bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Healthspan pathway maps in *C. elegans* and humans highlight transcription, proliferation/biosynthesis and lipids

Steffen Möller[1], Nadine Saul*[2], Alan A. Cohen[3], Rüdiger Köhling[4], Sina Sender[5], Hugo Murua Escobar[5], Christian Junghanss[5], Francesca Cirulli[6], Alessandra Berry[6], Peter Antal[7,8], Priit Adler[9], Jaak Vilo[9], Michele Boiani[10], Ludger Jansen[11,12], Dirk Repsilber[13], Hans Jörgen Grabe[14], Stephan Struckmann[1,15], Israel Barrantes[1], Mohamed Hamed[1], Brecht Wouters[16], Liliane Schoofs[16], Walter Luyten[17], Georg Fuellen[1,**]*

[1] Rostock University Medical Center, Institute for Biostatistics and Informatics in Medicine and Aging Research (IBIMA), Rostock, Germany
[2] Humboldt-University of Berlin, Institute of Biology, Berlin, Germany
[3] Department of Family Medicine, University of Sherbrooke, Sherbrooke, Canada
[4] Rostock University Medical Center, Institute for Physiology, Rostock, Germany
[5] Rostock University Medical Center, Klinik für Hämatologie, Onkologie und Palliativmedizin, Rostock, Germany
[6] Center for Behavioral Sciences and Mental Health, Istituto Superiore di Sanità, Italy
[7] Budapest University of Technology and Economics, Budapest, Hungary
[8] Abiomics Europe Ltd., Hungary
[9] Institute of Computer Science, BIIT research group, University of Tartu, Tartu, Estonia
[10] Max-Planck Institute for Molecular Biomedicine, Münster, Germany
[11] Ruhr-Universität Bochum, Philosophisch-Theologische Grenzfragen, Bochum, Germany
[12] Universität Rostock, Institut für Philosophie, Rostock, Germany
[13] School of Medical Sciences, University of Örebro, Sweden
[14] University Medicine Greifswald, Klinik und Poliklinik für Psychiatrie und Psychotherapie, Greifswald, Germany
[15] University Medicine Greifswald, Institute for Community Medicine, Greifswald, Germany
[16] KU Leuven, Department of Biology, Leuven, Belgium
[17] KU Leuven, Department of Pharmaceutical & Pharmacological Sciences, Leuven, Belgium
*co-first authors. **corresponding author: Georg Fuellen, Ernst-Heydemann-Str.8, 18057 Rostock, Germany; +49 381 494 7360; fuellen@uni-rostock.de.
Running title: Healthspan pathway maps in *C. elegans* and humans

Summary

The molecular basis of aging and of aging-associated diseases is being unraveled at an increasing pace. An extended healthspan, and not merely an extension of lifespan, has become the aim of medical practice. However, a precise definition of health and healthspan is not straightforward, and the causal molecular basis of health "per se" is largely unknown. Here, we define health based on the absence of diseases and dysfunctions. Based on an extensive review of the literature, in particular for humans and *C. elegans*, we compile a list of features of health and of the genes associated with them. Clusters of these genes based on molecular interaction data give rise to maps of healthspan pathways for humans, featuring the themes *transcription initiation, proliferation* and *cholesterol/lipid processing*, and for *C. elegans*, featuring the themes *immune response, mitochondrion* and *biosynthesis* based on genetic and compound intervention data, and *lipids, biosynthesis* and *transcription* based on WormBase compound intervention data. Overlaying healthspan-related gene expression data (describing effects of metabolic intervention associated with improvements in health) onto the aforementioned healthspan pathway maps, we observe the downregulation of Notch signalling in humans and of proliferation/cell-cycle in *C. elegans*. The former reflects the proinflammatory role of the Notch pathway. We identify *transcription, proliferation/biosynthesis* and *lipids* as a common theme on the annotation level, and proliferation-related kinases on the gene/protein level. Our literature-based data corpus, including visualization, is available as a reference for future investigations, at http://www.h2020awe.eu/index.php/pathways/.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

For a long time, an active, targeted intervention to maintain health into old age was *terra incognita*. It had no priority, and few, if any, reliable data were available to implement it in everyday life. Today, however, systematically established diagnostic hints become available for the individual, based on family history and biomarker data, including genetic variants (polymorphisms). To assess and prevent premature health deterioration successfully, it would therefore be useful (1) to dissect "health" into a set of its most important features, (2) to detail its molecular basis and to map out molecular "healthspan pathways", and (3) to specify biomarkers and corresponding supportive interventions for the various features of health and for health itself. Arguably, the increase in life expectancy in the last 100 years has not been accompanied by an increase in disease-free life expectancy (Crimmins, 2015), (Robine, Jagger, & Crimmins, 2013). Cardiovascular disease, type-2 diabetes and neurodegenerative disorders are highly prevalent in the elderly, and these disorders frequently coexist in the same aged individual, often with mutual reinforcement (Fillenbaum, Pieper, Cohen, Cornoni-Huntley, & Guralnik, 2000). Extending healthspan may thus enable economic, societal and individual gains on a large scale (Fuellen et al., 2016).

Intervention studies to prolong healthspan based on compound exposure in humans are limited to relatively few compounds. Resveratrol for instance, being one of the best-studied poly-phenols in humans and animals, has been tested in several clinical studies (Berman, Motechin, Wiesenfeld, & Holz, 2017). These include studies focused on biomarkers, like the level of blood glucose (Brasnyó et al., 2011) and cholesterol (S. Chen et al., 2015), or glutathione S-transferase expression (Ghanim et al., 2011). However, data about *long-term* effects on overall health in human are missing in general, and given the average human life expectancy, they are difficult to obtain. Therefore, model organisms are of great relevance to uncover the molecular basis of healthspan and to identify supporting compounds. The nematode *Caenorhabditis elegans (C. elegans)* is a widely used model organism for studying ageing which guided the discovery of fundamental ageing-related findings, e.g., on calorie restriction and Insulin/IGF-1 like signaling (Gruber et al., 2015). Last not least, around 40% of the genes found in *C. elegans* have human homologues (Shaye & Greenwald, 2011) and studies revealing the role of metabolism on health conducted in *C. elegans* have been

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

subsequently strengthened in murine models (Berry & Cirulli, 2013), rendering this nematode a valuable model for human ageing processes. Most recently, *C. elegans* has come to enjoy increasing popularity as a model for health (Sutphin et al., 2017), (Luyten et al., 2016), and an ever increasing number of compounds are tested in *C. elegans* for their anti-ageing and health effects (Chen, Barclay, Burgoyne, & Morgan, 2015), (Luo, Wu, Brown, & Link, 2009), (Collins, Evason, & Kornfeld, 2006) , (De Haes et al., 2014).

Here, we assemble and explore "healthspan pathway maps", that is, annotated sets of interacting genes implicated in health. To create these, we follow a stepwise procedure: first, we dissect health into its various features, with an emphasis on disease and dysfunction. Second, we compile lists of genes associated with health based on the literature, for humans and *C. elegans*. Third, we organize these genes into *maps of healthspan pathways,* based on gene/protein interaction and annotation data. Fourth, we create an overlay of health-related gene expression data onto the resulting healthspan pathway maps, highlighting corroborating knowledge that was not used as input. Finally, we investigate the overlap of the *healthspan pathways* in humans and *C. elegans.*

For humans, we consider that knowledge of the causal basis of health may be best derived from genetic association studies. Based on an extensive review of the literature, we identify a core set of 12 genes that are genetically associated with a lack of frailty (Mekli, Marshall, Nazroo, Vanhoutte, & Pendleton, 2015), (Ho et al., 2011) and the Healthy Aging Index (Sanders et al., 2014), and another set of 40 genes genetically associated with (a lack of) multiple diseases, or with longevity mediated by a lack of disease (see Supplementary Tables 1-3). In contrast to humans, genetic intervention studies on healthspan are available for *C. elegans*, as well as compound intervention data. A lack of dysfunction exemplified by stress resistance, locomotion, pharyngeal pumping and reproduction are taken as the key health features in *C. elegans* (Rollins, Howard, Dobbins, Washburn, & Rogers, 2017). On this basis, a core set of 11 genes is directly implicated in improvements of locomotion by genetics, and another set of 20 genes is indirectly implicated in improvements of the key health features by studies that investigate effects of compounds (see Supplementary Tables 4 and 5).

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We then place the genes implicated in health into context by adding gene/protein interaction and gene annotation knowledge. Specifically, we turn the lists of genes into gene/protein interaction networks, to which 20 closely interacting genes are added, employing GeneMania (Zuberi et al., 2013). Gene ontology annotation data are then used to annotate clusters of strongly connected genes within the network, employing AutoAnnotate (Kucera, Isserlin, Arkhangorodsky, & Bader, 2016). Then, we elaborate how the resulting healthspan pathways can be interpreted in plausible ways, specifically in the light of independent health-related gene expression data describing effects of caloric restriction and of rapamycin, and in the light of gene expression data describing aging and disease. We also predict microRNAs that may be potential regulators of healthspan. Finally, we find that if we construct an overlap between the healthspan pathways in *C. elegans* and humans, we find genes involved in transcription, proliferation/biosynthesis and lipids, but this overlap it is not straightforward to interpret in the light of the independent health-related gene expression data that we used to test plausibility of the single-species healthspan pathway maps.

All healthspan pathways discussed in this manuscript, as well as the overlaps we found between species, are available for interactive exploration at http://www.h2020awe.eu/index.php/pathways/.

Results

Studies of health. The main sources of knowledge about health, that is, about features, biomarkers and interventions regarding health-related phenotypes, are

(a) observational genetic investigations, usually in the form of genome-wide association studies, looking for associations between health and polymorphisms of specific genes,

(b) observational studies of non-genetic biomarkers, which are dynamic in time and are usually related to known canonical pathways, and their longitudinal or cross-sectional correlation with health,

(c) interventional studies, most often in model organisms, where interventions affecting health may be genetic or based on food or (pharmaceutical)

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

compounds, and the intervention effects are measured on the molecular level, implicating particular genes or pathways.

Like genetic studies, compound intervention studies can, in principle, elucidate the causative basis of health. Studies of type (b) may only be revealing correlative evidence and can sometimes not be linked to particular genes; therefore, we will not consider these further. Candidate biomarkers of health can be many kinds of features with the potential to predict future health better than chronological age (Fuellen et al., 2018); they may be genetic (polymorphisms; such biomarkers are essentially static over lifetime), molecular but not genetic (epigenetic or transcript or protein or metabolic markers, etc.), cellular (blood counts, etc.) or organismic (such as grip strength). Based on studies of types (a) and (c), in this work we will only deal with genes and sets of genes (that is, genes organized into networks or pathways) as candidate biomarkers of health.

Defining Health. *Health* is a term in biology and medicine that is hard to define. We propose that the best definition of health must be based on an aggregation of the literature, see also (Fuellen et al., 2018), (Luyten et al., 2016). Then, healthspan is simply the time spent in good health. Supplementary Tables 1-3 list features of *human* health as discussed in the literature, referring to lack of dysfunction, lack of *multiple* diseases, and lifespan/longevity mediated by lack of disease. In principle, at least for human, *dysfunction* can be operationalized with the help of a codified classification of function (such as the ICF, the International Classification of Functioning, Disability and Health, www.who.int/classifications/icf/en/). This classification provides criteria to establish that an individual is affected by a dysfunction. As described and discussed in (Fuellen et al., 2018), we can filter the “body function” part of the ICF by looking for follow-up in the literature on health and healthspan. The result is a pragmatic community consensus definition of dysfunction, centering around the lack of physiological, physical, cognitive and reproductive function; a lack of physiological, physical and cognitive functions is often called frailty. To a large degree, this consensus definition can be used for non-human species as well. Further, *disease* can also be operationalized by a codified classification (such as ICD-11, International Statistical Classification of Diseases and Related Health Problems, www.who.int/classifications/icd/en/). Again, the classification provides criteria to establish that an individual is affected by a disease. In this paper, affection by a single

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

disease is not considered, as in old age, single-disease morbidity rarely exists, and in terms of interventions, we are interested in preventing more than one disease. As described and discussed in (Fuellen et al., 2018), not all parts of the ICD feature diseases related to health and healthspan. However, we note that all diseases referred to in Supplementary Tables 1-3 qualify as age-associated diseases.

For *C. elegans*, Supplementary Tables 4-5 list features of health based on the literature, referring to lack of dysfunction in the form of stress resistance (in response to thermal and oxidative stress), (stimulated) locomotion, pharyngeal pumping, and reproduction. These features dominate the literature, and they cover the aspects of physiological function, physical and cognitive function, and reproductive function, as in human (Fuellen et al., 2018). Of note, genetic analyses of health in *C. elegans* have focused up to now mostly on (stimulated) locomotion. Stimulated locomotion integrates some aspects of strength (physical function) and cognition (cognitive function).

Genes associated with health. Our list of *health genes,* that is, genes with associations to health in humans are based on genetic association, and we can assume some probability of them being causal, leaving aside the intrinsic ambiguities to assign polymorphisms (in the form of SNPs, single-nucleotide polymorphisms) to genes, e.g. in intergenic regions or in intronic regions with overlapping non-coding RNA on the complementary strand (Schwarz et al., 2008). In turn, for *C. elegans*, few studies report effects of genetic interventions on health, though these are increasingly becoming available; (Sutphin et al., 2017) is the only large-scale genetic study that we could identify as of early 2018, even though it is essentially a small-scale study of healthspan based on a large-scale study of lifespan. Many more studies in *C. elegans* refer to canonical aging-related pathways, and in contrast to studies in humans, these studies often directly report the molecular effects of compound intervention. The *C. elegans* genes listed in the supplementary tables are thus based on the effects of genetic intervention and on the effects of compound intervention on the gene level, and we can assume a high probability of causality in both cases.

Additional genes associated with C. elegans health. For *C. elegans,* we generated an additional list of health-associated genes that cannot be generated for humans (see Experimental Procedures), using *WormBase* to systematically identify health-related compound interventions with associated gene expression data, and compiling the list

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of genes with strongest differential expression that are well-annotated by Gene Ontology terms.

From gene lists to maps of healthspan pathways. We used *Cytoscape* and some of its application plugins as the most straightforward tool for obtaining and annotating a connected network of the genes from Supplementary Tables 1-3 and from Supplementary Tables 4 & 5. Specifically, we used *GeneMANIA* to establish a gene/protein interaction network and to add connecting genes, and subsequently we clustered all genes based on their connectivity, and added GO-based annotations, using *AutoAnnotate*. The resulting healthspan pathway maps are presented in the following. Moreover, health-related gene expression data are overlaid onto all healthspan pathway maps and will be discussed as well; these data are describing the effects of caloric restriction (CR) in humans (Mercken et al., 2013) and of rapamycin in *C. elegans* (Calvert et al., 2016), as examples of health-promoting interventions, or they describe the effects of aging and disease in specific tissues.

For humans, the gene list derived from Supplementary Tables 1-3 (see Suppl. Table 6) yielded the network of Figure 1, where the two largest pathways/clusters (15 and 13 genes) are specifically labeled by NOTCH & transcription initiation, and by proliferation, and the smaller pathways/clusters (4, 3, 3 and 3 genes) are labeled by cholesterol & lipid processes, by thymus activation, by myotube (striate muscle) regulation, and by Wnt signalling. In Figure 1 bottom, the list of pathways/clusters is given, and the details of the largest pathway are zoomed in.

In the largest pathway/cluster, in light of the CR-triggered gene expression changes, the most prominent findings are an induced downregulation of NOTCH4 (and to a lesser extent of NOTCH 2&3), as well as of LRP1, and an upregulation of TOMM40 and CREBBP (also known as CBP). The family of NOTCH proteins has various functions, including a pro-inflammatory one (Balistreri, Madonna, Melino, & Caruso, 2016), (Zhang, Kuang, Wang, Xu, & Ren, 2017). NOTCH4 is upregulated in kidney failure (Liu et al., 2017), and promotes vascularization/angiogenesis, which includes its upregulation in malignancy (Zhang et al., 2017), (Kofler et al., 2011). A downregulation of NOTCH4 by CR can thus be taken as beneficial effect. This is less obvious for LRP1, the low-density lipoprotein receptor-related protein 1, which is responsible for membrane integrity and membrane cholesterol homeostasis, thus

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

being involved in proper myelination (Lin, Mironova, Shrager, & Giger, 2017) and vascular integrity (Strickland, Au, Cunfer, & Muratoglu, 2014). A downregulation of LRP1 during CR could therefore be seen as deleterious. However, LRP1 expression mainly depends on cholesterol levels (Llorente-Cortes, Otero-Nivas, Sanchez, Rodriguez, & Badimon, 2002) – and these are lower during fasting. Hence, lower LRP1 expression actually reflects a lower LDL level, which per se has been found to be protective, assuming brain levels to follow plasma levels. The upregulations observed for TOMM40 and CREBBP during CR can also be seen as protective. TOMM40 is part of a mitochondrial membrane protein translocase, supporting mitochondrial function (Zeitlow et al., 2017), and low expression and/or particular risk alleles of this protein are associated with Huntington's and Alzheimer's Disease (Shirendeb et al., 2011), (Chong et al., 2013). Of note, TOMM40 upregulation during CR goes together with APOE4 downregulation. Although both genes are closely located on chromosome 19, prompting the speculation that this linkage could imply concordant expression changes, this is obviously not the case here. CREBBP is a transcriptional co-activator with histone-acetyltransferase activity (Bedford & Brindle, 2012), acting primarily on histones 3 and 4, and thus it acts in concert with a range of transcription factors. Its downregulation is deleterious, resulting in, e.g., MHCII expression loss on lymphocytes (Hashwah et al., 2017), rendering the lymphocytes dysfunctional for antigen presentation, and in inflammatory signalling (Dixon et al., 2017). An upregulation of CREBBP by CR is thus likely beneficial. We further investigated the miRNAs that are statistically enriched in the largest healthspan pathway using the TFmir webserver (Hamed, Spaniol, Nazarieh, & Helms, 2015), revealing regulation of NOTCH genes implicated in the epithelial-mesenchymal transition, cancer, heart failure and obesity, see Supplementary Results. The genes in the next-largest pathway/clusters, related to cell proliferation and lipids, are also described there in detail, as well as further evidence provided by mapping aging- and disease-related gene expression data onto them, as published or collected by (Aramillo Irizar et al., 2018).

For *C. elegans*, the gene list derived from Supplementary Tables 4-5 (see Suppl. Table 7) yielded the network of Figure 2, where the largest clusters (9 and 6 genes, respectively) are labeled by immune response process and by terms related to the mitochondrion. Three clusters (of 4 genes each) specifically feature dauer/dormancy, regulation and hormone response. In Figure 2 bottom, the list of pathways/clusters,

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and the details of the largest pathway are zoomed in. Regarding the first pathway, rapamycin induces ets-7 transcription, which was shown to be also necessary for the healthspan-promoting effects of salicylamine (Nguyen et al., 2016). Furthermore, rapamycin upregulates the transcription factor daf-16 (a homolog to Foxo) and downregulates the daf-16 inhibitors akt-1 and akt-2, putatively leading to an improved stress- and immune-response and prolonged lifespan via the Insulin/IGF-1 pathway (Henderson & Johnson, 2001). Along the same lines, the akt-1 and akt-2 activator pdk-1 is also downregulated by rapamycin, further promoting daf-16 activity (Paradis, Ailion, Toker, Thomas, & Ruvkun, 1999). In contrast, the daf-16 inhibitor sgk-1 (a homolog to Nrf) is upregulated; however, its inhibitory role is subject of discussion (Mizunuma, Neumann-Haefelin, Moroz, Li, & Blackwell, 2014). Finally, the transcription factors hsf-1 and skn-1, both important in stress response processes (Morton & Lamitina, 2013); (Wang et al., 2010), are slightly downregulated in rapamycin-treated *C. elegans*. In the Supplementary Results, the next-largest pathway/clusters, related to the mitochondrion, to *dauer*/dormancy, to regulation, and to hormone response are described in detail.

For *C. elegans*, we also derived a gene list from *WormBase*, taking the genes that are most differentially regulated by healthspan-extending interventions and, at the same time, are annotated with a sufficient number of GO terms (see Experimental Procedures; Suppl. Table 8). We obtained the network of Figure 3. Curiously, the top healthspan pathways of 11, 9 and 8 genes are related to the *endoplasmic reticulum (ER)*, lipid & membrane, to the *peroxisome*, macrobody & ER, and to the *lysosome*. The endoplasmic reticulum, the peroxisome and the lysosome are part of the endomembrane system, together with the mitochondria, contributing to healthspan and longevity in mammals and beyond (Nisoli & Valerio, 2014). Peroxisomal function connects this pathway to dietary effects on lifespan (Weir et al., 2017), and to liver disease (Cai et al., 2018). The second tier of healthspan pathways (6 or 5 genes) are related to morphogenesis, biosynthesis and transcription.

For the *WormBase* data, the list of pathways/clusters, and the details of the largest pathway, are given in Figure 3, bottom. The ER/lipid-related pathway includes genes involved in fatty acid elongation/production (elo-1 to elo-9; let-767; art-1). Overlaying the rapamycin gene expression data, the well-characterized elo-1 and let-767 genes

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

show negligible downregulation. However, the importance of elongase genes for health maintenance in general was repeatedly documented. Vásquez and colleagues (Vásquez, Krieg, Lockhead, & Goodman, 2014) demonstrated the impairment of touch response in elo-1 mutants. They argue that elo-1 has a crucial role in the synthesis of C20 polyunsaturated fatty acids which are required for mechanosensation. Moreover, elo-1 mutants showed increased resistance to *Pseudomonas aeruginosa* infections due to the accumulation of gamma-linolenic acid and stearidonic acid (Nandakumar & Tan, 2008) and knockdown of elo-1 or elo-2 extend survival during oxidative stress (Reis et al., 2011). Finally, art-1 is a steroid reductase that is downregulated by rapamycin in our case, but also in long-lived eat-2 mutants (Yuan et al., 2012). In the Supplementary Results, the next-largest pathway/clusters, related to the ER/peroxisome, the lysosome, morphogenesis, biosynthesis and transcription, are described in detail.

Overlap between human and C. elegans health genes and healthspan pathways. Based on reciprocal best orthologs, we found no direct overlap between the human health genes based on genetic associations and the *C. elegans* healthspan genes based in part on genetic interventions, but mostly on expert analysis of intervention effects (Figure 2), or on gene expression changes related to healthspan-extending interventions (Figure 3). We found some hints at an overlap on the level of the healthspan pathway annotations, considering that “proliferation” is listed for human, and “biosynthesis” for *C. elegans*, and “transcription” as well as “lipid” for both.

In search for other modes of overlap, we therefore constructed and compared two interaction networks, based on mapping genes to their respective orthologs in the other species. Each of the two interaction networks is based on the *union set* of the health genes of human (based in turn on genetics, Supplementary Tables 1-3, Figure 1) and of *C. elegans* (based in turn on the gene expression analysis of healthspan-extending interventions using WormBase, Figure 3). Specifically, as outlined in Figure 4, we added the *C. elegans* orthologs of the human health genes to the list of *C. elegans* health genes and *vice versa,* yielding two separate input gene lists for GeneMANIA to enable the construction of the two interaction networks. We used strict ortholog mapping rules (only reciprocal best hits were accepted). By design, the two gene lists feature a high degree of overlap (with differences due to missing orthologs), and their

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

subsequent comparison, consisting of the partial network alignments that are based on ortholog mapping on the one hand and the species-specific network data on the other hand can only reveal *hypotheses for common healthspan pathways*, as long as explicit experimental evidence for a relation to health is only found for one species. Moreover, *interaction points between a healthspan pathway in one species and a healthspan pathway in the other species* may be revealed, if a partial alignment of the interaction networks consists of interacting genes for which the relationship to health was demonstrated only in one species for each pair of orthologs.

Of note, of the two interaction networks to be aligned, the first network is based on *C. elegans* health genes, the *C. elegans* orthologs of human health genes, and *C. elegans* gene interaction information provided by GeneMANIA. The second network is based on human health genes, the human orthologs of *C. elegans* health genes, and human gene interaction information provided by GeneMANIA. Despite using similar lists of genes (with differences due to missing orthologs and due to the genes added by GeneMANIA), we can expect that the two GeneMANIA networks are quite different because the interaction data sources employed by GeneMANIA are strongly species-specific. Moreover, we observe that in both cases, the 20 closely interacting genes added by GeneMANIA for one species included no orthologs of the other species. Nevertheless, to identify *joint healthspan pathways* and *interaction points between healthspan pathways,* we used GASOLINE (Micale, Continella, Ferro, Giugno, & Pulvirent, 2014) to align the two networks wherever feasible, obtaining two partial (subnetwork) alignments as output, as shown in Figure 5.

In the first alignment (Figure 5, left), we see an alternating pattern of demonstrated health-relatedness, since pak-2, sad-1 and pig-1 are considered health-related by gene expression analysis using WormBase, while CDKN2B and GSK3B are known to be human health genes (Supplementary Tables 1-3; GSK3B was implicated by a GWAS of the Healthy Aging Index, while CDKN2B was in fact one of the few genes implicated by two independent health studies). The *C. elegans* genes belong to three small clusters in the healthspan pathway map of Figure 3 (pak-2: lysosomal, sad-1: neural, pig-1: biosynthesis), while the human genes belong to one large (GSK3B: proliferation) and one small (CDKN3B: cyclin-dependent kinase) cluster in the human healthspan pathway map of Figure 1. Interactions in *C. elegans* are all based on shared

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

domains (kinase signaling, except for the predicted interaction of gsk-3 & C25G6.3, which is based on the Interologous Interaction Database), while interactions in human are based on shared domains, genetic interaction (i.e., large-scale radiation hybrid) and pathway data. Essentially, the healthspan pathway overlap suggested by our analysis involves serine/tyrosine kinase signalling (pak-2/sad-1/pig-1 & PAK4/BRSK2/MELK), Wnt signalling (GSK3) and cyclin-dependent kinase signalling (CDKN2B). Both alignments are described further in detail in the Supplementary Results.

Discussion

Healthspan pathway mapping, current possibilities and future opportunities. We took a pragmatic approach to investigate the molecular basis of health. For humans, there are only few alternatives to genetic association studies, to determine the genes responsible for the phenotype of lacking disease and dysfunction. Post-mortem analyses aside, studies in human are bound to the insights gathered from the etiology and treatment of diseases in patients. Genetic manipulation is possible in human cell lines, but this requires to break down the disease or dysfunction to a cellular phenotype. On the other hand, molecular insights can be gained with few ethical constraints in non-vertebrates such as *C. elegans.* A range of age-associated phenotypes can be observed in *C. elegans* that are similar to human, yet the nematode is still a much simpler organism with far fewer cells and a smaller genome. As noted, however, sequence similarity supports the mapping of ortholog genes across species, suggesting functional similarity or equivalence, and many core pathways are preserved. We discuss the biological interpretation of the lack of evolutionary conservation in the Supplementary Discussion.

Looking for models of human health, there is a choice of cell lines, organoids ("organ-in-a-dish"), *in silico* simulations and model organisms. A major disadvantage of cell lines is their artifactually isolated cultivation. This disadvantage is addressed by organoid-based approaches, which, however, are still far from the human situation they are supposed to model, in terms of size, complexity, and the duration of the investigation. *In silico* simulations are limited in complexity, and their faithfulness may be questioned on many accounts. This faithfulness is also a big issue with model

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

organisms, where ease of handling and a reasonable duration of the investigation usually result in difficult tradeoffs with faithfulness. A discussion of the various approaches just mentioned, and of the various choices of model organisms is beyond the scope of this paper. But it should be mentioned that while health can be defined and investigated in *C. elegans*, this model is also fraught with an array of issues, some shared with organoid-based approaches, but also including some particularities of the species (post-mitotic tissues, artificial cultivation, hermaphrodite, genetic uniformity of strains, lack of organs present in humans such as a heart) and of our modes of investigation (limitations of the data currently available, limitations of healthspan assays in terms of their design and the controlled environment in which these are conducted). Nevertheless, for studies in a living system to be completed in a short time-frame, *C. elegans* is arguably the best choice.

Given lists of genes, there is a plethora of possibilities to organize the genes into groups of related ones. Motivated by the idea of a “healthspan pathway”, we hypothesized that the genes should be known to interact based on functional gene/protein interaction data (provided by GeneMANIA). Here, as in most other studies, pathways are not assumed to be linear. The (higher-level) interaction among the clusters/healthspan pathways (i.e., the pathway map) is given by the individual gene/protein interactions that are shown *between* the clusters in Figures 1-3. However, we did not investigate these further.

The small amount of healthspan gene/pathway overlap that we found may be seen from a pessimistic or an optimistic perspective, depending in part on expectations. From the pessimistic perspective, the molecular processes may be completely different, and the *C. elegans* orthologs of the human health genes are involved in different processes as compared to the human health genes, and *vice versa*. From the optimistic perspective, it may just be that the number and scope of the investigations that yielded the health genes we studied is still insufficient, annotations are still incomplete, and considering only reciprocal best orthologs may be too restrictive. (We tried a less restrictive mapping of orthologs by relaxing the condition that orthologs must be reciprocal, but the overlap was still negligible; results not shown). Nevertheless, future genetic studies are expected to yield more health genes in both species, and their characterizations are expected to improve. Moreover, when we

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

analyze in detail the effects of intervention studies in *C. elegans*, we do find clear hints to some mechanisms that underlie healthspan also in human. For example, changes in the Ins/IGF-1 pathway genes daf-2 and daf-16 are found to be associated with many of the features described in Supplementary Table 5, suggesting a fundamental role for immune defense mechanisms (and proliferation) in health maintenance, as described by (Ermolaeva & Schumacher, 2014).

Of course, the precise definition of phenotype is crucial. If the samples are not really about (lack of) health, in human or in *C. elegans*, then any subsequent molecular or bioinformatics analyses will compare apples and oranges and may thus fail. Therefore, it is important to use a good phenotyping of health in human as well as in *C. elegans,* and on this basis, to collect data as genome-wide as possible. It is evident that there is no *C. elegans* counterpart to most of the (age-related) diseases that we use to define health in humans, and that the aging process that may underlie most of these age-related diseases is poorly characterized and hard to quantify in humans. Nonetheless, locomotion degrades with age in both species, due to changes at the muscle as well as neural level. Two related features of physical function, that is, grip strength (Leong et al., 2015) and the ability to sit and rise from the floor (Brito et al., 2014) are good predictors of all-cause mortality in humans. Likewise, both in humans and in *C. elegans*, the ability to withstand various forms of stress decreases with age (Dues et al., 2016), (Bajorat et al., 2014). Thus, at the level of organs or functional systems, both *C. elegans* and humans show age-related declines in performance, that may well be due to underlying processes that are similar at the cellular and molecular level. We claim that within the limitations of currently available data, the health genes we assembled, the healthspan pathways we constructed based on these, and the overlap we then found between species, are a first glimpse of the species-specific and cross-species molecular basis of health.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Experimental Procedures

Gene sets associated with health, literature-based. In this work, we conducted a semi-systematic review, using *health*, *healthspan* and *healthy aging,* for human and *C. elegans,* as search terms in *Google Scholar,* initially filtering for recent reviews and considering only the top hits. For humans, genetic studies of Supplementary Tables 1-3 are often not found using health-related keywords, so we included terms related to dysfunction (such as *frailty*) and disease (such as *multi-morbidity*) as well. For the genetics of human frailty, we identified two publications (Mekli et al., 2015), (Ho et al., 2011). Overall, a list of 52 genes (Supplementary Table 1, 12 genes; Supplementary Tables 2 and 3, 40 genes) was taken as the starting point in humans. For the genetics of *C. elegans* health, we followed a similar approach (Supplementary Table 4). For compound interventions in *C. elegans*, we identified a specific set of recent reviews (see Supplementary Table 5). Overall, a list of 31 genes (Supplementary Table 4, 11 genes; Supplementary Table 5, 20 genes) was taken as the starting point in *C. elegans*. From the original publications and reviews, we extracted the gene names, using *ihop* (Hoffmann & Valencia, 2004) to assign HUGO nomenclature names if necessary. In the Supplement, we describe in detail how a second set of health-associated genes in *C. elegans* was identified using WormBase.

Construction of maps of clusters/pathways. For all gene sets analyzed, we used the Cytoscape 3.5.1 application *GeneMANIA* (Zuberi et al., 2013), version 3.4.1, downloaded October 2017, with default settings, to create a functional interaction network that is complemented with the GeneMANIA default of 20 connecting genes. For clustering, and for annotating the clusters based on the “annotation name” column of GO annotations collected by GeneMANIA, we used *AutoAnnotate* (Kucera et al., 2016) v1.2, downloaded October 2017, in Quick start modus to enable to “layout network to prevent cluster overlap”, so that a map of disjoint clusters (healthspan pathways) was generated, supplemented by a second advanced annotation step to increase the “max. number of words per cluster label” to the largest possible value of 5. Cluster annotations were generated using WordCloud (Oesper, Merico, Isserlin, & Bader, 2011) v3.1.1, downloaded January 2018.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In the Supplement, we further describe in detail how we overlaid expression data onto the pathway maps, constructed the overlap of healthspan pathways in *C.elegans* and humans, and programmed the web presentation.

Acknowledgements

We thank Yasmeen Quawasmeh for technical assistance. We acknowledge assistance by Giovanni Micale in using GASOLINE. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No 633589 (Aging with Elegans). This publication reflects only the authors' views and the Commission is not responsible for any use that may be made of the information it contains. AAC is supported by a CIHR New Investigator Salary Award and is a member of the Fonds de recherche du Québec – Santé funded Centre de recherche du CHUS and Centre de re-cherche sur le vieillissement.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Study design: GF, WL, SM, NS. Collection of data: GF, NS, SM. Analysis of data: GF, NS, RK, SSe, HME, CJ, SS, IB, MH. Website: SM, PAd, JV. Manuscript writing: GF, NS, SM, AAC, RK, SSe, HME, LS, BW, FC, AB, PAn, HJG, DR, MB, LJ. All authors reviewed and approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References:

Aramillo Irizar, P., Schäuble, S., Esser, D., Groth, M., Frahm, C., Priebe, S., . . . Kaleta, C. (2018). Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nature communications, 9*(1), 327.

Bajorat, J., Oberacker, T., Ziola, S., Edgar, B., Gülow, K., & Krammer, P. (2014). AF-1 a novel regulator of the redox equilibrium during aging. *Free Radical Biology and Medicine, 75*, S22-S23.

Balistreri, C., Madonna, R., Melino, G., & Caruso, C. (2016). The emerging role of Notch pathway in ageing: Focus on the related mechanisms in age-related diseases. *Ageing research reviews, 29*, 50-65.

Bedford, D., & Brindle, P. (2012). Is histone acetylation the most important physiological function for CBP and p300? *Aging (Albany NY), 4(4)*, 247-255.

Berman, Y., Motechin, R., Wiesenfeld, M., & Holz, M. (2017). The therapeutic potential of resveratrol: a review of clinical trials. *NPJ Precis Oncol, 1*, pii 35.

Berry, A., & Cirulli, F. (2013). The p66(Shc) gene paves the way for healthspan: evolutionary and mechanistic perspectives. *Neurosci Biobehav Rev, 37*(5), 790-802.

Brasnyó, P., Molnár, G., Mohás, M., Markó, L., Laczy, B., Cseh, J., . . . Wittmann, I. (2011). Resveratrol improves insulin sensitivity, reduces oxidative stress and activates the Akt pathway in type 2 diabetic patients. *Br J Nutr, 106*(3), 383-389.

Brito, L., Ricardo, D., Araujo, D., Ramos, P., Myers, J., & Araujo, C. (2014). Ability to sit and rise from the floor as a predicator of all-cause mortality. *Eur J Prev Cardiol, 21*(7), 892-898.

Cai, M., Sun, X., Wang, W., Lian, Z., Wu, P., Han, U., . . . Zhang, P. (2018). Disruption of peroxisome function leads to metabolic stress, mTOR inhibition, and lethality in liver cancer cells. *Cancer Lett, 421*, 82-93.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Calvert, S., Tacutu, R., Sharifi, S., Teixeira, R., Ghosh, P., & de Magalhães, J. (2016). A network pharmacology approach reveals new candidate caloric restriction mimetics in C. elegans. *Aging cell, 15(2)*, 256-266.

Chen, S., Zhao, X., Ran, L., Wan, J., Wang, X., Qin, Y., . . . Mi, M. (2015). Resveratrol improves insulin resistance, glucose and lipid metabolism in patients with non-alcoholic fatty liver disease: a randomized controlled trial. *Dig Liver Dis, 47*(3), 226-232.

Chen, X., Barclay, J., Burgoyne, R., & Morgan, A. (2015). Using C. elegans to discover therapeutic compounds for ageing-associated neurodegenerative diseases. *Chem Cent J, 9*, 65.

Chong, M., Goh, L., Lim, W., Chan, M., Tay, L., Chen, G., . . . Lee, T. (2013). Gene expression profiling of peripheral blood leukocytes shows consistent longitudinal downregulation of TOMM40 and upregulation of KIR2DL5A, PLOD1, and SLC2A8 among fast progressors in early Alzheimer's disease. *J Alzheimers Dis, 34*(2), 399-405.

Collins, J., Evason, K., & Kornfeld, K. (2006). Pharmacology of delayed aging and extended lifespan of Caenorhabditis elegans. *Experimental gerontology, 41*(10), 1032-1039.

Crimmins, E. (2015). Lifespan and healthspan: past, present, and promise. *The Gerontologist, 55*(6), 901-911.

De Haes, W., Frooninckx, L., Van Assche, R., Smolders, A., Depuydt, G., Billen, J., . . . Temmerman, L. (2014). Metformin promotes lifespan through mitohormesis via the peroxiredoxin PRDX-2. *Proceedings of the National Academy of Sciences of the United States of America, 111*(24), E2501-2509.

Dixon, Z., Nicholson, L., Zeppetzauer, M., Matheson, E., Sinclair, P., Harrison, C., & Irving, J. (2017). CREBBP knockdown enhances RAS/RAF/MEK/ERK signaling in Ras pathway mutated acute lymphoblastic leukemia but does not modulate chemotherapeutic response. *Haematologica, 102*(4), 736-745.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Dues, D., Andrews, E., Schaar, C., Bergsma, A., Senchuk, M., & Van Raamsdonk, J. (2016). Aging causes decreased resistance to multiple stresses and a failure to activate specific stress response pathways. *Aging (Albany NY), 8*(4), 777.795.

Ermolaeva, M., & Schumacher, B. (2014). Insights from the worm: the C. elegans model for innate immunity. *Semin Immunol, 26*(4), 303-309.

Fillenbaum, G., Pieper, C., Cohen, H., Cornoni-Huntley, J., & Guralnik, J. (2000). Comorbidity of five chronic health conditions in elderly community residents: determinants and impact on mortality. *J Gerontol A Biol Sci Med Sci, 55*(2), M84-89.

Fuellen, G., Jansen, L., Cohen, A., Luyten, W., Gogol, M., Simm, A., . . . Möller, S. (2018). Health and Aging: Unifying Concepts, Scores, Biomarkers and Pathways. *Preprints, 2018080025*. doi:10.20944/preprints201808.0025.v1

Fuellen, G., Schofield, P., Flatt, T., Schulz, R. J., Boege, F., Kraft, K., . . . Simm, A. (2016). Living Long and Well: Prospects for a Personalized Approach to the Medicine of Ageing. *Gerontology, 62*(4), 409-416. doi:10.1159/000442746

Ghanim, H., Sia, C., Korzeniewski, K., Lohano, T., Abuaysheh, S., Marumganti, A., . . . Dandona, P. (2011). A resveratrol and polyphenol preparation suppresses oxidative and inflammatory stress response to a high-fat, high-carbohydrate meal. *J Clin Endocrinol Metab, 96*(5), 1409-1414.

Gruber, J., Chen, C., Fong, S., Ng, L., Teo, E., & Halliwell, B. (2015). Caenorhabditis elegans: what we can and cannot learn from aging worms. *Antioxidants & Redox Signaling, 23*(3), 256-279.

Hamed, M., Spaniol, C., Nazarieh, M., & Helms, V. (2015). TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. *Nucleic acids research, 43*(W1), W283-288.

Hashwah, H., Schmid, C., Kasser, S., Bertram, K., Stelling, A., Manz, M., & Müller, A. (2017). Inactivation of CREBBP expands the germinal center B cell compartment, down-regulates MHCII expression and promotes DLBCL growth. *Proceedings of the National Academy of Sciences of the United States of America, 114*(36), 9701-9706.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Henderson, S., & Johnson, T. E. (2001). daf-16 integrates development and environmental inputs to mediate aging in the nematode Caenorhabditis elegans. *Current biology : CB, 11*(24), 1975-1980.

Ho, Y., Matteini, A., Beamer, B., Fried, L., Xue, Q., Arking, D., . . . Walston, J. (2011). Exploring biologically relevant pathways in frailty. *J Gerontol A Biol Sci Med Sci, 66(9):*, 975-979.

Hoffmann, R., & Valencia, A. (2004). A gene network for navigating the literature. *Nature genetics, 36*(7), 664.

Kofler, N., Shawber, C., Kangsamaksin, T., Reed, H., Galatioto, J., & Kitajewski, J. (2011). Notch signaling in developmental and tumor angiogenesis. *Genes Cancer, 2*(12), 1106-1116.

Kucera, M., Isserlin, R., Arkhangorodsky, A., & Bader, G. (2016). AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. . *F1000Res, 5*, 1717.

Leong, D., Teo, K., Rangarajan, S., Lopez-Jaramillo, P., Avezum, A., Orlandini, A., . . . Lekishadi, R. (2015). Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet, 386*(9990), 266-273.

Lin, J., Mironova, Y., Shrager, P., & Giger, R. (2017). LRP1 regulates peroxisome biogenesis and cholesterol homeostasis in oli-godendrocytes and is required for proper CNS myelin development and repair. *eLife, 6*, 330498.

Liu, M., Liang, K., Zhen, J., Zhou, M., Wang, X., Wang, Z., . . . Yi, F. (2017). Sirt6 deficiency exacerbates podocyte injury and proteinuria through targeting Notch signaling. *Nature communications, 8*(1), 413.

Llorente-Cortes, V., Otero-Nivas, M., Sanchez, M., Rodriguez, C., & Badimon, L. (2002). Low-Density lipoprotein upregulates low-density lipoprotein receptro-related protein expression in vascular smooth muscle cells: possible involvement of sterol regulatory element binding protein-2-dependent mechanism. *Circulation, 106*(24), 3104-3110.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Luo, Y., Wu, Y., Brown, M., & Link, C. (2009). Caenorhabditis elegans Model for Initial Screening and Mechanistic Evaluation of Potential New Drugs for Aging and Alzheimer's Disease. In B. J (Ed.), *Methods of Behavior Analysis in Neuroscience - 2nd edition*: Boca Raton (FL): CRC Press/Taylor & Francis.

Luyten, W., Antal, P., Braeckman, B., Bundy, J., Cirulli, F., Fang-Yen, C., . . . Rattan, S. (2016). Ageing with elegans: a research proposal to map healthspan pathways. *Biogerontology, 17*(4), 771-782.

Mekli, K., Marshall, A., Nazroo, J., Vanhoutte, B., & Pendleton, N. (2015). Genetic variant of Interleukin-18 gene is associated with the Frailty Index in the English Longitudinal Study of Ageing. *Age Ageing, 44(6)*, 938-942.

Mercken, E., Crosby, S., Lamming, D., JeBailey, L., Krzysik-Walker, S., Villareal, D., . . . Fontana, L. (2013). Calorie restriction in humans inhibits the PI3K/AKT pathway and induces a younger transcription profile. *Aging cell, 12(4)*, 645-651.

Micale, G., Continella, A., Ferro, A., Giugno, R., & Pulvirent, i. A. (2014). eCollection 2014. GASOLINE: a Cytoscape app for multiple local alignment of PPI networks. *F1000Res, 3*, 140.

Mizunuma, M., Neumann-Haefelin, E., Moroz, N., Li, Y., & Blackwell, T. (2014). mTOR2-SGK-1 acts in two environmentally responsive pathways with opposing effects on longevity. *Aging cell, 13*(5), 869-878.

Morton, E., & Lamitina, T. (2013). Caenorhabditis elegans HSF-1 is an essential nuclear protein that forms stress granule-like structures following heat shock. *Aging cell, 12*(1), 112-120.

Nandakumar, M., & Tan, M. (2008). Gamma-linolenic and stearidonic acids are required for basal immunity in Caenorhabditis elegans through their effects on p38 MAP kinase activity. *PLoS genetics, 4*(11), e1000273.

Nguyen, T., Caito, S., Zackert, W., West, J., Zhu, S., Aschner, M., . . . Roberts, L. (2016). Scavengers of reactive gamma-ketoaldehydes extend Caenorhabditis elegans lifespan

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and healthspan through protein-level interactions with SIR-2.1 and ETS-7. *Aging (Albany NY), 8*, 1759-1780.

Nisoli, E., & Valerio, A. (2014). Healthspan and Longevity in Mammals: A Family Game for Cellular Organelles? Author(s): . *Current Pharmaceutical Design, 20*(35), 5663-5670.

Oesper, L., Merico, D., Isserlin, R., & Bader, G. (2011). WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol Med, 6*, 7.

Paradis, S., Ailion, M., Toker, A., Thomas, J., & Ruvkun, G. (1999). A PDK1 homolog is necessary and sufficient to transduce AGE-1 PI3 kinase signals that regulate diapause in Caenorhabditis elegans. *Genes Dev, 13*(11), 1438-1452.

Reis, R., Xu, L., Lee, H., Chae, M., Thaden, J., & Bharill, P. (2011). Modulation of lipid biosynthesis contributes to stress resistance and longevity of C. elegans mutants. *Aging (Albany NY), 3*(2), 125.

Robine, J.-M., Jagger, C., & Crimmins, E. (2013). *Healthy Longevity: A Global Approach*: Springer Publishing Company.

Rollins, J. A., Howard, A. C., Dobbins, S. K., Washburn, E. H., & Rogers, A. N. (2017). Assessing Health Span in Caenorhabditis elegans: Lessons From Short-Lived Mutants. *The journals of gerontology. Series A, Biological sciences and medical sciences, 72*(4), 473-480. doi:10.1093/gerona/glw248

Sanders, J., Minster, R., Barmada, M., Matteini, A., Boudreau, R., Christensen, K., . . . Newman, A. (2014). Heritability of and mortality prediction with a longevity phenotype: the healthy aging index. *J Gerontol A Biol Sci Med Sci, 69*(4), 479-485.

Shaye, D., & Greenwald, I. (2011). OrthoList: a compendium of C. elegans genes with human orthologs. *PloS one, 6*(5), e20085.

Shirendeb, U., Reddy, A., Manczak, M., Calkins, M., Mao, P., Tagle, D., & Reddy, P. (2011). Abnormal mitochandrial dynamics, mitochondrial loss and mutant huntingtin oligomers in Huntington's disease: implications for selective neuronal damage. *Hum Mol Genet, 20*(7), 1438-1455.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Strickland, D., Au, D., Cunfer, P., & Muratoglu, S. (2014). Low-density lipoprotein receptor-related protein-1: role in the regulation of vascular integrity. *Arterioscler Thromb Vasc Biol, 34*(3), 487-498.

Sutphin, G., Backer, G., Sheehan, S., Bean, S., Corban, C., Liu, T., . . . Korstanje, R. (2017). Caenorhabditis elegans orthologs of human genes differentially expressed with age are enriched for determinants of longevity. *Aging cell, 16*(4), 672-682.

Vásquez, V., Krieg, M., Lockhead, D., & Goodman, M. (2014). Phospholipids that contain polyunsaturated fatty acids enhance neuronal cell mechanics and touch sensation. *Cell reports, 6*(1), 70-80.

Wang, J., Robida-Stubbs, S., Tullet, J., Rual, F., Vidal, M., & Blackwell, T. (2010). RNAi screening implicates a SKN-1-dependent transcriptional response in stress resistance and longevity deriving from translation inhibition. *PLoS genetics, 6*(8), e1001048.

Weir, H., Yao, P., Huynh, F., Escoubas, C., Goncalves, R., Burkewitz, K., . . . Mair, W. (2017). Dietary Restriction and AMPK Increase Lifespan via Mitochondrial Network and Peroxisome Remodeling. *Cell metabolism, 26*(6), 884-896.

Yuan, Y., Kadiyala, C., Ching, T., Hakimi, P., Saha, S., Xu, H., . . . Feng, Z. (2012). Enhanced energy metabolism contributes to the extended life span of calorie-restricted Caenorhabditis elegans. *The Journal of biological chemistry, 287(37)*, 31414-31426.

Zeitlow, K., Charlambous, L., Ng, I., Gagrani, S., Mihovilovic, M., Luo, S., . . . Gottschalk, W. (2017). The biological foundation of the genetic association of TOMM40 with late-onset Alzheimer's disease. *Biochim Biophys Acta, 1863(11)*, 2973-2986.

Zhang, J., Kuang, Y., Wang, Y., Xu, Q., & Ren, Q. (2017). Notch-4 silencing inhibits prostate cancer growth and EMT via the NF-κB pathway. *Apoptosis, 22*(6), 877-884.

Zuberi, K., Franz, M., Rodriguez, H., Montojo, J., Lopes, C., Bader, G., & Morris, Q. (2013). GeneMANIA prediction server 2013 update. . *Nucleic acids research, 41(Web Server issue)*, W115-W122.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supporting Information listing

Supplementary Table 1. Features of human health, lack of dysfunction(s), and associated genes.

Supplementary Table 2. Features of human health, (lack of) multiple diseases, and associated genes.

Supplementary Table 3. Features of human health, lifespan/longevity mediated by lack of disease, and associated genes.

Supplementary Table 4. Features of *C. elegans* health, based on genetic studies of health, and associated genes.

Supplementary Table 5. Features of *C. elegans* health, based on compound intervention studies affecting health, and associated genes.

Supplementary Table 6. Human genes associated with health, listing all genes from Tables 1+3.

Supplementary Table 7. *C. elegans* genes associated with health, listing all genes from Tables 4+5.

Supplementary Table 8. *C. elegans* genes associated with health, listing genes based on WormBase gene expression data.

Supplementary Table 9: List of diseases and functional terms associated with the miRNAs enriched as regulators of the largest human healthspan pathway.

Supplementary Table 10: List of diseases associated with the miRNAs enriched as regulators of the second-largest human healthspan pathway.

Supplementary Figure 1: The regulatory interactions between the largest human healthspan pathway and the corresponding enriched miRNAs.

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplementary Figure 2: The regulatory interactions between the second-largest human healthspan pathway and the corresponding enriched miRNAs.

Supplementary Figure 3: Comparison of expression patterns in two aging tissues, largest human healthspan pathway.

Supplementary Figure 4: Comparison of expression patterns in two aging tissues, second-largest human healthspan pathway.

Supplementary Figure 5: Comparison of expression patterns in two aging tissues, third healthspan pathway.

Supplementary Figure 6: Comparison of expression patterns in three disease-affected tissues, top 3 healthspan pathways.

Supplementary References

Figure legends

Figure 1: A healthspan pathway map for humans, based on Supplementary Tables 1-3, including the list of pathways/clusters with their labels and size (number of genes), and a zoom into the first pathway. The size of a gene node is proportional to its GeneMANIA score, which indicates the relevance of the gene with respect to the original list of genes to which another 20 genes are added by GeneMANIA, based on the network data. Genes upregulated by CR are shown in yellow, downregulated genes are shown in blue, and grey denotes genes for which no expression values are available in the caloric restriction dataset (Mercken et al. 2013). The color of an edge refers to the source of the edge in the underlying network, that is co-expression (pink), common pathway (green), physical interactions (red), shared protein domains (brown), co-localization (blue), predicted (orange), and genetic interaction (green). The thickness of an edge is proportional to its GeneMANIA “normalized max weight”, based

bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

on the network data. Genes from the GeneMANIA input list feature a thick circle, while genes added by GeneMANIA do not.

Figure 2: A healthspan pathway map for C. elegans, based on Supplementary Tables 4 and 5. See also Figure 1. Gene expression data reflect the effect of rapamycin (Calvert et al. 2016).

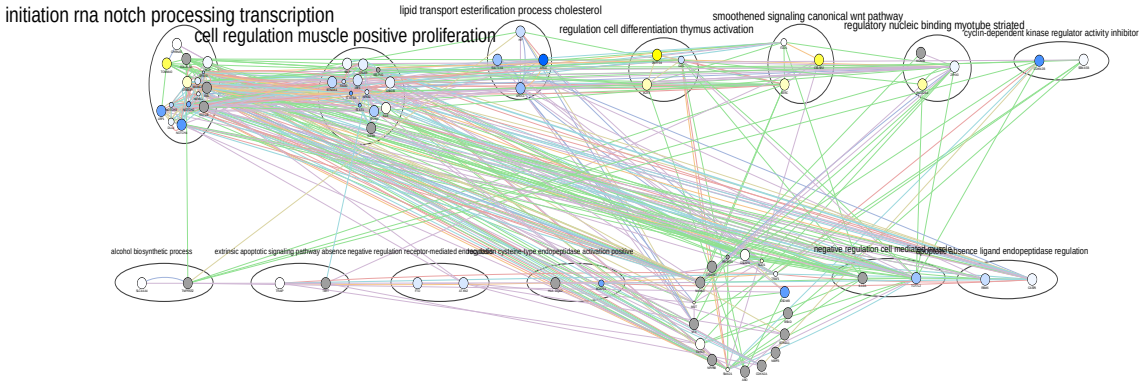
Figure 3: A healthspan pathway map for *C. elegans*, based on genes affected the most by healthspan-extending interventions, using WormBase gene expression data. See also Figures 1 and 2.

Figure 4: Workflow of the main analysis steps. First, 52 human health genes (Supplementary Tables 1-3) were processed with GeneMANIA and AutoAnnotate to determine the human healthspan pathway map (left, see also Figure 1). Analogously, 58 worm health genes (based on gene expression analysis using Wormbase) were studied, yielding the C. elegans healthspan pathway map (right, see also Figure 3). Then, to determine overlap across species, the gene lists were extended by the orthologs (calculated by WORMHOLE, see Supplemental Experimental Procedures) from the respective other species. We then employed GeneMANIA as before, to generate two interaction networks (one per list). and overlaps between these two networks of health genes were determined by GASOLINE (middle, see also Figure 5).

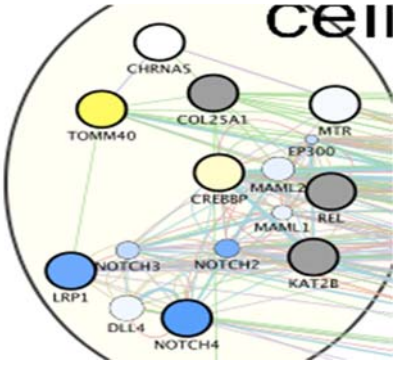
Figure 5: The two alignments demonstrating overlap of (putative) healthspan pathways in human and *C. elegans,* based on a GASOLINE alignment of the network of genes implicated in health-related gene expression changes in WormBase (top), and in human health based on genetic studies (bottom), and of corresponding orthologs. Dashed edges indicate orthologs, green edges indicate interactions based on GeneMANIA known for the respective species; the node shape is square if the gene originates from the original lists of health genes and it is circular if the gene is an ortholog, and node colors are based on gene expression changes triggered by rapamycin (in case of *C. elegans*) or by caloric restriction (in case of human), as in Figures 1-3.

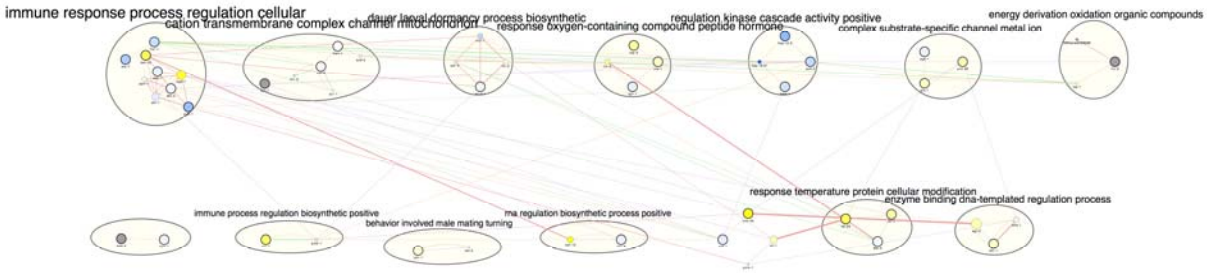
bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figures

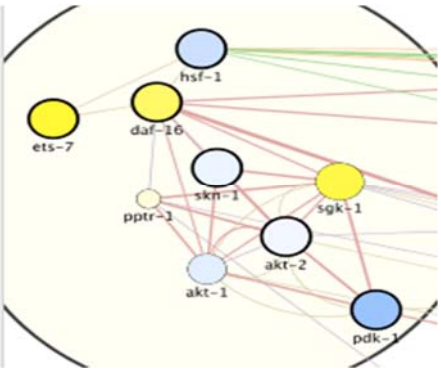


Cluster	...
initiation rna notch processing transcription	15
cell regulation muscle positive proliferation	13
lipid transport esterification process cholesterol	4
regulation cell differentiation thymus activation	3
regulatory nucleic binding myotube striated	3
smoothened signaling canonical wnt pathway	3
alcohol biosynthetic process	2
apoptotic absence ligand endopeptidase regulation	2
cyclin-dependent kinase regulator activity inhibitor	2
extrinsic apoptotic signaling pathway absence	2
negative regulation cell mediated muscle	2
negative regulation receptor-mediated endocytosis	2
regulation cysteine-type endopeptidase activation positive	2

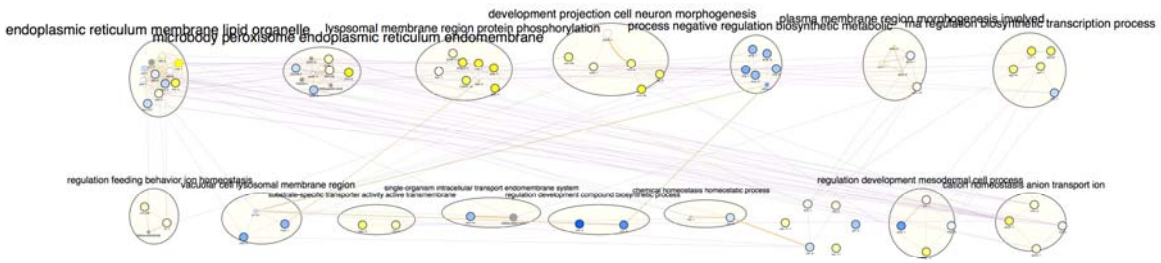




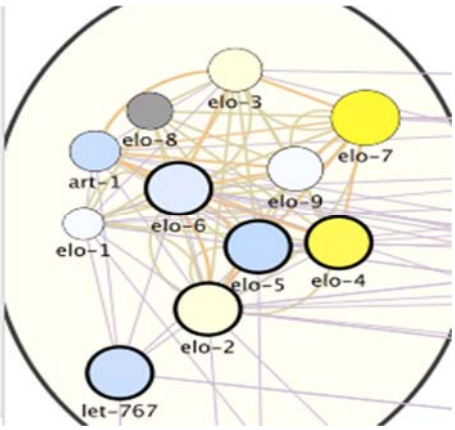
Cluster	...
immune response process regulation cellular	9
cation transmembrane complex channel mitochondrion	6
dauer larval dormancy process biosynthetic	4
regulation kinase cascade activity positive	4
response oxygen-containing compound peptide hormone	4
complex substrate-specific channel metal ion	3
energy derivation oxidation organic compounds	3
enzyme binding dna-templated regulation process	3
response temperature protein cellular modification	3
	2
behavior involved male mating turning	2
immune process regulation biosynthetic positive	2
rna regulation biosynthetic process positive	2

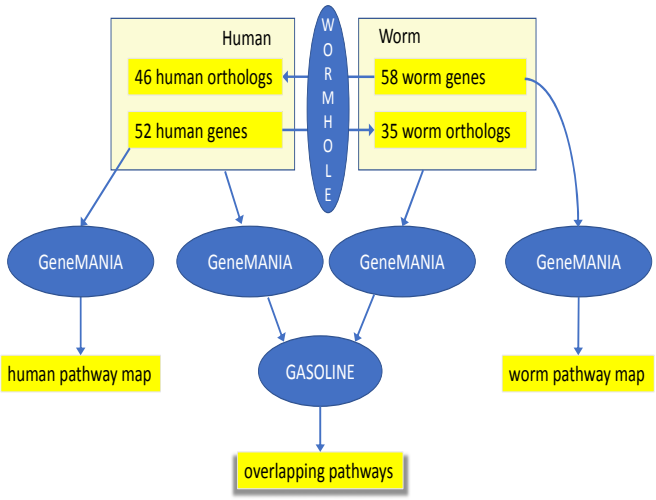


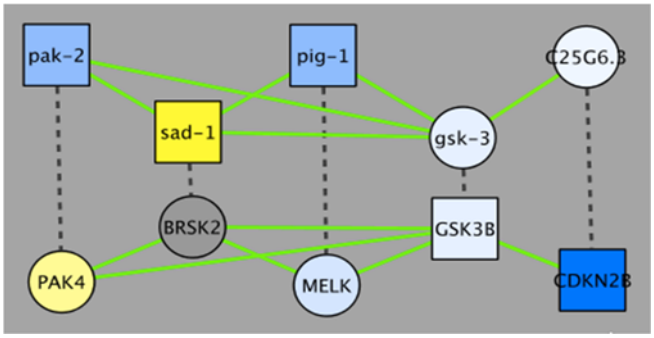
bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

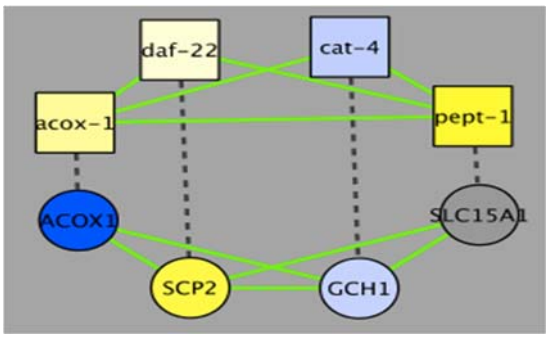


Cluster	...	...
endoplasmic reticulum membrane lipid organelle	11	
microbody peroxisome endoplasmic reticulum endomembrane	9	
lysosomal membrane region protein phosphorylation	8	
development projection cell neuron morphogenesis	6	
process negative regulation biosynthetic metabolic	6	
plasma membrane region morphogenesis involved	5	
rna regulation biosynthetic transcription process	5	
cation homeostasis anion transport ion	4	
regulation development mesodermal cell process	4	
regulation feeding behavior ion homeostasis	3	
vacuolar cell lysosomal membrane region	3	
chemical homeostasis homeostatic process	2	
regulation development compound biosynthetic process	2	
single-organism intracellular transport endomembrane syste...	2	
substrate-specific transporter activity active transmembrane	2	









bioRxiv preprint doi: https://doi.org/10.1101/355131; this version posted August 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.