E-06 | 2.3E-04 |
| | Kd | Cd111 | 1.80 | 2.5E-05 | 1.5E-03 | 9.1E-05 | 4.2E-03 | 3.3E-05 | 2.0E-03 |
| | Lv | Ca40 | 0.13 | 1.4E-02 | 1.5E-01 | 1.4E-03 | 3.0E-02 | 6.7E-02 | 3.5E-01 |
| | Lv | Zn64 | 0.24 | 5.7E-03 | 9.6E-02 | 6.8E-04 | 2.0E-02 | 7.6E-03 | 1.1E-01 |
| | Lv | Zn66 | 0.21 | 1.0E-02 | 1.3E-01 | 2.5E-03 | 4.6E-02 | 1.5E-02 | 1.6E-01 |
| | Lv | Se78 | -0.24 | 4.5E-02 | 3.2E-01 | 6.0E-03 | 8.1E-02 | 3.3E-02 | 2.8E-01 |
| | Lv | Cd111 | 1.42 | 2.4E-06 | 3.3E-04 | 9.9E-06 | 6.9E-04 | 2.1E-06 | 2.1E-04 |
| FTMres | Ht | B11 | 0.20 | 9.9E-03 | 1.3E-01 | 1.8E-03 | 3.5E-02 | 1.1E-02 | 1.4E-01 |
| | Ht | Na23 | 0.18 | 3.4E-02 | 2.7E-01 | 1.7E-02 | 1.4E-01 | 5.3E-02 | 3.4E-01 |
| | Ht | P31 | 0.07 | 1.6E-01 | 5.4E-01 | 1.1E-02 | 1.1E-01 | 3.5E-02 | 2.8E-01 |
| | Kd | Cd111 | 1.86 | 6.7E-03 | 1.0E-01 | 5.7E-03 | 7.9E-02 | 3.8E-03 | 7.0E-02 |
| | Lv | S34 | -0.05 | 1.1E-01 | 4.5E-01 | 1.4E-02 | 1.2E-01 | 1.1E-01 | 4.5E-01 |
| | Lv | Fe56 | 0.31 | 6.0E-02 | 3.6E-01 | 1.2E-02 | 1.2E-01 | 5.4E-02 | 3.4E-01 |
| | Lv | Fe57 | 0.31 | 6.0E-02 | 3.6E-01 | 1.0E-02 | 1.0E-01 | 5.2E-02 | 3.4E-01 |
| | Lv | Se78 | -0.29 | 5.2E-02 | 3.4E-01 | 7.2E-03 | 8.8E-02 | 7.3E-02 | 3.6E-01 |
| | Lv | Cd111 | 1.46 | 1.4E-03 | 3.4E-02 | 7.9E-04 | 2.1E-02 | 2.1E-03 | 5.5E-02 |

In terms of body mass, liver Zn level showed a strong positive correlation (Figure 4.8B; p value.robust = $4.15\times10^{-9}$ for $^{64}$Zn and $1.70\times10^{-8}$ for $^{66}$Zn; q value.robust < $10^{-5}$ for both), whereas negative correlations were observed for Fe, Mn, Co and Ni, especially in heart tissue (Figure 4.8C). In terms of longevity, Zn levels in liver and kidney showed strong, positive correlations with ML and FTM. However, the relationships became much weaker in MLres and FTMres, suggesting the observed correlation is largely due to the effects of body mass.

Cd levels in liver and kidney correlated positively with all four measures of longevity (Figure 4.8A), with a particularly robust relationship in liver (q value.robust < 0.05 in all four measures). While the Cd levels might potentially be affected by diets, the correlation remained statistically significant when we methodically left out each species one at a time. As there has been no known biological role for Cd in mammals, the result was somewhat unexpected. Since Cd mostly comes from the soil or food, one possibility is that longer lived/larger mammals simply consume a greater amount of food over their life time and, as they cannot efficiently excrete this metal, accumulate more Cd. Among our samples, the liver Cd levels were highest in horse and bear, both of which were large and long-lived. The naked mole rat, being the longest-lived rodent, also had a much higher liver Cd level than other rodents of comparable sizes.

Liver Se was the only element correlating negatively with all four measures of longevity, although the correlations were relatively weak (Figure 4.8E and Table 4.2; p value.robust = 0.013 for ML, 0.015 for MLres, 0.006 for FTM, and 0.007 for FTMres). Even when the points corresponding to the African mole rats were excluded, we still observed a negative correlation with ML (p value = 0.028). Selenoproteins are important cellular redox regulators, whose changes may affect redox homeostasis and DNA mutation rates. Interestingly, the long-lived naked mole rat may experience significant oxidative stress in captivity (Andziak et al., 2006) and limit both GPX1 expression and GPX activity (Andziak et al., 2005; Kasaikina et al., 2011; Kim et al., 2011). Excessive intake of Se is toxic and can lead to selenosis. Therefore, there is a delicate balance between the beneficial aspects of Se, its toxicity, and other systems that support maintenance functions.

# DISCUSSION

Prior to this study, it has been unclear how evolutionary processes adjust the ionomes across mammalian species and organs according to phylogeny, longevity and body mass. We were able to offer new insights on elemental composition of organs and species, identify common and distinct patterns of element utilization, and link these findings to physiological functions. The elements clustered as macronutrients, transition metals or toxic elements, suggesting they are handled by mammals using common strategies based on their biological functions. Some elements also showed lineage-specific changes, including decreased utilization of Se in African mole rats, and a strong link between the number of Sec residues in SelP and the kidney and liver Se levels across mammals. In addition, we found that lifespan of mammals was positively linked with Cd tissue levels and negatively linked with Se in liver. Some elements, such as Zn, showed a positive association with species body mass, whereas others, such as Ni, Co, Fe, and Mn, showed negative association.

The multiple associations and interactions among these elements revealed by our study would not have been possible without examining a wide spectrum of mammals and different organs. This approach can also offer valuable insights into the interaction between elements and environment, as previously demonstrated by similar studies in plants (Watanabe et al., 2007; White et al., 2012). When the same plant species were collected under different fertilizer treatments and environmental conditions, the concentrations of Ca, Zn, Mn and Mg were found to remain more closely linked to phylogeny. In contrast, the concentrations of Cu and Fe varied more strongly with environmental factors, suggesting different elements might be more or less responsive to external variations. In addition, related species may be exposed to similar environmental conditions, or may have similar detoxification abilities, either of which can drive the similarities in their ionomes and potentially confound the phylogenetic signals. Similar studies on a much larger scale will help determine the variability and elasticity of element levels in mammals under different dietary regimens. Overall, this study provided direct insights into how evolution may adjust the ionome of mammals according to organ physiology, phylogeny, environment, lineage specialization and life histories and may provide a useful predictive tool in future studies.

## EXPERIMENTAL PROCEDURES

### Biological samples and element quantifications

Inductively-coupled plasma mass spectrometry (ICP-MS) was applied to characterize element levels in the brain, heart, kidney, and liver of 26 mammalian species. The species were as described previously (Fushan et al., 2015; Ma et al., 2015). The animals were young adults and, except for horse and vervet, all were males, with two to four biological replicates (tissues from different animals) for most species. Whole liver, kidney, heart, or frontal parts of brain were frozen in liquid nitrogen and stored at -80ºC until further use.

The concentrations of Li (nM), B (nM), Na (mM), Mg (μM), P (mM), S (mM), K (mM), Ca (μM), Mn (nM), Fe (μM), Co (nM), Ni (nM), Cu (μM), Zn (μM), As (nM), Se (nM), Mo (nM), and Cd (nM), per gram of tissue digested, were quantified with spike-in 50 μg/L Ga as internal control, using the sample preparation and data collection method described previously (Malinouski et al., 2014). Two isotopes were measured for Fe ($^{56}$Fe, $^{57}$Fe), Cu ($^{63}$Cu, $^{65}$Cu), and Zn ($^{64}$Zn, $^{66}$Zn). The samples were subjected to four independent ICP-MS runs, each with three injections to each sample. Those injections with clearly abnormal results (i.e. reporting negative concentrations for most elements and differing significantly from the other runs of the same sample) were discarded (79 out of over 2600 injections, or ~ 3%, were discarded this way).

### Data processing

The results were log10-transformed and batch effect was removed using R package “sva” (Leek et al.) (Figure 4.2). The average values were computed across the four runs and the mean and standard errors were calculated across the replicates. For those without biological replicates, the standard error was taken as average standard error of the other samples. Standardized concentrations (i.e. scaled to mean = 0 and standard deviation = 1) were used for cross-element analysis.

### Element distribution in organs

Wilcoxon Rank-Sum Test was used to identify elements relatively enriched or depleted in a particular organ. Data from the same animal were considered as paired. For each element, the numbers of significant pair-wise comparisons ($p$ value $< 0.05$) were tabulated. Enrichment or

depletion was declared in an organ if at least two out of the three pair-wise comparisons were statistically significant.

**Phylogenetic signals and phylogenetic ANOVA**

Pagel's lambda and Blomberg's K were calculated using R package "phytools" (Revell, 2012), after incorporating standard error of measurement. Those elements with p value < 0.05 in both cases were considered to have significant phylogenetic signals. Phylogenetic ANOVA was performed using R package "geiger" (Harmon et al., 2008). The species were grouped and compared according to their taxonomical orders or families.

**Regression by generalized least square and robustness of results**

See Chapter 2. The cut-off for top hits was p value.robust < 0.02.

# REFERENCES


Amaravadi, R., Glerum, D.M., and Tzagoloff, A. (1997). Isolation of a cDNA encoding the human homolog of COX17, a yeast gene essential for mitochondrial copper recruitment. Hum. Genet. *99*, 329-333.

Andziak, B., O'Connor, T.P., and Buffenstein, R. (2005). Antioxidants do not explain the disparate longevity between mice and the longest-living rodent, the naked mole-rat. Mech. Ageing Dev. *126*, 1206-1212.

Andziak, B., O'Connor, T.P., Qi, W., DeWaal, E.M., Pierce, A., Chaudhuri, A.R., Van Remmen, H., and Buffenstein, R. (2006). High oxidative damage levels in the longest-living rodent, the naked mole-rat. Aging Cell *5*, 463-471.

Buffenstein, R. (2005). The naked mole-rat: a new long-living model for human aging research. J. Gerontol. A Biol. Sci. Med. Sci. *60*, 1369-1377.

Clemens, S., Aarts, M.G., Thomine, S., and Verbruggen, N. (2013). Plant science: the key to preventing slow cadmium poisoning. Trends Plant Sci. *18*, 92-99.

Eide, D.J. (2006). Zinc transporters and the cellular trafficking of zinc. Biochim. Biophys. Acta *1763*, 711-722.

Eide, D.J., Clark, S., Nair, T.M., Gehl, M., Gribskov, M., Guerinot, M.L., and Harper, J.F. (2005). Characterization of the yeast ionome: a genome-wide analysis of nutrient mineral and trace element homeostasis in Saccharomyces cerevisiae. Genome Biol. *6*, R77.

Fang, X., Seim, I., Huang, Z., Gerashchenko, M.V., Xiong, Z., Turanov, A.A., Zhu, Y., Lobanov, A.V., Fan, D., Yim, S.H.*, et al.* (2014). Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. Cell Rep. *8*, 1354-1364.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. Am. Nat. *125*, 1-15.

Fraga, C.G. (2005). Relevance, essentiality and toxicity of trace elements in human health. Mol. Aspects Med. *26*, 235-244.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Goldhaber, S.B. (2003). Trace element risk assessment: essentiality vs. toxicity. Regul. Toxicol. Pharmacol. *38*, 232-242.

Grafen, A. (1989). The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci. *326*, 119-157.

Gunshin, H., Mackenzie, B., Berger, U.V., Gunshin, Y., Romero, M.F., Boron, W.F., Nussberger, S., Gollan, J.L., and Hediger, M.A. (1997). Cloning and characterization of a mammalian proton-coupled metal-ion transporter. Nature *388*, 482-488.

Hamza, I., Chauhan, S., Hassett, R., and O'Brian, M.R. (1998). The bacterial irr protein is required for coordination of heme biosynthesis with iron availability. J. Biol. Chem. *273*, 21669-21674.

Hamza, I., Schaefer, M., Klomp, L.W., and Gitlin, J.D. (1999). Interaction of the copper chaperone HAH1 with the Wilson disease protein is essential for copper homeostasis. Proc. Natl. Acad. Sci. USA *96*, 13363-13368.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. Bioinformatics *24*, 129-131.

Hill, G.M., and Link, J.E. (2009). Transporters in the absorption and utilization of zinc and copper. J. Anim. Sci. *87*, E85-89.

Holm, R.H., Kennepohl, P., and Solomon, E.I. (1996). Structural and Functional Aspects of Metal Sites in Biology. Chem. Rev. *96*, 2239-2314.

Hurst, R., Siyame, E.W., Young, S.D., Chilimba, A.D., Joy, E.J., Black, C.R., Ander, E.L., Watts, M.J., Chilima, B., Gondwe, J.*, et al.* (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. Sci. Rep. *3*, 1425.

Ives, A.R., Midford, P.E., and Garland, T. (2007). Within-Species Variation and Measurement Error in Phylogenetic Comparative Methods. Syst. Biol. *56*, 252-270.

Kasaikina, M.V., Lobanov, A.V., Malinouski, M.Y., Lee, B.C., Seravalli, J., Fomenko, D.E., Turanov, A.A., Finney, L., Vogt, S., Park, T.J.*, et al.* (2011). Reduced utilization of selenium by naked mole rats due to a specific defect in GPx1 expression. J. Biol. Chem. *286*, 17005-17014.

Katoh, Y., Sato, T., and Yamamoto, Y. (2002). Determination of multielement concentrations in normal human organs from the Japanese. Biol. Trace Elem. Res. *90*, 57-70.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P.*, et al.* (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223-227.

Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R., and Gladyshev, V.N. (2003). Characterization of mammalian selenoproteomes. Science *300*, 1439-1443.

LaVinka, P.C., and Park, T.J. (2012). Blunted behavioral and c Fos responses to acidic fumes in the African naked mole-rat. PLoS One *7*, e45060.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2014). sva: Surrogate Variable Analysis. R package version 3.10.0.

Liuzzi, J.P., and Cousins, R.J. (2004). Mammalian zinc transporters. Annu. Rev. Nutr. *24*, 151-172.

Lobanov, A.V., Hatfield, D.L., and Gladyshev, V.N. (2008). Reduced reliance on the trace element selenium during evolution of mammals. Genome Biol. *9*, R62.

Ma, S., Yim, S.H., Lee, S.G., Kim, E.B., Lee, S.R., Chang, K.T., Buffenstein, R., Lewis, K.N., Park, T.J., Miller, R.A.*, et al.* (2015). Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity. Cell Metab. *22*, 332-343.

Mackenzie, B., Takanaga, H., Hubert, N., Rolfs, A., and Hediger, M.A. (2007). Functional properties of multiple isoforms of human divalent metal-ion transporter 1 (DMT1). Biochem. J. *403*, 59-69.

Maina, J.N., Gebreegziabher, Y., Woodley, R., and Buffenstein, R. (2001). Effects of change in environmental temperature and natural shifts in carbon dioxide and oxygen concentrations on the

lungs of captive naked mole-rats (Heterocephalus glaber): a morphological and morphometric study. J. Zool. *253*, 371-382.

Malinouski, M., Hasan, N.M., Zhang, Y., Seravalli, J., Lin, J., Avanesov, A., Lutsenko, S., and Gladyshev, V.N. (2014). Genome-wide RNAi ionomics screen reveals new genes and regulation of human trace element metabolism. Nat. Commun. *5*, 3301.

Malinouski, M., Kehr, S., Finney, L., Vogt, S., Carlson, B.A., Seravalli, J., Jin, R., Handy, D.E., Park, T.J., Loscalzo, J.*, et al.* (2012). High-resolution imaging of selenium in kidneys: a localized selenium pool associated with glutathione peroxidase 3. Antioxid. Redox Signal. *16*, 185-192.

Mendel, R.R., and Bittner, F. (2006). Cell biology of molybdenum. Biochim. Biophys. Acta *1763*, 621-635.

Ozaki, T., Enomoto, S., Minai, Y., Ambe, S., and Makide, Y. (2000). A survey of trace elements in pteridophytes. Biol. Trace Elem. Res. *74*, 259-273.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217-223.

Roth, J.A. (2006). Homeostatic and toxic mechanisms regulating manganese uptake, retention, and elimination. Biol. Res. *39*, 45-57.

Salt, D.E., Baxter, I., and Lahner, B. (2008). Ionomics and the study of the plant ionome. Annu. Rev. Plant Biol. *59*, 709-733.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-1033.

Waldron, K.J., and Robinson, N.J. (2009). How do bacterial cells ensure that metalloproteins get the correct metal? Nat. Rev. Microbiol. *7*, 25-35.

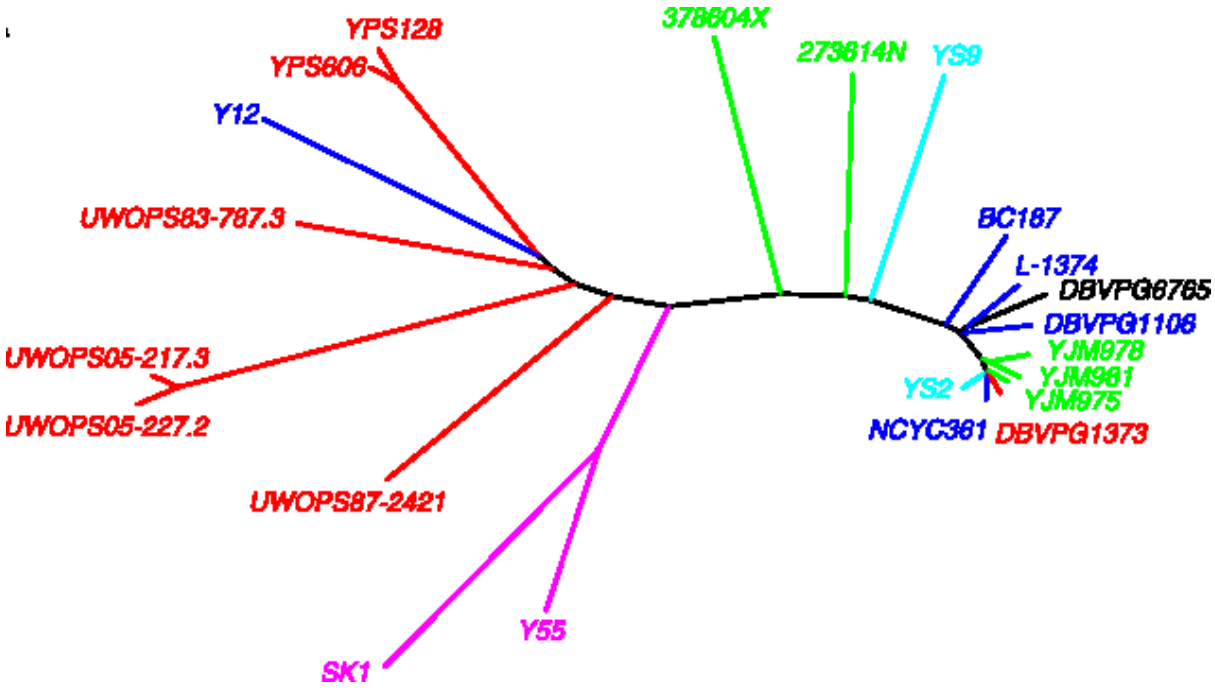
Watanabe, T., Broadley, M.R., Jansen, S., White, P.J., Takada, J., Satake, K., Takamatsu, T., Tuah, S.J., and Osaki, M. (2007). Evolutionary control of leaf element composition in plants. New Phytol. *174*, 516-523.

White, P.J., Broadley, M.R., Thompson, J.A., McNicol, J.W., Crawley, M.J., Poulton, P.R., and Johnston, A.E. (2012). Testing the distinctness of shoot ionomes of angiosperm families using the Rothamsted Park Grass Continuous Hay Experiment. New Phytol. *196*, 101-109.

Zhang, Y., and Gladyshev, V.N. (2009). Comparative genomics of trace elements: emerging dynamic view of trace element utilization and function. Chem. Rev. *109*, 4828-4861.

Zhang, Y., Rodionov, D.A., Gelfand, M.S., and Gladyshev, V.N. (2009). Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. BMC Genomics *10*, 78.

# Chapter 5 Yeast Transcriptome And Proteome




This chapter is based on the following publication:

Alaattin Kaya*, Siming Ma*, Brian Wasko, Mitchell Lee, Matt Kaeberlein, Vadim N. Gladyshev. **Defining molecular basis for longevity traits in natural yeast isolates.** *npj Aging and Mechanisms of Disease* 1 (2015), 15001.

* equal contribution.

**Author contributions**

A.K., and V.N.G. designed the experiments. A.K., B.W., and M.L performed the experiments. A.K., S.M., M.L., M.K., and V.N.G analyzed the data. A.K., S.M., M.K and V.N.G wrote the paper. All authors reviewed the paper.

# ABSTRACT

The budding yeast has served as a useful model organism in aging studies, leading to the identification of genetic determinants of longevity, many of which are conserved in higher eukaryotes. However, factors that promote longevity in a laboratory setting often have severe fitness disadvantages in the wild. To obtain an unbiased view on longevity regulation, we analyzed how a replicative lifespan is shaped by transcriptional, translational, metabolic, and morphological factors across 22 wild-type *Saccharomyces cerevisiae* isolates. We observed significant differences in lifespan across these strains and found that their longevity is strongly associated with up-regulation of oxidative phosphorylation and respiration and down-regulation of amino- acid and nitrogen compound biosynthesis. As calorie restriction and TOR signaling also extend the lifespan by adjusting many of the identified pathways, the data suggest that the natural plasticity of yeast lifespan is shaped by the processes that not only do not impose cost on fitness, but also are amenable to dietary intervention.

# INTRODUCTION

The idea of slowing aging and extending lifespan of organisms has attracted much attention, leading to the identification of numerous factors that mitigate the effects of the aging process. At the cellular level, the driving force behind aging may be the inevitable accumulation of a myriad different forms of molecular damage (Gladyshev, 2012). Many genetic and pharmacological interventions have been discovered that increase the lifespan of model organisms, including some with single gene effects (Finch and Ruvkun, 2001; Kenyon, 2010). In addition, diverse classes of genes have been reported to be involved in lifespan control, pointing to several key regulatory pathways. However, it remains to be seen whether similar strategies may be applied to combat aging in humans. A major challenge in the field is that many of the findings apply to model organisms in laboratory settings, but these longevity conditions may come at the expense of fitness, making them detrimental when organisms are in their natural environment.

Aging is a process that involves complex gene networks. While broad genome manipulation is not yet practical in higher eukaryotes, fine-tuning these gene networks by environmental or dietary factors may offer a solution. It has been shown that manipulations such as calorie restriction (CR), oxygen availability, pH, and alternative carbon sources can modulate gene expression and the aging process (Botta et al., 2011; McCay et al., 1935; Murakami et al., 2012; Pitt et al., 2014). CR is among the most studied and widely used longevity interventions, which can extend lifespan in almost all model organisms (Finch and Ruvkun, 2001). Although the precise mechanisms of CR-mediated lifespan extension remain debatable, it is known that CR causes a metabolic shift from fermentation to respiration in yeast, and that mitochondrial metabolism tends to increase in multicellular eukaryotes subjected to CR (Lin et al., 2002; Lopez-Lluch et al., 2006; Schulz et al., 2007). These findings also agree with the effects observed by manipulating various lifespan-regulating pathways, such as TOR (target of rapamycin) signaling (Johnson et al., 2013). Suppression of TOR signaling mimics the reduction of nutritional input under CR in yeast and extends lifespan while concomitantly increasing mitochondrial respiration (Bonawitz et al., 2007; Johnson et al., 2013). Taken together, these studies

link elevated mitochondrial function with lifespan, suggesting that a metabolic switch to oxidative metabolism is beneficial with regard to delaying aging.

The fact that metabolic pathways can be modulated by both CR and TOR inhibition suggests that complex processes such as aging may also be amenable to environmental and genetic manipulation. It is conceivable that the interaction between environmental factors and gene networks can explain the diverse phenotypes of species inhabiting different ecological niches. It is known that environmental adaptation and parallel evolution help create the genetic diversity for selection in natural populations (Reznick et al., 2001). By evaluating the lifespan differences among natural populations of closely related strains or species, one may obtain insights into the underlying mechanisms that modulate aging and longevity. Towards this goal, in the current work we employed a powerful aging model, the budding yeast. Analyses of the aging process in *Saccharomyces cerevisiae* have mostly been performed on a small number of laboratory-adapted strains, but whether the identified mechanisms can explain the lifespan variation across natural strains is unknown. We evaluated the lifespans of 22 natural isolates of *S. cerevisiae* (Liti et al., 2009) and used transcriptome, proteome, metabolome and morphology data (Skelly et al., 2013) to identify the signatures associated with natural lifespan variation. Our data suggest that increased replicative lifespan (RLS) in natural yeast populations is associated with increased oxidative phosphorylation and reduced amino acid biosynthesis. Our study also represents a new approach that combines phenotypic variation across yeast populations with high-throughput data to elucidate underlying molecular mechanisms driving this variation.

# RESULTS

## Variation in replicative lifespan and growth rate across natural yeast isolates

Phylogenetic analysis using complete genome sequence alignment of 22 natural *S. cerevisiae* isolates revealed a complex cladogram that could be divided into two main groups (Figure 5.1A). Assaying these isolates at 30 °C on standard YPD plates, we observed over 10-fold variation in RLS (Pearson correlation coefficient = 0.95 between mean and maximum lifespans; Figure 5.1B and Table 5.1). BC187 showed the largest number of cell divisions (mean = 39; maximum = 60); NCYC361 and YS2 had the fewest (mean = 3 for both; maximum = 7 for NCYC361 and 9 for YS2); and many strains produced on average 20-30 daughter cells, similar to BY4743, a standard laboratory diploid strain and the parental strain of the yeast ORF deletion collection (Table 5.1).

Changes in growth rate have previously been shown to affect mRNA, protein and metabolite levels (Brauer et al., 2008; Castrillo et al., 2007; Regenberg et al., 2006), and a recent study has reported a positive correlation between time spent in the G1 phase of the cell cycle and RLS in yeast (He et al., 2014). To determine a potential relationship between growth rate and lifespan, we monitored the growth of these strains by automated Bioscreen-C growth analyzer and calculated the doubling time in both glucose and glycerol medium. Of the 22 isolates, 21 grew faster than BY4743 strain, and four strains doubled in less than 50 min (Figure 5.1C, Table 5.1) in glucose medium. However, we found only a weak negative correlation between the doubling time and mean lifespan (Pearson correlation coefficient = -0.42). In addition, we observed that all strains can utilize glycerol as a carbon source, which indicates these strains are capable of mitochondrial respiratory metabolism (Table 5.1).

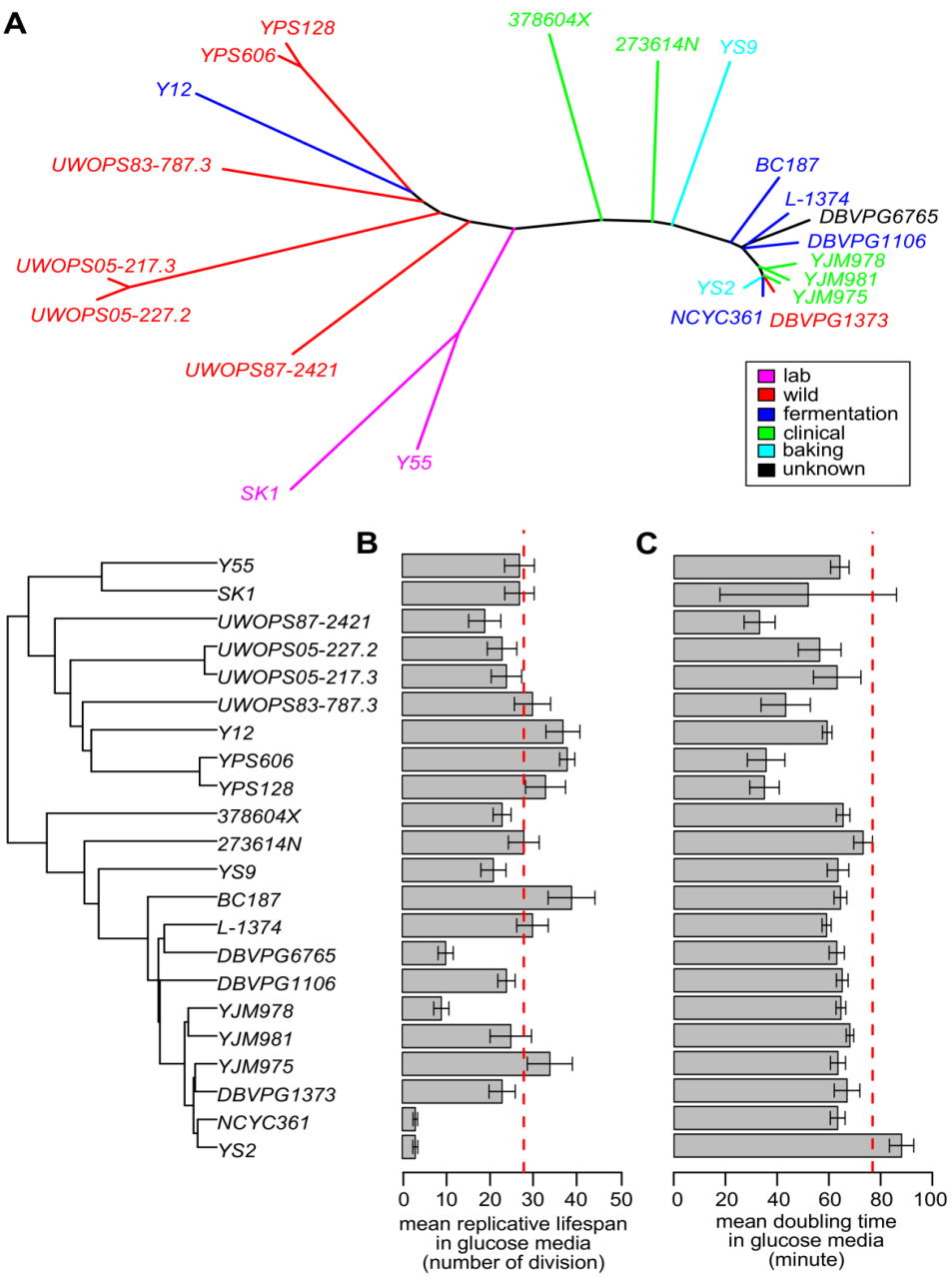



**Figure 5.1.Yeast strains examined in this study.**

**(A) Phylogenetic relationship.** The tree was constructed based on the alignment of complete genome sequences of the strains, using MEGA 6.06 (Tamura et al., 2013) and neighbor joining method (Saitou and Nei, 1987). The branches are colored according to strain types shown in the legend in the lower right corner.

**(B) Mean replicative lifespan and (C) mean growth rate (doubling time) of the strains in glucose media**. The strains are ordered by phylogeny. The error bars indicate standard error. Red dotted lines indicate the mean replicative lifespan (B) and doubling time (C) of the reference strain BY4743.

**Table 5.1. Strains used in this study.** Natural isolates of yeast strains are shown along with their wild niche, mean and maximum lifespan, maximum doubling time in glucose and glycerol media, relative coverage of mitochondrial DNA, and cell size. "S.E." denotes standard error of mean.

| Strain Name | Strain Type | Source | Replicative Lifespan (glucose) | | | Replicative Lifespan (glycerol) | | | Doubling Time in glucose (min) | | Doubling Time in glycerol (min) | | Mitochondrial DNA coverage | Cell Size (μm) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | max | mean | S.E. | max | mean | S.E. | mean | S.E. | mean | S.E. | | |
| Y55 | lab | grape | 47 | 27 | 3.43 | 57 | 35 | 2.25 | 63.79 | 3.56 | 110 | 1.13 | 4 | 10 |
| SK1 | lab | soil | 45 | 27 | 3.39 | -- | -- | -- | 51.62 | 33.95 | 190 | 11.74 | 5 | 13 |
| UWOPS87-2421 | wild | plant | 35 | 19 | 3.72 | -- | -- | -- | 32.91 | 5.96 | 107 | 0.74 | 4 | 12 |
| UWOPS05-227.2 | wild | bee | 37 | 23 | 3.39 | -- | -- | -- | 56.03 | 8.23 | 117 | 4.93 | 4 | 11 |
| UWOPS05-217.3 | wild | plant | 42 | 24 | 3.51 | -- | -- | -- | 62.77 | 9.14 | 116 | 5.52 | 4 | 11 |
| UWOPS83-787.3 | wild | fruit | 52 | 30 | 4.16 | -- | -- | -- | 42.96 | 9.47 | 107 | 1.46 | 4 | 10 |
| Y12 | fermentation | palm wine | 55 | 37 | 3.92 | 37 | 24 | 1.42 | 58.91 | 1.83 | 118 | 3.92 | 5 | 10 |
| YPS606 | wild | oak tree | 44 | 38 | 1.76 | 56 | 33 | 2.13 | 35.43 | 7.22 | 98.5 | 1.29 | 4 | 10 |
| YPS128 | wild | oak tree | 54 | 33 | 4.61 | -- | -- | -- | 34.78 | 5.68 | 97.7 | 1.83 | 3 | 10 |
| 378604X | clinical | sputum | 31 | 23 | 2.06 | -- | -- | -- | 65.04 | 2.64 | 205 | 10.88 | 4 | 12 |
| 273614N | clinical | fecal | 45 | 28 | 3.55 | 52 | 35 | 2.04 | 72.75 | 3.63 | 109 | 1.24 | 6 | 13 |
| YS9 | baking | unknown | 37 | 21 | 2.9 | -- | -- | -- | 63.08 | 4.15 | 131 | 0.91 | 3 | 14 |
| BC187 | fermentation | barrel | 60 | 39 | 5.39 | 92 | 43 | 4.86 | 64.03 | 2.46 | 118 | 0.35 | 7 | 10 |
| L-1374 | fermentation | must | 47 | 30 | 3.63 | -- | -- | -- | 58.68 | 1.75 | 105 | 1.34 | 5 | 12 |
| DBVPG6765 | unknown | unknown | 16 | 10 | 1.76 | 21 | 15 | 1.66 | 62.6 | 2.91 | 99.2 | 5.06 | 4 | 13 |
| DBVPG1106 | fermentation | grapes | 32 | 24 | 2 | 32 | 24 | 1.55 | 64.7 | 2.26 | 109 | 1.52 | 5 | 12 |
| YJM978 | clinical | vaginal | 18 | 9 | 1.76 | -- | -- | -- | 64.2 | 1.89 | 110 | 4.21 | 6 | 12 |
| YJM981 | clinical | vaginal | 50 | 25 | 4.78 | 30 | 20 | 1.51 | 67.68 | 1.43 | 102 | 3.16 | 8 | 14 |
| YJM975 | clinical | vaginal | 52 | 34 | 5.18 | 30 | 23 | 1.3 | 63.07 | 2.89 | 107 | 3.28 | 6 | 11 |
| DBVPG1373 | wild | soil | 35 | 23 | 3.02 | 49 | 30 | 1.86 | 66.58 | 4.93 | 123 | 2.48 | 5 | 11 |
| NCYC361 | fermentation | wort | 7 | 3 | 0.53 | -- | -- | -- | 62.97 | 2.84 | 118 | 0.75 | 5 | 12 |
| YS2 | baking | unknown | 9 | 3 | 0.57 | -- | -- | -- | 87.55 | 4.67 | 216 | 2.31 | 4 | 11 |
| BY4743 | lab | grape | 44 | 28 | 1.67 | -- | -- | -- | 76.38 | 2.89 | 164 | 7.17 | 4 | -- |

## Phenotypic variation across strains

Gene expression, proteomic, metabolomic, and morphological data for these 22 strains have been reported previously (Skelly et al., 2013). After our filtering and quality control, the dataset contained RNA-seq reads for 6207 transcripts; proteomic measurement of 6842 peptide fragments corresponding to 1643 unique genes; mass spectrometric quantification of 107 metabolites; and quantitative microscopy of 392 morphological phenotypes (Experimental Procedures). In particular, 1641 unique genes were represented by both transcripts and peptides, but the correlation between the transcript and peptide levels was not strong (median Spearman correlation coefficient = 0.31; Figure 5.2A). Similar conclusions were reached when we used the mean peptide values for each gene instead (median Spearman correlation coefficient = 0.28; Figure 5.2B).

To visualize phenotypic variation across these strains, we performed Principal Component Analysis on each type of the phenotypic data as well as on the combined data (Figure 5.3A, Figure 5.4A-D; the combined data excluded metabolites as values were not available for strain 378604X). The observed patterns resembled the phylogenetic relationship, with the first 3 Principal Components (PCs) explaining 36-53% of total variance (Figure 5.4E). Examination of the genes contributing to the first 3 PCs in the combined data revealed a distinctive set of GO terms and KEGG pathways, including oxidative phosphorylation (PC1), aerobic respiration (PC1), mitochondrion (PC1), response to temperature stimulus (PC1), ribosome (PC2), protein synthesis (PC2), regulation of translation (PC2 and PC3), ribonucleoprotein complex (PC3), and ribosome biogenesis (PC3). These results suggest that the strains predominantly differ in energy metabolism, protein synthesis, and ribosome regulation. Consistent with a previous report (Skelly et al., 2013), along PC1 the strains segregated largely according to their relative preferences for aerobic respiration or fermentation (Figure 5.3B).

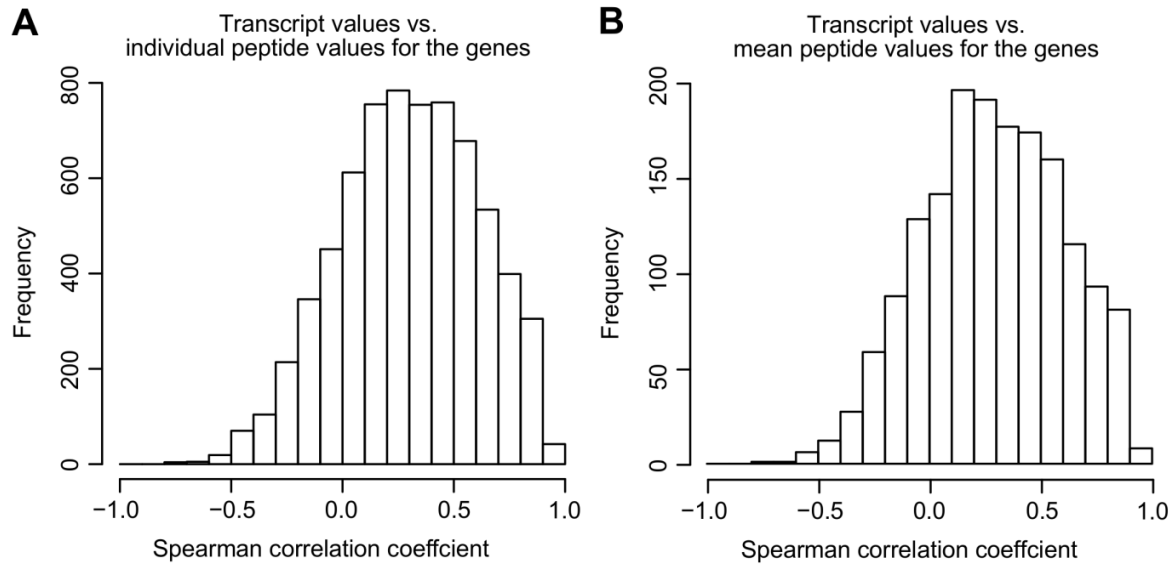



**Figure 5.2. Spearman correlation coefficients between transcript values and (A) individual peptide values, or (B) mean peptide values for genes.** The 25th, 50th, and 75th percentile values are: (A) 0.08, 0.31, 0.54, respectively; and (B) 0.05, 0.25, 0.51, respectively.

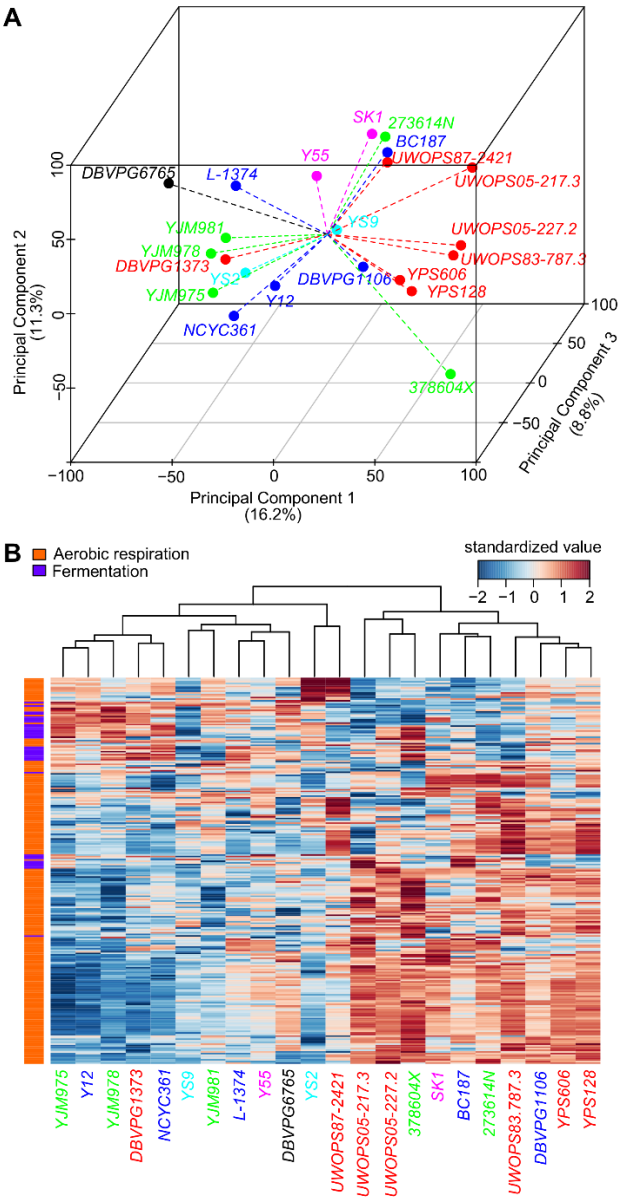



**Figure 5.3. Phenotypic variation across the strains.**

**(A) Principal Component Analysis (PCA) of combined data.** PCA was performed by combining transcripts, peptides, and morphology data (metabolite data were not available for strain 378604X and were omitted). Percentage variance explained by each Principal Component (PC) is shown in the parentheses.

**(B) Relative levels of transcripts and peptides involved in aerobic respiration or fermentation.** The heat map shows the transcripts and peptides with top contribution to Principal Component 1 and involved in aerobic respiration or fermentation. Hierarchical clustering was performed using complete linkage.

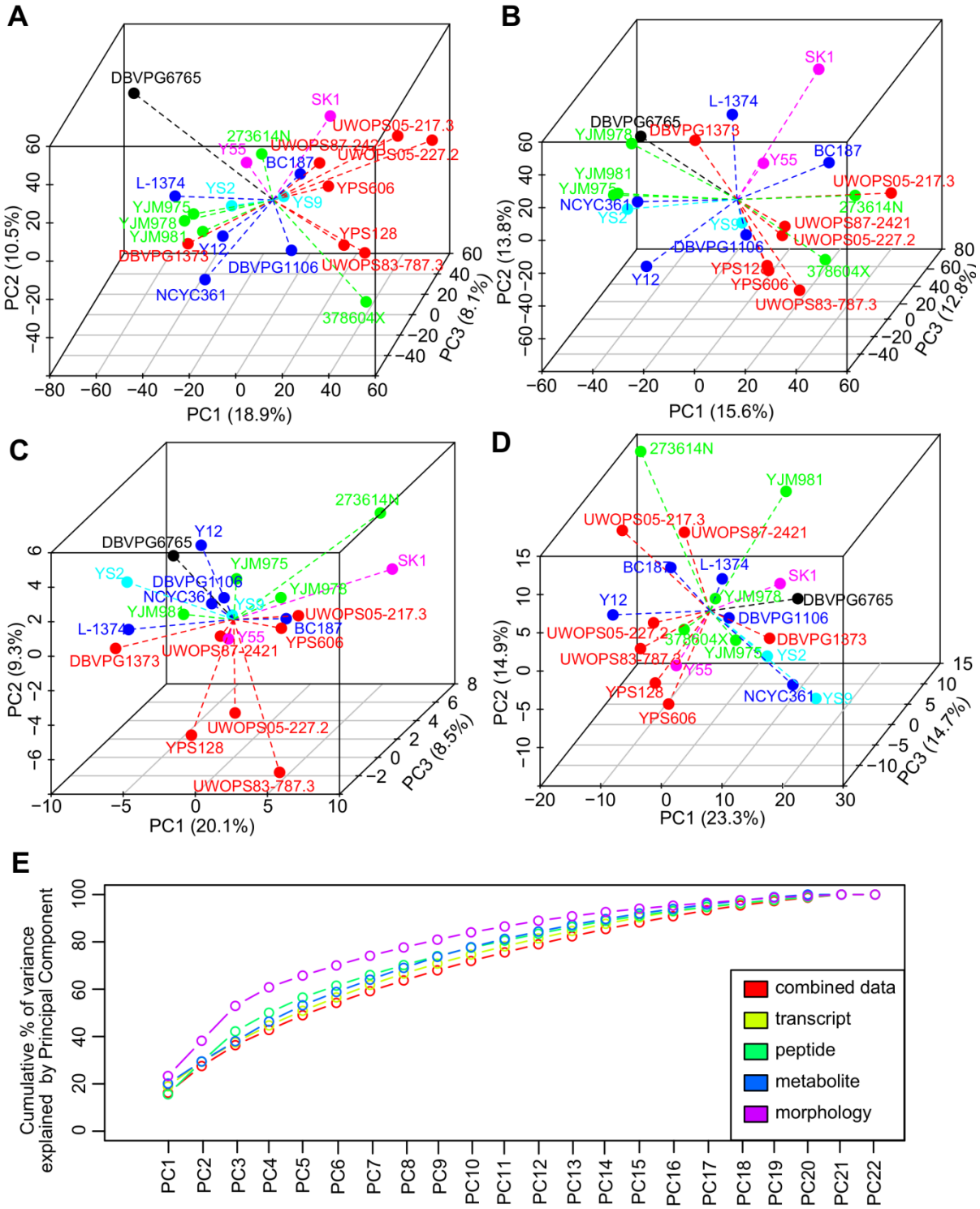



**Figure 5.4. Principal Component Analysis.**

**Principal Component Analysis of (A) transcripts, (B) peptides, (C) metabolites, and (D) morphology data.** Percentage variance explained by each Principal Component (PC) is shown in parentheses.

**(E) Cumulated percentage of variance explained by Principal Components.** Combined data: Figure 5.3A; transcripts, peptides, metabolites, and morphology: Figure 5.4A-D, respectively.

## Correlation between phenotype and lifespan

To identify a link between phenotypic variation and lifespan, we performed phylogenetic regression by generalized least squares, uncovering the phenotypes associated with longevity after accounting for the phylogenetic relationship of these strains (Felsenstein, 1985; Freckleton et al., 2002; Martins and Garland, 1991) (Chapter 2). Regression was performed between phenotypic values and any one of the following lifespan measurements: mean RLS, maximum RLS (Max RLS), log mean RLS (Log Mean RLS), and log maximum RLS (Log Max RLS). Different models of trait evolution were tested and the best-fit model was then selected based on maximal likelihood (Materials and Methods). To assess robustness of the relationship, we also left out one yeast strain at a time and re-calculated regression slopes using the remaining strains (Chapter 2). This ensured the overall relationship did not depend on a particular isolate.

The four different RLS measurements yielded very similar results, with Pearson correlation coefficients ranging between 0.90 and 0.98 for the regression slopes. We defined the top hits as phenotypes with statistically significant regression slopes under at least two different RLS measures, and identified 249 gene transcripts, 347 peptide fragments (representing 216 unique genes), 5 metabolites, and 43 morphology features (Table 5.2). Among the top gene transcripts and peptide fragments, only 10 unique genes were supported by both measures, consistent with the weak correlation between transcript and peptide levels noted above (Figure 5.2). When the mean peptide values were used for calculation, 88 genes reached statistical significance, 80 of which were also supported based on peptide fragments (Table 5.2).

With regard to morphology measures, features such as “maximal intensity of nuclear brightness divided by average”, “nucleus roundness in mother cell”, and “length from bud tip to mother cell’s short axis on nucleus C” showed significant negative correlation with RLS, whereas “fitness in nucleus C” correlated positively with longevity. Among the metabolite top hits, asparagine showed negative correlation with Max RLS (p value = 0.014) and Log Max RLS (p value = 0.017) (Figure 5.5A). A related amino acid, glutamine, also negatively correlated with Max RLS (p value = 0.042) and weakly with Log Max RLS (p value = 0.055) (Figure 5.5B). This was of note, since the

TOR pathway is known to be regulated by the levels of amino acids, especially intracellular glutamine (Martin and Hall, 2005). Treating yeast cells with methionine sulfoximine, an inhibitor of glutamine synthetase, has been shown to decrease both intracellular glutamine levels and TOR-dependent signaling (Crespo et al., 2002) while increasing RLS (Kaeberlein et al., 2005), whereas removal of either asparagine or glutamate from the medium produced a dose-dependent effect on chronological lifespan (CLS; CLS is the survival time of populations of nondividing cells, while RLS is the number of daughter cells produced by a mother cell prior to senescence; they are related but not identical) (Powers et al., 2006). We also found 2-octenoic acid to correlate negatively with Max RLS (p value =0.019) and Log Max RLS (p value =0.014) (Figure 5.5C). This compound is known to be elevated in mitochondria, but its effect on aging is not known. Some of the transcript and peptide top hits have also been implicated in lifespan regulation in yeast. For example, the protein levels of ADH1p (alcohol dehydrogenase, coded by *YOL086C*) correlated negatively with both Mean RLS and Max RLS, and deletion of *ADH1* was found to extend RLS by 23% in MATα and 15% in MATa (Smith et al., 2008). *DCW1* (also known as *YKL046C*, coding for a putative mannosidase in cell wall biosynthesis), whose transcript levels correlated negatively with all four RLS measurements, was previously identified in a genetic screen to increase yeast CLS when deleted (Matecic et al., 2010). In addition, a number of top hits correlating positively with longevity at the transcript (e.g. *VRP1* (*YLR337C*), *KGD1* (*YIL125W*)) and protein (e.g. PET9p (YBL030Cp), SP160p (YJL080Cp), GSY2p (YLR258Wp)) levels were previously shown to decrease RLS or CLS when deleted or mutated (Fabrizio et al., 2010; Laschober et al., 2010; Smith et al., 2008; Wang et al., 2008).

**Table 5.2. (see attached Excel file) Phenotypes with significant correlation to mean replicative lifespan (Mean RLS), maximum replicative lifespan (Max RLS), log mean replicative lifespan (Log Mean RLS), and log maximum replicative lifespan (Log Max RLS).** Only phenotypes with significant correlation to at least two of the RLS measurements are shown. "Best.Model" indicates the best-fit regression model; "Slope.coefficient" indicates the regression slope under the best-fit model (positive value indicates positive correlation; negative value indicates negative correlation); "p.value" indicates the regression slope p value; "RobustTest.Min", "RobustTest.Median", and "RobustTest.Max" indicate the minimal, median, and maximal p values when the regression is performed by leaving out one strain at a time. "peptide_mean" refers to the mean peptide values across all the peptide fragments for the unique genes. The "Note" column in "peptide_mean" indicates the number of peptide fragments for each unique gene.

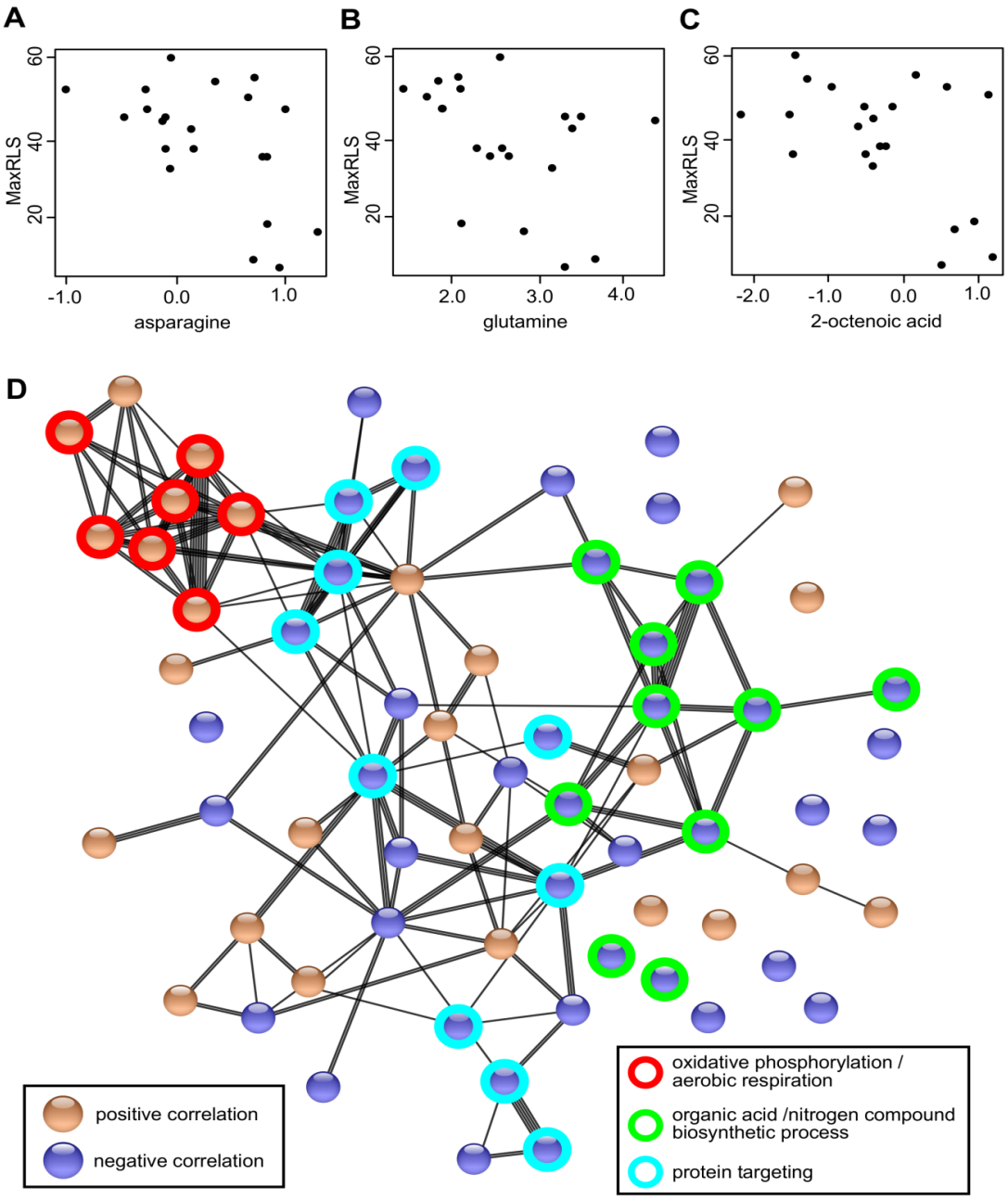



**Figure 5.5. Selected phenotypes correlating with replicative lifespan.**

**Levels of (A) asparagine, (B) glutamine, and (C) 2-octenoic acid negatively correlate with maximum replicative lifespan (MaxRLS).** Regression slope p values: (A) 0.014; (B) 0.042; (C) 0.019.

**(D) Protein-protein interaction network of the top hits identified by the mean peptide values.** The interaction network is based on STRING database (evidence view). Genes without interacting partners are omitted. Selected pathways are indicated by colored rings. Most of the peptides here showed significant correlation to all four RLS measures.

## Networks and pathways represented by top hits

To further understand the biological pathways underlying natural regulation of lifespan, we performed pathway enrichment analysis using DAVID (Huang da et al., 2009). The enrichment results for the peptide fragments were especially significant. Among the peptide fragments correlating positively with longevity, the enriched terms included “oxidative phosphorylation”, “mitochondrial respiratory chain”, “ion transport”, “hexose metabolic process”, “glucose metabolic process”, and “aerobic respiration”. On the other hand, for those correlating negatively with longevity, “amino acid biosynthesis”, “organic acid biosynthetic process”, “nitrogen compound biosynthetic process”, “nucleotide binding”, “cofactor binding”, and “glycolysis” were enriched. Many of these terms were similarly enriched when we carried out calculations using the mean peptide values. In comparison, the enrichment statistics were weaker for the transcripts, even though the numbers of top hits were similar. Among those with positive correlation, enrichment was observed for “ion transport”, “mitochondrial membrane part”, “ATP biosynthetic process”, “oxidative phosphorylation”, and “actin binding”. For the transcripts with negative correlation to lifespan, the enriched terms included “RNA polymerase II transcription factor activity”, “transcription regulator activity”, “microtubule”, “regulation of RNA metabolic process”, and “mRNA splicing”. Overall, the results suggest that the long-lived strains tend to up-regulate oxidative phosphorylation, aerobic respiration, and ion transport; and down-regulate transcription, splicing, and various biosynthetic processes (especially amino acid metabolism).

We visualised protein-protein interactions among the top hits using STRING (Jensen et al., 2009) and found the network is significantly enriched in interactions. The top hits identified using the mean peptide values were grouped into several prominent clusters, including oxidative phosphorylation and aerobic respiration (positive correlation); organic acid and nitrogen compound biosynthetic process (negative correlation); and protein targeting (negative correlation) (Figure 5.5D). Similar clusters of the top hits were observed for transcripts and peptide fragments data (Figure 5.6), suggesting that the top hits, rather than being a random collection of genes, represent interconnected nodes in regulatory networks and pathways.

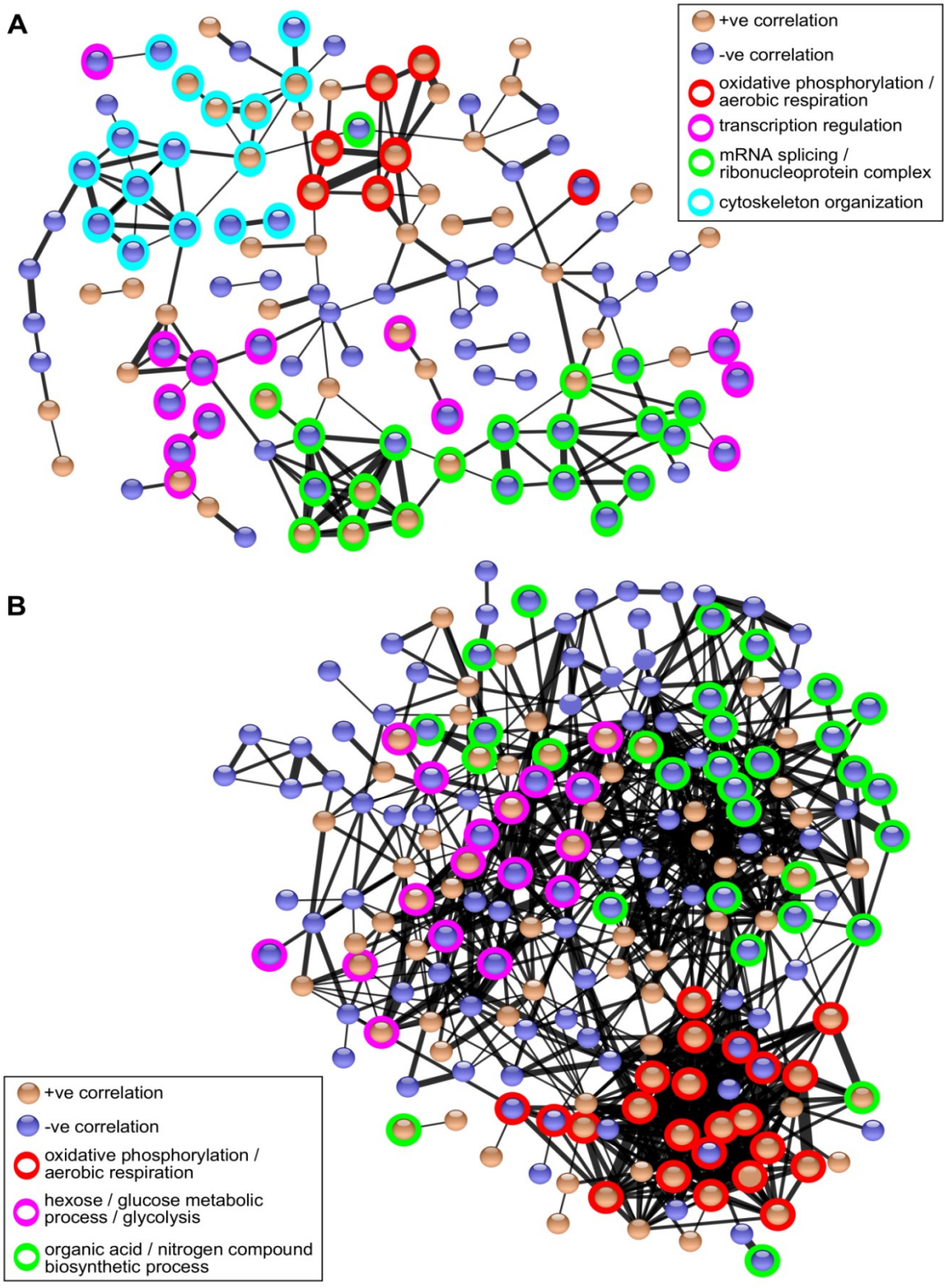



**Figure 5.6. Protein-protein interaction network of top hits identified based on (A) transcripts and (B) peptide fragments.** The interaction network is based on STRING database (evidence view, high confidence). Genes without interacting partners are omitted. Selected pathways are indicated by colored rings.

## Mitochondrial abundance and composition of the strains

Since the results suggested a relative up-regulation of oxidative phosphorylation and aerobic respiration among the long-lived strains, we examined more closely the nature of such differences. First, the genomic reads of these strains (Skelly et al., 2013) were used to calculate average coverage of the mitochondrial DNA relative to that of the chromosomes (Table 5.1, Figure 5.7A), as a proxy for mitochondria copy number. While the relative coverage was highest in YJM381 (8.0) and lowest in YPS128 (3.0), the values were relatively constant for most of the strains (4.0—5.0) and there was no overall correlation with longevity (Pearson correlation p value = 0.31 with Max RLS and 0.53 with Mean RLS). Moreover, Western blotting confirmed the similar expression of a mitochondrial marker protein Por1 in these strains (Figure 5.7B). In addition, the doubling times in glycerol media were also similar (100-120 min for most of the strains, with exception of >180 min for YS2, DBVPG1373, and Y55 strains; Figure 5.7C), suggesting the longevity variation across these strains could not be simply explained by total mitochondrial content or number.

However, when we examined the top hits based on mean peptide values (Table 5.2), a trend emerged. Approximately 1/3 of these peptides were related to mitochondria, with characteristic distribution patterns across the strains depending on their lifespans (Figure 5.8). For example, the longer-lived strains generally contained higher levels of proteins belonging to pyruvate dehydrogenase complex (PDH complex), Complex III, Complex IV, mitochondrial ATP synthase, inner membrane ADP/ATP carrier, as well as mitochondrial ribosomal proteins. On the other hand, long-lived strains had lower relative levels of outer membrane translocases, mitochondrial chaperonins, and certain metabolic enzymes (Figure 5.8B). The results suggest that the mitochondrial metabolism may vary widely across the strains according to their longevity. The longer-lived strains seem to enhance the electron transport chain and oxidative phosphorylation capacity, whereas the shorter-lived strains place more emphasis on protein folding and outer membrane transport. While the biological implications underlying these observations need to be further explored, the results show that distinct mitochondrial composition is associated with different yeast strains, and such patterns agree well with the observed lifespan variation.

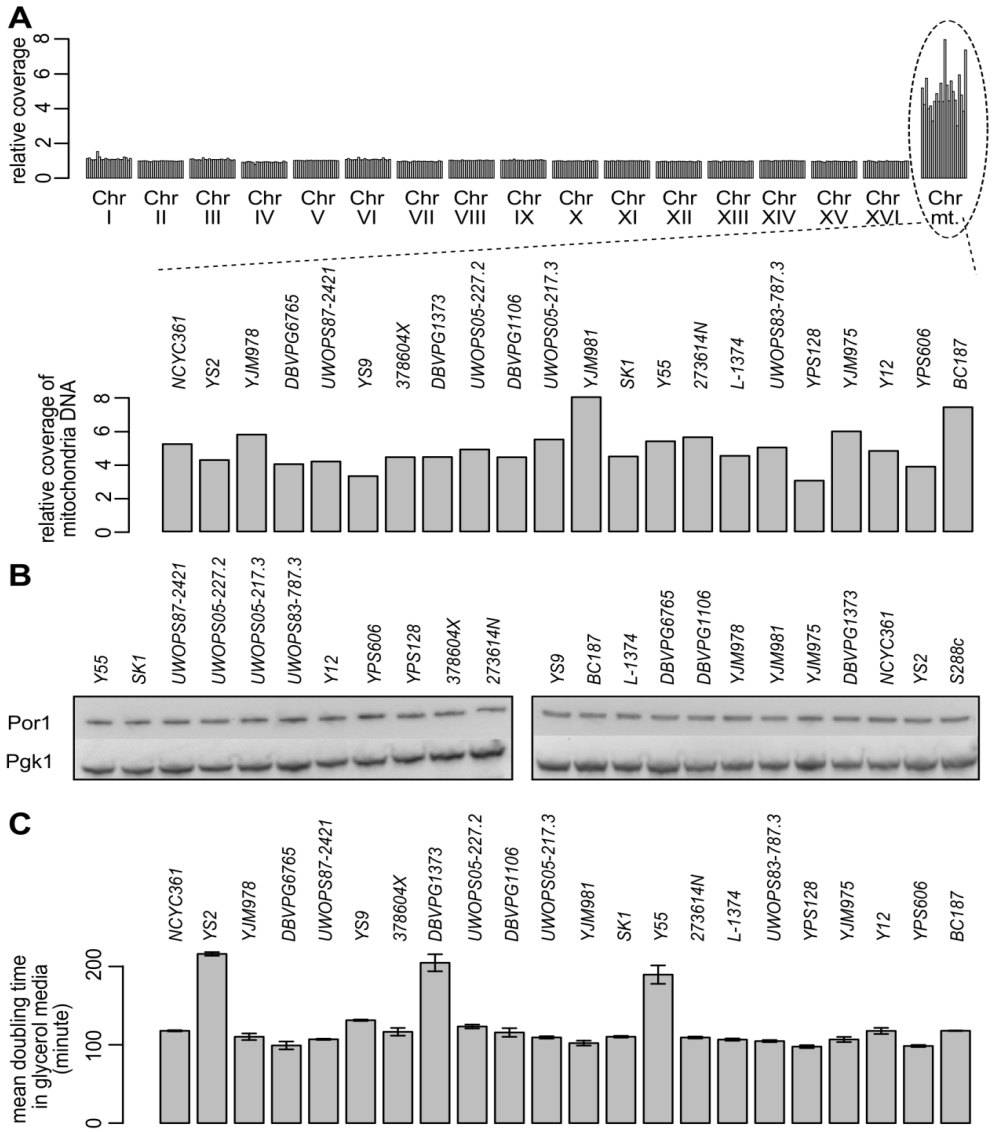



**Figure 5.7. Analysis of mitochondria across the strains.**

**(A) Relative coverage of the chromosomes and mitochondria DNA.** The upper panel shows the relative coverage for each of the chromosomes as well as mitochondria DNA (mean coverage across the chromosomes for each strain is set as 1.0). Each bar represents one chromosome in one strain. The strains are ordered by their mean replicative lifespan. The lower panel shows the enlarged view for mitochondria DNA.

**(B) Western blot shows the strains contain similar amount of mitochondria.** POR1 (coded by YNL055C): mitochondrial porin (voltage-dependent anion channel). PGK1 (coded by YCR012W): 3-phosphoglycerate kinase.

**(C) Mean doubling time in glycerol media. The error bars indicate standard error.**

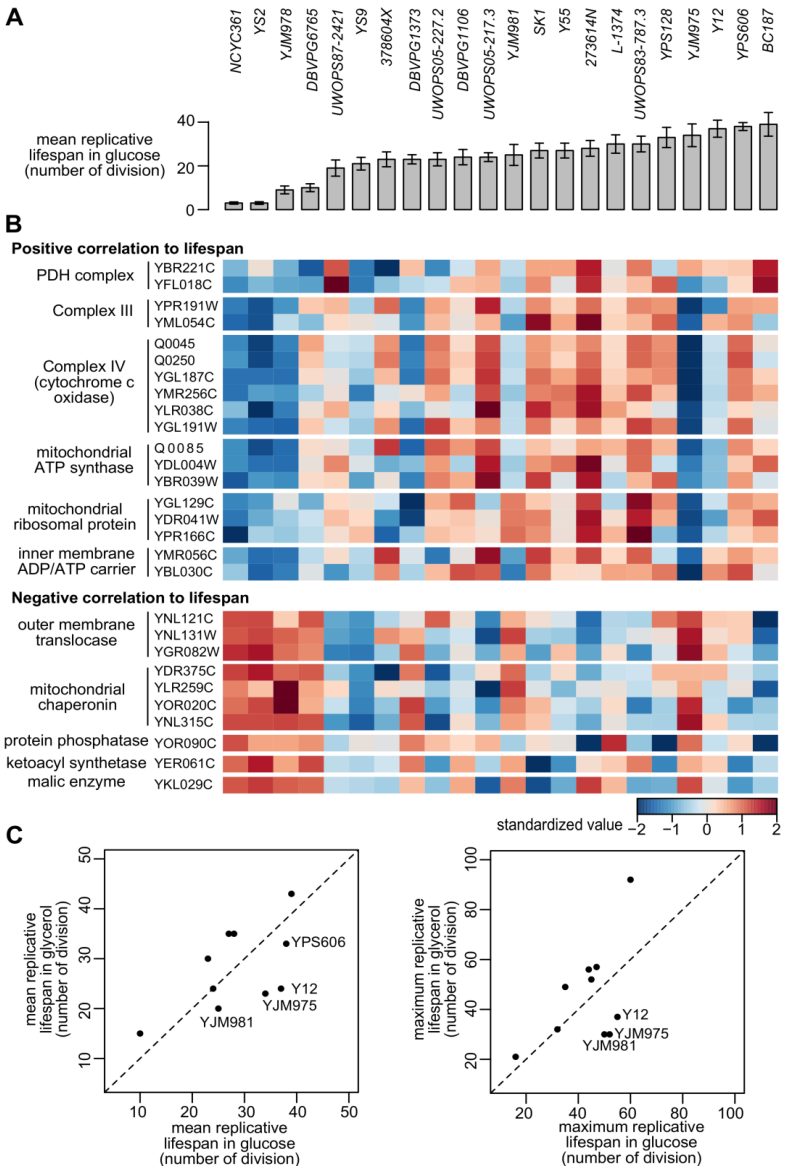



**Figure 5.8. Mitochondrial respiratory composition varies across the strains according to lifespan.**

**(A) Mean replicative lifespan of strains.** Strains are ordered according to their mean lifespan.

**(B) Levels of certain proteins correlate with lifespan.** The mean values of the selected proteins (related to mitochondrial function) are shown.

**(C) Effect of growth on a respiratory substrate on lifespan.** Replicative lifespan of 10 strains was tested on yeast peptone glycerol (3% YPG) plates and expressed as mean (left) and maximum (right) replicative lifespan. Except for the three long-lived outlier strains (YJM981, YJM975, and Y12), all strains either increased or did not change lifespan when their growth substrate was switched from glucose to glycerol..

## Comparison of related long-lived and short-lived strains

A number of our strains (YJM978, YJM981, YJM975, DBVPG1373, NCYC361, and YS2) are closely related to each other phylogentically (Figure 5.1A), but differ significantly in replicative lifespan (Figure 5.1B). In particular, they may be grouped into long-lived (YJM981, YJM975, and DBVPG1373) and short-lived (YJM978, NCY361, and YS2). If our findings above were valid, then we should observe similar sets of genes and pathways differentially expressed between these two groups. The analysis showed that the genes involved in "hexose metabolic process", "glucose metabolic process", and "glycolysis" were expressed highly in the long-lived strains, whereas those involved in "organic acid biosynthetic process", "amino acid biosynthesis", and "cofactor binding" were expressed at relatively low levels. Compared with the pathways we identified above, the genes involved in oxidative phosphorylation and aerobic respiration did not emerge as top hits, and there were not as many proteins related to mitochondria among these 6 strains. This is likely because all of these strains prefer fermentation over aerobic respiration (Figure 5.3B), and they already share similar mitochondrial composition profiles (Figure 5.8B). Among the strains designated as YJM are clinical isolates and their adaptation to longevity appears to be different from other strains. For example, YJM975 and YJM981 are long-lived, but their mitochondrial patterns are similar to the short-lived strains. Perhaps, their longevity is based on lineage-specific features that are not shared by other long-lived isolates. Nevertheless, among the long-lived strains we observed lower levels of expression of genes and proteins involved in biosynthetic processes (most of which were cytoplasmic), in agreement with our observations based on the 22 strains. This suggests that long lifespan can also arise without substantially altering the mitochondrial composition, although the reduction in biosynthesis seems to be a common feature.

# DISCUSSION

Availability of high quality genome sequence of *Saccharomyces cerevisiae* has made yeast an attractive model for dissecting complex traits associated with various phenotypes. Comparative genomics across multiple natural yeast isolates enabled the identification of extensive natural genetic variation at the nucleotide polymorphism (SNP) level and the elucidation of genotype to phenotype relation in several traits (Liti and Louis, 2012). Here, we ask: can similar strategies be applied to understand the common determinants of aging and longevity?

Using high throughput omics data, we examined 22 yeast natural isolates, which were found to vary over 10 fold in RLS. These isolates occupy diverse ecological niches and face different evolutionary pressures, so their natural lifespan variation must be encoded in their respective genomes. However, it has been challenging to characterize the cumulative effect of multiple alleles on a phenotype, especially if the underlying process involves a complex gene network. Alternatively, one may look at variation in transcriptome and proteome and correlate them and the associated pathways with the phenotypic traits, since the genotypic variation should be reflected in the expression variation in order to create the associated phenotypic differences (Brem et al., 2005).

To identify a link between transcript variation and lifespan, we performed phylogenetic regression and identified genes correlating with RLS, some of which were previously implicated in longevity regulation. Our pathway analysis showed that the long-lived strains tend to up-regulate oxidative phosphorylation, aerobic respiration, and ion transport; and down-regulate transcription, splicing, glycolysis and various biosynthetic processes, most notably amino acid synthesis. In particular, the variation in mitochondrial respiratory composition of these strains agrees well with their differences in lifespan. Mitochondria are at the heart of cellular metabolism and energy production, and increased mitochondrial respiratory capacity has been linked to longevity (Bonawitz et al., 2007; Pan and Shadel, 2009). We hypothesize that many of these natural isolates reside in the environments with low fermentable carbon sources, so that they undergo diauxic shift and metabolize respiratory carbon sources. Shifting from fermentable (glucose) to respiratory carbon sources is known to extend both replicative and chronological lifespan in yeast (Delaney et al., 2011).

Genetic variation responsible for lifespan differences may also affect metabolite levels and morphology. Among the examined metabolites, glutamine and asparagine showed strong negative correlation with RLS, which is consistent with the known inhibition of TOR activity and extension of chronological lifespan by removing glutamine or asparagine from yeast media (Powers et al., 2006) and extension of RLS by treating cells with methionine sulfoximine (Kaeberlein et al., 2005). In terms of cell morphology, a number of nuclear features such as brightness, roundness and distance to bud tip showed significant negative correlation with RLS, whereas "fitness in nucleus C" correlated positively with longevity. Interestingly, longer-lived strains tend to possess smaller mother cell volume (Figure 5.9), indicative of a potential compromise between mother cell size and lifespan, as has been previously observed for long-lived cells treated with ibuprofen (He et al., 2014). In agreement, inverse correlation between cell size and lifespan has been observed in yeast previously (Yang et al., 2011). Thus, here too, natural changes in a phenotype can be linked with longevity interventions and maybe used as aging biomarkers.

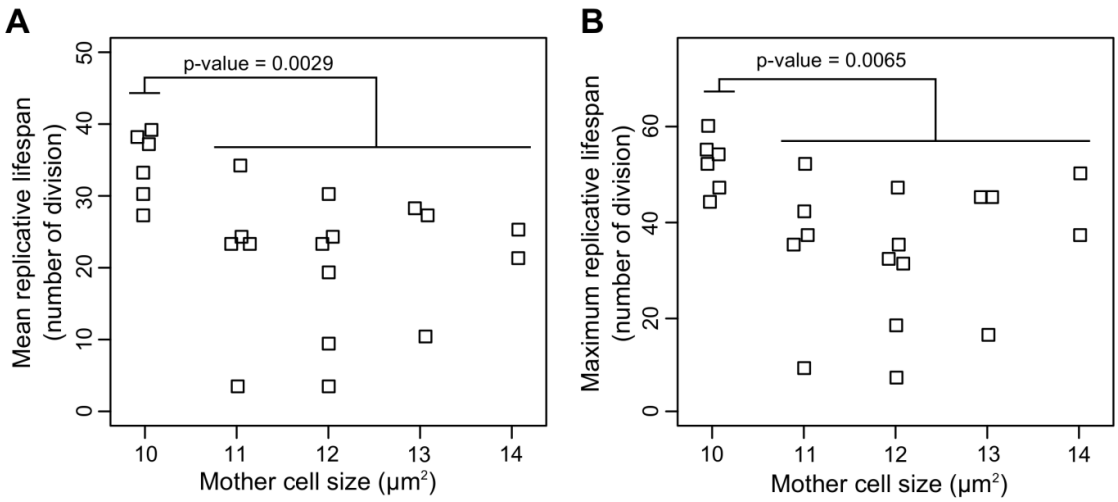



**Figure 5.9. Variation of (A) mean replicative lifespan and (B) maximum replicative lifespan according to mother cell size.** p values refer to ANOVA test comparing 10 $\mu m^2$ mother cells against 11-14 $\mu m^2$ mother cells. Overlapping points are jittered for visualization.

It should be noted that our method is limited to identifying the genetic and metabolic processes that show concerted changes in relation to longevity across these 22 strains, which are more likely to be generalizable and do not depend on single or a few strains. On the other hand, an exceptionally long-lived or short-lived strain can also have arisen due to certain strain-specific changes that are not shared by other isolates, and such changes may be missed by our method. Comparison among the 6 related long-lived and short-lived strains suggests there may be more than one way to achieve long lifespan, and it will be useful to examine strains across different evolutionary distances to identify the common features.

To our knowledge, this is the first report that analyzes inter-strain natural diversity of RLS at the population level using high throughput data. Natural isolates occupying diverse ecological niches may face different selection pressures and have evolved to adjust their gene expression, metabolism, longevity, and reproduction to ensure survival and propagation (Spor et al., 2009). While evolution can sometimes provide different solutions to the same challenge (Romano et al., 2010), our data suggest a consistent set of genes and pathways are responsible for modulating the lifespan trait across a broad diversity of wild yeast isolates.

Finally, it has been unclear whether the previous findings of various longevity regulator genes identified in the laboratory setting could be translated to the natural environment. A possibility is that these lifespan-extending interventions come at the expense of fitness. For example, many longest-lived *C. elegans* laboratory mutants tend to develop and move slowly and often show reduced fecundity, so they will probably be eliminated quickly for lack of competitive advantage in the wild. Consistent with this, 65% of long lived single gene deletion mutants in yeast demonstrated significantly reduced fitness compared to isogenic wild type cells (Delaney et al., 2011). Our results show that natural changes in lifespan need not compromise fitness, as longer-lived yeast isolates are presumably well adapted to their respective ecological niches. In addition, our analysis is unbiased with regard to the genes and pathways involved in lifespan control and supports a possibility that multiple correlates cumulatively contribute to the longevity phenotypes. Specifically, we found that the ability of yeast cells to rely more heavily on respiration and repress their anabolic programs, even under conditions of glucose excess, are among the key adaptations that lead to increased lifespan.

Importantly, since CR and TOR signaling are also known to extend lifespan by activating respiration and inhibiting biosynthetic processes, these data show that natural plasticity of yeast lifespan is shaped by pathways that both impose little cost on fitness and are amenable to dietary intervention. Thus, environment may be a trigger for changes associated with increased lifespan that are then fixed in the genomes.

## EXPERIMENTAL PROCEDURES

### Yeast strains

Diploid natural isolates of *S. cerevisiae* were obtained from the Sanger Institute and are summarized in Table 5.1. These strains are well-characterized (Bergstrom et al., 2014; Liti et al., 2009). The diploid laboratory strain BY4743 was purchased from ATCC.

### Replicative lifespan assay

For each strain, cells were freshly grown on YPD plates prior to dissections. Several colonies were streaked onto new YPD plates using pipette tips. After overnight growth, 40-50 dividing cells were lined up. Newborn daughter cells were chosen for RLS assays after the first division using a micromanipulator. Plates were incubated at 30 °C between dissections and left at 4 °C during night. All RLS assays were performed in standard YPD plates with 2% glucose as previously described (Steffen et al., 2009). For each natural isolate, at least two independent assays were performed. Each assay contained 20-40 mother cells.

### Phenotypic data

Growth rates were determined using a Bioscreen C MBR machine by analysis of optical density in the $OD_{420-580}$ range as previously described in combination with the YODA Software package (Olsen et al., 2010). The data on transcripts, peptides, metabolites and morphology were downloaded from Yeast Resource Center (http://www.yeastrc.org/g2p/download.do). Values corresponding to the 22 strains were extracted; metabolite data were not available for 378604X. Metabolites with missing values in more than one strain (other than 378604X) were discarded; the remaining missing values (6 out of 107 metabolites) were imputed based on 10 nearest neighbors. For comparison across the phenotypic data, the values were standardized across the strain by setting mean = 0 and standard deviation = 1. In addition, for genes represented by multiple peptides, we calculated the mean standardized values to perform the regression.

### Principal component analysis (PCA)

PCA was performed on standardized values using R package “stats” (R Development Core Team, 2013). To identify the underlying pathways, the factors in each of the first three principal

components (PCs) were ranked by their contributions, and pathway enrichment analysis was performed on the top 10% factors using DAVID after correcting for background.

**Phylogenetic regression by generalized least square**

See Chapter 2. The cut-off for top hits was p value $< 0.05$ in at least two RLS measurements.

**Relative coverage of mitochondrial DNA**

Genomic reads of strains examined in our study were downloaded from Yeast Resource Center and mapped to reference genome of *Saccharomyces cerevisiae* strain S288c. The average coverage per base across the chromosomes (excluding positions 45,000 to 50,000 of chromosome XII) was calculated using R package "ShortRead" for each strain. The relative coverage of mitochondrial DNA was expressed as the ratio of per-base coverage of mitochondrial DNA to per-base coverage of chromosomes.

**Differential expression between long-lived and short-lived groups.**

Six closely related strains were grouped into long-lived (YJM981, YJM975, DBVPG1373) and short-lived (YJM978, NCY361, YS2). Differential expressions of the phenotypic data were calculated using R package "limma" (R Development Core Team, 2013).

**Mitochondrial protein expression**

Logarithmically growing cells (5 ml and $OD_{600}$=0.6) were harvested and incubated in 150 µl extraction buffer (1.85 mM NaOH and 2% β-mercaptoethanol) on ice for 10 min. Then, 150 µl of 50% TCA (trichloroacetic acid) was added and incubated for 30 min on ice. After incubation, the cells were pelleted and supernatant aspirated. After 30 min of air drying, the pellets were heated at 60$^{o}$C in SDS loading buffer and 4 µl of each sample was analyzed by SDS-PAGE. To examine the expression of a mitochondrial protein, Western blotting was carried out with antibodies against mitochondrial outer membrane protein Por1 (Abcam). The membranes were stripped and developed with antibodies against phosphoglycerate kinase (Pgk1) (Life technologies) as an internal loading control.

# REFERENCES


Bergstrom, A., Simpson, J.T., Salinas, F., Barre, B., Parts, L., Zia, A., Nguyen Ba, A.N., Moses, A.M., Louis, E.J., Mustonen, V.*, et al.* (2014). A high-definition view of functional genetic variation from natural yeast genomes. Mol. Biol. Evol. *31*, 872-888.

Bonawitz, N.D., Chatenay-Lapointe, M., Pan, Y., and Shadel, G.S. (2007). Reduced TOR signaling extends chronological life span via increased respiration and upregulation of mitochondrial gene expression. Cell Metab. *5*, 265-277.

Botta, G., Turn, C.S., Quintyne, N.J., and Kirchman, P.A. (2011). Increased iron supplied through Fet3p results in replicative life span extension of Saccharomyces cerevisiae under conditions requiring respiratory metabolism. Exp. Gerontol. *46*, 827-832.

Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., Boer, V.M., Troyanskaya, O.G., and Botstein, D. (2008). Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Mol. Biol. Cell *19*, 352-367.

Brem, R.B., Storey, J.D., Whittle, J., and Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. Nature *436*, 701-703.

Castrillo, J.I., Zeef, L.A., Hoyle, D.C., Zhang, N., Hayes, A., Gardner, D.C., Cornell, M.J., Petty, J., Hakes, L., Wardleworth, L.*, et al.* (2007). Growth control of the eukaryote cell: a systems biology study in yeast. J. Biol. *6*, 4.

Crespo, J.L., Powers, T., Fowler, B., and Hall, M.N. (2002). The TOR-controlled transcription activators GLN3, RTG1, and RTG3 are regulated in response to intracellular levels of glutamine. Proc. Natl. Acad. Sci. USA *99*, 6784-6789.

Delaney, J.R., Murakami, C.J., Olsen, B., Kennedy, B.K., and Kaeberlein, M. (2011). Quantitative evidence for early life fitness defects from 32 longevity-associated alleles in yeast. Cell Cycle *10*, 156-165.

Fabrizio, P., Hoon, S., Shamalnasab, M., Galbani, A., Wei, M., Giaever, G., Nislow, C., and Longo, V.D. (2010). Genome-wide screen in Saccharomyces cerevisiae identifies vacuolar protein sorting, autophagy, biosynthetic, and tRNA methylation genes involved in life span regulation. PLoS Genet. *6*, e1001024.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. Am. Nat. *125*, 1-15.

Finch, C.E., and Ruvkun, G. (2001). The genetics of aging. Annu. Rev. Genomics Hum. Genet. *2*, 435-462.

Freckleton, R.P., Harvey, P.H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. *160*, 712-726.

Gladyshev, V.N. (2012). On the cause of aging and control of lifespan: heterogeneity leads to inevitable damage accumulation, causing aging; control of damage composition and rate of accumulation define lifespan. Bioessays *34*, 925-929.

He, C., Tsuchiyama, S.K., Nguyen, Q.T., Plyusnina, E.N., Terrill, S.R., Sahibzada, S., Patel, B., Faulkner, A.R., Shaposhnikov, M.V., Tian, R.*, et al.* (2014). Enhanced Longevity by Ibuprofen, Conserved in Multiple Species, Occurs in Yeast through Inhibition of Tryptophan Import. PLoS Genet. *10*, e1004860.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M.*, et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412-416.

Johnson, S.C., Rabinovitch, P.S., and Kaeberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. Nature *493*, 338-345.

Kaeberlein, M., Powers, R.W., 3rd, Steffen, K.K., Westman, E.A., Hu, D., Dang, N., Kerr, E.O., Kirkland, K.T., Fields, S., and Kennedy, B.K. (2005). Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science *310*, 1193-1196.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Laschober, G.T., Ruli, D., Hofer, E., Muck, C., Carmona-Gutierrez, D., Ring, J., Hutter, E., Ruckenstuhl, C., Micutkova, L., Brunauer, R.*, et al.* (2010). Identification of evolutionarily conserved genetic regulators of cellular aging. Aging Cell *9*, 1084-1097.

Lin, S.J., Kaeberlein, M., Andalis, A.A., Sturtz, L.A., Defossez, P.A., Culotta, V.C., Fink, G.R., and Guarente, L. (2002). Calorie restriction extends Saccharomyces cerevisiae lifespan by increasing respiration. Nature *418*, 344-348.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V.*, et al.* (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337-341.

Liti, G., and Louis, E.J. (2012). Advances in quantitative trait analysis in yeast. PLoS Genet. *8*, e1002912.

Lopez-Lluch, G., Hunt, N., Jones, B., Zhu, M., Jamieson, H., Hilmer, S., Cascajo, M.V., Allard, J., Ingram, D.K., Navas, P.*, et al.* (2006). Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. Proc. Natl. Acad. Sci. USA *103*, 1768-1773.

Martin, D.E., and Hall, M.N. (2005). The expanding TOR signaling network. Curr. Opin. Cell Biol. *17*, 158-166.

Martins, E.P., and Garland, T., Jr. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution *45*, 534-557.

Matecic, M., Smith, D.L., Pan, X., Maqani, N., Bekiranov, S., Boeke, J.D., and Smith, J.S. (2010). A microarray-based genetic screen for yeast chronological aging factors. In PLoS Genet., pp. e1000921.

McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The effect of retarded growth upon the length of life span and upon the ultimate body size. J. Nutr. *5*, 155-171; discussion 172.

Murakami, C., Delaney, J.R., Chou, A., Carr, D., Schleit, J., Sutphin, G.L., An, E.H., Castanza, A.S., Fletcher, M., Goswami, S.*, et al.* (2012). pH neutralization protects against reduction in replicative lifespan following chronological aging in yeast. Cell Cycle *11*, 3087-3096.

Olsen, B., Murakami, C.J., and Kaeberlein, M. (2010). YODA: software to facilitate high-throughput analysis of chronological life span, growth rate, and survival in budding yeast. BMC Bioinformatics *11*, 141.

Pan, Y., and Shadel, G.S. (2009). Extension of chronological life span by reduced TOR signaling requires down-regulation of Sch9p and involves increased mitochondrial OXPHOS complex density. Aging (Albany NY) *1*, 131-145.

Pitt, J.N., Leiser, S.F., and Kaeberlein, M. (2014). Oxygen and Aging. In Annual Review of Gerontology and Geriatrics (Springer Publishing Company), pp. 59-81.

Powers, R.W., 3rd, Kaeberlein, M., Caldwell, S.D., Kennedy, B.K., and Fields, S. (2006). Extension of chronological life span in yeast by decreased TOR pathway signaling. Genes Dev. *20*, 174-184.

R Development Core Team (2013). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Regenberg, B., Grotkjaer, T., Winther, O., Fausboll, A., Akesson, M., Bro, C., Hansen, L.K., Brunak, S., and Nielsen, J. (2006). Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol. *7*, R107.

Reznick, D., Buckwalter, G., Groff, J., and Elder, D. (2001). The evolution of senescence in natural populations of guppies (Poecilia reticulata): a comparative approach. Exp. Gerontol. *36*, 791-812.

Romano, G.H., Gurvich, Y., Lavi, O., Ulitsky, I., Shamir, R., and Kupiec, M. (2010). Different sets of QTLs influence fitness variation in yeast. Mol. Syst. Biol. *6*, 346.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Schulz, T.J., Zarse, K., Voigt, A., Urban, N., Birringer, M., and Ristow, M. (2007). Glucose restriction extends Caenorhabditis elegans life span by inducing mitochondrial respiration and increasing oxidative stress. Cell Metab. *6*, 280-293.

Skelly, D.A., Merrihew, G.E., Riffle, M., Connelly, C.F., Kerr, E.O., Johansson, M., Jaschob, D., Graczyk, B., Shulman, N.J., Wakefield, J.*, et al.* (2013). Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res. *23*, 1496-1504.

Smith, E.D., Tsuchiya, M., Fox, L.A., Dang, N., Hu, D., Kerr, E.O., Johnston, E.D., Tchao, B.N., Pak, D.N., Welton, K.L.*, et al.* (2008). Quantitative evidence for conserved longevity pathways between divergent eukaryotic species. Genome Res. *18*, 564-570.

Spor, A., Nidelet, T., Simon, J., Bourgais, A., de Vienne, D., and Sicard, D. (2009). Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of Saccharomyces cerevisiae. BMC Evol. Biol. *9*, 296.

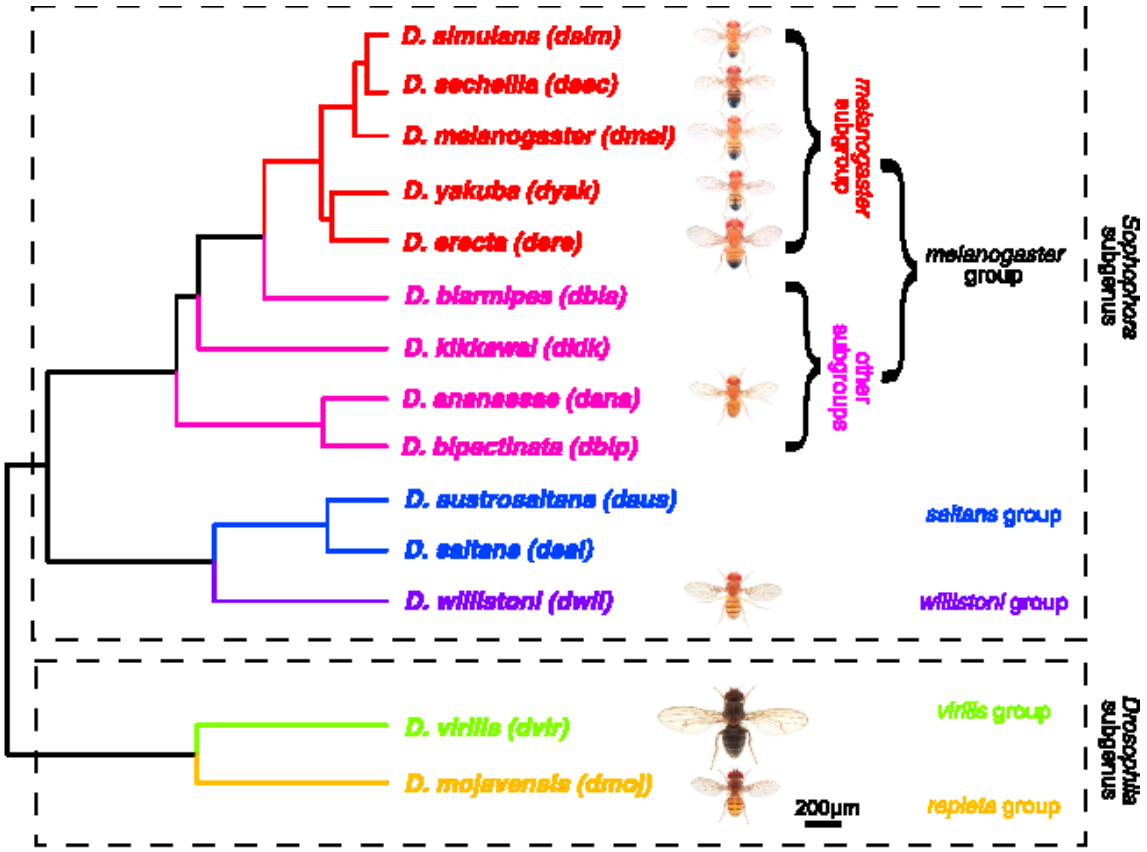
Steffen, K.K., Kennedy, B.K., and Kaeberlein, M. (2009). Measuring replicative life span in the budding yeast. Journal of visualized experiments.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. *30*, 2725-2729.

Wang, X., Salinas, K., Zuo, X., Kucejova, B., and Chen, X.J. (2008). Dominant membrane uncoupling by mutant adenine nucleotide translocase in mitochondrial diseases. Hum. Mol. Genet. *17*, 4036-4044.

Yang, J., Dungrawala, H., Hua, H., Manukyan, A., Abraham, L., Lane, W., Mead, H., Wright, J., and Schneider, B.L. (2011). Cell size and growth rate are major determinants of replicative lifespan. Cell Cycle *10*, 144-155.

# Chapter 6 Fly Transcriptome




This chapter is based on the following manuscript:

Siming Ma, Byung Cheon Lee, Marco Mariotti, Nadezhda Zemskaya, Roderic Guigo, Alexey A. Moskalev, Vadim N. Gladyshev. **Comparative transcriptomics across 14 *Drosophila* species reveals signatures of longevity.** *Manuscript submitted.*

# ABSTRACT

Lifespan varies dramatically among species, but the biological basis is not well understood. Previous studies in model organisms revealed the importance of nutrient sensing, mTOR, NAD/sirtuins and insulin/IGF1 signaling in lifespan control. By studying the life history traits and transcriptomes of 14 Drosophila species differing more than 6 fold in lifespan, we explored expression divergence across these flies and identified genes and processes that correlate with longevity. These longevity signatures suggested that longer-lived flies up-regulate fatty acid metabolism, down-regulate neuronal system development and activin signaling, and alter dynamics of RNA splicing. Interestingly, there was significant overlap between the genes correlating with natural lifespan and those found to influence lifespan in model organisms. Moreover, these gene expression patterns resembled those of flies under dietary restriction and several other lifespan-extending interventions. The data suggest that natural variation in longevity across species can be represented by gene expression patterns and is achieved via pathways amenable to dietary intervention.

# INTRODUCTION

Since the early 20th century, fruit flies have remained a vital tool in cell biology, genetics, development, and evolution. While the best known early example is probably the use of *Drosophila melanogaster* by Thomas H. Morgan to elucidate the chromosomal theory of inheritance, other species such as *D. pseudoobscura* and *D. virilis* have long been used to study evolution and speciation (Markow and O'Grady, 2007). The entire genus *Drosophila* contains over 2000 described species that occupy diverse ecological niches such as forests, deserts, and cosmopolitan areas (Markow and O'Grady, 2005; Schnebel and Grossfield, 1983). With the recent completion of full genome sequences of 12 *Drosophila* species (Clark et al., 2007), researchers are able to explore various aspects of their biology in much greater depth. Examination across multiple evolutionarily related lineages can reveal insights on the unique biology of flies, as well as new themes and biological mechanisms that apply across diverse life forms.

Given their relatively short life cycle, fruit flies are particularly suitable for studying longevity and aging. Under laboratory settings, the lifespan of *D. melanogaster* has been successfully increased by genetic manipulations (Clancy et al., 2001; Hwangbo et al., 2004; Kapahi et al., 2004; Lin et al., 1998; Orr and Sohal, 1994; Parkes et al., 1998; Sun et al., 2002; Tatar et al., 2001), dietary interventions (Chapman and Partridge, 1996; Grandison et al., 2009; Lee et al., 2014; Magwere et al., 2004; Mair et al., 2003; Min and Tatar, 2006), and pharmacological treatments (Bjedov et al., 2010; Danilov et al., 2013; Wang et al., 2013). These findings are similar to those reported in other model organisms and highlight the important role of nutrient sensing, mTOR, NAD/sirtuins, insulin/IGF1 signaling pathways and other systems in lifespan control (Fontana et al., 2010; He et al., 2014).

Lifespan, weight, time to maturity and other life history traits naturally differ across various *Drosophila* species as the result of millions of years of natural selection, drift and adaptation. Since the divergence from a common ancestor, *Drosophila* lifespan has increased along certain lineages but decreased in others (Schnebel and Grossfield, 1983), indicating that longevity can be modulated in both directions on the evolutionary time scale. The heritability and stability of species lifespan across generations indicates a genetic basis for the longevity determinant(s). Furthermore, the species

naturally differing in lifespan are fit and well adapted to thrive in their respective ecological niches, whereas many long-lived model organisms produced in the laboratory setting by genetic or dietary interventions are often less robust, suggesting that nature has managed to modulate longevity without compromising fitness.

By identifying the genes whose expression levels correlate with lifespan across the *Drosophila* lineage, one may obtain clues about the pathways involved and ultimately the mechanisms through which nature modulates longevity. Such gene expression patterns may also be compared with known lifespan-extension strategies to identify commonality in their effects. Here, we used 14 *Drosophila* species spanning 5 taxonomical groups and more than 30 million years of evolution, examining their lifespan, body mass, development time, and gene expression profiles. We explored the relationships among various life-history traits, identified the pathways that diverged significantly across these species, and observed the role of stabilizing selection in gene expression variation. We also identified the genes and pathways with significant positive or negative correlation to longevity, after taking into account the influence of phylogeny and body mass differences. Finally, we analyzed our list of genes against previously published lifespan extension data, offering various insights into regulation of longevity.

# RESULT

## Life history traits in *Drosophila*

The 14 species surveyed in this study fall into two subgenera, *Drosophila* and *Sophophora*, which diverged ~30 million years ago (Figure 6.1, Table 6.1). Subgenus *Drosophila* is represented by *D. virilis* and *D. mojavensis*, which together form the *virilis-repleta* radiation (Markow and O'Grady, 2007). Within subgenus *Sophophora*, *D. melanogaster* and eight related species belong to the *melanogaster* species group, whereas the other two species groups (*saltans* and *willistoni*) are represented by one to two members. The flies within the *melanogaster* species group can be classified further into *melanogaster* subgroup and other subgroups (Figure 6.1).

We first characterized lifespan, body mass, and developmental time of these flies. *D. virilis* was morphologically distinct, largest in size (almost 2 grams), and longest-lived (mean lifespan: male 52.8 days; female 62.3 days), whereas *D. sechellia*, *D. yakuba*, and *D. bipectinata* were among the shortest-lived (mean lifespan 8-16 days) and relatively small (0.5-0.8 grams) (Figure 6.2A and Table 6.1). For most of the other species, the mean lifespan ranged between 20 and 40 days, consistent with the literature records (Schnebel and Grossfield, 1983). Within each species, female flies were also generally larger in size than male flies (Table 6.1). When their body weights were plotted against median lifespan on log-scale, a strong positive correlation was observed (Figure 6.2B; Pearson correlation coefficient=0.59, p value=$8.7\times10^{-4}$). Furthermore, longer-lived species also developed more slowly (Figure 6.2C), suggesting the relationship between longevity and various life history traits previously observed in mammals and birds (Fushan et al., 2015; Speakman, 2005) also applies across the *Drosophila* species.

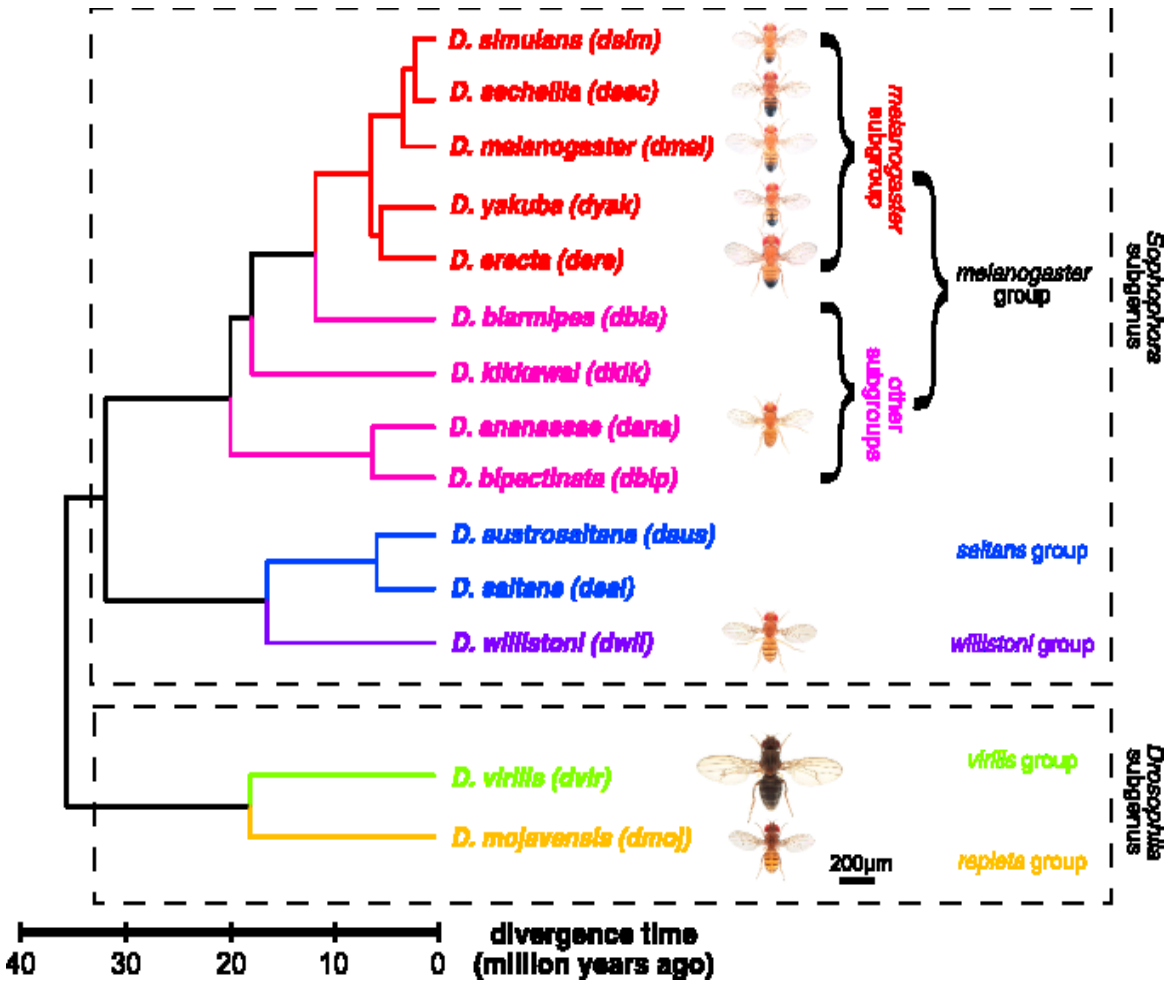



**Figure 6.1.** ***Drosophila*** **species surveyed in the current study.** The species are coloured by taxonomical grouping (abbreviations of species names shown in parentheses). The tree is based on amino acid sequences of orthologs and calibrated using previous estimates (Russo et al., 1995). Images of flies (all males) were obtained from Flybase (copyright: Nicolas Gompel).

**Table 6.1. Species information and life history.**

| Scientific Name | Abbre. | Stock Number | Reference Genome | Gender | No. of fly | Lifespan (day) | | Adult Weight (g) | Develop-ment time (day) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | mean ± SE | median | | |
| *Drosophila simulans* | dsim | 14021-0251.194 | Ensembl WUGSC1.21 | male | 102 | 37.44 ± 0.63 | 39.0 | 0.690 | 10 |
| | | | | female | 102 | 28.79 ± 0.53 | 30.0 | 0.843 | |
| *Drosophila sechellia* | dsec | 14021-0248.25 | Ensembl dsec_caf1.21 | male | 104 | 16.76 ± 1.05 | 12.0 | 0.600 | 12 |
| | | | | female | 105 | 8.51 ± 0.34 | 9.0 | 0.800 | |
| *Drosophila melanogaster* | dmel | 14021-0231.36 | Ensembl BDGP5.75 | male | 81 | 36.22 ± 1.10 | 36.0 | 0.723 | 10 |
| | | | | female | 100 | 26.22 ± 0.75 | 27.0 | 1.063 | |
| *Drosophila yakuba* | dyak | 14021-0261.01 | Ensembl dyak_r1.3_FB2008_07.21 | male | 106 | 9.37 ± 0.42 | 6.0 | 0.417 | 11 |
| | | | | female | 107 | 16.79 ± 0.59 | 18.0 | 0.767 | |
| *Drosophila erecta* | dere | 14021-0224.01 | Ensembl dere_caf1.21 | male | 89 | 21.03 ± 0.86 | 21.0 | 0.567 | 12 |
| | | | | female | 101 | 31.78 ± 0.96 | 33.0 | 0.873 | |
| *Drosophila biarmipes* | dbia | 14023-0361.09 | NCBI AFFD00000000.2 | male | 99 | 36.12 ± 0.91 | 39.0 | 0.543 | 11 |
| | | | | female | 89 | 31.42 ± 1.00 | 33.0 | 0.707 | |
| *Drosophila kikkawai* | dkik | 14028-0561.14 | NCBI AFFH00000000.2 | male | 58 | 25.40 ± 1.13 | 27.0 | 0.610 | 12 |
| | | | | female | 78 | 25.19 ± 0.71 | 27.0 | 0.897 | |
| *Drosophila ananassae* | dana | 14024-0371.13 | Ensembl dana_caf1.21 | male | 100 | 29.88 ± 0.75 | 33.0 | 0.707 | 10 |
| | | | | female | 100 | 25.98 ± 0.85 | 27.0 | 1.130 | |
| *Drosophila bipectinata* | dbip | 14024-0381.19 | NCBI AFFE00000000.2 | male | 103 | 12.29 ± 0.49 | 12.0 | 0.453 | 11 |
| | | | | female | 106 | 12.71 ± 0.34 | 12.0 | 0.760 | |
| *Drosophila austrosaltans* | daus | 14045-0881.00 | unpublished | male | 94 | 29.46 ± 1.09 | 28.5 | 0.917 | 11 |
| | | | | female | 99 | 23.61 ± 0.72 | 24.0 | 1.443 | |
| *Drosophila saltans* | dsal | 14045-0911.00 | unpublished | male | 104 | 27.00 ± 0.61 | 27.0 | 0.733 | 11 |
| | | | | female | 102 | 30.12 ± 0.88 | 30.0 | 1.183 | |
| *Drosophila willistoni* | dwil | 14030-0811.24 | Ensembl dwil_caf1.21 | male | 98 | 27.67 ± 0.86 | 30.0 | 0.533 | 13 |
| | | | | female | 99 | 21.91 ± 0.65 | 21.0 | 0.833 | |
| *Drosophila virilis* | dvir | 15010-1051.87 | Ensembl dvir_caf1.21 | male | 82 | 52.79 ± 2.29 | 61.5 | 1.957 | 18 |
| | | | | female | 79 | 62.28 ± 2.95 | 75.0 | 2.187 | |
| *Drosophila mojavensis* | dmoj | 15081-1352.22 | Ensembl dmoj_caf1.21 | male | 84 | 42.46 ± 1.99 | 42.0 | 0.860 | 12 |
| | | | | female | 88 | 36.31 ± 1.57 | 36.0 | 1.090 | |

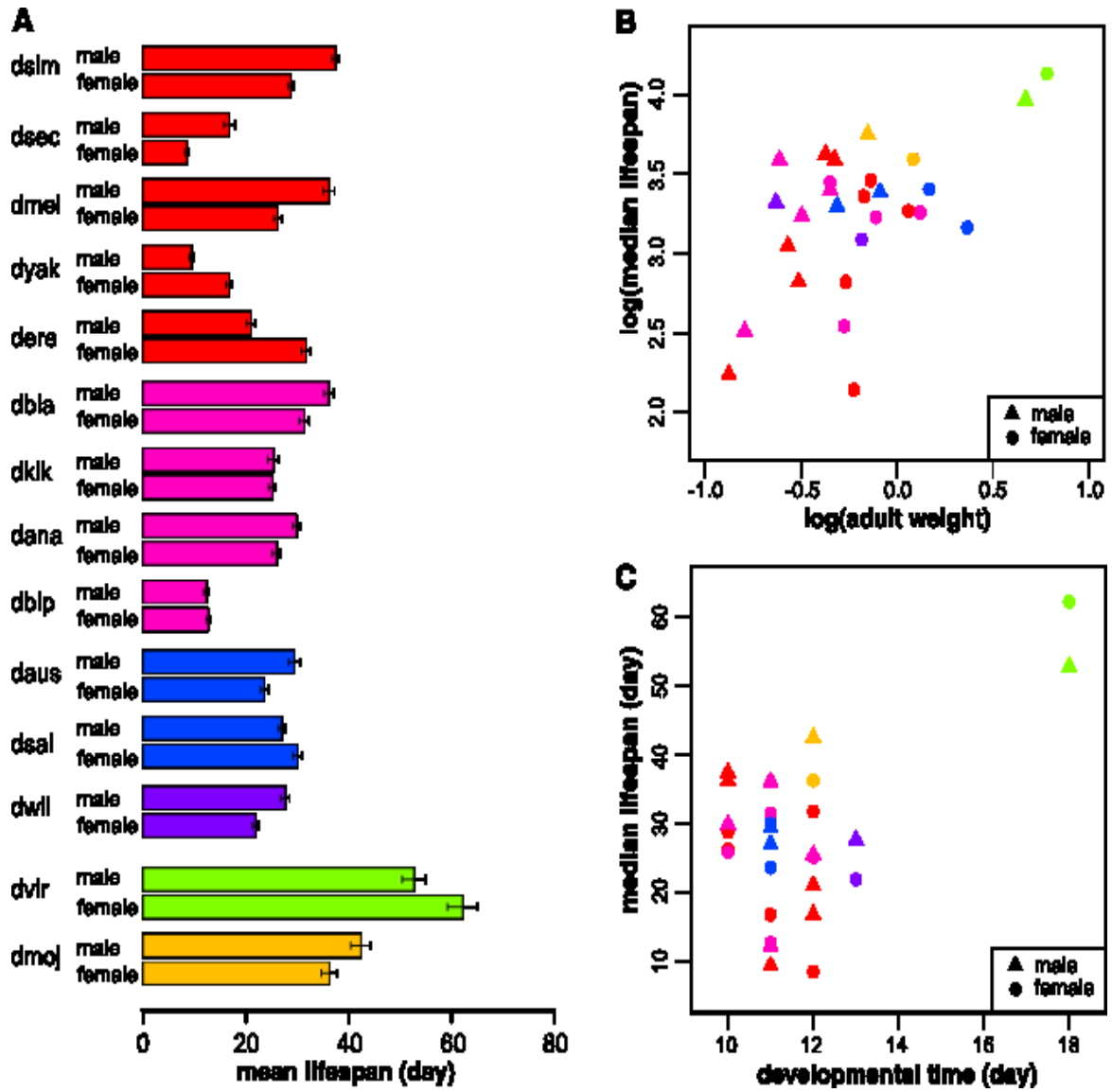



**Figure 6.2. Life history traits of examined species.**

**(A) Mean lifespan.** Error bars indicate 95% confidence intervals (C.I.) by Kaplan-Meier method.

**(B) Adult weight and (C) developmental time correlate positively with median lifespan.** Each point represents one species.

## Gene expression variation reflects evolutionary relationships

To examine the expression variation underlying the differences in life history traits, we subjected young adult male flies of each species to whole body RNA sequencing. After normalization across species and filtering out genes with low expression, log-RPKM values were calculated for 6510 gene ortholog sets to represent the expression levels. Overall, the expression profiles were similar to one another, with Spearman correlation coefficients of species pairs ranging between 0.68 and 0.90.

To determine whether the evolutionary relationship of these species could be recapitulated by their expression variation, gene expression phylograms were constructed using a distance matrix of 1 minus Spearman correlation coefficients and the neighbor-joining method (Brawand et al., 2011; Clark et al., 2007). The resulting topology was largely consistent with their phylogeny (Figure 6.3A). For example, there was a clear separation between subgenera *Drosophila* and *Sophophora*; all 9 species of the group *melanogaster* fell under a single clade; and the 2 species of the *saltans* group also clustered with each other. Most of the nodes received strong support in bootstrapping, indicating the segregation pattern was evident in many genes. However, *D. biarmipes* was placed within the subgroup *melanogaster*, which might reflect variation in biological sampling or actual deviation from phylogeny.

Principal Component Analysis also showed that the species clustered according to their lineages (Figure 6.3B), with the first three principal components (PCs) accounting for 45% of the total variance. To identify the biological processes that diverge significantly across these species, the genes were ranked by their contributions to each PC and pathway enrichment analysis was performed on the top 5% candidates in the first three PCs using DAVID (Huang da et al., 2009a, b). The enriched KEGG and GO terms in the PC1 included “oxidative phosphorylation” (KEGG), “respiratory electron transport chain” (Biological Processes), “NADH dehydrogenase activity” (Molecular Function), “ribosomal subunits” (Cellular Compartment), “structural constituent of ribosome” (MF), and “microtubule-based process” (BP). In PC2, we observed enrichment of “regulation of transcription” (BP), “positive regulation of biosynthetic process” (BP), “transcription factor complex” (CC), and

“ovarian follicle cell development” (BP). In PC3, “vitamin binding” (MF), “glycine, serine and threonine metabolism” (KEGG), “cellular amino acid biosynthetic process” (BP) and “oligopeptide transport” (BP) were enriched. The results suggest that the physiological, morphological, and life history diversities among these species may be attributed to their differences in transcription, translation, mitochondrial functions, and metabolic processes.

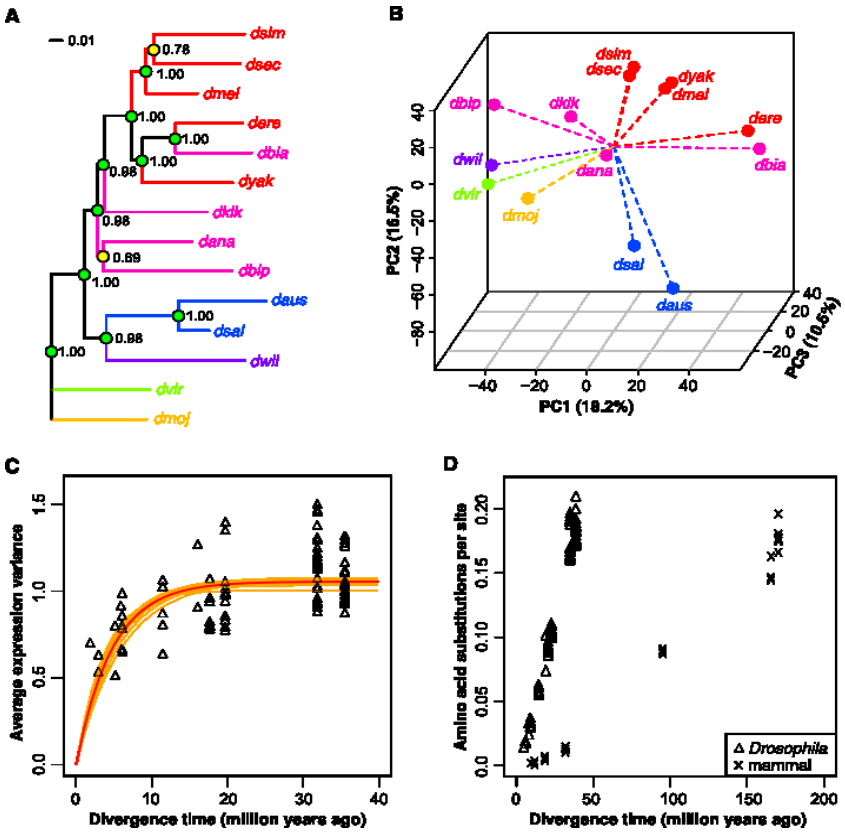



**Figure 6.3. Gene expression reflects evolutionary relationships.**

**(A) Phylogram constructed based on gene expression.** *D. virilis* was used as out-group. Reliability of the branching pattern was assessed by 1000-time bootstrap across the genes (bootstrap values next to the nodes; green: ≥0.9; yellow: 0.6-0.9).

**(B) Principal Component Analysis.** Proportion of variance explained by each principal component (PC) is indicated in parentheses.

**(C) Gene expression divergence reaches a plateau.** Each triangle represents a pair of species. The red curve represents the best-fit line based on the model previously described (Bedford and Hartl, 2009), with the following parameters: selection parameter $\alpha$=0.11 (95% C.I.: 0.10–0.11); drift parameter $\sigma^2$=0.22 (95% C.I.: 0.21–0.24) (Experimental Procedures). Orange curves represent the best-fit lines when each individual species was removed, one at a time ($\alpha$ ranged between 0.08 and 0.14).

**(D) Amino acid substitutions per site increase faster in *Drosophila* than in mammals.** Amino acid substitutions per site between species pairs were calculated based on concatenated, gap-free alignment of orthologs.

## Expression variation is best described by stabilizing selection

Since both life history traits and gene expression patterns are shaped by forces of evolution, we next examined the evolutionary models underlying expression variation across these species. In the absence of selective pressure, variation between a pair of species is expected to increase linearly with divergence time and can be modeled by a Brownian Motion (BM) process (Felsenstein, 1985). This has been observed for transcription of many genes in mammals (especially among primates), supporting a neutral model of evolution (Brawand et al., 2011; Khaitovich et al., 2004). However, while the genetic variation also increases with divergence time in *Drosophila*, previous studies based on genomes and transcriptomes observed that a large fraction of the genes in fruit flies were likely subjected to stabilizing selection (Bedford and Hartl, 2009; Clark et al., 2007; Kalinka et al., 2010; Rifkin et al., 2003). In particular, the increase in gene expression divergence between species pairs eventually reaches a plateau (Bedford and Hartl, 2009), which may be better described by Orstein-Uhlenbeck (OU) process (Butler and King, 2004; Martins and Hansen, 1997).

In agreement, we observed a non-linear relationship with a plateau phase when plotting the average expression variances of the *Drosophila* species pairs against their divergence time (Figure 6.3C). Fitting the data with a previously published equation (Bedford and Hartl, 2009) (Experimental Procedures), we confirmed the selection parameter α was significantly greater than 0 (α=0.11; 95% confidence interval: 0.10–0.11), and the observed relationship did not depend on any particular species (α ranged between 0.08 and 0.14 when each species was removed, one at a time). Data simulation also suggested that this trajectory resembled the pattern produced under an OU model more than that produced under a BM model, or it could be a mixture of both models with greater contribution from OU (Figure 6.4A). For comparison, a similar analysis was performed using the brain and liver data of 9 mammalian species (Brawand et al., 2011), but the plateau feature was not as strong (Figures 6.4B and 6.4C). Importantly, in *Drosophila* the average expression variance became saturated for species pairs that diverged more than 30 million years ago, yet in mammals the saturation was not evident among species that diverged within the last 100 million years. Indeed, when amino acid substitutions per site between species pairs were plotted against divergence time,

*Drosophila* produced a much steeper slope than the mammals (Figure 6.3D). This likely reflects the notion that the evolutionary divergence covered by the genus *Drosophila* equals or exceeds that of the entire mammalian radiation, probably due to the short generation time of fruit flies (Clark et al., 2007; Stark et al., 2007).

To examine the evolutionary models at the individual gene level, phylogenetic signals were measured using two metrics, Pagel's λ (Pagel, 1999) and Blomberg's K (Blomberg et al., 2003). These metrics are usually high for genes that follow BM model, but can be weakened by processes such as stabilizing selection. We found that the phylogenetic signals were low for many genes (Figure 6.4D; median values: $\lambda$=0.03, K=0.41), suggesting most of the variations observed were not fully accounted for by the BM model. In addition, when we compared the goodness of fit of individual gene expression under BM model against OU models with up to three optima (Butler and King, 2004; Kalinka et al., 2010), we found over 85% of the genes fitted better with one of the OU models than with the BM model (Figures 6.4E and 6.4F), similar to the percentage previously observed (Kalinka et al., 2010). Together, these data suggest that stabilizing selection likely plays an important role in influencing the gene expression patterns in *Drosophila* and may also affect the evolution of life history traits.

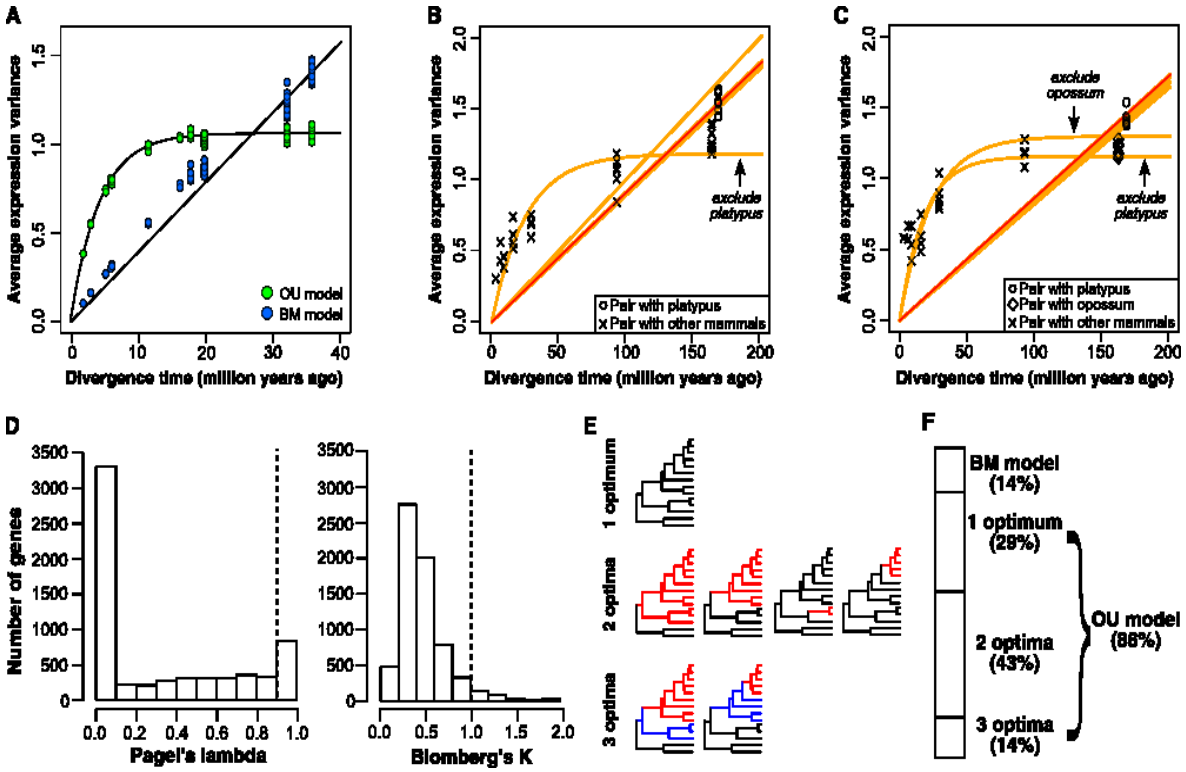



**Figure 6.4. Expression divergence and evolutionary model.**

**(A) Simulation under Orstein-Uhlenbeck (OU) and Brownian Motion (BM) models.** Data were simulated for 6000 genes using the phylogenetic tree in Figure 6.1, as well as the α (for OU only) and $\sigma^2$ (for both BM and OU) values estimated in Figure 6.3C.

**Average expression variance in (B) brain and (C) liver of 9 mammalian species.** The species are human, gorilla, bonobo, chimpanzee, orangutan, macaque, mouse, opossum, and platypus. Estimated parameters: (B) $\alpha<10^{-5}$ (95% C.I.: $10^{-5}$–$10^{-5}$), $\sigma^2=0.0092$ (95% C.I.: 0.0091–0.0092); (C) $\alpha<10^{-5}$ (95% C.I.: $10^{-5}$–$10^{-5}$), $\sigma^2=0.0087$ (95% C.I.: 0.0086–0.0087). Red lines indicate the best-fit lines using all the species. Orange lines indicate the best-fit lines when one of the species is excluded.

**(D) Pagel's λ and Blomberg's K were small in many genes.** Right side of the dotted lines indicates the genes with strong phylogenetic signals (i.e. Pagel's $\lambda > 0.9$ or Blomberg's $K > 1$).

**(E) Gene expression fitted by OU models with up to 3 optima.** The phylogenetic tree is the same as Figure 1, with the tip labels omitted.

**(F) Percentage of genes best fitted by each of the models.** The goodness of fit of the models was determined by likelihood ratio test.

## Gene expression and longevity

To identify the genes that correlate with species longevity, the phylogenetic generalized least squares approach was employed to adjust for the evolutionary relationship (Felsenstein, 1985; Freckleton et al., 2002; Grafen, 1989; Martins and Garland, 1991) (Chapter 2). Regression was performed between expression values and male median lifespan ("ML"), or male median lifespan with adult weight as covariant ("ML-aw"). Inclusion of adult weight in the equation accounted for the potential influence of body mass. Different models of trait evolution were tested and the best-fit model was then selected based on maximal likelihood (Experimental Procedures). Given that *D. virilis* was much larger in size and longer-lived than the other species, we also performed regression after excluding the *D. virilis* data. In addition, we examined the data after introducing 10% variation to the median lifespan values to account for possible inaccuracy in measurements (Experimental Procedures).

For regression with ML, we found 383 out of the 6510 genes with significant regression slope (p value < 0.05), among which 195 were in positive correlation and 188 were in negative correlation. 85% of them were unaffected by the introduction of variation in lifespan measurement, and 73% still remained significant when *D. virilis* was excluded. When the effect of body mass was accounted for (ML-aw), we obtained a shorter list of 172 top genes, with 80 in positive and 92 in negative correlation. Over 77% and 82% of these genes remained statistically significant after introducing variation and excluding *D. virilis*, respectively (Table 6.2).

To understand the biological pathways represented by these genes, enrichment analysis was performed using DAVID separately for those with positive and negative correlation, after adjusting for the full list of orthologs as background. Among those with positive correlation with ML, the top annotation clusters consisted of GO terms related to lipid synthesis and metabolism, including "organic acid biosynthetic process" (BP), "fatty acid metabolic process" (BP), "fatty acid beta-oxidation" (BP), and "lipid particle" (CC). Another group of genes were related to cofactors, such as "cofactor metabolic process" (BP), "cofactor biosynthetic process" (BP), "iron ion binding" (MF), "tetrapyrrole binding" (MF), and "heme binding" (MF). KEGG pathways "glyoxylate and

dicarboxylate metabolism" and "limonene and pinene degradation" were also enriched. When the influence of body mass was removed, there were fewer top genes and the enriched terms included "aspartate oxidase activity" (MF) and "disaccharide biosynthetic process" (BP) (Table 6.2). On the other hand, for genes with negative correlation to ML, significant enrichment was observed for several developmental processes, especially those related to the nervous system. The top terms included "neuron recognition" (BP), "synaptic transmission" (BP), "cell projection morphogenesis" (BP), "salivary gland development" (BP), and "visual behaviour" (BP). In addition, gene expression and post-translational modification of proteins were implicated, such as "protein-DNA complex" (CC), "DNA binding" (MF), "protein amino acid phosphorylation" (BP), and "protein kinase activity (MF). After accounting for body mass, enrichment was still observed for the developmental and behavioural processes, including "neuron recognition" (BP), "cell recognition" (BP), "axongenesis" (BP), "regulation of synapse structure and activity" (BP), and "response to light stimulus" (BP). Additionally, a number of the top genes were involved in "RNA processing" (BP), "nucleotide binding" (MF), "ATP binding" (MF), and "cellular macromolecular complex assembly" (BP), suggesting the potential regulation of alternative splicing (Table 6.2).

We visualized protein-protein interactions among top hits using STRING (Jensen et al., 2009) and found the network significantly enriched in interactions (p value $=3.9\times10^{-13}$; Figure 6.5). Among the genes positively correlating with lifespan, those found in lipid metabolic processes or lipid particles clustered together (Figure 6.5). For example, *Thiolase*, *Acox57D-d* (acyl-CoenzymeA oxidase at 57D distal), and *CG6543* (enoyl-CoA hydratase) are involved in fatty acid beta-oxidation, and two of them (*Thiolase* and *CG6543*) code for mitochondrial matrix proteins (St Pierre et al., 2014). *Taz* and *CG4585* are also involved in the metabolism of phospholipids. Several studies in *Drosophila* reported links between fatty acid oxidation and longevity control: flies overexpressing genes involved in beta-oxidation were longer-lived and more resistant to oxidative stress induced by paraquat treatment (Lee et al., 2012), while knockout of *Thiolase* significantly shortened lifespan (Kishita et al., 2012). Another smaller cluster consisted of genes implicated in cofactor binding and metabolism of small organic molecules, including *Dhfr* (dihydrofolate reductase), *CG18003* (glycolate oxidase), *Upgo* (uroporphyrinogen decarboxylase), *CG5854* (contains an NAD(P)-binding

domain), and *CG31673* (glyoxylate reductase activity). In addition, some of the top genes were previously found to influence longevity in flies. For example, cyclin-dependent kinase Cdk5 requires an activating subunit (p35) for its full biological function, and flies with p35 mutation had significantly shortened lifespan and age-dependent loss of motor function (Connell-Crowley et al., 2007). Another gene was *Gclm*, which codes for the modulatory subunit of glutamate-cysteine ligase (GCL), the rate-limiting enzyme in *de novo* glutathione biosynthesis. Global overexpression of GCLm in flies extended the mean lifespan by 24%, and neuronal overexpression of the catalytic subunit (GCLc) increased mean and maximum lifespans by up to 50% (Orr et al., 2005).

Many genes negatively correlating with lifespan were involved in nervous system development (Figure 6.5). Among them were several cell surface receptors and signaling molecules, such as *Fas2*, which codes for a cell adhesion protein Fasciclin 2 that interacts with semaphorine (Smad) and connectin to regulate axon fasciculation (Yu et al., 2000); *babo*, which codes for a type I activin receptor and regulates cell proliferation by stimulating Smad2-dependent pathways (Brummel et al., 1999); and *shark*, which codes for SH2 ankyrin repeat tyrosine-protein kinase and is required for dorsal closure during development (Fernandez et al., 2000). Also present were component of gap junction (*ogre*) and TGF-beta receptor (*tkv*). Down-regulation of activin signaling by forkhead transcription factor (FOXO) in muscle tissues of flies has been shown to improve muscle performance, reduce secretion of insulin peptides from brain, and extend lifespan (Bai et al., 2013). Flies with *babo* knockdown in muscle lived about 20% longer than wild type, and pathway analysis of FOXO gene targets revealed enrichment of processes involved in post-embryonic development, neuron differentiation, axongenesis, and regulation of transcription and growth (Bai et al., 2013), similar to the terms we observed here (Table 6.2). The low expression levels of development genes (especially those related to neuronal control) among the longer-lived species were consistent with their slower growth rates (Figure 6.2C). A study in mammals also revealed a significant negative correlation between species maturity time and amino acid levels in brain, whereas the relationship was not as strong in non-neuronal organs (Ma et al.). Another cluster included several genes affecting RNA polyadenylation (cleavage and polyadenylation specificity factor *Cpsf160* and *Cpsf73*) and RNA splicing (e.g. *CG6841*, *CG10333*, *CG6686*, *Ars2*, *CG7564*, and *l(2)35Df*) (St Pierre et al., 2014).

Many of them were also involved in the nervous system development, as alternative splicing has very important roles in modulating neuronal maturation and functions (Li et al., 2007).

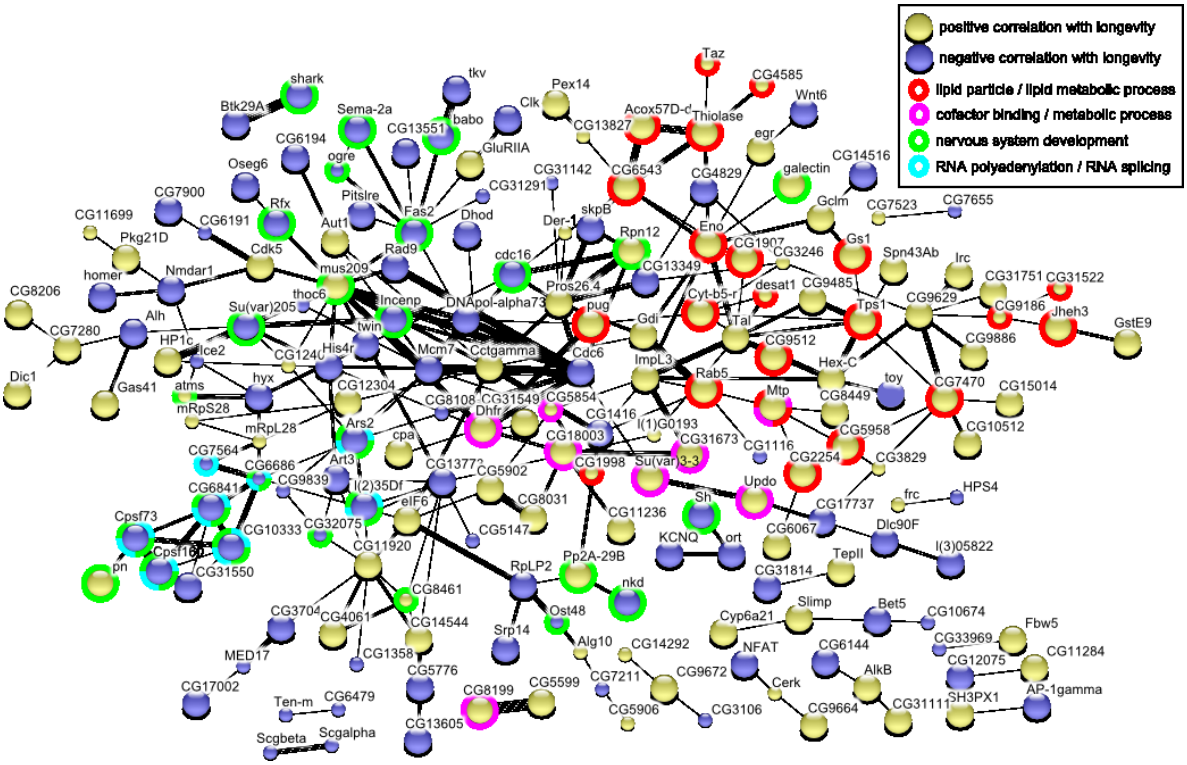



**Figure 6.5. Protein-protein interaction network among the top hits correlating with median lifespan.** The interaction network is based on STRING database (evidence view). Genes without interacting partners are omitted.

**Table 6.2. (see attached Excel file) Genes with significant correlation to median lifespan (ML) and median lifespan with adult weight as covariant (ML-aw).** "Best.Model" indicates the best-fit regression model; "Benjamini" indicates the p value adjustment for multiple testing (Benjamini and Hochberg, 1995); "p.val (excl D.virilis)" indicates the p values excluding the *D. virilis* data; "p.val (Variation 25 percentile)" and "p.val (Variation 75 percentile)" indicate the 25$^{th}$ and 75$^{th}$ percentile p values after adding ±10% variation to the median lifespan data. Enrichment analysis was performed using DAVID, separately for genes with positive and negative correlations.

## Changes in gene expression drive changes in species lifespan

Natural changes in gene expression observed across *Drosophila* may be related to both the direction of changes in these genes in response to changes in diet and environment, and the consequences of genetic manipulations in these genes in laboratory setting. To examine these possibilities, we searched GenAge and GenDR databases (de Magalhães et al., 2009; Plank et al., 2012) to test if any of our longevity signature genes had been experimentally confirmed for association with aging or dietary restriction in yeast, flies, worms, and mice. Specifically, we checked whether the direction of correlation was in agreement with the longevity effects revealed by the experiments, i.e. genes showing positive correlation in our list should have pro-longevity and those showing negative correlation should be anti-longevity effects. Among the ML gene list, 18 genes were found in the database and 14 of them were in the expected direction of correlation; whereas for ML-aw, 4 out of the 6 genes had the expected direction (Table 6.3). Among these genes, all correlations in flies were in the correct direction; inconsistencies primarily involved the data in worms. However, even with these cross-species differences, the overall result was statistically significant (binomial p value = 0.011).

We further compared our top genes against published microarray experiments in flies, in which changes in longevity were induced by dietary or genetic interventions (Table 6.4). Treatments included dietary restriction in different strains of *D. melanogaster*, overexpression of *Sir2*, mutation in *OvoD1* (to repress egg maturation in females), knockdown of tumor suppressor *p53*, and ablation of corpora allata (the endocrine gland that generates juvenile hormone). Again, we checked whether the direction of correlation of our top genes was consistent with the up- or down-regulation observed in these studies. The data indicated that the top genes in ML and ML-aw both showed significant similarity to the lifespan-modification treatments (Table 6.4). Simulation confirmed that these results could not have arisen by chance. The data suggest that there are commonalities between the mechanisms that nature employs to vary lifespan across species, and those that have been experimentally verified to influence longevity in model organisms.

**Table 6.3. Top hits that were previously reported to influence lifespan in model organisms.** The data were compiled primarily based on GenAge and GenDR databases, and were further supplemented with literature searches. The "matching direction" column indicates whether longevity effects are in the same or opposite direction of correlation in our top list.

| Trait | Fly Gene ID | Lifespan Corr. | Model Organism | Symbol | Longevity effect | Matching direction | Note | Reference |
|---|---|---|---|---|---|---|---|---|
| ML | FBgn0086687 | Positive | Worm | fat-7 | Pro-longevity | Same | RNAi decreased lifespan | (Van Gilst et al., 2005) |
| | FBgn0037020 | Positive | Yeast | PEX14 | Necessary for Fitness | Same | Deletion decreased lifespan | (Matecic et al., 2010) |
| | FBgn0025352 | Positive | Fly | Thiolase | Pro-longevity | Same | Knockout decreased lifespan | (Kishita et al., 2012) |
| | FBgn0014010 | Positive | Worm | rab-5 | Pro-longevity | Same | RNAi decreased lifespan | (Tacutu et al., 2012) |
| | FBgn0039636 | Positive | Yeast | ATG14 | Necessary for Fitness | Same | Deletion decreased lifespan | (Matecic et al., 2010) |
| | FBgn0000442 | Positive | Worm | pkg-1 | Anti-longevity | Opposite | Mutation increased lifespan | (Hirose et al., 2003) |
| | FBgn0013762 | Positive | Fly | Cdk5 | Pro-longevity | Same | Loss of function decreased lifespan | (Connell-Crowley et al., 2007) |
| | FBgn0036813 | Positive | Worm | atg-3 | Pro-longevity | Same | RNAi decreased lifespan | (Hashimoto et al., 2009) |
| | FBgn0046114 | Positive | Fly | Gclm | Pro-longevity | Same | Overexpression increased lifespan | (Orr et al., 2005) |
| | FBgn0004868 | Positive | Worm | gdi-1 | Pro-longevity | Same | RNAi decreased lifespan | (Samuelson et al., 2007) |
| | FBgn0011205 | Positive | Worm | pnk-1 | Pro-longevity | Same | RNAi decreased lifespan | (Lee et al., 2003) |
| | FBgn0038325 | Negative | Worm | atg-4.2 | Pro-longevity | Opposite | RNAi decreased lifespan | (Hashimoto et al., 2009) |
| | FBgn0023076 | Negative | Mouse | Clock | Pro-longevity | Opposite | Deletion decreased lifespan | (Dubrovsky et al., 2010) |
| | FBgn0011725 | Negative | Worm | ccr-4 | Pro-longevity | Opposite | RNAi decreased lifespan | (Samuelson et al., 2007) |
| | FBgn0003380 | Negative | Mouse | Kcna3 | Anti-longevity | Same | Deletion increased lifespan | (Tucker et al., 2008) |
| | FBgn0011300 | Negative | Fly | babo | Anti-longevity | Same | RNAi increased lifespan | (Bai et al., 2013) |
| | FBgn0020379 | Negative | Yeast | RFX1 | Anti-longevity | Same | Deletion increased lifespan | (Smith et al., 2008) |
| | FBgn0035918 | Negative | Yeast | CDC6 | Anti-longevity | Same | Knockout increased lifespan | (Sinclair and Guarente, 1997) |
| ML-aw | FBgn0030605 | Positive | Worm | D2030.4 | Anti-longevity | Opposite | RNAi increased lifespan | (Hamilton et al., 2005) |
| | FBgn0025352 | Positive | Fly | Thiolase | Pro-longevity | Same | Knockout decreased lifespan | (Kishita et al., 2012) |
| | FBgn0003380 | Negative | Mouse | Kcna3 | Anti-longevity | Same | Deletion increased lifespan | (Tucker et al., 2008) |
| | FBgn0011300 | Negative | Fly | babo | Anti-longevity | Same | RNAi increased lifespan | (Bai et al., 2013) |
| | FBgn0020379 | Negative | Yeast | RFX1 | Anti-longevity | Same | Deletion increased lifespan | (Smith et al., 2008) |
| | FBgn0010414 | Negative | Worm | Mod-5 | Pro-longevity | Opposite | Knockout decreased lifespan | (Murakami et al., 2012) |

**Table 6.4. Comparison between the top hits and previously published microarray experiments.** The data were obtained from Gene Expression Omnibus (GEO) database. Microarray experiments were carried out in *D. melanogaster* (female) with lifespan extension effects. Differentially expressed genes in each microarray experiment were compared to our list of top hits to determine if they were in the same or opposite direction of correlation. P values are based on binomial distribution (p value < 0.05 are in bold)

| Microarray (GEO accession) | Tissue | Strain | Treatment (versus wild-type) | Correlate with ML | | | Correlate with ML-aw | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Same direction | Opposite direction | P value | Same direction | Opposite direction | P value |
| GSE37537 | Whole body | -- | dietary restriction | 94 | 106 | $8.21\times10^{-1}$ | 57 | 38 | **$3.21\times10^{-2}$** |
| GSE26724 | Head/thorax | Canton-S | dietary restriction | 61 | 33 | **$2.54\times10^{-3}$** | 21 | 13 | $1.15\times10^{-1}$ |
| GSE48145 | Whole body | -- | ovoD mutant (sterile) | 145 | 85 | **$4.60\times10^{-5}$** | 79 | 28 | **$4.21\times10^{-7}$** |
| GSE48145 | Whole body | -- | Corpora allata knockout (CAKO) | 45 | 37 | $2.20\times10^{-1}$ | 29 | 8 | **$3.76\times10^{-4}$** |
| GSE48145 | Whole body | -- | ovoD mutant with CAKO | 146 | 88 | **$9.07\times10^{-5}$** | 74 | 31 | **$1.64\times10^{-5}$** |
| GSE26726 | Whole body | yw, w1118 | dietary restriction (10 days) | 112 | 75 | **$4.15\times10^{-3}$** | 61 | 34 | **$3.66\times10^{-3}$** |
| GSE26726 | Whole body | yw, w1118 | dietary restriction (40 days) | 48 | 29 | **$1.98\times10^{-2}$** | 23 | 14 | $9.39\times10^{-2}$ |
| GSE26726 | Whole body | Canton-S | dietary restriction (10 days) | 109 | 79 | **$1.71\times10^{-2}$** | 61 | 31 | **$1.16\times10^{-3}$** |
| GSE26726 | Whole body | Canton-S | dietary restriction (40 days) | 86 | 62 | **$2.92\times10^{-2}$** | 48 | 23 | **$2.03\times10^{-3}$** |
| GSE26726 | Whole body | yw, w1118 | p53 knockdown (10 days) | 43 | 20 | **$2.58\times10^{-3}$** | 21 | 10 | **$3.54\times10^{-2}$** |
| GSE26726 | Whole body | yw, w1118 | sir2 overexpression (10 days) | 91 | 40 | **$4.90\times10^{-6}$** | 46 | 15 | **$4.42\times10^{-5}$** |

# DISCUSSION

Fruit flies have contributed significantly to our understanding of genetics and developmental biology and remain a vital tool in the studies on aging and longevity. While most experiments in the aging field have been conducted using single species, we hypothesized that the comparative analysis of closely related species with varying natural lifespan may offer unbiased information on longevity mechanisms. By studying life history traits and transcriptomes of 14 *Drosophila* species, we identified the pathways that diverge across these species and confirmed the role of stabilizing selection in influencing their expression patterns. Similar to mammals and birds, fruit flies exhibit the typical positive correlation among longevity, body mass, and development time. By identifying the genes that correlate with longevity across these species, we found that longer-lived species up-regulate genes involved in lipid metabolism and down-regulate those involved in neuronal system development and activin signaling. The dynamics of RNA polyadenylation and splicing also differed across the species. Interestingly, some genes that showed significant association with longevity in our study were also found to influence lifespan in other model organisms and showed the same direction of change as lifespan extension experiments in flies. The results suggest that the natural variation of lifespan across closely related species under the forces of evolution may have been achieved via pathways that are also influenced by dietary restriction or other interventions. The data suggest that our approach offers an unbiased way to uncover genes and processes whose changes lead to changes in lifespan. More generally, the findings suggest the molecular mechanisms by which nature alters species longevity. Finally, our data may serve as a starting point for experimental analysis of genes, processes, diets and pharmacologic interventions that mimic natural changes in lifespan, thereby exhibiting longevity modification effects.

## EXPERIMENTAL PROCEDURES

### Fly stocks and husbandry

14 *Drosophila* species, *D. ananassae*, *D. austrosaltans*, *D. biarmipes*, *D. bipectinata*, *D. erecta*, *D. kikkawai*, *D. melanogaster*, *D. mojavensis*, *D. saltans*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. willistoni*, and *D. yakuba*, were purchased from UC San Diego Stock Center (La Jolla, CA, USA) (Table 6.1). Flies were maintained on corn meal food (85.7 g corn meal 'Aunt Jemina' (The Quaker Oats Company, Chicago, IL, USA), 50 ml golden A unsulfured molasses (Groeb Farms Inc, Onsted, MI, USA), 71.4 g Torula yeast (MP Biomedicals, Solon, OH, USA), 2.86 g p-hydroxybenzoic acid methyl ester (Sigma), 6.4 g agar (MoorAgar Inc, Loomis, CA, USA) and 5.7 ml propionic acid (Sigma) per litre water) and kept in a temperature-controlled incubator at 25 °C with 12-h light/dark cycle and ~60% humidity. Newly emerged flies were collected within 18 h at 18 °C, transferred to fresh corn meal food at density of 35 animals per vial, and allowed to mate for 1–2 days. Three-day-old mated flies were collected using $CO_2$, sorted by sex and then transferred to cages in the temperature-controlled incubator. Experimental flies were held on the designed diet and transferred to fresh vials without anaesthesia every 3 days. Dead flies were removed by aspiration and counted. Survival analyses were performed using R package "survival" (Kaplan and Meier, 1958; Therneau, 2014).

### RNA sequencing

Three-day-old male flies were placed in vials on the corn meal diet for 12 days, with three replica vials for each species. Fresh food was supplied every 3 days and dead flies were removed by aspiration. After 12 days, those flies were subject to total RNA extraction. RNA sequencing libraries were constructed using the Illumina mRNA-Seq Prep Kit and oligo(dT) magnetic beads were used to purify polyA containing mRNA molecules. mRNA was further fragmented and randomly primed during the first strand synthesis by reverse transcription and followed by second-strand synthesis with DNA polymerase I to create double-stranded cDNA fragments. Double stranded cDNA was subjected to end repair by Klenow and T4 DNA polymerases and A-tailed by Klenow lacking exonuclease activity. Ligation to Illumina Paired-End Sequencing adapters, size selection by gel electrophoresis

and then PCR amplification completed library preparation. The 200 bp paired-end libraries were sequenced using Illumina HiSeq according to the manufacturer's protocol.

**Ortholog set identification**

Ortholog sets across the species were identified by reciprocal best hits in BLAST. Briefly, we downloaded the genomes and annotation files for the species from Ensembl and NCBI and extracted all coding sequences ("Species CDS"). The genomes of *D. austrosaltans* and *D. saltans* were based on our unpublished data. As reference we extracted the longest open reading frame for each gene in *D. melanogaster* ("Dmel ORF"), after excluding those with multiple paralogs (i.e. genes with over 80% identity over 70% of length) or highly repetitive sequences. Mega BLAST (Morgulis et al., 2008) was performed to obtain reciprocal best hits between Dmel ORF and Species CDS. An ortholog set was declared if the orthologs were present in all 14 species. We confirmed our list had over 90% overlap with the curated ortholog list on Flybase (St Pierre et al., 2014) (which covered 9 of our species). To improve the quality of ortholog sequences, Trinity (Grabherr et al., 2011) was also used to *de novo* assemble transcriptomes from RNA-seq data. Poorly annotated Species CDS (e.g. those without proper start or stop codons) were replaced by Trinity transcripts where applicable and we ensured that the final list of orthologs contained at least 20% conserved blocks in multiple sequence alignment. The final ortholog sequences were mapped back to their respective genomes with GMAP(Wu and Watanabe, 2005) to generate customized GFF (General Feature Format) files.

**Data processing**

RNA-seq reads alignment was performed using TopHat (Trapnell et al., 2009) and read counting was performed using featureCounts (Liao et al., 2014). Ortholog sets with low expression (i.e. less than 3 counts in 3 or more species) were removed and counts were normalised by total library sizes with Trimmed Mean of M-values correction. The final list consisted of 6510 ortholog sets. The normalised counts were then converted to reads per kilobase per million mapped reads (PRKM) values and natural log transformed. For cross-species comparison, log-RPKM values of each ortholog set were standardised by setting mean as 0 and standard deviation as 1. Q-Q plot and Shapiro-Wilk test confirmed that normalcy was a valid assumption for 88% of the ortholog sets on log-scale.

**Divergence time and phylogram**

*Drosophila* species divergence time was estimated based on previous estimates and ortholog amino acid sequences. Briefly, ortholog sequences were aligned with Clustal Omega v1.2.0 (Sievers et al., 2011) and concatenated gap-free with Gblocks v0.91 (Castresana, 2000). The tree was constructed using neighbour-joining method (Saitou and Nei, 1987) in Mega 6.06 (Tamura et al., 2013) and calibrated using estimates of divergence time in the literature (Russo et al., 1995). The *Drosophila* expression phylogram was based on a distance matrix of 1 minus Spearman correlation coefficient and constructed by neighbour-joining method using R package "ape" (Paradis et al., 2004). Reliability of the branching pattern was assessed by 1000-time bootstrap across the genes. For the mammalian dataset (Brawand et al., 2011), the amino acid sequence alignments of 8 mammalian species (human, gorilla, chimpanzee, orangutan, macaque, mouse, opossum, and platypus; sequences for bonobo not available) were extracted from the 46-way multiple alignment in UCSC genome browser (Kuhn et al., 2013). Species divergence time was based on TimeTree database (Hedges et al., 2006).

**Principal Component Analysis (PCA) and heat maps**

PCA was performed on standardized expression values using R package "stats" (R Development Core Team, 2013). To identify the underlying pathways, the genes in each of the first three principal components (PCs) were ranked by their contributions and pathway enrichment analysis was performed on the top 300 (about 5%) genes in each PC using DAVID (Huang da et al., 2009a, b) after correcting for background. To generate the heat map, the genes (columns) were ordered by contributions and the species (rows) were ordered by projection values.

**Expression divergence**

Expression divergence was measured as average expression variance in standardized expression values across all ortholog sets between species pairs. The points were fitted by the model previously described (Bedford and Hartl, 2009): $y = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha x})$, where $x$ represents divergence time, $y$ represents expression divergence, $\alpha$ represents strength of selection, and $\sigma^2$ represents strength of drift. It can be shown that (in the case of a pure BM model): $\lim_{\alpha \to 0} y = \sigma^2 x$. Optimal values

of parameters were estimated by least squares method. Confidence intervals (C.I.) were estimated by 1000-time bootstrap. To test robustness of the relationship, individual species were removed one at a time and the remaining species were then subjected to the same analysis.

**Regression by generalized least square**

See Chapter 2. The cut-off for top hits was p value < 0.05.

**Test for robustness**

To confirm the robustness of the results, two methods were used. First, regression was performed by leaving out *D. virilis* (the most long-lived species) and computing p values using the remaining species. Second, median lifespan of species was randomly varied by ±10% (on log scale) and then used to compute regression p values. Simulation was performed 1000 times for each gene and the $25^{th}$ and $75^{th}$ percentiles of the p values were determined. This method imitated potential inaccuracy in our lifespan measurements.

**Comparison with GenAge/GenDR database and microarray data**

Our gene list was examined against GenAge/GenDR databases (de Magalhães et al., 2009; Plank et al., 2012) to determine how the longevity effects reported in model organisms relate to our gene dataset. To analyze microarray datasets, data were downloaded from Gene Expression Omnibus (GEO) database. Relevant comparisons of treatment versus control were selected and differentially expressed (DE) genes were identified using R package "limma" (Smyth, 2005). These DE genes were then compared with our list of top hits to determine if the direction of correlation was consistent. Two methods were employed to calculate p values. The first relied on binomial distribution, counting the number of match (i.e. same direction of correlation) and the number of mis-match (i.e. opposite direction of correlation), assuming equal probability of obtaining a match and a mismatch by chance. The second method relied on simulation, where the direction of correlation in our top list was shuffled randomly and compared with microarray experiments to calculate p values (by binomial distribution); this procedure was repeated 1000 times to generate an empirical distribution. The original p value was then compared to the empirical distribution. Both methods produced very similar results.

# REFERENCES


Bai, H., Kang, P., Hernandez, A.M., and Tatar, M. (2013). Activin signaling targeted by insulin/dFOXO regulates aging and muscle proteostasis in Drosophila. PLoS Genet. *9*, e1003941.

Bedford, T., and Hartl, D.L. (2009). Optimization of gene expression by natural selection. Proc. Natl. Acad. Sci. USA *106*, 1133-1138.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B *57*, 289-300.

Bjedov, I., Toivonen, J.M., Kerr, F., Slack, C., Jacobson, J., Foley, A., and Partridge, L. (2010). Mechanisms of life span extension by rapamycin in the fruit fly Drosophila melanogaster. Cell Metab. *11*, 35-46.

Blomberg, S.P., Garland, T., Jr., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution *57*, 717-745.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Brummel, T., Abdollah, S., Haerry, T.E., Shimell, M.J., Merriam, J., Raftery, L., Wrana, J.L., and O'Connor, M.B. (1999). The Drosophila activin receptor baboon signals through dSmad2 and controls cell proliferation but not patterning during larval development. Genes Dev. *13*, 98-111.

Butler, M.A., and King, A.A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am. Nat. *164*, 683-695.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540-552.

Chapman, T., and Partridge, L. (1996). Female fitness in Drosophila melanogaster: an interaction between the effect of nutrition and of encounter rate with males. Proc. Biol. Sci. *263*, 755-759.

Clancy, D.J., Gems, D., Harshman, L.G., Oldham, S., Stocker, H., Hafen, E., Leevers, S.J., and Partridge, L. (2001). Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science *292*, 104-106.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N.*, et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

Connell-Crowley, L., Vo, D., Luke, L., and Giniger, E. (2007). Drosophila lacking the Cdk5 activator, p35, display defective axon guidance, age-dependent behavioral deficits and reduced lifespan. Mech. Dev. *124*, 341-349.

Danilov, A., Shaposhnikov, M., Plyusnina, E., Kogan, V., Fedichev, P., and Moskalev, A. (2013). Selective anticancer agents suppress aging in Drosophila. Oncotarget *4*, 1507-1526.

de Magalhães, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics *25*, 875-881.

Dubrovsky, Y.V., Samsa, W.E., and Kondratov, R.V. (2010). Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. Aging (Albany NY) *2*, 936-944.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. Am. Nat. *125*, 1-15.

Fernandez, R., Takahashi, F., Liu, Z., Steward, R., Stein, D., and Stanley, E.R. (2000). The Drosophila shark tyrosine kinase is required for embryonic dorsal closure. Genes Dev. *14*, 604-614.

Fontana, L., Partridge, L., and Longo, V.D. (2010). Extending healthy life span--from yeast to humans. Science *328*, 321-326.

Freckleton, R.P., Harvey, P.H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. *160*, 712-726.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644-652.

Grafen, A. (1989). The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci. *326*, 119-157.

Grandison, R.C., Piper, M.D., and Partridge, L. (2009). Amino-acid imbalance explains extension of lifespan by dietary restriction in Drosophila. Nature *462*, 1061-1064.

Hamilton, B., Dong, Y., Shindo, M., Liu, W., Odell, I., Ruvkun, G., and Lee, S.S. (2005). A systematic RNAi screen for longevity genes in C. elegans. Genes Dev. *19*, 1544-1555.

Hashimoto, Y., Ookuma, S., and Nishida, E. (2009). Lifespan extension by suppression of autophagy genes in Caenorhabditis elegans. Genes Cells *14*, 717-726.

He, K., Zhou, T., Shao, J., Ren, X., Zhao, Z., and Liu, D. (2014). Dynamic regulation of genetic pathways and targets during aging in Caenorhabditis elegans. Aging (Albany NY) *6*, 215-230.

Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics *22*, 2971-2972.

Hirose, T., Nakano, Y., Nagamatsu, Y., Misumi, T., Ohta, H., and Ohshima, Y. (2003). Cyclic GMP-dependent protein kinase EGL-4 controls body size and lifespan in C elegans. Development *130*, 1089-1099.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

Hwangbo, D.S., Gershman, B., Tu, M.P., Palmer, M., and Tatar, M. (2004). Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body. Nature *429*, 562-566.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M.*, et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412-416.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. Nature *468*, 811-814.

Kapahi, P., Zid, B.M., Harper, T., Koslover, D., Sapin, V., and Benzer, S. (2004). Regulation of lifespan in Drosophila by modulation of genes in the TOR signaling pathway. Curr. Biol. *14*, 885-890.

Kaplan, E.L., and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. JASA *53*, 457-481.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., and Paabo, S. (2004). A neutral model of transcriptome evolution. PLoS Biol. *2*, E132.

Kishita, Y., Tsuda, M., and Aigaki, T. (2012). Impaired fatty acid oxidation in a Drosophila model of mitochondrial trifunctional protein (MTP) deficiency. Biochem. Biophys. Res. Commun. *419*, 344-349.

Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief. Bioinform. *14*, 144-161.

Lee, B.C., Kaya, A., Ma, S., Kim, G., Gerashchenko, M.V., Yim, S.H., Hu, Z., Harshman, L.G., and Gladyshev, V.N. (2014). Methionine restriction extends lifespan of Drosophila melanogaster under conditions of low amino-acid status. Nat. Commun. *5*, 3592.

Lee, S.H., Lee, S.K., Paik, D., and Min, K.J. (2012). Overexpression of fatty-acid-beta-oxidation-related genes extends the lifespan of Drosophila melanogaster. Oxid. Med. Cell. Longev. *2012*, 854502.

Lee, S.S., Lee, R.Y., Fraser, A.G., Kamath, R.S., Ahringer, J., and Ruvkun, G. (2003). A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. Nat. Genet. *33*, 40-48.

Li, Q., Lee, J.A., and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. Nat. Rev. Neurosci. *8*, 819-831.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Lin, Y.J., Seroude, L., and Benzer, S. (1998). Extended life-span and stress resistance in the Drosophila mutant methuselah. Science *282*, 943-946.

Ma, S., Yim, S.H., Lee, S.G., Kim, E.B., Lee, S.R., Chang, K.T., Buffenstein, R., Lewis, K.N., Park, T.J., Miller, R.A.*, et al.* (2015). Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity. Cell Metab. *22*, 332-343.

Magwere, T., Chapman, T., and Partridge, L. (2004). Sex differences in the effect of dietary restriction on life span and mortality rates in female and male Drosophila melanogaster. J. Gerontol. A Biol. Sci. Med. Sci. *59*, 3-9.

Mair, W., Goymer, P., Pletcher, S.D., and Partridge, L. (2003). Demography of dietary restriction and death in Drosophila. Science *301*, 1731-1733.

Markow, T.A., and O'Grady, P.M. (2005). Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. Annu. Rev. Genet. *39*, 263-291.

Markow, T.A., and O'Grady, P.M. (2007). Drosophila biology in the genomic age. Genetics *177*, 1269-1276.

Martins, E.P., and Garland, T., Jr. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution *45*, 534-557.

Martins, E.P., and Hansen, T.F. (1997). Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. Am. Nat. *149*, 646-667.

Matecic, M., Smith, D.L., Pan, X., Maqani, N., Bekiranov, S., Boeke, J.D., and Smith, J.S. (2010). A microarray-based genetic screen for yeast chronological aging factors. PLoS Genet. *6*, e1000921.

Min, K.J., and Tatar, M. (2006). Restriction of amino acids extends lifespan in Drosophila melanogaster. Mech. Ageing Dev. *127*, 643-646.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., and Schaffer, A.A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics *24*, 1757-1764.

Murakami, C., Delaney, J.R., Chou, A., Carr, D., Schleit, J., Sutphin, G.L., An, E.H., Castanza, A.S., Fletcher, M., Goswami, S.*, et al.* (2012). pH neutralization protects against reduction in replicative lifespan following chronological aging in yeast. Cell Cycle *11*, 3087-3096.

Orr, W.C., Radyuk, S.N., Prabhudesai, L., Toroser, D., Benes, J.J., Luchak, J.M., Mockett, R.J., Rebrin, I., Hubbard, J.G., and Sohal, R.S. (2005). Overexpression of glutamate-cysteine ligase extends life span in Drosophila melanogaster. J. Biol. Chem. *280*, 37331-37338.

Orr, W.C., and Sohal, R.S. (1994). Extension of life-span by overexpression of superoxide dismutase and catalase in Drosophila melanogaster. Science *263*, 1128-1130.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877-884.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289-290.

Parkes, T.L., Elia, A.J., Dickinson, D., Hilliker, A.J., Phillips, J.P., and Boulianne, G.L. (1998). Extension of Drosophila lifespan by overexpression of human SOD1 in motorneurons. Nat. Genet. *19*, 171-174.

Plank, M., Wuttke, D., van Dam, S., Clarke, S.A., and de Magalhaes, J.P. (2012). A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms. Mol. Biosyst. *8*, 1339-1349.

R Development Core Team (2013). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Rifkin, S.A., Kim, J., and White, K.P. (2003). Evolution of gene expression in the Drosophila melanogaster subgroup. Nat. Genet. *33*, 138-144.

Russo, C.A., Takezaki, N., and Nei, M. (1995). Molecular phylogeny and divergence times of drosophilid species. Mol. Biol. Evol. *12*, 391-404.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Samuelson, A.V., Carr, C.E., and Ruvkun, G. (2007). Gene activities that mediate increased life span of C. elegans insulin-like signaling mutants. Genes Dev. *21*, 2976-2994.

Schnebel, E.M., and Grossfield, J. (1983). A comparison of life span characteristics in Drosophila. Exp. Gerontol. *18*, 325-337.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Sinclair, D.A., and Guarente, L. (1997). Extrachromosomal rDNA circles--a cause of aging in yeast. Cell *91*, 1033-1042.

Smith, E.D., Tsuchiya, M., Fox, L.A., Dang, N., Hu, D., Kerr, E.O., Johnston, E.D., Tchao, B.N., Pak, D.N., Welton, K.L.*, et al.* (2008). Quantitative evidence for conserved longevity pathways between divergent eukaryotic species. Genome Res. *18*, 564-570.

Smyth, G.K. (2005). Limma: linear models for microarray data (New York, Springer).

Speakman, J.R. (2005). Body size, energy metabolism and lifespan. J. Exp. Biol. *208*, 1717-1730.

St Pierre, S.E., Ponting, L., Stefancsik, R., and McQuilton, P. (2014). FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res. *42*, D780-788.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N.*, et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219-232.

Sun, J., Folk, D., Bradley, T.J., and Tower, J. (2002). Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult Drosophila melanogaster. Genetics *161*, 661-672.

Tacutu, R., Shore, D.E., Budovsky, A., de Magalhaes, J.P., Ruvkun, G., Fraifeld, V.E., and Curran, S.P. (2012). Prediction of C. elegans longevity genes by human and worm longevity networks. PLoS One *7*, e48282.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. *30*, 2725-2729.

Tatar, M., Kopelman, A., Epstein, D., Tu, M.P., Yin, C.M., and Garofalo, R.S. (2001). A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science *292*, 107-110.

Therneau, T. (2014). A Package for Survival Analysis in S. http://CRAN.R-project.org/package=survival.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Tucker, K., Overton, J.M., and Fadool, D.A. (2008). Kv1.3 gene-targeted deletion alters longevity and reduces adiposity by increasing locomotion and metabolism in melanocortin-4 receptor-null mice. Int. J. Obes. *32*, 1222-1232.

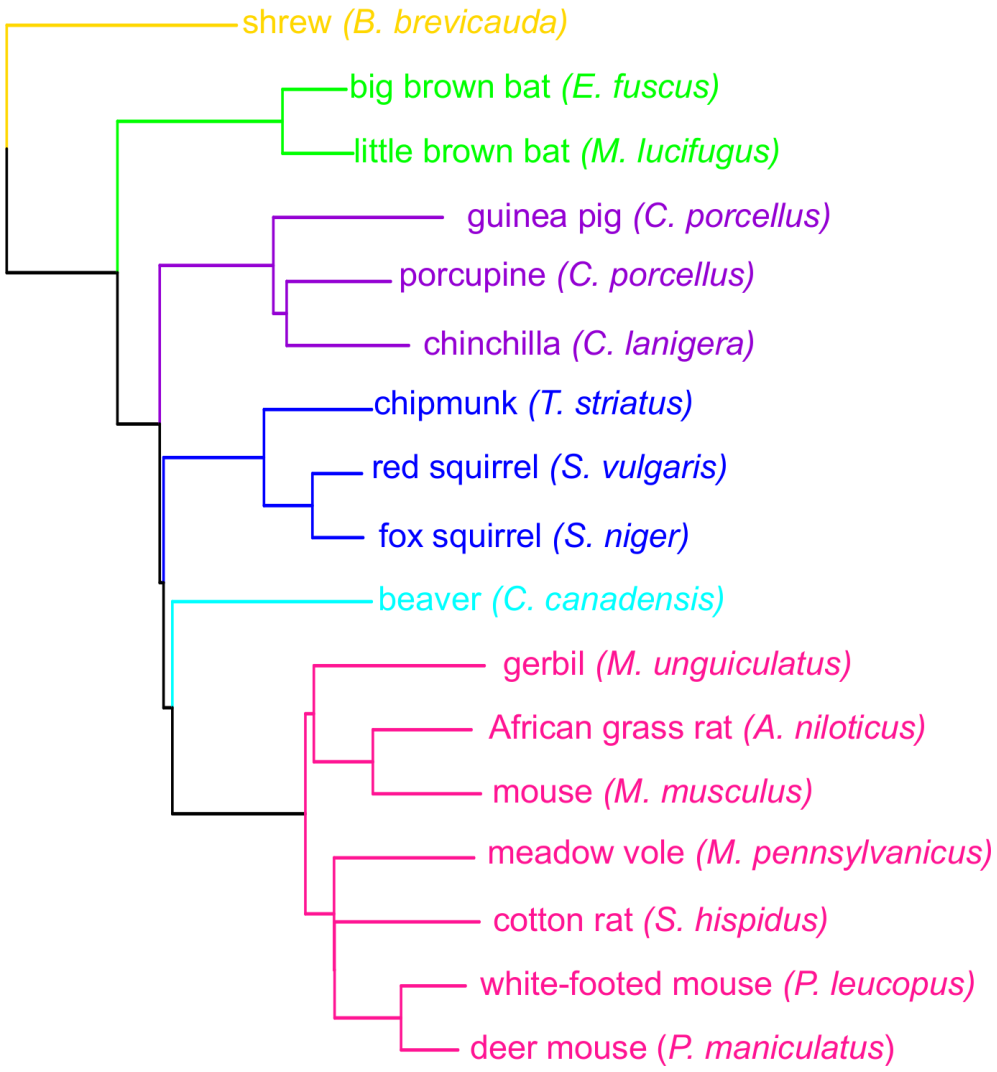
Van Gilst, M.R., Hadjivassiliou, H., Jolly, A., and Yamamoto, K.R. (2005). Nuclear hormone receptor NHR-49 controls fat consumption and fatty acid composition in C. elegans. PLoS Biol. *3*, e53.

Wang, C., Wheeler, C.T., Alberico, T., Sun, X., Seeberger, J., Laslo, M., Spangler, E., Kern, B., de Cabo, R., and Zou, S. (2013). The effect of resveratrol on lifespan depends on both gender and dietary nutrient composition in Drosophila melanogaster. Age (Dordr) *35*, 69-81.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859-1875.

Yu, H.H., Huang, A.S., and Kolodkin, A.L. (2000). Semaphorin-1a acts in concert with the cell adhesion molecules fasciclin II and connectin to regulate axon fasciculation in Drosophila. Genetics *156*, 723-731.

# Chapter 7 Rodent Fibroblasts




This chapter is based on the following manuscript:

Siming Ma, Akhil Upneja, Andrei Seluanov, Vera Gorbunova, Clary B. Clish, Richard A. Miller, Vadim N. Gladyshev. **Cell culture-based profiling across mammals reveals DNA repair and metabolism as determinants of species longevity.** *Manuscript under preparation.*

# ABSTRACT

The natural species lifespan across mammals differs by more than 100-fold, but the molecular signatures associated with such longevity differences are not yet fully understood. Cross-species analyses are also hampered by the inability to conduct these studies under controlled experimental settings. Here, we analyzed primary skin fibroblasts isolated from 16 species of mammals and maintained under identical conditions in cell culture. We further developed a pipeline for obtaining species-specific ortholog sequences, profiled gene expression by RNA sequencing and small molecules by metabolite profiling, and identified genes and metabolites correlating with species longevity. We found that cells from long-lived species up-regulate genes involved in DNA repair and glucose metabolism, down-regulate proteolysis and protein transport, and show high levels of amino acids but low levels of lysophosphatidylcholine and lysophosphatidylethanolamine. The study suggests that fibroblast profiling captures differences in longevity across mammals at the level of global gene expression and metabolism and reveals pathways that may define these differences.

# INTRODUCTION

The maximum lifespan of mammalian species differs by more than 100-fold, ranging from ~2 years in shrews to >200 years in bowhead whales (Tacutu et al., 2013). While it has long been observed that maximum lifespan tends to correlative positively with body mass and time to maturity, but negatively with growth rate, mass-specific metabolic rate, and number of offspring (Peters, 1986; Sacher, 1959; Western, 1979), the underlying molecular basis is only starting to be understood.

One way to study the control of longevity is to identify genes, pathways, and interventions capable of extending lifespan or delaying aging phenotypes in experimental animals. Studies using model organisms have uncovered several important conditions, such as knockout of insulin-like growth factor 1 (IGF-1) receptor (Friedman and Johnson, 1988; Holzenberger et al., 2003; Tatar et al., 2001), inhibition of mechanistic target of rapamycin (mTOR) (Harrison et al., 2009; Kenyon, 2010; Miller et al., 2014), mutation in growth hormone (GH) receptor (Coschigano et al., 2000), ablation of anterior pituitary (e.g. Snell dwarf mice) (Flurkey et al., 2002), augmentation of sirtuin family proteins functions (Chang and Guarente, 2013; Gomes et al., 2013; Mouchiroud et al., 2013; Wood et al., 2004), and restriction of dietary calorie intake (Guarente and Kenyon, 2000; Heilbronn and Ravussin, 2003; McCay et al., 1935; Weindruch et al., 1986). While many of these genes and pathways have been verified in yeast, flies, worms, and mice, the comparisons largely involve treatment and control groups of the same species, and the extent to which they explain the longevity variations across different species is unclear. For example, do the long-lived species have metabolic profiles resembling calorie restriction? Do they suppress IGF-1 or growth hormone signaling compared with the shorter-lived species? More generally, how the evolutionary strategies of longevity relate to the experimental strategies that extend lifespan in model organisms?

To begin address these questions, a popular approach has been to compare exceptionally long-lived species with closely related species characterized by more common lifespan, identifying features associated with exceptional longevity. Examples include amino acid changes in Uncoupling Protein 1 (UCP1) and production of high-molecular-mass hyaluronan in the naked mole rat (Kim et al., 2011; Tian et al., 2013); unique sequence changes in IGF1 and GH receptors in Brandt's bat (Seim

et al., 2013); gene gain and loss associated with DNA repair, cell-cycle regulation, and cancer, as well as alteration in insulin signaling in the bowhead whale (Keane et al., 2015; Seim et al., 2014); and duplication of the p53 gene in elephants (Abegglen et al., 2015). Again, it is important to ascertain whether these mechanisms are unique characteristics of certain exceptionally long-lived species, or they can also be extended to account for the general lifespan variation (Partridge and Gems, 2002).

An extension of this approach has been cross-species analyses in a larger scale. For example, several biochemical studies across multiple mammalian and bird species identified some features correlating with species lifespan. Thus, longevity of fibroblasts and erythrocytes *in vitro* (Rohme, 1981), poly (ADP-ribose) polymerase activity (Grube and Bürkle, 1992), and rate of DNA repair (Cortopassi and Wang, 1996) were found to be positively correlated with longevity, whereas mitochondrial membrane and liver fatty acid peroxidizability index (Pamplona et al., 2000; Pamplona et al., 1998), rate of telomere shortening (Haussmann et al., 2003), and oxidative damage to DNA and mitochondrial DNA (Adelman et al., 1988; Barja and Herrero, 2000) showed negative correlation. Recently, the advent of high throughput RNA sequencing (RNAseq) and mass spectrometry technologies has enabled the quantification of whole transcriptomes (Fushan et al., 2015), metabolomes (Ma et al., 2015b), and ionomes (Ma et al., 2015a), across multiple species and organs. These studies revealed the complex transcriptomic and metabolic landscape across different organs and species, as well as some overlaps with the changes observed in the long-lived mutants created in laboratory (Ma et al., 2015b).

In this regard, while molecular profiling of mammals at the level of tissues may better represent the underlying biology, profiling in cell culture represents more defined experimental conditions and allows further manipulation to alter the identified molecular phenotypes. In this study, we examined the transcriptomes and metabolomes of primary skin fibroblasts across 16 species of mammals, to identify the molecular patterns associated with species longevity. We report that the genes involved in DNA repair and glucose metabolism were up-regulated in the longer-lived species, whereas proteolysis and protein translocation activities were suppressed. In terms of metabolites, the longer-lived species had higher levels of amino acids, but lower levels of lyosphosphatidylcholine and

lysophosphatidylethanolamine. Thus, molecular insights into longevity may indeed come from defined cell culture systems in mammals.

# RESULTS

## Longevity trait variation among the species

To identify the molecular signatures associating with differences in longevity, we obtained primary skin fibroblasts from 13 species of rodents, and supplemented them with fibroblasts of 2 species of bats and 1 species of shrew which served as outgroups (Figure 7.1). These animals represented 3 taxonomic orders (Rodentia, Chiroptera, and Soricomorpha) and were characterized by a wide range of maximum lifespan (ML; from 2.2 years in shrew to 34.0 years in little brown bat) and adult weight (AW; from 10 g in little brown bat to 20 kg in beaver) (Figure 7.1, Table 7.1.). Female time to maturity (FTM) was included as an additional longevity trait, as it might be less prone to reporting bias than ML (Spearman correlation coefficient between ML and FTM was 0.87). In addition, since both ML and FTM increase with AW, we calculated the body mass adjusted residuals (i.e. MLres and FTMres), to represent the ratio between the observed longevity and the expected longevity based on body mass (Ma et al., 2015b; Tacutu et al., 2013). For instance, the little brown bat is small in size but exceptionally long-lived (Brunet-Rossinni, 2004), hence its positive log10 MLres value (Figure 7.1).

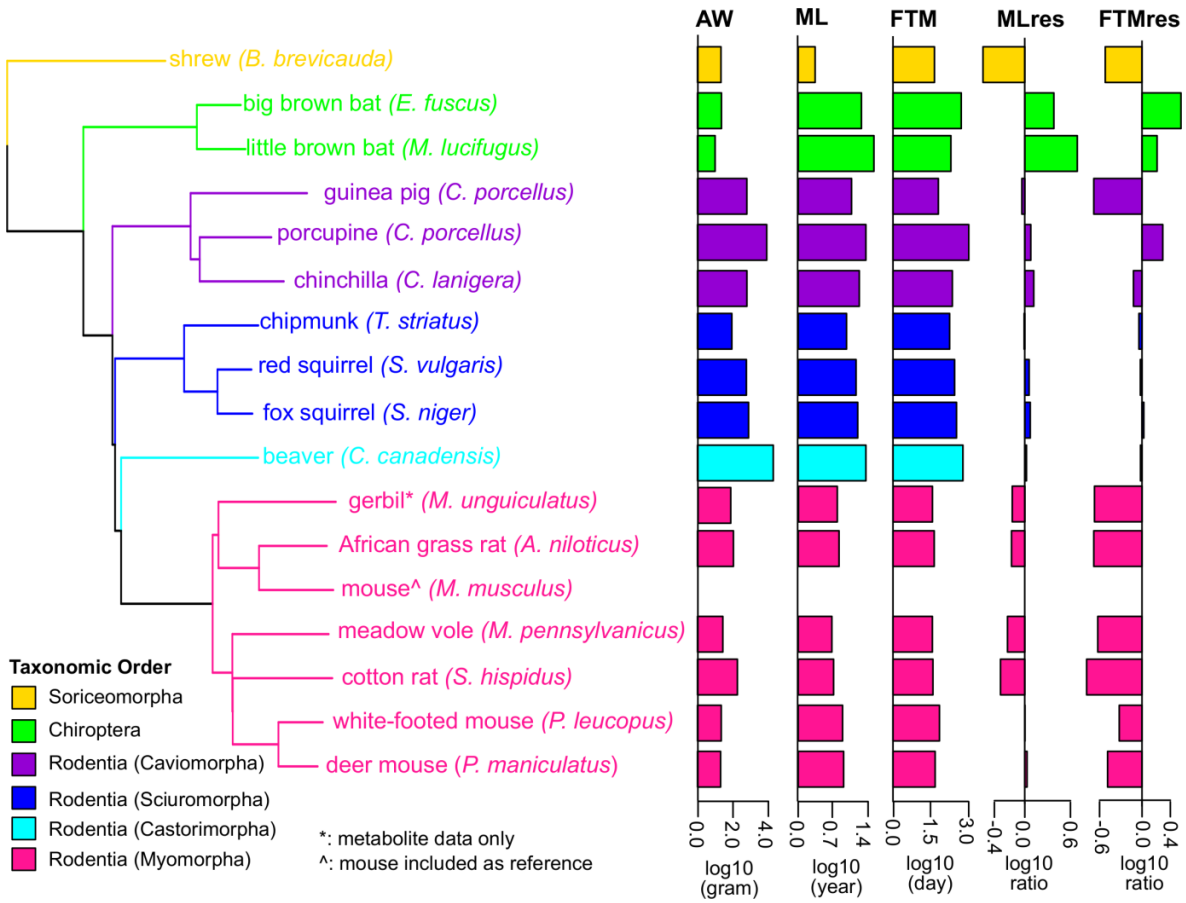



**Figure 7.1. Phylogenetic relationship among species used in the study.**

The tree was constructed using Neighbor-Joining method based on nucleotide sequences. Shrew was used as the out-group. Gerbil was collected for metabolite data only and mouse was included as reference. The species are colored by taxonomic order. Adult Weight (AW), Maximum Lifespan (ML), Female Time to Maturity (FTM), Maximum Lifespan Residual (MLres), and Female Time to Maturity Residual (FTMres) of these species are displayed in log10 scale.

**Table 7.1. Species and traits information.** The life history traits of adult weight (AW, in grams), maximum lifespan (ML, in years), and female time to maturity (FTM, in days) of these species were obtained from Anage database (Tacutu et al., 2013). Since the life history data were not available for meadow vole, the data of a related species *Microtus arvalis* were used instead. Maximum lifespan residual (MLres), and female time to maturity residuals (FTMres) were computed based on the allometric equations $MLres = ML/(4.88 \times AW^{0.153})$ and $FTMres = FTM/(78.1 \times AW^{0.217})$, respectively (Ma et al., 2015a). (*): longevity data based on a related species *Microtus arvalis.*

| Order | Family | Scientific Name | Common Name | AW | ML | FTM | MLres | FTMres |
|---|---|---|---|---|---|---|---|---|
| Soricomorpha | Soricidae | *Blarina brevicauda* | Shrew | 21.6 | 2.2 | 46 | 0.282 | 0.302 |
| Chiroptera | Vespertilionidae | *Eptesicus fuscus* | Big brown bat | 23 | 19 | 547 | 2.41 | 3.547 |
| Chiroptera | Vespertilionidae | *Myotis lucifugus* | Little brown bat | 10 | 34 | 210 | 4.898 | 1.631 |
| Rodentia | Caviidae | *Cavia porcellus* | Guinea pig | 639.1 | 12 | 66 | 0.915 | 0.208 |
| Rodentia | Erethizontidae | *Erethizon dorsatum* | Porcupine | 8600 | 23.4 | 1095 | 1.199 | 1.963 |
| Rodentia | Chinchillidae | *Chinchilla lanigera* | Chinchilla | 642.5 | 17.2 | 240 | 1.311 | 0.756 |
| Rodentia | Sciuridae | *Tamias striatus* | Chipmunk | 89.6 | 9.5 | 187 | 0.979 | 0.903 |
| Rodentia | Sciuridae | *Sciurus niger* | Fox squirrel | 800 | 16 | 353 | 1.179 | 1.06 |
| Rodentia | Sciuridae | *Sciurus vulgaris* | Red squirrel | 600 | 14.8 | 296 | 1.14 | 0.946 |
| Rodentia | Castoridae | *Castor canadensis* | Beaver | 20250 | 23.4 | 639 | 1.052 | 0.951 |
| Rodentia | Muridae | *Meriones unguiculatus* | Gerbil | 64.8 | 6.3 | 43 | 0.682 | 0.223 |
| Rodentia | Muridae | *Arvicanthis niloticus* | African grass rat | 110 | 6.7 | 45 | 0.669 | 0.208 |
| Rodentia | Muridae | *Microtus pennsylvanicus (*)* | Meadow vole | 27.5 | 4.8 | 38 | 0.592 | 0.237 |
| Rodentia | Muridae | *Sigmodon hispidus* | Cotton rat | 185 | 5.2 | 40 | 0.479 | 0.165 |
| Rodentia | Muridae | *Peromyscus leucopus* | White footed mouse | 22.3 | 7.9 | 73 | 1.007 | 0.477 |
| Rodentia | Muridae | *Peromyscus maniculatus brandii* | Deer mouse | 20.5 | 8.3 | 49 | 1.071 | 0.326 |

## Gene expression by RNA sequencing

We profiled gene expression by RNAseq on 28 samples representing 15 species, with 12-28 million reads per sample and read length of 50 or 100 nucleotides (Table 7.2). Since only 5 of the species had publicly available genomes, it posed some challenges for cross-species analysis. Reliable reference sequences (e.g. genome or transcriptome) are crucial for accurate RNAseq read alignment and read counting, and aligning reads to the genome of a related species is often far from ideal: only 13% of the reads of African grass rat could be uniquely mapped to the mouse genome (even though both species belong to the same Family Muridae), and the alignment rate was even lower for the red squirrel (about 5%). Furthermore, while the gene orthology information for many well-studied animals can be obtained from public databases (Blanchette et al., 2004; Remm et al., 2001; Vilella et al., 2009), such information is very limited or unavailable for the less common species.

To address these issues, we developed a pipeline to obtain species-specific ortholog sets (Figure 7.2A, Experimental Procedures). First, we defined a set of mouse reference sequences based on Ensembl by selecting the longest transcript per gene and removing highly repetitive or highly similar sequences. Then, from the raw RNAseq reads, the transcriptome was assembled *de novo* for each species. To identify the ortholog sets, BLAST was used to find reciprocal best hits between the assembled transcriptome (and published genome, if available) and the mouse reference (Altschul et al., 1997; Camacho et al., 2009; Tatusov et al., 1997). The reciprocal best hits were then trimmed down to open reading frame (ORF, i.e. coding sequence flanked by start and stop codons) and the quality of the ortholog sets was assessed by multiple sequence alignment.

With respect to the mouse reference sequences, the median nucleotide sequence identity for our ortholog sets ranged from 83.2% (shrew) to 95.0% (African grass rat), and protein sequence identity from 88.0% (little brown bat) to 96.8% (African grass rat) (Figure 7.2B), consistent with the evolutionary distance of the species to mouse. The read alignment rates to our ortholog sets were also largely consistent across samples (Figure 7.2C), and we observed no significant differences between the species with publicly available genomes and those without. For those with publicly available genomes, the read counts using the complete genome also agreed well with the read counts using our

ortholog sets (Pearson correlation coefficient 0.95-0.98 for log10 counts; Table 7.2). For a number of sequences with poor coverage, consensus sequences of closely related species were used instead (Figure 7.3A); this did not significantly affect the results (Figure 7.3B). After data filtering and normalization (Experimental Procedures), the expression of 9389 gene orthologs were reliably detected across the 28 samples.

**Table 7.2. RNA sequencing and read mapping.** The read mapping statistics were based on STAR. For those species with publically available genomes, the reads were also aligned to the full genomes for mapping rate comparison.

| Sample ID | | Read Counts | Average read length | Aligned to ortholog sets | Aligned to complete genome | Read counts correlation (mean) |
|---|---|---|---|---|---|---|
| Shrew.1 | | 13,936,898 | 97 | 43.7% | n/a | n/a |
| BigBrownBat.1 | | 20,591,634 | 97 | 42.1% | 84.9% | 0.968 |
| LittleBrownBat.1 | | 14,475,085 | 97 | 42.7% | 77.4% | 0.933 |
| LittleBrownBat.2 | | 16,713,612 | 97 | 41.1% | 77.6% | 0.931 |
| GuineaPig.1 | | 17,244,791 | 97 | 41.6% | 86.5% | 0.965 |
| GuineaPig.2 | | 22,611,505 | 195 | 37.8% | 89.1% | 0.972 |
| Porcupine.1 | | 16,797,544 | 97 | 37.0% | n/a | n/a |
| Porcupine.2 | | 15,861,088 | 97 | 19.7% | n/a | n/a |
| Chinchilla.1 | | 22,077,329 | 195 | 45.3% | 90.1% | 0.983 |
| Chipmunk.1 | | 16,203,627 | 97 | 24.6% | n/a | n/a |
| RedSquirrel.1A | Technical Replicates | 37,526,540 | 195 | 42.1% | n/a | n/a |
| RedSquirrel.1B | | 23,118,578 | 195 | 48.5% | n/a | n/a |
| RedSquirrel.2 | | 14,929,276 | 97 | 46.6% | n/a | n/a |
| FoxSquirrel.1 | | 14,759,305 | 97 | 47.4% | n/a | n/a |
| FoxSquirrel.2 | | 12,470,932 | 97 | 21.9% | n/a | n/a |
| Beaver.1 | | 12,907,469 | 97 | 34.9% | n/a | n/a |
| Beaver.2 | | 17,404,650 | 97 | 39.2% | n/a | n/a |
| AfricanGrassRat.1 | | 27,700,912 | 196 | 50.3% | n/a | n/a |
| MeadowVole.1 | | 15,904,868 | 97 | 33.8% | n/a | n/a |
| CottonRat.1 | | 18,164,822 | 97 | 44.7% | n/a | n/a |
| WhiteFootedMouse.1 | | 14,953,087 | 97 | 43.3% | n/a | n/a |
| DeerMouse.1 | | 17,338,832 | 97 | 47.8% | 85.8% | 0.954 |
| DeerMouse.2A | Technical Replicates | 16,152,677 | 97 | 38.0% | 81.8% | 0.948 |
| DeerMouse.2B | | 18,526,425 | 97 | 38.5% | 81.2% | 0.952 |
| DeerMouse.3A | | 18,696,359 | 97 | 42.9% | 84.8% | 0.950 |
| DeerMouse.4 | | 21,536,475 | 195 | 50.2% | 90.6% | 0.965 |
| DeerMouse.5 | | 18,726,959 | 97 | 41.5% | 82.4% | 0.949 |
| DeerMouse.6 | | 15,027,631 | 97 | 40.5% | 83.2% | 0.952 |

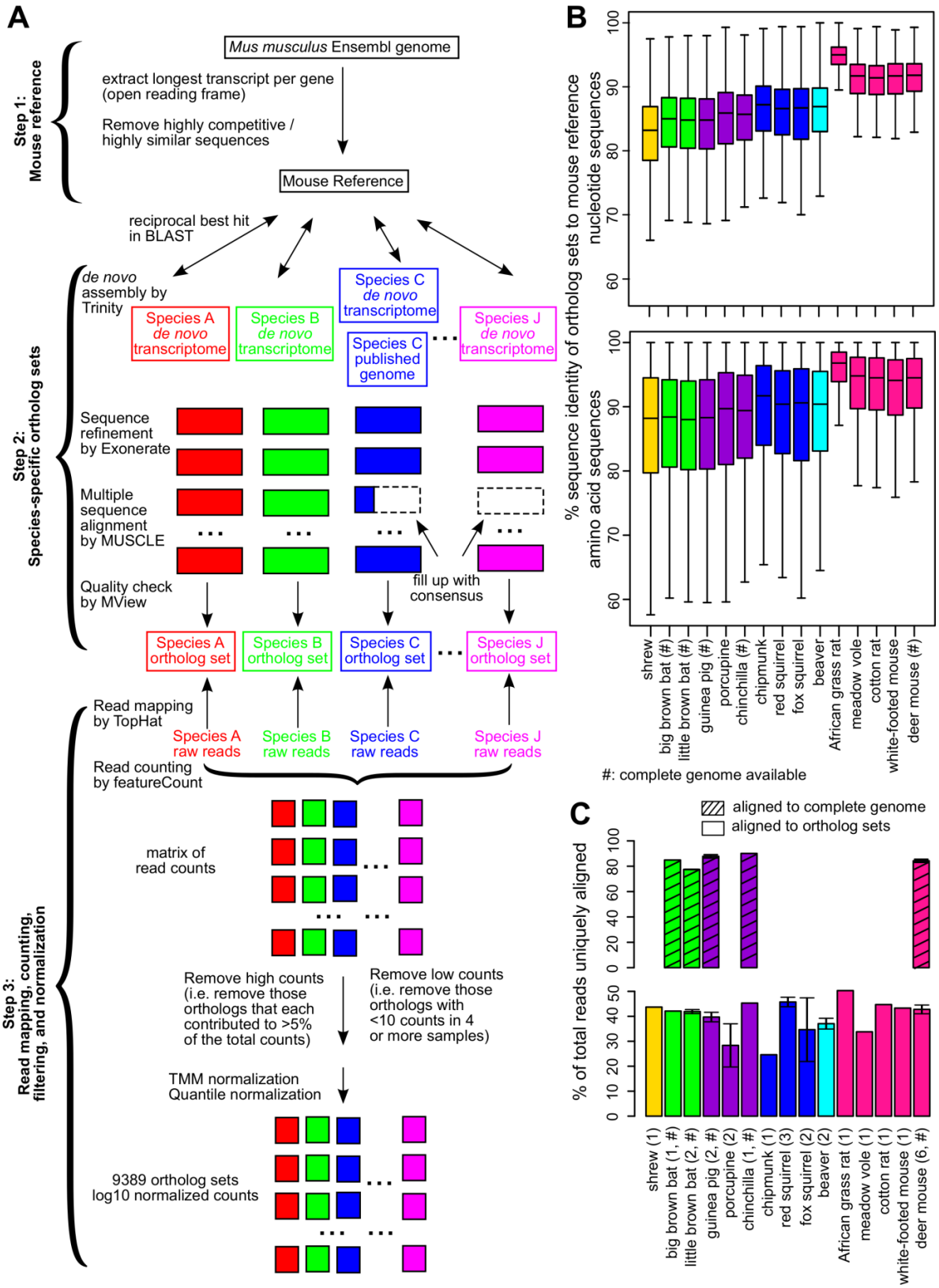
**Figure 7.2. Cross-species analysis of gene expression in cultured skin fibroblasts.**

**(A) Pipeline to obtain the species-specific ortholog sets and expression values.** See Experimental Procedures for a more detailed description of the methodology.

**(B) Sequence identity of ortholog sets compared to mouse.** Nucleotide and amino acid sequence identity of the ortholog sets in each species was compared to mouse reference (mouse was set as 100%). The ortholog sequences were based on *de novo* assembled transcriptomes, as well as NCBI genomes (if available; indicated by "#"). The box plot shows the distribution across the 9389 gene orthologs, with the central bars indicating median values.

**(C) Read alignment rates for mapping to complete genomes and ortholog sets.** Percent of total reads that could be uniquely aligned to the complete genomes (if available, indicated by "#"; shaded bars) or to the ortholog sets are shown. Error bars refer to standard error of mean. Number of samples (biological and technical replicates) per species is indicated in parenthesis.

**Figure 7.2 (Continued)**



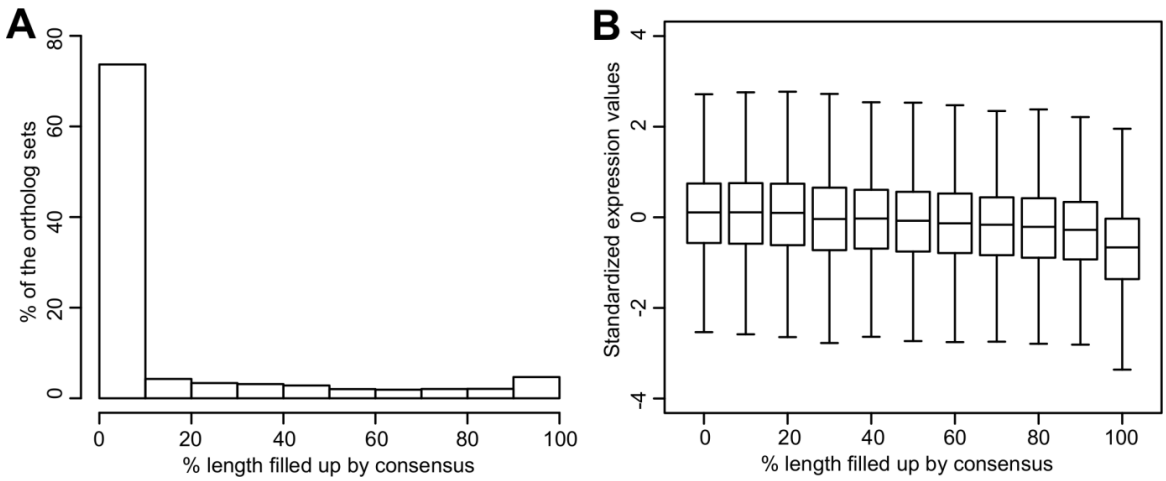



**Figure 7.3. Genes filled up with consensus sequences.**

**(A) Percentage of ortholog sets filled up using consensus.** The horizontal axis indicates the percentage of sequence length filled up by consensus. For example, 74% of the ortholog sets did not require filling up or were filled up < 10% of the sequence length. 5% of the ortholog sets were filled up 90% - 100% of the sequence length.

**(B) Standardized expression values of ortholog sets filled up using consensus.** Within each ortholog set, the expression values were standardized to mean = 0 and standard deviation = 1.

## Gene expression patterns in fibroblasts follow phylogeny

To assess gene expression patterns across species, we first performed Principal Component Analysis and projected the data on the first 3 Principal Components (Figure 7.4A). Samples segregated predominantly by taxonomic relationship. For example, the species belonging to the sub-orders Sciuromorpha (chipmunk, red squirrel, and fox squirrel), Hystricomorpha (guinea pig, porcupine, and chinchilla), and Myomorpha (African grass rat, meadow vole, cotton rate, white-footed mouse, and deer mouse) separated clearly from one another (Figure 7.4A). Furthermore, when we constructed a phylogram using the expression values, the topology was also similar to the tree based on nucleotide sequences (Figure 7.4B), suggesting the expression patterns are influenced by phylogeny. In addition, the biological and technical replicates of the respective species clustered together, confirming that the within-species variation was generally smaller than the cross-species variation (Ma et al., 2015b).

**Figure 7.4. Gene expression variation and correlation with longevity.**

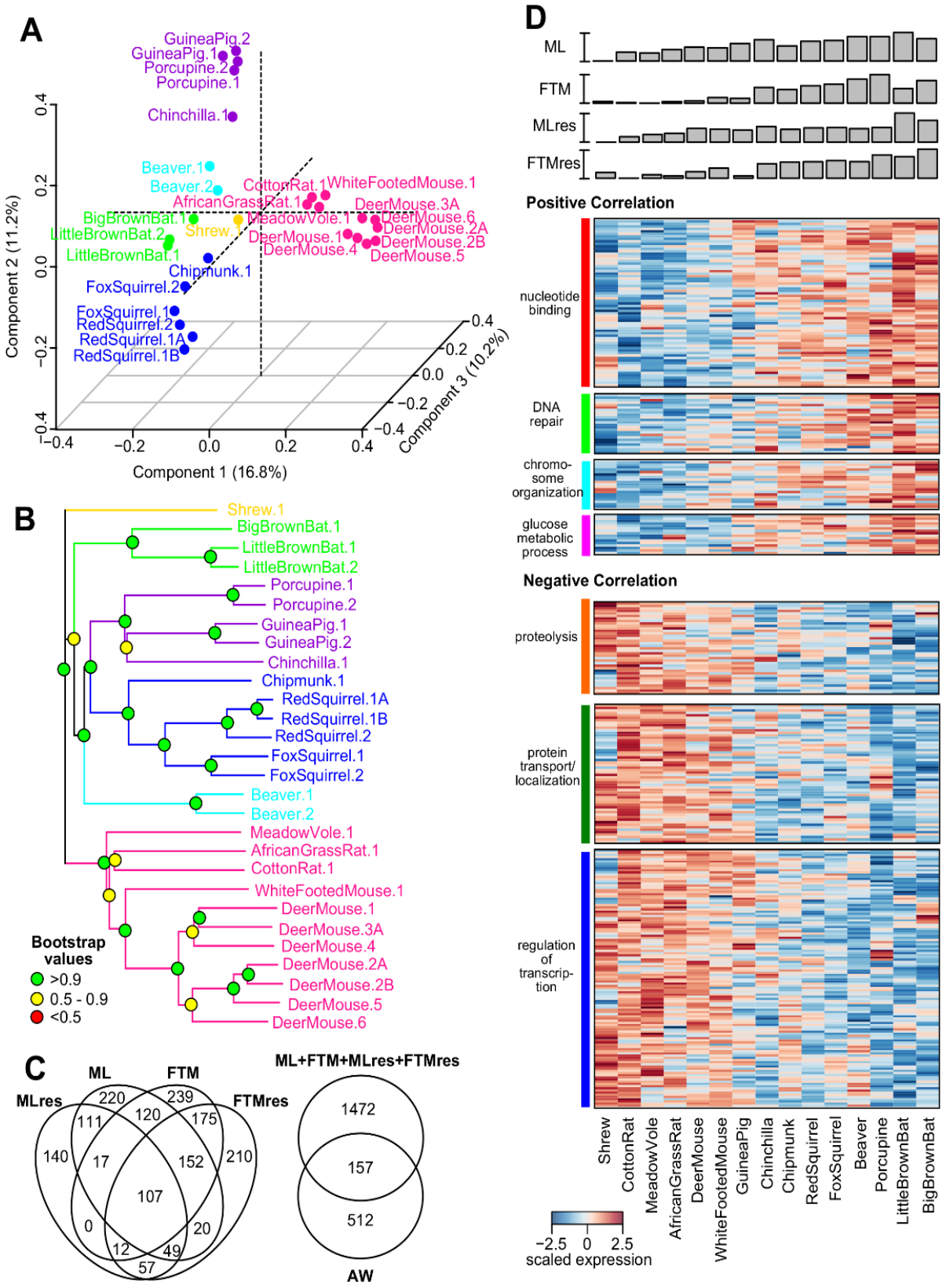
**(A) Projection of the first 3 Principal Components (PCs) in Principal Component Analysis.** Values in parenthesis indicate percentage of variance explained by each of the PCs. Points are colored by taxonomic order (same color scheme as in Figure 7.1)

**(B) Gene expression phylogram.** Color of the nodes indicates the result of 1000 times bootstrap.

**(C) Overlap of genes associating with Adult Weight and indicated longevity traits.** AW: Adult Weight; ML: Maximum Lifespan; FTM: Female Time to Maturity; MLres: Maximum Lifespan Residual; FTMres: Female Time to Maturity Residual.

**(D) Heat map showing expression patterns of the top enrichment pathways.** Species are arranged in the order of increasing longevity (the four longevity traits are scaled between 0 and 1).

**Figure 7.4 (Continued)**



## Expression of many genes correlates with longevity traits

To identify genes with significant correlation with longevity, we performed regression by generalized least squares between the gene expression values and AW, as well as the four longevity traits (ML, FTM, MLres, and FTMres). The phylogenetic relationship of the species was incorporated in the variance-covariance matrix and four different trait evolutionary models were tested to select the best models based on maximum likelihood (Experimental Procedures) (Lavin et al., 2008; Ma et al., 2015b). A two-step verification procedure was applied to assess robustness of the results, as previously described (Ma et al., 2015b). Briefly, in the first step, regression was performed by excluding the species with the largest residue error (e.g. a potential outlier) to report the regression slope p value ("p value.robust"); in the second step, regression was repeated by excluding each species, one at a time, to report the maximal (i.e. least significant) p value ("p value.max"), to ensure the overall relationship did not depend on any single species. The respective False Discovery Rates (i.e. "q value.robust" and "q value.max") were also computed.

We qualified as top hits those genes meeting both criteria of p value.robust < 0.01 (~11% FDR) and p value.max < 0.05. The numbers of top hits were 669 for AW, 796 for ML, 822 for FTM, 493 for MLres, and 782 for FTMres, with roughly equal proportions in positive and negative correlations. There was significant overlap among the top hits identified by the four longevity traits (Figure 7.4C). For most of the top hits, the directions of correlation were also consistent across the four longevity traits (even for those that failed to reach statistical significance). Therefore, our procedure was able to identify a core set of longevity-associated genes that behaved consistently across the life history traits, and minor inaccuracy in the reported lifespan data was unlikely to affect the overall results. On the other hand, the overlap with the hits identified by AW was much smaller (Figure 7.4C), suggesting the observed correlations were not driven mainly by body mass differences. For the 820 top hits supported by two or more longevity traits, we performed pathway enrichment analysis using DAVID (Table 7.3) (Huang da et al., 2009a, b) and visualized the results using heat map (Figure 7.4D) and STRING (Figure 7.5) (Jensen et al., 2009).

**Table 7.3. Pathway enrichment analysis of the genes with significant correlation with the longevity traits.** The genes were supported by at least 2 longevity traits (p value.robust < 0.01 and p value.max < 0.05). Pathway enrichment was performed using DAVID. Only selected pathways are shown here. GO (BP): Gene Ontology (Biological Process). SP/PIR: SwissProt and Protein Information Resource.

| Annotation Cluster | Enriched Terms and Genes | No. of Genes | p value |
|---|---|---|---|
| Positive Correlation Cluster No. 1 | GO (BP): adenyl nucleotide binding | 50 | $3.52\times10^{-4}$ |
| | GO (BP): nucleotide binding | 64 | $7.06\times10^{-4}$ |
| | *Acly, Atad2, Atp2b4, Cdk2, Cdk20, Chd7, Chek1, Chkb, Cpsf7, D2hgdh, Dgkq, Dhx58, Dock6, Ero1lb, Etnk1, Fastkd5, Fn3krp, Gnai1, Guk1, Hk1, Hmgcr, Hnrnpd, Hyou1, Insr, Madd, Map4k5, Mastl, Mlkl, Mov10, Msh6, Mx2, Nadsyn1, Oplah, Pdk1, Pfkp, Phka2, Phkg2, Pkmyt1, Pms2, Pnkp, Ppp2r4, Prkar1b, Qrsl1, Rbm10, Rbm15b, Rbm38, Rhot2, Rnasel, Rps6ka2, Sacs, Sirt3, Slirp, Smarca1, Smarca5, Srsf9, Stk19, Stk36, Tbrg4, Tesk2, Thnsl1, Tia1, Top3a, Trpm4, Ttf2, Tyk2, Vps4a, Ythdc2* | | |
| Positive Correlation Cluster No. 2 | SP/PIR: DNA damage | 13 | $8.37\times10^{-4}$ |
| | SP/PIR: DNA repair | 11 | $3.66\times10^{-3}$ |
| | GO (BP): cellular response to stress | 15 | $4.55\times10^{-2}$ |
| | *Bnip3, C17orf70, Chek1, Dtx3l, Ercc1, Errfi1, Fancg, Hif1a, Mapkbp1, Msh6, Myd88, Pms2, Pnkp, Prdx3, Prpf19, Pttg1, Rif1, Rnaseh1, Slx4, Tdp2, Terf1, Tinf2, Top3a, Wrap53* | | |
| Positive Correlation Cluster No. 3 | SP/PIR: chromatin regulator | 11 | $7.41\times10^{-4}$ |
| | GO (BP): chromosome organization | 17 | $1.13\times10^{-2}$ |
| | *Bnip3, Cenph, Chd7, Dtx3l, Ercc1, Hdac2, Jade1, Kdm5d, Kmt2c, Pttg1, Rcor1, Rrp8, Smarca1, Smarca5, Smyd3, Terf1, Tinf2, Wdr5, Wrap53* | | |
| Positive Correlation Cluster No. 5 | GO (BP): glucose metabolic process | 11 | $7.34\times10^{-4}$ |
| | GO (BP): hexose metabolic process | 11 | $3.01\times10^{-3}$ |
| | GO (BP): generation of precursor metabolites and energy | 14 | $3.52\times10^{-3}$ |
| | *Aldh5a1, Atp6v0d1, Atp6v0e2, Ero1lb, Fads1, Gbe1, Gpi1, Hk1, Ndufa8, Pdk1, Pfkp, Phka2, Phkb, Phkg2, Sdhaf3, Tpi1* | | |
| Negative Correlation Cluster No. 1 | GO (BP): modification-dependent protein catabolic process | 26 | $3.81\times10^{-5}$ |
| | SP/PIR: ubiquitin conjugation pathway | 25 | $1.33\times10^{-4}$ |
| | GO (BP): proteolysis | 36 | $1.83\times10^{-3}$ |
| | *Adamts2, Agtpbp1, Anapc4, Atg10, Atg4a, Atg7, Btbd1, Ctsl, Ctsz, Dcaf10, Dda1, Dpp8, Fbxl17, Fbxl20, Fbxo18, Fbxw2, Kcmf1, Map1lc3b, Med8, Mmp2, Mycbp2, Oma1, Pcsk5, Pgpep1, Pmepa1, Ppp2r5c, Rad18, Rfwd2, Rnf14, Rnf2, Rnf6, Sumo3, Tpp2, Ube2b, Ube2v1, Ufm1, Vhl* | | |
| Negative Correlation Cluster No. 2 | GO (BP): protein localization | 38 | $4.86\times10^{-7}$ |
| | GO (BP): protein transport | 34 | $1.08\times10^{-6}$ |
| | *Agap1, Akap7, Ap3d1, Atg10, Atg4a, Atg7, Bax, Cav1, Clpx, Cnih1, Col4a3bp, Cry2, Dirc2, Ergic2, Fdx1l, Fkbp15, Gabarapl2, Gdi2, Gm10273, Golt1b, Hspa9, Ift46, Ipo4, Kif1bp, Kpna4, Laptm4a, Lrp4, mt-Nd4, Mtch1, Ndel1, Ndufb11, Necap1, Ppp3ca, Rab18, Rab2a, Rab6a, Rhot1, Sar1a, Sec22a, Sec31a, Sec62, Slc25a12, Slc29a1, Slc33a1, Slc35a4, Snx12, Snx13, Stx17, Timm8a1, Tomm6, Trappc6b, Trp53, Tsg101, Vps36, Vps53, Ywhag* | | |
| Negative Correlation Cluster No. 4 | GO (BP): regulation of transcription | 73 | $2.07\times10^{-5}$ |
| | SP/PIR: transcription regulation | 55 | $2.45\times10^{-4}$ |
| | *Actl6a, Ak6, Anp32a, Anp32e, Atf6b, Bckdha, Bmi1, Ccdc59, Cd3eap, Cdc5l, Cggbp1, Clk2, Cnbp, Cops7a, Crtc3, Cry2, Csrp2, Ebna1bp2, Ehmt2, Elk4, Ergic2, Fbxo18, Fip1l1, Fosb, Foxo3, Gatad2b, Gid8, Gmcl1, Gtf2h1, Gtf2h2, Gtf2h5, Harbi1, Hlx, Hmga1-rs1, Hnrnpab, Hnrnpf, Ift57, Ing2, Ints4, Ipo4, Jund, Klf11, Klf2, Klf4, Klf9, Kpna4, Mafb, Mapk1, Mdm4, Med16, Med17, Med31, Med8, Mef2a, Mettl8, Mmp2, Mnt, Morf4l2, Mta1, Mtdh, Mxd1, Mycbp2, Nabp2, Ncor2, Neo1, Nfe2l2, Nr1d2, Papd4, Parp2, Phf12, Phlpp1, Pkig, Pomp, Pop5, Ppp1r8, Ppp2r5c, Ppp3ca, Ptbp1, R3hdm4, Rab18, Rad18, Rbbp4, Rfwd2, Rnf14, Rnf2, Rnf6, Rps6ka4, Rrs1, Sap30l, Sav1, Scoc, Sfmbt1, Sin3b, Snrk, Sqstm1, Srpk2, Ssbp1, Tep1, Tgfbr3, Trim35, Trip6, Trp53, Tsg101, Ube2b, Ube2v1, Ubtf, Ufm1, Vhl, Vps36, Wiz, Xrcc5, Yeats4, Zbtb14, Zfp414, Zfp637, Zfp655, Zfp710, Zfp821* | | |

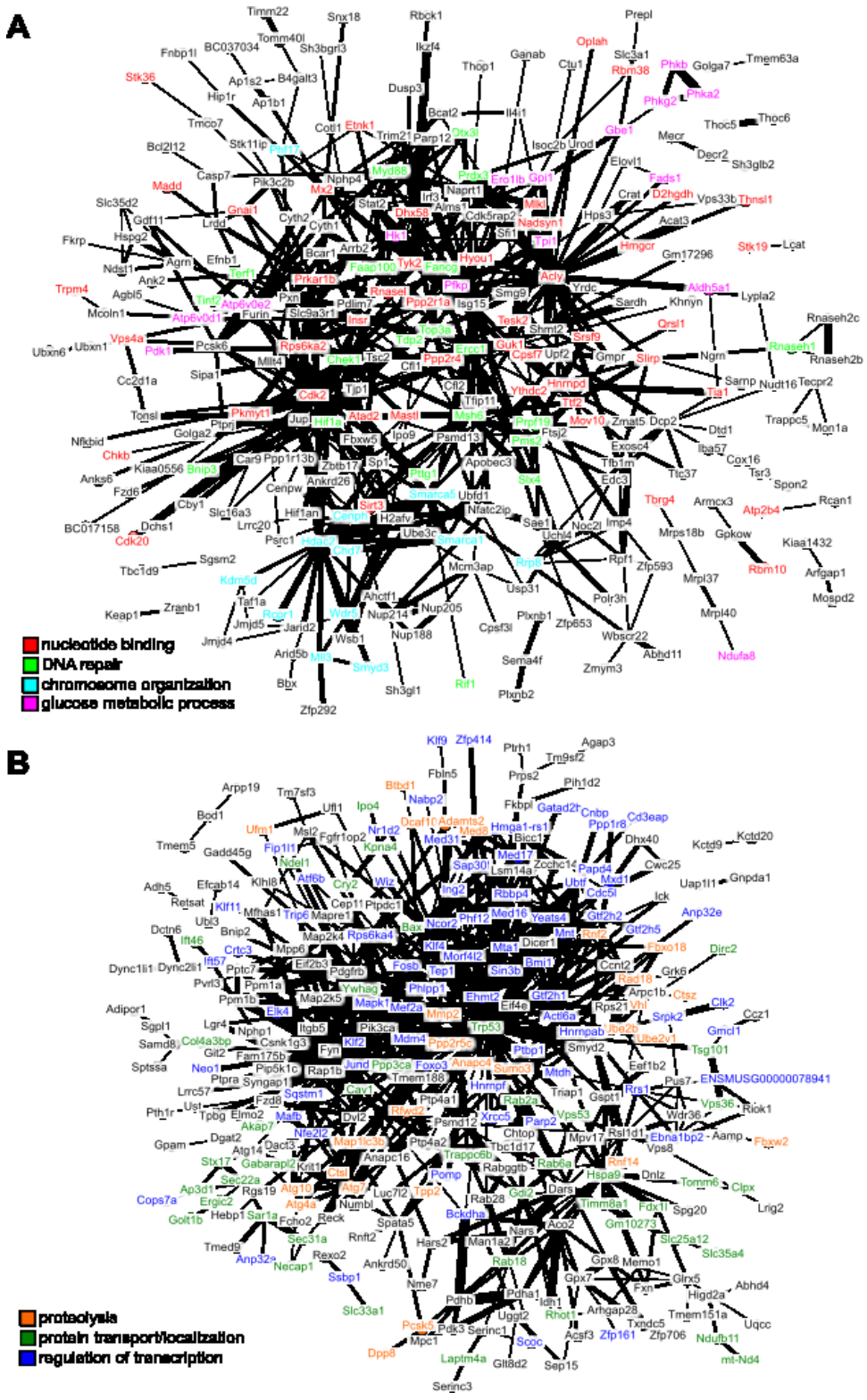



**Figure 7.5. Interaction network among the top hits in (A) positive and (B) negative correlation with longevity.** The lines represent interaction based on "evidence view" of STRING database. Selected gene names are colored based on the enriched pathways (see Table 7.3).

## Genes showing positive correlation with lifespan

Genes showing positive correlation were enriched for pathways that included "nucleotide binding", "DNA repair", "chromosome organization", and "glucose metabolic process" (Table 7.3, Figure 7.4D). "DNA repair" genes included DNA mismatch repair (*Msh6, Pms2*), nucleotide excision repair (*Ercc1, Pnkp*), Fanconi anemia-associated DNA damage response network (*C17orf70, Fancg*), and protection of telomeres (*Rif1, Terf1, Tinf2*). The products of checkpoint kinase *Chek1* and anaphase promoting complex substrate *Pttg1* were regulators of cell cycle.

Among other genes, *Hif1a* encodes the alpha subunit of hypoxia-inducible factor 1 (HIF-1), a key transcription factor in mediating the metabolic responses to hypoxia, whereas *Prdx3* encodes mitochondrial peroxiredoxin that regulates redox homeostasis. In particular, *Pnkp* (Figure 7.6A), *Prdx3*, and *Rif1* reached statistical significance in all four longevity traits. Consistent with the findings, over-expression of *hif-1* in *C. elegans* was shown to promote longevity (Zhang et al., 2009), whereas deletion of *rif1* and *msh6* in yeast (Austriaco and Guarente, 1997; Laschober et al., 2010), knockout of *prdx3* in *C. elegans* (Ha et al., 2006), and disruption of *Ercc1* in mouse (Weeda et al., 1997) were all detrimental and led to decreased lifespan. Several previous studies also suggested that long-lived species generally had enhanced DNA repair capacity (Cortopassi and Wang, 1996), higher poly (ADP-ribose) polymerase activity (Grube and Bürkle, 1992), up-regulation of genes in base-excision repair and superoxide metabolic process (Fushan et al., 2015), as well as reduced free radical production (Perez-Campo et al., 1998), reduced oxidant generation (Sohal et al., 1995), and less oxidative damage to nuclear DNA (Adelman et al., 1988) and mitochondrial DNA (Barja and Herrero, 2000), although the degree of contribution towards the observed differences in lifespan varied and had potential confounding effects (Debrabant et al., 2014; Montgomery et al., 2012; Promislow, 1994).

"Glucose metabolic process" included gene products of hexokinase *Hk1*, glucose phosphate isomerase *Gpi1*, triose phosphate isomerase *Tpi1*, phosphofructose kinase *Pfkp*, and pyruvate dehydrogenase kinase (*Pdk1*) which are involved in glycolysis/gluconeogenesis, glucan branching enzyme (encoded by *Gbe1*) and several phosphorylase kinases (encoded by *Phka2, Phkb, Phkg2*), which regulate the metabolism of glycogen. In addition, the genes coding for NAD synthetase

(*Nadsyn1*), which is involved in converting nicotinate adenine dinucleotide (NaAD) to nicotinamide adenine dinucleotide (NAD), also showed positive correlation with all four longevity traits (Figure 7.6B). Previously, it was observed that $NAD^+$ levels declined with age and affected SIRT1 functions, whereas supplementation with $NAD^+$ precursors reversed the aging phenotypes in mouse muscle (Gomes et al., 2013) and overexpression of SIRT1 in mouse brain could protect against aging-dependent circadian changes (Chang and Guarente, 2013). Calorie restriction also increases the $NAD^+$/NADH ratio in yeast (Lin et al., 2004). As our study did not directly quantify the $NAD^+$/NADH ratio, it remains to be seen how high *Nadsyn1* expression in long-lived species affects these metabolites.

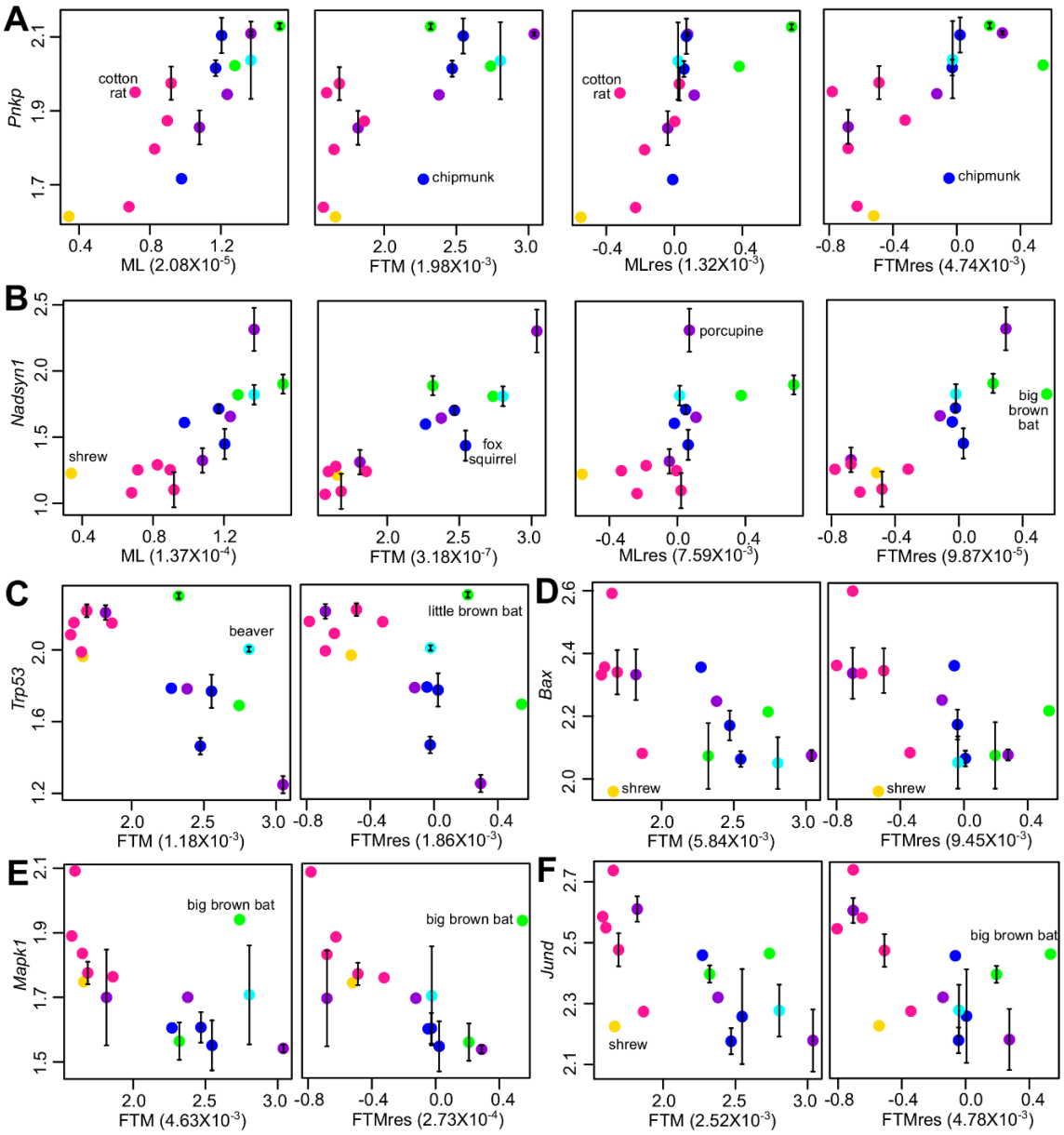



**Figure 7.6. Selected genes with significant correlation with longevity.**

**(A) *Pnkp* and (B) *Nadsyn1* show positive correlation with the longevity traits. (C) *Trp53*, (D) *Bax*, (E) *Mapk1*, and (F) *Jund* show negative correlation with the longevity traits.** In each plot, the vertical axis indicates the average log10 expression value; the horizontal axis indicates the log10 longevity traits (ML: Maximum Lifespan; FTM: Female Time to Maturity; MLres: Maximum Lifespan Residual; FTMres: Female Time to Maturity Residual); the numbers in parenthesis indicate p value.robust; the labeled species is the one with the largest residual error (which was removed to calculate p value.robust). Error bars indicate standard error of mean. Points are colored by taxonomic group (same color scheme as in Figure 7.1).

## Expression of genes showing negative correlation with lifespan

With regard to the top hits showing negative correlation, the major enriched pathways included "proteolysis", "protein transport/localization", and "regulation of transcription" (Table 7.3, Figure 7.4D). For "proteolysis", we observed relatively low expression of genes coding for E2 ubiquitin-conjugating enzyme (*Ube2b*, *Ube2v1*), E3 ubiquitin-protein ligase (*Rad18*, *Mycbp2*), ubiquitin-like modifier (*Sumo3*, *Ufm1*), as well as several proteins containing RING finger domain (*Rnf2*, *Rnf6*, *Rnf14*, *Rfwd2*) or F-box domain (*Fbxl17*, *Fbxl20*, *Fbxo18*, *Fbxw2*), both of which are known to be involved in the ubiquitination pathway. Also, low expression was observed for the genes encoding autophagy related proteins (*Atg4a*, *Atg7*, *Atg10*) and lysosomal cysteine proteinases (*Ctsl*, *Ctsz*). Genes implicated in "protein transport/localization" included several vesicle trafficking proteins (*Sec22a*, *Sec31a*, *Sec62, Golt1b*), mitochondrial membrane translocases (*Timm8a1*, *Tomm6*), and nuclear transport receptors (*Ipo4*, *Kpna4*). As for "regulation of transcription", we observed down-regulation of genes coding for mediator complex subunits (*Med8*, *Med16*, *Med17*, *Med31*), zinc finger proteins (*Zfp414*, *Zfp655*, *Zfp637*, *Zfp710*, *Zfp821*), Kruppel-like factors (*Klf2, Klf4, Klf9, Klf11*) and members of the MYC/MAX/MAD network of transcription factors (*Mxd1*, *Mnt*). Interestingly, the tumor suppressor TP53 (encoded by *Trp53*) and its regulator MDM4 (encoded by *Mdm4*), apoptosis regulator BAX (encoded by *Bax*), transcription activator of apoptosis FOXO3 (encoded by *Foxo3*), transforming growth factor beta (TGF-β) receptor (encoded by *Tgfbr3*), mitogen-activated protein (MAP) kinase (encoded by *Mapk1*), and transcription factor JunD (encoded by *Jund*) were all lower in longer-lived species (Figures 7.6C-F).

Overall, the results suggested that fibroblasts of longer-lived species had lower protein degradation, lower protein transport activities, and subdued signaling for growth and apoptosis. The link between proteolysis/autophagy and aging has been proposed by a number of authors, as generally the proteolytic functions decline and oxidized proteins increase with age, and autophagy genes are required for the lifespan extension effects of Insulin/IGF-1 signaling and dietary restriction (Chondrogianni and Gonos, 2008; Hansen et al., 2008; Kenyon, 2010; Kevei and Hoppe, 2014; Low, 2011; Melendez et al., 2003; Rubinsztein et al., 2011; Starke-Reed and Oliver, 1989; Vernace et al.,

2007). Activation of proteasome or autophagy has also been shown to extend lifespan in *C. elegans* (Chondrogianni et al., 2015; Ghazi et al., 2007), yeast (Kruegel et al., 2011), and flies (Simonsen et al., 2008). Immunoproteasome and proteasome activity was also elevated in the livers of long-lived Snell dwarf mice and in mice exposed to drugs known to extend lifespan (Pickering et al., 2015). On the other hand, our results suggest that across the species with natural lifespan variations, the longer-lived animals actually have lower expression levels of genes involved in proteolysis and autophagy, and of the tumor suppressor TP53. While our study examined only mRNA levels and disregarded potential differences in protein levels, coding sequences, and enzymatic activities, one could speculate that the naturally long-lived species need less protein degradation, perhaps due to lower damage generation or more efficient repair mechanisms. In fact, in a study examining gene expression in mammalian organs (Fushan et al., 2015), down-regulation of the ubiquitin ligase complex was observed in the liver of longer-lived species.

## Metabolites correlating with longevity traits

For 12 of the rodent species, we also profiled metabolites (Townsend et al., 2013). After data filtering and normalization, 144 water soluble metabolites and 82 lipids were reliably detected across 22 biological samples. Principal Component Analysis (Figure 7.7A) and the phylogram based on metabolite levels (Figure 7.7B) both indicated that the metabolic profiles of these species, like gene expression, segregated according to phylogeny, although the patterns were less clear-cut than those based on the RNAseq dataset. This might be partly due to the much fewer metabolites detected compared to the genes (226 metabolites vs. 9389 genes), but the degree of deviation from phylogeny might also depend on tissue origin (Ma et al., 2015b). Nevertheless, the biological and technical replicates clustered together (Figure 7.7B), suggesting the within-species variation was relatively small.

To identify metabolites with significant correlation with longevity traits, we also applied the phylogenetic regression method described above. Since there were fewer species for the metabolite dataset, the statistical power was weaker than the gene expression dataset. At the cut-off of p value.robust < 0.01 (~11% FDR) and p value.max < 0.05, 12 metabolites showed significant correlation with AW, 24 metabolites with ML, 18 metabolites with FTM, 16 metabolites with MLres, 17 metabolites with FTMres, and 22 of these metabolites were supported by 2 or more longevity traits (Figure 7.7C). Pathway analysis revealed enrichment of “common amino acids” among the top hits with positive correlation, and “glycerophospholipids” among the top hits with negative correlation. In particular, tryptophan, glutamate, leucine, arginine, lysine, and histamine showed positive correlation with multiple longevity traits; so did a number of nucleotides/nucleosides including ADP, GDP, and adenosine. This stands in contrast with the observation that the longer-lived mammalian species tend to have lower levels of amino acids in their brains (Ma et al., 2015b). On the other hand, a number of lysophosphatidylchonline (LPC; e.g. C16:0 LPC,

C18:0 LPC, C18:1 LPC) and lysophosphatidylethanolamine (LPE; e.g. C20:4 LPE, C22:6 LPE) showed negative correlation (Figure 7.7D), which were consistent with the previous report of low LPC and LPE in long-lived mammals (Ma et al., 2015b). LPC levels were also previously reported to decrease with age but maintained in mice under caloric restriction (De Guzman et al., 2013). LPC and LPE are generated by phospholipase-dependent hydrolysis of phosphatidylcholine and phosphatidylethanolamine, respectively. Elevated phospholipase A2 (PLA2) activity can lead to inflammatory response and is linked to coronary artery disease in humans (Rosenson and Stafforini, 2012). If we relaxed the cut-off criteria to p value.robust < 0.05, the patterns of positive correlation with amino acids and negative correlation with LPC and LPE were still supported across multiple longevity traits (Table 7.4).

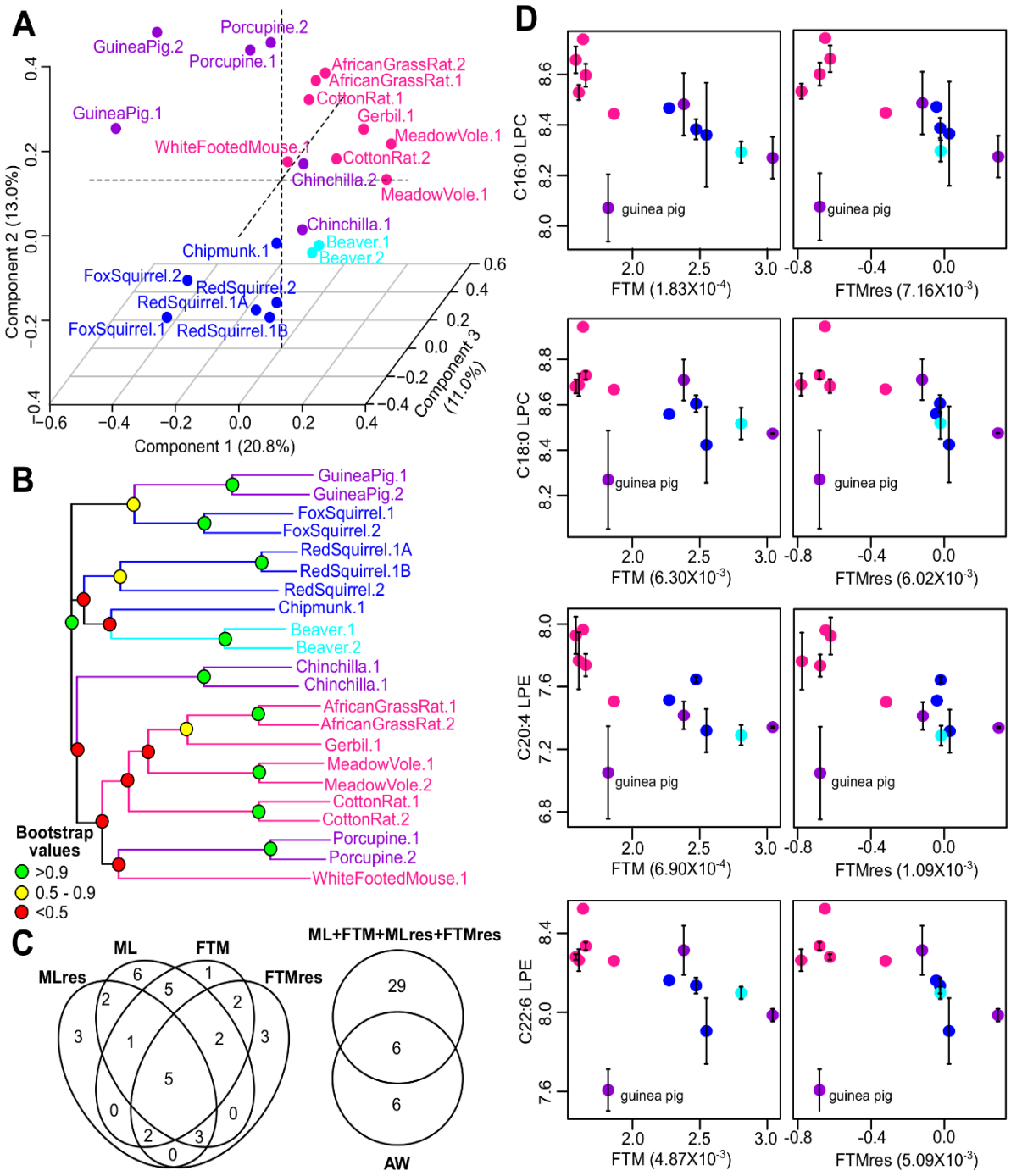



**Figure 7.7. Metabolite variation and correlation with longevity.**

**(A) Projection of the first 3 Principal Components (PCs) in Principal Component Analysis.**

Values in parenthesis indicate percent of variance explained by each of the PCs.

**(B) Metabolite phylogram.** Color of the nodes indicates the result of 1000 times bootstrap.

**(C) Overlap of metabolites associating with Adult Weight and longevity traits.**

**(D) Selected metabolites with significant correlation with longevity.**

**Table 7.4. Top hits identified by 2 or more longevity traits, using cut-off of p value.robust < 0.05.** The p value.robust against each of the four longevity traits (ML, FTM, MLres, and FTMres) as well as adult weight (AW) are shown.

| Metabolite | AW | ML | FTM | MLres | FTMres |
|---|---|---|---|---|---|
| valine | 3.40E-02 | 2.32E-02 | 3.45E-02 | 5.95E-02 | 4.86E-02 |
| C36.1.PC | 5.52E-02 | 4.68E-02 | 2.11E-02 | 1.08E-01 | 3.77E-02 |
| lactose | 4.16E-02 | 2.06E-02 | 2.63E-02 | 8.61E-02 | 1.01E-01 |
| lactate | 3.87E-02 | 1.91E-02 | 1.91E-02 | 1.05E-01 | 5.75E-02 |
| histidine | 1.32E-01 | 5.21E-02 | 4.90E-02 | 1.75E-02 | 2.22E-02 |
| X2.hydroxyglutarate | 2.38E-01 | 9.81E-02 | 5.34E-02 | 1.72E-02 | 2.42E-02 |
| ADMA.SDMA | 6.25E-02 | 1.77E-02 | 4.95E-02 | 1.36E-02 | 4.65E-02 |
| acetylcholine | 5.82E-03 | 1.28E-02 | 4.64E-02 | 1.95E-01 | 1.99E-01 |
| C34.2.PC | -1.53E-01 | 1.26E-01 | 1.94E-02 | 2.74E-01 | 1.27E-02 |
| trimethylamine.N.oxide | 1.82E-01 | 3.05E-02 | 2.31E-02 | 2.34E-02 | 1.27E-02 |
| phenylalanine | 8.50E-02 | 1.27E-02 | 2.18E-02 | 1.29E-02 | 2.25E-02 |
| UDP | 2.57E-01 | 4.82E-02 | 1.04E-01 | 1.24E-02 | 5.99E-02 |
| NMMA | 2.24E-02 | 4.04E-02 | 1.89E-02 | 1.74E-02 | 1.19E-02 |
| oxalate | 1.75E-02 | 1.12E-02 | 1.18E-02 | 2.68E-02 | 3.14E-02 |
| cotinine | 1.12E-01 | 1.15E-02 | 5.88E-02 | 1.02E-02 | 4.07E-02 |
| C32.2.PC | -2.81E-01 | 2.51E-01 | 3.00E-02 | 1.65E-01 | 9.37E-03 |
| hippurate | 1.50E-01 | 1.70E-02 | 8.03E-02 | 9.15E-03 | 1.03E-01 |
| methionine | 1.10E-02 | 7.87E-03 | 2.47E-02 | 1.90E-02 | 5.75E-02 |
| CDP | 6.62E-02 | 7.62E-03 | 2.65E-02 | 1.41E-02 | 2.23E-02 |
| tryptophan | 1.29E-01 | 6.71E-03 | 6.45E-03 | 1.78E-02 | 1.74E-02 |
| kynurenine | 2.91E-02 | 8.41E-02 | 6.11E-03 | 2.50E-01 | 3.52E-02 |
| creatine | 6.35E-02 | 8.83E-03 | 9.80E-02 | 5.64E-03 | 1.22E-01 |
| tyrosine | 3.13E-02 | 1.28E-02 | 4.06E-02 | 4.56E-03 | 1.13E-01 |
| proline | 7.77E-03 | 4.26E-03 | 1.52E-02 | 1.34E-02 | 3.81E-02 |
| glutamate | 2.35E-02 | 4.13E-03 | 2.44E-02 | 8.22E-03 | 5.31E-03 |
| leucine | 1.09E-02 | 4.65E-03 | 3.70E-03 | 3.49E-02 | 1.02E-02 |
| arginine | 1.78E-02 | 3.94E-03 | 2.97E-03 | 1.38E-02 | 4.33E-03 |
| histamine | 4.02E-02 | 2.93E-03 | 4.61E-03 | 4.90E-03 | 8.47E-03 |
| C34.3.PC | -1.85E-01 | 1.70E-01 | 3.30E-02 | 3.82E-02 | 2.52E-03 |
| sarcosine | 8.57E-02 | 1.73E-02 | 1.02E-01 | 2.49E-03 | 1.06E-01 |
| C34.2.DAG | -5.30E-02 | 2.23E-01 | 3.03E-02 | 7.34E-02 | 1.43E-03 |
| ADP | 5.95E-02 | 1.57E-03 | 7.51E-03 | 1.43E-03 | 3.35E-03 |
| pyroglutamic.acid | 1.87E-02 | 1.22E-03 | 1.51E-02 | 1.30E-02 | 2.79E-01 |
| lysine | 1.88E-04 | 1.77E-03 | 1.06E-03 | 7.04E-02 | 9.22E-03 |
| GDP | 4.53E-01 | 4.55E-03 | 1.55E-02 | 3.68E-04 | 3.11E-03 |
| sorbitol | 2.53E-02 | 3.22E-02 | 2.26E-03 | 2.23E-03 | 2.69E-04 |
| adenosine | 9.69E-02 | 4.04E-03 | 8.61E-04 | 4.12E-03 | 2.31E-04 |
| X3.phosphoglycerate | -7.79E-01 | -8.37E-03 | -3.49E-02 | -1.57E-04 | -5.21E-03 |
| C16.0.LPC | -5.44E-03 | -7.79E-04 | -1.83E-04 | -1.15E-02 | -7.16E-03 |
| C20.4.LPE | -1.30E-02 | -4.99E-04 | -6.90E-04 | -2.21E-03 | -1.09E-03 |
| alpha.glycerophosphate | -5.34E-01 | -8.13E-02 | -7.55E-03 | -5.47E-04 | -1.61E-03 |
| X2.deoxycytidine | -7.02E-03 | -6.22E-04 | -5.75E-04 | -1.89E-03 | -1.30E-03 |
| malondialdehyde | -9.58E-03 | -7.66E-04 | -7.59E-03 | -5.23E-03 | -6.23E-02 |
| cytidine | -1.98E-04 | -1.46E-03 | -1.70E-03 | -1.05E-01 | -1.68E-01 |
| C36.4.PC.B | -1.81E-02 | -2.35E-03 | -7.29E-03 | -4.80E-03 | -2.02E-02 |
| C36.4.PC.A | -3.73E-01 | -7.46E-02 | -1.65E-02 | -8.33E-02 | -3.32E-03 |
| X2.aminoadipate | -7.94E-02 | -4.49E-02 | -3.54E-03 | -3.82E-01 | -1.09E-02 |
| C18.1.LPC | -9.03E-03 | -4.39E-03 | -6.89E-03 | -4.05E-01 | -5.88E-02 |
| C22.6.LPE | -4.15E-02 | -2.23E-02 | -4.87E-03 | -7.04E-02 | -5.09E-03 |
| C18.0.LPC | -5.06E-02 | -2.69E-02 | -6.30E-03 | -6.26E-02 | -6.02E-03 |

**Table 7.4 (Continued)**

| Metabolite | AW | ML | FTM | MLres | FTMres |
|---|---|---|---|---|---|
| C18.0.SM | -4.41E-03 | -8.57E-03 | -7.48E-03 | -3.16E-01 | -3.19E-02 |
| kynurenine.vs.tryptophan | -3.45E-02 | -9.08E-03 | 2.16E-02 | -6.06E-02 | 5.54E-02 |
| C18.2.LPC | -4.00E-02 | -3.18E-02 | -1.26E-02 | -5.89E-02 | -1.04E-02 |
| putrescine | -1.57E-01 | -1.30E-01 | -1.85E-02 | -6.68E-02 | -1.12E-02 |
| dCMP | -1.32E-02 | -1.18E-02 | -3.08E-02 | -5.34E-02 | -8.54E-02 |
| X4.pyridoxate | -2.83E-02 | -7.87E-02 | -1.32E-02 | -1.43E-01 | -2.39E-02 |
| C16.carnitine | -9.17E-02 | -2.64E-02 | -5.50E-02 | -1.54E-02 | -4.94E-02 |
| glycocholate | -1.28E-02 | -1.62E-02 | -3.80E-02 | -1.02E-01 | -1.29E-01 |
| X3.methyladipate.pimelate | -2.08E-01 | -5.84E-02 | -3.41E-02 | -3.04E-02 | -1.80E-02 |
| C5.carnitine | -1.53E-01 | -2.02E-02 | -2.42E-02 | -1.83E-02 | -2.09E-02 |
| cystathionine | -4.40E-02 | -4.16E-02 | -2.22E-02 | -1.48E-01 | -4.57E-02 |
| C16.0.LPE | -5.08E-02 | -2.62E-02 | -3.87E-02 | -2.32E-01 | -1.23E-01 |
| C20.4.LPC | -7.02E-03 | -2.75E-02 | -2.67E-02 | -1.80E-01 | -5.59E-02 |
| cytosine | -6.78E-04 | -3.49E-02 | -2.90E-02 | -3.08E-01 | -2.87E-01 |
| C22.6.LPC | -2.66E-02 | -3.92E-02 | -3.12E-02 | -2.55E-01 | -9.34E-02 |
| C38.5.PC | -2.81E-03 | -4.05E-02 | -4.93E-02 | -2.45E-01 | -2.58E-01 |
| C14.0.SM | -4.41E-02 | -1.06E-01 | -4.30E-02 | -1.87E-01 | -4.81E-02 |
| C52.6.TAG | 5.06E-01 | -1.99E-01 | -1.29E-01 | -4.93E-02 | -4.99E-02 |

# DISCUSSION

All mammals descended from the same common ancestor ~230 million years ago, and since then they have undergone remarkable diversification in body size, metabolic rate, fertility, and longevity, with the corresponding changes in the gene expression and metabolite landscape (Fushan et al., 2015; Ma et al., 2015b). As fibroblasts can be obtained without sacrificing animals and can be cultured under standardized conditions, it is of great interest to determine if their gene expression and metabolite patterns represent lifespan across mammals. In addition, fibroblasts are amenable to experimental manipulation. On the other hand, cross-species gene expression analyses are often hampered by the lack of publicly available genomes and gene orthology information, especially for those species not commonly studied. Using the primary fibroblasts from 16 species of rodents, bats, and shrew, we developed a pipeline for generating species-specific ortholog sets, profiled gene expression by RNAseq and the metabolites by mass spectrometry, and identified the molecular features associated with the longevity traits.

Our pipeline can be easily extended for a larger number of species. We took an approach of defining gene orthology based on reciprocal best hit in BLAST (Tatusov et al., 1997) and ignored the issues of gene duplication and gene loss. We also filled up sequence fragments and missing sequences using consensus of the other species, so as to avoid significant length differences within the ortholog sets. While these steps unavoidably introduced inaccuracy within our species-specific ortholog sequences, they did not significantly affect the read alignment result (Figures 7.2B-C, Figure 7.3) and would be much preferred to aligning all the reads to a single reference species (e.g. mouse).

The gene expression findings revealed a clear segregation based on phylogeny (Figures 7.4A-B), suggesting the evolutionary relationship significantly influenced the expression patterns. On the other hand, the metabolite patterns were less clear-cut (Figures 7.7A-B), which might be attributed to the fewer species and metabolites, and/or stronger environmental influences. Using phylogenetic regression and the two-step verification procedure, we identified a list of genes and metabolites with significant correlations to multiple longevity traits. Importantly, although many longevity traits correlate positively with body mass, our lists of genes and metabolites were not driven primarily by

adult weight differences (Figures 7.4C and 7.7C). In terms of gene expression, the pathways of "nucleotide binding", "DNA repair", "chromosome organization", and "glucose metabolic process" were enriched among the genes with positive correlation with longevity, whereas "proteolysis", "protein transportation/localization" and "regulation of transcription" were enriched for the negative correlation. A previous study on gene expression variation in mammalian organs also reported up-regulation of "base excision repair" and down-regulation of "ubiquitin ligase complex" among long-lived species (Fushan et al., 2015), suggesting that some of the longevity signatures were also evident in fibroblasts. Interestingly, proteolysis and autophagy functions of an organism generally decline with age, and restoration of these functions have been shown effective in extending lifespan (Chondrogianni et al., 2015; Ghazi et al., 2007; Kruegel et al., 2011; Simonsen et al., 2008). Furthermore, genes coding for the tumor suppressor TP53, apoptosis regulator BAX, and several growth and proliferation signaling pathways were all down-regulated in longer-lived species (Figure 7.6C-F). One possible interpretation may be that these species generate less damage and/or have better repair mechanisms, so that the cells rely less on proteolysis, autophagy and apoptosis. In agreement, previous studies reported enhanced DNA repair mechanism and reduced oxidative damage in longer-lived species (Adelman et al., 1988; Cortopassi and Wang, 1996; Grube and Bürkle, 1992; Perez-Campo et al., 1998; Sohal et al., 1995). The metabolite dataset, on the other hand, had weaker statistical power due to the fewer species and samples. The pattern of low LPC and LPE among long-lived species was consistent with previous reports, although the positive correlation between amino acids and longevity was contrary to the observation based on mammalian brain tissues (De Guzman et al., 2013; Ma et al., 2015b).

Overall, our study is consistent with the idea that gene expression, and to some degree metabolite levels, in fibroblast cultures can uncover the cell states that correspond to longer life. Apparently, these expression patterns are preserved when the intraorganismal environment is removed and cells instead subjected to standardized cell culture conditions in the lab setting. This makes fibroblasts a particularly attractive experimental system to examine and manipulate molecular patterns, with gene expression (or a combination of gene expression and metabolite patterns) as a readout. While our study represents an initial study, this approach can be extended to a larger group of

species and samples, refining the molecular signatures and then manipulating them via genetic and environmental manipulations. Ultimately, this should reveal the genetic basis for differences in species longevity and lead to new strategies for targeting them, thereby shifting cells, and ultimately organisms to the state of related longer-lived species.

# EXPERIMENTAL PROCEDURES

**Sample collection and RNA sequencing**

Primary skin fibroblast samples were collected from shrew (*Blarina brevicauda*), big brown bat (*Eptesicus fuscus*), little brown bat (*Myotis lucifugus*), guinea pig (*Cavia porcellus*), porcupine (*Erethizon dorsatum*), chinchilla (*Chinchilla lanigera*), chipmunk (*Tamias striatus*), fox squirrel (*Sciurus niger*), red squirrel (*Sciurus vulgaris*), beaver (*Castor canadensis*), gerbil (*Meriones unguiculatus*), African grass rat (*Arvicanthis niloticus*), meadow vole (*Microtus pennsylvanicus*), cotton rat (*Sigmodon hispidus*), white-footed mouse (*Peromyscus leucopus*), and deer mouse (*Peromyscus maniculatus brandii*). Biological replicates (i.e. tissues from different individuals) and technical replicates were collected on selected species.

RNAseq libraries were prepared as previously described (Fushan et al., 2015). Paired end sequencing was done on the Illumina HiSeq2000 platform generating approximately 30 to 75 million reads per sample, with read length 50 or 100 nucleotides. The raw data were processed by Cutadapt (Martin, 2011) to remove low quality reads.

**Species specific ortholog sets and expression values**

Reference genomes were publicly available for 5 species (*Eptesicus fuscus, Myotis lucifugus, Cavia porcellus, Chinchilla lanigera, Peromyscus maniculatus brandii*). To ensure consistency across the entire dataset, we developed the following pipeline to identify species-specific ortholog sets, map the reads and obtain expression values.

Step 1: generate mouse reference. Based on the *Mus musculus* Ensembl genome and annotation (release 78), the longest transcript was extracted for each protein-coding gene locus, after confirming the presence of start and stop codons and the proper reading frame. Those transcripts containing highly repetitive or highly similar sequences were identified and removed using BLAST (at e-value cut-off $10^{-6}$) (Camacho et al., 2009). This generated the Mouse Reference, representing the coding sequences of 16,816 unique protein-coding genes.

Step 2: identify species-specific ortholog sets. For each species, the transcriptome was assembled *de novo* from the RNAseq reads using Trinity (Grabherr et al., 2011). BLAST (with "dc-

megablast" option) was performed between Mouse Reference and the assembled transcriptome (and the published genome, if available) of each species to identify the reciprocal best hits (Tatusov et al., 1997). The sequences were trimmed down to open reading frame (i.e. flanked by start and stop codons) using Exonerate (Slater and Birney, 2005). Within each ortholog sets, multiple sequence alignment was performed using MUSCLE (Edgar, 2004) and the percentage of sequence identity was assessed by MView (Brown et al., 1998). For the sequence fragments or missing sequences due to poor coverage, they were filled up using the consensus. We confirmed the filling up procedure did not significantly affect the read counting results. 74% of the ortholog sets did not require filling up or were filled up < 10% of the sequence length, whereas 5% of the ortholog sets were filled up 90% - 100% of the sequence length. When the expression values were standardized to mean = 0 and standard deviation = 1 within each ortholog set, there was no significant bias against those ortholog sets with high percentage of filling up.

Step 3: read mapping, counting, filtering and normalization. The RNAseq reads were mapped to the species-specific ortholog sets using STAR (Dobin et al., 2013), with an average read alignment rate of ~ 40%. As comparison, read mapping to publically available genomes achieved an average alignment rate of ~85%. The lower alignment rate to the species-specific ortholog sets was likely due to the exclusion of 5' and 3' untranslated regions, repetitive or highly similar sequences, and introns. Nevertheless, the alignment rates were largely similar across the samples and species. Read counting was performed by featureCounts (Liao et al., 2014) and those ortholog sets with too high counts (i.e. read counts contributing to >5% of the total counts; 3 orthologs were removed this way) or too low counts (i.e. < 10 counts in 4 or more samples) were discarded. The library sizes were then scaled by trimmed mean of M-values (TMM) method, log10-transformed, and quantile-normalized. The final expression set consisted of 9389 gene orthologs across 28 samples.

**Metabolite profiling and data processing**

The metabolite levels were quantified by mass spectrometry as previously described (Townsend et al., 2013). From the raw metabolite measurements, we kept only those metabolite with < 10% missing values. The raw values were normalized separately for the 3 collection modes (water soluble positive ionization mode "HILIC-pos", water soluble negative ionization mode "HILIC-neg",

and lipid mode “C8-pos”), first by the internal standards, and then by the total signals within each mode. The data were then log10-transformed and quantile normalized. The final expression set consisted of 226 metabolites across 22 samples.

**Principal Component Analysis and Phylograms**

Principal Component Analysis was performed on the standardized expression values or metabolite values and the first 3 Principal Components were extracted. The phylograms were constructed using neighbor joining method (Saitou and Nei, 1987), based on the distance matrix of 1 minus Pearson correlation coefficient of the standardized expression values or metabolite values. The reliability of the branching patterns was assessed by 1000 times bootstrap.

**Phylogenetic regression and two-step verification procedure**

See Chapter 2, Phylogenetic Regression. To qualify as a top hit, we required a gene to have p value.robust < 0.01 and p value.max < 0.05. For pathway enrichment purposes, we further required that the genes were identified as a top hit in 2 or more longevity traits (ML, FTM, MLres or FTMres).

**Pathway enrichment analysis and interaction network**

For the genes, pathway enrichment analysis was performed using DAVID (Huang da et al., 2009a, b). For those genes showing positive and negative correlation with longevity (supported by 2 or more longevity traits), we queried Gene Ontology (“GO Term”; Biological Process and Molecular Functions only), SwissProt and Protein Information Resource (“SP PIR Keywords”), and Kyoto Encyclopedia of Genes and Genomes (“KEGG Pathway”). STRING (Jensen et al., 2009) was used to visualize the interaction network among the top hits. Selected nodes were highlighted based on the enriched pathways. For the metabolites, pathway information was obtained from ConsensusPathDB (Kamburov et al., 2009) and Human Metabolome Database (HMDB) (Wishart et al., 2013). For ConsensusPathDB, only pathways with known KEGG IDs were incorporated. Analysis was performed on pathways with at least 5 but less than 100 metabolites. Enrichment statistics was based on a hypergeometric distribution. Odd ratios and expected counts were calculated as previously described (Gentleman et al., 2013).

# REFERENCES


Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S.*, et al.* (2015). Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA *314*, 1850-1860.

Adelman, R., Saul, R.L., and Ames, B.N. (1988). Oxidative damage to DNA: relation to species metabolic rate and life span. Proc. Natl. Acad. Sci. USA *85*, 2706-2708.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389-3402.

Austriaco, N.R., Jr., and Guarente, L.P. (1997). Changes of telomere length cause reciprocal changes in the lifespan of mother cells in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *94*, 9768-9772.

Barja, G., and Herrero, A. (2000). Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. FASEB J. *14*, 312-318.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D.*, et al.* (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. *14*, 708-715.

Brown, N.P., Leroy, C., and Sander, C. (1998). MView: a web-compatible database search or multiple alignment viewer. Bioinformatics *14*, 380-381.

Brunet-Rossinni, A.K. (2004). Reduced free-radical production and extreme longevity in the little brown bat (Myotis lucifugus) versus two non-flying mammals. Mech. Ageing Dev. *125*, 11-20.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Chang, H.C., and Guarente, L. (2013). SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. Cell *153*, 1448-1460.

Chondrogianni, N., Georgila, K., Kourtis, N., Tavernarakis, N., and Gonos, E.S. (2015). 20S proteasome activation promotes life span extension and resistance to proteotoxicity in Caenorhabditis elegans. FASEB J. *29*, 611-622.

Chondrogianni, N., and Gonos, E.S. (2008). Proteasome activation as a novel antiaging strategy. IUBMB life *60*, 651-655.

Cortopassi, G.A., and Wang, E. (1996). There is substantial agreement among interspecies estimates of DNA repair activity. Mech. Ageing Dev. *91*, 211-218.

Coschigano, K.T., Clemmons, D., Bellush, L.L., and Kopchick, J.J. (2000). Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. Endocrinology *141*, 2608-2613.

De Guzman, J.M., Ku, G., Fahey, R., Youm, Y.H., Kass, I., Ingram, D.K., Dixit, V.D., and Kheterpal, I. (2013). Chronic caloric restriction partially protects against age-related alteration in serum metabolome. Age (Dordr) *35*, 1091-1104.

Debrabant, B., Soerensen, M., Flachsbart, F., Dato, S., Mengel-From, J., Stevnsner, T., Bohr, V.A., Kruse, T.A., Schreiber, S., Nebel, A.*, et al.* (2014). Human longevity and variation in DNA damage

response and repair: study of the contribution of sub-processes using competitive gene-set analysis. Eur. J. Hum. Genet. *22*, 1131-1136.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792-1797.

Flurkey, K., Papaconstantinou, J., and Harrison, D.E. (2002). The Snell dwarf mutation Pit1(dw) can increase life span in mice. Mech. Ageing Dev. *123*, 121-130.

Friedman, D.B., and Johnson, T.E. (1988). A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics *118*, 75-86.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Gentleman, R., Falcon, S., and Sarkar, D. (2013). Category: Category Analysis. R package.

Ghazi, A., Henis-Korenblit, S., and Kenyon, C. (2007). Regulation of Caenorhabditis elegans lifespan by a proteasomal E3 ligase complex. Proc. Natl. Acad. Sci. USA *104*, 5947-5952.

Gomes, A.P., Price, N.L., Ling, A.J., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P.*, et al.* (2013). Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. Cell *155*, 1624-1638.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644-652.

Grube, K., and Bürkle, A. (1992). Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span. Proc. Natl. Acad. Sci. USA *89*, 11759-11763.

Guarente, L., and Kenyon, C. (2000). Genetic pathways that regulate ageing in model organisms. Nature *408*, 255-262.

Ha, M.K., Soo Cho, J., Baik, O.R., Lee, K.H., Koo, H.S., and Chung, K.Y. (2006). Caenorhabditis elegans as a screening tool for the endothelial cell-derived putative aging-related proteins detected by proteomic analysis. Proteomics *6*, 3339-3351.

Hansen, M., Chandra, A., Mitic, L.L., Onken, B., Driscoll, M., and Kenyon, C. (2008). A role for autophagy in the extension of lifespan by dietary restriction in C. elegans. PLoS Genet. *4*, e24.

Harrison, D.E., Strong, R., Sharp, Z.D., Nelson, J.F., Astle, C.M., Flurkey, K., Nadon, N.L., Wilkinson, J.E., Frenkel, K., Carter, C.S.*, et al.* (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature *460*, 392-395.

Haussmann, M.F., Winkler, D.W., O'Reilly, K.M., Huntington, C.E., Nisbet, I.C., and Vleck, C.M. (2003). Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. Proc. Biol. Sci. *270*, 1387-1392.

Heilbronn, L.K., and Ravussin, E. (2003). Calorie restriction and aging: review of the literature and implications for studies in humans. Am. J. Clin. Nutr. *78*, 361-369.

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloen, A., Even, P.C., Cervera, P., and Le Bouc, Y. (2003). IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature *421*, 182-187.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M.*, et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412-416.

Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2009). ConsensusPathDB--a database for integrating human functional interaction networks. Nucleic Acids Res. *37*, D623-628.

Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I.*, et al.* (2015). Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. *10*, 112-122.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Kevei, E., and Hoppe, T. (2014). Ubiquitin sets the timer: impacts on aging and longevity. Nat. Struct. Mol. Biol. *21*, 290-292.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P.*, et al.* (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223-227.

Kruegel, U., Robison, B., Dange, T., Kahlert, G., Delaney, J.R., Kotireddy, S., Tsuchiya, M., Tsuchiyama, S., Murakami, C.J., Schleit, J.*, et al.* (2011). Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLoS Genet. *7*, e1002253.

Laschober, G.T., Ruli, D., Hofer, E., Muck, C., Carmona-Gutierrez, D., Ring, J., Hutter, E., Ruckenstuhl, C., Micutkova, L., Brunauer, R.*, et al.* (2010). Identification of evolutionarily conserved genetic regulators of cellular aging. Aging Cell *9*, 1084-1097.

Lavin, S.R., Karasov, W.H., Ives, A.R., Middleton, K.M., and Garland, T., Jr. (2008). Morphometrics of the avian small intestine compared with that of nonflying mammals: a phylogenetic approach. Physiol. Biochem. Zool. *81*, 526-550.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Lin, S.J., Ford, E., Haigis, M., Liszt, G., and Guarente, L. (2004). Calorie restriction extends yeast life span by lowering the level of NADH. Genes Dev. *18*, 12-16.

Low, P. (2011). The role of ubiquitin-proteasome system in ageing. General and comparative endocrinology *172*, 39-43.

Ma, S., Lee, S.G., Kim, E.B., Park, T.J., Seluanov, A., Gorbunova, V., Buffenstein, R., Seravalli, J., and Gladyshev, V.N. (2015a). Organization of the Mammalian Ionome According to Organ Origin, Lineage Specialization, and Longevity. Cell Rep. *13*, 1319-1326.

Ma, S., Yim, S.H., Lee, S.G., Kim, E.B., Lee, S.R., Chang, K.T., Buffenstein, R., Lewis, K.N., Park, T.J., Miller, R.A.*, et al.* (2015b). Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity. Cell Metab. *22*, 332-343.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10-12.

McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The Effect of Retarded Growth Upon the Length of Life Span and Upon the Ultimate Body Size: One Figure. J. Nutr. *10*, 63-79.

Melendez, A., Talloczy, Z., Seaman, M., Eskelinen, E.L., Hall, D.H., and Levine, B. (2003). Autophagy genes are essential for dauer development and life-span extension in C. elegans. Science *301*, 1387-1391.

Miller, R.A., Harrison, D.E., Astle, C.M., Fernandez, E., Flurkey, K., Han, M., Javors, M.A., Li, X., Nadon, N.L., Nelson, J.F.*, et al.* (2014). Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. Aging Cell *13*, 468-477.

Montgomery, M.K., Buttemer, W.A., and Hulbert, A.J. (2012). Does the oxidative stress theory of aging explain longevity differences in birds? II. Antioxidant systems and oxidative damage. Exp. Gerontol. *47*, 211-222.

Mouchiroud, L., Houtkooper, R.H., Moullan, N., Katsyuba, E., Ryu, D., Canto, C., Mottis, A., Jo, Y.S., Viswanathan, M., Schoonjans, K.*, et al.* (2013). The NAD(+)/Sirtuin Pathway Modulates Longevity through Activation of Mitochondrial UPR and FOXO Signaling. Cell *154*, 430-441.

Pamplona, R., Portero-Otin, M., Riba, D., Requena, J.R., Thorpe, S.R., Lopez-Torres, M., and Barja, G. (2000). Low fatty acid unsaturation: a mechanism for lowered lipoperoxidative modification of tissue proteins in mammalian species with long life spans. J. Gerontol. A Biol. Sci. Med. Sci. *55*, B286-291.

Pamplona, R., Portero-Otin, M., Riba, D., Ruiz, C., Prat, J., Bellmunt, M.J., and Barja, G. (1998). Mitochondrial membrane peroxidizability index is inversely related to maximum life span in mammals. J. Lipid Res. *39*, 1989-1994.

Partridge, L., and Gems, D. (2002). Mechanisms of ageing: public or private? Nat. Rev. Genet. *3*, 165-175.

Perez-Campo, R., Lopez-Torres, M., Cadenas, S., Rojas, C., and Barja, G. (1998). The rate of free radical production as a determinant of the rate of aging: evidence from the comparative approach. J. Comp. Physiol. B *168*, 149-158.

Peters, R.H. (1986). The ecological implications of body size, Vol 2 (Cambridge University Press).

Pickering, A.M., Lehr, M., and Miller, R.A. (2015). Lifespan of mice and primates correlates with immunoproteasome expression. J. Clin. Invest. *125*, 2059-2068.

Promislow, D.E. (1994). DNA repair and the evolution of longevity: a critical analysis. Journal of theoretical biology *170*, 291-300.

Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. Journal of molecular biology *314*, 1041-1052.

Rohme, D. (1981). Evidence for a relationship between longevity of mammalian species and life spans of normal fibroblasts in vitro and erythrocytes in vivo. Proc. Natl. Acad. Sci. USA *78*, 5009-5013.

Rosenson, R.S., and Stafforini, D.M. (2012). Modulation of oxidative stress, inflammation, and atherosclerosis by lipoprotein-associated phospholipase A2. J. Lipid Res. *53*, 1767-1782.

Rubinsztein, D.C., Marino, G., and Kroemer, G. (2011). Autophagy and aging. Cell *146*, 682-695.

Sacher, G.A. (1959). Relation of Lifespan to Brain Weight and Body Weight in Mammals. In Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing), G.E.W. Wolstenholme, and M. O'Conner, eds. (Chichester, John Wiley & Sons, Ltd), pp. 115-141.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L.*, et al.* (2013). Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat. Commun. *4*, 2212.

Seim, I., Ma, S., Zhou, X., Gerashchenko, M.V., Lee, S.G., Suydam, R., George, J.C., Bickham, J.W., and Gladyshev, V.N. (2014). The transcriptome of the bowhead whale Balaena mysticetus reveals adaptations of the longest-lived mammal. Aging (Albany NY) *6*, 879-899.

Simonsen, A., Cumming, R.C., Brech, A., Isakson, P., Schubert, D.R., and Finley, K.D. (2008). Promoting basal levels of autophagy in the nervous system enhances longevity and oxidant resistance in adult Drosophila. Autophagy *4*, 176-184.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

Sohal, R.S., Sohal, B.H., and Orr, W.C. (1995). Mitochondrial superoxide and hydrogen peroxide generation, protein oxidative damage, and longevity in different species of flies. Free Radic. Biol. Med. *19*, 499-504.

Starke-Reed, P.E., and Oliver, C.N. (1989). Protein oxidation and proteolysis during aging and oxidative stress. Arch. Biochem. Biophys. *275*, 559-567.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-1033.

Tatar, M., Kopelman, A., Epstein, D., Tu, M.P., Yin, C.M., and Garofalo, R.S. (2001). A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science *292*, 107-110.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. Science *278*, 631-637.

Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., Mao, Z., Nevo, E., Gorbunova, V., and Seluanov, A. (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. Nature *499*, 346-349.

Townsend, M.K., Clish, C.B., Kraft, P., Wu, C., Souza, A.L., Deik, A.A., Tworoger, S.S., and Wolpin, B.M. (2013). Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin. Chem. *59*, 1657-1667.

Vernace, V.A., Arnaud, L., Schmidt-Glenewinkel, T., and Figueiredo-Pereira, M.E. (2007). Aging perturbs 26S proteasome assembly in Drosophila melanogaster. FASEB J. *21*, 2672-2682.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327-335.

Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C.J., Nigg, A., van Steeg, H., Bootsma, D., and Hoeijmakers, J.H. (1997). Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. Curr. Biol. *7*, 427-439.

Weindruch, R., Walford, R.L., Fligiel, S., and Guthrie, D. (1986). The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. J. Nutr. *116*, 641-654.

Western, D. (1979). Size, life history and ecology in mammals. Afr. J. Ecol. *17*, 185-204.

Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E*., et al.* (2013). HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. *41*, D801-807.

Wood, J.G., Rogina, B., Lavu, S., Howitz, K., Helfand, S.L., Tatar, M., and Sinclair, D. (2004). Sirtuin activators mimic caloric restriction and delay ageing in metazoans. Nature *430*, 686-689.

Zhang, Y., Shao, Z., Zhai, Z., Shen, C., and Powell-Coffman, J.A. (2009). The HIF-1 hypoxia-inducible factor modulates lifespan in C. elegans. PLoS One *4*, e6348.

# Chapter 8 Mammalian Transcriptome

Previously, our laboratory carried out RNAseq-based gene expression analyses in brain, kdiney, and liver of 33 mammalian species (Fushan et al., 2015). We have now analyzed additional species and incorporated RNAseq data from databases, generating a gene expression dataset consisting of 383 biological samples across brain, cerebellum, heart, kidney, liver, and testis in chicken and 41 mammalian species. This chapter presents initial analyses of this dataset, revealing insights into how Nature adjusts lifespan of mammals by altering gene expression in an organ-, pathway- and gene-specific manner.

## PRELIMINARY RESULTS AND DISCUSSION

### Biological samples and RNA sequencing

We compiled RNAseq data for chicken (used as an out-group) and 41 mammalian species in brain (37 species), kidney (37 species), liver (42 species), cerebellum (11 species), heart (14 species), and testis (11 species). In this dataset, 48 samples were newly sequenced and 335 samples corresponded to datasets previously published by us and other laboratories (Brawand et al., 2011; Fushan et al., 2015; Merkin et al., 2012). Biological replicates (i.e. samples from multiple individuals of a species) were available for most species. These mammals represent 12 taxonomic orders, over a wide range of longevity-associated traits (e.g. adult weight ranged from 7.0 g in the Brandt's bat to 100 tons in the bowhead whale; maximum lifespan ranged from 3.2 years in the shrew to 211 years in the bowhead whale) (Tacutu et al., 2013) (Figure 8.1, Table 8.1). 29 of these species were represented by publicly available genomes in NCBI or Ensembl, whereas the other 13 species required *de novo* assembly of the transcriptomes. To ensure consistency of read alignment and counting across the samples, a pipeline was developed to identify the sets of ortholog sequences in each species (Chapter 2, "Species Without Reference Genomes"). We confirmed that the read alignment rates to the ortholog sets were consistent across the samples with and without complete genomes (Figure 8.2A), and the Spearman correlation coefficients for the read counts from alignment to ortholog sets and from alignment to complete genomes were > 0.95 for most of the samples (Figure 8.2A). For those

species with Ensembl annotations for gene orthologs, 90-99% of our definition of ortholog sets agreed with the Ensembl annotation (Chapter 2, Table 2.5). After data filtering and normalization, 13,784 genes were reliably detected across 383 samples. For cross-organ analysis, the expression values were standardized to mean 0 and standard deviation 1. Mean values and standard errors were calculated across biological replicates.

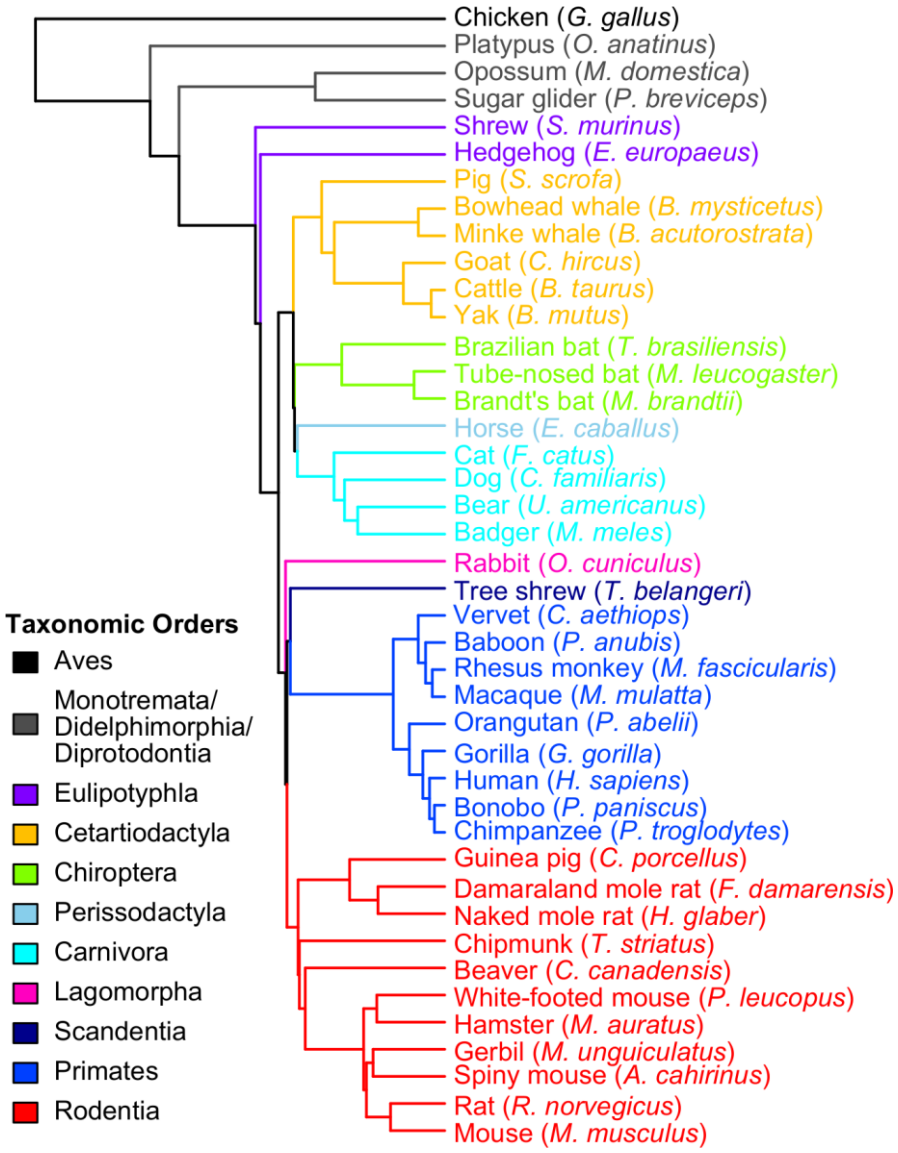



**Figure 8.1. The 42 species examined in this study.** The tree is based on nucleotide sequences and constructed by neighbor joining method (Saitou and Nei, 1987). Chicken is used as the out-group. The tips are colored by taxonomic orders. For simplicity, the same color is used for Monotremata (platypus), Didelphimorphia (opossum), and Diprotodontia (sugar glider).

**Table 8.1. 42 species and life history traits.** Adult Weight (AW), Maximum Lifespan (ML) and Female Time to Maturity (FTM) were obtained from AnAge database (Tacutu et al., 2013). Maximum Lifespan Residuals (MLres) and Female Time to Maturity Residuals (FTMres) were computed using the following allometric equations: $MLres = ML/(4.88 \times AW^{0.153})$; $FTMres = FTM/(78.1 \times AW^{0.217})$. The numbers of samples in brain (Br), cerebellum (Cb), heart (Ht), kidney (Kd), liver (Lv) and testis (Ts) are shown.

**Table 8.1 (Continued)**

| Common Name | Class | Order | Br | Cb | Ht | Kd | Lv | Ts | AW | ML | FTM | MLres | FTMres |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chicken | Aves | Galliformes | 2 | 2 | 2 | 2 | 2 | 2 | 779.8 | 30.0 | | | |
| Platypus | Mammalia | Monotremata | 2 | 2 | 2 | 2 | 2 | 2 | 1,250.0 | 22.6 | 548 | 1.56 | 1.49 |
| Opossum | Mammalia | Didelphimorphi | 2 | 2 | 2 | 2 | 2 | 2 | 105.0 | 5.1 | 122 | 0.51 | 0.57 |
| SugarGlider | Mammalia | Diprotodontia | 2 | 0 | 0 | 2 | 2 | 0 | 129.3 | 17.8 | 236 | 1.73 | 1.05 |
| Shrew | Mammalia | Eulipotyphla | 0 | 0 | 0 | 0 | 3 | 0 | 39.7 | 3.2 | 36 | 0.37 | 0.21 |
| Hedgehog | Mammalia | Eulipotyphla | 2 | 0 | 0 | 2 | 2 | 0 | 1,213.5 | 11.7 | 253 | 0.81 | 0.69 |
| Pig | Mammalia | Cetartiodactyla | 2 | 0 | 0 | 2 | 2 | 0 | 135,000.0 | 27.0 | 334 | 0.91 | 0.33 |
| BowheadWhale | Mammalia | Cetartiodactyla | 0 | 0 | 0 | 4 | 3 | 0 | 100,000,000.0 | 211.0 | 8,212 | 2.58 | 1.93 |
| MinkeWhale | Mammalia | Cetartiodactyla | 1 | 0 | 1 | 1 | 1 | 0 | 7,500,000.0 | 50.0 | 2,740 | 0.91 | 1.13 |
| Goat | Mammalia | Cetartiodactyla | 2 | 0 | 0 | 2 | 2 | 0 | 61,000.0 | 20.8 | 406 | 0.79 | 0.48 |
| Cattle | Mammalia | Cetartiodactyla | 4 | 0 | 2 | 4 | 4 | 0 | 548,500.0 | 20.0 | 548 | 0.54 | 0.40 |
| Yak | Mammalia | Cetartiodactyla | 1 | 0 | 1 | 0 | 1 | 0 | 667,000.0 | 26.3 | 738 | 0.69 | 0.51 |
| BrazilianBat | Mammalia | Chiroptera | 0 | 0 | 0 | 0 | 3 | 0 | 13.3 | 12.0 | 273 | 1.66 | 1.99 |
| TubeNosedBat | Mammalia | Chiroptera | 2 | 0 | 0 | 1 | 3 | 0 | 12.8 | 9.0 | 180 | 1.25 | 1.33 |
| BrandtsBat | Mammalia | Chiroptera | 3 | 0 | 0 | 6 | 6 | 0 | 7.0 | 41.0 | | 6.24 | |
| Horse | Mammalia | Perissodactyla | 2 | 0 | 0 | 3 | 3 | 0 | 300,000.0 | 57.0 | 914 | 1.70 | 0.76 |
| Cat | Mammalia | Carnivora | 3 | 0 | 0 | 3 | 3 | 0 | 3,900.0 | 30.0 | 289 | 1.73 | 0.62 |
| Dog | Mammalia | Carnivora | 2 | 0 | 0 | 2 | 2 | 0 | 10,148.9 | 21.8 | 274 | 1.09 | 0.47 |
| Bear | Mammalia | Carnivora | 2 | 0 | 0 | 3 | 2 | 0 | 154,250.0 | 34.0 | 1,278 | 1.12 | 1.22 |
| Badger | Mammalia | Carnivora | 2 | 0 | 0 | 2 | 2 | 0 | 11,050.0 | 18.6 | 365 | 0.92 | 0.62 |
| Rabbit | Mammalia | Lagomorpha | 2 | 0 | 0 | 2 | 2 | 0 | 2,167.9 | 9.0 | 730 | 0.57 | 1.77 |
| TreeShrew | Mammalia | Scandentia | 1 | 0 | 1 | 1 | 1 | 1 | 200.0 | 11.1 | 90 | 1.01 | 0.37 |
| Vervet | Mammalia | Primates | 1 | 0 | 0 | 1 | 1 | 0 | 5,620.0 | 30.8 | 1,034 | 1.68 | 2.03 |
| Baboon | Mammalia | Primates | 6 | 0 | 0 | 6 | 6 | 0 | 18,000.0 | 37.5 | 1,514 | 1.72 | 2.31 |
| RhesusMonkey | Mammalia | Primates | 0 | 6 | 0 | 5 | 6 | 0 | 6,362.5 | 39.0 | 1,238 | 2.09 | 2.37 |
| Macaque | Mammalia | Primates | 3 | 2 | 2 | 2 | 2 | 2 | 8,235.0 | 40.0 | 1,231 | 2.06 | 2.23 |
| Orangutan | Mammalia | Primates | 2 | 1 | 2 | 2 | 2 | 0 | 64,475.0 | 59.0 | 2,555 | 2.22 | 2.96 |
| Gorilla | Mammalia | Primates | 2 | 2 | 2 | 2 | 2 | 1 | 139,842.0 | 55.4 | 2,829 | 1.85 | 2.77 |
| Human | Mammalia | Primates | 5 | 1 | 2 | 2 | 2 | 2 | 62,035.0 | 122.5 | 4,745 | 4.64 | 5.54 |
| Bonobo | Mammalia | Primates | 3 | 2 | 2 | 2 | 2 | 1 | 39,925.0 | 55.0 | 3,194 | 2.23 | 4.10 |
| Chimpanzee | Mammalia | Primates | 6 | 2 | 2 | 2 | 2 | 1 | 44,983.5 | 59.4 | 3,376 | 2.36 | 4.23 |
| GuineaPig | Mammalia | Rodentia | 3 | 0 | 0 | 3 | 3 | 0 | 639.1 | 12.0 | 66 | 0.92 | 0.21 |
| DamaralandMoleRat | Mammalia | Rodentia | 4 | 0 | 0 | 4 | 4 | 2 | 131.3 | 15.5 | 511 | 1.51 | 2.27 |
| NakedMoleRat | Mammalia | Rodentia | 4 | 0 | 0 | 4 | 4 | 0 | 35.3 | 31.0 | 228 | 3.68 | 1.35 |
| Chipmunk | Mammalia | Rodentia | 2 | 0 | 0 | 2 | 2 | 0 | 89.6 | 9.5 | 187 | 0.98 | 0.90 |
| Beaver | Mammalia | Rodentia | 2 | 0 | 0 | 0 | 2 | 0 | 20,250.0 | 23.4 | 639 | 1.05 | 0.95 |
| WhiteFootedMouse | Mammalia | Rodentia | 3 | 0 | 0 | 2 | 3 | 0 | 22.3 | 7.9 | 73 | 1.01 | 0.48 |
| Hamster | Mammalia | Rodentia | 3 | 0 | 0 | 3 | 3 | 0 | 108.2 | 3.9 | 48 | 0.39 | 0.22 |
| Gerbil | Mammalia | Rodentia | 3 | 0 | 0 | 3 | 3 | 0 | 64.8 | 6.3 | 43 | 0.68 | 0.22 |
| SpinyMouse | Mammalia | Rodentia | 0 | 0 | 0 | 0 | 3 | 0 | 42.0 | 5.9 | 58 | 0.68 | 0.33 |
| Rat | Mammalia | Rodentia | 3 | 0 | 0 | 3 | 3 | 0 | 117.0 | 4.2 | 90 | 0.42 | 0.41 |
| Mouse | Mammalia | Rodentia | 6 | 3 | 3 | 6 | 6 | 2 | 18.0 | 4.0 | 42 | 0.53 | 0.29 |

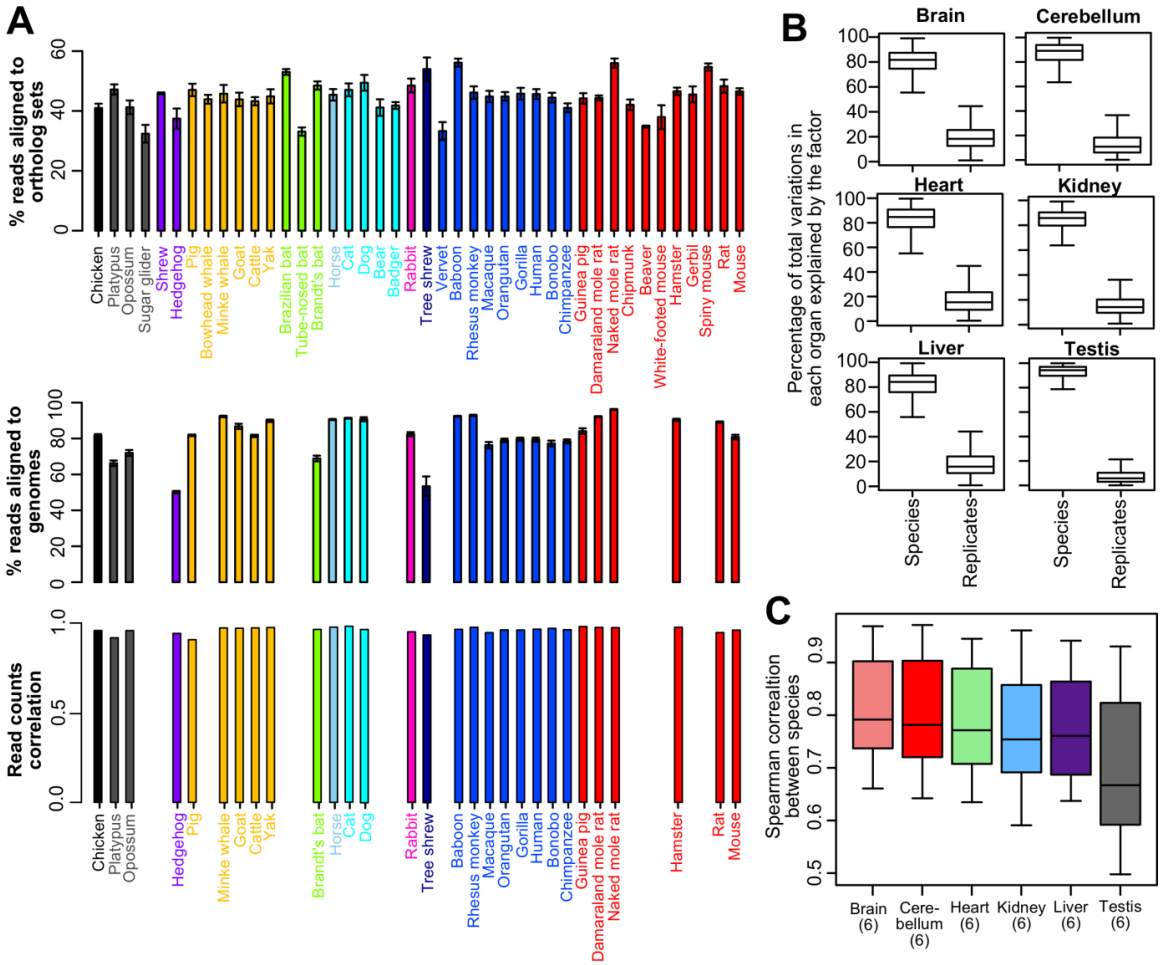



**Figure 8.2. Data quality assessment**

**(A) Comparison of read alignment to ortholog sets and to genome.** Percentage of reads for each species aligned to the ortholog sets are shown (top). For species with complete genomes, the reads were also aligned to the respective genomes (middle), and the average Spearman correlation coefficient between the read counts of ortholog set alignment and the read counts of genome alignment was calculated (bottom). See Table 2.5 for more details.

**(B) Percentage of total variation explained by species and replicates.** The box plots show the results of ANOVA in each organ.

**(C) Gene expression patterns are least conserved in testes.** The pairwise Spearman correlation coefficients between the species were calculated for each organ. Here, only six species were used: bonobo, chicken, chimpanzee, gorilla, human, macaque, mouse, opossum, and platypus, because data are available for these species in each of the organs.

## Segregation of samples by organ origin

Principal Component Analysis (PCA) revealed that the samples segregated predominantly by their organ origin (Figure 8.3A). The cerebellum samples clustered closely with the brain (cortex) samples, whereas the testis samples were distant from those of the other organs. Analysis of Variance (ANOVA) confirmed that organ and species were the predominant sources of variation and together accounted for 87% of the total variance, whereas variation among replicates accounted for only 13%. Within individual organs, the variance due to species was 82-94% and due to replicates 6-18% (Figure 8.2B). To determine how well the gene expression patterns were conserved across the species, we calculated the Spearman correlation coefficients among the species in each organ, using the mean expression values across the replicates. This revealed that the neural tissues such as brain and cerebellum had high degree of conservation (Brawand et al., 2011; Fushan et al., 2015; Ma et al., 2015b), whereas the expression patterns in testis were very diverse across the species (Figure 8.3B). Similar results were obtained if we included only those species with data in all 6 organs (Figure 8.2C), consistent with the notion that testis is a rapidly evolving tissue under the impact of sex-related evolutionary forces (Brawand et al., 2011; Kaessmann, 2010; Khaitovich et al., 2006).

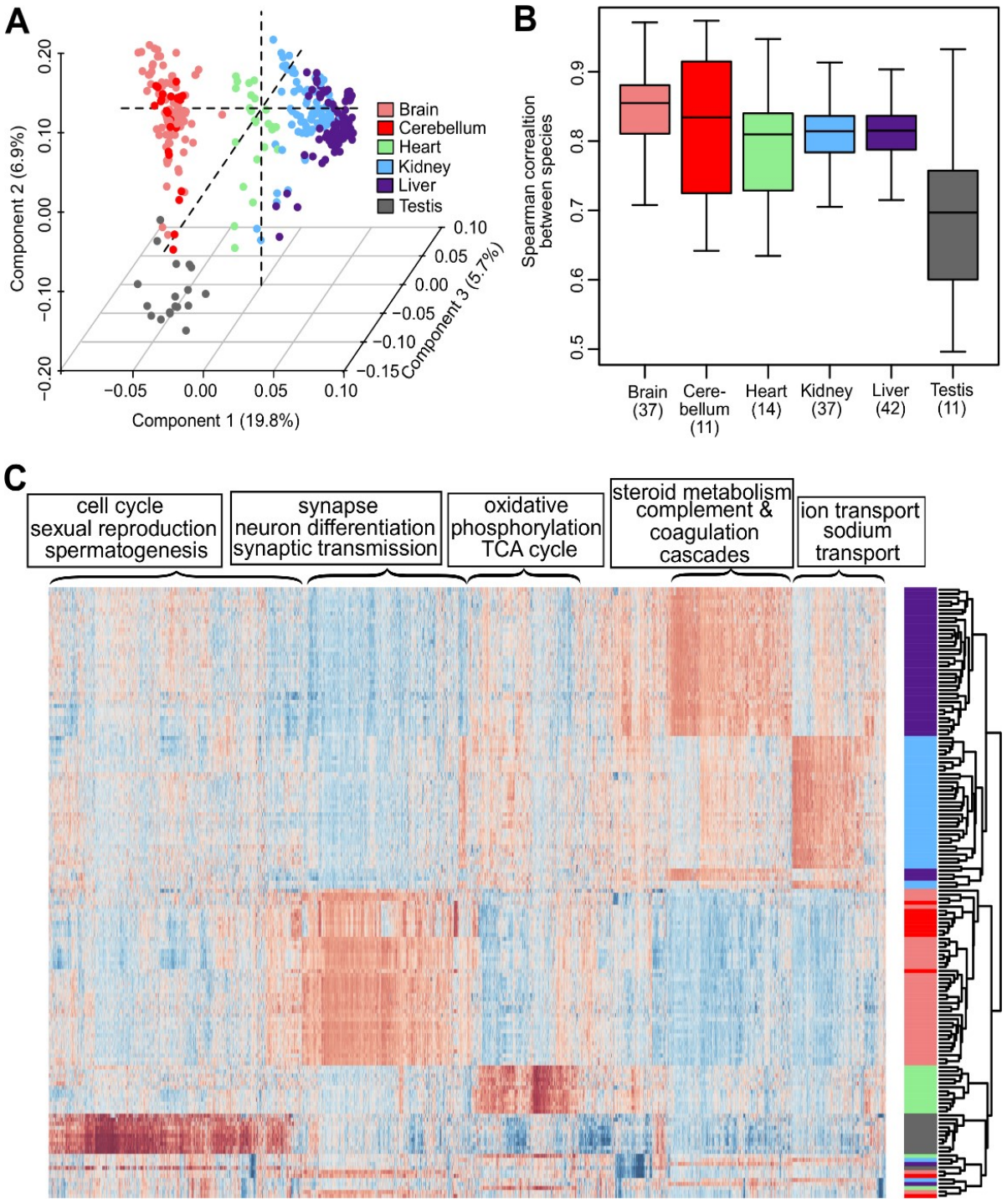
**Figure 8.3. Sample segregation and organ-specific expression patterns.**

**(A) Samples cluster by organ origin in Principal Component Analysis.** The samples (biological replicates treated as individual points) were projected on the first 3 Principal Components (PCs). The number in the parenthesis indicates the percentage of total variance explained by each PC. The points are colored by organ origin.

**(B) Gene expression patterns are least conserved in testis.** Pairwise Spearman correlation coefficients between the species were calculated for each organ. The results are displayed in a box plot. The numbers in the parenthesis indicate the number of species available in each organ.

**(C) Heat map showing organ-specific expression patterns.** The heat map shows 6,050 genes significantly enriched or depleted in one organ relative to the others. Rows represent biological samples and are colored by organ origin (the same color scheme as (A)). Columns represent the genes with significant enrichment or depletion. Pathway enrichment analysis was performed on selected genes. The top enriched terms are shown.

**Figure 8.3 (Continued)**



## Organ-specific expression patterns

To check that the expression patterns of the organs were consistent with their biological functions, we used Wilkoxon rank-sum test to identify 6,050 genes significantly enriched or depleted (p value < 0.01) in one organ relative to the others. The results were visualized on a heat map and the genes were analyzed for the enriched pathways (Figure 8.3C). In kidney, genes involved in ion transport and sodium transport (e.g. genes coding for aquaporins and many solute carrier family proteins) were overexpressed, consistent with its functions in ultra-filtration and selective reabsorption. In liver, genes coding for proteins involved in steroid metabolism and detoxification (e.g. hydroxysteroid dehydrogenases, cytochrome P450 family proteins, apolipoproteins), and in complement and coagulation cascades (e.g. coagulation factors, complement components) were found expressed at high levels, as these processes are largely liver-specific. Consistent with their high energy demand, the heart exhibited high expression of genes related to mitochondria and oxidative phosphorylation (e.g. those coding for mitochondrial $H^+$ transporting ATP synthase, different subunits of NADH dehydrogenases, and several mitochondrial ribosomal proteins), as well as enzymes of the tricarboxylic acid cycle. On the other hand, the brain and cerebellum specifically expressed genes involved in synaptic transmission (e.g. those coding for cholinergic receptors, gamma-aminobutyric acid (GABA) receptors, and glutamate receptors) and neuronal differentiation and development (e.g. a number of homeobox proteins). The testes were unique for their high expression levels of genes coding for cyclins, centrosomal proteins, and spermatogenesis-associated proteins, underlying their roles in sexual reproduction. Overall, the organ-specific expression patterns were in-line with their biological roles, and most of the samples clustered according to organ origin. The exceptions were the samples from chicken and platypus, which clustered by species and away from the rest of the samples (Figure 8.3C, the bottom rows of the heat map), probably due to their significant evolutionary distance from the rest of the examined mammalian species.

## Genes associated with longevity

For the brain, kidney and liver samples, we performed phylogenetic regression by generalized least squares (Chapter 2, "Phylogenetic Regression") to identify the genes with significant association with maximum lifespan (ML), female time to maturity (FTM), as well as the body-mass adjusted maximum lifespan residual (MLres) and female time to maturity residual (FTMres). Four different models of trait evolution were tested and the best models were selected by maximum likelihood (Chapter 2, "Phylogenetic Regression"). A two-step verification procedure was also applied to assess robustness of the data (Chapter 2, "Phylogenetic Regression"). Briefly, in the first step, the regression slope p value ("p value.robust") was calculated by excluding species with the largest residue error (e.g. a potential outlier); in the second step, regression was repeated by excluding each species, one at a time, to report the maximal (i.e. least significant) p value ("p value.max"), so that the overall relationship did not depend on any single species. The corresponding False Discovery Rates (i.e. "q value.robust" and "q value.max") were also computed. In each organ, the cut-off of p value.robust < 0.01 and p value.max < 0.05 was used to define the top hits (corresponding to ~ 8% false discovery rate).

Across the organs, depending on the longevity trait, ~ 800-2,000 genes showed significant correlation, with slightly more genes with negative correlation than with positive correlation (Table 8.2). Since ML and FTM correlate with each other (Pearson correlation coefficient 0.84), many of the genes showed the same direction of correlation across all four longevity traits. To identify the pathways represented by the top genes in each organ, we applied a stringent criterion of considering only those genes with significant correlations to both ML and MLres, or to both FTM and FTMres (Table 8.2), so that the results would not be due purely to body mass differences. To ensure the overall consistency of the results, we also pooled together the top candidates from all 3 organs. Pathway enrichment analysis revealed a number of common pathways across brain, kidney and liver, as well as some organ-specific processes (Figure 8.4).

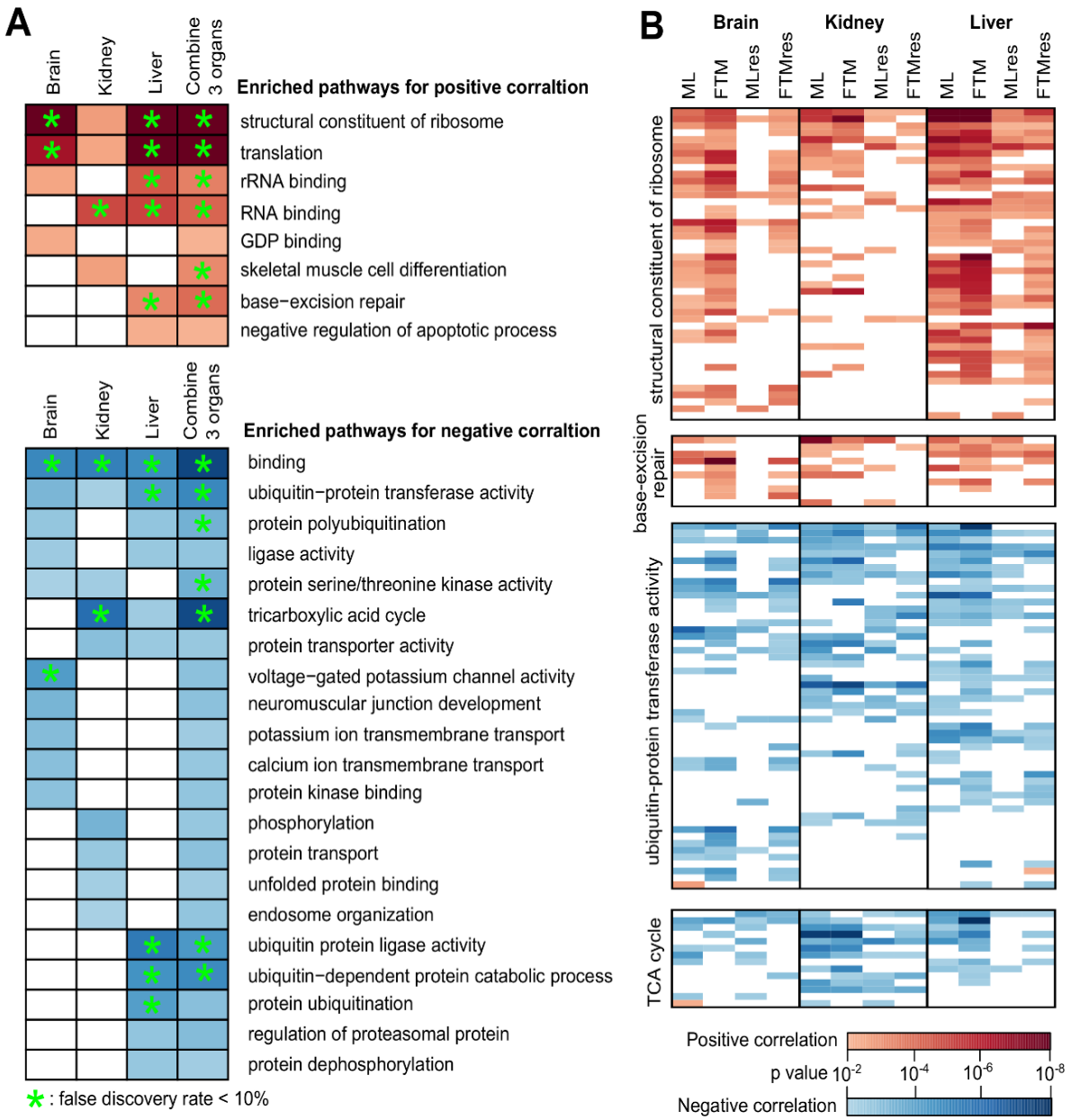



**Figure 8.4. Pathways with significant correlation to longevity traits.**

**(A) Enriched pathways for genes with positive (top, red) and negative (bottom, blue) correlation to the longevity traits.** The grid indicates the p values for enrichment analysis. Only p value < 0.01 are shown in color. Green asterisks indicate false discovery rate < 0.1. Input genes were those with significant correlation to both ML and MLres, or to both FTM and FTMres.

**(B) Genes in the selected pathways.** The grids showed the p value.robust for phylogenetic regression against the longevity traits. Only p value.robust < 0.01 are shown in color.

**Table 8.2. Number of genes with significant correlation to longevity traits.** Genes with significant correlation to both ML and MLres, or to both FTM and FTMres, are selected for pathway enrichment analysis.

| Analysis | Longevity trait | Brain | | Kidney | | Liver | | Combine 3 organs | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg |
| Phylogenetic Regression | ML | 679 | 844 | 581 | 887 | 751 | 1098 | N/A | N/A |
| | FTM | 1071 | 1393 | 673 | 983 | 1908 | 1308 | N/A | N/A |
| | MLres | 355 | 424 | 460 | 684 | 374 | 570 | N/A | N/A |
| | FTMres | 690 | 838 | 673 | 907 | 548 | 808 | N/A | N/A |
| Pathway Enrichment | Both ML and MLres | 157 | 225 | 187 | 352 | 197 | 337 | 472 | 746 |
| | Both FTM and FTMres | 393 | 527 | 293 | 472 | 325 | 507 | 869 | 1251 |

In terms of positive correlation, the most significant common pathways were "structural constituent of ribosome" (enrichment p value = $2.67\times10^{-9}$ in brain, $1.48\times10^{-3}$ in kidney, $4.99\times10^{-22}$ in liver) and "translation" (p value = $2.80\times10^{-7}$ in brain, $2.70\times10^{-3}$ in kidney, $7.54\times10^{-20}$ in liver), consisting of ~40-50 ribosomal proteins (both large subunits; e.g. *Rpl3, Rpl6, Rpl8, Rpl10a, Rpl11, Rpl13, Rpl18a, Rpl22, Rpl23, Rpl24, Rpl26; Rpl28; Rpl30, Rpl31, Rpl32, Rpl35a, Rpl37, Rpl37a, Rpl38, Rpl39*; and small subunits, e.g. *Rps5, Rps9, Rps11, Rps15, Rps15a, Rps16, Rps21, Rps27*) (Figure 8.5A) and mitochondrial ribosomal proteins (large subunits such as *Mrpl20, Mrpl21, Mrpl27, Mrpl37, Mrpl41, Mrpl43*; and small subunits such as *Mrps2, Mrps11, Mrps12, Mrps15, Mrps24*) (Figure 8.5B). Ribosomal proteins are key components of ribosomes, the cellular machinery for protein synthesis, and higher expression of ribosomal proteins may imply adjusted levels of protein synthesis, larger amounts of ribosomes, and/or greater turn-over of ribosomes in the longer-lived species. Interestingly, among the genes with negative correlation to longevity were those involved in "ubiquitin-protein transferase activity" (p value = 6.87×10-4 in brain, 7.75×10-3 in kidney, 8.35×10-5 in liver) and "ubiquitin protein ligase activity" (p value = 3.92×10-6 in liver) (Figure 8.4). These processes included a number of genes coding for ubiquitin protein ligases (e.g. *Ube2a, Ube2g1, Ube2k, Ube2s, Ube2w, Ube3b, Ube4a, Ube4b, Ubr1, Ubr3, Ubr4, Ubr5, Smurf2*), kelch-like proteins (*Klkl13, Klhl17, Klhl24, Klhl38*), and ring finger proteins (*March2, March5, March6, Rnf4, Rnf10, Rnf13, Rnf115, Rnf139, Rnf144b*) (Figure 8.5C), which are involved in protein degradation. In addition, the genes involved in "base excision repair" (e.g. *Ogg1* coding for 8-oxoguanine DNA-glycosylase 1; *Mpg* coding for N-methylpurine-DNA glycosylase; *Rpa2* coding for replication protein

A2; *Sirt6* coding for sirtuins 6) were up-regulated in liver of the long-lived species, whereas those in "tricarboxylic acid cycle" (e.g. aconitase *Aco1, Aco2*; isocitrate dehydrogenase *Idh1, Idh2, Idh3a*; succinate dehydrogenase complex *Sdha, Sdhb*; succinate-CoA ligase *Sucla2, Suclg1*) were down-regulated in kidney of these species (Figure 8.5D).

Together, the results suggested that, among the long-lived species, protein degradation activities decreased whereas the transcription of ribosomal proteins increased. This stood in contrast with a number of lifespan experiments in model organisms, wherein depletion of ribosomal proteins was found to extend lifespan in yeast (Heeren et al., 2009; Steffen et al., 2008), and activation of proteasome or autophagy could extended lifespan in *C. elegans* (Chondrogianni et al., 2015; Ghazi et al., 2007), yeast (Kruegel et al., 2011), and fruit flies (Simonsen et al., 2008). On the other hand, analysis of gene expression in primary fibroblasts of 15 species of rodents, bats, and shrew also found that the levels of genes involved in ubiquitin-mediated protein degradation were lower in the long-lived species than in the short-lived ones, whereas the genes involved in DNA repair were high in the long-lived species (Chapter 7). Similar relationships were also observed in the previous gene expression study involving a smaller subset of mammalian species (Fushan et al., 2015). Furthermore, a recent study comparing gene expression in the muscle tissues of hibernating black bears and arctic ground squirrels with the summer active animals revealed a similar signature of up-regulation of ribosomal proteins and down-regulation of oxidation-reduction and glucose metabolism during hibernation (Fedorov et al., 2014). While more data would be required to guide interpretation, one might speculate that the longer-lived species exhibit lower levels of protein degradation, probably due to higher accuracy of protein synthesis, lower level of damage, or better maintenance of the translation machinery.

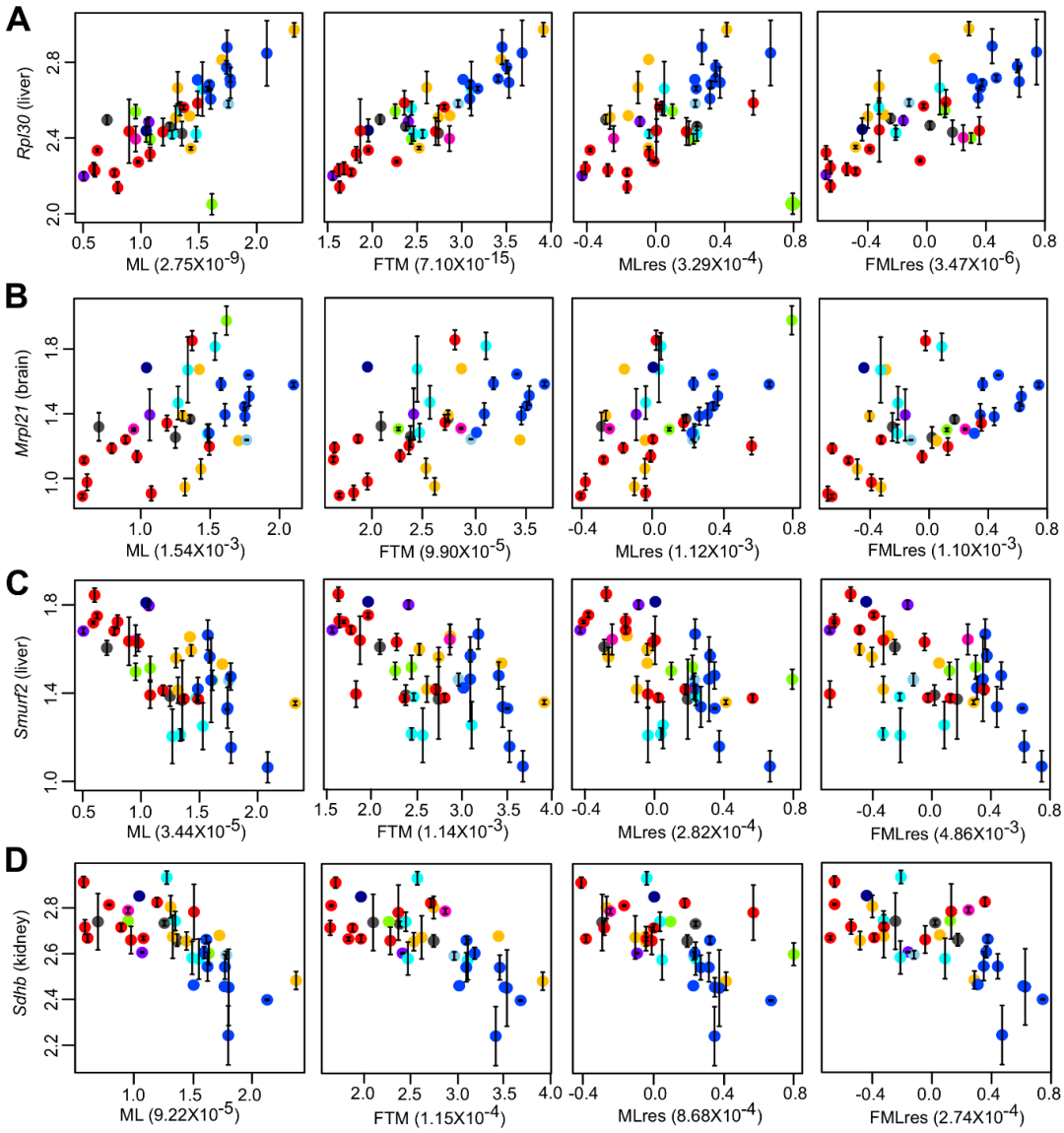



**Figure 8.5. Selected genes with significant correlation with longevity.**

**(A) Ribosomal protein large subunit *Rpl30* and (B) mitochondrial ribosomal protein large subunit *Mrpl21* show positive correlation with the longevity traits. (C) E3 ubiquitin ligase *Smurf2* and (D) succinate dehydrogenase complex subunit *Sdhb* show negative correlation with the longevity traits.** In each plot, the vertical axis indicates the average log10 expression value in the organ; the horizontal axis indicates the log10 longevity traits (ML: Maximum Lifespan; FTM: Female Time to Maturity; MLres: Maximum Lifespan Residual; FTMres: Female Time to Maturity Residual); the numbers in parenthesis indicate p value.robust. The error bars indicate standard error of mean. The points are colored by taxonomic group (same color scheme as in Figure 8.1).

## Gene expression and metabolic pathways

Previously, we performed metabolic profiling on a subset (26 species) of the species in this study and reported a number of metabolites with signficant correlation with species longevity (Ma et al., 2015a). In particular, we observed that the urate:allantoin ratio in kidney showed positive correlation with the longevity traits, suggesting that the long-lived species had higher urate and/or lower allantoin levels. Furthermore, the liver levels of two tryptophan degradation products, kynurenine and anthranilic acid, as well as the kynurenine:tryptophan ratio and the anthranilic acid:tryptophan ratio, showed negative correlation with longevity. Here, we related these metabolic changes with gene expression patterns.

In mammals, urate is produced from the degradation of purines and is then converted to allantoin by the enzyme uric acid oxidase (also known as uricase) and secreted into urine (Ngo and Assimos, 2007) (Figure 8.6A). In humans and other hominoids, the gene coding for uric oxidase is a pseudogene and urate is excreted instead (Oda et al., 2002; Wu et al., 1989; Wu et al., 1992). Our analyses showed that the expression levels of *Uox* in liver correlated negatively with FTM and FTMres (p value.robust = $5.26\times10^{-3}$ for FTM and $1.94\times10^{-3}$ for FTMres; Figure 8.6A), and to a lesser extent, with ML and MLres (p value.robust = 0.021 for ML and 0.013 for MLres). As expected, the read counts in human, bonobo, orangutan, chimpanzee and gorilla were all much lower (~10-1000 times) than those of other primate species (including macaque, rhesus monkey, vervet, and baboon), suggesting the gene (in the form of pseudogene) was not well transcribed among the hominoids. Consistent with our previous observations (Ma et al., 2015a), *Uox* expression was also low in naked mole rats and Damaraland mole rats (Figure 8.6A).

In terms of the tryptophan degradation products, we did not observe longevity correlation for the genes coding for tryptophan 2,3-dioxygenase (TDO) and indoleamine 2.3-dioxygenase (IDO), the enzymes that mediate the first step of degradation. On the other hand, quinolinic acid, which is at the end of the degradation pathway, can be converted to nicotinamide adenine dinucleotide (NAD) via a number of intermediates (Figure 8.6B). It was observed previously in mouse muscles that $NAD^+$ levels declined with age and affected SIRT1 functions (which consumed NAD in deacetylation),

whereas supplementation with nicotinamide mononucleotide (NaMN) reversed the aging phenotypes (Gomes et al., 2013). Overexpression of SIRT1 in mouse brain could protect against aging-dependent circadian changes (Chang and Guarente, 2013). Here, we observed that the gene expression levels of *Qprt* (coding for quinolinic acid phosphoribosyltransferase) and *Nadsyn1* (coding for NAD synthetase 1) in brain correlated positively with FTM (p value.robust = $6.84\times10^{-3}$ for *Qprt* and $4.58\times10^{-3}$ for *Nadsyn1*; Figure 8.6B), but the correlations were generally weak with the other longevity traits and in the other organs (*NMNAT1* did not show signficant correlation either). More data would be needed to determine whether the species longevity variation was linked to NAD synthesis.

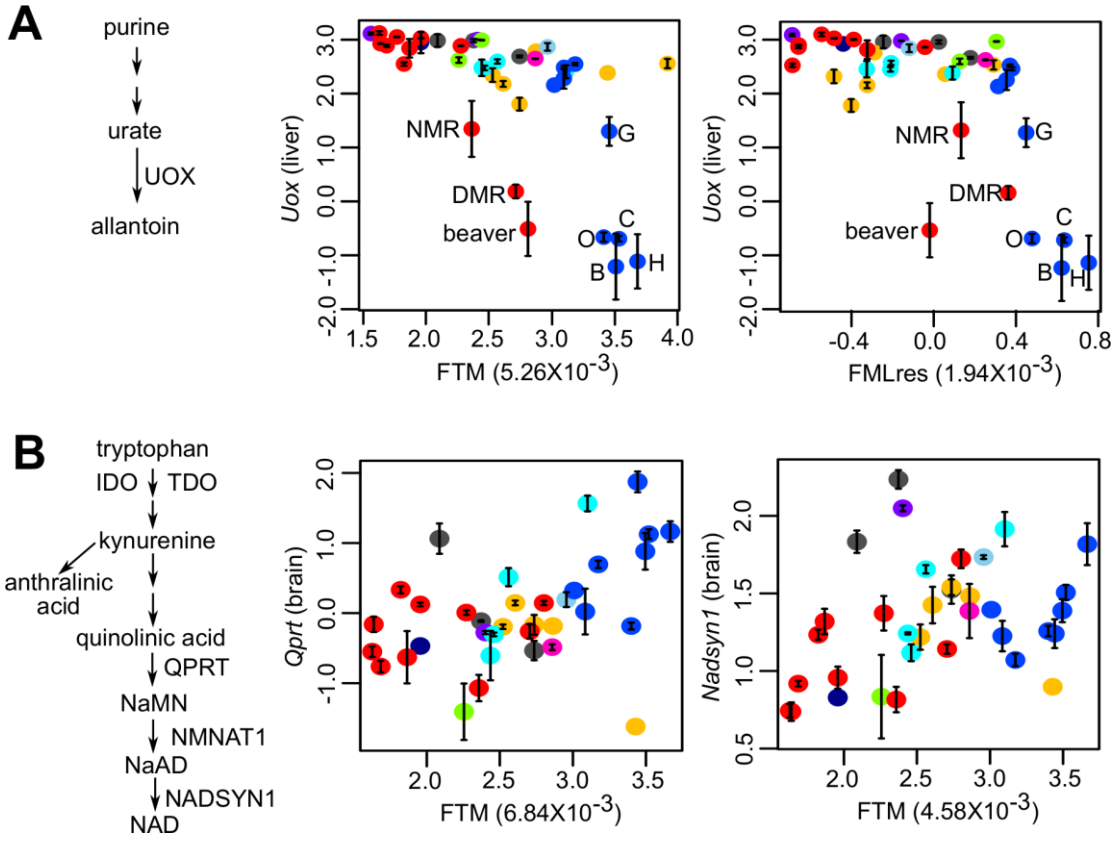



**Figure 8.6. Expression of genes in longevity-related metabolic pathways.**

**(A) Urate oxidase (*Uox*) expression in liver correlates negatively with longevity.** The hominoids are indicated (H: human; C: chimpanzee; O: orangutan; B: bonobo; G: gorilla). The points representing naked mole rat (NMR), Damaraland mole rat (DMR) and beaver are indicated.

**(B) Quinolinate phosphoribosyl transferase (*Qprt*) and NAD synthetase 1 (*Nadsyn1*) expression in brain correlates positively with female time to maturity (FTM).**

# REFERENCES


Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Chang, H.C., and Guarente, L. (2013). SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. Cell *153*, 1448-1460.

Chondrogianni, N., Georgila, K., Kourtis, N., Tavernarakis, N., and Gonos, E.S. (2015). 20S proteasome activation promotes life span extension and resistance to proteotoxicity in Caenorhabditis elegans. FASEB J. *29*, 611-622.

Fedorov, V.B., Goropashnaya, A.V., Stewart, N.C., Toien, O., Chang, C., Wang, H., Yan, J., Showe, L.C., Showe, M.K., and Barnes, B.M. (2014). Comparative functional genomics of adaptation to muscular disuse in hibernating mammals. Molecular ecology *23*, 5524-5537.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H.*, et al.* (2015). Gene expression defines natural changes in mammalian lifespan. Aging Cell *14*, 352-365.

Ghazi, A., Henis-Korenblit, S., and Kenyon, C. (2007). Regulation of Caenorhabditis elegans lifespan by a proteasomal E3 ligase complex. Proc. Natl. Acad. Sci. USA *104*, 5947-5952.

Gomes, A.P., Price, N.L., Ling, A.J., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P.*, et al.* (2013). Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. Cell *155*, 1624-1638.

Heeren, G., Rinnerthaler, M., Laun, P., von Seyerl, P., Kossler, S., Klinger, H., Hager, M., Bogengruber, E., Jarolim, S., Simon-Nobbe, B.*, et al.* (2009). The mitochondrial ribosomal protein of the large subunit, Afo1p, determines cellular longevity through mitochondrial back-signaling via TOR1. Aging (Albany NY) *1*, 622-636.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. *20*, 1313-1326.

Khaitovich, P., Enard, W., Lachmann, M., and Paabo, S. (2006). Evolution of primate gene expression. Nat. Rev. Genet. *7*, 693-702.

Kruegel, U., Robison, B., Dange, T., Kahlert, G., Delaney, J.R., Kotireddy, S., Tsuchiya, M., Tsuchiyama, S., Murakami, C.J., Schleit, J.*, et al.* (2011). Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLoS Genet. *7*, e1002253.

Ma, S., Lee, S.G., Kim, E.B., Park, T.J., Seluanov, A., Gorbunova, V., Buffenstein, R., Seravalli, J., and Gladyshev, V.N. (2015a). Organization of the Mammalian Ionome According to Organ Origin, Lineage Specialization, and Longevity. Cell Rep. *13*, 1319-1326.

Ma, S., Yim, S.H., Lee, S.G., Kim, E.B., Lee, S.R., Chang, K.T., Buffenstein, R., Lewis, K.N., Park, T.J., Miller, R.A.*, et al.* (2015b). Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity. Cell Metab. *22*, 332-343.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593-1599.

Ngo, T.C., and Assimos, D.G. (2007). Uric Acid nephrolithiasis: recent progress and future directions. Rev. Urol. *9*, 17-27.

Oda, M., Satta, Y., Takenaka, O., and Takahata, N. (2002). Loss of urate oxidase activity in hominoids and its evolutionary implications. Mol. Biol. Evol. *19*, 640-653.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Simonsen, A., Cumming, R.C., Brech, A., Isakson, P., Schubert, D.R., and Finley, K.D. (2008). Promoting basal levels of autophagy in the nervous system enhances longevity and oxidant resistance in adult Drosophila. Autophagy *4*, 176-184.

Steffen, K.K., MacKay, V.L., Kerr, E.O., Tsuchiya, M., Hu, D., Fox, L.A., Dang, N., Johnston, E.D., Oakes, J.A., Tchao, B.N.*, et al.* (2008). Yeast life span extension by depletion of 60s ribosomal subunits is mediated by Gcn4. Cell *133*, 292-302.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-1033.

Wu, X.W., Lee, C.C., Muzny, D.M., and Caskey, C.T. (1989). Urate oxidase: primary structure and evolutionary implications. Proc. Natl. Acad. Sci. USA *86*, 9412-9416.

Wu, X.W., Muzny, D.M., Lee, C.C., and Caskey, C.T. (1992). Two independent mutational events in the loss of urate oxidase during hominoid evolution. J. Mol. Evol. *34*, 78-84.

# Chapter 9 Conclusions

*"Today is the oldest you've ever been*

*and the youngest you'll ever be again."*

*-- Eleanor Roosevelt*

Photo credit: Dr. Alaattin Kaya.

## SUMMARY OF MAIN FINDINGS

Since their decent from a common ancestor ~210 million years ago, mammals have undergone remarkable diversification in body size, morphology, diet, habitat, reproduction strategies, and lifespan. From the small shrews weighing under 2 grams and living about 3 years to the gigantic bowhead whales weighing over 100 tons and surviving more than 2 centuries, evolution has acted like a Master Scientist to bring about so many species with wide variation in their life history traits. Just as much of our biology knowledge has come from studying different model organism mutants in the laboratory, we may also be able to gain a better understanding of the control of lifespan by examining natural variation across different species. Much of the prior effort was limited by technical constraints and focused on a few traits across a small set of species. As the high throughput 'omics approaches have become technically accessible and cost-effective, we now have the opportunity to examine the full transcriptomes, metabolomes, ionomes, and proteomes, across multiple organs and across a significant number of species. This dissertation presents a few of such snapshots – they illustrate the highly complex landscape of changes associated with species longevity (Table 9.1) and may inform the future research with regard to the strategies for lifespan control. In addition, the organ- and lineage-specific patterns of gene expression, metabolite, and element distribution were examined (Tables 9.2 and 9.3) – some are consistent with known physiology of the tissues and animals, while other patterns can offer new insights into various aspects of biology.

Moving from the short-lived mammals to the long-lived ones, the general trends of larger body mass, longer female time to maturity, fewer offspring with better survival rates, reduced damage generation, and enhanced repair and maintenance capacities are well known (Adelman et al., 1988; Barja and Herrero, 2000; Cortopassi and Wang, 1996; Peters, 1986; Sacher, 1959; Western, 1979). In our work, we identified the corresponding changes on the molecular levels, including higher expression of genes involved in DNA repair (liver and fibroblast), up-regulation of ribosomal proteins (brain and liver), down-regulation of protein ubiquitination pathways (brain, kidney, liver, and fibroblast), lower expression of genes involved in apoptosis (fibroblast), and reduced levels of oxidation-prone triacylglycerols with polyunsaturated fatty acid side chains (PUFA TAGs; kidney).

Other molecular signatures previously implicated in aging and longevity studies, such as high sphingomyelins (SM; brain, heart, kidney, and liver), high expression of NAD synthase 1 (*NADSYN1*) (brain and fibroblast), low lysophosphatidylcholine (LPC; brain, heart, kidney) and lysophosphatidylethanolamine (LPE; brain, kidney), low tryptophan degradation products (liver), and high urate and low allantoin levels (brain, kidney, liver), were also observed (Capuron et al., 2011; Cutler, 1984; De Marte and Enesco, 1986; Frick et al., 2004; Gonzalez-Covarrubias et al., 2013; Kaur and Halliwell, 1990; Rosenson and Stafforini, 2012; Segall and Timiras, 1976). Overall, the patterns suggest a set of coordinated changes across multiple organs and tissues, underlying the theme of better maintenance and repair and less damage generation among the long-lived mammals. In terms of the non-mammalian datasets, the *Drosophila* data highlighted the key roles of fatty acid metabolism in lifespan variation, whereas the yeast data suggested that the replicative lifespan was influenced by mitochondria composition rather than by mitochondria number. Furthermore, a number of these longevity-associated genes in *Drosophila* were also shown in previous experiments to affect lifespan, suggesting our method has the potential to identify new nodes of lifespan regulation.

**Table 9.1. Summary of the molecular patterns and signatures associated with longevity.**

| Dataset | Positive correlation with longevity | Negative correlation with longevity |
|---|---|---|
| **Mammalian Metabolome** | • ↑ sphingomyelin (brain, heart, kidney, liver)<br>• ↑ urate:allantoin ratio (kidney, liver) | • ↓ polyunsaturated triacylglycerols (kidney)<br>• ↓ lysophosphatidylcholine (LPC; brain, heart, kidney)<br>• ↓ lysophosphatidylethanolamine (LPE; brain, kidney)<br>• ↓ amino acids (brain)<br>• ↓ allantoin (brain, kidney, liver)<br>• ↓ kynurenine:tryptophan ratio (liver)<br>• ↓ anthranilic acid:tryptophan ratio (liver) |
| **Long-lived Mouse Models Metabolome** | • ↑ sphingomyelin (liver)<br>• ↑ methionine (brain) | • ↓ polyunsaturated triacylglycerols (liver) |
| | Note:<br>• Brain metabolites remain largely unchanged under caloric restriction (CR), acarbose treatment (ACA), rapamycin treatment (RAP), and growth hormone receptor knock-out (GHRKO). Only Snell dwarf mice (Snell) exhibited significant changes in brain.<br>• CR and ACA produced very similar changes in liver metabolites but were rather different from the profiles of RAP and GHRKO. The profile of Snell shared similarity with CR and ACA. | |
| **Mammalian Ionome** | • ↑ zinc (kidney and liver; effects likely due to body mass differences)<br>• ↑ cadmium (kidney and liver) | • ↓ selenium (liver; weak effects) |

**Table 9.1. (Continued)**

| Dataset | Positive correlation with longevity | Negative correlation with longevity |
|---|---|---|
| **Fly Transcriptome** | • ↑ organic acid biosynthetic process<br>• ↑ fatty acid beta oxidation<br>• ↑ cofactor biosynthetic process<br>• ↑ glyoxylate and dicarboxylate metabolism | • ↓ neuron recognition / axongenesis<br>• ↓ synaptic transmission<br>• ↓ DNA binding<br>• ↓ protein kinase activity<br>• ↓ RNA processing |
| | Note:<br>18 of the genes correlating with maximum lifespan in fly were previously reported to affect lifespan; 14 of them had the expected direction of correlation (i.e. pro-longevity genes were expressed at high levels in long-lived *Drosophila* species; anti-longevity genes were expressed at low levels in long-lived *Drosophila* species) | |
| **Yeast Transcriptome and Proteome** | Gene expression:<br>• ↑ oxidative phosphorylation / aerobic respiration / mitochondrial respiratory chain<br>• ↑ ion transport<br>• ↑ hexose/glucose metabolic process<br>• ↑ actin binding<br><br>Mitochondrial composition:<br>• ↑ Pyruvate dehydrogenase complex<br>• ↑ Complex III<br>• ↑ Complex IV<br>• ↑ Mitochondrial ATP synthase<br>• ↑ Mitochondrial ribosomal proteins | Metabolites:<br>• ↓ asparagine, glutamine, 2-octenoic acid<br><br>Gene expression:<br>• ↓ protein targeting<br>• ↓ transcription regulation<br>• ↓ mRNA splicing/ ribonucleoprotein complex<br>• ↓ cytoskeleton organization<br>• ↓ nitrogen compound biosynthetic process<br><br>Mitochondrial composition:<br>• ↓ outer membrane translocases<br>• ↓ mitochondrial chaperonine<br>• ↓ metabolic enzymes |
| **Rodent Fibroblasts** | Gene expression:<br>• ↑ adenyl nucleotide binding<br>• ↑ DNA repair / DNA damage<br>• ↑ chromatin regulator / chromosome organization<br>• ↑ glucose / hexose metabolic process<br><br>Metabolites:<br>• ↑ amino acids | Gene expression:<br>• ↓ proteolysis / ubiquitin conjugation pathway<br>• ↓ protein transport<br>• ↓ transcription regulation<br><br>Metabolites:<br>• ↓ glycerophospholipids (LPC, LPE) |
| **Mammalian Transcriptome** | • ↑ structural constituent of ribosome (brain, kidney liver)<br>• ↑ translation (brain, kidney, liver)<br>• ↑ rRNA binding (brain, liver)<br>• ↑ RNA binding (kidney, liver)<br>• ↑ base-excision repair (liver)<br>• ↑ negative regulation of apoptotic process (liver) | • ↓ ubiquitin-protein transferase activity (brain, kidney, liver)<br>• ↓ protein polyubiquitination (brain, liver)<br>• ↓ protein serine/threonine kinase activity (brain, kidney)<br>• ↓ tricarboxylic acid cycle (kidney, liver)<br>• ↓ protein transport (kidney, liver)<br>• ↓ voltage-gated potassium channel activity (brain)<br>• ↓ neuromuscular junction development (brain)<br>• ↓ potassium / calcium ion transport (brain)<br>• ↓ unfolded protein binding (kidney) |

**Table 9.2. Organ-specific changes in metabolome, ionome, or transcriptome.** Red upward arrows denote relatively high level; blue downward arrows indicate relatively low level.

| Organ | Mammalian Metabolome | Mammalian Ionome | Mammalian Transcriptome |
|---|---|---|---|
| **Brain** | Metabolites diverge least<br>↑ glycerophospholipids<br>↑ sphingomyelins<br>↑ neurotransmitters | ↑ potassium (K)<br>↑ phosphorous (P) | ↑ synaptic transmission<br>↑ neuron differentiation |
| **Heart** | ↑ acylcarnitines<br>↑ triacylglycerols | | ↑ oxidative phosphorylation<br>↑ tricarboxylic acid cycle |
| **Kidney** | ↑ most amino acids<br>↓ glutamine | ↑ lithium (Li)<br>↑ sodium (Na)<br>↑ calcium (Ca)<br>↑ selenium (Se)<br>↑ cadmium (Cd) | ↑ ion transport<br>↑ sodium transport |
| **Liver** | ↑ amino acids<br>↑ glycerophospholipids<br>↑ carbohydrates<br>↑ steroids<br>↑ sucrose / lactose<br>↑ bile acid components | ↑ iron (Fe)<br>↑ molybdenum (Mo)<br>↑ manganese (Mn)<br>↑ selenium (Se)<br>↑ cadmium (Cd)<br>↑ copper (Cu)<br>↑ zinc (Zn) | ↑ steroid metabolism<br>↑ complement and coagulation cascades |
| **Testis** | | | Metabolites diverge most<br>↑ cell cycle<br>↑ spermatogenesis<br>↑ steroid metabolism<br>↑ sexual reproduction |

**Table 9.3. Lineage specific changes or correlation patterns.**

| Dataset | Lineage specific changes / Correlation patterns |
|---|---|
| **Mammalian Metabolome** | • ↓ methionine sulfoxide in bats<br>• ↓ allantoin and ↓ urate oxidase in African mole rats<br>• Bile acid conjugation consistent with diet preference: carnivores prefer taurine-conjugation; herbivores prefer glycine-conjugation |
| **Mammalian Ionome** | • Divalent ions such as iron, molybdenum, manganese, nickel and cobalt have similar distribution patterns<br>• Copper and zinc have similar distribution patterns<br>• ↑ magnesium, ↑ sulfur, ↑ phosphorus, and ↑ potassium in brain of rodents<br>• ↓ selenium in kidney and liver of African mole rats<br>• The numbers of selenocysteine residues in selenoprotein P of the species agree with their kidney and liver selenium levels. |
| **Yeast Transcriptome and Proteome** | • Pathways that are differentially expressed across the yeast strains:<br>  ○ oxidative phosphorylation<br>  ○ ribosome biogenesis<br>  ○ ribonucleoprotein complex<br>  ○ protein synthesis<br>  ○ translation regulation. |
| **Fly Transcriptome** | • Pathways that are differentially expressed across the *Drosophila* species:<br>  ○ oxidative phosphorylation<br>  ○ electron transport chain, NADH dehydrogenase activity<br>  ○ structural constituents of ribosomes<br>  ○ transcription regulation<br>  ○ amino acid biosynthesis.<br>• The expression divergence of the flies indicate strong stabilizing selection. |

## WITHIN-SPECIES VS. CROSS-SPECIES LIFESPAN VARIATIONS

In Chapter 1 of this dissertation, a distinction was made between the pathways underlying lifespan extensions within single species and those underlying lifespan changes across different species, and questions were raised regarding the extent of their overlap. With hindsight, this issue can now start to be addressed.

Comparison with the metabolic profiles of the 5 long-lived mouse models suggested that a number of changes (including high SMs and low PUFA TAGs) were common in both within-species and cross-species lifespan variations. The up-regulation of DNA repair pathways among long-lived mammals is also a key feature of some of the known within-species lifespan extension strategies. On the other hand, several cross-species longevity signatures seem to contradict some of the established paradigms of within-species lifespan extension. For example, autophagy is required for the lifespan extension via dietary restriction and subdued IGF-1 signaling (Chondrogianni and Gonos, 2008; Hansen et al., 2008; Kenyon, 2010; Kevei and Hoppe, 2014; Low, 2011; Melendez et al., 2003; Rubinsztein et al., 2011; Starke-Reed and Oliver, 1989; Vernace et al., 2007) and activation of proteasome or autophagy can extend lifespan in *C. elegans*, yeast, and flies (Chondrogianni et al., 2015; Ghazi et al., 2007; Kruegel et al., 2011; Simonsen et al., 2008), yet the genes involved in proteolysis and ubiquitination were found to be expressed at lower levels among long-lived species. Depletion of ribosomal proteins in yeast has lifespan extension effect (Heeren et al., 2009; Steffen et al., 2008), but the long-lived mammals also express high levels of ribosomal proteins. Based on these results, it seems that there are indeed some overlaps between the within-species lifespan extension pathways and the cross-species longevity signatures, but there are also clear differences in many other aspects. Although up-regulation of genes involved in DNA repair pathways are observed in both cases, more data will be needed to clarify whether they are brought about by the same mechanism. For example, the up-regulation of DNA repair for within-species lifespan extension is likely controlled by transcription factors such as FOXO (Tran et al., 2002) and is largely a temporary response to dietary restriction or drug treatment, whereas the high expression of DNA repair genes in long-lived species may be the result of more permanent changes (e.g. changes in promoter sequences).

Importantly, the results here also suggest the various within-species lifespan extension strategies may act through different molecular pathways and have different effects on the organs. The metabolic profiles of the 5 long-lived mouse models reveal that caloric restriction and acarbose treatment produce similar metabolic shifts, whereas rapamycin treatment, growth hormone receptor knock-out, and Snell dwarf mice seem to affect different sets of metabolic pathways. It was previously shown that caloric restriction could extend the lifespan of both wild-type mice and long-lived Ames dwarf mice (Bartke et al., 2001), even though the gene expression changes were rather divergent (Masternak et al., 2004), indicative of potentially different pathways of lifespan controls. Furthermore, although there were extensive metabolic shifts in liver of the long-lived mouse models, only the Snell dwarf mice (which have ablation of anterior pituitary) exhibited significant changes in brain metabolite levels, suggesting the brain metabolites might be more stable and refractory to fluctuation than the liver metabolites. At the moment, most of the gene expression and metabolite data of lifespan extension strategies are reported independently by different research groups, often using different animal models and in different organs and tissues. While some attempts were made to perform meta-analysis on these datasets to identify the common features (de Magalhães et al., 2009; Swindell, 2007), it was rather difficult to distinguish the variations due to actual biological differences of the various lifespan extension pathways, from those due to different genetic backgrounds and experimental set ups. The organ context of the data was not sufficiently explored either. On the other hand, by subjecting the mice of identical genetic background to different lifespan extension treatments within the same experiment set up and collecting the data across multiple organs, we were able to directly observe and compare the different biological effects of these treatments. In future, it will be helpful to measure the gene expression profiles of these mouse models, so as to systematically assess the extent of overlap among these lifespan extension pathways, as well as with the cross-species longevity signatures.

Conceptually, it is possible, and even probable, that within-species lifespan extension and cross-species lifespan variation are governed by different mechanisms: many of the changes observed in within-species lifespan extension (e.g. up-regulation of anti-oxidant enzymes, up-regulation of autophagy, suppression of anabolic metabolism and reproduction, and induction of stress responses)

are suited for dealing with acute stress and short-term adversity, whereas the long-term, sustained alteration of lifespan across species may involve other permanent and coordinated modifications. Furthermore, many of the long-lived models produced in laboratory settings may not be fit to compete in the wild, whereas the naturally long-lived species are well adapted to their respective ecological niches. Furthermore, some of the cross-species trends may not apply to the within-species comparison. While larger mammals are known to live longer than small mammals, within species patterns are often opposite, e.g. smaller dogs live longer than large dogs, and dwarf mice live longer than control mice. Therefore, the validity of these observations here may be dependent on the levels of comparison, and further research efforts will be needed to determine where the line can be drawn. It should be noted that the key pathways for within-species lifespan extension have been experimentally validated in yeast, flies, worms, and mice, whereas the cross-species longevity signatures identified here are yet verified (except the "natural evolutionary experiment" that gave rise to these species). It cannot be excluded that the cross-species signatures represent such fundamental re-wiring and re-programming of the biological systems that they will not be applicable within the context of a single species, at least using the currently available methods.

## POTENTIAL PITFALLS

While the molecular patterns and signatures of longevity presented here satisfy stringent statistical criteria, one should keep in mind of several potential pitfalls when interpreting the data.

First, all these relationships are based solely on correlation and do not prove causation. On the one hand, there are consistent and coordinated changes in gene expression and metabolite profiles in relation to longevity variation, and they are likely modulated by the same set of underlying forces and factors. Some of these changes may be the result, and not the cause, of a long lifespan, though it is often difficult to differentiate between the two. For example, we found that the levels of cadmium in brain, heart, kidney and liver showed strong positive correlation with longevity, and the relationship was not affected by differences in body mass. Since cadmium is a heavy metal, is toxic at high levels, and plays no known biological functions in mammals, it is hard to perceive the potential longevity benefit of high cadmium. Rather, the observation may simply reflect the passive accumulation of cadmium from food intake, as the longer-lived mammals consume a greater amount of food over their life time. Further experimental evidence will be required to pinpoint the causality.

Second, although the enrichment analyses provide useful summaries of the major pathways related to longevity variation, more detailed and careful interpretation of the results will be needed to fully delineate the principles governing lifespan regulation. For example, while "lipid metabolic process" was shown to have positive correlation with *Drosophila* lifespan (Chapter 6), it is important to differentiate the genes involved in fatty acid beta-oxidation from those involved in fatty acid synthesis. Although as a group the PUFA TAGs showed significant negative correlation to longevity, individually they have different biological functions depending on the numbers and positions of the double bonds. While many ribosomal proteins showed positive correlation with longevity, it may be useful to carefully examine and map the ribosomal proteins on the ribosome structure. If the identified proteins are mostly in the external shell, then they may indicate improved accuracy of protein synthesis. Beside the annotated pathways in the databases, the roles of the various metabolites should also be viewed in terms of the respective organs and biological origins, as the same metabolite (e.g. methionine) may also affect very diverse functions depending on organ and tissue. The levels of

trimethylamine N-oxide (TMAO) in brain and heart showed positive correlation with longevity, but it is also a product of gut microbiome, and such aspects deserve further exploration. Additionally, epigenetic changes such as DNA methylation and histone modifications also show important age- and longevity-related trends, but they are not yet reflected in the current datasets.

Third, the results should be viewed in the light of possible sample bias and variation. The gene expression and metabolite data presented here are cross-sectional in nature: they represent a snapshot measurement across various species at a single point in time. While they provide very useful information on the cross-species variation, they cannot inform us on the within-species variation, e.g. due to differences in fed or fasting state, diets, and genders. In many long-lived animal models, it has been shown that the dietary restriction and other lifespan extension strategies are often more effective in females than in males (Burger and Promislow, 2004; Magwere et al., 2004; Miller et al., 2014), although the mammalian and fly data presented here came mostly from males. It was also pointed out that the gene expression profiles of the primary fibroblasts might be confounded by culture conditions, as the fibroblasts from the long-lived species might survive better in cell culture than those from the short-lived species, and the observed expression patterns reflect more of the cellular states in culture than the cellular states *in vivo*. As for the *Drosophila* dataset, the different species of flies were fed a diet optimal for *D. melanogaster*, which might not be ideal for the non-conventional species. While we have shown that the variation due to biological replicates within a species is generally much smaller than the variation in the longevity traits across the species, it will be very informative to collect more samples to formally assess the variation in gene expression and metabolites due to differences in gender, diet, feeding state, and cell culture conditions.

# FUTURE DIRECTIONS

Overall, these studies provide the starting points for discovering and exploring new mechanisms and paradigms of lifespan regulation. To experimentally apply these findings, one can make use of the Transgenic RNAi Project (TRiP) library of *D. melanogaster* to test the longevity effects of knocking out specific genes of interest (Ni et al., 2008). The recently developed CRISPR-Cas system can also be used to generate more stable deletion lines and ameliorate the off-target effects of RNAi (Jinek et al., 2012; Sander and Joung, 2014). Most importantly, one may query the Connectivity Map Database to identify small chemical molecules capable of producing the desired molecular signatures (Lamb et al., 2006) and test their longevity effects on laboratory animals (Calvert et al., 2016). A similar approach may involve genetic manipulations as well as screens. For example, resemblance of the gene expression pattern resulting from gene knockout to our evolutionary signature will suggest that this gene knockout is a candidate for lifespan extension. Due to the inter-connected and systemic nature of lifespan regulation, one may need to simultaneously perturb multiple pathways using a combination of gene manipulation and small pharmacological molecules to bring about the optimal lifespan extension effects. As recently demonstrated in *C. elegans*, targeting 3 network modules of dietary restriction simultaneously could produce exceptionally long-lived animals (Hou et al., 2016).

At the moment, it is not known whether the different species follow the same trajectory of aging (i.e. age at the same rate). For example, it was shown in *C. elegans* that interventions as diverse as changes in diet, temperature, exposure to oxidative stress, and disruption of genes all altered lifespan distributions by an apparent stretching or shrinking of time (Stroustrup et al., 2016). In addition to the cross-sectional data presented here, one should also obtain longitudinal data from animals of different ages, to assess the gene expression and metabolite changes from young to old and determine whether they are related to the cross-species longevity signatures identified here. It will also be useful to obtain samples from animals of different genders and feeding states, to formally determine the degree of data variation. A number of studies in plants (Watanabe et al., 2007; White et al., 2012) showed that when the same plant species were collected under different fertilizer treatments

and environmental conditions, the concentrations of Ca, Zn, Mn and Mg were found to remain more closely linked to phylogeny, but Cu and Fe varied more strongly with environmental factors, suggesting different elements might be more or less responsive to external variations. Similarly, some metabolites and genes in mammals may be more resistant to dietary and gender variation than others and may therefore serve as more useful and stable longevity signatures.

As high throughput methods become increasingly accessible, we will likely see more and more cross-species expression and metabolomics studies. These studies bring with them their own set of unique challenges, including the phylogenetic relationship of samples and identification of accurate ortholog sequences (often without the reference genomes available). The phylogenetic regression method was originally developed by evolutionary biologists to study co-evolution of a few traits. Here, we applied it on hundreds of metabolites and tens of thousands of genes, and although we adjusted for false discovery rate using the conventional approach, more careful thoughts shall be given to better understand the issue of multiple testing. Furthermore, the phylogenetic regression method here assumes normal distribution (which was valid for ~80% of all genes and metabolites on log scale), but other distribution models (e.g. negative binomial) may be more appropriate for RNAseq data, so there will be much room for improvement of the phylogenetic regression model. In terms of ortholog identification, we assembled our pipeline using a number of publicly available software and packages. Development of a single tool, with scalable inputs, standardized outputs, and more robust quality controls will significantly facilitate cross-species comparative studies.

Lastly, it will also be fruitful to obtain the gene expression profiles of various long-lived mouse models, and examine them against each other as well as against the cross-species longevity signatures identified here. It may be possible that the longevity signatures will be represented by multiple dietary, pharmacological and/or genetic interventions, which again will suggest that such a combination may result in lifespan extension. Ultimately, these approaches should lead to defining the gene expression space (and similarly the metabolite space) that is associated with the most robust longevity effects.

# REFERENCES


Adelman, R., Saul, R.L., and Ames, B.N. (1988). Oxidative damage to DNA: relation to species metabolic rate and life span. Proc. Natl. Acad. Sci. USA *85*, 2706-2708.

Barja, G., and Herrero, A. (2000). Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. FASEB J. *14*, 312-318.

Bartke, A., Wright, J.C., Mattison, J.A., Ingram, D.K., Miller, R.A., and Roth, G.S. (2001). Extending the lifespan of long-lived mice. Nature *414*, 412.

Burger, J.M., and Promislow, D.E. (2004). Sex-specific effects of interventions that extend fly life span. Sci Aging Knowledge Environ *2004*, pe30.

Calvert, S., Tacutu, R., Sharifi, S., Teixeira, R., Ghosh, P., and de Magalhaes, J.P. (2016). A network pharmacology approach reveals new candidate caloric restriction mimetics in C. elegans. Aging Cell *15*, 256-266.

Capuron, L., Schroecksnadel, S., Feart, C., Aubert, A., Higueret, D., Barberger-Gateau, P., Laye, S., and Fuchs, D. (2011). Chronic low-grade inflammation in elderly persons is associated with altered tryptophan and tyrosine metabolism: role in neuropsychiatric symptoms. Biol. Psychiatry *70*, 175-182.

Chondrogianni, N., Georgila, K., Kourtis, N., Tavernarakis, N., and Gonos, E.S. (2015). 20S proteasome activation promotes life span extension and resistance to proteotoxicity in Caenorhabditis elegans. FASEB J. *29*, 611-622.

Chondrogianni, N., and Gonos, E.S. (2008). Proteasome activation as a novel antiaging strategy. IUBMB life *60*, 651-655.

Cortopassi, G.A., and Wang, E. (1996). There is substantial agreement among interspecies estimates of DNA repair activity. Mech. Ageing Dev. *91*, 211-218.

Cutler, R.G. (1984). Urate and ascorbate: their possible roles as antioxidants in determining longevity of mammalian species. Arch. Gerontol. Geriatr. *3*, 321-348.

de Magalhães, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics *25*, 875-881.

De Marte, M.L., and Enesco, H.E. (1986). Influence of low tryptophan diet on survival and organ growth in mice. Mech. Ageing Dev. *36*, 161-171.

Frick, B., Schroecksnadel, K., Neurauter, G., Leblhuber, F., and Fuchs, D. (2004). Increasing production of homocysteine and neopterin and degradation of tryptophan with older age. Clin. Biochem. *37*, 684-687.

Ghazi, A., Henis-Korenblit, S., and Kenyon, C. (2007). Regulation of Caenorhabditis elegans lifespan by a proteasomal E3 ligase complex. Proc. Natl. Acad. Sci. USA *104*, 5947-5952.

Gonzalez-Covarrubias, V., Beekman, M., Uh, H.W., Dane, A., Troost, J., Paliukhovich, I., van der Kloet, F.M., Houwing-Duistermaat, J., Vreeken, R.J., Hankemeier, T.*, et al.* (2013). Lipidomics of familial longevity. Aging Cell *12*, 426-434.

Hansen, M., Chandra, A., Mitic, L.L., Onken, B., Driscoll, M., and Kenyon, C. (2008). A role for autophagy in the extension of lifespan by dietary restriction in C. elegans. PLoS Genet. *4*, e24.

Heeren, G., Rinnerthaler, M., Laun, P., von Seyerl, P., Kossler, S., Klinger, H., Hager, M., Bogengruber, E., Jarolim, S., Simon-Nobbe, B.*, et al.* (2009). The mitochondrial ribosomal protein of the large subunit, Afo1p, determines cellular longevity through mitochondrial back-signaling via TOR1. Aging (Albany NY) *1*, 622-636.

Hou, L., Wang, D., Chen, D., Liu, Y., Zhang, Y., Cheng, H., Xu, C., Sun, N., McDermott, J., Mair, W.B.*, et al.* (2016). A Systems Approach to Reverse Engineer Lifespan Extension by Dietary Restriction. Cell Metab. *23*, 529-540.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kaur, H., and Halliwell, B. (1990). Action of biologically-relevant oxidizing species upon uric acid. Identification of uric acid oxidation products. Chem. Biol. Interact. *73*, 235-247.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Kevei, E., and Hoppe, T. (2014). Ubiquitin sets the timer: impacts on aging and longevity. Nat. Struct. Mol. Biol. *21*, 290-292.

Kruegel, U., Robison, B., Dange, T., Kahlert, G., Delaney, J.R., Kotireddy, S., Tsuchiya, M., Tsuchiyama, S., Murakami, C.J., Schleit, J.*, et al.* (2011). Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLoS Genet. *7*, e1002253.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N.*, et al.* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929-1935.

Low, P. (2011). The role of ubiquitin-proteasome system in ageing. General and comparative endocrinology *172*, 39-43.

Magwere, T., Chapman, T., and Partridge, L. (2004). Sex differences in the effect of dietary restriction on life span and mortality rates in female and male Drosophila melanogaster. J. Gerontol. A Biol. Sci. Med. Sci. *59*, 3-9.

Masternak, M.M., Al-Regaiey, K., Bonkowski, M.S., Panici, J., Sun, L., Wang, J., Przybylski, G.K., and Bartke, A. (2004). Divergent effects of caloric restriction on gene expression in normal and long-lived mice. J. Gerontol. A Biol. Sci. Med. Sci. *59*, 784-788.

Melendez, A., Talloczy, Z., Seaman, M., Eskelinen, E.L., Hall, D.H., and Levine, B. (2003). Autophagy genes are essential for dauer development and life-span extension in C. elegans. Science *301*, 1387-1391.

Miller, R.A., Harrison, D.E., Astle, C.M., Fernandez, E., Flurkey, K., Han, M., Javors, M.A., Li, X., Nadon, N.L., Nelson, J.F.*, et al.* (2014). Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. Aging Cell *13*, 468-477.

Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., Booker, M., Perkins, L., and Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nat Methods *5*, 49-51.

Peters, R.H. (1986). The ecological implications of body size, Vol 2 (Cambridge University Press).

Rosenson, R.S., and Stafforini, D.M. (2012). Modulation of oxidative stress, inflammation, and atherosclerosis by lipoprotein-associated phospholipase A2. J. Lipid Res. *53*, 1767-1782.

Rubinsztein, D.C., Marino, G., and Kroemer, G. (2011). Autophagy and aging. Cell *146*, 682-695.

Sacher, G.A. (1959). Relation of Lifespan to Brain Weight and Body Weight in Mammals. In Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing), G.E.W. Wolstenholme, and M. O'Conner, eds. (Chichester, John Wiley & Sons, Ltd), pp. 115-141.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347-355.

Segall, P.E., and Timiras, P.S. (1976). Patho-physiologic findings after chronic tryptophan deficiency in rats: a model for delayed growth and aging. Mech. Ageing Dev. *5*, 109-124.

Simonsen, A., Cumming, R.C., Brech, A., Isakson, P., Schubert, D.R., and Finley, K.D. (2008). Promoting basal levels of autophagy in the nervous system enhances longevity and oxidant resistance in adult Drosophila. Autophagy *4*, 176-184.

Starke-Reed, P.E., and Oliver, C.N. (1989). Protein oxidation and proteolysis during aging and oxidative stress. Arch. Biochem. Biophys. *275*, 559-567.

Steffen, K.K., MacKay, V.L., Kerr, E.O., Tsuchiya, M., Hu, D., Fox, L.A., Dang, N., Johnston, E.D., Oakes, J.A., Tchao, B.N*., et al.* (2008). Yeast life span extension by depletion of 60s ribosomal subunits is mediated by Gcn4. Cell *133*, 292-302.

Stroustrup, N., Anthony, W.E., Nash, Z.M., Gowda, V., Gomez, A., Lopez-Moyado, I.F., Apfeld, J., and Fontana, W. (2016). The temporal scaling of Caenorhabditis elegans ageing. Nature *530*, 103-107.

Swindell, W.R. (2007). Gene expression profiling of long-lived dwarf mice: longevity-associated genes and relationships with diet, gender and aging. BMC Genomics *8*, 353.

Tran, H., Brunet, A., Grenier, J.M., Datta, S.R., Fornace, A.J., Jr., DiStefano, P.S., Chiang, L.W., and Greenberg, M.E. (2002). DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein. Science *296*, 530-534.

Vernace, V.A., Arnaud, L., Schmidt-Glenewinkel, T., and Figueiredo-Pereira, M.E. (2007). Aging perturbs 26S proteasome assembly in Drosophila melanogaster. FASEB J. *21*, 2672-2682.

Watanabe, T., Broadley, M.R., Jansen, S., White, P.J., Takada, J., Satake, K., Takamatsu, T., Tuah, S.J., and Osaki, M. (2007). Evolutionary control of leaf element composition in plants. New Phytol. *174*, 516-523.

Western, D. (1979). Size, life history and ecology in mammals. Afr. J. Ecol. *17*, 185-204.

White, P.J., Broadley, M.R., Thompson, J.A., McNicol, J.W., Crawley, M.J., Poulton, P.R., and Johnston, A.E. (2012). Testing the distinctness of shoot ionomes of angiosperm families using the Rothamsted Park Grass Continuous Hay Experiment. New Phytol. *196*, 101-109.