**Comparative analysis of molecular basis underlies mammalian lifespan**

Dongming Xu1, Shaogang Qu1

1 Central Laboratory and Department of Neurology, Shunde Hospital, Southern Medical University (The First People’s Hospital of Shunde Foshan), Foshan, Guangdong, China

**Abstract**

The maximum lifespan varies dramatically across mammals. However, the genetics underly lifespan determination is poorly understood. Here, we genomic-wide explored the correlative genes between protein evolutionary rate and lifespan in 13381 one-to-one orthologue genes among 74 mammals. We identified 370 negative correlative genes and 300 positive correlative genes. Enrichment analyses showed that the positive correlative genes and the negative correlative genes were not only severally enriched in many canonical pathways that regulation of lifespan and aging, such as DNA repair and energy metabolism, but also were overrepresented in several closely relative or adverse categories simultaneously, such as purine metabolism and purinergic nucleotide receptor signaling pathway, oxygen utilization. Further evolutionary analyses suggested that most of positive correlative genes were driven by relaxed selection. Finally, combing the results of correlative and evolutionary force analyses, our signed functional interaction network analyses displayed several important modules and highlighted the core genes in each module that might play central role in regulation of mammalian lifespan. Our study uncovered many important pathways and hub genes that might universally regulate lifespan among mammals, which may contribute us to meet the challenges of human aging and aging-accompanied diseases.

**Highlights**

**Largest genomic data ever used to mammalian lifespan study, including 13381 one-to-one orthologous genes across 74 mammals.**

**Positively correlative genes and negatively correlative genes enriched in closely relative or adverse categories, simultaneously, such as purine metabolism and purinergic nucleotide receptor signaling pathway, oxygen utilization.**

**Most positively correlative genes were driven by relaxed selection.**

**Signed** functional interaction **network analyses showed important modules and core genes that might play central role in regulation of mammalian lifespan.**

**Introduction**

Aging and the accompany diseases are major threaten for human health and society economics and is affected by inherently complex process1, 2. However, the genetic mechanisms of aging determination are still poorly understood3. Most of our knowledge about the genetic mechanisms that govern aging were obtained by studying genetic manipulations in short-lived laboratory animal models2, 4. It is unclear if these insights can be transferred to long-lived mammals like humans and it is difficult to know to what extent these represent insights into the universal mechanisms of longevity regulation rather than species-specific adaptation or coincidental neutral changes5-7.

Fortunately, the rate of aging varies dramatically across wild mammals8. Maximum lifespan (MLS), which can reflect the inherent longevity and “rate of aging” in organisms9, is positively correlative with their body size10. However, many mammals repeatedly evolved to deviate their expected MLS values given their body mass11. For example, the naked-mole rat and Brandt’s bat (Myotis) can live at least 31 and 41 years, over 4 and 10 times longer with respect to the predicated lifespan based on their body mass, respectively12. On the contrary, the rat, mouse and shrew can only live about half years of their expected MLS even in the well-cared laboratory. More importantly, these long-lived or short-lived species exhibit delayed or accelerated age-associated physiological declines. For example, the long-lived naked-mole rat exhibit little age-specific hazard of mortality13 and fertility and the longest-lived genus (Myotis) maintains the length of their telomeres with age without developing cancer14. While cancer-related mortality can reach up to 90% in short-lived mice2, 15. So, the repeatedly changes of MLS in mammals might provide a good opportunity for exploring the underlying molecular mechanisms of regulation of lifespan and aging.

The evolution of protein-coding genes play an important role in the phenotypic evolution. Comparative genomics study of the evolution of protein-coding genes have provide a fruitful avenue for uncovering the genetic basis underlying mammalian lifespan regulation3, 16. However, previous comparative genomics studies are mainly focus on site-based methods and focus on only a few lineages. For example, comparative genomics analyses of the positively selected genes (PSGs) in rodents3 and identified the convergent evolution genes in long-lived primates16. These comparative genomic studies have many limits, such as, sites-based methods are hard to detect genes that respond to convergent changes through nonidentical changes in the same gene17. And these studies only can focus on a few species or groups in one study, which might not differentiate the uncovering molecular mechanisms is lineage-specific or universal across mammals6.

Recently, a new method which tests for association between relative evolutionary rates of genetic elements and the evolution of traits across a phylogeny (Reconverge)17, 18 has been developed. This method can search for convergent shifts in evolutionary rates of individual protein-coding genes that response for convergent phenotypes. The evolutionary rates of protein sequences are useful for linking phenotypes to genes because they can reflect their evolutionary constraints. An increased selective constraint can lead to a slower evolutionary rate of protein-coding genes, whereas faster evolutionary rates can result from a relaxation of constraint or from intensified selection (positive selection)18. A previous study used RERconverge methods to genomic-wide identify genes that negatively correlated with the MLS and found that genes related to DNA repair, immunity, cell cycle, cell death, the IGF1and NFKB-related pathways and are under increased evolutionary constraint in long-lived mammals6. However, the authors only used 35 mammals in their mainly analyses and they did not explore positively correlative genes because it was difficult to differentiate the evolutionary driven forces (relaxed selection or intensified selection)6.

Uncovering the evolutionary forces of the candidate genes underlying phenotypic evolution will deepen our understanding of the detailed molecular mechanisms and benefit us for further verification by functional experiments. Generally, intensified selection means intensify or change the function of the gene will be benefit to the phenotypic evolution, while relaxed selection means weaken or eliminate the function of the gene will be benefit to the phenotypic evolution. Relaxed selection in protein-coding genes is difficult to evaluated, however, a recently developed framework, RELAX19, can determine the relaxed selection and intensified selection in protein-coding gene according to the distribution of the nonsynonymous to synonymous substitution rates (ω or dN/dS) in the codon sites. In addition, branch-site model is usually used to identified positively selective genes20.

Here, we integrated the lifespan phenotypic data from HAGR (Human Ageing Genomic Resources) database12 and the orthologous genes from OrthoMaM database21 to obtain the largest genomic data ever used to explore the molecular basis underlying regulation of mammalian lifespan. We genomic-wide identify correlative genes between relative evolutionary rate and lifespan across the mammals and explored their evolutionary forces, especially for positively correlative genes. In addition, we constructed a signed functional interaction network to explore the important modules and core genes that might play central role in regulation of mammalian lifespan. Our study will deepen our knowledges of the universal molecular mechanisms underlying regulation of mammalian lifespan and provide candidate genes for further experiment verification and drug development.

**MATERIALS AND METHODS**

Integration of lifespan phenotypes and genomic data

The protein alignment of one-to-one orthologous genes was downloaded from OrthoMaM database (V10b)21, which includes 14509 genes among 116 species. The data of observed maximum lifespan and adult weight was extracted from the Human Aging Genomic Resource (HAGR) database12, 999 mammals include both traits. The expected maximum lifespan was calculated according to their adult weight following the formula *expect lifespan (years)=* *3.34\*(adult weight (g)^0.193)*10 and the Longevity quotients (LQ) value was considered as the ratio of observed maximum lifespan to expected maximum lifespan. Finally, 74 species with LQ values and genomics data were used in this study (**Figure 1 and Supplementary Table 1**). HAGR database not only records the lifespan phenotypes, but also collects many genes that affect longevity and aging. We downloaded the longevity-associated genes from 4 databases in HAGR, includes LongevityMap (human genetic variants associated with longevity), AnAge (Curated genes affect ageing and life history in animal), CellAge (genes affect Cell Senescence) and GenDR (Different expression genes after Dietary Restriction) and compared them with our significantly correlative genes.

Filtration of orthologous genes and alignment regions

To exclude the fake orthologous sequences in the one-to-one orthologous genes, we downloaded the protein references of human and mouse from the UCSC database. Then, we searched each protein sequence in each one-to-one orthologous genes to the protein references of human and mouse using blastp, respectively. If the best hit of protein sequence in one-to-one orthologous gene is different with that of human protein when searched to the human protein reference or is different with that of mouse protein when searched to the mouse protein reference, then we removed the protein sequence from the one-to-one orthologous genes (**Supplementary Table 2**).

Next, to exclude the poor alignment regions in one-to-one orthologous genes, we used *trimAL*22 to filter the original alignments with parameter “-resoverlap 0.70 -seqoverlap 50-automated1 -colnumbering”. Finally, only the orthologous with more than 50 species and with longer than 50 amino acids were remained. The distribution of specie number and alignment length in one-to-one orthologous before and after filtering can be seen in **Supplementary Figure 1**.

Identification of significantly correlative genes

The phylogenetic relationship of 74 species was according to the OrthoMaM database (v10b)21 (**Figure 1a**). The *nw\_prune* module in *newick-utils* (v1.6) package23 was used to generate the gene tree for each one-to-one orthologous gene based on the phylogenetic tree of 74 mammals and the missing specie. Next, the *codeml* module from the *PAML* (V4.7) packages24 was used to calculate the branch lengths with Empirical + F model. Then, we used R package *RERconverge* (v0.1.0)18 to compute the association statistic between LQ value changes and relative evolutionary rates (RER) for each orthologous gene. Briefly, RERconverge read all gene trees with their branch length values and calculate the average rate for each branch in the tree (**Supplementary Figure 2**). Then, RERconverge calculate gene-specific rates of evolution, termed relative evolutionary rates (RER) through average rate normalization. This correction can remove the non-specific factors affecting divergence on the branch such as time since speciation and mutation rate. Also, *RERconverge* read all LQ values of 74 mammals as continuous phenotypes and inferred the change of LQ value in each branch (**Supplementary Figure 3**) using maximum likelihood method. Finally, *RERconverge* computed the Pearson correlation between LQ value changes and relative evolutionary rates (RER) with default parameters, excluding the minimum number of species in the gene tree was 50. The Pearson correlative coefficient, termed Rho (the correlation between relative evolutionary rates of genes and the evolution of traits across a phylogeny). Positive Rho value means the relative evolutionary rates of gene is positively correlative with the LQ value changes across phylogenetic lineages, while negative Rho value means the relative evolutionary rates of gene is negatively correlative with the LQ value changes. Rough P value at 0.05 was used as threshold for determining the significantly correlative genes. we called the significantly correlative genes with positive Rho value as positively correlative genes, and the significantly correlative genes with negative Rho value as negatively correlative genes (**Supplementary Table 3-4**).

Enrichment analyses

Firstly, we used ClusterProfiler package25 to searched the overrepresented GO Biological Process (GO BP) terms in significantly correlative genes. The positively correlative genes and negatively correlative genes were used as foreground genes, respectively. All used one-to-one orthologous genes were used as background genes. We removed the overrepresented GO BP terms which includes more than 300 background genes, because these GO BP terms are usually at lower GO BP level and can give us little information. The overrepresented GO BP terms can be seen in **Supplementary Table 6-7**. We also searched the relatively overrepresented GO BP terms in positively correlative genes when compared with negatively correlative genes, and reverse. Fisher’s single-tailed test with *P value* at 0.05 was used as threshold for statistical significance (**Supplementary Table 8-9**). In addition, we used EVIGO26 to remove the redundantly significant GO BP terms with medium cutoff standard (allowed similarity=0.7) (**Figure 3**).

Determine of evolutionary forces

The coding sequence alignments of 74 mammals using in this study were download and extracted from OrthoMaM database (v10b)21. Next, we removed the non-orthologous sequences and bad alignment regions which dropped by previous protein alignment filtering. Next, we considered the species with top 10 largest LQ values as foreground group, and the other species as background group. *Hyphy* package developed by RELAX framework19 was used to determine the evolutionary selective forces in the foreground group according to the distribution of the nonsynonymous to synonymous substitution rates (ω or dN/dS) in the codon sites. Intensified selection (positive selection) will push all ω categories away from neutral evolution (leading to parameter k > 1), whereas relaxed selection will push all ω categories toward neutral evolution (k < 1). Finally, the maximum likelihood method was used to evaluate whether alternative hypothesis is better than the null hypothesis, the null hypothesis indicates no different evolutionary selection between foreground group and background group. Two-sided Chi-square was used to statistically significant test. Rough P value at 0.05 was used as threshold for statistical significance (**Supplementary Table 10**).

In addition, we also used branch-site model27 to detect positive selection among all significantly correlative genes in long-lived species. The species with top 10 largest LQ values were label as foreground. Two-sided Chi-square was used to statistically significant test. Rough P value at 0.05 was used as threshold for statistical significance.

Network analyses

To draw the signed functional interaction network, we extracted genes and their links in the directed interactions recorded by Reactome Functional Interactions database28 (2020) from all significantly correlative genes. The database is designed to find pathways and network patterns related to cancer and other types of diseases. The directed interactions include catalyze, inhibit, activate and expression regulate. Next, we drew the signed network of remained genes using *ReactomeFIViz* module in software *Cytoscape* (v3.8.0)29. We dropped the predict interactions in the network. Then, we added the information of the results of correlative and evolutionary forces analyses manually (**Figure 6**).

**RESULTS AND DISCUSSION**

Mammals with genomic data and lifespan phenotype

The values of observed maximum lifespan and adult weight of mammals were extracted from HAGR database12. There are 999 mammals include both traits. The expected maximum lifespans were calculated according to their adult weights based on previous studies10. The Longevity quotients (LQ) is considered as the ratio of the observed maximum lifespan to the expected maximum lifespan16. The original protein alignment of one-to-one orthologous were downloaded from the OrthoMaM database (v10b)21, there are included 14509 genes span 116 mammals, Finally, 74 mammals include both genomic sequences and lifespan phenotypes were used in following study (**Supplementary Table 1 and Figure 1a**). The adult weight of our species ranges from 4,800,000g (elephant) to 7g (Brandt's bat) and the maximum lifespan ranges from 122.5 years (human) to 2.5 years (Star-nosed mole), which can cover the major diversity of mammals (**Figure 1b**).

Filtration of one-to-one orthologous sequences

The accuracy of calculating the evolutionary rate is largely depended on the quality of the orthologous identification and the alignment accuracy. To exclude the mendacious orthologous sequences in the original OrthoMaM orthologous genes, we searched each protein sequence in each orthologous gene to the human and mouse protein reference database using *blastp* tool. If the best hit of the sequence was not consistency with the best hit which human sequence in the same orthologous gene searched to human protein reference or was not consistency with the best hit which mouse sequence searched to mouse protein reference, the sequence will be dropped (**seen methods**). Totally, we removed 7838 sequences from 1169 one-to-one orthologous genes (**Supplementary Table 2**). Then, *TrimAL*22 was used to remove the poorly aligned fragments which may be not accurately aligned. Then, we dropped the orthologous genes with less than 50 species or with alignment length shorter than 50 amino acids. Finally, 13381 orthologous genes were remained. The distribution of the number of the species and the protein alignment length of each orthologous before and after filtering process were shown in **Supplementary Figure 2a and Supplementary Figure 2b**.

Identification of significantly correlative genes

We used the *codeml* model in *PAML* package (V4.7)24 to calculate the branch lengths and then used RERconverge package18 to screen genes which showed significant correlation between relative evolutionary rates and LQ value across the phylogeny (**seen methods**). Totally, we identified 370 negatively correlative genes (**Supplementary Table 3**) and 300 positively correlative genes (**Supplementary Table 4**). Although these genes are not enriched in longevity genes collected by HAGR database12 (**Supplementary Table 5**), we indeed found that the most significantly correlative genes might involve in regulation of lifespan or aging. For example, A minor SNP in ADAMTS6 (ADAM metallopeptidase with thrombospondin type 1 motif, 6), the mostly positively correlative gene (**Figure 2a**), has protective effective on aging in a GWAS study30. Another locus (rs12199884) in PKHD1 (polycystic kidney and hepatic disease 1), the mostly negatively correlative gene (**Figure 2d**), is negatively associated with longevity in males from Han Chinese population GWAS study31. Deletion of YbeY (YbeY Metalloendoribonuclease), the second mostly negatively correlative gene (**Figure 2e**), will shorten the bacteria longevity32. COL4A2 (Collagen type IV alpha2), the third mostly positively correlative gene (**Figure 2c**), was positively selected in long-live naked mole rat and down-expression in ant-aging mutant mice2. SLC11A1 (Solute Carrier Family 11 Member 1), the third mostly negatively correlative gene (**Figure 2f**), was different expression and methylation in multiple human aging tissues33, 34.

Enrichment analysis

Biological Process of Gene Ontology (GO BP) enrichment analyses showed that the most enrichment categories of negatively correlative genes are related to cell division, cell cycle, DNA repair, DNA recombination and p53 signal pathway (**Supplementary Table 6 and 8, Figure 3a**), which was almost identical to the previously similar study6. DNA repair has been hypothesized to be a longevity determination35, genes involved in DNA repair were experienced positive selection3, convergent evolution16, up expression36 and regulated by higher expressed miRNA37 in long-lived mammals. Our results suggested that DNA repair associated genes were experienced more evolutionary constraints in long-lived species. Negatively correlative genes were enriched in “double-strand break repair”, “DNA synthesis involved in DNA repair”, “recombinational repair”, “regulation of DNA repair” and so on, either comparing with all one-to-one orthologous genes or comparing with positively correlative genes (**Supplementary Table 6 and 8, Figure 3a**). p53 signal pathway also plays a critical role in the regulation of aging and longevity in worms, flies, mice, and humans38, 39. Our analyses suggested negatively correlative genes were enriched in “regulation of signal transduction by p53 class mediator” (10 genes, P=0.0037) (**Supplementary Table 6 and Figure 3a**), which was significantly more than that of positively correlative genes (1 gene, P=0.0170) (**Supplementary Table 8 and Figure 3a**).

GO Biological Process (BP) enrichment analyses of positively relative genes also supported many classical pathways that regulation of lifespan. For example, the most well-known and effective way to extend lifespan in invertebrate and vertebrate is dietary restriction2. In our study, we found that the energy metabolic genes were one of the most enrichment categories in our positively correlative genes. There were 4 positively correlative genes were involved in categories of “positive regulation of glucose import” (P=0.0047), “positive regulation of cellular carbohydrate metabolic process” (P=0.0195) and “tricarboxylic acid metabolic process” (P=0.0053), respectively (**Supplementary Table 7 and Figure 3b**), while no negatively correlative gene was involved in these GO BP terms (P=0.039) (**Supplementary Table 9 and Figure 3b**). Another energy-associated pathway that relative to regulation of lifespan is insulin/IGF-1 signaling pathway. Although mutations that reduce insulin/IGF-1 signaling were well confirmed to extend lifespan in invertebrate, the extent to which the insulin/IGF-1 signaling pathway regulate the lifespan in vertebrate is controversial40, 41. Our results supported genes involved in insulin/IGF-1 signaling pathway might play important role in universal regulation of lifespan in mammals. For example, 3 positively correlative genes were involved in category “positive regulation of insulin-like growth factor receptor signaling pathway” (P=0.0016) (**Supplementary Table 7**), while no negatively correlative genes were involved in this pathway. In addition, 7 positively correlative genes were involved in “response to insulin”, which was significantly more than that of negatively correlative genes (2 genes, P=0.0466) (**Supplementary Table 9**).

Beside energy metabolisms, positively correlative genes were also enriched in many other GO BP terms. Previous studies showed that down regulation of neural excitation and synaptic function is associated with extend lifespan in human42. Also, neural excitation increases with age and inhibition of excitation glutamatergic increases longevity in nematode42. In our study, synaptic transmission, especially excitatory neurotransmitter, glutamate was the most overrepresented GO BP categories in positively correlative genes. For example, 5 positively correlative genes involved in ionotropic glutamate receptor signaling pathway (P=1.30E-04) (**Supplementary Table 7 and Figure 3b**), which is significant more than negatively correlative genes (0 gene, P=0.017) (**Supplementary Table 9 and Figure 3b**). Proteolysis plays important role in regulation of longevity. Proteolysis mainly includes ubiquitin-dependent pathway and autophagy-lysosomal mediate pathway. Previous studies showed that genes related to autophagy-lysosomal pathway are up expression in centenarians43, while genes involved in ubiquitin-dependent pathway is down-expression in long-live mammals44. Here, we found that positively correlative genes were significantly enriched in “positive regulation of ubiquitin-dependent protein catabolic process” (7 genes), either related to all orthologous genes (P=0.0014) (**Supplementary Table 7 and Figure 3b**) or comparing with the negatively correlative genes (0 genes, P=0.0034) (**Supplementary Table 9 and Figure 3b**).

Most interesting, we found several closely relative or adverse categories that were enriched by positively correlative genes and negatively correlative genes, simultaneously. For example, 4 negatively correlative genes (P2RY2/P2RY10/GPR171/P2RY14) were associated with “G-protein coupled purinergic nucleotide receptor signaling pathway” (P=2.38E-04) (**Supplementary Table 6**), while no positively correlative genes involved in this pathway (P=0.0929) (**Figure 4a**). Interesting, all these 4 genes belong to P2Y receptors family. Although, we know little about the relationship between P2Y receptors and lifespan, P2Y receptors play import roles in aging-related disease, such as Alzheimer disease45, 46. In contrast, positively correlative genes were enriched in many other purine-related categories, such as, “purine ribonucleoside diphosphate metabolic process”, the relative evolutionary rate of 7 genes displays positively correlative with the maximum lifespan (P=0.0076) (**Supplementary Table 7**), which was significantly more than negatively correlative gene (1 gene, P=0.0167) (**Supplementary Table 9 and Figure 4a**). In addition, positively correlative genes were also enriched in regulation of purine nucleotide biosynthetic process and purine-containing compound biosynthetic process (**Supplementary Table 9 and Figure 4a**). Previous studies showed that addition of dietary adenine can shorten the lifespan by 48%47 and blocks the longevity effects of dietary restriction48. These results suggested that genes involved in purine metabolic and signaling pathway might play more complex roles in regulation of lifespan or aging.

Another adverse category was involved in oxygen utilization. 7 positively correlative genes were associated with “aerobic respiration” (P=2.80E-04) (**Supplementary Table 7**), which was significantly more than negatively correlative genes (0 gene, P=0.0062) (**Supplementary Table 9 and Figure 4b**). In adverse, 4 negatively correlative genes (CYBB/PSMB11/STOX1/BCL2) were associated with category “cellular response to hypoxia”, while no positively correlative genes involved in this pathway (P=0.0929) (**Figure 4b**). Several studies have suggested that hypoxia may affect the longevity. For example, hypoxia can extend the lifespan in fruit fly49 and nematode50. Also, elderly people living in hypoxia in the Tibetan Plateau tend to have a longer life than similarly aged people in other Chinese regions51. In addition, 5 negatively correlative genes (CYBB/RECQL5/BLM/ADCY6/SPIDR) were enriched in category “cellular response to alcohol” (P=0.0096) (**Supplementary Table 6 and Figure 4b**), while no positively correlative gene involved in this pathway (P=0.0511) (**Figure 4b**). Alcohol metabolism can cause oxygen deficiencies or hypoxia, which in turn impedes ATP production and thus contributes to cell death52. These results suggested that genes related to oxygen utilization might play more important and universal roles in lifespan regulation.

Determine of evolutionary forces

We used RELAX method to uncover the evolutionary forces underlying the evolution of significantly correlative genes, especially for positively correlative genes. We totally identified 2224 genes were experienced relaxed selection (**Supplementary Table 10**) and 682 genes were experienced intensified selection in long-lived mammals (**Supplementary Table 11**). Our results showed that only 3 (1%) positively correlative genes showed intensified selection in long-lived mammals, which was significantly less than that of significantly correlative genes (485/11690) and that of negatively correlative genes (26/370) (**Figure 5a**). Branch-site model also found the ratio of positively selected genes for long-lived mammals in positively correlative genes was significantly less than that of negatively correlative genes (**Supplementary Figure 4 and Supplementary Table 12**). In adverse, we found 193 (64.3%) positively correlative genes showed relaxed selection in long-lived mammals, which was almost 3 times higher than that of non-significantly correlative genes (P=3.03E-11) and was over 6 times more than that of negatively correlative genes (P=1.84E-52) (**Figure 5b**). These results suggested that most of the positively correlative genes might be driven by relaxed selection in the long-live mammals, which indicates that knock-out or down-regulation of these genes might contribute to extend lifespan or enhance anti-aging in mammals.

Network analyses

To further explore the interactions between significantly correlative genes, we constructed a signed functional interaction network with their evolutionary driven forces (**Figure 6**). From the signed network, we can identifyat least five modules (**Figure 6 Module1-5**). In each module, most of their genes were either negatively correlative genes or positively correlative genes, except module 1. Module 1 was involved in regulation of insulin signaling pathway. Most of genes directly activate or regulated by insulin signaling pathway were positively correlative genes and many of them were experienced relaxed selection in long-lived mammals, such as INSR, SH3GL2, ANGPT1 and PRKCZ (**Figure 6 Module 1**). The insulin receptor (INSR) was the hub gene in this module, a recent study also suggested that INSR was experienced rapid evolution in long-lived mammals53. Disruption of INSR in mice adipose tissue can extend their lifespan but did not appear to delay their ageing54. However, we also found 5 genes in the upstream of regulating insulin pathways which were consistently experienced increasing evolutionary constraints in long-lived mammal, especially three of them (MUC20/EPHB2/SYNJ1) were experienced positive selection. Positively selected gene were relatively rare either in our network or in our significantly correlative genes. These genes might also regulate lifespan or aging in mammals. For example, MUC20 was one the most prominent calorie restriction (CR)-associated genes in mice55. The complicated evolution in insulin/IGF-1 associated genes suggested 41, 56. The complicated evolution of genes in this module may explain the controversial roles of the extent to which insulin/IGF-1 signaling pathway involved in regulation of mammalian lifespan and aging

In Module2, GO analyses showed that all 6 genes were involved in rRNA metabolism and ribosome biogenesis. 5 of these 6 genes were experienced more evolutionary constraint in long-lived mammals (**Figure 6 module 2**). Previous founds suggested that a fundamental mechanism across eukaryotic species for extending lifespan is to slow down or halt the expenditure of cellular energy that is normally absorbed by the manufacturing and assembly of new ribosomes57. The Module 2 also suggested that WD Repeat-Containing Protein 46 (WDR46) might be the hub gene of this module. All genes in this module activated or catalyzed protein WDR46 (**Figure 6 module 2**). Previous study showed that depletion of wdr-46 reduced the lifespan in nematode58. So, WDR46 might also play an important role in regulation of lifespan in mammals.

Genes in Module 3 involved in DNA metabolic and DNA repair, 14 of 15 genes in this module were negatively correlative genes (**Figure 6 module 3**), which was consistent with our GO analyses (**Figure 3a**). The important of DNA repair capability in regulation of lifespan and aging in vertebrate and invertebrate is undoubtedly35. Here, we found that the genes in this module mainly activated FANCG, FANCC and PRIM1. Thus, these 3 genes may play central role in mediation of DNA repair pathway and regulation of lifespan and aging in mammals (**Figure 6 module 3**). Module 5 like a protein complex and mainly consist of negatively correlative genes and involved in cell division. Genes in this module may combine other protein, such as NUP88 and NUP155 to activate CPSF3 and NCBP1 (**Figure 6 module 5**).

Module 4 mainly consisted of positively correlative genes and these genes were mainly involved in cell adhesion, cell matrix interaction, extracellular matrix organization and so on. This module might be regulated by insulin/IGF-1 signaling pathway (**Figure 6 module 4**). In this module, Protein Tyrosine Kinase 2 Beta (PTK2B) catalyzed 8 genes, which regulated largest number of genes in our signed network. ITGB7 was activated by four genes and was the major target of this module (**Figure 6 module 4**). However, we known little information about how PTK2B and ITGB7 affect lifespan or aging.

Beside five modules above, other genes might also involve in regulation of lifespan or aging in our network. For example, cyclin D1 (CCND1) was regulated by largest number of genes in our signed network (**Figure 6**). Overexpression of CCND1 will lead to cell senescence59. Another most targeted gene was apoptosis regulator (BCL2), which was regulated by 4 genes. BCL2 is a negative regulator of autophagy, disruption of the beclin 1-Bcl2 autophagy regulatory complex promotes longevity in mice60 and down-regulation of Bcl-2 expression controls murine dendritic cell longevity61.

**CONCLUSIONS**

Our study collected the largest genomic data ever to systematically uncover the universal molecular mechanisms underlying the evolution of mammalian lifespan. We identified 370 negatively correlative genes and 300 positively correlative genes which showed convergent shift between protein evolutionary rate and maximum lifespan. Enrichment analyses not only confirmed many canonical pathways known to regulation of longevity and aging, but also found many other pathways might also play important roles, such as synaptic transmission, ubiquitin-dependent proteolysis, oxygen utilization, purine metabolisms and purinergic nucleotide receptor signaling pathway and so on. Further our evolutionary analyses suggested that most of positively correlative genes were driven by relaxed selection. Finally, our signed functional interaction network analyses of significantly correlative genes and their evolutionary forces highlighted several modules and their core genes in each module. Our study uncovered many important pathways and genes that might universal regulation of longevity and aging across mammals. Deepen studies of the molecular mechanisms of these genes and pathways might contribute us to meet the challenges of aging and aging-accompanied diseases.

**ACKNOWLEDGMENTS**

This work was supported by the China Postdoctoral Science Foundation (No. 2019M652951).

**AUTHOR CONTRIBUTIONS**

D.-M.X. designed the project. S.-G.Q. supervised the project. D.-M.X. performed the genomic analyses and wrote the paper.

**Competing interests**: The authors declare no competing interests.

Reference

1 Jin K *et al.* *Aging Dis* **6**:1-5 (2015).

2 Tian X *et al.* *Trends Endocrinol Metab* **28**:722-734 (2017).

3 Sahm A *et al.* *PLoS Genet* **14**:e1007272 (2018).

4 Folgueras AR *et al.* *Circ Res* **123**:905-924 (2018).

5 Austad SN. *J Gerontol A Biol Sci Med Sci* **64**:199-201 (2009).

6 Kowalczyk A *et al.* *Elife* **9**(2020).

7 Ma S *et al.* *Semin Cell Dev Biol* **70**:190-203 (2017).

8 Austad SN. *Exp Gerontol* **36**:599-605 (2001).

9 JE B. Living Longer, in Good Health to the End. The New York Times 25 August 2008:p. D7.

10 de Magalhaes JP *et al.* *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **62**:149-160 (2007).

11 Nowak RM *et al.* Walker's Mammals of the World: JHU press (1999).

12 Tacutu R *et al.* *Nucleic Acids Res* **46**:D1083-D1090 (2018).

13 Buffenstein R. *J Comp Physiol B* **178**:439-445 (2008).

14 Foley NM *et al.* *Sci Adv* **4**:eaao0926 (2018).

15 Lipman R *et al.* *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **59**:B977-B983 (2004).

16 Muntane G *et al.* *Mol Biol Evol* (2018).

17 Partha R *et al.* *Mol Biol Evol* **36**:1817-1830 (2019).

18 Kowalczyk A *et al.* *Bioinformatics* (2019).

19 Wertheim JO *et al.* *Mol Biol Evol* **32**:820-832 (2015).

20 Zhang J *et al.* *Mol Biol Evol* **22**:2472-2479 (2005).

21 Scornavacca C *et al.* *Mol Biol Evol* **36**:861-862 (2019).

22 Capella-Gutierrez S *et al.* *Bioinformatics* **25**:1972-1973 (2009).

23 Junier T *et al.* *Bioinformatics* **26**:1669-1670 (2010).

24 Yang ZH. *Mol Biol Evol* **24**:1586-1591 (2007).

25 Yu GC *et al.* *Omics-a Journal of Integrative Biology* **16**:284-287 (2012).

26 Supek F *et al.* *PLoS One* **6**:e21800 (2011).

27 Zhang J *et al.* *Mol Biol Evol* **22**:2472-2479 (2005).

28 Wu G *et al.* *Genome Biol* **11**:R53 (2010).

29 Shannon P *et al.* *Genome Res* **13**:2498-2504 (2003).

30 Shi H *et al.* *Neurobiol Aging* **33**:1849 e1845-1818 (2012).

31 Zeng Y *et al.* *JAMA Netw Open* **1**:e181670 (2018).

32 Yin L *et al.* *mBio* **10**(2019).

33 Reynolds LM *et al.* *Nat Commun* **5**:5366 (2014).

34 Shokhirev MN *et al.* *Aging Cell* **20**:e13280 (2021).

35 Tian X *et al.* *Cell* **177**:622-638 e622 (2019).

36 MacRae SL *et al.* *Aging (Albany NY)* **7**:1171-1184 (2015).

37 Huang Z *et al.* *BMC Genomics* **17**:906 (2016).

38 Feng Z *et al.* *Genes Cancer* **2**:443-452 (2011).

39 Seim I *et al.* *NPJ Aging Mech Dis* **2**:16014 (2016).

40 van Heemst D. *Aging Dis* **1**:147-157 (2010).

41 Barbieri M *et al.* *Am J Physiol Endocrinol Metab* **285**:E1064-1071 (2003).

42 Zullo JM *et al.* *Nature* **574**:359-364 (2019).

43 Xiao FH *et al.* *Genome Res* **28**:1601-1610 (2018).

44 Fushan AA *et al.* *Aging Cell* **14**:352-365 (2015).

45 Ajit D *et al.* *Mol Neurobiol* **49**:1031-1042 (2014).

46 Erb L *et al.* *Biol Cell* **107**:1-21 (2015).

47 Lang S *et al.* *PLoS Genet* **15**:e1008318 (2019).

48 Stenesen D *et al.* *Cell Metab* **17**:101-112 (2013).

49 Harrison JF *et al.* Moderate hypoxia (10 kPa) extends lifespan with adult but not juvenile exposure in the fruitfly, Drosophila melanogaster. Wiley Online Library 2008.

50 Mehta R *et al.* *Science* **324**:1196-1198 (2009).

51 Li Y *et al.* *Cell Res* **27**:302-305 (2017).

52 Cunningham CC *et al.* *Alcohol Res Health* **27**:291-299 (2003).

53 Yu Z *et al.* *The Innovation* **2**(2021).

54 Bluher M *et al.* *Science* **299**:572-574 (2003).

55 Kok DEG *et al.* *J Nutr Biochem* **56**:152-164 (2018).

56 Barker DJP. *Br Med J* **311**:171-174 (1995).

57 MacInnes AW. *Wiley Interdiscip Rev RNA* **7**:198-212 (2016).

58 Leung CK *et al.* *Free Radic Biol Med* **52**:937-950 (2012).

59 Atadja P *et al.* *Exp Cell Res* **217**:205-216 (1995).

60 Fernandez AF *et al.* *Nature* **558**:136-140 (2018).

61 Nopora A *et al.* *J Immunol* **169**:3006-3014 (2002).