EXTENDING HEALTHY LIFESPAN BY SYSTEMATICALLY TARGETING AGING PATHWAYS SYNERGIES

DIOGO GONÇALVES BARARDO

NATIONAL UNIVERSITY OF SINGAPORE

2021

EXTENDING HEALTHY LIFESPAN BY SYSTEMATICALLY TARGETING AGING PATHWAYS SYNERGIES

DIOGO GONÇALVES BARARDO

(M. Sc., BSc.)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2021

Supervisors:

Associate Professor Jan Gruber, Main Supervisor

Associate Professor Thilo Hagen, Co-Supervisor

Examiners:

Professor Brian Kennedy

Dr Cheung Chun Yue Maurice

Associate Professor Rattan Suresh, Aarhus University

Declaration Page

I hereby declare that this thesis is my original work, and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Diogo Gonçalves Barardo

January 2021

Acknowledgements

I would like to thank my supportive family, my dear supervisor Jan, my patient lab manager Li Fang.

A special appreciation goes to the most valuable thing in the universe: M.C.

Table of Contents

Summary

Aging is the main risk factor underlying the dramatic increase in incidence of cancer, cardiovascular and neurodegenerative diseases in aging populations. However, evidence over the last three decades has revealed that aging rate and individual aging-trajectory is more than previously appreciated. This fact opens the possibility for therapeutically relevant pharmaceutical interventions to ameliorate the impact of population aging in terms of disease burden and to potentially even delay the aging process itself.

Recent insights from work on drugs and genes affecting lifespan in model organisms has revealed that the most significant lifespan-effects are the result of pathway synergy and drug-drug interactions. While exciting, it is currently unclear how common such synergistic interactions are or how to best leverage them therapeutically. There is currently no validated approach to construct such interactions rationally and systematic screening for such interactions is challenging due to the combinatorial explosion in search space.

I have created and validated an automated high-throughput screening platform for the identification of healthspan-extending drugs in Caenorhabditis elegans. Using this system, conducted a large pilot screen for combinatorial drug benefits, identifying several new drug pairs with additive or synergistic benefits in terms of healthspan and lifespan. Furthermore, I generated and analysed transcriptional signatures (RNA-Seq) for each drug in my dataset of lifespan-extending drugs. This transcriptomics information, coupled with the results of the screen, were then used to test the nature of drug synergies, and explore potential approaches to predicting beneficial drug-drug interactions based on

transcriptional data on individual drugs. This analysis reveals that drugs interact mainly non-linearly, and that, in contrast to our previous results, dissimilar drugs are not more likely to be synergistic.

List of Tables

List of Figures

List of Symbols

α-KG	alpha-ketoglutarate
ACE	angiotensin-converting enzyme
AMPK	AMP-activated protein kinase
cDNA	complementary DNA
C. elegans	*Caenorhabditis elegans*
CI	confidence interval
CITP	Caenorhabditis Intervention Testing Program
CR	caloric restriction
DEG	differentially expressed gene
DMSO	dimethyl sulfoxide
DR	dietary restriction
EGCG	epigallocatechin gallate
ER	endoplasmic reticulum
FOXO	forkhead box protein O
GRN	gene-regulatory network
GSEA	gene set enrichment analysis
GSK-3	glycogen synthase kinase-3
HIF	hypoxia inducible factor
HSP	heat-shock protein

HTS	high-throughput screening
IIS	insulin/IGF-1
ITP	Interventions Testing Program
JNK	c-Jun N-terminal kinase
LFC	log-fold changes
LB	Luria-Bertani
MoA	mode of action
NAC	N-acetyl-L-cysteine
NAD	nicotinamide adenine dinucleotide
NGM	nematode growth medium
NRF-2	nuclear factor erythroid 2-related factor
PC	principal component
PCA	Principal Component Analysis
PCR	polymerase chain reaction
RIF	rifampicin
RNAi	RNA interference
RNA-Seq	RNA sequencing
TOR	target of rapamycin
YFP	yellow fluorescent protein

Chapter 1 - Introduction

1.1 Biogerontology

Although the main culprit of population aging is a decline of fertility rate in many societies[1], this socio-demographic crisis arguably is aggravated by the undeniable achievements of modern medicine that have significantly increased the average human life expectancy without always extending healthspan as much as lifespan[2]. This discrepancy, together with increasingly costly solutions in healthcare result in socioeconomic as well as ethical challenges that will impact more and more countries as populations around the world continue to age[3].

The prevalent disease-centric paradigm of modern medicine relies on diagnosis and treatment of specific, well-defined diseases. However the incidences of cancer, cardiovascular and neurodegenerative diseases all show a dramatic and simultaneous increase with age[4]. Therfore, approaches that targets one disease at a time, while likely leading to an extension of lifespan[5], will not necessarily result in similar increases in healthspan[6]. This culminates in an accumulation of individuals which are physically unable[7] to be active members of society and this may result in an increasing burden to the public healthcare system[8].

It is clearly documented that the aging rate may differs between individual animals of the same species and data from model organisms suggest that ageing is a highly malleable process[9,10]. Age is also the main risk factor responsible for the diseases underlying the aged population crisis, suggesting that if we were to delay the aging process itself, the prevalence of age-related diseases would be

delayed to our later years, in other words, we would achieve healthspan extension. This paradigm forms the basis of Biogerontology[11,12].

Of the intervention types discovered to drastically prolong the lifespan in model organisms, pharmacological interventions are the ones with the most potential to be eventually adapted at the population scale. With the exception of one compound, metformin, which has reached the clinical trial stage[13], most research in this area remains pre-clinical.

1.2 - Caenorhabditis elegans

The process of drug discovery typically starts by screening large libraries of candidate compounds, for their effect on the phenotype or disease model of interest. Although the use of mammalian models might be more immediately relevant for estimating the effects of longevity drugs in humans, there is a need for a model organism that are suitable for large-scale screening and target validation. The desiderata of an ideal model organism for high-throughput screening (HTS) for longevity drugs are: complex enough so that the different aging pathways and hallmarks are homologous to the ones in humans; short-lived, easy to handle and cheap enough so that HTS is possible. *Caenorhabditis elegans* (*C. elegans*) is a model organism that fulfills these requirements.

This free-living non-parasitic soil nematode enjoys several additional characteristics that made it one of the most popular model organism for assaying potential genetic[14] and pharmaceutical[15] longevity interventions. It takes only 3 days from hatching to adulthood. Additionally, its mean lifespan is about 20 days. Furthermore, being the first multicellular organism with its genome sequenced[16] it is easy to manipulate genetically using existing and well-

established tools[17] and, coupled with being monogenic (reduced inter-individual variability in populations of hermaphrodites), this facilitates the rapid elucidation of new longevity pathway or drug mode of action. As a case in point, the first aging gene ever discovered was identified more than three decades ago in *C. elegans*[18].

Importantly, research on this and related pathways since then suggests that the gene regulatory network of ageing and longevity determination is largely evolutionarily conserved between *C. elegans* mammals. As a consequence of this conservation, several lifespan-extending compounds such as rapamycin and metformin have been identified that appear to work in worms and also in mice[19].

1.2.1- High-throughput Lifespan Assays

The effects of candidate longevity interventions are assessed by survival analysis. Traditionally, acquiring a survival curve in *C. elegans* requires daily manual observation of animals cultured on solid agar in Petri dishes using a dissecting microscope. In more detail, for each petri dish, individual animal death is recognized by failure to react to a probing wire[20,21]. This routine protocol requires manual interaction with each individual animal in the cohort for every timepoint and is highly labor-intensive and therefore not amenable to HTS. Not surprisingly, several attempts have been made to allow HTS of longevity interventions in *C. elegans*.

Microfluidic based systems[22–24] have been developed but have the major drawback that they impose an unnatural culturing environment, as *C. elegans* are not aquatic animals[25]. This limitation makes the alternative class of

approaches more popular[26]. The alternative methods are based on high-throughput time-lapse microphotography or recording of animals, maintained on solid medium, throughout their lifespan.

The Lifespan Machine[27] is one such system. It is a modular system of modified commercial flatbed office scanners that repeatedly scan low profile petri dishes, and automatically generate survival curves. Unfortunately, this automated scoring involves four 10-minute, high intensity scans per hour, which may potentially to influence the lifespan of worms[26].

Another system of this class is the WormMotel[28]. This system consists of a plate handling robot coupled with a camera and custom-designed worm plates, that record the movement of individually housed worms longitudinally. The WormMotel also requires custom hardware that we did not have access to. I therefore decided to develop my own system based on hardware and expertise available in our group.

1.2.2- Gene Synergies in *C. elegans*

According to my literature review, the two largest lifespan-extending interventions ever reported in *C. elegans* belong to the *age-1(mu44)* mutant strain[29] and gonad-ablated *daf-2* mutants[30]. These interventions result in 10- and 6-fold lifespan extension relative to wild-type, respectively. However, in both cases, the individuals are vigorous and metabolic active but sterile. On the other hand, the third, fourth and fifth largest lifespan-extending interventions belong to the double mutants *daf-2 rsks-1*[31], *daf-2 clk-1*[32] and *daf-2 daf-12*[33], respectively. Unlike *age-1(mu44)* and gonad-ablated *daf-2*, these strains are

fertile and without severe fitness trade-offs (although this was not deeply tested).

A difference between the two sets of genetic interventions is that they are all cases of synergy. For example, the *daf-2* and *rsks-1* single mutants exhibited a 169% and 40% (or less) increase in lifespan, but the *daf-2 rsks-1* strain displayed a 454% longer lifespan than the wild-type N2. In other words, if there was independency between the modes by which *daf-2* and *rsks-1* prolong lifespan, we would expect additivity of effects (~209%), but, the lifespan of the double mutant strain vastly surpasses that.

Although highly desirable, synergistic anti-aging interactions are unpredictable. For example, in mice, testing all possible combinations of 3 gene therapies targeting longevity genes resulted in the identification of a synergistic pair, and a toxic pair and triple combinations[34].

1.2.3- Known Drug Synergies

In terms of pharmaceutical interventions, the same difficulty is generally observed. In crickets, combining metformin with aspirin prolonged lifespan to a lesser degree than any of the single drugs interventions[35]. In mice, when metformin is added to the lifespan-extending rapamycin, it slightly increases effect size[19]. In the rotifer *Brachionus manjavacas*, rapamycin and a c-Jun N-terminal kinase (JNK) inhibitor extended mean lifespan 65% more than either compound alone[36]. In *Drosophila*, combining rapamycin and/or lithium and/or trametinib[a] creates combinations that are better than any of the monotherapies[37].

[a] Trametinib is a mitogen-activated protein kinase (MEK) inhibitor.

In nematodes, we have previously reported a systematic combinatorial drug screen, and this result is discussed in more detail in section 1.5.1.

Possibly due to the complex and non-linear underlying mechanisms, drug synergies do not seem to be robust. For example, two anti-oxidants (resveratrol and n-acetyl-l-cysteine) could additively extend the lifespan of worms, but only if their dosage is carefully calibrated (revealing a U-shape dose-response curve)[38]. Furthermore, in fruit flies, two synergistic drugs pairs only worked on females[39]. Even in the unicellular yeast model organism, synergistic drug pairs were highly sensitive to the concentrations of drugs in the medium[40,41].

All the approaches discussed (which to best of my knowledge constitute all the known life-extending drug synergies) were the result of drug pairs that were constructed in a hypothesis-driven manner (based on detailed hypotheses regarding modes of actions of individual interventions and likely interactions between these modes). An obvious limitation of this approach is that it introduces bias in so far as novel, unexpected or counter-intuitive synergies would likely never be assessed (and therefore would remain undiscovered). To efficiently exploit drug-drug interactions, it would also be desirable design or predict candidate synergistic drug combinations in a hypothesis-free fashion, so that large libraries of compounds without known modes of action can be screened.

1.3 - RNA-seq

1.3.1 - Introduction

This brief introduction will be limited to the information necessary for the current dissertation and will be limited to the approach taken by me. Other types of transcriptomics paradigms and analyses such as: single-cell transcriptome sequencing[42], long-read transcriptome sequencing[42], de novo (or reference-based) transcriptome assembly[43], allele-specific expression analyses[44], expression quantitative trait loci mapping[45], splicing[46] could be explored in future.

Part of the new omics, the transcriptome refers to the set of transcripts present in a biological sample. In the current project, transcriptomics will be applied for the quantification of genome-wide expression changes of each transcript[a] between the control and treatment conditions. More concretely, I will make use of the latest[b] widely adopted high-throughput DNA sequencing method that allows mapping and quantifying transcriptomes, named RNA sequencing (RNA-Seq).

A basic step-by-step overview of an RNA-Seq experiment is depicted in Figure 1.1. One starts with the extraction and purification of RNA from the samples, followed by enrichment of target RNAs. In my case, poly(A) capture was used to select for polyadenylated RNAs[c]. The resulting RNAs are then fragmented to

[a] To be accurate, ultimately, I will compare gene-level expression changes derived by the respective transcript-level expression changes.
[b] Its original protocols date the year 2008[302,303].
[c] This is the most commonly used technique. Another popular alternative is ribosomal depletion.

the appropriate size. In the previous paragraph, I designated RNA-Seq as a high-throughput DNA sequencing method. This stems from the fact that the current market leading system, and the one I will use – Illumina – only sequences DNA. As consequence, the single-stranded RNA molecules need to be reverse-transcribed to cDNA (first strand), this is immediately followed by the degradation of the RNA, which then allows the first strand of cDNA to be complemented into a double strand. This cDNA synthesis step if followed by one last step before sequencing - adapter ligation and polymerase chain reaction (PCR). The final goal is to amplify the cDNA library by PCR, and for that the ends of every double-stranded cDNA must be standardized. This standardization consists in flanking each end with and adapter sequence that was ligated to either the 3' or 5' end, and, therefore, will work as primer in the reverse transcription reaction.

In my chosen technology, the final cDNA library is loaded into a flow cell of the sequencing machine, where through complementarity the adapter sequences allow the binding of cDNAs to short oligonucleotides. The process of bridge amplification is then leveraged to achieve dense clonal clusters of each cDNA loaded[47]. The elegant process of sequencing by synthesis is then used to determine the sequence of each cluster[48].

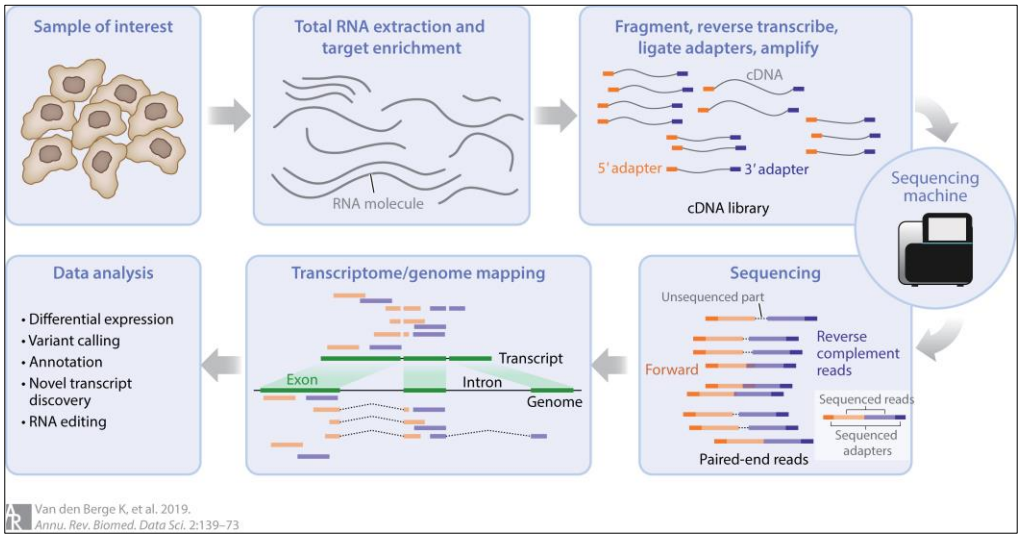

Figure 1.1 - Overview of the experimental steps in an RNA-seq protocol.
The cDNA library is generated from the RNA targets isolated from the samples of interest, and then the paired-end reads[a] are mapped against a reference transcriptome[b]. This is followed by data analysis, of which some possibilities are enumerated. Source of the picture acknowledge in the left bottom corner.

Regarding the design of RNA-Seq experiments, the intuitive basic rules of scientific experimental design apply. For example, randomizing and blocking over batches are recommended and critical for the subsequent statistical modelling[49]. Furthermore, there are specific design aspects that should be considered, such as the number of replicates, sequencing depth and length[50]. In praxis, however, the first limiting factor in the experimental design is budget[42].

Several sample size calculators for detecting differential expression are available to guide the experimental design upon the user definition of input parameters such as the expected alignment rate, the desired statistical power, significance level, and log-fold changes (LFC) of differentially expressed genes (DEG). Unfortunately, these tools are of limited use due to considerable

[a] As opposed to single-end. They are so called this way because both ends of the cDNA insert are sequenced, yielding one read from each end, in opposite orientation; and this is advantageous[50].
[b] This is in my case.

disagreement among their output[51], and difficulty in defining the outcome with the required level of precision on the part of the user. The latter originates from the fact that, such as the origins of the previously generated data that I will be using, most RNA-Seq experiments are exploratory in nature.

1.3.2- Alignment

The product of the sequencing machine is a set of files containing data of billions of short cDNA fragments. The first step to translate this output into quantitative biological information is the alignment to a reference genome.

There has been a panoply of alignment software available[52], but until very recently this process would be time consuming. Such was the magnitude of this hindrance that an alternative approach allowing to circumvent the alignment step altogether was proposed. Ultimately, this approach, based on the extraction of *k*-mers from reads followed by their exact matching using a hash table, culminated in the Kallisto software[53]. This is the software that I selected, as without any sacrifice in performance, it reduces the computation time by approximately 3 orders of magnitude[53].

The speed of Kallisto results from the fact that it does the pseudoalignment of reads and fragments, that is, it only focuses on identifying the transcripts from which the reads could have originated, without attempting to pinpoint the exact alignment of the sequences of reads and transcripts.

Also, part of the alignment process are sequence read quality-control (on raw and filtered sequenced data), trimming and filtering (of low-quality bases and

calls marked as N). The FastqPuri is an RNA-Seq specific software solution for these steps and it especially leverages paired-end sequencing data[54].

1.3.3- Quantification and DEG Analysis

Transcript to Gene-Level

After obtaining the abundance levels of the transcripts, several distinct analysis paths can be undertaken. This choice should be dictated by the underlying biological granularity most likely to yield a satisfactory answer to our biological question. Transcript-level analysis is used in differential transcript expression and in differential transcript/exon usage (differential splicing) analysis pipelines, which assess whether individual transcripts are differentially expressed between conditions and if for a given gene its expressed isoforms composition change, respectively. Gene-level analysis is the choice in the canonical DEG pipeline.

I am interested in the changes in the overall expression output of genes to obtain the holistic transcriptome differences induced by drugs, so, accordingly DEG will be my sole focus. Nonetheless, it has recently been shown that incorporating transcript-level estimates leads to slightly more accurate DEG analysis results compared to traditional simple counting apporaches[55]. An R language[56] package, *tximport*[55] was created and it will be leveraged to this effect in my pipeline.

Gene-length and GC-content Biases

Inherent to the method of RNA-Seq are biases that should be ameliorated by proper data normalization. In other words, variation due to technical bias and

limitations should have a minimal impact on the results, and therefore deliberate accounting and removal of these systemic effects should be undertaken.

Several methods for normalizing data in terms of how to best express the aligned counts in terms of their biological assumptions and statistical approach are available[57,58]. Notwithstanding, there are well-defined biases that should take priority in terms of normalization.

In a DEG-based pipeline, the priority should be accounting for gene-length bias. This bias has recently been shown to be the major responsible for the functional misinterpretation of RNA-Seq data[59].

Gene-length bias is the overrepresentation of long genes in the DEG subset of genes. This happens because longer genes in fact tend to get more counts than equally expressed shorter genes. The main cause of this effect is the experimental step in which molecules are fragmented prior to sequencing, which is used by current RNA-Seq protocols to gain sequence coverage of the whole transcript. The result of this step is that a longer transcript will have more reads mapping to it compared to a shorter gene of otherwise equivalent expression, and since the sampling size is proportional to statistical power, there is more power to detect DEG that are longer in length.

There are two R packages shown to be effective in removing effect[59], of those I will choose to use the *EDASeq* package[60], as it has the additional functionality of accounting for GC-content bias, which is the other well-defined bias that was hinted above.

The causes of GC-content bias appear more complex[61], but from a statistical modelling perspective this bias is said to be in effect when the read counts

display a unimodal distribution according to their GC-content, in which GC-poor and GC-rich fragments are statistically under-represented.

The EDASeq package normalization approach consists of two steps: within-lane normalization followed by between-lane normalization[60]. The first step adjusts for within-lane gene-specific effects, such as the aforementioned gene length or QC-content biases, and the second step deals with between-lane distributional differences, like sequencing depth.

For the first step, the authors implemented four within-lane normalization methods. I will use the full-quantile normalization because this was the normalization used to effectively remove gene-length bias effects[59]. In my case, this normalization will stratify the genes into *K* equally sized bins according to their length. The quantiles of the read counts distribution are then matched between "length-bins" , by sorting counts within bins and then taking the median of quantiles across bins (this is analogous to the between-lane normalization of Bullard *et al*.[62]). Full-quantile normalization is also my choice among the three available methods for between-lane normalization.

Differential Expressed Genes Detection

After the gene counts have been normalized for known biases, one can finally proceed to filtering, modelling, estimation, and statistical inference of the gene expression changes between conditions. There are a multitude of approaches and software to do so, but I picked the *DESeq2*[63] package for three reasons: it is easily integrated with the output from the EDAseq package (it can use its normalizing factors directly); it is shown to be the most robust to outliers and low replicate number (which is my case)[64]; its robustness can be further refined,

namely for low-expressed genes, by its integration a method that uses a heavy-tailed Cauchy prior distribution for effect sizes[65] (which I will describe shortly).

1.3.4- Aging Gene Sets

A crucial method to increase the biological interpretability and relevance of the results to a given research question is to restrict the universe of genes being tested to the ones that are known to causally influence longevity or for which the gene expression significantly changes with age. I will make use of one gene set of the first type and two of the second.

The GenAge database of aging-related genes[14] is the largest compilation of genetic interventions that are experimentally shown to influence the longevity of model organism. In other words, it compiles lifespan-extending and lifespan-decreasing genetic interventions These are mostly cases of gene knockout and overexpression experiments.

Tarkhov *et al.* gathered 60 publicly-available age-dependent transcriptomes of *C. elegans* and after scaling gene expression variation by median lifespan variation and applying dimensionality reduction found a signature of 327 aging-related genes[66].

1.4 - Dimensionality Reduction Methods

As described in the previous section, the output of an RNA-Seq experiment is a matrix of n samples and their respective p gene expression features (which are usually counts or logarithmic fold changes relative to a baseline). It is almost always the case that the number of these features is on the order of three

magnitudes larger than the sample size, convincingly setting the RNA-Seq data analysis in the high-dimensional paradigm of $p \gg n$.

Clustering techniques (namely hierarchical clustering) are a great way to perform quality control and classifying new samples into previously established clusters, but not very elucidative in terms of the biological underpinnings of such categorization. And this is the reason dimensionality reduction methods are the preferred methods in the more final phases of RNA-Seq data analysis pipelines.

Dimensionality reduction methods aim to describe the gene expression changes of a dataset in a lower-dimensional space. They are based on creating new features, according to well-defined criteria, that incorporate most of the information of the known features p, and, therefore, bypassing the $p \gg n$ issues altogether.

Moreover, these new features are tractable in their aggregation formula and therefore further inspection of how each of the aggregated p features that constitute them is being weighted. The weights of the p features can then be used for biological insights, for example, as a gene importance measure.

1.4.1 - Principal Component Analysis for Biologists

There are two reasons for preferring Principal Component Analysis (PCA) in place of other dimensionality reduction technique: compared to specialized techniques like Independent Principal Component Analysis[a], PCA is general in

[a] which better models the statistical distribution of gene expression originating from microarrays[304].

its application, widely popular and well characterized; in contrast most of other popular techniques like Independent Component Analysis[67], the PCA calculation is deterministic, and its independence on random seeds makes it robust to one source of the "Reproducibility Crisis"[68].

Instead of discussing in detail linear algebra, its long history and implications, I will rather fully describe the intuition behind its motivation and innerworkings, with only the minimal formalism necessary.

Let us assume a dataset of the expression of 2 genes in n samples. One could do the scatter (XY) plot of the samples according to the expression of these two genes – each of the genes constituting an axis – and see how the distinct samples relate together in this bidimensional space. If the dataset was expanded to 3 genes, the same inference requires an extra axis for the extra gene expression values. For this case, a tridimensional plot would be required (3 axes) to gain intuition about how the samples relate to each other in a single graph. The cruse of high dimensionality is felt there are several thousands of genes. Humans can only visuality up to 3 dimensions, so the one axis per gene correspondence is no longer a viable approach. What PCA allows is to define a new chosen number of axes, that are created from the linear combination of gene expressions and that still retain most of the sample information, to cluster the samples in a more human-friendly low-dimensional space. The goal is to bypass the limit imposed by the one-to-one gene-axis correspondence, while at the same time retaining or even increasing[a] the interpretability of the dataset.

[a] When there is significant noise present in the gene expression measurements.

Each PCA axis is called a principal component (PC). For a two-dimensional PCA-plot, the first of the axes, PC1, will stretch out in the direction where there is the most variance in the dataset; and the next axis, PC2, will have the restriction of having to be orthogonal to the first, and extend in the direction where there is the second most spread of variance. From here, it can be seen that an additional benefit of using PCA is that the axes are ranked in terms of explanatory power, while in the second toy example it was not clear which of the 3 genes was the most discriminatory (and this would be exponentially aggravated if an entire transcriptome was to be used).

Assuming p genes and n samples, a step-by-step intuition of how PCA is calculated goes as follows[a]:

1. Center and scale the data – although optional, this is a recommended step. Centering is done so that there will be a well-defined origin point for the axes, which corresponds to the mean of all gene expressions (mean centering). Scaling the data is a way to standardize gene expression across all genes, so that a two-fold increase of a lowly expressed gene is weighted the same as a two-fold increase of a highly expressed gene. This is accomplished by dividing each expression value of a given gene by the standard deviation from that gene's mean expression.
2. Find the PC1 – in this step, the line that passes the origin and better fits the data points (samples) is calculated. This is the standard line fitting procedure consisting of calculating the line that maximizes the sum of

[a] Based on the singular value decomposition algorithm for the sake of simplicity, as the more commonly used NIPALS algorithm is more complex.

the squared distances from the projection points (the projection of the samples on the line) to the origin[a]. As a result, the formula of PC1 is a linear combination (weighted sum) of the expression of all the genes expression values. This vector is called the eigenvector, and the proportions of each gene contribution are called loadings (or loading score). It is these loadings that latter allow biologist to gain insights into the gene sets that explain most of the variation in the dataset. Lastly, the sum squared of distance used in PC1 is called the eigenvalue for PC1[b].

3. Find PC2 – this will be the best fitting line that passes on the origin and is perpendicular to PC1. The imposing of this perpendicular relationship to PC1 is called orthogonality. Apart from the orthogonality constraint, the fitting procedure is the same used for calculating PC1.
4. Find the rest of PCs – repeat step 3 until $n - 1$ PCs have been found[c].
5. Find the relative importance of each PC – this is accomplished by calculating the amount of variation that is accounted by each PC. In PCA terminology, these are called the PC loadings and can be inferred by dividing the eigenvalues by $n - 1$.

With slightly more formulism, an alternative formulation of PCA[69], that is helpful for comparison to the dimensionality reduction technique mentioned in the next section, is one in which the principal component vectors are given by the eigenvectors of the non-singular portion of the covariance matrix C:

[a] Which is equivalent to minimize the distance between the points and the line.
[b] And the square root of the eigenvalue is called the singular value of PC1.
[c] In the paradigm of $p \gg n$. In general, the maximum number of PCs is $\min(p, n) - 1$.

$$C = \frac{1}{n-1} X^T C_n$$

where X is a $n \times p$ data matrix (of n samples with p gene expression features), $C_n$is the $n \times n$ centering matrix, and X^T is the transpose of matrix X. Furthermore, as previously hinted, the loading vectors of C, denoted by $L_1,\ldots,L_n$, are given by:

$$L_i = \sqrt{\lambda_i} e_i \qquad\qquad i = 1, \ldots, n$$

where $e_1,\ldots,e_n$ are the eigenvectors and $\lambda_1,\ldots,\lambda_n$ are the eigenvalues of C.

1.5 - This Thesis

1.5.1 - Context

We have previously published a pilot study[70] that serves as the basis of my PhD project.

In a hypothesis-driven fashion, an initial drug library of 11 compounds was selected, targeting 4 distinct but connected well-known longevity pathways, of which only 5 reproducible extended the lifespan of *Caenorhabditis elegans* in our lab.

Based on the intuition that drugs regulating highly connected pathways are more likely to produce synergistic effects, we tested all the possible pair-wise combinations of these 5 compounds. We found that 2 (rapamycin with rifampicin, and psora-4 with rifampicin) out of the 10 drug pairs acted synergistically[70]. Furthermore, some triple combinations were tested (addition of a third drug to the 2 synergistic pairs) and displayed further lifespan-

extension. These triple drug combinations have the largest reported effect size for any adult-onset pharmacological intervention in a model organism. Moreover, the triple drug combination interventions also doubled healthspan[70]. In other words, we maximized effect size without compromising translatability potential.

1.5.2 – Hypotheses and Goals

Hypothesis 1 – Combinations of dissimilar drugs are synergistic.

As discussed in section 1.2.31.2.2 and in the previous paragraph, it has been suggested that combining drugs targeting distinct subsets of the aging regulatory network is more likely to result in synergistic pharmaceutical interventions. Albeit the previously discovered anti-aging drug synergies were designed assuming that drugs with dissimilar modes of action are likely to be synergistic, careful observation of the tested interventions reveals that a large part of them were not synergistic.

It is my hypothesis that in a larger size combinatorial drug screen, I will be able to test if combinations of dissimilar drugs are more likely to be synergistic. However, this also requires a quantitative metric of "drug similarity", something that I will further explore in Chapter 6.

Hypothesis 2 – Gene expression changes in known ageing genes are predictive of effect size for monotherapies.

Another implicit assumption in the cases of all the previously synergistic pharmaceutical longevity interventions is that simply targeting additional nodes in the aging regulatory network leads to increased effects size (sections 1.2.3

and 1.5.1). Albeit, once again, there were a large number of cases of lack of synergistic interaction even though distinct aging related gene sets were being targeted.

One hypothesis to test was that combining drugs that target additional known aging genes will result in a concomitant increase in effect size. In other words, that targeting more aging genes (and/or having a bigger impact in terms of favorable fold-changes) would result in larger therapeutic effect.

Hypothesis 3 – Combinatorial interventions are linear combinations of their monotherapies.

A corollary of this assumption is that transcriptional changes of a drug combination is assumed to be mostly the result of a linear combination of the changes seen for drugs that constitute it. At least to first approximation, perturbations involving a pair of drugs are assumed to be modelled as linear super-position of the individual drug effects. For two drugs (Drug A and Drug B), the DEG for the drug pair (AB) needs to be, at least to first approximation:

$$AB \sim \alpha * A + \beta * B$$

Where alpha and beta are real-valued factors allowing for linear interactions between genes (e.g., saturation effects or changes in effective *in vivo* drug concentration due to global effects on drug detoxification and transport pathways). This linearity assumption is a strong assumption, and I will explicitly test how far from realistic it is (see sub-chapter 6.3).

Goal 1 – Design and validate an automated HTS for longevity interventions in C. elegans

Both of my hypotheses require a larger number of drug combinations to be assayed (as well as the respective monotherapies). The lack of a satisfactory solution (section 1.2.1) led me to design and validate a new automated HTS for longevity pharmaceutical intervention in *C. elegans* (Chapter 3).

Goal 2 – Conduct a combinatorial HTS to identify new drug synergies.

After having created a functional automated HTS, I used to identify new synergistic drug combinations (Chapter 5).

Goal 3 – Generate and analyze a larger RNA-Seq dataset of lifespan-extending pharmaceutical intervention.

Hypothesis 2 requires me to know which known aging-genes are target by my drugs. I found out this information by generating and analyzing the transcriptome of each of my drugs (Chapter 4).

Goal 4 – Test my hypotheses using drugs' transcriptomic profiles and screen results.

I tested the **similarity assumption** (hypothesis 1) and **linearity assumption** (hypothesis 2) by combined computational analysis of the data previously generated (Chapter 6).

Chapter 2 - General Experimental Methods

2.1 - Drug Library Selection

I wanted to maximize the chance of identifying novel drug-drug interactions and I therefore created a careful set of criteria which candidate drugs had to fulfill before being considered for inclusion in my test set.

According to the DrugAge (build 3) database of aging-related drugs[a], at the time there were 567 compounds that had been shown to extend the lifespan of model organisms. By far the most popular among these model organisms is *C. elegans* with 395 compounds tested through a total of 970 individual lifespan assays.

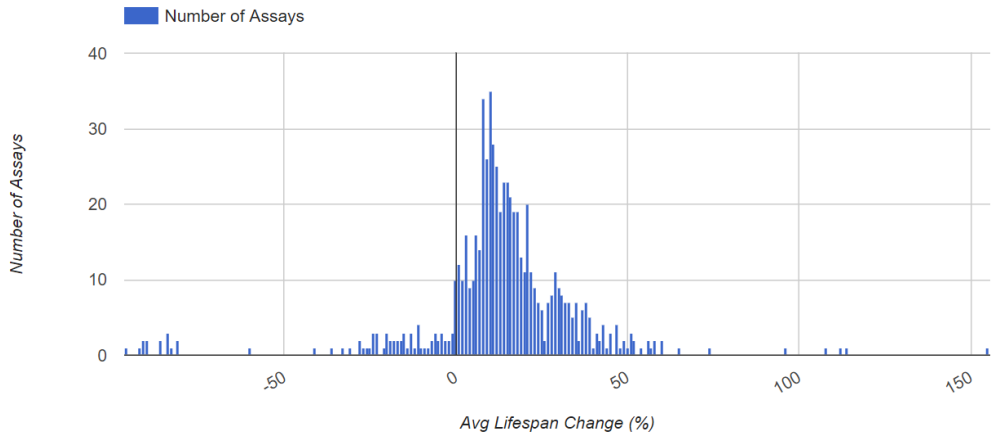

Figure 2.1 – Distribution of average lifespan change for *C. elegans* in GenAge. Distribution of number of GenAge lifespan assays that resulted in a given magnitude of average lifespan change in worms.

The first criterion that I applied was based on effect size displayed by a compound. The mode of the distribution of average lifespan changes reported in DrugAge[71] for *C. elegans* is 10% with a standard deviation of 20%. To maximize my statistical power, I set an effect size threshold of 30% (equivalent

[a] Freely available at https://genomics.senescence.info/drugs/index.php .

to one standard deviation over the mode). That is, I considered only compounds that at least in one lifespan assay statistically significantly extended the average lifespan of *C. elegans* by 30%. I further prioritized compounds that work at the standard temperature of 20 C and excluded those for which control lifespan was abnormally short. Specifically, I required the control group having a mean lifespan of at least 17 days.

Some compounds (for example some FDA-approved drugs) require the mode of administration to be intravenous administration to have therapeutic effects. This makes these compounds potentially hard to further investigate in vertebrates and I therefore also only selected compound that could be delivered through oral administration.

For similar reasons, I excluded compounds shown to extend lifespan in worms but that had already failed to do so in mammals.

Lifespan studies can be notoriously difficult to reproduce and show large lab-to-lab variability. I therefore excluded compounds that other members of our lab had tested previously but for which they were unable to reproduce published lifespan benefits.

Furthermore, to increase the reproducibility and tractability of my results, natural extracts or multi-compound formulations were rejected as possible candidate pharmaceutical interventions.

Lastly, and taking in consideration the goals of my project, after looking at the remaining candidate drugs, I gave priority to compounds with published information suggesting a clearly defined and distinct modes of action and large effect size. Applying these criteria, I arrived at final drug library comprising 15

drugs in total: alpha-ketoglutarate, aspirin, captopril, curcumin, epigallocathecin-3-gallate, DL-alpha-lipoic acid, icariin, lithium chloride, n-acetyl-l-cysteine, myricetin, piceatannol, resveratrol, spermidine, thioflavin-T and ursolic acid. For each of these compounds I conducted a full literature review to further refine / inform my understanding of mode of action and potential or reported interactions.

2.2 - Drugs Literature Review

In the following subsections, I will analyze the available literature for each drug that is part of my library plus all 5 drugs that were used in our previous work[70a]. The methodology for identifying the relevant literature for each compound was the following: I started by extracting the literature referred in DrugAge[71]; I then checked the abstracts of literature that cite the previously selected DrugAge references; additionally, I included new literature that is yet to be included in DrugAge but that I was aware of.

To acquire expertise is not sufficient to exhaustively read the necessary literature. One must also examine it in a critical fashion, namely, through the lens of consilience[72]. In our context, consilience is applied by searching for agreeable and disagreeable evidences among distinct research papers, and checking if there is the emergence of a stronger conclusion[73]. Even if there is no general convergency of evidence, this exercise allows the development of an unbiased holistic view of the current state of knowledge.

[a] This is because I will be joining the RNA-Seq samples that we had previously generated with the ones from this project.

2.2.1 - Allantoin

According to the DrugAge[15] database (build 3), the only time that allantoin was subjected to a lifespan assay was in *C. elegans*[74]. In this paper, the authors selected allantoin with the hypothesis that it would function as a caloric restriction mimetic. The evidence suggesting this hypothesis was that allantoin is one of the drugs inducing a transcriptional profile in human cell lines that most resemble the transcription profile induced by caloric restriction (CR). Since the transcription profiles used came from Connectivity Map[75], and were of mammalian origin, the authors then proceeded to test allantoin in *C. elegans*.

Allantoin, at a dosage of 250μM starting with adulthood, significantly extended mean lifespan by 20.1% and 4.2% on wild-type and *eat-2*[a] (DA465 strain) background, respectively. I interpret this result as showing that allantoin is at best a partial CR mimetic, as supported by the additional lifespan increase on the *eat-2* strain.

Evidence against my interpretation is provided by the authors[74] when they show that allantoin significantly extended the lifespan in *daf*-16(mgDf50) worms by 19.7%. The rationale for testing in this strain is that it is known that lifespan extension by CR does not require DAF-16[76]. Nonetheless, I counter-argue that it might just be that the only commonality between CR and allantoin is to the extent that they simply extend lifespan in a *daf-16* independent manner.

From a healthspan perspective, allantoin displays mixed effects of its components in N2 worms[74]: it increased healthspan if measured by pharyngeal

[a] *eat-2* strains are the standard genetic model of CR. They exhibit reduced food intake due to impaired pharyngeal pumping[305].

pumping rate[77]; but had no effect on healthspan as measured by movement (body bends per minute).

In our recent paper, we show that for another strain of EAT-2[a], allantoin fails to extend lifespan. Furthermore, we replicated the DAF-16 independence of allantoin, albeit using the *daf-16*(mu86) strain. Additionally, we demonstrate that allantoin does not extend the mean or maximum lifespan of the *C. elegans* transforming growth factor beta homologue DAF-7[b] mutants.

It is worth noticing that the anti-aging translation potential of allantoin is supported by our results on *Drosophila melanogaster*[70]. In more detail, we show that at the same previously used dosage, the mean lifespan of wild-type Oregon-R male fruit flies is significantly extended by 15%.

2.2.2 - Alpha-ketoglutarate

To the best of my knowledge, the first time in which alpha-ketoglutarate (α-KG) was tested in a lifespan assay was in 2014[78]. It was a success (and lifespan extension has since been replicated in two other studies[79,80]), with the magnitude of lifespan extension shown to be independent of whether the treatment was started at the egg stage or adulthood. In metrics of healthspan, α-KG did not influence the rate of egg laying, total progeny, and pumping rate.

Regarding the mode of action (MoA), the authors suggest that α-KG specifically disrupts the energetic flow in the complex V of the electron transport chain by binding to the ATP synthase subunit beta, and therefore, being an uncompetitive

[a] *eat-2*(ad1116) strain.
[b] *daf-7*(e1372) mutants.

inhibitor of it. Namely, the longevity benefits of α-KG require ATP-2[a] because there was no lifespan extension in the already longer lived worms that had the expression of *atp-2* gene knockdown by RNA interference (RNAi)[78].

Additionally, five lines of evidence were provided to make a compelling case that α-KG is a CR mimetic. Firstly, as it is the case with *eat-2;daf-2* double mutants[78], α-KG fed *daf-2* worms live longer than their respective *daf-2* controls, that is, the longevity mechanism of α-KG appears to be independent of the insulin/IGF-1 (IIS) pathway. Secondly, similarly to *eat-2* mutants α-KG fails to increase the lifespan of *CeTOR*(RNAi) animals. This suggests that α-KG requires TOR/*let-363*. Thirdly, the lower ATP content present in α-KG fed worms may partial being inducing a CR-like state, and this is supported by the failing to extend the lifespan of *eat-2*(ad1116) animals. Regarding this point, the authors fail to mention that the early mortality (corresponding to approximately the first half of their lifespan) of N2+α-KG is much less than the one achieved even in *eat-2*+α-KG[78]. Fourthly, the FoxA transcription factor PHA-4[81] is required by α-KG to prolong longevity, just like what happens in the case of dietary restriction (DR) and reduced CeTOR signaling. The last evidence consists that in the same vein as TOR inhibited, dietary restricted and *atp*-2(RNAi) animals, α-KG fed wild-type worms exhibit increased autophagy, while α-KG had no further effect on the autophagy levels of *atp-2*(RNAi) and *CeTOR*(RNAi) animals.

In this very same work[78], the authors were further able to reveal that α-KG is partially dependent on the AMP-activated protein kinase (AMPK) and forkhead

[a] the orthologue of ATP5B in *C. elegans*.

box protein O (FoxO)[a] pathways, as judging by its significant but smaller magnitude of lifespan extension on the *aak-2*[b] and *daf-16* strains. In contrast, another canonical aging pathway, the hypoxia inducible factor (HIF-1) pathway, does not seem to be involved, as α-KG extends the lifespan in loss-of-function *hif-1*, *egl-9,* and *vhl-1* mutants.

Latter work shows that the pro-longevity effect of α-KG is evolutionary conserved in *Drosophila melanogaster*[82]. Although, in this model organism reproduction is compromised for some dosages but not others. Additionally, α-KG effects on other fly healthspan assays were null (normal tolerance to oxidative stress, starvation, and desiccation) or positive (enhanced climbing ability and increased heat stress resistance).

In great support of the evolutionary conservation of the MoA proposed in the aforementioned worm manuscript, in flies α-KG also: activates leads to a reduction of ATP and ATP/ADP ratio, AMPK, inhibits the target of rapamycin (TOR) pathway and activates autophagy-associated genes[82]. In sum, the entire biological chain of events elicited by α-KG seems to be entirely conserved. This is interesting, considering that the α-KG performance in the Caenorhabditis Intervention Testing Program (CITP) was positive in 3 different *C. elegans* wild-type strains but neutral or even detrimental to the longevity of 3 wild-type strains of *C. briggsae*[83].

[a] FOXO in humans, is the O subclass of the forkhead family of transcription factors. These transcription factors have a fork head protein domain.
[b] *aak-2* carry a deletion on the gene that encodes the catalytic (α) subunit of AMPK, the cellular sensor of low energy levels.

2.2.3 - Aspirin

Aspirin or acetylsalicylate is rapidly converted *in vivo* to salicylate, and therefore it is considered its pro drug. Accordingly, I analyzed the literature that directly administered either compound, to present a holistic view of aspirin as a potential pro-longevity intervention.

In *D. melanogaster*, aspirin increases the minimum and maximum lifespan in both genders (albeit at different dosages), with no clear effect on spontaneous activity and activity as tested by negative geotaxis[84]. In this paper, it exists sexual dimorphism in other measures of healthspan: in males aspirin increases the resistance to heat, paraquat, and starvation, while in females it is of benefit only in starvation conditions. The results on the stress-resistance of female flies were largely contradicted by a more recent study that shows that they do display increased resistance to heat-stress, to two types of oxidative stressors and to starvation[85]. Nonetheless, both works agree in that there is a decrease in the number of eggs laid[84,85].

Perhaps the most striking fact is that aspirin-fed flies live longer although their food consumption increases[84]. I might be the first considering this free-lunch a signature of aspirin treatment, but my reasoning is that it is also occurs in crickets (on top of extending mean and maximum lifespan in both genders)[86]. Furthermore, the fact that the lifespan of female flies is increased whether they are fully-fed[84,85] or food-restricted[85].

In *C. elegans*, two studies report an extension of mean lifespan of N2 but not of *daf-16* mutants, accompanied by an improvement in motility and pharyngeal pumping rate[87,88]. Specific to each study, it is also shown that aspirin reduces

the formation of aging-associated protein aggregates[87], increases resistance to oxidant[87] and heat-stress[88]. The robust lack of longevity effect on the *daf-16* strain, coupled with the absence of changes to peroxide resistance[87] and to the expression of the *sod*-3 antioxidant gene (which is heavily induced by aspirin in the wild-type[87,88] and a known DAF-16 target[89]), suggests that most or all the pro-longevity effects of aspirin are mediated by the DAF-16/FOXO pathway.

Like α-KG, the MoA of aspirin is suggested to consist in the increase of the AMP to ATP ratio[a], which leads to an increase of AMPK. This is supported by the fact that aspirin does not extend the lifespan of *aak-2* or *par-4* worms[b88]. This can be considered the main candidate explanation for aspirin effects, as it is the same mechanism suggested to be at play in flies[85].

Alternative modes of action explaining the pro-longevity effects of aspirin in *C. elegans* are: the reduction of mitochondrial respiration (evidenced by the lack of longevity effects on *clk-1*[c] and *isp-1*[d] mutant strains[88]); inhibition of proliferative germline stem cells leading to the activation of DAF-16 (supported by the absence of increased longevity and heat-stress resistance in *glp-1* germline deficient mutants fed aspirin[90]); and competitive inhibition of EP300, which induces autophagy (namely mitophagy) in worms and mice[91]. In contrast, the hypothesis that the aspirin-mediated extension of lifespan is the result of DAF-16 activation by the binding of silent information regulator 2 (SIR2) is

[a] which is taken to be equivalent to the reduction of the ATP to AMP ratio.
[b] AMP can stimulate AMPK by two mechanisms: allosteric activation, and, mainly, by binding to AMPK and promote its phosphorylation by LKB1. This latter mechanism is conserved from worms to humans[306], with PAR-4 being the worm homolog of LKB1.
[c] strain defective in the enzyme involved in ubiquinone synthesis.
[d] strain with a defect on a component of the respiratory chain complex III.

rejected empirically, as aspirin treatment increases the lifespan of the null mutant worm strain *sir-2.1*(ok434)[88].

Curiously, aspirin slightly extended the lifespan of *daf-2*(e1370) animals[88], suggesting only a partial dependence on the IIS pathway. Although, the small magnitude of lifespan extension (approximately 3%) might just be an artifact due to a residual expression of DAF-2, since *daf-2*(e1370) is not a null mutant. If this is the case, aspirin could fully depend on the IIS pathway.

Unfortunately, under the recent gold standard of CITP, aspirin treatment did not altered the longevity of any of the species (and respective strains) tested[83]. Furthermore, studies using *Mus musculus* conclusively show that aspirin does not extend the longevity of females[92–94]. Regarding male mice, in an initial report, aspirin extended average lifespan (no effect on maximum lifespan)[93], but it subsequently failed to do so under the American National Institute of Aging Interventions Testing Program (ITP), leading the authors to consider their first report a false positive[94].

Taking into consideration the lack of reproducibility in mammals and under the CITP, the best way to translate this intervention to humans might be on a case-by-case basis[95]. A cheerful example of oral aspirin benefits in a well-defined human subpopulation is the reduction in all-cause mortality observed in a cohort of 226 centenarians[96].

2.2.4 - Captopril

Captopril is an FDA-approved treatment for high blood pressure. It is an angiotensin-converting enzyme (ACE) inhibitor, which is of importance as the

ACE gene has been conserved from bacteria to mammals[97] and it is one of the few genes with genetic polymorphisms robustly associated with longevity in humans[98].

It has only been explored once under the biogerontology paradigm and was shown to extend the mean and maximum lifespan of *C. elegans*[97]. Moreover, its pro-longevity effect was robust in 3 distinct culture temperatures and to live and heat-killed bacteria feeding conditions; but with no effect on brood size, reproductive span, and pharyngeal pumping rate. The same authors replicated these effects using *acn-1*[a] RNAi-treated worms, with the exception that the RNAi treatment was able to increase the pharyngeal pumping rate from days 12th to 20th of adulthood.

There is substantial evidence that the pro-longevity MoA of captopril is the inhibition of the ACN-1 gene: combined treatment of *acn-1* RNAi with captopril, did not have and addictive effect on lifespan and both interventions have the same results in distinct genetic backgrounds. Going into more detail on the later point, reducing the activity of ACN-1: has additive pro-longevity benefits with caloric restriction (*eat-2*), mitochondrial insufficiency (*isp-1*) and with long-lived IIS pathway mutants (*daf-2* and *age-1*); it is a lifespan-extension pathway independent of TOR (*rict-1*), proteotoxic stress (*hsf-1*)[b] and of *sir-2.1* activity; and it is toxic to *daf-16* worms, reducing their lifespan.

[a] the *C. elegans* homologue of the ACE gene.
[b] although it is also revealed that *acn-1* RNAi increases resistance to heat-stress.

Even though captopril has not been assayed for longevity in mammals, its pro-longevity MoA is tractable and evolutionary conserved in humans[98] and flies[a], robust to different conditions and additive to the main pro-longevity genetic interventions (which indicates its potential to be combined with other drugs).

2.2.5 - Curcumin

Curcumin is a component of turmeric in Indian curry. The metabolite of curcumin, tetrahydrocurcumin, extends the mean and maximum lifespan of male C57BL/6 mice when treatment is initiated at 13-months of age, but is of no effect when started at the 19th month of age[99]. However, in under the ITP these benefits were not reproduced[100]. Despite failing in the gold standard testing program, curcumin is still consideration from a biogerontology perspective due to its very robust results in all the other model organisms tested, e.g. in worms, curcumin does promote lifespan[101b]. The rest of supporting literature was conducted in flies.

Curcumin robustly extends lifespan of male and female D. melanogaster from two distinct wild-type strain, confers additional protection against oxidative stress and improves climbing ability[102]. In yet another fly strain, a female and a male group of flies exhibited increased mean lifespan and activity of the superoxide dismutase[103]. Moreover, several works reproduced the gender and strain-independent pro-longevity effect of curcumin in flies[104–106]. Other beneficial effects include enhanced progeny viability and parental reproductive

[a] another FDA-approved ACE inhibitor, lisinopril, extends the lifespan of 3 *D. melanogaster* strains through the same MoA[307].
[b] I am disregarding results in which the control group had a mean lifespan of less than 9 days. Albeit there is still lifespan-extension in this work[308].

fitness[107]. Larval feeding of curcumin in flies did not yield any additional benefit on CR flies, which indicates potential overlapping modes of action between these two pro-longevity interventions. Furthermore, age-specific lifespan assays reveal that curcumin treatment effects are incredibly age-dependent, going from benefic to harmful; accordingly to if the treatment is stated at the healthspan to the senescent span, respectively[104,106]. An additional factor influencing the magnitude of lifespan-extension caused by curcumin treatment is temperature. Through regulation of heat shock proteins, the higher the environmental temperature, in other words, heat stress, the larger the effect size relative to untreated flies, independently of gender[108].

2.2.6 - Epigallocathecin gallate

The first trial of the popular green tea polyphenol epigallocatechin gallate (EGCG), resulted in no effects on the survival of worms, but improvement in healthspan-associated traits including augmented resistance to oxidative stress, attenuated decline of pharyngeal pumping rate and enhance chemotaxis index in old age[109]. Three years after, the potent oxidative stress resistance induced by EGCG was replicated, but this time it was accompanied the a reported mean lifespan-extending effect of 10%[110].

The unclear role of EGCG as a lifespan-extending compound in worms was resolved by Liu Gui Xiong *et al.*[111] which showed that this effect is robust to different temperatures and genetic background[a], but highly sensitive to

[a] two distinct wild-type strains were used.

concentration[a]. EGCG induces longevity by causing a transient increase in reactive oxygen species, that stimulates the endogenous detoxification defense system. More importantly for my approach, co-administration of n-acetyl-l-cysteine abolished EGCG-mediated lifespan and oxidative stress resistance benefits and further supports that EGCG is a case of mitohormesis[112]. Due to the transient induction of reactive oxygen species, and consequent defense response, declining with age, EGCG treatment is progressively blunted with age. This work also reiterated the requirement of the DAF-16 pathway for the EGCG-induced longevity that was reported in a previous manuscript[113].

The proposed mode of action of EGCG seems to be conserved in mammals, as male weaning Wistar rats fed EGCG display increased median lifespan, improved age-associated oxidative stress and superior activation of the FoxO3a longevity factor[114].

For the sake of completeness I must mention that there is an alternative mode of action for EGCG based on results obtained in flies that proposes that reduction in glucose metabolism is the key contributor for the superior fitness and lifespan observed in EGCG-treated *D. melanogaster*[115].

2.2.7 - Icariin

At the optimal dosage of 45μM when under 25ºC, icariin extends the mean lifespan of worms by 25%[116]. Icariin is not the active biological form, and this flavonoid is eventually hydrolyzed to icariside II. The direct administration of

[a] the reader might notice that this is common for compounds that act mainly by antioxidant mechanisms. For the relevant literature please read the subsection about n-acetyl-l-cysteine.

icariside II resulted in increased tolerance to thermo and oxidative stress and slowed the locomotion (as measured by swimming bends) decline in late adulthood. In terms of the genetic pathways involved, both icariin and icariside II fail to extend the mean lifespan of *daf-16*(mu86) and *daf-2*(e1370) mutants, but they did prolong the lifespan of *eat-2*(ad1116) and *rsks-1*(ok1255) strains; suggesting that icariin depends on the IIS pathway and it is not a caloric restriction mimetic[116]. Furthermore, the expression of SOD-3 increased as confirmed by PCR and fluorescent marking.

The results from mice are in agreement with the ones obtained in worms: there is a mean lifespan extended by 8% but no maximal lifespan; and the SOD gene expression is increased, moreover it is indistinguishable from youth levels[117]. Treated mice were lighter than the control individuals, even though they consumed more calories. Healthspan improved as measured by the Morris water maze, rotarod and bone mineral density (worth notice that bone mineral density is conserved at youth levels). The icariin-treated mice also display less DNA damage as indicated by the decreased expression of gamma-H2AX.

2.2.8 - Lipoic Acid

Lipoic acid was first shown to extend the lifespan of female and male *D. melanogaster*, by 12% and 4%, respectively[118]. Very recently, this effect on female flies was replicated even when started only at day 26 (mid-aged)[119]. The pro-longevity effect of lipoic acid are associated with the prevention of age-associated functional decline and hyperproliferation of intestinal stem cells, through the activation of the endocytosis-autophagy network[119].

The literature in *C. elegans* reveals that lipoic acid administration results in an increase in the mean and maximum lifespan, with no effect on pharyngeal pumping rate, but with the enhancement of the chemotaxis index[a] in aged worms. The lifespan-extending effect of lipoic acid on this model organism has been confirmed only in regard to mean lifespan, but not maximum[120].

Male rats fed an *ad libitum* diet supplemented with lipoic acid display similar survival profiles as the control animals[121].

2.2.9 - Lithium

The first initial report of lithium as a lifespan-extending drug showed that this effect was independent of DR, FOXO/IIS or germline signaling pathways[122]. Additionally, it was observed a trade-off between longevity and fertility, as lithium treated animals produced less eggs, and most eggs that were laid subsequently failed to hatch. Due to the lithium status as an FDA-approved drug for the treatment of psychiatric disorders[123] as a classic glycogen synthase kinase-3 (GSK-3) inhibitor, the authors conducted lifespan assays of lithium under different genetic interventions in the worm orthologue of this gene. Inhibition of GSK-3β is not sufficient for lifespan extension (mutants live shorter), and lifespan extension by lithium actually requires GSK-3β (because it further decreases the short lifespan of *gsk3* mutants).

The beneficial effects of lithium on the mean lifespan of worms were reproduced, and there were no changes seen in maximum lifespan. Although,

[a] it measures the fraction of worms able to display goal-oriented motor responses from the transformation of specific sensory stimuli.

lithium treatment did not significantly slow the rate of aging as measured by mortality rate doubling time. The slower decline in locomotor phenotype at all ages, together with the previous mention studies, shows that lithium treatment in worms as a mixed effect on healthspan assays. Moreover, the authors create a dynamical model and supporting it by several lines of evidence conclude that lithium elicits a greater increase in autophagy than in mitochondrial biogenesis, and that this dictates the cellular respiratory capacity by influencing the ratio of functional and dysfunctional mitochondria. In sum, the increase in the ATP levels by lithium is due to a higher ratio of functional to dysfunctional mitochondria[124].

The lifespan-extending of lithium in *C. elegans* were reproduced in yet two other studies and with a similar dose-response curve[125,126], giving rise to the view that lithium is a robust pro-longevity treatment. Lithium therapy as a biphasic dose-dependent effect on lifespan, seen at similar concentrations in *in vitro*[127], worms and flies[128]. After a minimum threshold to elicit a therapeutic effect, lithium treatment eventually becomes toxic at higher concentrations[122].

On top of the biphasic lifespan extension effects being reproduced in flies, in this model organism these are independent from gender and genetic background[128]. Lithium is arguably even more benefic than in worms because it elicits a significant improvement and protection against age-related locomotor decline, but this time without compromising feeding behavior or fecundity. With relevancy for the translational potential of lithium as an anti-aging therapy, this work validated that lithium extends lifespan even when administration starts only in mid-life or just with a short-term treatment in the case of young flies. Like in worms, lithium extends the lifespan of DR flies, but it seems that it does

so by inhibiting the fly orthologue of GSK-3. Also, in opposition to what was suggested based on worm experiments, lithium does not induce or require autophagy to promote longevity in flies. These authors propose a model in which lithium inhibits GSK-3, which indirectly leads to the activation of the transcription factor nuclear factor erythroid 2-related factor (NRF-2)[a].

Very exciting form a translational point-of-view, are two studies showing correlation evidence that lithium concentration in drinking water is associated with reduced all-cause mortality, in distinct human populations, and with a similar dosage therapeutic window to the aforementioned pre-clinical results[125,129].

As discussed in the known drug synergies section (section X), lithium as a pro-longevity treatment can be paired with rapamycin or trametinib for additional benefits, which can be even more powerful by combining all these drugs as a triple pharmaceutical treatment[37].

2.2.10 - N-acetyl-L-cysteine

N-acetyl-L-cysteine (NAC) is a natural source of cysteine which in turn is used in the synthesis of glutathione, and as such, it is a potent antioxidant *in vivo*. It was initially discovered to increase the mean and maximum lifespan of flies by 26.6%[130]. Albeit this effect seems to be only conserved in the male gender of genetically heterogenous mice, and it is marked by an accentuated weight loss, which might hint at the induction of a partially DR-like state[131].

[a] since GSK-3 is an NRF-2 inhibitor.

In *C. elegans*, it was first shown that liposomal delivery of NAC improves survival, in opposition with the lack of effect seen from its direct administration[132]. In contrast, it was later shown that direct NAC administration extended the mean and maximum lifespan, by up to 30.5% and 8 days, respectively. Furthermore, it was found that NAC increases the number of progeny, and resistance to radiation and heat shock stress[133]. These discrepant evidence can be conciliated by the recent results showing that NAC has a highly sensitive dose-response curve. NAC is a canonical antioxidant and as such its efficacy is slave to the inverted U-shaped dose-response relation between reactive oxygen species levels and lifespan[38].

2.2.11 - Metformin

Metformin is a biguanide drug commonly used to treat type-2 diabetes.

In worms it extends median lifespan[134–137] and delays the age-related decay of mobility[134,138][2,11]. The lifespan extension seems to be independent of the IIS pathway (DAF-16, DAF-2, and AGE-1). The lack of effect of metformin in *eat-2*(ad1116) mutants coupled with the display of several phenotypes associated with DR (lower lipofuscin accumulation[138], slimmer bodies, extended period of egg-laying) and the genetic requirement of SKN-1 and AAK-2 strongly suggests that metformin is a CR mimetic. Important from a translational perspective, metformin did not impaired feeding[134]. An alternative model for the mode of action of metformin is mitohormesis. Metformin increases reactive oxygen species production, metabolic heat production and respiration, and, crucial to our paradigm of combinatorial drug interventions, the antioxidant

NAC abolishes its benefits[135a]. These are the two main ones, but not the only models seeking to elucidate the mode of action of metformin[137,138].

In an attempt to confirm that metformin promotes longevity in an evolutionary conserved manner in flies, indeed it was observed a robust activation of AMPK and reduced body fat, but no lifespan extension in either gender. Moreover, metformin was toxic in some of the concentrations tested[139].

Metformin has also been tested in some less common animal models. In crickets metformin extends survivorship and maximal longevity[86]. Metformin significantly reduced growth rates and delayed maturation in crickets of both genders. The same authors also tried the aspirin plus metformin drug pair, but the lifespan-extension was less than in any of the single drugs. Furthermore, metformin prolongs the lifespan of in short-lived fish and delays several markers of aging, including lipofuscin, inflammaging[b], cell senescence and cognitive decline[140].

Results in middle-aged male mice of two different strains are more encouraging[141]. Metformin improves general fitness and lifespan at a low dosage[c], but it is significantly nephrotoxic (renal failure) at a higher one. Animals in the low-dose long-term metformin treatment are initially slimmer (even though they consumed more calories than the control animals, like in worms), but tended to preserve their bodyweight with advancing age. Intermittent treatment regimens of every-other week or two consecutive weeks

[a] the requirement of AAk-2 would equally be justified under this alternative model.
[b] the chronic low-grade inflammation that its characteristic of advanced age.
[c] this dosage still corresponds to serum levels an order of magnitude higher than those used in the treatment diabetic human patients.

per month were tested, with the aim of bypassing the toxicity associated with high levels (relative to the human therapeutic dosages) of long-term metformin treatment. Unfortunately, neither extend mean or maximum lifespan in mice[142].

In rats, metformin supplementation replicates the reduced food intake and body weight seen in the caloric restricted group but did not extended lifespan at any quantile[143].

2.2.12 - Myricetin

The two works with lifespan assays using the naturally occurring flavonol myricetin were conducted in worms. Their findings agree: on its pro-longevity effects (18%[144] and 33%[145] mean lifespan increase, 22% maximum lifespan increase[144]), on its anti-oxidant capacity and that it elicits the nuclear translocation of the daf-16 protein. They do diverge, however, in the effects of myricetin on the daf-16(mu86) genetic background. On one there is a pro-longevity effect on this background[144], while on the other this is completely abolished[145]. It is suggested that the differences might be consequence of the markedly different culturing conditions used, solid medium at 20ºC versus liquid medium at 25ºC, respectively.

The great anti-oxidant potential of myricetin might explain its reduction on the accumulation of lipofuscin[145], a biomarker of aging, if one takes into consideration that lipofuscin are highly oxidized crossed-linked proteins.

Additionally, myricetin does not increase tolerance to heat-stress, pharyngeal pumping rate or body size, which are all evidence against a possible role as a caloric restriction mimetic[145].

2.2.13 - Piceatannol

In the sole research paper assessing piceatannol through lifespan assays in worms, it is shown that this natural stilbene increases median lifespan by 18%[146]. Excitingly, piceatannol does so without signs of toxicity (no alteration of growth rate, worm size or progeny production), and with a wide range of positive effects on healthspan markers (delayed age-related decline of pumping rate and locomotive activity; and increase resistance to heat and oxidative stress).

The fact that enhanced stress-resistance and lifespan-extension is lost in the daf-16(mu86) mutants, coupled with fluorescent marker evidence of daf-16 protein nuclear translocation, and increase gene expression of its downstream targets, logically led to the suggestion that piceatannol acts via DAF-16. Through further lifespan assays additional genetic background requirements were discovered, including DAF-2, AGE-1 and EAT-2 but not CLK-1 (for which it extended median lifespan by 22%)[146].

2.2.14 - Psora-4

Psora-4 is an inhibitor of the potassium channel Kv1.3. In the only time that it was assayed for longevity, besides the work from our group, Psora-4 prolonged *C. elegans* longevity by 42% - the second largest effect size present in the large-scale screen on which it was initially identified as a lifespan-extending compound[147]. It was also shown to have no significant effect on survival under conditions of oxidative stress.

2.2.15 - Rapamycin

Rapamycin is arguably the most famous anti-aging drug; therefore, this FDA-approved immunosuppressant drug is vastly studied, with one of the major aging pathways being named after it - mammalian target of rapamycin (mTOR). A systematic review of rapamycin potential in biogerontology is outside the scope of this dissertation, and this section provides only the upmost relevant information. The curious reader is advised to read the cited body of work.

In fruit flies, rapamycin slightly extends mean lifespan in a gender-independent manner, although at a significant fecundity cost[148]. Both these effects have been reproduced and shown to be modulated specifically through the complex 1 branch of the TOR pathway (TORC1), including the downstream upregulation of autophagy[149]. Furthermore, it is accompanied by increase stress resistance and it is efficacious even on long/lived IIS mutant and DR flies.

The literature in *C. elegans* shows that the lifespan extension from rapamycin treatment[150–154] is dependent on SKN-1, but not on DAF-16[155,156] or DAF-2[157]. Additionally it induces mitochondrial unfolded protein response and increases respiration[158], and ameliorates the age-related decline of pharyngeal pumping rate[156]. The fact that rapamycin treatment does not extend lifespan of eat-2 mutants and that it elicits a gene expression profile resembling that of a DR state, might suggest that it could be considered a calorie restriction mimetic[156][6], although evidence in mice disputes this line of reasoning[159][13].

Surprisingly, the majority of studies with rapamycin as an anti-aging intervention have been done on *Mus musculus* (e.g.[159,160]).

On the ITP, whether fed at 270 or 600 days of age, rapamycin extends median and maximal lifespan in both genders of genetically heterogenous mice[161]. In old mice, rapamycin restores the self-renewal and hematopoiesis of hematopoietic stem cells and boosts immune function[162].

Considering that mTOR is an evolutionary conserved aging pathway, in mice too, rapamycin administration reduces its activity, albeit not in all tissues[163]. Nonetheless, it is important to mention that muscle mass was maintained[163]. Interestingly, since rapamycin improves age-related phenotypes even in young mice, its lifespan-extension might be dissociated from the aging process itself[164].

Due to the side effects of inducing insulin resistance and immune suppression, several intermittent schemes of rapamycin have been attempted. A 2 weeks per month intermittent administration, is sufficient to increase lifespan, inhibit age-related weight gain, and delay spontaneous cancer incidence in inbred female mice[165]. Moreover, a once every 5 days rapamycin treatment regimen was shown to have no impact on blood glucose, while it still resulted on lifespan extension in 20-months old female mice[166]. Although, such regiment was elsewhere shown to not completely negate neither the immunosuppressive effects of rapamycin, nor the associated decrease in testis weight[167]. Hopefully complementary to the intermittent regimens, a 3-months transient rapamycin treatment extends the median lifespan of middle-age mice by up to 60%[168].

2.2.16 - Resveratrol

Resveratrol is a natural polyphenol found in grapes and red wine, and it is the compound for which there are the most lifespan assays[169].

Resveratrol activates sirtuin genes and extends lifespan in yeast[170–175], worm[38,154,182,183,158,170,176–181], flies[170,184], fish[185–188] and bees[189]. When nutrients are restricted, this effect is abrogated which suggests that it is related with caloric restriction[170,171].

In worms the lifespan extension is fully dependent on SIR-2.1[176,179,190] (at least sometimes[177]), AAK-2[190] and autophagy[178], but it is independent from DAF-16[176,190]. In this model organism, resveratrol treatment also induces mitonuclear protein imbalance and activates the mitochondrial unfolded protein response[158].

In *Nothobranchius furzeri*, a short-lived fish, it causes a 56% and 59% increase in median and maximum lifespan, respectively. Additionally, the treated group exhibits ameliorated decay of locomotor and cognitive function. The initial lower survival of the treated group for the first few weeks, led the authors to propose that resveratrol might be an hormetic compound[185]. In another vertebrate model, the fish *Nothobranchius guentheri*, the lifespan-extending effects and increased cognitive and locomotor function caused by resveratrol administration were conserved. Additionally, it was shown that there was less accumulation of lipofuscin and senescence (as measured by beta-galactosidase activity)[186]. In this species too, resveratrol decreases oxidative stress[188].

An additional mode of action to the hypothesis that resveratrol is a CR mimetic acting through activating sirtuins, namely because it was contested that

resveratrol can increase sirtuins *in vivo* in either worms or flies[191a], is as an antioxidant[188]. For example, in worms resveratrol increased mean and maximum lifespan and also oxidative stress resistance[183] (these effects have been reproduced elsewhere[179]), but also there is work showing that resveratrol does not have free-radical scavenging activity in vivo[180] (albeit in this particular work it also did not extended lifespan).

This antioxidant role it is not how resveratrol acts in bees, as the mean and maximum lifespan extension are abolished in hyperoxic stress and the honey bees ingest fewer quantities of food[189].

After the previous brief description of the lifespan-extending effects of resveratrol, it is also now crucial to underline the plethora of negative results.

In the ITP and in the only lifespan assay in a mammal, resveratrol did not extended lifespan[100]. Also, in the model crustacean Daphnia, resveratrol has no effect or even significantly decreases lifespan[192]. Furthermore, even though resveratrol was tested in both genders of a tephritid fruit fly species and under 24 different diets it still did not elicit any lifespan extension[193]. In D. melanogaster, using different strains, gender and diets, the pro-longevity effects of resveratrol were very variable, and eventually consider dubious[177]. Wang *et al.*, reported that resveratrol has effect only on female D. melanogaster and depending on the diet in question[184]. Staas *et al.* that dietary resveratrol had absolutely no effect on male and female *w^{1118} D. melanogaster* in terms of

[a] In yeast that is clearly not the case as resveratrol does increase SIRT1 12-fold[172], and certainly there is evidence that allows a counter-argument to be made[170,190].

lifespan, locomotor activity, body composition, stress response and longevity-associated gene expression[194].

In sum, I consider that resveratrol achieves its pro-longevity effects through caloric restriction mimicry or antioxidant capacity, in a species-specific manner. Moreover, I consider it perhaps the most unclear anti-aging compound: the mode of action and efficacy are highly debated, but either way it is extremely sensitive to experiment conditions. The incredibly unreliability of resveratrol treatment can be predicted if its main mode of action is as an antioxidant or mitohormetic compound, as both of these are known to has an inverted U-shaped dose-response on longevity[38].

2.2.17 - Rifampicin

The FDA-approved antibiotic rifampicin (RIF) or rifampin extends the lifespan of *C.elegans* by almost 45%, even in heat-killed bacterial lawns, which clearly indicates that its pro-longevity effects are not due to its bactericidal properties[195]. The very same work showed that rifampicin acts as a potent glycation inhibitor *in vivo*, reducing the age-relate accumulation of advanced glycation end products. It is worth mentioning that rifampicin treatment extended lifespan even when initiated at only day 9 of adulthood.

Besides its anti-glycation effects, rifampicin pro-longevity effects depend on the DAF-16 gene, as it was shown to activate daf-16 protein translocation into the nucleus, had no influence on the longevity of the daf-16(mgdf50) null mutant strain, and led to an increase in the expression of SOD-3 (which is a direct target of DAF-16). Although the targets downstream of DAF-16

modulated by rifampicin seem to be a separate subset from those regulated by the IIS pathway[195].

2.2.18 - Spermidine

Spermidine is a polyamine naturally present in humans, and its intracellular levels are known to decline during the human aging process. Therapeutic quantities are hard to be obtain through diet alone, but its supplementation is regarded as safe[196].

Research of spermidine in murine models is compelling. Four-months old C57BL/6J wild-type female mice subjected to a lifelong supplementation of spermidine in their drinking water, have their median lifespan extended. Even more relevant from a translational medicine perspective, there is still a 10% increase of the median lifespan of a group of male and female mice when spermidine supplementation is started later in life, on the 18th month[197]. Spermidine caused no changes in food and water consumption, and on body weight and composition, so the possibility of CR-like state being in play can be rejected. Instead, spermidine delays the aging process (namely cardiac aging) by eliciting autophagy. The same authors also report an inverse correlation between dietary spermidine intake and human cardiovascular disease[197]. I note that the corresponding spermidine human dosage to the one used in this mice study would be too high to be achieved easily by diet[198], which suggests that extra spermidine supplementation might have unexplored cardioprotective potential in humans. The pro-longevity effects of life-long intake of spermidine, at the same dosage, were replicated on a different wild-type mouse strain with

a median lifespan extension of up to 25%, and the autophagy-based mode of action was further substantiated[199]. In middle-aged Sprague-Dawley male rats spermidine supplementation fails to extend mean or maximum lifespan[200], but the dosage given was only 72% of the converted mice dosage[198]. Nonetheless, it still extended healthspan (if only very slightly), as measured by improved kidney tubules, liver, and heart morphology; increased exploratory behavior; and diminished expression of neuroinflammatory markers (once again the autophagy process was enhanced)[200].

Regarding other model organisms, spermidine retards chronological aging and rejuvenates replicative old yeast cells, and enhances the lifespan in *D. melanogaster*, *C. elegans* and human peripheral blood mononuclear cells. In yeast, mice and human cells, this spermidine-induced longevity is correlated with hypoacetylation of histone H3. Furthermore, and in completely agreement with the rest of the literature, it is shown that spermidine induces autophagy, and that this process is required for its lifespan-extension effect in yeast, flies and worms[201].

2.2.19 - Thioflavin-T

The intuition that compounds traditionally used in histopathology to stain amyloid in tissues might be candidate drugs for delaying the aging process comes from evidence that such compounds do bind and slow the aggregation of such protein aggregates *in vitro*[202], and that loss of proteostasis is one of the main hallmarks of aging[203]. Thioflavin T is an amyloid-binding dye that was first explored under this rationale in *C. elegans*, with an impressive 70% median

lifespan extension and 43%-78% maximum lifespan extension[204a]. These are perhaps the largest drug-induced pro-longevity effects ever reported for an adult-onset monotherapy at the standard culturing conditions of 20ºC in solid medium[15]. The efficacy of thioflavin T requires HSF-1 and SKN-1, but it is partly independent of caloric restriction, as it prolongs lifespan (albeit not to the same degree) in nutrient-based models of CR and *eat-2*(ad1116) mutants.

With this effect size on lifespan, it is not very surprising that thioflavin T was the most robust compound assayed in the CITP, showing significant media lifespan-extension in 5 of the 6 strains of *Caenorhabditis* tested[83]. Not only that, but it was also the most potent convincingly reproducing the initial result. The robustness and potency of thioflavin T suggest that its mode of action is well-conserved and of major importance among aging processes, respectively.

2.2.20 - Ursolic Acid

Ursolic acid is a lipophilic pentacyclic triterpenoid of botanical origins that was first shown to increase the mean and maximum lifespan of treated worms by up to 30%, although these effects were relative to a control group that had a mean lifespan of less than 15 days[205]. Later, the same group, with equally relatively short-lived controls, replicated their initial results and complement them with healthspan assays[206]. These latter also showcase the anti-aging potential of ursolic acid including in delaying the accumulation of lipofuscin, protecting against heat-shock, achieving an healthier chemotaxis index, and improving motility. Regarding the mode of action, since there is similar lifespan extension

[a] it also decreased age-specific mortality across all ages.

on the *daf-16*(mgDf50) and *daf-2*(e1370) genetic backgrounds, the mode of action is thought to be independent of the IIS pathway. Instead, the failure to be of benefit to a JNK-1 mutant strain and its predicted binding affinity with the jnk-1 protein supports that it is through the modulation of JNK-1 that ursolic acid exerts its pro-longevity effect[205]. Another additional mode of action is the initiation of a DR-like, supported by the observation that treated worms are slimmer and that there are no effects on the lifespan of *eat-2*(ad1116) animals[206].

In *w*^{1118} *Drosophila melanogaster* flies, ursolic acid is of benefit to males only. In this gender it increases the mean and maximum lifespan and healthspan (as measured by climbing ability and immune function), without any fertility trade-off. However, it does not increase oxidative stress resistance or affect gut health. Interestingly, total body weight remains the same even though ursolic acid-fed male flies ate significantly more food than controls. This makes sense in light that removal of the microbiome negates the anti-aging effects of ursolic acid. Which to me suggests that the treated flies could be benefiting from a microbiome-derived metabolite that is produced from the extra calories and activated by ursolic acid.

2.2.21 - Summary

While doing the literature review in the previous sections of this sub-chapter, I kept track of the gene epistasis of each drug. In more detail, I annotated the genetic background (mutants or RNAi) and the effect of the drug regarding lifespan. Table 2.1 summarizes the result of this annotating procedure. It only displays genetic epistasis information regarding genetic backgrounds that were

used to test at least two drugs. This criterion is to display only genetic intervention that might have discriminatory power, as it was my intention to check if the drugs could be clustered based on this information only. In case of information available from multiple sources, the ternary classification scheme used in Table 2.1, only features values for which there was agreement between sources. Furthermore, I collapsed the data from multiple loss-of-function mutants for the same gene and RNAi knockdown together, because otherwise the table would be too large and sparse.

Drug	daf-16	eat-2	daf-2	sir-2.1	clk-1	age-1	aak-2	skn-1	isp-1	hsf-1	daf-7	pha-4	nuo-6	sbp-1	jkn-1	jkk-1	mev-1
Lithium	1	1	1	1	1												
Ursolic acid	1	0	1	1				0		1		1			0	0	
Metformin	1	0	1			1	0	0									
aKG	1	0	1				1					0					
Rapamycin	1	0	1					0			0						
Psora	1	0	1								0			0			
Thioflavin T	1	0				1		0		0							
Resveratrol	1			0	-1		0		-1				-1				
Curcumin	1			0		0		0									1
Allantoin	1										0						
Piceatannol	0	1	0	0	1	0											
Icariin	0	1	0							0					0	0	
Rifampicin	0	1									1			1			
Aspirin	0	0	1	1	0		0		0								
EGCG	0		1	0	-1	1	0		0				0				
Lipoic acid	0																
Captopril	-1	1	1	1		1			1	1							
Myricetin																	1

Table 2.1 – Drugs genetic epistasis.
Based on the literature mentioned in the review. Columns represent distinct genetic intervention backgrounds. 1 – drug extends lifespan on that genetic background. 0 – drug has no effect on the lifespan. -1 – drug shortens average lifespan.

As it can be observed in Table 2.1, the drugs genetic epistasis information is overly sparse, not allowing a granular clustering of drugs.

Additionally, it is evident from Table 2.1, the strong bias for favoring known genetic backgrounds, for example, an anti-aging drug is almost always tested on daf-16 genetic backgrounds. This is important to keep in mind in Chapter 4, where I annotate the RNA-Seq differentially expressed genes based on the

available literature. The more popular pathways will keep being mentioned reflecting this bias.

2.3 - *C. elegans* Culturing

2.3.1 - Standard Conditions

For this dissertation only the Bristol N2 wilt-type and GRU101 mutant strains were used. The N2 strain was initially obtained from the Caenorhabditis Genetics Center Populations and the GRU101 strain was created in our laboratory[126]. These strains were growth and maintained on nematode growth medium (NGM) agar plates, incubated at 20ºC, and using *E. coli* OP50-1 as bacterial food source. The integrity of the populated plates was checked weekly, and the population was transferred to new plates as needed (but at maximum every other week).

In case of drug treatments, there was a standardization of the reagents, mediums and materials used, whenever possible. For example, in experiments, the NGM agar and bacterial food were sourced from the same batches.

2.3.2 - Bacterial food source maintenance and preparation

The *E. coli* strain OP50-1 was initially obtained from the Caenorhabditis Genetics Center, and subsequently stored in small aliquots at -80ºC. This is a streptomycin-resistant strain that was used to bacterial lawns that worms fed on. In other words, the maintenance of selective growth for this bacterial strain requires streptomycin.

I created my own stock of OP50-1 *E. coli* bacteria from the aliquoted frozen stock. Under sterile conditions, an aliquoted stock was thawed and streak into Luria-Bertani (LB) agar plates. The streaked plates were then incubated at 37º for 16h, and then stored at 25ºC. Plates older than a month were discarded and this procedure was repeated as needed.

To create frozen stocks of OP50-1 *E. coli* bacteria, single colonies of *E. coli* OP50-1 were transferred to 15ml tubes containing 5ml of LB broth with 200 mg/ml streptomycin., under sterile conditions. To generate starter cultures, incubate the tubes overnight at 37 ºC and 200 rpm. The starter cultures were then transferred to large flasks (between 1 L to 3 L of volume), that were filled up to 60% by LB broth with 200 mg/ml streptomycin. Then these flasks too were incubated overnight at 37 ºC and 200 rpm. In the following day, concentrated *E. coli* cells were collected by centrifugation at 5000 g, for 10 minutes, at 4 ºC. The bacterial pellets were then resuspended in M9 buffer. After serial dilution the number of bacterial forming units was determined spectrophotometrically, at 600 nm wavelength. Using this estimation, the concentration of the solution of bacteria in M9 buffer was standardized to 10^{10} cells per ml. The standardized solutions constitute my *E. coli* stock and were stored at -80 ºC until needed.

2.3.3 - Buffers and Mediums preparation

The LB broth was created by dissolving 10g of Bacto-tryptone, 5g of Bacto-yeast and 5g of NaCl into 1L of pH neutral MiliQ water and sterilizing the solution by autoclaving.

To prepare the M9 buffer I mixed 3 g of KH_2PO_4, 6 g of Na_2HPO_4, 5 g of NaCl, 1 ml of 1 molar $MgSO_4$ into 1 L of MiliQ water. The final solution was immediately sterilized by autoclaving.

2.3.4 - Preparation of LB agar plates

I prepared fresh LB agar plates each time I wanted to culture OP50-1. This is required because streptomycin loses potency in aged LB agar plates. For each culture plate[a], I mixed 0.75 g of LB miller and 0.75 g of agar into 50 mL of pH neutral MiliQ water. The mixture was then sterilized by autoclaving. After moving to sterile conditions and allowing a slightly cooldown, 50 µL streptomycin was added to the liquid LB agar, at a concentration of 50 mg/ml. While still in liquid form, the mixture was poured into plates, that after solidification and cool-down were ready to be used. If needed, LB agar plates were stored at 4 ºC. Stored plates older than a month were discarded.

2.3.5 - Preparation of NGM petri plates

For each liter of NGM medium the following procedure was followed. In a 2 L clean, autoclaved and oven-dried flask 23 g of agar, 3 g of NaCl and 2.5 g of Bacto-peptone were mixed. To the mixture 1 L of pH neutral MiliQ water was added, and the flask was immediately submitted to autoclaving.

As soon as the autoclaving strep finished, the flask was transferred to a water bath that has been pre-heat to 55 ºC. After the media cools down to 55 ºC, the flask was brought to a hood and the following ingredients were added to the

[a] scale the numbers of the recipe by the intended number of LB agar plates.

warm solution: 1 ml of 1 M $MgSO_4$, 1 ml of 1 M $CaCl_2$, 1 ml of 5 mg/ml cholesterol, 25 ml of 1 M potassium phosphate buffer (pH 6) and, if required, 250 µL of 50 mg/ml FUdR. The NGM medium as then poured into plates. After the NGM in the plates solidified, the plates were covered with a lid, wrapped with 3 layers of parafilm and incubated at 20 ºC for 24 h.

On the following day, the plates were taken back to sterile conditions (hood) and seeded with OP50-1. Once the bacterial lawn dried and thicken, the plates were once again covered with their lid and wrapped in 3 layers of parafilm and incubated for a day at 20 ºC.

After this procedure, the plates are ready to be populated with worms (the ideal scenario) or are stored at 4 ºC for later use. Stored NGM plates older than a month were discarded.

2.3.6 - Maintaining age-synchronized worm populations

All the *C. elegans* experiments required synchronous populations of worms. These were obtained from the eggs extracted from gravid worms by hypochlorite treatment[207]. Using M9 buffer I washed a population of gravid adult animals off my plates. I collected them in 15ml centrifuge tubes. When the resulting solution was not clear (due to the presence of bacterial food), the worms were given enough time to settle to the bottom and the top of the solution was then removed. New M9 was then added, and this procedure was repeated until a clear solution was achieved. To the final volume of 7ml of the clear M9 buffer solution, I added 2ml of household bleach (5% sodium hypochlorite solution) and 1ml of 5N NaOH. The tubes with 10ml of the resulting solution

each, were incubated for up to 10 minutes at room temperature (20ºC). During this period, they were vortexed for 20s every 2 minutes until all the nematodes were lysed (this process usually took 4 to 5 minutes). This was followed by centrifugation at 1500g for 1 minute, at the same temperature. Before the 10 minutes mark, the tubes were brought to a sterile environment, their liquid content discarded, and the solid pellet of concentrated eggs was resuspended with M9 buffer. The tubes were then taken to centrifuge as before, and this washing step was repeated 3 times. In the final resuspension the egg pellets were suspended in just 0.5ml to 1ml volume of M9 and dispersed into new NGM seeded plates. Unless otherwise specified, the eggs were allowed to hatch and the animals to grow.

2.3.7 - Compound preparation

Each compounds stock was kept into the specific conditions required for its stability. For example, resveratrol is light-sensitive and therefore was stored in the dark, covered by aluminum foil and locked inside a box.

For each drug experimental condition, a compound working solution was prepared from the stock solutions at the desired concentration and with the solvent of choice. This working solution was then added to bacteria seeded NGM plates. After the addition of the experimental compound or compound combination, the plates were allowed to dry in the hood. As soon as drying was achieved, the plates were cover with the respective lids and wrapped in 3 layers of parafilm. They were then incubated for 24h at 20 ºC. After incubation the plates were either used or stored at 4 ºC.

2.3.8 - Specifications of 96-plates used in high-throughput screens

After producing warm liquid NGM flasks as normally, the NGM medium was divided equally among pre-heat 50 ml centrifuge tubes set on a heat block, at 55 ºC. According to each experimental condition's specifications, a solution of drug or drugs was added to each of the 50 ml tubes, at the targeted final concentration. The tubes were shook to homogenize the solution, and then closed, only to be opened sequentially as needed. After use they were discarded.

The NGM from the single open tube was transferred into each well of a 96-wells plate, at a volume of 200 µL per well. This transferred was made by using a single channel micropipette. After all the wells have been filled with medium, the 96-well plate was left in the hood until the NGM solidified. After solidification had been achieved, the plates were covered with their lid, wrapped into 3 layers of parafilm. All the wells were checked for bubbles and unlevelled surface and marked for future censoring accordingly. After this step the plates were incubated at 20 ºC for 24 h.

On the following day, the plates were returned to sterile conditions (hood) and seeded with OP50-1. The required total volume of OP50-1 bacteria was divided into 1.5 ml tubes. To each of these tubes, as required, a solution of drug or drugs was added, until the concentration of pharmaceuticals in the bacteria was 3 times[a] the concentration of pharmaceuticals in the NGM, for the respective conditions.

[a] the three times more concentration is to account for the fact that in the next I would add 10 µL of worms in a water solution to the top of the wells. As I would not add drugs in that next step, I add to account for the extra total well volume cause by the worm transferring at the current step.

As before, manually, using a single channel micropipette, 5 µL of bacterial lawn were deposited in the middle of each well. Once the bacterial lawn dried and thicken, the plates were once again covered with their lid and wrapped in 3 layers of parafilm and incubated for a day at 20 ºC.

The next day, the plates were returned to a hood so that the young adult worms could populate the wells or discarded otherwise (no storing).

Stock populations of GRU101 strain worms were grown to maturity and bleached to obtained synchronized eggs. The eggs were transferred to agar growth plates with OP51 bacteria as food. When most of the animals reached adulthood, they were transferred to previously prepared 96-well plates.

The worms were washed from the plate with pH neutral MiliQ water and pipetted into 50 ml tubes. From the center of the filled part of the tubes, 10 µL of water with about half-a-dozen worms (estimated visually) was manually dropped in the center of each well. This liquid transfer was done by using a single channel micropipette. Care was taken to avoid piercing the solid medium. After drying the 96-wells plates were closed with their lid and sealed with 3 layers of parafilm. These populated plates were then incubated at 20 ºC and only leaving the incubation for the daily imaging scanning. Regardless of the location, the plates were kept at no more than 21 ºC and protected from light.

At every 10 days after being populated, 96-wells plates' seal were checked for integrity, and extra layers of parafilm were added as needed.

2.4 - RNA-Seq data

2.4.1 - Acquisition

Stock populations of wild-type N2 were grown to maturity and bleached to obtained synchronized eggs. The eggs were transferred to fresh agar growth plates with OP51 bacteria as food. When the animals reached the L4 stage of development, they were transferred to new plates with added compounds, creating biological samples for each of the following experimental conditions: aspirin, curcumin, captopril, DMSO (the negative control), epigallocathecin, icariin, alpha-ketoglutarate, lipoic acid, lithium, myricetin, n-acetyl cysteine, piceatannol, resveratrol, spermidine, thioflavin-T and ursolic acid. Between three to five 15 cm plates were used per condition. FUdR was added to the NGM media to maintain synchronized populations. On day 2 of adulthood, the worms were washed off the plates into 15 ml tubes and the adults were isolated by floatation. The suspension with adult worms was then washed several times with fresh M9 buffer until a clear solution was obtained. Clean worm pellets were then frozen and later used for RNA extraction. Total RNA was isolated using the Qiagen RNAeasy micro kit (Qiagen, Hilden, Germany) following the manufacturer's protocol and the resulting samples were sent for quality control, library preparation and sequencing. Quadruplicate samples for each condition were sent, and the three replicates with the highest QC scores were sequenced. There were 6 samples that although being among the respective three highest quality samples, did not meet the minimum quality standard and were excluded from sequencing. Due to high number of samples I had to do the above procedure in two batches.

Libraries were sequenced using the Illumina HiSeq4000 sequencing platform (Illumina, San Diego, CA, USA) in a paired-end read approach at a read length of 150 nucleotides. Each of the two batches of samples were sequenced in the three parallel lanes of the HiSeq4000. All the samples belonging to a given experimental condition were sequenced in the same batch but distributed across the three lanes[a]. There was the added consideration that each lane had a sample belonging to the control group, and no lane could have more than 9 samples[b]. The final sequenced sample data was extracted and returned to me in FastQ format.

2.4.2 - Quality-control and Alignment

I started by doing a benchmark quality control check using the FastQC software[208].

The next step consisted in using the FastqPuri software[54] for the removal of low-quality reads from a given pair of pair-ended sequencing files. This filtering procedure was composed by several criteria. First it looks for low quality (below 25) base callings at the beginning and at the end of the read, and iteratively trims them at both ends until the quality is above the threshold. Furthermore, the reads are discarded if there are more than 25% low-quality nucleotides. Additionally,

[a] the distribution across different lanes is an experimental design decision with the goal of better accounting for lane-to-lane variation. Albeit two of the samples required additional sequencing, in a lane that it is not the original, and this created an additional source of nose.

[b] The 9-sample limit was a recommendation from the manufacturer in order to guarantee enough reads per sample.

if a read is less than 31 nucleotides long, it is discarded[a]. The resulting quality filtered sequenced samples files were then used in the next step.

For the alignment step, the reference transcriptome use was the *C. elegans* WBcel235 downloaded from the Ensembl release 100.

As mentioned in the first chapter, the alignment software used was kallisto. The sequenced sample files were processed leveraging their pair-ended nature, with sequence-bias correction. The output of this step was bam format pseudoalignment files.

The pseudoalignment bam format files were process by the picard software[209] in order to remove optical duplicates. In detail, the "MarkDuplicates" function was used with "OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500" and "VALIDATION_STRINGENCY=LENIENT" settings, as recommended for data prevenient from Hiseq4000.

As the output of picard is in the bam format, I then used the samtools software[210] to convert the deduplicated files for fastq format.

Lastly, I realigned the deduplicated fastq data with kallisto, in the same way as before. The quality of the filtering, deduplication and alignment pipeline was assessed by quality checking, each sample file, using the FastQC software. All the treated files were of sufficient quality.

[a] this specific length threshold is more conservative than the default, but I chose it because it is the maximum size of k-mer that the kallisto alignment software can make use of. In any case, it is worth reiterating that the HiSeq4000 generated reads 150 nucleotides long, and, therefore, the length threshold parameter should not be of much importance.

2.4.3 - DEG Analysis

In the R environment (v3.6.1), the biomaRt package (v2.40.0)[211] was used for downloading identifiers of genes and specific genomic features (gene length and GC-content).

The estimated counts produced by kallisto (see previous section) were imported to the R environment using the *tximport* package (v1.12.3)[55], which summarized the transcript-level abundance estimates to gene-level values (and requires the gene mapping identifiers obtained using the biomaRt package).

The *EDASeq* package (v2.18.0)[60] was then used to generate the correction factors to correct for length- and GC-content[a] bias that will be used by DESeq2 for modelling differentially expressed genes. The bias plot of the raw gene counts always displayed a length-bias, with the short genes, comprising a few hundredths of nucleotides long counted less than the rest of the genome. After correction, samples that still displayed significant length-bias were excluded. As a byproduct of correcting for the length-bias using the EDASeq package, GC-bias was also corrected.

The normalization factors were taken to be the *DESeq2 package* (v1.24.0)[63] size factors. This was followed by dispersion estimation.

On the last step, the DESeq2 package was then used to model the gene expression of experimental drugs, taking into consideration batch effects, in one of two ways.

[a] this is why these genomic features annotated were downloaded using the biomRt package.

One of the pipeline variants was based on the Wald test for the significance of coefficients in a negative binomial generalized linear model. The two terms of the model were the experimental condition and the batch that each sample belonged to. After this modelling, the log foldchanges for each experimental condition were obtained by comparing with the control group using the adaptive t prior shrinkage estimator[65]. After obtaining the foldchanges of each experimental group versus the control, the list of DEGs was obtain based on a significance threshold α less than the threshold calculated by applying the heuristic for controlling the optimal false discovery rate. This threshold takes the value of 2^{-r}, where r is the number of replicates[212].

On the other pipeline variant, the likelihood ratio test (chi-square test) between generalized linear models was used instead. This is a statistical test for significance of change in deviance values between a full and reduced model. The full model used had two terms the experimental condition and batch, awhile the reduced model had batch as its single term. The null hypothesis is that there is no significant difference between the two models, i.e., the simpler model is sufficient to explain the variation in gene expression between the samples. After this modelling, the log foldchanges for each experimental condition were obtained by comparing with the control group using the adaptive t prior shrinkage estimator[65]. After obtaining the foldchanges of each experimental group versus the control, the list of DEGs was obtain based on a significance threshold $\alpha < 0.005$.

2.5 - Healthspan/Survival Analysis

2.5.1 - Plotting

Survival plots depict the relationship between the relative survival and time. That is the ratio of living individuals to the initial population for each time point.

The survival plots present in this dissertation were generated using the first or second versions of the OASIS software. Both, OASIS[213] and OASIS 2 (online application for survival analysis 2)[214] are freely available as web applications[a]. The choice between the two is based on my personal judgment of which of them produces more easily discernable plots. I am using this software which outputs survival plots, but because I am measuring healthspan and not lifespan the survival plots are actually plotting the ratio of voluntarily moving animals (at a given time point) by the starting population of moving worms.

2.5.2 - Statistical Testing

As the reader will learn in the upcoming chapters the drug screen assay is based on healthspan (specifically, the presence of movement) and not lifespan. The longitudinal distribution of movement in a worm population as not, to my knowledge, being statistically parameterized. Since such effort is beyond the scope of my project, I will make use of metrics and tests used in survival analysis to characterize my results. This choice can be justified by the fact that

[a] The first version available at https://sbi.postech.ac.kr/oasis/surv/ and the second at https://sbi.postech.ac.kr/oasis2/surv/ .

the lifespan distribution is a bound on the healthspan distribution. That is being alive is a requirement to movement, and no dead animals are capable of moving.

In more detail, I will use two metrics to judge my drug screen results: the restricted mean lifespan and the log-rank statistical test.

The restricted mean lifespan μ_τ was created by [215]Irwin (and also described by Kaplan and Meier[216]) and is estimated with the following formula:

$$\mu_\tau = \int_0^\tau S(t)dt$$

where $S(t)$ is the survival function, t is time, τ is the largest observed time. The Kaplan-Meier method is used to estimate $S(t)$ and it can be calculated in the following way:

$$S(t) = \prod_{j:t_j \leq t_i} (1 - \frac{d_j}{n_j})$$

where j is a time interval, d is the number of deaths and n is the size of the population at risk.

Based on the previously described rationale, I will use the value of the restricted mean as my mean healthspan. What I am doing is using a different definition of "population" and "death". With my population being moving worms and deaths being the loss of moving worms.

To complement the use of the restricted mean, I will use the log-rank test. This non-parametric test takes the whole time period into account and with equal weight to any time period. The null hypothesis is that the there is no difference in the probability of an event at any given time point for the two populations

being compared. In survival analysis, the notion of event corresponds to death, but in my case, it will mean animal immobility. In other words, the log-rank test will be used to compare the relative healthspan throughout experimental time, between two conditions. It is beyond the scope of my project, but the interested reader can consult the relevant literature for the computation and assumptions of the log-rank test[217].

2.5.3 - Definition of synergy

There are several definitions of synergy available in the literature. I chose to define synergy by setting two necessary conditions. First, following the Higher Single Activity (HSA) model of synergy[218a], I require that synergistic healthspan-extending drug pairs must elicit a mean healthspan larger than both of the constituent single drugs or observed in the respective negative controls. Second, all the log-rank tests for healthspan between the combined intervention and the monotherapies and controls conditions must return a p-value less than 0.05. This enforces that the drug pair displays a unique hazard function, and it is inspired by the synergy definition used to describe non-linear interactions in dynamical systems[219,220b]. A more stringent definition of synergy requires that: "The whole be largest than the sum of its parts". Following this definition, aging drug synergy can be defined more narrowly by requesting that the lifespan benefits of a pair of drugs should be statistically significantly larger than the

[a] this model considers drug combinations synergistic if the effect under its treatment is statistically significantly greater than the effect of the more efficacious of the two individual drugs.
[b] in this field, synergy can be defined as a beneficial interaction where the whole is different from the sum of its parts.

sum of lifespan benefits of the two constituent drugs. I will refer to cases where this is true explicitly but will in general use the less stringent definition above.

Chapter 3 - Automated high-throughput healthspan drug screening in *C. elegans*

3.1 - Introduction

After being faced with the limitations of current high-throughput screening (HTS) methods for longevity interventions in *C. elegans* (section 1.2.1), I decided that to eventually generate the scale of data needed to test my hypothesis, I had to create my own HTS method.

There were certain criteria that dictated my initial design choices. I wanted a method of screening that used solid medium, to avoid the main drawback of the HTS microfluidics systems (section 1.2.1). Since, my goal is to create a system from the screening of drug combinations, I need to maximize the number of interventions. This is because even a small library of drugs can be used to generate hundredths of distinct combination pairs interventions (combinatorial explosion). Lifespan assays in petri dishes allow a sufficiently large population of worms to be assayed (which is crucial for statistical power) but cannot scale with the number of conditions. The alternative was to assay animals living in well plates. These allow more interventions while still being easy to manipulate. Of the available sizes, 96-well plates seemed to strike a balance between statistical power and scalability. Moreover, a 96-well plates solid medium HTS had just been published[221]. It did not feature automated scoring, so I could only use some of its specifications (e.g. volume of bacterial food per well) as a guide.

3.2 - Results

3.2.1 - Validation of the GRU101 strain

Automated lifespan studies at the scale required for this project involves large cohort sizes (typically several thousand animals). While commercially available systems for process automation, such as our automated plate pourer, robotic dissection microscope and computer controlled precision stages can be used to automate cohort generation and image acquisition, the analysis of the resulting images remains a rate-limiting step in automated lifespan studies, as well as a major source of error (see second class of systems described in section 1.2.1). Following initial protocol development, I realized that I needed to develop an image processing pipeline to automated scoring live and dead animals, ideally with minimal manual input. A typical lifespan study, in my case comprising approximately 1000 wells (repeats of experimental conditions), would require tracking upwards of 15,000 animals over approximately of 30 days. To analyze such an experiment therefore requires close to half a million life/dead decisions. Considering all available solutions (see section 1.2.1), I decided to take a different approach instead of using the standard N2 wild-type strain by taking advantage of a *C. elegans* strain carrying a transgenic fluorescence marker. The rationale was that by leveraging the GFP signal for tracking and survival scoring, it should be possible to improve signal-to-noise ratio in the detection of dead animals by the use of fluorescence microscopy imaging. I estimated this to be advantageous because it would allow better foreground/background discrimination and remove artefacts e.g. due to misidentification of inanimate features on the plate for worms.

To test my reasoning, I chose the GRU101 *C. elegans* strain. The GRU101 strain is a vector-only control strain that was previously generated in our lab. The strain carries only a *myo-2::yfp* construct (a pharyngeal-specific yellow fluorescence marker) and is used routinely in the lab as non-transgenic vector control[222]. When imaged using fluorescence microscopy with an excitation wavelength tuned specifically to the excitation wavelength of yellow fluorescent protein (YFP) and using an emission filter suitable for GFP/YFP fluorescence, the pharynx of worms can be easily identified. This approach should improve foreground/background separation and allows better discrimination of worms from erroneous signals derived from plate contamination, bubbles or cracks in the plate and other optical artifacts. Previous evidence from out laboratory suggested that the YFP transgene expressed by the GRU101 strain did not impact lifespan, with GRU101 consistently exhibiting a lifespan that was indistinguishable from that of WT[222,223a]. To further assess the healthspan of GRU101 strain, I took the results of an egg-laying assay[224] and movement ABC scoring[225]. In the ABC scoring system: class A worms move spontaneously in a sinusoidal manner; class B worms only move upon prodding; class C only exhibit head or tail movement even after poking; and class D are dead. I considered the ratio of animals in class A or B (capable of movement, hence healthy) to those in class C (considerably restrained movement, therefore unhealthy but alive) as a proxy for the health of the population.

[a] 16 to 18 days of median lifespan in standard conditions.

For the movement ABC scoring, I analyzed unpublished data from the lab (kindly provided by Dr. Ng Li Fang) and the two-sided two-sample Kolmogorov-Smirnov test did not reveal a statistical difference (p-value = 0.59) between the ratio of moving worms (classes A + B) to the total of worms in the plate, between N2 and GRU101 strains. Indicating that healthspan was indistinguishable between GRU101 and N2 WT animals under the conditions used in our laboratory.

For the egg-laying assay, I generated the data myself. Briefly, two age-synchronized populations, one for each strain, of N2 wild-type and GRU101, were bleached through hypochlorite bleaching and maintained on nematode growth medium (NGM) plates (section 2.3.6). On day 3 post-bleaching, young adult animals[a] were picked and transferred, each to a new seeded NGM plate. Every 24 hours, I transferred the adult worms to new seeded NGM plates and then proceeded to count the number of eggs laid in the respective preceding plates. I did this until reproduction ceased in all the animals.

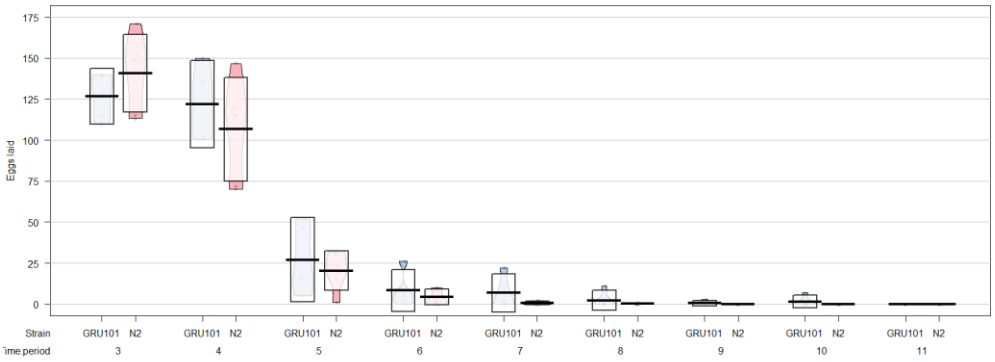

Figure 3.1 – N2 and GRU101 egg-laying assay.
Blue violin plots correspond to the GRU101 strain. Red violin plots correspond to the N2 strain. Each pair of box plots represent the distribution of eggs laid for a given day.

[a] after censoring, there are 6 N2 animals, and 5 GRU101 worms assayed.

I compared the number of eggs laid at every assayed day of the assay. This is a highly granular statistical approach because it takes into consideration both the shape of the egg-laying curve across the reproductive span and the aggregate fecundity, thereby allowing detection of difference in absolute fecundity and peak fecundity as well as potential

As observed in Figure 3.1, there is no statistical differences for any of the days according to the two-sample t-test. In other words, populations of N2 and GRU101 show similar developmental schedule and have comparable reproductive fitness under the conditions used in our lab, at least in terms of viable eggs produced on each day.

These data are consistent with previous experiments in our group that similarly showed that during direct scramble competition between GRU101 and wild-type N2, both strains have similar evolutionary fitness, with both genotypes persisting in direct competition at stable percentages over many generations.

Having convinced myself that the GRU101 strain did not show abnormalities in lifespan, healthspan, life-history traits or evolutionary fitness that might confound its use for lifespan studies, I next set out to confirm that use of this fluorescent strain was indeed advantageous for use in automated lifespan assays in terms of facilitating the discrimination of live individuals from background. Two commonly encountered problems during lifespan studies is plate contamination (growth of fungi) and bubbles or cracks on the plate. Both problems can make it difficult to accurately quantify live worms on affected plates. To test if GRU101 might be used to overcome these challenges, I transferred GRU101 young adults to two solid medium 96-well plates. I then

deliberately exposed one of these plates to unfiltered room air, by leaving it unsealed overnight. In my experience, plates exposed in this way will inevitably get contaminated with fungal growth. Cracking of NGM is due to dehydration and occurs over time or when plates are left unsealed with filtered air (e.g. inside of a horizontal flow hood) cycling over the wells. I deliberately induced cracking in this way, allowing the NGM agar to dehydrate by until serious cracking was observable (Figure 3.2). I then used our robotic microscope to take images of these plates, with and without making use of the fluorescence settings. Visual inspection confirmed that, compared to counting worms under light microscopy alone, using fluorescence microscopy facilitated the detection of animals, reduced ambiguity and revealed missing animals, thereby confirming the usefulness of the GRU101 strain. Figure 3.2, displays examples of wells from this experiment, showing two plates subject to significant contamination and cracking. These images illustrate two facts: even if the agar is heavily contaminated, adult (and even L1 stage) animals can immediately be detected by visual inspection[a] using the fluorescence signal. Furthermore, if by the end of an experiment, the agar gets considerably dry and cracks, burrowed animals can be hard to identify / track using ordinary microscopy, but such animals can still be identified using the fluorescence signal. While problems of this type are not ideal or desired experimental scenarios, they represent worst-case scenarios and cannot always be completely avoided, especially during large cohorts studies that, by necessity, run over several weeks. These data therefore serve as

[a] which for now I will take as a proxy to the easiness of automated worm counting.

stress test to establish the robustness of worm detection by fluorescence microscopy.

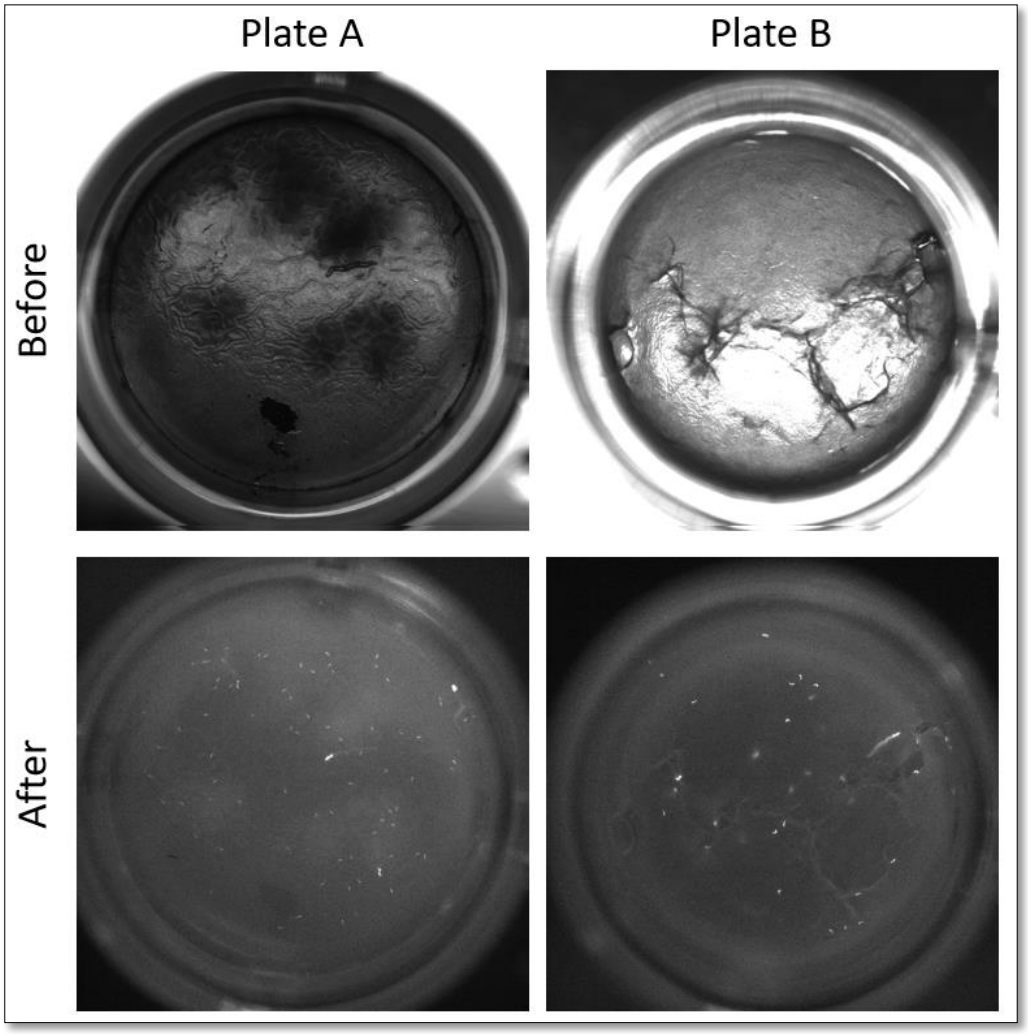

Figure 3.2 – Compromised wells under light versus fluorescence microscopy.
The rows of "Before" and "after" pictures, show the same wells under light and fluorescence microscopy, respectively. The well from Plate A is contaminated, and the well from Plate B has a cracked solid medium.

As described in detail in the next section the worm detection, and therefore scoring will be based on if the worm has moved in two consecutive plate scans. This means that I am actually using a metric of healthspan for drug screening. To assess how far this healthspan-based metric is from measuring lifespan (dead/alive), I killed a population of worms and quantify how long would it take

for the fluorescence marker to fade to baseline levels. In other words, for how long after a worm death can I still detect its cadaver through fluorescence microscopy.

An age-synchronized population GRU101 animals, was bleached through hypochlorite bleaching and maintained on a nematode growth medium (NGM) plate (section 2.3.6). On day 5 post-bleaching, the plate populated by young adult animals was put under fluoresce microscopy. I measured the intensity of fluorescence detected (time point zero in Figure 3.3), and immediately poured 70% ethanol solution to kill the worm population. After 7 minutes elapsed the entire population was death and the fluorescence intensity had already declined (second time point in figure). The worm cadavers lost fluorescence at an exponential rate, and after 6h were visually indistinguishable from the background. In other words, 6h after death worms will be undetectable based on fluorescence microscopy.

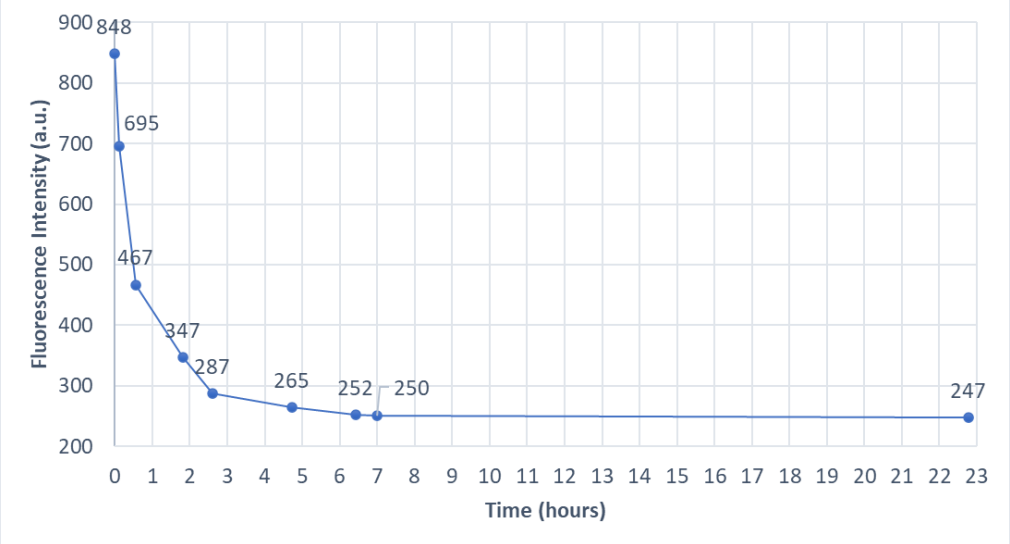

Figure 3.3 – Fluorescence intensity decay after worms death.
Ethanol was poured immediately after time point 0. The value 695 was taken 7 minutes after and it is the time point of a totally dead population. Fluorescence intensity is in arbitrary units.

3.2.2- Automated high-throughput image acquisition and processing

After, generating a population of L4/young adult GRU101 worms (see section 2.3 for detailed protocol), I used liquid transfer to populate the wells of two 96-well plates by suspending worms in M9 buffer (see section 2.3.8 for details).

I generated standardized fluorescence (YFP) images from these plates by taking consecutive plate scans. I aligned these images using automatic labelling, generating paired consecutive frames of each well. These were foundational for my automated detection of actively moving animals.

Essentially, the system determined pixel-flips (changes in location of objects emitting YFP fluorescence) between two images 3.5 minutes apart[a]. Next, I developed a filter able to identify objects with visual features (size, dimensions) consistent with moving worms. The first step that must be undertaken is to filter visual features by size / number of pixels that change (move). The approach I took for this is to define a bounding box. The dimensions of this bounding box must be suitable for capturing a single worm[b]. To accomplish this in a robust manner, I took 12 pictures of each of the 80 wells (of two 96-well plates) that I set up the previous day. These plates serve as ground truth, as I deliberately picked an exact number of worms for each well (and confirmed their survival on the imaging day), by loosely following the expected distribution expected throughout a lifespan screen (Figure 3.4). In more detail, a well starts with some worms and this number will longitudinally decrease until it reaches zero.

[a] The plates were scanned continuously, but because the microscope mounted camera moved and took a single image well-by-well, frames of a given well were taken 3.5 minutes apart.
[b] not the whole body of a worm but only its fluorescent pharynx.

Moreover, even after being considered deprived of alive worms, a well will continue to be imaged, because the imaging is decided at the plate level, in other words, as long as there are any alive worms in the plate, the plate will be scanned in its totality. To increase robustness to experimental conditions and introduce more variability into the ground truth image set, between each scan, I introduced a small variation in focus of the microscope and slightly changed the illumination power. The resulting dataset was a set of 1920 pictures of wells, subject to realistic variation in image quality.

The pictures corresponding to wells that had a single worm were used to find the bounding box dimensions. Bounding boxes are image regions that may contain an object of interest (worm). These are represented two center coordinates and height and width values. In my case, the centers were detected my thresholding and a bounding box of the chosen dimensions was established around it.

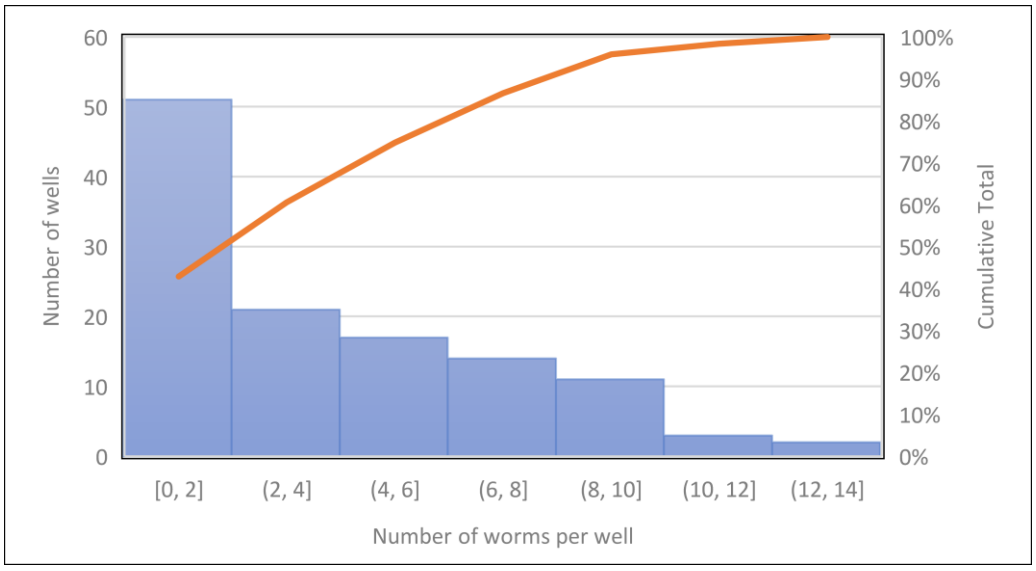

Figure 3.4 – Histogram of wells based on their amount of worms.
It depicts the distribution of the number of wells with a given amount of worms.

The single worms were detected by using the imager R package[a]. In detail, after automatic intensity-based thresholding all images[b], all possible pairs of tile scans of a given well were subtracted to remove static features, that is, assuming that worms are moving they will constitute the only differentially detected pixels (Figure 3.5).

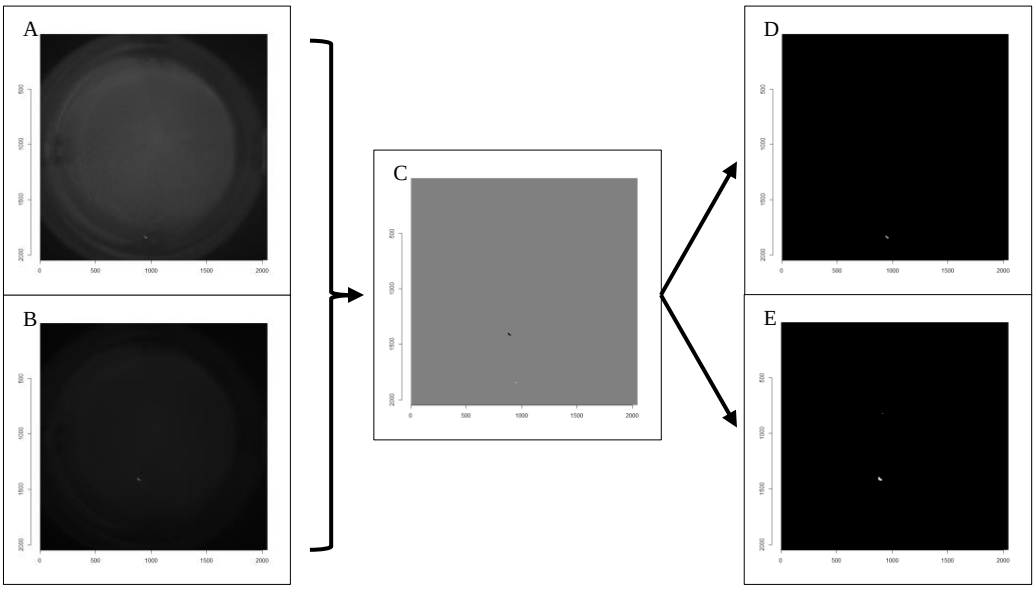

Figure 3.5 – Frames subtraction and thresholding.
A and B are two subsequent frames of a well. The pixel intensity values of B are subtracted from A, to generate the frame C. After thresholding image C, frames D and E can be obtained displaying the moving worm and nothing else.

The obtained images were then cleaned and the rectangular dimensions of the bounding boxes that incapsulated each worm found. I defined my square bounding box based on the median area of these 1182 boxes, which corresponds to a worm size of approximately 35 pixels long. This size seems to work in practice, based on visual confirmation that regardless of the direction and

[a] https://cran.r-project.org/package=imager
[b] thresholding consists in setting all values below a threshold to 0, and all above it to 1. The auto thresholding used was adjusted by 2.0.

position of the worms, the full fluorescence marked body area (pharynx) is contained within the bounded area (Figure 3.6).

Visual inspection suggested that using this traditional image processing algorithm, I was able to detect many of worms on most plates correctly. However, it was my hypothesis that the worm detection pipeline could benefit from a computer vision classification algorithm (deep learning[a]) on top of the filtering step. To test this each of the two plates were considered training and test plates. The filtering step, which I will delve deeper in the next paragraph, was applied to the plates and the produced 35x35 pixels images were put on the corresponding plate data subset and furthered split into "positive" and "negative" accordingly to if the filtering consider that the images were of alive worms or not, respectively. In this way I effectively generated a labelled training and testing set for the supervised deep learning step.

The complete filtering procedure consists in thresholding[b] and cleaning the tile scan images, much like before when I was choosing the bounding box dimensions, with the added steps of: transforming the differential pixels into pixel sets[c]; finding the center of each pixel set and considering it the center of a bounding box that will be a saved filtered image. This image is considered to

[a] Deep learning algorithms in particular because they can even outperform humans in computer vision tasks similar to mine (e.g. [309]).

[b] this time the threshold was set to the 99.87 percentile of pixel intensity based on an estimation that even if the image had 18 worms, all of them would be detected by this filtering:

$$\frac{bounding\ box\ \times number\ of\ worms \times area\ of worm\ body}{well\ dimensions} = \frac{35 \times 35 \times 18 \times \frac{1}{4}}{2024 \times 2024} \sim 99.87\ percentile$$

[c] pixels that are adjacent are considered connected and therefore part of the same pixel set.

have a worm ("positive") if the pixel set on which it was centered is one of the largest n^{th} pixel sets, with n being the true number of worms in the well; and it will be classified into the "negative" subset otherwise.

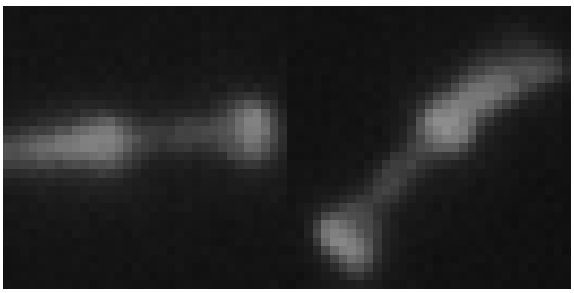

Figure 3.6 - Samples of bounding boxes of single worms.
Independently of the orientation or bend, the full pharynx of the worm fits in a bounding box of the used dimensions.

The following subsets were automatically created: 2835 positive training samples; 346 negative training samples; 2669 positive testing samples; 674 negative testing samples. To test the accuracy of this automated approach, and, therefore, the need further processing by a machine learning classifier, I manually curated the samples, checking each candidate to determine if it indeed showed a worm. Having manually curated these datasets, I was able to determine that using only the parameter based (traditional) image processing algorithm, would result in the following error rate for the training set: 24% false negative (objects falsely rejected as worm) and 41% false positive (object falsely classified as worms). Similarly, for the testing set, I identified 15% false negative and 42% false positives. This means that, if the parameter-based filtering process was my only data analysis step, the pipeline accuracy[a] would

[a] $accuracy = \frac{correctly\ classified\ samples}{total\ samples}$

be approximately 62%. That is, I would accurately classify bounding boxes has having a worm or not 62% of the times.

For training and testing a deep learning algorithm, I imported the respective two datasets of images to the R environment. I then used the *keras* (v2.2.4.1.9001) R package[226] to do data augmentation. The data augmentation was based on rotating the original images and flipped versions by right angles. In this way not only did my datasets increased 8 times in their number of images, training on the augmented dataset should provide some translational-invariance[227]. In other words, the classification of the algorithm will be unbiased to the orientation of the worms. I then trained a convolutional neural network[228] for the binary classification task of detection if a given image has a worm or not. The final model achieved 0.93 accuracy on the augmented testing dataset[a].

After validating my automated lifespan assay and image processing pipeline, I next carried out a validation study using a drug with known lifespan effect.

I tested two necessary criteria: (1) that under my experimental conditions (solid medium 96-well plates, fluorescence microscopy, automated scoring) a known lifespan-extending drug would extend lifespan compared to the control group; and (2) that survival curves resulting from automated scoring were not statistically different from survival curves from manually counted scores.

According to the protocol previously described, I prepared two 96-well plates with rapamycin (experimental condition) and DMSO (control condition). The rapamycin experimental condition is considered a positive control, as it has

[a] Testing various models of deep learning is beyond the scope of this project, and it here deep learning was chosen because even an unoptimized algorithm already achieves sufficient performance.

worked in our hands before, and the DMSO is its respective negative control group. Already with the intention of in later screen having DMSO being the default organic solvent most often used in lifespan studies of drugs that are relatively insoluble in water, I manually scored the lifespan assay pictures of these two conditions. The image acquisition part was conducted as previously described, and I acquired two consecutive plate scans almost[a] daily from day 1 of adulthood, to day 55. A short custom R script was created to iterate across all wells of each condition and on each day for which there was data available, display a picture of each of the two pictures (per time point) and ask how many worms I could observe. The survival curve was then obtained by aggregating all the wells across time. As seen in Figure 3.7, rapamycin can extend lifespan under my experimental conditions (log-rank test p-value= 0.0064). The magnitude of lifespan extension based on the restricted mean was approximately 18%.

[a] My goal was to take daily scans, but due to the Covid-19 pandemic I was quarantine in the middle of the experiment. This is what produced the missing days in the survival curves.

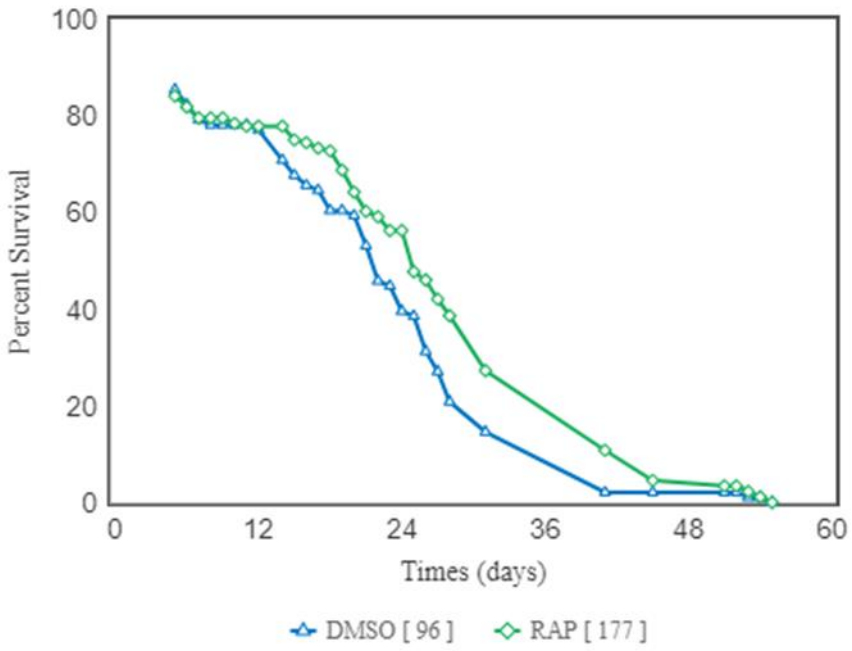

Figure 3.7 – Survival plot of rapamycin and DMSO control.
Rap = rapamycin. Rapamycin concentration at 0.1 mM. DMSO was standardized in both conditions to 0.1% v/v. Values inside [] indicate the sample size.

To test the second criterion, I wrote an R script that queries a human operator for input on how many live worms there are present. This is done following the image processing step in which there is a subtraction of pixel intensity values of two consecutive plate scans, asking the human operator to evaluate the pixel flip mask and determine how many live worms are present. The obtained healthspan curves were not statistically different (log-rank test p-value= 0.8254) and the healthspan restricted mean difference was less than 9% (Figure 3.8). This means that the final image processing pipeline fully agrees (to within statistical significance) with the human operator.

I use the log-rank test because that is the standard statistical test for survival curve differences in lifespan assays, and, mainly, because “healthspan survival curves” have not been parameterized. In other words, it is not being derived the

statistical distribution of the population cumulative loss of healthspan (based on voluntary mobility) throughout *C. elegans* lifespan.

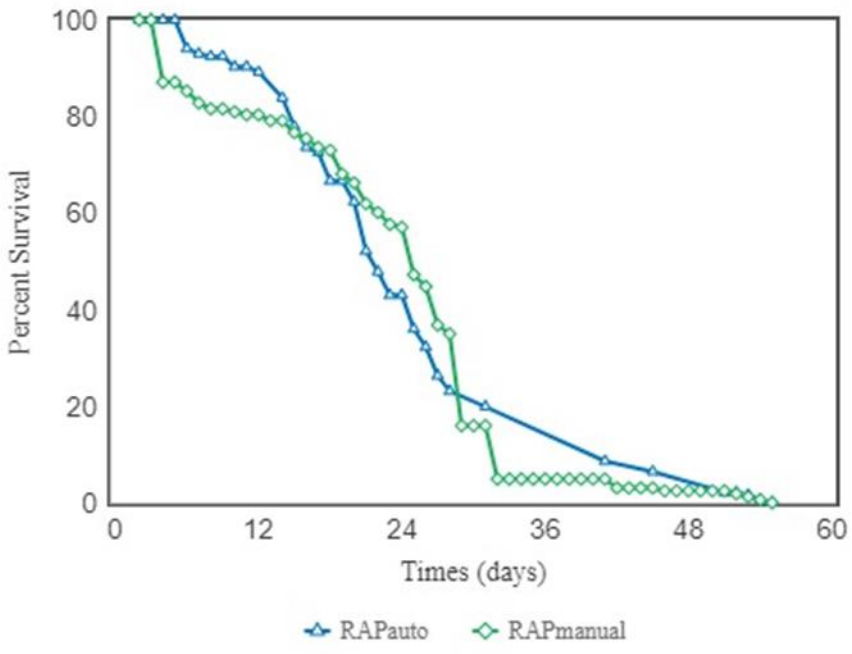

Figure 3.8 – Survival plots of rapamycin scored automatically and manually. This intervention consisted in rapamycin concentrated at 0.1 mM and diluted in 0.1% DMSO. RAPauto = rapamycin intervention automatically scored. RAPmanual = rapamycin intervention manually scored.

3.2.3- Preliminary drug screening

After building and initial validating the lifespan system, workflow and image-processing pipeline, I validated the overall approach by carrying out a small scale drug screen. For my initial screen, I focused on a set of diverse drugs, requiring different solvents, and also attempted to test some drug pairs. While in the context of my workflow this is a comparatively small-scale screen, this trial involves 23 experimental conditions overall. By comparison, our previously largest single experiment comprised fewer than 10 separate

conditions[70], meaning that even the validation screen would certainly be considered a large-scale experiment if it had been conducted manually.

Using the DrugAge database, I selected a set of drugs that had previously been reported to extend lifespan. I chose my set of test drugs from DrugAge, based on reported effect size, diversity of mode of action and solubility (with the aim of testing the robustness of my system relative to different solvent systems). Specifically, I selected: alpha-ketoglutarate (α-KG; section 2.2.2), aspirin[87], N-acetyl-L-cysteine (NAC; section 2.2.10), nicotinamide adenine dinucleotide (NAD)[229], doxycycline[158], ethosuximide[230], lipoic acid (section 2.2.8), lithium chloride (section 2.2.9), oleanolic acid[231], propyl gallate[120], rapamycin (considered the positive control; section 2.2.15), resveratrol (section 2.2.16) and thioflavin-T (section 2.2.19). For each drug, I used the specific solvent for which the largest lifespan extension had been reported in the literature. I tested three differ solvents commonly used in the literature – water, ethanol and dimethyl sulfoxide (DMSO). For each solvent I standardized concentration and included a solvent-only negative control in my screen. Because the eventual goal was to validate a system for the identification of drug interactions, I also tested a selected number of candidate drug combinations. In addition, I tested the following set of novel combinations for potential beneficial interactions: α-KG and lithium; NAC and lithium; NAC and NAD; RAP and lithium; RAP and NAC; resveratrol and NAC; resveratrol and RAP. These pairs were selected randomly.

The experimental setup was as previously described, with the drug treatments were applied from the onset of adulthood onwards. The plates were scanned until 55 days have passed since the transfer of worms.

Regarding lifespan-extending drugs which were applied as an aqueous solution, all were detected to extend healthspan, when considering a multi-hypothesis adjusted log-rank test significance threshold of less than 0.05 (Figure 3.9). In more detail and taking into consideration the log-rank test between experimental conditions too, the drugs can be separated into 3 classes according to their effect in healthspan: alpha-ketoglutarate (1 mM), doxycycline (30 µg/ml) and NAC (5 mM) extend healthspan slightly by approximately 14%, 11% and 9%, respectively; NAD (0.1 mM) displays larger effect and extends the healthspan restricted mean by approximately 26%; lithium (10 mM) is the drug that among these extends healthspan the most, by approximately 35%.

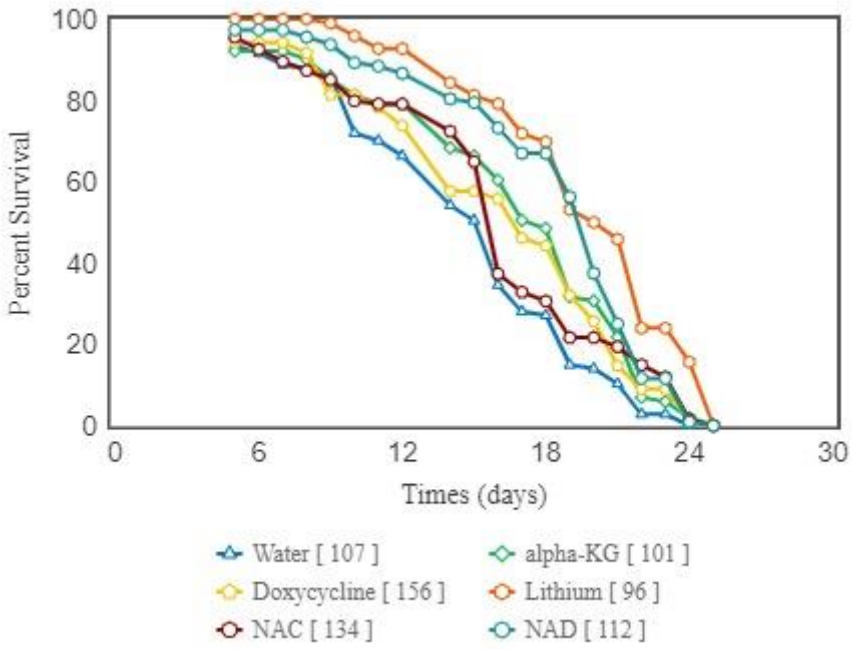

Figure 3.9 – Healthspan of water-soluble drugs.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. alpha-KG = alpha-ketoglutarate. Values inside [] indicate the sample size.

For the drugs that were dissolved in ethanol, I surprisingly found that none elicited a statistically significant lifespan effect (Figure 3.10). This was despite the fact that the control group (only ethanol at 100 mM) did not display an abnormally high restricted mean healthspan (approximately 16 days), so this cannot be the reason for the lack of reproducibility of these drugs. It might be the case that these drugs extend lifespan without a concomitant increase in healthspan.

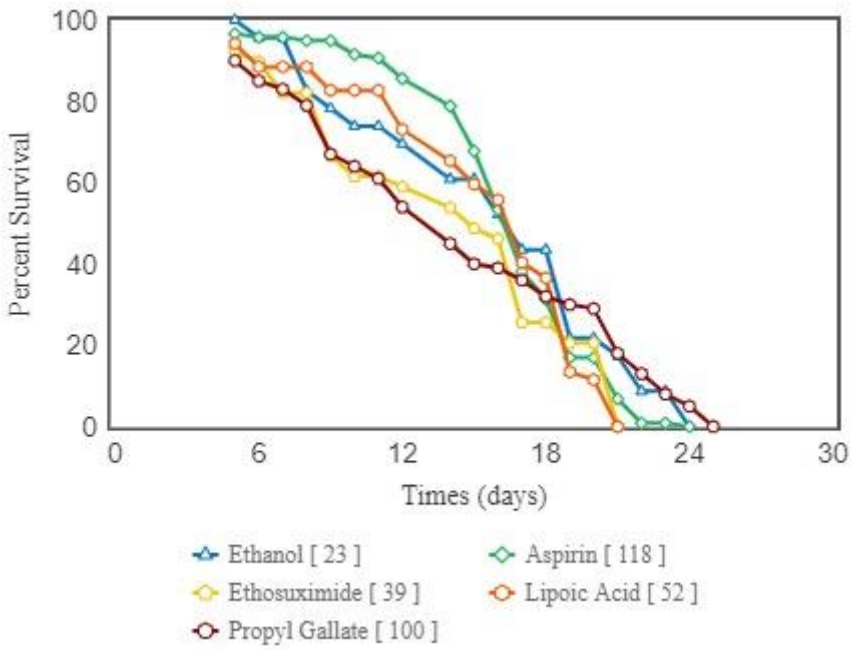

Figure 3.10 – Healthspan of ethanol-soluble drugs.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. The following final drug concentrations were used: aspirin at 0.5 mM, ethosuximide at 4 mg/ml, lipoic acid at 0.024 mM, and propyl gallate at 1.3 mM. Values inside [] indicate the sample size.

Finally, regarding the subset of my drug library for which the solvent was 0.1% DMSO (Figure 3.11), rapamycin (0.1 mM) and oleanolic acid (0.3 mM) extended the restricted mean healthspan slightly by approximately 16% and 14%. The approximately 5% healthspan extension by resveratrol (0.05 mM) did

not reach statistical significance after adjusting the p-value of the log-rank test for multiple hypothesis[a]. Too much surprise, thioflavin-T (0.05 mM) treatment displayed a decreased "healthspan survival curve" relative to the DMSO negative control group. The restricted mean healthspan was shortened by approximately 26%. This is notable considering that thioflavin-T was the most robust drug assayed in the Caenorhabditis Intervention Testing Program (CITP)[83]. After examined the raw pictures and their processed counterparts, I found that this effect was due to erroneous automated scoring, that originates from the fact that thioflavin-T emits fluorescence on its own under my experimental conditions. In other words, the separation of worm pharynxes from the background was compromised because the entire surface of the medium was being mapped by high intensity pixel values. I solved this issue in my next (and final) screen by adjusting the intensity of the light emitting source, as a lower light intensity leads to more selective fluorescence by my fluorescence marker protein (see sub-chapter 5.2).

[a] The lack of healthspan effect of resveratrol treatment will be reproduced in my final screen.

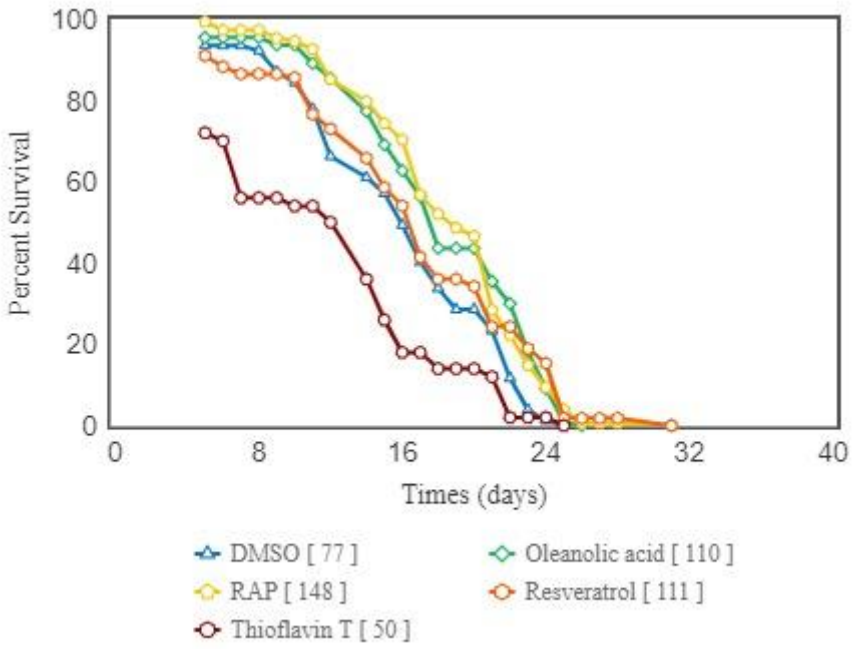

Figure 3.11 – Healthspan of DMSO-soluble drugs.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. RAP = rapamycin. Values inside [] indicate the sample size.

After contemplating the results of the single drug interventions present in this small-scale validation trial, and all the previous evidence, I consider my methodology capable of detecting meaningful changes in healthspan and, therefore, successfully validated.

As mentioned above, I took advantage of the easily scalable nature of my screening methodology to, even in this preliminary trial, screen some drug pairs. The selection of drug combinations was taken lightly and based on my knowledge of the literature. If my drug screening methodology works, then the results regarding the frequency of synergistic healthspan-extending drug pairs found, can be taken as the relative expectation of synergistic pairs in a human

biased and non-exhaustive drug screen[a]. For example, I suspected that lithium would be one of the best performing drugs, based on its reported healthspan-extending effect (in an Alzheimer's disease worm strain)[126] and therefore paired it several times.

The addition of alpha-ketoglutarate to lithium was toxic relatively to lithium monotherapy (Figure 3.12), in a statistically significant manner, with multi-comparison adjusted log-rank p-value$= 0.0229$. The combination intervention group displayed a healthspan restricted mean extension of approximately 27% relative to the control group. This effect is of intermediate magnitude to the ones exhibited by the single drug conditions.

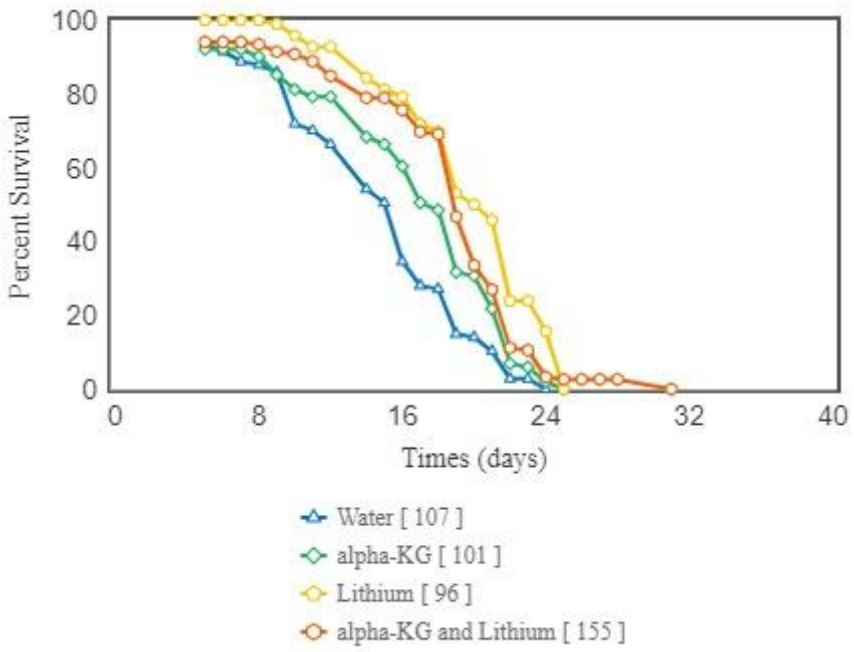

Figure 3.12 – Healthspan of alpha-ketoglutarate and/or lithium.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. alpha-KG = alpha-ketoglutarate. Values inside [] indicate the sample size.

[a] The reader has not to worry about what would be the frequency in a screen with a deeply curated drug selection and that assays drug combinations exhaustively, because those results will be revealed in an upcoming chapter.

The combination of lithium and NAC is also toxic compared to the lithium-only condition (Figure 3.13), in a statistically significant manner, with multi-comparison adjusted log-rank p-value= 0.0001. The combination intervention group displayed a healthspan restricted mean extension of approximately 19% relative to the control group. This effect is of intermediate magnitude to the ones exhibited by the single drug conditions.

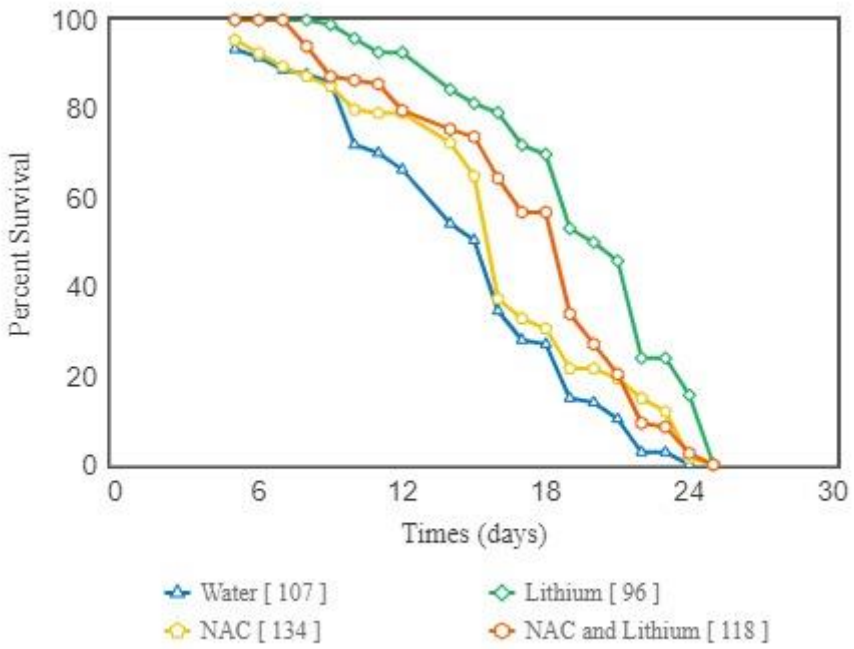

Figure 3.13 – Healthspan of NAC and/or lithium.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. Values inside [] indicate the sample size.

The final water-soluble drug pair that I assayed was NAD and NAC. The combination of these drugs results in a therapy with intermediate efficacy compared to the single drug interventions (Figure 3.14). The NAC and NAD statistically significantly extended the restricted mean healthspan by

approximately 10% compared to the water control group and was not significantly different from any of the single drug interventions (as quantified by the adjusted p-value of the log-rank test between these conditions).

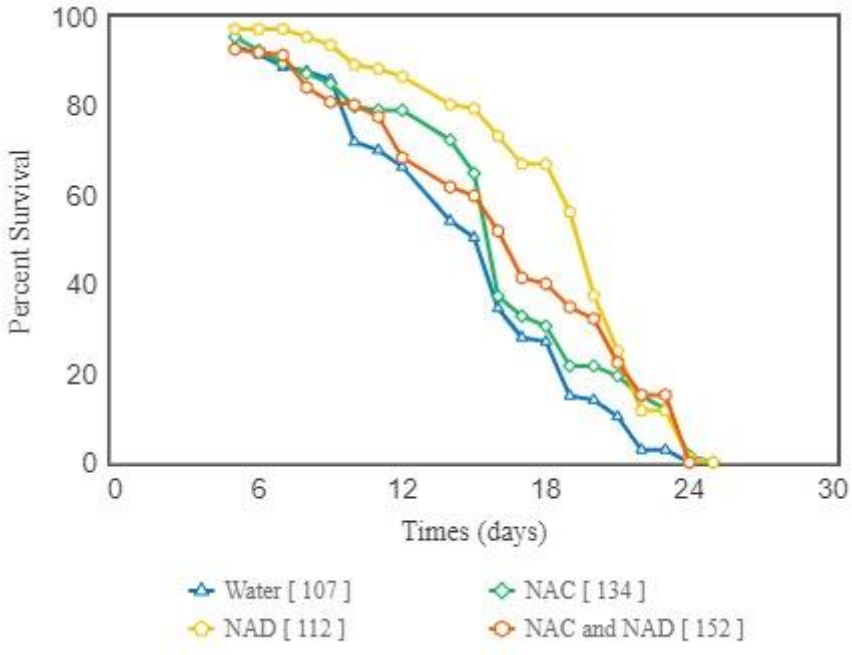

Figure 3.14 – Healthspan of NAC and/or NAD.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. Values inside [] indicate the sample size.

Another drug that I paired with lithium was rapamycin. In this case, the combination therapy is not statistically distinct from any of the monotherapies, neither are the monotherapies among themselves (Figure 3.15). The magnitude of extension of the mean healthspan versus the DMSO treatment is approximately 16%, 24% and 18%; by the rapamycin, lithium, and combination (rapamycin and lithium) experimental groups, respectively[a]. The healthspan

[a] Please notice that lithium was dissolved in water. It is being compared to the DMSO group only because the rapamycin and lithium condition was dissolved in this solvent.

extension of these 3 groups is statistically significant even after adjusting for multiple comparison (see appendix for the precise values).

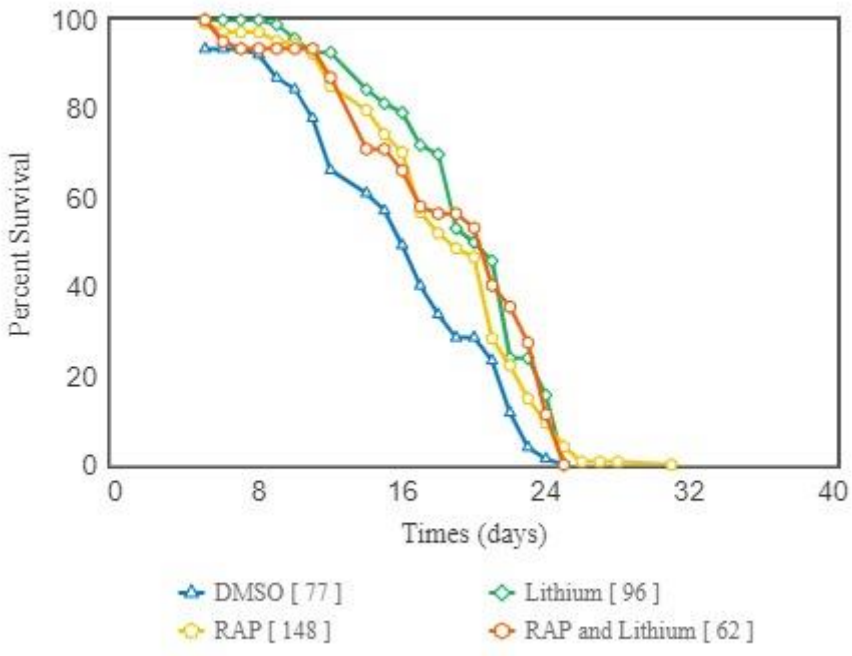

Figure 3.15 – Healthspan of Rapamycin and/or lithium.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. RAP = rapamycin. Values inside [] indicate the sample size.

Finally, I also tested all the possible pairings of resveratrol, NAC and rapamycin. Although, NAC is water-soluble, due to being paired with DMSO-soluble drugs[a], the drug pairs were dissolved in DMSO. Figure 3.16 depicts the results of all these conditions[b]. Of the combination interventions, only rapamycin plus resveratrol extend healthspan (adjusted log-rank p-value= 7×10^{-6}) compared to the DMSO control group. The healthspan restricted

[a] more precisely with drugs that were reported to achieve their maximum lifespan extension when diluted in DMSO.
[b] Figure 3.16 is stylishly dissimilar to the previous survival curve plots because, due to the elevated number of conditions being depicted simultaneously, I used the first version of the OASIS graphing software[213], which is freely available at https://sbi.postech.ac.kr/oasis/surv/ .

mean for the rapamycin plus resveratrol intervention group, versus the DMSO-treated worms, is approximately 22%. This is larger than the approximately 16% increase elicited by rapamycin (which is the best of this two drugs), but the log-rank test between the pair and rapamycin only conditions does not reveal a statistically significant difference.

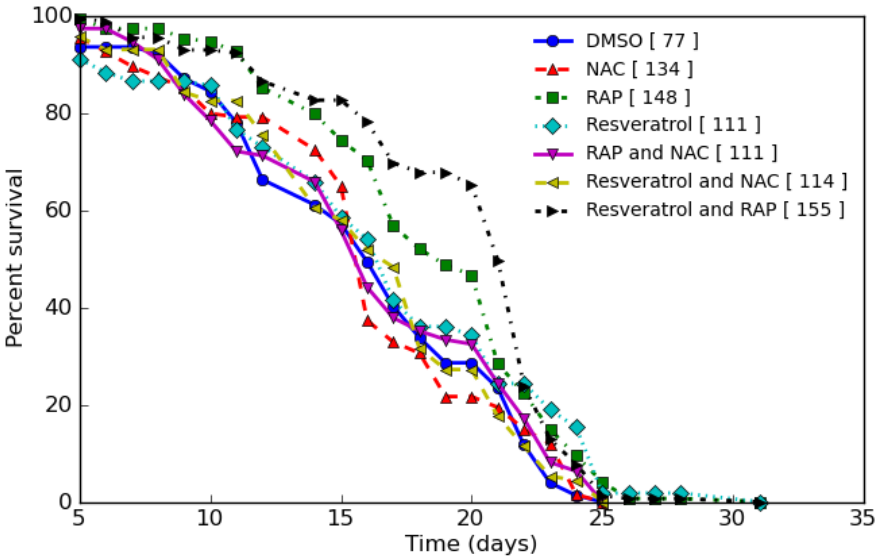

Figure 3.16 – Healthspan of RAP, NAC, resveratrol, and their pairs.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. RAP = rapamycin. Values inside [] indicate the sample size.

3.3 - Discussion

In this chapter, I created and validated a healthspan-based drug screen that is easily scalable and largely automated. Even though I ended up conducting a preliminary small-scale screen for validation purposes, this already constitutes, to the best of my knowledge, the second largest combinatorial anti-aging drug screen in *C. elegans*. The largest belongs to our previously published work[70], but it was highly laborious. If considering healthspan as being the main measure of interest, then this is indeed the largest screen in the literature.

If one considers this screen, with all its limitations, to represent the frequency of drug synergies that extend healthspan; then the estimation could not be farther from some authors' expectation[37]. Namely, among the 7 drug pairs tested I found no drug synergies. It is important to reiterate that I am effectively measuring healthspan, so there is a possibility that things could be different regarding lifespan.

Chapter 4 - Transcriptomics Data Analysis

4.1 - Introduction

To test my hypothesis that pairings of dissimilar drugs are more likely to be synergistic I would need to know the mode of action of each drug in my library. Almost all the drugs have several proposed mechanisms of action (subchapter 2.2) and do not seem to be comparable based on their known genetic dependencies (Table 2.1). Therefore, I undertook the hypothesis-free and systematic identification of their differentially expressed genes (DEGs).

Filtering the resulting DEGs for known aging-genes, it is also a way to test the linearity assumption (see section 1.5.2).

4.2 - Results

4.2.1 - RNA-Seq Data Analysis

Worm cohorts were treated with drugs from young adult stage (3 days after hatching) before harvesting on day 5 of life (day 2 of adulthood)[a]. Following RNA extraction, reverse transcription, sequencing (see section 2.4.1), quality control / filtering (QC) and alignment (see section 2.4.2), abundance data for 40 samples were imported to the R language environment. Seven samples had to be rejected either during the initial pre-sequencing QC stage or during the RNA-Seq data analysis step (due to severe gene-length bias). Overall, I obtained gene

[a] This choice was made also in our previous work, and it was repeated here to minimize protocol differences. This is crucial, as later (Chapter 6) I will be joining these two RNA-Seq datasets together.

expression data for 20 drug treatment conditions plus the control conditions, with an average of 2.86 repeats per condition.

The differential expressed gene (DEG) analysis followed the protocol established in section 2.4.3. In this specific analysis I used the variant of the protocol that is based on the Wald test for the significance of coefficients in a negative binomial generalized linear model. The control condition was 0.1% DMSO solvent, and all the drug conditions were standardized to this. After obtaining the foldchanges of each experimental group (treatment condition) versus the control, the list of DEGs was obtain based on a significance threshold of $\alpha < 0.138$. This unusual significance value is based on the heuristic by Lamarre, S. *et al.* for controlling the optimal false discovery rate of 2^{-r}, where r is the number of replicates[212].

4.2.2- Enrichment Analysis on GenAge

After identifying the DEGs, I was faced with the challenge of how to extract biological insights from this information. A major challenge in this context is dimensionality reduction. Approximately half of my drugs resulted in gene expression changes to more than 1000 genes (they have over 1000 DEGs – see Table 4.1). This is expected for three reasons. Aging is a systemic, multifactorial process, and by pre-selecting drugs that are known to influence aging, I am implicitly selecting for drugs likely to have a broad effect on the transcriptome. Secondly, in pharmacological interventions, as opposed to genetic interventions that target a specific gene or gene-regulatory network, off-target effects tend to

be more pervasive. Thirdly, previous experimental evidence also shows that the most efficacious drugs often change several thousands of genes[232].

When extracting and prioritizing biological knowledge from a large list of DEGs in a data-driven way, an initial approach is typically gene set enrichment analysis (GSEA). Gene Set Enrichment Analysis was actually the name given to the first of such methods[233], but its modern usage refers to the category of methods that search a set of genes (e.g. DEGs) for significantly over-represented subsets or categories of genes. The important concept is the subset of genes, rather than individual genes, being the unit of interest. The rationale of changing the level of analysis is that there should be a dimensionality reduction and that gene sets are more interpretable than individual genes. The sets of genes are annotated in databases, and group gene sets that belong to the same biological pathway (available from databases such as KEGG[234] and Reactome[235]), molecular function or cellular structure (compiled in the Gene Ontology[236]). Depending on the nature of the gene annotations available and the goal of the scientific inquiry, different methods can be used[a]. As there are no detailed aging-specific annotations and most *C. elegans* genes are automatically curated by extrapolation from human and mice data, I employed a different methodology.

I reduced my large number of DEGs to the lifespan-extending *C. elegans* genes present in the GenAge database[237] of ageing-related genes. This filtering process has the advantage that for each gene in GenAge, there is published data showing that the gene is causally involved in aging and lifespan determination.

[a] for an up-to-date comparison see [310].

GenAge contains experimental data regarding pro- and anti- longevity effects observed following genetic interventions (knock down, mutations, over-expression) targeting each gene. The GenAge filtering scheme allows sorting of observed gene expression changes into those predicted to increase lifespan and those predicted to shorten lifespan, and this can be used to further classify DEGs with respect to their predicted overall lifespan effects.

In more detail, GenAge classifies the lifespan-extending genes in two categories: genes that have their gene expression level correlated with their lifespan effect are referred to as pro-longevity genes; genes that display a negative correlation between their transcriptional abundance and elicited longevity are called anti-longevity genes. Pro-longevity genes must be overexpressed to be lifespan-extending, and, in opposition, anti-longevity genes must be down-regulated.

Theoretically, consideration of direction of the fold changes is an improvement over the standard GSEA approaches. The GenAge approach is a more nuanced, because knowing only if a gene is present in a gene set of interest is not sufficient to predict functional impact. A toy example: if DAF-2 was the most significant DEG in one of my lifespan-extending drug conditions, it would be present in GenAge and likely be considered the strongest candidate for the mode of action of that drug. But it could be the case that the drug led to a large overexpression of DAF-2, which would convey the exactly opposite biological insight. DAF-2 cannot be the mode of action of my drug because DAF-2, being an anti-longevity gene, only leads to lifespan-extension if its expression is down-regulated. Since it is over-expressed in this drug, and this drug extends lifespan, DAF-2 modulation cannot be the direct mode of action of my drug.

As a result of the filtering process, I collected all the GenAge lifespan-extending DEGs likely to directly contribute to the mode of action for each of my drugs. Henceforth, throughout this dissertation, I will refer to the set of lifespan-extending DEGs that are overexpressed and pro-longevity or down-regulated and anti-longevity as DEGs in the "**correct direction**" or "**correctly modulated**". The number of correctly modulated DEGs for each of my drugs is shown in the "GenAge DEGs" column of Table 4.1.

Drug	Total DEGs	GenAge DEGs	Unique GenAge DEGs	Drug Classification
Thioflavin-T	10188	181	100	unique
alpha-Ketoglutarate	6888	105	11	unique
NAC	6248	89	13	unique
Lithium	3630	82	13	unique
EGCG	3188	75	20	unique
Spermidine	1240	24	13	unique
Captopril	1315	24	1	unique
Icariin	196	8	2	unique
Myricetin	78	3	0	group-dominated
Lipoic acid	26	3	0	dominated
Piceatannol	13	3	0	dominated
Ursolic acid	57	2	0	dominated
Resveratrol	30	2	0	dominated
Aspirin	23	2	0	dominated
Curcumin	17	2	0	dominated

Table 4.1 – Number of DEGs and Dominance-based classification of drugs.
Total DEGs indicate the total number of DEGs for a drug intervention. GenAge DEGs are the total correctly modulated GenAge DEGs. Unique GenAge DEGs are specifically target by a drug intervention. Drug Classification corresponds to the class of drug according to the dominance framework (section 4.2.4).

4.2.3 - Common Lifespan-extending Drug Targets

Of all the candidate lifespan-extending GenAge genes, only two are common to more than 5 drugs. These are the stress-induced chaperone *hsp-16.48* and *hsp-16.49*. Surprisingly, these two genes up-regulated by 14 of the 15 drugs I tested, with only myricetin failing to do so. Furthermore, these two genes are the entire

set of lifespan-extending GenAge DEGs correctly modulated by four of the drugs I tested (aspirin, curcumin, resveratrol and ursolic acid).

4.2.4- Drug Dominance

The fact that the two of the genes are present in the DEGs set of 14 drugs out of 15 drugs and constitute the full set of 4 of them led me formulate the concept of "**drug dominance**". In decision theory, a random variable is said to be state-wise dominant over another random variable if it gives at least as good[a] a result in every state, and a strictly better result in at least one state[238]. I imported this notion into my methodology in the following way: Drug A is said to be dominant over drug B, if the set of GenAge lifespan-extending DEGs correctly modulated by drug B is a strict subset of the ones correctly modulated by drug A. Putting into a more practically way, drug A is dominant over drug B, if it targets all of drug B's known (correct) lifespan-extending genes, plus at least one more.

The concept of drug dominance is useful as an heuristic by which to categorize candidate drugs (if drug A is dominant over drug B, it might be useful to prioritize it over drug B, because it has more therapeutic potential) and to design drug combinations (assuming equal side-effects, if drug A is dominant over B, it can replace drug B in any combination, with additional therapeutic potential[b]).

We can already infer that aspirin, curcumin, resveratrol and ursolic acid are dominated drugs. It is worth notice that due to the restricted set of candidate

[a] "good" is context specific. For example, in a gamble it could be defined as a larger expected reward.
[b] that is, it makes drug B obsolete.

genes and the simplification of not considering the magnitude of gene expression changes, my classification approach does not perfectly categorize drugs. Case in point, theses 4 drugs all have the same lifespan-extending GenAge DEGs in the correct direction, but while resveratrol and ursolic acid did not extend healthspan in my high-throughput assay (see sub-chapter 1275.2), aspirin and captopril prolonged healthspan significantly by the same magnitude (23%), suggesting that either the latter have additional modes of action (benefit) that are not captured in GenAge or that the former are subject to detrimental changes offsetting beneficial ones.

In addition to the two mentioned genes, lipoic acid correctly targets one additional GenAge gene (*Y54G9A.7*) but is dominated by lithium. Piceatannol also targets one more gene (*lys-10*), and it is dominated by alpha-ketoglutarate and captopril.

Myricetin treatment only regulates 3 lifespan extending GenAge DEGs in the correct direction, and although none of them are unique to it, it is unique in that no other drug modulates all three of these genes. Therefore, by the previous definition, it cannot be dominated by any other of my drugs. Assuming linear superposition (see section 1.5.2), a drug combination that together targets all three of the genes correctly changed by myricetin can be designed[a] and at the treatment level it would dominate the myricetin treatment. As consequence, myricetin is a case of what I coin "**group-dominated**". The gene set modulated by myricetin cannot be jointly modulated by any other drug in the set, but there

[a] conditional to the assumption that the gene expression changes elicit by a drug combination is a linear combination of the gene expression changes elicited by its constituent drugs.

are other drugs in the set that can modulate each of them independently. This additional definition, of group-dominated drug is important because it reveals that drugs in this class leverage common pathways in a unique way. From an information theoretic perspective, drugs in this class might help to detect high-level gene-gene interactions because they allow to model the therapeutic potential from a common set of genes that it is jointly unique. In my case, myricetin treatment extended healthspan by 25%, so it might be the case that myricetin's 3 DEGs hits act synergistically.

Finally, from a therapeutic perspective, it might be better to not replace a group-dominated drug by a set of drugs that jointly dominates it, because multiple drugs might increase the risk of off-target and side-effects, or at the very least, increase the possibility of interacting in unpredictable, non-linear ways.

In an increasing order of correctly modulated lifespan extending GenAge DEGs set size, after mentioning all the drugs with 2 and 3 DEG hits, the next drug is icariin, with 8 hits. Of the 5 additional DEG hits, 2 G protein[a] alpha subunits (*gpa-5* and *gpa-6*) are modulated exclusively by icariin. This makes icariin a "**unique drug**". This was the last of the three classes of drugs according to group dominance, and the classification of my entire drug library is displayed in Table 4.1.

With a total of 24, both captopril and spermidine belong to the category of unique drugs. Captopril has a single exclusive DEG hit (*B0250.5*) and spermidine has 13. Continuing, EGCG correctly modulates 72 GenAge genes,

[a] G proteins are a family of guanine nucleotide-binding proteins and hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP).

of which 20 are uniquely targeted by this compound. Another unique drug is lithium, which has 13 exclusive DEGs among its 82 hits. With 13 unique DEGs out of 89 total hits, NAC is also a unique drug. Alpha-ketoglutarate correctly modulates 105 lifespan-extending GenAge genes, and of these 11 are specific to this drug.

The drug with the most targets both in general and in my gene set of interest is thioflavin-T. It correctly modulates 181 lifespan-extending GenAge genes, and its profile is unique because there is a subset of 100 DEG genes that are not influenced by any of the other drugs.

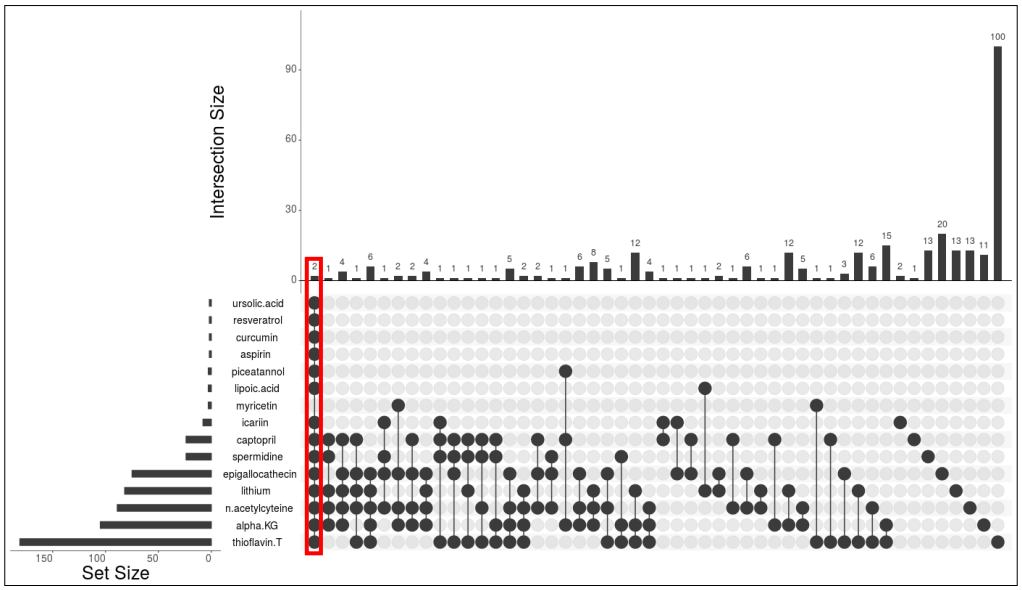

Figure 4.1 – UpSet plot of lifespan-extending GenAge DEGs.
The red box refers to the example mentioned in the text. On the bottom left it is depicted the correctly-modulate GenAge DEGs set size for each drug. On the bottom graph it is represented the number genes in the subset that it is exclusively shared by the drugs marked with a dark circle and none others.

Using Venn diagrams to convey the unique and shared genes among such a high number of drugs would be very confusing. Therefore, I used the UpSetR[239]

Shiny App[a] to create an UpSet plot (Figure 4.1). This plot displays on the bottom left the set sizes of the correctly modulated DEGs for each of the drugs, in an ordered manner. The bottom right part exhibits an existing gene membership profile, and it is aligned with the top frequency bar plot showing the number of correctly modulated DEGs displaying it. This way of displaying the data makes it noticeably clear that (as previously mentioned) there are 2 DEGs that are correctly modulated by all of the drugs except myricetin (see red box in Figure 4.1). Such plots also allow quick identification of unique drugs, as they result in rows with unconnected circles.

A complementary way of displaying the DEGs correctly modulated by the drugs is called a Bertin plot. Using the seriation R package (version 1.2-8)[240] I ordered the correctly modulated hits of each of the drugs (Figure 4.2). The Bertin plot complements the previous plot because it makes the amount of correctly modulated DEGs by each drug and the drug-wise overlap even more explicit.

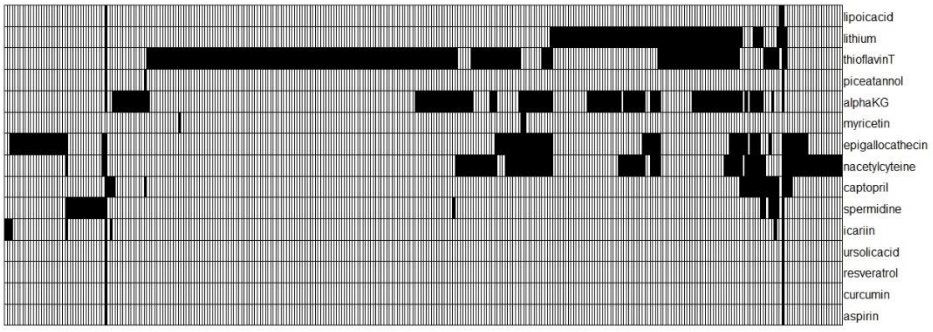

Figure 4.2 – Bertin plot of lifespan-extending GenAge DEGs.
Only correctly modulated DEGs are considered. Each column pertains to a specific GenAge gene, and it is a DEGs for the drugs that name the rows with black cells.

[a] Freely available online at https://gehlenborglab.shinyapps.io/upsetr/ .

Analysis of transcriptional effects of drugs based on this drug dominance framework, allows to understand how drugs relate among themselves (assess similarity and differences). Again, assuming linearity of addition (see section 1.5.2) a combination of EGCG, NAC, lithium, alpha-ketoglutarate, captopril, icariin, spermidine, and thioflavin-T, would target[a] all the DEGs targeted by my entire drug library. That is, I can greatly reduce my drug library from 15 to 8 drugs without abdicating targeting any of the GenAge lifespan-extending genes.

In addition to assuming that gene-expression changes would add linearly, the group dominance framework has further limitations in that I only considered the binary case, that is a gene is correctly modulated or not, regardless of their relative magnitude of gene expression changes. However, this framework might be of value conceptually as it allows classification of drugs by similarity and overlap in terms of likely mode of action.

4.3 - Manual interpretation of mode of action based on DEG data

Next, I attempted to relate significant transcriptional changes to biological effects on lifespan and putative mode of action. To do so, I read the bibliography referenced by GenAge for each of the genes, and attempted to relate the findings between a gene manipulation and: the phenotypic effects of my drugs (assuming that if a single gene is the main mode of action of one of my drugs, then they should have similar phenotypes); the gene epistasis therapeutic dependency of a drug treatment (based on Table 2.1); and the magnitude of transcriptional changes elicit by my drug (based on the simplification that strongly drug-

[a] assuming solely additive effects, with no possibility for non-linear interactions.

modulated genes are more likely to be constitute the drug mode of action). Additionally, after reading the literature regarding the correctly modulated DEGs for a given drug, I tried to find commonalities between genes and conceptually cluster them[a].

Besides the assumptions touched in the last paragraph, the disadvantage of this procedure is that it is biased towards explanations that are more popular in the literature[b]. This can be evident, for example, by the information in Table 2.1, which shows that almost all my drugs have been assay in DAF-16 mutants, while less than half have been examined on AGE-1 mutants. The advantage is that I manually curated each of my results, assuring a higher quality then automatically extrapolated gene sets[c].

Apart from myricetin, all the drugs activate the classical heat-shock protein 16 (HSP-16) stress response. I can make four remarks regarding this result.

First, extra-copies of the *hsp-16* gene have been shown to increase lifespan[241], and its down-regulation by RNAi shortens lifespan[242]. This causal and positively correlated relationship with longevity is nonetheless of small magnitude[d] and therefore might only explains partially the pro-longevity effects of the more powerful of these drugs.

Second, it might just be that most of the drugs stress the organism to some extent and this pathway is being activated indirectly. It is a well-established result that a mild heat-shock stimulus extends lifespan in young worms and this response

[a] the gene sets annotated in databases are nothing more than clusters of genes.
[b] I omitted the time requirements of manual curation as a disadvantage.
[c] obviously, the manually curation also allowed me to acquire significantly more knowledge regarding aging-related pathways.
[d] approximately from -11% to $+11\%$ of lifespan difference.

requires the HSP-16 pathway[243]. This leads to the hypothesis that most of my anti-aging drugs act partially through hormesis[112], which in the specific case of HSP-16 leads to up-regulation of mechanisms preventing protein aggregation[244] (an hallmark of aging[245]).

Third, the fact that aspirin, curcumin, resveratrol and ursolic acid have this pathway as their only hit from my methodology based on GenAge is an interesting result. Aspirin and curcumin extended mean healthspan by around 28%, while resveratrol and ursolic acid failed to have an effect. I checked if there is a correlation between the activation strength (gene expression fold changes) of said genes and the magnitude of healthspan-extension, but the result was insignificant. This could have been a possible explanation for the two drug pairs difference in their therapeutic potential. An alternative explanation is that there are more DEGs, outside the set of GenAge, that are involved in this biological mechanism and dictate the therapeutic effect of activating the HSP response pathway. Some support for this explanation comes from the observation that for the pair of healthspan-extending drugs (aspirin and curcumin) the HSP response has been implied in their mode-of-action, in contrast with the pair of drugs of no therapeutic benefit (resveratrol and ursolic acid). In worms, aspirin increases resistance to heat-shock[87] and reduces the formation of aging-associated protein aggregates[88] (which are phenotypic responses typical of HSP activation). In flies, the effect size of curcumin treatment is larger, at higher environmental temperature (heat stress), and it is concomitant with the increase in the expression of heat-shock proteins[108]. To best of my knowledge, the effect of ursolic acid on HSP-60 has never been studied before, however an increase in heat-shock resistance[206] has been

reported in worms treated with ursolic acid. In the case of resveratrol, my findings, seem to initially contradict existing literature which showed that juglone-induced HSP-60 protein expression is ameliorated by resveratrol in a dose-dependent manner[246]. Albeit it might be the case that HSP-60 is increased in response to juglone-induced reactive oxygen species damage and that resveratrol protects from this stress, therefore leading to a decrease in the HSP-60 levels, as there will be less stress.

Fourth, myricetin seems to be a unique drug in this regard. My literature review did reveal that myricetin-fed worms have been reported not to be more resistance to heat-stress than untreated controls[145], which supports the transcriptomics results that show lack of heat-shock protein response.

Inspired by the concept of stochastic dominance from decision theory, I created a simple framework to categorize drugs based on their gene expression profiles in the subset of DEGs of therapeutic relevance. I further shown that under this categorization[a], a set of 8 drugs is sufficient to target the totality of the lifespan-extending GenAge DEGs modulated by the entire drug library.

The categorization procedure under the drug dominance framework is not without limitations. Its binary nature, in which genes are either targeted or not is the maximum simplification of the continuous nature of gene expression modulation. Surprisingly, binarization of gene expression activation was proven to generate a more accurate transcriptomic aging clock than using continuous values[247]. The other limitation of this framework is that it completely disregards

[a] which makes assumptions, including being limited to binary variables (a DEG is or is not being targeted).

the hierarchical nature of gene regulatory networks (down-stream and up-stream gene relationships).

In the following sections, I will individually discuss the most relevant unique genes of this set of 8 drugs, because by definition (under the same simplifying assumptions) they will be responsible for the added therapeutic potential specific to the addition of each drug to the set of the remaining 7 unique drugs.

4.3.1 - Captopril

The only lifespan-extending gene specifically modulated by captopril is the probable mitochondrial 3-hydroxyisobutyrate dehydrogenase *B0250.5* (-14% expression). It encodes an aggregation-prone protein, and its knockdown causes lifespan extension, which the authors imply might be due to reduced protein aggregation[248].

4.3.2 - Icariin

The two genes that are exclusively induced by icariin correspond to two subunits alpha of the G protein. After analyzing the cited source reference[249], I noticed that there is a mistake in the GenAge database, the *gpa-6(pk480)* mutants do not live longer as reported in GenAge[a]. As a result, the only lifespan-extending GenAge correctly modulated by icariin treatment is *gpa-5* (0.5% decrease in gene expression). This gene is involved in olfaction, and its expression level is negatively correlated with lifespan[249].

[a] their lifespan is actually less than the control group, although this result is not statistically significant.

4.3.3 - Alpha-ketoglutarate

Like icariin, alpha-ketoglutarate also modulates a sub-unit alpha of the G protein involved in olfaction. With the difference that the expression of the *gpa-2* gene (+47% gene expression changes) is positively correlated with longevity[249].

It has previously been reported that alpha-ketoglutarate is able to further extend the lifespan of long-lived *daf-2(e1370)* mutants[78]. One of the exclusively alpha-ketoglutarate upregulated DEGs which might be involved in this effect is the *jnk-1* (c-Jun N-terminal kinase) gene (38% more expression). The double mutant of a JNK overexpression line (*lpIn1*) with *daf-2(e1370)*, lives much longer (an additional 21%) than the DAF-2 control mutants[250]. The JNK pathway is a classical aging pathway, and the observation that alpha-ketoglutarate is a JNK activator is an interesting starting point for further studies. Nonetheless, I expect that the JNK pathway might at best only explain part of the alpha-ketoglutarate pro-longevity effects, because alpha-ketoglutarate also extends the lifespan of *daf-16* knockout mutants (although only by a small magnitude)[78], but the pro-longevity effects of JNK activation are known to be abolished in this genetic background[250].

4.3.4 - Spermidine

Of the 15 drugs in my drug library, spermidine is the least understood regarding its genetic epistasis of lifespan effects in worms. Consequently, I will delve deeper in my findings, as they are all new.

The pro-longevity effects of spermidine in an SKN-1[a] genetic background should be examined. My suggestion originates from the fact that both spermidine and *skn-1* RNAi[b] down-regulate the anti-longevity *lys-6* (lysozyme) and *cyp-13B1* (cytochrome P450 subfamily) genes[251], by 28% and 9%, respectively.

Furthermore, analysis suggests that the effect of spermidine can be expected to be DAF-1-dependent. Out of the 13 DEGs solely affected by spermidine, 3 have been reported to be dependent on or downstream of DAF-16. These are the *egl-27* (a GATA transcription factor[252c]), *dod-19* ("downstream of daf-16"[89]) and *sams-1* (S-adenosyl methionine synthetase[253]) genes; 20%, -38% and -61% expression changes, respectively. Interestingly, the latter has been reported to be involved in DR[253d], and since it is implied in two of the major aging-associated pathways it would by itself be interesting to investigate the joint effect of spermidine and *sams-1* RNAi. In more detail regarding the role of *sams-1* in DR, *sams-1* RNAi treated animals resemble DR animals in their slightly reduced brood size, marginally delayed reproductive timing, and slenderness. Furthermore, *eat-2* mutants display a 3-fold reduction in *sams-1* mRNA levels[253].

[a] orthologous to the mammalian Nrf (Nuclear factor-erythroid-related factor) transcription factor and its overexpression increases lifespan[311].
[b] the transcription factor *skn-1* has a complex effect on the longevity of worms because, on another hand, the decrease of its expression by RNAi or loss-of-function mutations leads to a shortened lifespan[311,312].
[c] it belongs to the family of transcription factors characterized by their affinity to bind with "GATA" parts of the DNA[313].
[d] more precisely, sams-1 RNAi failed to further extend the long lifespan of the eat-2(ad1116) mutant worms.

In addition of *sams-1*, two more lifespan-extending GenAge DEGs correctly modulated only by spermidine have been previously associated with DR. *nhx-2* is a Na+/H+ exchanger expressed exclusively at the intestinal epithelium that spermidine inhibits by 27%. It has been reported that *nhx-2* RNAi treated worms display a CR-like phenotype, including fat loss and 40% increased lifespan[254]. The other gene is the ATP synthase active-subunit of the respiratory complex V – *atp-2* (-21% expression change). Post-developmental *atp*-2 RNAi, results in an impressive 43% lifespan extension[255]. Contrary to *nhx-2*[254], *atp-2* knockout mutants do display reduced pharyngeal pumping rates[256] (therefore this might be a discriminant healthspan assay to try[257]). It is important to keep in mind that *atp-2* encodes an aggregation prone protein, and it might just be less accumulation of its aggregates that results in lifespan extension[248].

If spermidine manages to extend the lifespan of both DR and *daf-16* knockout models, it might be the case that it is acting by downregulating the *kynu-1* gene (-24% gene expression). This gene encodes what is thought to be an ortholog of the human kynureninase enzyme. Decreasing its gene expression by *kynu-1* RNAi is known to increase lifespan, largely independent from the previously mentioned pair of pathways[258].

In addition to the complex V, Spermidine treatment also targets the complex 1 of the respiratory chain. It does so by being the only one of my drugs capable of modulating the *C18E9.4* gene (NADH CoQ oxidoreductase B12 subunit). This DEG could be of interest because its RNAi leads to a 60% mean lifespan extension[259] and it has been replicated[260]. Unfortunately, spermidine only decreases its expression by 9%.

The *bcat-1* (branched-chain amino acid transferase-1) gene is an evolutionary conserved lifespan and healthspan-extending gene[261] that is downregulated by spermidine treatment (-44% abundance).

4.3.5 - NAC

NAC is another drug for which the genetic underpinnings of its action in worms remain unknown. From the only 13 lifespan-extending GenAge DEGs correctly modulated by NAC, it is very surprisingly that two of them are the classical aging-associated genes *daf-2* and *glp-1*. NAC-treated worms display 37% and 24% gene expression reduction, of *daf-2* and *glp-1*, respectively. I have touched on the DAF-2 pathway before, and the *glp-1* gene encodes the receptor for a germ-line proliferation signal that is produced by the somatic gonad cells[262]. The effect of both anti-longevity genes, which NAC treatment downregulates, have been widely replicated. Another commonality between these pair of genes is that even their post-developmental RNAi increases lifespan[255].

Besides the previously mentioned two canonical aging-related genes, there are several more DAF-16 dependent DEGs unique to NAC treatment. The *hrp-1* gene encodes a human HnRNP A1 homolog, and its post-developmental RNAi leads to a lifespan-extension of approximately 25% (which is in-between the magnitudes caused by post-developmental *daf-2* and *glp-1* RNAi)[255]. The *spe-6* gene, when silenced, specifically affects the development of viable spermatids[263,264]. The multi-PDZ domain-containing protein[a], encoded by the *mpz-1* gene, is thought to act downstream of DAF-2 and upstream of DAF-16[265].

[a] the PDZ domain is a common evolutionary conserved protein domain.

The expression of an anti-longevity DEG encoding a lipid binding protein, *lbp-7*, is also down-regulated in DAF-2 (mutants and RNAi) and up-regulated in *daf-16;daf-2* double mutants[89]. This indicates that expression of the *lbp-7* gene is elicited by insulin-like growth factor signaling. Lastly, the *pbs-5* gene encodes a catalytic 20S subunit orthologue of the human core proteasome beta-5 subunit. It acts upstream of DAF-16, and its pro-longevity effects also depend on the SKN-1 and HSF-1 pathways[266]. Furthermore, it is related to the first two mentioned DEGs, in that it is upregulated in *glp-1* mutants and its overexpression further extends the lifespan of *daf-2(e1370)* mutants. The lifespan-extension resulting from activation of the *pbs-5* gene is accompanied by increased resistance to polyglutamine and A-beta proteotoxicity[a], as expected from the elevated capacity of preserving proteostasis[266].

In regards to DEGs related to the DAF-2 pathway (and not confirmed to be daf-16 dependent), the transcription elongation encoding gene *hmg-4* is thought to directly interact with DAF-2[267], and the *gst-10* is partially-required for DAF-2 lifespan extension. It encodes a glutathione transferase that interacts with products of lipid peroxidation, that when overexpressed further increases anti-oxidant defenses of the organism, culminating in lifespan-extension[268,269].

The specific up-regulation of the *alh-6* aldehyde dehydrogenase gene expression observed in NAC treatment extends lifespan depending on the specific bacteria food source[270], so it might not be a robust mechanism for lifespan extension.

[a] the authors tested on disease model strains.

4.3.6 - Lithium

Of the 13 lifespan-extending GenAge DEGs that among my drugs are uniquely modulated by lithium, I will focus on only on those for which I found more information.

As mentioned in the literature review (section 2.2.9), lithium is an FDA-approved GSK-3 inhibitor. Contrary to this hypothesis that lithium's pro-longevity mode of action is inhibition of GSK-3[37], there are several DEGs associated with the mTOR pathway. For example, the *pdk-1* gene (PDK-class protein kinase 1) is conserved in humans and it is down-regulated by 21% in lithium-treated animals[271]. Another, perhaps even more fundamental mTOR lifespan-extending GenAge DEG modulated by lithium is the conserved Rag GTPases coding gene *raga-1*. Moreover, down-regulation of *raga-1* increases healthspan as measured by locomotion, just like lithium treatment[272]. Furthermore, identical to lithium treatment[124], inhibition of *raga-1* gene expression also leads to a redistribution of *lgg-1*:GFP from diffused to punctate foci[272]. The GFP-fused vacuolar protein LGG-1 marks autophagic vesicles[273], and induction of autophagy is one of the ways by which the inhibition of the mTOR prolongs longevity[274]. In sum, this is another line of evidence supporting lithium as an mTOR inhibitor. This mode of action would agree with the results of my healthspan high-throughput screen. In my combinatorial drug trial, lithium synergized with rifampicin but not psora-4; and this is the same profile displayed by rapamycin in our previous work[70]. In other words, there is some support to the hypothesis that rapamycin and lithium share the same mode of

action, because it showcases that lithium treatment by itself already down-regulates the TORC1 sub-pathway[a].

Curiously, *raga-1* RNAi increases the lifespan of the already long-lived *glp-1* knockout mutants; and *glp-1* was a unique DEG down-regulated by NAC (see the previous section). This leads me to suggest that a combined lithium plus NAC treatment might be even more beneficial than any of the individual drugs. Not supporting all of the above, is the fact that the magnitude of *raga-1* expression inhibition (-13% expression) by lithium treatment is small.

The *ifta-2* gene encodes the intraflagellar transport associated protein 2, which is homologous to the mammalian Rab-like 5 protein. Lithium treatment inhibited its transcripts abundance by 25%. Modulation of *ifta-2* expression has been shown to extend lifespan in a DAF-16 dependent manner[275]. Lithium pro-longevity effects have no such restriction (see literature section X), and therefore this is unlikely to be the main mode of action.

Taking the known genetic epistasis of lithium in *C. elegans*, I found only one gene that when it is targeted by RNAi, elicits lifespan-extension in a comparable manner: the *tars-1* threonyl amino-acyl tRNA synthetase gene. Like lithium treatment, knockdown of *tars-1* expression further prolongs the longevity of *daf-2* and *eat-2* knockout mutants and is independent of the DAF-16 pathway[276]. Worms treated with *tars-1* RNAi also displayed reduced fertility. Unfortunately the magnitude of *tars-1* inhibition caused by lithium treatment is small (16% reduction).

[a] and rapamycin is a canonical mTOR inhibitor.

4.3.7 - EGCG

The *atf-6* gene is modulated (-25% expression) by EGCG, in the same way first assigned (and thought to be unique) to astragalus polysaccharide[a] treated-worms[277]. It would be interesting to find the drug dominance relationship between these two compounds. Furthermore, EGCG down-regulates the expression of *che-3* (-66%), che-*11* (-78%) and *daf-10* (-62%), which are the known genes for which their knockout mutants are long-lived[278] and have reduced or irregular cilia[b]. The fact that all 3 of them are targets of EGCG makes it more likely that this cluster of genes is involved in the increased lifespan-extension elicited by EGCG.

Lifespan extension by EGCG is DAF-16 dependent. Consistently, several of its targets DEGs are gene knowns to be regulated by DAF-16. The genes *ttr-5*[279] (transthyretin-related family domain) and ammonia permease *amt-2*[279] belong to this category, and are down-regulated by EGCG by 25% and 81%, respectively. Furthermore, both DEGs *sca-1* (sarco-endoplasmic reticulum ATPase)[280] and *par-5* (abnormal embryonic partitioning of cytoplasm)[281] are direct targets of DAF-16. While the expression of the first is decreased by EGCG (-35%), the latter is up-regulated by EGCG (55% increase in expression). Another DAF-16 dependent DEG modulated by EGCG is *odr-3* (odorant response abnormal). EGCG treatment reduced *odr-3* expression by 32%. This is of interest because decreasing the expression of this gene has been shown to prolong the lifespan of DAF-2 mutants[249], and it is known that EGCG

[a] is a traditional Chinese medicine obtained from the herb *Astragalus membranaceus*[314].
[b] cilia function as sensory receptors.

also further extended DAF-2 knockout mutants lifespan. There is another DEG, that among my drugs it is uniquely modulated by EGCG, *tax-6* (abnormal chemotaxis), that reproducibly extends the lifespan of DAF-2 mutants[157,282]. EGCG reduces *tax*-6 transcript abundance by 36%. Unfortunately, the longevity effect of *tax-6* expression inhibition is only partially independent (smaller magnitude of effect) on DAF-16 (it increases the longevity of DAF-16 knockout mutants), which does not match the genetic epistasis of EGCG treatment, which is of no benefit to the longevity of DAF-16 mutants (see Table 2.1).

Nonetheless, EGCG treatment still has commonalities with DAF-2 knockout mutants. For example, the *gcy-18* (guanylyl cyclase)[89] gene is only modulated by EGCG. Its expression is reduced by 65% due to EGCG and it is also known to be down-regulated in DAF-2 mutants.

Interestingly, there are 3 egg-laying defective DEGs among the ones uniquely modulated by EGCG *egl-8*, *elg-30,* and *elg-9*; their expression is reduced by 36%, 30% and 41%, respectively. The results showing the first to be an anti-longevity lifespan-extending gene[283] has been reproduced[284]. The *elg-30* gene activates *egl-8*. Interesting, decreasing the expression of either of this genes has opposite effects on the lifespan of DAF-16 mutants; knockdown of the former extends it, while the latter shortens it[283].

The EGL-9 protein hydroxylates HIF-1, which is then ubiquitinated and targeted for proteosome degradation. HIF-1 is a hypoxia inducible transcription factor that has a complex role in lifespan with both its over-expression[285] and deficiency[286] have previously been shown to extend lifespan. The down-regulation of the *egl-9* by EGCG, might lead to a more powerful hypoxic

response (due to accumulation of HIF-1), and constitute a pro-longevity pathway that is independent of dietary restriction and the IIS pathway. Moreover, *egl-9* RNAi confers resistance to amyloid beta toxicity[287].

4.3.8 - Thioflavin-T

In the literature review section of this drug (section X), I stated that it has the largest effect size of any known adult-onset pharmacological intervention in *C. elegans* and attributed its robustness in the Caenorhabditis Intervention Testing Program (CITP) to this. Now that I have transcriptomics data of worms treated with thioflavin-T This provides me with the observation that can just as well explain the robustness of thioflavin-T treatment. Thioflavin-T it is not only the drug with the most DEGs, but more importantly, that is my drug with the larger set of correctly modulate lifespan-extending GenAge DEGs (by an order of magnitude higher than any of my drugs).

Among the 100 lifespan-extending GenAge DEGs correctly modulated by thioflavin-T, at least 12 of them encode aggregation-prone proteins[248] and are down-regulated by this drug. This could imply a remarkable complementary role for thioflavin-T in the maintenance of proteostasis: not only it binds to the amyloid directly[a202], it also acts indirectly by reducing the expression of 12 genes that encode proteins that form aggregates (see Table 4.2).

[a] recall that thioflavin-T is used as an amyloid-binding dye (see section 2.2.19).

Gene symbol	Relative expression (%)
rps-23	-52
rps-5	-40
dlat-1	-38
cct-1	-31
tbb-2	-26
pab-1	-21
rpl-17	-19
cpn-3	-17
rpl-10	-17
F01G6.4	-15
gdh-1	-9
rpl-31	-6

Table 4.2 – DEG coding aggregation-prone proteins unique to thioflavin-T.
The left column displays the gene symbol, and the left the relative expression of transcript abundance relative to the DMSO negative control group.

The dozen mentioned genes are[248]: *cct-1* (chaperin containing T-complex protein 1 subunit alpha), *cpn-3* (calponin), *dlat-1* (dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex), *F01G6.4* (mitochondrial phosphate carrier protein), *gdh-1* (glutamate dehydrogenase), *pab-1* (polyA binding protein)[288]; *rpl-10*[289], *rpl-17* and *rpl-31* (60S ribosomal large subunit proteins); *rps-5*[264] and *rps-23*[264] (two ribosomal small subunit proteins); *tbb-2* (tubulin beta-2 chain). Moreover, most of these genes have orthologs in humans.

Thioflavin-T pro-longevity has been shown to be DAF-16 independent, and partially overlapping with DR. From the lifespan-extending DEGs unique to thioflavin-T for which there is lifespan data in *C. elegans*, I will now discuss the ones that are neither totally or partially dependent on DAF-16, nor on its downstream targets. Two of them[253], *ril-1* and *ril-2* (RNAi-induced longevity) encode genes that have no obvious homologue, so might not be the most

applicable in mammals. I found an additional 12 DAF-16 independent DEGs[264]: *F13B6.1*, *F49C12.9*, *mrpl-12*; *mrps-9* and *mrps-33* (two mitochondrial ribosomal proteins); *T28D6.4*, *tba-7* (alpha tubulin), *ZK809.3*, *Y56A3A.19*[255], *rec-8*[290], *qars-1* (glutaminyl(Q) tRNA synthetase), *lpd-5* (lipid depleted)[276], *cyc-1* (cytochrome c1)[253,259]. The last three also extend lifespan in DAF-2 mutants.

Gene symbol	Relative expression (%)
tba-7	-42
Y56A3A.19	-40
F13B6.1	-39
rec-8	-34
qars-1	-34
lpd-5	-30
T28D6.4	-29
ril-2	-25
F49C12.9	-24
cyc-1	-24
mrps-9	-21
mrpl-12	-18
mrps-33	-15
ZK809.3	-10
ril-1	-5

Table 4.3 – The 12 DAF-16 independent DEGs unique to thioflavin-T.
The left column displays the gene symbol, and the left the relative expression of transcript abundance relative to the DMSO negative control group.

Another alternative processes that I found to be caused by DEGs unique to thioflavin-T are mitonuclear protein imbalance[151] and mitochondrial unfolded protein response[291].

Under a different paradigm, but equally interesting would be to explore if the progeny of thioflavin-T treated worms has a longer lifespan. This might actually be the case because thioflavin-T correctly modulates two DEGs that are both capable of this[292,293].

The previously mentioned pathways are highly distinct, and this may explain the robustness of thioflavin-T in the CITP. But there is an even stronger possibility: thioflavin-T is an incredibly robust intervention because it has among its uniquely modulated DEGs 5 of the canonical aging-genes: *sir-2.1* (35% expression increases), *daf-7* (-23% expression change), *skn-1* (29% expression increase), *hsf-1* (52% expression increase) and *clk-1* (-9% expression change).

Chapter 5 – Larger combinatorial healthspan drug screen

5.1 - Introduction

In - Automated high-throughput healthspan drug screening in *C. elegans*Chapter 3, I validated all the steps of my proposed methodology in isolation and integrated. In Chapter 4, I confirmed that the selected drug library features drugs which target a diverse set of known lifespan-extending genes. Therefore, I will now apply the drug screen methodology to my drug library.

As mentioned before, the selection of drugs was based on previously described criteria (see sub-chapter 2.1), and an up-to-date literature review regarding each of the drugs in the drug library is provided in sub-chapter 2.2.

5.2 - Monotherapy results

Even for conditions testing only water-soluble drugs, DMSO was added to a final concentration of 0.1%. This was done so that the concentration of DMSO was identical in each well (apart from psora-4 which required a higher DMSO concentration, see below), allowing direct comparison of efficacy of drug combinations with each other and with single drug and untreated controls.

I included rapamycin amongst the drugs for which I investigated interactions, but I am not including these results here because of an unexpected problem with solubility that was consistently observed across all rapamycin conditions of this screen. I observed that wells containing rapamycin showed visible precipitates (crystals). In agreement with this observation, indicating a problem with rapamycin solubility, rapamycin treated cohorts in this trial were indistinguishable in terms of lifespan from controls and, for each drug tested together with rapamycin, the observed effects were identical to treatment with

that single drug alone. This failure of rapamycin to elicit any lifespan benefits, either alone or in combination with other drugs stands in direct contradiction to my previous results which, in fact, successfully used rapamycin as the positive control. The cause of this might be because I used an older stock, but I did not have time to fully investigate the cause of this failure.

Drugs	Restricted Mean (days)	Magnitude %	Bonferroni p-value
alpha-KG	**21**	**5**	**1**
Aspirin	**25**	**23**	**0.00003**
Captopril	**25**	**24**	**1.20E-07**
Curcumin	**25**	**23**	**0.0008**
EGCG	**19**	**-5**	**1**
Icariin	**22**	**8**	**0.0695**
Lipoic acid	**23**	**13**	**0.471**
Lithium	**22**	**8**	**1**
Myricetin	**25**	**25**	**<1E-10**
NAC	26	28	<1E-10
Piceatannol	**24**	**22**	**<1E-10**
Psora-4	23	11	0.**0087***
Resveratrol	20	0	1
Rifampicin	22	8	0.**041**
Spermidine	23	17	0.**0078**
Thioflavin T	22	8	1
Ursolic acid	22	8	1

Drugs and their used concentrations are reported in the first column. "Mean healthspan" is quantified by the restricted mean (sub-chapter 2.5). The third column indicates the magnitude of mean healthspan changes relative to the negative control group.
*due its solubility psora-4 was diluted in 1% DMSO and the values reported are relative to the 1% DMSO concentration.

I separated the above conditions, except for psora-4, into two survival separate plots, one for water soluble and the other for water insoluble (DMSO) drugs. This separation is somewhat arbitrary because even the water soluble compounds were tested in the presence of DMSO to allow more direct

comparison. For consistent visual inferences, I included the same negative (0.1% DMSO) and positive (rifampicin) control conditions in each plots. The rifampicin positive control did extend mean healthspan in statistically significant manner by about 8%[a].

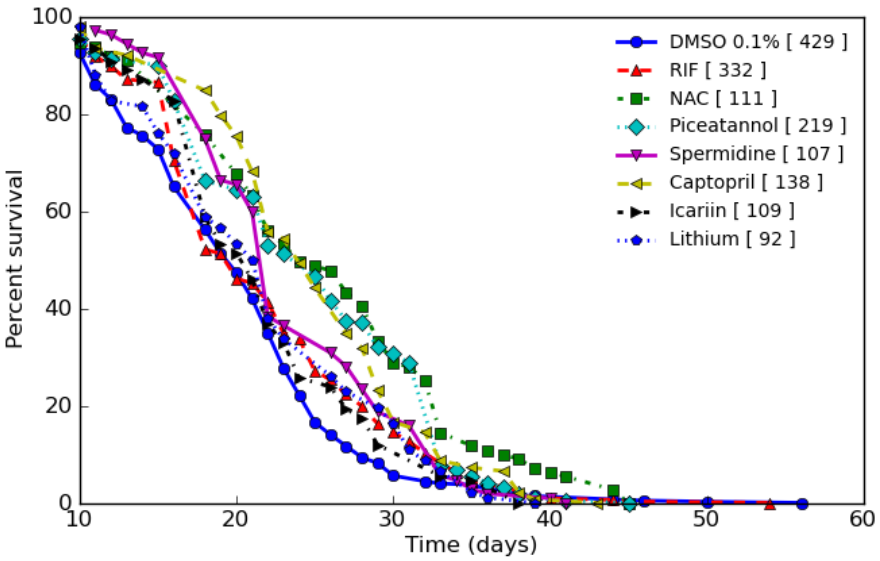

Figure 5.1 – Survival plot of water-soluble single drug interventions.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

[a] this is especially acceptable if one considers that the mean healthspan of the 0.1% DMSO negative control group was 20 days.

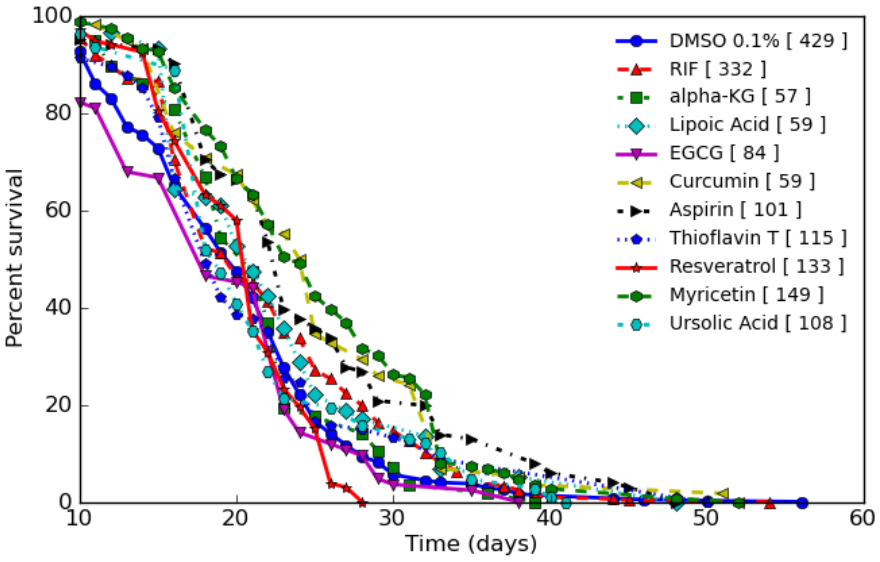

Figure 5.2 – Survival plot of DMSO-soluble single drug interventions.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

5.3 - Combinatorial healthspan interventions

The key objective of this screen was to systematically test every single drug in my drug library when combined with rapamycin, rifampicin, or psora-4 (3 pairing of each drug). I chose this design as an extension of our previous work, identifying two synergies involving rapamycin, rifampicin and psora-4[70]. By investigating all possible pairs of drugs involving one of these three compounds, I aimed to answer the question how common synergistic interactions between drugs are. The intention was for the two known synergistic pairs (rifampicin with psora-4 and rifampicin with rapamycin) also served as positive controls for synergy while the (non-synergistic) rapamycin and psora-4 combination would serve as the negative control. Unfortunately, as mention in the previous section for the monotherapy, there was a problem with rapamycin solubility in this trial, and the there was no difference between each single drug and their

corresponding pairs with rapamycin [a]. Due to the failure of rapamycin in this trial, only the single combination (rifampicin with psora-4) was used as positive control.

5.3.1 - Rifampicin Pairs

As in the previous section, Figure 5.3 and Figure 5.4 show all rifampicin drug pairs with drugs that are either water or DMSO soluble, respectively. For the sake of comparison, I included the negative (0.1% DMSO) and positive (rifampicin only) control conditions in both plots.

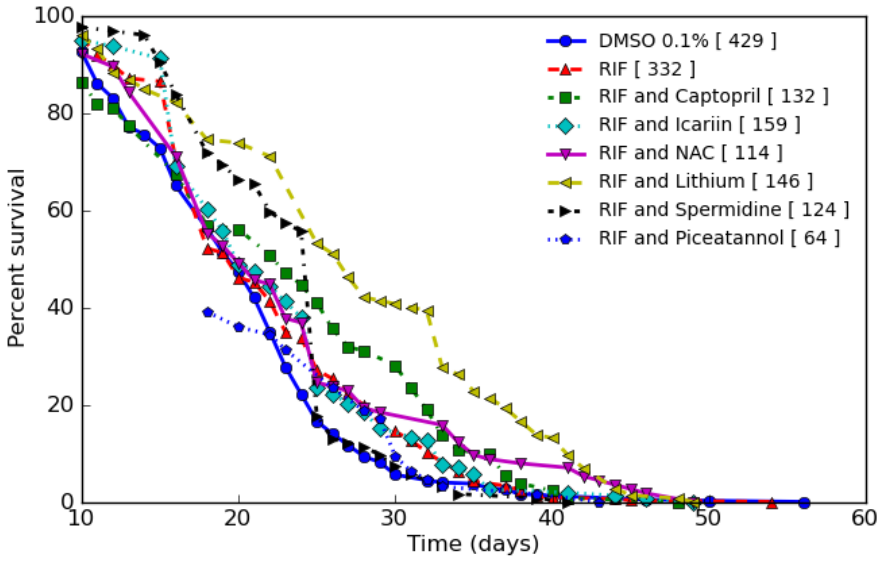

Figure 5.3 – Survival plot of rifampicin paired with water-soluble drugs.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

[a] This suggests that the results regarding the monotherapies are robust, which is in itself evidence for the validity of my pipeline.

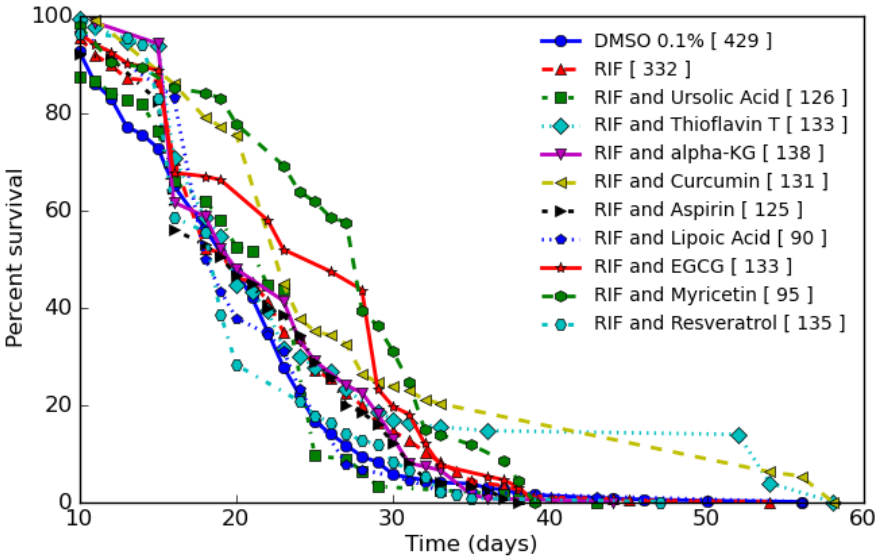

Figure 5.4 – Survival plot of rifampicin paired with DMSO-soluble drugs.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

For a better systematic comparison, the healthspan results of each rifampicin pair is depicted in Table 5.1. The combinatorial interventions are named in the first column and each row's data belong to the named combinations. The second column presents the mean healthspan (measured as explained in section 672.5.2) in days, rounded to unit (nearest day). The rows of the table are sorted in decreasing order. The values that are in bold signify that the restricted mean healthspan achieved by the respective combination is larger than the reference restricted mean healthspan of the synergy control pair rifampicin and psora-4 of approximately 27 days. The left 3 columns contrast the relative difference between the restricted mean healthspan of the drug pair intervention versus 0.1% DMSO, rifampicin and the monotherapy of the drug that is being paired with rifampicin in the combination, respectively. Therefore, positive values indicate an extension of mean healthspan relative to the 0.1% DMSO control

group. Finally, bold values correspond to comparisons for which the log-rank test shows statistical significance (for an alpha significance threshold level of 0.05), meaning that the drug combination has a different healthspan relative to the group named in the column (see section 2.5.3). Positive bold values indicate that healthspan is statistically significantly larger than the respective column reference group, and vice-versa for negative bold values. There are three bold drug pair to indicate that these interventions achieve statistical significance across all the table comparisons. They are colored green because they are cases of positive synergy. Notice that in this case the comparisons between each drug pair with the respective monotherapies interventions is precise, because all the conditions involved were assayed simultaneously, at a standard DMSO concentration of 0.1%.

Interestingly, there were no drugs that when paired with rifampicin decrease the mean healthspan relative to the DMSO only control group (no directly toxic interactions). However, Rifampicin combined with alpha-ketoglutarate, piceatannol, lipoic acid, aspirin, ursolic acid and resveratrol did not extend healthspan compared to the DMSO control condition. Although there are no toxic drug interactions compared to the negative control group, there are negative drug interactions when compared to the mean healthspan of the treatment using the best of the two single drugs in each pair. Adding rifampicin to aspirin significantly decreases mean healthspan (by 15%) relative to aspirin alone, also adding resveratrol to rifampicin significantly decreases the mean healthspan relative to rifampicin monotherapy (by 7%).

Drug Pairs [N]	Mean Healthspan (days)	Relative mean healthspan extension (%)		
		vs 0.1% DMSO	vs RIF	vs Monotherapy
RIF and Curcumin [131]	**29**	**45**	**34**	17
RIF and Lithium [146]	**28**	**37**	**27**	**28**
RIF and Myricetin [95]	26	**32**	**22**	6
RIF and Thioflavin T [133]	25	**25**	**16**	**16**
RIF and EGCG [133]	24	**20**	**11**	**27**
RIF and Captopril [132]	23	**14**	**6**	-8
RIF and Spermidine [124]	23	**14**	5	-3
RIF and NAC [114]	23	**13**	4	-12
RIF and Icariin [159]	22	**11**	2	2
RIF and alpha-KG [138]	22	10	2	5
RIF and Piceatannol [64]	22	10	2	-10
RIF and Lipoic Acid [90]	21	5	-3	-7
RIF and Aspirin [125]	21	5	-3	**-15**
RIF and Ursolic Acid [126]	20	2	-6	-5
RIF and Resveratrol [135]	20	0	**-7**	0

Table 5.1 – Summary of the longevity effects of rifampicin drug pair combinations.
The first column identifies the drug pair and sample size (value inside []). **Green bold** drug pairs are cases of positive synergies. The last three column display the magnitude of healthspan extension relative to the DMSO, rifampicin and paired drug interventions, respectively. **Bold** values represent a statistically significant log-rank test. RIF = rifampicin.

Overall, out of the total 15 rifampicin pairs tested, I found nine drug pairs that significantly extended mean healthspan relative to the 0.1% DMSO group. When rifampicin was combined with spermidine, NAC and icariin, the effect of the joint interventions was not different from the effect of the better of the two single interventions that comprised it. In the cases of rifampicin paired with curcumin, myricetin and captopril, there was a statistically significant increase of the mean healthspan relative to the rifampicin only group. But for these drug pairs the effect does not reach statistical significance when compared to the other drug of the pair (curcumin, myricetin and captopril, respectively). Another way to summarize the results for these 6 pairs is to say that, there was no benefit

to combining compounds in these cases. That is, applying the best of the monotherapies would suffice to elicit the same healthspan extension as the combination treatment.

However, among the rifampicin-based pair interventions are three examples where combination treatment significantly extended mean healthspan compared to the 0.1% DMSO negative control group. One of these pairs even extends healthspan, in a statistical significantly manner, when compared to the rifampicin group and the respective monotherapy condition. These 3 cases of drug synergy are achieved when rifampicin is paired with lithium, thioflavin-T and EGCG. I will refer to these three pairs as “synergistic”, following the definition introduced in section 2.5.3.

Of the 3 synergistic rifampicin-based pairs, rifampicin, and lithium (Figure 5.5) is the only intervention that displays a mean healthspan that is significantly larger than the sum of the benefits of the two constituent drugs (Table 5.1). As it can be seen in Figure 5.5, combining rifampicin and lithium resulted in an intervention that robustly extended lifespan at all time points. Rifampicin combined with lithium is therefore the only combination that is synergistic in the most stringent sense – namely that the whole is significantly larger than the sum of its parts.

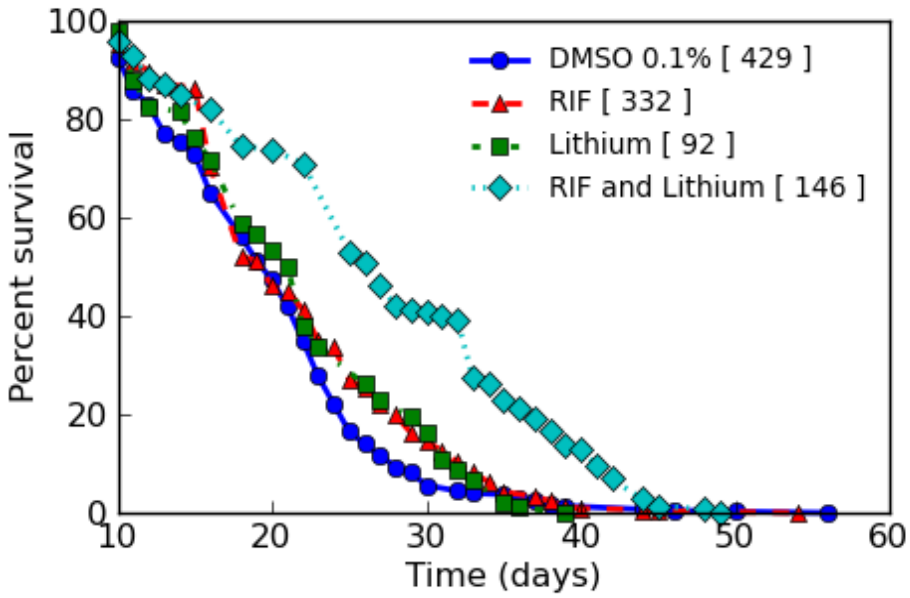

Figure 5.5 – Healthspan of rifampicin and/or lithium.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

By contrast, the combination of rifampicin and thioflavin-T resulted in a survival curve with distinct hazard ratios depending on the lifespan stage (Figure 5.6). This intervention reduced the early decline of healthspan at the population level until about day 15. From then until about day 30, it is like the rifampicin monotherapy. Interestingly, from this point on, the 20% surviving population display almost no loss of healthspan (and therefore also no mortality) for over 20 additional days, suggesting that a small subset of the population benefited disproportionally from this treatment combination. Moreover, I know that this is not some artifact (false positives) of automatizing the scoring because while I was blinded and for several days, I took note of the observation that a subset of wells had a surprisingly substantial number of worms still alive. I later found out, that these wells belonged mainly to the rifampicin with thioflavin-T paired intervention.

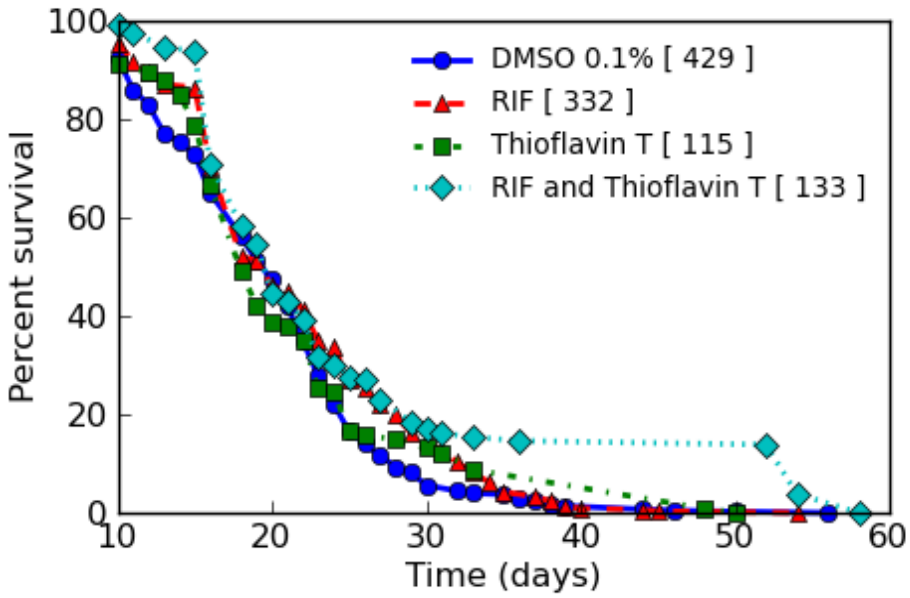

Figure 5.6 – Healthspan of rifampicin and/or thioflavin-T.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

Combining rifampicin and EGCG results in an intervention that matches the rifampicin monotherapy until about 30% of the population has stopped moving. From there until about day 35 (when 90% of the population has stopped moving), the pair intervention significantly extends healthspan. This is an interesting case of synergy because EGCG monotherapy does not extend healthspan relative to the negative control group (Table 5.1).

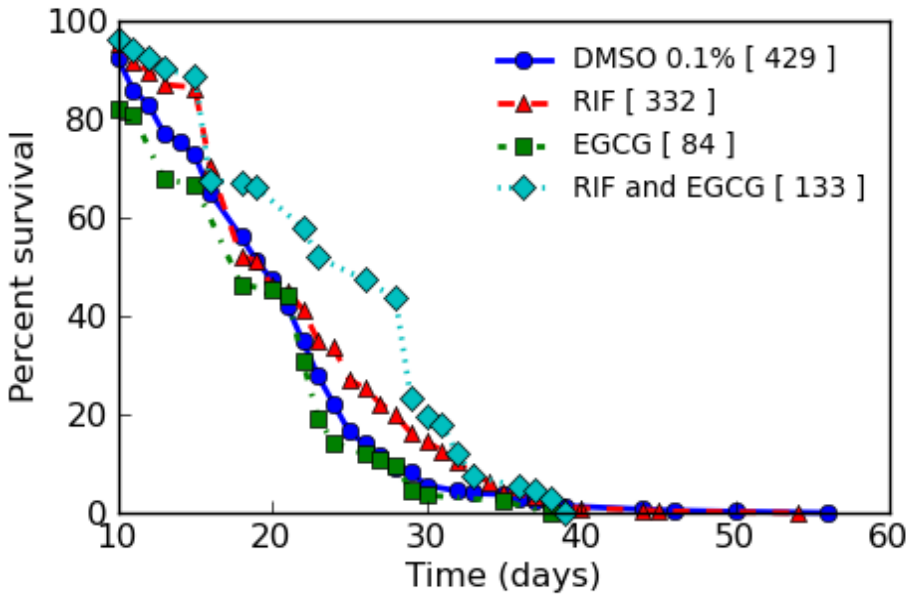

Figure 5.7 – Healthspan of rifampicin and/or EGCG.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

5.3.2 - Psora-4 pairs

Due to psora-4 requiring a final DMSO concentration of 1% only now, that I am analyzing the drug pairs that have 1% DMSO, do I display its "healthspan" survival curve (Figure 5.8). Psora-4 was my other positive control condition and it was valid as such because it increased mean healthspan by 11% compared to the 1% DMSO control. What in rifampicin was a somewhat arbitrary bipartition of drug pairs into water-soluble and DMSO-soluble drugs, is more significant when analyzing pairs based on psora-4. The reason is that when paired with water-soluble drugs (Figure 5.8) the final DMSO concentration was the same as for psora-4 by itself (1% DMSO). But when paired with other DMSO-only soluble drugs, the final concentration reached 1.1% DMSO, and, accordingly, I created a 1.1% DMSO negative control group (Figure 5.9).

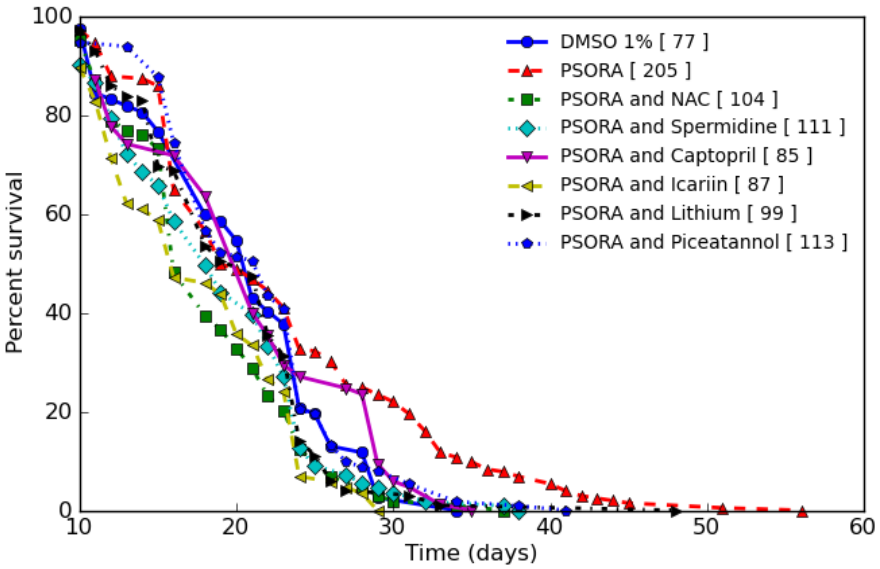

Figure 5.8 – Survival plot of psora-4 paired with water-soluble drug.
Numbers inside square brackets indicate the sample size. Percent survival represents the percentage of moving worms per time point relative to the starting population.

The positive (known synergy) control (psora-4, rifampicin) is among the psora-4 pairs with DMSO soluble drugs and this condition indeed shows synergistic benefits (Figure 5.9). The high-throughput automated screen was, therefore, able to clearly identify this synergy as a statistically significant increase of mean healthspan of 21% over the 0.1% DMSO negative control group.

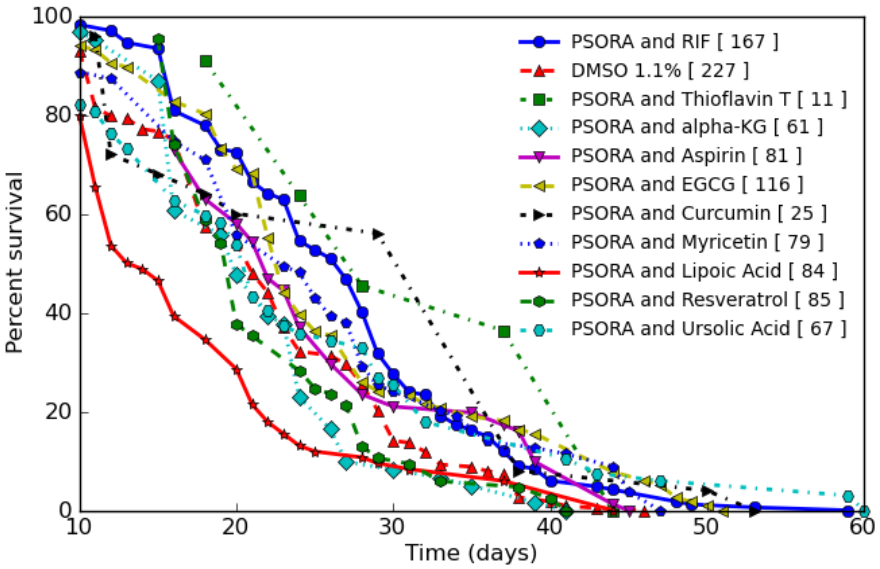

Figure 5.9 – Survival plot of psora-4 paired with DMSO-soluble drug.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population. RIF = rifampicin.

Following the same conventions established for Table 5.1 in the previous section, Table 5.2 summarizes the results of the psora-4 drug pairs with some additional considerations. The results of the comparison with the negative group, in column "vs DMSO", are values obtained using the respective DMSO control concentration (1% or 1.1%). It is also important to be mindful that psora-4 was assayed at 1% DMSO and, therefore, effects observed for psora-4 and its pairs are not fully comparable with drug pairs that were tested at 0.1% DMSO (see previous section). Similarly, comparisons between each drug pairs and respective monotherapy of the drug being paired with psora-4 (column "vs Monotherapy"), are not standardized for final DMSO concentration, as all the monotherapies were assayed at the lower concentration of 0.1% DMSO (matching the final concentration of DMSO used in the rifampicin drug pairs). Finally, there are now three conditions with red-colored names. These represent truly toxic interventions. That is interventions that decrease healthspan relative

to the DMSO (what % - given that we are now considering 0.1, 1 and 1.1.% ;)) control group. Adding NAC, icariin or lipoic acid to psora-4 resulted in intervention that were healthspan decreasing relative to untreated control (DMSO only). These conditions are therefore directly toxic to animals, even though each of the individual compounds has beneficial effects on healthspan.

Drug Pairs [N]	Mean Healthspan (days)	Relative mean healthspan extension (%)		
		vs DMSO	vs PSORA	vs Monotherapy
PSORA and Thioflavin T [11]	**33**	**48**	**44**	**51**
PSORA and Curcumin [25]	**29**	**32**	**27**	**17**
PSORA and EGCG [116]	26	**18**	**14**	**36**
PSORA and Myricetin [79]	25	**14**	**10**	0
PSORA and Aspirin [81]	25	**13**	9	0
PSORA and Ursolic Acid [67]	24	**10**	6	**12**
PSORA and Resveratrol [85]	22	0	-4	**9**
PSORA and alpha-KG [61]	21	-3	-7	1
PSORA and Piceatannol [113]	21	4	**-7**	**-14**
PSORA and Captopril [85]	21	2	**-8**	**-16**
PSORA and Lithium [99]	20	-3	**-13**	-8
PSORA and Spermidine [111]	19	-8	**-17**	**-20**
PSORA and NAC [104]	18	**-11**	**-20**	**-29**
PSORA and Icariin [87]	18	**-14**	**-23**	**-19**
PSORA and Lipoic Acid [84]	17	**-21**	**-23**	**-23**

Table 5.2 - Summary of the longevity effects of psora-4 drug pair combinations. The first column identifies the drug pair and sample size (value inside []). **Green bold** drug pairs are cases of positive synergies. **Red bold** drug pair names are toxic interactions. The last three column display the magnitude of healthspan extension relative to the DMSO, rifampicin and paired drug interventions, respectively. **Bold** values represent a statistically significant log-rank test. RIF = rifampicin.

When piceatannol, captopril, lithium or spermidine were added to psora-4, healthspan benefits resulting from exposure to each of the single drugs were lost, resulting in cohort healthspans that were not statistically significantly different from the respective (1% or 1.1%) DMSO control group (Table 5.2). In other words, in the case of these drug pairs, the effects of the individual drugs appear to cancel each other.

As it happened with the examined rifampicin drug pairs, there is a subset of psora-4 pairs that do not extend healthspan beyond the best of the single drugs that composes it (Table 5.2). Interventions in which psora-4 is being combined with myricetin, aspirin, ursolic acid, resveratrol, and alpha-ketoglutarate, do not elicit healthspan benefits that exceed those of the individual.

More encouraging, 3 drugs resulted in significant additional benefits when paired with psora-4. These combinations increase mean healthspan, in a statistically significant manner, relative to negative control (DMSO only) and when compared to either of the two monotherapies (Table 5.2).

Thioflavin-T and psora-4 were one such synergistic healthspan-extending pair. However, for reasons that are not clear, sample size for this condition was unusually small, with lifespan data for only 11 individual nematodes being recorded. Nonetheless, the healthspan extension achieved is of such magnitude (larger than the sum of the two individual drug interventions) and longitudinal consistency (Figure 5.10) that it is one of the most exciting results. It is worth to bear in mind that, albeit allowing quite granular survival data, the nature of the purposed high-throughput screen is as an initial screen to reliable detect promising healthspan-extending interventions. The successful candidate combinations ought to be subsequentially assayed in standard conditions[294].

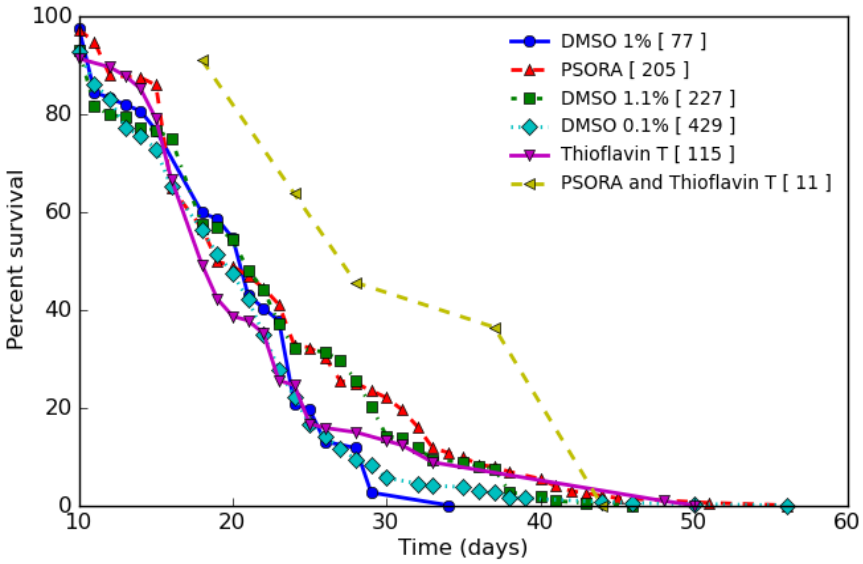

Figure 5.10 – Healthspan of psora-4 and/or thioflavin-T.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population.

Another synergistic drug combination was the combination of psora-4 with curcumin. This is the other rare case of small sample size (Figure 5.11). The combination of psora-4 and curcumin elicits no apparent healthspan-extension until about 40% of the population stopped moving (day 20). From then onwards, the healthy survivors saw their remaining healthspan extended significantly. This intervention too, exhibits a mean healthspan extension larger than the positive control synergy of rifampicin and psora-4.

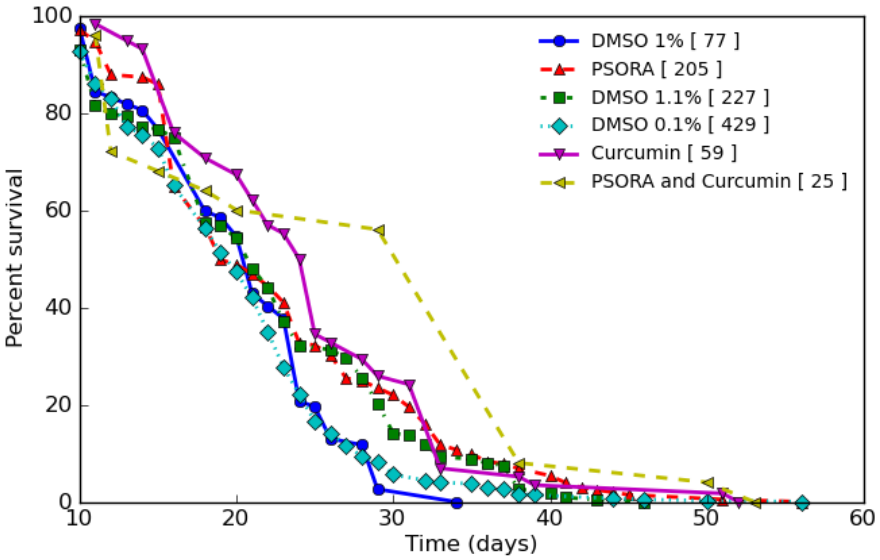

Figure 5.11 – Healthspan of psora-4 and/or curcumin.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population.

The final new drug synergy that I discovered was psora-4 with EGCG. The survival curve of this synergy indicates that at in the early time points, up to day 15, this combination achieves about as much healthspan-extension as the best of its single drug interventions (psora-4). From then until day 50, the group on this synergistic pair of drugs displays significantly better healthspan than any other drug in this trail (Figure 5.12). Notice that EGCG, when by itself, was ineffective in extending healthspan under these conditions. This, therefore, is another example of a combination that was synergistic under the most stringent definition that the combination treatment results in benefits that were larger than the sum of the individual benefits.

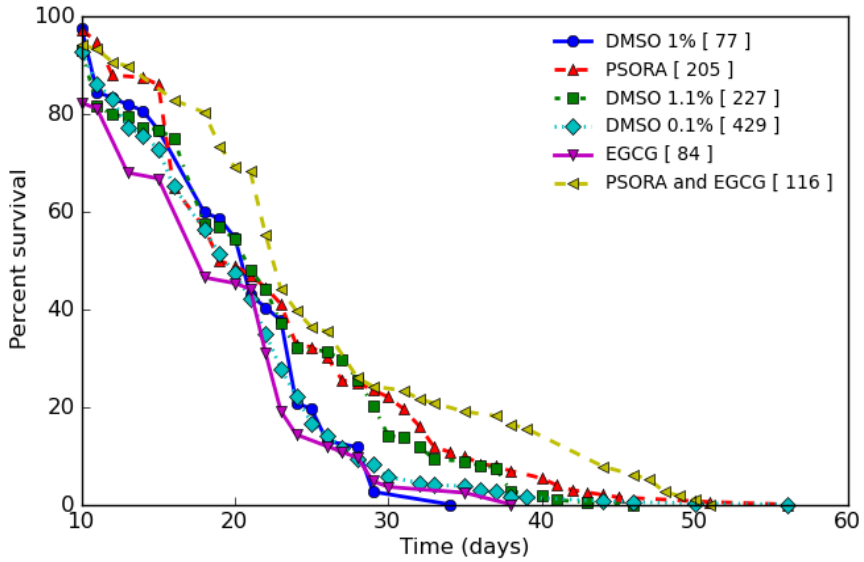

Figure 5.12 – Healthspan of psora-4 and/or EGCG.
Numbers inside square brackets indicate the sample size. Percent survival represent the percentage of moving worms per time point relative to the starting population.

5.4 - Discussion

This chapter exemplifies the capabilities of my methodology and data-processing pipeline. Overall, I tested 69 different conditions, including testing 18 single drugs, 48 drug combinations and 3 distinct negative control conditions. The total number of animals for which I obtained lifespan data was approximately 5985 (this number does not include the dropped conditions that featured rapamycin), while a typical manual lifespan experiment rarely exceeds 400 animals. Leveraging the system I developed, I was able to conduct what to the best of my knowledge is the largest combinatorial anti-aging drug screen to date. In this single screen, I exhaustively crossed all my compounds with the 3 compounds previously known to be involved in synergistic combinations[70].

I discovered 6 novel anti-aging drug synergies. Prior to this screen, there were only a handful of know synergistic interaction.

The scalability capabilities of the proposed screen allow a greater standardization among the experimental conditions. In other words, these drugs were assayed simultaneously and with specifically designed control groups if necessary. This allowed me for example to confidently use the log-rank test between single and dual drug interventions (namely in the rifampicin-based drug pairs) to obtain a more formal definition of synergy[a].

The statistical power of the proposed methodology is enhanced in two ways. First, the for a high-throughput screen the proposed methodology matches the temporal granularity (daily measures) used in the standard manual lifespan assays[295]. This has only been achieve recently[27], with a few measurements still being the norm in *C. elegans* drug screens (e.g.[[221]]). The second characteristic that increases statistical power is the considerable number of individuals per experimental condition. Indeed, the average sample size in this screen is characteristic of manual lifespan assays. Case in point, it is even slightly larger than our previously manual-based screen[70]. In total, the screens presented here generated almost 200000 frames, tracking 1632 individual plate wells for over 60 days. These advances were only possible due to the use of automated image processing and lifespan scoring.

Regarding our previous work, among the discovered synergistic lifespan-extending drug pairs there was the rifampicin and psora-4 combination. I used this known synergy as a positive control for synergistic interactions, and as a byproduct, obtained its healthspan. Our work reported a mean lifespan between

[a] I considered synergistic a drug pair that exhibited larger restricted mean healthspan larger and statistically significant log-rank test p-value that the respective control and monotherapies interventions.

28 to 31 days for this intervention, and my current work reveals a mean healthspan of about 27 days. This are encouraging results, suggesting that the almost the totality of the lifespan of worms treated with these pairs of drugs was spent in good health. This is consistent with our previous results on healthspan effects of synergistic drug combinations, both in models of neurodegeneration[126] and in wild-type animals[70]. The magnitude of healthspan-extension is comparable to that reported in the ABC scoring healthspan assays that we have previously reported in our triple synergistic combinations[70].

Among the discovered new anti-aging drug synergies some of them display impressive effect size. For example, the combination of rifampicin and lithium extends mean healthspan by 37%, compared to untreated control group. This mean locomotion-based healthspan of 28 days is at least as large as the one observed in 4 canonical long-lived mutant strains[296]. Another example, the new candidate drug synergy of rifampicin and thioflavin-T displays a 90^{th} percentile of healthspan (the healthspan-based analogous to the 90^{th} percentile of lifespan usually defining maximum lifespan) of approximately 58 days. This sets this the drug pair intervention with the largest maximum lifespan (because the measured healthspan implies a lifespan at least as large) ever reported in *C. elegans,* being surpassed only by our previously discovered triple drug combinations[70]. However, these results need to be confirmed using more traditional, manual lifespan assays.

Furthermore, it is worth noting that beneficial synergies, in particular synergies in the strictest sense (with combination treatment resulting in healthspan benefits that exceed the sum of the benefits from the two individual drugs) are rare. Of the 30 new pairs tested, only 5 showed (strict) synergistic benefits and

only one was narrowly synergistic. By contrast, in 21 cases combining two compounds resulted in diminished or null benefits and 3 drug pairs resulted in toxicity. This is an important observation as it suggests that simply combining compounds, drugs and supplements that are beneficial individually is unlikely to result in beneficial combinations and may result in diminishing returns or toxic interactions.

Chapter 6 – Transcriptomics and Synergistic Potential

6.1 – Introduction

Throughout my attempts of rationalizing, explaining, and predicting drug-drug interactions (additive or detrimental interactions, drug-drug synergies), I only make use on information about the transcriptional effects or published mode of action of the single compounds. This is because, due to the problem of combinatorial explosion in search space, systematically generating DEG epistasis data on drug combinations is as infeasible (at least given current technological limitations) as generating lifespan data for every possible drug pair. Tools for prediction of drug-drug interactions, therefore, by necessity attempt to infer interactions between drugs based on information derived from known effects of single drugs (on transcription, dependence on key pathways). This implicitly or explicitly requires the assumption that, at least to first approximation, perturbations involving a pair of drugs can be modelled as linear super-position of the individual drug effects. For example, considering two drugs (Drug A and Drug B) with vectors of DEGs (A and B), I will assume that the DEG for the drug pair (AB) can be approximated by:

$$AB \sim \alpha * A + \beta * B$$

Where alpha and beta are real-valued factors allowing for linear interactions between genes (e.g. saturation effects or changes in effective *in vivo* drug concentration due to global effects on drug detoxification and transport pathways). This linearity assumption is a strong simplification of how complex biological system works. In fact, we know that in many specific cases, especially where pathways intersect, the assumption is explicitly violated.

However, a more complete analysis of drug-drug interaction would require construction of an explicit gene-regulatory network (GRN) of aging genes, which is well beyond the scope of this study. It is therefore worth exploring to what extent, if any, drug-drug interactions can be understood and/or predicted without the ability to explicitly model the GRN determining aging. At the very least, this will allow me to evaluate to what extend non-linear interactions drive drug synergies. Our previous work shows that drugs that are separated in a transcriptional PCA space were more likely to interact synergistic[70]. This observation led us to the assumption that compounds that are dissimilar in terms of their mode of action may be more likely to interact additively or synergistically. However, given the small size of all previous screens (typically less than 10 drug pairs) and the limited number of known drug-drug synergies affecting aging (9 if one defines synergy as more than the best of the single drugs)[37,38,70], none of these observations could be supported statistically, making them hypotheses based on anecdotal evidence, at best.

Interestingly, these observations were based on simple PCAs, and PCA by design, is a linear technique (see sub-chapter 0), that is unable to model non-linear effects. This reasoning suggested that the linearity assumption of drug-drug interaction effects might preserve some predicative power and further encouraged me to explore simple gene-set based metrics.

Taking into consideration all of the above, in this chapter I will jointly analyze my RNA-Seq dataset (for single drugs) with the one generated in our previous work (which contains a limited number of examples where we collected data on both the single drugs (A and B) and their interactions (AB)[70]. I will complement that information with the healthspan results obtained in my high-throughput

assay (Chapter 5) as well as some of the available literature data (e.g. 70) to explore the nature of synergy from a transcriptomic perspective.

6.2 – GenAge genes expression and healthspan

First, I combined the RNA-Seq samples that I generated with the samples from our previous work[70]. In other words, my working dataset for this chapter comprises the aligned samples (FastQ format) of section 4.2.1 plus the aligned samples[a] of our previous work. As a result, from the initial transcriptome of 21923 genes, I found that 13862 genes pass the significance threshold, and can be considered what I call "druggable genes" by our drug library. I am using the term "druggable", in the context of my experiment, to refer to genes for which there are significant transcriptional changes for at least one condition (and that cannot be explained by batch effect). In other words, I consider genes "druggable" if I have evidence that they can be impacted transcriptionally by at least one drug intervention. I then mapped my druggable genes to the GenAge (see section 1.3.4) gene for *C. elegans*. After converting IDs and data cleaning, this procedure resulted in a final gene set of 407 druggable aging genes, that is, genes that are impacted by at least one of my drugs and that are experimentally validated to be involved in lifespan determination in *C. elegans*.

I hypothesised that the expression changes of GenAge genes should be correlated with the mean healthspan changes for a given drug intervention. To

[a] to minimize inconsistence between datasets, I realigned the raw (FastQ) samples from our previous work using my custom-made pipeline (see section 2.4.2).

test this, I built several ways to count gene and score expression changes within GenAge genes. I will describe one of the more complex ones in detail and summarize observations for the complete set.

To determine predicted lifespan benefits of "correct" GenAge DEGs (see section 4.2.2), I extracted all DEGs of GenAge genes that significantly change in the correct direction (expected to extend lifespan). I consider these correct changes positive (+1). Depending on the exact scoring approach, I weighted these correct genes by their fold-change (emphasizing numerically larger changes in the right direction) and/or by their known max effect size (emphasizing genes with larger impact on lifespan – e.g. upon knockout). In a way analogous to the definition of correctly modulated DEGs, I will also account for the lifespan-decreasing DEGs, scoring changes within GenAge that were known to shorten lifespan. Finally, DEGs that belong to GenAge but went into the opposite direction predicted to influence lifespan were set to zero. The subset of lifespan-decreasing GenAge DEGs includes anti-longevity lifespan-decreasing genes with increased expression upon drug treatment and pro-longevity lifespan-decreasing genes that are down-regulated. My goal in constructing this scoring function was to compute a sum of all the lifespan-extending and lifespan-decreasing DEGs (either unweighted, or weighted by fold change and/or known effect size) for each of the single drugs. By relating this GenAge score to the observed lifespan effects for each of the monotherapies, I was hoping to identify a scoring function that might then be explored further in the prediction of drug-drug interactions. The importance of being able to do a weighted sum based on known effect size can be made evident, for example, by the fact that there are anti-longevity genes that when

knockout increase lifespan by only 10% while others increase lifespan by up to 100%. Conceptually, I am adding the potential lifespan-extending effects of a drugs and subtracting its potential lifespan-decreasing effects:

$$estimated\ effect\ size = \sum lifeExtGenes - \sum lifeDecrGenes$$

For this more complex example (among the literally tenths tested[a]), I weight the $\log_2$ foldchanges of a gene with the respective genetic intervention effect size reported in GenAge. Additionally, I calibrated and scaled the effect size by considering -1 $\log_2$ foldchange to be equivalent of a RNAi knockdowns condition while I considered a -2.5 $\log_2$ foldchanges as knockouts and assumed that genetic overexpression experiments would result at minimum in 2 $\log_2$ foldchanges. I also capped these values so that at ± 2.5 $\log_2$ foldchanges, the maximum effect size was achieved.

As is evident in Figure 6.1, there was no correlation between the DEG based single-drug GenAge score and the observed lifespan effects for this scoring function. It might be argued that this could be because this particular score is calculated as the difference between to scores (the sums over beneficial and detrimental impacts). Since both scores are subject to significant error, it could be argued that the difference between them, being, by design smaller than either factor but subject to the sum of both error terms, loses predictive power. However, I exhaustively explored variations on this theme, for example only considering beneficial effects, reducing the score to binary counts (ignoring fold

[a] For example, I tried all the combinations: with or without weighing gene effect sizes; instead of a sum using the ratio between lifespan-extending and lifespan-decreasing genes; instead of the total sum use robust measures like the median value; disregard the magnitude of fold changes and considered just if it is relevant (binary); etc.

changes and/or known effect sizes) and even testing an “impact score” by counted all changes to GenAge genes as signs of effects on aging, even if changes were “incorrect” (predicted to shorten lifespan). Disappointingly, none of the scoring function displayed any significant correlation with mean healthspan changes as observed experimentally. In other words, regardless of the scoring function used, the number, degree, or identity of GenAge genes affected was not predictive of observed lifespan effects. Drugs that affected more GenAge genes, affected them more severely or more correctly (or were free of detrimental penalties) were not more likely to robustly extend lifespan than drugs that resulted in less pronounced or less “correct” changes to GenAge genes. This, of course, poses a significant challenge to any attempt to use similar scoring functions for the prediction of drug-drug interactions. Given that even single drugs, for which DEGs are accurately reflect transcriptional effects of each intervention, cannot be predicted by any of the scoring function quantifying impact on the GenAge set, attempts to predict drug-drug interactions, that would have to be based on a linear (additive) approximation of their combined transcriptional effect would have to be based on a different rational.

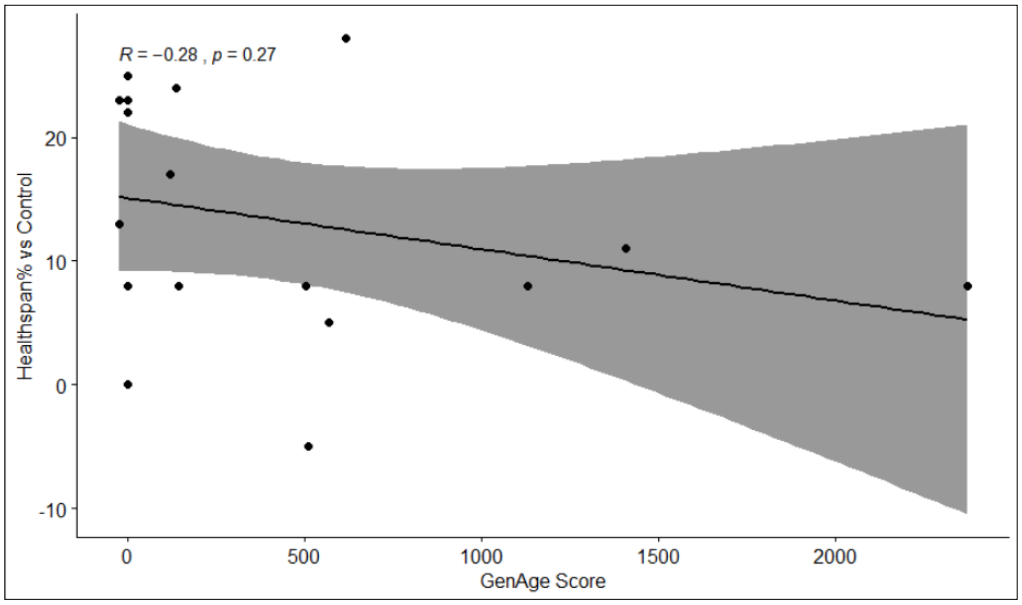

Figure 6.1 – Linear correlation plot between the GenAge Score and relative healthspan extension.
The healthspan extension is in relation to the DMSO negative control condition. The correlation is not statistically significant.

6.2.1 - Drug similarity in GenAge does not predict drug synergy

To test this intuition, I revisited the score based on the modulation of lifespan-extending and lifespan-decreasing GenAge genes (sub-chapter 6.2), but, as expected, found that the same approaches that failed to predict individual drug effects was also unable to predict drug synergistic interactions (for an example see Figure 6.2).

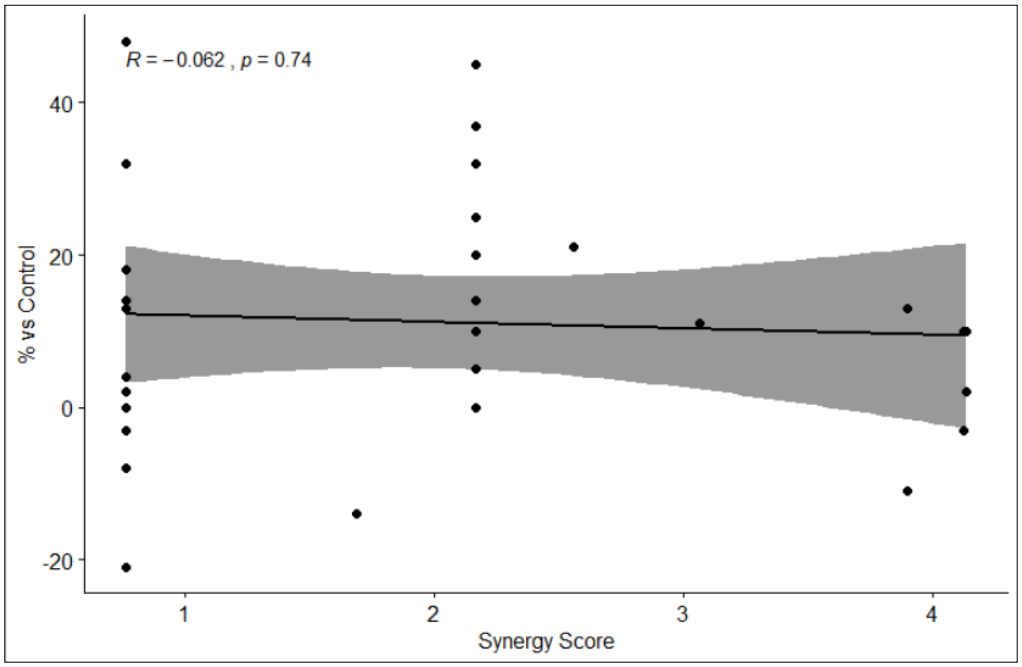

Figure 6.2 - Linear correlation plot between the Synergy Score and relative healthspan extension.
The synergy score is larger the stronger the modulation of lifespan-extending GenAge DEGs and the weaker the targeting of lifespan-decreasing GenAge DEGs. The two clear clusters of aligned points correspond to combinations in which psora-4 and rifampicin were much stronger than the paired drugs in modulating GenAge genes. The "Synergy Score" is uncorrelated with the magnitude of mean healthspan extension of the drug combinations relative to the negative control group.

6.2.2– Monotherapies transcriptional similarity

Considering this lack of predictive power of the GenAge gene set, I next explored if our previous hypothesis, that more dissimilar drugs are more likely to interact beneficially, could be confirmed quantitatively in my dataset. I previously mentioned the fact that there is a significant bias in the literature, regarding evidence on epistasis of lifespan effects, with most anti-aging drugs assayed under hypothesis connected to a limited number of popular aging-related genes and pathways (see section 2.2.21). This means that these data are of limited use for the characterization of mode action because there is only an exceedingly small number of genes for which epistasis information is available

for more than a few drugs in my dataset (see Table 2.1). I therefore decided that it would be valuable to conduct the analysis in a hypothesis-free fashion. To this end, I constructed two different drug similarity scores, intended to compare the degree of similarity between the mode of action of drugs based on the impact on the transcriptome.

The most obvious way to compare mode of action, as pertaining to aging, might be to compare the set of GenAge genes impacted by each drug. I counted the amount of GenAge DEGs[a] in common between each two drug transcriptional profiles of my screened drug combinations. Then I correlated this amount to the mean healthspan extension displayed by the respective drug pair interventions. The rationale is that if dissimilar drugs are more likely to be synergistic, then drugs with more shared genes target the known aging-related genes in a similar fashion and therefore will tend to not be synergistic. Figure 6.3 displays the results of this approach. The correlation between this measure of drug-drug similarity and effect size of drug pairs was not significant.

[a] I used the DEGs specific to each drug, from the RNA-Seq pipeline explained in detail in section 6.2.3.

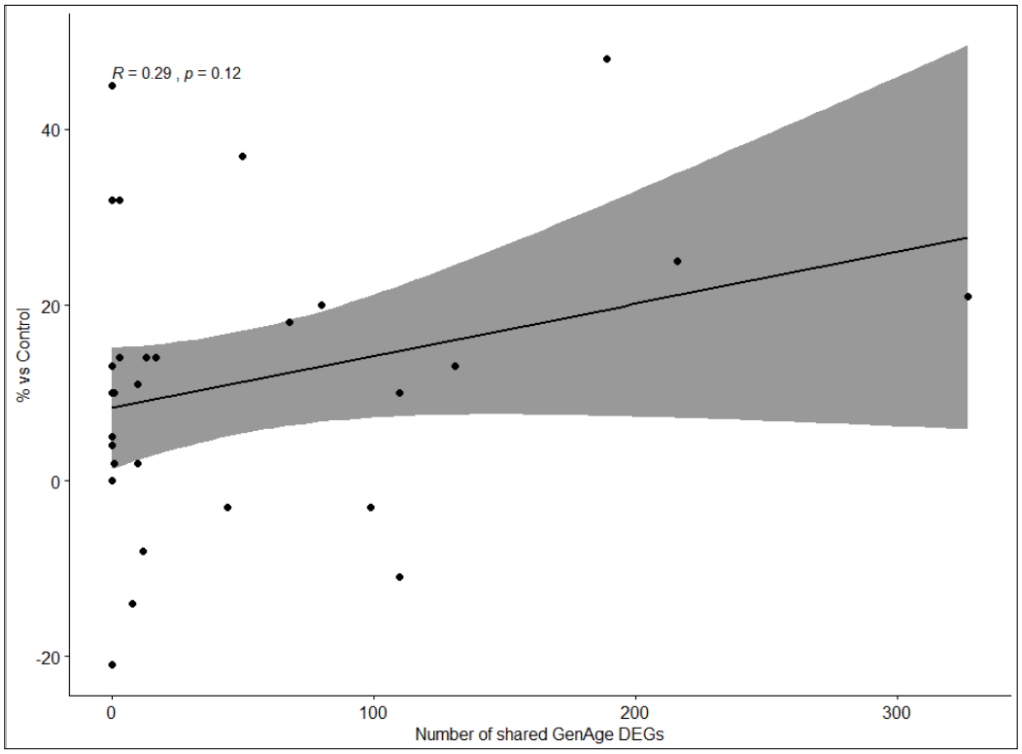

Figure 6.3 - Linear correlation plot between the number of shared GenAge DEGs and relative healthspan extension.
The healthspan extension is in relation to the DMSO negative control condition. The correlation is not statistically significant.

Given the failure of the GenAge based impacted scores to predict single drug lifespan benefits, it is not self-evident that the GenAge gene set is the complete or sufficient to explain aging-related effects in general. I therefore next constructed considered using if an alternative aging-related gene set[66] or even not filtering by gene set would be better.

6.2.3– PCA-based decomposition of transcriptional profiles

Following the failure of similarity comparison bases on simple gene-set analysis (within GenAge) to predict beneficial interactions, I next attempted to construct a more informative way to classify similarity and differences in mode of action between drugs. For this purpose, I constructed an algorithm based on using

principal component analysis (PCA) to express my drug transcriptional profiles in a much lower number of dimensions (see sub-chapter 0).

Intrinsically, I am modelling a very noisy gene expression dataset due to the lack of a more stringent pre-filtering. This is a fair observation, as I had pre-filtered based on the results of the likelihood ratio test. This test defined as DEG, any gene that changes in at least one condition more than what can be attributed to the batch effect. Since I am comparing drug profiles, I needed to move to an approach that considers DEG at the drug level (instead of at the drug set level), and identifies which genes are differentially expressed by each specific drug. This required a change of the RNA-Seq analysis workflow. More precisely, I will still use the same samples as before (where I combined the single drugs samples of our previous work[70] and my dataset) and analyze it in the same way, up to the point of testing for DEGs. The change happens in this step, as I will use the Wald test for the GLM coefficients instead. This is the same test used in a previous chapter (section 4.2.1) to obtain the drug-specific DEGs to characterize the drugs in my dataset. Since I have 60 samples over 21 conditions, based on the heuristic by Lamarre, S. *et al.*[212], I set the significance threshold of the adjusted p-value at approximately 0.138^{a}. For each set of drugs considered in the analysis that follow, I kept only the gene expression values of DEG genes, and the remaining values were set to zero.

Principal Component Analysis (PCA) can be used to explore the similarity between drug profiles in a hypothesis-free way (see section 1.4.1). The foldchanges of druggable genes from the drug conditions were used to build a

[a] Precisely, it was set at $2^{-60/21}$.

PCA. This allowed me to reduce the dimensionality of the dataset to a n-dimensional space, with n corresponding to the number of drugs gene expression profiles. Furthermore, the linear combination of gene expressions that constitute each of the PCs was used to quantify the novelty of a new drug expression signature. In more detail, these PCs were used as the input of a linear regression model to predict the gene expression signature of a drug that was not part of the n initial drugs. I used as proxy for the novelty of the unseen drug gene expression profile the amount of variance explained by the linear regression as quantified by R-squared[a].

From a biological standpoint, I am assuming that there are only a finite number of independent ways in which healthspan can be extended by known pharmaceutical interventions and that a drug working through a completely new mode of action would result in a gene expression profile which would remain largely unexplained by the known linear combination of existing modes.

I formalized the concept of the "least informative drug" – for a set of n drug gene expression signatures, the least informative $drug_i$ in the set is the drug with index i, which maximizes the R-squared value of explained variance of the linear regression:

$$drug_i = \sum_{n-1} PCs$$

with PCs representing the principal components resulting from the $n-1$ drugs gene expression signatures PCA. I then calculated the least informative drug using a backward pass procedure starting from the all the possible sets of $n-1$

[a] I did not use the adjusted R-squared because it can take negative values.

cardinality. At each iteration in the backward pass the least informative drug of the initial set was removed, and the next least informative drug was identified. I called “PCA saturation plot”, to the plot that displays the results of my procedure. Assuming a minimum set size of 2 (the required minimum to apply the PCA technique), the PCA saturation plot displays on the x-axis the number of terms used in the PCs linear regression, and on the y-axis the explained variance (on the scale from 0 to 1) of the least informative drug found for the drug gene set of that size.

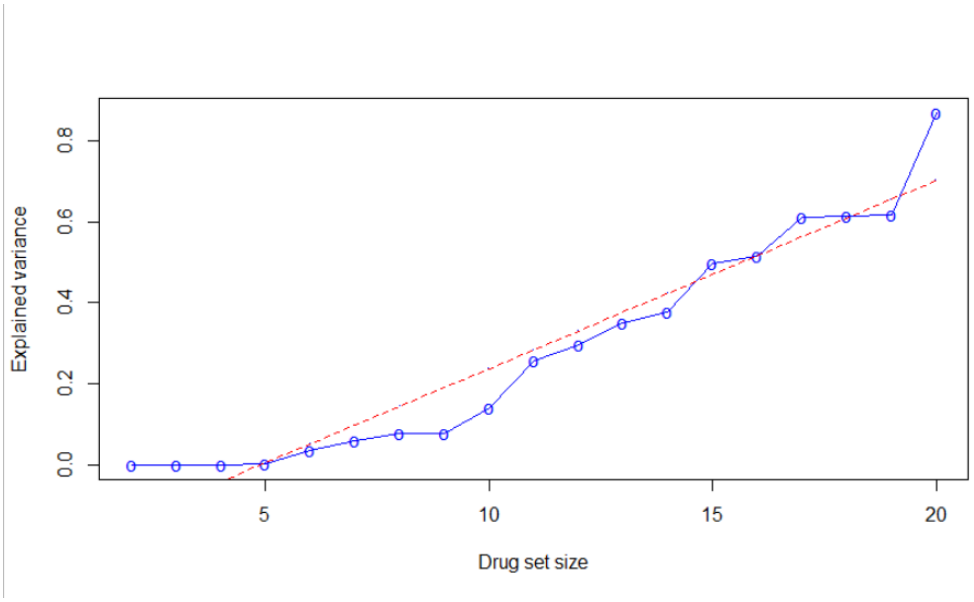

Figure 6.4 – PCA saturation plot for the joint set of 20 drug transcriptional signatures.
The horizontal axis displays the drug set size with the added unseen drug. The vertical axis displays the relative amount of variance of the new drug’s gene expression changes that can be decomposed by as linear combination of the existing drug transcriptional signatures. The red line is the line with the best linear.

Then the *REAT* package (v3.0.2)[297] was used to check if the type of relationship between drug set size and explained variance of the least informative drug. The best curve fit was of linear nature[a] with an adjusted R-squared of 0.93.

Essentially, this approach attempts to express the transcriptional changes of every new drug (in terms of the sequence with which they are added to the set) as linear combinations of the DEG vectors of "known" drugs (drugs already part of the set).

Since the maximum combined explanation power of the existing set of drug transcriptomic signature to explain an unseen drug expression only reaches 87% (Figure 6.1), I modeled this relationship as a linear regression and calculated that for a 26-dimensional gene expression backward pass the least informative drug would be totally explained by the rest of the drugs, that is, the PCA gene expression drug space would reach saturation.

Because I do not have more drugs, I cannot create a saturated reduced-space to work with by adding extra transcriptomic drug profiles. Therefore, I took the approach of further reducing the transcriptomic drug space to a smaller gene set relevant to my paradigm and repeat the backward pass procedure. I used two aging-related gene sets, the gene set of *C. elegans* present in GenAge[237], and the MetaWorm[66] gene set. The results were robust to gene set choice (both PCPA plots reach saturation), so for brevity I will explain the procedure using just GenAge data.

From GenAge (build 20) GenAge[237], I imported 872 unique genes. Among these 13 genes had evidence of being both anti-aging and pro-aging genes. I

[a] compared with power, logistic, and exponential growth fits.

removed these 13 “controversial” genes from the gene set, and after further cleaning the dataset, I ended up with 754 genes (85% of the initial GenAge worm genes).

In contrast to the use of the entire transcriptome or all the DEGs, for the GenAge DEG gene set the gene expression space is practically saturated, with the first least informative drug of the backward pass having 99.9% of its variance explained by the PCs resulting from the gene signatures of the other drugs.

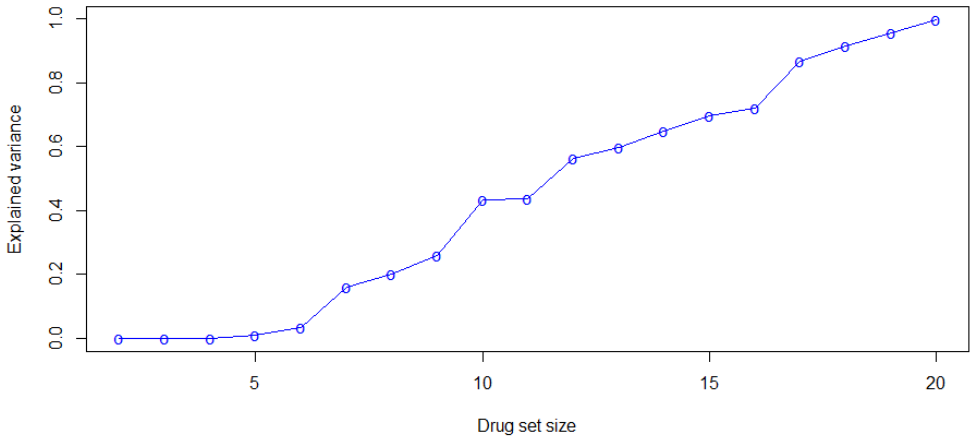

Figure 6.5 - PCA saturation plot for the joint set of 20 drug transcriptional signatures in the GenAge gene set.
The horizontal axis displays the drug set size with the added unseen drug. The vertical axis displays the relative amount of variance of the new drug’s gene expression changes that can be decomposed by as linear combination of the existing drug transcriptional signatures.

Proceeding in the same as before, the best curve fit was of linear nature with an adjusted R-squared of 0.98. Furthermore, the linear modelling of this relationship predicts that indeed such as in my case a drug gene expression set of 19 drugs are sufficient to saturate the reduced-dimension transcriptional space. This means that for gene expression changes impacting only GenAge

genes, there is only a limited subset of genes that are impacted, and this subspace can be completely constructed using fewer drugs as base than are in the entire dataset. In other words, DEGs of the last drug can be almost completely expressed as linear combination of all previous drugs. This is true although the total set of GenAge genes impacted by the set of drugs is composed by 652 genes, which is substantially larger than the number of drugs (20) in the set. One way to interpret this observation is that there is only a limited number of mechanisms explored collectively by the drug set as a whole.

6.2.4 – Hierarchical clustering of drugs transcriptional profile

In the previous chapter, I showed that, restricted to the gene set of GenAge DEGs and given enough drug profiles (which I did obtain), the transcriptomic signature of a pharmaceutical intervention can be modelled by a linear combination of other pharmaceutical interventions. This suggests the notion of a drug-drug similarity. The set of n vectors spanning the transcriptional space explored collectively by all n drugs in the set can be considered a base system for the effect of each drug. Each individual drug then has coordinates within that base set. Below is one way to illustrate this idea (Figure 6.6). Drugs more similar in mode of action are closer to each other in this highly reduced (18 dimensional) space.

To display the relatedness between all the drugs in my dataset, I reduced the previous sparse matrix to the drugs that I assayed in my healthspan screen and built a PCA. Then I took the coordinates of each drug in the PCA

transcriptional-space and used the *optCluster* R package (v1.3.0) to build the an heatmap based on optimal[a] hierarchical clustering.

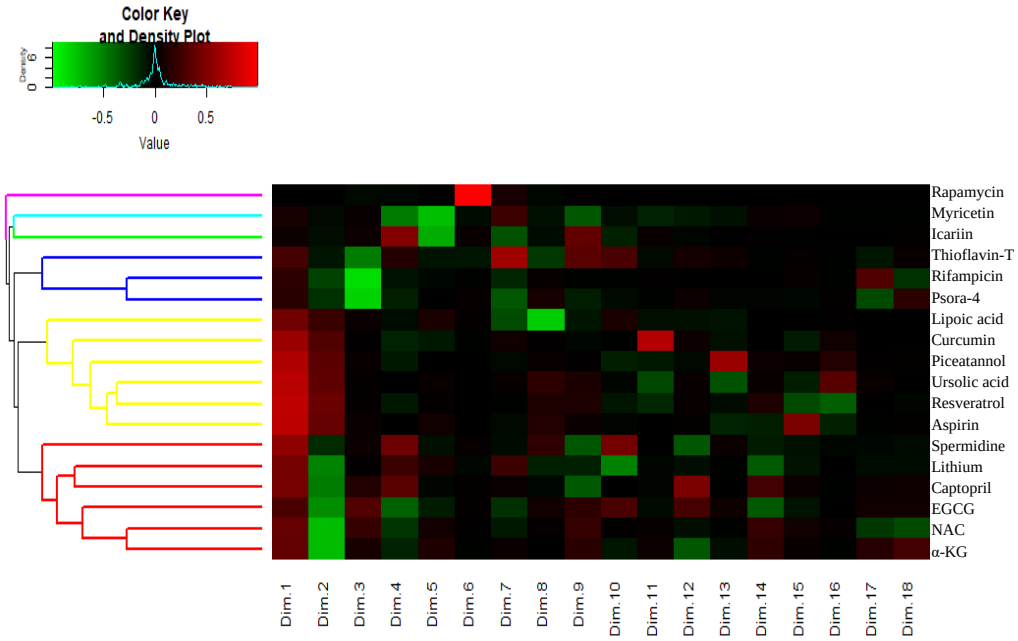

Figure 6.6 – Heatmap and hierarchical clustering of drugs based on their PC coordinates.
The columns represent the principal components of the PCA of drugs DEGs. Each row corresponds to a drug. The dendrogram on the left is the result of optimal aggregated clustering and it separates the drugs into 6 clusters (one of each color).

After testing all the available methods and for 2 to 8 possible number of existing clusters, the highest quality clustering separated by PCA transcriptional space into 6 parts. Rapamycin, myricetin and icariin are unique, and then 3 clusters of drugs can be observed.

Once each drug effect has been mapped into this 19 dimensional space, their similarities can be evaluated using a simple clustering algorithm. Interestingly

[a] The *optCluster* package uses different hierarchical clustering algorithms and tries a range of values for the number of clusters, and then choses the optimal clustering based on the rank aggregation performance across several clustering metrics.

As can be seen, rapamycin almost completely defines the mode of action along dimension 6 (Figure 6.6). Interestingly, no other drug has a significant weight (coordinate) along this axis. Another interesting observation is that several polyphenols and antioxidants cluster closely with each other, suggesting that the base system indeed captures some of the relevant biology.

6.2.5– PCA-based independence does not predict synergy

My PCA-based approach was able to cluster drugs by similarity in terms of transcriptional impact, using a system that appeared to capture key aspects of mode of action. Since my working hypothesis was that drugs with distinct mode of action might be more likely to interact synergistically, I next used this PCA-based approach to systematically evaluate pairwise similarity and tested this score as an alternative synergy prediction criterion.

Under my paradigm, the notion of drug independence is analogous to compute the cosine similarity between drug gene expression profiles in the PCA-derived space. While this vector space in my case is 19 Dimensional, the concept of cosine similarity is analogous to the case of vectors in ordinary 2D or 3D space. Two very similar drugs have coordinate vectors that are nearly parallel in this space, resulting a small angle between the coordinate vectors and therefore returning a cosine value of close to 1. Two vectors (gene expression profiles) that are orthogonal (completely independent from each other) return a cosine similarity of zero[a]. The cosine similarity therefore varies between 1 (identical

[a] The cosine similarity is directly applicable to very high-dimensional spaces although it is not the best metric (and improvements have been suggested e.g.[315]). In my case, the coordinates vector is created based on PCA, and therefore, I am not in a high-dimensional setting.

gene expression profile) and -1 (opposite gene expression profile); with a value of 0 indicating that the drugs profiles are orthogonal or decorrelated (independent).

Using this definition, the cosine similarity between a drug A and drug B, is given by:

$$cosine\ similarity\ = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

where i denotes the features of each drug profile. In our case, these might represent gene expression abundance values or PC coordinates.

The cosine similarity is a special case of a weighted cosine similarity in which all the features i (coordinate axis/principal components) are equally weighted[a]. It is often useful to weight features according to their relevancy to the task in hand, and in this case the weighted cosine similarity simply adds a feature-wise weight w_i, that represents percentage of variance explained by PC_i, as follows:

$$weighted\ cosine\ similarity\ = \frac{\sum_{i=1}^{n} w_i A_i B_i}{\sqrt{\sum_{i=1}^{n} w_i A_i^2} \sqrt{\sum_{i=1}^{n} w_i B_i^2}}$$

In my case, for standardization purposes and before computing a weighted cosine similarity, I will always normalize the w_i vector so that its elements add up to one[b].

[a] weighting scheme based on the one in the widely used SciPy[316].
[b] The total explained variance by the n principal components of a set of n drugs transcriptional profiles would add to 1, unless due to floating point imprecision.

There will be times that I am interested in the statistical testing of orthogonality[a]. I will feature engineer[b] the cosine similarities previously mentioned into a decorrelation score using the formula:

$$decorrelation\ score = 1 - abs(x)$$

with x being a cosine similarity value and abs representing the absolute value function. The decorrelation score achieves its minimum of 0 when the cosine similarity is 1 or -1, as these values are the farthest away from being orthogonal.

Using the low-dimensional gene expression space of the GenAge set of druggable genes I attempted the creation of a synergistic drug-drug interaction measure or metric. The choice to restrict my analysis to an aging-related gene set is consequence of aging being an extraordinarily complex process without a well-defined module. In other words, in *praxis*, the GenAge set of druggable genes represents my disease module[c].

As a benchmark, I calculated the cosine similarity and decorrelation score directly from the gene expression values between all possible drug pairs for which I have the healthspan effects (obtained from my high-throughput screen in Chapter 5). Subsequently, I tested the linear correlation between the cosine similarity or decorrelation score and two measures of the magnitude of healthspan. The magnitude of healthspan effects can be calculated relative to the control group or relative to the best effect among the two single drugs. The later corresponds to the higher single activity model[298] definition of synergy that

[a] I can promptly test if a correlation is statistically significant, what is not available is a test to detect statistical orthogonality or decorrelation.
[b] this is just a heuristic, and I claim no statistical properties.
[c] with the previous chapter showing how it was found to be the best candidate for this role.

used in our previous work[232], and I should henceforth refer to it as the "synergy score". When referring to the former I will simply call it "healthspan score".

I proceeded by repeating the DEG pre-filtering procedure, based on the Wald test variant of my pipeline[a], on the set of 17 drugs for which I have healthspan measures[b], and restricted to the GenAge set of genes. Unfortunately, the pre-filtering step, I that not all drugs had DEGs in the 859 genes constituting of the GenAge set. Therefore, aspirin, curcumin, lipoic acid, piceatannol and resveratrol were removed from my drug profiles set.

I continued by computing the cosine similarity and decorrelation score on the PC coordinates of the remaining 12 drug DEG profiles and testing if it is correlated with the two measure of healthspan. None of the relationships were found to be statistically significant.

Furthermore, I need to take into consideration the direction of the modulation of the gene expression. This more stringent and biological precise filtering procedure still keeps all the 12 drug profiles, meaning that my all the drugs in this set have at least one possible lifespan-extending GenAge gene which can explain their therapeutic effects. Computing the cosine similarity based on the gene expression of this sparse DEG matrix or the weighted cosine similarity based on its PC coordinates have no statistically significant linear correlation with either measures of healthspan. Neither does their respective decorrelation score. This means that drugs with a mode of action more alike (as evaluated

[a] Although, the results are robust to the pipeline variant choice.
[b] I restricted myself to use these 17 drugs, because later in this chapter I will have the possibility to make use of the healthspan information. Furthermore, it allows a fairer comparison between the results originating from this and the previous RNA-Seq workflows.

using my PCA based decomposition of transcriptional change) are no more likely than drugs that are less alike to interact beneficially when combined.

6.3 – Drug-drug interactions are predominantly non-linear

Throughout my analysis, a significant potential confounder for predicting drug-drug interactions was the fact that in most cases we do not have information regarding the transcriptional changes elicited by pairs of drugs. In constructing scoring functions, I therefore often assumed implicitly or explicitly that transcriptional changes of pairs of drugs could be approximated as linear combination of the changes of individual drugs. To explicitly evaluate the degree to which this assumption is incorrect, I took advantage of a small number of drug-drug pairs for which we had previously determined transcriptional changes of both individual drugs and drug pairs[70]. Taking advantage of this dataset, I first filtered each list of DEGs by applying an adjusted p-value threshold of α less than 0.05. This information was compiled into a matrix with the 14524 rows representing the union gene set of significant DEGs from all drugs and drug-pairs in this set. Lastly, I used linear regression to construct the transcriptional profile of each drug combination, based on the linear combination of the transcriptional signatures of the single drugs involved (see sub-chapter 6.1).

My results are displayed in Table 6.1. Modelling by linear regression is the formal way to test the linearity assumption. The multiple R-square value the regressions was use as the metric of the explanatory power of the linearity assumption, as it measures the explained variance of the drug combination transcriptional profile.

Drug A	Coefficient A	Drug B	Coefficient B	Drug C	Coefficient C	Explained Variance
Rap	0,313	Rif	0,762			0,49
Allan	-0,028	Rif	0,889			0,408
Allan	-0,38	Rap	0,263	Rif	0,586	0,401
Allan	-0,204	Psora	1,209			0,366
Psora	0,204	Rif	0,468			0,275
Allan	0,132	Rap	0,836			0,22
Allan	-0,351	Psora	-0,003	Rif	0,348	0,155
Psora	-0,03	Rap	0,261	Rif	0,374	0,149
Psora	0,275	Rap	0,084			0,116
Met	0,954	Rap	0,596			0,114

Table 6.1 – Summary of linear modelling of composite drug interventions.
Allan = allantoin. Met = metformin. Psora = psora-4. Rap = rapamycin. Rif = rifampicin.

As shown in Table 6.1, limiting the modelling of drug combinations to linear combinations of their composing drugs can explain less than half of the experimentally observed transcriptional profiles following exposure with the pair of drugs. In other word, most of the gene expression changes observed when combining two drugs cannot be re-created by linearly combining the gene expression changes of the constituent drugs. Additional evidence for the presence of significant non-linearity is the fact that some of the single drug profiles required a negative coefficient to maximize the quality of the fit. Conceptually, this is equivalent to adding less than zero of one of the two drugs. It would be expected that all the drug profiles present a positive coefficient, as negative values are non-sensical from a biological context (what does it mean to be exposed to less than zero allantoin ?).

In sum, a linear combination of drug transcriptional profiles can only explain the minority of the transcriptional changes of their combined intervention, emphasizing the fact that a more biologically realistic, although much harder to

generate, network-level model of the aging GRN is most likely needed to fully predict drug-drug interactions.

6.4 – Conclusion

In this chapter, using two related approaches, I demonstrated that dissimilar drugs are not more likely to be synergistic. One class of methods was to use scoring functions based on the current best knowledge of genes involved in lifespan determination in *C. elegans*. The other method was hypothesis-free and based on constructing a low-dimensional transcriptional space suitable for clustering drugs by similarity in terms of transcriptional effects. My results contradict our previous hypothesis that broadly dissimilar drugs are more likely to interact synergistically than drugs that overlap in their target gene set and mode of action. However, both of these approaches assume that comparing transcriptional effects of single drugs allows prediction of their interaction without knowledge of a detailed model of the GRN governing aging (both assume that linear effects dominate or at least play a significant role). I tested this assumption and showed that gene expression changes of drug combinations are substantially different from the best linear combination of individual effects. Statistically, less than half of all changes can be explained by linear combination. Looking into some of the unexplained changes in detail, further reveals evidence for new/different pathways being recruited when exposing organism to pairs of drugs but also for pathways targeted by individual drugs being, less intuitively, lost from the mode of action of pairs[70].

This implies that a regulatory gene network description of aging is needed to model the non-linearity of anti-aging drug interactions.

Chapter 7 Conclusion

7.1 Discussion

I designed, assemble, validated and finally used a novel, fully automated high-throughput lifespan screening system (Chapter 3). I also carefully selected my drug library (see through and up-to-date literature review in sub-chapter 2.2) in order to guarantee significant effect size, diversity of modes of action (which I hypothesized to lead to drug synergies) and effect size, among other factors (sub-chapter 2.1). Using the system in conjunction with the library, I validated each of these drugs for healthspan effects and then carried out a drug-drug interaction screen. During the screen, I discovered 6 novel drug pairs that interact synergistically to extend healthspan in *C. elegans*. Moreover, by using these drugs, I was able to extend mean healthspan by approx. 30% or more, even thought drugs were administered strictly in adulthood (Chapter 5). This is contrast to 28% healthspan extension for the "best" single drug (curcumin).

While I was working on this project another high-throughput system (HTS) was published – the WormBot[299]. The WormBot is an open-source robotic platform capable of taking time-lapse bright field images and real-time video micrographs of up to 144 parallel *C. elegans* experiments. Like my system, it also used solid agar medium and well-tissue plates. It solves the same issues that my system does (see section for the previous systems limitation and drawbacks). Nonetheless, my HTS should be more scalable regarding the number of conditions, especially in terms of volume of medium (and drug stocks) required. For this fact one can envision that for an even larger combinatorial genetic or drug screen, an ideal procedure might be to conduct a two-step screening: the first step using my system to score a large number of

interventions; followed by a second step consisting in the validation of successful hits using the WormBot.

To be able to understand the effects of my drug library in worms, I generated RNA-Seq data for each of the lifespan-extending drugs I tested (Chapter 4). Together, these datasets are the largest systematic set of transcriptomics data for drugs with associated standardized (in a single experiment) healthspan information. I then used this data to test several of my hypotheses regarding drug-drug interactions.

It has been suggested previously that drugs that target known aging genes or pathways can be combined to target a larger set of these genes or pathways (section 1.2.3). It is therefore logical that drugs that modulate more of these genes and/or by a larger amount should elicit larger effect sizes. I implemented several scoring functions to test this assumption, and, surprisingly, there was no correlation between individual drugs ability to modulate known aging genes and the magnitude of their healthspan extension efficacy (Figure 6.1).

Attempting to further explore this conundrum, I developed a hypothesis-free scoring function based on principal component analysis of the full transcriptome changes. This function does not explicitly attempt to predict effect size by based on transcriptional changes but functions as a similarity score, allowing comparison of drugs by mode of action.

This classifier successfully separated drugs into different classes (clusters) that are broadly consistent with their likely mode of action. Using my classifier, as well as a simpler binary gene-by-gene (Venn diagram) comparison of GenAge genes impacted, I then tested the assumption that differences in terms of mode

of action would increase the probability of synergistic interactions. However, this similarity assumption also failed to correlate with the likelihood of drug pairs being synergistic.

The lack of synergy predictability based on drug similarity lead me to further reflect on my hypothesis that a drug combination can be explained by the sum of the effects of the drugs that constitutes (linearity assumption). I revisited data that I had analyzed as part of a previous manuscript that reported lifespan and transcriptomic data for 5 lifespan-extending drugs and most of their corresponding paired interventions[70]. I revisited this data to test the linearity assumption. I found that it is false. In more detail, when drugs are combined the transcriptional profile of their joint intervention cannot usefully be modelled as linear combination of the profiles of its single drugs interventions (Table 6.1).

It is surprising that despite having thorough knowledge of all the gene expression changes of lifespan determining genes (GenAge), I was unable to predict synergies or explain individual drug effects. This suggests that I am working under the wrong paradigm. In other words, it seems that the capacity to predict synergistic drug combinations will only be achieved when we obtain an aging gene-regulatory network capable of informing the modelling of non-linear interactions. This is consistent with the observation that all known drug synergies were either found by screening or constructed in a hypothesis driven fashion (see section 1.2.3).

However, conceptionally, another important insight is that synergies are relatively rare. Out of the 48 pairs tested in the final combinatorial screen only 6 were convincingly synergistic. By contrast 39 either showed no additional or

diminishing benefits and 3 were even toxic. This is a critical point in the context of translation as supplements and even some drugs are commonly combined into “stacks” under the assumption that “more-helps-more”. This is another example of an implicit assumption of linearity, namely that benefits of several drugs or supplements will be at least partially additive. My results suggest that diminishing or zero effects are the norm and that negative drug interactions (including pure toxicity and drugs that nullify each other’s pro-longevity effects) appear more frequently than beneficial synergies.

7.2 Future Work

I will leave as open avenues of investigation the following suggestions.

Besides my proposed HTS being extendable to video tracking it might also be able to directly measure lifespan (and bypass the need for pre-filtering the images) by using more refined deep learning algorithms (e.g.[300]). Furthermore, the use of strains that have a fluorescence marker that is related to the aging process might provide additional mode of action insights, on top of being used for tracking purposes. An example would be *hsp-16.1*[301], for which levels of fluorescence intensity would be related to aging whether being larger[241] or smaller[242] than wild-type.

Some of my newly discovered synergistic drug pairs mainly extended mean healthspan, while others worked incredibly well only when most of the population had already perished (see sub-chapter 5.3). This suggests the interesting possibility that shifting drug interventions according to the period of life might maximize effect size.

Bibliography

1. Bloom, D. E., Canning, D. & Fink, G. Implications of population ageing for economic growth. *Oxford Rev. Econ. Policy* **26**, 583–612 (2010).

2. Zhavoronkov, A. & Litovchenko, M. Biomedical progress rates as new parameters for models of economic growth in developed countries. *Int. J. Environ. Res. Public Health* **10**, 5936–5952 (2013).

3. Petsko, G. A. A seat at the table. *Genome Biol.* **9**, 113 (2008).

4. Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Current Biology* **22**, (2012).

5. Vaupel, J. W. Biodemography of Human Ageing. *Nature* **464**, 536–42 (2010).

6. Zuyun Liu, Ling Han, Xiaofeng Wang, Qiushi Feng, T. M. G. Disability Prior to Death Among the Oldest-Old in China. *Journals Gerontol. Ser. A* (2017). doi:10.1093/neuonc/nox188/4265638

7. Rechel, B. *et al.* Ageing in the European Union. *Lancet* **381**, 1312–1322 (2013).

8. Howse, K. Healthy ageing: the role of health care services. *Perspect. Public Health* **132**, 171–7 (2012).

9. Luyten, W. *et al.* Ageing with elegans: a research proposal to map healthspan pathways. *Biogerontology* (2016). doi:10.1007/s10522-016-9644-x

10. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A C.

elegans mutant that lives twice as long as wild type. *Nature* **366**, 461–4 (1993).

11. Rattan, S. I. S. Biogerontology: From here to where? The Lord Cohen Medal Lecture-2011. *Biogerontology* **13**, 83–91 (2012).

12. Longo, V. D. *et al.* Interventions to slow aging in humans: Are we ready? *Aging Cell* **14**, 497–510 (2015).

13. Barzilai, N. *et al.* Metformin as a Tool to Target Aging. *Cell Metab.* **23**, 1060–1065 (2016).

14. Tacutu, R. *et al.* Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1042

15. Barardo, D. *et al.* The DrugAge database of aging-related drugs. *Aging Cell* **16**, 594–597 (2017).

16. Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science* (1998). doi:10.1126/science.282.5396.2012

17. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature* (1998). doi:10.1038/35888

18. Kenyon, C. The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **366**, 9–16 (2011).

19. Strong, R. *et al.* Longer lifespan in male mice treated with a weakly estrogenic agonist, an antioxidant, an α-glucosidase inhibitor or a Nrf2-inducer. *Aging Cell* **15**, 872–884 (2016).

20. Wilkinson, D. S., Taylor, R. C. & Dillin, A. Analysis of Aging in Caenorhabditis elegans. in *Methods in Cell Biology* (2012). doi:10.1016/B978-0-12-394620-1.00012-6

21. Sutphin, G. L. & Kaeberlein, M. Measuring Caenorhabditis elegans life span on solid media. *J. Vis. Exp.* (2009). doi:10.3791/1152

22. Saberi-Bosari, S., Huayta, J. & San-Miguel, A. A microfluidic platform for lifelong high-resolution and high throughput imaging of subtle aging phenotypes in C. elegans. *Lab Chip* (2018). doi:10.1039/c8lc00655e

23. Banse, S. A., Blue, B. W., Robinson, K. J., Jarrett, C. M. & Phillips, P. C. The stress-chip: A microfluidic platform for stress analysis in Caenorhabditis elegans. *PLoS One* (2019). doi:10.1371/journal.pone.0216283

24. Xian, B. *et al.* WormFarm: A quantitative control and measurement device toward automated Caenorhabditis elegans aging analysis. *Aging Cell* (2013). doi:10.1111/acel.12063

25. Schulenburg, H. & Félix, M. A. The natural biotic environment of Caenorhabditis elegans. *Genetics* (2017). doi:10.1534/genetics.116.195511

26. Banse, S. A. *et al.* Automated lifespan determination across Caenorhabditis strains and species reveals assay-specific effects of chemical interventions. *GeroScience* (2019). doi:10.1007/s11357-019-00108-9

27. Stroustrup, N. *et al.* The Caenorhabditis elegans Lifespan Machine. *Nat.*

Methods **10**, 665–70 (2013).

28. Churgin, M. A. *et al.* Longitudinal imaging of caenorhabditis elegans in a microfabricated device reveals variation in behavioral decline during aging. *Elife* (2017). doi:10.7554/eLife.26652

29. Ayyadevara, S., Alla, R., Thaden, J. J. & Shmookler Reis, R. J. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging Cell* **7**, 13–22 (2008).

30. Arantes-Oliveira, N., Berman, J. R. & Kenyon, C. Healthy animals with extreme longevity. *Science* **302**, 611 (2003).

31. Chen, D. *et al.* Germline Signaling Mediates the Synergistically Prolonged Longevity Produced by Double Mutations in daf-2 and rsks-1 in C.elegans. *Cell Rep.* **5**, 1600–1610 (2013).

32. Burgess, J., Hihi, A. K., Benard, C. Y., Branicky, R. & Hekimi, S. Molecular mechanism of maternal rescue in the clk-1 mutants of Caenorhabditis elegans. *J. Biol. Chem.* **278**, 49555–49562 (2003).

33. Larsen, P. L., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in Caenorhabditis elegans. *Genetics* **139**, 1567–1583 (1995).

34. Davidsohn, N. *et al.* A single combination gene therapy treats multiple age-related diseases. *Proc. Natl. Acad. Sci. U. S. A.* (2019). doi:10.1073/pnas.1910073116

35. Hans, H., Lone, A., Aksenov, V. & Rollo, C. D. Impacts of metformin and aspirin on life history features and longevity of crickets: trade-offs

versus cost-free life extension? *Age (Omaha).* (2015). doi:10.1007/s11357-015-9769-x

36. Snell, T. W., Johnston, R. K., Rabeneck, B., Zipperer, C. & Teat, S. Joint inhibition of TOR and JNK pathways interacts to extend the lifespan of Brachionus manjavacas (Rotifera). *Exp. Gerontol.* (2014). doi:10.1016/j.exger.2014.01.022

37. Castillo-Quan, J. I. *et al.* A triple drug combination targeting components of the nutrient-sensing network maximizes longevity. *Proc. Natl. Acad. Sci. U. S. A.* (2019). doi:10.1073/pnas.1913212116

38. Desjardins, D. *et al.* Antioxidants reveal an inverted U-shaped dose-response relationship between reactive oxygen species levels and the rate of aging in Caenorhabditis elegans. *Aging Cell* (2017). doi:10.1111/acel.12528

39. Danilov, A. *et al.* Selective anticancer agents suppress aging in Drosophila. *Oncotarget* **4**, 1507–1526 (2013).

40. Huang, X. *et al.* Reducing signs of aging and increasing lifespan by drug synergy. *Aging Cell* (2013). doi:10.1111/acel.12090

41. Dakik, P. *et al.* Pairwise combinations of chemical compounds that delay yeast chronological aging through different signaling pathways display synergistic effects on the extent of aging delay. *Oncotarget* (2019). doi:10.18632/oncotarget.26553

42. Van den Berge, K. *et al.* RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annu. Rev. Biomed. Data Sci.* (2019).

doi:10.1146/annurev-biodatasci-072018-021255

43. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nature Reviews Genetics* (2011). doi:10.1038/nrg3068

44. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* (2015). doi:10.1186/s13059-015-0762-6

45. Sun, W. & Hu, Y. eQTL Mapping Using RNA-seq Data. *Stat. Biosci.* (2013). doi:10.1007/s12561-012-9068-3

46. Alamancos, G. P., Agirre, E. & Eyras, E. Methods to study splicing from high-throughput RNA sequencing data. *Methods Mol. Biol.* (2014). doi:10.1007/978-1-62703-980-2_26

47. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* (2008). doi:10.1038/nature07517

48. Ju, J. *et al.* Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0609513103

49. Lazic, S. E. Experimental design for laboratory biologists: Maximising Information and Improving Reproducibility. *Book* (2016).

50. Williams, A. G., Thomas, S., Wyman, S. K. & Holloway, A. K. RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. *Curr. Protoc. Hum. Genet.* (2014).

doi:10.1002/0471142905.hg1113s83

51. Poplawski, A. & Binder, H. Feasibility of sample size calculation for RNA-seq studies. *Brief. Bioinform.* (2018). doi:10.1093/bib/bbw144

52. Baruzzo, G. *et al.* Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat. Methods* (2017). doi:10.1038/nmeth.4106

53. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–7 (2016).

54. Pérez-Rubio, P., Lottaz, C. & Engelmann, J. C. FastqPuri: High-performance preprocessing of RNA-seq data. *BMC Bioinformatics* (2019). doi:10.1186/s12859-019-2799-0

55. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research* (2016). doi:10.12688/F1000RESEARCH.7563.2

56. Team, R. C. R: A Language and Environment for Statistical Computing. *Vienna, Austria* (2019).

57. Abrams, Z. B., Johnson, T. S., Huang, K., Payne, P. R. O. & Coombes, K. A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinformatics* (2019). doi:10.1186/s12859-019-3247-x

58. Evans, C., Hardin, J. & Stoebel, D. M. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* (2018). doi:10.1093/bib/bbx008

59. Mandelboum, S., Manber, Z., Elroy-Stein, O. & Elkon, R. Recurrent

functional misinterpretation of RNA-seq data caused by sample-specific gene length bias. *PLoS Biol.* (2019). doi:10.1371/journal.pbio.3000481

60. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics* (2011). doi:10.1186/1471-2105-12-480

61. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks001

62. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* (2010). doi:10.1186/1471-2105-11-94

63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014). doi:10.1186/s13059-014-0550-8

64. Lin, B. & Pang, Z. Stability of methods for differential expression analysis of RNA-seq data. *BMC Genomics* **20**, 1–13 (2019).

65. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* (2019). doi:10.1093/bioinformatics/bty895

66. Tarkhov, A. E. *et al.* A universal transcriptomic signature of age reveals the temporal scaling of Caenorhabditis elegans aging trajectories. *Sci. Rep.* (2019). doi:10.1038/s41598-019-43075-z

67. Sompairac, N. *et al.* Independent component analysis for unraveling the complexity of cancer omics datasets. *International Journal of Molecular Sciences* (2019). doi:10.3390/ijms20184414

68. Crane, M. Questionable Answers in Question Answering Research: Reproducibility and Variability of Published Results. *Trans. Assoc. Comput. Linguist.* (2018). doi:10.1162/tacl_a_00018

69. Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K. & Narasimhan, G. So you think you can PLS-DA? in *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)* (ed. IEEE) (2018). doi:10.1109/iccabs.2018.8542038

70. Admasu, T. D. *et al.* Drug Synergy Slows Aging and Improves Healthspan through IGF and SREBP Lipid Signaling. *Dev. Cell* (2018). doi:10.1016/j.devcel.2018.09.001

71. Barardo, D. *et al.* The DrugAge database of aging-related drugs. *Aging Cell* **16**, (2017).

72. Whewell, W. *The Philosophy of the Inductive Sciences, Founded upon their History*. (1847). doi:10.5840/eps201647141

73. Tucker, A. *A Companion to the Philosophy of History and Historiography*. *A Companion to the Philosophy of History and Historiography* (2009). doi:10.1002/9781444304916

74. Calvert, S. *et al.* A network pharmacology approach reveals new candidate caloric restriction mimetics in C. elegans. *Aging Cell* **15**, 256–266 (2016).

75. Lamb, J. *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science (80-.).* (2006). doi:10.1126/science.1132939

76. Houthoofd, K., Braeckman, B. P., Johnson, T. E. & Vanfleteren, J. R. Life extension via dietary restriction is independent of the Ins/IGF-1 signalling pathway in Caenorhabditis elegans. *Exp. Gerontol.* (2003). doi:10.1016/S0531-5565(03)00161-X

77. Hosono, R., Sato, Y., Aizawa, S. I. & Mitsui, Y. Age-dependent changes in mobility and separation of the nematode Caenorhabditis elegans. *Exp. Gerontol.* (1980). doi:10.1016/0531-5565(80)90032-7

78. Chin, R. M. *et al.* The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. *Nature* (2014). doi:10.1038/nature13264

79. Edwards, C. *et al.* Mechanisms of amino acid-mediated lifespan extension in Caenorhabditis elegans. *BMC Genet.* (2015). doi:10.1186/s12863-015-0167-2

80. Fu, X. *et al.* 2-hydroxyglutarate inhibits ATP synthase and mTOR Signaling. *Cell Metab.* (2015). doi:10.1016/j.cmet.2015.06.009

81. Gaudet, J. & Mango, S. E. Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science (80-.).* (2002). doi:10.1126/science.1065175

82. Su, Y. *et al.* Alpha-ketoglutarate extends Drosophila lifespan by inhibiting mTOR and activating AMPK. *Aging (Albany. NY).* (2019).

doi:10.18632/aging.102045

83. Lucanic, M. *et al.* Impact of genetic background and experimental reproducibility on identifying chemical compounds with robust longevity effects. *Nat. Commun.* (2017). doi:10.1038/ncomms14256

84. Danilov, A. *et al.* Influence of non-steroidal anti-inflammatory drugs on Drosophila melanogaster longevity. *Oncotarget* (2015). doi:10.18632/oncotarget.5118

85. Song, C. *et al.* Metabolome analysis of effect of aspirin on Drosophila lifespan extension. *Exp. Gerontol.* (2017). doi:10.1016/j.exger.2017.04.010

86. Hans, H., Lone, A., Aksenov, V. & Rollo, C. D. Impacts of metformin and aspirin on life history features and longevity of crickets: trade-offs versus cost-free life extension? *Age (Omaha).* (2015). doi:10.1007/s11357-015-9769-x

87. Ayyadevara, S. *et al.* Aspirin inhibits oxidant stress, reduces age-associated functional declines, and extends lifespan of caenorhabditis elegans. *Antioxidants Redox Signal.* (2013). doi:10.1089/ars.2011.4151

88. Wan, Q. L., Zheng, S. Q., Wu, G. S. & Luo, H. R. Aspirin extends the lifespan of Caenorhabditis elegans via AMPK and DAF-16/FOXO in dietary restriction pathway. *Exp. Gerontol.* (2013). doi:10.1016/j.exger.2013.02.020

89. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. *Nature* (2003).

doi:10.1038/nature01789

90. Huang, X. B. *et al.* Aspirin increases metabolism through germline signalling to extend the lifespan of Caenorhabditis elegans. *PLoS One* (2017). doi:10.1371/journal.pone.0184027

91. Pietrocola, F. *et al.* Aspirin Recapitulates Features of Caloric Restriction. *Cell Rep.* (2018). doi:10.1016/j.celrep.2018.02.024

92. Hochschild, R. Effects of various drugs on longevity in female C57BL/6J mice. *Gerontology* **19**, 271–280 (1973).

93. Strong, R. *et al.* Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice. *Aging Cell* (2008). doi:10.1111/j.1474-9726.2008.00414.x

94. Miller, R. A. *et al.* Glycine supplementation extends lifespan of male and female mice. *Aging Cell* (2019). doi:10.1111/acel.12953

95. Karmali, K. N. & Huffman, M. D. I do not have heart disease-should i be taking aspirin? *JAMA Cardiology* (2017). doi:10.1001/jamacardio.2017.0294

96. Agüero-Torres, H., Viitanen, M., Fratiglioni, L. & Louhija, J. The effect of low-dose daily aspirin intake on survival in the Finnish centenarians cohort. *Journal of the American Geriatrics Society* (2001). doi:10.1046/j.1532-5415.2001.4911264.x

97. Kumar, S., Dietrich, N. & Kornfeld, K. Angiotensin Converting Enzyme (ACE) Inhibitor Extends Caenorhabditis elegans Life Span. *PLoS Genet.* (2016). doi:10.1371/journal.pgen.1005866

98. Revelas, M. *et al.* Review and meta-analysis of genetic polymorphisms associated with exceptional human longevity. *Mechanisms of Ageing and Development* (2018). doi:10.1016/j.mad.2018.06.002

99. Kitani, K., Osawa, T. & Yokozawa, T. The effects of tetrahydrocurcumin and green tea polyphenol on the survival of male C57BL/6 mice. *Biogerontology* (2007). doi:10.1007/s10522-007-9100-z

100. Strong, R. *et al.* Evaluation of resveratrol, green tea extract, curcumin, oxaloacetic acid, and medium-chain triglyceride oil on life span of genetically heterogeneous mice. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2013). doi:10.1093/gerona/gls070

101. Li, J., Cui, X., Wang, Z. & Li, Y. rBTI extends Caenorhabditis elegans lifespan by mimicking calorie restriction. *Exp. Gerontol.* **67**, 62–71 (2015).

102. Lee, K. S. *et al.* Curcumin extends life span, improves health span, and modulates the expression of age-associated aging genes in drosophila melanogaster. *Rejuvenation Res.* (2010). doi:10.1089/rej.2010.1031

103. Shen, L. R. *et al.* Curcumin-supplemented diets increase superoxide dismutase activity and mean lifespan in Drosophila. *Age (Omaha).* (2013). doi:10.1007/s11357-012-9438-2

104. Soh, J. W. *et al.* Curcumin is an early-acting stage-specific inducer of extended functional longevity in Drosophila. *Exp. Gerontol.* (2013). doi:10.1016/j.exger.2012.09.007

105. Chandrashekara, K. T., Popli, S. & Shakarad, M. N. Curcumin enhances

parental reproductive lifespan and progeny viability in Drosophila melanogaster. *Age (Omaha).* (2014). doi:10.1007/s11357-014-9702-8

106. Arking, R. Independent chemical regulation of health and senescent spans in Drosophila. *Invertebr. Reprod. Dev.* (2015). doi:10.1080/07924259.2014.978028

107. Chandrashekara, K. T., Popli, S. & Shakarad, M. N. Curcumin enhances parental reproductive lifespan and progeny viability in Drosophila melanogaster. *Age (Omaha).* **36**, 9702 (2014).

108. Chen, Y. *et al.* Curcumin supplementation increases survival and lifespan in Drosophila under heat stress conditions. *BioFactors* (2018). doi:10.1002/biof.1454

109. Brown, M. K., Evans, J. L. & Luo, Y. Beneficial effects of natural antioxidants EGCG and α-lipoic acid on life span and age-dependent behavioral declines in Caenorhabditis elegans. *Pharmacol. Biochem. Behav.* (2006). doi:10.1016/j.pbb.2006.10.017

110. Abbas, S. & Wink, M. Epigallocatechin gallate from green tea (Camellia sinensis) increases lifespan and stress resistance in Caenorhabditis elegans. *Planta Med.* (2009). doi:10.1055/s-0028-1088378

111. Xiong, L. G. *et al.* Epigallocatechin-3-gallate promotes healthy lifespan through mitohormesis during early-to-mid adulthood in Caenorhabditis elegans. *Redox Biol.* (2018). doi:10.1016/j.redox.2017.09.019

112. Ristow, M. & Schmeisser, K. Mitohormesis: Promoting Health and Lifespan by Increased Levels of Reactive Oxygen Species (ROS). *Dose.*

Response. **12**, 288–341 (2014).

113. Bartholome, A., Kampkötter, A., Tanner, S., Sies, H. & Klotz, L. O. Epigallocatechin gallate-induced modulation of FoxO signaling in mammalian cells and C. elegans: FoxO stimulation is masked via PI3K/Akt activation by hydrogen peroxide formed in cell culture. *Arch. Biochem. Biophys.* (2010). doi:10.1016/j.abb.2010.05.024

114. Niu, Y. *et al.* The phytochemical, EGCG, extends lifespan by reducing liver and kidney function damage and improving age-associated inflammation and oxidative stress in healthy rats. *Aging Cell* **12**, 1041–1049 (2013).

115. Wagner, A. E. *et al.* Epigallocatechin gallate affects glucose metabolism and increases fitness and lifespan in Drosophila melanogaster. *Oncotarget* (2015). doi:10.18632/oncotarget.5215

116. Cai, W. J. *et al.* Icariin and its derivative icariside II extend healthspan via insulin/IGF-1 pathway in C. elegans. *PLoS One* (2011). doi:10.1371/journal.pone.0028835

117. Zhang, S.-Q. *et al.* Icariin, a natural flavonol glycoside, extends healthspan in mice. *Exp. Gerontol.* (2015). doi:10.1016/j.exger.2015.06.020

118. Bauer, J. H., Goupil, S., Garber, G. B. & Helfand, S. L. An accelerated assay for the identification of lifespan-extending interventions in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0403493101

119. Du, G. *et al.* Lipoic acid rejuvenates aged intestinal stem cells by preventing age-associated endosome reduction. *EMBO Rep.* (2020). doi:10.15252/embr.201949583

120. Benedetti, M. G. *et al.* Compounds that confer thermal stress resistance and extended lifespan. *Exp. Gerontol.* (2008). doi:10.1016/j.exger.2008.08.049

121. Merry, B. J., Kirk, A. J. & Goyns, M. H. Dietary lipoic acid supplementation can mimic or block the effect of dietary restriction on life span. *Mech. Ageing Dev.* (2008). doi:10.1016/j.mad.2008.04.004

122. McColl, G. *et al.* Pharmacogenetic analysis of lithium-induced delayed aging in Caenorhabditis elegans. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M705028200

123. Cipriani, A., Pretty, H., Hawton, K. & Geddes, J. R. Lithium in the prevention of suicidal behavior and all-cause mortality in patients with mood disorders: A systematic review of randomized trials. *American Journal of Psychiatry* (2005). doi:10.1176/appi.ajp.162.10.1805

124. Tam, Z. Y., Gruber, J., Ng, L. F., Halliwell, B. & Gunawan, R. Effects of lithium on age-related decline in mitochondrial turnover and function in caenorhabditis elegans. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2014). doi:10.1093/gerona/glt210

125. Zarse, K. *et al.* Low-dose lithium uptake promotes longevity in humans and metazoans. *Eur. J. Nutr.* (2011). doi:10.1007/s00394-011-0171-x

126. Teo, E., Fong, S., Tolwinski, N. & Gruber, J. Drug synergy as a strategy

for compression of morbidity in a Caenorhabditis elegans model of Alzheimer's disease. *GeroScience* (2020). doi:10.1007/s11357-020-00169-1

127. Suganthi, M., Sangeetha, G., Gayathri, G. & Ravi Sankar, B. Biphasic dose-dependent effect of lithium chloride on survival of human hormone-dependent breast cancer cells (MCF-7). *Biol. Trace Elem. Res.* (2012). doi:10.1007/s12011-012-9510-x

128. Castillo-Quan, J. I. *et al.* Lithium Promotes Longevity through GSK3/NRF2-Dependent Hormesis. *Cell Rep.* (2016). doi:10.1016/j.celrep.2016.03.041

129. Fajardo, V. A., Leblanc, P. J. & Fajardo, V. A. Trace lithium in texas tap water is negatively associated with all-cause mortality and premature death. *Appl. Physiol. Nutr. Metab.* (2018). doi:10.1139/apnm-2017-0653

130. Brack, C., Bechter-Thüring, E. & Labuhn, M. N-Acetylcysteine slows down ageing and increases the life span of Drosophila melanogastar. *Cell. Mol. Life Sci.* (1997). doi:10.1007/PL00013199

131. Flurkey, K., Astle, C. M. & Harrison, D. E. Life extension by diet restriction and N-acetyl-L-cysteine in genetically heterogeneous mice. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2010). doi:10.1093/gerona/glq155

132. Shibamura, A., Ikeda, T. & Nishikawa, Y. A method for oral administration of hydrophilic substances to Caenorhabditis elegans: Effects of oral supplementation with antioxidants on the nematode

lifespan. *Mech. Ageing Dev.* (2009). doi:10.1016/j.mad.2009.06.008

133. Oh, S., Park, J. & Park, S. Lifespan extension and increased resistance to environmental stressors by N-Acetyl-L-Cysteine in Caenorhabditis elegans. *Clinics* **70**, 380–386 (2015).

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

135. De Haes, W. *et al.* Metformin promotes lifespan through mitohormesis via the peroxiredoxin PRDX-2. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1321776111

136. Dehghan, E. *et al.* Hydralazine induces stress resistance and extends C. elegans lifespan by activating the NRF2/SKN-1 signalling pathway. *Nat. Commun.* (2017). doi:10.1038/s41467-017-02394-3

137. Wu, L. *et al.* An Ancient, Unified Mechanism for Metformin Growth Inhibition in C. elegans and Cancer. *Cell* (2016). doi:10.1016/j.cell.2016.11.055

138. Chen, J. *et al.* Metformin extends C. Elegans lifespan through lysosomal pathway. *Elife* (2017). doi:10.7554/eLife.31268

139. Slack, C., Foley, A. & Partridge, L. Activation of AMPK by the Putative Dietary Restriction Mimetic Metformin Is Insufficient to Extend Lifespan in Drosophila. *PLoS One* **7**, (2012).

140. Wei, J. *et al.* Effects of Metformin on Life Span, Cognitive Ability, and Inflammatory Response in a Short-Lived Fish. *Journals Gerontol. Ser. A*

XX, 1–9 (2020).

141. Martin-Montalvo, A. *et al.* Metformin improves healthspan and lifespan in mice. *Nat Commun.* **4**, 2192 (2013).

142. Alfaras, I. *et al.* Health benefits of late-onset metformin treatment every other week in mice. *npj Aging Mech. Dis.* (2017). doi:10.1038/s41514-017-0018-7

143. Smith, D. L. *et al.* Metformin supplementation and life span in fischer-344 rats. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2010). doi:10.1093/gerona/glq033

144. Grünz, G. *et al.* Structural features and bioavailability of four flavonoids and their implications for lifespan-extending and antioxidant actions in C. elegans. *Mech. Ageing Dev.* (2012). doi:10.1016/j.mad.2011.11.005

145. Büchter, C. *et al.* Myricetin-mediated lifespan extension in caenorhabditis elegans is modulated by DAF-16. *Int. J. Mol. Sci.* (2013). doi:10.3390/ijms140611895

146. Shen, P. *et al.* Piceatannol extends the lifespan of Caenorhabditis elegans via DAF-16. *BioFactors* (2017). doi:10.1002/biof.1346

147. Ye, X., Linton, J. M., Schork, N. J., Buck, L. B. & Petrascheck, M. A pharmacological network for lifespan extension in Caenorhabditis elegans. *Aging Cell* (2014). doi:10.1111/acel.12163

148. Wu, Q. *et al.* 2,5-Dimethyl-Celecoxib Extends Drosophila Life Span via a Mechanism That Requires Insulin and Target of Rapamycin Signaling. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2017).

doi:10.1093/gerona/glw244

149. Bjedov, I. *et al.* Mechanisms of Life Span Extension by Rapamycin in the Fruit Fly Drosophila melanogaster. *Cell Metab.* **11**, 35–46 (2010).

150. Wu, D., Rea, S. L., Cypser, J. R. & Johnson, T. E. Mortality shifts in Caenorhabditis elegans: remembrance of conditions past. *Aging Cell* **8**, 666–75 (2009).

151. Houtkooper, R. H. *et al.* Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* (2013). doi:10.1038/nature12188

152. Chen, D. *et al.* Germline Signaling Mediates the Synergistically Prolonged Longevity Produced by Double Mutations in daf-2 and rsks-1 in C.elegans. *Cell Rep.* (2013). doi:10.1016/j.celrep.2013.11.018

153. Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* (2009). doi:10.1038/nature08221

154. Mi, X. N. *et al.* Methyl 3,4-dihydroxybenzoate enhances resistance to oxidative stressors and lifespan in C. elegans partially via daf-2/daf-16. *Int. J. Mol. Sci.* **19**, (2018).

155. Robida-Stubbs, S. *et al.* TOR signaling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO. *Cell Metab.* (2012). doi:10.1016/j.cmet.2012.04.007

156. Calvert, S. *et al.* A network pharmacology approach reveals new candidate caloric restriction mimetics in C. elegans. *Aging Cell* (2016). doi:10.1111/acel.12432

157. Chen, D. *et al.* Germline Signaling Mediates the Synergistically Prolonged Longevity Produced by Double Mutations in daf-2 and rsks-1 in C.elegans. *Cell Rep.* **5**, 1600–1610 (2013).

158. Houtkooper, R. H. *et al.* Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* **497**, 451–7 (2013).

159. Miller, R. A. *et al.* Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. *Aging Cell* (2014). doi:10.1111/acel.12194

160. Fok, W. C. *et al.* Mice fed rapamycin have an increase in lifespan associated with major changes in the liver transcriptome. *PLoS One* (2014). doi:10.1371/journal.pone.0083988

161. Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).

162. Chen, C., Liu, Y., Liu, Y. & Zheng, P. MTOR regulation and therapeutic rejuvenation of aging hematopoietic stem cells. *Sci. Signal.* (2009). doi:10.1126/scisignal.2000559

163. Zhang, Y. *et al.* Rapamycin extends life and health in C57BL/6 mice. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2014). doi:10.1093/gerona/glt056

164. Neff, F. *et al.* Rapamycin extends murine lifespan but has limited effects on aging. *J. Clin. Invest.* (2013). doi:10.1172/JCI67674

165. Anisimov, V. N. *et al.* Rapamycin increases lifespan and inhibits spontaneous tumorigenesis in inbred female mice. *Cell Cycle* (2011).

doi:10.4161/cc.10.24.18486

166. Apelo, S. I. A., Pumper, C. P., Baar, E. L., Cummings, N. E. & Lamming, D. W. Intermittent administration of rapamycin extends the life span of female C57BL/6J Mice. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2016). doi:10.1093/gerona/glw064

167. Arriola Apelo, S. I. *et al.* Alternative rapamycin treatment regimens mitigate the impact of rapamycin on glucose homeostasis and the immune system. *Aging Cell* **15**, 28–38 (2016).

168. Bitto, A. *et al.* Transient rapamycin treatment can increase lifespan and healthspan in middle-aged mice. *Elife* (2016). doi:10.7554/eLife.16351

169. Barardo, D. *et al.* The DrugAge database of aging-related drugs. *Aging Cell* (2017). doi:10.1111/acel.12585

170. Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).

171. Howitz, K. T. *et al.* Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. *Nature* (2003). doi:10.1038/nature01960

172. Yang, H., Baur, J. A., Chen, A., Miller, C. & Sinclair, D. A. Design and synthesis of compounds that extend yeast replicative lifespan. *Aging Cell* (2007). doi:10.1111/j.1474-9726.2006.00259.x

173. Sun, K. *et al.* Anti-aging effects of hesperidin on saccharomyces cerevisiae via inhibition of reactive oxygen species and UTH1 gene expression. *Biosci. Biotechnol. Biochem.* (2012).

doi:10.1271/bbb.110535

174. Sun, K. *et al.* A steroidal saponin from Ophiopogon japonicus extends the lifespan of yeast via the pathway involved in SOD and UTH1. *Int. J. Mol. Sci.* (2013). doi:10.3390/ijms14034461

175. Orlandi, I., Stamerra, G., Strippoli, M. & Vai, M. During yeast chronological aging resveratrol supplementation results in a short-lived phenotype Sir2-dependent. *Redox Biol.* (2017). doi:10.1016/j.redox.2017.04.015

176. Viswanathan, M., Kim, S. K., Berdichevsky, A. & Guarente, L. A role for SIR-2.1 regulation of ER stress response genes in determining C. elegans life span. *Dev. Cell* (2005). doi:10.1016/j.devcel.2005.09.017

177. Bass, T. M., Weinkove, D., Houthoofd, K., Gems, D. & Partridge, L. Effects of resveratrol on lifespan in Drosophila melanogaster and Caenorhabditis elegans. *Mech. Ageing Dev.* (2007). doi:10.1016/j.mad.2007.07.007

178. Morselli, E. *et al.* Caloric restriction and resveratrol promote longevity through the Sirtuin-1-dependent induction of autophagy. *Cell Death Dis.* (2010). doi:10.1038/cddis.2009.8

179. Kim, J. *et al.* The natural phytochemical dehydroabietic acid is an anti-aging reagent that mediates the direct activation of SIRT1. *Mol. Cell. Endocrinol.* (2015). doi:10.1016/j.mce.2015.05.006

180. Fischer, N. *et al.* The resveratrol derivatives trans-3,5-dimethoxy-4-fluoro-4′-hydroxystilbene and trans-2,4′,5-trihydroxystilbene decrease

oxidative stress and prolong lifespan in Caenorhabditis elegans. *J. Pharm. Pharmacol.* **69**, 73–81 (2017).

181. Pannakal, S. T. *et al.* Longevity effect of a polysaccharide from Chlorophytum borivilianum on Caenorhabditis elegans and Saccharomyces cerevisiae. *PLoS One* (2017). doi:10.1371/journal.pone.0179813

182. Lim, Y. P. *et al.* Synthetic Enzymology and the Fountain of Youth : Repurposing Biology for Longevity. (2018). doi:10.1021/acsomega.8b01620

183. Gruber, J., Soon, Y. T. & Halliwell, B. Evidence for a trade-off between survival and fitness caused by resveratrol treatment of Caenorhabditis elegans. in *Annals of the New York Academy of Sciences* **1100**, 530–542 (2007).

184. Wang, C. *et al.* The effect of resveratrol on lifespan depends on both gender and dietary nutrient composition in Drosophila melanogaster. *Age (Omaha).* (2013). doi:10.1007/s11357-011-9332-3

185. Valenzano, D. R. *et al.* Resveratrol prolongs lifespan and retards the onset of age-related markers in a short-lived vertebrate. *Curr. Biol.* (2006). doi:10.1016/j.cub.2005.12.038

186. Yu, X. & Li, G. Effects of resveratrol on longevity, cognitive ability and aging-related histological markers in the annual fish Nothobranchius guentheri. *Exp. Gerontol.* (2012). doi:10.1016/j.exger.2012.08.009

187. Genade, T. & Lang, D. M. Resveratrol extends lifespan and preserves

glia but not neurons of the Nothobranchius guentheri optic tectum. *Exp. Gerontol.* (2013). doi:10.1016/j.exger.2012.11.013

188. Liu, T. *et al.* Resveratrol Attenuates Oxidative Stress and Extends Life Span in the Annual Fish *Nothobranchius guentheri*. *Rejuvenation Res.* **18**, 225–233 (2015).

189. Rascón, B., Hubbard, B. P., Sinclair, D. A. & Amdam, G. V. The lifespan extension effects of resveratrol are conserved in the honey bee and may be driven by a mechanism related to caloric restriction. *Aging (Albany. NY).* (2012). doi:10.18632/aging.100474

190. Lee, J., Kwon, G., Park, J., Kim, J. K. & Lim, Y. H. Brief Communication: SIR-2.1-dependent lifespan extension of Caenorhabditis elegans by oxyresveratrol and resveratrol. *Exp. Biol. Med.* **241**, 1757–1763 (2016).

191. Kaeberlein, M. *et al.* Substrate-specific activation of sirtuins by resveratrol. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M500655200

192. Kim, E., Ansell, C. M. & Dudycha, J. L. Resveratrol and food effects on lifespan and reproduction in the model crustacean Daphnia. *J. Exp. Zool. Part A Ecol. Genet. Physiol.* (2014). doi:10.1002/jez.1836

193. Zou, S. *et al.* The prolongevity effect of resveratrol depends on dietary composition and calorie intake in a tephritid fruit fly. *Exp. Gerontol.* (2009). doi:10.1016/j.exger.2009.02.011

194. Staats, S. *et al.* Dietary resveratrol does not affect life span, body composition, stress response, and longevity-related gene expression in

Drosophila melanogaster. *Int. J. Mol. Sci.* (2018). doi:10.3390/ijms19010223

195. Golegaonkar, S. *et al.* Rifampicin reduces advanced glycation end products and activates DAF-16 to increase lifespan in Caenorhabditis elegans. *Aging Cell* **14**, 463–473 (2015).

196. Schwarz, C. *et al.* Safety and tolerability of spermidine supplementation in mice and older adults with subjective cognitive decline. *Aging (Albany. NY).* (2018). doi:10.18632/aging.101354

197. Eisenberg, T. *et al.* Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nat. Med.* (2016). doi:10.1038/nm.4222

198. Nair, A. & Jacob, S. A simple practice guide for dose conversion between animals and human. *J. Basic Clin. Pharm.* **7**, 27 (2016).

199. Yue, F. *et al.* Spermidine prolongs lifespan and prevents liver fibrosis and hepatocellular carcinoma by activating MAP1S-mediated autophagy. *Cancer Res.* (2017). doi:10.1158/0008-5472.CAN-16-3462

200. Filfan, M. *et al.* Long-term treatment with spermidine increases health span of middle-aged Sprague-Dawley male rats. *GeroScience* (2020). doi:10.1007/s11357-020-00173-5

201. Eisenberg, T. *et al.* Induction of autophagy by spermidine promotes longevity. *Nat. Cell Biol.* **11**, 1305–14 (2009).

202. Porat, Y., Abramowitz, A. & Gazit, E. Inhibition of amyloid fibril formation by polyphenols: Structural similarity and aromatic interactions as a common inhibition mechanism. *Chemical Biology and Drug Design*

(2006). doi:10.1111/j.1747-0285.2005.00318.x

203. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–217 (2013).

204. Alavez, S., Vantipalli, M. C., Zucker, D. J. S., Klang, I. M. & Lithgow, G. J. Amyloid-binding compounds maintain protein homeostasis during ageing and extend lifespan. *Nature* (2011). doi:10.1038/nature09873

205. Negi, H., Shukla, A., Khan, F. & Pandey, R. 3β-Hydroxy-urs-12-en-28-oic acid prolongs lifespan in C. elegans by modulating JNK-1. *Biochem. Biophys. Res. Commun.* **480**, 539–543 (2016).

206. Negi, H., Saikia, S. K. & Pandey, R. 3β-Hydroxy-urs-12-en-28-oic Acid Modulates Dietary Restriction Mediated Longevity and Ameliorates Toxic Protein Aggregation in C. elegans. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2017). doi:10.1093/gerona/glx118

207. Lewis, J. A. & Fleming, J. T. Basic culture methods. *Methods Cell Biol.* (1995).

208. Andrews, S. FastQC - A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. *Babraham Bioinforma.* (2010).

209. Broad Institute. Picard toolkit. *Broad Institute, GitHub repository* (2019).

210. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp352

211. Smedley, D. *et al.* The BioMart community portal: an innovative

alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589-98 (2015).

212. Lamarre, S. *et al.* Optimization of an RNA-seq differential gene expression analysis depending on biological replicate number and library size. *Front. Plant Sci.* (2018). doi:10.3389/fpls.2018.00108

213. Yang, J. S. *et al.* OASIS: Online application for the survival analysis of lifespan assays performed in aging research. *PLoS One* (2011). doi:10.1371/journal.pone.0023525

214. Han, S. K. *et al.* OASIS 2: Online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research. *Oncotarget* (2016). doi:10.18632/oncotarget.11269

215. Irwin, J. O. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *J. Hyg. (Lond).* (1949). doi:10.1017/S0022172400014443

216. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* (1958). doi:10.1080/01621459.1958.10501452

217. Bland, J. M. & Altman, D. G. Statistics Notes: Survival probabilities (the Kaplan-Meier method). *BMJ* (1998). doi:10.1136/bmj.317.7172.1572

218. Borisy, A. A. *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. U. S. A.* (2003). doi:10.1073/pnas.1337088100

219. Andersen, P. W. More is different. *Science (80-.).* (1972).

doi:10.1126/science.177.4047.393

220. Strogatz, S., Friedman, M., Mallinckrodt, A. J. & McKay, S. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. *Comput. Phys.* (1994). doi:10.1063/1.4823332

221. Lucanic, M., Garrett, T., Gill, M. S. & Lithgow, G. J. A simple method for high throughput chemical screening in caenorhabditis elegans. *J. Vis. Exp.* (2018). doi:10.3791/56892

222. Fong, S. *et al.* Energy crisis precedes global metabolic failure in a novel Caenorhabditis elegans Alzheimer Disease model. *Sci. Rep.* **6**, (2016).

223. Teo, E. *et al.* A high throughput drug screening paradigm using transgenic Caenorhabditis elegans model of Alzheimer's disease. *Transl. Med. Aging* **4**, 11–21 (2020).

224. Keith, S. A., Amrit, F. R. G., Ratnappan, R. & Ghazi, A. The C. elegans healthspan and stress-resistance assay toolkit. *Methods* **68**, 476–486 (2014).

225. Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans. *Nature* **419**, 808–14 (2002).

226. Chollet, F. & Allaire, J. J. R Interface to Keras. *GitHub. https://github.com/rstudio/keras* (2017).

227. Shin, H.-C., Orton, M., Collins, D. J., Doran, S. & Leach, M. O. Organ Detection Using Deep Learning. in *Medical Image Recognition, Segmentation and Parsing* (2016). doi:10.1016/b978-0-12-802581-9.00007-x

228. Albawi, S., Mohammed, T. A. & Al-Zawi, S. Understanding of a convolutional neural network. in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017* (2018). doi:10.1109/ICEngTechnol.2017.8308186

229. Hashimoto, T., Horikawa, M., Nomura, T. & Sakamoto, K. Nicotinamide adenine dinucleotide extends the lifespan of Caenorhabditis elegans mediated by sir-2.1 and daf-16. *Biogerontology* (2010). doi:10.1007/s10522-009-9225-3

230. Evason, K., Huang, C., Yamben, I., Covey, D. F. & Kornfeld, K. Anticonvulsant medications extend worm life-span. *Science* **307**, 258–262 (2005).

231. Zhang, J., Lu, L. & Zhou, L. Oleanolic acid activates daf-16 to increase lifespan in Caenorhabditis elegans. *Biochem. Biophys. Res. Commun.* **468**, 843–849 (2015).

232. Dessale, T. *et al.* Doubling healthy lifespan using drug synergy. *BioRxiv* (2017).

233. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0506580102

234. Kanehisa, M. *et al.* Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gkt1076

235. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids*

Res. (2020). doi:10.1093/nar/gkz1031

236. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).

237. Tacutu, R. *et al.* Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids Res.* **46**, (2018).

238. Quirk, J. P. & Saposnik, R. Admissibility and measurable utility functions. *Rev. Econ. Stud.* (1962). doi:10.2307/2295819

239. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* (2014). doi:10.1109/TVCG.2014.2346248

240. Hahsler, M., Hornik, K. & Buchta, C. Getting things in order: An introduction to the R package seriation. *J. Stat. Softw.* (2008). doi:10.18637/jss.v025.i03

241. Walker, G. a & Lithgow, G. J. Lifespan extension in C. elegans by a molecular chaperone dependent upon insulin-like signals. *Aging Cell* **2**, 131–139 (2003).

242. Hsu, A. L., Murphy, C. T. & Kenyon, C. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science (80-.).* (2003). doi:10.1126/science.1083701

243. Olsen, A., Vantipalli, M. C. & Lithgow, G. J. Lifespan extension of Caenorhabditis elegans following repeated mild hormetic heat treatments. *Biogerontology* (2006). doi:10.1007/s10522-006-9018-x

244. Yee, Z., Lim, S. H. Y., Ng, L. F. & Gruber, J. Inhibition of mTOR

decreases insoluble proteins burden by reducing translation in C. elegans. *Biogerontology* (2020). doi:10.1007/s10522-020-09906-7

245. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* (2013). doi:10.1016/j.cell.2013.05.039

246. Chen, W., Rezaizadehnajafi, L. & Wink, M. Influence of resveratrol on oxidative stress resistance and life span in Caenorhabditis elegans. *J. Pharm. Pharmacol.* (2013). doi:10.1111/jphp.12023

247. Meyer, D. H. & Schumacher, B. A transcriptome based aging clock near the theoretical limit of accuracy. *bioRxiv* (2020). doi:10.1101/2020.05.29.123430

248. Reis-Rodrigues, P. *et al.* Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. *Aging Cell* (2012). doi:10.1111/j.1474-9726.2011.00765.x

249. Lans, H. & Jansen, G. Multiple sensory G proteins in the olfactory, gustatory and nociceptive neurons modulate longevity in Caenorhabditis elegans. *Dev. Biol.* (2007). doi:10.1016/j.ydbio.2006.11.028

250. Oh, S. W. *et al.* JNK regulates lifespan in Caenorhabiditis elegans by modulating nuclear translocation of forkhead transcription factor/DAF-16. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0500749102

251. Oliveira, R. P. *et al.* Condition-adapted stress and longevity gene regulation by Caenorhabditis elegans SKN-1/Nrf. *Aging Cell* (2009). doi:10.1111/j.1474-9726.2009.00501.x

252. Xu, X. & Kim, S. K. The GATA Transcription Factor egl-27 Delays

Aging by Promoting Stress Resistance in Caenorhabditis elegans. *PLoS Genet.* (2012). doi:10.1371/journal.pgen.1003108

253. Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a Caenorhabditis elegans genomic RNAi screen. *PLoS Genet.* (2005). doi:10.1371/journal.pgen.0010017

254. Nehrke, K. A Reduction in Intestinal Cell pHi Due to Loss of the Caenorhabditis elegans Na+/H+ Exchanger NHX-2 Increases Life Span. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M307351200

255. Curran, S. P. & Ruvkun, G. Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS Genet.* **3**, 0479–0487 (2007).

256. Tsang, W. Y., Sayles, L. C., Grad, L. I., Pilgrim, D. B. & Lemire, B. D. Mitochondrial Respiratory Chain Deficiency in Caenorhabditis elegans Results in Developmental Arrest and Increased Life Span. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M103999200

257. Keith, S. A., Amrit, F. R. G., Ratnappan, R. & Ghazi, A. The C. elegans healthspan and stress-resistance assay toolkit. *Methods* **68**, 476–86 (2014).

258. Sutphin, G. L. *et al.* Caenorhabditis elegans orthologs of human genes differentially expressed with age are enriched for determinants of longevity. *Aging Cell* (2017). doi:10.1111/acel.12595

259. Zuryn, S., Kuang, J., Tuck, A. & Ebert, P. R. Mitochondrial dysfunction

in Caenorhabditis elegans causes metabolic restructuring, but this is not linked to longevity. *Mech. Ageing Dev.* (2010). doi:10.1016/j.mad.2010.07.004

260. Hamilton, B. *et al.* A systematic RNAi screen for longevity genes in C. elegans. *Genes and Development* **19**, 1544–1555 (2005).

261. Mansfeld, J. *et al.* Branched-chain amino acid catabolism is a conserved regulator of physiological ageing. *Nat. Commun.* (2015). doi:10.1038/ncomms10043

262. Arantes-Oliveira, N., Apfeld, J., Dillin, A. & Kenyon, C. Regulation of life-span by germ-line stem cells in Caenorhabditis elegans. *Science (80-.).* (2002). doi:10.1126/science.1065768

263. Van Voorhies, W. A. Production of sperm reduces nematode lifespan. *Nature* (1992). doi:10.1038/360456a0

264. Kim, Y. & Sun, H. Functional genomic approach to identify novel genes involved in the regulation of oxidative stress resistance and animal lifespan. *Aging Cell* **6**, 489–503 (2007).

265. Palmitessa, A. & Benovic, J. L. Arrestin and the multi-PDZ domain-containing protein MPZ-1 interact with phosphatase and tensin homolog (PTEN) and regulate Caenorhabditis elegans longevity. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M110.104612

266. Chondrogianni, N., Georgila, K., Kourtis, N., Tavernarakis, N. & Gonos, E. S. 20S proteasome activation promotes life span extension and resistance to proteotoxicity in Caenorhabditis elegans. *FASEB J.* (2015).

doi:10.1096/fj.14-252189

267. Tacutu, R. *et al.* Prediction of C. elegans Longevity Genes by Human and Worm Longevity Networks. *PLoS One* **7**, (2012).

268. Ayyadevara, S. *et al.* Life span and stress resistance of Caenorhabditis elegans are differentially affected by glutathione transferases metabolizing 4-hydroxynon-2-enal. *Mech. Ageing Dev.* (2007). doi:10.1016/j.mad.2006.11.025

269. Ayyadevara, S. *et al.* Lifespan and stress resistance of Caenorhabditis elegans are increased by expression of glutathione transferases capable of metabolizing the lipid peroxidation product 4-hydroxynonenal. *Aging Cell* (2005). doi:10.1111/j.1474-9726.2005.00168.x

270. Pang, S. & Curran, S. P. Adaptive capacity to bacterial diet modulates aging in C. elegans. *Cell Metab.* (2014). doi:10.1016/j.cmet.2013.12.005

271. Paradis, S., Ailion, M., Toker, A., Thomas, J. H. & Ruvkun, G. A PDK1 homolog is necessary and sufficient to transduce AGE-1 PI3 kinase signals that regulate diapause in Caenorhabditis elegans. *Genes Dev.* (1999). doi:10.1101/gad.13.11.1438

272. Robida-Stubbs, S. *et al.* TOR signaling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO. *Cell Metab.* **15**, 713–724 (2012).

273. Meléndez, A. *et al.* Autophagy genes are essential for dauer development and life-span extension in C. elegans. *Science (80-.).* (2003). doi:10.1126/science.1087782

274. Kim, Y. C. & Guan, K. L. MTOR: A pharmacologic target for autophagy regulation. *Journal of Clinical Investigation* (2015). doi:10.1172/JCI73939

275. Schafer, J. C. *et al.* IFTA-2 is a conserved cilia protein involved in pathways regulating longevity dauer formation in Caenorhabditis elegans. *J. Cell Sci.* (2006). doi:10.1242/jcs.03187

276. Chen, D., Pan, K. Z., Palter, J. E. & Kapahi, P. Longevity determined by developmental arrest genes in Caenorhabditis elegans. *Aging Cell* **6**, 525–533 (2007).

277. Wang, N. *et al.* miR-124/ATF-6, a novel lifespan extension pathway of Astragalus polysaccharide in Caenorhabditis elegans. *J. Cell. Biochem.* (2015). doi:10.1002/jcb.24961

278. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in Caenorhabditis elegans. *Nature* (1999). doi:10.1038/45544

279. Kim, Y. & Sun, H. Functional genomic approach to identify novel genes involved in the regulation of oxidative stress resistance and animal lifespan. *Aging Cell* (2007). doi:10.1111/j.1474-9726.2007.00302.x

280. Seung, W. O. *et al.* Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nat. Genet.* (2006). doi:10.1038/ng1723

281. Wang, Y. *et al.* C. elegans 14-3-3 proteins regulate life span and interact with SIR-2.1 and DAF-16/FOXO. *Mech. Ageing Dev.* (2006). doi:10.1016/j.mad.2006.05.005

282. Dong, M. Q. *et al.* Quantitative mass spectrometry identifies insulin signaling targets in C. elegans. *Science (80-.).* (2007). doi:10.1126/science.1139952

283. Ch'ng, Q., Sieburth, D. & Kaplan, J. M. Profiling synaptic proteins identifies regulators of insulin secretion and lifespan. *PLoS Genet.* (2008). doi:10.1371/journal.pgen.1000283

284. Kawli, T., Wu, C. & Tan, M. W. Systemic and cell intrinsic roles of Gqα signaling in the regulation of innate immunity, oxidative stress, and longevity in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.0914715107

285. Zhang, Y., Shao, Z., Zhai, Z., Shen, C. & Powell-Coffman, J. A. The HIF-1 hypoxia-inducible factor modulates lifespan in C. elegans. *PLoS One* (2009). doi:10.1371/journal.pone.0006348

286. Chen, D., Thomas, E. L. & Kapahi, P. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in Caenorhabditis elegans. *PLoS Genet.* (2009). doi:10.1371/journal.pgen.1000486

287. Mehta, R. *et al.* Proteasomal regulation of the hypoxic response modulates aging in C. elegans. *Science (80-.).* (2009). doi:10.1126/science.1173507

288. Pan, K. Z. *et al.* Inhibition of mRNA translation extends lifespan in Caenorhabditis elegans. *Aging Cell* (2007). doi:10.1111/j.1474-9726.2006.00266.x

289. Hansen, M. *et al.* Lifespan extension by conditions that inhibit translation

in Caenorhabditis elegans. *Aging Cell* **6**, 95–110 (2007).

290. Ayyadevara, S. *et al.* Rec-8 dimorphism affects longevity, stress resistance and X-chromosome nondisjunction in C. elegans, and replicative lifespan in S. cerevisiae. *Front. Genet.* (2014). doi:10.3389/fgene.2014.00211

291. Bennett, C. F. *et al.* Activation of the mitochondrial unfolded protein response does not predict longevity in Caenorhabditis elegans. *Nat. Commun.* (2014). doi:10.1038/ncomms4483

292. Greer, E. L. *et al.* Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. *Nature* (2011). doi:10.1038/nature10572

293. Greer, E. L. *et al.* Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. *Nature* (2010). doi:10.1038/nature09195

294. Amrit, F. R. G., Ratnappan, R., Keith, S. A. & Ghazi, A. The C. elegans lifespan assay toolkit. *Methods* **68**, 465–475 (2014).

295. Petrascheck, M. & Miller, D. L. Computational analysis of lifespan experiment reproducibility. *Front. Genet.* (2017). doi:10.3389/fgene.2017.00092

296. Bansal, A., Zhu, L. J., Yen, K. & Tissenbaum, H. A. Uncoupling lifespan and healthspan in Caenorhabditis elegans longevity mutants. *Proc. Natl. Acad. Sci.* **112**, E277–E286 (2015).

297. Wieland, T. Reat: A regional economic analysis toolbox for r. *Region* (2019). doi:10.18335/region.v6i3.267

298. Borisy, A. a *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7977–7982 (2003).

299. Pitt, J. N. *et al.* WormBot, an open-source robotics platform for survival and behavior analysis in C. elegans. *GeroScience* (2019). doi:10.1007/s11357-019-00124-9

300. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2016). doi:10.1109/CVPR.2016.91

301. David, H. E. *et al.* Construction and evaluation of a transgenic hsp16-GFP-lacZ Caenorhabditis elegans strain for environmental monitoring. *Environ. Toxicol. Chem.* (2003). doi:10.1002/etc.5620220114

302. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* (2008). doi:10.1038/nmeth.1226

303. Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* (2008). doi:10.1038/nmeth.1223

304. Yao, F., Coquery, J. & Lê Cao, K. A. Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets. *BMC Bioinformatics* (2012). doi:10.1186/1471-2105-13-24

305. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13091–6

(1998).

306. Hawley, S. A. *et al.* Use of cells expressing γ subunit variants to identify diverse mechanisms of AMPK activation. *Cell Metab.* (2010). doi:10.1016/j.cmet.2010.04.001

307. Gabrawy, M. M. *et al.* Lisinopril Preserves Physical Resilience and Extends Life Span in a Genotype-Specific Manner in Drosophila melanogaster. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* (2019). doi:10.1093/gerona/glz152

308. Liao, V. H. C. *et al.* Curcumin-mediated lifespan extension in Caenorhabditis elegans. *Mech. Ageing Dev.* (2011). doi:10.1016/j.mad.2011.07.008

309. Hekler, A. *et al.* Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur. J. Cancer* (2019). doi:10.1016/j.ejca.2019.06.012

310. Mathur, R., Rotroff, D., Ma, J., Shojaie, A. & Motsinger-Reif, A. Gene set analysis methods: A systematic comparison. *BioData Min.* (2018). doi:10.1186/s13040-018-0166-8

311. Tullet, J. M. A. *et al.* Direct Inhibition of the Longevity-Promoting Factor SKN-1 by Insulin-like Signaling in C. elegans. *Cell* (2008). doi:10.1016/j.cell.2008.01.030

312. Bishop, N. A. & Guarente, L. Two neurons mediate diet-restriction-induced longevity in C. elegans. *Nature* (2007). doi:10.1038/nature05904

313. Ko, L. J. & Engel, J. D. DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* (1993). doi:10.1128/mcb.13.7.4011

314. Zheng, Y. *et al.* A Review of the Pharmacological Action of Astragalus Polysaccharide. *Frontiers in Pharmacology* (2020). doi:10.3389/fphar.2020.00349

315. Sohangir, S. & Wang, D. Improved sqrt-cosine similarity measurement. *J. Big Data* (2017). doi:10.1186/s40537-017-0083-6

316. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* (2020). doi:10.1038/s41592-019-0686-2