Cross-species functional modules link proteostasis to human normal aging

Andrea Komljenovic[1,2], Hao Li[3], Vincenzo Sorrentino[3], Zoltán Kutalik[2,4], Johan Auwerx[3], Marc Robinson-Rechavi[1,2*]

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Laboratory of Integrative Systems Physiology, EPFL, Lausanne, Switzerland
[4]Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Lausanne 1010, Switzerland

*Correspondence author: marc.robinson-rechavi@unil.ch

Running title: Conserved functional modules link proteostasis to aging

Abstract

The evolutionarily conserved nature of the few well-known anti-aging interventions that affect lifespan, such as caloric restriction, suggests that aging-related research in model organisms is directly relevant to human aging. Since human lifespan is a complex trait, a systems-level approach will contribute to a more comprehensive understanding of the underlying aging landscape. Here, we integrate evolutionary and functional information of normal aging across human and model organisms at three levels: gene-level, process-level, and network-level. We identify evolutionarily conserved modules of normal aging across diverse taxa, and importantly, we show that proteostasis involvement is conserved in healthy aging. Additionally, we find that mechanisms related to protein quality control network are enriched in 22 age-related genome-wide association studies (GWAS) and are associated to caloric restriction. These results demonstrate that a systems-level approach, combined with evolutionary conservation, allows the detection of candidate aging genes and pathways relevant to human normal aging.

Highlights

- Normal aging is evolutionarily conserved at the module level.

- Core pathways in healthy aging are related to mechanisms of protein quality network
- The evolutionarily conserved pathways of healthy aging react to caloric restriction.
- Our integrative approach identifies evolutionarily conserved functional modules and showed enrichment in several age-related GWAS studies.

Introduction

Aging is a process that affects all living organisms and results in a progressive decline in life function and a gain in vulnerability to death (Jones *et al.*, 2014). In humans, aging is the main risk factor in a wide spectrum of diseases. The recent increase in human healthspan, also called 'normal', 'disease-free', or 'healthy' aging, is mostly due to improved medical care and sanitation (Greene, 2001; Rappuoli *et al.*, 2011).

Major strides have been made in understanding the main molecular pathways underpinning the aging phenotype, leading to the definition of a number of "hallmarks" of aging, that may be common between species (López-Otín *et al.*, 2013). Mitochondrial dysfunction and loss of proteostasis are two such conserved hallmarks of aging. Indeed, many comparative studies have shown mitochondrial dysfunction as a common feature of aging across species. Shared gene signatures in aging of *D. melanogaster* and *C. elegans* are linked to mitochondrial oxidative respiration, and similar results are observed in primates, including humans (McCarroll *et al.*, 2004; de Magalhães, Curado and Church, 2009; Alexey A. Fushan *et al.*, 2015). Collapse of proteostasis is another hallmark of aging that was shown to be important not only in short-lived species, but also in long-lived ones (Tian, Seluanov and Gorbunova, 2017). Loss of proteostasis is related to major human pathologies, such as Alzheimer's and Parkinson's disease, offering an opportunity to detect conserved candidate genes important in those age-related diseases (Labbadia and Morimoto, 2015; Sorrentino *et al.*, 2017). The proteostasis network consists of three major mechanisms: protein synthesis, autophagy and the proteasome complex (Kaushik and Cuervo, 2015). Recent studies on the long-lived naked mole rat showed maintenance of proteasome activity throughout life (Rodriguez *et al.*, 2012). Perturbations of components of the proteostasis network have already been observed in other species, such as mice (Pyo *et al.*, 2013). Notably, caloric restriction, defined as a reduction of regular caloric intake by 20-40%, extends lifespan and delays the onset of age-related diseases in many species (Lee *et al.*, 2006; Selman and Hempenstall, 2012; Bass *et al.*, 2015; Mattison *et al.*, 2017), in part through effects on mitochondria and proteostatic networks.

Although significant efforts have been made to uncover the identity of genes and pathways that affect lifespan, it is unclear to what extent the functional information of aging obtained from model organisms can contribute to human aging. Focusing on the process of aging in healthy individuals should improve the discovery of pathways important in natural aging. In addition, systems-level analysis of large datasets has emerged as an important tool for identifying relevant molecular mechanisms, as single gene-based methods are not sufficient to elucidate complex processes such as aging. The integration of various data types contributes to identify pathways and marker genes associated with specific phenotypes (Baumgart *et al.*, 2016; Hasin, Seldin and Lusis, 2017). Notably, co-expression network analyses can help to elucidate the underlying mechanisms of various complex traits (Xue *et al.*, 2007; van Dam *et al.*, 2017).

To incorporate evolutionary and functional age-related information, we integrated transcriptome profiles of four animal species from young and old adults: *H. sapiens*, *M. musculus*, *D. melanogaster* and *C. elegans*. As a source of gene expression, we used human data from the large-scale Genotype-Tissue expression (GTEx) project (Mele *et al.*, 2015), together with aging transcriptomes of model organisms. We identified the functional levels of conserved genetic modifiers important during normal aging, and related them to caloric restriction experiments and enrichments in age-related genome-wide association studies (GWAS). We used gene families as evolutionary information across distant species in a two-step approach to observe age-related conserved mechanisms. Our results show the contribution of age-related mechanisms from model organisms to human normal aging, with notably a demonstration of the conserved role of proteostasis in normal aging and in the reaction to dietary restriction.

Results

Data-driven integrative evolutionary approach to healthy aging

We used a three steps-approach to integrate transcriptomes across distant species (human and model organisms) and to identify evolutionarily conserved mechanisms in normal aging (Figure 1A). In the first step, we performed differential expression analysis between young and old samples in two tissues, skeletal muscle and hippocampus, from humans (*Homo sapiens*) and mice (*Mus musculus*), and in whole body for the fly (*Drosophila melanogaster*) and the worm (*Caenorhabditis elegans*). We also used transcriptome datasets related to caloric restriction in these species for validation. In the second step, we obtained 3232 orthologous sets of genes, 'orthogroups', across those four species (see Methods). Each orthogroup (OG) is defined as the set of the orthologous and paralogous genes that descended from a single ancestral gene in the last common ancestor to those four species (*H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*) and an outgroup species (*Amphimedon queenslandica*). Each orthogroup can contain a different number of genes, and was treated as a single functional meta-gene common to four species. We corrected for the orthogroup sizes by applying Bonferroni correction on the gene p-values from differential expression analysis within the orthogroup. Then, we selected a representative gene per species within orthogroups. We took the minimum Bonferroni adjusted p-value of a species-specific age-related gene from differential expression analysis. This allowed us to build 'age-related homologous quadruplets' (see Details in Figure S1A). The four p-values within each quadruplet were then summarized into a single p-value per quadruplet, by using Fisher's combined test. We obtained 2511 gene quadruplets in skeletal muscle, 2800 in hippocampus, and 1971 in caloric restriction experiments (Table S4). We characterized their biological relevance by functional enrichment. In the third and final step, those quadruplets of age-related genes were used to build a co-expression network per species (Figure S1B). These networks were then integrated together using order statistics into one cross-species age-related network. We performed community search algorithm on this network to obtain age-related and evolutionarily conserved modules. The modules were then tested for functional enrichment and for enrichment in GWAS hits.

Age-related gene expression patterns in four species

To study normal aging, we restricted ourselves to transcriptomic studies with at least one young adult and one old adult time-point, adult being defined as after sexual maturity (Figure 1B). Transcriptomes had to come from control samples (model organism datasets) or relatively healthy individuals (GTEx dataset). We defined young and old adults across species as follows: young: 3-4 months for *M. musculus*, 2-10 days for *D. melanogaster*, 3-6 days for *C. elegans*; old: 18-24 months for *M. musculus*, 20-50 days for *D. melanogaster*, 10-15 days for *C. elegans*. For the GTEx data, samples from all adults (20-70 years old) were taken into account in a linear model to detect differentially expressed genes. In human and mouse, we focused on two tissues, skeletal muscle and hippocampus, because they are known to be profoundly affected by aging. During aging, skeletal muscle is affected by sarcopenia (Marzetti and Leeuwenburgh, 2006). Changes

in hippocampus function have a significant impact on the memory performances in elderly people (Driscoll *et al.*, 2003). Thus both tissues are susceptible to aging-related diseases. For human, we used transcriptomes of 361 samples from skeletal muscle tissue and 81 samples from hippocampus from GTEx V6p. For the other species we used diverse publicly available transcriptomic datasets (Table S1). The sample sizes for model organisms were variable, from 3 to 6 replicates per time-point. In order to compare samples between young and old age groups, we fitted linear regression models for each dataset. In addition, in the GTEx dataset we controlled for covariates and hidden confounding factors to identify genes whose expression is correlated or anti-correlated with chronological age, taking into account all samples (see Methods).

We observed uneven distributions of up- and down-regulated genes with aging across different species and datasets (Figure 1C, Table S2), suggesting variable responses to aging and different power of datasets. The human hippocampus shows substantially more age-related gene expression change than skeletal muscle (6083 *vs*. 5053 differentially expressed genes, FDR < 0.1). However, mouse hippocampus shows less gene expression change than skeletal muscle (1639 *vs*. 2455 differentially expressed genes, FDR < 0.1). These differences are due in part to the smaller sample size of the mouse skeletal muscle study. We limited our analysis to genes that were expressed in at least one age group, leading to detection of 15-40 % of genes that exhibits age-related gene expression changes. Of note, these changes are often very small, typically less than 1.05 fold in humans and less than 2-fold in animal models.

It has been previously reported that there is a small overlap of differentially expressed genes among aging studies (de Magalhães, Curado and Church, 2009; Yang *et al.*, 2015). To make results easily comparable across species, the young and old adults of one species should correspond to young and old adults of another species (Flurkey, M. Currer and Harrison, 2007). Our clustering shows good consistency across age groups of samples between species, based on one-to-one orthologous genes with significant age variation (FDR < 0.05) (Figure 2A, Figure S2). Yet there is a low overlap of one-to-one orthologous genes with significant expression change in aging (Table S3). This observation is in line with two studies showing that the overlap between individual genes associated with aging did not reach the level of significance (Smith *et al.*, 2008; Alexey A. Fushan *et al.*, 2015). To go beyond this observation, we correlated log-transformed fold change (old/young; or log of α age-related regression coefficient in human) between human and model organisms. We observed weak pairwise correlations (Figure S3) when comparing single genes. This indicates that most transcriptional changes on the gene level are species-specific, and that there is little evolutionary conservation to be found at this level.

Cross-species integration at the process-level reveals proteostasis-linked age-related mechanisms

To assess the age-related gene expression changes on a functional level in healthy individuals per species, we performed gene set enrichment analysis (GSEA) (Subramanian *et al.*, 2005) using gene ontology (GO) annotations (Gene Ontology Consortium *et al.*, 2000; The Gene Ontology Consortium, 2017). We then selected significant GO terms (FDR < 0.20) that we grouped into broader categories.

All species showed a general pattern of down-regulation of metabolic processes, such as mitochondrial translation (GO:0032543) in human GTEx skeletal muscle tissue, nucleotide metabolic process (GO:0009117) in mouse muscle tissue, cellular respiration (GO:0045333) in fly whole body, and oxoacid metabolic process (GO:0043436) in worm whole body (Figure S4). The pattern of metabolic down-regulation was stronger in muscle for both human and mouse. The processes that were down-regulated in hippocampus were related to behavior (GO:0007610), cognition (GO:0050980) and neurotransmitter secretion (GO:0007269) in human, and to synaptic signaling (GO:0099536) and axonogenesis (GO:0007409) in mouse. This confirms that there is a tissue-specific signal in normal aging. Due to small samples size of the mouse skeletal muscle dataset, we were able to detect only down-regulated metabolic processes. In addition to metabolism, we observe strong immune systems response to aging, such as regulation of cytokine production (GO:0001817) in human hippocampus or leukocyte-mediated immunity (GO:0002443) in mouse hippocampus. These results are consistent with known links between metabolism, immunity and aging (Lanna *et al.*, 2017).

We aggregated processes on the functional level across four species using evolutionary information to observe common age-related mechanisms rather than tissue-specific mechanisms. We integrated differential expression analysis from each species, as described above. We obtained 2010 genes in skeletal muscle / whole body, 2075 genes in hippocampus / whole body, and 1962 genes in caloric restriction experiments (Fisher combined tests, FDR < 0.10) (Table S4). We examined their biological relevance using Gene Ontology enrichment analysis (GEA) based on human annotation (Figure 2B). We did not take into account whether the processes that are shared across species are regulated in the same direction, but rather whether they are consistently perturbed during aging.

We obtained 100 significant GO terms (FDR < 0.05) related to biological processes, and aggregated them into broader GO categories. While our species-specific analysis mostly shows tissue-specific pathways, we found that terms with an evolutionarily conserved relation to normal aging are strongly enriched for processes involved in proteostasis, or protein homeostasis. The proteostasis-linked processes are more conserved than expected by chance (Figure S6). The other conserved processes are related to transport, translation, transcription and post-transcriptional modifications, and protein ubiquitination (Figure 2B, Table S5). We also confirmed previously known evolutionarily conserved age-related pathways, such as cellular respiration and immune response. Integrating caloric restriction datasets across the four species showed enrichments in similar processes (Figure 2B).

While most of the shared processes have been previously linked to aging, we focused on proteostasis and related processes. To characterize in more detail the specificity of proteostasis-linked processes, we investigated their enrichment strength in the large human GTEx dataset (Figure 2C). Since proteostasis perturbation is detected both through the GO domains of cellular localization and of biological process, we investigated these two domains, and obtained similar enrichments for both skeletal muscle (Figure 2C), and in hippocampus (Figure S6, Table S6). The most enriched cellular component terms in skeletal muscle were related to proteasome complex (GO:0000502, enrichment score: 1.99) and to mitochondrial matrix (GO:0005759, enrichment score: 1.38). We also observed strong enrichment of ribosomal large (GO:0000027, enrichment score: 1.57) and small subunit (GO:0000028, enrichment

score: 1.84), of protein homotetramerization (GO:0051289, enrichment score: 1.18), and of GO biological processes that are part of the protein quality control network. Overall, the translation and proteasome complexes appear to be the parts of the protein quality control network whose involvement in aging is both evolutionarily conserved across different species, and significantly enriched in human healthy aging. Interestingly, we also detect the mRNA splicing pathway as a part of the conserved processes between species.

The direction of the changes in conserved proteostasis processes in humans is consistent with a relation between loss of proteostasis and healthy aging (Figure 3). Although macroautophagy did not show a strong enrichment score in the Figure 2C (GO:0016236, enrichment score: 0.90), there is down-regulation of the conserved genes associated with macroautophagy (Figure 3A), translation (Figure 3B), and the proteasome complex (Figure 3C), which are important in the protein quality network. Similar results are observed in hippocampus, although not with a strong signal as in skeletal muscle (Figure S7). The changes during healthy aging in both tissues are rather subtle but significant (Figure 3, Table S7).

Functional characterization of cross-species age-related network identifies candidate genes related to healthy aging

To characterize age-related processes at a systems-level and to prioritize conserved marker genes associated with normal aging, we constructed probabilistic networks. These were based on prioritization of co-expression links between conserved age-related genes across four species. These genes became nodes in the multi-species network. Thus the connections between the conserved age-related genes are based on evolutionary conservation, and prioritized according to the their co-expression in each species.

Our integrative network analysis initially identified 20 and 14 modules for skeletal muscle and hippocampus, respectively. We randomized our networks 100 times based on the same number of conserved genes per experiment and obtained significantly higher numbers of gene-gene connections than in the original network (permutation test, $p = 0.0198$) (Figure S9). Thus aging networks appear to be lowly connected. We focused only on the modules larger than 10 genes; there were 12 such modules per tissue. These modules ranged in size from 16 (M7 hippocampus) to 191 genes (M12 hippocampus) (Figure 4A and 4B, Table S8). The networks were summarized to module level (module as a node), and we observed strong inter-modular associations. This analysis provided several levels of information. First, it provided a small number of coherent gene modules that represent distinct transcriptional responses to aging, confirming the existence of a conserved modular system. Second, it detected conserved marker genes affected during aging, discussed below.

To determine which of the conserved aging-associated modules are related to the main components of the proteostasis network, we carried out functional enrichment analysis on these modules, based on human gene annotations. The enrichments were highly significant for all modules (FDR < 0.01), and confirmed the inter-modular associations (Table S8). Not all of the modules were related to proteostasis. Interestingly, M1, M10 and M5 in the skeletal muscle network share strong associations with mitochondrion

organization and distribution, regulation of cellular amino acid metabolic process and ubiquitin protein catabolic process, while M2 and M3 in hippocampus share associations with different types of protein transport. Other modules (M1, M6, M7, M8, M11, M12 in skeletal muscle; M2, M3, M4, M5, M12 in hippocampus) support the impact of healthy aging on genes related to the proteostasis-linked processes. This included processes related to protein polyubiquitination (GO:0000209), translational initiation (GO:0006413), protein transport (GO:0015031), regulation of macroautophagy (GO:0016241), and proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161). In skeletal muscle tissue there were also a strong enrichment in splicing process (M3). Moreover, the connection between M2, M10 and M6 in hippocampus, and between M1, M5 and M12 in skeletal muscle indicates that there is a connection between mitochondrial and proteostasis-related processes, recently shown to occur also in amyloid-beta proteotoxic diseases (Sorrentino *et al.*, 2017), and during mitochondrial stress(Labbadia and Morimoto, 2015; D 'amico, Sorrentino and Auwerx, 2017; Sorrentino, Menzies and Auwerx, 2018).

To investigate the relevance of proteostasis-linked modules to age-related diseases, we performed enrichment analysis based on genes coming from 22 GWAS studies (See Methods, Table S9). M3, M4, M5 and M12 of skeletal muscle showed enrichment in coronary artery disease, triglycerides, 2hr glucose, multiple sclerosis and cholesterol-related diseases, while M4 and M6 of hippocampus showed enrichment in coronary artery disease and fasting proinsulin, respectively. Skeletal muscle module M12 is particularly interesting because its genes are not only enriched in GWAS studies but also have strong involvement in proteostasis (Figure S10A). Similarly, hippocampus module M4 is interesting due to enrichment in both GWAS and in one of the proteostasis processes (Figure 5B).

To further characterize these modules, we studied how conserved modular genes associated with proteostasis and age-related GWAS diseases are changed in expression in humans, as a long-lived species. We looked deeper into the gene composition of two modules, M1 associated with SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (79 genes) and M4 associated with positive regulation of telomerase RNA localization to Calaj body (155 genes) from the skeletal muscle and hippocampus networks, respectively. We defined network hubs, genes that exhibit a significantly high number of connections with other genes in the network, for each of these modules in muscle (Figure 5A, S10A) and hippocampus (Figure 5B, S10B). We focused on the hubs with the highest scores in each module and examined their neighborhood. The top ranked genes in M1 of the skeletal muscle were *CTSK, UBE2L3 and CPA3* (Figure 5A). They are associated with protein quality network, related to protein degradation. Interestingly, the neighboring genes *PSMB2* and *PSMA1* are associated with the proteasome complex (Figure 5A). The top ranked genes in M4 in skeletal muscle were related to the translational initiation process, with *MAPRE3*, *SPTBN2* and *ATP6V0A1* as hub genes. Their network neighbors were tightly connected to the cytoskeleton and protein transportation (Figure 5B).

Other modules also show links to metabolism and to proteostasis. For example muscle module M12 and hippocampus module M3 are associated with the protein polyubiquitination process (Figure S10). The top-ranked hub genes in muscle M12 were *DDX3X*, *KIF5B* and *USP7* (Figure S9A). Those genes are related to DNA damage, translation and transport regulation in the cell. In the hippocampus module M3 (Figure 9B), the three hub genes (*PPP3CB*, *DNM1L* and *ITFG1*) are involved in hydrolase

activity, apoptosis and programmed necrosis and modulating T-cell function. Although the hub genes with the highest scores were strongly related to metabolism and to tissue-specific functions in each of these two modules, their network neighborhood is associated with the protein quality control network. More specifically, the *PSMB5* and *PSMD3* genes are related to the proteasome complex and are connected to hub genes.

We combined this hub gene analysis with GWAS association gene scores, and observed that *PSMB5, UBE2L3*, and *PSMD3* (Figure 5C, Table S9) are important in many age-related diseases or phenotypes, such as Alzheimer's disease, HDL cholesterol, LDL cholesterol, triglycerides, and insulin resistance. Other genes related to translation and proteasome complex were also strongly associated to such diseases, such as *PSMB5* with multiple sclerosis (Pascal (Lamparter *et al.*, 2016) gene score: *p-value* = 0.0348) and HDL cholesterol (Pascal gene score: *p-value* = 0.0155). Finally, we observed that the prioritized genes associated with age-related diseases from conserved functional modules change in opposite directions with healthy aging and with caloric restriction (Figure 5D). This differential expression is consistent with a causal role in these age related diseases, given the attenuating effect of caloric restriction on aging.

Validation of marker genes using independent mouse studies

We analyzed the association of the expression levels of candidate genes with lifespan in different tissues of mouse recombinant inbred lines used for population genetics analyses, such as the BXD (Andreux *et al.*, 2012) and LXS (Liao *et al.*, 2010) strains. We observed an inverse correlation between transcript levels of *PSMB5* (Figure 5E) in the spleen of the BXD strains (average age at the time of transcript analysis 78 days; $p = 7.14 \times 10^{-5}$) and in the prefrontal cortex of LXS lines (average age of 72-days; $p = 0.03$), and lifespan longevity. This correlation was consistent even after correction for the population structures with mixed models (Kang *et al.*, 2008). Thus lower expression of PSMB5 is linked to lifespan. Consistent with this, the GSEA showed down-regulation of the proteasome complex during the lifespan of the mice (Figure 5E, left panel).

Discussion

The challenge of detecting underlying mechanisms of healthy aging that are evolutionarily conserved is thought to be a key impediment for understanding human aging biology (Fontana *et al.*, 2010). In this work, we coupled evolutionary and functional information of healthy aging gene expression studies to identify conserved age-related systems-level changes. We identified conserved functional modules by integration of co-expression networks, and we prioritized genes highlighted by GWAS of age-related diseases and traits. The observations on several functional levels allowed us to highlight the role of proteostasis, which includes all processes related to protein quality control network, as a strong core process associated with normal aging.

Previous observations restricted to a small number of evolutionarily conserved genes with large effects in aging, or in age-related diseases, provided some evidence that aging mechanisms might be conserved among animals (de Magalhães, Curado and Church, 2009). However, transcriptome level correlations of expression changes in aging between species are very low in our gene-level results, in accordance with other

studies (Zahn *et al.*, 2006; Smith *et al.*, 2008; Alexey A Fushan *et al.*, 2015). Yet the process of aging appears overall conserved, with notably common effects of interventions, such as caloric restriction, showing similar effects across species ranging from nematodes, flies, to mammals (Gems and Partridge, 2013). The solution to this apparent paradox seems to be that pathways are evolutionarily conserved in aging (Smith *et al.*, 2008), even when single genes are not. Indeed, we have found strong similarities in age-related gene sets between human and other species.

Beyond individual pathways, the modular nature of aging has been previously reported at several levels, such as by protein-protein interaction network analysis during human and fruit fly brain aging (Xue *et al.*, 2007), human longevity network construction and identifying modules (Budovsky *et al.*, 2006), mouse age-related gene co-expression modules identification (Southworth, Owen and Kim, 2009), or aging and age-related diseases cluster detection in human aging (Fernandes *et al.*, 2016). Integrating co-expression networks across species, we identified 10 and 13 evolutionarily conserved functional modules for skeletal muscle and hippocampus, respectively. These conserved modules are not only enriched in processes known to be involved in healthy aging, such as immune-related pathways, they significantly overlap with results from age-related GWASs. The latter is of particular relevance, since finding causality for aging in GWAS is difficult, given its highly multifactorial nature (McDaid *et al.*, 2017). Of note, these modules can be tissue-specific, for example related to energy and amino acids in muscle (Figure 5A). Thus, aging is an evolutionarily conserved modular process, and this modularity is tissue-specific.

An advantage of our approach is that it allows us to detect with good confidence processes whose changes in aging are quite subtle. This is important because healthy aging is not a dramatic process, akin to embryonic development or cancer, but a gradual change in tissues and cell types which keep their defining characteristics. In other words, old muscle and young muscle are very similar at the molecular level, as shown, e.g., by the log-fold change scale in Fig. 3: a log2 age-related regression coefficient (Formula 1) of -0.005 corresponds to a decrease of only 1.0035 fold. Yet we are able to detect processes associated to these changes with strong confidence, and these processes are mostly known in to be age-related. The largest changes, thus easiest to detect, include metabolism (Finkel, 2015), transcription (Roy *et al.*, 2002), translation (Steffen and Dillin, 2016), and immune response. Changes in expression for proteostasis-related genes are weaker, yet integrating at a systems level between species provided us with a strong signal.

More broadly, our results strengthen the case for further investigation into the molecular program that links proteostasis to healthy aging. This is in line with “loss of proteostasis” as one of the nine proposed hallmarks of aging (López-Otín *et al.*, 2013; Walther *et al.*, 2015). Aging involves a deregulation of the protein quality control network, and this is conserved between distant species. Changes in protein synthesis and protein degradation processes have already been linked to several age-related diseases, most notably Alzheimer’s and Parkinson’s disease (Morimoto and Cuervo, 2014). They may be fundamental to the response to normal aging because the accumulation of somatic and germline mutations can alter fine modulation of the protein homeostasis network and produce pathological alterations (Woodruff and Thompson, 2003; Khodakarami *et al.*, 2015). Thus proteostasis provides a link between somatic genome-level changes and the phenotypic impact of aging. Our results show that during healthy or normal aging, the alterations in proteostasis network are rather subtle and discrete, by contrast

to the strong down-regulation of metabolic processes. This suggests that perhaps there is a cascade of triggered pathways as aging proceeds (Tomaru *et al.*, 2012) . Moreover, we detect evolutionarily conserved links inside modules between mitochondrial deregulation (hub genes) and protein homeostasis (neighboring genes) in normal aging, consistent with recent advances in the field (D 'amico, Sorrentino and Auwerx, 2017; Labbadia *et al.*, 2017; Sorrentino, Menzies and Auwerx, 2018).

The main evolutionarily conserved gene candidates from proteostasis, *PSMB5* and *PSMD3,* are related to the proteasome. These two genes were tightly connected to metabolic hub genes in skeletal muscle and to filament organization genes in the hippocampus. The proteasome complex is down-regulated during aging in our results, and in a transgenic mouse mutant proteasome dysfunction led to shorter lifespan (Schmidt and Finley, 2014). In the database of gene expression Bgee (Bastian *et al.*, 2008), human PSMB5 and PSMD3 have top expression in gastrocnemius muscle, with weaker expression in old age. Moreover, both genes showed significant association in GWAS studies with metabolic and disease traits. The *PSMB5* gene was validated by comparing mice strains, and the *PSMD3* gene was related with coronary artery disease, HDL cholesterol and fasting proinsulin, all indicators of healthspan, and would also be worthwhile to explore further.

The association with caloric restriction studies strengthens the functional contribution to aging of the processes we identified. We observed that the gene-set signals were both evolutionarily conserved in caloric restriction, and shared between healthy aging and caloric restriction experiments. Genes related to proteostasis showed opposite directions in expression changes between human healthy aging and caloric restriction. This indicates that these functions are maintained during caloric restriction in humans but lost during aging, and reinforces the case for a causal link between proteostasis and healthy aging. Our observations are consistent with previous research in *C. elegans*, reporting improvement of proteostasis during caloric restriction treatments and extension of the lifespan (Depuydt *et al.*, 2013; Chondrogianni *et al.*, 2015). Notably, *PSMB5* and *PSMD3* follow this trend in caloric restriction relative to healthy aging, further suggesting that they are prime candidates to study genes underlying functional modules in healthy aging.

Integrating biological processes based on evolutionary conservation allows distinguishing relevant signals from noise, despite the weak patterns in aging transcriptomes. Moreover, the fact that a process is similarly involved in aging in very different species strengthens the case for causality. This provides a promising foundation to search for relevant biomarkers of healthy aging of specific tissues, e.g. further analysis of directions of change in homologous tissues, in different model organisms.

In summary, the large-scale, comprehensive gene expression characterization in our study provides insights in underlying evolutionarily conserved mechanisms in normal aging. While metabolic and certain tissue-specific pathways play a crucial role in aging, processes affecting the protein quality control network also show very consistent signal. Using both evolutionary and functional information, we detected conserved functional modules that allowed us to identify core proteostasis-related genes. These genes were implicated as important hits in age-related GWAS studies (Gomes, 2013). Together, the integrative systems-level approach facilitated the identification of conserved modularity of aging, and of candidate genes for future normal aging biomarkers.

Figures

Figure 1. Study design and differential gene expression analysis.
(A) An overview of the integration process based on transcriptomes across the species. (I) Analysis starts at the single-gene level by performing differential expression analysis per species between young and old adults (all samples in case of GTEx human data), and determining the orthogroups across species. (II) The orthogroups (OG) are summarized to single genes that represent age-associated conserved genes. (III) The same genes are then used to build the co-expression networks per species and being integrated in the final cross-species network. (See Methods, Figure S1A and S1B)
(B) The species used in the study with their phylogenetic relations and the alignment of their ages categories.
(C) Barplots representing the numbers of significantly differentially expressed age-related genes (FDR < 0.1) in old healthy individuals in each dataset used. Blue (resp. red) bars represent genes significantly up- (resp. down-) regulated in old adults.

Figure 2. Functional enrichment analysis of integrated age-associated conserved genes.
(A) Clustering of the age-related samples between human (20-30y; 61-70y) and mouse. The heatmaps show good concordance between the young and old samples between species based on the 1-1 orthologous genes that are differentially expressed.
(B) Bubble plot showing the number of GO categories with conserved change of expression in aging between species. The analysis only includes categorized GO terms that are significant (FDR < 0.05) and unique to the homologous quadruplets enrichment.
(C) GO enrichment of genes involved in processes related to proteostasis based on cellular component (CC) and biological process (BP). Lengths of bars represent GO log2-transformed enrichment scores.

Figure 3. Gene expression changes in the main aspects of the proteostasis network in healthy aging human skeletal muscle.
Conserved genes from macroautophagy (A), translation (B) and proteasome complex (C) in GTEx skeletal muscle data. Grey, conserved genes that are not significant (FDR > 0.05) in human GTEx skeletal muscle data. The x-axis of the volcano plots shows the log2 of age-regression coefficient (log2 slope, Formula 1) across the samples in GTEx data (see Methods; Formula 1), named log2 fold-change.
(D) Schematic outline of the gene expression direction of the proteostasis-linked processes in aging human muscle.

Figure 4. Cross-species aging-associated skeletal muscle and hippocampus functional modules and GO enrichments.
Module networks of skeletal muscle (A) and hippocampus (B) with GO and GWAS enrichments for modules of size greater than 10. The tables on the right show top GO BP terms (FDR < 0.1) enriched in the skeletal (upper panel) and hippocampus (lower panel) modules. The GWAS-associated disease column in the same table contains associations to the module passing a threshold of FDR < 0.2.

Figure 5. Module architectures and prioritization of candidate genes

(A-B) Architecture of modules related to protein polyubiquitination (M2; A) and positive regulation of telomerase RNA (M4; B) with hub genes (in red) and their neighboring genes (in black), in skeletal muscle (A) and hippocampus (B).
(C) GWAS heatmap of the conserved proteostasis-related genes that were prioritized in modules. The heatmap shows the strength of association of each gene (hubs and neighbouring genes from the interested modules) with GWAS.
(D) Volcano plot of the prioritized and conserved genes in human dietary restriction dataset.
(E) Validation plots for *PSMB5* gene in independent mouse studies, taken at 72 and 78 days of age. The x-axis represents the expression values of the gene in 35 strains of LXS (upper scatterplot) and BXD (lower scatterplot), and y-axis maximum (upper scatterplot) and median (lower scatterplot) lifespan of that strain. The left panel shows GSEA enrichment relation between proteasome complex and lifespan.

Supplemental Data

Supplement Figures are in Supplemental document.

Table S1. Expression datasets used in aging and caloric restriction analysis. This table contains 2 sheets, corresponding to aging and dietary restriction experiments.

Table S2. Differential expression statistics in skeletal muscle (human, mouse), hippocampus (human, mouse), whole body (fly, worm) for age-related experiments and skeletal muscle (human, mouse) and whole body (fly, worm) for dietary restriction. This table contains 6 sheets, each sheet corresponds for tissue and species. In each sheet, rows correspond to genes with no cutoffs applied. The columns provide differential expression statistics for all the samples (GTEx) and two-group comparisons (model organisms).

Table S3. Overlap between the 1-to-1 conserved age-related orthologs between human and model organisms.

Table S4. List of orthologous genes from integrative analysis. This table contains 3 sheets, corresponding to muscle, hippocampus and dietary restriction experiments that were integrated based on orthologous groups. The columns represent name of orthogroups, combined p-values across species from Fisher's combined probability test, original p-values from differential expression analysis per species and annotations of genes. The rows contain genes that are representative per orthologous group for each species.

Table S5. Summarized clusters based on GO semantic similarity method. This table contains 3 sheets, corresponding to muscle, hippocampus and dietary restriction GO analysis. The file shows the GO enrichments and categorization to higher (more general) GO terms.

Table S6. Proteostasis-linked processes enriched in 2 tissues and dietary restriction experiments. This table contains 3 sheets, corresponding to muscle, hippocampus and dietary restriction GO analysis for proteostasis-linked processes.

Table S7. Significant conserved genes from human GTEx in proteostasis quality network for skeletal muscle and hippocampus. This table contains 6 sheets for each part of the protein quality network (macroautophagy, translation and proteasome complex) per tissue.

Table S8. Summary of the statistics from network analysis. This table contains 5 sheets of the information about the sizes of the all modules and GO and GWAS enrichments in each tissue for proteostasis-linked modules.

Table S9. Summary of mapping the GWAS traits for selected modules. This table contains the gene-level p-values from the PASCAL tool for the heatmap of Figure 5C for selected 22 GWAS age-related studies.

METHODS

Data selection. To obtain a representative set of aging gene expression experiments, a set of raw RNA-seq and microarray datasets of four species (*H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*) were downloaded from the GEO database (Barrett *et al.*, 2013) and SRA database (Leinonen *et al.*, 2011) (Table S1). For observing aging gene expression signatures in human and mouse, we selected hippocampus and skeletal muscle tissues. The aging gene expression experiments for fly and worm were available as whole-body experiments. All the healthy or control samples came from two extreme age groups (young and old adults) that are counted from sexual maturity. This corresponds to 20-30 years old humans, 3-4 months old mice, 4-5 days old flies and 3-6 days old worms (see Figure 1B) in young adults. In old adult age group, this corresponds to 60-70 years old humans, 20-24 months old mice, 40-50 days old flies and 12-14 days old worms. The sample size per age group was 3-6 replicates. The GTEx V6p read counts were used as *H. sapiens* aging experiment (V6p dbGaP accession phs000424.v6.p1, release date: October, 2016). The information about the sample ages was obtained through dbGAP annotation files of the GTEx project (restricted access). Two RNA-seq datasets were matched for *M. musculus* and *C. elegans*; and the microarray platforms included were from Affymetrix: Mouse 430 A/2.0, GeneChip Drosophila Genome array and *C. elegans* Genome array.

GTEx v6p analysis. From the downloaded GTEx V6p data, we extracted the gene read counts values for protein-coding genes by using Ensembl (release 91). For each tissue, the lowly expressed genes were excluded from data analysis according to the GTEx pipeline (Mele *et al.*, 2015). Prior to the age-related differential expression analysis, we used the PEER algorithm (Stegle *et al.*, 2012) in a two-step approach to account for known covariates as well as for hidden factors present in GTEx V6p data per tissue. From covariate files (Brain_Hippocampus_Analysis.covariates.txt and Muscle_Skeletal_Analysis.covariates.txt), we used information about the three genotype principal components. From phenotype file (phs000424.v6.pht002742.v6.p1.c1.GTEx_Subject_Phenotypes.GRU.txt), we used information about age, gender, ischemic time and BMI information. From attribute file (phs000424.v6.pht002743.v6.p1.c1.GTEx_Sample_Attributes.GRU.txt), we extracted information about the sample associations with interested tissues, hippocampus and skeletal muscle. In the first step, the PEER algorithm discovers patterns of common

variation; it created 15 and 35 assumed global hidden factors for hippocampus and skeletal muscle, respectively. In addition to global hidden factors, we provided age, BMI, sex and ischemic time as known covariates in PEER model. In the second step those hidden factors (gene expression principal components) that showed significant Pearson's correlation coefficient with age (p-value < 0.05) were excluded. The number of hidden factors that did not significantly correlate in hippocampus was 7/15 and in skeletal muscle were 22/35 that were selected for further linear model analysis. The sum of remaining hidden factors and known covariates were included in a linear regression model to obtain the genes differentially expressed during age in GTEx V6p data for each tissue (Formula 1).

$$Y_{ji} = \mu_0 + \alpha_j Age_i + \gamma_j Sex_i + \beta_j BMI_i + \theta_j Ishemic\ time_i + \sum_{k=1}^{n} \delta_j PC_{ki} + \epsilon_i \quad [1]$$

where, Y_{ji} is the expression of a gene *j* in a sample *i*, where *Age*, *Sex*, *BMI*, *Ischemic time* of sample *i*, with their regression coefficients α, γ, β, θ. PC_{ki} ($1 < k < n$) is the value of the *k*-hidden factors for the *i*-th sample with regression coefficient δ; *n* is a total number of factors that was not correlated with age, ε_i is the error term, and μ_0 is the regression intercept. If $\alpha > 0$, gene *j* was treated as up-regulated, otherwise gene *j* was treated as down-regulated. The linear model (Formula 1) was performed in *limma voom*, and the p-values were corrected for multiple testing by performing false discovery rate (FDR) correction using Benjamini-Hochberg method.

Aging datasets microarray analysis. For microarray datasets (both aging and caloric restriction experiments) from skeletal muscle of *M. musculus* and whole-body of *D. melanogaster*, raw Affymetrix .CEL files were downloaded from the GEO database and preprocessed using RMA normalization algorithm (Irizarry *et al.*, 2003) (Table S1). In case of multiple probes mapping to the genes on the array, the average of the probes was taken in further analysis. The annotation was used from Ensembl release 91. In order to identify the features that exhibit the most variation in the dataset, principal component analysis (PCA) was performed on the expression matrices to detect outlier samples, gender and other batches.

Aging datasets RNA-seq analysis. For RNA-seq datasets from two model organisms, *M. musculus* and *C. elegans*, the .sra files were downloaded from the SRA database (Leinonen *et al.*, 2011). Both datasets were sequenced on Illumina HiSeq 2000 with read length 50nt. The reads were mapped to species-specific reference genomes (*M. musculus*: GRCm38.p5, *C. elegans*: WBCel235) using kallisto v0.43.1 (for index building: kallisto index –i genome.idx genome.cdna.all.fa (k-mer = 31, default option); for mapping: kallisto quant -i genome.idx –o output.file –single –l 200 –s 20 single.end.fastq.file) (Bray *et al.*, 2016). Both *M. musculus* and *C. elegans* had single-end RNA-seq libraries in the experiments (Table S1.). The transcript abundances were summarized at the gene-level (Soneson, Love and Robinson, 2015). For both species, we used GTF gene annotation files that were downloaded from Ensembl ftp site (release 91) (Aken *et al.*, 2016). The transcript abundances were summarized at the gene-level to lengthscaledTPMs using tximport v1.6.0 (Soneson, Love and Robinson, 2015) and used as an input to *limma voom*. The gene-level read counts were further analyzed in R v3.4.3. The read counts were normalized by total number of all mappable reads (library size) for each gene. The *limma voom* results in a matrix of normalized gene expression values on log2 scale. The counts and normalized log2 *limma voom* expression values were used as a raw input for all the analysis. Outlier samples were checked by principal

678 component analysis. For each species, genes that showed expression below 1 count per
679 million (cpm < 1) in the group of replicates were excluded from downstream analysis.
680
681 **Identification of age-related differentially expressed genes.** To be able to obtain
682 differentially expressed genes from different experiments that were normalized, we had
683 to account for the possible batches present. Since we are not aware of all the batches in
684 the studies, we used Surrogate Variable Analysis (SVA) to correct for batches (Leek and
685 Storey, 2007) in microarray data analysis. The SVA method borrows the information
686 across gene expression levels to estimate the large-scale effects of all factors absent
687 from the model directly from the data. After species-specific expression matrices were
688 corrected, they served as input into linear model analysis implemented in *limma*
689 (Affymetrix) or *limma voom* (RNA-seq) (Law *et al.*, 2014), for finding age-related
690 differentially expressed genes between two extreme aging groups, young and old.
691 Briefly, *limma* uses moderate t-statistics that includes moderated standard errors across
692 genes, therefore effectively borrowing strength from other genes to obtain the inference
693 about each gene. The statistical significance of putatively age-dependent genes was
694 determined with a false discovery rate (FDR) of 10%.
695
696 **Caloric restriction datasets microarray analysis.** The GEO database was used to
697 download caloric restriction datasets (Table S1). Only muscle tissue was available in *H.*
698 *sapiens*, therefore we selected correspondingly muscle tissue in mouse, but whole body
699 in fly and worm. The datasets were normalized using RMA normalization algorithm
700 (Irizarry *et al.*, 2003) (Table S1). In case of multiple probes mapping to the genes on the
701 array, the average of the probes was taken in further analysis. The annotation was used
702 from Ensembl release 91. To call differentially expressed genes, we used *limma*
703 between caloric restriction and control samples. The statistical significance of putatively
704 age-dependent genes was determined with a false discovery rate (FDR) of 5%.
705
706 **Age group alignments between species.** For deriving one-to-one orthologs, human
707 genes were mapped to the homologs in the respective species using biomaRt v2.34.2.
708 After detection of significant age-associated differentially expressed genes, we
709 overlapped one-to-one orthologous genes between the species in order to observe the
710 consistency of age groups between species. We took the *limma voom* corrected
711 expression matrix for GTEx V6p and the expression matrices of model organisms, and
712 selected only genes that were differentially expressed with an FDR of 5%. We then
713 accounted for the laboratory batch effect by applying Combat on expression matrices
714 (Leek *et al.*, 2012).
715
716 **Gene-level analysis.** To examine the relationship between aging in human and model
717 organisms on single-gene level, we mapped one-to-one orthologous genes from human
718 to model organisms and between the organisms downloaded from Ensembl (Aken *et al.*,
719 2016). We calculated Spearman correlations between sets of matched differentially
720 expressed orthologous genes, between log2 fold-changes (Supplementary Figure S2).
721 No cutoff for fold change was used.
722
723 **Constructing homologous quadruplets and enrichment analysis.** We downloaded
724 hierarchical orthologous groups (HOGs, in further text referring to orthologous groups
725 (OG)) across four species from the OMA (orthologous matrix analysis) database
726 (Altenhoff *et al.*, 2015) at the Bilatera level (*Amphimedon queenslandica* (*Cnidaria*) was
727 used as a metazoan outgroup), which resulted in 3232 orthologous groups. Briefly,
728 hierarchical orthologous groups are gene families that contain orthologs (genes related

by speciation) and in-paralogs (genes related by duplication) at the taxonomic level which orthologous groups were defined. The sizes of orthologous groups in this study range from 4 to 246 genes. We filtered age-related genes per orthologous group per species in order to obtain representative species-specific genes per group. The genes within orthologous group were selected according to the *P* values from differentially expression analysis (Rittschof *et al.*, 2014). We applied Bonferroni correction on each orthologous group to the differential expression *P* values in order to correct for the size of the orthologous group. We then combined the corrected differential gene expression *P* values across species using Fisher's combined probability test generating a new *P* value from χ^2 distribution with 2k degrees of freedom (Formula 2).

$$-2\sum_{i=1}^{k}\ln(P_i) \sim \chi^2_{2k} \qquad [2],$$

where P_i is species-specific gene *P* value from differential expression analysis within a OG.

We adjusted combined Fisher *P* values for multiple testing, and filtered orthologous groups with FDR of 10% for further analysis. This resulted in 2010 and 2075 common OGs for skeletal muscle and hippocampus, respectively. In caloric restriction experiments, we detected 1962 common OGs.

We performed general GO enrichment analysis using Fisher's test (topGO R package) on significant orthologous group genes and based on human gene set annotation to find functional enrichment of OGs in GO 'biological process' terms. To summarize the significantly enriched top 100 GO terms into main ones, we used the Wang GO semantic similarity method (Wang *et al.*, 2007) that takes into account the hierarchy of gene ontology, and performed hierarchical clustering (11 clusters for skeletal muscle and 13 clusters for hippocampus, 10 clusters for caloric restriction) on the semantic matrix for both aging and caloric restriction experiments (Table S5). The clusters were then named according to the common term of the cluster. We associated proteostasis-linked processes to GO terms associated with 'translation', 'protein folding', 'proteasome assembly', 'macroautophagy', 'proteasome complex', 'endoplasmic reticulum', 'lysosome' and others.

To perform the randomizations, we selected random genes from the differential expression matrices with the same number as the number of orthologous groups selected for skeletal muscle and hippocampus. The p-values associated with the random genes per species were then combined with the Fisher's combined test. The GO enrichment analysis was performed as for the observed data with focus on the 'biological process' and based on the human annotation. The procedure was repeated 100 times (Figure S6).

Prioritization of OG gene pairs in multi-species co-expression network. We aimed to detect gene sets that are perturbed in aging in different species. We selected the genes from previously formed significant age-related OGs per species and constructed the species-specific co-expression networks by calculating Pearson correlation coefficient between age-related OGs genes. In the resulting species-specific co-expression network, nodes represent genes and edges connect genes that are above a set significant threshold from Pearson correlation calculation (*P* value < 0.05). Only positively correlated genes were taken into account, while the negatively correlated genes and genes correlating under the threshold were set to zero. Negatively correlated genes might be interesting to detect complex regulatory patterns, but are beyond the

scope of this study. The cross-species network was obtained as follows (Stuart *et al.*, 2003). Each co-expression link was assigned a rank within the species according to the Pearson correlation value. We then divided the species-specific ranks by the total number of OGs per tissue to normalize the ranks across the species (Formula 3, example for human, but same for other species).

$r_n = \frac{r_{cxh}}{N_{eog}}$ [3], where r_n is normalized gene pair rank, r_{cxh} is the rank of co-expression link in human and N_{eog} is the number of common evolutionary orthologous groups selected for tissue.

The final gene-pair list was then obtained by integrating human, mouse, fly and worm ranked lists using robust aggregation, originally made for comparing two lists (Kolde *et al.*, 2012). Briefly, using beta probability distribution on order statistics, we asked how probable is the co-expression link by taking into account the ranks of all four species. This method assigns a *P* value to each co-expression link in an aggregated list, indicating how much better it is ranked compared to the null model (random ordering). This yielded networks with 2887 and 3353 significant gene-pairs (edges) (*P* value < 0.001) for skeletal muscle and hippocampus, respectively.
To confirm that the integrated age-related multi-species networks are significant, we selected randomly collected genes from each species. The numbers of selected genes was the same as in the OGs. We then formed the quadruplets and performed the same integration analysis as before. We repeated the procedure 100 times, and obtained 100 randomly integrated multi-species networks (Figure S7). In both cases, random and original analysis, the annotation was based on human.

Clustering the integrated cross-species network. In order to identify aging-associated functional modules, we created networks containing 1142 nodes (2887 edges) in skeletal muscle and 1098 nodes (3353 edges) in hippocampus, from our prioritized gene pair list based on orthology and all edges between them. The negative logarithm (base 10) of *P* values from aggregated list was assigned as edge weights in both integrated networks. We decomposed the skeletal muscle and hippocampus integrated networks into components and the further analysis was restricted to analysis of a giant component. The giant component contained 1050 genes (nodes) in skeletal muscle and 1067 genes (nodes) in hippocampus. As before, we used human annotation. The modules within the cross-species networks of each tissue were obtained by using a multilevel community algorithm that takes into account edge weights (Yang, Algesheimer and Tessone, 2016) from igraph (Csárdi & Nepusz 2006). Briefly, the multilevel algorithm (Blondel *et al.*, 2008) takes into account each node as its own and assigns it to the community with which it achieves the highest contribution to modularity. To obtain Figure 4, we summarized groups of module nodes to single meta-nodes according to their multilevel-algorithm calculated module membership, and showed the inter-modular connectivity using a circular layout. We selected the modules with size greater than 10, which returned 12 modules per tissue-specific cross-species network. We checked the functional enrichment of genes within selected modules in every network using Gene Ontology through topGO R package (See Figure 4).
Moreover, we downloaded the pre-calculated file of gene-level summary statistics from 37 GWASs from the Pascal method (Lamparter *et al.*, 2016). We selected 22 out of 37 GWAS studies (Marbach *et al.*, 2016) (Table S9) that are associated with metabolic and neurological age-related diseases. To perform enrichment of the module genes within GWAS age-related diseases categories, we selected top-ranking genes (GWAS gene

score < 0.1) within each disease and formed the categories for enrichment. We ran enrichment analysis on final network modules to find disease-related modules (adjusted p-value < 0.2). The human genome was used as a background gene set.
Finally, we used Kleinberg's hub centrality score to determine the hub genes within interested modules and observed the hub-gene neighborhood. The final genes were then selected to show their *P* value association within GWAS studies (Figure 5C, Table S9).

LXS and BXD mouse data. Male and female mice from those strains were fed with normal ad libitum diet, and median and maximum lifespan were calculated to represent longevity across strains. Microarray data as well as lifespan data were downloaded from GeneNetwork.org. Microarray data from prefrontal cortex of LXS mice was generated by Dr. Michael Miles using animals with the average age of 72 days (GN Accession: GN130). Microarray data from spleen of BXD mice was generated by Dr. Robert W. Williams using animals with the average age of 78 days (GN Accession: GN283). Microarray data from hippocampus of BXD mice was generated by Dr. Gerd Kempermann and Dr. Robert W. Williams using animals with the average age of 70 days (GN Accession: GN110). To correct for the population structure within the strains, a linear mixed model approach was applied. For enrichment analysis, genes were ranked based on their Pearson correlation coefficients with the lifespan data of the BXD strains, and Gene Set Enrichment Analysis (GSEA) was performed to find the enriched gene sets correlated with the lifespan (Subramanian et al., 2005).

Author Contributions

Conceptualization, A.K. and M.R.R.; Methodology, A.K.; Investigation, A.K.; Preprocessing datasets: A.K, Formal Analysis, A.K.; Validation analysis: H.L.; Writing – Original Draft, A.K.; Writing – Review & Editing, A.K., H.L., V.S., J.A., Z.K. and M.R.R.; Funding Acquisition, M.R.R.; Supervision, M.R.R.
The authors declare that they have no conflict of interest.

Acknowledgments

We thank Natasha Glover for help with the OMA database. We thank AgingX collaborators, Matthias Morf and all members of the Robinson-Rechavi group for helpful comments. We thank Thomas Flatt for critical reading of the manuscript. We are grateful to the teams that made their data available through Gene expression omnibus (GEO) and the database of Genotypes and Phenotypes (dbGAP) repositories. The computations were performed at the Vital-IT (http://www.vital-it.ch) center for high-performance computing of the SIB Swiss Institute of Bioinformatics. This work was supported by the AgingX project from SystemsX.ch, the EPFL, the Velux Stiftung (J.A.), and the NIH (R01AG043930).

References

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., Vogel, J.-H., White, S., Zadissa, A., Flicek, P. and Searle, S. M. J. (2016) 'The Ensembl gene annotation system', *Database*. Oxford University Press, 2016, p. baw093. doi: 10.1093/database/baw093.

Altenhoff, A. M., kunca, N., Glover, N., Train, C.-M., Sueki, A., Pili ota, I., Gori, K., Tomiczek, B., Muller, S., Redestig, H., Gonnet, G. H. and Dessimoz, C. (2015) 'The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements', *Nucleic Acids Research*. Oxford University Press, 43(D1), pp. D240–D249. doi: 10.1093/nar/gku1158.

Andreux, P. A., Wiliams, E. G., Koutnikova, H., Houtkooper, R. H., Champy, M.-F., Hugues, H., Schoonjans, K., Williams, R. W. and Auwerx, J. (2012) 'Systems Genetics of Metabolism: The Use of the BXD Murine Reference Panel for Multiscalar Integration of Traits', *Cell*. Cell Press, 150(6), pp. 1287–1299. doi: 10.1016/J.CELL.2012.08.012.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. and Soboleva, A. (2013) 'NCBI GEO: Archive for functional genomics data sets - Update', *Nucleic Acids Research*, 41(D1), pp. 991–995. doi: 10.1093/nar/gks1193.

Bass, T. M., Grandison, R. C., Wong, R., Martinez, P., Partridge, L. and Piper, M. D. W. (2015) 'Europe PMC Funders Group Optimization of Dietary Restriction Protocols in Drosophila', 62(10), pp. 1071–1081.

Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V. and Robinson-Rechavi, M. (2008) 'Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species', in *Data Integration in the Life Sciences*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 124–131. doi: 10.1007/978-3-540-69828-9_12.

Baumgart, M., Priebe, S., Groth, M., Hartmann, N., Menzel, U., Pandolfini, L., Koch, P., Felder, M., Ristow, M., Englert, C., Guthke, R., Platzer, M. and Cellerino, A. (2016) 'Longitudinal RNA-Seq Analysis of Vertebrate Aging Identifies Mitochondrial Complex I as a Small-Molecule-Sensitive Modifier of Lifespan', *Cell Systems*. Elsevier, 2(2), pp. 122–132. doi: 10.1016/j.cels.2016.01.014.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008) 'Fast unfolding of communities in large networks'. doi: 10.1088/1742-5468/2008/10/P10008.

Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L. (2016) 'Near-optimal probabilistic RNA-seq quantification', *Nature Biotechnology*, 34(5), pp. 525–527. doi: 10.1038/nbt.3519.

Budovsky, A., Abramovich, A., Cohen, R., Chalifa-Caspi, V. and Fraifeld, V. (2006) 'Longevity network: Construction and implications'. doi: 10.1016/j.mad.2006.11.018.

Chondrogianni, N., Georgila, K., Kourtis, N., Tavernarakis, N. and Gonos, E. S. (2015) '20S proteasome activation promotes life span extension and resistance to

914 proteotoxicity in Caenorhabditis elegans.', *FASEB journal : official publication of the*
915 *Federation of American Societies for Experimental Biology*. Federation of American
916 Societies for Experimental Biology, 29(2), pp. 611–22. doi: 10.1096/fj.14-252189.
917 Csárdi, G. and Nepusz, T. (no date) 'The igraph software package for complex
918 network research'. Available at:
919 http://www.necsi.edu/events/iccs6/papers/c1602a3c126ba822d0bc4293371c.pdf
920 (Accessed: 2 April 2017).
921 D 'amico, D., Sorrentino, V. and Auwerx, J. (2017) 'Cytosolic Proteostasis Networks
922 of the Mitochondrial Stress Response'. doi: 10.1016/j.tibs.2017.05.002.
923 van Dam, S., Võsa, U., van der Graaf, A., Franke, L. and de Magalhães, J. P. (2017)
924 'Gene co-expression analysis for functional classification and gene–disease
925 predictions', *Briefings in Bioinformatics*, p. bbw139. doi: 10.1093/bib/bbw139.
926 Depuydt, G., Xie, F., Petyuk, V. A., Shanmugam, N., Smolders, A., Dhondt, I., Brewer, H.
927 M., Camp, D. G., Smith, R. D. and Braeckman, B. P. (2013) 'Reduced insulin/insulin-
928 like growth factor-1 signaling and dietary restriction inhibit translation but
929 preserve muscle mass in Caenorhabditis elegans.', *Molecular & cellular proteomics :*
930 *MCP*. American Society for Biochemistry and Molecular Biology, 12(12), pp. 3624–
931 39. doi: 10.1074/mcp.M113.027383.
932 Driscoll, I., Hamilton, D. A., Petropoulos, H., Yeo, R. A., Brooks, W. M., Baumgartner, R.
933 N. and Sutherland, R. J. (2003) 'The Aging Hippocampus: Cognitive, Biochemical and
934 Structural Findings', *Cerebral Cortex*. Oxford University Press, 13(12), pp. 1344–
935 1351. doi: 10.1093/cercor/bhg081.
936 Fernandes, M., Wan, C., Tacutu, R., Barardo, D., Rajput, A., Wang, J., Thoppil, H.,
937 Thornton, D., Yang, C., Freitas, A. and de Magalhães, J. P. (2016) 'Systematic analysis
938 of the gerontome reveals links between aging and age-related diseases', *Human*
939 *Molecular Genetics*. Oxford University Press, 25(21), p. ddw307. doi:
940 10.1093/hmg/ddw307.
941 Finkel, T. (2015) 'The metabolic regulation of aging', *Nature Medicine*, 21, pp. 1416–
942 1423. Available at:
943 https://www.nature.com/nm/journal/v21/n12/pdf/nm.3998.pdf (Accessed: 23
944 May 2017).
945 Flurkey, K., M. Currer, J. and Harrison, D. E. (2007) 'Chapter 20 – Mouse Models in
946 Aging Research', in *The Mouse in Biomedical Research*, pp. 637–672. doi:
947 10.1016/B978-012369454-6/50074-1.
948 Fontana, L., Partridge, L., Longo, V. D. and Longo4, V. D. (2010) 'Extending Healthy
949 Life Span—From Yeast to Humans IHHl^H^HHHilH^^^HKJSBI^ Extending Healthy
950 Life Span -From Yeast to Humans', *Source: Science, New Series*, 328(5976), pp. 321–
951 326. doi: 10.1126/science.1172539.
952 Fushan, A. A., Turanov, A. A., Lee, S.-G., Kim, E. B., Lobanov, A. V, Yim, S. H.,
953 Buffenstein, R., Lee, S.-R., Chang, K.-T., Rhee, H., Kim, J.-S., Yang, K.-S. and Gladyshev,
954 V. N. (2015) 'Gene expression defines natural changes in mammalian lifespan.',
955 *Aging cell*, 14(3), pp. 352–65. doi: 10.1111/acel.12283.
956 Fushan, A. A., Turanov, A. A., Lee, S. G., Kim, E. B., Lobanov, A. V., Yim, S. H.,
957 Buffenstein, R., Lee, S. R., Chang, K. T., Rhee, H., Kim, J. S., Yang, K. S. and Gladyshev, V.
958 N. (2015) 'Gene expression defines natural changes in mammalian lifespan', *Aging*
959 *Cell*, 14(3), pp. 352–365. doi: 10.1111/acel.12283.

960 Gems, D. and Partridge, L. (2013) 'Genetics of Longevity in Model Organisms:
961 Debates and Paradigm Shifts', *Annu. Rev. Physiol*, 75, pp. 621–44. doi:
962 10.1146/annurev-physiol-030212-183712.
963 Gene Ontology Consortium, T., Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D.,
964 Butler, H., Michael Cherry, J., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris,
965 M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E.,
966 Ringwald, M., Rubin, G. M. and Sherlock, G. (2000) 'Gene Ontology: tool for the
967 unification of biology', *Nature genetics*, 25(1), pp. 25–29. Available at:
968 https://www.nature.com/articles/ng0500_25.pdf (Accessed: 20 November 2017).
969 Gomes, A. V (2013) 'Genetics of proteasome diseases.', *Scientifica*. Hindawi, 2013, p.
970 637629. doi: 10.1155/2013/637629.
971 Greene, V. W. (2001) 'Personal hygiene and life expectancy improvements since
972 1850: Historic and epidemiologic associations', *American Journal of Infection*
973 *Control*. Mosby, 29(4), pp. 203–206. doi: 10.1067/MIC.2001.115686.
974 Hasin, Y., Seldin, M. and Lusis, A. (2017) 'Multi-omics approaches to disease',
975 *Genome Biology*, 18(1), p. 83. doi: 10.1186/s13059-017-1215-1.
976 Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. (2003)
977 'Summaries of Affymetrix GeneChip probe level data', *Nucleic acids research*, 31(4),
978 p. e15. doi: 10.1093/nar/gng015.
979 Jones, O. R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C. G., Schaible, R., Casper,
980 B. B., Dahlgren, J. P., Ehrlén, J., García, M. B., Menges, E. S., Quintana-Ascencio, P. F.,
981 Caswell, H., Baudisch, A. and Vaupel, J. W. (2014) 'Diversity of ageing across the tree
982 of life'. doi: 10.1038/nature12789.
983 Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J. and Eskin,
984 E. (2008) 'Efficient Control of Population Structure in Model Organism Association
985 Mapping', *Genetics*, 178(3), pp. 1709–1723. doi: 10.1534/genetics.107.080101.
986 Kaushik, S. and Cuervo, A. M. (2015) 'Proteostasis and aging', *Nature Medicine*.
987 Nature Research, 21(12), pp. 1406–1415. doi: 10.1038/nm.4001.
988 Khodakarami, A., Saez, I., Mels, J. and Vilchez, D. (2015) 'Mediation of organismal
989 aging and somatic proteostasis by the germline.', *Frontiers in molecular biosciences*.
990 Frontiers Media SA, 2, p. 3. doi: 10.3389/fmolb.2015.00003.
991 Kolde, R., Laur, S., Adler, P. and Vilo, J. (2012) 'Robust rank aggregation for gene list
992 integration and meta-analysis', *Bioinformatics*. Oxford University Press, 28(4), pp.
993 573–580. doi: 10.1093/bioinformatics/btr709.
994 Labbadia, J., Brielmann, R. M., Neto, M. F., Lin, Y.-F., Haynes, C. M. and Morimoto, R. I.
995 (2017) 'Mitochondrial Stress Restores the Heat Shock Response and Prevents
996 Proteostasis Collapse during Aging', *Cell Reports*. Cell Press, 21(6), pp. 1481–1494.
997 doi: 10.1016/J.CELREP.2017.10.038.
998 Labbadia, J. and Morimoto, R. I. (2015) 'The Biology of Proteostasis in Aging and
999 Disease', *Annual Review of Biochemistry*. Annual Reviews , 84(1), pp. 435–464. doi:
1000 10.1146/annurev-biochem-060614-033955.
1001 Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., Bergmann, S. and Kutalik, Z.
1002 (2016) 'Fast and Rigorous Computation of Gene and Pathway Scores from SNP-
1003 Based Summary Statistics', *PLOS Computational Biology*. Edited by J. Listgarten.
1004 Public Library of Science, 12(1), p. e1004714. doi: 10.1371/journal.pcbi.1004714.
1005 Lanna, A., Gomes, D. C. O., Muller-Durovic, B., McDonnell, T., Escors, D., Gilroy, D. W.,

Lee, J. H., Karin, M. and Akbar, A. N. (2017) 'A sestrin-dependent Erk–Jnk–p38 MAPK activation complex inhibits immunity during aging', *Nature Immunology*. Nature Research, 18(3), pp. 354–363. doi: 10.1038/ni.3665.

Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014) 'voom: precision weights unlock linear model analysis tools for RNA-seq read counts', *Genome Biology*, 15, p. R29. doi: 10.1186/gb-2014-15-2-r29.

Lee, G. D., Wilson, M. A., Zhu, M., Wolkow, C. A., De Cabo, R., Ingram, D. K. and Zou, S. (2006) 'Dietary deprivation extends lifespan in Caenorhabditis elegans', *Aging Cell*, 5(6), pp. 515–524. doi: 10.1111/j.1474-9726.2006.00241.x.

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. and Storey, J. D. (2012) 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinformatics*. Oxford University Press, 28(6), pp. 882–883. doi: 10.1093/bioinformatics/bts034.

Leek, J. T. and Storey, J. D. (2007) 'Capturing heterogeneity in gene expression studies by surrogate variable analysis', *PLoS Genetics*, 3(9), pp. 1724–1735. doi: 10.1371/journal.pgen.0030161.

Leinonen, R., Sugawara, H., Shumway, M. and International Nucleotide Sequence Database Collaboration (2011) 'The sequence read archive.', *Nucleic acids research*. Oxford University Press, 39(Database issue), pp. D19-21. doi: 10.1093/nar/gkq1019.

Liao, C. Y., Rikke, B. A., Johnson, T. E., Diaz, V. and Nelson, J. F. (2010) 'Genetic variation in the murine lifespan response to dietary restriction: From life extension to life shortening', *Aging Cell*, 9(1), pp. 92–95. doi: 10.1111/j.1474-9726.2009.00533.x.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. and Kroemer, G. (2013) 'XThe hallmarks of aging', *Cell*, 153(6). doi: 10.1016/j.cell.2013.05.039.

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 (Oxford, England)*, 25(7), pp. 875–81. doi: 10.1093/bioinformatics/btp073.

Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. and Bergmann, S. (2016) 'Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases.', *Nature methods*, 13(January), pp. 1–44. doi: 10.1038/nmeth.3799.

Marzetti, E. and Leeuwenburgh, C. (2006) 'Skeletal muscle apoptosis, sarcopenia and frailty at old age', *Experimental Gerontology*, 41(12), pp. 1234–1238. doi: 10.1016/j.exger.2006.08.011.

Mattison, J. A., Colman, R. J., Beasley, T. M., Allison, D. B., Kemnitz, J. W., Roth, G. S., Ingram, D. K., Weindruch, R., de Cabo, R., Anderson, R. M., Gibbs, R. A., Zimin, A. V., Bowden, D. M., Williams, D. D., Uno, H., Hudson, J. C., Baum, S. T., Frye, D. M., Roecker, E. B., Kemnitz, J. W., Pandya, J. D., Ramsey, J. J., Laatsch, J. L., Kemnitz, J. W., Ngwenya, L. B., Heyworth, N. C., Shwe, Y., Moore, T. L., Rosene, D. L., Masoro, E. J., Yu, B. P., Bertrand, H. A., Weindruch, R., Walford, R. L., Fligiel, S., Guthrie, D., Masoro, E. J., Gems, D., Partridge, L., Guarente, L., Houthoofd, K., Vanfleteren, J. R., Kennedy, B. K., Steffen, K. K., Kaeberlein, M., Mair, W., Dillin, A., Bodkin, N. L., Alexander, T. M., Ortmeyer, H. K., Johnson, E., Hansen, B. C., Colman, R. J., Colman, R. J., Mattison, J. A., Weindruch, R., Walford, R. L., Lane, M. A., Kemnitz, J. W., Clarke, M. R., O'Neil, J. A.,

Kanthaswamy, S., Genome, S. R. M., Colman, R. J., Anderson, R. M., Hadfield, R. M., Mattison, J. A., Ottinger, M. A., Powell, D., Longo, D. L., Ingram, D. K., Kemnitz, J. W., Kirkland, J. L., Peterson, C., Hansen, B. C., Bodkin, N. L., Rodriguez, N. A., Uno, H., Laakso, M., Kuusisto, J., Dailey, G., Polewski, M. A., Grove, K. L., Fried, S. K., Greenberg, A. S., Xiao, X. Q., Clegg, D. J., Fontana, L., Heilbronn, L. K., Lefevre, M., Racette, S. B., Redman, L. M., Weiss, E. P., Edwards, I. J., Gresl, T. A., Rezzi, S., Yamada, Y., Kennedy, B. K., Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., Kroemer, G. and Saville, D. J. (2017) 'Caloric restriction improves health and survival of rhesus monkeys', *Nature Communications*. Nature Publishing Group, 8(May 2016), p. 14063. doi: 10.1038/ncomms14063.

McCarroll, S. A., Murphy, C. T., Zou, S., Pletcher, S. D., Chin, C.-S., Jan, Y. N., Kenyon, C., Bargmann, C. I. and Li, H. (2004) 'Comparing genomic expression patterns across species identifies shared transcriptional profile in aging.', *Nature genetics*, 36(2), pp. 197–204. doi: 10.1038/ng1291.

Mcdaid, A. F., Joshi, P. K., Porcu, E., Komljenovic, A., Li, H., Sorrentino, V., Litovchenko, M., Bevers, R. P. J., Rüeger, S., Reymond, A., Bochud, M., Deplancke, B., Williams, R. W., Robinson-Rechavi, M., Paccaud, F., Rousson, V., Auwerx, J., Wilson, J. F. and Kutalik, Z. (2017) 'ARTICLE Bayesian association scan reveals loci associated with human lifespan and linked biomarkers', *Nature Communications*, 8. doi: 10.1038/ncomms15842.

Mele, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., Johnson, R., Segre, A. V., Djebali, S., Niarchou, A., Consortium, T. G., Wright, F. A., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E. T., Ardlie, K. G. and Guigo, R. (2015) 'The human transcriptome across tissues and individuals', *Science*, 348(6235), pp. 660–665. doi: 10.1126/science.aaa0355.

Morimoto, R. I. and Cuervo, A. M. (2014) 'Proteostasis and the Aging Proteome in Health and Disease', *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. Oxford University Press, 69(Suppl 1), pp. S33–S38. doi: 10.1093/gerona/glu049.

Pyo, J.-O., Yoo, S.-M., Ahn, H.-H., Nah, J., Hong, S.-H., Kam, T.-I., Jung, S. and Jung, Y.-K. (2013) 'Overexpression of Atg5 in mice activates autophagy and extends lifespan', *Nature Communications*. Nature Publishing Group, 4, p. ncomms3300. doi: 10.1038/ncomms3300.

Rappuoli, R., Mandl, C. W., Black, S. and De Gregorio, E. (2011) 'Vaccines for the twenty-first century society', *Nature Reviews Immunology*. Nature Publishing Group, 11(12), p. nri3085. doi: 10.1038/nri3085.

Rittschof, C. C., Bukhari, S. A., Sloofman, L. G., Troy, J. M., Caetano-Anollés, D., Cash-Ahmed, A., Kent, M., Lu, X., Sanogo, Y. O., Weisner, P. A., Zhang, H., Bell, A. M., Ma, J., Sinha, S., Robinson, G. E., Stubbs, L., Kelly, D. and White, K. P. (2014) 'Neuromolecular responses to social challenge: Common mechanisms across mouse, stickleback fish, and honey bee', *Proceedings of the national Academy of Sciences*, 111(50), pp. 17929–17934. doi: 10.1073/pnas.1420369111.

Rodriguez, K. A., Edrey, Y. H., Osmulski, P., Gaczynska, M. and Buffenstein, R. (2012) 'Altered Composition of Liver Proteasome Assemblies Contributes to Enhanced Proteasome Activity in the Exceptionally Long-Lived Naked Mole-Rat', *PLoS ONE*.

Edited by J. L. Brodsky, 7(5), p. e35890. doi: 10.1371/journal.pone.0035890.

Roy, A. K., Oh, T., Rivera, O., Mubiru, J., Song, C. S. and Chatterjee, B. (2002) 'Impacts of transcriptional regulation on aging and senescence.', *Ageing research reviews*, 1(3), pp. 367–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12067592 (Accessed: 23 May 2017).

Schmidt, M. and Finley, D. (2014) 'Regulation of proteasome activity in health and disease.', *Biochimica et biophysica acta*. NIH Public Access, 1843(1), pp. 13–25. doi: 10.1016/j.bbamcr.2013.08.012.

Selman, C. and Hempenstall, S. (2012) 'Evidence of a metabolic memory to early-life dietary restriction in male C57BL/6 mice.', *Longevity & healthspan*, 1, p. 2. doi: 10.1186/2046-2395-1-2.

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., Promislow, D. E. L., Thomas, J. H., Kaeberlein, M. and Kennedy, B. K. (2008) 'Quantitative evidence for conserved longevity pathways between divergent eukaryotic species', *Genome Research*, 18(4), pp. 564–570. doi: 10.1101/gr.074724.107.

Soneson, C., Love, M. I. and Robinson, M. D. (2015) 'Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.', *F1000Research*, 4(0), p. 1521. doi: 10.12688/f1000research.7563.2.

Sorrentino, V., Menzies, K. J. and Auwerx, J. (2018) 'Repairing Mitochondrial Dysfunction in Disease', *Annual Review of Pharmacology and Toxicology*. Annual Reviews , 58(1), pp. 353–389. doi: 10.1146/annurev-pharmtox-010716-104908.

Sorrentino, V., Romani, M., Mouchiroud, L., Beck, J. S., Zhang, H., D'Amico, D., Moullan, N., Potenza, F., Schmid, A. W., Rietsch, S., Counts, S. E. and Auwerx, J. (2017) 'Enhancing mitochondrial proteostasis reduces amyloid-β proteotoxicity', *Nature*. Nature Publishing Group, 552(7684), p. 187. doi: 10.1038/nature25143.

Southworth, L. K., Owen, A. B. and Kim, S. K. (2009) 'Aging mice show a decreasing correlation of gene expression within genetic modules', *PLoS Genetics*, 5(12). doi: 10.1371/journal.pgen.1000776.

Steffen, K. K. and Dillin, A. (2016) 'A Ribosomal Perspective on Proteostasis and Aging'. doi: 10.1016/j.cmet.2016.05.013.

Stegle, O., Parts, L., Piipari, M., Winn, J. and Durbin, R. (2012) 'Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses.', *Nature protocols*. Europe PMC Funders, 7(3), pp. 500–7. doi: 10.1038/nprot.2011.457.

Stuart, J. M., Segal, E., Koller, D. and Kim, S. K. (2003) 'R ESEARCH A RTICLES A Gene-Coexpression Network', *October*, 302(October), pp. 249–255. doi: 10.1126/science.1087447.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. a, Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–50. doi: 10.1073/pnas.0506580102.

The Gene Ontology Consortium (2017) 'Expansion of the Gene Ontology knowledgebase and resources', *Nucleic Acids Research*, 45(D1), pp. D331–D338. doi: 10.1093/nar/gkw1108.

Tian, X., Seluanov, A. and Gorbunova, V. (2017) 'Molecular Mechanisms Determining Lifespan in Short- and Long-Lived Species.', *Trends in endocrinology and metabolism: TEM*. Elsevier, 28(10), pp. 722–734. doi: 10.1016/j.tem.2017.07.004.

Tomaru, U., Takahashi, S., Ishizu, A., Miyatake, Y., Gohda, A., Suzuki, S., Ono, A., Ohara, J., Baba, T., Murata, S., Tanaka, K. and Kasahara, M. (2012) 'Decreased Proteasomal Activity Causes Age-Related Phenotypes and Promotes the Development of Metabolic Abnormalities', *The American Journal of Pathology*, 180(3), pp. 963–972. doi: 10.1016/j.ajpath.2011.11.012.

Walther, D. M., Kasturi, P., Zheng, M., Pinkert, S., Vecchi, G., Ciryam, P., Morimoto, R. I., Dobson, C. M., Vendruscolo, M., Mann, M. and Hartl, F. U. (2015) 'Widespread Proteome Remodeling and Aggregation in Aging C. elegans', *Cell*. Elsevier, 161(4), pp. 919–932. doi: 10.1016/j.cell.2015.03.032.

Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. and Chen, C.-F. (2007) 'A new method to measure the semantic similarity of GO terms', *Bioinformatics*, 23(10), pp. 1274–1281. doi: 10.1093/bioinformatics/btm087.

Woodruff, R. C. and Thompson, J. N. (2003) 'The Role of Somatic and Germline Mutations in Aging and a Mutation Interaction Model of Aging', *Journal of Anti-Aging Medicine*, 6(1), pp. 29–39. doi: 10.1089/109454503765361560.

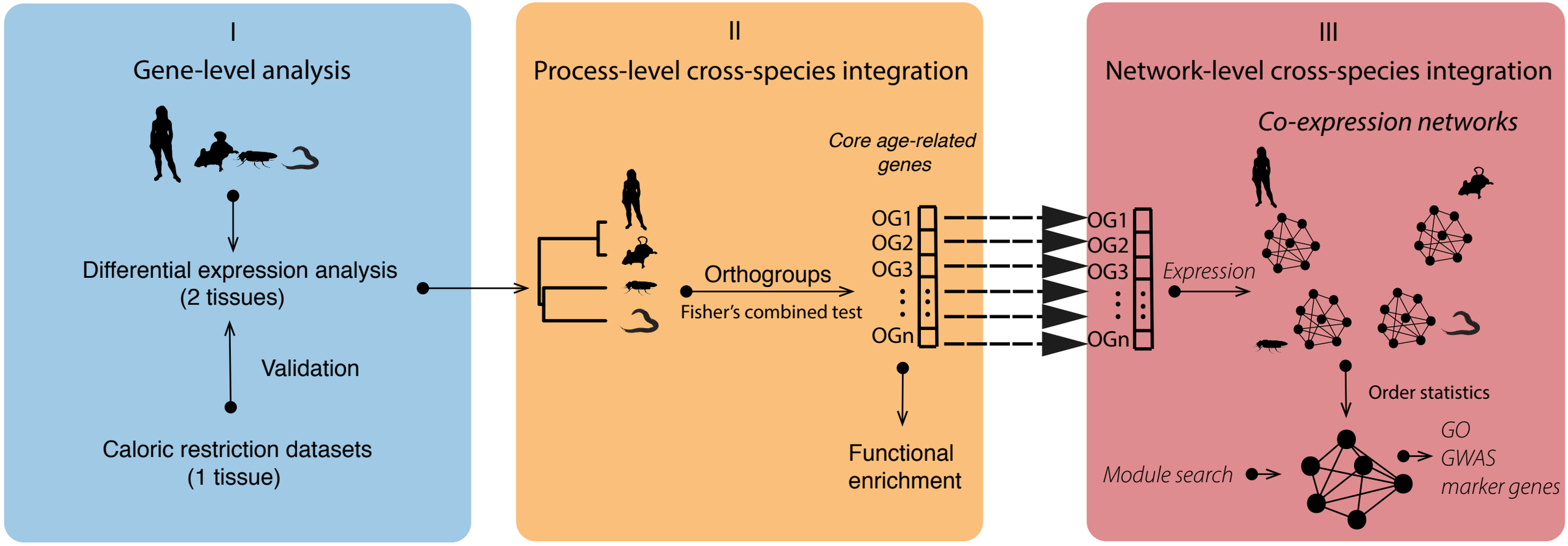
Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S., Zhang, Z., Hou, L., Zhang, Q., Zhang, Y. and Han, J.-D. J. (2007) 'A modular network model of aging.', *Molecular systems biology*, 3(147), p. 147. doi: 10.1038/msb4100189.

Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., Zhao, Y., Mobbs, C. V, Schadt, E. E., Zhu, J. and Tu, Z. (2015) 'Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases.', *Scientific reports*, 5(October), p. 15145. doi: 10.1038/srep15145.

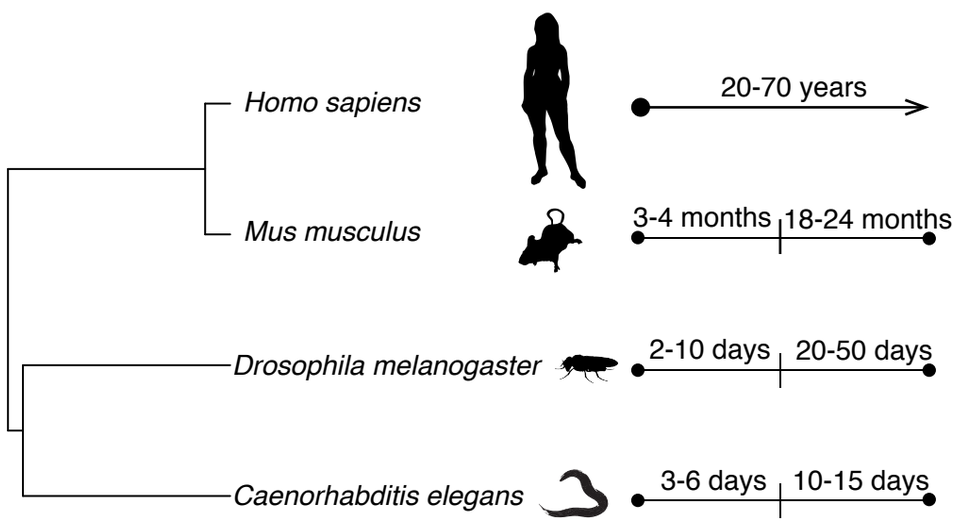
Yang, Z., Algesheimer, R. and Tessone, C. J. (2016) 'A Comparative Analysis of Community Detection Algorithms on Artificial Networks', *Scientific Reports*. Nature Publishing Group, 6(July), p. 30750. doi: 10.1038/srep30750.

Zahn, J. M., Sonu, R., Vogel, H., Crane, E., Mazan-Mamczarz, K., Rabkin, R., Davis, R. W., Becker, K. G., Owen, A. B. and Kim, S. K. (2006) 'Transcriptional Profiling of Aging in Human Muscle Reveals a Common Aging Signature', *PLoS Genetics*. McGraw-Hill, 2(7), p. e115. doi: 10.1371/journal.pgen.0020115.

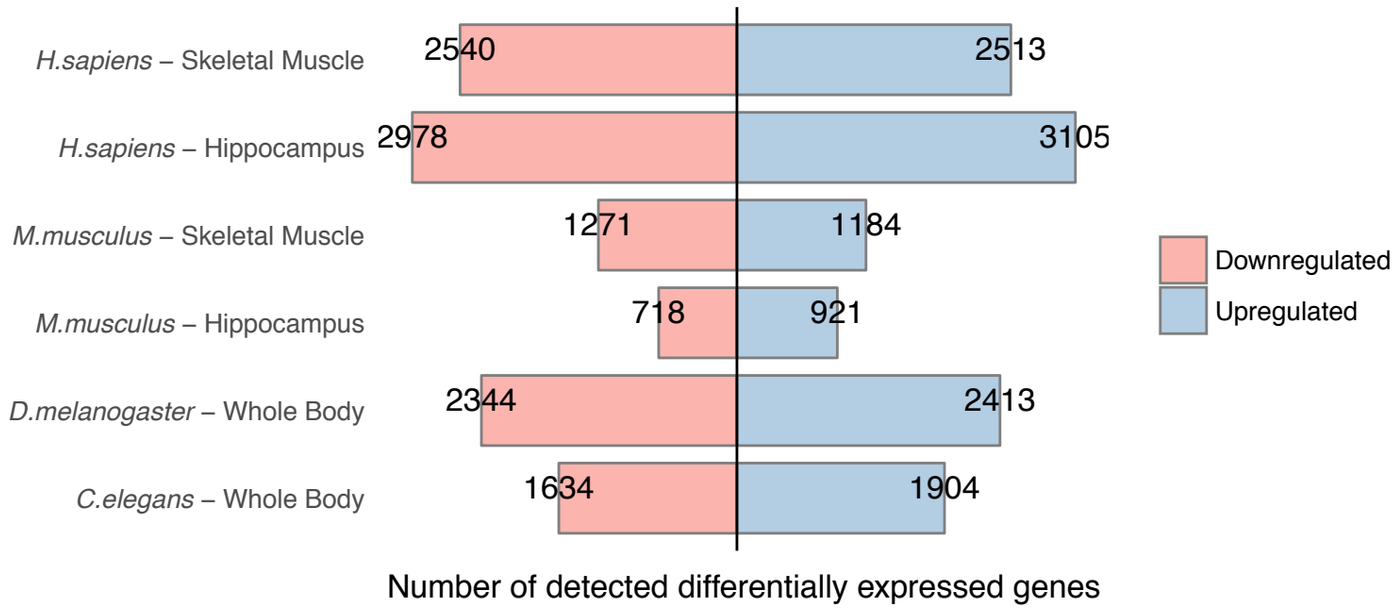
A



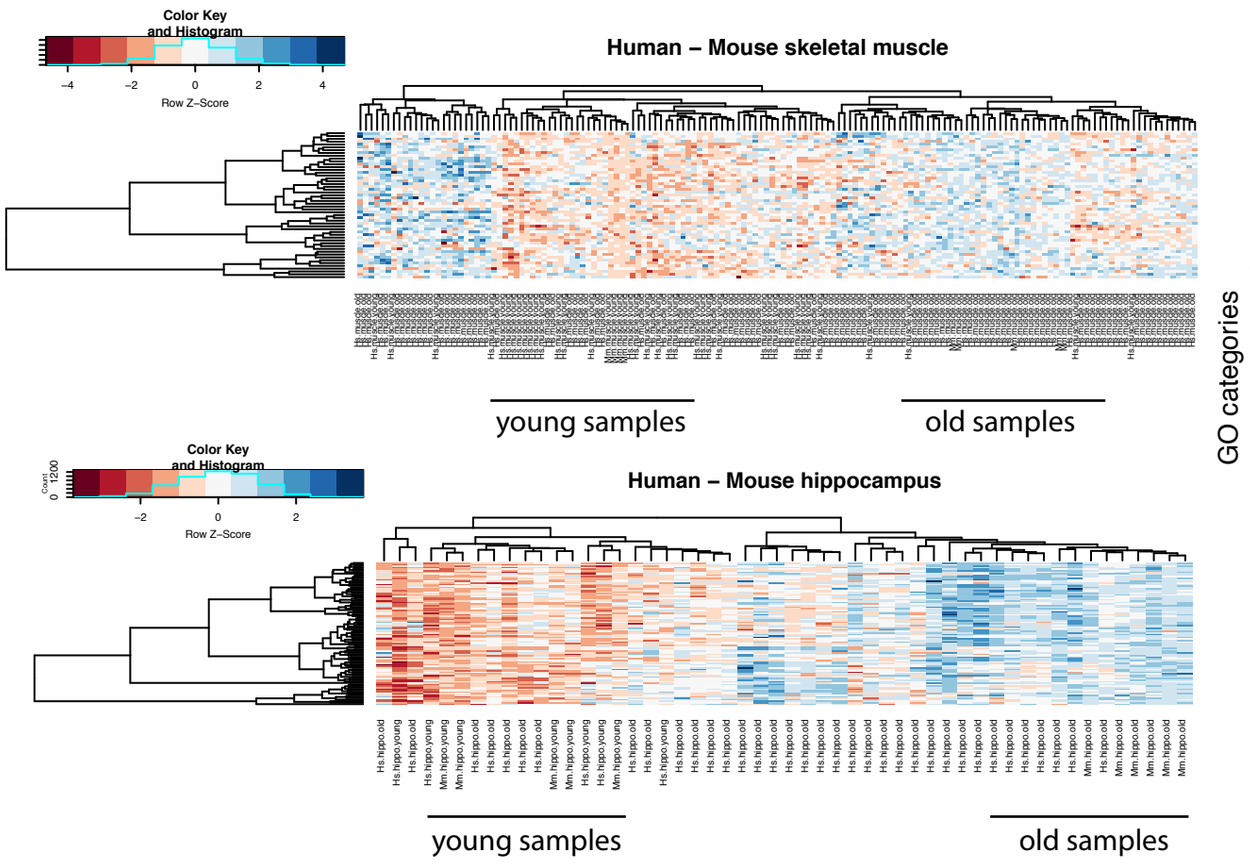
B



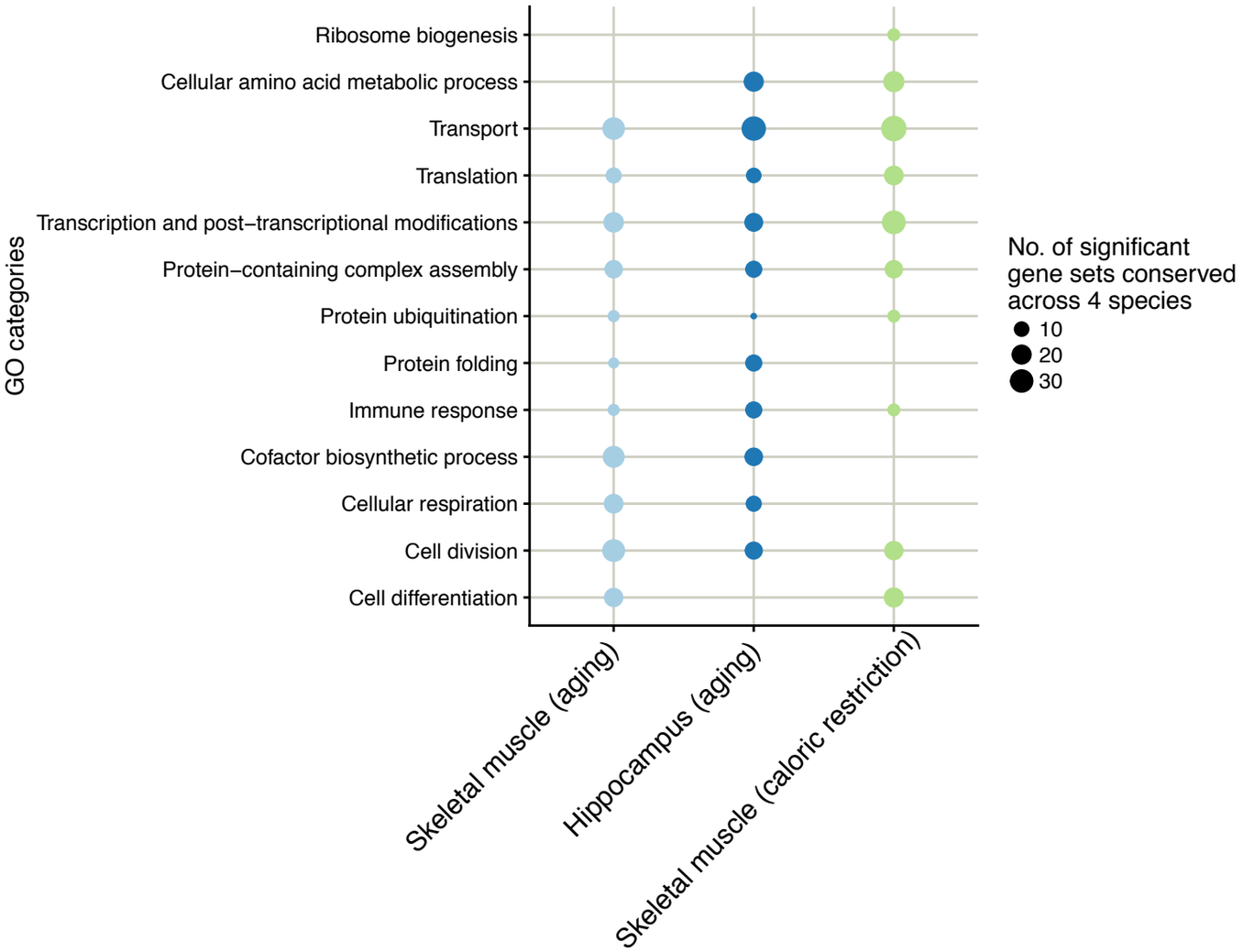
C



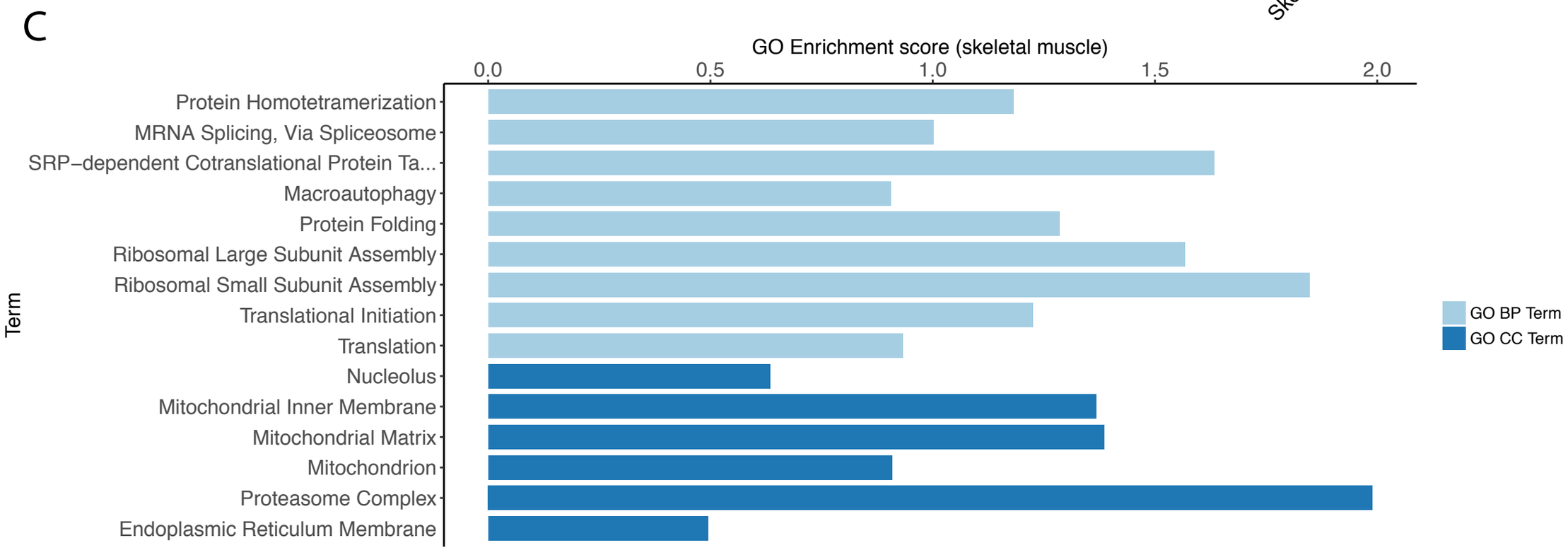
A



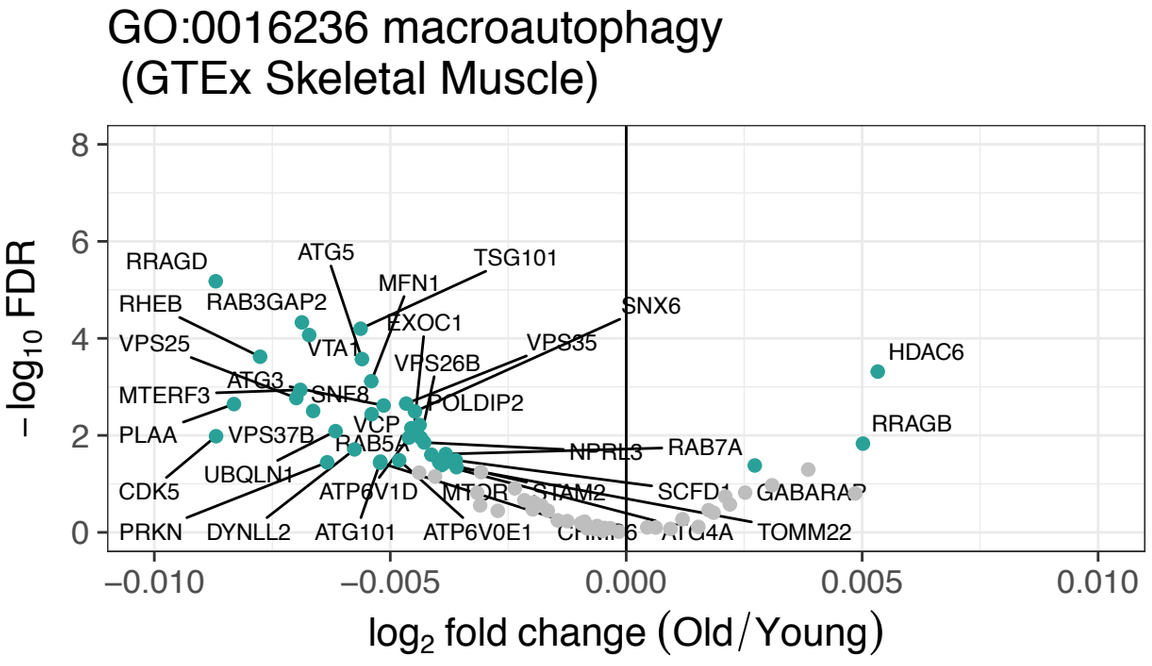
B



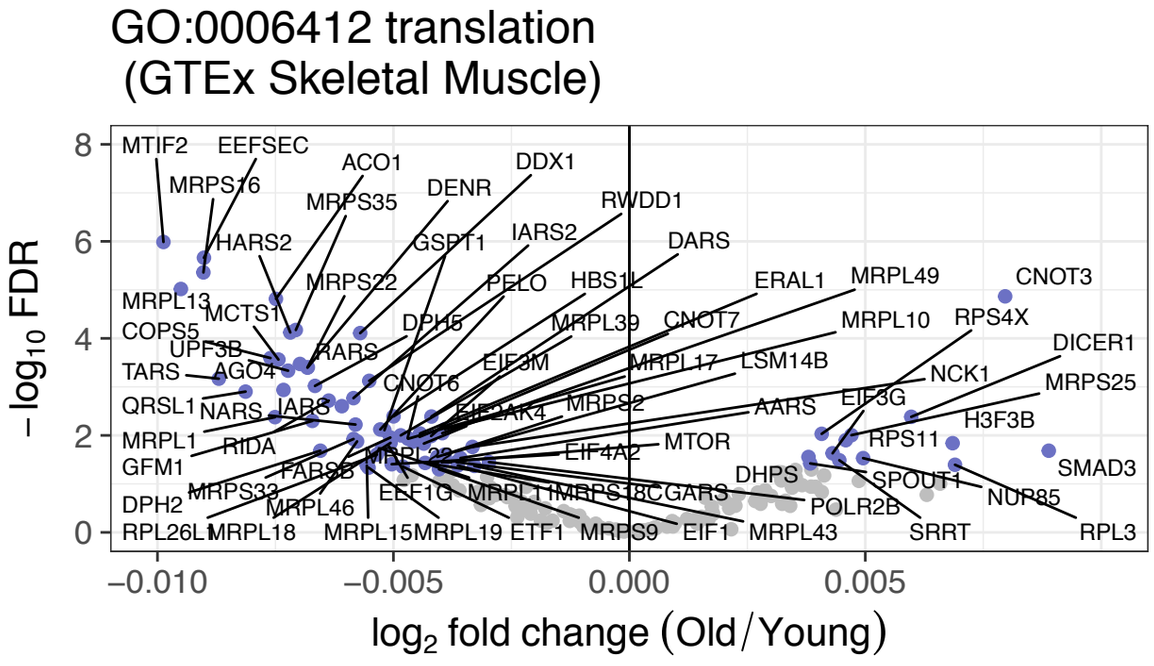
C



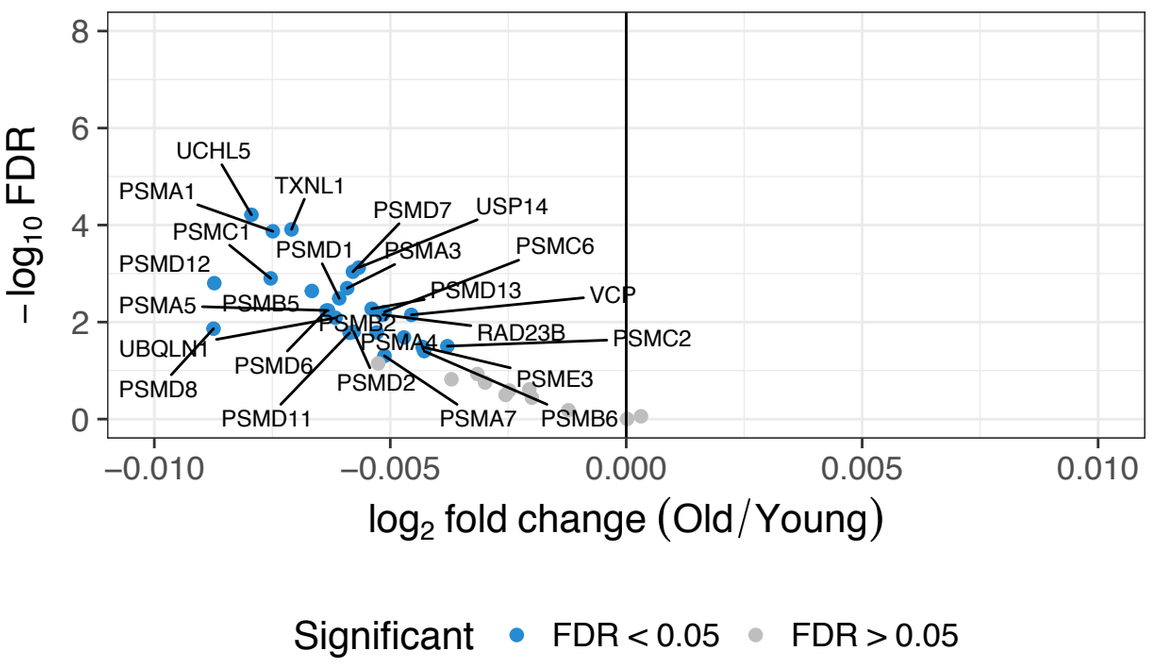
A



B



C



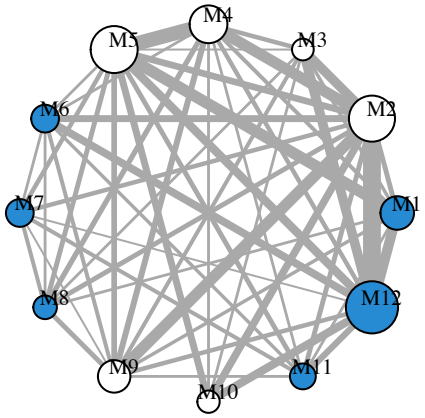
D

Proteostasis network (*H. sapiens*)

- Macroautophagy
- Protein synthesis
- Ubiquitin-proteasome system

A

Skeletal Muscle

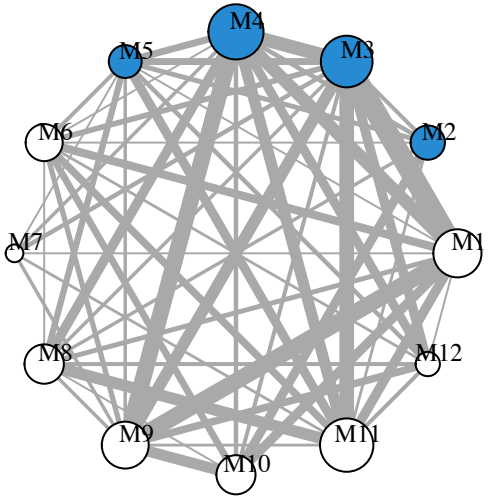


Proteostasis-associated processes

Module	GO BP terms	GWAS-associated disease
M1	GO:0010972 negative regulation of G2/M transition of mitotic cell cycle GO:0031146 SCF-dependent proteasomal ubiquitin-dependent protein catabolic process GO:0006521 regulation of cellular amino acid metabolic process	
M2	GO:0043488 regulation of mRNA stability GO:0006458 'de novo' protein folding GO:0032212 positive regulation of telomere maintenance via telomerase	
M3	GO:0000398 mRNA splicing, via spliceosome GO:00321124 mRNA 3'-end processing	LDL cholesterol, Total cholesterol
M4	GO:0006094 gluconeogenesis GO:0061621 canonical glycolysis	2hr glucose, multiple sclerosis, triglycerides
M5	GO:0006099 tricarboxylic acid cycle GO:0032981 mitochonodrial respiratory chain complex I assembly	Insulin resistance
M6	GO:0016241 regulation of macroautophagy GO:0042147 retrograde transport, endosome to Golgi	
M7	GO:0006413 translational initiation GO:0006364 rRNA processing	
M8	GO:0006457 protein folding GO:0006283 transcription-coupled nucleotide-excision repair	
M9	GO:0006189 'de novo' IMP biosynthetic process GO:0009113 purine nucleobase biosynthetic process	
M10	GO:0097194 execution phase of apoptosis GO:0048312 intracellular distribution of mitochondria	
M11	GO:0042274 ribosomal small subunit biogenesis GO:0006605 protein targeting	
M12	GO:000209 protein polyubiquitination	Coronary artery disease

B

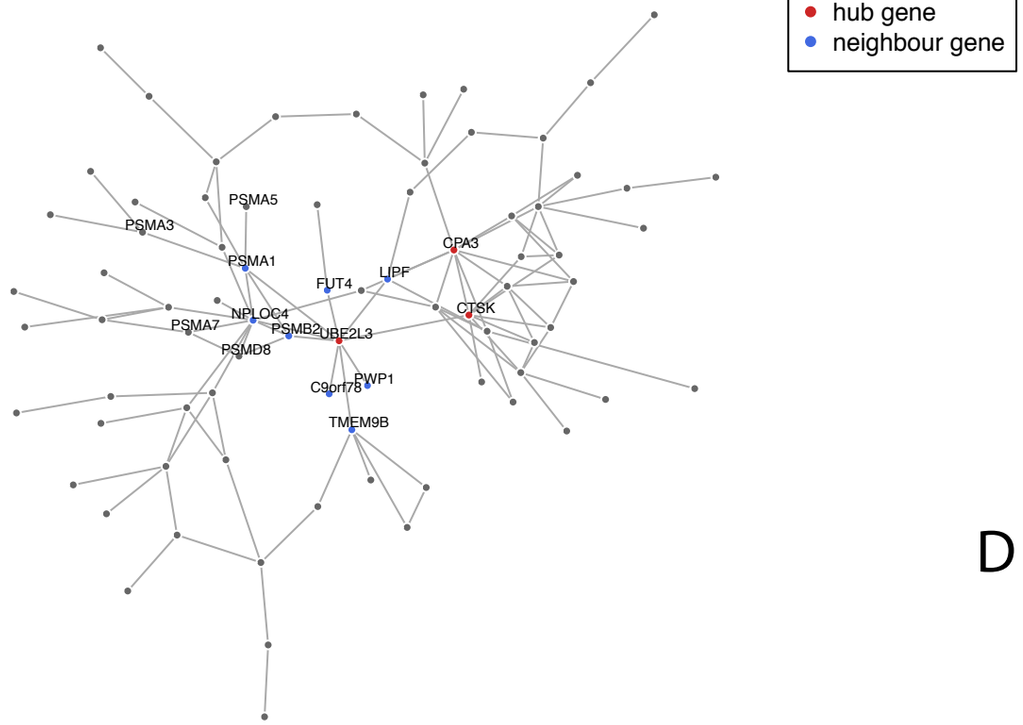
Hippocampus



Module	GO BP terms	GWAS-associated disease
M1	GO:0006099 tricarboxylic acid cycle GO:0006734 NADH metabolic process	
M2	GO:0006413 translational initiation GO:0006364 rRNA processing GO:0006614 SRP-dependent cotranslational protein targeting to membrane	
M3	GO:0006886 intracellular protein transport GO:0000209 protein polyubiquitination	
M4	GO:1904874 positive regulation of telomerase RNA localization to Cajal body	Coronary artery disease
M5	GO:0014850 response to muscle activity GO:0018344 protein geranylgeranylation	
M6	GO:0055114 oxidation-reduction process	Fasting proinsulin
M7	GO:0010976 positive regulation of neuron projection development	
M8	GO:1902001 fatty acid transmembrane transport	
M9	GO:0050806 positive regulation of synaptic transmission	
M10	GO:0006120 mitochondrial electron transport, NADH to ubiquinone GO:0032981 mitochondrial respiratory chain complex I assembly	
M11	GO:0022618 ribonucleoprotein complex assembly GO:0000082 G1/S transition of mitotic cell cycle	
M12	GO:0006189 'de novo' IMP biosynthetic process GO:0009113 purine nucleobase biosynthetic process	

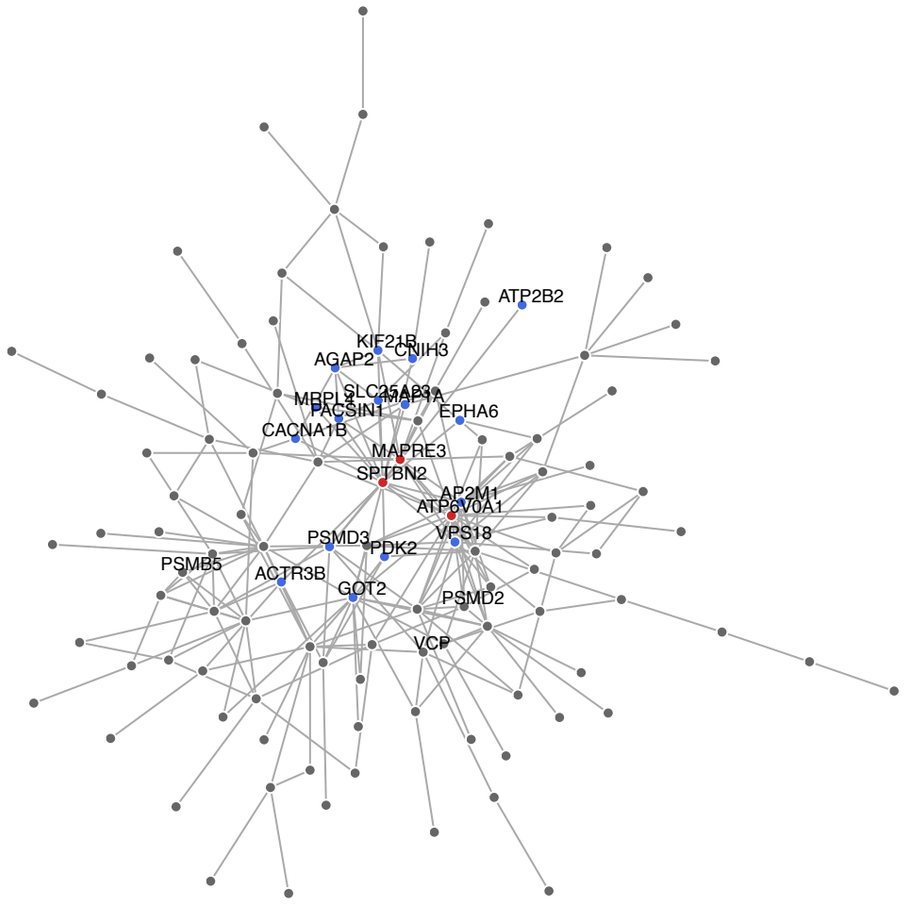
A

Skeletal muscle M1
(GO:0031146 SCF–dependent proteasomal ubiquitin–dependent protein catabolic process

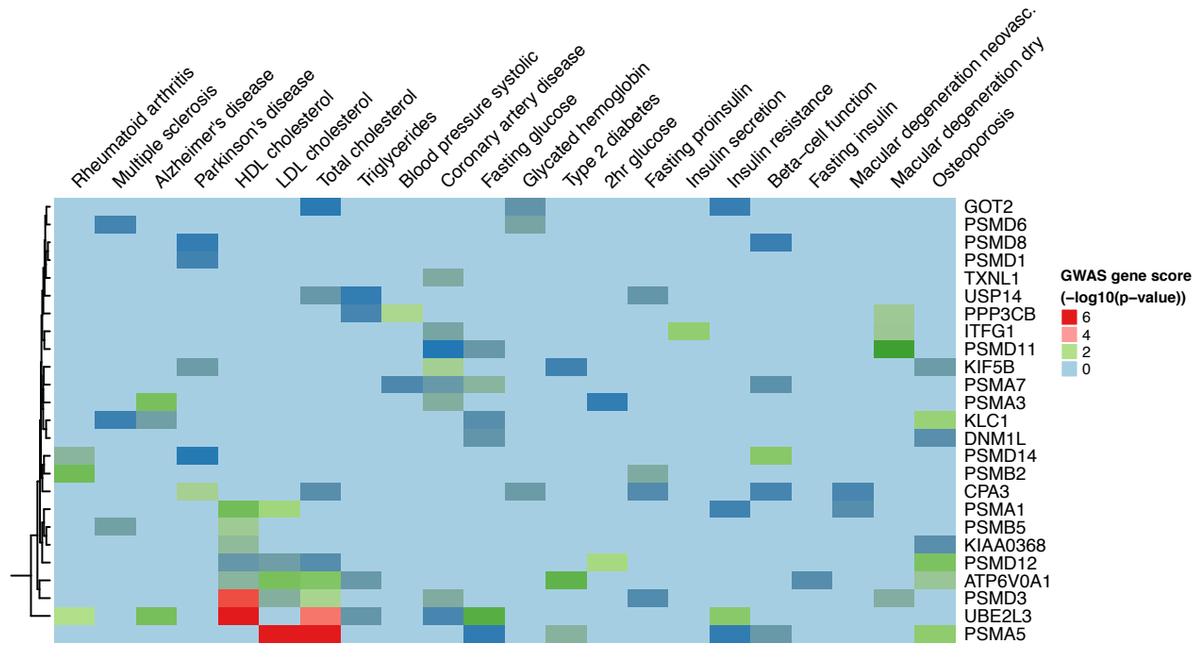


B

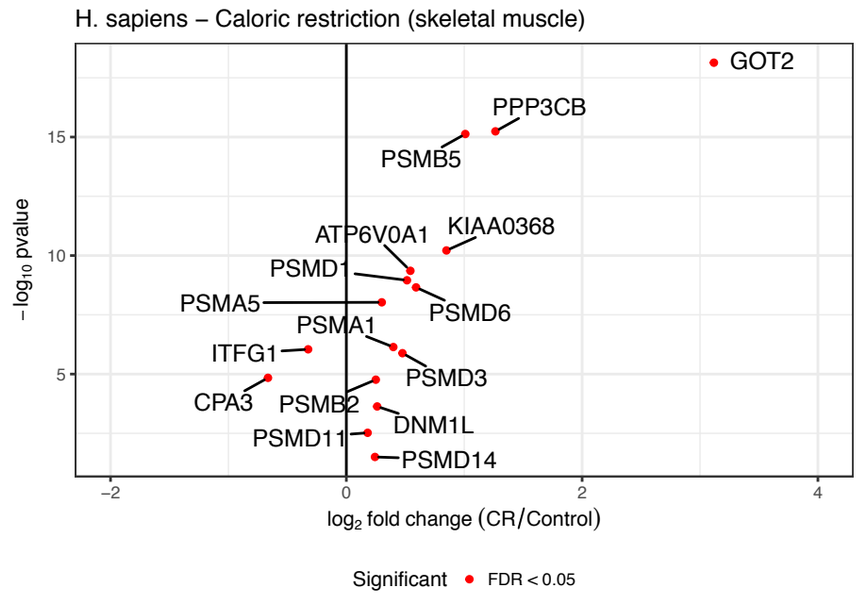
Hippocampus M4
(GO:1904874 positive regulation of telomerase RNA localization to Calaj body)



C



D



E