Understanding longevity sub-networks using network propagation algorithms

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Manopriya Chokkalingam
aus Coimbatore, India
2018

Berichterstatter: Prof. Dr. Andreas Beyer
(Gutachter)

Prof. Dr. Achim Tresch

April 23, 2018

Acknowledgements

First and foremost I would like to extend my sincere thanks to my advisor Prof.Dr.Andreas Beyer for his boundless support. This thesis would not have been possible without his guidance. Most often the discussions with him were new oppurtunities to tackle obstacles and vanishing the disappointments. He has also facilitated many collaborations that has deepened my research experience.

My sincere thanks to Prof.Dr.Achim Tresch for the joined lab-meetings with his group and the valuable suggestions on the mathematical problems. I would like to thank Prof.Dr.Thorsten Hoppe for kindly joining my PhD thesis committee.

Many thanks to all the collaborators and co-authors for their pleasant cooperation and successful projects. Especially Luke Tain, Özlem Karaley, Alessandro Ori, Roman Ulrich Müller and Reza for enriching me with their expert knowledge.

I thank all the group members Cédric, Robert, Viera, Jan, Tim, Maria, Micheal Kaminski, Konstantina, Ana, Antonis, Oliver, Janis, Corinna and Kay for creating such a supportive and pleasant working environment. I would also like to thank the previous lab members Mathiéu and Marius for helping me especially during my starting days of PhD. I am also grateful to Ronja Johnen for her research support during the final and crucial phase of my PhD.

My sincere thanks to the coordinators of the graduate school for biological sciences, especially Kathy and Dr. Isabell Witt for their constant support starting from my early days of PhD untill now.

I would like to thank all my Indian friends, Kavya, Kanimozhi, Hemal, Subha and Ramesh for providing the homely feel and making the missed festivals a momentous ones in Germany.

My family, Amma, Appa, Ram, Santhosh, Nithu, Periamma and Periappa for their trust, support, love and confidence always kept me pushing to achieve the best in me. I would like to mention my friend, who is now my better half in all sense, Bala. His boundless encouragements and his passion in science has been a true inspiration for me. I dedicate this to my entire family.

Summary

Pioneering works on the laboratory model organisms has identified multiple evolutionarily conserved signalling pathways in regulation of longevity. High-throughput experiments have identified the genes and pathways that are regulated by these longevity interventions. However the high-throughput experiments yields a long list of significantly differentially regulated genes posing the challenge in identifying the causal genes of the phenotype. Network approaches are powerful resources in discovering genes and modules that are associated with the phenotype. Network propagation is a powerful systems biology approach that works through the principle that genes driving the same phenotype tend to interact closely, thus the gene signal is spread on the network to amplify the phenotype associated genes and modules. It is a global scoring method with various applications such as protein function prediction, inferring condition specifically altered sub-networks and prioritising the genes.

Various mathematical formulations for network propagation exists, including Random walks, Random walk with restart and Heat diffusion. In this thesis we have systematically analysed the performance of RWR and HD algorithms using the *Rattus norvegicus* ageing mRNA and protein abundance data from two different metabolically active tissues. We observed that depending on the network normalization approach and the nature of input scores, the propagated scores are biased by topology of the network ("topology bias"). In the algorithms of network propagation, spreading coefficient (α or 't') governs the amount and distance of signal spread in the network, is a tuning parameter thus it becomes extremely important to assess its impact on the propagated scores. In this study we have compared the two algorithms by employing a wide range of α and 't' parameters and demonstrate the existence of optimal spreading coefficients and their dependence on the input scores (initial states of the walker). Furthermore, we exemplified its utility and robustness in finding the altered sub-networks during ageing with gene expression and protein abundance datasets from brain and liver tissues.

Using *C. elegans* as a model we investigated the transcriptional responses to perturbations of insulin signalling, germline signalling, calorie uptake, and hypoxia. So far we lack insight into the extent to which these pathways impact on common molecular endpoints relevant for ageing or lifespan. Using traditional methods for the interpretation of transcriptomics data focussing on the responses of individual genes we observed only little similarity between

the perturbations. We thus employed network propagation for the detection of molecular networks that are consistently affected across conditions. This method rests on the notion that even if the same cellular function is targeted by different perturbations we often do not observe responses of the same genes. Instead, different perturbations may lead to the alteration of different genes acting in a common molecular sub-network. Our analysis revealed molecular sub-networks that were relevant for lifespan across multiple pathway perturbations. These networks included proteins involved in transcription, tRNA & rRNA processing, chromatin remodelling, stress resistance and reproduction and collagen suggesting the existence of common modules and converging downstream mechanisms that are involved in the lifespan control of *C. elegans*.

Furthermore we aimed to understand the tissue specific responses to reduced IIS (rIIS) and pinpoint the lifespan specific molecular mechanisms that are mediated by the transcription factor dfoxo using network propagation. To address this, dfoxo dependent and independent proteins were classified and the individual classes of proteins differential protein abundance p-values were propagated on the protein-protein interaction network of *Drosophila*. Subsequently, dfoxo dependent and independent network modules were revealed. Besides that network propagation was used to identify the common molecular signatures of rIIS from two genetic models that are associated with longevity.

Zusammenfassung

Wegweisende Arbeiten an Modellorganismen haben mehrere evolutionär konservierte Signalwege in der Regulation von Langlebigkeit identifiziert. Hochdurchsatz-Experimente haben Gene und Pfadwege identifiziert, die durch diese Langlebigkeits-Interventionen reguliert werden. Diese Hochdurchsatz-Experimente liefern lange Listen von signifikant differentiell regulierten Genen, wodurch die Identifizierung der kausalen Gene des Phänotyps eine Herausforderung darstellt. Netzwerk-Ansätze sind leistungsfähige Methoden zur Entdeckung von Genen und Modulen, die direkt mit dem Phänotypen assoziiert sind. Network Propagation ist ein systembiologischer Ansatz, der darauf basiert, dass Gene, die den gleichen Phänotyp auslösen, dazu neigen, eng zusammenzuarbeiten: das Signal einzelner Gene wird auf ein Netzwerk abgebildet, um die mit dem Phänotypen assoziierten Gene und Module zu verstärken. Network Propagation ist damit eine globale Scoring-Methode mit verschiedenen Anwendungen, wie der Vorhersage von Proteinfunktionen, der Identifizierung von spezifisch veränderten Sub-Netzwerken sowie Genprioritisierung. Es existieren verschiedene mathematische Formulierungen von Network Propagation, unter anderem Random Walks, Random Walks mit Neustart (RWR) und Heat Diffusion (HD). In dieser Arbeit haben wir systematisch die Leistung von RWR und HD Algorithmen unter Verwendung von *Rattus norvegicus* altersassoziierten mRNA- und Proteinexpressionsdaten in zwei metabolisch aktiven Geweben analysiert. Wir beobachteten, dass die propagierten Scores - abhängig von Netzwerk-Normalisierung und Art der Eingabedaten - von der Topologie des Netzwerkes beeinflusst werden ("Topologie-Bias"). In den Algorithmen der Network Propagation bestimmt der Streukoeffizient ("spreading coefficient", α oder "t") die Stärke und die Reichweite der Signalausbreitung im Netzwerk. Daher ist es wichtig, den Einfluss dieses Parameters auf die propagierten Scores einzuschätzen. In dieser Studie haben wir die beiden Algorithmen unter einer breiten Auswahl von α und 't' Parametern verglichen. Wir demonstrieren die Existenz von optimalen Streukoeffizienten und ihre Abhängigkeit von den Eingabewerten (Anfangszustände des Random Walks). Darüber hinaus zeigen wir die Anwendbarkeit und Robustheit von Network Propagation bei der Identifizierung veränderter Subnetzwerke während des Alterns, anhand von Genexpressions- und Proteinexpressionsdaten aus Gehirn- und Lebergeweben.

Im Modellorganismus *C. elegans* haben wir die Antwort des Transkriptoms auf Störungen der Insulin-Signalgebung, der Keimbahn-Signalübertragung, der

Kalorienaufnahme und der Hypoxie untersucht. Es ist eine offene Frage, inwiefern diese Pfadwege auf gemeinsame molekulare Endpunkte wirken, die für das Altern oder die Lebensspanne relevant sind. Unter Verwendung traditioneller Methoden zur Interpretation von Transkriptomdaten, welche sich auf die Antwort einzelner Gene konzentrierten, beobachteten wir nur eine geringe Ähnlichkeit zwischen unterschiedlichen Perturbationen. Wir verwendeten daher Network Propagation für die Identifizierung von molekularen Netzwerken, die konsistent durch die betrachteten Faktoren beeinflusst werden. Diese Methode beruht auf der Annahme, dass selbst bei Ansprache derselben zellulären Funktion durch Perturbationen häufig nicht dieselben Gene verändert werden. Stattdessen können verschiedene Perturbationen zur Veränderung verschiedener Gene führen, die in einem gemeinsamen molekularen Subnetzwerk agieren. Unsere Analyse identifizierte molekulare Subnetzwerke, welche die Lebensspanne durch Perturbation verschiedener Signalwege beeinflussen. Diese Netzwerke enthielten Proteine, die an Transkription, tRNA- und rRNA-Prozessierung, Chromatin-Remodellierung, Collagen, Stressresistenz und Reproduktion beteiligt sind, was auf die Existenz gemeinsamer Module und konvergenter Downstream-Mechanismen der Kontrolle der Lebensspanne in *C. elegans* hindeutet.

Ein weiteres Ziel dieser Arbeit war die Identifizierung gewebespezifischer Reaktionen auf reduzierte Aktivität des Insulin-Signalweges IIS (rIIS). Mithilfe von Network Propagation sollten die molekularen Mechanismen, die über den Transkriptionsfaktor dfoxo die Lebensspanne modulieren, präzise bestimmt werden. Dafür wurden Proteine in dfoxo-abhängige und unabhängige Proteine klassifiziert und die stärke der differentiellen Expression (p-Werte) wurden in einem Protein-protein-Interaktionsnetzwerk von *Drosophila* propagiert. Anschließend wurden dfoxo-abhängige und unabhängige Netwerkmodule bestimmt. Darüber hinaus wurde Network Propagtion verwendet, um die gemeinsamen molekularen Signaturen von rIIS in zwei mit Langlebigkeit assoziierten genetischen Modellen zu bestimmen.

Contents

List of Figures

List of Tables

1. Introduction

Ageing refers to decline in organismal fitness or tissue/cell health with time. It is a complex process caused by accumulation of molecular, cellular and organ damage leading to loss of physiological function and increased vulnerability to diseases,[56].[25] Inspite the complexity of aging, research works on laboratory model organisms has shown that single gene mutations can extend healthy lifespan through ameliorating age dependent loss of function,[62].[86] The first gene that was identified to increase lifespan of *C.elgegans* was age-1, catalytic subunit of class-1 phosphatidylinositol 3-kinase (PI3K),[46].[27] Most of the insights in ageing were obtained from the laboratory model organisms including budding yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* and the mouse. *C.elegans* is convenient model organism for studying the genetic basis of ageing as we have the potential to down-regulate the gene expression by feeding them bacteria that express double stranded RNA copies of gene of interest[16] and most interestingly 80% of its genes have human ortholog.[49] It has a normal lifespan of 2-3 weeks, due to its short lifespan several lifespan-extending mutants have been identified. The normal lifespan of *Drosophila melanogaster* is comparatively shorter living for 3 months and is easy to culture and maintain and possess powerful genetic tools and most importantly the full genomic sequence has been published (Drosophila melanogaster sequencing consortium 2000).

Research on the model organisms has demonstrated that there are commonalities of the aging process in these very different organisms. The same type of interventions can extend lifespan through improving health and function during ageing as well as protecting the animal from age-related diseases. Insulin/IGF-1 signaling, Target of Rapamycin (TOR) pathway and dietary restriction modulate lifespan in *C.elegans* and are conserved in higher organisms,[41],[50],[12],[108],[4].[89] This evolutionary conservation has robustly acclaimed for effects of nutrition and the molecular mechanisms involved in sensing of nutrients,[25],[42].[28] However the insulin/IGF-1 signaling regulate diverse physiological functions such as development, growth, stress resistance, metabolism, reproduction and most interestingly lifespan,[25].[42] This pleotropic effect of the lifespan extending pathways poses challenge in dissecting and identifying the mechanisms that are particularly associated in ameliorating aging.

1.1. High-throughput gene expression studies in ageing

Transcriptome profiling of *C.elegans* long lived mutants have identified the genes and the molecular mechanisms associated to lifespan extension,[33],[68],[85].[35] The gene expression studies of longevity mutants with the epistatic conditions has revealed molecular signatures such as stress resistance, antimicrobial response and reduced energy metabolism that ameliorate ageing. These high-throughput assays yields a long list of significantly differentially regulated genes posing a challenge in identifying the causal genes of the phenotype. Though these studies have pinpointed cellular functions that are specifically altered in long lived mutants, the common molecular responses across these longevity pathways still remains elusive. Quantitative proteomics study of germline loss mutants of *C.elegans* has found proteins involved in RNA processing, translation, protein folding and proteolysis were decreased and collagen proteins and innate immune response proteins were increased in their abundance.[77] Despite recent improvements, MS-based shotgun proteomics does not cover all proteins. In addition, measurements are affected by technical and biological noise. Network approaches are powerful resources in discovering genes and modules that are associated with the phenotype,[83].[8]

1.2. Network approaches

A simple and straight forward network analysis approaches are to predict all the neighbouring genes of regulated gene in the network as being associated to the phenotype[71] or calculation of the shortest path between potential and known disease proteins.[26] However these naive and simple approach would lead to false predictions i.e. genes connected to regulated genes through irrelevant interactions. Furthermore with this appraoches one would leave relevant genes (false negatives) that are not directly interacting with the regulated genes, eventhough such genes are well connected to regulated genes through multiple long distance interactions. Thus global network-similarity appraoches outperform these local distance measures.[43] This study focused on a method that accounts for the global network structure: network propagation has been studied in great detail. Network propagation relies on the principle that genes exerting the same phenotype are closely interacting,[61] hence spreading the signal on the network allows the identification of the altered pathways in a condition of study. The large protein-protein interaction data produced through high-throughput assays such as yeast two-hybrid assay and tandem

affinity purification (TAP) followed by mass spectrometry serve as a powerful resource for global network-based approaches.[87]

1.3. Network propagation

Network propagation works by combining each gene's score with scores of neighbouring genes in the network, it considers all possible paths between genes. Therefore gene prioritization with network propagation can overcome the false positive prediction and enrich the true causal genes that are connected through multiple paths.[18] Network propagation starts with the definition of query nodes i.e nodes that are altered in the condition of study from the experimental measurements and subsequently propagating certain fraction of query node scores to neighbouring nodes. The amount or fraction of spread is defined by the spreading coefficients. The spreadng coefficient corresponds to the amount and distance of signal spread in the global network. Propagation is done iteratively until the propagated scores converge on the network. Thus nodes without prior information will also gain scores after spreading and can be associated with the phenotype. Network propagation offers several advantages such as scoring the distant gene that that are not direct neighbors but still relevant for the phenotype and high scoring the genes that are well connected through many short paths to the regulated genes. Through signal spreading to the neighbouring genes network propagtion can be applied for imputing missing values,[20] main applications of network propagation are prioritization of causal genes with the input of known phenotype associated genes,[104] boosting signal-to-noise ratio by amplifying the signal and reducing the noise as well as inferring sub-networks or modules that are associated with the phenotype,[93].[57]

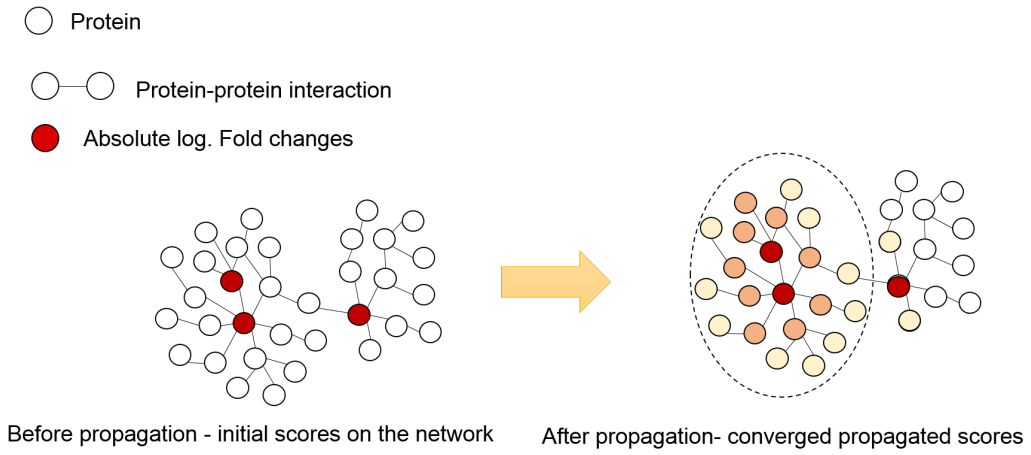

Figure 1: Illustration of network with scores before and after network propagation

1.4. Thesis aims

Exploring the different mathematical formulation of network propagation and understanding the nuances between the random walk with restart and heat diffusion formulations of network propagation. Further to study the impact of various graph normalization approaches with these formulations, finding the appropriate graph transformation for each formulation. To demonstrate the power of network propagtion in different application, we considered the identification of common functional modules of longevity using the transcriptome data from longevity mutants of *C.elegans*, besides that the candidate genes identified through propagation were experimentally validated with the support from our collaborators. Additionally to illustrate the specificity of network propagation in identifying the tissue specifically modified protein sub-networks of reduced insulin/IGF-1 genetic models of longevity from *Drosophila*.

1.4.1. Overview of the thesis

Chapter 2 covers the comparison of the two network propagation algorithms, random walk with restart and heat diffusion. The two algorithms were explored in great detail using the published ageing datasets from Ori et al., 2015.[70] R package BioNetSmooth was developed for implementing random walk with restart algorithm based network propagation.

Chapter 3 highlights the application of random walk with restart network propagation algorithm with longevity RNA-Sequencing datasets. We identified

the lonegivty sub-network of *C.elegans* and subsequently the funtional modules association with lifespan extension were experimentally validated using the RNAi assays.

Chapter 4 focuses on understanding the transcription factors interaction in regulation of lifespan extension in *C.elegans*. For this study the ChIP-Seq datasets of the X transcription factors were processed and targets of were predicted using the TFTargetCaller R package. The transcription factors activities were estimated using their target genes expression in the longevity mutants. Thus we identified the transcription factors that may be associated with the lifespan extension in worms.

Chapter 5 describes the application of network propagation in determining the tissue specific proteome alterations in reduced insulin signaling mutants of *Drosophila melanogaster*. Furthermore the integrated analysis of two reduced insulin signaling mutants pinointed the lifespan associated modules.

2. Network propagation algorithms: a detailed exploration

2.1. Introduction

Network propagation has become a common systems biology tool with wide range of applications such as protein function prediction, inferring condition specifically altered sub-networks and prioritising the disease genes.[18] It starts with the definition of query nodes, the genes that are known to be associated with the phenotype. Once the query nodes are assigned with a score on the apriori network, certain fraction of query node scores are spread to neighbouring nodes through the edges in network. The magnitude of spread to adjacent nodes is defined with spreading coefficient. As a result of spreading, genes neighbouring the query genes tend to accumulate similar scores. Subsequently the nodes propagated scores can be examined to identify altered sub-networks where the query nodes are closely connected to other nodes in the network. Furthermore network propagated scores can be ranked to prioritize the phenotype causal genes. There are various mathematical formulations exist in implementing the network propagation, such as random walk with restart (RWR),[43] heat diffusion (HD).[15] There is a vast list of tools available for network propagation, they include geneMania,[66] TieDie,[75] Diffusion[15] etc.

2.1.1. Random walk with restart

Random walk on graphs is an iterative walker that transitions randomly to the neighbouring node on graph with a probability α (spreading coefficient). In the weighted network, the transition to adjacent node is proportional to their weights (for STRING functional interaction network the combined scores can be used as edge weights). Random walk with restart is a variant of random walk with an option to transition back to the starting node at every time step that is given by a restart probability $(1 - \alpha)$. In the network propagation framework restart probability defines the magnitude of signal that has to be retained at the query nodes, can be seen as a dampening factor on long walks. RWR is an iterative step process, the node scores are propagated on network until the scores converge on the network.[104]

$$F^t = \alpha \cdot W \cdot F^{t-1} + (1 - \alpha) \cdot F^0. \quad (1)$$

Where F^0 is a vector representing the initial node scores, α is the spreading coefficient representing the fraction of spread, W is the normalized matrix of the interaction network, t represents the number of iterations of spread and F^t is the vector representing the converged propagated scores. The convergence is estimated using the norm of $F^t - F^{t-1}$ and when it is well below 10^{-6}.[37]

2.1.2. Heat diffusion

Heat diffusion is a continuous-time analogue of lazy random walks without restart and the spreading of signal is controlled by the time parameter. The nodes are assigned with the certain ratios of changes in their expression and at every time step the information from these nodes flows to the adjacent nodes. In heat diffusion the fraction of spread is defined with 't' time parameter, at every time step the node scores are propagated to their adjacent nodes.[15] When the time parameter is 0 the node scores are not propagated to adjacent nodes and when the 't' is infinite the solutions would represent the network topology with a loss in initial information. The heat diffusion equation is given by

$$F = F^0 \cdot exp^{-W \cdot t} \quad (2)$$

where h is the vector of initial query node scores, L is the normalized matrix representation of the interaction network, t is the time diffusion and is a tunable parameter and d is the vector representing network propagated node scores.

2.1.3. Study Aims

Here in this chapter we have systematically analyzed the performance of these two algorithms using the *Rattus norvegicus* ageing mRNA and protein abundance data from two different metabolically active tissues: Brain and liver at two age groups.[70] Firstly we have demonstrated the influence of the graph normalization techniques on the propagated scores with the utilization of two network propagation algorithms. Secondly, transcriptome replicates were utilized to illustrate the effect of spreading parameters in reducing the variances between the replicates and increasing the consistent signals. Finally

the significance of choosing the optimal spreading parameters were illustrated by showing the increase in consistency between transcriptome and proteomic changes of tissues as well as amplification of more tissue specific and condition specific cellular functions.

2.2. Datasets

For this detailed study on the network propagation algorithms, the published young and old tissues transcriptome and proteome datasets from *Rattus norvegicus* were used.[70] The accession number for the transcriptome is GEO: GSE66715 and for the proteome ProteomeXchange: PXD002467. The data were derived from young (6 months) and old (24 months) Rat liver and brain samples with three biological replicates for each age. The transcriptome profiling was from the entire tissues whereas the proteomics measurements were from four sub-cellular fractions (nuc:nuclei, pn1:mitochondrial, pn2:cytoplasmic membrane, sol:soluble cytosolic proteins). Thus to collate the mRNA and protein abundance changes, the protein log fold changes from four sub-cellular fractions of a tissue were combined by calculating the weighted means.

$$\text{weighted LFC}_{\text{fraction1}} = \frac{\text{LFC from fraction1}}{\text{no. of proteins measured in fraction1}} \tag{3}$$

$$\text{weighted average LFC} = \text{mean (weighted LFC from fraction1:4)} \tag{4}$$

These weighted average log2 fold changes were propagated on the network and subsequently mRNA and protein propagated scores were compared.

2.2.1. STRING functional interaction network

The functional interaction network for *Rattus norvegicus* was retrieved from STRING database version 10.5.[91] The generic network was filetered with the combined scores threshold of above 900 for high confidence interactions. This step of filtering was essential to avoid the flow of signal to the false positive edges. The filetered network had 9,747 nodes and 8,78,886 interactions, this network was employed for diffusing the transcriptome and proteome signals.

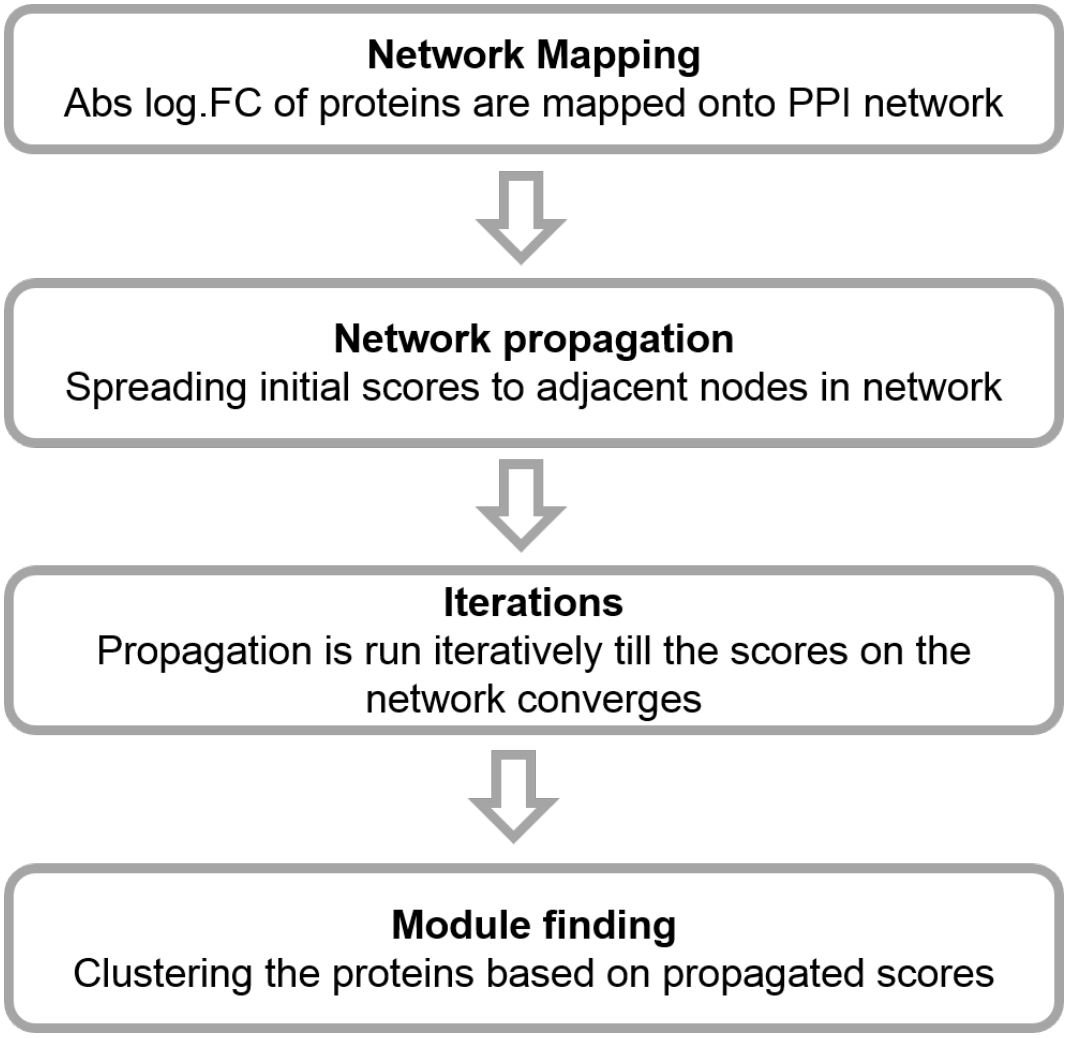

Figure 2: Workflow of network propagation analysis

2.3. Methods

2.3.1. Random Walk with Restart

Network propagation with random walk with restart was performed by using the R package BioNetSmooth, version 1.0.0. First the propagation starts by mapping the log fold changes of expression data or protein abundances on the protein-protein interaction network of *Rattus norvegicus* with the function network_mapping. Subsequently the function network_propagation propagates the mapped log fold changes on the network according to the equation 1. The progation runs iteratively with t=[1, 2, 3, ...] until convergence. The final scores are reached when the matrix norm of F^{t+1} - F^{t} fell below 10^{-6}.

2.3.2. Heat diffusion

Network propagation with heat diffusion was done according to the equation 2. First the network matrix W was multiplied with spreading paramter t with negative sign. Further exponential of the multiplied matrix was computed using the expm function from Matrix R package, version 1.2-10. The propgated scores were obtained by multiplying the log fold changes with the exponential of the matrix.

2.3.3. Graph normalization methods

For the network matrix W different graph normalizations can be employed. The simplest representation of a graph is the adjacency matrix $A = [a_{ij}]$. The entries a_{ij} of the matrix are defined by

$$a_{ij} = \begin{cases} 1, & \text{if } v_i \text{ is adjacent to } v_j \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

when G = (V,E) is a connected undirected graph with V as vertex set.

The adjacency matrix can be normalized by the degrees of nodes. In this case the entries are

$$a_{ij} = \begin{cases} \frac{1}{d_i}, & \text{if } v_i \text{ is adjacent to } v_j \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where d_i denotes the degree of vertex v_i.

The Laplacian transformation $L = [l_{ij}]$ of the graph can be used. It is defined by $L = D - A$, where A is the adjacency matrix and D is a diagonal matrix containing the nodes degrees. The entries l_{ij} of the matrix L are filled by the following equation

$$l_{ij} = \begin{cases} d_i, & \text{if } v_i = v_j \\ -1, & \text{if } v_i \neq v_j \text{ and } v_i \text{ is adjacent to } v_j \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

The Laplacian matrix is normalized by the degrees of the interacting nodes. In this case the entries l_{ij} of the matrix are defined by

$$l_{ij} = \begin{cases} 1, & \text{if } v_i = v_j \\ \dfrac{-1}{\sqrt{d_i \cdot d_j)}}, & \text{if } v_i \neq v_j \text{ and } v_i \text{ is adjacent to } v_j \\ 0, & \text{otherwise} \end{cases} . \quad (8)$$

2.3.4. Gene ontology analysis

topGO was employed for GO term enrichment analysis and *R.norvegicus* genes and protein annotations were retrieved from biomaRt R package, version 2.35.13. To identify enriched GO terms, the one-sided elim Fisher procedure was used ($\alpha \leqslant 0.05$; (Alexa, RahnenfÃijhrer, and Lengauer 2006). The enrichment score of a GO term is defined as log2 (#Detected significant genes/ #Expected significant genes). The network propagation results were interpreted by selecting the top 10% high scored genes based on the propagated scores from mRNA and protein log fold changes separately. For this analysis the network genes were ussed as reference background. Further the GO terms were filtered to contain atleast 10 genes in a term and having log2 enrichment scores above 1. The barplots represent the number of enriched GO terms after filtering based on above conditions from each method.

2.4. Results

2.4.1. Effects of graph normalization and topology bias

We have assessed the influence of different graph normalization schemes on the propagated scores with two different network propagation algorithms: Random walk with restart (RWR) and Heat diffusion (HD). For our study, the functional interaction network G = (V,E) of *Rattus norvegicus* was retrieved from STRING database and filtered for high confidence edges using combined scores ≥ 900. To propagate the gene scores, the network has to be normalized using their degrees or interacting nodes degrees. The network was first converted to an adjacency matrix (A) and subsequently normalized with the following schema: Laplacian, normalized Laplacian and degree normalized adjacency. Laplacian of the functional interaction network was created by subtracting the Degree matrix with adjacency matrix of G. The normalized Laplacian was created by dividing the adjacency matrix with square root product of degrees of interacting nodes. Degree normalized adjacency of the network was created by multiplying

the adjacency matrix with inverse of the column sums on the diagonal. For heat diffusion the Laplacian transformed network (D-A) was utilized and the log2 fold changes of genes expression were propagated with spreading parameter 't'=0.6. To determine the influence of graph normalization scheme on the propagated scores, hub and non-hub nodes were classified based on their degrees. The top 10% high degreed nodes were classified as hub nodes and the bottom 10% low degreed nodes as non-hub nodes. Then hub and non-hub nodes starting scores and propagated scores were compared. The Laplacian transformation has an intrinsic nature of biasing the node scores in favour of hub nodes due to their higher connections in network. Hub nodes on average tend to gather higher scores than initial scores just by topology of network (Figure 1A). Similarly for heat diffusion, degree normalized adjacency of the network (A/D) was employed for propagating genes expression changes with spreading parameter 't'=0.6. Followed by network propagation, the hub and non-hub node scores were evaluated. Most interestingly the degree normalized adjacency transformed network has not lead to topology bias on the propagated scores (Figure 1B). Further we evaluated the graph normalization approaches with random walk with restart algorithm, the normalized Laplacian of the filtered functional interaction network was employed to propagate scores using spreading coefficient of X until convergence. Subsequently hub and non-hub nodes scores were compared, the hub nodes tend to accumulate higher scores compared to the non-hub nodes (Figure 1C). Additionally we tested degree normalized adjacency for RWR, interestingly the topology bias was not observed on the propagated scores (Figure 1D). Therefore for successive analysis we have employed degree normalized adjacency of the network for propagating expression changes of the old tissues of rat.

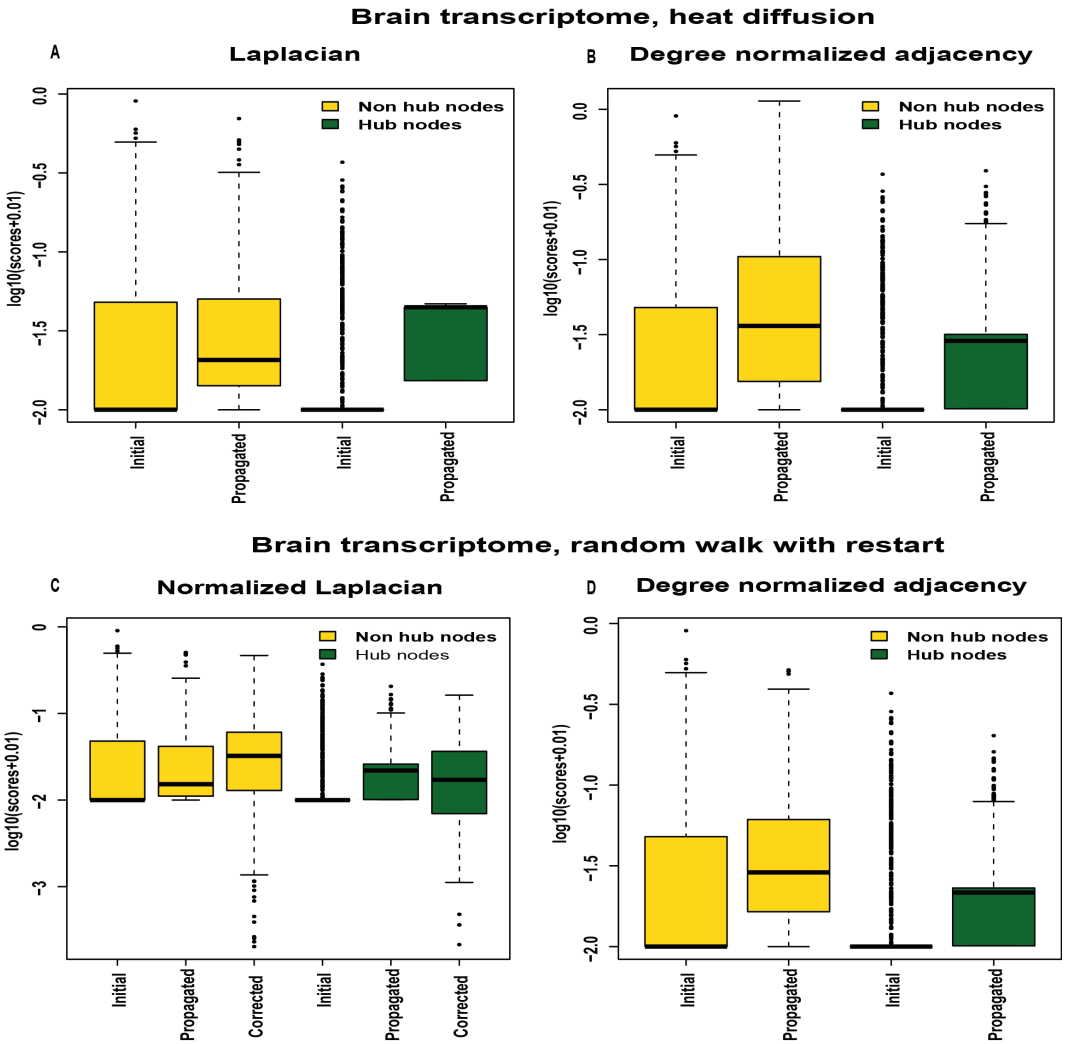

Figure 3: For RWR and HD, depending on the graph normalization scheme the network propagated scores are biased in favour of hub nodes. A) For the heat diffusion algorithm, use of Laplacian transformed network leads to accumulation of higher scores on hub nodes just by topology of the network. As on average the network propagated scores on hub nodes are much higher compared to the propagated scores on non-hub nodes. B) Whereas employing degree normalized adjacency (AD−1) matrix of network has similar propagated scores on hub and non-hub nodes on average. C) For the RWR algorithm, using normalized Laplacian matrix representation of network for propagation leads to a bias on hub nodes as illustrated. D) Whereas with the utilization of degree normalized adjacency (AD−1) of network has similar propagated scores on hub and non-hub nodes on average. Thus the degree normalized adjacency transformed matrix was utilized for propagation of log2 fold changes for both the algorithms and subsequently the results were compared.

Table 1: Different graph normalization approaches and their impact on propagated scores.

Method	Graph normalization	Topology bias	Correction
HD	laplacian	Yes	-
HD	normalized laplacian	No	-
HD	degree normalized adjacency	No	-
RWR	laplacian	-	-
RWR	normalized laplacian	Yes	mean smoothed scores
RWR	degree normalized adjacency	No	-

2.4.2. Finding optimal spreading parameters- Within-dataset consistency

The crucial parameter in propagation is the spreading coefficient, which governs the amount and distance of signal spread on interaction network. Since the spreading parameter is tunable for both algorithms we have determined the impact of spreading parameter on propagated scores. For random walk with restart, parameter α defines the fraction of node scores that has to be propagated to adjacent nodes and it ranges between 0-1. Likewise in heat diffusion the parameter 't' corresponds to the fraction of diffusion and it ranges from 0 to infinity. Reduction in variances of the replicates from transcriptome has been utilized as a criterion in estimating the impact of spreading parameters on the propagated scores. Most often we only have one level of expression data, either transcriptome or proteome thus it becomes essential to find approaches for assessing the effect of spreading parameters with the given datasets. Therefore in this study we have utilized the transcriptome data alone and used variances within the replicates as a measure to assess the range of spreading parameters.

Transcriptome of brain and liver tissues at their young and old ages were profiled with three biological replicates by Ori et al 2015. For the differential expression analysis of old vs young tissue samples a generalized linear model

was constructed using DESeq2. Similarly for every replicate of a condition differential expression analysis was carried out with all replicates of other condition (old1 vs young), this way individual replictaes variances are retained. Subsequently the mean log fold changes and the replicate-wise log fold changes for each tissue were propagated on the network. For this propagation process, the heat diffusion and random walk with restart algorithms were employed with varying spreading parameters α and 't'. To assess the influence of spreading parameters the propagated scores from mean log fold changes were correlated with the replicate-wise propagated scores for each α and 't' respectively for RWR and HD (Figure 2). With heat diffusion of brain log fold changes, the maximum correlation between replicates was observed at the spreading parameter t of 0.6. Suggesting that the consistency between the replicates were improved by reducing the replicate specific variances. Additionally suggesting the maximum limit for propagation of brain datasets with heat diffusion (Figure 2A). For the liver samples, there is a little gain of information after propagation as we observe only a slight improvement in the correlation with increasing signal spread on the network (Figure 2B). With RWR, for Brain samples the maximum correlation is reached only near 1, this may indicate that propagation has not improved the consistency between the replicates (Figure 2C). For liver samples replicate-wise propagated scores have maximal correlation at α 0.5. Suggesting this could be the optimal spreading coefficient for liver samples (Figure 2D). Thus this analysis provides a valuable information on the amount of propagation that has to be done for the datasets.

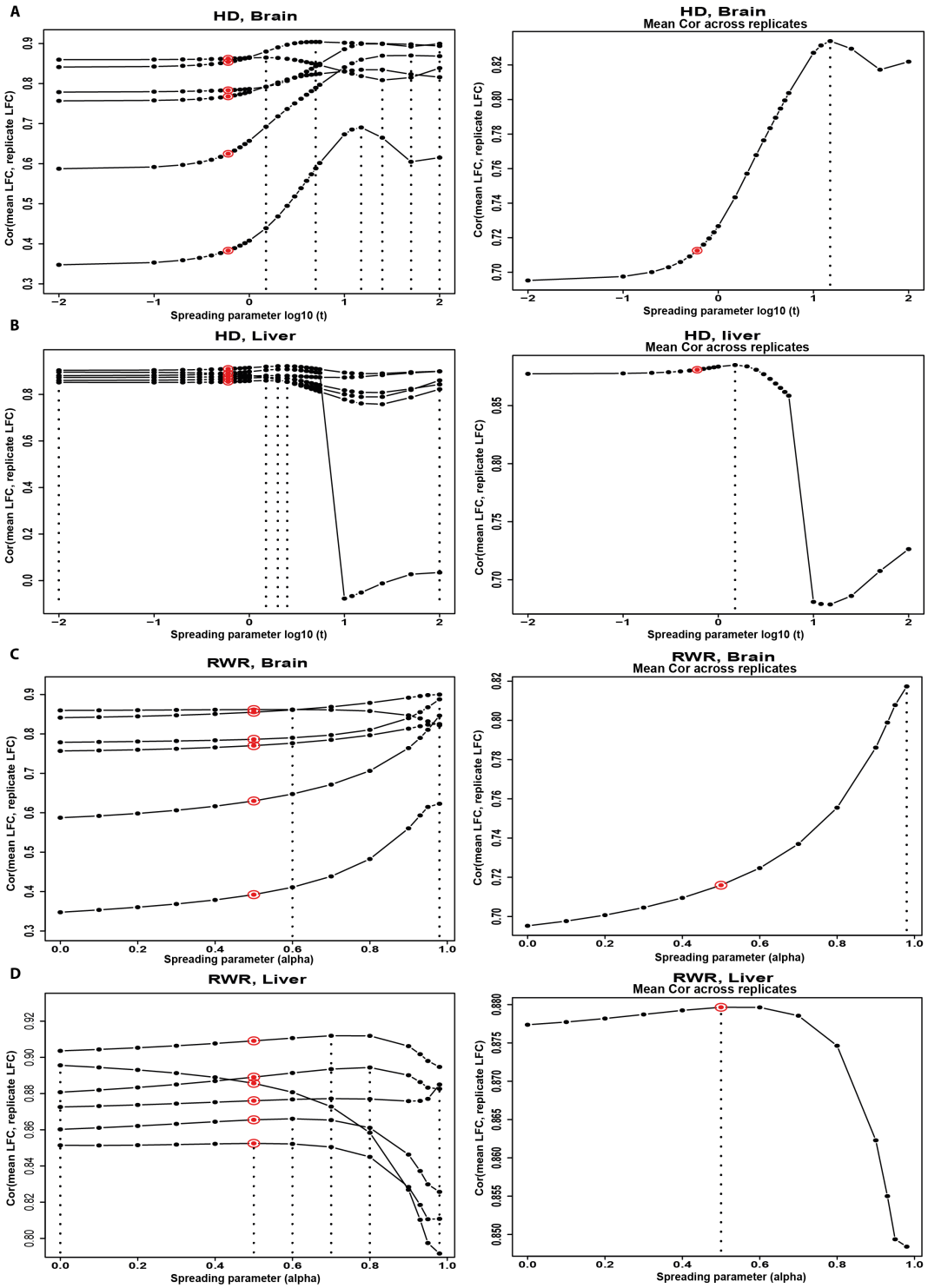

Figure 4: continued on the following page

Figure 4: Correlation of replicate-wise propagated log fold changes with propagated avg.log2 fold changes. A.Correlation values from replicate-wise propagated log fold changes with mean log2 fold changes of Brain samples with varying t in heat diffusion. B.Correlation values from heat diffusion of replicate-wise propagated scores and mean propagated scores of Liver samples. C.Brain samples RWR propagated scores from replicates and all old vs young correlation values. D. Liver samples correlation values across varying α. The red dots denote the optimal spreading parameters from the between dataset i.e. mRNA and protein propagated scores correlations, dotted lines represent the individual replicates maximal correlation, the right panel represents the average correlations from replicates at each spreading parameter for HD and RWR.

2.4.3. Finding optimal spreading parameters- Between-dataset consistency

The relationship between protein and mRNA levels in the ageing tissues of Rattus norvegicus is utilized as an yet another criterion in assessing the impact of the spreading parameters.[54] To collate the mRNA and proteins expression changes of old tissues vs young, the protein log2 fold changes from four sub-cellular fractions of a tissue were combined by calculating the weighted means (LFC Fraction1= LFC from fraction1 / # proteins measured in fraction1 subsequently mean lfc values were estimated across the fractions). For each tissue mRNA and protein actual log2 fold changes were propagated independently on the degree normalized adjacency of network with varying spreading parameters, for RWR α starting from 0 to 0.98 and for HD 't' starting from 0 to 200 were tested. When the spreading parameter is 0, the scores are not propagated on the network. The correlation between mRNA and protein levels of ageing brain is 0.179 without network propagation. The correlations were calculated with the propagated mRNA and protein log2 fold changes of brain with varying spreading parameters for each propagation algorithm. Most interestingly we observe that the correlation of mRNA and protein changes in old brain tissue was improved with network propagation for a certain range of spreading coefficients (HD_Cor = 0.183; RWR_Cor = 0.184) (Figure 2A,B). Similarly for the liver tissue the correlation of transcriptome and proteome was enhanced after propagation, initial_Cor = 0.26 and after propagation HD_Cor = 0.263 and RWR_Cor = 0.264 (Figure 2C,D). This suggests that through propagating the signals from old tissues vs young, the network regions

that are consistently changing in their expression and protein abundance were amplified. However as the spreading coefficients increases we tend to lose the biological signal and amplify the network topology, thus the correlations once again increases as we are nearly close to 100% spreading of the initial signal on the network (most likely to be noise-noise correlation). Thus we hypothesize that the local maximum correlated spreading coefficients could be the optimal spreading parameters for these datasets on this interaction network.

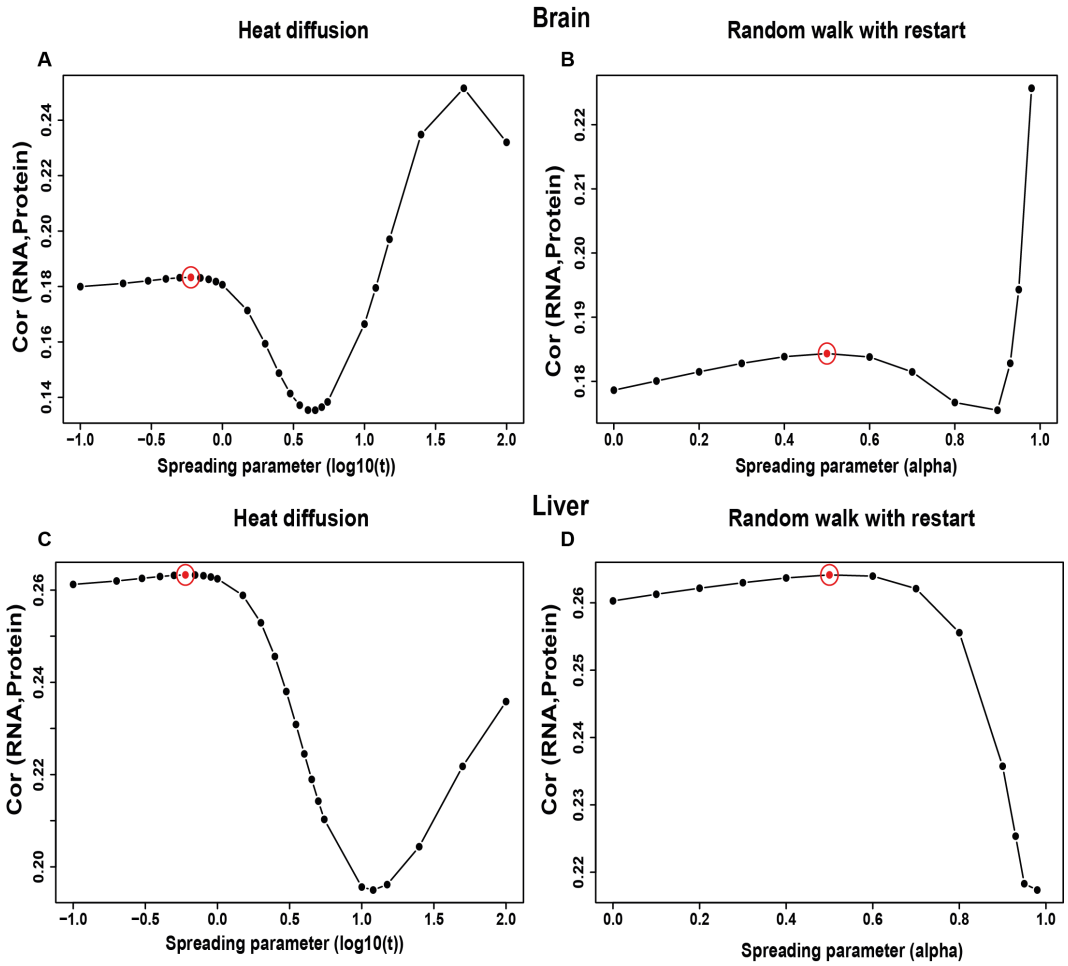

Figure 5: continued on the following page

Figure 5: Network propagation improves the correlation between mRNA and protein levels of ageing tissues (brain and liver). A & B. plot shows the correlations between propagated scores from mRNA log2 fold changes and protein weighted mean log2 fold changes from old brain samples with varying spreading coefficients for random walk with restart and heat diffusion. C & D. Correlations of propagated scores from mRNA and protein log fold changes of liver samples with HD and RWR. Correlations were calculated for the genes that are expressed and quantified with RNA-Sequencing and MS proteomics as well as present in the functional interaction network (n= 1741 genes).

2.4.4. Functional interpretation of the network propagation results

The increase in consistency between mRNA and protein levels with network propagation is likely due to amplification of genes involved in cellular functions that are specific for the tissue. In particular we expect that for brain samples the amplified genes should be relevant for brain functions like synaptic transmission and for liver tissue the amplified genes are expected to be relevant for liver specific functions such as fatty acid metabolism and amino acid catabolism. To assess this notion the GO enrichment analysis was performed on the top 10% of the high scored genes from network propagation separately for each tissue mRNA and proteome. The number of significantly enriched GO terms from network analysis were compared to the significantly enriched GO terms from the differentially expressed genes. We observed more number of significant GO terms with the network amplified genes compared to terms enriched from differentially expressed genes (Figure 3A). The increase in number of significant GO terms is expected, as we used a functional interaction network for node scores propagation. More interestingly, GO enrichment of high scored genes from propagation using optimal spreading parameters clearly showed more tissue specifically altered functions. Furthermore we conducted overlap analysis of top10% genes from propagation with known ageing related genes of Rat from JenAge database (http://agefactdb.jenage.de/). This intersection analysis also revealed that network amplified genes are indeed enriched for ageing associated genes compared to the differentially regulated genes. Likewise for the liver tissue we observe after network propagation, more number of significantly enriched tissue specific functional GO terms as well as more number of known ageing associated genes (Figure 3B). These observations supports that spreading parameters (t=0.6 in HD and α =0.5 in RWR) are the optimal spreading

parameters for these tissue expression and protein profiles with STRING functional interaction network.

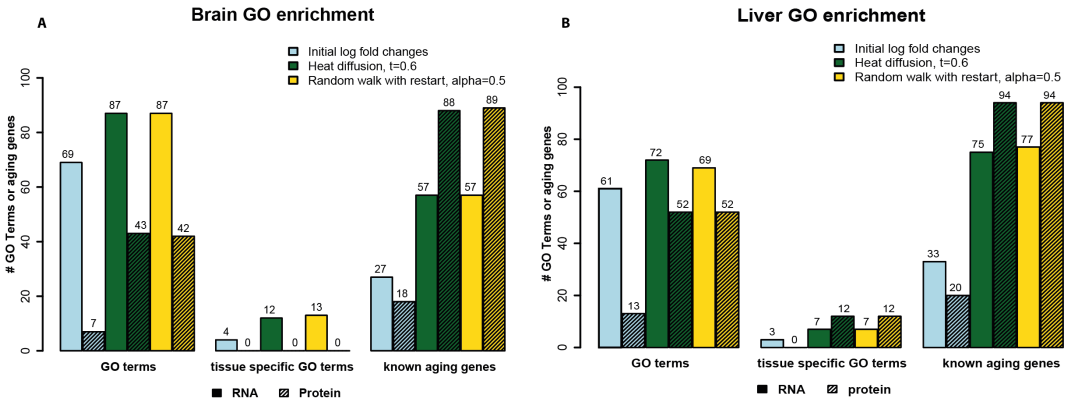

Figure 6: Amplification of more tissue specific GO terms and known ageing associated genes. A. Bar plot represents number of significantly enriched GO terms, number of tissue specific terms and no of ageing related genes in the top10% genes before and after network propagation with two algorithms for old brain samples. B. Liver samples number of significantly enriched GO terms, number of tissue specific terms and no of ageing related genes in the top10% genes before and after network propagation with two algorithms. The unshaded bars represent the no. of functional terms enriched in the transcriptome and shaded bars represent no. of functional terms enriched with the proteome.

2.4.5. Analysis of age-related proteome and transcriptome changes

The actual log fold changes of transcripts and proteins were propagated on the network, most of the genes retained their direction of change after propagation. Interestingly we observe that the correlation between mRNA and protein changes were slightly improved after propagation with use of a certain spreading parameters on both algorithms (Fig.5).

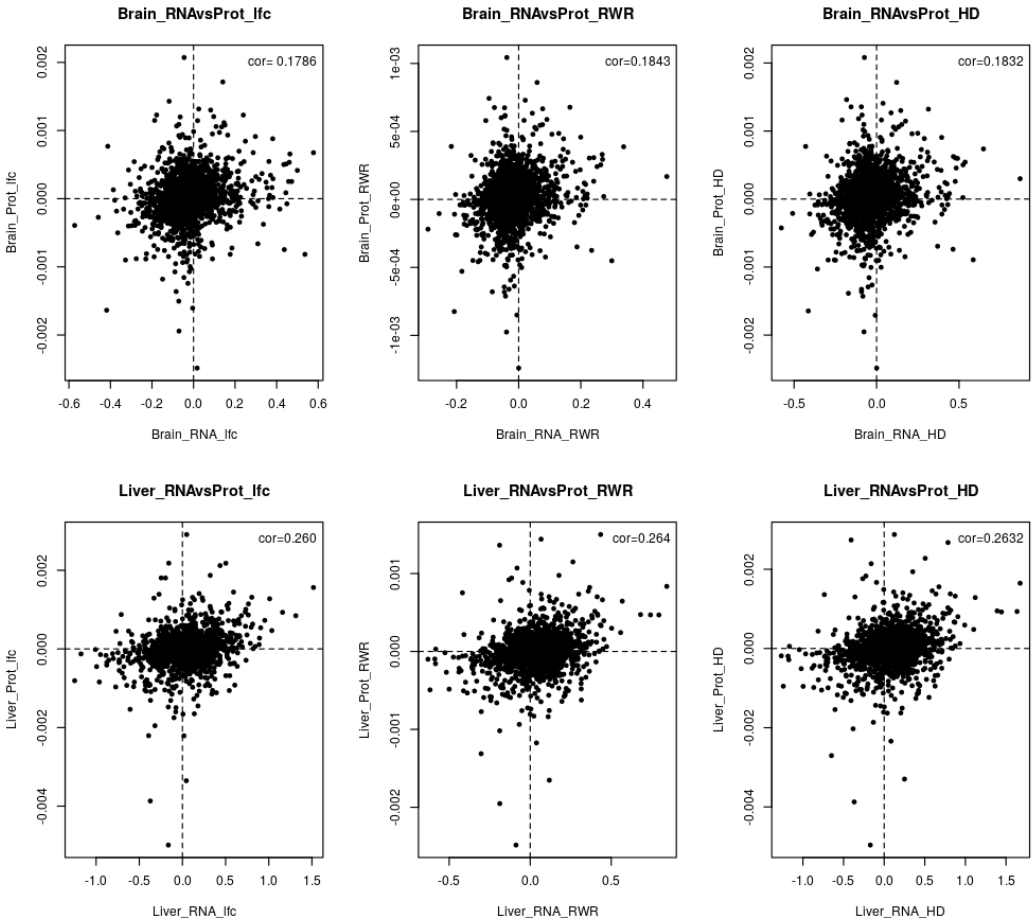

Figure 7: Correlation of mRNA and protein initial scores and propagated scores from RWR and HD. The top panel represents the mRNA and protein log2 fold changes from brain old vs young samples. The bottom panel corresponds to the mRNA and protein log2 fold changes from liver samples

Based on actual log fold changes of transcripts and proteins, the genes were categorized into concordant and discordant classes. Concordant genes are the genes whose mRNA log fold changes signs are matched with protein direction of change and discordant genes have the non-matching signs of log2 fold changes. Subsequently we performed gene ontology enrichment analysis on the individual catogories of genes to understand cellular functions that are concordantly changed and discordantly changed in individual tissues (Fig.6A). Glutathione metabolism (glutathione has an antioxidant role and might be activated as a response to increased stress during ageing) and response to toxic substance were consistently activated in both brain and liver old samples. Cellular amino acid metabolism was activated at both mRNA and protein levels

for old brain samples. Interestingly we observed that mitochondrial function related terms were inactivated in old brain samples. Network propagation improved the consistency between mRNA and protein changes in both tissues. However, certain sub-networks might still show discordant behaviour i.e. mRNA positively regulated and protein negatively regulated and vice versa. To check this notion we have categorized the genes into two classes: concordant and discordant genes. This time genes were categoried based on signs of propagated scores. From the Gene ontology enrichment analysis of the individual categories of genes, the top 3 terms based on their lowest p-value from elim fisher test are represented in the heatmap (Fig.6B). The genes with increased mRNA expression and protein abundances across the sub-cellular fractions correspond to processes such as Glutathione metabolism, cellular oxidant detoxification, aging and regulation of insulin secretion. Genes with decreased mRNA and protein abundances are enriched for: mito.respiratory complex I, mitochondrial ETC (mitochondrial number and function has shown to decline with age) this suggests that in the old brain samples there is reduced mitochondrial function. Genes with increased expression and decreased protein abundances in old brain samples are enriched for electron transport chain, mitochondrion organization and response to hormone. Genes with decreased expression and increased protein abundances mainly correspond to spliceosome complex and splicing.

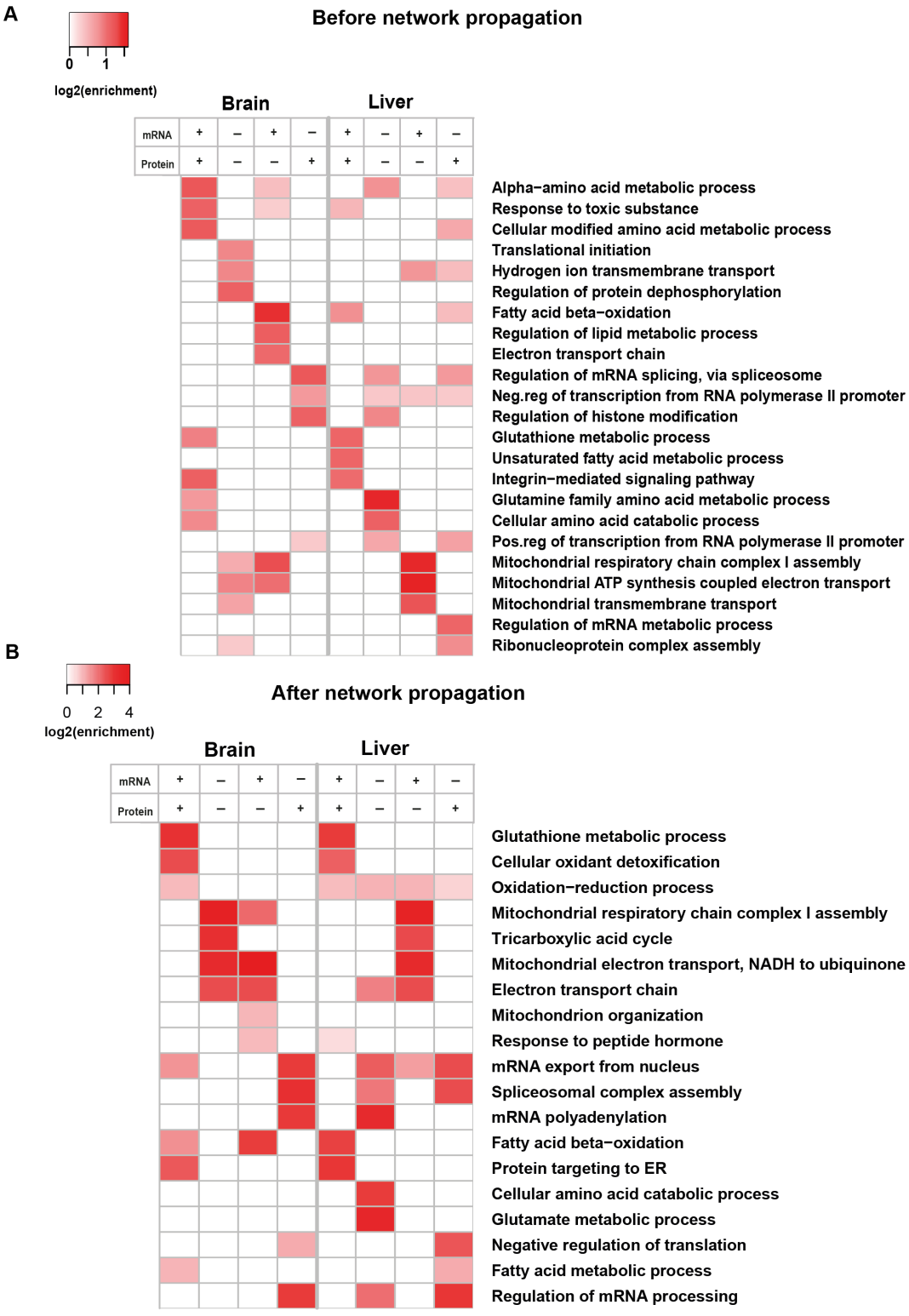

Figure 8: Gene Ontology enrichments from the concordant and discordant genes.

For propagating the protein lfc, the lfc from fractions were weighted and

Figure 8: A) Based on genes actual mRNA and protein log fold changes they were categorized into concordant and discordant groups. GO enrichment analysis were carried on each quadrant of genes and the top 3 GO terms based on their lowest p-value from elim fisher test on the biological processes category are represented. B) After network propagation with HD and RWR, the genes were classified similarly but this time with their propagated scores. GO enrichment of the different classes of genes and top 3 terms are represented

averaged. Thus we wanted to check whether the splicing complex proteins were increased in all the sub-cellular fractions as this would have different implications ex:if cytosol fraction has increased splicing proteins then it would imply enhanced production of splicing proteins, if nucleus has increased protein abundance it would mean more stable assembly splicing complex and efficient splicing. The log2 fold changes of the spliceosome complex genes from RNA-seq and proteomics on sub-cellular fractions indicate that the splicing proteins abundances are indeed increased in the nuclear fraction of the old brain samples (Fig.7).

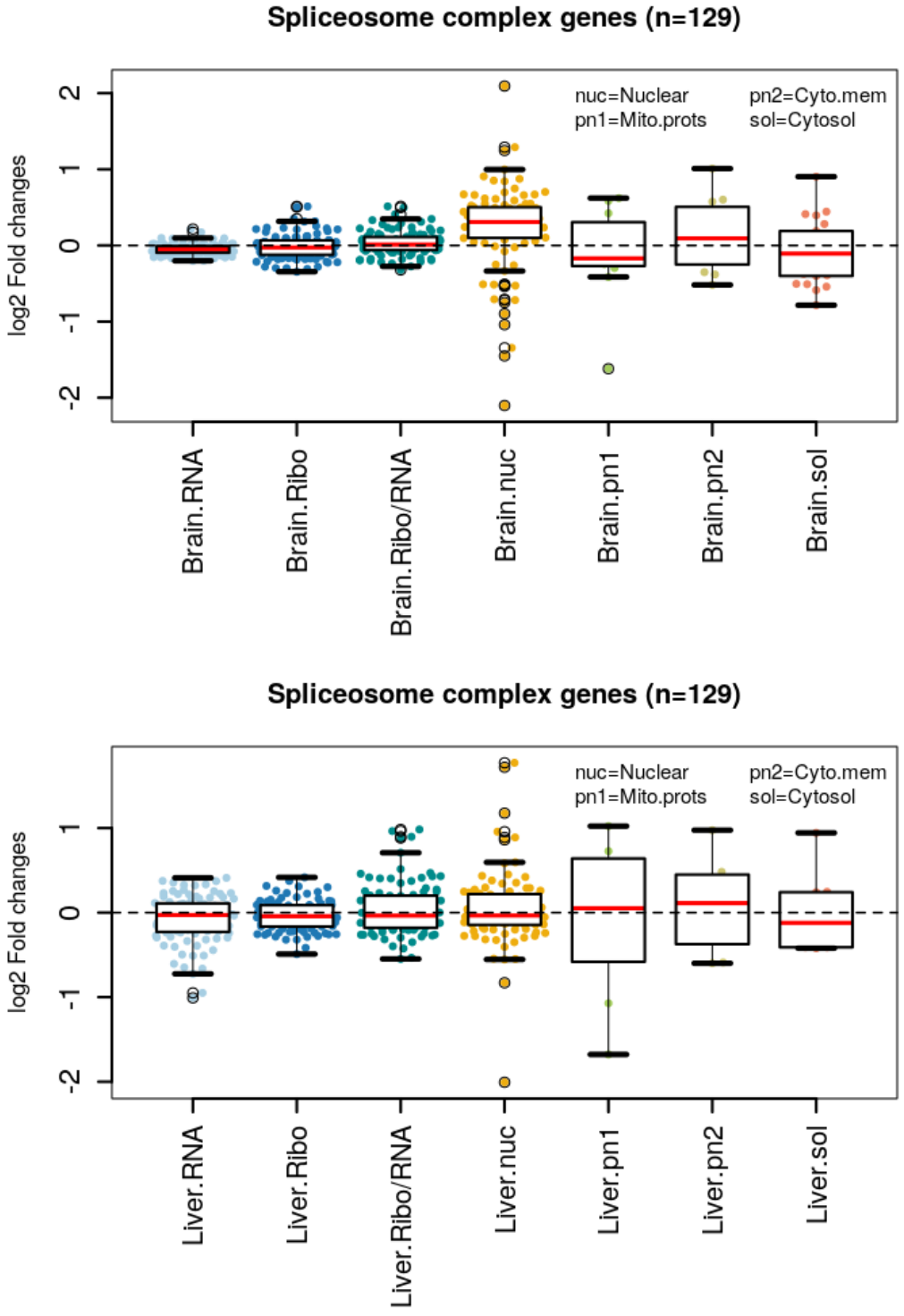

Figure 9: Splicing complex genes had increased in their protein abundances from nuclear fractions of old brain samples. A) Splicing complex genes mRNA expression and protein changes in old vs young tissues are represented.

2.5. Discussion

Our studies on detailed exploration of the network propagation algorithms, Random Walk with restart (RWR) and Heat Diffusion (HD) has provided the following insights. Firstly we have tested different graph normalization techniques for network propgation and from this analysis we identified that degree normalized adjacency representation of the network does not induce any topology bias on the propagated scores. With RWR, use of the normalized Laplacian network and for HD employing Laplacian network leads to accumulation of higher scores on the hub nodes just by the topology of the network though those nodes were not high scored initially. Gene function prediction following the principle of guilt by association was shown to be having similar performance after pruning the edges based on the node degrees.[30] This is in line with the idea that functionally relevant information are localized on a small fraction of the interactions of the network.

Secondly we have assessed the impact of the spreading parameters with different approaches, using transcriptome replicates vairance as a criterion and using both mRNA and protein measurements consistency as yet another measure. From the replicates analysis we observed that the different datasets require different amounts of spreading additionally for brain datasets there was not any further improvements upon network propagation. Previously RWR was used in genome wide association studies for identifying candidate genes of crohn's disease.[78] In this study the impact of the parameter α has been assessed using the internal consistency of gene ranks by employing 0.2,0.4,0.6,0.8,0.9 values and determined that 0.85 was optimal for their datasets. To our knowledge this is the first study to assess the impact of spreading parameters α and 't' in RWR and HD systematically with two different approaches.

Using Gene ontology enrichment analysis of the high scored nodes with use of optimal spreading parameters revealed enrichment of tissue function specific GO terms as well as known aging related genes. Further examination of the concordant and discordant genes before and after network propagation revealed that mitochondrial function related terms were reduced in their protein abundances for both the tissues. Mitochondrial functional decline is one of the well known hallmarks of aging,[56].[90] Furthermore we also observe that spliceosome complex proteins abundances were increased specifically in the brain nuclear fractions, this could possible due to enhanced stability of the spliceosome proteins in the nucleus to reduce the splicing errors with age or

possibly a result of efficient relocalization of splicing complex subunits from cytosol to nucleus. Experiments are needed to validate the possible reasons for spliceosome complex increased abundances in brain samples. Interestingly overexpression of one of the splicing subunit in lab model organisms has shown to be associated with lifespan extension.[36]

2.6. Author contribution and acknowledgements

Andreas Beyer and I conceived the idea. Andreas Beyer, Ronja johnen and I designed the analysis. Ronja johnen and I performed the analysis. Andreas Beyer and I wrote the related manuscript.

3. Systems insights from *C.elegans* longevity interventions

3.1. Introduction

Seminal works in laboratory model organisms has identified evolutionarily conserved signaling pathways in regulation of longevity.[42] These conserved pathways include reduced insulin/IGF-1 like signaling,[41] dietary restriction,[50] signals from the germline,[38],[5],[52] and the hypoxia signaling. These lifespan extensions are mediated through single gene mutations,[27] importantly these longevity traits are regulated through specific transcription factors. In C.elegans weaker mutations in daf-2 gene, encoding a hormone receptor orthologous to insulin or IGF-1 receptor has shown to double the lifespan by activating the transcription factors daf-16 and skn-1,[68],[4145] . The eat-2 mutants, genetic mimetic of dietary restriction, whose lifespan extension is dependent on activity of transcription factors nhr-62, pha-4 and skn-1,[35],[73].[9] Ablation of germline precursors of C.elegans results in lifespan extension, glp-1 mutants longevity is mediated by nuclear hormone receptors and transcription factors such as daf-12, daf-16, HSF-1, nhr-80, pha-4 and nhr-49,[38],[31].[80] Deletion of vhl-1, a cullin E3 ubiquitin ligase leads to lifespan extension and hif-1 transcription factor is epistatic to vhl-1.[60]

Transcriptome profiling of C.elegans long lived mutants have identified the genes and the molecular mechanisms associated to lifespan extension,[33],[68],[85].[35] The gene expression studies of longevity mutants with the epistatic conditions has revealed molecular signatures such as stress resistance, antimicrobial response and reduced energy metabolism that ameliorate ageing. Although these independent studies have pinpointed cellular functions that are specifically altered in long lived mutants, the common molecular responses across these longevity pathways still remains elusive. Network based approaches are one of the efficient methods for such integrated analysis.

In this study, with the utilization of a robust systems biology approach, network propagation coupled with experimental studies we have uncovered common and specific downstream processes that are causally associated with lifespan extension of *C.elegans*. The transcriptome of long lived mutants, eat-2, daf-2, glp-1 and vhl-1 with their epistatic double mutants were characterized using RNA- sequencing. The genes that were commonly differentially regulated in all the four longevity mutants were revealed through overlap analysis. Fur-

thermore network propagation of the transcriptional changes has pinpointed the sub-networks comprising nine functional modules that are altered and ameliorate aging in *C.elegans*.

3.1.1. Study aims

There are various molecular pathways that are implicated to extend life span across different species. The molecular responses upon modulation of these pathways in isolation have been studied in detail. However the common molecular responses across these longevity pathways remains elusive. Therefore in this study we wanted to understand the convergent molecular mechanisms that are co-occurring on modulating these lifespan extending pathways in *C.elegans*. Thus We performed RNA-sequencing on different lifespan extending mutant worm populations that covers the four signaling pathways such as insulin signaling, dietary restriction, germline ablation and hypoxia signaling.

3.2. Datasets

To study the insulin signalling mediated lifespan extension we have daf-2 (e1370: nucleotide substitution mutant for insulin signalling receptor), compared to the wild type (N2) and daf-2 (e1370);daf-16 (mgDf50), longevity phenotype of daf-2 mutants are dependent on daf-16 transcription factor. Therefore we have the double knockout mutants as a second control and comparison of the daf-2 transcriptome with these strains allow us to pinpoint the lifespan specific molecular changes.

To identify the molecular changes of dietary restriction mediated lifespan extension, eat-2 (ad465: mutant with reduced pharyngeal pumping mimicking dietary restriction), a genetic mimetic of DR is compared to wild type (N2) and eat-2 (ad465);nhr-62 (tm1818) as longevity phenotype is partially dependent on nhr-62 transcription factor.

Similarly for germline ablation signalling, glp-1 (e2141:mutant that have impaired germ cell proliferation) is compared to wild type (N2) and glp-1 (e2141); daf-16 (mu86), longevity phenotype of glp-1 is mediated through daf-16.

For hypoxia signalling, vhl-1 mutant worms (ok161: mutation in subunit of E3-ubiquitin ligase complex) that extend lifespan is compared to wild type and vhl-1 (ok161); hif-1 (ia4) as vhl-1 mediated longevity is dependent on the hypoxia response transcription factor hif-1.

3.2.1. STRING functional interaction network

The functional interaction network for *Caenorhabditis elegans* was retrieved from STRING database version 10.5.[91] To limit the analysis only to the high confidence interactions the generic network was filtered using combined scores threshold $\geqslant 700$. This step of filtering was essential to avoid the flow of signal to the nodes connected through spurious edges. The filtered network had 10,735 nodes with 4,05,668 edges and this network was employed for diffusing the transcriptome signals from each longevity mutant (negative log transformed p-values).

3.3. Methods

3.3.1. RNA-seq read mapping and differential expression analysis

The read qualities were assessed using the FastQC. For the libraries that had adaptor sequences contamination and poor quality reads, raw sequence reads were trimmed using the cutadapt (V1.12).[58] The trimmed reads were mapped to the *C.elegans* reference genome assembly from Ensembl WBcel235. For aligning the reads to the reference genome Tophat2 (V2.1.1)[44] was used. Subsequently the aligned reads were counted over the protein coding genes (# total protein coding genes) using bedtools multicov (V2.26.0).[79] Differentially expressed genes (DEGs) were determined using DESeq2[55] using Wald test with pairwise contrasts of longevity mutant vs N2 strains and longevity mutant vs double knockout strains. p-values were adjusted for multiple hypothesis testing using BH method. For each mutant, adjusted P-values of genes were -log transformed and used for network propagation.

3.3.2. Network propagation of differential expression scores with Random Walk with Restart

Network propagation was done using the filtered *C.elegans* STRING functional interaction network. First, the network was converted to an adjacency matrix and normalized with Laplacian transformation (using the graph.laplacian function in the igraph R package). The p-values of genes that resulted from differential expression analysis were negative log transformed and subsequently superimposed on the corresponding nodes in network separately for every longevity mutant. After mapping, the transformed p-values on the individual nodes were diffused to their adjacent nodes using the spreading coefficient of

0.6 (corresponds to the percentage of sharing to neighbours). This spreading changed the scores of all genes in the network, and it was iteratively repeated until gene scores do not change anymore. The convergence of scores was estimated by calculating the maximum absolute column sums between the current propagated scores matrix (F^t) and previous iteration propagated scores matrix (F^{t-1}) using the norm function in R and when norm of the matrices falls below 1×10^{-6} the scores are considered for subsequent exploration,[37].[104]

$$F^t = \alpha W F^{t-1} + (1 - \alpha)Y$$

Where F^t is the matrix of network propagated scores at 't' iteration, column corresponds to the conditions considered for the study and rows of the matrix corresponds to network genes. W is an asymmetric square matrix representing the laplacian transformed network. α is the spreading coefficient corresponds to the fraction of the node scores to be shared with adjacent nodes in network. We noticed that the network topology (i.e. the network structure) induces a bias: certain regions of the network tend to accumulate high scores during the propagation even if there is no actual signal. In order to correct for this bias, we computed a node-specific topology bias in the following way: we started with equal scores on each node (e.g. each node has a score of 1). Then, we perform the propagation as described above. The resulting scores reflect the bias and are subtracted from the node scores after propagating the actual nodes scores. Subsequently, the corrected propagated scores were clustered based on their Euclidean distances (using the hclust function in R) allowing for the identification of clusters of genes that has accumulated higher signals through propagation consistently across lifespan extending mutants.

3.3.3. Gene ontology term enrichment

topGO was employed for Gene Ontology enrichment analysis and *C.elegans* genes and protein annotations were retrieved from gene ontology consortium (www.geneontology.org). To identify enriched GO terms, the one-sided elim Fisher procedure was used ($\alpha \leqslant 0.05$).[1] The enrichment score of a GO term is defined as log2 (#Detected significant genes/ #Expected significant genes). From the clustering of corrected propagated scores of genes, the individual cluster genes were tested for GO term enrichment against the network genes as reference background. Furthermore to functionally characterize the consistently high score accumulating cluster genes were pooled and tested for gene enrich-

ment again against the network genes. The dotplots shows the most specific significantly enriched categories with a minimum of five associated genes.

3.3.4. Selection of candidate genes for lifespan screening

For the experimental validation of the network propagation highlighted modules, we have selected module members based on their propagated scores and log2 fold changes. For each module the constituent genes median log2 fold changes were calculated across the seven comparisons (every pro-longevity mutant compared to wild type and double mutants, except glp-1 which is only compared to glp-1;daf-16). From each of the longevity module, the genes that had the negative median log2 fold changes were filtered. As these genes were down-regulated in atleast 50% of the analysed comparisons with the logic that knocking down these genes in the wt strains will allow us to score their functional relevance with lifespan extension.

3.4. Results

3.4.1. Transcriptional changes in the longevity mutants of *C.elegans*

To identify the longevity genes mutation specific expression profiles of *C.elegans* we characterised the transcriptomes of wild type and longevity strains at their young adult stage. For the detailed characterization of the insulin signaling mediated changes in gene expression daf-2 (lowered Insulin signaling pathway) mutants were studied. Similarly other pro-longevity mutants, eat-2 (genetic mimetic of Dietary restriction), glp-1 (germline-less mutant) and vhl-1 (hypoxia mutant) were characterized using the RNA-Sequencing. First we compared the transcriptional profiles of long lived strains with wild type N2 strains and double mutants (daf-2;daf-16 for daf-2 mutants, eat-2;nhr-62 for eat-2 mutants, vhl-1;hif-1 for vhl-1 mutants).

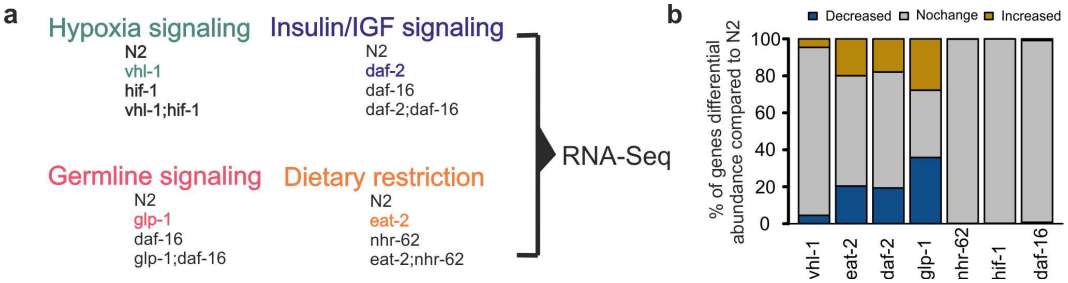

Figure 10: Transcriptional response to longevity interventions. (a) Schematic overview of the expreimental design (b) Percentage of genes that are differentially regulated in lifespan extending mutants with respect to N2 strains of *C.elegans*.

3.4.2. Consistently regulated genes in longevity mutants

To identify the genes that are affected commonly in the pro-longevity mutants, the overlap of top 1,000 most strongly affected genes were examined. From this overlap analysis, we observed only six were commonly affected by the lifespan extending interventions (Figure 2a). These six genes include fib-1, is a rRNA methyltransferase involved in nucleolar size control, ruvb-2 has DNA helicase activity, glh-1 is a RNA helicase and involved in P-granule maintenance, retr.1 is a retrotransposon like protein 1, Y66D12A.9 is an human ortholog of PSME (proteasome activator complex), and E04D5.1 is an ortholog of eukaryotic translation initiation factor 2A. All these 6 common genes were consistently down-regulated in all the longevity mutants. In order to determine the cellular functions which were consistently affected, we analyzed genes that were differentially regulated in at least three out of the four mutants. Functional enrichment analysis of these genes identified defense response, RNA processing, translation, processes linked to reproduction among others (Figure 2b). This functional enrichment analysis suggested that common molecular sub-networks could be affected in the lifespan extending mutants even though the specific genes responding to the knockouts are not identical. Further to systematically explore the notion of consistently regulated functions we utilized network propagation. For propagating the differential regulation scores, we have used a functional gene interaction network from STRING database, that links genes with common molecular functions. Thus, network propagation will lead to increased gene scores in sub-networks that are strongly affected by the mutation. We performed network propagation independently for each of the four mutants

and selected again the top 1,000 genes, but this time with the largest network propagation scores and not those with the most significant individual changes. Using this approach we obtained 16 genes that were affected commonly in all four mutants (Figure 2c). Once again we performed functional enrichment analysis on gene sets affected by at least three mutations, which identified twice as many significant GO terms as before (16 versus 8). The increase in number of significant GO terms is expected, because we used a functional interaction network for propagating the scores. Some functions were commonly identified with both approaches, such as defense response, posttranscriptional regulation, and developmental processes related to reproduction. Importantly, the network approach lead to the identification of several other cellular processes that are known to be important for lifespan, such as proteolysis and amino acid metabolism. Our results indicate that network propagation identifies molecular networks that are consistently altered in longevity mutants.

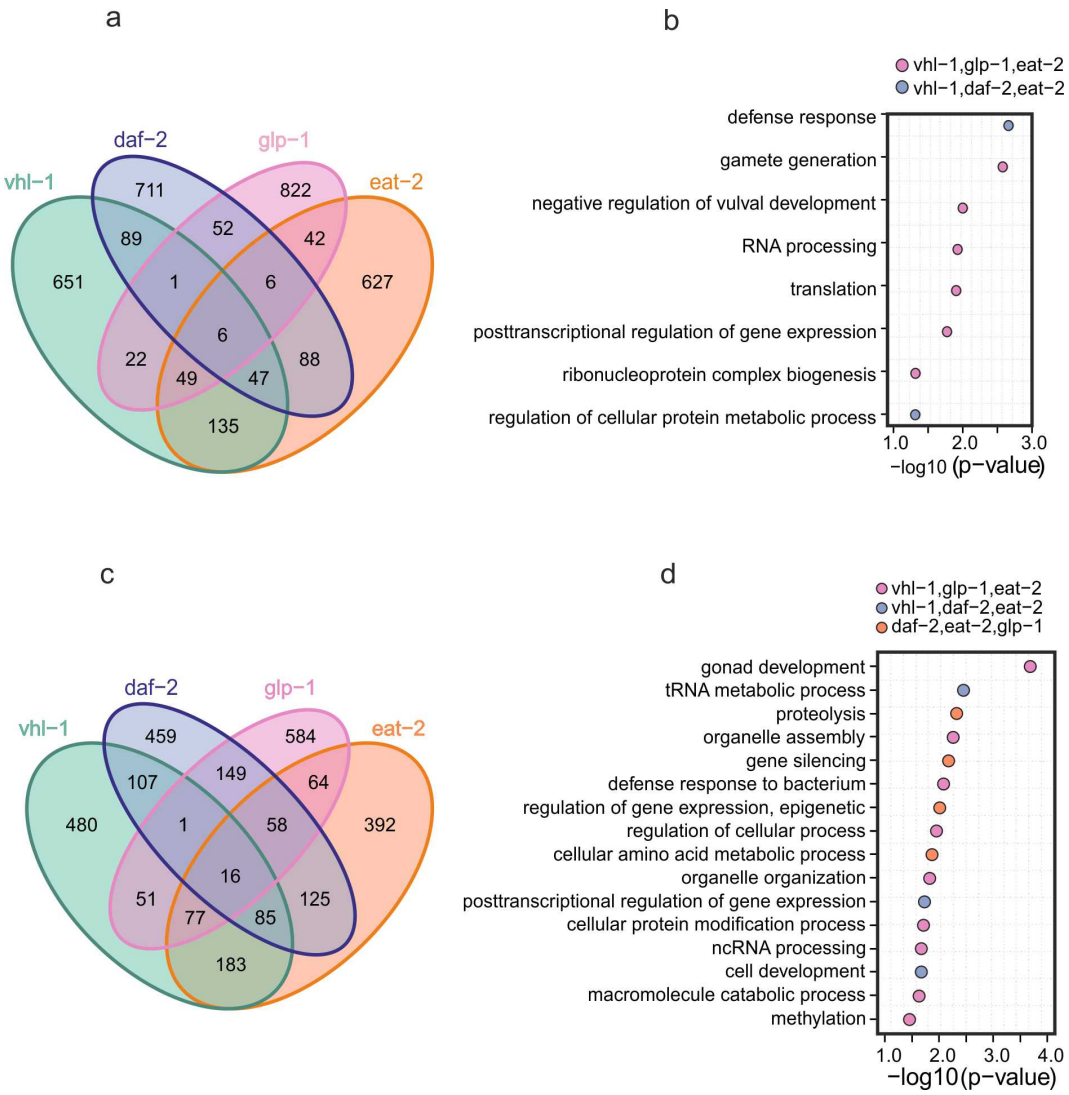

Figure 11: Comparison of conventional approach to network propagation approach. Venn diagram depicts the overlap of genes that are differentially regulated in different lifespan extending conditions. (a) Number of uniquely and commonly differentially expressed genes from top1000 significant genes for every longevity mutants transcriptomes compared to wild type: vhl-1, daf-2, glp-1 and eat-2 versus N2 using Deseq2. (b) GO enrichment results from differentially regulated genes in at-least three conditions using topGO. The elim algorithm was used and the GO terms were filtered to have minimum of five genes. (c) Common and unique genes from top 1000 genes ranked based on network propagated scores. With the conventional differential expression analysis using RNA-seq data revealed very few genes that are commonly regulated in different longevity mutants. However, using network propagation approach much more genes that are functionally related are found to be commonly regulated in longevity mutants. (d) GO enrichment results from common top 1000 highly ranked genes from at least three mutants after network propagation.

3.4.3. Consistently altered functional sub-network of *C.elegans* longevity

Next we sought to identify the subnetworks that are altered in longevity mutants, the genes were clustered based on their network propagated scores. This resulted in five gene clusters, among them three clusters were consistently high-scoring in at least three longevity mutants (Figure 3). We hypothesized that these three clusters contain genes that are particularly relevant for lifespan regulation. In total 2,782 genes were contained in these three clusters. In order to obtain a smaller network containing genes most relevant for lifespan regulation we further filtered these 2,782 genes. First, we performed GO enrichment analysis, which identified nine cellular functions that were already known to be important for lifespan: reproduction, nucleosome assembly and disassembly, determination of adult lifespan, collagens, protein metabolism, oxidation-reduction, transcription, rRNA and tRNA processing and stress resistance among others. We have extracted the genes annotated in these Gene Ontology Terms and constructed the longevity sub-network with 784 genes.

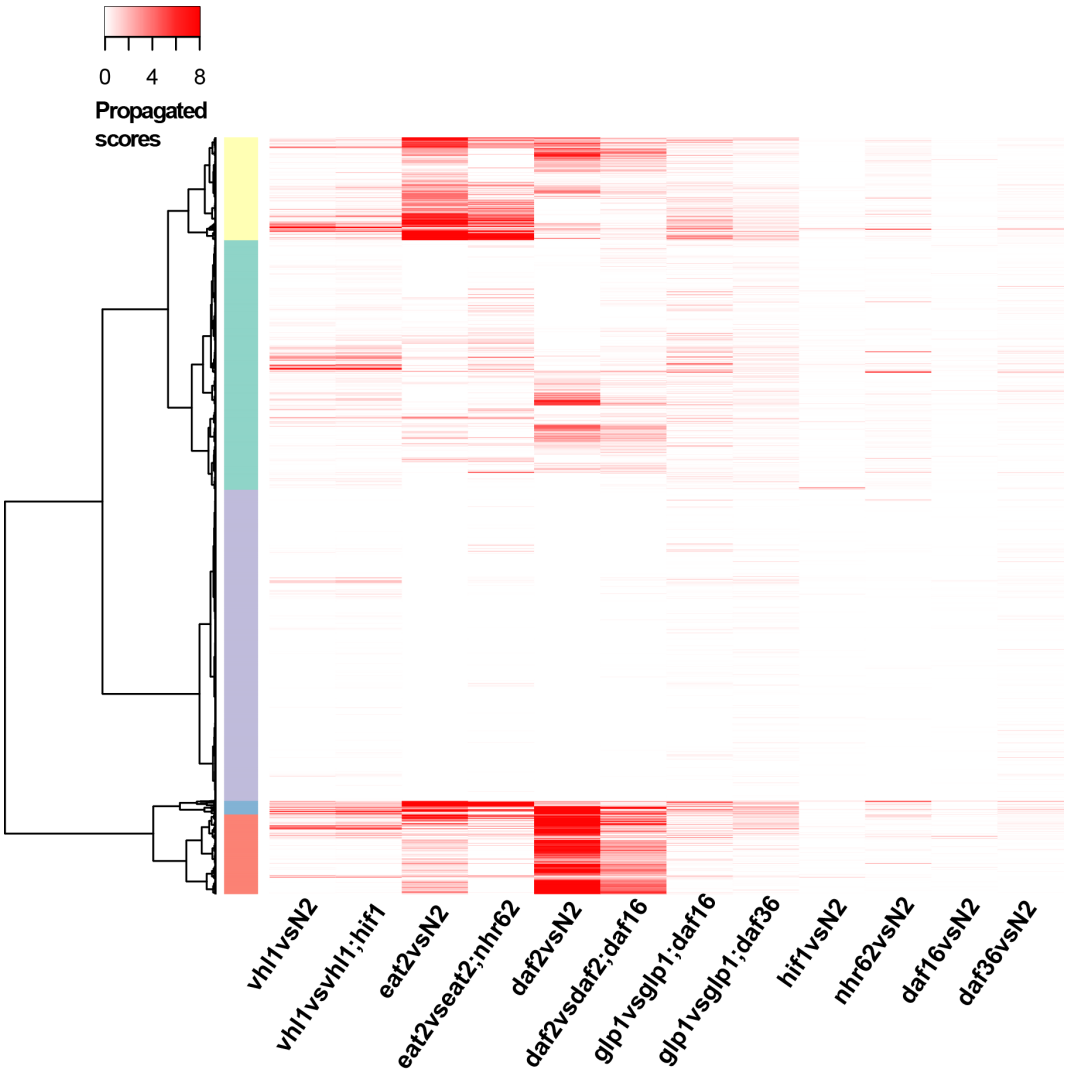

Figure 12: Heatmap of network propagated scores. Network propagated scores for each condition after topology bias correction. The clusters 1, 4 & 5 are commonly enriched in lifespan extending conditions from both comparisons (Wt and double KO).

This sub-network was further filtered for genes with at least three neighbours to extract highly connected modules, resulting in 410 nodes in the filtered sub-network. Moreover for visual simplicity, the nodes that had more than three edges were shown. The node colours represent their functions, the node size corresponds to the no.of nodes that are merged together by their gene name ex. col-1,2,10 were merged to one single node for viewing. All the subsequent analyses are restricted to these 410 genes with constituting the different functional modules (Figure 4).

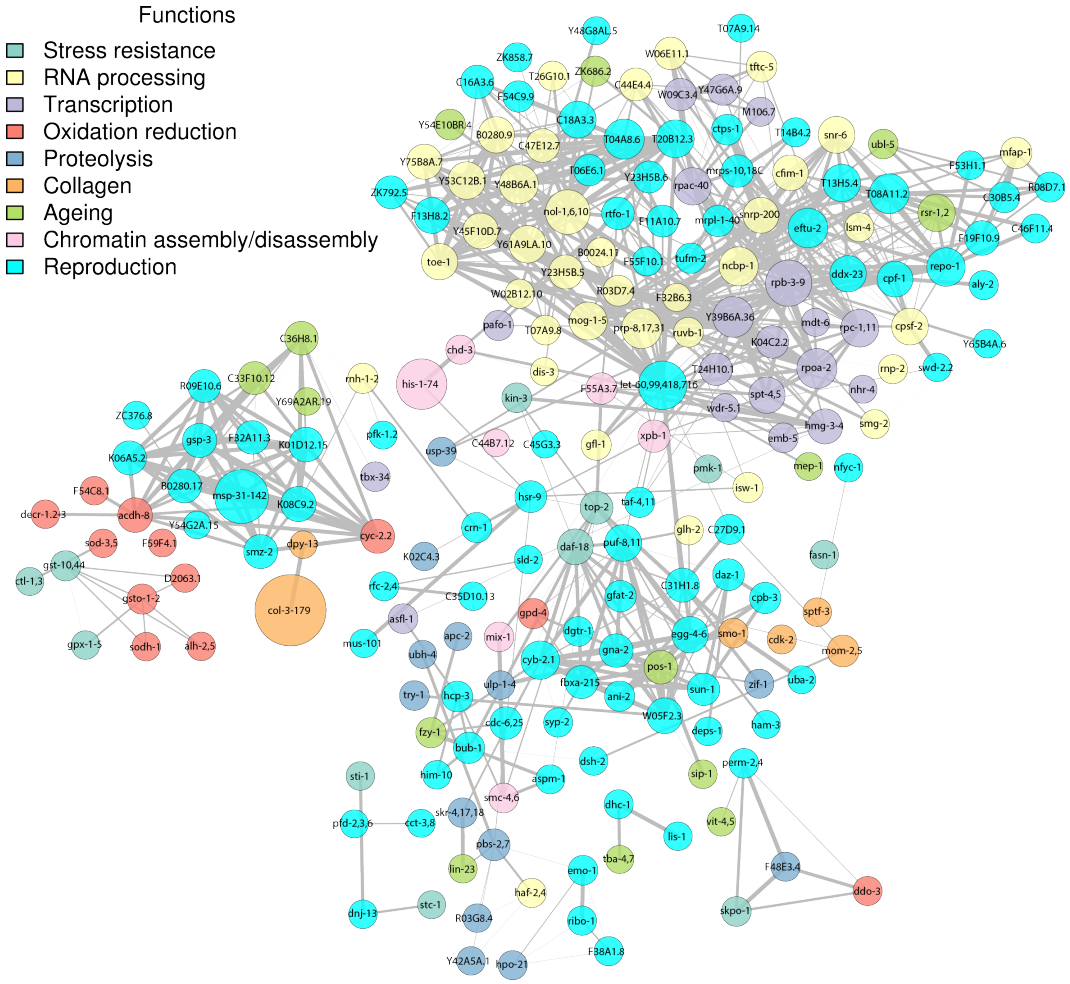

Figure 13: Longevity sub-network and the functional modules revealed through network propagation. Sub-network of genes that are consistently high scored in longevity mutants. The node colors correspond to the primary function of the gene, the nodes betweenness were log transformed and represented as the node sizes. The width of edges correspond to their confidence of interaction from the STRING database.

To gain further insights from the identified longevity sub-network, modules (densely interconnected communities) were determined using the Girvan-Newman algorithm. This algorithm finds communities by iteratively removing the edges with high betweenness thus eventually resulting in 8 individual network components with more than 15 genes in each community. Based on the previous studies on these mutants we expected that these modules are important for lifespan, but their activity could differ across the mutants. In order to explore this hypothesis we used the average expression of genes in these modules as a proxy for module activity (Fig5). Some of the modules showed

highly consistent behaviour across the lifespan extending mutants. For example, the modules 'transcription', 'RNA processing', and 'proteolysis' were down-regulated in all mutants with extended lifespan, while they were up-regulated in mutants with reduced lifespan. However, other network modules showed more condition-specific regulation. For example, 'Nucleosome assembly' was specifically down-regulated in the eat-2 mutant, but not in the other lifespan extending mutants.

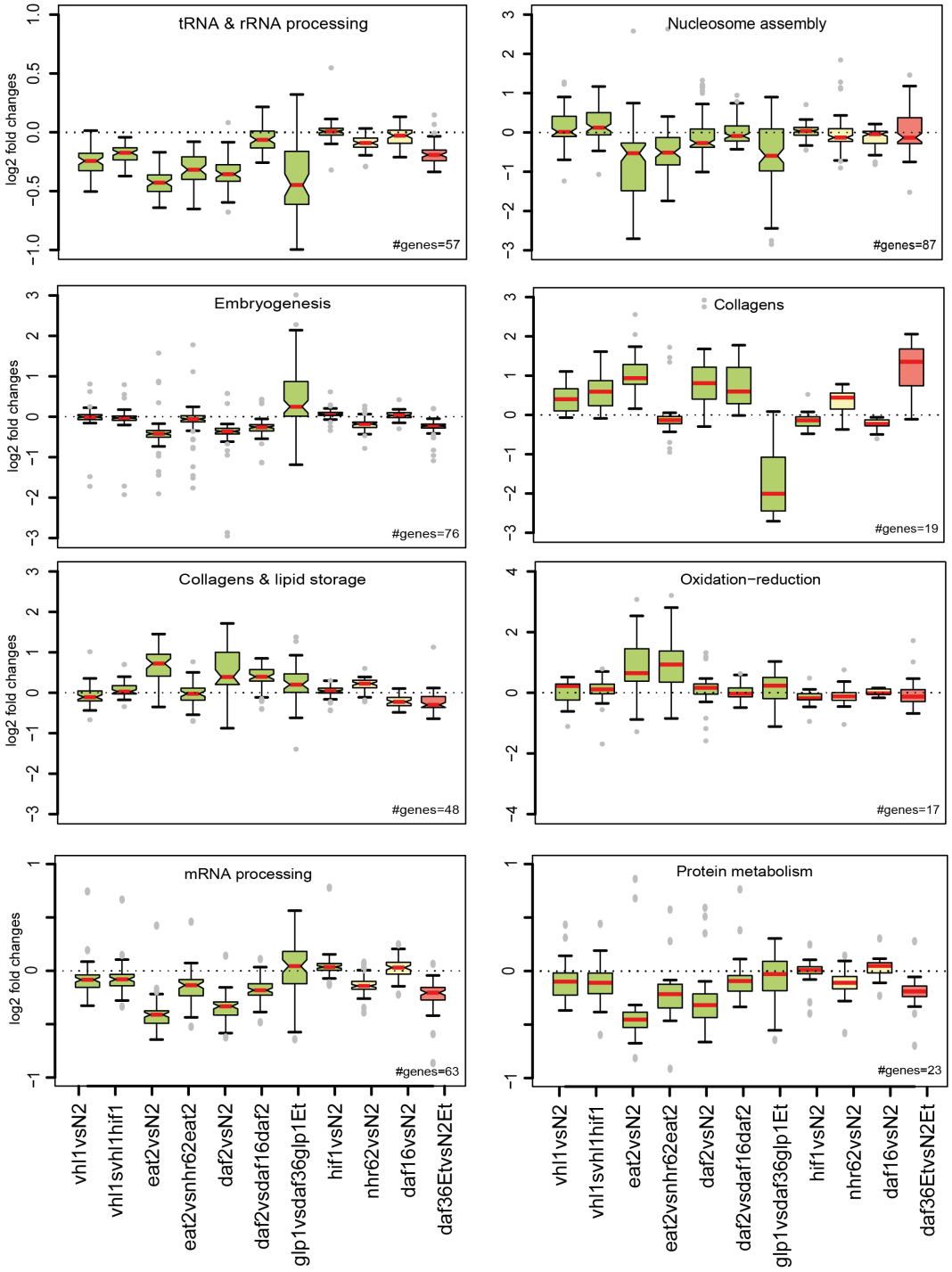

Figure 14: Functional modules from the network propagation identified subnetwork. The interconnected modules in subnetwork were found by Girvan-Newman algorithm based on the edge betweenness measures. Activity of the network modules across the longevity mutants.

3.4.4. Selection of module members for experimental validation

All of the above analyses were based on correlating gene expression of network modules with lifespan phenotypes. This does not establish a causal relationship between the activation (or inactivation) of individual modules and lifespan. Thus, we selected module members whose repression with RNAi would likely extend lifespan. Specifically, we selected 156 genes from the network that had negative median expression fold changes across the four lifespan extending conditions compared to wild type and double knockouts (Ref. Table of selected genes). Further, we added the 6 and 16 genes identified by overlapping the four longevity conditions before and after network propagation (Figure 2). In total we have selected 172 genes as 6 genes that were common from both Venn analysis, were differentially regulated and also high scored from smoothing. We have employed RNAi to understand the functional relevance of these modules in longevity. We observed an interesting overlap of the selected genes with the GenAge ageing related genes.[92] Network propagated scores guided the selection of candidate genes for longevity.

3.4.5. Causal association of network propagation identified genes in longevity of C.elegans

To score the functional relevance of selected candidate genes in lifespan extension, the survival ratios of the RNAi strains were analysed. The survival ratios were calculated by counting the number of worms survived at day 25 in RNAi treated compared to mock treated worms. The ratios above 1 imply that RNAi treatment has a positive effect on the survival of worms. The maximum lifespan extension was observed for the positive genes (Fig. 6a). To our surprise we found two positive genes that were longer lived than daf-2 mutants (rpb-5 and glh-1). rbp-5 is a shared component of RNA polymerases I, II and III and it has been reported to have a role in transcriptional activation of yeast.[64] glh-1 is a RNA helicase and is consistently down-regulated in all longevity mutants subsequently identified through direct intersection of regulated genes. To further understand the modules involvement in lifespan regulation the modules percentage of survival responses were calculated. For each module the fraction of lifespan associated genes were counted among the selected candidate genes (number of genes with survival ratios above 1/ number of genes selected from a module * 100). From this analysis we observed that among the selected network modules, collagen modules had the maximum positive survival effect

followed by oxidation-reduction module (Figure 6).

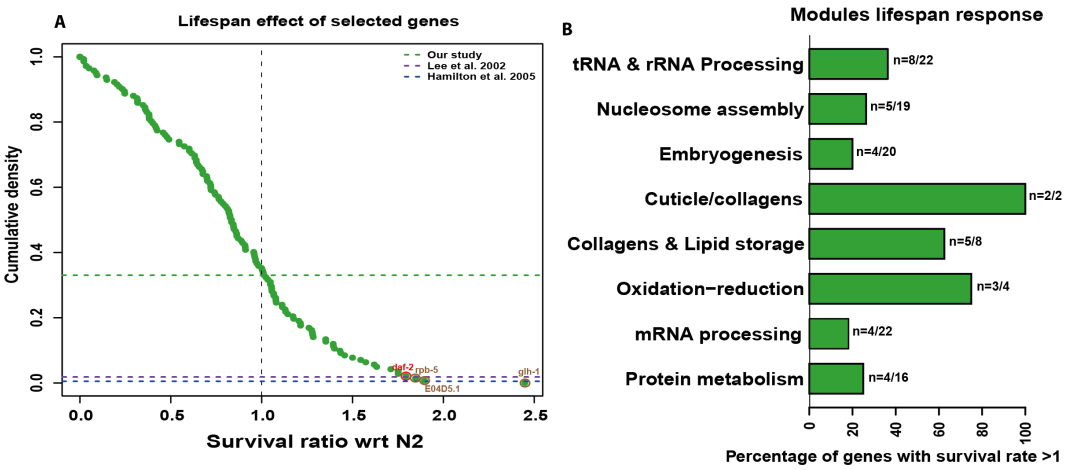

Figure 15: Survival response of selected candidate genes through network propagation. (a) Cumulative density of survival ratios of the candidate genes. (b) Individual network modules survival responses in percentages, n corresponds to no of genes with improved survival/ no of genes selected from this module.

Table 2: Comparison of lifespan screening studies of *C.elegans*

Lifespan study by	No.of RNAi clones screened	No.of strains reproducibly lifespan extending	% of lifespan extending strains from screened strains
Hansen et al. 2005	70% of *C.elegans* ORFs	94 candidates of which only 29 could be confirmed	
Lee et al. 2002	2,663 genes	52 genes	1.8%
Hamilton et al. 2005	16,475 genes	89 genes	0.54%
Network propagation guided study	147 genes	27 genes (If mean lifespan response is considered-49 genes)	18.4%,(If mean lifespan response is considered-33%)

3.4.6. Dichotomous expression behaviour of collagen genes in c.elegans longevity mutants

In *C.elegans*, collagens are a family of structural proteins that constitute the extracellular matrix. The collagen gene family is quite large, comprising of

more than 150 gene members. From the previous studies on daf-2 mutants it was observed that many collagen gene members are overexpressed in lifespan extending conditions.[23] Therefore we checked the expression changes of all the cuticle forming collagen genes in all the longevity mutants (Figure.7). In our analysis we also observed that most of the collagen genes were overexpressed across the lifespan extending mutants. Most interestingly, we also observed that few collagen gene members were consistently down-regulated in the lifespan extending mutants. Furthermore these collagens were high scored after network propagation implying that they themselves and their functional neighbors are also differentially regulated. Thus we got more curious to look in detail on these few collagen gene members such as col-19, col-20, col-92, col-93, col-119, col-178 and col-179. We also studied the role of these down-regulated collagens in lifespan extension. For that we have made RNAi knockdowns of each of these collagens in N2 strains and observed the number of worms that are alive at day 25. From this lifespan screening experiments we found that all the selected collagens were consistently having higher survival ratios compared to the untreated worms except col-19 and col-20. In order to avoid miscounting the offsprings these lifespan experiments were performed in the presence of a chemical FUDR which blocks the germline development. To make sure that the survival response of these collagens are not a side effect of the FUDR treatment we have also performed lifespan assays in the absence of FUDR and observed that the longevity effect on knocking these selected collagens are reproducible in the absence of FUDR treatments.

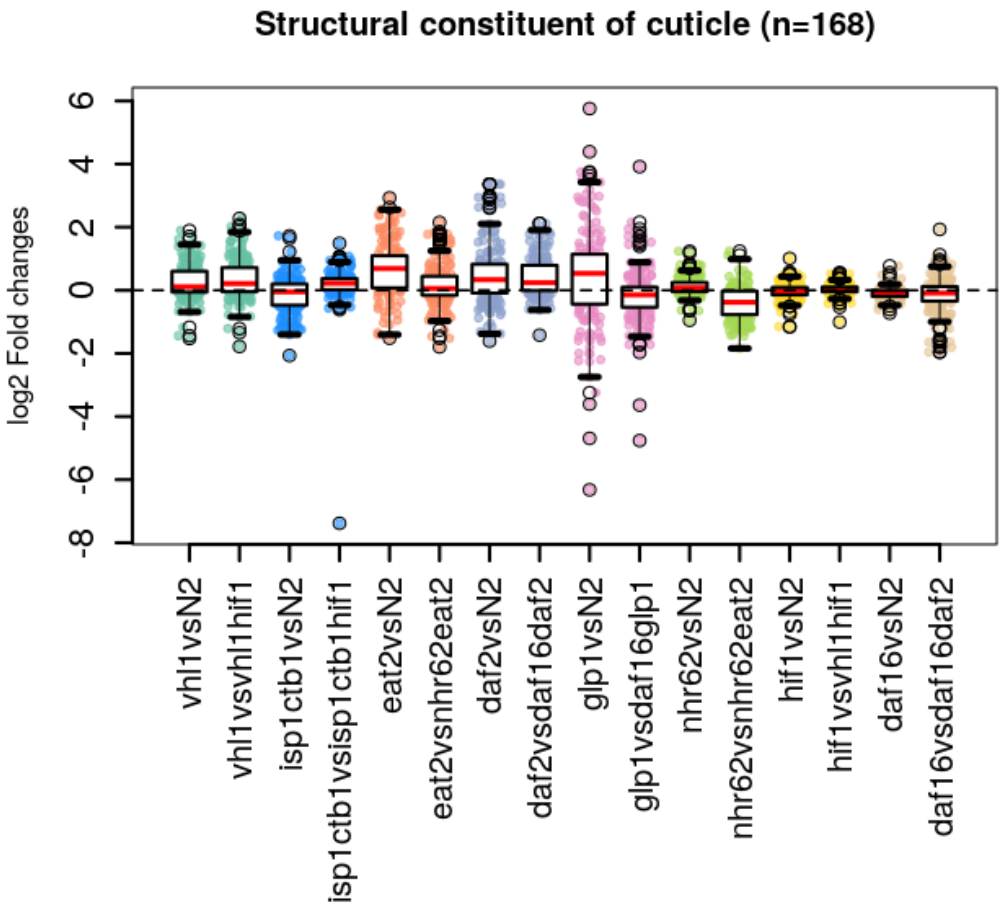

Figure 16: Expression changes of all the collagens that are constituent of cuticle in longevity mutants of *C.elegans*

Most interestingly, we also observed that few collagen gene members were consistenly down-regulated in the longevity mutants. Furthermore these collagens were high scored after network propagation implying that they themselves and their functional neighbors are also differentially regulated in the lifespan extending conditions.

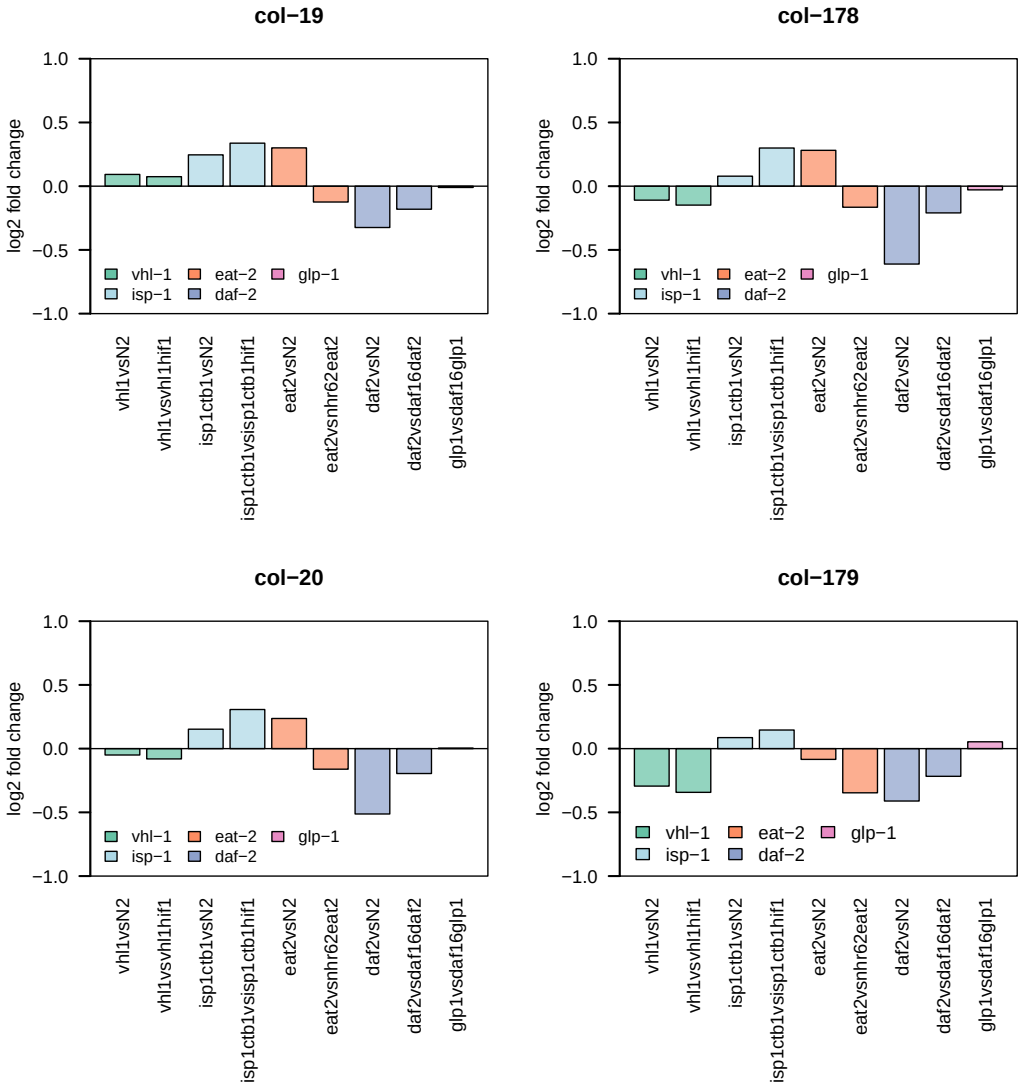

Figure 17: Collagen members that show divergent pattern of expression compared to other members of collagens. Furthermore these collagens inactivation is functionally linked to lifespan extension in *C.elegans*

Thus we got more curious to look in detail on these few collagen gene members such as col-19, col-20, col-92, col-93, col-119, col-178 and col-179. We also studied the role of these down-regulated collagens in lifespan extension. For that we have made RNAi knockdowns of each of these collagens in N2 strains and observed the number of worms that are alive at day25. From this lifespan screening experiments we found that all the selected collagens except of col-19 and col-20 were consistently having higher survival ratios compared to the untreated worms. In order to avoid miscounting the offsprings these lifespan

experiments were performed in the presence of a chemical FUDR which blocks the germline development. To make sure that the survival response of these collagens are not a side effect of the FUDR treatment we have also done lifespan experiments in the absence of FUDR and observed that the longevity effect on knocking these selected collagens are reproducible.

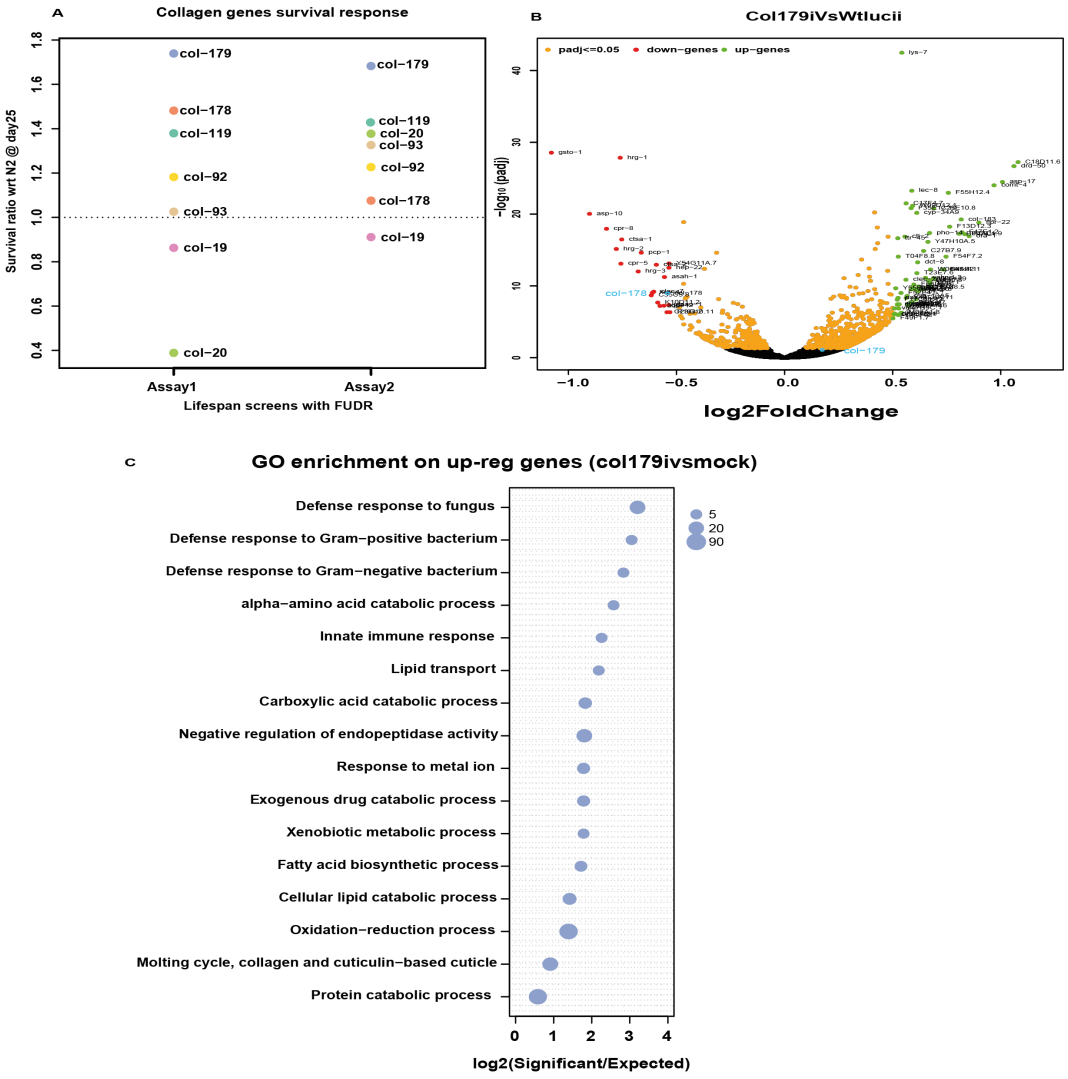

Figure 18: Collagens association with lifespan extension in *C.elegans*. a. Survival ratios of selected collagen gene members upon RNAi treatment in the N2 strains. b. Differentially regulated genes in col179 RNAi strains compared to untreated N2 strains of worms. c. Biological processes that are activated in col-179 RNAi treated worms.

3.4.7. Col-179 RNAi worms exhibits enhanced defense response

We observed that col-179 RNAi treated worms were reproducibly lifespan extending in both presence and in absence of FUDR. Thus we were curious to uncover the molecular changes that are associated with lifespan extension. Hence we performed RNA-sequencing of both col-179 RNAi treated worms and the untreated worms. To validate the knock-down of col-179 transcripts in RNAi treated worms we checked the expression changes of col-179 RNAi worms in comparison to N2 mock treated worms. We observed that col-179 transcripts were expressed at the same levels in RNAi treated as of the mock treated worms. This caught our attention to check for the sequence identity of col-179 with other collagen gene members. From the sequence analysis we found that col-178 and col-179 share high sequence identity, 93% of the col-179 nulceotide sequences are identical to col-178 sequence. Therefore it is likely that our RNAi treatment could be knocking down both col-179 & col-178 transcripts. On examining the expression changes of col-178 transcripts in RNAi treated worms with respect to mock treated worms we observed that only col-178 expression was knocked down with RNAi treatment but not both col-178 and col-179 (Figure 8). Further to gain more molecular insights on these stains we identified the differentially regulated genes that might causally be linked to lifespan extension. For this analysis, relative abundance of transcripts were estimated by comparing the transcriptome of RNAi treated worms with mock treated worms (Figure 9b).

Gene ontology (GO) enrichment was performed separately for the up-regulated and down-regulated genes in RNAi treated worms this allowed us to uncover the cellular functions that are affected and might be associated to lifespan extension. From this analysis we found that defense response and immune response genes were particularly over expressed in addition to other processes including lipid transport, protein and amino acid catabolic processes, carboxylic acid catabolism and oxidation-reduction process. Furthermore we also noticed that other collagen gene members were over-expressed. Upon GO enrichment analysis of the down-regulated genes we determined many biological processes that are strongly linked to reproduction such as embryonic pattern specification, oocyte maturation, single fertilization, hatching, hermaphrodite genitalia development, regulation of vulval development etc. From the previous studies on longevity interventions, lifespan extention through somatic maintenance are often are associated with reduced reproductive outputs.[74] Thus we speculate

that our RNAi worms could also have a delayed reproductive age. In addition to these cellular functions we also observed glycolysis related genes, protein N-linked glycosylation, proteasome dependent protein catabolism and other collagen gene members.

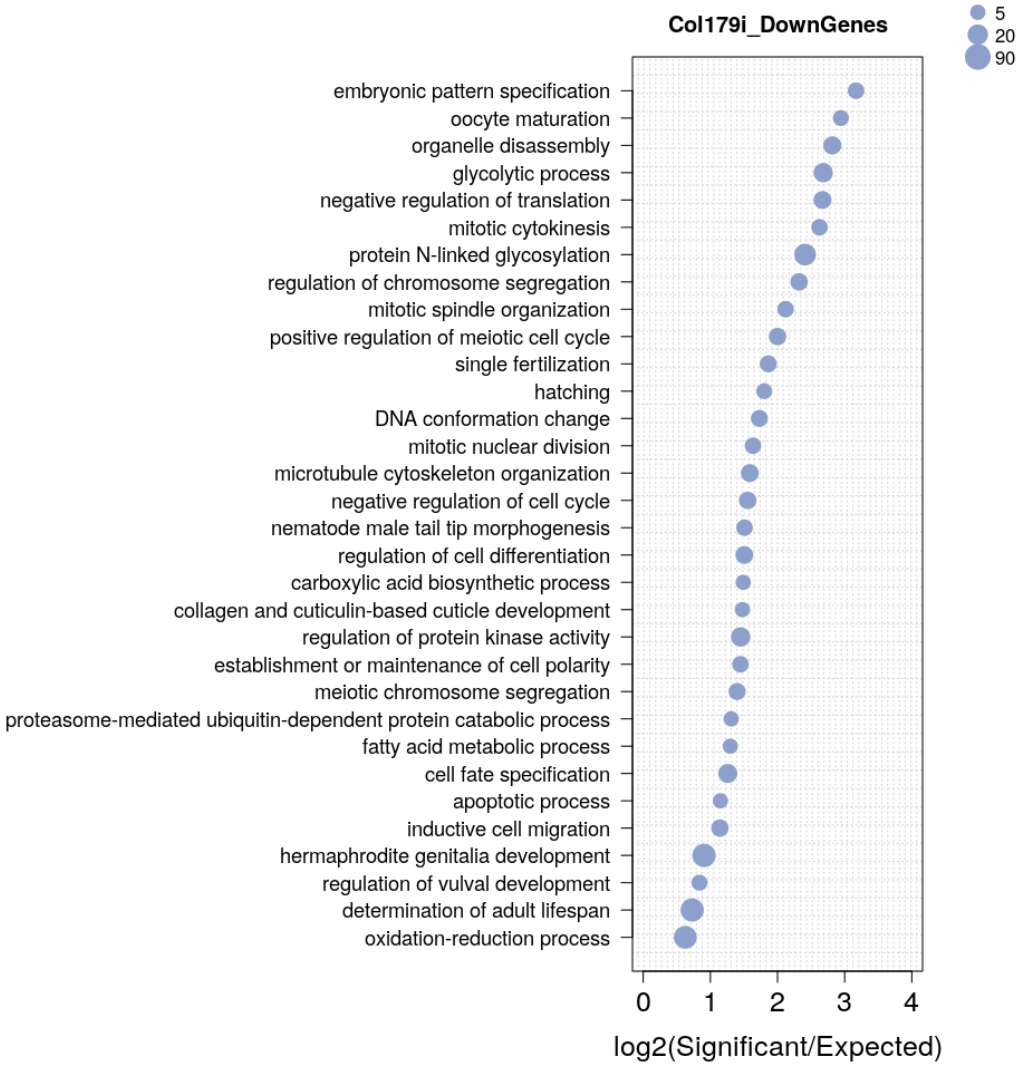

Figure 19: Biological processes that are down-regulated in col-179 RNAi treated worms

3.5. Discussion

Lowered insulin signaling, dietary restriction, germline ablation as well as hypoxia signaling pathways ameliorates aging effects and extends lifespan. These conserved signaling mediated changes in the transcriptome has been

studied in detail only in isolation. However the transcriptional changes across these pro-longevity pathways has not been systematically characterised. This is essential to pinpoint intervention specific and common alterations across these pro-longevity mutants as they would allow us to delineate the cellular functions that are associated with lifespan extension. Here we have identified the sub-networks that are consistently altered and validated their causal association with longevity.

3.5.1. Transcriptome profiling of longevity mutants of *C.elegans*

Transcriptomics studies on reduced insulin signaling in *C.elegans* has identified the molecular signatures such as stress resistance, antimicrobial response, lipid, protein and energy metabolism being associated with longevity,[94].[33] We also observed that these cellular functions are altered in the reduced insulin signaling mutants. The dietary restriction mutant, eat-2 has reduced pharyngeal pumping, further gene expression studies on DR has identified the processes associated to longevity: fatty acid localization, protein metabolism and nucleosome assembly and organization.[35] From our transcriptome characterization of DR we also observed pharyngeal development and pumping genes were regulated thus confirming DR phenotype, additionally nucleosome organization, protein metabolism and fatty acid biosynthesis were regulated in alignment with previous findings. Proteomic quantification studies on germline loss mutants, glp-1 revealed protein metabolism, immune responses being associated with longevity.[77] Our transcriptome studies of glp-1 compared to glp-1;daf-16 indicates immune responses, carbohydrate and protein metabolism being differentially regulated, the gene expression changes were consistent with the protein abundances of glp-1. Hypoxic signaling,[17],[6010] is mediated in both dependent and independent manner of HIF-1 transcription factor and also observed to induce changes in extracellular matrix genes. Our transcriptome analysis on vhl-1 indicates ER unfolded protein response, innate immune response were activated along with the extracellular matrix remodelling.

3.5.2. Identification of commonly regulated genes mediating lifespan extension

From the direct intersection of differentially regulated genes from all four pro-longevity mutants we identified that six genes were commonly regulated. Common genes include fib-1 a nucleolar protein, is involved in regulation of

nucleolar size and was observed that reduced nucleolar size is a hallmark of longevity.[97] glh-1 is germline specific ATP dependent RNA helicase and in our lifespan screening assays ie was observed to be consistently having much higher survival rates compared to mock treated. E04D5.1 is predicted to have trnaslation initiation activity and in our lifespan screening analysis was found to be 80% better surviving. Other genes ruvb-2, retr-1 and Y66D12A.9 were not screened for their survival response as these genes clones we were not able to obtain their RNAi clones.

3.5.3. Convergent modules of longevity identified through network propagation

Employing network propagation on the RNA-Seq datasets from the longevity mutants aids in discovering the lifespan associated subnetwork in *C.elegans*. The subnetwork had 410 genes after filtering for nodes with three or more edges. Upon application of edge-betweenness community finding approach we pinpointed 8 functional modules. The modules activities were estimated using their constituent genes expression changes in the respective mutants. On observing the activity profiles of the modules, we grouped the modules into two classes: condition- specific modules (nucleosome assembly module to be specifically down-regulated in DR mutant), lifespan related modules (as these modules were oppositely regulated in lifespan decreasing mutants eg: mRNA processing). We also observed that tRNA and rRNA processing module was inactivated in all pro-longevity mutants. RNA polymerase III generates short non-coding RNAs including tRNAs and 5S rRNAs. Target of rapamycin kinase complex 1 is an important determinant of lifespan and regulates RNA pol III activity.[40] Study on yeast, worms and flies has shown that reduction in RNA polymerase III extends chronological and organismal lifespan respectively.[24] Our results are in good agreement with these previous observations on RNA polymerase III activity involvement in lifespan extension.

3.5.4. Dampened collagen expression associated with enhanced immune response

Collagen genes were found to be overexpressed in rIIS condition and germline ablated worms more interestingly a molecular signature of longevity.[23] Surprisingly we identified that knocking the expression of our collagen module members were promoting longevity. Later for detailed study, we focused on

one of the collagen module member, col-179 as this gene knock-down was sufficient to extend lifespan and ameliorate aging effects. Col-179 was observed to be consistently down-regulated in day 1,6 and 10 of daf-2 mutants.[33] Upon gene expression quantification studies of col-179 RNAi strains, we found that immune response module was activated. Additionally col-179 was altered as a down-stream effector of the longevity interventions. Most of the longevity interventions activate immune responses in *C.elegans* through activation of daf-16,[65],[76].[111] These observations suggests that col-179 activates immune module as a secondary downstream effect of the longevity interventions. The activated innate immune response in *C.elegans* is important for both anti-viral and anti-microbial defense response. The col-179 RNAi strains could impart resistance to pathogens thus provide better survivability compared to mock treated strains.

3.6. Author contribution and acknowledgements

Andreas Beyer and Adam Antebi conceived the idea. I implemented the network propagation algorithm in R and analysed the RNA-sequencing data. Andreas Beyer, Adam Antebi, Roman Ulrich Müller and I designed the experiments. Özlem Karaley performed the lifespan experiments. I would like to thank Cédric Debès and Marius Garmhausen for their valuable suggestions on network propagation.

4. Interplay of transcription factors in regulation of longevity in *C.elegans*

4.1. Introduction

Lifespan of the organism is influenced by genetic and environmental factors. In *C.elegans* multiple molecular pathways has discovered in regulating lifespan.[103] The signaling pathways include reduced insulin/IGF-1 like signaling, dietary restriction,[50] signals from the germline,[38],[52] and the hypoxia signaling. These lifespan extensions are exhibited through single gene mutations, importantly the longevity trait is regulated through specific transcription factors. In C.elegans weaker mutations in daf-2 gene, encoding a hormone receptor orthologous to insulin or IGF-1 receptor has shown to double the lifespan by activating the transcription factors daf-16 and skn-1,[68],[41].[45] In C.elegans the maximum lifespan is observed with daf-2 mutants i.e. by reducing insulin/IGF-1 signaling.[41] Using microarray experiments the downstream genes of daf-16 has been identified and found to be responsible for lifespan extension.[68] With the advancements in the sequencing platforms, the targets of daf-16 transcription factor has been profiled using ChIP-Sequencing,[69].[48] These target genes exert various functions including stress-resistance, anti-microbial response and metabolic functions.

Similarly the dietary restriction induced lifespan extension is regulated by specific transcription factor, pha-4 which has important role in embronic development of the foregut. Later in the life of worm pha-4 is involved in glycogen production and glucose homeostasis during the caloric restriction conditions.[73] The binding sites of pha-4 is specific for the different lifestages, the binding sites were dramatically shifted between the embro stage and starved larval condition. The binding sited from embronic stage corresponded to genes involved in organ development and binding sites from starved larval condition represented genes involved in metabolism.[112]

Genes differential expression reflects the regulatory networks that operative in the organism under particular condition. Precise and comprehensive characterization of mRNA transcript levels is essential in understanding the molecular changes responsible in dictating a phenotype.

4.1.1. Study aims

In this study the RNA-sequencing data from the lifesan extending and decreasing mutants has been exploited to understand the transcription factors regulatory interactions in *C.elegans*. The ChIP-seq profiles of 176 transcription factors were processed and target genes of the individual transcription factors at their lifestages were predicted. Subsequently the target genes expression profiles in the lifespan extending and decreasing mutants were analysed, the mean expression change of the target genes of a transcription factor in the particular condition is estimated as TF activity. Further TF activities of the mutants were correlated with lifespan change of mutant in order to associate TF with lifespan change in *C.elegans*.

4.2. Datasets

4.2.1. ChIP-Seq datasets

Transcription factors ChIP-seq datasets were retrieved from the modENCODE database (ftp://data.modencode.org/). The transcription factors were profiled at different lifestages, only for few transcription factors all the lifestages were covered. There were two biological replicates for every experiment, in total 176 experiments raw sequence files were retrived.

4.2.2. RNA-seq datasets

To study the insulin signalling mediated lifespan extension we have daf-2 (e1370: nucleotide substitution mutant for insulin signalling receptor), compared to the wild type (N2) and daf-2 (e1370);daf-16 (mgDf50), longevity phenotype of daf-2 mutants are dependent on daf-16 transcription factor. Therefore we have the double knockout mutants as a second control and comparison of the daf-2 transcriptome with these strains allow us to pinpoint the lifespan specific molecular changes.

To identify the molecular changes of dietary restriction mediated lifespan extension, eat-2 (ad465: mutant with reduced pharyngeal pumping mimicking dietary restriction), a genetic mimetic of DR is compared to wild type (N2) and eat-2 (ad465);nhr-62 (tm1818) as longevity phenotype is partially dependent on nhr-62 transcription factor.

Similarly for germline ablation signalling, glp-1 (e2141:mutant that have impaired germ cell proliferation) is compared to wild type (N2) and glp-1

(e2141); daf-16 (mu86), longevity phenotype of glp-1 is also mediated through daf-16. In addition to daf-16 transcription factor, Myc-like complex (MML-1/MXL-2) and hlh-30 were found to be associated with germline induced lifespan extension. Thus these datasets were retrieved from ArrayExpress database with the accession: E-MTAB-3686.

For hypoxia signalling, vhl-1 mutant worms (ok161: mutation in subunit of E3-ubiquitin ligase complex) that extend lifespan is compared to wild type and vhl-1 (ok161); hif-1 (ia4) as vhl-1 mediated longevity is dependent on the hypoxia response transcription factor hif-1.

4.3. Workflow

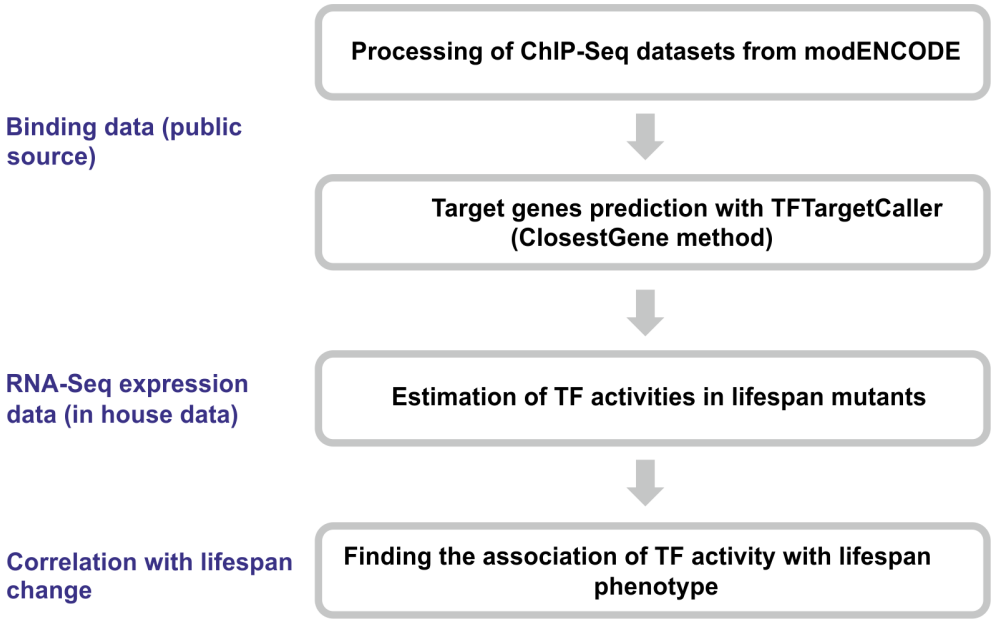

Figure 20: Transcription factors activity analysis pipeline. The transcription factors ChIP-Seq datasets mapped read files were retrieved from modENCODE database. The genome aligned reads were used for peak calling with MACS2 software. Subsequently the peak called files were used for target gene prediction with TFTargetCaller R package. Once target genes of the transcription factors were predicted, the target genes log fold changes in the lifespan increasing and decreasing mutants were computed. Subseqently the target genes mean expression changes were estimated for transcription factor's activity in the respective mutant. Calculated TF activities were correlated with the lifespan change of the mutants from geneAge database.

4.4. Methods

4.4.1. modENCODE ChIP-Seq data processing

Raw sequence fastQ files for every experiment were retreived from modENCODE database. The read qualities were assessed using the FastQC. For the libraries that had adaptor sequences contamination and poor quality reads, raw sequence reads were trimmed using the cutadapt (V1.12).[58] The trimmed reads were mapped to the *C.elegans* reference genome assembly from Ensembl WBcel235. For aligning the reads to the reference genome Bowtie 2[51] was used. Subsequently the duplicaed reads were removed using MACS2 filterdup tool and filtered reads were used for peak calling with MACS2 callpeak command. Further peakcalled files were processed using TFTargetCaller R package.[82]

4.4.2. Target gene prediction using TFTargetCaller

Transcripts gene positions were retrieved from the Ensembl64 database (www.ensembl.org) using the R package biomaRt.[21] All genes of *C.elegans* that are described as protein-, miRNA-, or lincRNA-coding were included for target gene prediction. For genes with multiple transcripts, the most 5âĂš TSS position of the transcripts was considered as the representative TSS of the gene. The peakcalled files of the transcription factors were imported to R environment and each peak was assigned to its closest gene, the peak that is closest to that gene was considered. Subsequently the peaks were scored with peak-to-gene distance distribution.[82] Through randomizations the q-values were calculated for every gene, representing the probability of the particular gene being a target of the transcription factor. Genes with $q - value \leqslant 0.1$ were predicted as target genes of a TF and were considered for subsequent analysis.

4.5. Results

4.5.1. Known longevity associated transcription factors

Firstly as a proof of principle the known longevity transcription factors were analysed. Expression changes of target genes in a condition would indicate the transcription factor's activity in the given condition. We have exploited this notion in assessing the activity of the transcription factors in the lifespan related mutants. For each TF the target genes were predicted and their log2 fold changes in the mutants were calculated. DAF-16 is known to be active in the reduced insulin signaling mutant (daf-2), thus we expected that the predicted target genes of daf-16 should be overexpressed in the daf-2 mutant. As anticipacted the target genes of daf-16 were overexpressed in the daf-2 mutant, where it has to be active. Interestingly we also observed that on average the target genes of daf-16 were overexpressed in lifespan extending mutants and down-regulated in lifespan decreasing mutants. We reasoned that daf-16 regulates a large classe of genes that are associated in lifespan extension and it is likely that some of these genes will be expressed in the other longevity mutants through target sharing with other TFs. However the strength of target genes expression in daf-2 is much higher compared to their expression in other longevity mutants, thus still maintaining the specificity of daf-16 in insulin signaling.

Further we assessed pha-4 TF activity in the lifespan related mutants, similarly target genes of pha-4 were predicted and their expression changes in the lifespan related mutants were computed. Dietary restriction (DR) induced lifespan extension is mediated through pha-4 hence the target genes are anticipated to be overexpressed in eat-2, a genetic mimetic of DR. We observed that on average pha-4 target genes were overexpressed in eat-2 mutants and in other longevity mutants, interestingly the target genes were not expressed or down-regulated in the lifespan decreasing mutants. pha-4 target genes expression correlation with lifespan change intrigued us, hence we estimated the TF activity as mean log fold changes of target genes in a mutant and the TF activities of the mutants were correlated with the lifespan change in fraction. The lifespan changes of the mutants were retrieved from the genAge database (http://genomics.senescence.info/genes/). On correlating the daf-16 activities with lifespan change of the mutants we observed a positive correlation of 0.842. For pha-4 we observed a correlation of 0.88.

4. Interplay of transcription factors in regulation of longevity in C.elegans

HLH-30 is one of the 42 HLH transcription factors in *C.elegans*, it regulates the expression of multiple autophagy-related and lysosomal genes. HLH-30 is shown to be required for the lifespan extension in all *C.elegans* longevity mutants.[52] Moreover overexpression of hlh-30 in control strains extends lifespan, all these findings suggests that hlh-30 activity may be a universal mechanism of longevity in *C.elegans*. To test our approach we predicted the target genes of hlh-30 profiled at the L4 larval stage and the targets expression change in the mutants were computed. We observed that hlh-30 target genes were over-expressed in the longevity mutants suggesting that hlh-30 is active in these mutants. hlh-30 targets were not changing in their expression or down-regulated in the lifespan decreasing mutants. Subsequently the hlh-30 activities in the mutants were estimated and correlated with their lifespan change, the correlation was 0.786. These findings confirmed our approach of estimating TF activities and subsequently identifying their association with lifespan.

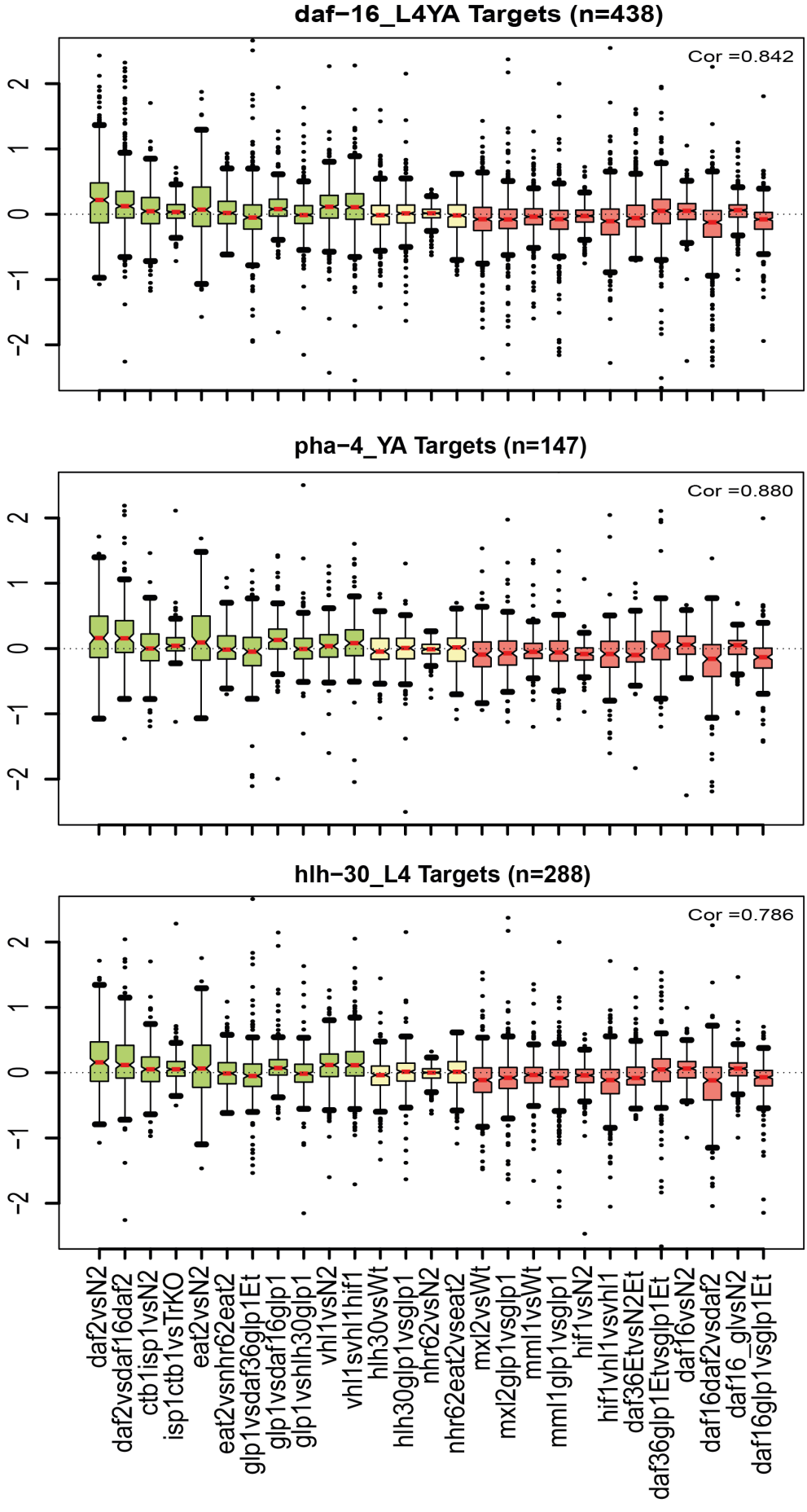

Figure 21: continued on the following page

Figure 21: Transcription factors: daf-16,pha-4 and hlh-30 target genes expression changes in the lifespan extending and decreasing mutants of *C.elegans*. A) daf-16 ChIP-Seq data was retrieved from modENCODE database and its target genes were predicted using TFTargetCaller R package. Target genes expression changes in the lifespan increasing and decreasing mutants. B) pha-4 ChIP-Seq data was processed and the targets were predicted, target genes expression changes in the lifespan mutants of *C.elegans*. C) hlh-30 target genes log2 fold changes in the mutants of worm. The green boxes represents the log fold changes from the lifespan extensing mutants, yellow boxes represent mutants that have no change in their lifespan and saffron denotes the lifespan decreasing mutants. On each plot correlation of TF activity with the lifespan change is indicated.

4.5.2. Common targets of DAF-16:PHA-4:HLH-30

DAF-16, PHA-4 and HLH-30 target genes were overepressed in the longevity mutants still showing their specificity to their respecitive signaling pathways as the lfc signal is particularly stronger in the respective mutants. In daf-2, daf-16 target genes has higher expression signal similarly in eat-2, pha-4 target genes has higher strength of expression in eat-2 as well as in daf-2. We reasoned that daf-16 regulates a large classe of genes that are associated in lifespan extension and it is likely that some of these genes are expressed in the other longevity mutants through sharing of targets with other TFs. In ordder to test this notion we interesected the target genes of DAF-16, PHA-4 and HLH-30 and we found that there were 39 genes that were commonly regulated by these three longevity associated transcription factors. Further functional enrichment of these common target genes showed significant enrichments for the followinf cellular functions: amino acid biosynthesis, male sex differentiaion,response to growth stimulus, lipid metabolism and regulation of gene expression this includes other down-stream transcription factors that are regulated through these three master transcription factors. Then we estimated the mean log fold changes of these 39 common targets of daf-16, pha-4 and hlh-30 and correlated with lifespan change of the mutants, we observed that these common targets were also strongly positively associated with lifespan of *C.elegans* with correlation value of 0.787. This intersection and correlation analysis of the shared targets suggests that these genes could be the common denominators of lifespan extension.

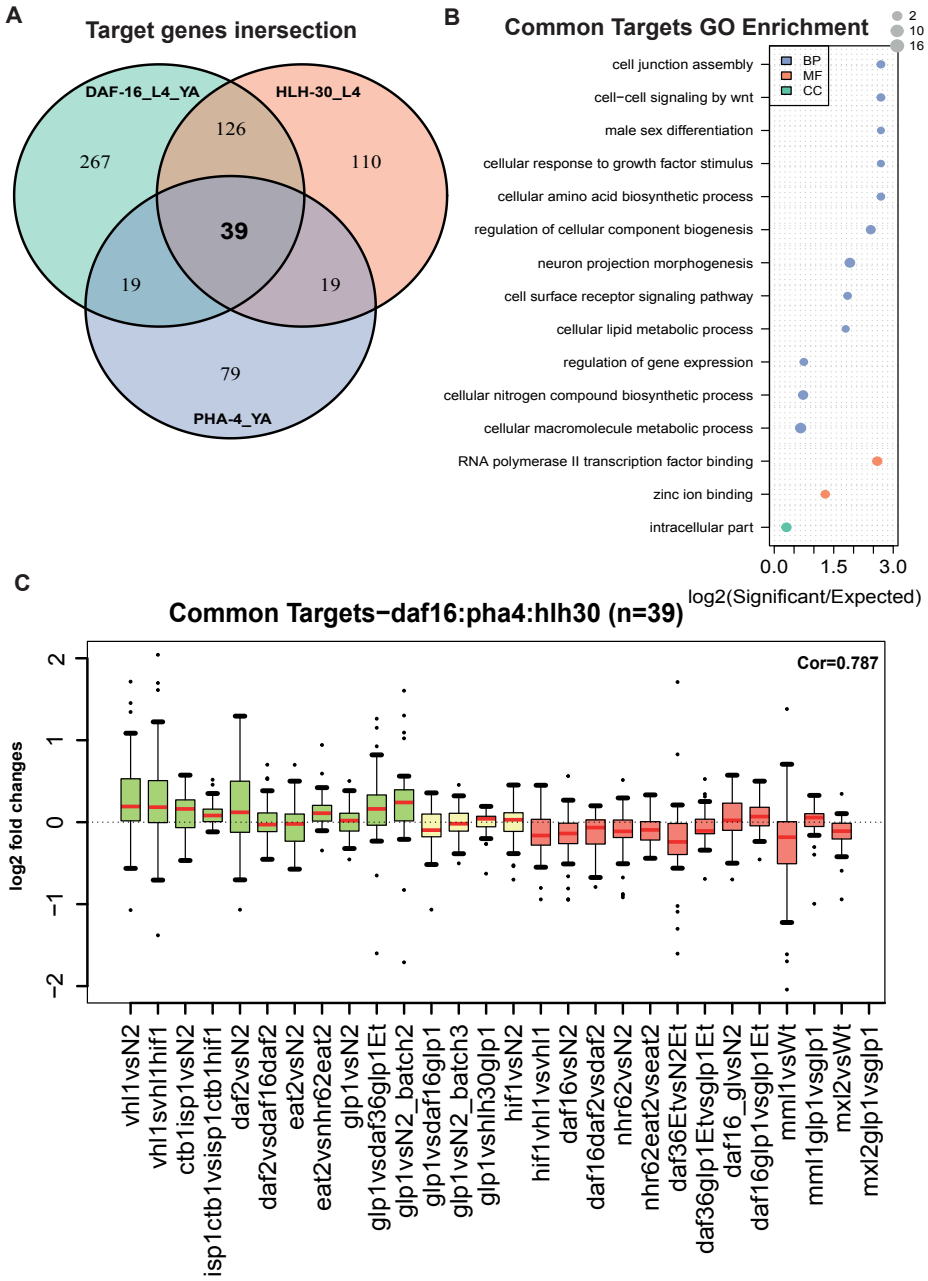

Figure 22: Transcription factors: daf-16:pha-4:hlh-30 common target genes expression changes in the lifespan extending and decreasing mutants of *C.elegans*. A) Venn diagram representing the overlap of the target genes of the three important longevity transcription factors of *C.elegans*. B) GO enrichment of the common target genes. C) Common target genes log2 fold changes in the respective mutants of *C.elegans*. The green boxes represents the log fold changes from the lifespan extensing mutants, yellow boxes represent mutants that have no change in their lifespan and saffron denotes the lifespan decreasing mutants. On each plot correlation of TF activity with the lifespan change is indicated.

4.5.3. Experimental design for validation of TFs interaction

We found that the most studied transcription factors of longevity, daf-16, pha-4 and hlh-30 commonly regulate a set of genes and they are associated with lifespan extension. Our analysis suggests that these transcription factors are in functional interaction. Most importantly HLH-30 is required for lifespan extension in all *C.elegans* longevity mutants[52] irrespective of the longevity mediating signaling pathway. Moreover overexpression of hlh-30 in control strains extends lifespan, all these findings suggests that hlh-30 activity may be a universal mechanism of longevity in *C.elegans*. Based on our observations we hypothesize that based on the up-stream longevity intervention, hlh-30 is activated through activation of the respective signaling pathway master transcription factor. For example in the reduced insulin/IGF-1 signaling pathway (daf-2) daf-16 becomes activated to induce hlh-30 activation and reguates common target genes that are causal for lifespan extension, similarly during dietary restriction pha-4 is activated and in turn activates hlh-30 to drive the expression of lifespan causal genes. In order to test this hypothesis we have an co-immunoprecipation assay for hlh-30 in the different genetic backgrounds, N2 control strain, daf-2 mutant strain and eat-2 mutant strain. The interactome of the hlh-30 can be captured in these genetic backgrounds and among the differentially interacting proteins we expect daf-16 in daf-2 mutants and pha-4 in eat-2 mutants.

4.5.4. Estimation of TF activities

To assess activities of all ChIP-seq profiled transcription factors, the target genes of the TFs from different lifestages were predicted. The predicted target genes expression change, log fold changes in the mutants were computed. The TF activity in each mutant was estimated as mean log2 fold change of targets in the mutant. Further the TF activities were correlated with lifespan change of mutants. This resulted in TFs that are strongly positively associated with lifespan ($Cor \geqslant 0.5$), TFs that are slightly associated with lifespan ($\text{Cor} < 0.5$ & > 0), TFs that are negatively associated with lifespan ($\text{Cor} < 0$). Among the strongly lifespan associated transcription factors we observed the already known longevity TFs such as JUN-1,[102] EOR-1,[53] SKN-1,[101],[100],[11] DVE-1.[96] Using our approach we have identified the other novel transcription factors that are associated with lifespan extension.

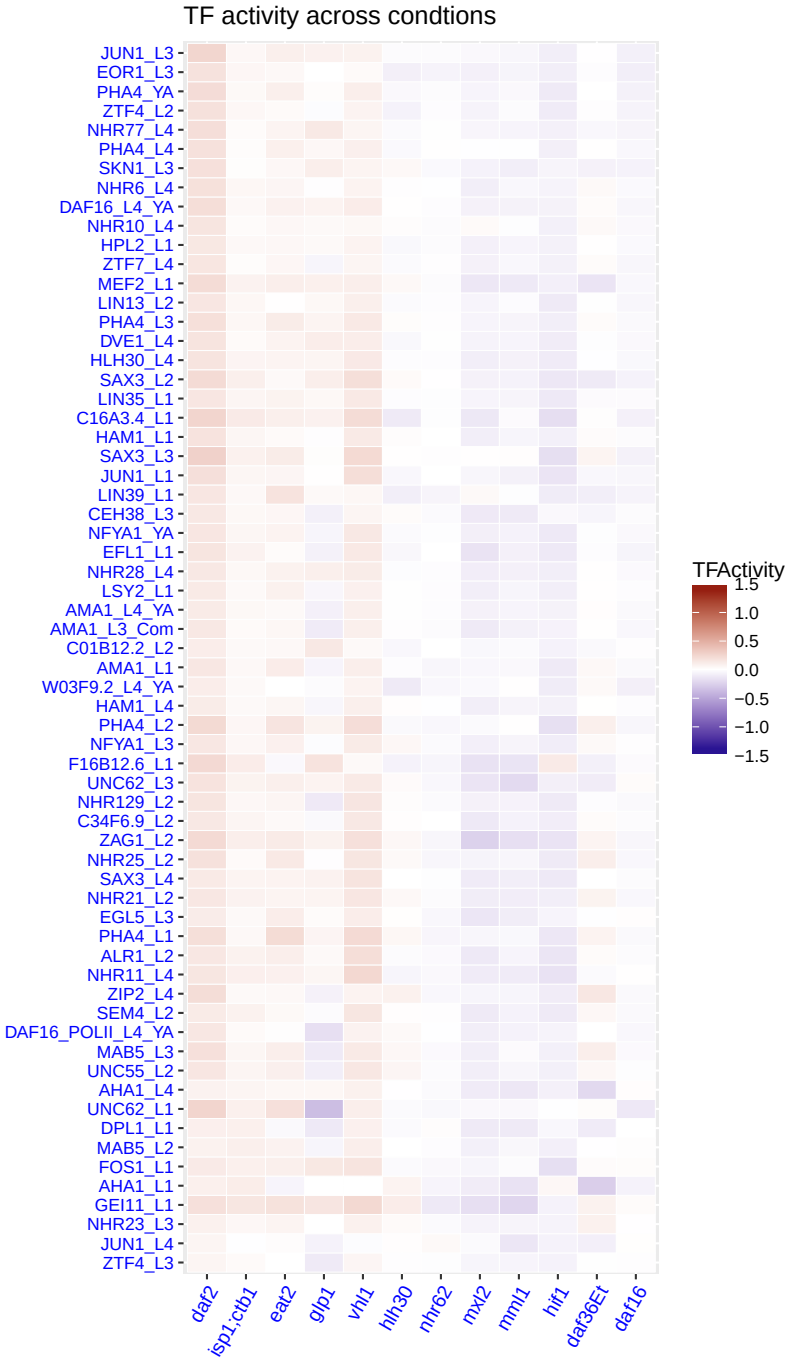

Figure 23: Transcription factors robustly associated with lifespan of *C.elegans*: Identified through their activity correlation with lifespan change. The transcription factors activities are estimated as mean log fold changes of their predicted target genes in each mutant and are represented. The TF activities are correlated with lifespan change of the respective mutants, in the plot the transcription factors are sorted based on their correlation values.

5. Altered tissue specific sub-networks in reduced insulin/IGF-1 signaling of *Drosophila*

(Part of this chapter has been published in: Luke S Tain, Robert Sehlke, Chirag Jain, Manopriya Chokkalingam, Nagarjuna Nagaraj, Paul Essers, Mark Rassner, Sebastian Grönke, Jenny Froelich, Christoph Dieterich, Matthias Mann, Nazif Alic, Andreas Beyer, and Linda Partridge. A proteomic atlas of insulin signalling reveals tissue-specific mechanisms of longevity assurance. Molecular Systems Biology, 13(9), September 2017.)

5.1. Introduction

Insulin/IGF-1 signaling (IIS) has essential roles in growth, metabolism, stress resistance and lifespan. In the rececnt studies reduced activities of insulin/IGF-1 signaling has shown to increase lifespan of laboratory model organisms such as yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* as well as in mice.[14] In *Drosophila melanogaster* the IIS pathway consists of 7 insulin like peptide ligands (DILP1-7), an Insulin receptor (DInR), an insulin receptor substrate and a downstream forkhead box O transcription factor (dFOXO). Each of the DILPs shows characteristic spatio-temporal expression patterns.[32] DILP1,2,3 and 5 are expressed in brain median neurosecretory cells (mNSCs). Partial ablation of median neurosecretory cells of *Drosophila* adult brain leads to reduced activity of IIS pathway and extension of lifespan.[13]

Long-lifespan of the reduced IIS pathway is mediated by a transcription factor, dfoxo in *Drosophila melanogaster* and daf-16 in *C.elegans*,[41],[109].[84] Lowered IIS mediated longevity and xenobiotic resistance phenotypes are regulated by the transcription factor dfoxo and other reduced IIS phenotypes are regulated by other transcription factors.[84] Thus dfoxo directs gene expression changes that are potentially causal for lifespan extension and therefore aid in pinpointing the lifespan extension specific mechanisms in *Drosophila melanogaster*. In *C.elegans* transcriptome studies on reduced IIS mutants has revealed lifespan specific molecular mechansims,[68],[23].[33] Similarly in *Drosophila* gene expression studies has revelaed the lifespan associated biological processes,[95].[3] Various post transcriptional mechanisms are known to modify the protein abundances.[54]

Thus quantification of protein abundances are more representative of the functional state of the organism.

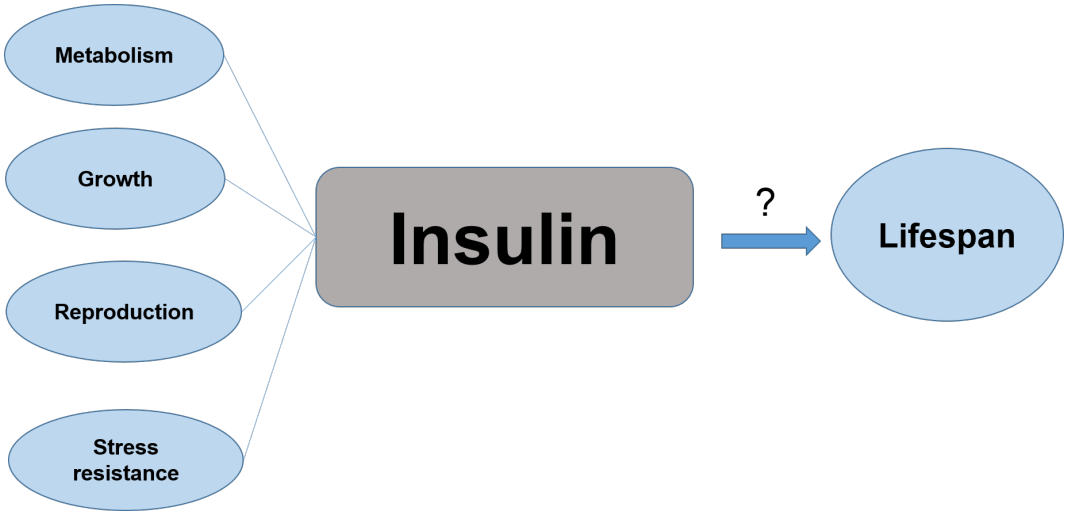

Figure 24: Pleotropic effects of insulin signaling network.

In this study as a first genetic model, we profiled the proteomes of the mNSC ablated flies in the presence and absence of the transcription factor dfoxo. Modulation of the IIS in a tissue-specific manner is also associated to lifespan extension in *C.elegans* and *Drosophila*,[107],[29],[19].[2] Induction of dfoxo in midgut and fat body activates the transcription of dilp6 in fat body. However dfoxo represses the activin ligand dowdle in muscle and these signals have an effect on the mNSCs thus dilp2 peptide levels are reduced in circulation,[7].[6] Furthermore activation of dilp6 is essential for the lifespan extension mediated by dfoxo.[7] Hence to pinpoint the tissue specific mechanisms that are mediated by lowered IIS through dfoxo we have characterized the proteome of four metabolically active tissues of *Drosophila*, brain, gut, fat body and thorax.

Interestingly the lifespan effect in *Drosophila* is also modulated by an intracellular symbiont *Wolbachia pipientis*, is a maternally transmitted bacteria,[63].[98] Furthermore *Wolbachia* has been shown to increase IIS in *Drosophila*.[39] Tetracycline treatment of wolbachia carrying fly strains leads to normal lifespan. Intriguingly dilp2-3, 5 mutants in the presence of wolbachia has shown extreme lifespan however *wolbachia* did not have any effect on the wDah control strains, suggesting specific interaction between IIS pathway and emphwolbachia. Therefore in this study as a second genetic model we have characterized the proteome of dilp2-3,5 mutants, WDah control strains in the presence and absence of *Wolbachia*. These proteomes were also quantified for the same four tissues as of

ablation model: brain, gut, fat body and thorax.

5.1.1. Study aims

In this study we aimed to understand the tissue specific responses to reduced IIS (rIIS) and pinpoint the lifespan specific molecular mechanisms that are mediated by the transcription factor dfoxo. To address this, dfoxo dependent and independent proteins were classified and the individual classes of proteins differential protein abundance p-values were propagated on the protein-protein interaction network of *Drosophila*. Subsequently the dfoxo dependent and independent network modules were revealed. Furthermore with the application of network propagation we unravelled the common molecular signatures of rIIS from two genetic models as well as lifespan mediating protein modules.

5.2. Datasets

For this comprehensive study on reduced insulin/IGF-1 signaling in *Drosophila melanogaster*. For abaltion genetic model, the systemic insulin responsive tissues brain, gut, fat body and thorax proteomes were characterized for mNSCs ablated fly strains and wDah control strains in the presence and absence of the transcription factor dfoxo. The data is already published and can be retrieved from PRIDE archive database with the accession PXD006225. For the dilp genetic model of rIIS, the dilp2-3,5 mutants and wDah control strains proteomes for the same four tissues were characterised with the presence and absence of the endosymbiont *wolbachia pipientis*. Thus for each genetic model of rIIS for each tissue we have two strains ablated or dilp mutant with control strain and two conditions either presence or absence of dfoxo or *Wolbachia*.

5.2.1. Drosophila interaction network (DroID)

The protein-protein interaction network of *D.melanogaster* was obtained from Drosophila interaction database (DroID).[67] It is a comprehensive database designed uniquely for the organism. In this database the protein-protein interactions are assembled from various sources such as flybase experimentally derived from physical interactions, protein interactions determined in large-scale co-affinity purification (co-AP)/MS screens from Perimon Lab, co-AP/MS screens by the Drosophila interaction mapping project, protein interaction generated from yeast two-hybrid system of Finley lab etc. The network was

further filtered for high confidence edges (40% confidence, n = 34,866).[110]

5.3. Methods

5.3.1. Proteome differential expression analysis

The proteins with differential abundances were tested against null hypothesis of no change, for ablation model mNSCs abaltion induced changes were evaluated by comparing long-lived vs wDah control strains. The differentially expressed proteins were subsequently classified into dfoxo dependent and dfoxo independent, dfoxo dependent proteins are differentially expressed in long-lived vs wDah control strains and also in the interaction term (long-lived - wDah vs ablated-dfoxo - wDah-dfoxo). dfoxo independent proteins were differentially expressed in long-lived vs wDah control strains and required to be equivalent with regard to differential expression change in the interaction term analysis.

For dilp2-3,5 rIIS model, reduced insulin signaling induced changes in protein abundances we compared long-lived (dilp2-3,5 mutants + wol) vs wDah control strains. For wolbachia induced changes in dilp2-3,5 mutants protein abundances were compared between long-lived (dilp2-3,5 mutants + wol) vs dilp2,3-5 mutants in the absence of *Wolbachia* (dilp2-3,5 mutants - wol).

5.3.2. Network propagation

The network was converted to an adjacency matrix and normalized with the Laplacian transformation (using the graph.laplacian function in the igraph R package). Differentially regulated proteins were classified into dfoxo-dependent and dfoxo-independent. For the analysis of dfoxo-dependent proteins, the p-values of proteins not belonging to this group were excluded (set to 1). Likewise, for the evaluation of the dfoxo-independent set, p-values of proteins not detected as such were excluded. Finally, the p-values were -log2 transformed and mapped to the network. After mapping, the transformed p-values on the individual nodes were diffused to their adjacent nodes using the spreading coefficient of 0.8 (corresponds to the percentage of sharing to neighbours). Subsequently, the corrected propagated scores were clustered based on their Euclidean distances (using the hclust function in R) allowing for the identification of enriched clusters for each tissue. For visualizing the enriched clusters from dfoxo-dependent and dfoxo-independent proteins on the same heatmap, the propagated profiles were subtracted (dfoxo-dependent propagated scores and dfoxo-independent

propagated scores).

5.3.3. Functional enrichment analysis

topGO was employed for Gene Ontology enrichment analysis and *D.melanogaster* genes and protein annotations were retrieved from gene ontology consortium (www.geneontology.org). To identify the significantly enriched Gene Ontology terms, one-sided elim Fisher procedure was employed ($\alpha \leqslant 0.05$).[1] The enrichment scores were calculated as log2 (#Detected significant genes/ #Expected significant genes). From the clustering of corrected propagated scores of genes, the individual cluster genes were tested for GO term enrichment against the network genes as reference background. Furthermore to functionaly characterize the consistently high score accumulating cluster genes were pooled and tested for gene enrichment again against the network genes. The significantly enriched GOTerms from each cluster was represented on the heatmap.

5.4. Results

5.4.1. Tissue specifically altered network modules in lowered insulin signaling induced by mNSC ablation

To identify the candidate proteins that could be causal for lifespan extension, first we have classified the differentially expressed proteins in ablated mNSCs strains with respect to wDah control strains into dfoxo dependent and independent proteins. Subsequently the dfoxo dependent proteins P-values from long-lived vs control strains differnetial expression analysis were propagated on the network with spreading coefficient of 0.8 for each tissue. Similarly for dfoxo independent proteins the P-values were transformed and propagated on the network until convergence. The propagated scores were combined and then clustered the dfoxo-dependent and dfoxo-independent responses to reduced IIS in each tissue and identified functional categories of these clusters with GO enrichment analysis.

The dfoxo dependent clusters were particularly interesting as dfoxo transcription factor mediates longevity and xenobiotic stress resistance phenotypes of rIIS pathway and other phenotypes are mediated in a dfoxo independent manner.Thus these clusters should include the processes that are causally associated with lifespan extension. In brain samples the dfoxo dependent protein modules were enriched for mitochondrial electron transport chain, mRNA

splicing and nucleosome assembly. Whereas gut samples showed enrichment for proteasome and ubiquitin mediated protein catabolism in a dfoxo dependent way. In fat body samples mitochondrial electron transport chain, ribosomal contituents and nucleosome assembly were enriched in dfoxo dependent clusters. Thus network propagation of scores from differentially expressed proteins in rIIS condition in the generic protein-protein interaction network clearly suggests tissue-specifically altereted protein modules. Further these results were experimentally validated for their association in lifespan extension.[93]

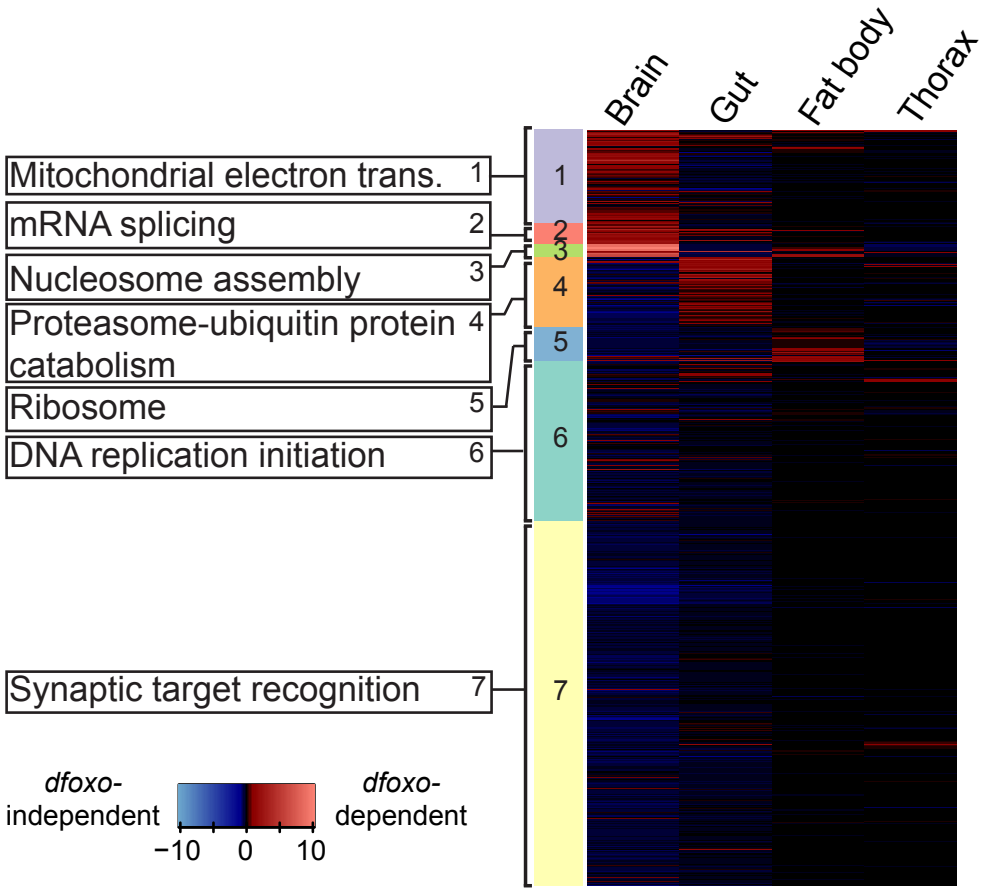

Figure 25: Hierarchical clustering and Gene Ontology enrichment analysis of dfoxo-dependent and dfoxo-independent rIIS-mediated regulation of the proteome. The P-vales from ablation vs wDah dfoxo dependent proteins were propagated on the DroID filtered network separately and likewise for the dfoxo-independent proteins P-values. For visualizing dfoxo dependent and independent protein modules, dfoxo-dependent_propagated scores were subtracted from dfoxo-independent_propagated scores

$$dfoxoDependentScore - dfoxoIndependentScore$$

and clustered based on their propagated scores. Heatmap of significantly regulated dfoxo-dependent (red) and dfoxo-independent (blue) proteins in response to reduced IIS. Coloured side bars represent network clusters and associated most significantly enriched GO terms.

5.4.2. Network integration of two lowered insulin signaling models

To investigate consistently affected processes in reduced IIS condition from dilp2-3,5 and ablation models, the network propagation results from the ablation and dilp were integrated. From the ablation proteomics, the absolute log2 fold changes from long-lived vs wDah control and absolute log fold changes from ablated longlived vs ablated foxo Null strains were propagated independently on the network to identify the rIIS induced changes and rIIS induced changes in foxo background respectively. Similarly from dilp2-3,5 model the log fold changes from dilp2-3,5 mutant vs wDah control strains and dilp2-3,5 mutant vs dilp2-3,5 in the absence of wolbachia were propagated on the network. Thus each protein in the network has received four propagated scores, the minimum score for each protein was calculated. If the minimum score of the protein is positive then it is initially altered (either by its own alteration of protein expression or through its neighbours in network) in both dilp and ablation in foxo and wolbachia dependent manner respectively. Hence these positive scored proteins are most likely causally linked to lifespan extension.

The heatmap represents the proteins with minimal positive scores from network propagation of both models for each tissue. In total there were 2040 proteins that were lifespan associated revealed from both dilps and ablation in any of the four tissues. GO enrichment analysis was done on the individual cluster proteins with all network genes as reference set. From this integration analysis, translation and related terms of translation were enriched in different tissues (ex: tRNA metabolic processes is specifically enriched in the Gut, rRNA processing being enriched in both Fat and Gut) and thus we focused on translation.

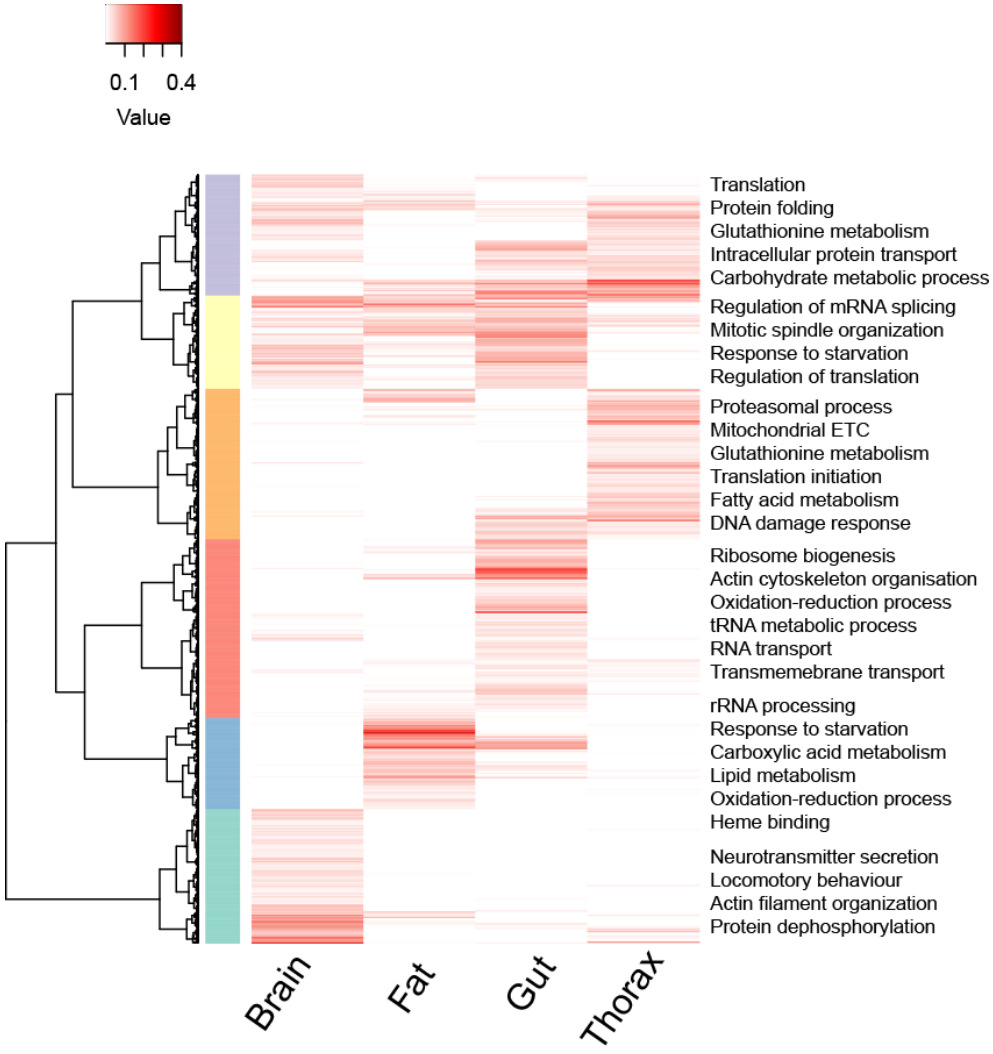

Figure 26: Hierarchical clustering and Gene Ontology enrichment analysis of consistently high scored proteins from ablation and dilp2-3,5 models of rIIS pathway. The proteins absolute log fold changes from ablated vs wDah control strains comparison and ablated vs ablated-dfoxo comparison of ablation genetic model were propagated on the DroID filtered network. Similarly from the dilp2-3,5 genetic model proteins log fold changes from dilp mutants vs control strains and dilp mutants vs dilp-wolbachia strains were propagated on the network. This way both the rIIS responses in foxo and wolbachia mediated changes can be captured. The heatmap represents the proteins which are positively scored from all the four propagations. Coloured side bars represent network clusters and associated most significantly enriched GO terms are listed sideby.

5.4.3. Translation: a common denominator of rIIS in *Drosophila*

Subsequently the proteins direction of change for individual components of translation were analysed. protein synthesis requires production and proper assembly of ribosomes and also synthesis of enough tRNAs matching the codons of the expressed mRNAs. Thus we have separately investigated the log fold changes of proteins in different GOterms associated to ribosome (rRNA processing, ribosome biogenesis, all the ribosomal proteins as such, nucleolus) and for tRNA metabolism and genes that are annotated as regulating protein synthesis. In ablated fat body samples there is foxo-dependent reduction in expression of ribosomal proteins and all the allied processes (rRNA processing, ribosomal biogenesis). This observation is experimentally confirmed by S35 incorporation assay to show the de novo protein synthesis rates in different tissues and the results are very much the same: In the fat bodies there is reduced incorporation of labelled amino acids (implying reduced protein synthesis rates could be as a result of reduced ribosomal proteins) and this phenotype is dfoxo dependent.[93]

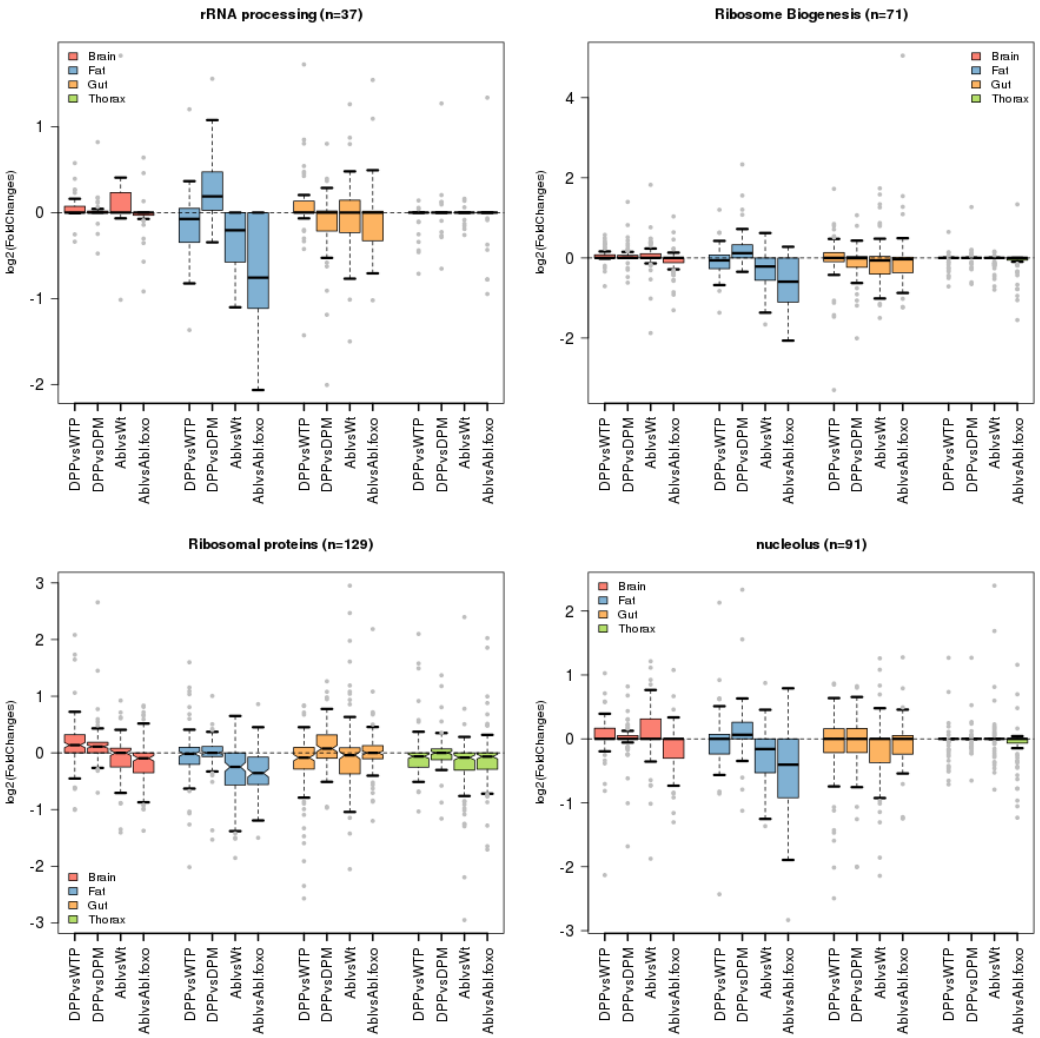

Figure 27: Ribosome associated proteins expression in reduced IIS conditions (both dilps and Ablation). Genes from the GOterms rRNA processing, ribosome biogenesis, ribosome and nucleolus were retrieved. Proteins of these individual terms log2 fold changes are plotted. Of dilp rIIS model, DPP corresponds to dilp2-3,5 mutant in the presence of wolbachia, WTP corresponds to control strains in the presence of wolbachia and DPM corresponds to dilp2-3,5 in the absence of wolbachia. Of mNSCs ablation rIIS model, Abl corresponds to ablated strains, Wt corresponds to wDah control strains, Abl.foxo corresponds to ablated dfoxo null strains.

Intriguingly the tRNA metabolism proteins were also reduced in long-lived flies brain, gut and fat bodies in a foxo dependent manner. In fat bodies the reduced tRNA metabolism proteins levels are reduced as a consequence of reduced ribosomal proteins and both together contributing to reduced protein

synthesis rates. In gut samples the tRNA metabolism proteins were particularly reduced in their abundances in ablated strains. This suggests that each tissue responds differently for rIIS pathway, i.e. in fat bodies both ribosomal proteins and tRNA metabolism proteins are strikingly down regulated whereas in gut samples the ribosomal proteins are not consistently down regulated but tRNA metabolism proteins are strikingly down-regulated.

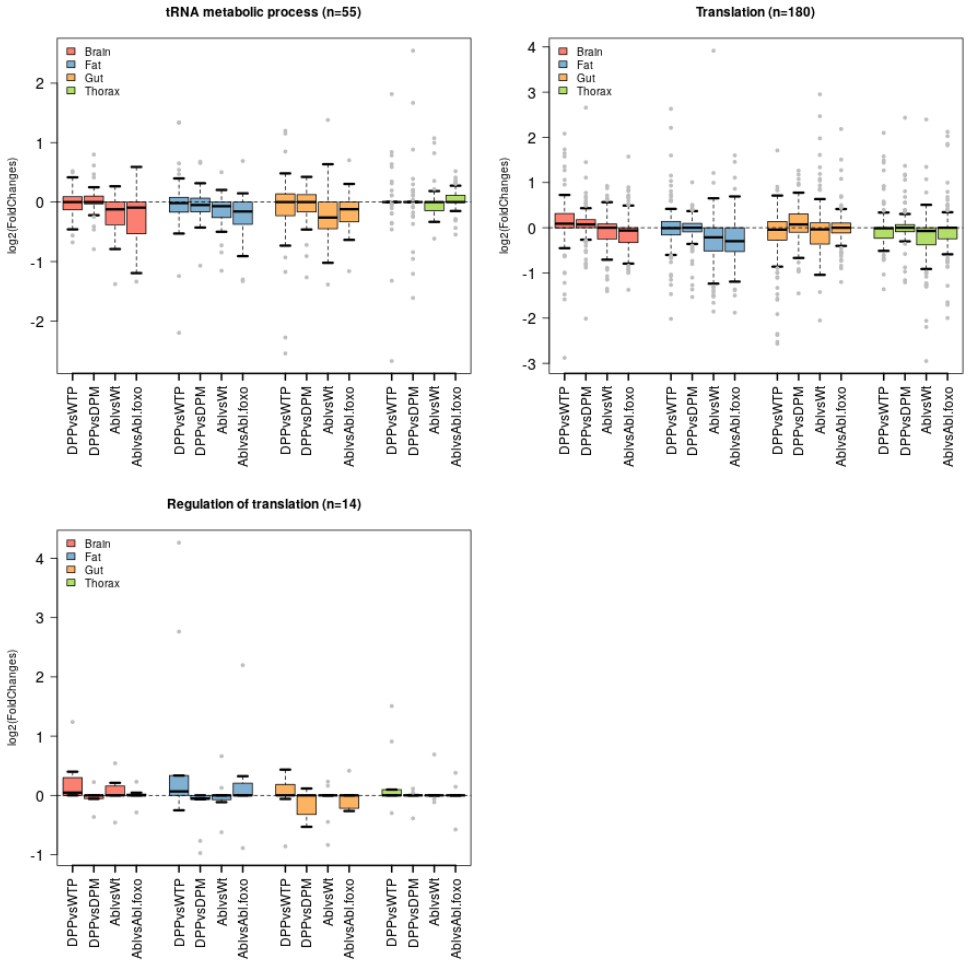

Figure 28: The tRNA metabolism and other translation regulating proteins expression in reduced IIS conditions (both dilps and Ablation). Genes from the GOterms tRNA metabolic process, translation and regulation of translation were retrieved. Proteins of these individual terms log2 fold changes are plotted. Of dilp rIIS model, DPP corresponds to dilp2-3,5 mutant in the presence of wolbachia, WTP corresponds to control strains in the presence of wolbachia and DPM corresponds to dilp2-3,5 in the absence of wolbachia. Of mNSCs ablation rIIS model, Abl corresponds to ablated strains, Wt corresponds to wDah control strains, Abl.foxo corresponds to ablated dfoxo null strains. n represents no of proteins annonated to have the respective function

5.5. Discussion

Lowered activity of insulin/IGF-1 signaling is an evolutionarily conserved pathway that extends lifespan and healthspan.[25] With the network propaga-

tion of the differentially expressed proteins from mNSCs ablated *Drosophila* strains we observed that the proteome has tissue speific responses. Each tissue has responded by altering specific protein modules for example mitochondrial electron transport processes is altered in brain and fat bodies. Further the experimental assays indicates that fat body specific increased respiration is causally associated to lifespan extension in the rIIS mutants of fruit fly. Increasing mitochondrial biogenesis has been shown to extend lifespan in both fat bodies and gut tissues of *Drosophila*.[81] Though the exact mechanism of how increased mitochondrial respiration is mediating lifespan extension is still unclear.

Loss in protein homeostasis is one of the hallmarks of ageing,[56].[47] We observed that in the rIIS fly strains the proteasome proteins were increased in their abundances only in the gut tissues. Proteasomes are complex of proteins that functions in degrading of proteins that are damaged through proteolysis. During aging decreased proteasome function is associated with decline in protein homeostasis.[99] Increased expression of proteasome sub-unit RPN-6 provides resistance to proteotoxic stress and extends lifespan in *C.elegans*.[105] Lifespan extension mediated by over expression of the proteasome sub-unit is likely due to maintenance of the cellular proteome. The increased proteasome activity in gut was recapitulated by expressing RPN-6 subunit specifically in the gut tissues of wDah strains and the causal association with longevity was demonstrated.[93]

From integrating the rIIS responses of mNSCs ablation model and dilp2-3,5 mutant model we observed that translation is reduced in the fat bodies and also in gut tissues of rIIS mutants. Global reduction in translation can extend lifespan of *C.elegans*,[7234] and *Drosophila*.[106] Furthermore reduced translation is a conserved response to reduced insulin/IGF-1 signaling pathway,[88],[59].[22] In this study we found that translation is reduced through different mechanisms in each tissue for example in the fat body the ribosome biogenesis and other ribosome related proteins are reduced in their abundances and in the gut tissue tRNA metabolism is reduced. This suggests that in the fat body there could be a overall reduction in the protein synthesis due to reduced ribosomes. Whereas in gut there lies a possibility that only a selected class of proteins synthesis rates might be altered as a result of reduced tRNA metabolism and not necessarily exhibiting an overall reduction in translation. However this observation has to be validated with tRNA quantification and translation activity assays in gut tissues in rIIS condition.

5.6. Author contribution and acknowledgements

Luke S Tain and Linda Partridge designed the experiments. Andreas Beyer, Luke S Tain, Linda Partridge, Robert Sehlke and I designed the data analysis. Robert Sehlke and I performed the data analysis.

A. Appendix

A.1. Candidate genes selected for lifespan screening from network propagation results from *C.elegans*

Gene name	Wb id	Clone name	Screen1	Screen2	Mean
alh-10	WBGene00000116	C54D1.4	1.34680135	1.21457490	1.28068812
alh-2	WBGene00000108	K04F1.15	1.35802469	0.68940493	1.02371481
aly-2	WBGene00000121	F23B2.6	1.58024691	0.64406780	1.11215736
asfl-1	WBGene00007277	C03D6.5	0.86956522	0.79722222	0.83339372
aspm-1	WBGene00008107	C45G3.1	1.03225806	1.12272727	1.07749267
B0035.6	WBGene00007109	B0035.6	0.15384615	0.60774411	0.38079513
C02B10.6	WBGene00015331	C02B10.6	1.53846154	1.32352941	1.43099548
C04E6.11	WBGene00015425	C04E6.11	0.20689655	1.05769231	0.63229443
C18A3.3	WBGene00015941	C18A3.3	1.18863049	0.63127690	0.90995370
C44E4.4	WBGene00016653	C44E4.4	0.38314176	0.35256410	0.36785293
C45G3.3	WBGene00001590	C45G3.3	0.35555556	0.47107438	0.41331497
cct-8	WBGene00021934	Y55F3AR.3	0.15384615	0.25541796	0.20463206
cdc-25.2	WBGene00000387	F16B4.8	0.62222222	0.77727273	0.69974747
cht-3	WBGene00016084	C25A8.4	0.83333333	0.98214286	0.90773810
col-119	WBGene00000693	C53B4.5	1.37931034	1.41025641	1.39478338
col-178	WBGene00000751	C34F6.2	1.48148148	1.07561930	1.27855039
col-179	WBGene00000752	C34F6.3	1.73913043	1.68269231	1.71091137
col-19	WBGene00000608	ZK1193.1	0.86274510	0.91093117	0.88683814
col-20	WBGene00000609	F11G11.11	0.38888889	1.40156454	0.89522671
col-92	WBGene00000667	W05B2.6	1.18203310	1.22679045	1.20441177
col-93	WBGene00000668	W05B2.5	1.02564103	1.32625995	1.17595049
cpar-1	WBGene00010036	F54C8.2	1.56862745	0.85953878	1.21408312
cra-1	WBGene00020068	R13F6.10	0.82978723	1.28296146	1.05637435
ctps-1	WBGene00012316	W06H3.3	0.00000000	0.00000000	0.00000000
cul-5	WBGene00000840	ZK856.1	1.85714286	0.84860174	1.35287230
D2045.2	WBGene00008422	D2045.2	1.22222222	0.88811995	1.05517109
ddo-3	WBGene00017648	F20H11.5	0.90476190	0.39108062	0.64792126
dis-3	WBGene00001001	C04G2.6	0.39682540	0.56285178	0.47983859

dnj-13	WBGene00001031	F54D5.8	0.78348214	0.86659664	0.82503939
E04D5.1a.1	WBGene00008480	E04D5.1a.1	1.10204082	2.68888889	1.89546485
eftu-2	WBGene00001166	ZK328.2	1.05050505	0.69090909	0.87070707
emb-5	WBGene00001259	T04A8.14	0.04273504	0.00000000	0.02136752
emo-1	WBGene00001303	F32D8.6	0.00000000	0.00000000	0.00000000
F25G6.9	WBGene00017800	F25G6.9	1.17857143	0.83877996	1.00867569
F38A1.8	WBGene00009521	F38A1.8	0.15686275	0.03140496	0.09413385
F48E3.4	WBGene00018605	F48E3.4	1.30370370	0.83041958	1.06706164
F53H1.1	WBGene00018776	F53H1.1	0.00000000	0.00000000	0.00000000
F55F10.1	WBGene00018898	F55F10.1	1.88679245	1.61538462	1.75108853
F57B9.3	WBGene00018997	F57B9.3	0.33333333	0.04668305	0.19000819
fzy-1	WBGene00001511	ZK177.6	1.00000000	1.26461039	1.13230519
gfat-2	WBGene00009035	F22B3.4	1.33333333	0.95297806	1.14315569
glh-1	WBGene00001598	T21G5.3	2.33333333	2.56381798	2.44857566
glh-2	WBGene00001599	C55B7.1	1.33333333	0.21983471	0.77658402
gst-10	WBGene00001758	Y45G12C.2	0.89540816	0.80000000	0.84770408
gst-44	WBGene00001792	F13A7.10	0.97222222	1.05224964	1.01223593
haf-2	WBGene00001812	F43E2.4	0.48484848	0.71969697	0.60227273
ham-3	WBGene00044072	ZK1128.5	0.99206349	0.40196078	0.69701214
hcp-3	WBGene00001831	F58A4.3	1.00217865	1.15407407	1.07812636
hil-4	WBGene00001855	C18G1.5	0.56565657	0.71459695	0.64012676
hil-5	WBGene00001856	B0414.3	0.89743590	0.54666667	0.72205128
hmg-3	WBGene00001973	C32F10.5	1.08747045	0.59941521	0.84344283
hpo-21	WBGene00012550	Y37D8A.10	0.10526316	0.07189542	0.08857929
hsr-9	WBGene00002027	T05F1.6	1.41414141	0.59549746	1.00481944
K04C2.2	WBGene00019380	K04C2.2	0.04115226	0.00000000	0.02057613
K07A12.5	WBGene00010623	K07A12.5	1.23456790	0.99358974	1.11407882
K09E4.1	WBGene00010719	K09E4.1	1.69934641	1.56695157	1.63314899
kin-3	WBGene00002191	B0205.7	0.58823529	0.00000000	0.29411765
let-60	WBGene00002335	ZK792.6	0.29629630	0.00000000	0.14814815
let-716	WBGene00002850	C16A3.3	1.35265700	0.56561086	0.95913393
let-99	WBGene00002368	K08E7.3	0.66666667	0.42424242	0.54545455
lin-23	WBGene00003009	K10B2.1	0.25000000	0.23965142	0.24482571
lsm-4	WBGene00003078	F32A5.7	0.88888889	0.20493066	0.54690978

mog-4	WBGene00003392	C04H5.6	0.00000000	0.00000000	0.00000000
mog-5	WBGene00003393	EEED8.5	1.24528302	0.23529412	0.74028857
mrpl-24	WBGene00019076	F59A3.3	0.68027211	0.52199074	0.60113142
msp-50	WBGene00003443	C34F11.4	0.71428571	1.22041420	0.96734996
nol-1	WBGene00021073	W07E6.1	1.38271605	0.96153846	1.17212726
pafo-1	WBGene00008338	C55A6.9	0.41025641	0.81972265	0.61498953
pbs-7	WBGene00003953	F39H11.5	0.00000000	0.00000000	0.00000000
perm-2	WBGene00016636	C44B12.1	0.97435897	0.74025974	0.85730936
perm-4	WBGene00016638	C44B12.5	0.28368794	0.56583072	0.42475933
pfd-2	WBGene00019220	H20J04.5	1.96551724	0.90588235	1.43569980
pfd-3	WBGene00006889	T06G6.9	0.77777778	0.62539185	0.70158481
R03D7.2	WBGene00010989	R03D7.2	0.00000000	0.91564928	0.45782464
R03G8.4	WBGene00010999	R03G8.4	0.30769231	0.45454545	0.38111888
repo-1	WBGene00008683	F11A10.2	0.49382716	0.00000000	0.24691358
retr-1	WBGene00018416	F44E2.2	0.72222222	0.89869281	0.81045752
ribo-1	WBGene00020683	T22D1.4	0.09523810	0.05956113	0.07739961
rnp-2	WBGene00004385	K08D10.4	1.47126437	0.05956113	0.76541275
rpac-40	WBGene00019275	H43I07.2	0.88888889	0.71180556	0.80034722
rpb-5	WBGene00019246	H27M09.2	1.86000000	1.83333333	1.84666667
rpb-6	WBGene00007355	C06A1.5	0.00000000	0.62809917	0.31404959
rpb-8	WBGene00017830	F26F4.11	1.43089431	1.11952862	1.27521146
rpc-11	WBGene00022309	Y77E11A.6	0.84051724	0.80000000	0.82025862
rpoa-2	WBGene00008781	F14B4.3	0.90534979	1.02564103	0.96549541
rsr-1	WBGene00004706	F28D9.1	0.14035088	0.32590051	0.23312570
rtfo-1	WBGene00009103	F25B3.6	1.20192308	0.46723647	0.83457977
ruvb-1	WBGene00007784	C27H6.2	0.57657658	0.22813036	0.40235347
skpo-1	WBGene00009897	F49E12.1	1.12592593	0.31404959	0.71998776
skr-17	WBGene00004823	C06A8.4	1.20000000	0.52435065	0.86217532
sld-2	WBGene00020466	T12F5.1	0.36363636	0.27094474	0.31729055
smo-1	WBGene00004888	K12C11.2	0.00000000	0.06902357	0.03451178
spt-5	WBGene00005015	K08E4.1	0.00000000	0.00000000	0.00000000
stc-1	WBGene00006059	F54C9.2	1.94444444	1.55773420	1.75108932
sti-1	WBGene00019983	R09E12.3	0.81250000	1.31808279	1.06529139
sun-1	WBGene00006311	F57B1.2	0.50000000	0.47979798	0.48989899

T01C3.7.1	WBGene00001423	T01C3.7.1	0.39130435	0.39454806	0.39292621
T06E6.1	WBGene00011538	T06E6.1	1.75925926	1.15384615	1.45655271
T14B4.2	WBGene00020499	T14B4.2	0.65359477	0.66191833	0.65775655
T20B12.3	WBGene00020601	T20B12.3	0.66666667	0.68578256	0.67622461
T22D1.10.1	WBGene00020687	T22D1.10.1	1.00000000	2.00000000	1.50000000
T24H10.1	WBGene00012000	T24H10.1	0.95238095	0.68439108	0.81838602
T26G10.1	WBGene00012059	T26G10.1	1.64874552	1.13888889	1.39381720
tba-4	WBGene00006530	F44F4.11	0.61538462	0.09965035	0.35751748
toe-1	WBGene00022739	ZK430.1	0.42042042	0.72888889	0.57465465
tofu-5	WBGene00012167	W01A8.5	0.64000000	0.11912226	0.37956113
top-2	WBGene00010785	K12D12.1	1.13378685	0.96172839	1.04775762
try-1	WBGene00006619	ZK546.15	0.78431373	0.74025974	0.76228673
uba-2	WBGene00006700	W02A11.4	1.49494949	1.02602603	1.26048776
ubc-12	WBGene00006707	R09B3.4	0.23188406	0.48885077	0.36036741
ubh-4	WBGene00006724	C08B11.7	0.66666667	0.67489712	0.67078189
ulp-1	WBGene00006736	T10F2.3	0.81871345	0.11909949	0.46890647
ulp-2	WBGene00006737	Y38A8.3	0.00000000	0.10123457	0.05061728
ulp-4	WBGene00006739	C41C4.6	0.00000000	0.06902357	0.03451178
unc-85	WBGene00006817	F10G7.3	0.64000000	0.79966330	0.71983165
vit-5	WBGene00006929	C04F6.1	1.02564103	0.92319749	0.97441926
W06E11.1	WBGene00021061	W06E11.1	0.98765432	1.07954545	1.03359989
Y23H5B.5	WBGene00021276	Y23H5B.5	0.28846154	0.00000000	0.14423077
Y48B6A.1	WBGene00012978	Y48B6A.1	0.72222222	1.43790850	1.08006536
Y53C12B.1	WBGene00013143	Y53C12B.1	1.11111111	1.31578947	1.21345029
Y62H9A.5	WBGene00013393	Y62H9A.5	0.28301887	1.29411765	0.78856826
Y62H9A.6	WBGene00013394	Y62H9A.6	1.06122449	1.50608519	1.28365484
Y65B4A.6	WBGene00022029	Y65B4A.6	0.63461538	0.64705882	0.64083710

Y75B8A.7	WBGene00013544	Y75B8A.7	1.46572104	1.33547009	1.40059556
yars-1	WBGene00013677	Y105E8A.19	0.96428571	0.85834334	0.91131453
ZK686.2	WBGene00022792	ZK686.2	1.51300236	1.19230769	1.35265503
ZK792.5	WBGene00014078	ZK792.5	0.75555556	0.50617284	0.63086420
B0391.11	WBGene00007166	B0391.11	0.85849057	0.79463364	0.82656210
nhr-150	WBGene00007367	C06B8.1	0.87500000	1.03773585	0.95636792
E02C12.11	WBGene00017096	E02C12.11	0.55519481	1.44432773	0.99976127
F02E9.1	WBGene00008528	F02E9.1	0.00000000	1.35294118	0.67647059
str-112	WBGene00006163	F10D2.4	0.84375000	0.88235294	0.86305147
F11F1.4	WBGene00008716	F11F1.4	0.65217391	0.83877996	0.74547693
F11F1.5	WBGene00008717	F11F1.5	1.01694915	0.41301627	0.71498271
sup-9	WBGene00006318	F34D6.3	0.00000000	0.63437140	0.31718570
nhr-111	WBGene00003701	F44G3.9	1.53333333	1.70588235	1.61960784
F47B8.10	WBGene00009811	F47B8.10	0.30508475	1.61764706	0.96136590
F53C3.5	WBGene00018749	F53C3.5	0.00000000	0.69327731	0.34663866
F57F4.2	WBGene00019016	F57F4.2	0.43636364	1.52941176	0.98288770
F59E12.8	WBGene00019123	F59E12.8	0.00000000	1.69117647	0.84558824
cyp-25A6	WBGene00019438	K06B9.1	1.40000000	0.85834334	1.12917167
sup-10	WBGene00006319	R09G11.1	0.00000000	0.43333333	0.21666667
nhr-271	WBGene00011396	T03E6.3	1.87500000	1.27100840	1.57300420
T15D6.11	WBGene00011785	T15D6.11	0.63725490	1.47058824	1.05392157
col-164	WBGene00000737	T21D9.1	1.01694915	1.07843137	1.04769026
srh-60	WBGene00005282	W10G11.9	0.00000000	0.83877996	0.41938998
ZK675.4	WBGene00014067	ZK675.4	1.85000000	1.24019608	1.54509804

A.2. Transcription factors activity with lifespan change of *C.elegans* mutants

TF lifestage	slope	R	R.squared	adj.R.squared	p-value
JUN1 L3	5.9786	0.9315	0.8677	0.8545	0
EOR1 L3	7.0998	0.9278	0.8608	0.8469	0
PHA4 YA	6.0508	0.8808	0.7759	0.7535	2e-04
ZTF4 L2	6.9497	0.8801	0.7745	0.752	2e-04
NHR77 L4	6.1603	0.8668	0.7514	0.7265	3e-04
PHA4 L4	7.3594	0.8564	0.7334	0.7067	4e-04
SKN1 L3	6.1317	0.8509	0.7241	0.6965	4e-04
NHR6 L4	7.1038	0.8502	0.7228	0.695	5e-04
DAF16 L4 YA	6.1016	0.8426	0.71	0.681	6e-04
NHR10 L4	8.9626	0.8419	0.7088	0.6797	6e-04
HPL2 L1	7.7324	0.837	0.7006	0.6707	7e-04
ZTF7 L4	7.4764	0.8156	0.6652	0.6317	0.0012
MEF2 L1	4.3968	0.8114	0.6583	0.6241	0.0014
LIN13 L2	6.9992	0.8107	0.6573	0.623	0.0014
PHA4 L3	5.7327	0.7975	0.636	0.5996	0.0019
DVE1 L4	5.832	0.7873	0.6198	0.5818	0.0024
HLH30 L4	5.6059	0.7856	0.6172	0.5789	0.0025
SAX3 L2	4.1474	0.7784	0.6059	0.5665	0.0029
LIN35 L1	5.8416	0.7684	0.5904	0.5495	0.0035
C16A3.4 L1	3.4777	0.7649	0.585	0.5435	0.0038
HAM1 L1	6.046	0.7606	0.5785	0.5363	0.0041
SAX3 L3	3.8295	0.7605	0.5783	0.5362	0.0041
JUN1 L1	4.4556	0.7512	0.5643	0.5207	0.0049
LIN39 L1	5.0444	0.7482	0.5599	0.5158	0.0051
CEH38 L3	5.8547	0.7379	0.5445	0.499	0.0061
NFYA1 YA	5.2135	0.7294	0.5321	0.4853	0.0071
EFL1 L1	4.6456	0.7252	0.5259	0.4785	0.0076
NHR28 L4	5.4556	0.7214	0.5204	0.4725	0.0081
LSY2 L1	6.2148	0.7163	0.5131	0.4645	0.0088
AMA1 L4 YA	6.4729	0.7013	0.4918	0.4409	0.0111
AMA1 L3	5.4653	0.6999	0.4899	0.4388	0.0113

C01B12.2 L2	6.8133	0.6982	0.4875	0.4362	0.0116
AMA1 L1	5.3647	0.6913	0.4779	0.4257	0.0128
W03F9.2 L4 YA	6.1714	0.685	0.4693	0.4162	0.014
HAM1 L4	6.3874	0.685	0.4693	0.4162	0.014
PHA4 L2	3.5904	0.6798	0.4622	0.4084	0.015
NFYA1 L3	5.318	0.6772	0.4586	0.4045	0.0156
F16B12.6 L1	3.3271	0.6755	0.4564	0.402	0.0159
UNC62 L3	3.5894	0.6718	0.4513	0.3964	0.0167
NHR129 L2	4.4132	0.6661	0.4437	0.3881	0.018
C34F6.9 L2	4.9889	0.6576	0.4325	0.3757	0.0201
ZAG1 L2	2.7599	0.6568	0.4314	0.3745	0.0203
NHR25 L2	4.1717	0.6485	0.4206	0.3626	0.0225
SAX3 L4	4.3348	0.6428	0.4132	0.3546	0.0242
NHR21 L2	4.4415	0.6427	0.413	0.3543	0.0242
EGL5 L3	4.8183	0.6274	0.3936	0.3329	0.029
PHA4 L1	3.3668	0.6254	0.3911	0.3302	0.0296
ALR1 L2	3.7507	0.624	0.3894	0.3284	0.0301
NHR11 L4	3.2438	0.6149	0.3781	0.3159	0.0333
ZIP2 L4	3.9553	0.6143	0.3773	0.3151	0.0336
SEM4 L2	4.4426	0.6125	0.3751	0.3126	0.0343
DAF16 POLII L4 YA	4.5055	0.6076	0.3692	0.3061	0.0361
MAB5 L3	3.8769	0.6064	0.3677	0.3044	0.0366
UNC55 L2	4.2407	0.5967	0.3561	0.2917	0.0405
AHA1 L4	3.771	0.5734	0.3288	0.2617	0.0513
UNC62 L1	2.2689	0.5717	0.3268	0.2595	0.0522
DPL1 L1	4.1824	0.5602	0.3138	0.2451	0.0582
MAB5 L2	5.0452	0.5578	0.3111	0.2422	0.0595
FOS1 L1	3.4424	0.5499	0.3023	0.2326	0.064
AHA1 L1	3.0301	0.537	0.2883	0.2172	0.0718
GEI11 L1	2.1037	0.5238	0.2744	0.2018	0.0805
NHR23 L3	5.2779	0.515	0.2653	0.1918	0.0866
JUN1 L4	5.5507	0.5107	0.2608	0.1869	0.0898
ZTF4 L3	6.0461	0.5064	0.2565	0.1821	0.0929
GEI11 L3	2.8632	0.4895	0.2396	0.1636	0.1062

ELT1 L3	3.4684	0.4882	0.2384	0.1622	0.1073
FOS1 L2	6.1303	0.4832	0.2335	0.1568	0.1115
NHR77 L1	2.4258	0.482	0.2323	0.1556	0.1126
ZTF4 L1	3.2711	0.4701	0.221	0.1431	0.1231
UNC62 YA	12.0665	0.4639	0.2152	0.1367	0.1287
MML1 L3	1.4994	0.4432	0.1964	0.116	0.1491
NHR6 L2	3.2289	0.4395	0.1931	0.1124	0.1529
LSY2 L4	4.3551	0.4244	0.1801	0.0981	0.1691
R02D3.7 L3	2.719	0.4104	0.1684	0.0853	0.1851
F45C12.2 L1	2.9081	0.3988	0.159	0.0749	0.1991
DPL1 L4	3.7508	0.3984	0.1587	0.0746	0.1996
LIN13 L1	0.7913	0.3888	0.1512	0.0663	0.2116
FKH2 L3	1.2252	0.387	0.1498	0.0647	0.2139
GEI11 YA	1.9887	0.3318	0.1101	0.0211	0.292
DAF12 L4	0.4463	0.3036	0.0922	0.0014	0.3373
RPC1 YA	2.0617	0.3009	0.0905	-4e-04	0.3419
F23B12.7 YA	1.6305	0.2916	0.085	-0.0065	0.3579
PES1 L4	1.5636	0.2726	0.0743	-0.0182	0.3913
F45C12.2 L3	1.5458	0.2565	0.0658	-0.0276	0.4209
EFL1 YA	1.3054	0.254	0.0645	-0.029	0.4257
NHR76 L3	1.2884	0.2472	0.0611	-0.0328	0.4386
NHR23 L2	0.7818	0.2192	0.048	-0.0472	0.4937
NHR11 L1	1.6733	0.2139	0.0458	-0.0497	0.5043
NHR28 L1	0.4565	0.164	0.0269	-0.0704	0.6105
CEH16 L2	0.665	0.1538	0.0237	-0.074	0.6332
LIN13 L4	1.5187	0.1537	0.0236	-0.074	0.6334
GEI11 L2	0.5645	0.1473	0.0217	-0.0761	0.6478
NHR11 L3	0.1959	0.1412	0.0199	-0.0781	0.6616
CES1 L3	0.7468	0.0954	0.0091	-0.09	0.7681
NHR77 L2	0.3021	0.0616	0.0038	-0.0958	0.8491
NHR67 L3	0.1816	0.0607	0.0037	-0.0959	0.8514
CEH38 L4	0.2352	0.06	0.0036	-0.096	0.853
FOS1 L4	0.2428	0.0581	0.0034	-0.0963	0.8577
UNC62 L2	0.0377	0.013	2e-04	-0.0998	0.9681
NHR77 L3	-0.0041	8e-04	0	-0.1	0.998

LSY2 L2	-0.0422	0.0057	0	-0.1	0.9859
NHR76 L4	-0.0547	0.0119	1e-04	-0.0998	0.9707
ELT1 L2	-0.1783	0.0368	0.0014	-0.0985	0.9095
F45C12.2 L2	-0.3046	0.0453	0.002	-0.0977	0.8889
CES1 L4	-0.1333	0.0461	0.0021	-0.0977	0.887
NHR12 L2	-0.2933	0.0532	0.0028	-0.0969	0.8695
LIN35 YA	-0.1585	0.0723	0.0052	-0.0942	0.8232
DPL1 YA	-0.2138	0.0897	0.008	-0.0911	0.7816
HPL2 YA	-0.4248	0.114	0.013	-0.0857	0.7242
R02D3.7 L2	-0.2507	0.1662	0.0276	-0.0696	0.6056
SKN1 L4	-0.4672	0.1991	0.0396	-0.0564	0.535
CES1 L1	-0.9528	0.2012	0.0405	-0.0555	0.5307
NHR11 L2	-0.6167	0.202	0.0408	-0.0551	0.5289
FOS1 L3	-1.698	0.2462	0.0606	-0.0333	0.4405
LIN15B L4	-0.5224	0.2483	0.0617	-0.0322	0.4364
ALY2 L3	-1.3871	0.2697	0.0727	-0.02	0.3967
NHR237 L2	-0.7414	0.2779	0.0772	-0.0151	0.3819
FKH10 L4	-0.8457	0.2809	0.0789	-0.0132	0.3764
ZAG1 L3	-1.1511	0.2882	0.083	-0.0086	0.3637
ZAG1 L4	-0.7622	0.3175	0.1008	0.0109	0.3147
ALY2 L1	-2.5342	0.3293	0.1085	0.0193	0.2959
NHR237 L1	-0.7128	0.3366	0.1133	0.0247	0.2847
ZAG1 L1	-1.053	0.4477	0.2004	0.1205	0.1444
NHR28 L3	-0.793	0.4755	0.2261	0.1487	0.1182
R02D3.7 L4	-2.5654	0.4886	0.2387	0.1626	0.107
TLP1 L1	-2.8204	0.4979	0.2479	0.1727	0.0995

A.3. Software Versions

- Bowtie 2.2.9
- TopHat 2.1.1
- macs2 2.1.1.2
- bedtools 2.26.0
- samtools 1.3.1
- fastQC 0.11.5
- R 3.3.2
- TFTargetCaller 0.7
- biomaRt 2.30.0
- BioNetSmooth 1.0.0
- DESeq2 1.14.1

A.4. Supplements

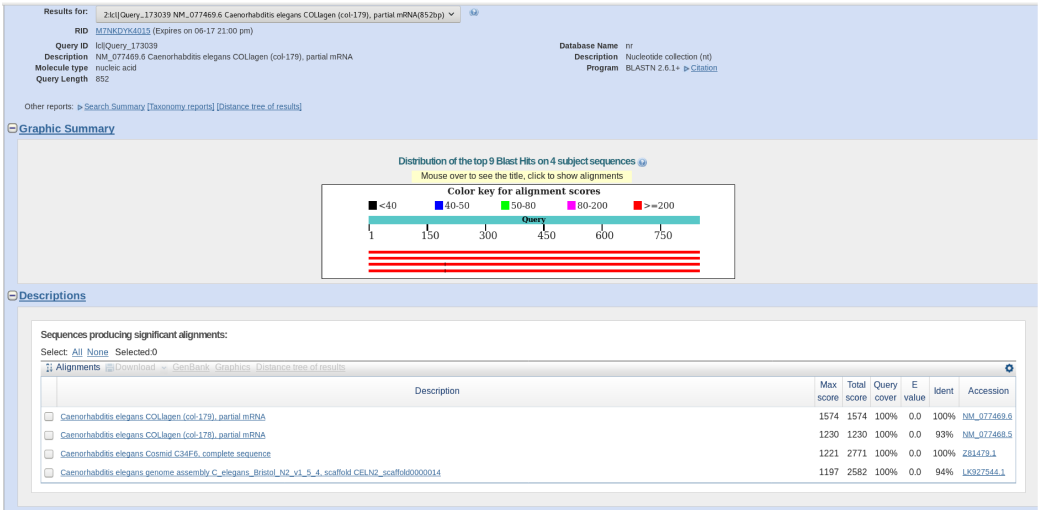

Figure 29: Col-179 nucleotide sequence identity with col-178.

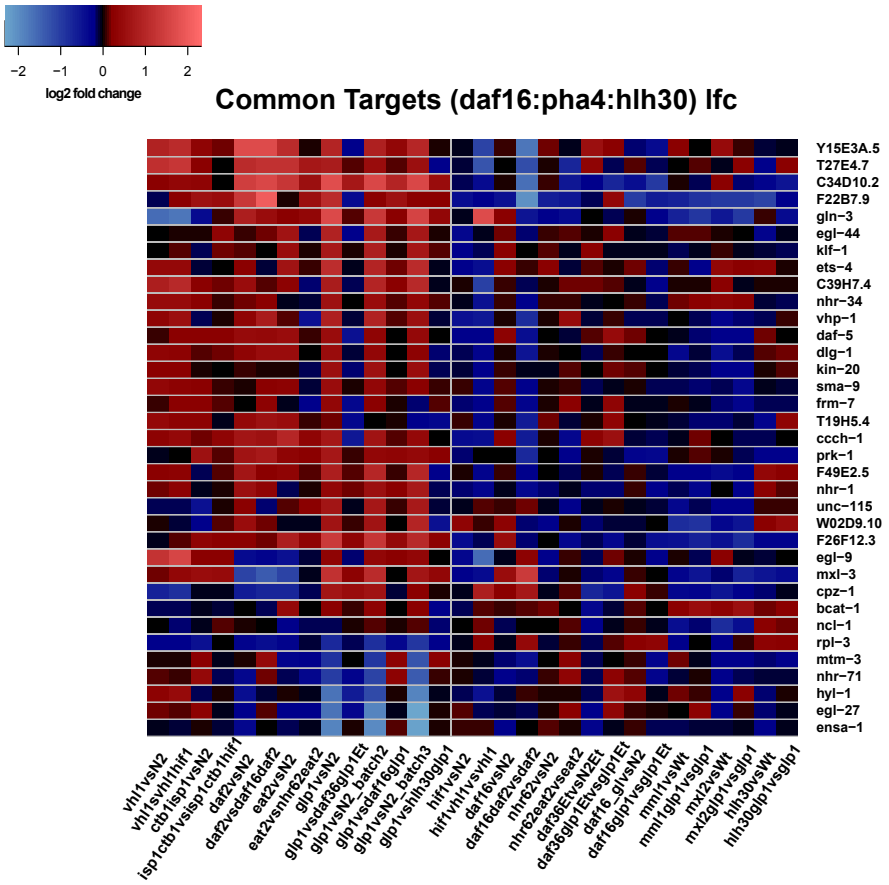

Figure 30: Common target genes of transcription factors DAF-16,PHA-4 and HLH-30 log2 fold changes in the lifespan increasing and decreasing mutants of *C.elegans*.

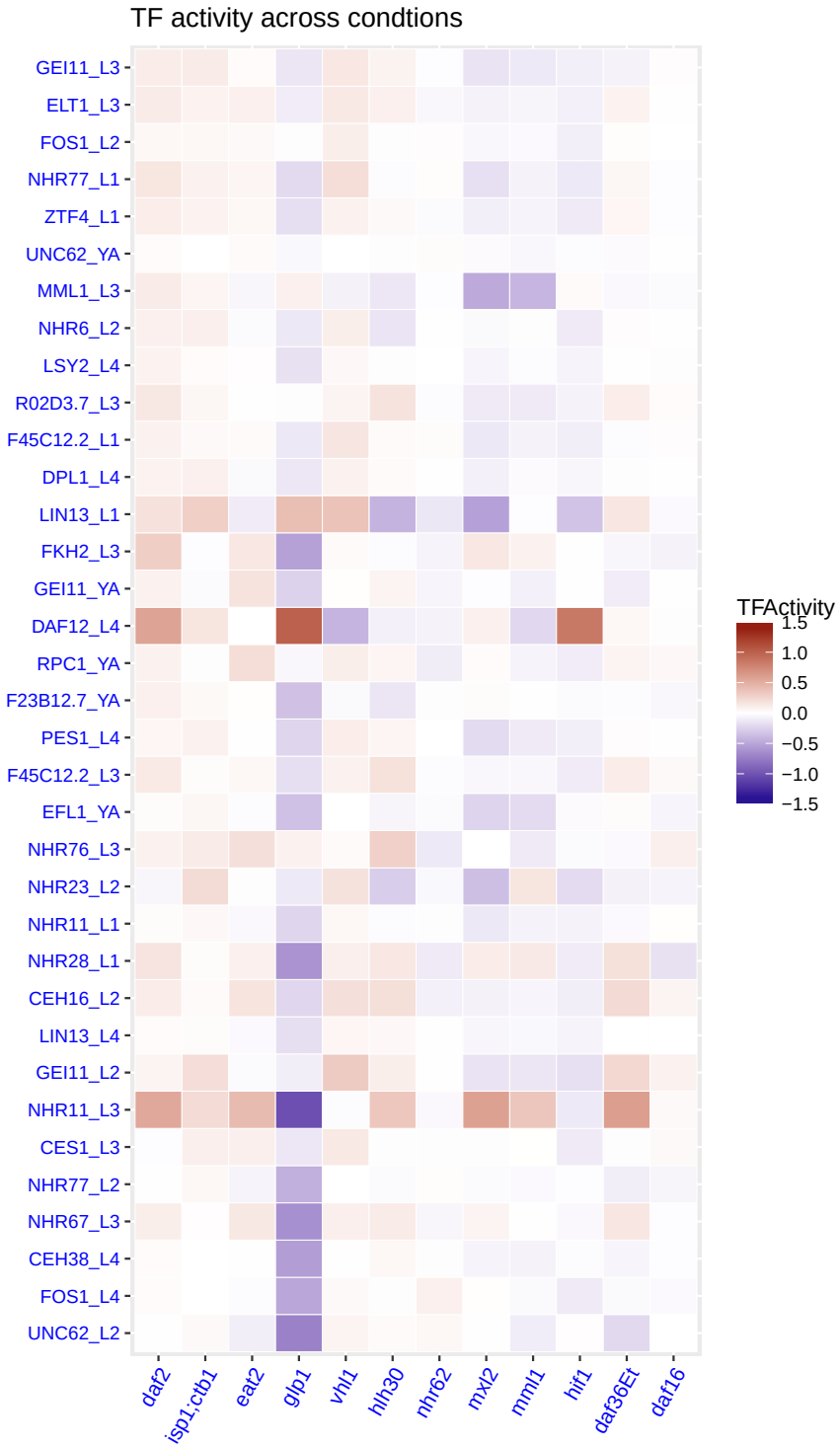

Figure 31: Transcription factors that are moderately positively associated with lifespan change in *C.elegans*

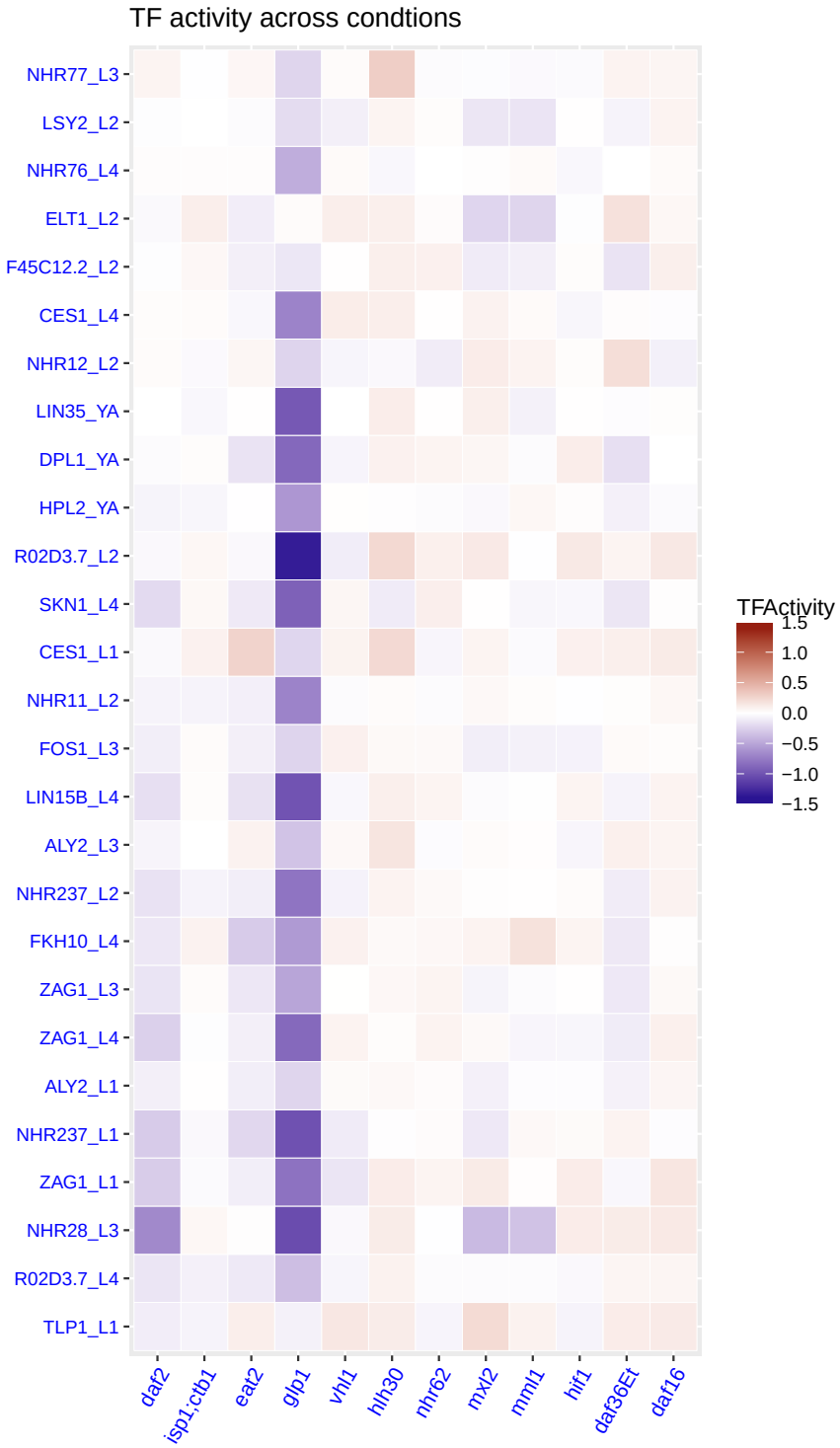

Figure 32: Transcription factors that negatively associated with lifespan change in *C.elegans*

References

[1] Adrian Alexa, JÃűrg RahnenfÃijhrer, and Thomas Lengauer. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics (Oxford, England)*, 22(13):1600–1607, July 2006.

[2] N. Alic, T. D. Andrews, M. E. Giannakou, I. Papatheodorou, C. Slack, M. P. Hoddinott, H. M. Cocheme, E. F. Schuster, J. M. Thornton, and L. Partridge. Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. *Molecular Systems Biology*, 7(1):502–502, April 2014.

[3] Nazif Alic, T. Daniel Andrews, Maria E. Giannakou, Irene Papatheodorou, Cathy Slack, Matthew P. Hoddinott, Helena M. CochemÃĺ, Eugene F. Schuster, Janet M. Thornton, and Linda Partridge. Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. *Molecular Systems Biology*, 7:502, June 2011.

[4] Rozalyn M. Anderson, Dhanansayan Shanmuganayagam, and Richard Weindruch. Caloric restriction and aging: studies in mice and monkeys. *Toxicologic Pathology*, 37(1):47–51, January 2009.

[5] Adam Antebi. Regulation of longevity by the reproductive system. *Experimental Gerontology*, 48(7):596–602, July 2013.

[6] Hua Bai, Ping Kang, Ana Maria Hernandez, and Marc Tatar. Activin signaling targeted by insulin/dFOXO regulates aging and muscle proteostasis in Drosophila. *PLoS genetics*, 9(11):e1003941, November 2013.

[7] Bai Hua, Kang Ping, and Tatar Marc. Drosophila insulinâĂŘlike peptideâĂŘ6 (dilp6) expression from fat body extends lifespan and represses secretion of Drosophila insulinâĂŘlike peptideâĂŘ2 from the brain. *Aging Cell*, 11(6):978–985, September 2012.

[8] Albert-LÃąszlÃş BarabÃąsi, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, January 2011.

[9] Nicholas A. Bishop and Leonard Guarente. Two neurons mediate diet-restriction-induced longevity in *C. elegans*. *Nature*, 447(7144):545–549, May 2007.

[10] Tammie Bishop, Kah Weng Lau, Andrew C. R. Epstein, Stuart K. Kim, Min Jiang, Delia O'Rourke, Christopher W. Pugh, Jonathan M. Gleadle, Martin S. Taylor, Jonathan Hodgkin, and Peter J. Ratcliffe. Genetic Analysis of Pathways Regulated by the von Hippel-Lindau Tumor Suppressor in Caenorhabditis elegans. *PLOS Biology*, 2(10):e289, September 2004.

[11] T. Keith Blackwell, Michael J. Steinbaugh, John M. Hourihan, Collin Y. Ewald, and Meltem Isik. SKN-1/Nrf, stress responses, and aging in Caenorhabditis elegans. *Free Radical Biology & Medicine*, 88(Pt B):290–301, November 2015.

[12] Matthias BlÃijher, Barbara B. Kahn, and C. Ronald Kahn. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science (New York, N.Y.)*, 299(5606):572–574, January 2003.

[13] S.J. Broughton, C. Slack, N. Alic, A. Metaxakis, T. M. Bass, Y. Driege, and L. Partridge. DILP-producing Median Neurosecretory Cells in the Drosophila Brain Mediate the Response of Lifespan to Dietary Restriction. *Aging cell*, 9(3):336–346, June 2010.

[14] Susan Broughton and Linda Partridge. Insulin/IGF-like signalling, the central nervous system and aging. *Biochemical Journal*, 418(1):1–12, February 2009.

[15] Daniel E. Carlin, Barry Demchak, Dexter Pratt, Eric Sage, and Trey Ideker. Network propagation in the cytoscape cyberinfrastructure. *PLOS Computational Biology*, 13(10):e1005598, October 2017.

[16] R. W. Carthew. Gene silencing by double-stranded RNA. *Current Opinion in Cell Biology*, 13(2):244–248, April 2001.

[17] Di Chen, Emma Lynn Thomas, and Pankaj Kapahi. HIF-1 Modulates Dietary Restriction-Mediated Lifespan Extension via IRE-1 in Caenorhabditis elegans. *PLOS Genetics*, 5(5):e1000486, May 2009.

[18] Lenore Cowen, Trey Ideker, Benjamin J. Raphael, and Roded Sharan. Network propagation: a universal amplifier of genetic associations. *Nature Reviews Genetics*, 18(9):551–562, September 2017.

[19] Fabio Demontis and Norbert Perrimon. FOXO/4e-BP signaling in Drosophila muscles regulates organism-wide proteostasis during aging. *Cell*, 143(5):813–825, November 2010.

[20] David van Dijk, Juozas Nainys, Roshan Sharma, Pooja Kathail, Ambrose J. Carr, Kevin R. Moon, Linas Mazutis, Guy Wolf, Smita Krishnaswamy, and Dana Pe'er. MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. *bioRxiv*, page 111591, February 2017.

[21] Steffen Durinck, Yves Moreau, Arek Kasprzyk, Sean Davis, Bart De Moor, Alvis Brazma, and Wolfgang Huber. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16):3439–3440, August 2005.

[22] Paul Essers, Luke S. Tain, Tobias Nespital, Joana Goncalves, Jenny Froehlich, and Linda Partridge. Reduced insulin/insulin-like growth factor signaling decreases translation in *Drosophila* and mice. *Scientific Reports*, 6:30290, July 2016.

[23] Collin Y. Ewald, Jess N. Landis, Jess Porter Abate, Coleen T. Murphy, and T. Keith Blackwell. Dauer-independent insulin/IGF-1-signalling implicates collagen remodelling in longevity. *Nature*, 519(7541):97–101, March 2015.

[24] Danny Filer, Maximillian A. Thompson, Vakil Takhaveev, Adam J. Dobson, Ilektra Kotronaki, James W. M. Green, Matthias Heinemann, Jennifer M. A. Tullet, and Nazif Alic. RNA polymerase III limits longevity downstream of TORC1. *Nature*, 552(7684):263–267, December 2017.

[25] Luigi Fontana, Linda Partridge, and Valter D. Longo. Extending Healthy Life SpanâĂŤFrom Yeast to Humans. *Science*, 328(5976):321–326, April 2010.

[26] Lude Franke, Harm van Bakel, Like Fokkens, Edwin D. deÂăJong, Michael Egmont-Petersen, and Cisca Wijmenga. Reconstruction of a Functional

Human Gene Network, with an Application for Prioritizing Positional Candidate Genes. *American Journal of Human Genetics*, 78(6):1011–1025, June 2006.

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

[28] David Gems and Linda Partridge. Genetics of longevity in model organisms: debates and paradigm shifts. *Annual Review of Physiology*, 75:621–644, 2013.

[29] Maria E. Giannakou, Martin Goss, Martin A. JÃijnger, Ernst Hafen, Sally J. Leevers, and Linda Partridge. Long-lived Drosophila with overexpressed dFOXO in adult fat body. *Science (New York, N.Y.)*, 305(5682):361, July 2004.

[30] Jesse Gillis and Paul Pavlidis. âĂIJGuilt by AssociationâĂİ Is the Exception Rather Than the Rule in Gene Networks. *PLOS Computational Biology*, 8(3):e1002444, March 2012.

[31] JÃľrÃťme Goudeau, StÃľphanie Bellemin, Esther Toselli-Mollereau, Mehrnaz Shamalnasab, Yiqun Chen, and Hugo Aguilaniu. Fatty Acid Desaturation Links Germ Cell Loss to Longevity Through NHR-80/HNF4 in C. elegans. *PLOS Biology*, 9(3):e1000599, March 2011.

[32] Sebastian GrÃűnke, David-Francis Clarke, Susan Broughton, T. Daniel Andrews, and Linda Partridge. Molecular evolution and functional characterization of Drosophila insulin-like peptides. *PLoS genetics*, 6(2):e1000857, February 2010.

[33] Julius Halaschek-Wiener, Jaswinder S. Khattra, Sheldon McKay, Anatoli Pouzyrev, Jeff M. Stott, George S. Yang, Robert A. Holt, Steven J. M. Jones, Marco A. Marra, Angela R. Brooks-Wilson, and Donald L. Riddle. Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. *Genome Research*, 15(5):603–615, May 2005.

[34] Malene Hansen, Stefan Taubert, Douglas Crawford, Nataliya Libina, Seung-Jae Lee, and Cynthia Kenyon. Lifespan extension by conditions that inhibit translation in Caenorhabditis elegans. *Aging Cell*, 6(1):95–110, February 2007.

[35] Bree N. Heestand, Yidong Shen, Wei Liu, Daniel B. Magner, Nadia Storm, Caroline Meharg, Bianca Habermann, and Adam Antebi. Dietary Restriction Induced Longevity Is Mediated by Nuclear Receptor NHR-62 in Caenorhabditis elegans. *PLOS Genetics*, 9(7):e1003651, July 2013.

[36] Caroline Heintz, Thomas K. Doktor, Anne Lanjuin, Caroline C. Escoubas, Yue Zhang, Heather J. Weir, Sneha Dutta, Carlos Giovanni Silva-GarcÃŋa, Gitte H. Bruun, Ianessa Morantte, Gerta Hoxhaj, Brendan D. Manning, Brage S. Andresen, and William B. Mair. Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in *C. elegans*. *Nature*, 541(7635):102–106, January 2017.

[37] Matan Hofree, John P. Shen, Hannah Carter, Andrew Gross, and Trey Ideker. Network-based stratification of tumor mutations. *Nature Methods*, 10(11):1108–1115, November 2013.

[38] Honor Hsin and Cynthia Kenyon. Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature*, 399(6734):362–366, May 1999.

[39] Tomoatsu Ikeya, Susan Broughton, Nazif Alic, Richard Grandison, and Linda Partridge. The endosymbiont Wolbachia increases insulin/IGF-like signalling in Drosophila. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674):3799–3807, November 2009.

[40] Brian K. Kennedy and Dudley W. Lamming. The Mechanistic Target of Rapamycin: The Grand ConducTOR of Metabolism and Aging. *Cell Metabolism*, 23(6):990–1003, June 2016.

[41] C. Kenyon, J. Chang, E. Gensch, A. Rudner, and R. Tabtiang. A C. elegans mutant that lives twice as long as wild type. *Nature*, 366(6454):461–464, December 1993.

[42] Cynthia J. Kenyon. The genetics of ageing, March 2010.

[43] Sebastian KÃűhler, Sebastian Bauer, Denise Horn, and Peter N. Robinson. Walking the Interactome for Prioritization of Candidate Disease Genes. *The American Journal of Human Genetics*, 82(4):949–958, April 2008.

[44] Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. TopHat2: accurate alignment of transcriptomes

in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14:R36, April 2013.

[45] K. D. Kimura, H. A. Tissenbaum, Y. Liu, and G. Ruvkun. daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science (New York, N.Y.)*, 277(5328):942–946, August 1997.

[46] M. R. Klass. A method for the isolation of longevity mutants in the nematode Caenorhabditis elegans and initial results. *Mechanisms of Ageing and Development*, 22(3-4):279–286, August 1983.

[47] Hiroshi Koga, Susmita Kaushik, and Ana Maria Cuervo. Protein homeostasis and aging: The importance of exquisite quality control. *Ageing Research Reviews*, 10(2):205–215, April 2011.

[48] Neeraj Kumar, Vaibhav Jain, Anupama Singh, Urmila Jagtap, Sonia Verma, and Arnab Mukhopadhyay. Genome-wide endogenous DAF-16/FOXO recruitment dynamics during lowered insulin signalling in C. elegans. *Oncotarget*, 6(39):41418–41433, November 2015.

[49] C. H. Lai, C. Y. Chou, L. Y. Ch'ang, C. S. Liu, and W. Lin. Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. *Genome Research*, 10(5):703–713, May 2000.

[50] Bernard Lakowski and Siegfried Hekimi. The genetics of caloric restriction in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 95(22):13091–13096, October 1998.

[51] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, April 2012.

[52] Louis R. Lapierre, Sara Gelino, Alicia MelÃĺndez, and Malene Hansen. Autophagy and lipid metabolism coordinately modulate life span in germline-less C. elegans. *Current biology: CB*, 21(18):1507–1514, September 2011.

[53] Gang Liu, Jason Rogers, Coleen T Murphy, and Christopher Rongo. EGF signalling activates the ubiquitin proteasome system to modulate C. elegans lifespan. *The EMBO Journal*, 30(15):2990–3003, August 2011.

[54] Yansheng Liu, Andreas Beyer, and Ruedi Aebersold. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, 165(3):535–550, April 2016.

[55] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014.

[56] Carlos LÃşpez-OtÃŋn, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The Hallmarks of Aging. *Cell*, 153(6):1194–1217, June 2013.

[57] Marie-Therese Mackmull, Bernd Klaus, Ivonne Heinze, Manopriya Chokkalingam, Andreas Beyer, Robert B. Russell, Alessandro Ori, and Martin Beck. Landscape of nuclear transport receptor cargo specificity. *Molecular Systems Biology*, 13(12):962, December 2017.

[58] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12, May 2011.

[59] Joshua J. McElwee, Eugene Schuster, Eric Blanc, Matthew D. Piper, James H. Thomas, Dhaval S. Patel, Colin Selman, Dominic J. Withers, Janet M. Thornton, Linda Partridge, and David Gems. Evolutionary conservation of regulated longevity assurance mechanisms. *Genome Biology*, 8(7):R132, 2007.

[60] Ranjana Mehta, Katherine A. Steinkraus, George L. Sutphin, Fresnida J. Ramos, Lara S. Shamieh, Alexander Huh, Christina Davis, Devon Chandler-Brown, and Matt Kaeberlein. Proteasomal regulation of the hypoxic response modulates aging in C. elegans. *Science (New York, N.Y.)*, 324(5931):1196–1198, May 2009.

[61] JÃűrg Menche, Amitabh Sharma, Maksim Kitsak, Susan Ghiassian, Marc Vidal, Joseph Loscalzo, and Albert-LÃąszlÃş BarabÃąsi. Uncovering disease-disease relationships through the incomplete human interactome. *Science (New York, N.Y.)*, 347(6224):1257601, February 2015.

[62] Athanasios Metaxakis, Luke S. Tain, Sebastian GrÃűnke, Oliver Hendrich, Yvonne Hinze, Ulrike Birras, and Linda Partridge. Lowered Insulin

Signalling Ameliorates Age-Related Sleep Fragmentation in Drosophila. *PLOS Biology*, 12(4):e1001824, April 2014.

[63] Kyung-Tai Min and Seymour Benzer. Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. *Proceedings of the National Academy of Sciences*, 94(20):10792–10796, September 1997.

[64] Takenori Miyao and Nancy A. Woychik. RNA polymerase subunit RPB5 plays a role in transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, 95(26):15281–15286, December 1998.

[65] Sachiko Miyata, Jakob Begun, Emily R. Troemel, and Frederick M. Ausubel. DAF-16-Dependent Suppression of Immunity During Reproduction in Caenorhabditis elegans. *Genetics*, 178(2):903–918, February 2008.

[66] J. Montojo, K. Zuberi, H. Rodriguez, F. Kazi, G. Wright, S. L. Donaldson, Q. Morris, and G. D. Bader. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics (Oxford, England)*, 26(22):2927–2928, November 2010.

[67] Thilakam Murali, Svetlana Pacifico, Jingkai Yu, Stephen Guest, George G. Roberts, and Russell L. Finley. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Research*, 39(Database issue):D736–743, January 2011.

[68] Coleen T. Murphy, Steven A. McCarroll, Cornelia I. Bargmann, Andrew Fraser, Ravi S. Kamath, Julie Ahringer, Hao Li, and Cynthia Kenyon. Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature*, 424(6946):277–283, July 2003.

[69] Seung Wook Oh, Arnab Mukhopadhyay, Bharat L. Dixit, Tamal Raha, Michael R. Green, and Heidi A. Tissenbaum. Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nature Genetics*, 38(2):251–257, February 2006.

[70] Alessandro Ori, BrandonÂăH. Toyama, MichaelÂăS. Harris, Thomas Bock, Murat Iskar, Peer Bork, NicholasÂăT. Ingolia, MartinÂăW. Hetzer,

and Martin Beck. Integrated Transcriptome and Proteome Analyses Reveal Organ-Specific Proteome Deterioration in Old Rats. *Cell Systems*, 1(3):224–237, September 2015.

[71] M. Oti, B. Snel, M. A. Huynen, and H. G. Brunner. Predicting disease genes using protein-protein interactions. *Journal of Medical Genetics*, 43(8):691–698, August 2006.

[72] Kally Z. Pan, Julia E. Palter, Aric N. Rogers, Anders Olsen, Di Chen, Gordon J. Lithgow, and Pankaj Kapahi. Inhibition of mRNA translation extends lifespan in Caenorhabditis elegans. *Aging Cell*, 6(1):111–119, February 2007.

[73] Siler H. Panowski, Suzanne Wolff, Hugo Aguilaniu, Jenni Durieux, and Andrew Dillin. PHA-4/Foxa mediates diet-restriction-induced longevity of *C. elegans*. *Nature*, 447(7144):550–555, May 2007.

[74] Linda Partridge, David Gems, and Dominic J. Withers. Sex and Death: What Is the Connection? *Cell*, 120(4):461–472, February 2005.

[75] Evan O. Paull, Daniel E. Carlin, Mario Niepel, Peter K. Sorger, David Haussler, and Joshua M. Stuart. Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics (Oxford, England)*, 29(21):2757–2764, November 2013.

[76] Mark W. Pellegrino, Amrita M. Nargund, Natalia V. Kirienko, Reba Gillis, Christopher J. Fiorese, and Cole M. Haynes. Mitochondrial UPR-regulated innate immunity provides resistance to pathogen infection. *Nature*, 516(7531):414–417, December 2014.

[77] Yuan-Zhu Pu, Qin-Li Wan, Ai-Jun Ding, Huai-Rong Luo, and Gui-Sheng Wu. Quantitative proteomics analysis of Caenorhabditis elegans upon germ cell loss. *Journal of Proteomics*, 156:85–93, 2017.

[78] Yu Qian, SÃÿren Besenbacher, Thomas Mailund, and Mikkel Heide Schierup. Identifying disease associated genes by network propagation. *BMC Systems Biology*, 8(1):S6, January 2014.

[79] Aaron R. Quinlan. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics*, 47:11.12.1–34, September 2014.

[80] Ramesh Ratnappan, Francis R. G. Amrit, Shaw-Wen Chen, Hasreet Gill, Kyle Holden, Jordan Ward, Keith R. Yamamoto, Carissa P. Olsen, and Arjumand Ghazi. Germline Signals Deploy NHR-49 to Modulate Fatty-Acid Îš-Oxidation and Desaturation in Somatic Tissues of C. elegans. *PLOS Genetics*, 10(12):e1004829, December 2014.

[81] Michael Rera, Sepehr Bahadorani, Jaehyoung Cho, Christopher L. Koehler, Matthew Ulgherait, Jae H. Hur, William S. Ansari, Thomas Lo, D. Leanne Jones, and David W. Walker. Modulation of longevity and tissue homeostasis by the Drosophila PGC-1 homolog. *Cell Metabolism*, 14(5):623–634, November 2011.

[82] Weronika Sikora-Wohlfeld, Marit Ackermann, Eleni G. Christodoulou, Kalaimathy Singaravelu, and Andreas Beyer. Assessing Computational Methods for Transcription Factor Target Gene Identification Based on ChIP-seq Data. *PLOS Computational Biology*, 9(11):e1003342, November 2013.

[83] GÃąbor I SimkÃş, DÃąvid GyurkÃş, DÃąniel V Veres, Tibor NÃąnÃąsi, and Peter Csermely. Network strategies to understand the aging process and help age-related drug design. *Genome Medicine*, 1(9):90, September 2009.

[84] Cathy Slack, Maria E. Giannakou, Andrea Foley, Martin Goss, and Linda Partridge. dFOXO-independent effects of reduced insulin-like signaling in Drosophila. *Aging Cell*, 10(5):735–748, October 2011.

[85] Heehwa G. Son, Mihwa Seo, Seokjin Ham, Wooseon Hwang, Dongyeop Lee, Seon Woo A. An, Murat Artan, Keunhee Seo, Rachel Kaletsky, Rachel N. Arey, Youngjae Ryu, Chang Man Ha, Yoon Ki Kim, Coleen T. Murphy, Tae-Young Roh, Hong Gil Nam, and Seung-Jae V. Lee. RNA surveillance via nonsense-mediated mRNA decay is crucial for longevity in *daf-2*/insulin/IGF-1 mutant *C. elegans*. *Nature Communications*, 8:14749, March 2017.

[86] Geneva M. Stein and Coleen T. Murphy. The Intersection of Aging, Longevity Pathways, and Learning and Memory in C. elegans. *Frontiers in Genetics*, 3:259, 2012.

[87] Ulrich Stelzl and Erich E. Wanker. The value of high quality protein-protein interaction networks for systems biology. *Current Opinion in Chemical Biology*, 10(6):551–558, December 2006.

[88] Gerdine J. Stout, Edwin C. A. Stigter, Paul B. Essers, Klaas W. Mulder, Annemieke Kolkman, Dorien S. Snijders, Niels J. F. van den Broek, Marco C. Betist, Hendrik C. Korswagen, Alyson W. Macinnes, and Arjan B. Brenkman. Insulin/IGF-1-mediated longevity is marked by reduced protein metabolism. *Molecular Systems Biology*, 9:679, July 2013.

[89] Yousin Suh, Gil Atzmon, Mi-Ook Cho, David Hwang, Bingrong Liu, Daniel J. Leahy, Nir Barzilai, and Pinchas Cohen. Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3438–3442, March 2008.

[90] Nuo Sun, Richard J. Youle, and Toren Finkel. The Mitochondrial Basis of Aging. *Molecular Cell*, 61(5):654–666, March 2016.

[91] Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P. Tsafou, Michael Kuhn, Peer Bork, Lars J. Jensen, and Christian von Mering. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43(Database issue):D447–452, January 2015.

[92] Robi Tacutu, Thomas Craig, Arie Budovsky, Daniel Wuttke, Gilad Lehmann, Dmitri Taranukha, Joana Costa, Vadim E. Fraifeld, and JoÃčo Pedro de MagalhÃčes. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Research*, 41(Database issue):D1027–1033, January 2013.

[93] Luke S Tain, Robert Sehlke, Chirag Jain, Manopriya Chokkalingam, Nagarjuna Nagaraj, Paul Essers, Mark Rassner, Sebastian GrÃűnke, Jenny Froelich, Christoph Dieterich, Matthias Mann, Nazif Alic, Andreas Beyer, and Linda Partridge. A proteomic atlas of insulin signalling reveals tissueâĂŘspecific mechanisms of longevity assurance. *Molecular Systems Biology*, 13(9), September 2017.

[94] Marc Tatar, Andrzej Bartke, and Adam Antebi. The endocrine regulation of aging by insulin-like signals. *Science (New York, N.Y.)*, 299(5611):1346–1351, February 2003.

[95] Aurelio A. Teleman, Ville Hietakangas, Aram C. Sayadian, and Stephen M. Cohen. Nutritional control of protein biosynthetic capacity by insulin via Myc in Drosophila. *Cell Metabolism*, 7(1):21–32, January 2008.

[96] Ye Tian, Gilberto Garcia, Qian Bian, Kristan K. Steffen, Larry Joe, Suzanne Wolff, Barbara J. Meyer, and Andrew Dillin. Mitochondrial Stress Induces Chromatin Reorganization to Promote Longevity and UPRmt. *Cell*, 165(5):1197–1208, May 2016.

[97] Varnesh Tiku, Chirag Jain, Yotam Raz, Shuhei Nakamura, Bree Heestand, Wei Liu, Martin SpÃđth, H. Eka D. Suchiman, Roman-Ulrich MÃijller, P. Eline Slagboom, Linda Partridge, and Adam Antebi. Small nucleoli are a cellular hallmark of longevity. *Nature Communications*, 8:16083, August 2017.

[98] Janne M. Toivonen, Glenda A. Walker, Pedro Martinez-Diaz, Ivana Bjedov, Yasmine Driege, Howard T. Jacobs, David Gems, and Linda Partridge. No influence of Indy on lifespan in Drosophila after correction for genetic and cytoplasmic background effects. *PLoS genetics*, 3(6):e95, June 2007.

[99] Eleni N. Tsakiri, Gerasimos P. Sykiotis, Issidora S. Papassideri, Vassilis G. Gorgoulis, Dirk Bohmann, and Ioannis P. Trougakos. Differential regulation of proteasome functionality in reproductive vs. somatic tissues of Drosophila during aging or oxidative stress. *The FASEB Journal*, 27(6):2407–2420, June 2013.

[100] Jennifer M. A. Tullet, James W. Green, Catherine Au, Alexandre Benedetto, Maximillian A. Thompson, Emily Clark, Ann F. Gilliat, Adelaide Young, Kathrin Schmeisser, and David Gems. The SKN-1/Nrf2 transcription factor can protect against oxidative stress and increase lifespan in C.Âăelegans by distinct mechanisms. *Aging Cell*, 16(5):1191–1194, October 2017.

[101] Jennifer M. A. Tullet, Maren Hertweck, Jae Hyung An, Joseph Baker, Ji Yun Hwang, Shu Liu, Riva P. Oliveira, Ralf Baumeister, and T. Keith

Blackwell. Direct inhibition of the longevity promoting factor SKN-1 by Insulin-like signaling in C. elegans. *Cell*, 132(6):1025–1038, March 2008.

[102] Masaharu Uno, Sakiko Honjoh, Mitsuhiro Matsuda, Haruka Hoshikawa, Saya Kishimoto, Tomohito Yamamoto, Miki Ebisuya, Takuya Yamamoto, Kunihiro Matsumoto, and Eisuke Nishida. A Fasting-Responsive Signaling Pathway that Extends Life Span in C.Âăelegans. *Cell Reports*, 3(1):79–91, January 2013.

[103] Masaharu Uno and Eisuke Nishida. Lifespan-regulating genes in *C. elegans*. *npj Aging and Mechanisms of Disease*, 2:16010, June 2016.

[104] Oron Vanunu and Roded Sharan. A Propagation-based Algorithm for Inferring Gene-Disease Assocations. pages 54–52, January 2008.

[105] David Vilchez, Ianessa Morantte, Zheng Liu, Peter M. Douglas, Carsten Merkwirth, Ana P. C. Rodrigues, Gerard Manning, and Andrew Dillin. RPN-6 determines C. elegans longevity under proteotoxic stress conditions. *Nature*, 489(7415):263–268, September 2012.

[106] Daoyong Wang, Yuanping Cui, Zhixia Jiang, and Wei Xie. Knockdown expression of eukaryotic initiation factor 5 C-terminal domain containing protein extends lifespan in Drosophila melanogaster. *Biochemical and Biophysical Research Communications*, 446(2):465–469, April 2014.

[107] C. A. Wolkow, K. D. Kimura, M. S. Lee, and G. Ruvkun. Regulation of C. elegans life-span by insulinlike signaling in the nervous system. *Science (New York, N.Y.)*, 290(5489):147–150, October 2000.

[108] Jason G. Wood, Blanka Rogina, Siva Lavu, Konrad Howitz, Stephen L. Helfand, Marc Tatar, and David Sinclair. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature*, 430(7000):686–689, August 2004.

[109] Rochele Yamamoto and Marc Tatar. Insulin receptor substrate chico acts with the transcription factor FOXO to extend Drosophila lifespan. *Aging Cell*, 10(4):729–732, August 2011.

[110] Jingkai Yu and Russell L. Finley. Combining multiple positive training sets to generate confidence scores for protein-protein interactions. *Bioinformatics (Oxford, England)*, 25(1):105–111, January 2009.

[111] Elad Yunger, Modi Safra, Mor Levi-Ferber, Anat Haviv-Chesner, and Sivan Henis-Korenblit. Innate immunity mediated longevity and longevity induced by germ cell removal converge on the C-type lectin domain protein IRG-7. *PLOS Genetics*, 13(2):e1006577, February 2017.

[112] Mei Zhong, Wei Niu, Zhi John Lu, Mihail Sarov, John I. Murray, Judith Janette, Debasish Raha, Karyn L. Sheaffer, Hugo Y. K. Lam, Elicia Preston, Cindie Slightham, LaDeana W. Hillier, Trisha Brock, Ashish Agarwal, Raymond Auerbach, Anthony A. Hyman, Mark Gerstein, Susan E. Mango, Stuart K. Kim, Robert H. Waterston, Valerie Reinke, and Michael Snyder. Genome-Wide Identification of Binding Sites Defines Distinct Functions for Caenorhabditis elegans PHA-4/FOXA in Development and Environmental Response. *PLOS Genetics*, 6(2):e1000848, February 2010.

B. Curriculum Vitae

Personal data

M.Sc. Manopriya Chokkalingam

Born on 11.04.1990

Mommsenstr.31A
50935 Köln

Mobil: 01578-3368 147

email: manopriya.chokkalingam@uni-koeln.de

Education

Since June 2014	Universität zu Köln PhD in Computational Biology
2011–2013	Madurai Kamaraj University, India M.Sc in Computational Biology
2007–2011	Tamil Nadu Agricultural University, India B.Tech in Biotechnology

Research experiences

2012–2013	Madurai Kamaraj University, India Master Thesis: Computational prediction of exoproteome in Aspergillus: Comparative profiling of pathogenic vs non-pathogenic species
May 2012– July 2012	Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany Research internship: Screening of A.thaliana mutants that are defective in growth promotion response to Raoultella terrigena
2010–2011	TamilNadu Agricultural University, India Bachelor thesis: Molecular and biochemical profiling of

Kavuni rice (Oryza sativa)- towards developing Anti-diabetic rice variety

July 2010- Sep 2010 University of Agricultural Sciences, GKVK, India
Research internship: Analysis of Genetic Variations in Oryza sativa Mutants (Nagina 22) differing in Water Use Efficiency Using CAPS Markers.

List of Publications

1 Luke S Tain, Robert Sehlke, Chirag Jain, Manopriya Chokkalingam, Nagarjuna Nagaraj, Paul Essers, Mark Rassner, Sebastian Grönke, Jenny Froelich, Christoph Dieterich, Matthias Mann, Nazif Alic, Andreas Beyer, and Linda Partridge. A proteomic atlas of insulin signalling reveals tissue-specific mechanisms of longevity assurance. Molecular Systems Biology, 13(9), September 2017.

2 Marie-Therese Mackmull, Bernd Klaus, Ivonne Heinze, Manopriya Chokkalingam, Andreas Beyer, Robert B. Russell, Alessandro Ori, and Martin Beck. Landscape of nuclear transport receptor cargo specificity. Molecular Systems Biology, 13(12):962, December 2017.

C. Erklärung zur Dissertation

Ich versichere, dass ich die von mir vorlegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation nach keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Andreas Beyer betreut worden.

Ort, Datum	Unterschrift