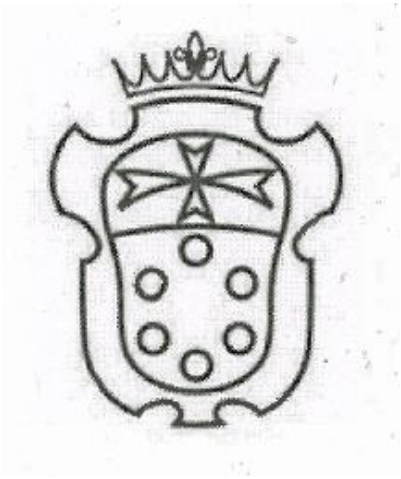


SCUOLA NORMALE SUPERIORE

CLASSE DI SCIENZE

TESI DI PERFEZIONAMENTO IN NEUROSCIENZE

MARIATERESA MAZZETTO

UNRAVELING THE MOLECULAR MECHANISMS OF AGING IN *NOTHOBRANCHIUS FURZERI* USING HIGH-THROUGHPUT TECHNIQUES

RELATORI

PROF. ALESSANDRO CELLERINO

DR. ALESSANDRO ORI

CANDIDATA

MARIATERESA MAZZETTO

ANNO ACCADEMICO 2019/2020

Contents

Introduction

1. Aging and proteostasis

1.1. What is aging?

Aging is defined as age-dependent increase in mortality. The improvement of medical cares, as well as living conditions, have increased health and life expectancy, from more or less 50 years in the early 1900s to over 80 years today. However, the aging rate (defined as the doubling time of the mortality risk) remained stable. Aging is also the main risk factor for the prevalent diseases of developed countries: cancer, cardiovascular disease and neurodegeneration (**Figure 1**). Natural selection is expected to exert a negative selection against ageing (because an organism that does not age would leave more offspring increasing fitness), but extrinsic factors such as parasites, mortality and accidents curb natural lifespan. Therefore, the pressure of selection against mutations that cause detrimental phenotypes at advanced age, like in Huntington's disease, is negligible and mutations with late phenotypes can accumulate as quasi-neutral mutations (Medawar's theory). An alternative theory posits that natural selection exerts a

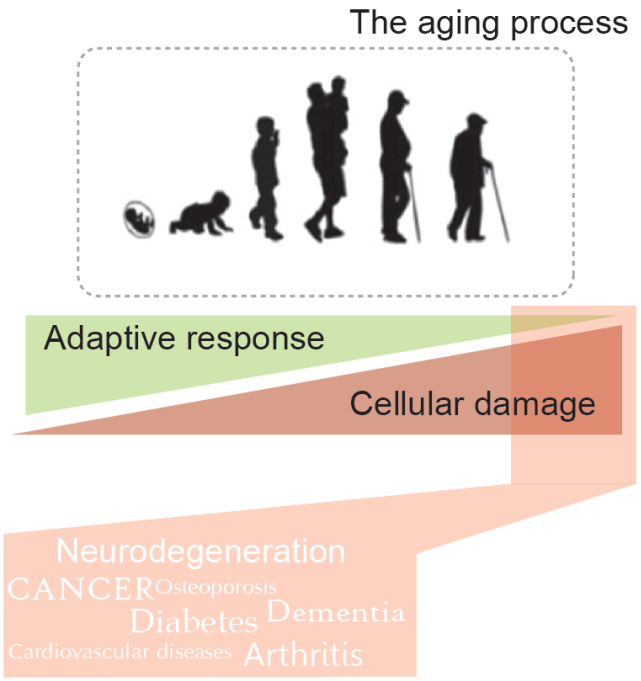


Figure 1. Aging and disease. The aging process is characterized by the decrease in the adaptive response and in the increase of cellular damage, which lead to a higher probability of developing different kinds of diseases.

positive selection for alleles that increase fitness effects in youth, but the same alleles have deleterious phenotypes later in life (antagonistic pleiotropy).

Different types of interventions have been shown to prolong lifespan and retard physiological decay in model organisms: dietary restriction (DR) can increase lifespan in diverse species, from yeast to primates (Omodei and Fontana, 2011). Rapamycin, an inhibitor of the mTOR pathway, was observed to increase lifespan in worms, flies and mice (Sharp and Richardson, 2011; Robida-Stubbs et al., 2012) through the active upregulation of autophagy, while metformin, an AMPK activator, can extend lifespan in mice and *C.elegans* (Onken and Driscoll, 2010). More recently also rotenone, a small molecule which inhibits Complex I, has been shown to extend lifespan in *Nothobranchius furzeri* at low concentrations, adding the oxidative phosphorylation pathway among the processes that have an important impact on lifespan (Baumgart, Priebe, Groth et al., 2016). Several other conserved traits, like mitochondrial activity, DNA damage response, telomere shortening and autophagy are associated with ageing and seem to play a prominent role in disease development: nutrient-sensing pathways and mitochondria maintain metabolic and energy homeostasis, while DNA repair and autophagy are deputed to repair damages (Niccoli and Partridge, 2012).

1.2. The proteostasis network and the proteasome

Loss of protein homeostasis (proteostasis) is among the most common features of aging and age-associated disease, in particular neurodegeneration: it is characterized by the appearance of non-native protein aggregates in various tissues with consequences for cognitive decline. Since many proteins have a spontaneous tendency to form aggregates, protein quality control mechanisms avoid aggregation in physiological condition mainly by the actions of the proteostasis network (PN), an assemble of macromolecular machines which operate in different ways, in order to maintain proteome integrity across subcellular compartments to ensure a healthy life span (Powers et al., 2009). This definition encompasses the translational machinery, molecular chaperones and co-chaperones, the ubiquitin–proteasome system (UPS), and the autophagy machinery (**Figure 2**).

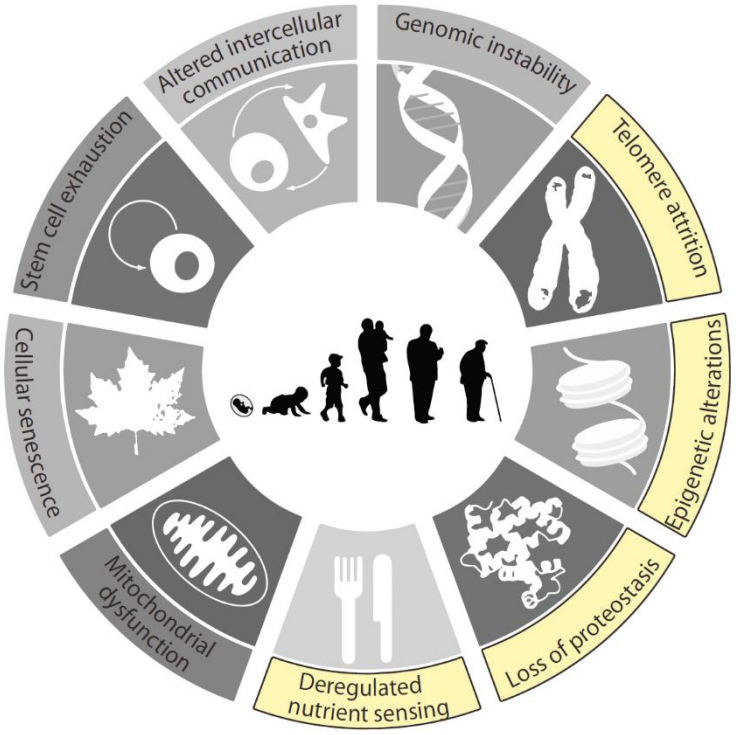


Figure 2. Graphical representation of the 9 hallmarks of aging, containing also "loss of proteostasis" category (modified from Harel et al., 2015).

Molecular chaperones are central to the function of the PN (Haslbeck et al., 2005); they "can act alone or in various combinations with different co-chaperones to regulate client–substrate interactions, folding, disaggregation, degradation, and trafficking within the cell" (Labbadia and Morimoto, 2015). When the functions of a protein cannot be restored from a misfolded or aggregated states, chaperones intervene, in order to redirect non-native proteins towards degradation pathways. Misfolded proteins are degraded by proteasomes (Finley, 2009) or lysosomes (Yiang and Klionsky, 2010).

The proteasome is composed by two sub-complexes (**Figure 3**): a core particle with catalytic functions (CP; also known as the 20S proteasome) and one or two 19S regulatory particle(s) (RP), with a mass of 700 kDa (so called PA700) that work as proteasome activators (Tanaka, 2009; Demartino and Gillette, 2007).

The 20S core particle is a protein complex with a sedimentation coefficient of 20S and a molecular mass of 750 kDa. It forms a packed particle, which is the result of the stacking of two outer β-rings and two inner α-rings, forming a $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$ structure. The 20S proteasome has a conserved role of proteolysis in eukaryotes, "differing from proteasomes in prokaryotes that mainly consists of homo-hepatmeric α- and β-rings of the same α and β subunits, respectively, i.e., the $\alpha\beta\beta\alpha$ structure". (Labbadia and Morimoto, 2015)

The proteasome, in its enzymatically active form, is normally capped on one or both ends of the central 20S proteasomal core by proteins with regulatory activity. The 19S regulatory particle recognizes non-native proteins tagged with poly-ubiquitin chains, removes the chain and traps the protein portion, unfolds the substrate peptide, opens the α-ring, and transfers the unfolded peptides into the core particle for degradation. It comprises approximately 20 different subunits that can be sub-classified into "Regulatory particle of triple-ATPase" (Rpt) subunits and "Regulatory particle of non-ATPase" (Rpn) subunits, both of which containing multiple proteins with variable molecular mass (10-110 kDa).

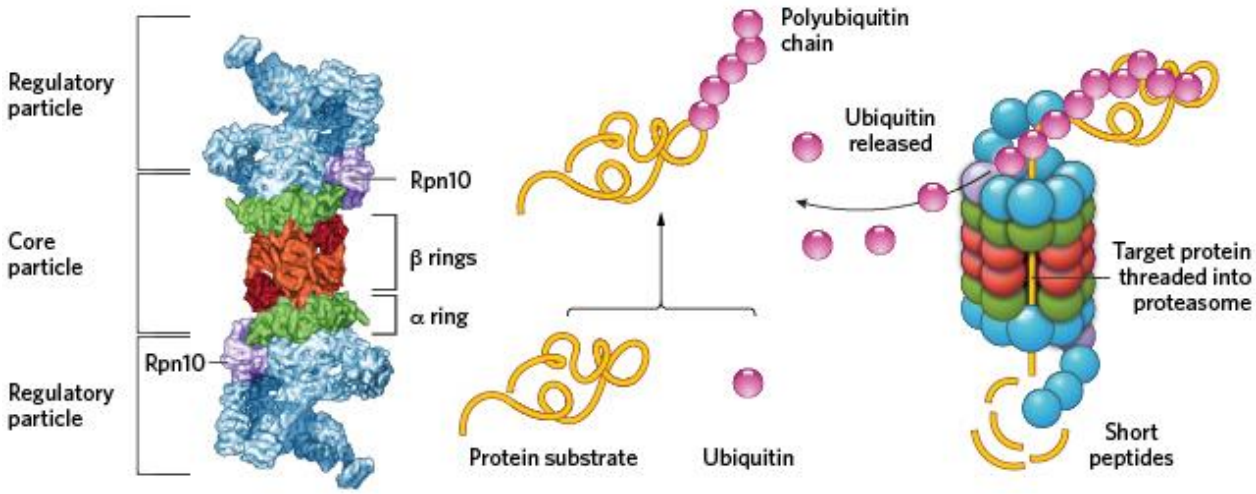


Figure 3. Proteasome structure and function. After binding to the misfolded protein ubiquitin is responsible for the targeting to the proteasome complex, with subsequent degradation of the protein into short peptides (modified from Tanaka et al., 2009).

Ubiquitinated substrates are then recognized by ubiquitin-specific receptors and targeted to the 26S proteasome for destruction. The receptors can be classified in relation to the association with the proteasome: there are proteasome intrinsic receptors which are subunits of the 19S (such as Rpn13, Rpn1 and Rpn10; Husnjak, Elsasser, Zhang et al., 2008; Shi, Chen, Elsasser et al., 2016; Fu, Sadis, Rubin et al., 1998), and also extra-proteasomal proteins (also called non-proteasomal ubiquitin-associated shuttle proteins) that bind ubiquitinated substrates as free entities and shuttle them to the 26S proteasome (such as Rad23, Dsk2 and Ddi1, which have been well characterized in yeast).

1.3. The proteostasis network during aging

A large number of human disorders can be related to disruptions of the proteostasis network, and this highlights its importance in human health. Protein aggregation is categorized as one of the main hallmarks of neurodegenerative diseases, and is characterized by the increasing appearance of detergent-insoluble inclusions in the nucleus and cytoplasm of neurons. These structures contain amyloid fibrils of cross-β-sheet-enriched proteins, and are observed in Huntington's Disease (HD), Parkinson's Disease (PD), Alzheimer's Disease (AD), and Amlylotrophic Lateral Sclerosis (ALS). Similar aggregate structures have also been detected in type II diabetes and in a range of amyloid disorders affecting peripheral organs: a total 50 human diseases have been linked to amyloid formation (Knowles, Vendruscolo and Dobson, 2014). The common basis of these observations has led to the fundamental questions of how and why proteostasis collapse occurs and how this can affect the onset and progression of neurodegenerative diseases. A lot of studies performed in worms suggest that the chronic expression of aggregation-prone proteins can reduce the folding capacity of the proteostasis network, resulting in the misfolding of proteins (Morley, Brignull, Weyers and Morimoto, 2002; Gidalevitz, Ben-Zvi, Ho, Brignull and Morimoto, 2006). Although these observations support a dysregulation of the proteostasis network as an important feature in the development of neurodegenerative disease, these studies were not successful in finding which components of the proteostasis machinery are compromised.

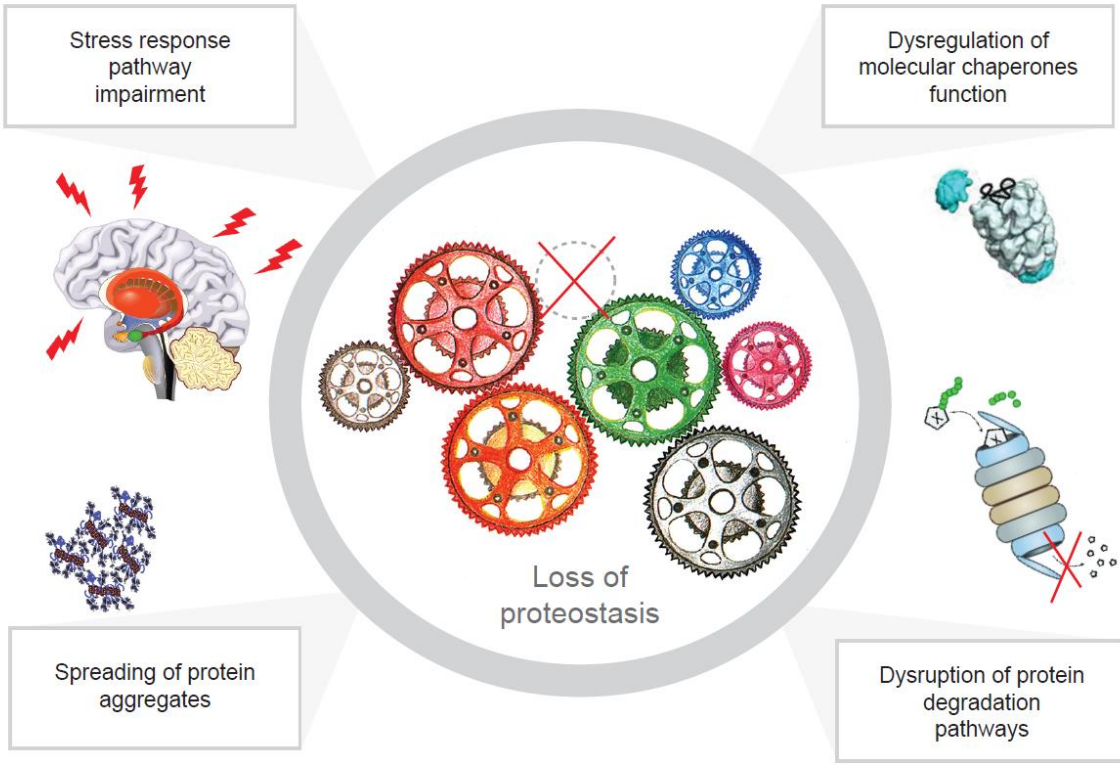


Figure 4. Mechanisms of loss of proteostasis. The disruption of the proteostasis network is generally related to four mechanisms: the impairment of the stress response pathway, the dysregulation of chaperones, to the disruption of protein degradation pathways and also to the accumulation of aggregates.

In general, proteostasis network disruption has been reported to be consequence of (**Figure 4**):

- **The dysregulation of molecular chaperones activity**. Studies of chaperone levels showed that a decline in the levels of HSP70 and DNAJ/HSP40, as well as some co-chaperones correlates with protein aggregation (Hay et al., 2004; Kim et al., 2002). This suggests that perturbation of molecular chaperones may be central to global proteostasis collapse under pathological conditions;

- **The disruption of protein degradation pathways.** One of the principal features of protein folding diseases is the ubiquitination of proteins which are contained in aggregates; the presence of inclusions that are ubiquitinated suggests that disease-related proteins marked for degradation may be inefficiently targeted to proteasomes or are resistant to degradation or an impairment of the proteolytic activity of the UPS could underlie the accumulation of ubiquitylated proteins and therefore significantly contribute to multiple neurodegenerative diseases;

- **The impairment of the Stress Response Pathway.** Increase in the dysregulation of the heat shock response (HSR) during disease progression has been observed to be associated with toxicity not only in tissue culture, *D. melanogaster* and mouse models, but also in mammalian cells (Chan et al., 2011; Olzscha et al., 2011). Few studies have investigated the relationship between the UPR pathway and neurodegenerative diseases, although several groups have reported that ER stress is an early feature of neurodegenerative and prion diseases (Hetz and Mollereau, 2014). Disruption of the ATF6 arm of the UPR response is reported "to occur in mouse models of HD and a cell model of ALS based on expression of the disease gene VAPB", suggesting that differential changes in UPR arms may be a feature of disease progression (Gkogkas et al., 2008; Fernandez-Fernandez et al., 2011);

- **Spreading of protein aggregates.** The spreading of aggregates was originally demonstrated for Prions. In addition to disruptions in the proteostasis network, several observations indicate that mutant aggregates show a spreading behaviour in cultured cells, but also in rodent and patient brains. Although the spreading of aggregates appears to be a common feature of these diseases, "the mechanism by which internalization and transmission occur may be disease specific and has been proposed to occur through endosomal pathways, secretory vesicles, and macropinocytosis in models of PD, HD, and ALS, respectively";

Frequently, cells may be able to adequately handle aberrant protein species for long periods, sometimes decades, as suggested by the fact that even the inherited forms of neurodegenerative disease, such as HD, do not show a clinical phenotype until advanced age.

Indeed, specific modulation of PN components can impact both aggregate morphology and lifespan in model systems, paving the way for therapeutic intervention (Balch et al., 2008; Powers et al., 2009). Expression of chaperones and co-chaperones of different classes have consistently resulted in a decrease in disease-aggregate toxicity and even increased lifespan (Auluck et al., 2002; Hoshino et al., 2011; Chafekar et al., 2012). Analogous to cellular stress responses, strategies for therapeutic treatments for neurodegenerative diseases that are associated with protein aggregation have focused on preventing the increase in misfolded entities, stabilization of the properly folded ones, and clearance of existing aggregates (Balch et al., 2008; Calamini et al., 2011). Indeed, small molecules that prolong translation attenuation on stress, and which also stabilize mutant proteins against aggregation, have been identified (Tsaytler et al., 2011), such as transthyretin. Other molecules previously found are related to folding and trafficking defects in specific disease associated proteins, such as mutant cystic fibrosis transmembrane conductance regulator (Baranczak and Kelly, 2016), and increase the degradation of toxic protein species through activation of the UPS (Lee et al., 2010) or autophagy (Sarkar et al., 2009). Because of the broad range of components which are captured by protein aggregates, improvement of the activity of endogenous stress-response pathways has been particularly useful in extending lifespan and health of the proteostasis network (Sittler et al., 2001; Mu et al., 2008; Akerfelt et al., 2010; Kumsta et al., 2017).

2. *Nothobranchius furzeri*: an emerging model for aging research

2.1. Natural history

Annual killifish are a group of teleost fishes from the cyprinodont clade, a taxon with Gondwanan distribution; they belong to the Cyprinodontiforms order, which comprises small fishes which regularly occupy marginal habitats, often inhospitable for other teleost species (Naiman, Gerking and Ratcliff, 1973; Taylor et a., 2008).

Most of the African annual killifishes belong to the genus *Nothobranchius*, which contains over 60 species (Froese and Pauly, 2014).

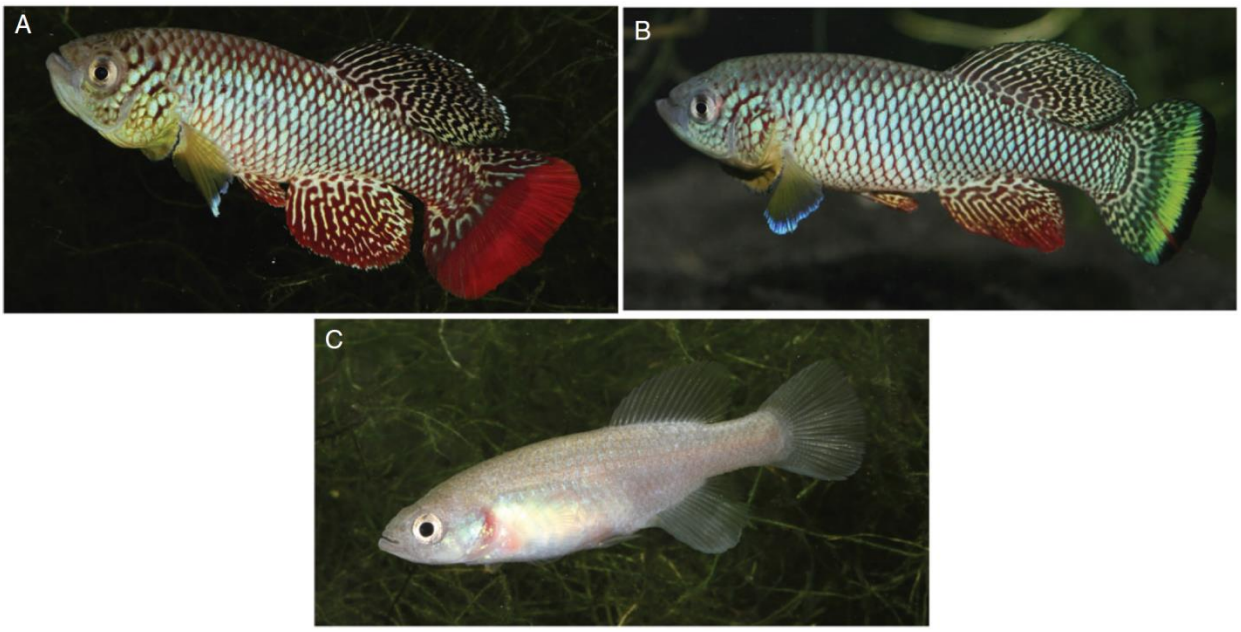


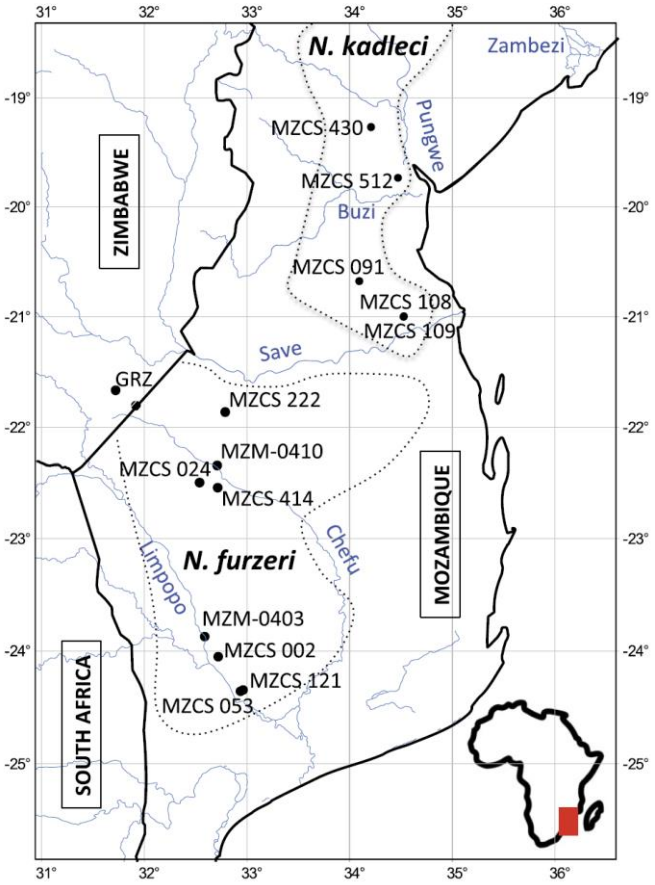
Figure 5. Sexual dimorphism in *N.furzeri* (modified from Cellerino, Valenzano & Reichard, 2016). Adult males of (A) red and (B) yellow morphs, and (C) female.

African annual killifishes are small and characterized by marked sexual dimorphism and dichromatism: males are robust and colourful while females are dull. The colour of the male is sexually selected and species-specific (Haas, 1976; Wildekamp, 2004). Males of multiple species, including *N.furzeri*, can happen in two or more colour forms that are sympatric or allopatric. In *N.furzeri*, red and yellow morphs are present, differing mainly in the colouration of the caudal fin (**Figure 5**). Female *Nothobranchius* are always smaller than males, their fins are translucent and the body is pale brown.

The distribution of the genus goes from Sudan (North Africa) to KwaZulu Natal in South Africa (Wildekamp, 2004), and is composed by four different phylogenetic clades that are "almost exclusively allopatric" (Dorn et al., 2014; Cellerino, Valenzano & Reichard, 2016; **Figure 6**):

- The northern clade occupies mainly dry regions in northern Kenya and Somalia;

- The southern clade is distributed in the southern part of the Zamberi River, going from the humid coast to dry habitats, and has also been the most studied;
- The inland clade is distributed in a region at high altitude between Lake Victoria (Uganda) and Kafue basin (Zambia) and is separated from the coastal clade by rift valleys;
- The coastal clade is distributed in basins of the coast of Southern Kenya, Tanzania and Northern Mozambique.



In the genus *Nothobranchius*, *N.furzeri* has been particularly studied: it goes from the area between the Save river and the Lebombo ridge. The type locality is the Sazale Pan (Gonarezhou National Park), Zimbabwe, where the species was first collected in 1968 (Jubb, 1971). Its range encompasses a strong cline of aridity and rainfall unpredictability (Cellerino, Valenzano & Reichard, 2016; Terzibasi Tozzini et al., 2013).

Figure 6. Localization of Nothobranchius genus across Africa (Cellerino, Valenzano & Reichard, 2016).

The most arid weather conditions are found in the hinterland, at higher altitude and furthest from the ocean; the region has irregular precipitations and is subject to stronger evaporation (Mazuze, 2007; Terzibasi et al., 2008).
Nothobranchius, including *N.furzeri*, normally feed on macroinvertebrates; on the taxonomic point of view, their diet is opportunistic and depends on prey availability (Polacik and Reichard, 2010). All *Nothobranchius* tend to be generalist predators; captive *N. furzeri*, like other species of the same genus, quickly eat live and frozen dipteran larvae of the genera *Chironomus* and *Chaoborus* (Genade, 2005), Live *Tubifex* sp. (Oligochaeta), *Artemia salina* and *Daphnia* spp.

2.2. Life cycle, lifespan and aging in *N.furzeri*

The ability of *Nothobranchius* embryos to enter diapause seems to be an ancient character state that has been lost multiple times during the evolution of African and Neotropical killifish clades (Murphy and Collier, 1997; Hrbek and Larson, 1999).

After the flooding of the pools, eggs hatch and the juveniles fishes can develop quickly in order to to reach sexual maturity. The duration of the habitat is normally of 2-3 months (Terzibasi Tozzini et al., 2013; Vrtilek et al., 2018a) and eggs stay in diapause for 10 months (as shown in the canonical developmental pathway, **Figure 7**); some embryos do not enter into diapause and go through direct development, hatching if the habitat is filled again during the same season (*alternative pathway*), otherwise they die. Some eggs have to remain in diapause for longer than one year in order to survive the occasional dry years. Variability of the diapause period length can represent an important strategy to survive an unpredictable environment (Polacik et al., 2014). After hatching juveniles are at once able to feed actively, and, under natural conditions, they can reach sexual maturity in about 12 days (Vrtilek et al., 2018b).

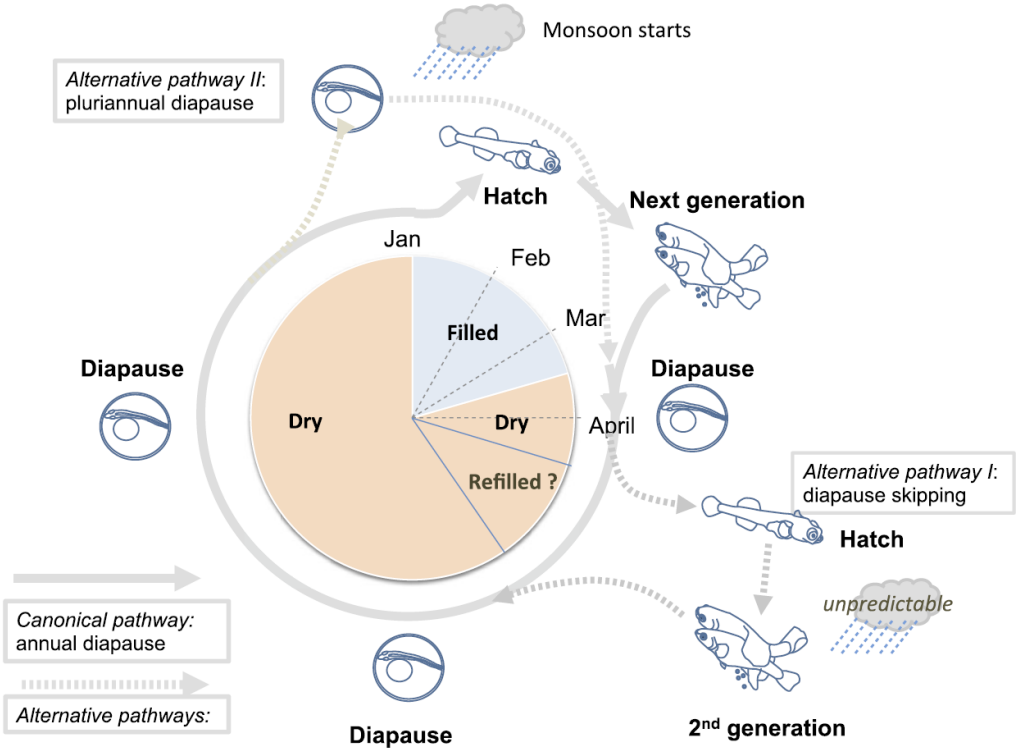


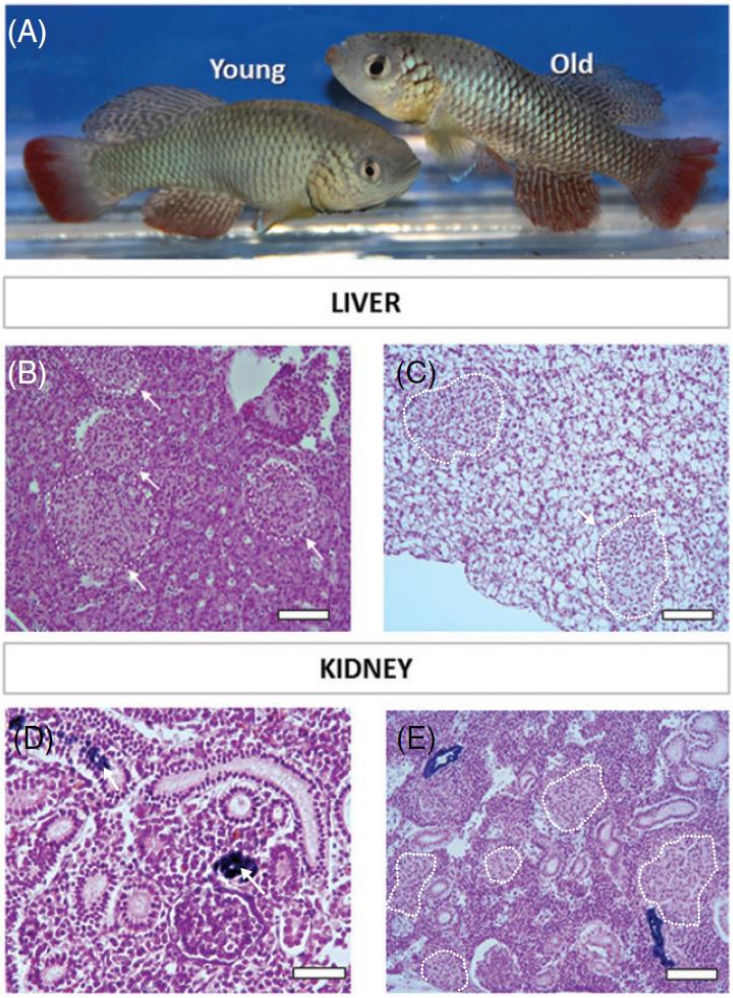
Figure 7. Schematic representation of the life cycle of *Nothobranchius* (Cellerino, Valenzano & Reichard, 2016), containing the canonical pathway (annual diapause) as well as the alternative pathways.

Several species of the *Nothobranchius* genus have been studied in captivity and all show short captive lifespan varying between 3 and 18 months (Valdesalici and Cellerino, 2003; Lucas-Sanchez et al., 2011; Baumgart et al., 2015).

Given the extremely short lifespan of *Nothobranchius*, and in particular *N.furzeri*, as compared to other vertebrates, an important part of studies in this species was focused on the confirmation that the short lifespan was a consequence of accelerated aging. Old *N.furzeri* shows a typical phenotype of emaciation, spinal curvature, and reduced colouration in males (**Figure 8**), known to be present also in *Danio rerio* and *Oryzias latipes* (Gerhard et al., 2002; Hatakeyama et al., 2008). At the behavioural level there is a generalized reduction in spontaneous locomotor activity, with older *N.furzeri* individuals which spend less time exploring, in comparison to young ones (Genade et al., 2005); this phenotype was reported also in *N.korthausae*, in addiction to the disruption of the circadian rhythm (Lucas-Sanchez et al., 2011). *N.furzeri* shows an impairment of learning performance during ageing, in particular when tested using an active avoidance task: the performance of older individuals in associating a conditioned stimulus with an unconditional stimulus was significantly lower than the one of younger animals (Valenzano et al., 2006; Tezibasi et al., 2008).

Histopathological examinations have revealed age-related decay of organs (Di Cicco et al., 2011). *Post-mortem* analyses revealed mainly lesions in kidney (**Figure 8**), liver and heart. Kidneys normally suffer tubule dilatation and crystal deposition (Di Cicco et al., 2011), which are present also in the human geriatric phenotype (Silva, 2005), while cardiac lesions included hypertrophy of the cardiomyocytes (another typical aspect of vertebrate ageing; Woodhead, 1984; Dai et al., 2012).

Figure 8. Aging phenotype of *N.furzeri* (Cellerino, Valenzano & Reichard, 2016; Di Cicco et al., 2011). (A) Representation of the aging phenotype, with discolouration and emaciation of the older individual. (B-C) Liver histopathological phenotype, in particular (B) hepatomas and (C) hepatocarcinomas (both lighlighted by white arrows and dotted lines). (D-E) Kidney histopathological phenotype, in particular (D) nephrocalcinosis (indicated by white arrows) and (E) neoplasias (highlighted by dotted lines).



Cellular ageing has been deeply investigated in liver, skin and brain of several *Nothobranchius* species. Age-dependent accumulation of lipofuscin was observed in the liver also of wild *Nothobranchius* spp. (Terzibasi Tozzini et al., 2013); lipofuscin is an auto fluorescent pigment that accumulates with age in many organisms from nematodes to humans and is a key point for ageing intervention studies (Valenzano et al., 2006a; Hsu and Chiu, 209; Terzibasi et al., 2009; Yu and Li, 2012). In the liver, increased apoptosis is observed during aging of *N.furzeri* and also medaka (Di Cicco et al., 2011; Ng'oma et al., 2014; Ding et al., 2010).

In the skin, there is evidence that ageing is associated with accumulation of cells that may have irreversibly lost their capacity of entering the cell cycle: human senescent cells can be identified *in vivo* by the expression of β-galactosidase (β-Gal) marker. Age-dependent increase of β-Gal staining has been reported in *N.furzeri* and other two *Nothobranchius* species (Genade et al., 2005; Hsu et al., 2008; Liu et al., 2012) as well as up-regulation of the cell-cycle inhibitor *CDKN1A*.

N.furzeri brain was studied in relation to its morphology and gene expression patterns, as reported by D'Angelo et al., 2012, 2014; D'Angelo, 2013. The major cellular phenotypes that have been observed in ageing are the following:

- the important reduction of stem cell activity, observed by Tozzini et al.(2012), which accompaines the age-dependent reduction of adult neurogenesis described in mammals (Kempermann, 2011);
- glial hypertrophy or gliosis, as over-expression of glial fibrillary acidic protein (GFAP) (Tozzini et al., 2012), another typical phenotypic response of mammalian glia to injury,

neurodegenerative disease and ageing (Norton et al., 1992, Bronson, Lipman and Harrison, 1993);

- neuronal degeneration, as measured by Fluoro-Jade B staining in *N.furzeri* and *N.guentheri* (shown by Valenzano et al., 2006b; Terzibasi et al., 2009; Liu et al., 2012);
- accumulation of lipofuscin (Terzibasi et al., 2009; Terzibasi Tozzini et al., 2013).

At a molecular level, the major changes described in *Nothobranchius* species are:

- telomere erosion, observed in three different species and multiple organs (Hsu et al., 2008; Hartmann et al., 2009; Liu et al., 2012);
- reduced mitochondrial DNA quantity and mitochondrial function of the muscular mitochondria (Hartmann et al., 2011);
- Reduction of polyunsaturated fatty acids that are peroxidation-sensitive (in *N.korthausae*; Lucas-Sanchez et al., 2011) and modulation of mitochondrial fatty acid contents (*N.rachovii*), in relation to the increased damage to mitochondrial lipids (Lucas-Sanchez et al., 2014).

2.3. Genetic differences of lifespan across Turquoise killifish species

For my work, I decided to focus on *N.furzeri*: median lifespan in this species can be of 3 months for the Gonarezhou (GRZ) inbred strain (Valdesalici and Cellerino, 2003). Records for wild-derived strains from semi-arid or more humid regions in Mozambique (e.g., MZM-0403 and MZM-0410) and also a locality at the borderland between Mozambique and Zimbabwe (MZZW-0701) within less than 30 km from the collection point of the GRZ strain have a longer median lifespan of ~6-8 months and maximum lifespans exceeding one year (Terzibasi et al., 2008, Tozzini et al., 2013; Baumgart et al., 2016; Blazek et al., 2016); these observations suggest that the short lifespan of the GRZ strain is an extreme phenotype within the *N.furzeri* species itself and raises the question as to whether it is the result of captive breeding rather than a natural phenotype. Indeed, the strains cited above were established only several years ago and are genetically heterogeneous to various extents, whereas GRZ is highly inbred (Reichwald et al., 2009; **Table 1**). The females are entirely inbred (with a 100% homozygosity in 152 tested microsatellite markers) and males show heterozygosity exclusively at the sex-linked markers, as reported by Valenzano et al. (2009) and confirmed by Kirschner et al., where the average heterozygosity was found to be 0.01 (Kirschner et al., 2012). Genetic mapping studies did not detect a single locus of major effect for lifespan determination, suggesting its complex multigenic nature and excluding the possibility of it as result of a single recessive mutation (Kirschner et al., 2012; Valenzano et al., 2015).

Microsatellite markers in *N. furzeri* and *N. kunthae*

locus	primers	repeat motif	alleles [repeat number]		
			N. furzeri [1]		*N. kunthae* [2]
			GRZ	MZM-0403	
Nofu0001	F: GCTACGAGGTTCTGCAGTCA	ATT	17	16, 17,19, 28	7, 8, 9
	R:CACATCCACCTACTAACCAGCA				
Nofu0002	F: CACCAACTGTGGAGTAGTGC	AC	33	15,20,24,33	21, 26, 32, 33
	R: CCCAACATGAACAAAGACGC				
Nofu0003	F: TTGGTCACACCTCTGGTTTC	AT	22	20,22,33,38,41	28, 30
	R: CCCAAAGCCAGCTATTAGTCC				
Nofu0004	F: ACCTATTCCACCTTCCTCAG	TTC	12	10,11,12	5, 6, 9
	R: TCAGAGCTGGGATCTAACC				
Nofu0005	F: AGTGACCTTGGGTGTTCTG	ATT	14	11,13,14,20	n.a.
	R: GCTAGGCGGTTAAAGTGTC				
Nofu0006	F: GGACCACAGAGCAAAAGGAG	AC	39	21,22,27,34,39,44	8, 12, 17, 24
	R: AGTTTTTGCCCCACTGTACG				
Nofu0007	F: GACTGACTCACAGGGTCAC	AATC	9	5,9	5, 7, 15
	R: CGAATGACCACAGTCACCAA				
Nofu0008	F: GTTAATCTGGCCCTGCTTGT	GT	34	34,40,41,49,50,53	n.a.
	R: TGGCACAGGTGAAAGCACTA				

[1] two pools of ten specimens of *N. furzeri* strain GRZ as well as ten specimens of *N. furzeri* strain MZM-0403 were analyzed; all markers were monomorphic in *N. furzeri* GRZ
[2] ten specimens of a laboratory strain of *N. kunthae* were analyzed

Table 1. Genotypes of microsatellites analyzed in *N. furzeri* strains GRZ and MZM-0403 and the closely related species *N. kunthae* (from Reichwald et al., 2009).

The extreme short lifespan observed in the GRZ strain has been shown experimentally to be coupled with an accelerated aging profile: a point of particular interest that has been studied is age-dependent neoplasia. Tumour onset is an age-dependent feature in humans, but in teleosts it is rare; on the other hand, a high incidence of spontaneous neoplasias in liver and kidney was observed in *N. furzeri* (Di Cicco et al., 2011). The onset of liver neoplasias was dependent on the longevity of the strain, and it was accelerated in the GRZ. Age-related histological markers have also been investigated in GRZ, as well as wild-derived strains (MZM-0403 and MZM-04010): lipofuscin accumulation in liver and brain was shown to be accelerated in the GRZ strain in comparison to the age-matched MZM-0403 animals, and also neurodegeneration was more pronounced.

2.4. Wild populations: growth and mortality

Captive conditions are very different from those in the wild environment, and important factors such as stress, diet composition and predator absence may affect ageing. All these parameters certainly influence gene expression. Although such differences are unavoidable, their identification is mandatory in order to generalize conclusions based on analysis of captive specimens, put them into evolutionary and natural-history perspectives and frame their interpretation. RNA-seq studies of age-associated gene expression in the wild are currently available only for bats (Huang et al., 2019) and wolves (Charruau et al., 2016). Remarkably, no datasets are available for wild populations of mice, fruit flies or nematode worms that represent the models of choice for experimental studies of aging.
Food availability and taxonomical diversity of prey, but also oscillations of environmental temperature, individual interactions and sex differences in survival are the more observed differences between wild and captive populations of *N.furzeri* (as well as other *Nothobranchius* species).
Food availability in the laboratory is normally discontinuous during the day, depending on the facility operator, while wild fishes tend to eat throughout the day: in addiction, the wild consumes different types of food (Polacik and Reichard, 2010) compared to the uniform diet used in captivity. Possibly as a result of different diets, growth rates of wild *N. furzeri* are remarkably higher than in captivity (Blazek, Polacik & Reichard, 2013).
Water temperature vary diurnally in the wild with a maximum amplitudes of 15°C (Reichard et al., 2009), inducing important changes in gene expression patterns (Podrabsky and Somero, 2004); while temperature fluctuation is buffered in larger and deeper habitats (**Figure 9**), *N.furzeri* often inhabit narrow pools which are exposed to extreme fluctuations in environmental temperature.

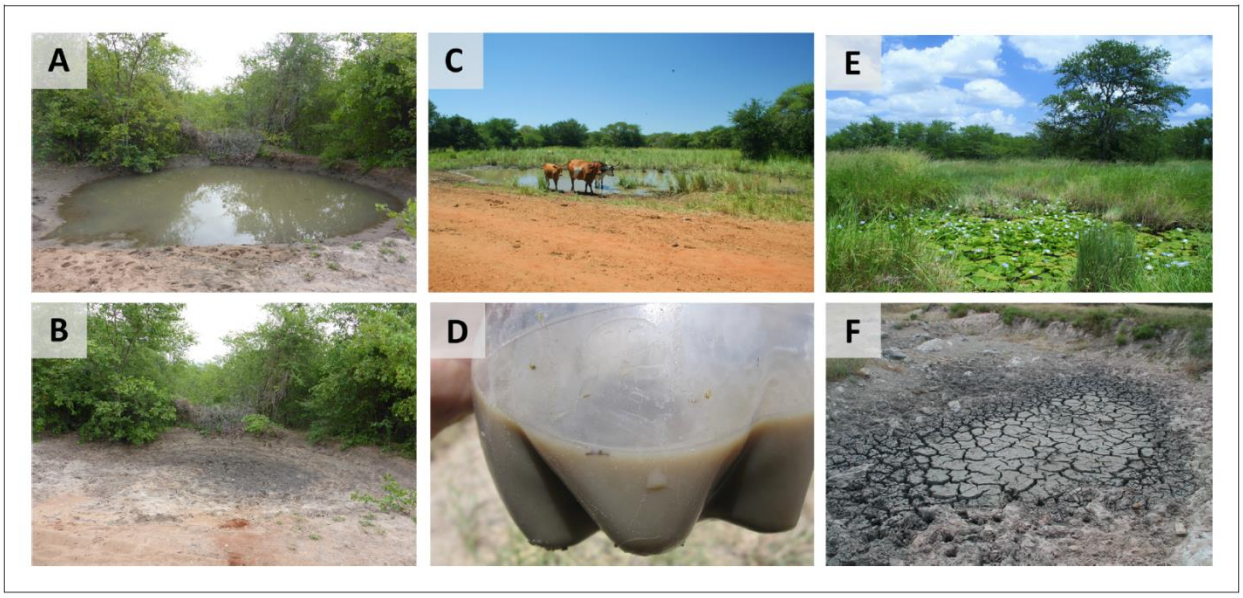


Figure 9. Natural habitat of N.furzeri (from Reichard & Polacik, 2019). (A) Habitat of *N.furzeri* one week after filling with water. (B) Photo of the same habitat three weeks after (A). (C) Domestic cattle that normally visits *N.furzeri* habitats. (D) Turbid water from *N.furzeri* habitat discoloured by dissolved fine sediment particles. (E) Complex habitat with aquatic vegetation. (F) Desiccated pool sediment.

Social structure in the laboratory populations is relatively stable, since each organism has opponents or sexual partners that are fewer, in comparison to the wild environment; however, these differences are not dramatic between captive and wild populations.

Another important point to discuss is natural mortality: the most striking difference between wild and captive populations in this case is the selective mortality of wild males, not observed in the laboratory (Reichard et al., 2014). Higher *extrinsic* mortality, which is the mortality related to external factors such as predation, implies that fewer individuals survive to reproduce at later ages, with a subsequent weaker selection for longevity; high extrinsic mortality should then result in a rapid and dramatic ageing phenotype, as deterioration of vital function, as postulated by classical theories (Medawar, 1952; Williams, 1957; Kirkwood, 1977). Inter-sexual differences in mortality rates in wild populations are linked to extrinsic mortality; in particular increased male mortality clearly arises from that (most likely predation; Haas, 1976; Reichard et al., 2014). In a population the sexes may differ in the "relative importance of intrinsic and extrinsic mortality, making inter-sexual comparison in ageing and survival a possible avenue for ecological tests of survival" (Cellerino, Valenzano & Reichard, 2016).

3. High-throughput techniques for quantification of biomolecules

3.1. Transcriptomics

3.1.1 Generalities

Transcriptome analysis technologies are the techniques used to study an organism's transcriptome, which corresponds to the set of all of its RNA transcripts. Sequencing-based techniques can be used both to identify transcripts in an unbiased way and to quantify their abundance. A transcriptome sequencing experiment therefore isolates the total transcripts present in a cell in a defined condition.

The first studies on the whole transcriptome began in the early 1990s, and technological advances since then have made transcriptomics a widely known methodology. There are two key techniques in the field nowadays: microarrays, which quantify a set of sequences which are predetermined, and RNA sequencing (RNA-seq), which uses high-throughput sequencing in order to obtain the number of counts of all expressed genes.

The measure of the expression of an organism's transcriptome in different tissues, conditions, or at different ages gives information on how genes are regulated and reveals details of the organism itself. It can also help to guess the functions of genes that are not yet annotated. Transcriptomic analysis has allowed the study of gene expression changes in different organisms and has been a key point in the understanding of the molecular basis of human diseases.

3.1.2. Data gathering

RNA-seq

All gene expression data analyzed for this thesis originate from RNA-seq, therefore this technique will be presented here in detail.

RNA-Seq refers to the "combination of a high-throughput sequencing methodology with computational methods to capture and quantify transcripts present in an RNA extract" (**Figure 10**) (Lowe et al., 2017; Ozsolak and Milos, 2011). The nucleotide sequences that are generated are

normally around 100 bp in length, but can range from 30 bp to over 10,000 bp, depending on the sequencing method which is used. RNA-Seq is based on the alignment of reads to a reference genome or to each other, in order to reconstruct the original RNA transcript (Wang, Gerstein and Snyder, 2009). The typical range of 7 orders of magnitude for RNA-Seq is an important benefit over microarray transcriptomes. In addition, input RNA amounts are much lower for RNA-Seq (ng) in comparison to microarrays (mg), which allow a more detailed examination of cellular structures and down to the single-cell level (Hashimshoni et al., 2012). Theoretically, there is no upper limit of quantification in RNA-Seq as sensitivity and precision increase by increasing sequencing redundancy, and background signal is very low for 100 bp reads in non-repetitive regions (Ozsolak and Milos, 2011). RNA-Seq can be used to identify genes in a previously annotated genome or identify which genes are active at a particular point in time, and read counts can be used to see the relative gene expression level. RNA-Seq methodology has constantly improved to increase throughput, accuracy, and read length. Since the first descriptions in 2006 and 2008 (Bainbridge et al., 2006; Nagalashkmi et al., 2008), RNA-Seq has been rapidly adopted and overtook microarrays as the dominant technique to quantify gene expression in 2015 (Su et al., 2014).

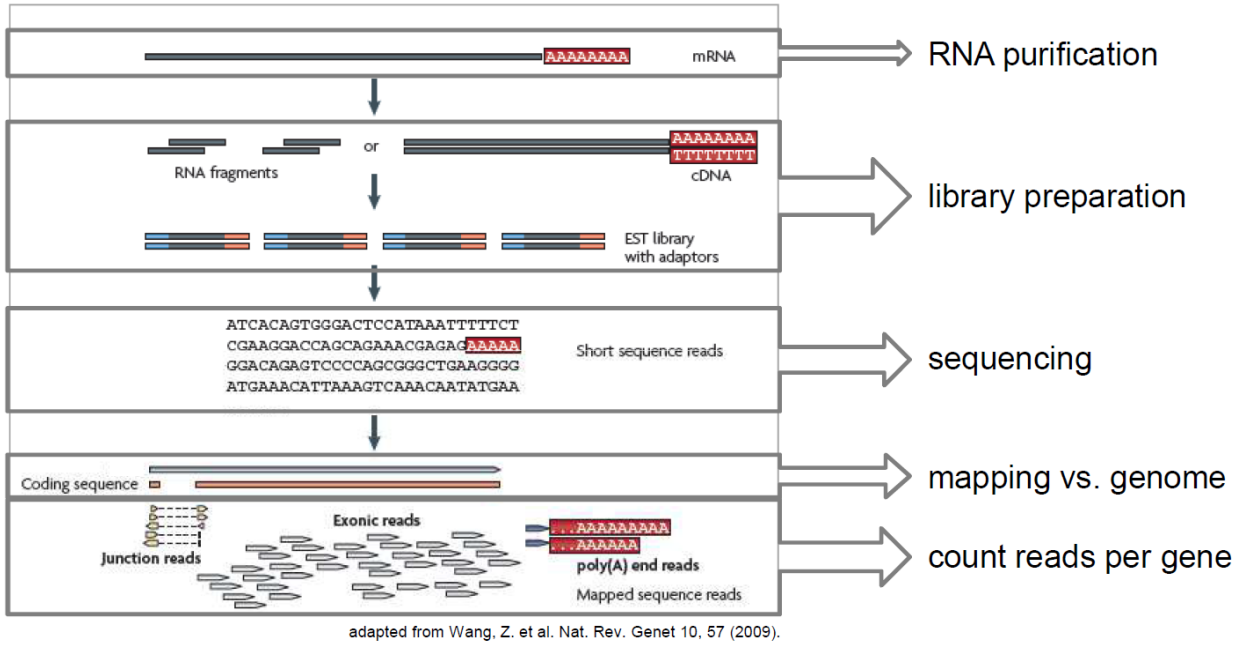


Figure 10. Workflow of an RNA-seq experiment, which is performed through RNA purification, library preparation, sequencing, mapping and counting critical steps (adapted by Wang et al., 2009).

RNA-Seq was established with the rapid development of multiple high-throughput DNA sequencing technologies for the alignment of short sequences (Shendure and Ji, 2008). Before the actual sequencing steps, there are different methods for transcript enrichment, fragmentation, amplification, single or paired-end sequencing that can be considered.

We can increase the sensitivity of an RNA-Seq with the enrichment of particular classes of RNA that are of interest, but also deleting known abundant RNAs. We can also perform ribo-depletion to remove abundant but uninformative ribosomal RNAs (rRNAs) using specific probes. However, this method can also lead to the depletion of off-target transcripts (Lahens et al., 2014). Small RNAs such as microRNAs can be purified using gel electrophoresis and extraction.

Since mRNAs are longer than the read-lengths of typical high-throughput sequencing methods, transcripts need to be fragmented before sequencing: this is a key aspect of the construction of a sequencing library (Knierim et al., 2011). It may incorporate chemical hydrolysis, nebulization, or sonication of RNA, but also make use of simultaneous fragmentation.

During the preparation, the copies of transcripts (as cDNA) can be amplified by PCR to enrich fragments with 5′ and 3′ adapter sequences (Parekh et al., 2016). Amplification is also used to allow sequencing of very low-input amounts of RNA.

When the molecules have been prepared, they can be sequenced in one direction (single-end) or both directions (paired-end). The first method is faster and less expensive, and also sufficient for quantification of gene expression levels, but the second one produces more robust alignments and/or assemblies, important in particular for gene annotation and transcript isoform discovery (Wang, Gerstein and Snyder, 2009). Strand-specific RNA-sequencing preserves the strand information of a sequenced transcript (Levin et al., 2010): without strand information, reads do not inform in which direction the gene is transcribed. This method is really useful for taking information about transcription for genes that overlap in different directions, and to make a more robust gene prediction in non-model organisms (Levin et al., 2010).

Currently, RNA-Seq is based on the conversion of RNA molecules into cDNA molecules before sequencing, making the platforms for transcriptomic and genomic data the same (**Figure 10B**). Consequently, the development of DNA sequencing technologies has been a defining feature of RNA-Seq (Liu et al., 2012; Loman et al., 2012; Goodwin, McPherson and McCombie, 2016). Direct sequencing of RNA using nanopore sequencing represents a current RNA-Seq technique in its infancy (in pre-release beta testing as of 2016) (Garalde et al., 2016; Loman et al., 2015). However, this sequencing method can detect modified bases that would not be found otherwise when sequencing cDNA and eliminates amplification steps that could otherwise introduce bias (Morozova, Hirst and Marra, 2009; Ozsolak et al., 2009).

The sensitivity and accuracy of an RNA-Seq experiment are dependent on the number of reads obtained from each sample. A large number of reads are needed, in order to have sufficient coverage of the transcriptome, permitting the detection of low abundance transcripts. However, the most effective way to improve the detection of differential expression in genes with low expression is to add more biological replicates, rather than adding more reads (Rapaport et al., 2013).

3.1.3. Data analysis

RNA-Seq experiments allow the generation of a large number of raw sequence reads, which have to be processed to obtain useful information: this step usually requires a combination of bioinformatic tools that can be used for different aims and with different experimental designs. The process can be broken down into the following stages: quality control, alignment, quantification, detection of differentially expressed genes and clustering (Van Verk, Hickman, Pieterse and Van Wees, 2013). Most popular RNA-sequencing softwares work from a command-line interface, either in a Unix environment or within the R/Bioconductor statistical environment (Huber et al., 2015).

Quality control step

Another important step is the quality control process, where raw data are examined for the detection of errors or artifacts (Conesa et al., 2016). There are several options for sequence quality analysis, including the FastQC and FaQCs software packages (Lo CC and Chain, 2014). Abnormalities identified can be removed by trimming or tagged for special treatment.

Alignment

Normally transcripts are aligned to a reference genome, or de novo aligned to one another if no reference is available. The key challenges for this type of process includes the speed to allow more than 10^9 of short sequences to be aligned in a meaningful way, the flexibility to recognize intron splicing regions, and the correct assignment of reads that map to multiple locations. Software advances have greatly addressed these issues; a list of currently available high-throughput sequence aligners is maintained by the EBI (Fonseca, Rung, Brazma and Marioni, 2012).

Alignment of eukaryotic transcript mRNA sequences to a reference genome requires previous handling of the intron sequences, which are absent in mature mRNA. Short read aligners perform an additional round of alignments for the identification of splice junctions; the identification of intron

splice junctions prevents reads from being aligned in an erroneous way or discarded, permitting to more reads to be aligned to the reference genome and improving the accuracy of the estimates.

In order to align reads to one another without the use of a reference genome a de novo assembly strategy can be used (**Table 2A**) (Miren, Koren and Sutton, 2010). An initial characterization of the *N. furzeri* transcriptome indeed required de novo transcriptome assembly (Baumgart et al., 2014) and transcriptome assembly was used to compared transcript sequences across Nothobranchius species differing in lifespan (Sahm et al., 2017). Once assembled de novo, the product of the process can be used as a reference for other sequence alignment methods and quantitative gene expression analysis.

A

Software (Manufacturer)	Released	Last Updated	Resource load	Strengths and weaknesses
Velvet-Oases [100][101]	2008	2011	Heavy	The original short read assembler, now largely superseded.
SOAPdenovo-trans [102]	2011	2015	Moderate	Early short read assembler, updated for transcript assembly.
Trans-ABySS [103]	2010	2016	Moderate	Short reads, large genomes, MPI-parallel version available.
Trinity [104][105]	2011	2017	Moderate	Short reads, large genomes, memory intensive.
miraEST [106]	1999	2016	Moderate	Repetitive sequences, hybrid data input, wide range of sequence platforms accepted.
Newbler [107]	2004	2012	Heavy	Specialised for Roche 454 sequence, homo-polymer error handling.
CLC genomics workbench (Qiagen—Venlo, Netherlands) [108]	2008	2014	Light	Graphical user interface, hybrid data.

MPI, Message Passing Interface; RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t003

B

Software	Environment	Specialisation
Cuffdiff2 [111]	Unix-based	Transcript analysis at isoform-level
EdgeR [112]	R/Bioconductor	Any count-based genomic data
DEseq2 [113]	R/Bioconductor	Flexible data types, low replication
Limma/Voom [114]	R/Bioconductor	Microarray or RNA-Seq data, isoform analysis, flexible experiment design

RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t004

Table 2. Common softwares used for alignment and differential analysis (adapted from Lowe et al., 2017). (A) Commonly used softwares for transcriptome *de novo* assembly. (B) Most used softwares for detection of differentially expressed genes.

Quantification

Quantification of sequence alignments may be performed at different levels (gene, exon, or transcript). Typical outputs include a table of reads counts for each feature: gene and exon read counts can be calculated using the HTSeq software package (Anders, Pyl and Huber, 2015), and reads that align equally to multiple locations must be identified and then removed, aligned to one of the possible locations, or aligned to the most probable location.

Detection of differentially expressed genes

Once quantitative counts of each transcript are available, differential gene expression is then tested. Detection of differential gene expression lies at the center of RNA-seq but is non-trivial due to the specific nature of the data. The first step is the normalization that is required to correct for the fact that different samples differ in the total amount of reads analyzed. The second step is the modelling of the probability distribution function (negative binomial) that underlies actual data distribution. Since typically only a handful of samples are available these do not allow gene-specific estimates of the parameters and *a priori* assumptions are necessary to model data distribution and perform statistical test. Examples of dedicated software are described in **Table 2B,** the use of generalized mixed linear models as statistical framework has become predominant since it allows to test for the effects of different covariates and also to introduce random effects in longitudinal designs.

Most read a table of read counts as their input, but some (such as cuffdiff) accept binary alignment map format read alignments. The final outputs of these analyses are "gene lists with associated pair-wise tests for differential expression between treatments and the probability estimates of those differences" (from Wikipedia).

Gene expression clustering methods

An additional step for the inspection of the data is clustering. Several methods can be used in order to subset the data and find a structure inside them (on the assumption that genes with similar expression profiles must share similar biological properties; Cellerino & Sanguanini, 2018; Rodriguez et al, 2019).

Gene expression clustering methods focus on the partition of a group of genes into sub-sets, in order to reduce the numerosity of the data. The critical steps in this type of analysis are defining a measure of distance between the genes, apply a similarity criterion (which varies from method to method) to partition the data, and define a cluster. The four more used types of analysis are the hierarchical clustering, the k-means clustering, the fuzzy c-means clustering and the Self Organizing Maps.

- *Hierarchical clustering*: it organizes the dataset as a dendrogram (where the root is the whole dataset, and the branches define different clusters) and it follows three steps. The first one is the quantification of the distance between each pair of genes (or pairwise distance), which gives an upper-diagonal similarity matrix as a result. After the highest value is found the two most connected genes are grouped into one single cluster and the similarity matrix is recomputed until only the root remains.

- *K-means clustering*: it subdivides the genes into a predefined number of clusters which centroids are randomly located (*seeding* step). The genes are then assigned to the closest centroid (through minimization of a distance measure) and the updated centroids are computed. The analysis is reiterated until the assignment of the genes to each cluster remains stable.
- *Fuzzy c-means clustering*: this method overcomes one of the limitations of k-means clustering, which is that a given gene can be included in only one cluster, and assigns to each gene a membership value (from 0 to 1) for each centroid. The algorithm continues until convergence (as in the k-means clustering).
- *Self-Organizing Map*: is a particular topologically-oriented 2D neural network, formed by a 2D grid of interconnected artificial neurons for mapping particular features in the dataset.

Knowledge-based clustering methods

A complementary method for the reduction of the data is the clustering of genes into given gene sets previously defined (by *a priori* knowledge of the function); normally this kind of analysis is particular useful when tested on an overrepresented list of genes (such as differentially expressed genes), but also to test a gene set in accordance to its up- or down-regulation to a given condition (as in the "Gene Set Enrichment Analysis" analysis).

- *Gene Overrepresentation (GO analysis)*: this type of analysis is done comparing a gene set to a background gene list and is based on the hypergeometric distribution

$$p(k|n,K,N) = \frac{\binom{K}{n}\binom{N-K}{n-k}}{\binom{N}{n}}$$

where *n* is the total number of genes in the experimental gene set and *N* the total number of genes in the background list, whereas *k* is the number of DEGs in the experimental gene set and *K* the number of DEGs in the background list. The method provides also the measurement of an enrichment score, as ratio between observed frequency (*n/k*) and expected frequency (*N/K*).

$$E = \frac{\frac{n}{k}}{\frac{N}{K}}.$$

- *Gene Set Enrichment Analysis (GSEA)*: GSEA provides a statistical framework in order to test whether a given set of genes is up or down-regulated in accordance to a given condition. The most used GSEA approach is called *Generally Applicable Gene Enrichment (GAGE)*, where a meta-analysis (for the combination of different observations) is used as main approach. Given the gene set (S), the background list (B) and two individual samples, the mean and variance of all the fold changes contained in each sample are compared with a Student's t-test, and this is repeated for all the pairwise comparisons. The process allows the measurement of the mean of log(pvalues), shown here,

$$x_j = -\frac{1}{n}\sum_{i=1}^{n}\log\left(p_{ij}\right)$$

and a new variable is obtained, which follows a Γ distribution, and corresponds to the enrichment of the gene set S.

$$P(y > X) = \Gamma_{m,1}(y > X)$$

3.1.4. Validation of the results

Differential expression detected by transcriptomic analyses may be validated using an independent technique, for example, quantitative PCR (qPCR), which is widely used and statistically assessable (Fang and Cui, 2011). Independent validation may entail the analysis of the same RNA samples analyzed by RNA-seq (technical validation) or of a new set of samples (biological validation).

In qPCR, expression of the genes of interest is quantified relative to the expression of a control (so called house-keeping) gene. The measurement by qPCR refers to the amplification cycle when the signal becomes detectable. However, qPCR specifically targets amplicons smaller than 300 bp, usually toward the 3′ end of the coding region, avoiding the 3′ untranslated region (3′UTR) (Ramsköld, Wang, Burge and Sandberg, 2009) and therefore cannot be used to asses differential transcript isoforms. A key disadvantage of qPCR is that is relies on the arbitrary selection of the normalizer genes and erroneous selection of nonstable normalizers invalidates all downstream analyses. The quantification of multiple control genes along with the genes of interest produces a

more stable reference within a biological context as reference genes may be stable under certain conditions but not others (Valdesompele et al., 2001).

RNA-seq experiments normally suggest some functional implication of differentially expressed genes or gene sets in the biological phenomenon under study. Assessing this assumption through experimental manipulations of the expression of some key genes (functional validation) has become an almost mandatory step in transcriptome analysis. Observed gene expression patterns may be functionally linked to a phenotype by an independent knock-down/rescue study in the cells or organism of interest. Since these experiments are typically expensive and time demanding, they currently represent the bottleneck of transcriptomic studies.

3.2. Proteomics

3.2.1 Generalities

Proteomics is the study of protein products expressed by the genome, and has become one of the leading technologies in the post-genomic era due to the central role of proteins and protein–protein interactions in the cell (Tyers and Mann, 2003; Cox and Mann, 2007). The term proteomics for the global study of proteins was first coined in the early 1990s by Marc Wilkins and can be equally used with respect to a whole organism, a tissue, a biological fluid, a cell, or even an organelle (Wilkins et al., 1996a,b). Proteomics is only made possible by the coordinated integration of many fields of scientific endeavor. Most importantly, but not exclusively, these include genome sequencing, protein separation science (and protein biochemistry in general), mass spectrometry (MS), and bioinformatics, and these have been described as the four pillars on which proteomics technologies stand (Tyers and Mann, 2003).

There is growing attraction to the concept of comprehensive or system-wide analyses of proteins and the 2000s have seen tremendous advances in proteomics technologies and the quality of data produced to address this question. These technologies offer considerable opportunities for improved biological understanding of a particular system in health or disease (see Cravatt et al., 2007 for some notable examples). In addition, another major potential application of proteomics is biomarker discovery and considerable resource has been directed in this direction as well over the past decade.

These advances in proteomics have taken place in coincidence with tremendous progress in all areas of functional genomics (understanding the function of genes and their corresponding proteins on a global scale). Mass spectrometry (MS) is the central technological basis of proteomics and this is

particular true of protein identification and quantification (Aebersold and Mann, 2003; Yates et al., 2005; Domon and Aebersold, 2006).

The first applications MS related to the identification of specific proteins but soon extended to the quantification of protein abundance. Later applications include the analysis of post-translational modifications (PTMs, such as lipid modification, glycosylation, phosphorylation, and many others), proteolytic processing, and association with other proteins or different types of molecules. All of these components are crucial for the description of biological systems but also significantly complicate the experimental analysis of proteins (Aebersold, 2003).

Protein discovery is a major opportunity afforded by proteomics technologies. This might apply to biomarker discovery in clinical contexts or the identification of new proteins of interest in basic science projects. Such protein discovery approaches also offer clear potential for the generation of new hypotheses. The technology is also suited to asking much more defined questions of a handful or even individual proteins. Protein discovery studies might arbitrarily be categorized as being comprehensive, broad scale, or focused. Comprehensive approaches are typically qualitative and are focused on the enumeration of the components of a biological system. A good example of this is the Human Proteome Organization (HUPO) Plasma Proteome Project (PPP), which aims to identify every protein in human plasma (Omenn et al., 2005). Proteomics technologies can also be used to assay specific known entities. An example of such an experiment would be the measurement of biomarker components in plasma within a clinical setting (Hoofnagle et al., 2008; Keshishian et al., 2009). These assay approaches can also be used for biomarker candidate validation and might be used for individual peptide quantification in a range of studies (Mallick and Kuster, 2010).

The field of proteomics spans a wide range of research topics and distinct approaches need to be applied depending on the question being asked. These vary widely in their versatility, technical maturity, and difficulty and, as a consequence, some questions are much harder to answer than others. Here we describe approaches for protein identification and quantification that are widely used today and reflect on the tremendous progress that has been made in these areas of analysis in the past decade.

3.2.2 Workflows for proteome quantification

Proteins are derived from entire cells or from biological fluids and the totality of the proteins can be processed (**Figure 12**). However, it is often preferable to perform subcellular fractionation to enrich for proteins of particular biological interest and achieve localization information. One further option with biological fluids is protein depletion in order to more readily analyze the lower abundance

proteins. Protein solubilization is usually performed in a single step but differential detergent solubilization is possible using detergents with different properties. Proteins can be separated by two-dimensional electrophoresis (2DE). In this workflow, an individual protein "is removed as a gel plug, trypsin digested and the resulting peptides are typically separated based on relative hydrophobicity by nanoscale liquid chromatography (LC) before tandem mass spectrometry (MS/MS)" (Labbadia and Morimoto, 2015).

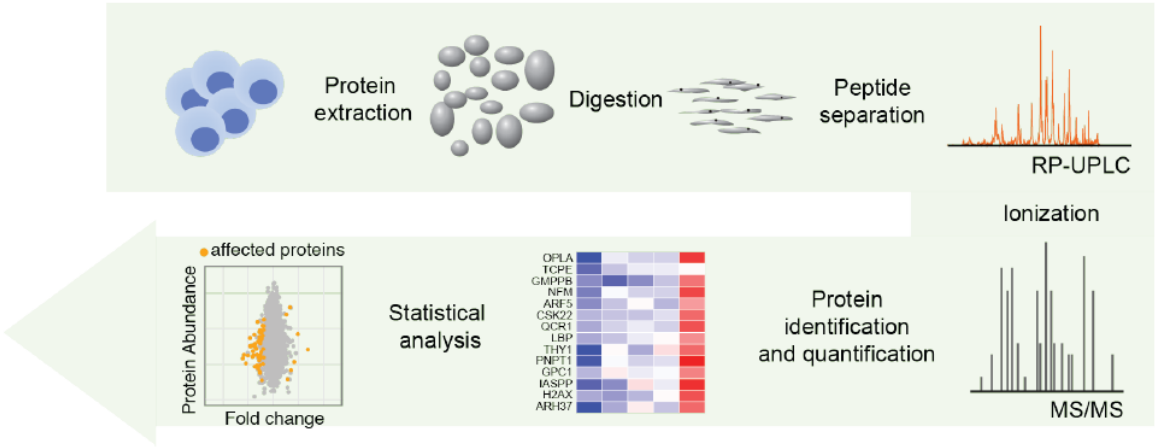


Figure 12. Typical proteome quantification workflow. A typical MS experiment consists in extraction of proteins (from cells or tissues), protein digestion, peptide separation, ionization, protein quantification and final computational analysis.

The resulting MS data is matched to existing protein databases; the method of choice for 2DE protein quantification is difference gel electrophoresis where samples are labeled with different fluorescence substances before mixing together and 2DE. For global analysis we rather digest with trypsin the entire solubilized protein mixture. Peptides are then separated by liquid chromatography on their hydrophobicity and often charge as a multidimensional separation. Then a first MS step analyzes m/z rations in a mixture of tryptic peptides and then individual peptides are broken down to the single amino acids in the second MS step. Database searches are then performed to assign the peptide to described proteins. Based on the intensity of the signal derived from peptides of a given protein, protein abundance in the original sample can be estimated. Label-free MS quantification can also be performed using spectral counting or ion current measurements. Quantifications can be performed using internal standards in the form of labelled peptides or by metabolic labelling of the whole proteome (stable isotope labelling by aminoacids in cell culture, SILAC). The most developed technique involves the multiplexing of peptides labelled with isobaric labels (tandem mass tag, TMT)

that co-elute during the first MS step and can be deconvoluted in the second MS step. This latter technique is a prominent source of the data analyzed in the present thesis and is described in detail.

3.2.3 The TMT10 technique

Several labeling techniques can be used to measure the relative abundance of proteins in different biological samples by mass spectrometry. Two general labeling strategies—*in vivo* and *in vitro* labeling—have been widely applied to quantitative protein analysis. The first strategy normally uses heavy isotopes, which are then incorporated into the proteins of the organisms: this step is performed through feeding in the medium or the food (Oda et al., 1999). This approach has limitations: it requires a large amount of growth medium, and there are difficulties in the control of labeling efficiency (Thompson et al., 2003). *In vitro* labeling passes these limitations by chemically labeling any peptide sample with high efficiency: an example is the isotope-coded affinity tag (ICAT) technique (Gygi et al., 1999). With this method, probes that are the same on the chemical point of view, but with different masses, are used to label two peptide samples. The relative abundances are then determined from the intensities of MS1 peaks that are distinct in their mass to charge ratio (m/z). Other techniques started using isobaric tags, such as TMT and iTRAQ (Thompson et al., 2003; Ross et al., 2004): the reagents used are normally composed of a mass reporter, a mass normalizer and an amine reactive group. The first two parts are important for the incorporation of stable isotopes in multiple configurations such that each mass reporter's mass can be seen in a MS2 spectrum. The intact mass of each isobaric tag variant, however, is the same. With this labeling strategy, digested peptides from multiple samples are first labeled in parallel, then they are then mixed and analyzed using reversed phase high performance liquid chromatograph (HPLC) coupled with a mass spectrometer capable of tandem MS analysis (**Figure 13**).

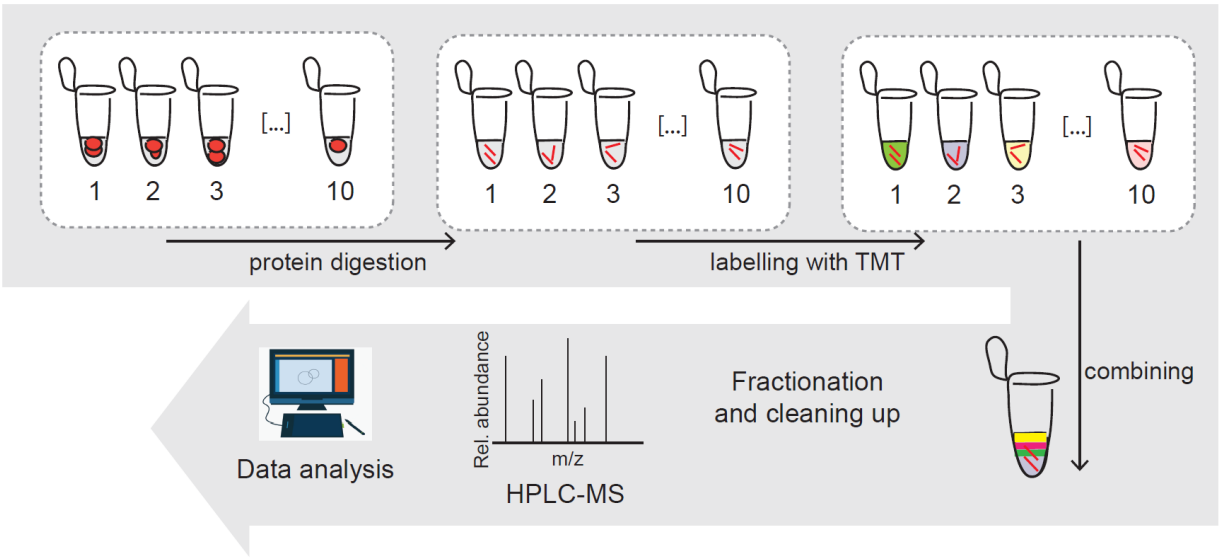


Figure 13. Workflow of TMT technique. After the first step of protein digestion, peptides are labelled and then all the samples are combined in one tube; after fractionation and cleaning a mass spectrometry workflow is followed and the data are analysed.

The mass reporter can be divided from the peptide through collision-induced dissociation (CID) or through higher energy collision dissociation (HCD) before detection; the relative quantification of the peptide is measured as the intensity of the reported ions in the MS2 spectrum. The isobaric and chemically identical nature of isobaric tags ensure that "identical peptides labeled with different tag variants will have the same chromatographic elution profile, and experience identical ionization processes in the mass spectrometer" (Zhang and Elias, 2017), with a more accurate final quantification. Another important strength of this technique is the improvement of the signal-to-noise ratio by enabling quantification at the MS2 level, in comparison to the MS1 scan of the peptides. However, in highly complex peptide mixtures, co-isolation and co-fragmentation of multiple ions is increasingly likely to occur with MS2, resulting in distorted ratios of isobaric tag reporter ions (Ting, Rad, Gygi and Haas, 2011). Sometimes, when working with complex peptide mixtures, the co-isolation and co-fragmentation of multiple ions normally happens with MS2, with the subsequent distortion of isobaric tag reporter ions: in this case, the ratios can't reflect anymore the true proportions of a selected peptide precursor's components. Combining ion trap and orbitrap technologies measures reporter signals with a third stage of ion isolation and fragmentation (MS3), reducing the ratio distortion resulting from interfering ions (Ting, Rad, Gygi and Haas, 2011). Another strategy, called "MultiNotch MS3" method, was shown to increase detection sensitivity after selecting multiple fragment ions from MS2 to enhance reporter ion intensities at the MS3 level (McAlister et al., 2014).

3.2.4 Thermal Proteome Profiling

In this thesis the Thermal Proteome Profiling (TPP) technique was adopted to highlight potential differences in protein stability between young and old animals: this method focuses on the differential stability of proteins after a heat shock (assuming that proteins become more resistant to heat-induced unfolding when in complex with a ligand), can be applied on live cells and doesn't require labelling.

As shown by the principle of thermal shift assays, when proteins are subjected to a heat shock they generally irreversibly unfold (exposing their hydrophobic core and subsequently aggregating); however, their melting temperature (i.e. the temperature at which unfolding happens) can be increased after the binding with a ligand, as shown by structural biology (Vedadi et al., 2006; Ericsson et al., 2006) and drug discovery fields (Pantoliano et al., 2001; Senisterra et al., 2006).

In particular, the TPP technique combines this principle with mass spectrometry-based proteomics: the workflow of the experiment generally consists of (1) preparation of the cells for the experiment, (2) heating procedure, (3) extraction of soluble proteins, (4) protein digestion with peptide labelling, (5) mass spectrometry and (6) data processing (**Figure 14**).

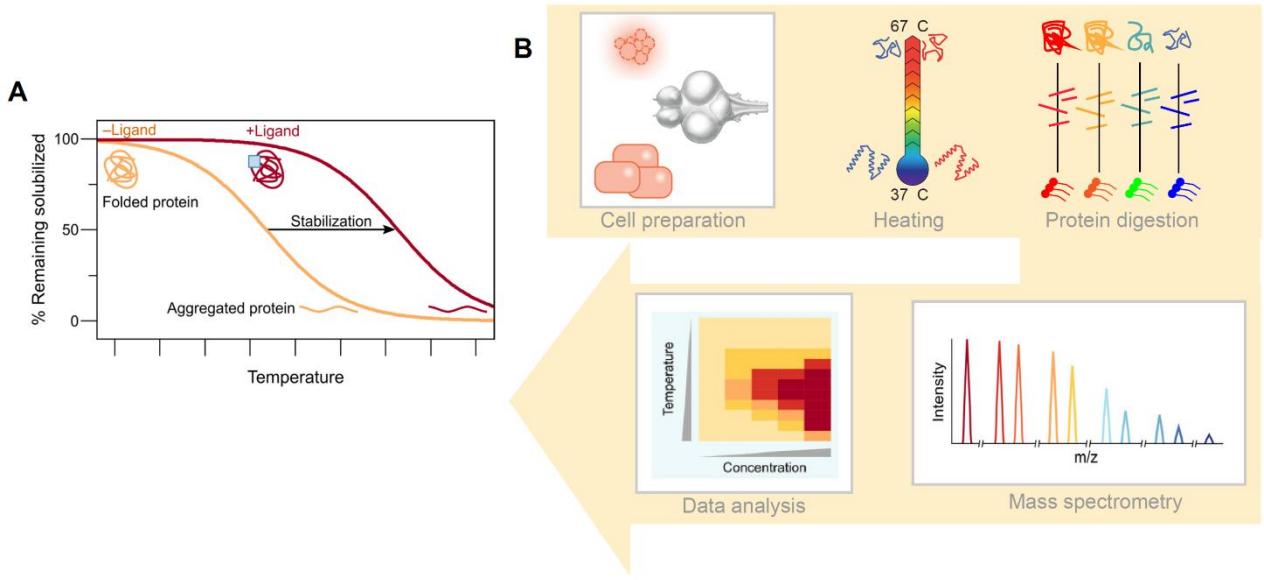


Figure 14. (A) Principle of the thermal shift assay (adapted from Mateus, Määttä & Savitzki, 2017). (B) General workflow of the Thermal Proteome Profile (TPP) technique.

TPP can be performed on cells extracts, intact cells or tissues: cell lysis (prior to TPP) dilutes metabolites, proteins and co-factors, in order to stop the metabolism of the cell (so that stabilization of the protein should just be the result of binding to the ligand). After this preparation, cells can be

incubated with a drug in order to identify candidates: the choice of using a single compound concentration or a range of concentrations is linked with the next step, the heating process.

Following the heat treatment cells are lysed and proteins that are denatured (and aggregated) are removed using ultracentrifugation (Savitzki et al., 2014); new protocols have shown that the use of mild detergents results in the inclusion of membrane proteins in the analysis without affecting heat-induced aggregation. Once the soluble proteins are collected, they are digested following a normal proteomics workflow (**Figure 12**) and the resulting peptides from each condition are labelled using TMT and combined into a single sample for the mass-spectrometry analysis (**Figure 13**).

4. *N.furzeri* as a model for computational genomics approach to ageing

4.1. Genome-related studies

In the genomic era the availability of a genome reference sequence is really important for the recognition of an organism as model for scientific research. The first insights were provided by cytogenetics and genome survey Sanger sequencing (Reichwald et al., 2009); the *N.furzeri* genome contains 19 chromosomes and was estimated to be extremely repeat rich, particularly in satellite sequences (**Figure 15**). Crosses of the shortest-lived strain GRZ and more recently wild-derived strains (already mentioned in paragraph 2.3) provided the first genetic maps of the *N.furzeri* genome and lifespan controlling quantitative trait loci (QTLs; Kirschner et al., 2012), showing that lifespan determination is polygenic. In particular, four QTLs regulating lifespan were identified, located at linkage groups (LG) 9, 11, 14 and 17, starting from 22 LGs. Both studies identified males as the heterogametic sex, concordant with an XY sex-determining (SD) system.

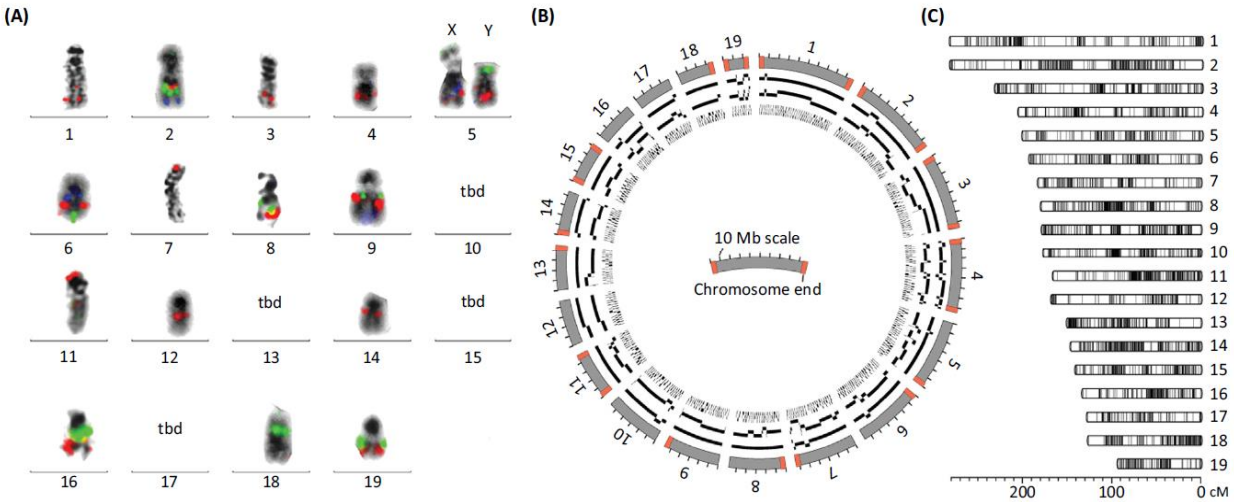


Figure 15. Genome architecture of *N.furzeri* (from Platzer & Englert, 2016). (A) Composite male GRZ karyotype. (B) Stepwise assembly of the 19 sgrs of the reference, with scaffolds obtained by sequence assembly (inner circle), super scaffolds built on integration of optical map (second circle), genetic scaffolds generated by linkage map integration (third circle), and sgrs defined on analyses of synteny in medaka and stickleback (outer circle). (C) High-density restriction-site-associated DNA (RAD) map of the 19 linkage groups.

Two independent *N.furzeri* genome assemblies were performed and published in parallel, thanks to the new possibilities given by modern sequencing technologies (Reichwald et al., 2015; Valenzano et al., 2015). In the first study, "to reach chromosome-scale long-range continuity, a five-step strategy was used compromising sequence assembly, scaffold/gap filling, integration of optical and genetic linkage maps, and finally comparative synteny mapping in two closely related fish species" (Platzer & Englert, 2016). The second study used RNA-seq and a high-density restriction-site-associated DNA (RAD) linkage map; this strategy resulted in an improved contiguity and the assignment of sequence scaffolds to chromosome-scale linkage groups.

The two studies succeed in finding genomic regions enriched in aging-related genes with these two complementary methods; the first used the long-range contiguity of the reference sequence and performed a genome-wide positional gene-enrichment analysis for differentially expressed genes (DEGs) in *N.furzeri* strains with different lifespan rates, detecting seven genome-wide positional regions. On the contrary, the second study made use of QTL mapping crossing short- and long-lived strains, identifying one genome-wide significant lifespan QTL (which was shown to be located on the sex chromosome, consistently with the previous findings reported by Kirschner et al., and was found to be enriched for known aging-related genes).

Both studies also searched their protein-coding annotations for signs of positive selection, in order to identify candidates which could lead to the short lifespan of this species, with 497 *N.furzeri* genes with at least one site under positive selection identified by Valenzano et al., and only seven in *N.furzeri* and one in *N.pienarii* (which shows convergent evolution of very short lifespan) by Reichwald et al.

The two studies focused on substitutions observed along phylogenetic branches of very different lengths; Valenzano et al. made use of a branch of 50-70 million years, while Reichwald et al. of a branch of <1 million years.

Currently the reference sequences of the *N.furzeri* genome are provided online by two dedicated browsers (http://www.nothobranchius.info/NFINgb and http://africanturquoisekillifishbrowser.org), and efforts are under way to integrate these resources and maintain an improved genome assembly at the Ensembl and UCSC genome browsers.

4.2. Transcriptome-related studies

The transcriptome of *N.furzeri* was recently sequenced (Petzold et al., 2013); using Sanger, 454/Roche and Solexa/Illumina technology more than 19,000 protein-coding genes with sequence homology to other vertebrate species were identified.

This catalogue was used as reference for quantifying the expression of protein-coding genes in the ageing brain (Baumgart et al., 2014): comparison with human datasets showed that down-regulated genes which are conserved code for proteins which are related to the RNA processing and protein biosynthesis categories, and, on the other hand, that up-regulated genes code for ribosomal and lysosomal proteins. On another level, results from *N.furzeri* were compared to human, rodent and primate tissues, with up-regulation of cytosolic ribosome and lysosome categories, and down-regulation of mitochondrion and collagen categories (Baumgart et al., 2014). Down-regulated genes showed to have different temporal profiles: neurogenesis and extracellular matrix-related genes displayed a rapid decrease, while genes coding for axonal and synaptic proteins showed a progressive decrease (Baumgart et al., 2014; Baumgart, Priebe, Groth et al., 2016). Network analysis showed extensive co-regulation of genes related to cell-cycle and DNA synthesis categories with the zinc-finger protein ZNF367 as a central hub (Baumgart et al., 2014; Naef et al., 2018); the same gene showed to be also strongly downregulated during ontogeny in the prefrontal cortex in human (Colantuoni et al., 2011). Later studies in *Xenopus* demonstrated that this gene is causally involved in neurogenesis.

The short lifespan of *N.furzeri* allowed also longitudinal studies in order to investigate molecular pathways involved in lifespan and to correlate quantitative variations in gene expression during early adult life with lifespan (Baumgart, Priebe, Groth et al., 2016); individual *N.furzeri* of different

lifespans differ in their transcript levels at an early adult age and the rate of age-dependent gene modulation is the lowest in longest-lived individuals, suggesting that they are characterized globally by a slower aging rate.

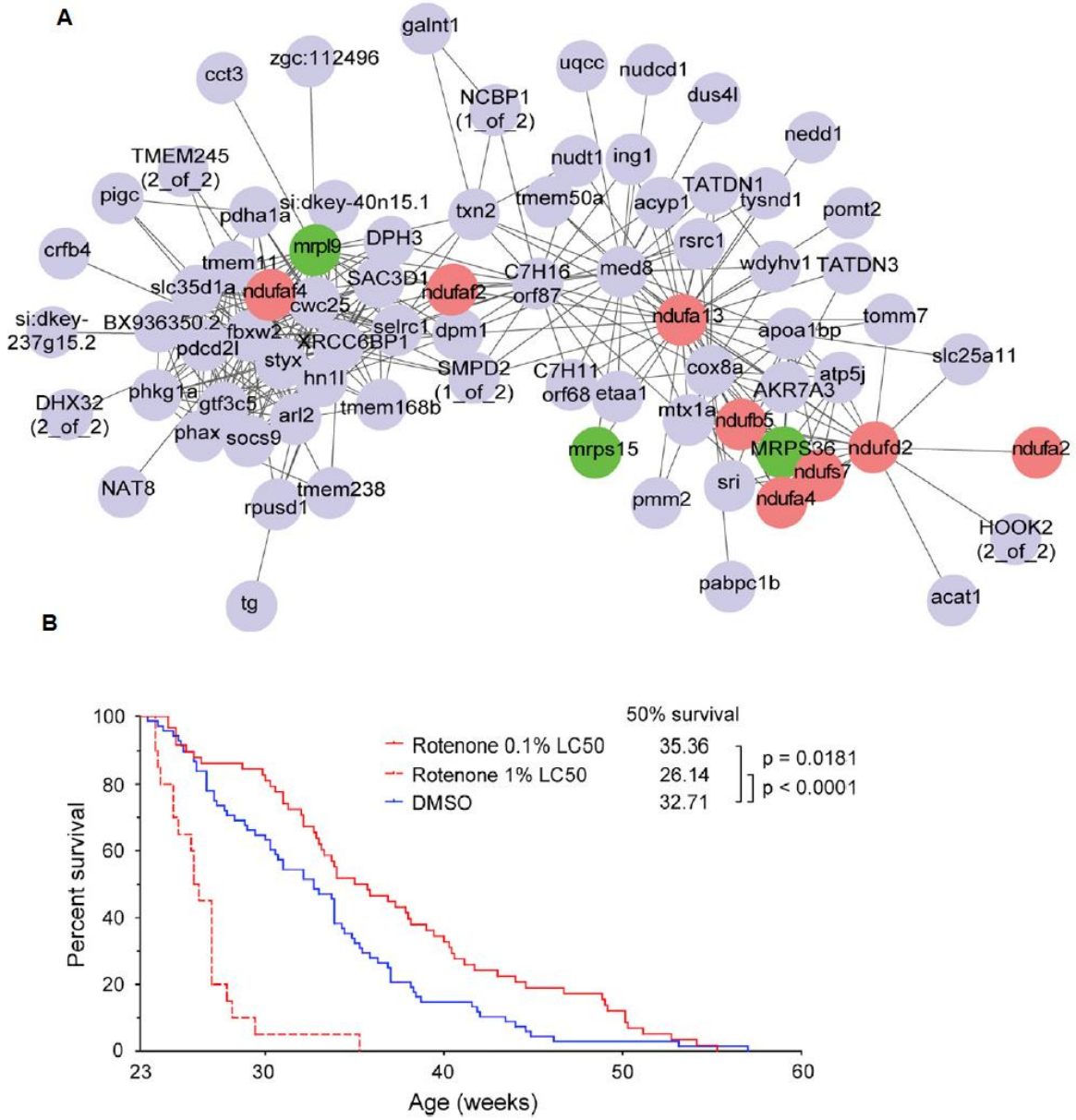


Figure 16. Complex I as central regulator of lifespan (adapted from Baumgart, Priebe, Groth et al., 2016). (A) Visualization of the module with the highest correlation with age of death, obtained through network analysis (with the WGCNA method): in green are highlighted mitochondrial ribosome subunits, in red subunits of Complex IV. (B) Effects of rotenone on *N.furzeri* lifespan: animals were treated with DMSO, as well as rotenone at two concentrations (0.1% and 1%).

Baumgart, Priebe, Groth et al. identified also Complex I of the respiratory chain as a central hub for the regulation of lifespan: the identified module showed the highest negative correlation with individual lifespan and contained both mitochondrial ribosomal as well as Complex I genes as hubs

(**Figure 16**). The consequential treatment with Rotenone (a selective inhibitor of Complex I) at low doses resulted in a significant prolongation of *N.furzeri* lifespan, and RNA-seq on rotenone-treated vs. control animals demonstrated that rotenone fights the effects of aging, suggesting Complex I as a new potential target of age-related dysfunctions.

4.3. MiRNome-related studies

MicroRNAs are highly conserved entities and can be easily quantified also when a reference genome is absent; analysis of microRNAs in the aging brain of *N.furzeri* showed down-regulation of oncogenic miRNAs, and the up-regulation of tumour suppressor microRNAs. In particular, it has been observed that age-regulated microRNAs are part of a gene regulatory network centred on the antagonistic action of the tumour suppressor tumour protein 53 (also known as TP53) and the prototypic oncogene v-myc avian myelocytomatosis viral oncogene homolog (MYC); microRNAs that are upregulated show positive interactions with TP53 and negative interactions with MYC, and down-regulated miRNAs follow the opposite pattern (Baumgart et al., 2012), hypothesizing that they might be involved in controlling the neurogenic activity of neuronal stem cells.
In addition, age-dependent regulation of microRNAs appeared conserved between *N.furzeri*, rodent and primates (Baumgart et al., 2012), with wide deregulation in the GRZ strain and 5 wph expression pattern more similar to the MZM 11 wph expression pattern, indicating that the expression profiles of the two strains diverge prior to sexual maturity (Baumgart et al., 2012).
In a recent study (Ripa et al., 2017), microRNA-29 family members (which are known to be upregulated in the central nervous system of different vertebrates) where shown to be actively transcribed in *N.furzeri* and upregulated with age in neurons. In particular, the pri-miR-29-2 genomic cluster showed the largest age-dependent up-regulation (with an increase in expression of more than 15 times between 3 and 27 weeks of age), and miRNA-29 targets displayed an excess of negative correlation with miR-29 expression, suggesting that it could be a regulator of ag-dependent gene expression (as previously reported in the prefrontal cortex of humans and macaques). Neuronal-specific deletion of miR-29 led to an induction of the aging phenotype (both at the gene expression and histological level), highlighting the possibility that miR-29 upregulation could be an evolutionary conserved aging signature.

Aim of the work

Large datasets of RNA-seq and mass-spectrometry-based proteomics were generated in *N. furzeri*. The focus my research was an in-depth analysis of these datasets to get novel insights into the biological mechanisms controlling aging.

My thesis is divided into three different sections, each dealing with a specific dataset aimed at capturing a particular aspect of killifish aging.

In the first part, a proteomic approach was used to unravel the molecular features of *N. furzeri* brain aging. A complete characterization of the *N. furzeri* brain proteome lacks in literature and different techniques were employed in order to obtain complete information on protein expression, protein stability and protein aggregation. The proteome regulation was also compared to transcriptome regulation in order to quantify the impact of post-transcriptional mechanisms during aging.

In the second part, a comprehensive RNA-seq dataset from two strains differing in lifespan was analyzed to identify a multi-tissue transcriptional signature of lifespan differences. These results were compared with predictive factors identified in a longitudinal RNA-seq dataset using Cox-Hazard Model and with RNA-seq data obtained in the naked mole rat in order to identify conserved aging factors.

In the third part, the same computational approach was used in order to compare the transcriptomic correlates of aging in wild animals and captive populations highlighting both common and condition-specific pathways in brain aging.

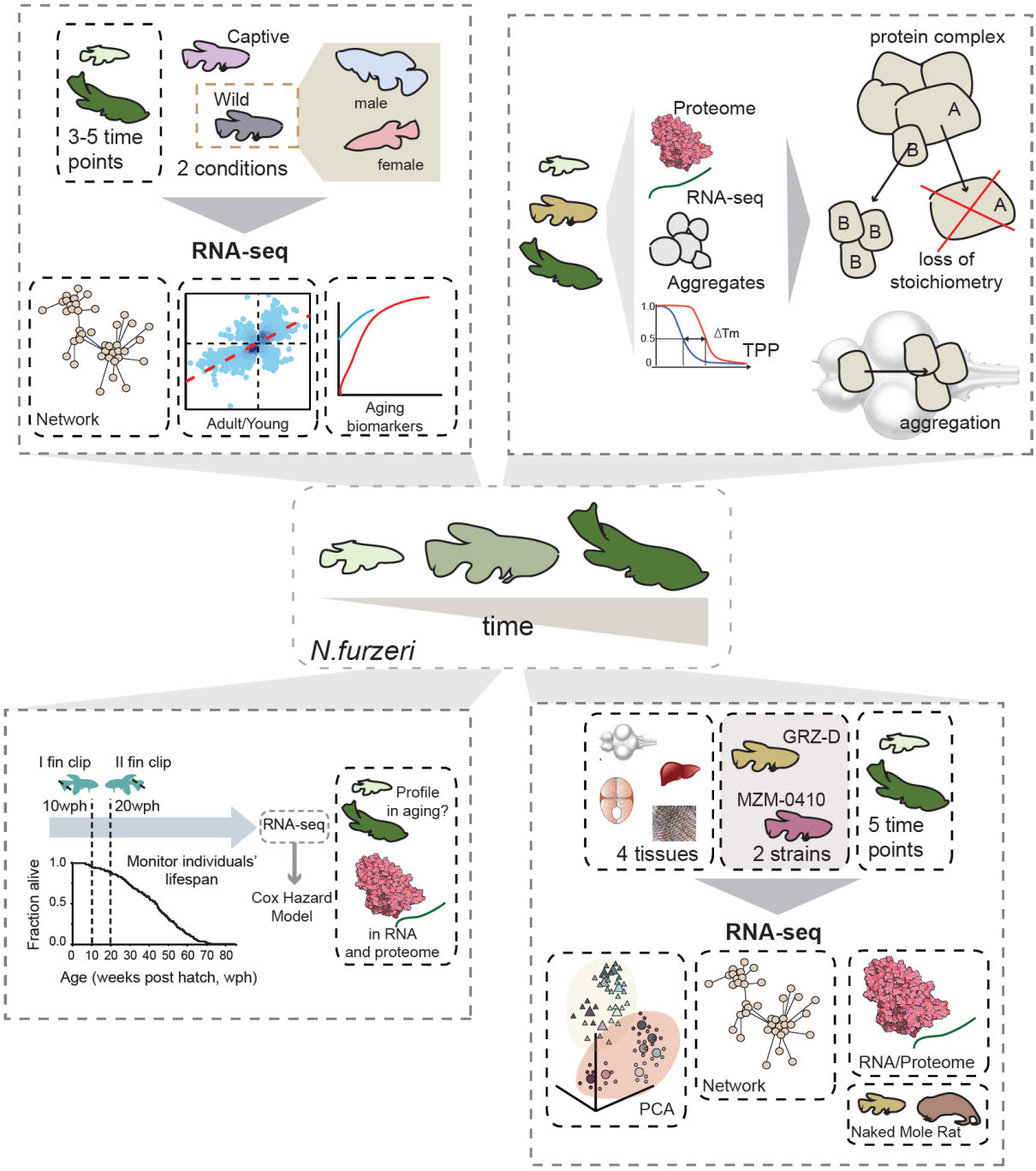


Figure 17. Workflow of the thesis. *Nothobranchius furzeri* aging was investigated using different types of datasets; each project is contained in a single box.

List of datasets used

Type	Description	Variables	N	Public yes/no	GEO/PRIDE number	PMID
RNA	JenAge dataset, *N.furzeri* aging (MZM)	5 tissues, 5 time points per tissue	75	YES	GSE52462	25059688
RNA	JenAge dataset, *N.furzeri* aging (GRZ)	5 tissues, 5 time points per tissue	75	NO	none	none
RNA	*N.furzeri* aging (MZM), independent dataset	2 tissues, 4 time points per tissue	39	NO	none	none
RNA	*N.furzeri* aging (GRZ), independent dataset	2 tissues, 4 time points per tissue	54	NO	none	none
RNA	Longitudinal *N.furzeri* dataset	2 fin clips	45	YES	GSE66712	27135165
RNA	Longitudinal *N.furzeri* dataset	2 fin clips	107	NO	none	none
RNA	Wild *N.furzeri* data	Brain, females and males, 3 time points each sex	22	NO	none	none
RNA	Naked Mole Rat/Guinea Pig data	2 tissues, young/old status	-	YES	GSE98744	30068331
smallRNA	JenAge miRNA dataset, *N.furzeri* aging (MZM)	3 tissues, 5 time points per tissue	75	YES	GSE92854	28874118
smallRNA	JenAge miRNA dataset, *N.furzeri* aging (GRZ)	3 tissues, 5 time points per tissue	75	YES	GSE92854	28874118
smallRNA	*N.furzeri* brain aging (MZM), independent dataset	Brain, 3 time points	12	NO	GSE125373/PXD012314	none
Proteome	*N.furzeri* aging (MZM), TMT-10	Brain, 3 time points	12	NO	GSE125373/PXD012314	none
RNA	*N.furzeri* brain aging (MZM), independent dataset	Brain, 3 time points	12	NO	GSE125373/PXD012314	none
Proteome	Thermal Proteome Profiling data	Brain, young/old status	4 pools	NO	GSE125373/PXD012314	none
Proteome	Aggregates	Brain, young/old status	-	NO	GSE125373/PXD012314	none

Materials and methods

1. Transcriptome - proteome comparison

1.1. Abundance comparison

To obtain the correlation between transcriptome and proteome for the different time points normalized counts were used. RPKM were used for transcripts, as raw counts normalized on transcript length, and IBAQ were used for proteins, as total intensities divided by the identified peptides for one protein. Spearman correlation was measured for every sample, then the correlations values were plotted as boxplot. For additional information take a look at the related R Markdown document (Supplementary Methods 1).

1.2. Differential expression analysis

For transcriptomic differential expression analysis DESeq2 package was used (Love MI, Huber W, Anders S. et al., 2014): raw counts were used to obtain DEGs (differentially expressed genes), and then the results were filtered on padj (adjusted pvalue) < 0.05. For the visualization of the different comparisons the intersection of the differentially expressed case was showed, since only genes differentially expressed in both the comparisons (with p adjusted < 0.05) were considered.

1.3. Metanalysis

For GO analysis (and also GAGE) another approach was used and metanalysis was applied in order to combine the results from the different comparisons. In particular Fisher's Method was performed (also known as Fisher's Combined probability test), where pvalues from each test where combined into one statistic using the formula:

$$X^2_{2k} \sim -2\sum \ln(p_i)$$

The resulted pvalue was then adjusted using FDR and significant cases were then isolated for adjusted pvalue < 0.05.

2. Enrichment analysis

2.1. GO analysis

Gene Ontology (GO) analysis was performed using the WebGestalt online tool, and in particular the 2017 version (Zhang, B., Kirov, S.A., Snoddy, J.R. et al., 2005): only Over-Representation Analysis (ORA) was performed, thus using the gene list with the relative background. Enriched categories where identified with KEGG and GO Process and filtered for FDR < 0.05, and then used for visualization as barplots.

2.2. Generally Applicable Gene-set Enrichment (GAGE)

GAGE was performed using the gage package (Luo, Weijun, Friedman, Michael, Shedden, Kerby, Hankenson, Kurt, Woolf, Peter, 2009): as the analysis required Entrez IDs also the biomaRt package was used to obtain them from Gene Symbols IDs (Durinck S, Spellman P, Birney E, Huber W, 2009). The enrichment analysis was performed using the *H.sapiens* background and enriched categories were filtered for qvalue < 0.05. GO process-related categories were considered for plotting with the Revigo tool (Supek F, Bošnjak M, Škunca N, Šmuc T., 2011).

3. miRNA analysis

In order to access posttranscriptional-dependency in *N.furzeri* proteomic data two independent analyses were performed and then combined. On one hand miRNAs affected by aging were detected for the two aging comparisons (as explained in the "Results" section) using DESeq2 package, filtering significant cases for padj < 0.05. On the other hand proteins affected by aging for the two comparisons were divided into different categories: (a) cases where the transcript was not detected; (b) cases where the transcript was significantly concordant to the protein; (c) cases where the transcript was concordant to the protein but not significant; (d) cases where the transcript was significantly discordant to the protein or not affected. miRNAs affecting proteins contained in the last category where selected, and then among these only the ones affected by aging (upregulated) where considered.

4. Loss of stoichiometry analysis

2.1. Complex analysis (Ori et al.)

Complex analysis already reported by Ori at al. (2016) was used to access loss of stoichiometry during aging in *N.furzeri* brain proteome: to measure stoichiometry changes a protein-complex based normalization was performed (using the median complex abundance) and a manually-curated complex resource was used. Significant cases were filtered on padj < 0.05 and absolute(logFC) > 0.5 and complexes with more than 2 members affected by loss of stoichiometry considered.

2.2. IQR analysis

In order to access differences in complex stoichiometry another strategy was also used and the InterQuantile Range (IQR) for each complex was obtained from the different distributions using the formula:

$$F^{-1}(0.75) - F^{-1}(0.25)$$

In order to perform the analysis logFoldChanges obtained for the Limma output were used and then IQRs for the different comparisons were plotted as boxplot, removing all the values equal to 0.

5. Principal Component Analysis (PCA)

For the visualization of the data PCA was performed using the prcomp() function in order to examine the covariance/correlations between individuals (through the singular value decomposition). Normalized values (as rlog obtained with DESeq2) were used (filtered for different feature selections) using the formula:

$$y = \log_2(n + n_0)$$

(where n_0 is a positive costant)

and the Principal Components were plotted as 3D-PCA using the rgl() package. Centroids were calculated as arithmetic mean of the replicates of a group and plotted together with the single values. For additional information take a look at the related R Markdown documents.

6. Network analysis

6.1. Consensus analysis on multiple datasets

Network analysis was performed using Weighted Gene Co-expression Network Analysis (WGCNA) method (Langfelder P., Horvath S., 2008).

Network analysis was performed through different steps:

- Setting of the soft threshold, coefficient necessary for the adjacency matrix construction, as shown in the formula:

$$a\mathbf{ij}=|cor(x\mathrm{i},x\mathrm{j})|\beta$$

- Adjacency matrix and TOM (Topological Overlap Matrix), defined as:

$$TOMij=(\sum aiu*auj+aij)/\min(ki,kj)+1-aij$$

- Hierarchical clustering and modules detection after measuring the module eigengenes; every module is characterized by a color (as the module which has been studied for the analysis, defined by the turquoise color)
- Module-trait relationship table construction, as correlation between single gene expression and external trait (in this case aging/development)
- Module membership plot (as correlation between single gene expression and module eigengene)
- Visualization with Cytoscape software (Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T., 2003).

7. Survival analysis

In order to isolate predictive factors of aging survival analysis was performed to correlate lifespan (as age of death of the animals) to gene expression. Cox-Hazard Model (also called cox regression) was applied using the formula:

$$h_j(t) = exp\ (\alpha + \beta_j x_{i1} + \beta_j x_{i2\ +\ \ldots\ +} \beta_j x_{ik}) = {}_e\alpha \text{ x } {}_e\beta_1 x_{i1} \text{ x } {}_e\beta_2 x_{i2\ \ldots} \text{ x } {}_e\beta_k x_{ik}$$

The analysis was performed separately for 10 wph and 20 wph samples: raw counts (obtained from the longitudinal experiment shown in Baumgart, Priebe, Groth et al., 2016) were considered and significant factors were filtered for pvalue < 0.05, then detrimental and beneficial factors were selected on the basis of the coefficient, where the first cases had $\beta>0$ and the second cases had $\beta<0$.

8. RNA isolation

8.1. Protocol

RNA isolation was performed using the TRIZOL protocol of extraction from brain *N.furzeri* tissues. First of all tissues were homogenized in 1 mL TRIZOL reagent (per 50-100 mg of tissue) using a power homogenizer; then, for the separation of the phases, 200 uL of chloroform (for each 1 mL of TRIZOL) were added, vortexed vigorously for 30 seconds and incubated at RT for 2-3 minutes. The samples were then centrifuged at 20000 x g for 20 minutes at 2-8°C. To precipitate the RNA a reaction containing isopropyl alcohol, sodium acetate and GlycoBlue (in order to see the pellet) was prepared, then the acqueous phase (upper one) was taken and mixed to the reaction. The samples were then incubated at 15-30°C for 10 minutes and centrifuged at 20000 x g for 30 minutes. After removing the supernatant the pellet was washed twice with 1 mL of 80% ethanol, vortexed and centrifuged at 7500 x g for 5 minutes at 2-8°C, and finally, after removing all the ethanol, the samples were left to dry (no more than 5-10 minutes) and resuspended in 30 uL of Nuclease-free Water.

8.2. Recipes

Reagent	Recipe	Storage
TRIZOL reagent	-	RT (under the hood)
RNA precipitation mix	Isopropyl alcohol (550 uL), sodium actate (80 uL), GlycoBlue (1 uL)	Fresh-made

9. Retro-transcription

9.1. Genomic DNA removal

Removal of genomic DNA was performed using the DNase I protocol from Thermo Scientific. An amount of 1 ug of RNA was used to which 10X Reaction Buffer (with $MgCl_2$), DNase I and Nuclease-free Water were added to have a final volume of 10 uL per sample: then the reaction were incubated at 37°C for 30 minutes and, after adding EDTA, were incubated again for 10 minutes at 65°C.

9.2. Retro-transcription

For the second step of the experiment the Superscript IV First-Strand cDNA synthesis reaction protocol from Invitrogen was used. To anneal the primers to the template RNA a reaction containing 1 uL of 50 uM random hexamers, 1 uL of 10 mM dNTPs, template RNA (10 pg – 500 ng) and water was prepared to the final volume of 14 uL for each sample, and the primer-RNA mix was heated at 65°C for 5 minutes. After incubating the reactions in ice for at least one minute the RT reaction mix was prepared, to which the RNA was added. The reaction was so incubated at 23°C for 10 minutes, at 50-55°C for 10 minutes and at 80°C for 10 minutes.

9.3. Recipes

Reagent	Recipe	Storage
10X reaction buffer with $MgCl_2$	100 mMTris-HCl (pH 7.5 at 25°C), 25 mM $MgCl_2$, 1 mM $CaCl_2$	-20°C
RT Reaction Mix	5X SSIV Buffer (4 uL), 100 mM DTT (1 uL), SuperScript IV Reverse Transcriptase (200 U/uL, 1 uL)	Fresh-made

9.4. Catalog numbers

Component	Concentration	N.°
10X reaction buffer with $MgCl_2$	1X	#EN0521
DNase I, RNase-free	1U	#EN0521
DEPC-treated Water	-	#R0601

EDTA	50 mM	#EN0521
Random hexamers	50 uM	Cat. no. 18418-020
dNTP mix	10 mM	Cat. no. 18427-013
DEPC-treated water	-	Cat. no. 10813-012
SuperScript IV Reverse Transcriptase		Cat. no. 18090-010

10. PCR technique

10.1. Primers used

Gene	Sequence (Primer Forward)	Sequence (Primer Reverse)
SPECC1L	GCGTCTGCATGTTTACTGGTG	GGCACTCAGACTTGGTTGCT
PPP2R2D	AGCAGTAGCAAAGGCACCAT	CTTCGAGTGCCTATCGCAGA
WTAP	GCGACTTCTCTGCTATACCCA	CGTGAGCCCCAAACTGTTAAAAT
GOLGA5	TTTCTTTAGTAGGGGTGTAGGTGC	AGTTCTCTCTCTCACACACACATT
WEE2	AACCTGTATGTAAGCGTCTGCT	AACATTCATGCGGGAGGAGG
FTL	ACCAACAAGATGGCAGAGTACC	AATCAGGTGAAGCAGAAGCTCA
LAMTOR1	TAACTAGCCAGCCCCACCAA	TCTGTGTGTGATTTACCTGCTGAA
CSMD3	ACTGTTTGGGATGGACCCGAC	ATTGGCAGTGGTGGAAACGC
XRCC2	TCTTCTGGTTTGGATCACGGTT	ACCCCATCAGAACCAGCTCTA
DNA2	TACTGTCACCTCTGTCTCCTAGA	ACAAACTGGATCACGCTACCT
PRIM1	GGCGTTTCTGTTCTTGTACCAC	CCCAACCAAATGAAATTGAGCTGAT
ANLN	GCCATCACAAAAAGCGCTCA	AGTTACTGGTCCGAACAGCG
KIF11	AGCAAAAGTTTCAGGTCGGATTG	GTCCTGATTACAGCCCAGTACATA

10.2. PCR

For the PCR technique OneTaq Polymerase Kit from NEB was used: a reaction containing 5X *OneTaq* Standard Reaction Buffer, 10 mM dNTPs, forward and reverse primer, OneTaq DNA polymerase, template DNA and Nuclease-free Water for a final volume of 25 uL for each sample.

The thermocycling conditions for a routine PCR were the following:

STEP	**TEMP**	**TIME**
Initial denaturation	94°C	30 seconds

30 Cycles	94°C	15-30 seconds
	45°C-68°C	15-60 seconds
	68°C	1 minute per Kb
Final Extension	68°C	5 minutes
Hold	4-10°C	

- 10.2.1. Recipes

Reagent	Recipe	Storage
5X One*Taq* Standard Reaction Buffer	20 mM Tris-HCl (pH 8.9 at 25°C), 1.8 mM $MgCl_2$, 22 mM NH_4Cl,22 mM KCl,0.06% IGEPAL CA-630, 0.05% Tween 20	-20°C
PCR Reaction Mix	5X One*Taq* Standard Reaction Buffer (5 uL), 10 mM dNTPs (0.5 uL), 10 mM Forward Primer (0.5 uL), 10 mM Reverse Primer (0.5 uL), One*Taq* DNA polymerase (0.125 uL), template DNA (1 uL) and Nuclease-free Water	Fresh-made

- 10.2.2. Catalog numbers

Component	Concentration	N.°
One*Taq* DNA polymerase Kit	-	#M0480S
dNTPs	10 mM	#N0447

10.3. Quantitative PCR

In order to perform the quantitative PCR the Quantinova SYBR Green Kit (Cat. no. 2018152) was applied. A reaction containing 2X Quantinova SYBR Master Mix, forward and reverse primer, cDNA and RNase-free Water was prepared for a final volume of 10 uL (384-well plates)/ 20 uL (96-well plates) for each sample.

N.B: cDNA volume must not exceed the 10% of the reaction.

For the cycler the following conditions were considered:

Step	Time	Temperature	Comments
Initial Activation Step	2 min	95°C	Activation of Quantinova DNA polymerase
Denaturation	5 sec	95°C	
Annealing/extension	10 sec	60°C	Fluorescent data collection
Number of cycles	35-40		Depending on the amount of DNA

o 10.3.1. Recipes

Reagent	Recipe	Storage
2X Quantinova SYBR Master Mix	QuantiNova DNA Polymerase, QuantiNova SYBR Green RT-PCR Buffer, dNTP mix	-20°C
qPCR Reaction Mix	2X One*Taq* Standard Reaction Buffer (5 uL/10 uL), 10 mM Forward Primer (0.7 uL/1.4 uL), 10 mM Reverse Primer (0.7 uL/1.4 uL), cDNA (1 uL, to add later) and RNase-free Water (up to 10/20 uL)	Fresh-made

11. Proteomic techniques

11.1. Sample processing for MS

Proteins were processed from tissues and cells (MEFs): PBS and lysis buffer were added to the samples in order to have a final proteinaceous concentration of 1 ug/uL and then the samples were biorupted twice: between the two cycles the samples were also boiled for 10 minutes at 95°C to facilitate the denaturation. IAA (200 mM) was then added to the samples for alkylation and they were then incubated for 30 minutes at RT in the dark. For the precipitation 4 volumes of ice-cold acetone were added, vortexed and placed at -20°C overnight, and then the samples were centrifuged for 30 minutes at 14000 rpm (4°C). For acetone removal two washed with 400 uL of 80% acetone were done, the first one centrifuging at 14000 for 10 minutes (4°C) and the second one centrifuging at 14000 for 2 minutes (4°C). After drying the pellet the digestion buffer was prepared and, after adding them to the sample, it was biorupted. For the digestion two different treatments were applied, the first one adding 1 uL of 0.5 ug/uL LysC per sample with an incubation of 4 h (37°C) and the second one adding 25 uL of water and 1 uL of 0.5 ugL Trypsin with o/n incubation at 37°C. The samples were

acidified with 10% TFA and after checking the pH the Clean-Up Oasis Kit was performed: they were finally eluited in MS Buffer A, in order to have a final concentration of 1 ug/uL.

o 11.1.1. Recipes

Reagent	Recipe	Storage
Lysis Buffer	100 mM HEPES (pH 8.5), 50 mM DTT, 4% SDS and MilliQ-water	Fresh-made
Digestion Buffer	3M Urea in 100 mM HEPES (pH 8)	Fresh-made
Buffer A (OASIS CleanUp Kit)	0.05% Formic Acid, MilliQ water	RT
Buffer B (OASIS CleanUp Kit)	80% Acetonitrile, 20% MilliQ water, 0.05% Formic Acid	RT

11.2. Data analysis

For the analysis of the raw data MaxQuant application was used: MaxQuant is a quantitative proteomics software package designed for analyzing large mass-spectrometric data sets and is specifically aimed at high-resolution MS data. For the analysis Trypsin/P was considered as type of digestion, while oxidation and acetyl (Protein N-term.) were counted as modifications. For the identification of differentially expressed proteins Limma and fdrtool packages were used: log(values) were considered as normalized values for the pipeline and significant cases were isolated on adjusted pvalue < 0.05.

11.3. Thermal Proteome Profiling

Thermal Proteome Profiling (TPP) was performed on old and young mouse brains: lysis buffer containing 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate and 0.1% SDS was added to the samples to lyse the tissue, and then biorupting was performed. After spinning to pellet debris (3000g, 5min, 4°C) 10 uL of protein were kept for EZQ measurements and supernatant was distributed into different tubes for heat treatment. The tubes were heated at different temperatures (37, 41, 44, 47, 50, 53, 56, 59, 63, 67°C), incubated at RT for 3 minutes and then snap-freezed with liquid nitrogen. The samples were transferred to polycarbonate thick-walls tubes and

then UC was performed at 100000g for 20 minutes at 4°C; the supernatant was transferred to a new tube and 10 ug of protein were used for an SDS-gel (following Coomassie staining). Volumes for each supernatant (equivalent to 100 ug) were taken and RIPA buffer was added in order to have a concentration of 2ug/uL of protein; samples were reduced with 10 mM DTT at 45°C for 15 minutes and alkylated with 15 mM IAA at room temperature for 30 minutes in the dark. In order to precipitate the proteins four volumes of ice-cold acetone were added and the samples were placed at -20°C O/N. After centrifuging them at 14000 rpm for 30 minutes (4°C) acetone was removed with two washes, the first one adding 500 uL of cold 80% acetone and centrifuging at 14000 rpm for 10 minutes (4°C) and the second one adding the same amount of 80% acetone and centrifuging at 14000 rpm for 2 minutes (always 4°C). The proteins were then digested using Digestion Buffer in two steps: the first one adding LysC (1 ug) and incubating for 4 hours at 37°C (shaking at 650 rpm), the second one adding trypsin (1 ug) diluted in water and incubating O/N at 37°C (always shaking at 650 rpm). Samples were subsequently acidified with TFA and spinned at 37°C for 30 minutes, and peptides were isolated with the OASIS CleanUp Kit. Peptides were resolved in 25 uL of Buffer A, while for MS analysis 0.5 uL with 9 uL of Buffer A were prepared and of them 2.5 uL were loaded for Digest Check. For TMT labelling samples were resuspended in additional 25 uL of 0.2 M HEPES (pH 8.5) and biorupted: 20 ug of material were used for labelling in two steps. For the label check 1 uL of sample was added to 9 uL of MS Buffer A, and finally 5 uL were injected: from the remaining labelled peptides 5 uL were pooled and cleaned (using the OASIS Kit), and finally resuspended in MS Buffer A. For the fractionation 50 uL were injected onto HpH and 16 fractions were collected.

o 11.3.1. Recipes

Reagent	Recipe	Storage
1X RIPA Buffer	50 mM Tris-HCl (pH 8), 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS	Fresh-made
Digestion Buffer	3M Urea, 1% SDC, 0.1 M HEPES	Fresh-made

12. Cell culture

12.1. Thawing cells

Vials with frozen cells were thawed at 37°C in a water bath for 1-2 minutes and decontaminated with 70% ethanol, then the vial content was transferred in a 15 mL falcon with 10 mL of growth medium (containing 4.5 g/L DMEM, FBS and L-Glutamine). The cells were pelleted at 500g for 5 minutes (20°C) and the supernatant was aspirated; the pellet was then resuspended in 10 mL of growth medium and seeded in p100 dishes.

12.2. Splitting cells

After removing the medium the dishes were washed with 7 mL of PBS; to separate the cells from the plate surface 2 mL of Trypsin (0.05%) were added and the cells were incubated for 1 minute at 37°C. After adding 6 mL of medium the cells were spinned at 500g for 5 minutes (20°C) and fresh medium was added for resuspension, before plating in a new p100 dish (0.5 mL of cells and 9.5 mL of medium).

12.3. Harvesting cells

Harvesting of cells was performed after treatment with Actinomycin D: medium was removed from the wells and cells were washed with PBS. Trypsin was added and plates were incubated from 1 minute (37°C). After adding the medium cells were transferred in a tube and centrifuged for 5 minutes at 500g (20°C), and then resuspended in PBS for counting. After transfer in a new tube cells were spinned down (5-10 minutes at 500g), then the supernatant was removed and the tubes were frozen in liquid nitrogen.

Results

1. Ribosome stoichiometry loss and aggregation in the aging brain of *N.furzeri*

1.1. Decoupling between transcriptome and proteome during aging

Whole brains from animals of three different age groups (5 wph, 12 wph and 39 wph) were analysed in the lab by mass spectrometry using a TMT method; RNA-seq data were already available and stored in the JenAge database (Baumgart et al., 2014). A total of 8885 protein groups were quantified with at least two proteotypic peptides, of which 7200 were quantified in both experiments, while for the transcriptome a total of 20018 RNAs were found (with number of raw counts > 0). Another RNA-seq dataset (obtained after rRNA depletion) and a smallRNA-seq were obtained from the same samples used for this work. For each sample, absolute protein abundances estimated from peptide intensities (IBAQ values, (Schwanhäusser et al., 2011)) were correlated with the corresponding transcript levels obtained by RNAseq (RPKM values), obtaining global protein-transcript correlation values for each sample separately: a progressive age-dependent reduction of protein-transcript correlation values was observed, consistent with a decoupling between RNA transcripts and proteins during brain aging (Wei et al., 2015). Decoupling was observed also when analysing the RNA-seq dataset from JenAge, an RNA-seq dataset from polyA+ RNA for animals of the same age groups (Baumgart et al., 2014). Fold-changes of genes differentially expressed in the two RNAseq datasets were strongly correlated. For further analysis, I also focused on the dataset with higher sequencing depth and larger number of replicates, for which the absolute number of differentially expressed genes was higher.

Direct comparison of protein and mRNA fold changes across age groups revealed discrepancies between RNA and protein regulation (**Figure 18D-E**). Protein and transcript changes were significantly correlated in the adult vs. young fish comparison (**Figure 18D**), but the correlation was reduced in the old vs. adult comparison (**Figure 18E**), further supporting a progressive decoupling between transcript and protein regulation. For validation, I analysed proteins previously identified to

be very long-lived in rodent brain (Toyama et al., 2013), including histones, collagens and myelin proteins. For these proteins, I found that transcript, but not protein levels, were generally decreased in old fish, indicating that protein stability contributes to the observed discrepancies between transcripts and proteins (**Figure 19A**). In contrast, immunoreactivity for the glial fibrillary acidic protein (GFAP), but not its RNA levels, were shown to increase significantly in the aging brain (Tozzini et al., 2012), as we confirmed in our data (**Figure 19B**).

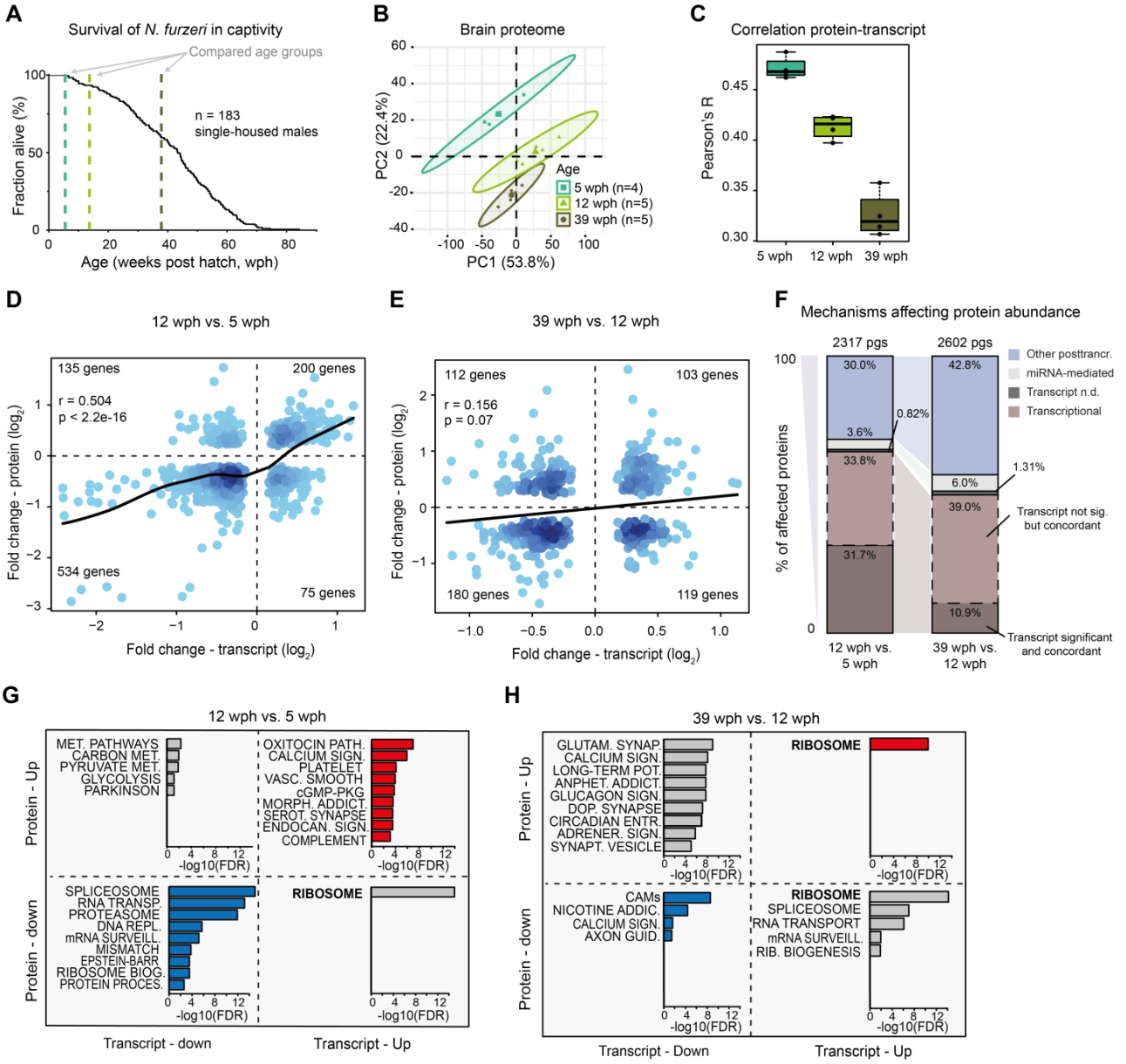


Figure 18. Transcript and protein levels become decoupled during *N. furzeri* brain aging. (A) Survival curve of *N. furzeri* in the FLI facility. Recording of deaths starts at age of 5 wph, which corresponds to sexual maturity, and the colored dashed lines indicate the three age groups analysed in this study (5 animals/group), namely 5 weeks post hatching (wph, young, sexual maturity), 12 wph (adult) and 39 wph (old, past median lifespan) of a wild-derived strain that exhibits a median lifespan of 7-8 months. **(B)** Principal component analysis (PCA) of brain samples based on the abundance of all proteins identified by label-free mass spectrometry. The smaller dots represent individual samples

and the larger dots the centroids of each age-matched group. Ellipses represent 95% confidence intervals. The percentage of variance explained by the first two PC axes is reported in the axis titles. **(C)** Global protein-transcript correlation for each sample, grouped by age. RPKM and iBAQ values were used to estimate transcript and protein levels from matched RNAseq and TMT-based proteomics data obtained from the same animal. An ANOVA test was performed to evaluate significance among the age groups (mean correlation at 5 wph: 0.48; at 12 wph: 0.43; and at 39 wph: 0.33; p = 3.05e-07). **(D and E)** Scatter plot of $\log_2$ fold changes for genes differentially expressed both at transcript and protein levels (adj p < 0.05). The color gradients indicate gene density in the regions where individual points overlap. Numbers of genes in each quadrant and the value of Pearson's coefficient of correlation, r, are reported for each graph. Solid lines represent a spline fit (r = 0.5 for genes significantly affected at both transcript and protein levels, $p < 2.2 \times 10^{-16}$, D) (r = 0.156, p = 0.07, E). **(F)** Mechanisms affecting protein abundance during aging. Bar plots are based on all the proteins affected in either one of the age comparisons (adj p < 0.05). Proteins were divided in the following five groups: (i) proteins and transcripts with significant and consistent changes (dark brown), (ii) proteins with significant changes, and with consistent changes of the transcripts (light brown), (iii) proteins with no transcripts detected (dark gray), (iv) proteins with transcripts whose translation is potentially regulated by miRNAs (light gray), as assessed by the workflow displayed in (Figure S2C), (v) all the remaining proteins that we classified as regulated by other post-transcriptional mechanisms (violet). pgs = protein groups. **(G and H)** Bar plots representing enriched KEGG pathways among genes that showed significant changes at both transcript and protein levels in aging. Genes were grouped according to the four possible patterns of transcript and protein regulation, as visualized by their positions in the four quadrants shown in **(D)** and **(E)**, respectively. Only pathways significantly enriched (FDR < 0.05) are shown. The proteomic dataset was generated by Kelmer Sacramento E. and Kirkpatrick J.; the transcriptomic data were generated by Kelmer Sacramento E., Baumgart M., and Matz S. Analyses performed by Mazzetto M. Figure generated by Mazzetto M. (C-H), Ori A. (B) and Baumgart M. (A).

To find out whether microRNAs could contribute to transcript-protein decoupling, I first analysed miRNA expression levels across the same three age groups: in particular, I selected miRNAs which were affected by aging (obtained with DESeq2 analysis) in the two comparisons, and filtered them by padj < 0.05. In the meantime, the targets of age-affected miRNAs were mapped to our proteome data: only cases were the protein was affected by aging and the transcript was either discordant or not affected were considered, as shown in **Figure 19C**, showing the computational approach. By taking into account potential regulation mediated by miRNAs, I defined a subset of proteins whose abundance is affected by aging via mechanisms independent of both transcript level and miRNA-mediated posttranscriptional regulation (**Figure 18F**). This subset accounted for 30% of the affected proteins in the adult vs. young fish comparison, and increased up to 43% in the old vs. adult comparison.

To clarify whether the transcript-protein decoupling preferentially affects some specific pathways, I classified age-affected genes according to their respective transcript and protein fold changes, and

performed pathway overrepresentation analysis. In the comparison adult vs. young fish, pathways related to the complement coagulation cascade and synaptic function/plasticity were overrepresented in concordantly increased transcript and protein levels, in agreement with the notion that synaptogenesis continues during this phase of residual brain growth (Tozzini et al., 2012). By contrast, genes coding for biosynthetic pathways such as RNA transport, splicing and surveillance of RNA, ribosome biogenesis, and protein processing in the endoplasmic reticulum (ER) were overrepresented in concordantly decreased proteins/transcripts (**Figure 18G**). When looking at old vs. adult fish categories related to synaptogenesis were decreasing at transcript level but still increasing at proteomic level, while the biosynthetic pathways cited above were found to be decreasing at the proteomic level but increasing at the transcriptomic level (**Figure 18H**). Another important result shown in this analysis was that some of the categories (in particular calcium signalling and ribosome) go into a process of loss of stoichiometry, were some subunits are shown to be upregulated and some others are shown to be downregulated, a process which will be explained deeply in the next paragraph.

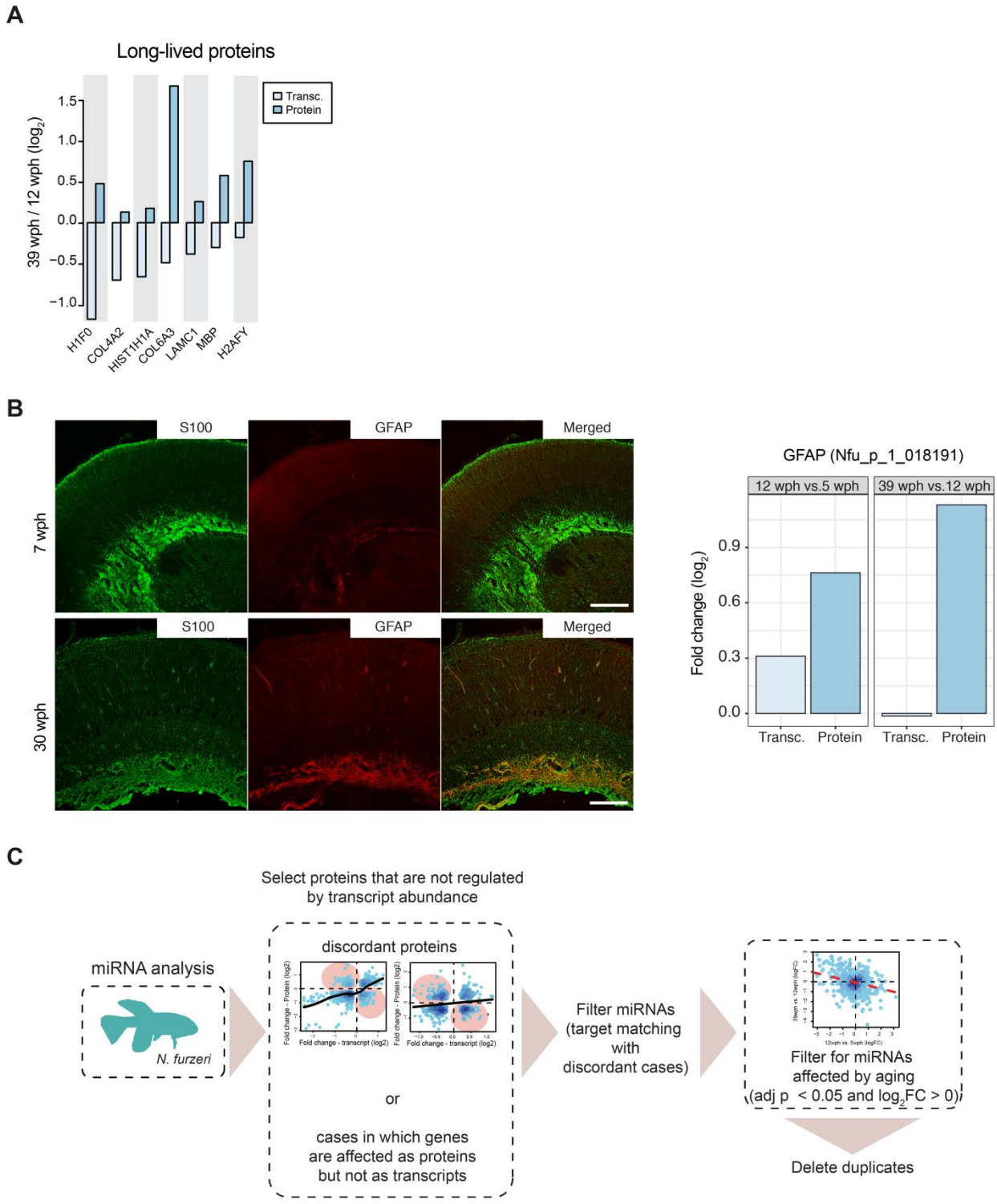


Figure 19. Validation of discordant transcript and protein changes during aging. (A) Analysis of age-related changes of abundance for long-lived proteins. Average fold changes (log2) are displayed as bars for transcripts (light blue) and proteins (dark blue), for a subset of extremely long-lived proteins identified by (Toyama et al., 2013). (B) Validation of an increased level of Glial Fibrillary Acidic Protein (GFAP) in old killifish brain that manifests exclusively at the protein level. Double immunostaining for S100 (green) and GFAP (red) in the central region of the optic tectum of 7 wph fish (upper panels) vs. 30 wph fish (lower panels). Scale bar = 100 μm. Average transcript and protein fold changes (log2) are displayed as bars for transcript (light blue) and 20 protein (dark blue) levels of GFAP. Experiment performed by Terzibasi Tozzini E. and Bagnoli S. (C) Workflow for the

identification of proteins potentially affected by miRNA regulation during aging. For this analysis, proteins were divided into two groups: (i) proteins affected by aging (adj $p < 0.05$) whose abundance change could be explained by transcript expression change, and (ii) proteins affected by aging whose transcript was either regulated but with opposite fold change (discordant cases) or not regulated at all. miRNA analysis was performed only on the latter. To obtain a list of miRNAs affected by aging, differential expression was performed on both the age comparisons (12 wph vs. 5 wph and 39 wph vs. 12 wph) separately. Figures generated by Mazzetto M.

1.2. Stoichiometry imbalance in protein complexes during aging

Amongst genes showing opposite transcript and protein changes already in the adult vs. young fish comparison, we identified 13 genes encoding ribosomal proteins with transcript levels being significantly increased and protein abundances decreased (**Figure 20A** and **Figure 21A**). As mentioned before, fold changes of genes encoding for ribosomal proteins split into two groups in the old vs. adult fish comparison (**Figure 18H**): while transcript levels increase in a concordant way, similarly to the adult vs. young fish comparison, ribosomal proteins show either increased (13 proteins, e.g., RPS20, RPL8 and RPL21) or decreased (14 proteins, e.g., RPS6, RPLP2 and RPL22L1) abundance (**Figure 20A** and **21A**). These findings indicate a loss of stoichiometry (i.e. an imbalance in their relative levels) of ribosomal proteins during aging, which is likely to impair ribosome assembly and to create a pool of orphan proteins at risk of aggregation. When mapped on the ribosome structure (Khatter et al., 2015), age-affected proteins form clusters of either consistently increased or decreased abundance (**Figure 19B**). Since transcript changes are consistent, while ribosomal protein levels are not (**Figure 20A**), the loss of ribosome stoichiometry must result from an alteration of post-transcriptional mechanisms mediating protein homeostasis.

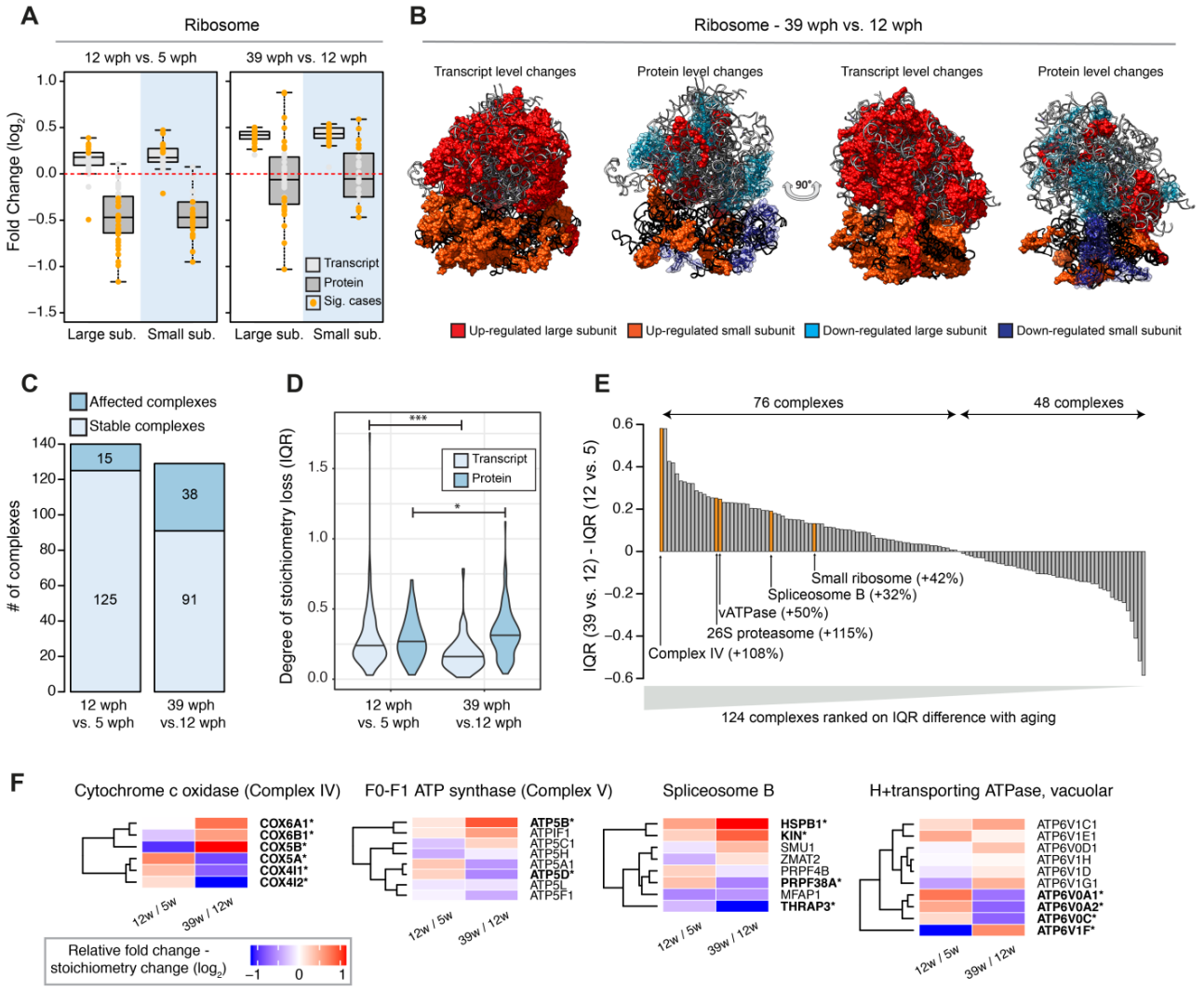


Figure 20. Global stoichiometry loss in protein complexes during aging. (A) Abundance changes of ribosomal proteins and their transcripts during aging. Cytoplasmic ribosomal proteins of large and small subunits are displayed separately and changes are shown for both the age comparisons as box plots. Transcripts are displayed as light blue and proteins as dark blue boxes. Changes of individual proteins are displayed as dots, orange dots identify significant cases (adj $p < 0.05$). **(B)** Visualization of age-related changes of proteins and transcripts projected on the 80S ribosome complex structure. Ribosomal RNA are depicted in ribbon form: 28S rRNA, 5S rRNA, 5.8S rRNA of large subunit are depicted in light gray and 18S rRNA of small subunit is depicted in black. Ribosomal proteins are depicted as molecular surfaces and shown only if significant changes in the level of corresponding mRNA or protein were detected. Affected proteins of large and small subunits are visualized in 2 different shades of red (up regulated), or blue (down regulated). For clarity, down regulated components are displayed as transparent molecular surfaces. Visualization was performed with USCF Chimera program (version 1.12), according to Protein Data Bank archive - human 80S ribosome 3D model: 4UG0. Performed by A. Bartolome **(C)** Statistics of protein complexes undergoing stoichiometry changes with aging. Only protein complexes that had at least 5 members quantified were considered for each comparison. Complexes were considered affected if at least two members showed significant stoichiometry change (adj $p < 0.05$ and absolute $\log_2$ fold change > 0.5). The complete list of stoichiometry changes is available in Table S6. **(D)** Violin plots depicting interquartile ranges (IQRs) of individual members of protein complexes during aging. The IQR for each protein

complex considered in **C** was calculated using transcript (light blue) or protein (dark blue) log_2 fold changes between two age groups. * $p < 0.05$, *** $p < 0.001$, Wilcoxon Rank Sum test. **(E)** Bar plot showing the ranking of protein complexes based on the difference in protein level IQR for each complex between the 39 wph vs. 12 wph and 12 wph vs. 5 wph comparison. Selected complexes are highlighted and the percent of IQR increase between the two age comparisons is indicated in brackets. **(F)** Heat map showing relative protein fold changes for members of selected complexes affected by aging. Names of significantly affected members in the 39 wph vs. 12 wph comparison (adj $p < 0.05$ and absolute log_2 fold change > 0.5) are highlighted in bold with a star. Analyses performed by Ori A. and Mazzetto M.; figures generated by Mazzetto M. (A, E), Bartholome A. (B) and Ori A. (C, D, F).

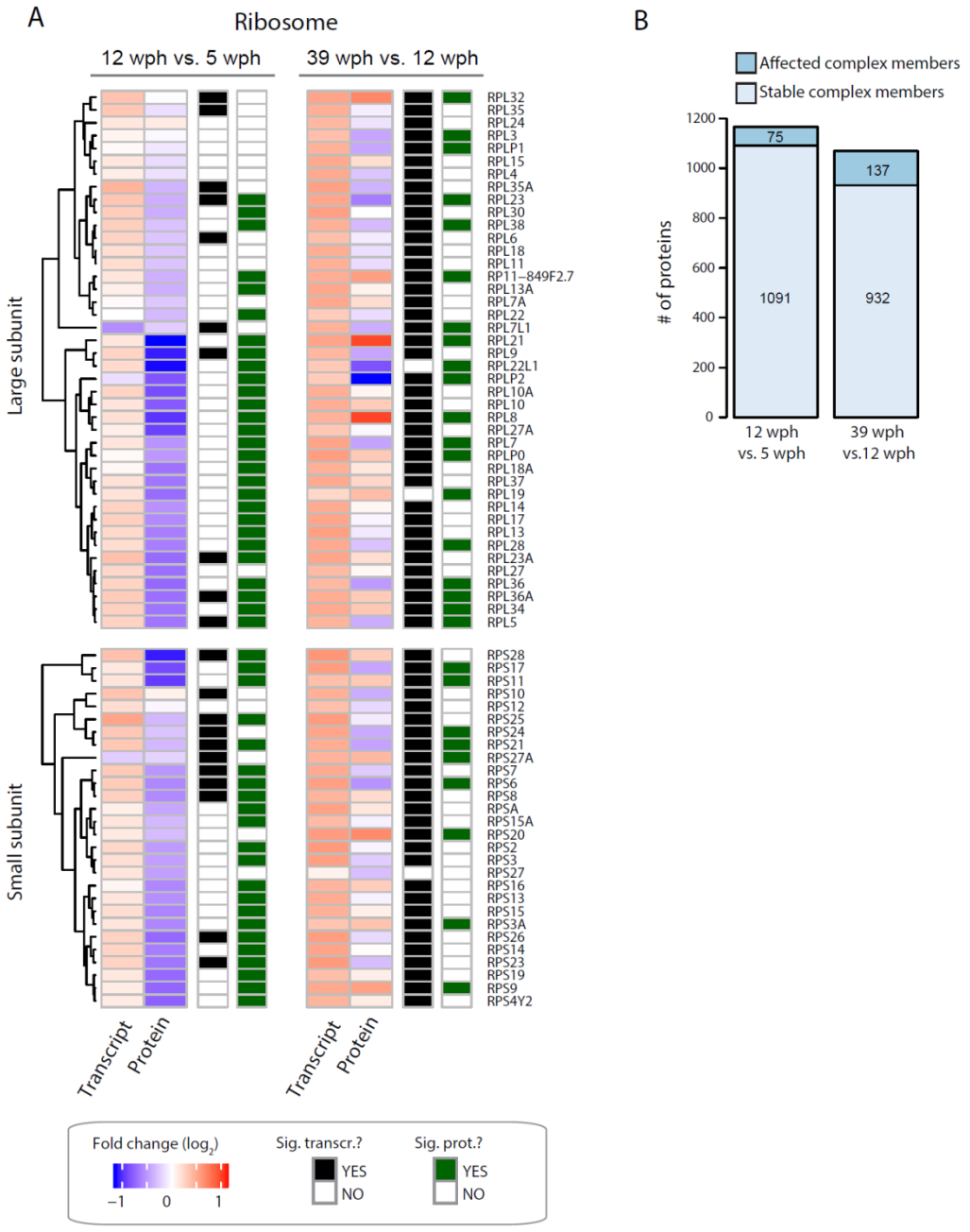


Figure 21. Detailed view of age-related abundance changes for ribosomal transcript and proteins. (A) Heatmap showing transcript and protein fold changes for members of the cytoplasmic

ribosome. Genes are annotated according to whether they are significantly affected at the transcript (adj $p < 0.05$, black and white heatmap) or protein (adj $p < 0.05$, black and green heatmap) level. (B) Statistics of members of protein complexes undergoing stoichiometry changes with aging. Members were considered as affected when their abundance differed significantly (adj $p < 0.05$) from the mean of the protein complex to which they belong, as described in (Ori et al., 2016). Only protein complexes that had at least 5 members quantified were considered for each comparison. Figure generated by Ori A.

We next asked whether the age-related loss of stoichiometry described above occurs more widely in the proteome. We thus analysed all annotated protein complexes in the two age group comparisons (Ori et al., 2013; Ori et al., 2016) and we found that the number of complexes undergoing stoichiometry changes increases from 11% (16 out of 140) between 5 and 12 wph, to 30% (39 out of 129) between 12 and 39 wph (**Figure 20C**). Loss of stoichiometry was confirmed by an alternative metric, namely an increase in the inter-quantile range (IQR) of fold changes of protein complex members (Janssens et al., 2015) in the old vs. adult fish comparison, as measure to estimate the distribution variability during aging. By contrast, the IQRs of transcripts coding for the same complex members were decreased (**Figure 20D**), demonstrating that post-transcriptional mechanisms are responsible for the age-related global loss of stoichiometry in protein complexes.

When individual complexes were ranked according to the difference of IQR between the two age comparisons, the majority of the complexes showed an increase in IQR (76 out of 124, 61%, Figure 2E and Table S6). The most affected complexes included Complex IV and Complex V of the mitochondrial respiratory chain, but not Complex I and Complex III, the cytoplasmic ribosome, the 26S proteasome, the B complex of the spliceosome, and the lysosomal V-type ATPase (**Figure 20E-20F**). These complexes take part in biological processes known to be causative in aging (Carmona-Gutierrez et al., 2016; Chondrogianni et al., 2014; Dillin et al., 2002; Heintz et al., 2017; Steffen and Dillin, 2016), and the expression of transcripts coding for them were previously shown to be correlated with individual lifespan in a longitudinal RNAseq study in *N. furzeri* (Baumgart et al., 2016). Many of these complexes are affected by stoichiometry changes already at the adult stage, identifying these alterations as early events during aging progression.

1.3. Thermal Proteome Profiling reveals minor changes in protein stability during aging

A major question arising from our findings is whether the stoichiometric changes described here influence complex stability or other features. To answer this question, protein stability was globally evaluated in young and old brain lysates using thermal proteome profiling (TPP) (Becher et al., 2018; Savitski et al., 2014) (**Figure 22-23**). Melting temperature from four independent pools of young and old brains were measured, and two biological replicates were analysed. The T_m values obtained in the two biological replicates were strongly correlated (**Figure 22A**).

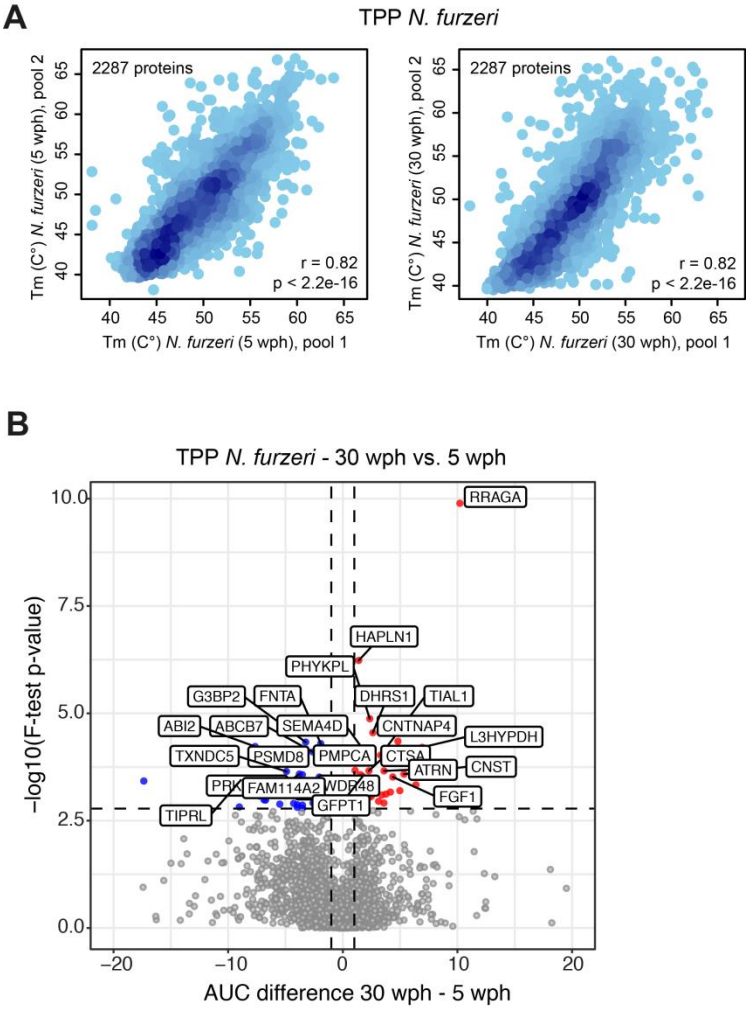


Figure 22. Thermal Proteome Profiling in young and old killifish brain and proteins affected by aging. (A) Scatter plot depicting the comparison of melting temperatures (Tm) values between biological replicates from for N. furzeri. (B) Volcano plot showing age-related changes of protein thermal stability in N. furzeri. Proteins displaying a significant increase (red) or decrease (blue) of thermal stability with age are highlighted. Not affected proteins are plotted in grey. The gene names of selected affected proteins are shown. Horizontal dashed line indicates a p value cut-off corresponding to an adj F test p value cut off of 0.1, and vertical lines a AUC difference cut off of ± 1. Protein quantification was based on samples obtained from 2 independent biological replicates for each group. Experiment performed by Di Sanzo S. and Kirkpatrick J; analyses performed by Mazzetto M. an Ori A. Figures generated by Mazzetto M. (A) and Ori A. (B).

The experimentally derived T_m values were then related to biophysical properties known to influence protein folding. We found that the predicted molecular chaperone requirement for folding is correlated with thermal stability in brain lysates; in particular, proteins that do not depend on

molecular chaperones to fold show a statistically significant increased average T_m both in young and old animals (**Figure 23B).**

An analysis to identify gene sets enriched among low- or high-thermally stable proteins was performed. To increase the number of analysed proteins, ranking was based on the area under the curve (AUC) obtained from the melting curve of each individual protein (Childs et al., 2019). I could rank 6337 protein groups. The ribosome, t-RNA biosynthesis, and pathways related to synaptic plasticity (e.g., "Long-term depression" and "Long term potentiation") were significantly enriched among the less thermally stable proteins, and proteins of higher thermal stability were enriched for lysosomal and SNARE proteins (**Figure 23C**). I found that ribosomes and the 26S proteasome are among the least thermally stable complexes (**Figure 23D**). I observed that median T_m values of protein complexes were highly correlated between young and old animals with the exception of the lysosomal vATPase, which showed higher thermal stability in old killifish brain (**Figure 23E**).

Age-related changes of thermal stability for individual proteins were then globally assessed. The distribution of AUC differences between young and old samples was symmetric around zero (**Figure 23F**), indicating that aging does not cause a global reduction of protein thermal stability. A total number of 47 proteins significantly affected were identified (**Figure 22B**). Among the top hits, mTOR Complex 1 (mTORC1) regulator RRAGA, which is also known to have a direct interaction with the lysosomal vATPase, showed an increased stability (Zoncu et al., 2011) (**Figure 23G**). Taken together, these data show that the thermal stability of few proteins is affected by aging and that there are no major changes in the stability of protein complexes, despite alteration of their stoichiometry, with the possible exception of vATPase and its interactors, indicating that once formed, protein complexes tend to be similarly stable at all ages.

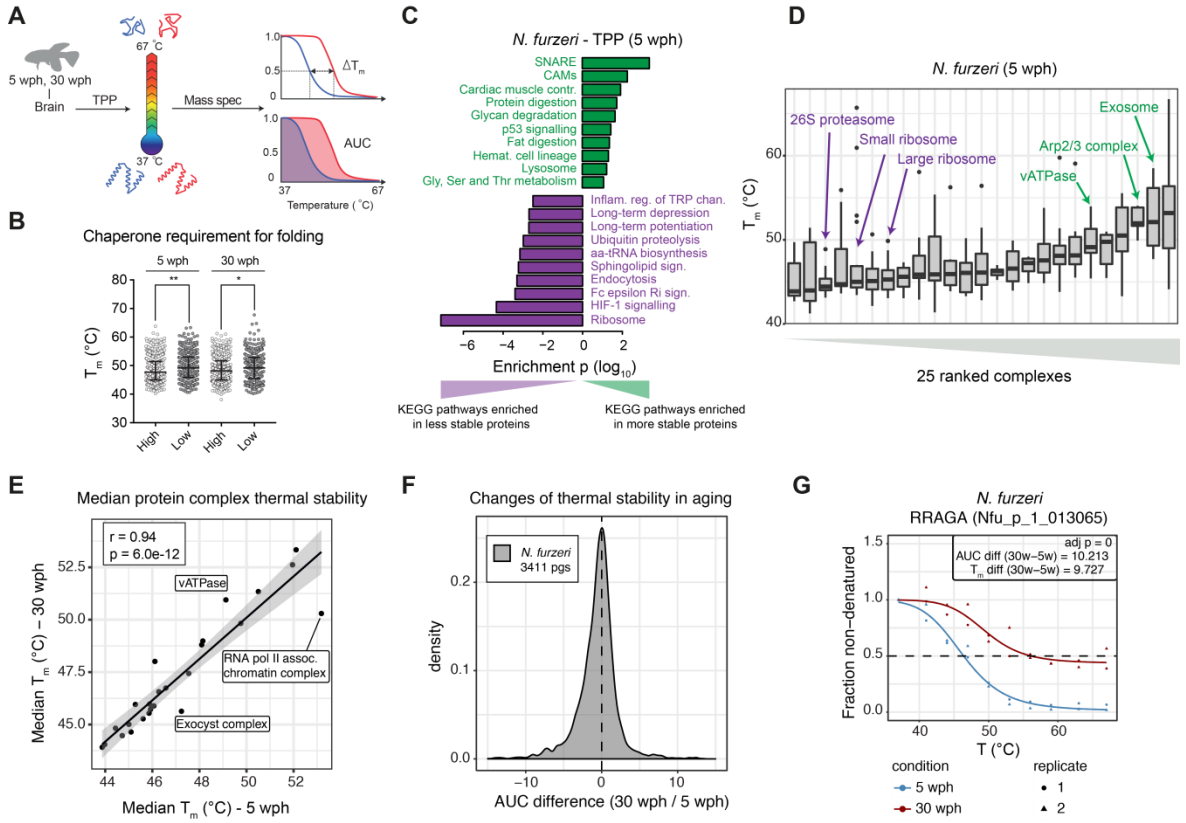


Figure 23. Changes of protein thermal stability during aging. (A) Schematic workflow of the thermal proteome profiling (TPP) procedure. The age groups compared are indicated. Melting point (T_m) and area under the curve (AUC) are illustrated. For each age group, two independent samples were analysed. **(B)** Plot representing T_m distributions of proteins with high requirement of molecular chaperones for folding (top 20% of the cleverSuite scores) or self-folding (bottom 20%) in either young or aged *N. furzeri*. The bar shows the median of the dataset, and the whiskers point the interquartile range. *: p<0.05, **: p<0.01, ***: p<0.001, Kolmogorov-Smirnov test. **(C)** Gene set enrichment analysis for TPP results. For each comparison, proteins were ranked according to the calculated AUC values and the enrichment of KEGG pathways calculated. Data obtained for young samples (5 wph) are shown. The 10 most significantly enriched pathways among more stable (high AUC, green) or least stable proteins (low AUC, violet) are shown ($p < 0.05$). **(D)** Box plot of complexes ranked according to the complex median T_m obtained from TPP experiments. Only complexes with at least 5 members quantified were considered. Selected complexes showing low or high thermal stability are highlighted in violet and green, respectively. **(E)** Comparison of median T_m values of protein complexes between young and old samples. The black solid line indicates the fitted linear model and the grey shade the 0.95 confidence interval of the model. Selected outlier complexes are indicated (r=0.94, p=6.0e-12). **(F)** Distribution of AUC differences between old and young and samples in *N. furzeri*. pgs = protein groups. **(G)** Melting curves for RRAGA obtained by TPP from

young (blue) and old (red) killifish brains. Related to Figure 17. Figures generated by Di Sanzo S. (A), Mazzetto M. (B-C), Ori A. (E-G) and Sanguanini M. (B).

1.4. Old protein aggregates are enriched in ribosomal proteins

Since protein aggregates are known to be SDS-insoluble (Reis-Rodrigues et al., 2012), SDS-insoluble fractions from brain homogenates of young and old animals were compared (**Figure 24A** and **Figure 25A**). Mice were used for this analysis because of the larger brain size that allows retrieval of sufficient amount of aggregates for proteomic analysis. As expected, the yield of SDS-insoluble protein aggregates was significantly higher from old animals, confirming that aging is associated with an increase in protein aggregation (**Figure 24B** and **Figure 25B**). We then analysed the aggregate composition by quantitative mass spectrometry to identify proteins enriched in these aggregates as compared to the starting total brain homogenates. A number of 965 protein groups across three independent experiments were identified, and identified 56 significantly enriched proteins.

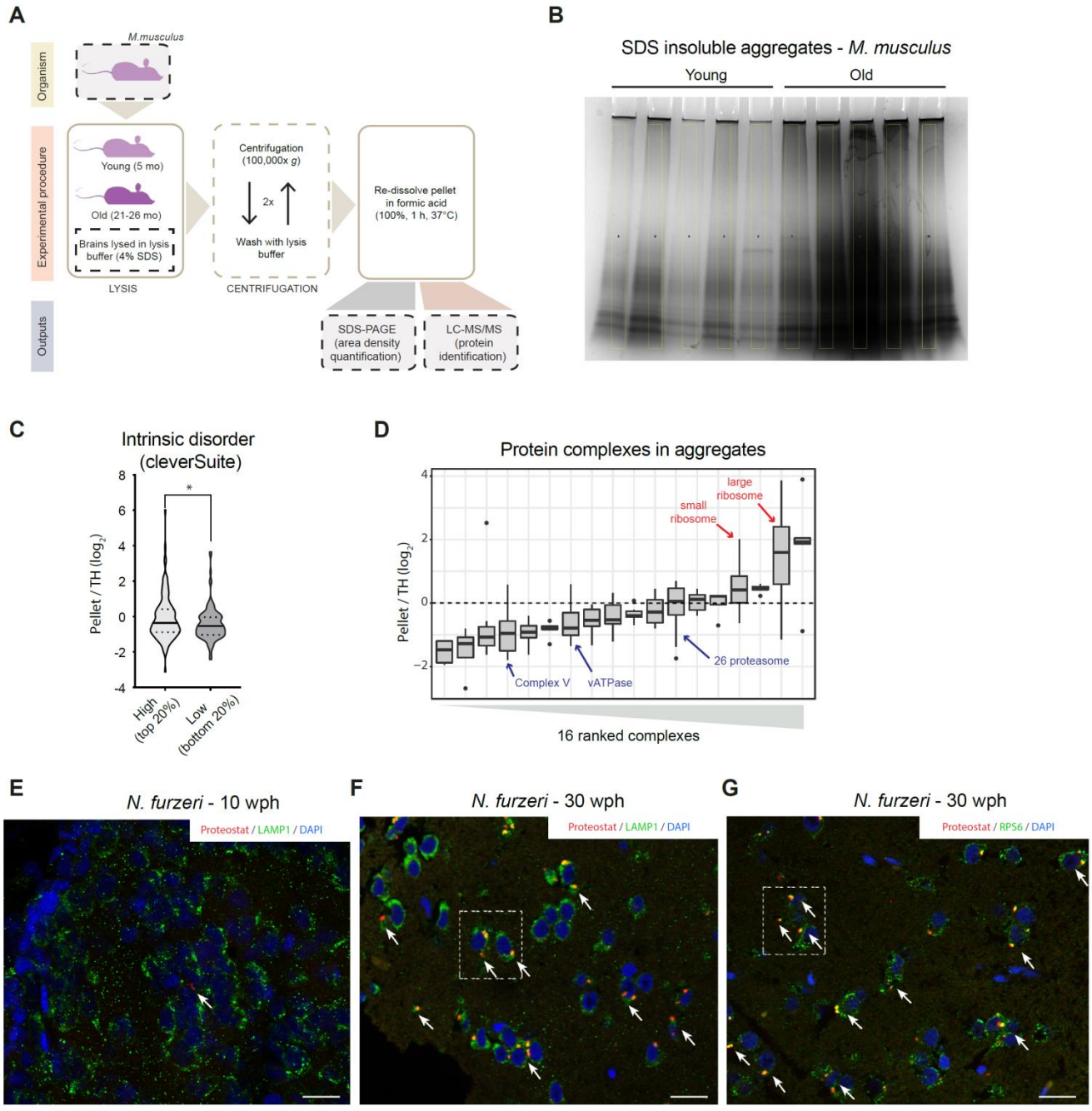


Figure 24. Biophysical properties of protein enriched in protein aggregates and validation of ribosome aggregation by immunofluorescence in old killifish brain. (A) Workflow for the isolation of SDS insoluble aggregates from mouse brain lysates. (B) Coomassie stained SDS-PAGE gel used for quantification of SDS-insoluble aggregates obtained from young and old mouse brain show 24 in Figure 4B. For each sample, a region of interest of the same area was selected for densitometry quantification (highlighted in yellow). (C) Plot representing the fold enrichment in the insoluble fraction of protein with either high (top 20% scores) or low (bottom 20% scores) content of intrinsically disordered region, as predicted with the cleverSuite classifier. Violin plots: the solid line shows the median, the dotted lines the interquartile ranges. *: $p<0.05$, Kolmogorov-Smirnov test. (D) Boxplots of complexes ranked according to the median enrichment of protein complex members in the aggregates. Only complexes with at least 3 members quantified were considered. (E, F and G) Double-labeling of telencephalic sections of N. furzeri with Proteostat as a marker of aggregated

proteins (red) and anti-LAMP1 as lysosomal marker (E/F, green, young versus old fish) and RPS6 to evidence the ribosomal component of protein aggregates (G, green, old fish). Nuclear counterstaining was performed with DAPI (blue). Scale bars = 20 μm. (H) Magnification of the selected area in G showing a detail of the co-localization between lysosomal structures (green) and protein aggregates (red) in the old brain. Scale bar = 10 μm. Aggregates isolation was performed by Kelmer Sacramento E., and proteomic dataset was generated by Kelmer Sacramento E. and Kirkpatrick J. Immunofluorescence (E-G) experiments were performed by Terzibasi Tozzini E. and Bagnoli S. Figures were generated by Kelmer Sacramento E. (B), Mazzetto M. (A, D), Terzibasi Tozzini E. (E-G) and Sanguanini M. (C).

Enriched proteins showed a predicted higher molecular chaperone requirement for folding, and were richer in intrinsically disordered regions (**Figure 25C**). Among the enriched proteins, we found collagens (Col1a1, Col1a2 and Col4a2), which are well-known to undergo age-dependent crosslinking, and ferritins (Fth1 and Ftl1), whose aggregation is linked to the age-dependent brain accumulation of intracellular iron (Ripa et al., 2016; Zecca et al., 2004) (Figure 4D). Protein aggregates were also enriched for ribosomal proteins (**Figure 24D-25D**) that are both characterised by protein transcript decoupling (**Figure 18H**) and loss of complex stoichiometry (**Figure 20A-20B**) during aging. Other protein complexes that displayed loss of stoichiometry (i.e., Complex V and vATPase) did not show significant enrichment in aggregates, indicating that stoichiometry imbalances do not always correlate with protein aggregation (Figure S5D).

To confirm the aggregation of ribosomal proteins in killifish, a staining of young (7-10 wph) and old (27-30 wph) brain slices using Proteostat was performed, an amyloid-specific dye (Shen et al., 2011). As expected, we detected lysosomal aggregates in old but not in young brains (**Figure 24E, 24F** and **24H**). These aggregates appeared to contain the ribosomal protein RPS6 (**Figure 24G-25D**). Taken together, these data demonstrate that an age-dependent aggregation of ribosomal proteins is conserved in both old fish and old mice.

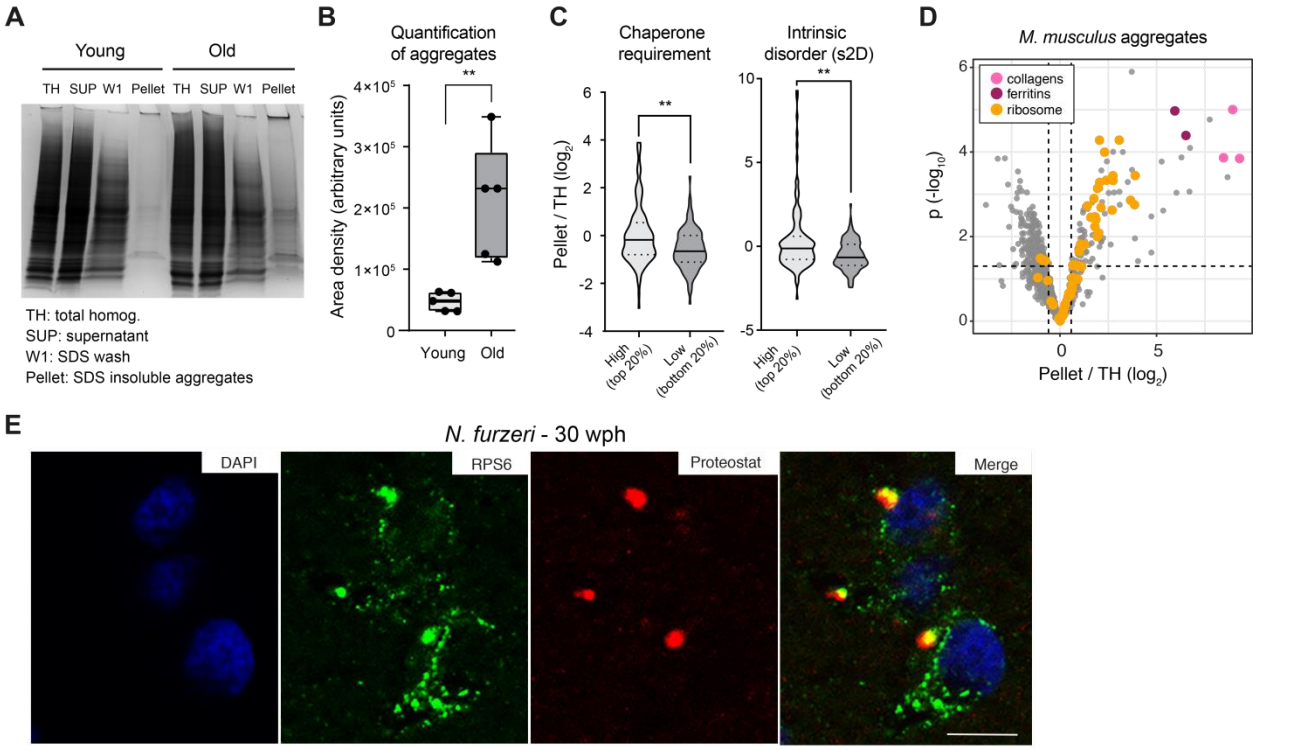


Figure 25. Aggregation of ribosomes during brain aging. (A) Representative Coomassie-stained SDS-PAGE gel showing the isolation of SDS insoluble aggregates from mouse brain lysates. TH = total homogenate, SUP = supernatant, W1 = SDS-soluble fraction, Pellet = formic-acid soluble fraction (see Figure S5A). **(B)** Quantification of the yield of SDS-insoluble aggregates from young and old brain lysates was based on densitometry analysis of Coomassie-stained gel bands obtained from different animals, n=5 per age group (Figure S5B); ** $p < 0.01$, unpaired t-test. **(C)** Proteins enriched in aggregates show a predicted higher molecular chaperone requirement for folding (top vs. bottom 20% p=0.0055, Kolmogorov-Smirnov test), and are richer in intrinsically disordered regions (s2D-derived scores, top vs. bottom 20% p=0.0019, Kolmogorov-Smirnov test). Violin plots: the solid line shows the median, the dotted lines the interquartile ranges. The same result was obtained with cleverSuite-derived scores (top vs. bottom 20% p=0.0201, Kolmogorov-Smirnov test) (Figure S5C). **(D)** Volcano plot based on protein quantification by label-free mass spectrometry depicting the enrichment of specific proteins in protein aggregates. The x-axis indicates the $\log_2$ ratio between protein abundance in aggregates (Pellet) and starting total homogenate (TH). The horizontal dashed line indicates a p value cut-off of 0.05 and vertical lines a $\log_2$ fold change cut off of ± 0.5. Selected proteins are highlighted as colored dots as indicated in the figure legend. Protein quantification was based on samples obtained from 3 independent isolations. **(E)** Double-labelling of telencephalic sections of *N. furzeri* with anti-RPS6 (green) as ribosomal marker and Proteostat as a marker for aggregated proteins (red). Nuclear counterstaining was performed with DAPI (blue). Scale bar = 10 µm. Immunostaining performed by Terzibasi Tozzini E. and Bagnoli S. (E); figures generated by Kelmer Sacramento E. (A-B), Ori A. (D), Terzibasi Tozzini E. (E) and Sanguanini M. (C).

1.5. Reduced proteasome activity promotes loss of protein stoichiometry

Since protein stoichiometry loss could be due to decreased proteolysis rates, we focused on the proteasome, which is one of the main degradation machineries and itself undergoes stoichiometry loss upon aging (**Figure 20E**). First, we observed a decrease in the levels of both proteasome proteins and their transcripts in adult killifish, which was followed by a stoichiometry imbalance that manifested in old fish exclusively at the protein level. In particular, we observed an imbalance between proteins belonging to the 19S and the 20S complexes, with the latter being exclusively upregulated in old fish (**Figure 26A-26B**).

To further investigate the proteasome functional status in killifish brains of different age groups, we performed native gel electrophoresis of proteasomes accompanied by in-gel proteasome activity assays (Chondrogianni et al., 2015). A significant decrease in the levels of both 30S (double-capped) and 26S (single-capped) proteasomes was revealed already in adult animals (**Figure 26C** and **28A**). This decrease was accompanied by a significant reduction of all three proteasome activities in adult and old samples compared to young samples (Figure **26D** and **28B**).

To assess whether this decline is relevant for lifespan determination, I re-analyzed a longitudinal study of RNAseq, where transcripts from fin biopsies were quantified at 10 and 20 wph, and variations in gene expression were correlated with individual variations in lifespan (Baumgart et al., 2016). I observed that the shortest-lived individuals of the cohort were characterized by a larger age-dependent downregulation of transcripts coding for proteasomal proteins, which was absent in the longest-lived individuals (**Figure 26E**), indicating that the rate of proteasome downregulation in early adult life is predictive of lifespan in killifish.

I then applied Cox-Hazard model on the dataset using 20 wph/10 wph expression ratios as value of differential expression: this statistical analysis correlates the probability of one or more events to occur with an independent variable. In this type of analysis, the event was death and the variable was gene expression at either 10 weeks or 20 weeks (**Figure 27B**). To better visualize in which pathways risk and protective factors were enriched during development I performed GAGE and collected the results in a barplot: as shown in **Figure 27C** factors that are detrimental for lifespan are enriched in ribosomal complex and Rap1 signaling, while factors that are beneficial for lifespan are enriched in the proteasome complex, in consistency with the previous results (**Figure 26E**). When looking at lifespan curves I observed that individuals that express more proteasome subunits are more likely to live longer, in contrast with individuals that do not express them (who are more likely to live less, **Figure 27D**).

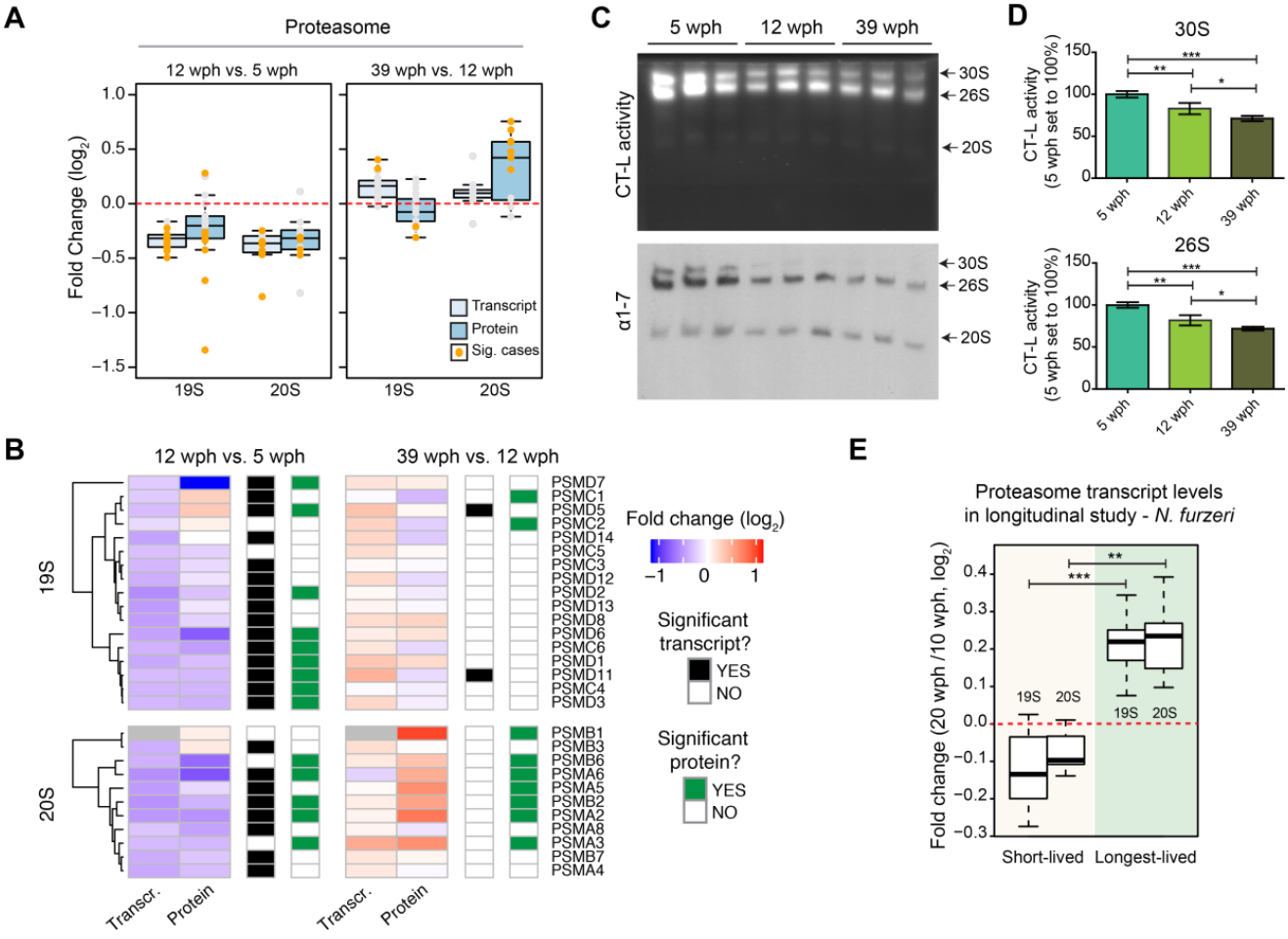


Figure 26. Reduced proteasome activity and assembly in old brains. (A) Abundance changes of proteasome proteins and their transcripts during aging. Members of the 19S and 20S complex are displayed separately and changes are shown for both the age comparisons as box plots. Transcripts are displayed as light blue and proteins as dark blue boxes. Changes of individual proteins are displayed as dots, orange dots represent significant cases (adj. $p < 0.05$). **(B)** Heat map showing transcript and protein fold changes for members of the 26 proteasome (19S and 12S complexes). Genes are annotated according to significance of their changes at the level of transcript (adj. $p < 0.05$: black, adj $p > 0.05$: white) or protein (adj. $p < 0.05$: green, adj. $p > 0.05$: white). **(C)** In-gel proteasome assay following native gel electrophoresis (top) and immunoblotting of proteasome complexes (30S, 26S and 20S) (bottom) in young (5wph), adult (12 wph) and old (39 wph) killifish brains. **(D)** Bar plots depicting the quantification of chymotrypsin-like (CT-L) activity from native gels calculated for doubly capped (30S) or singly capped (26S) proteasomes. n >=5 per sample group; error bars indicate standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, unpaired t-test. For each sample group, the mean value of activity in young samples (5 wph) was set to 100%. **(E)** Downregulation of proteasome transcripts in killifish longitudinal study. Boxplot depicts average fold-changes between 10 and 20 wph for 25 fish that lived shorter than the median lifespan (short-lived) and the 25 longest-lived fish of a cohort of 152 fish (longest-lived). Statistical significance calculated by Wilcoxon Rank Sum test. Expression data were obtained from (Baumgart, Priebe, Groth et al., 2016). Experiments (B-D)

performed by Chondrogianni N.; figures generated by Chondrogianni N. (C-D), Mazzetto M. (A, E) and Ori A. (B).

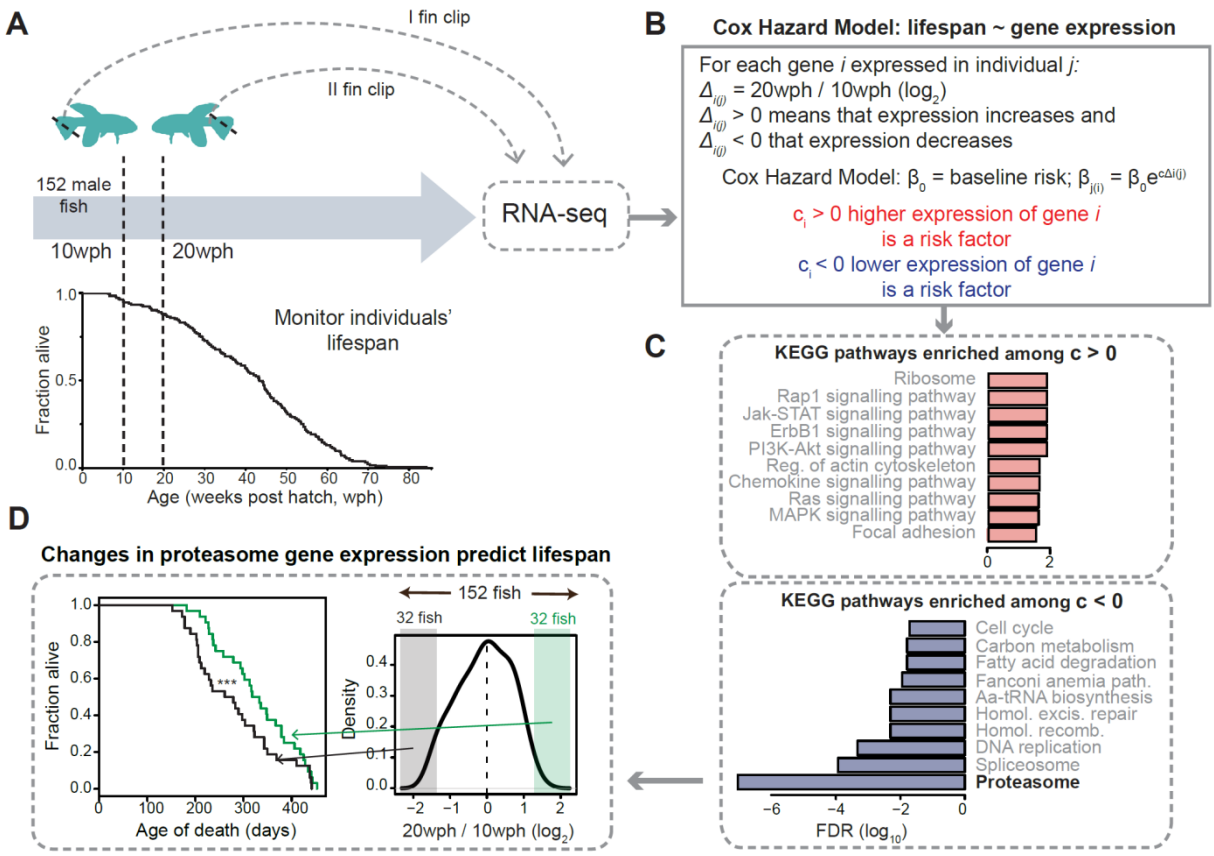


Figure 27. Longitudinal study in 152 killifish identifies early in life decrease of proteasome transcripts as a major risk factor for reduced lifespan. (A) Study design. Two fin clips were taken at 10 and 20 weeks post hatching (wph) from 152 male killifish and individual lifespans monitored. RNA sequencing was employed to compare transcriptome changes between 10 and 20 wph for each individual fish. (B) A Cox Hazard model was used to correlate lifespan to gene expression changes. Two groups of genes where identified: (i) genes that are risk factors (i.e., associated to reduced lifespan) when their expression increases between 10 and 20 wph (red), and (ii) genes that are risk factors when their expression decreases between 10 and 20 wph (blue). (C) KEGG pathways enriched among genes predictive of lifespan. Only pathways with FDR<0.05 are shown. (D) The average change in expression for proteasome transcripts was analysed across the entire cohort of 152 fish. Lifespan was compared among fish that showed extreme changes in proteasome levels between 10

and 20 wph (32 fish showing the most pronounced decreases, shown in black vs. 32 fish showing the most pronounce increases). *** p= log-rank test. Figures generated by Mazzetto M. and Ori A.

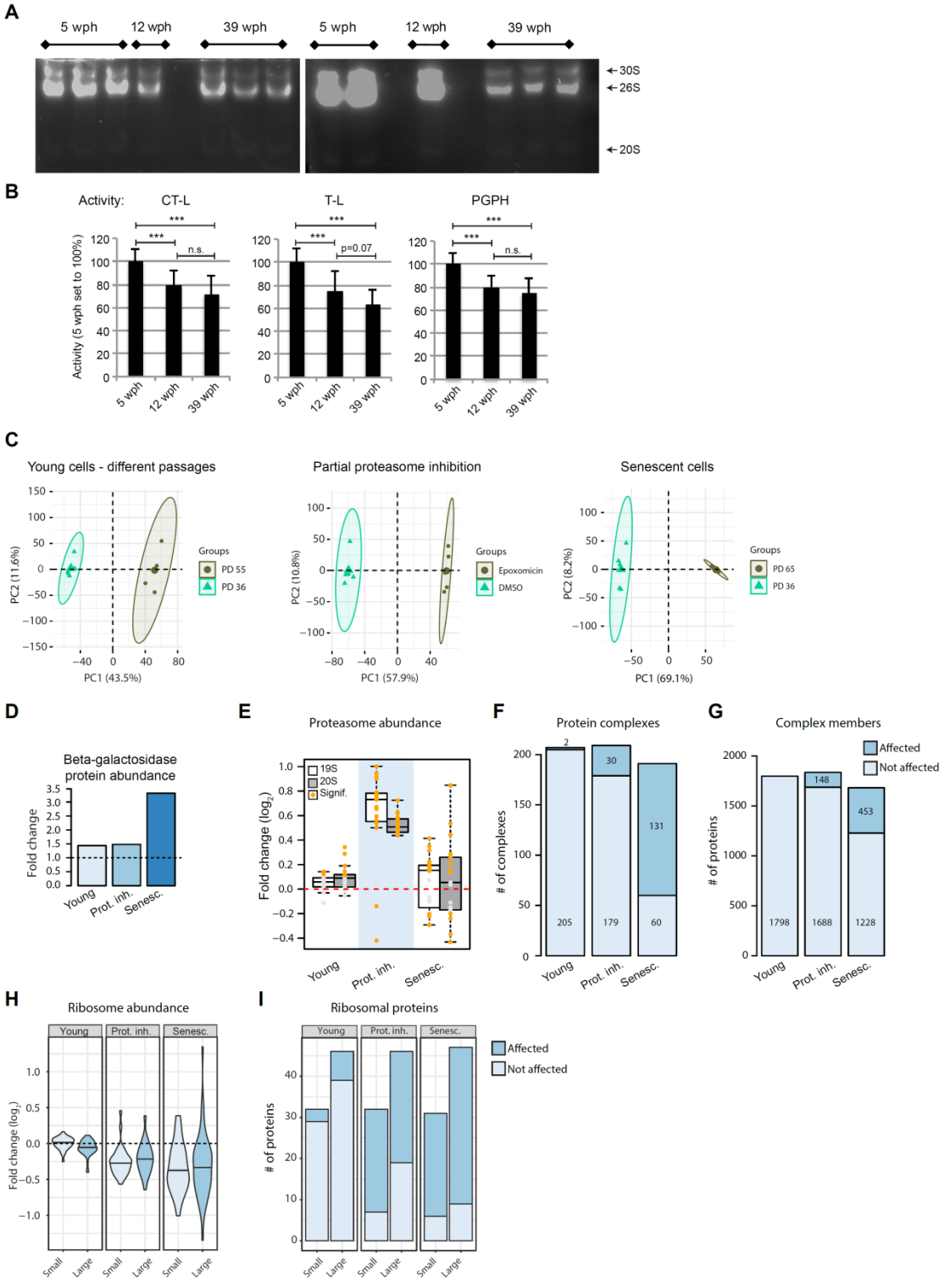


Figure 28. Loss of protein complex stoichiometry by partial proteasome inhibition in primary human fibroblasts. (A) Principal component analysis (PCA) of HFL-1 fibroblast samples based on proteome profiles obtained by Tandem Mass Tag (TMT) quantification for fibroblasts harvested at different passages ("Diff. pas."), fibroblast treated with the proteasome inhibitor epoxomicin for four days vs. vehicle control ("Prot. Inh."), and senescent vs. young fibroblast ("Senescent"). The smaller

dots represent individual samples and the larger dots the centroids of each age-matched group. The percentage of variance explained by the first two PC axes is reported in the axis titles. (B) Barplot depicting the abundance changes of GLB1 protein, whose enzymatic activity is responsible for the senescence-associated beta galactosidase staining, across the three sets of samples analysed. Abundance changes were obtained from the quantitative mass spectrometry experiments. (C) Proteasome protein abundance changes in human fibroblast treated with proteasome inhibitor. Fold changes (log2) for members of the 19S and 20S proteasome are shown as boxplots for the three different comparisons. Significantly affected 27 cases are highlighted as orange dots (adj $p < 0.05$). (D) Statistics of protein complexes undergoing stoichiometry changes upon proteasome inhibition and in senescent fibroblasts. Complexes were considered as affected if at least two members showed significant stoichiometry change (adj $p < 0.05$ and absolute log2 fold change > 0.5). (E) Statistics of members of protein complexes undergoing stoichiometry changes upon proteasome inhibition and in senescent fibroblasts. Members were considered as affected when their abundance differed significantly (adj $p < 0.05$ and absolute log2 fold change > 0.5) from the mean of the protein complex to which they belong, as described in (Ori et al., 2016). (F) Fold changes for proteins belonging to small and large ribosome subunit are displayed in light and dark grey, respectively. The middle line in the violin plot indicates the median fold change of the complex members. (G) Stacked barplots indicate the number of ribosomal proteins quantified (grouped as member of the small or large ribosomal subunit) and the fraction of significantly affected members (adj $p < 0.05$) for each comparison. (H) Ribosome abundance in young, proteasome-inhibited and senescent cells: fold changes (log2) were used for the analysis. Large (dark blue) and small (light blue) subunits' proteins are shown. (I) Number of ribosomal proteins affected by loss of stoichiometry in young, proteasome-inhibited and senescent cells. Large (dark blue) and small (light blue) subunits' proteins are shown. Figures generated by Ori A.

We then asked whether reduced proteasome function could induce stoichiometry loss of protein complexes. To answer this question, we used HFL-1 human primary lung fibroblasts as a model system for the following reasons: (i) these cells are a standard *in vitro* model for aging studies because they undergo replicative senescence (Hayflick, 1965), and (ii) senescent fibroblasts have been shown to possess reduced proteasome activities (Chondrogianni et al., 2003). We treated young fibroblasts with epoxomicin, a selective and irreversible proteasome inhibitor, for 4 days at a dosage that induces 50% reduction of proteasome activity, resembling the reduction of activity exhibited by senescent HFL-1 fibroblasts (Chondrogianni et al., 2003). We then analyzed proteome changes in epoxomicin- vs. vehicle control-treated HFL-1 cells, and compared it to changes induced by replicative senescence. In order to take into account technical and biological variability, we included an additional comparison of non-senescent cells at different passages (**Figure 29A**). Principal component analysis on protein abundance profiles revealed that samples cluster according to their condition (**Figure 28C**). Increased protein levels GLB1, the enzyme responsible for the senescence associated beta-galactosidase, confirmed the senescent phenotype of old fibroblasts (**Figure 28D**). Increased levels of proteasome subunits, a known compensatory reaction to proteasome inhibition

(Wojcik and DeMartino, 2002), were detected in epoxomicin-treated cells (**Figure 28E**). In both epoxomicin-treated cells and senescent fibroblasts, we detected changes in protein complex stoichiometries (**Figure 28F-G**) and a global increase of IQRs for members of protein complexes (**Figure 29B**), resembling the aging phenotype observed in killifish brain (**Figure 20C-20D**). Proteasome inhibition and replicative senescence both led to decreased levels of ribosomal proteins that affected most, but not all, the components of the ribosome (**Figure 25H-I**). As a consequence, we observed a significant increase in the IQRs of both small and large subunits of the ribosome (**Figure 29C**), indicating loss of ribosome stoichiometry. Taken together, these data demonstrate that a partial reduction of proteasome activity occurs in killifish brain starting from adult age, and that inducing a comparable reduction of proteasome activity in primary human fibroblast is sufficient to induce a loss of protein complex stoichiometry that is also observable in senescent cells.

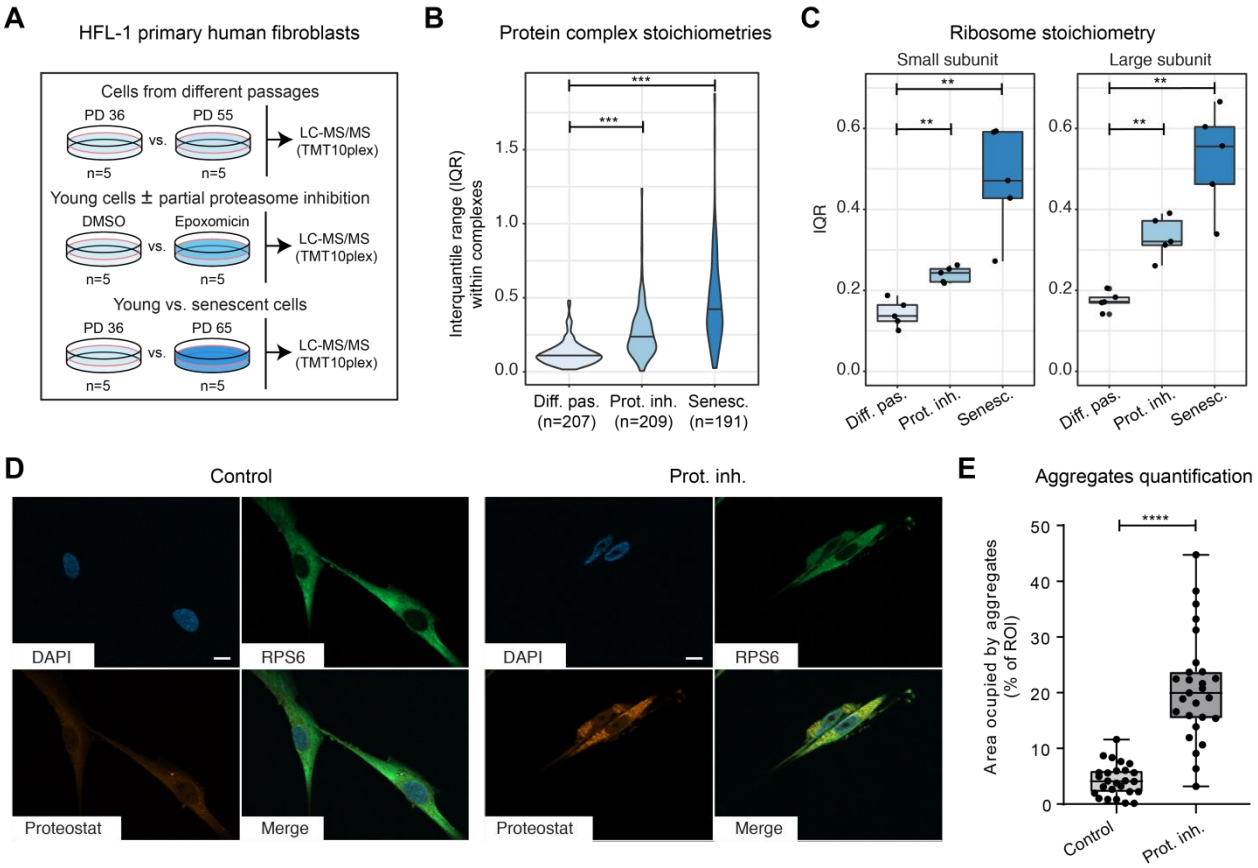


Figure 29. Partial inhibition of the proteasome is sufficient to induce loss of stoichiometry of protein complexes and aggregation in primary human fibroblasts. (A) Schematic workflow depicting the proteome comparisons performed in HFL-1 fibroblasts. Three comparisons were performed: (i) fibroblasts harvested at different passages ("Diff. pas."); (ii) fibroblast treated with the proteasome inhibitor epoxomicin for four days vs. vehicle control ("Prot. inh."); (ii) senescent vs. young fibroblast ("Senescent"). For each comparison, five biological replicates were analysed. **(B)**

Violin plots depicting interquartile ranges (IQRs) for protein complexes. IQRs were calculated for each complex from the protein fold changes of the respective members as in Figure 2D. *** $p < 0.001$, Wilcoxon Rank Sum test. **(C)** Boxplots depict the IQR for large and small ribosome subunit derived from protein fold changes for each sample set. ** $p < 0.01$, Wilcoxon Rank Sum test. **(D)** HFL-1 cells treated for four days with vehicle control ("Control", left) or epoxomicin ("Prot. inh.", right) stained with anti-RPS6 (green) and Proteostat, as a marker for aggregated proteins (red). Nuclear counterstaining was performed with DAPI (blue). Scale bar = 10 μm. **(E)** Quantification of aggregates in epoxomicin treated cells. Quantification was performed using the Proteostat channel. Control (n=26 cells), Prot. inh. (n=27), **** $p < 0.0001$, Unpaired t test. Proteasome inhibition experiments performed by Kelmer Sacramento E. and Bartholome A; immunostaining performed by Caterino C. an Matz S. Figures generated by Caterino C. (D), Kelmer Sacramento E. (E), Ori A. (A-C).

2. Conservative mechanisms of aging across genetic differences in *N.furzeri*

2.1. Genetic-related signature is already visible at embryonic level in *N.furzeri*

In order to identify the transcriptional signature associated to lifespan differences, I analysed 211 RNA-seq libraries obtained from two strains differing in lifespan (MZM-0410 and GRZ-D), four tissues (brain, liver skin, muscle) and five time points. The time points collected correspond to sexual maturity (5 weeks) and (for MZM) to young adult (12 weeks), adult (20 weeks), old (27 weeks), and geriatric (39 weeks), with the exception of muscle for which only young (9 weeks) and old (27 weeks) samples were available. For GRZ, the time points collected correspond to 5 wph, 7 wph, 10 wph, 12 wph, and 14 wph. In addition, we analysed one embryonic stage in two conditions (diapause and direct development) (**Figure 30A**).

Differential expression analysis was performed with DESeq-2 (Love MI, Huber W, Anders S. et al., 2014) using generalized linear models that separate the effects of different independent variables on gene expression. The three variables analysed here are age, tissue and strain. In a first analysis, I computed the effects of strain for each tissue separately. Effects of strain were large, with close to ten thousand differentially-expressed transcripts (DEGs) detected in the three tissues (brain, liver, skin). Notably, the effect of strain was already detectable in the embryo samples, with more than six hundred DEGs (**Figure 30B**). I then repeated the analysis across tissues and computed the DEGs and DEMs relative to age and strain (**Figure 30B-C**): the analysis revealed that the effects of strain were larger than the effect of age in terms of number of DEGs. This difference was particularly evident in the liver. Therefore, the effects of age and strain are at least partially independent.

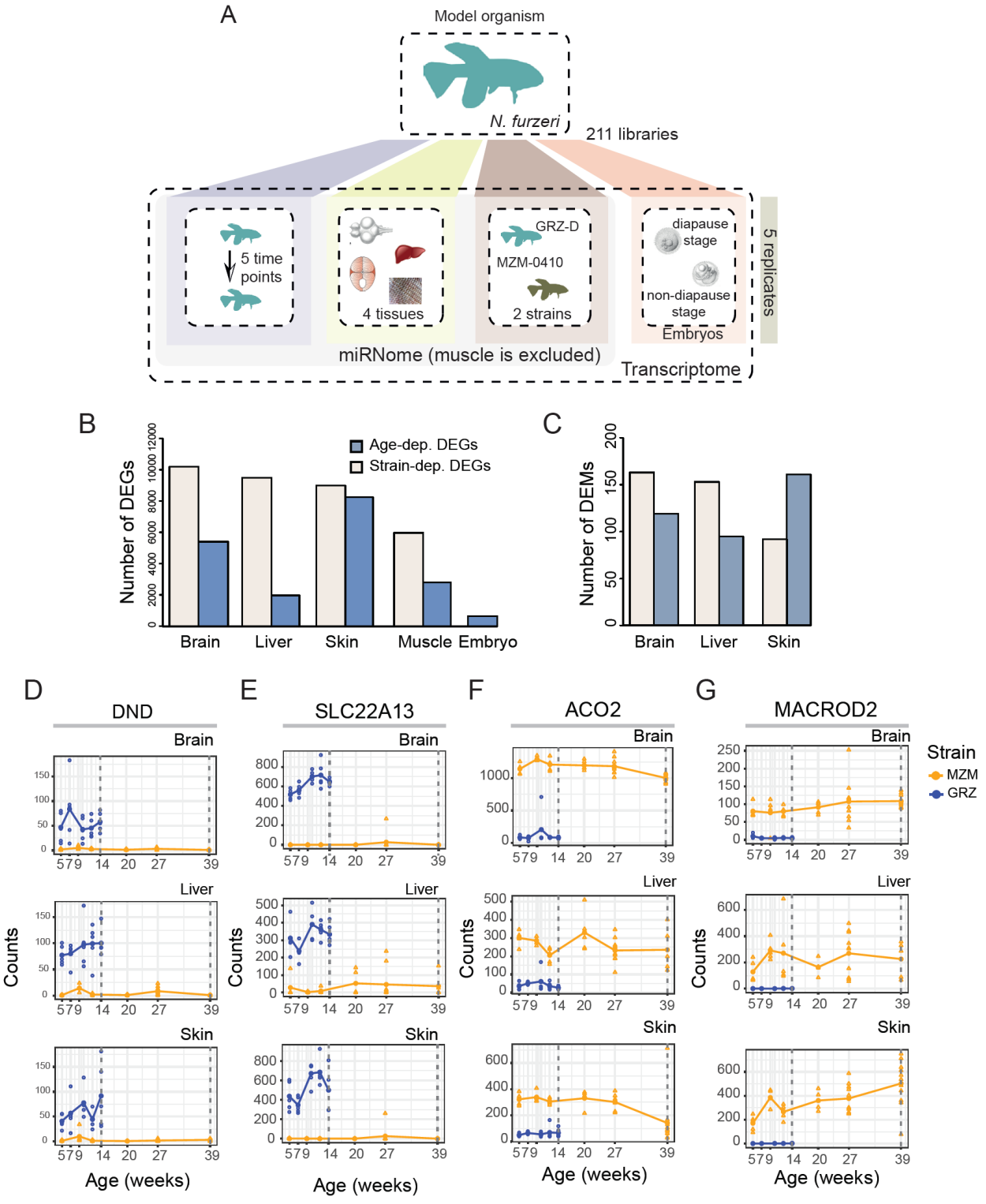


Figure 30. Description of the data. (A) Sample description. For this work two datasets were used: the first one is an RNA-seq dataset comprehensive of 5 tissues (brain, liver, skin, muscle and embryo), 2 strains (MZM-0410 and GRZ-D), 5 different time points (for brain, liver and skin) and 2 different stages (for embryo), with 5 replicates per sample. The second one is a small RNA-seq dataset (Baumgart, 2017) containing 3 different tissues (brain, liver and skin) and up to 6 different time points. (B) Number of differentially expressed (DE) transcripts for different variables: DE genes (DEGs) were computed for each tissue separately and number of age-dependent (plum) and strain-dependent (blue) DE cases are plotted for each tissue. (C) Number of DE miRNAs for different

variables: DEMs were computed for each tissue separately and number of age-dependent (plum) and strain-dependent (blue) DE cases are plotted for each tissue. (D-G) Expression profile of some of the most DEGs between the two strains: DND (D), SLC22A13 (E), ACO2 (F) and MACROD2 (G). Normalized pseudo-counts were used for plotting expression strength; each point represents an individual case, the line connects the medians of each age group, MZM-0410 (orange) and GRZ (blue). Transcriptomic datasets obtained by Baumgart M., Matz S. and Groth M. Analyses and figures generated by Mazzetto M. (A-C) and Bagnoli S. (D-G).

Some genes showed extreme differences in expression between the two strains, for example the genes DND, that is normally expressed only during early embryonic development (Gross-Thebing et al., 2017), and the transporter SLC22A13, whose expression is normally restricted to the kidney (Bahn et al., 2008), have negligible expression in MZM-04010, but are strongly upregulated in all tissues of the GRZ strain (**Figure 30D-30E**). Examples of the opposite pattern are ACO2, coding for aconitase 2, and MACROD2, a gene whose deletion causes chromosome instability and represents a cancer risk (Sakthianandeswaren et al., 2018; **Figure 30F-30G**). To identify gene sets that are differentially expressed in the two strains, I performed generally applicable gene set enrichment (GAGE) analysis (Luo, Weijun, Friedman, Michael, Shedden, Kerby, Hankenson, Kurt, Woolf, Peter, 2009) excluding the embryo samples. GAGE detected significant enrichment for a large number of gene sets, so these were further grouped into semantically similar clusters using REVIGO (Supek F, Bošnjak M, Škunca N, Šmuc T., 2011; **Figure 31**). Categories related to development, RNA processing, protein-targeting, -transport or -modification, autophagy and membrane organization were upregulated in the short-lived strain (GRZ-D). Genes upregulated in the long-lived strain were found to be enriched in categories related to DNA replication/cell cycle and immune response/protein activation cascade. These results highlight profound and unexpected differences in gene expression between these two strains.

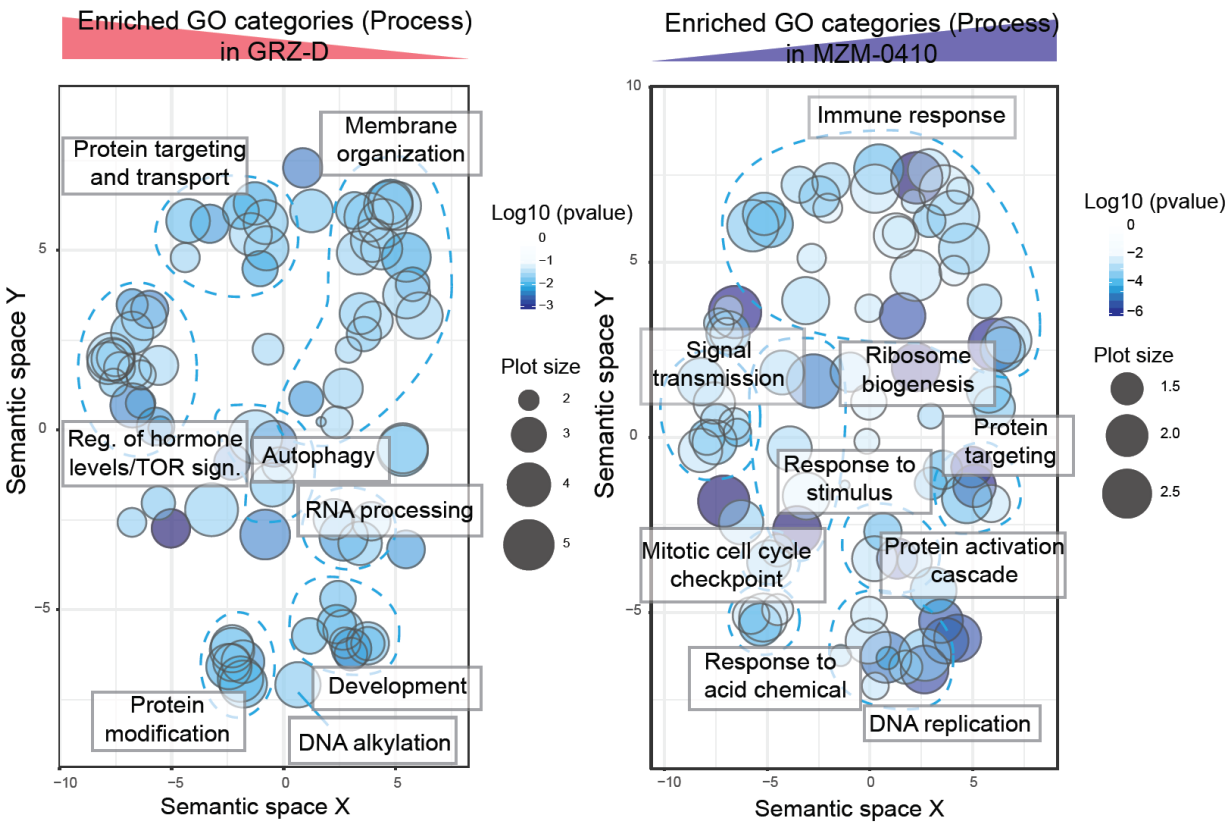


Figure 31. Gene enrichment results for strain-dependent transcripts. Enrichment analysis was performed with the gage() package and enriched categories were then displayed using REVIGO (Supek, 2011). Significance of the single categories is displayed by color code; GO terms were clustered according to semantic similarity (light blue dashed lines). Analysis and figure by Mazzetto M.

In order to visualize a gene expression signature associated to the different genotypes and common to different tissues, I performed principal component analysis (PCA) using as features the 3768 genes that in the linear model including all tissues were significantly affected by genotype. PCA analysis clearly revealed that the first two components separate samples according to the tissue of origin and the third component (12% of variance explained) separates the samples according to the strain (**Figure 32A**). Further, the vector in PCA space that connects the GRZ-D and MZM-0410 samples has a similar direction for all tissues, included embryo. This result demonstrates that strain-dependent differences in gene expression are conserved across tissues and are detectable already during embryonic development.

To corroborate these results, I analysed 150 libraries of miRNA-seq (Baumgart et al., 2017) obtained from a subset of the same adult samples using generalized linear models. In this case, 101 miRNAs were found to be differentially expressed according to strain. PCA analysis, using the expression level of miRNAs differentially expressed according to the strain variable as features, again reveals the existence of a shared transcriptional signature associated with strain (**Figure 32B**).

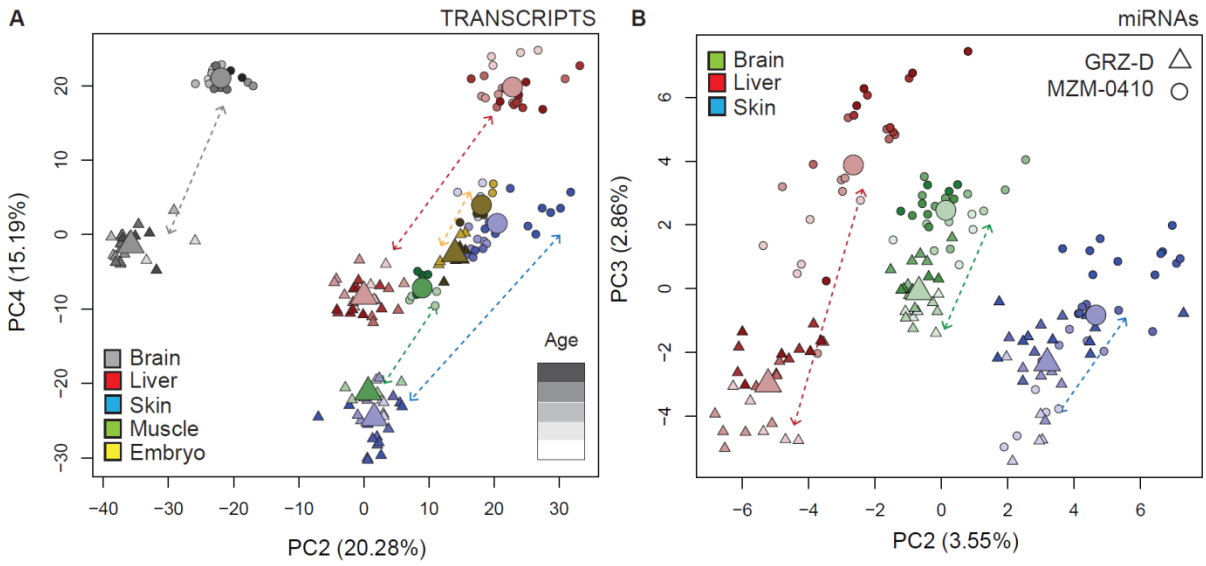


Figure 32. **Visualization of the sample relationships by PCA.** DEGs were obtained with the DESeq2 package both for the transcriptome (A) and for the miRNome (B). In the PCA 3D plot, different strains are coded by shape (triangles for GRZ-D and circles for MZM-0410), different tissues by color and progression from early to older time points is displayed as increasing saturation. Individual samples are represented by small symbols and large symbols represent the centroid of the samples sharing the same strain. Arrows connect clusters corresponding to samples from the same tissue and different strains. In Fig. 3A PC1 was not displayed due to the separation of the samples between embryos and the other tissues observed along this axis. Analysis and figure by Mazzetto M.

2.2. Network analysis reveals splicing and protein modification as central hubs for genetic differences across tissues

Network analysis can be useful to identify hub genes of gene co-regulation. I applied the weighted gene co-expression network analysis (WGCNA) method, which groups genes (or transcripts or proteins) into discrete modules based on their co-expression patterns quantified as topological overlap in a network space where edges are weighted by a power function of correlation of expression levels (Langfelder P., Horvath S., 2008). WGCNA also computes the eigengene for each module as a summary vector of the consistent regulation of genes within a module, which can be related to an external trait (in this case, strain or age). In this work, I decided to perform a variant of the WGCNA method, that is the consensus analysis among different datasets in order to find conserved modules (and genes), and I chose to compare all the available tissues (brain, liver, skin and muscle). I analyzed 4247 transcripts that were detectable in all tissues (with counts per sample > 100); from the analysis, we obtained 10 modules, and for each of them I computed the association with the "strain" variable. I focused our attention on module "turquoise" (**Figure 33A**), which had the highest association with strain (computed from Fisher's metanalysis among each tissue correlation value). The module was shown to be enriched in mRNA processing, transport-related and histone modification categories, as shown by the over-representation analysis (**Figure 33B**). In order to validate these results, I selected

genes with highest intra-modular connectivity and gene-trait significance which are highlighted in **Figure 33A** (orange points), which are related to splicing, (such as SNW1, SF1 and WTAP), and to maintenance of organelle structure (such as GOLGA5). I validated the differential expression of these genes using an independent unpublished RNA-seq dataset comprehensive of 54 GRZ samples and 39 MZM-0403 (another longer-lived strain of *N.furzeri*; Terzibasi et al., 2008) samples (**Figure 33E**), as shown in **Figure 33F**; the differential expression of SPECC1L (involved in the organization of the actin cytoskeleton) and PPP2R2D (a phosphatase involved in the control of mitosis entry and exit) were also additionally validated by qPCR using an independent biological sample of MZM-0410 and GRZ brains (**Figure 33C-D**).

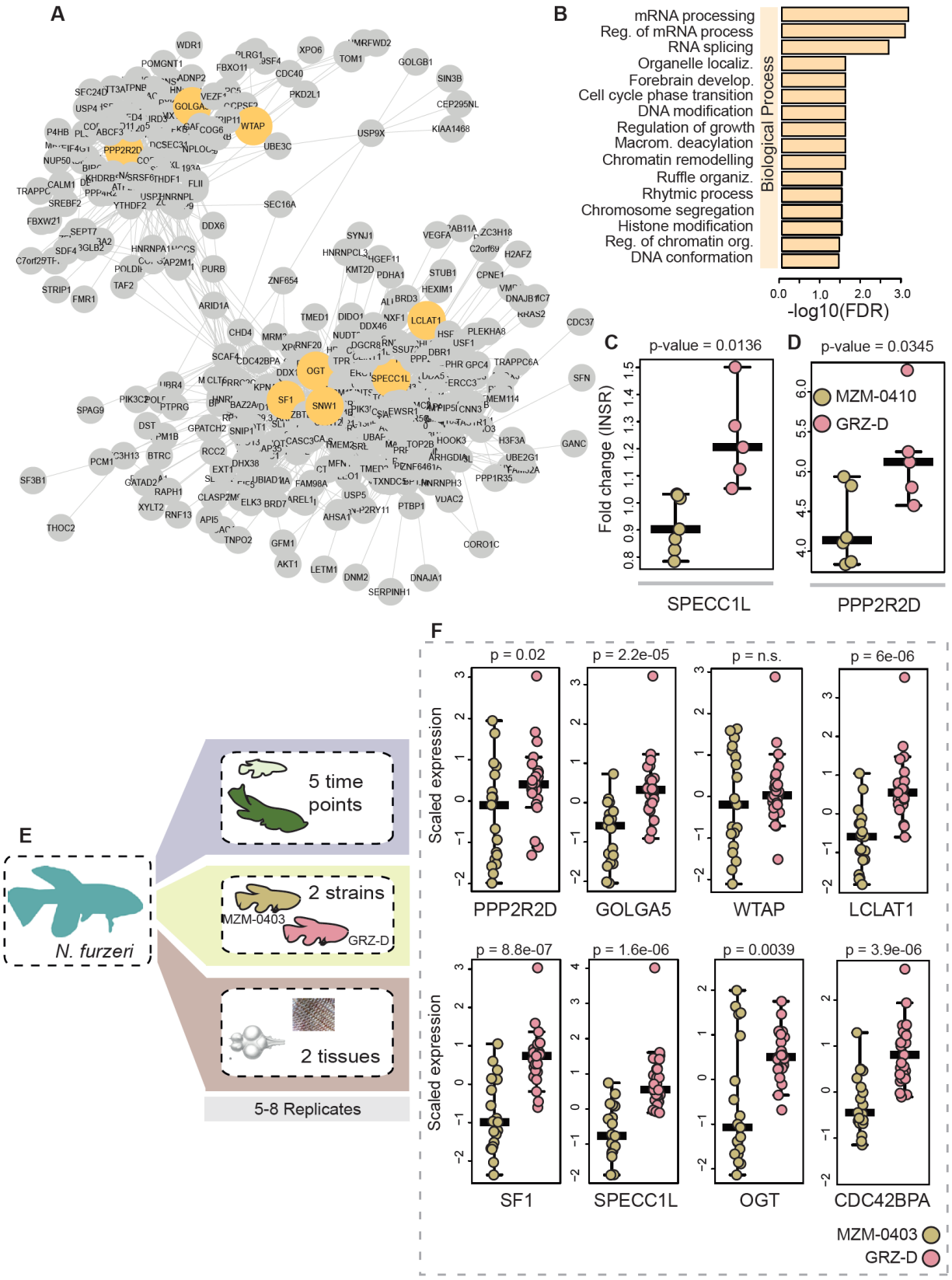


Figure 33. **Consensus network analysis across tissues.** (A) Overview of the module "turquoise": only the center of the module, containing nodes with connection weight > 0.08, was considered for visualization. Nodes highlighted in orange are the genes tested for validations. Embryo was excluded from the above analysis. (B) Over-representation analysis results obtained from WebGestalt: all the categories with FDR < 0.05 are shown for Biological Process and Cellular Component. (C-D)

Validation of hub genes contained in the module, in particular for SPECC1L (C) and PPP2R2D (D): qPCR was performed on MZM and GRZ brain samples (5-weeks old) and Cq values were normalized on INSR. Significance was determined with the Student t-test. (E) Description of the independent RNAseq used for validation. (F) Validation of the genes highlighted in (A) using the RNA-seq dataset shown in Fig. 31E; scaled expression was used for plotting. Statistical significance was computed with Wilcoxon test. Only brain samples were considered for the analysis. Experiments performed by Mazzetto M.; analyses and figures generated by Mazzetto M.

2.3. Young GRZ has an expression profile typical of older MZM-0410

To investigate the effects of age and strain on global gene expression and their relationship, I analysed, for each tissue separately, the transcripts that were detected as differentially expressed both during aging in MZM-0410 and between MZM-0410 and GRZ-D at the first time point analysed (9 weeks for muscle, 5 weeks for the other three tissues) and plotted the log2(fold change) for aging on the X axis and for strain on the Y axis. In all four tissues, the DEGs showed preferentially the same direction of regulation in the two conditions and the data points were therefore concentrated in quadrants I and III (i.e. genes down-regulated with age are also lower-expressed in GRZ and *vice versa*, Fisher's exact test: pvalue < 2.2e-16 for all the 4 tissues) (**Figure 34A-34D).** Similar positive correlations between the effects of age and strain was obtained also when only differential expression according to age in MZM-0410 was used as filtering criterion and in the analysis of miRNA expression (**Figure 36A-36D**, **Figure 37A-37C**).

In order to identify the biological processes that are differentially expressed according to strain and age, I performed GO over-representation analysis for the genes in the four quadrants of **Figure 34A-34D**. Results of the three individual tissues were combined using the Fisher's method of metanalysis where the p-value of each GO represents a summary of the p-value for over-representation in the individual tests. Results are displayed in **Figure 34E**. The genes down-regulated in the short-lived strain and during aging were enriched in categories related to cell cycle, DNA replication and cell proliferation and differentiation, indicating that age-dependent decrease in mitotic activity is anticipated in GRZ-D. Genes up-regulated in GRZ and during aging were found to be enriched in autophagy (**Figure 34F**), protein localization and I-kB/NF-kB signalling categories. Interestingly, genes up-regulated during aging and down-regulated in GRZ show a highly significant enrichment for ncRNA processing and defence response. Genes down-regulated during aging and up-regulated in GRZ are enriched in terms related to development.

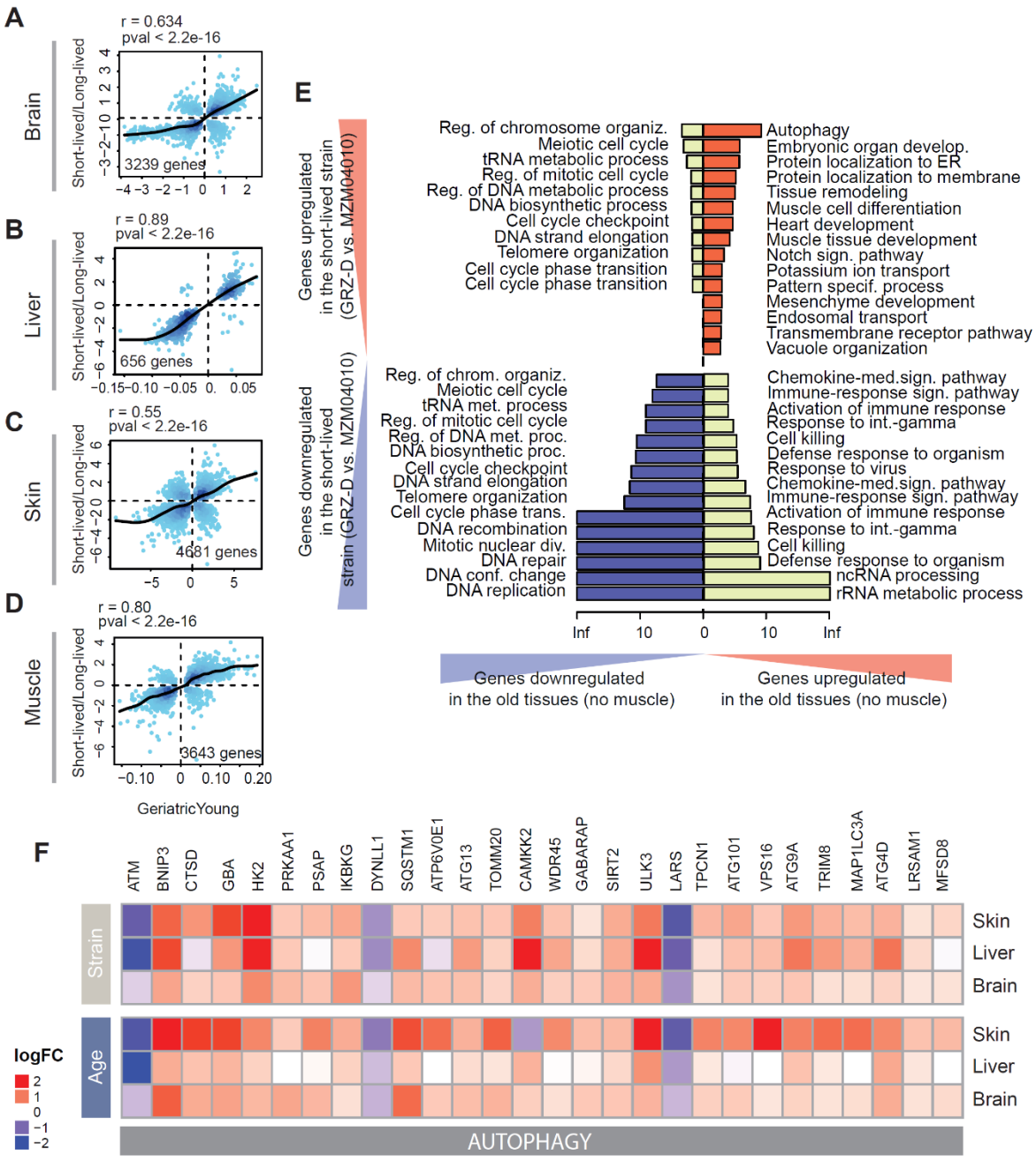


Figure 34. **Relationship between ageing and genetic differences.** (A-D) Comparison between strain and age-dependent DEGs in the transcriptome from A) brain, B) liver, C) skin and D) muscle. Gene regulation is plotted as Log_2FC for Old vs. Young comparison on the X axis and Short-lived vs. Long-lived comparison on the Y axis. Only genes differentially expressed in both conditions (intersection) are shown. Pearson's correlation coefficient, total number of DEGs and number of genes per quadrant are reported. The density of genes is coded by the intensity of the blue hues. (E) GO analysis of the genes contained in the 4 quadrants of the plots showed in Fig. 3A-3D. Gene enrichment was performed for every tissue separately (using GO Biological Process as a functional database) and then a metanalysis was performed using Fisher's method. The categories are plotted as –log10 of the false discovery rate (FDR). (F) Heatmap of autophagy genes across brain, liver and skin in the two strains. Upregulated transcripts are red, downregulated are blue. Analyses and figures by Mazzetto M.

In order to validate these findings, I used the independent RNA-seq dataset used also for **Figure 33E**. This transcriptomic dataset contains samples related to two different strains at different time points: GRZ and MZM-0403 (a genetically different strain). I first used this independent data to validate the top differentially expressed genes between GRZ and MZM-0410 shown in **Figure 30A**. The results are shown in **Figure 35A-D**: the selected genes indeed show a similar expression profile during aging; DND and SLC22A13 are overall more expressed in the GRZ strain, on the other hand ACO2 and MACROD2 are more expressed in the MZM-0403 strain, consistently with the previously results. I then repeated the analysis shown in **Figure 34A-D**. In this case, I analysed the transcripts that were detected as differentially expressed both during aging in MZM-0403 (comparing 287 days-old animals to 21 days-old animals) and between MZM-0403 and GRZ-D at the first time point analysed (21 dph) and plotted the log2(fold change) for aging on the X axis and for strain on the Y axis (**Figure 35B**). In both the comparisons, the DEGs showed preferentially the same direction of regulation in the two conditions (as when comparing GRZ with MZM-0410) and the data points were therefore concentrated in quadrants I and III, with a strong positive correlation (0.5127 for GRZ vs. MZM-0403; 0.604 for GRZ vs. MZCS-0403)**.** The genes down-regulated in the GRZ strain and in old animals (for both the comparisons) were enriched in categories related to genome maintenance and development (reproducing he results shown in **Figure 34**), while genes up-regulated for both the variables were enriched in rRNA metabolic process and protein localization to ER, but the autophagy category does not appear (**Figure 35F, Figure 35I**).

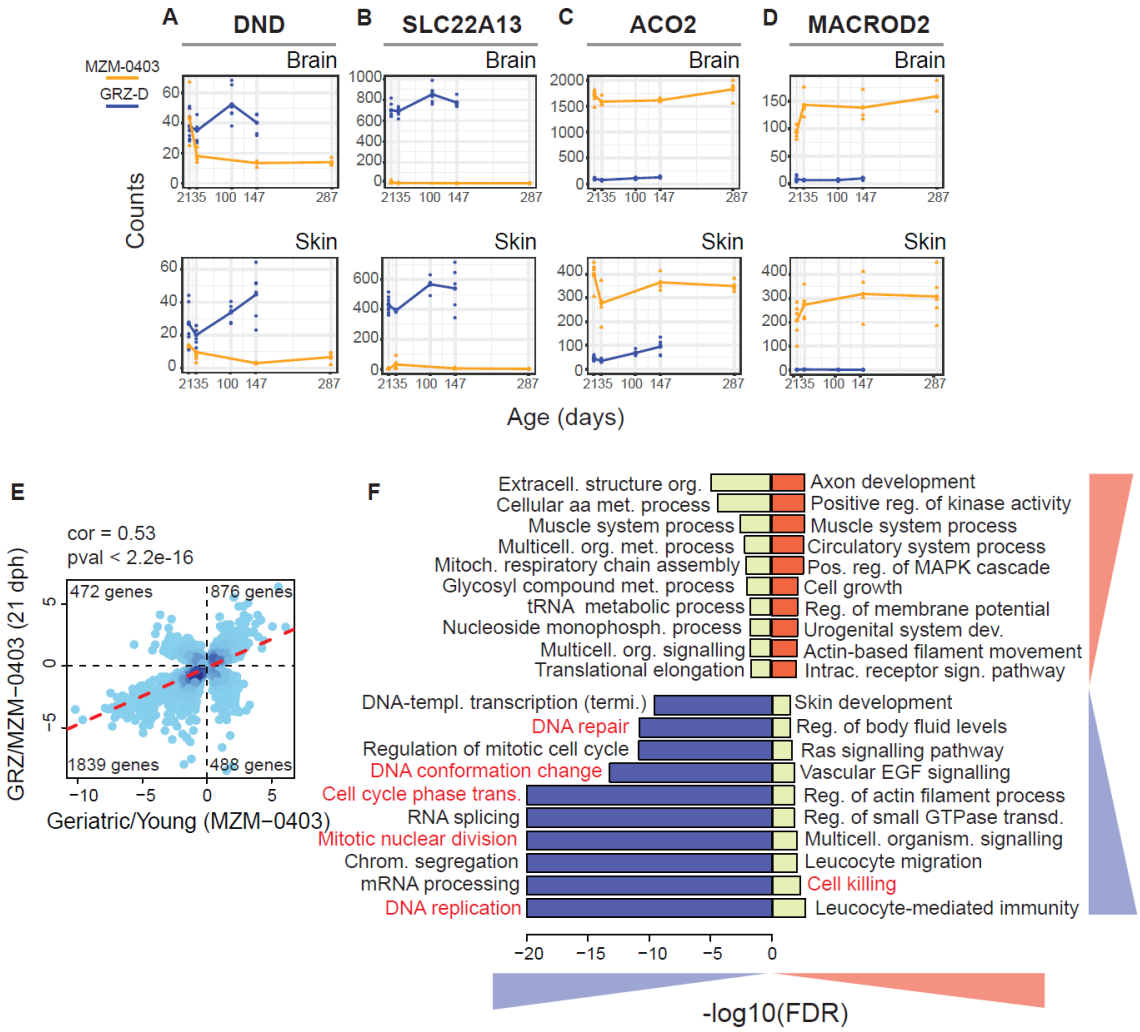


Figure 35. Comparison between strain and age-dependent DEGs in an independent RNA-seq dataset. (A-D) Expression profile of the top strain-specific DEGs using an independent RNAseq dataset: DND (A), SLC22A13 (B), ACO2 (C) and MACROD2 (D). Normalized pseudo-counts were used for plotting expression strength; each point represents an individual case, the line connects the medians of each age group, MZM-0403 (orange) and GRZ (blue). Both brain (upper) and skin (lower) expression data are shown. (E) Comparison between strain and age-dependent skin DEGs of an independent RNA-seq dataset. Gene regulation is plotted as Log_2FC for Old vs. Young comparison (287 days-old MZM-0403 vs. 35 days-old MZM-0403) on the X axis and Short-lived (GRZ) vs. Long-lived (MZM-0403) comparison on the Y axis (at 35 dph). Only genes differentially expressed in both conditions (intersection) are shown. Pearson's correlation coefficient and number of genes per quadrant are reported. The density of genes is coded by the intensity of the blue hues. (F) GO analysis of the genes contained in the 4 quadrants of the plots showed in Fig. 33E. Gene enrichment was performed using WebGestalt software (2017 version). The categories are plotted as –log10 of the false discovery rate (FDR). Common categories between this GO plots and the one showed in Fig. 32E are highlighted in red. Transcriptomic dataset generated by Reichwald K., Koch P. and Groth M. Analyses and figures generated by Mazzetto M.

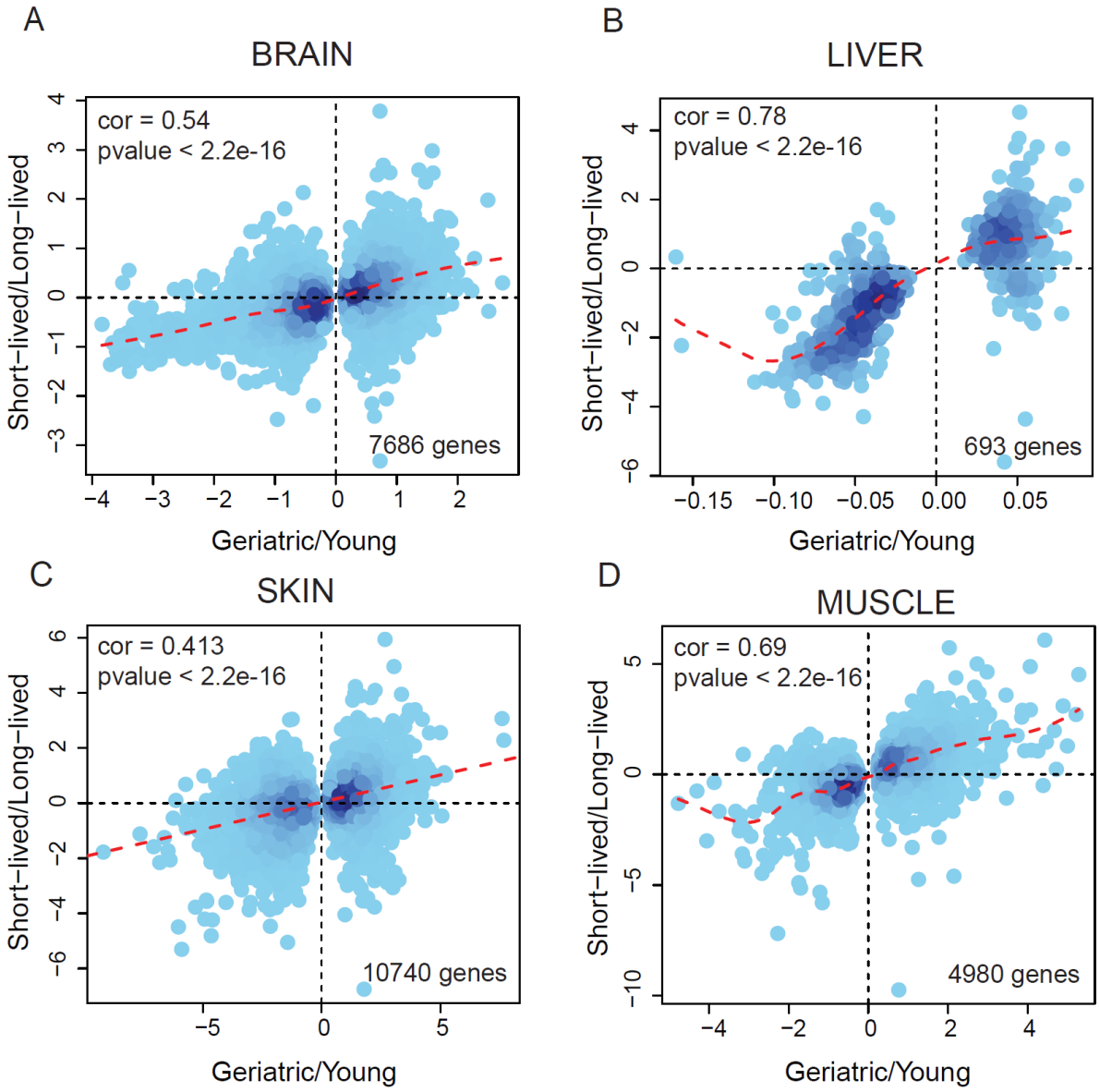


Figure 36. Relationship between ageing and genetic differences using metanalysis. A-D) Scatter plots of aging-dependent fold-changes in MZM-0410 (39 wph vs. 5 wph) vs. strain-dependent fold-changes in the young (5 wph). Only DEGs for the first comparison (Old vs. Young) are plotted and DEGs were computed for each tissue separately. The density of genes is coded by the intensity of the blue hues, the number of genes in each quadrant is reported in the respective vertex. (A) Brain, (B) Liver, (C) Skin and (D) Muscle. Analyses and figures generated by Mazzetto M.

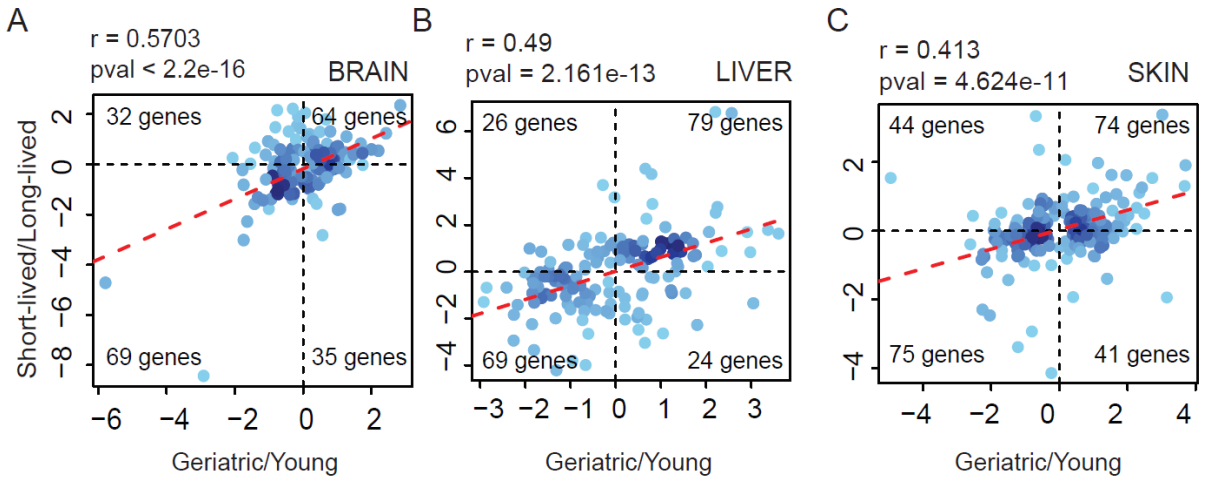


Figure 37. Comparison between strain and age-dependent DEMs in the miRNome of brain (A), liver (B) and skin (C). Gene regulation is plotted as Log_2FC for Old vs. Young comparison on the X axis and Short-lived vs. Long-lived comparison on the Y axis. Genes differentially expressed in either of the two conditions (union) are shown. Pearson's correlation coefficient, total number of DEGs and number of genes per quadrant are reported. The density of genes is coded by the intensity of the blue hues. Analysis and figure by Mazzetto M.

In addition, for brain, liver and skin, I analysed age-dependent gene-expression in both strains during the same chronological interval by intersecting DEGs at 5 weeks vs. 12 weeks in either strains and performing a linear correlation analysis. The large majority of genes were again observed in quadrants III, resulting in highly significant positive correlations (Pearson R: brain = 0.66, liver = 0.29 and skin = 0.631). In all three cases, the slope of the regression line was clearly smaller than 1, indicating that –within the analysed time window- the rate of age-dependent gene expression is smaller in GRZ (**Figure 38A-38C**).

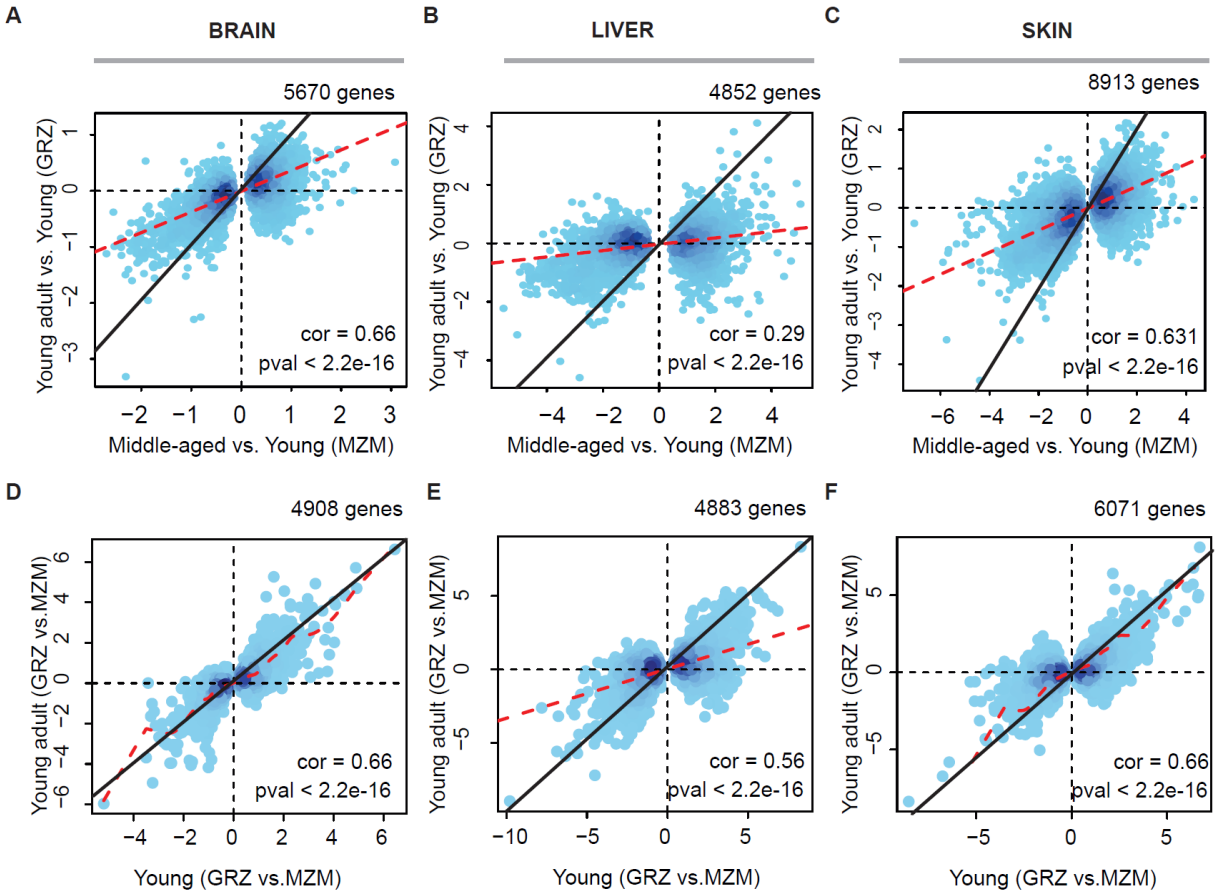


Figure 38. Additional comparisons. A-C) Scatter plots of aging-dependent fold-changes in MZM-0410 vs. aging-dependent fold-changes in GRZ. Only DEGs for the first comparison (Middle-aged vs. Young) are plotted and DEGs were computed for each tissue separately. The density of genes is coded by the intensity of the blue hues, the number of genes in each quadrant is reported in the respective vertex. (A) Brain, (B) Liver and (C) Skin. D-F) Scatter plots of strain-dependent fold-changes in the young (5 wph) vs. strain-dependent fold-changes in the middle-aged (12 wph). Only DEGs for the first comparison (Young GRZ vs. Young MZM) are plotted and DEGs were computed for each tissue separately. The density of genes is coded by the intensity of the blue hues, the number of genes in each quadrant is reported in the respective vertex. (D) Brain, (E) Liver and (F) Skin. Analysis and figure by Mazzetto M.

2.4. Proteome analysis confirms differential expression results

In order to corroborate these analyses at the protein level, I performed mass-spectrometry-based proteomics on an independent sample of five young (5 weeks-old) GRZ brains and I compared strain-dependent genes at the transcriptomic and proteomic level. The results indicate a clear positive correlation between differentially expressed proteins and transcripts in the comparison of the two strains at 5 weeks (**Figure 39A**).

I then performed GO over-representation analysis for the genes in the four quadrants of **Figure 39A**; the results of the enrichment are shown in **Figure 39B**. Genes that are upregulated both at the transcriptomic and proteomic level in the GRZ strain show enrichment in splicing and mRNA

processing, consistently with what shown previously at the RNA level: these pathways, which show to have a central role in genetic differences, have the same role also at the protein level.

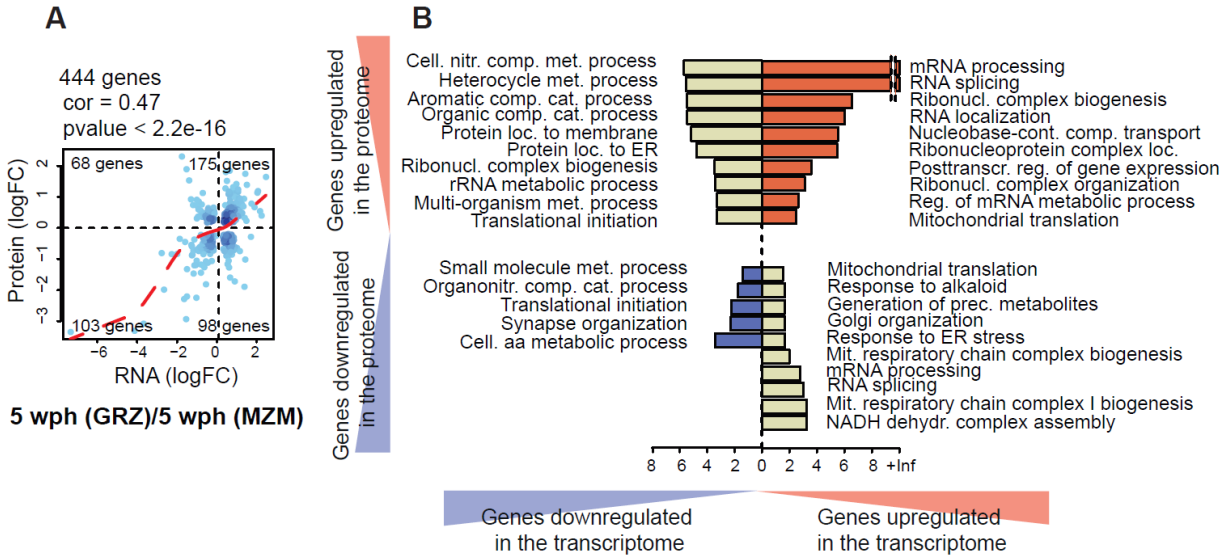


Figure 39. Relationship between proteome and transcriptome of strains with different genetic backgrounds. (A) Comparison between strain-dependent DE transcripts and proteins in the brain of *N.furzeri*. LogFC values for DE transcripts are shown on the X axis and for DE proteins on the Y axis. Only genes that are DE both at the transcript and protein level were plotted. Pearson's correlation and total number of genes, as well as number of genes per quadrant are shown. The density of genes is coded by the intensity of the blue hues. (B) GO analysis of the genes contained in the four quadrants of Fig. 37A. Gene enrichment was performed using GO Biological Process as a functional database. The categories are plotted as –log10 of the false discovery rate (FDR).

I then plotted age and strain-related differentially expressed proteins, as done previously at the transcriptomic level (**Figure 34A-D**): I used adult (12 weeks-old) and young (5 weeks-old) brain samples from the long-lived strain for the aging comparison, and young GRZ vs. MZM-0410 brain samples for the strain comparison. The results indicate a positive correlation between differentially-expressed age-dependent and strain-dependent proteins (**Figure 40A**); GO over-representation analysis for the genes in the four quadrant showed enrichment in autophagy and endosomal transport, consistently with what shown previously at the RNA level (**Figure 40B**). These pathways, which have a central position in genetic differences, have the same differential expression at the protein level.

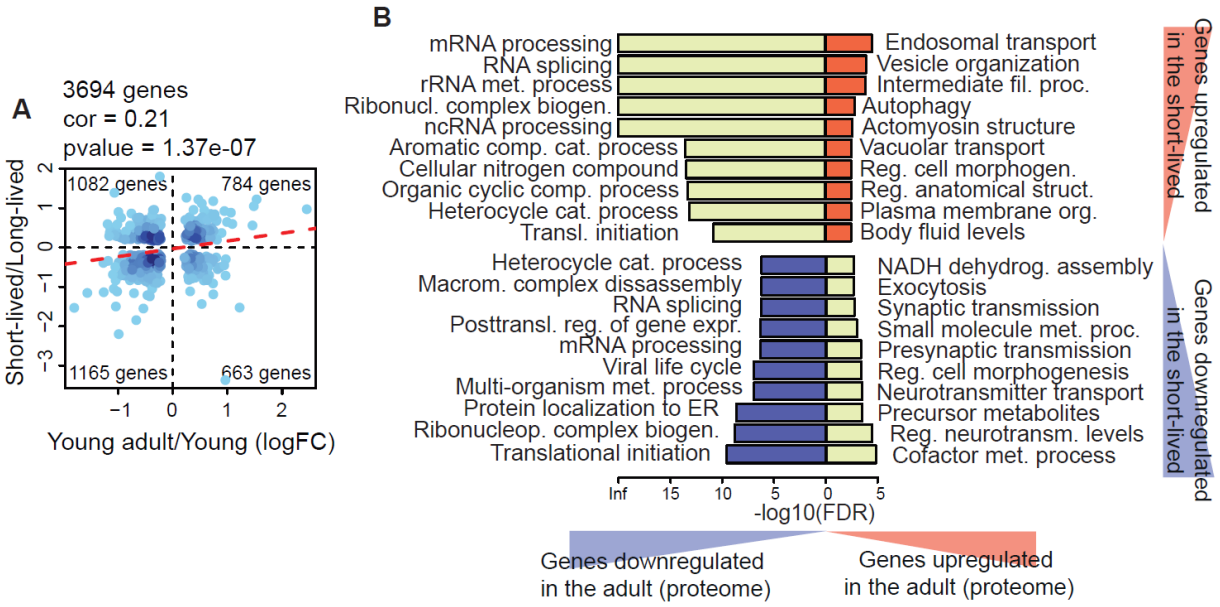


Figure 40. Relationship between *N.furzeri* ageing and genetic differences at the proteomic level. (A) Comparison between age and strain-dependent proteins in the brain of *N.furzeri*. LogFC values for age-dependent proteins are shown on the X axis and for strain-dependent proteins on the Y axis. Only proteins that are significant for both the comparisons were plotted. Pearson's correlation and total number of genes, as well as number of genes per quadrant are shown. The density of genes is coded by the intensity of the blue hues. (B) GO analysis of the genes contained in the four quadrants of Fig. 38A. Gene enrichment was performed using GO Biological Process as a functional database. The categories are plotted as –log10 of the false discovery rate (FDR). Proteomic datasets generated by Kelmer Sacramento E. and Kirkpatrick J. Analyses and figures by Mazzetto M.

2.5. Aging biomarkers reveal an "anticipated" aging profile in GRZ strain

In order to highlight possible differences in expression among aging biomarkers between the two strains, transcripts with either negative or positive monotonic expression profile were analyzed: in Baumgart, Priebe, Groth et al. paper (2016) a fuzzy c-means clustering was performed in order to isolate transcripts with monotonic expression profile in at least 2 tissues from the longer-lived strain. I decided to use the gene list of each cluster and plotted the scaled expression value for each strain, as shown in **Figure 41**. Transcript expression profiles show that MZM biomarkers in GRZ have an "anticipated aging" profile: in particular when looking at the expression level in the first time point (5 wph) we could strikingly notice that the transcripts display a lower expression in GRZ strain for negative biomarkers (**Figure 41A-41B**), but a higher expression when looking in positive biomarkers (**Figure 41C-41D**), whereas slopes are reduced in both cases.

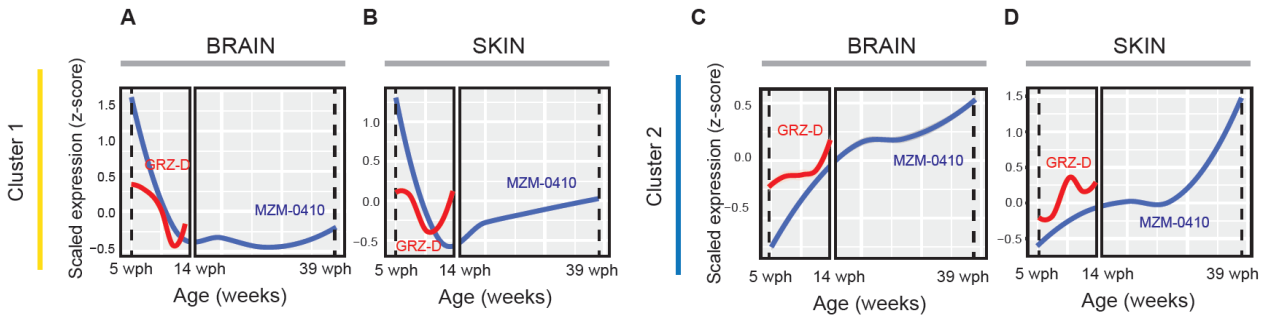


Figure 41. Comparison of MZM-specific aging biomarkers in both the strains (GRZ-D coded as red and MZM-0410 coded as blue) for brain (A, C) and skin (B, D). The gene list for each cluster was obtained from Baumgart, Priebe, Groth et al.: clusters 1 (A-C) shows negative monotonic expression profile in all the analysed tissues, while clusters 2 (B-D) shows positive monotonic expression profile. Scaled expression values are plotted with the regression line using ggplot2 package. Analyses and figures by Mazzetto M.

2.6. Cox-Hazard model reveals genes with antagonistic effects on lifespan which show a strain/age-dependent expression profile

In order to isolate predictive factors for lifespan, I reanalyzed the longitudinal dataset described in Chapter 1.5.

Cox-Hazard model was applied separately at the time points where fin clips were performed and then the beta coefficient obtained from the analysis were compared (**Figure 42A**): as shown from the plot, the majority of factors were concentrated in quadrant II and IV, with a β of opposite signs at the two tested ages. Since factors with positive sign (as log(b)) are risk factors (i.e. higher expression is associated with higher mortality risk) and factors with negative sign are protective factors (i.e. lower expression is matched with higher mortality risk), this means that the expression at the majority of observed factors is beneficial at early time points, but detrimental at later time points (quadrant II), or the opposite (quadrant IV).

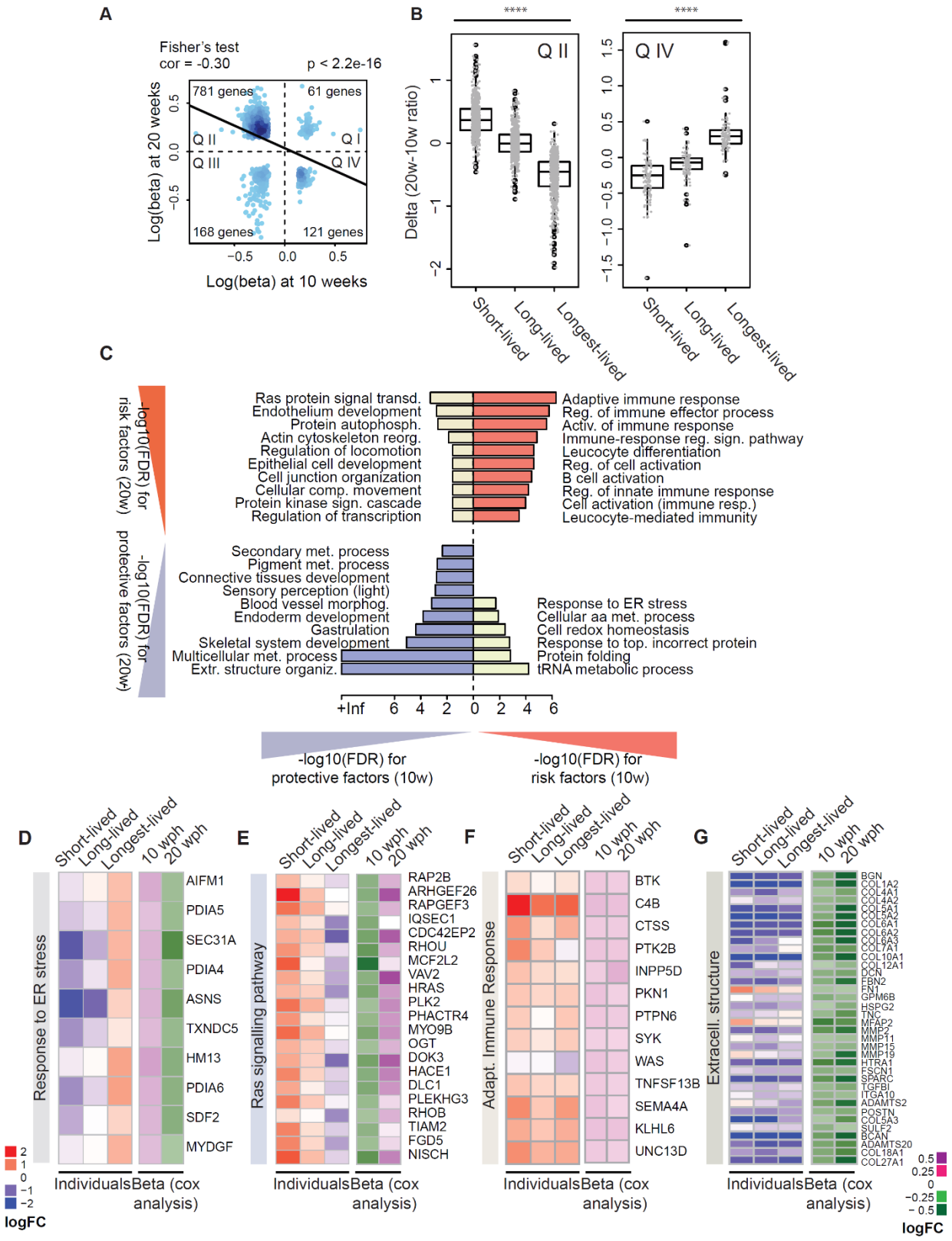


Fig. 42. Cox-Hazard model on a longitudinal dataset. (A) Cox-Hazard coefficients visualization for analyses at 10 wph and 20 wph: significant factors for both the analyses were filtered on pvalue<0.05. Fisher's test results as well as total number of factors are displayed. (B) Expression profile (shown as 20w/10w ratio) of factors contained in Quadrant II (left plot) and factors contained in Quadrant IV (right plot) in different classes of individuals (from Baumgart, Priebe, Groth et al.): short-lived, long-lived and longest-lived animals were taken into account, as individuals with lifespans respectively between 28-36 wph (short-lived), 45-50 wph (long-lived) and 57-71 wph

(longest-lived). Statistical analysis was performed with ANOVA test. (C) GO analysis of the genes contained in the 4 quadrants of the plot showed in Fig. 18B. Gene enrichment was performed using GO Process as a functional database; the categories are plotted as –log10(FDR). (D-G) Enriched categories for factors with opposite effects on mortality risk. Categories which were found to be enriched both from Cox regression analysis are displayed as heatmaps, in particular response to ER stress (D), Ras signaling pathway (E), adaptive immune response (F) and extracellular structure organization (G). Expression (as ratio 20w/10w) in short-lived, long-lived and longest-lived animals is displayed as gradient from red (upregulated) to white (neutral) to blue (downregulated), while beta coefficients (from Cox analysis) are displayed as gradient from purple (beta>0) to green (beta<0). Analyses and figures by Mazzetto M.

I also analyzed the expression profile (as 20w-10w ratio) of genes in Quadrant II or IV for different classes of individuals (**Figure 42B**). As shown in Baumgart, Priebe, Groth et al. paper, animals were categorized into three different classes related to their lifespan: short-lived with age of death between 28 and 36 weeks, long-lived with age of death between 45 and 50 weeks and finally longest-lived with age of death between 57 and 71 weeks. Genes contained in Quadrant II of **Figure 42A** increased expression with age in short-lived animals and decreased expression in the longest-lived animals; on the other hand, genes contained in Quadrant IV displayed decreased expression with age in short-lived animals and increased expression in longest-lived animals, similar to what described in Chapter 1.5 for the proteasome.

GO analysis was then performed on the factors contained in each quadrant of the plot (**Figure 42C**). Risk factors concordant between the two analyses (quadrant I) were enriched in categories related to immune response (as expected, since upregulation of the immune system at old age can induce chronic inflammation with increased mortality; Katharina Simon et al., 2015). On the other hand, beneficial/protective factors (quadrant III) were found to be highly enriched in extracellular structure organization. Factors that showed opposite behaviour between the two time points were enriched in response to ER stress, protein folding and homeostasis (quadrant IV), and Ras protein signal transduction (quadrant II), as shown also when regulation of all genes of the single categories are displayed (**Figure 42D-42F**): categories that have beneficial effect at early time points and detrimental effects at later time points show a progressive decrease in expression in long- and longest-lived animals (Ras signalling and protein phosphorylation, **Figure 42E-42F**) while categories that show the opposite effect on lifespan are progressive increased in long- and longest-lived individuals (response to ER stress, **Figure 42D**).

To identify risk/protective factors that are age-dependent, the results obtained from this analysis were combined with the ones obtained from the JenAge RNAseq dataset (Baumgart, Priebe, Groth et al.), as shown in Part 2 of the Results.

The genes shown in **Figure 42** were displayed on the plots shown in **Figure 34** for brain, liver and skin (**Figure 43A, 43C** and **43D**). From the plots, it is evident that factors contained in Quadrant IV

are downregulated both in old animals and in short-lived animals; on the other hand, factors with the opposite profile (Quadrant II) show an upregulation both in old and short-lived animals. These results could also be confirmed using the second dataset presented in **Figure 33E** (**Figure 43E**).

As control for the analysis, I also looked for factors that show a concordant effect on lifespan during aging (i.e. genes found in Quadrant I and III of **Figure 42A**) in the same transcriptomic dataset. The results are shown in **Figure 44**: a lower number of genes could be detected for all the tissues, with no particular enrichment in one specific quadrant of the plot.

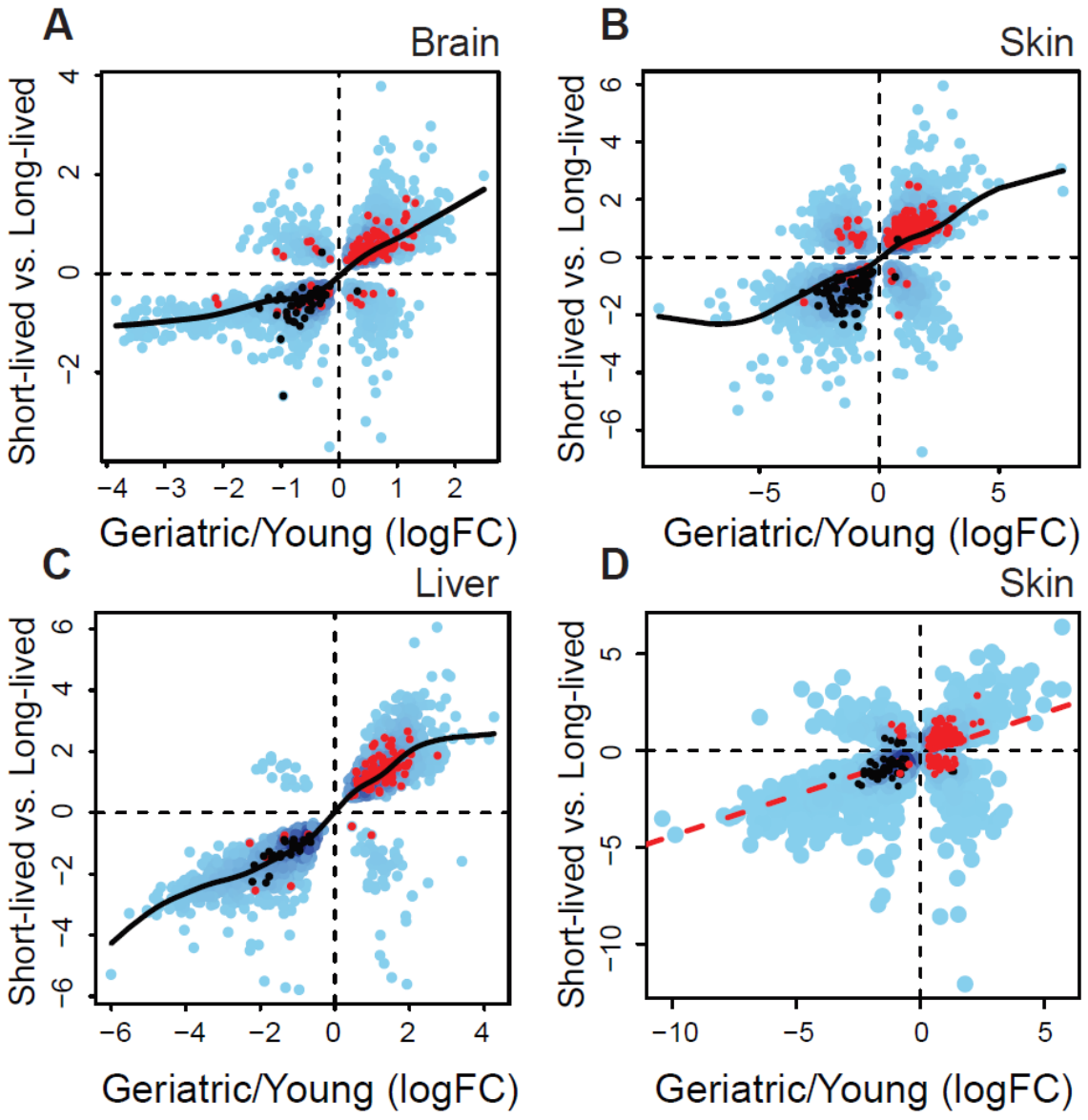


Figure 43. Aging- and strain-dependence of factors with discordant effects on lifespan. (A-D) Comparison between strain and age-dependent DEGs in the (A) brain transcriptome, (B) brain proteome, (C) liver transcriptome and (D) skin transcriptome and visualization of factors with opposite effects on mortality risk. The JenAge transcriptomic dataset (Fig. 21A) was used. (D) Comparison between strain and age-dependent DEGs in the skin transcriptome of the dataset presented in **Fig. 33E**. For all the plots LogFC value for Old vs. Young and Short-lived vs. Long-

lived samples is displayed. The density of genes is coded by the intensity of the blue hues, while factors in quadrant II and IV (Fig. 40A) are plotted as red and black dots respectively. Only DEGs for both the variables (intersection) were considered for plotting. Analyses and figures by Mazzetto M.

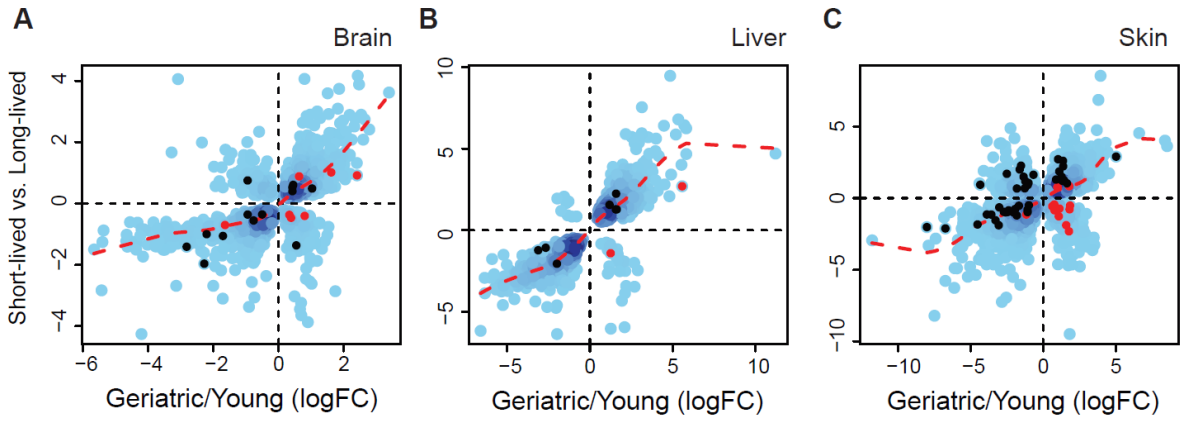


Figure 44. Aging- and strain-dependence of factors with concordant effects on lifespan. (A-C) Comparison between strain and age-dependent DEGs in the (A) brain, (B) liver and (C) skin transcriptome and visualization of the factors contained in Quadrant I (red) and Quadrant III (black). LogFC value for Old vs. Young and Short-lived vs. Long-lived samples is displayed. Metanalysis was performed in order to obtain a higher number of genes: only cases with pvalue<0.05 (combined and adjusted for FDR) were used for plotting. Analyses and figures by Mazzetto M.

2.7. Comparison with a long-lived model: the Naked Mole Rat

The previous results suggest that baseline molecular differences between short- and long-lived strains are correlated with age-dependent differences in the longer-lived strain or species. To test for the generality of the conclusions above, I analyzed RNA-seq data from skin of naked mole rat (*Heterocephalus glaber*), a notorious example of exceptional longevity and compared these with RNA-seq data for the guinea pig, which represents a shorter-lived close relative. Indeed, a recent paper centered on analysis of liver proteome in these two species revealed a correlation pattern similar to one we report in **Figure 35A-35D**. I therefore analysed a public dataset of aging DEGs in the naked mole rat and DEGs obtained with the contrast naked mole rat vs. guinea pig at young age in liver and skin and displayed these as 2D plots of the log2(fold change) in either condition obtaining a significant positive correlation (**Figure 45**). The same result was obtained by analysing the union of the DEGs for skin (**Figure 46A**).

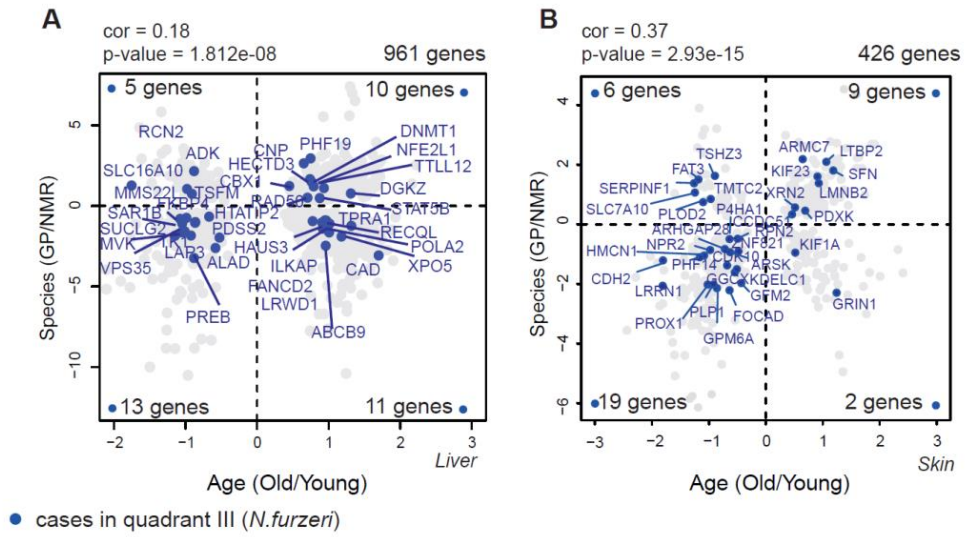


Figure 45. Comparison between *N.furzeri* and *Heterocephalus glaber* in the liver (A) and in the skin (B). (A) Scatterplot of age and strain DEGs in the Naked mole rat taken from liver samples (from Heinze, Bens, Calzia et al., 2018); LogFC for Old vs. Young and Long-lived vs. Short-lived comparisons are displayed. Genes with padj < 0.01 for both the comparisons were considered as significant. Pearson correlation and total number of significant genes are also shown in the plot. Colored dots correspond to genes found in *N.furzeri* comparisons in the same tissue: in particular blue dots are genes downregulated in both the variables. (B) Scatterplot of age and strain DEGs in the Naked mole rat taken from skin samples; LogFC for Old vs. Young and Long-lived vs. Short-lived comparisons are displayed. Genes with padj < 0.01 for both the comparisons were considered as significant. Pearson correlation and total number of significant genes are also shown in the plot. Colored dots correspond to genes found in *N.furzeri* comparisons in the same tissue: blue dots are genes downregulated in both the variables. Analyses and figures by Mazzetto M.

In order to compare the relationship between aging and genetic background between *N.furzeri* and Naked mole rat, I highlighted the genes that in *N.furzeri* were significantly downregulated both during aging and in GRZ-D strain (blue dots) (**Figure 35A-35D**). These are more likely to be also downregulated during aging in the Naked mole rat and in the guinea pig (quadrant II), with a Fisher's test pvalue = 2.45e-03 (**Figure 45, Figure 46A**). The mentioned conserved genes (quadrant II in **Figure 35B**) seem to be related to development and growth (such as PROX1 and NPR2), cell division/cytokinesis (such as HMCN1), neural stem cells differentiation and quiescence (such as GPM6A and CDH2), nervous system structure and development (such as PLP1 and LRRN1), and signal transduction (such as ARHGAP28).

In order to identify the biological processes that are differentially expressed according to genetic background and age in the Naked Mole Rat, I performed GO over-representation analysis for the genes in the four quadrants of **Figure 46A**. The results (**Figure 46B**) show that genes down-regulated in the Guinea Pig and in old animals are enriched in categories related to membrane potential and metabolism. Interestingly, genes that are up-regulated during aging and in Guinea Pig show a highly significant enrichment for RNA processing and splicing (**Figure 46C**). These results are consistent

with the other results obtained from the killifish transcriptome: RNA splicing is shown not only as a central pathway when comparing strain with different lifespan (**Figure 33**), but it also reproduces the results shown in **Figure39**, which shows splicing as top upregulated category also in the proteome.

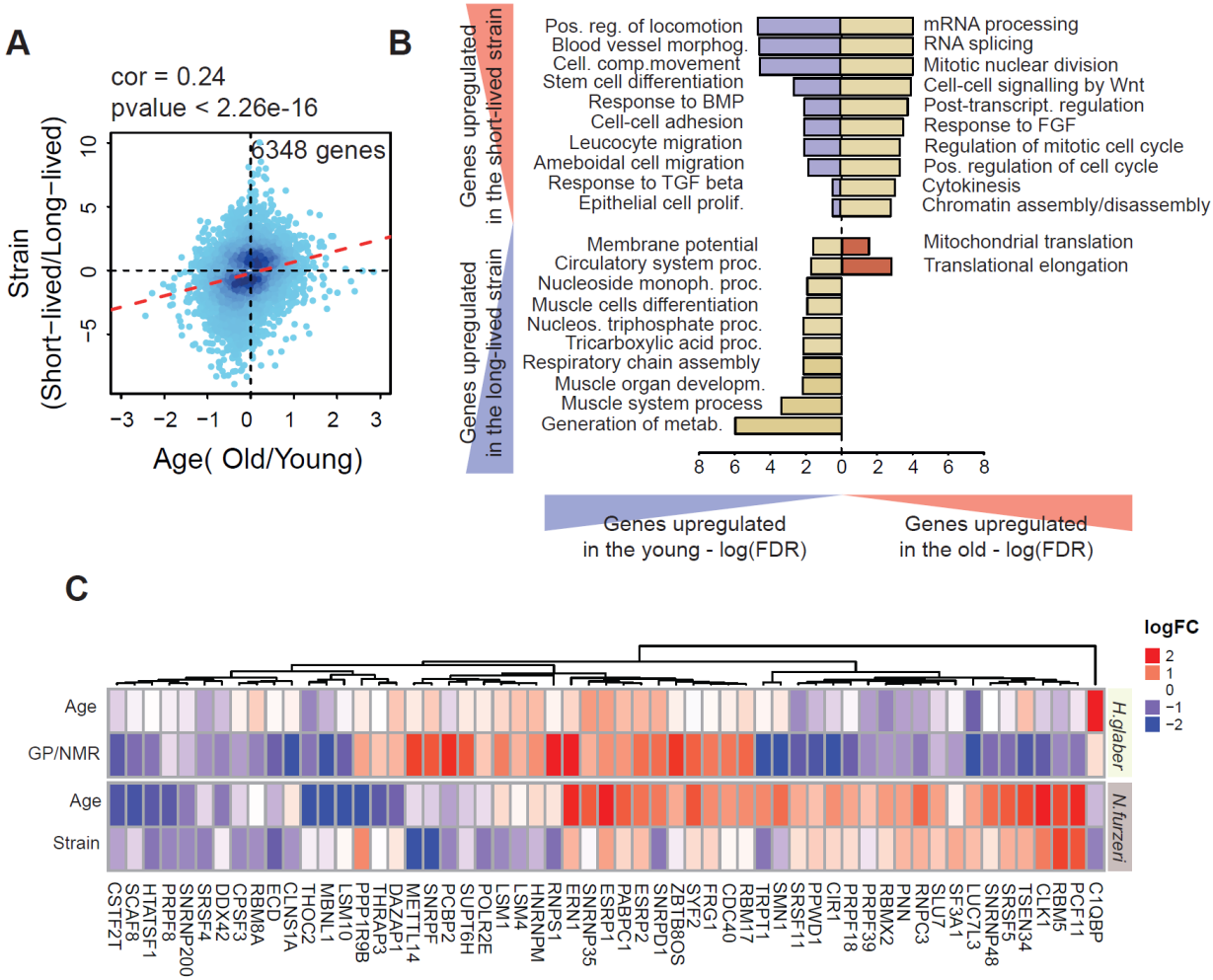


Figure 46. Enriched categories in the Naked Mole Rat skin transcriptomes across ageing and genetic differences. (A) Comparison between *N.furzeri* and Naked Moled Rat DEGs in the skin (A). LogFC value for Old vs. Young and Short-lived vs. Long-lived samples is displayed: filtering was performed on the combined pvalue between the two variables using meta-analysis. Pearson correlation and total number of genes, as well as number of genes per quadrant are shown. The density of genes is coded by the intensity of the blue hues. (B) GO analysis of the genes contained in the four quadrants of Fig. S7A. Gene enrichment was performed using GO Biological Process as a functional database. The categories are plotted as –log10 of the false discovery rate (FDR). (C) Heatmap of splicing-related genes in *Heterocepalous glabrus* and *Nothobranchius furzeri*. Only genes significant for both the comparisons and in both species are displayed. Upregulated genes (with logFC>0) are coloured as red boxes while downregulated genes (with logFC<0) are displayed as blue boxes. Analyses and figures by Mazzetto M.

3. Conserved and divergent correlates of brain aging between wild and captive *N.furzeri*

3.1. Differences in growth rate between wild and captive

For this study, I decided to focus on the *N.furzeri* A41 population (Ch1, described by Vrtilek et al., 2018), which belongs to the Chefu phylo-geographic lineage and it therefore genetically closest to MZM-0410. The pool was sampled at different times during the season (**Figure 47**). The time of flooding could be determined by direct observation and animals of specified ages could therefore be collected. Brains were collected in the wild and preserved, RNA-seq was performed and the results compared with those of the MZM-0410 aging dataset already mentioned in Part 2.

A striking difference between wild and captive fish is the much larger size of the former at young ages (5 wph) and an early cessation of growth, as shown in **Figure 48**: at 39 days post hatching (dph) wild animals already cease growing, and this process is shown in both females and males, while captive animals show a continuous growth (until 35 weeks, as shown in **Figure 48A**).

I therefore proceed to sequence a dataset obtained from wild *N.furzeri* brains comprehensive of three time points (39 dph, 73 dph and 108 dph) from both sexes.

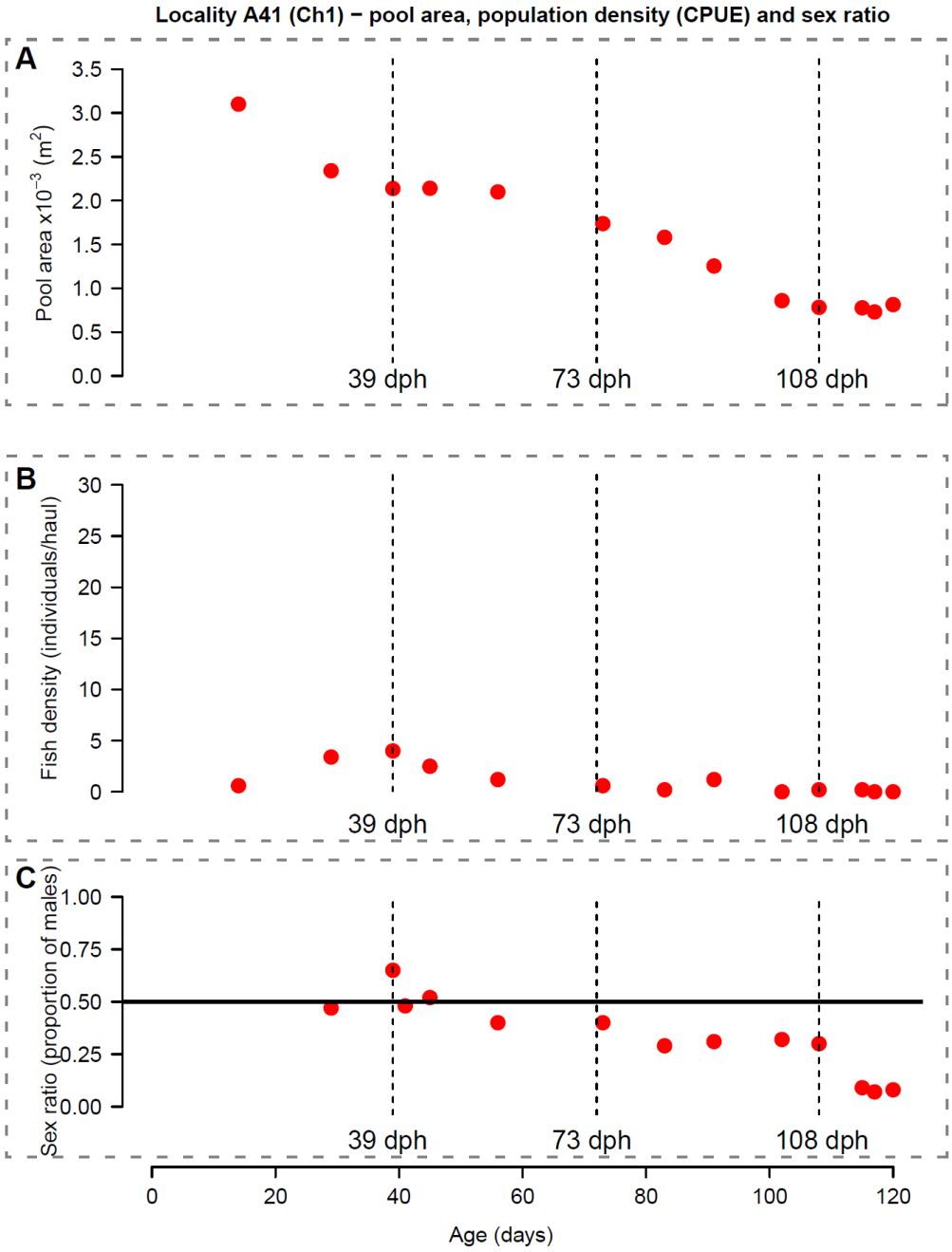


Figure 47. **Information on pool area (A), fish density (B) and sex ratio (C) across time.** Dashed lines indicate the three groups of samples (39dph, 73dph and 108dph, both males and females) in all the plots. Figure generated by Reichard M.

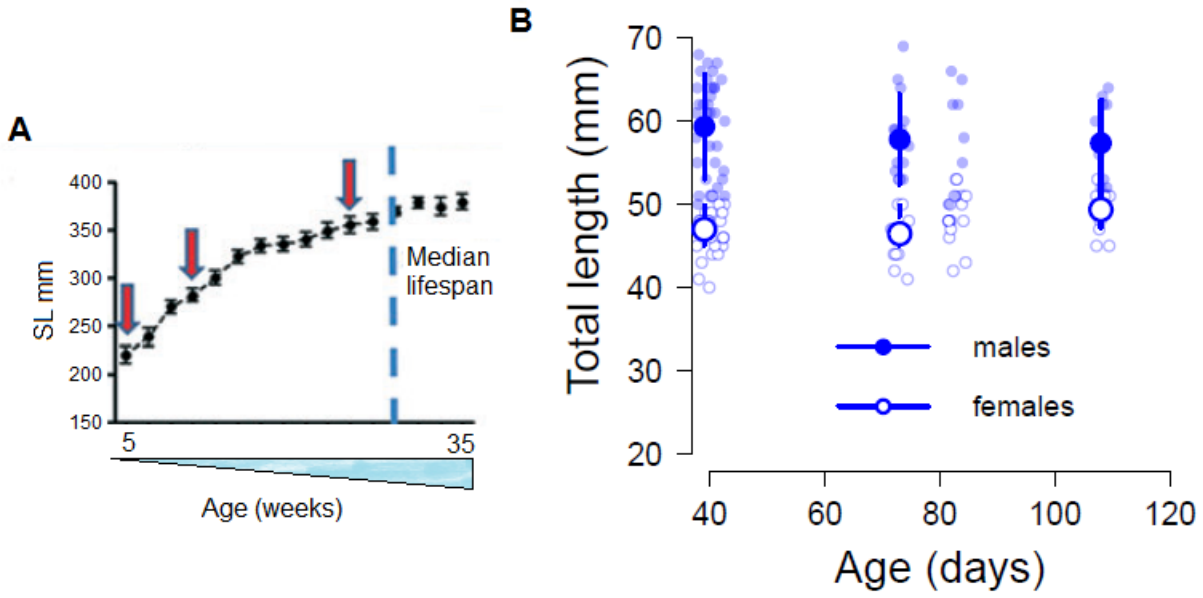


Figure 48. Growth rate in captive and wild animals during aging. (A) Somatic growth in captive *N.furzeri* males, modified from Tozzini et al., 2012. (B) Growth shown as body length at different time points in Ch1 (blue) wild *N.furzeri* populations. Both females and males are shown. Figures generated by Terzibasi Tozzini E. (A) and Reichard M. (B).

3.2. Sex differences in gene expression are minor in wild animals

Since a sex bias in mortality was observed in wild populations of *N.furzeri* (Vrtilek et al., 2018), I computed differentially expressed genes (DEGs) in all the pairwise comparisons (73 dph vs. 39 dph, 108 dph vs. 73 dph and 108 dph vs. 39 dph) for each sex separately, and no major gender-related differences in the number of DEGs transcripts could be detected (**Figure 49A**). I then computed putative aging biomarkers defined as genes with a monotonic trend for each sex independently. Markers selected in one sex have a very similar expression profile in the opposite sex (**Figure 49C**), demonstrating that sex has little influence on the genome-wide regulation of transcripts by age in wild animals.

When a generalized linear model is applied to isolate the effect of sex (independent of the aging variable), only 36 sex-dependent DEGs were detected. Among these, we found SYBU (syntabullin), which is the nearest-neighbour gene of GDF6 (growth differentiation factor 6), the sex-determining locus of *N. furzeri* (**Figure 49B**), and CYP19A1, a gene that catalyzes the formation of aromatic C18 estrogens from C19 androgens, demonstrating that the significant cases are mostly sex-related genes.

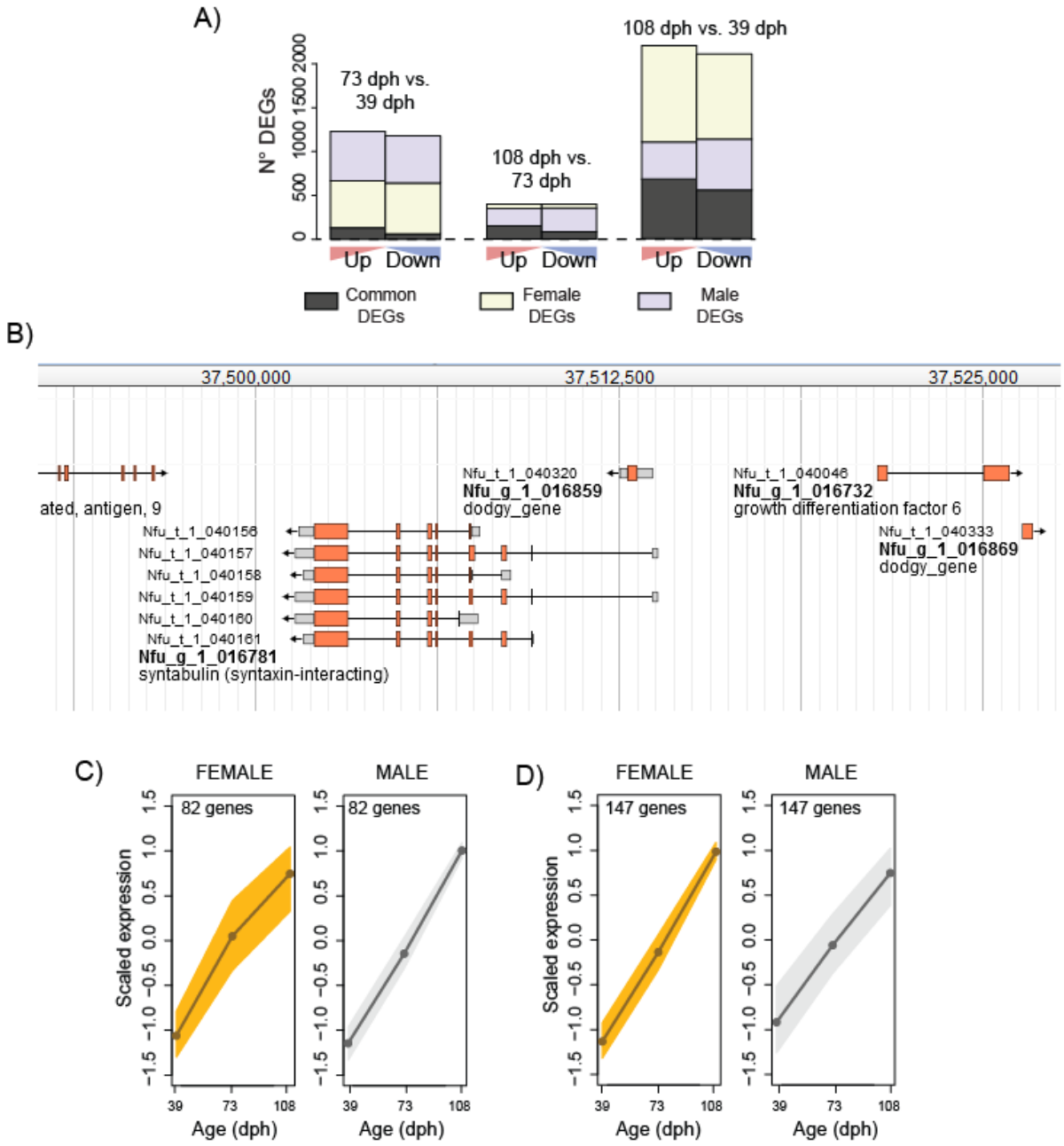


Figure 49. Comparison between sexes in *N.furzeri* wild brain. (A) Number of DE transcripts for different aging comparisons in male and female; three aging comparisons were considered (73 dph vs. 39 dph, 108 dph vs. 73 dph and 108 dph vs. 39 dph), and DE cases were isolated for padj < 0.05. Common cases are displayed as grey, female DE cases are displayed as yellow and male DE cases are displayed as violet. (B) *N.furzeri* genome localization of SYBU. (C-D) Aging biomarkers (as transcripts with a positive monotonic trend) in male (C) and female (D). Significant cases were obtained with pvalue < 1e-04 from Cuzick analysis. Analyses and figures by Mazzetto M.

3.3. Aging is characterized by the downregulation of mitosis-related and the upregulation of DNA replication genes in wild animals

Two aging comparisons were performed in the wild animals: for each sex separately and combining male and female samples. I analysed brain genes that were detected as differentially expressed both during maturation (73 dph vs. 39 dph) and during aging (108 dph vs. 73 dph) and plotted the log2 (fold change) for the first comparison on the X axis and the log2 (fold change) for the second one on the Y axis. The DEGs showed a positive correlation between the two variables and the data points were shown to be more concentrated in quadrants I and III (i.e. genes with negative monotonic trend and *vice versa*, Fisher's exact test: pvalue < 2.2e-16) (**Figure 50A**), indicating the progressive nature of age-dependent expression changes

In order to identify the biological processes that are most affected during aging of wild fish, I performed GO overrepresentation analysis for the genes in the four quadrants of **Figure 50A**. Results are displayed in **Figure 50B**. The genes downregulated during both maturation and aging were enriched in categories related to cell cycle, mitotic nuclear division and synapse organization, indicating that there is an age-dependent decrease in mitotic activity and synaptogenesis. Genes with opposite behaviour also showed overrepresentation of specific categories: genes up-regulated during aging and down-regulated in development for mRNA processing, RNA splicing, and protein folding, and genes downregulated during aging and up-regulated in development for CNS development. Among the genes with negative monotonic trend I found CDC20, which codes for a cytoskeletal protein required for ubiquitin ligase activity of the anaphase promoting complex/cyclosome (APC/C), and SMC2, which codes for a nuclear protein which is central in the condensing complex (a complex required for conversion of interphase chromatin into mitotic-like condense chromosomes) (**Figure 50C**). Among the genes with positive monotonic trend I noticed ATR, coding for a serine-threonine protein kinase acting as central DNA damage sensor, MCMBP, coding for an associated component for the MCM complex that acts as regulator of DNA replication, and GINS1, required for the correct functioning of the GINS complex (a complex that plays an essential role in the initiation of DNA replication) (**Figure 50D**).

I then performed a consensus analysis of gene co-expression networks among wild and captive aging data in order to find conserved modules (and genes). I analyzed 4751 transcripts that were expressed in both datasets (with counts > 500 to avoid low-count cases); gene co-expression modules were identified using weighted gene co-expression analysis (WGCA) and for each module obtained from the analysis we computed the correlation between the eigengene of the module and the "age" variable. I focused our attention on the module which had the highest correlation with age (**Figure 50E**). The

module was shown to be enriched in RNA splicing, mRNA processing and autophagy categories (**Figure 50B**), as shown from the over-representation analysis (**Figure 50F**). In order to validate these results I selected two genes, WTAP and GOLGA5 (**Figure 50G**), and quantified their expression by qPCR.

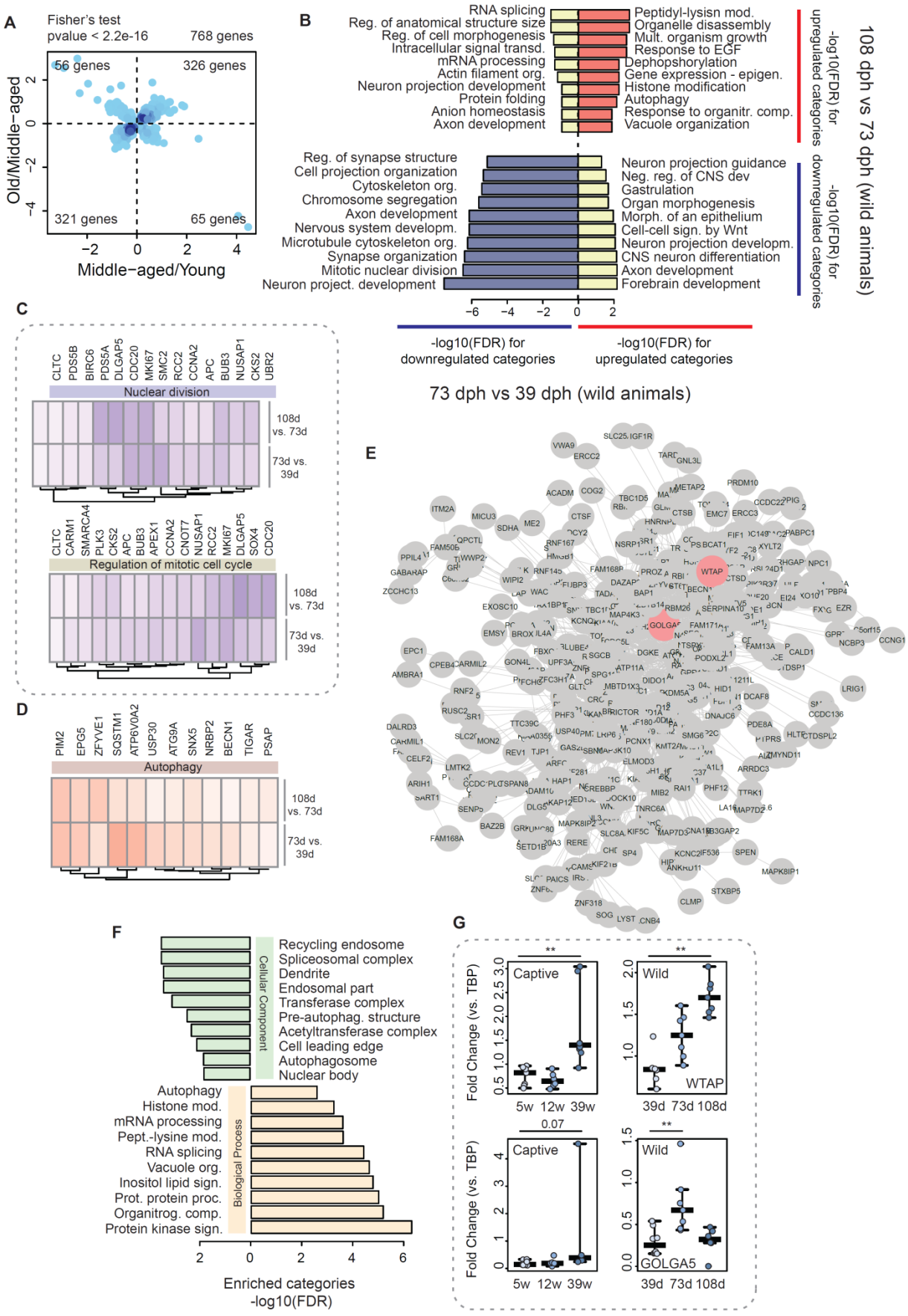


Figure 50. Aging in *N.furzeri* wild brain. (A) LogFC value for 73 dph vs. 39 dph DE transcripts (x variable) and 73 dph vs. 108 dph DE transcripts (y variable) is displayed. Fisher's test results and total number of genes, as well as number of genes per quadrant are shown. The density of genes is

coded by the intensity of the blue hues. Only significant DE cases for both the comparisons was considered, filtering on padj<0.05 (B) GO analysis of the genes contained in the 4 quadrants of the plot showed in Fig. 14A. Gene enrichment was performed using GO Process as a functional database; the categories are plotted as –log10(FDR). (C) Heatmaps of transcripts contained in quadrant I related to "Nuclear Division" and "Regulation of mitotic cell cycle" GO categories; logFC values of transcripts with padj<0.05 for both the comparisons were considered for plotting. (D) Heatmaps of transcripts contained in quadrant III related to "Autophagy" and "DNA replication" GO categories; logFC values of transcripts with padj<0.05 for both the comparisons (red means upregulated, blue downregulated). (E) Overview of the module "turquoise": only the center of the module, containing nodes with connection weight > 0.1, was considered for visualization. Nodes highlighted in orange are the genes tested for validations. (F) Over-representation analysis results obtained from WebGestalt: all the categories with FDR < 0.05 are shown for Biological Process and Cellular Component. (G) Validation of selected genes. WTAP (upper box) and GOLGA5 (lower box) were chosen for technical validation; captive and wild samples are shown at all the time points, as well as t-test statistical results. Experiments performed by Mazzetto M.; analyses and figures generated by Mazzetto M.

3.4. A scission between S and M phase categories during maturation is observed comparing wild and captive animals

GAGE was applied in order to detect pathways whose expression is condition-dependent between wild and captive animals: genes upregulated in wild animals showed an enrichment of categories related to protein synthesis, oxidative phosphorylation and antigen processing. On the other hand, genes upregulated in captive animals were found to be enriched in categories related to synapse formation and organization, as well as synaptic transmission and protein folding (**Figure 51A**). Among the top differentially regulated genes, I found WEE2, a protein tyrosine kinase which plays as a key regulator during meiosis, FTL, a ferritin component, LAMTOR1 and CSMD3 (involved in dendrite development), whose differential expression was validated through quantitative PCR (**Figure 51B-E**).

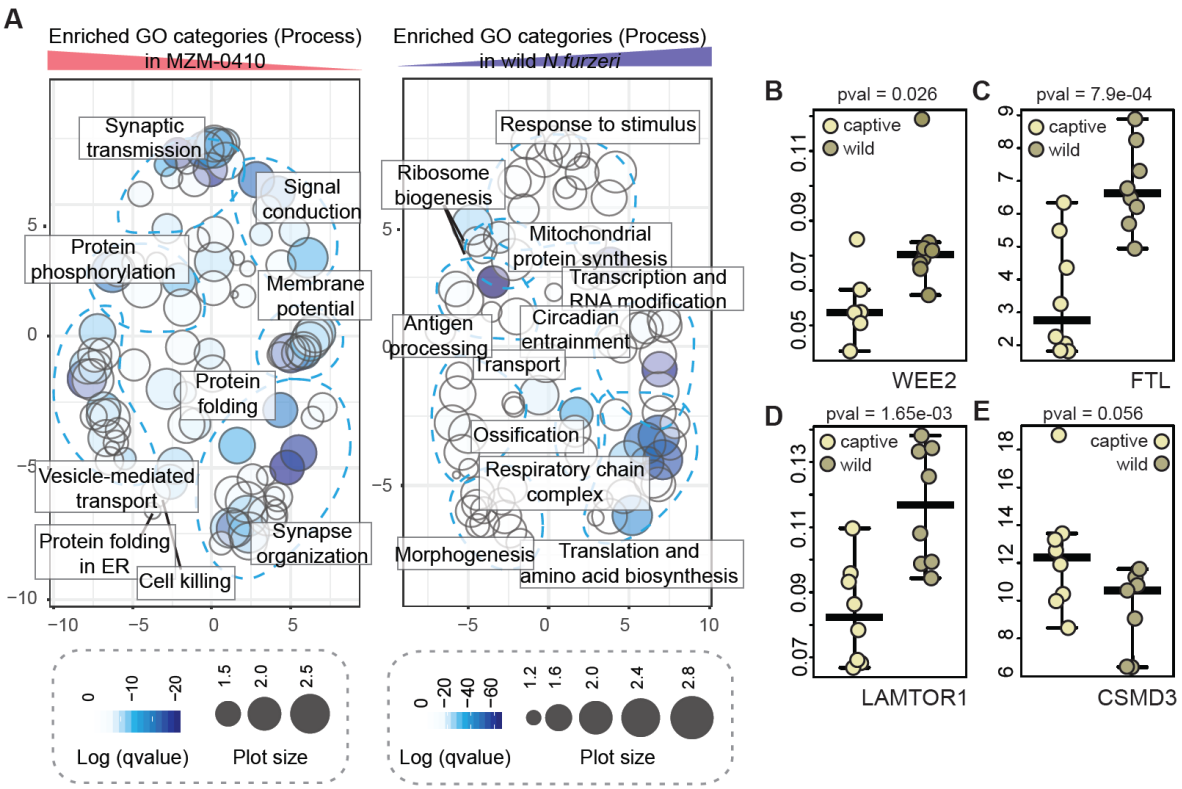


Figure 51. Comparison between wild and captive *N.furzeri* brain. (A) GAGE analysis of differentially expressed transcripts between wild and captive brain independent of the age variable. GO process categories are displayed; significant pathways were obtained filtering for qvalue < 0.05. (B-E) Validation of WEE2 (B), FTL (C), LAMTOR1 (D), and CSMD3 (E) was performed by qPCR. Captive and wild samples were used independently from the time points. Student t-test pvalues are also shown. Experiments performed by Caterino C. and Mazzetto M.; analyses and figures generated by Mazzetto M.

Principal Component Analysis (PCA) was then performed in order to visualize relationships between samples and the effects of age on global gene expression. As expected, the first component separates the conditions; the second component identifies an effect of age on global gene expression that is similar in the two conditions. It should also be noted that samples corresponding to the youngest wild animals are shifted towards more negative values of PC2 that correspond to older ages for the captive animals (**Figure 52A**).

In order to compare aging in the two conditions, I analysed genes that were detected as differentially expressed during aging in both the conditions and plotted the log2 (fold change) for aging in the captive animals on the X-axis and for aging in the wild animals on the Y-axis: in particular, I plotted the two aging comparisons with the highest correlation. The DEGs showed preferentially the same direction of regulation in the two conditions and the data points were concentrated in quadrants I and III with a high positive correlation (**Figure 52B**).

In addition, I performed GO overrepresentation analysis for the genes in the four quadrants of **Figure 52B**. Results are displayed in **Figure 52C**: the genes upregulated during aging in both wild and captive animals were enriched in categories related to autophagy, and vacuole organization. Genes that are upregulated in the captive but downregulated in the wild during aging were found to be enriched in protein secretion and cell organization categories.

Interestingly, genes down-regulated during aging in both the conditions show a highly significant enrichment for proteasome processes, cytokinesis and mitotic nuclear division categories, while, on the other hand, genes downregulated in the captive but upregulated in the wild during aging are enriched in DNA replication, DNA repair and telomere organization categories, highlighting a scission process between S phase and M phase categories in the wild during adult life (**Figure 52C** and **Figure 53**).

Among the genes that are downregulated in both the conditions, I focused my attention on ANLN, required for cytokinesis and essential for the structural integrity of the cleavage furrow, and KIF11, a motor protein required for establishing a bipolar spindle during mitosis (**Figure 53F-G**).

Among the genes that are downregulated in captive animals but upregulated in wild animals I focused on DNA2 (a key enzyme involved in DNA replication and DNA repair in nucleus and mitochondrion), XRCC2 (protein involved in the homologous recombination repair pathway of double-stranded DNA) and PRIM1 (subunit of the DNA primase complex) (**Figure 53B-D**). Differential expression of these genes was technically validated by qPCR on an independent biological sample (**Figure 53B-D; 53F-G**).

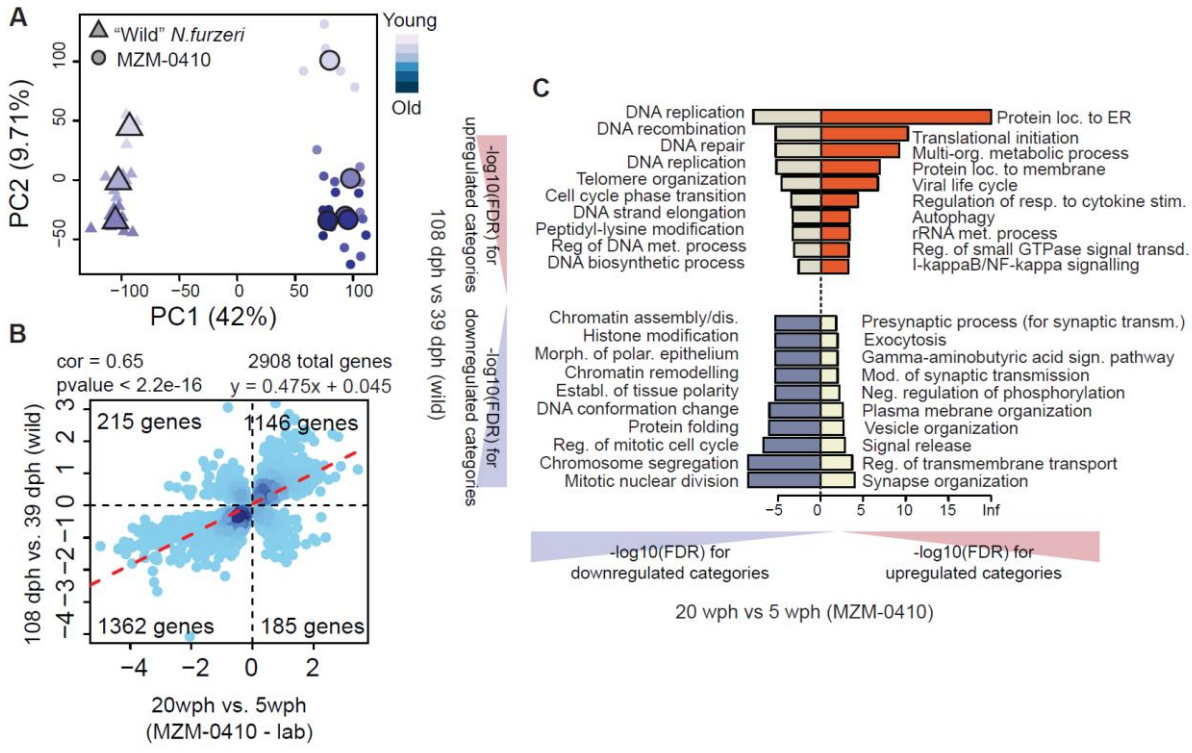


Figure 52. Comparison between captive and wild ageing. (F) PCA of wild and captive samples. Samples were corrected for batch effect and visualized using prcomp() function; wild (circles) and captive (triangles) samples are shown together, while ages are visualized as color shades. Centroids are displayed as bigger points, while the replicates are the smaller points. (G) Scatterplot of DE transcripts between wild and captive. Aging pairwise comparisons were considered for the two conditions (12 wph vs 5 wph for captive and 108 dph vs. 39 dph for wild) and only the intersection of the cases (so transcripts that were found to be significant for both the comparisons, with padj<0.05) was considered. Pearson correlation as well as total number of transcripts is displayed. (H) GO analysis of the genes contained in the 4 quadrants of the plot showed in Figure 45G. Gene enrichment was performed using GO Process as a functional database; the categories are plotted as –log10(FDR). Analyses and figures by Mazzetto M.

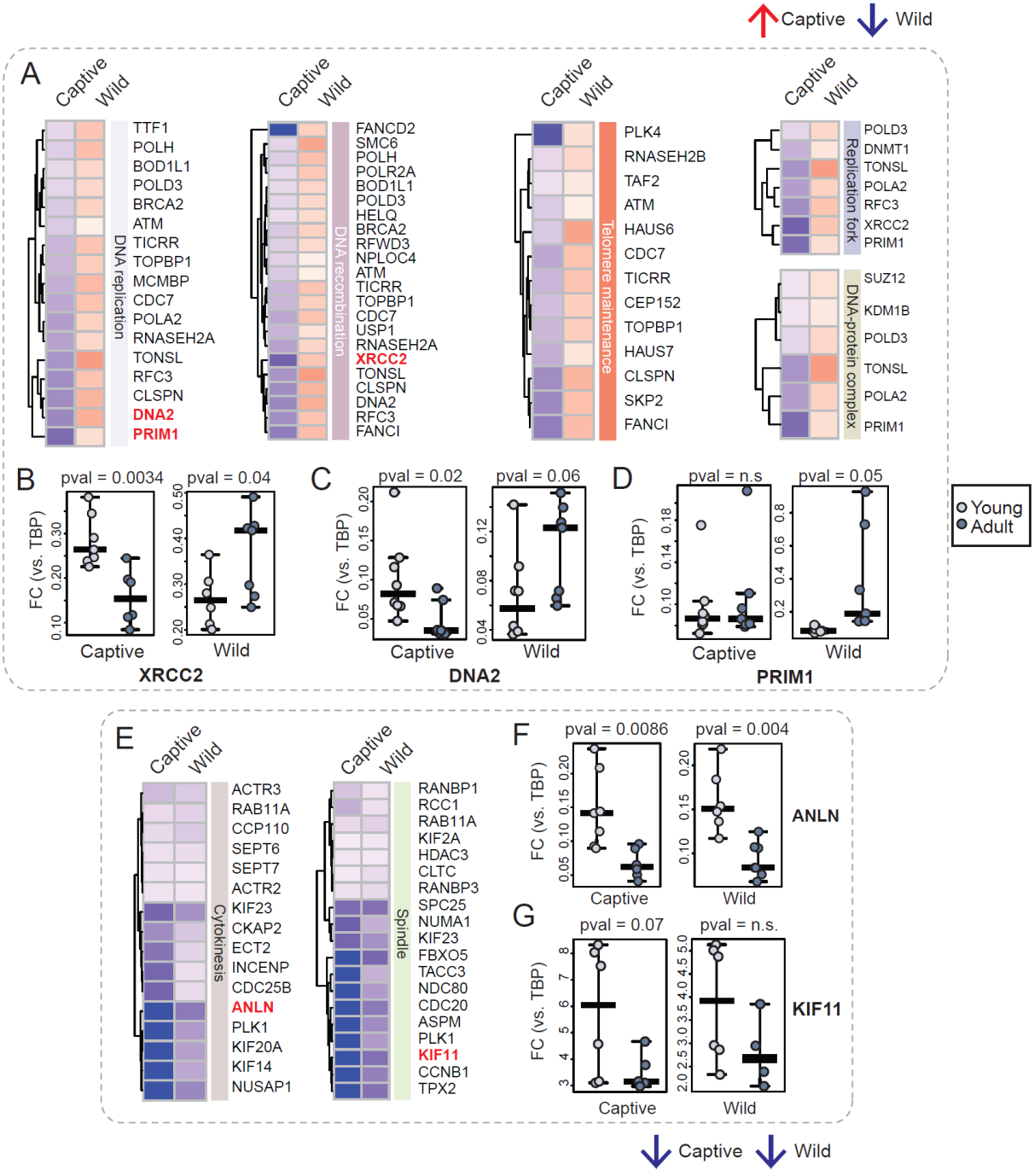


Figure 53. Visualization of interesting pathways related to Fig. 16. (A) Heatmap of pathways (DNA replication, DNA recombination and Telomere maintenance) and components (Replication

fork and DNA-protein complex) that are downregulated in the captive but upregulated in the wild during aging. (B) Heatmap of pathways (Cytokinesis) and components (Spindle) that are downregulated both in the captive and in the wild during aging. (C) Validation of XRCC2 (C), ANLN (D), KIF11 (E), DNA2 (F), and PRIM1 (G) was performed by qPCR. Captive and wild samples were used independently from the time points. T-test pvalues are also shown. Experiments performed by Mazzetto M.; analyses and figures by Mazzetto M.

3.5. Aging biomarkers reveal an "anticipated" aging profile in wild animals

In order to highlight possible differences in expression among aging biomarkers between the two conditions, transcripts with either negative or positive monotonic expression profile were considered: for captive animals the clusters obtained from Baumgart, Priebe, Groth et al. were considered, for wild animals monotonic genes were computed using the Cuzick method (**Figure 54**). I decided to use the gene list of each cluster and plotted the scaled expression value for each condition. Transcript expression profiles show that captive biomarkers in wild animals have a similar trend (**Figure 54A-B**): however, wild samples show an "anticipated aging" profile in relation to the captive. When looking at the expression level in the first time point (5 wph), I can notice that the transcripts display a lower expression in wild samples for negative biomarkers (**Figure 54A**), and more clearly a higher expression for positive biomarkers (**Figure 54B**).

Plotting the expression of wild aging biomarkers in captive animals reveals that they show an "accelerated aging" profile in comparison to the captive animals, so that the slope of change is steeper both for positive and negative biomarkers (**Figure 54D**).

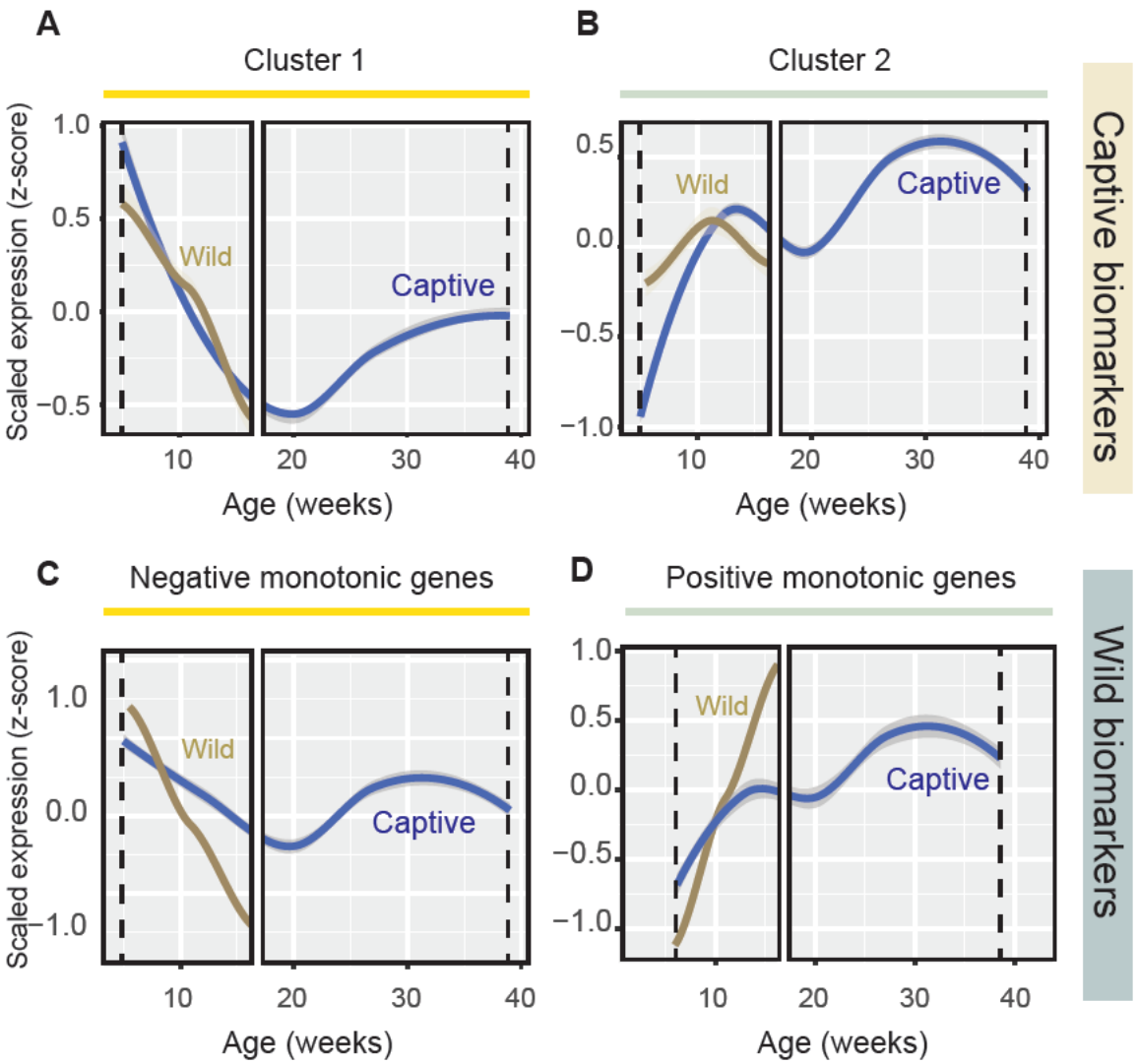


Figure 54. Comparison of captive-specific aging biomarkers in both wild and captive animals ("wild" coded as gold and "captive" coded as blue). (A-B) Expression profile of captive biomarkers in both the conditions. The gene list for each cluster was obtained from Baumgart, Priebe, Groth et al.: clusters 1 (A) shows negative monotonic expression profile, while clusters 2 (B) shows positive monotonic expression profile. Scaled expression values are plotted with the regression line using ggplot2 package. (C-D) Expression profile of wild biomarkers in both the conditions. Monotonic genes were obtained using the Cuzick method; clusters 1 (C) shows negative monotonic expression profile, while clusters 2 (D) shows positive monotonic expression profile. Scaled expression values are plotted with the regression line using ggplot2 package. Analyses and figures by Mazzetto M.

Discussion

1. Characterizing *N.furzeri* brain proteome: a molecular timeline for aging

The results shown in Part 1 of the Results delineate a timeline of events associated with loss of protein quality control upon aging. An early event, detectable already in adult fish, is a decreased proteolytic activity of the proteasome, which is driven by a downregulation of transcripts coding for components of the 19S and 20S complexes in adult fish brain as compared to young fish (**Figure 26A** and **26B**). The amplitude of this downregulation correlated with individual lifespan and could therefore represent an early biomarker of aging (**Figure 26E**). The reduction of proteasome activity precedes chronologically the decoupling of transcript/protein levels, suggesting a causative role for this aspect of the aging process. Decreased proteasome activity can lead to the accumulation of proteins that are synthesized in excess relative to their binding partners, thus causing a stoichiometric imbalance of protein complexes (McShane et al., 2016). For instance, deletion of the ubiquitin ligase Tom1 (yeast homologue of Huwe1), which is responsible for the labelling for degradation of overproduced ribosomal proteins, leads to accumulation of multiple ribosomal proteins in detergent-insoluble aggregates in yeast (Sung et al., 2016). More recently, it was shown that the aggregation of non stoichiometric complex members is an alternative process to compensate dosage imbalance in yeast and that aggregation and proteolytic processing of excess proteins are mutually exclusive (Brennan et al., 2019). Accumulation of ribosomes in detergent-insoluble aggregates (David et al., 2010) and a loss of stoichiometry in the proteasome (Walther et al., 2015) were previously reported to occur during aging in *C. elegans.* Our work demonstrates the conservation of these mechanisms in the vertebrate brain, by showing alteration of stoichiometry in several large protein complexes (**Figure 20C** and **20D**) and aggregation of ribosomes in old brains (**Figure 25D** and **25E**). Specifically, we establish a mechanistic link between the partial reduction of proteasome activity observed in adult brains and the loss of stoichiometry of protein complexes (**Figure 29B** and **29C**) and of down-regulation of proteasome transcripts and individual lifespan, demonstrating that these processes have consequences for the maintenance of the whole organism fitness.

Later in life, the stoichiometry imbalance in protein complexes contributes to exacerbate the loss of protein homeostasis. Proteasome activity is further reduced in old brains in coincidence with an increased imbalance between the 19S and 20S complexes (**Figure 26B** and **26C**). In addition, altered stoichiometry of the ribosome can underlie both the reduction and the qualitative changes of protein synthesis observed during brain aging (Ori et al., 2015). An alteration of the stoichiometry between membrane-bound and cytosolic components of the lysosomal v-type ATPase (**Figure 20F**) might influence the acidification of lysosomes and the activation of mTORC1 (Zoncu et al., 2011), thus hampering the clearance of protein aggregates. These aggregates in turn may further impair the proteasome activity (Grune et al., 2004), thus creating a negative feedback loop. Other key pathways implicated in aging are also affected by loss of stoichiometry: in particular, alterations of respiratory chain complexes (particularly Complex IV and V) might contribute to their decreased activity and increased ROS production in old brain (Stefanatos and Sanz, 2018), and changes in multiple spliceosome complexes might underlie the higher intron retention rate that is observed in the aging brain (Ori et al., 2015). Studies that are more mechanistic are needed to demonstrate that the alterations that we have described are sufficient to impair the function of these protein complexes during aging.

It remains to be determined which mechanisms promote the early decrease of proteasome activity during early adult life. Our data point to multiple processes being involved, including in particular: (i) decreased levels of rate-limiting proteasome members for the production of 20S assembled/functional proteasomes (e.g., PSMB5 or PSMB6, (Chondrogianni et al., 2015)), which we found to be significantly decreased already in adult fish (**Figure 26B**); (ii) changes in abundance of proteasome proteins that are important for the assembly and activity of the 19S proteasome complex, such as PSMD8. PSMD8 was shown to bridge the lid and the base of the 19S complex and to mediate the lid-base joining, thus being a crucial factor for the assembly of the 30S/26S proteasome (Tomko and Hochstrasser, 2011). We found that PSMD8 is destabilized in old killifish (**Figure 26B**), suggesting a change in its assembly status that could explain the loss of 30S/26S complexes observed in the adult and old fish (**Figure 26C**). PSMD5 has been also shown to inhibit the assembly and activity of the 26S proteasome, and this proteasome member has been shown to be induced by inflammation (Shim et al., 2012). Accordingly, we detected also in killifish an activation of inflammation-related pathways (not shown) and, importantly, we identified PSMD5 as one of the few proteasome members to be upregulated in the adult fish (**Figure 26B**). Finally, PSMD11 (known as RPN-6 in *C. elegans*), which has been shown to be responsible for increased proteasome assembly and activity in human embryonic stem cells and in *C. elegans* (Vilchez et al., 2012), was significantly downregulated already in adult fish (**Figure 26B**). Identifying ways of counteracting these

mechanisms might provide new avenues to delay organ dysfunction in aging and to increase lifespan. In this context, multiple studies have reported that transgenic animals (from various species) engineered to have enhanced proteasome activity show increased health- and lifespan (Augustin et al., 2018; Chondrogianni et al., 2015; Vilchez et al., 2012), and correspondingly we have shown that the rate of decline in the expression of proteasome genes predicts life expectancy in killifish (**Figure 26E**).

In conclusion, this work identifies the maintenance of proteasome activity upon aging as being critical to ensure the correct stoichiometry of protein complexes, thus preserving key biological functions, primarily protein synthesis, degradation, and energy production and ultimately organismal fitness.

2. Investigating genetic and environmental differences in *N.furzeri*

N.furzeri can be cultured in captivity and shares many mammalian typical features of aging (Cellerino et al., 2016). The recent sequencing of its genome (Reichwald et al., 2015; Valenzano et al., 2015) and the availability of multi-omics datasets (Baumgart et al., 2014; Baumgart, Priebe, Groth et al., 2016) makes it a convenient model not only for experimental validations but also for computational analyses in the context of aging research. In the second and third part of the results, I have used *Nothobranchius furzeri* to specifically analyze:

- The impact of genetic differences on aging using laboratory-derived strains with large heritable differences in lifespan (Part 2 of the Results);
- The differences in the transcriptional signature of aging between wild and captive populations (Part 3).

Mechanisms underlying short lifespan in *N.furzeri*: "accelerated" vs. "anticipated" aging profiles

Here, I report that the highly inbred and extremely short-lived GRZ-D strain and the heterogenetic and longer-lived MZM-0410 strains show marked differences in their gene expression profiles and these differences are detectable already at the embryo stage. Some of these differences are observed in all tissues analyzed and can be of very large magnitude so that the number of genes whose expression is dependent on the genotype is larger than the number of age-dependent genes in many tissues. In order to validate these results, a second longer-lived strain was analyzed and this approach

demonstrated that GRZ-D is the outlier. These results indicate that the gene expression patterns of the GRZ-D differs from both wild-derived strains and that DEGs are highly correlated showing that GRZ-D is an outlier, and that the GRZ-D strain has diverged from what can be considered the natural aging pattern of *N. furzeri,* possibly because of bottlenecks related to its captive history. This is linked to what appears as a deranged expression profile.

One of the main results of this part is also that the majority of genotype-related transcript changes are observed during the early phases of development: genetic differences affect gene expression already at embryonic stage and seem to show their biggest effect at 5 weeks post hatching (which, as discussed previously, corresponds to sexual maturity). On the other hand, transcript-based differences appear smaller at young adult stage (12 wph, following MZM-0410 timeline), and at that time point, shorter-lived individuals show a slower rate of transcript change in comparison to the longer-lived companion. This pattern is observed also in the aging biomarkers whose expression is down-regulated with age: the greatest differences in expression are observable in the first time point, but at 12 weeks post hatching (wph) these are no more detectable.

In addition, I also show that part of the genes that are differentially expressed as function of the genotype are also age-regulated in the wild-derived strain. The analysis of this specific gene set demonstrates that shorter lifespan is accompanied with an anticipated aging profile at the molecular level; by this I indicate that for this genes set, differences in gene expression between the two strains are detected already at young age and that young GRZ-D individuals have a gene expression profile closer to older MZM-0410. This pattern is conserved across the four tissues analyzed and is replicated in the second dataset used (shown in **Figure 33E**), suggesting that indeed genetic differences result in an altered aging profile (**Figure 34**). Strikingly, a positive correlation between species differences and age-related regulation was observed in a context of extended lifespan: the comparison between naked mole rat and guinea pig, suggesting that this is a general characteristic of genetic architecture of lifespan in vertebrates.

The anticipated aging process, which I computationally observed, is related to enrichment in autophagy (referred mostly to genes related to mitochondrion degradation) and cell cycle-related categories. Both age-dependent decrease in cell cycle, development, and upregulation of autophagy categories have been extensively reported in human fibroblasts (Marthandan et al., 2016; Hernandez-Segura et al., 2017). On the contrary, in bats, that represent a model of extended lifespan, cell cycle and DNA repair categories have been shown to be upregulated with aging (Huang et al., 2019), highlighting the importance of these pathways for lifespan determination.

The concept of anticipated aging is also supported by the analysis of transcriptional aging biomarkers in the GRZ-D strain (**Figure 41**). These marker genes show monotonous increase of expression with

age in the MZM-0410 strain. An accelerated aging model would require that the GRZ and MZM-04010 profiles diverge progressively with age because the age-dependent rate of regulation is higher in GRZ. What I observed, on the other hand, is that the age-dependent slope is similar in the two strains but GRZ has higher initial values. These markers were previously reported to be enriched in transcripts coding for ribosomal and lysosomal proteins (Baumgart et al., 2014).

To demonstrate that these gene expression differences are causally involved in aging, we compared our results with those of a longitudinal dataset where individual gene expression at young age is correlated with individual variations in lifespan. We observed that genes upregulated in GRZ-D and up-regulated with age are also enriched for genes whose expression is predictive of lifespan, in particular, with genes whose expression levels have positive prognostic values at young age and negative prognostic value at older ages (**Figure 42A, Quadrant II**). The opposite is true for genes consistently down regulated with age and in GRZ-D. This demonstrate that short lifespan of GRZ is due in part to precocious regulation of genes whose expression has a causative role in aging, with a more prominence of the "anticipated aging" model over the "accelerated aging" one.

The situation appears different when comparing wild and captive killifish populations (**Figure52**). The PCA shown in **Figure 52A** shows that samples corresponding to the youngest wild animals correspond to older ages for the captive animals (as a shift to more negative values in PC2). However, a closer inspection of aging biomarkers specific for the two conditions shows again an "accelerated" aging profile of wild animals, as evidenced by the higher slope of biomarkers change in the wild population.

Splicing and protein modification as central hubs for genetic differences in lifespan

Another key result is the centrality of RNA splicing category: network analysis revealed overrepresentation of this category in a co-expression module conserved in brain, liver, skin and muscle and whose expression is positively correlated with genetic differences and age. Comparison of age -dependent DEGs in a long-lived species, the Naked Mole Rat (*Heterocephalous glabrus*) with differences in gene expression between naked mole rat and guinea pig highlights again this category as particularly overrepresented in the genes regulated both during aging and versus the Guinea Pig. Consistently with the results in killifish, these genes are up-regulated in the shortest-lived species, these results are also supported by proteomic analysis. The age-dependent regulation appears to be opposite in the two comparisons as splicesome is up-regulated with age in naked mole rat and is down-regulated with age in killifish. The age-dependent regulation appears to be opposite in the two datasets, as splicesome is up-regulated with age in naked mole rat and is down-regulated with age in killifish. The upregulation of the spliceosome could represent one of the mechanisms that ensure longevity in the naked-mole rat. Indeed, analysis of age-dependent regulation in both naked mole rats

and bats has revealed that key pathways that are down-regulated in other vertebrate species are up-regulated in these species (Heinze et al, 2018; Huang et al., 2019). In the transcriptome of human fibroblasts, spliceosome was found to be downregulated with senescence (Marthandan et al., 2016), consistently spliceosome is down-regulated with age in *Nothobranchius furzeri* both at transcript level (Baumgart et al., 2014) and at the proteome level (**Figure 18** and **Figure 40**). Support in favor of the hypothesis that spliceosome down-regulation shortens lifespan comes from these observations: i) the amplitude of age-dependent spliceosome down-regulation is negatively correlated with individual lifespan in killifish (**Figure 27**), ii) the induction of life extension by calorie restriction (CR) in *Caenorhabditis elegans* depends of expression of splicing factors (Heintz et al., 2017; Tabrez et al., 2017), iii) CR was found to up-regulate RNA processing and to direct alternative splicing usage in the liver of monkeys (Rhoads et al., 2018).

Regarding the longitudinal work presented by Baumgart, Priebe, Groth et al., the third result shown in this thesis highlights the presence of factors that follow antagonistic pleiotropic profile. This theory posits that genes control multiple traits. If selection is driven mainly by the trait that has the major effect on the organism fitness the effects of selection can be detrimental for the other traits. Therefore, an allele that increases fitness early in life at the expenses of fitness late in life will be positively selected. I report that, for the majority of the genes whose expression is predictive of lifespan, the direction of the effect is opposite at the two ages examined. This would suggest that positive selection on the promoters of these genes exerts an antagonistic effect. It is interesting that these genes are highly enriched in the set of those with concordant regulation according to age and genotype, and that these results are consistent when looking at the brain proteome. On the other hand, genes with similar prognostic value at both ages are not enriched. These observations lend support to the antagonistic pleiotropic theory in aging research and define a selection pipeline to identify candidate genes for genetic interventions on lifespan.

Taken together, these results show that lifespan-associated genetic variability has a large impact on gene expression, that a genetic background which favors a shorter lifespan is associated with premature detrimental regulation of genes that are causative for aging and that alterations in the splicing process can be affected by genetic differences. Conversely, age-dependent up-regulation of spliceosome genes may be correlated with evolution of longevity in longer-lived species.

Common signatures of aging in wild and captive populations

As already mentioned in the Introduction (**Chapter 2.4**), gene expression levels can change in response to different environmental conditions, as when comparing population raised in the

laboratory with populations in their natural environment. The transcriptomic analysis of aging in wild populations has been limited to wolves (Charrau et al., 2016) and bats (Huang et al., 2019) so far.
In this thesis, I compared the brain transcriptome of wild and captive killifish populations in order to investigate common signatures of aging. The comparison of condition-dependent differentially expressed genes showed the upregulation of immune response-related categories and the downregulation of protein folding and synaptic transmission; the activation of immune response was recently demonstrated in wild mice after comparing serological, cellular and functional immune parameters of wild and laboratory populations (Abolins et al., 2017).
Age-dependent differentially expressed genes obtained from the two populations, on the other hand, showed a similar expression profile (displayed as a high positive correlation between fold changes, **Figure 52**), with an upregulation of autophagy and a downregulation of mitosis and protein folding in both the conditions.
Interestingly, if mitotic-related categories are consistently downregulated in wild and captive populations, at the same time categories related to DNA replication and DNA repair show downregulation in captive animals, but upregulation in wild animals. Upregulation of DNA repair-related genes reflects a more efficient genome repair machinery; a previous comparative study on the transcriptome of mice, human and naked mole-rat showed the upregulation of DNA damage and DNA repair genes in long-lived species, confirmed also at the level of fibroblast transcriptomes using 16 different mammalian species (Tiao, Seluanov & Gorbunova, 2017). The categories shown in Quadrant II of **Figure 52**, however, are also part of the cell cycle S phase, while categories in Quadrant III are M phase-related categories: studies on the molecular mechanisms underlying differential regulation of the distinct phases of cell cycle during aging still lack however in literature.

In conclusion, the data presented in my thesis show that:
- proteasome inhibition is a driver for loss of stoichiometry in the brain proteome of *N.furzeri*;
- genetic differences affecting lifespan can be reproduced at the transcriptomic level;
- during aging factors that affect fitness and follow an antagonistic pleiotropic profile show differential regulation.

Contributions

Section (Results)	Type of experiment	Performer
Section 1	Proteomic datasets generation	Kelmer Sacramento E, Kirkpatrick J
	Transcriptomic data generation	Kelmer Sacramento E, Baumgart M
	RNA-sequencing	Baumgart M, Groth M
	Thermal Proteome Profiling	Di Sanzo S, Kirkpatrick J
	Aggregates isolation	Kelmer Sacramento E, Kirkpatrick J
	Proteasome activity assay	Chondrogianni N
	Immunostainings (*N.furzeri*)	Terzibasi Tozzini E, Bagnoli S
	Proteasome inhibition	Kelmer Sacramento E, Bartholome A
	Immunostainings (HLF-1)	Caterino C
Section 2	qPCR	Mazzetto M
	Transcriptomic data generation	Reichwald K
	RNA-sequencing	Koch P, Groth M
Section 3	Transcriptomic data generation	Reichard M
	RNA-sequencing	Reichard M., Baumgart M.
	qPCR	Mazzetto M, Caterino C

References

Abolins S, King EC, Lazarou L, Weldon L, Hughes L, Drescher P, Raynes JG, Hafalia JCR, Viney ME, Riley EM. **The comparative immunology of wild and laboratory mice, *Mus musculus domesticus***. Nature communications, 2017 May 3; 8(14811).

Aebersold R, Mann M. **Mass spectrometry-based proteomics**. Nature, 2003 Mar 13; 422(6928):198-207.

Akerfelt M, Morimoto RI, Sistonen L. **Heat Shock Factors: integrators of cell stress, development and lifespan**. Nat Rev Mol Cell Biol. 2010 Aug; 11(8):545-555.

Anders S, Pyl PT, Huber W. **HTSeq--a Python framework to work with high-throughput sequencing data**. Bioinformatics, 2015 Jan 15; 31(2):166-9.

Augustin H, McGourty K, Allen MJ, Adcott J, Wong CT, Boucrot E, Partridge L. **Impact of insulin signaling and proteasomal activity on physiological output of a neuronal circuit in aging *Drosophila melanogaster***. Neurobiology of Aging, 2018 Jun; 66:149-157.

Auluck PK, Chan HY, Trojanowski JQ, Lee VM, Bonini NM. **Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson disease**. Science 2002 Feb 1; 295(5556):865-8.

Bahn A, Hagos Y, Reuter S, Balen D, Brzica H, Krick W, Burckhardt BC, Sabolic I, Burckhardt G. **Identification of a new urate and high affinity nicotinate transporter, hOAT10 (SLC22A13).** J Biol Chem, 2008 Jun 13; 283(24):16332-41.

Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJM. **Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach.** BMC Genomics, 2006 Sep 29; 7(246)

Balch WE, Morimoto RI, Dillin A, Kelly JW. **Adapting proteostasis for disease intervention**. Science 2008 Feb 15; 319(5865):916-9.

Baranczak A, Kelly JW. **A current pharmacologic agent versus the promise of next-generation therapeutics to ameliorate protein misfolding and/or aggregation diseases.** Current Opinion in Chemical Biology, 2016 Feb 6; 32:10-21.

Baumgart M, Barth E, Savino A, Groth M, Koch P, Petzold A, Arisi I, Platzer M, Marz M, Cellerino A. **A miRNA catalogue and ncRNA annotation of the short-living fish *Nothobranchius furzeri***. BMC Genomics, 2017; 18:693.

Baumgart M, Priebe S, Groth M, Hartmann N, Menzel U, Pandolfini L, Koch P, Felder M, Ristow M, Englert C, Guthke R, Platzer M, Cellerino A. **Longitudinal RNA-Seq Analysis of Vertebrate Aging identifies Mitochondrial Complex I as a small-molecule-sensitive Modifier of Lifespan**. Cell Systems 2016 Feb 24; 2(2):122-32

Baumgart M, Di Cicco E, Rossi G, Cellerino A, Tozzini ET. **Comparison of captive lifespan, age-associated liver neoplasias and age-dependent gene expression between two annual fish species: Nothobranchius furzeri and Nothobranchius korthause**. Biogerontology, 2015 Feb; 16(1):63-9.

Baumgart M, Groth M, Priebe S, Savino A, Testa G, Dix A, Ripa R, Spallotta F, Gaetano C, Ori M, Terzibasi Tozzini E, Guthke R, Platzer M, Cellerino A. **RNA-seq of the aging brain in the short-lived fish N. furzeri - conserved pathways and novel genes associated with neurogenesis.** Aging Cell, 2014 Dec; 13(6):965-74.

Baumgart M, Groth M, Priebe S, Appelt J, Guthke R, Platzer M, Cellerino A. **Age-dependent regulation of tumor-related microRNAs in the brain of the annual fish *Nothobranchius furzeri*.** *Mechanisms of Ageing and Development, 2012 May; 133(5):226-233.*

Becher I, Andres-Pons A, Romanov N, Stein F, Schramm M, Baudin F, Helm D, Kurzawa N, Mateus A, Mackmull MT, Typas A, Muller CW, Bork P, Beck M, Savitski MM. **Pervasive Protein Thermal Stability Variation during the Cell Cycle.** Cell, 2018 May 31; 173(6):1495-1507.

Blažek R, Polacik M, Reichard M. **Rapid growth, early maturation and short generation time in African annual fishes**. EvoDevo, 2013; 4, 24.

Brennan CM, Vaites LP, Wells JN, Santaguida S, Paulo JA, Storchova Z, Harper JW, Marsh JA, Amon A. **Protein aggregation mediates stoichiometry of protein complexes in aneuploidy cells**. Genes Dev, 2019 Aug 1; 33(15-16):1031-47.

Bronson RT, Lipman RD, Harrison DE.. **Age-related gliosis in the white matter of mice.** Brain Research, 1993; 609:124–128.

Calamini B, Silva MC, Madoux F, Huff DM, Khanna S, Chalfant MA, Saldanha SA, Hodder F, Tait BD, Garza D, Balch WE, Morimoto RI. **Small-molecule proteostasis regulators for protein conformational diseases.** Nat Chem Biol. 2011Dec 25; 8(2):185-96.

Carmona-Gutierrez D, Hughes AL, Madeo F, Ruckenstuhl C. **The crucial impact of lysosomes in aging and longevity**. Ageing Res Rev, 2016 Dec; 32:2-12.

Cellerino A, Sanguanini M. **Transcriptome Analysis**. Lecture Notes (Scuola Normale Superiore), 2018.

Cellerino A, Valenzano DR, Reichard M. **From the bush to the bench: the annual N*othobranchius* fishes as a new model system in biology**. Biol. Rev. 2015;000-000.

Chafekar SM, Duennwald ML. **Impaired heat shock response in cells expressing full-length polyglutamine-expanded huntingtin**. PloS One 2012; 7(5):e37929.

Chan SW, Lim CJ, Chong YF, Pobbati AV, Huang C, Hong W. 2011. ***Hippo pathway-independent restriction of TAZ and YAP by angiomotin.*** *J Biol Chem* ***286****: 7018–7026.*

Charruau P, Johnston RA, Stahler DR, Lea A, Snyder-Mackler N, Smith DW, vonHoldt BM, Cole SW, Tung J, Wayne RK. ***Pervasive effects of Aging on Gene Expression in Wild Wolves****. Molecular Biology and Evolution, 2016 Aug; 33(8):1967-78*

Childs D, Bach K, Franken H, Anders S, Kurzawa N, Bantscheff M, Saviski M, Huber W. **Non-parametric analysis of thermal proteome profiles reveals novel drug-binding proteins**. Molecular & Cellular Proteomics, 2019 Oct 3.

Chondrogianni N, Voutetakis K, Kapetanou M, Delitisikou V, Papaevgeniou N, Sakellare M, Lefaki M, Filippopoulou K, Gonos ES. **Proteoasome activation: an innovative promising approach for delaying aging and retarding age-related diseases**. Ageing Res Rev, 2015 Sep; 23(Pt A):37-55.

Chondrogianni N, Petropoulos I, Grimm S, Georgila K, Catalogol B, Friguet B, Grune T, Gonos ES. **Protein damage, repair and proteolysis**. Mol Aspects Med, 2014 Feb; 35:1-71.

Chondrogianni N, Gonos ES. **Proteasome inhibition induces a senescence-like phenotype in primary human fibroblasts cultures**. Biogerontology, 2004; 5(1):55-61.

Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahloun AG, Herman MM, Weinberger DR, Kleinman JE. ***Temporal dynamics and genetic control of transcription in the human prefrontal cortex****. Nature, 2011 Oct 26; 478(7370):519-23.*

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Mortazavi A. ***A survey of best practices for RNA-seq data analysis****. Genome Biology, 2016 Jan 26; 17(13).*

Cox J, Mann M. ***Is proteomics the new genomics?*** *Cell, 2007 Aug 10; 130(3):395-8.*

Cravatt BF, Simon GM, Yates GR 3rd. ***The biological impact of mass-spectrometry-based proteomics****. Nature, 2007 Dec 13; 450(7172):991-1000.*

D'Angelo L, Castaldo L, Cellerino A, de Girolamo P, Lucini C. ***Nerve growth factor in the adult brain of a teleostean model for aging research:*** **Nothobranchius furzeri***. Annals of Anatomy, 2014 Jul; 196(4):183-191*

D'Angelo L. ***Brain Atlas of an emerging teleostan model:*** **Nothobranchius furzeri***. The Anatomical Record, 2013 Apr; 296(4):681-691*

D'Angelo L, de Girolamo P, Cellerino A, Tozzini ET, Castaldo L, Lucini C. ***Neurotrophin Trk receptors in the brain of a teleost fish,*** **Nothobranchius furzeri.** *Microscopy & research technique, 2012 Jan; 75(1):81-88*

Dai DF, Chen T, Johnson SC, Szeto H, Rabinovitch PS. **Cardiac aging: from molecular mechanisms to significance in human health and disease**. Antioxidants & Redox Signaling 2012; 16:1492–1526.

David DC, Ollikainen N, Trinidad JC, Cary MP, Burlingame AL, Kenyon C. **Widespread Protein Aggregation as an Inherent Part of Aging in *C.elegans*.** PLoS Biology, 2010; 8(8).

Demartino GN, Gillette TG. **Proteasomes: machines for all reasons**. Cell 2007 May 18; 129(4):659-62

Di Cicco E, Tozzini ET, Rossi G, Cellerino A. **The short-lived annual fish Nothobranchius furzeri shows a typical teleost aging process reinforced by high incidence of age-dependent neoplasias**. Exp Gerontol. 2011 Apr; 46(4):249-56

Dillin A, Hsu AL, Arantes-Oliveira N, Leher-Graiwer J, Hsin H, Fraser AG, Kamath RS, Ahringer J, Kenyon C. **Rates of behavior and aging specified by mitochondrial function during development**. Science, 2002 Dec 20; 298(5602):2398-401.

Ding L, Kuhne WW, Hinton DE, Song J, Dynan WS. **Quantifiable Biomarkers of Normal Aging in the Japanese Medaka Fish (Oryzias latipes)**. PloS One, 2010 Oct 11; 5(10):e13287.

Domon B, Aebersold R. **Mass spectrometry and protein analysis**. Science, 2006 Apr 14; 312(5771):12-7.

Dorn A, Musilovà Z, Platzer M, Reichwald K, Cellerino A. **The strange case of East African annual fishes: aridification correlates with diversification for a savannah acquatic group?** BMC Evolutionary Biology, 2014; 14(210)

Durinck S, Spellman PT, Birney E, Huber W. **Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt**. Nature Protocols, 2009; 4(8):1184-9.

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. **Thermofluor-based high-throughput stability optimization of proteins for structural studies**. Anal. Biochem. 2006; 357:289–298.

Fang Z, Cui X. **Design and validation issues in RNA-seq experiments**. Brief Bioinform. 2011 May; 12(3):280-7.

Fernandez-Fernandez MR, Ferrer I, Lucas JJ. **Impaired ATF6a processing, decreased Rheb and neuronal cell cycle re-entry in Huntigton's disease**. Neurobiol. Dis. 2011; 41:23-32.

Finley D. **Recognition and processing of ubiquitin-protein conjugates by the proteasome**. Annu Rev Biochem. 2009; 78:477-513

Fonseca NA, Rung J, Brazma A, Marioni JC. **Tools for mapping high-throughput sequencing data**. Bioinformatics, 2012 Dec 15; 28(24):3169-77.

Froeze R, Pauly D. **FishBase: World Wide Web electronic publication**. 2014 Nov.

Fu H, Sadis S, Rubin DM, Glickman M, van Nocker S, Finley D, Verstra RD. **Multiubiquitin chain binding protein degradation are mediated by distinct domanis within the 26S proteasome subunit Mcb1**. J Biol Chem 1998 Jan 23; 273(4):1970-81

Garalde DR, Snell EA, Jachimoviz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, Jordan M, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace EJ, Jayasinghe L, Wright C, Blasco J, Young S, Brocklebank D, Juul S, Clarke J, Heron AJ, Turner DJ. **Highly parallel direct RNA sequencing on an array of nanopores**. Nature Methods, 2018; 15:201-206

Genade T, Benedetti M, Terzibasi E, Roncaglia P, Valenzano DR, Cattaneo A, Cellerino A. **Annual fishes of the genus *Nothobranchius* as a model system for aging research**. Aging Cell, 2005 Aug 22.

Gerhard GS, Kauffman EJ, Wang X, Steward R, Moore JL, Kasales CJ, Demidenko E, Cheng KC. **Lifespans and senescent phenotypes in two strains of Zebrafish (*Danio rerio*)**. Exp Gerontol, 2002 Aug-Sep; 37(8-9):1055-68.

Gidalevitz T, Ben-Zvi A, Ho KH, Brignull HR, Morimoto RI. **Progressive disruption of cellular protein folding in models of Polyglutamine Disease**. Science 2006 May 10; 311(5766):1471-1474.

Gkogkas C, Middleton S, Kremer AM, Wardrope C, Hannah M, Gillingwater TH, Skehel P. **VAPBinteracts with and modulates the activity of ATF6**. Hum. Mol. Genet. 2008; 17:1517-1526.

Graf M, Hartmann N, Reichwald K, Englert C. **Absence of replicative senescence in cultured cells from the short-lived killifish *Nothobranchius furzeri*.** Experimental gerontology, 2013 Jan; 48(1):17-28.

Gross-Thebing T, Yigit S, Pfeiffer J, Reichman-Fried M, Bandemer J, Ruckert C, Rathmer C, Goudarzi M, Stehling M, Tarbashevich K, Seggewiss J, Raz E. **The Vertebrate Protein Dead End Maintains Primordial Germ Cell Fate by Inhibiting Somatic Differentiation**. Developmental Cell, 2017Dec 18; 43(6):704-715.

Grune T, Jung T, Merker K, Davies KJA. **Decreased proteolysis caused by protein aggregates, inclusion bodies, plaques, lipofuscin, ceroid, and "aggresomes" during oxidative stress, aging, and disease.** The international Journal of Biochemistry & Cell Biology, 2004 Jul 2; 36(12):2519-30.

Gygi SP, Rochon Y, Franza BR, Aebersold R. **Correlation between protein and mRNA abundance in yeast**. Mol Cell Biol, 1999 Mar; 19(3):1720-30.

Haas R. **Behavioral biology of the annual killifish, *Nothobranchius guentheri*.** Copeia, 1976; 80-91

Hartmann N, Reichwald K, Wittig I, Drose S, Schmeisser S, Luck C, Hahn C, Graf M, Gausmann U, Terzibasi E, Cellerino A, Ristow M, Brandt U, Platzer M, Englert C. **Mitochondrial DNA copy number and function decrease with age in the short-lived fish *Nothobranchius furzeri***. Aging Cell, 2011 Oct; 10(5):824-31.

Hartmann N, Reichwald K, Lechel A, Graf M, Kirschner J, Dorn A, Terzibasi E, Wellner J, Platzer M, Rudolph KL, Cellerino A, Englert C. **Telomeres shorten while Tert expression increases during ageing of the short-lived fish *Nothobranchius furzeri*.** Mech Ageing Dev, 2009 May; 130(5):290-6

Hashimshony T, Wagner E, Sher N, Yanai I. **CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification**. Cell Reports, 2012 Sep 27; 2(3):666-73.

Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. **Some like it hot: the structure and function of small heat-shock proteins.** Nat Struct Mol Biol. 2005 Oct; 12(10):842-6

Hatakeyama H, Nakamura KI, Izumiyama-Shimomura N, Ishii A, Tsuchida S, Takubo K, Ishikawa N. **The teleost *Oryzias latipes* shows telomere shortening with age despite considerable telomerase activity throughout life**. Mechanisms of Ageing and Development, 2008 Sep; 129(9):550-557.

Hay DG, Sathasivam K, Tobaben S, Stahl B, Marber M, Mestril R, et al. **Progressive decrease in chaperone protein levels in a mouse model of Huntington's disease and induction of stress proteins as a therapeutic approach**. Hum Mol Genet. 2004;13:1389–405.

Hayflick L. **The limited *in vitro* lifetime of human diploid cell strains**. Experimental Cell Research, 1965; 37(3):614-36.

Heintz C, Doktor TK, Lanjuin A, Escoubas C, Zhang Y, Weir HJ, Dutta S, Silva-Garcia CG, Bruun GH, Morantte I, Hoxhaj G, Manning BD, Andresen BS, Mair WB. **Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in *C.elegans*.** Nature, 2017 Jan 5; 541(7635):102-106.

Heinze I, Bens M, Calzia E, Holtze S, Dakhovnik A, Sahm A, Kirkpatrick JM, Szafranski K, Romanov N, Sama SN, Holzer K, Singer S, Ermolaeva M, Platzer M, Hildebrandt T, Ori A. **Species comparison of liver prtoeomes reveals links to naked mole-rat longevity and human aging**. BMC Biology, 2018 Aug 2; 16(82).

Hernandez-Segura A, de Jong TV, Melov S, Guryev V, Campisi J, Demaria M. **Unmasking Transcriptional Heterogeneity in Senescent Cells.** Current Biology, 2017 Sep 11; 27(17):2652-2660.

Hetz C, Mollereau B. **Disturbance of endoplasmic reticulum proteostasis in neurodegnerative diseases**. Nat Rev Neurosci. 2014 Apr; 15(4):233-49

Hoofnagle AN, Heinecke JW. **Lipoproteomics: using mass spectrometry-based proteomics to explore the assembly, structure, and function of lipoproteins**. J Lipid Res. 2009 Oct; 50(10):1967-75.

Hrbek T, Larson A. **The evolution of diapause in the killifish family Rivulidae (Atherinomorpha, Cyprinodontiforms): a molecular phylogenetic and biogeographic perspective**. Evolution, 1999 Aug; 53(4):1200-16

Hsu CY, Chiu YC. **Ambient temperature influences aging in an annual fish (*Nothobranchius rachovii*)**. Aging Cell, 2009; 8:726-737.

Hsu CY, Chiu YC, Hsu WL, Chan YP. ***Age-related markers assayed at different developmental stages of the annual fish*** **Nothobranchius rachovii**. *The journals of Gerontology, 2008 Dec; 63(12):1267-76.*

Huang Z, Whelan CV, Foley NM, Jebb D, Touzalin F, Petit EJ, Puechmaille SJ, Teeling EC. ***Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats.*** *Nat Ecol Evol, 2019 Jul; 3(7):1110-20.*

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan

M. **Orchestrating high-throughput genomic analysis with Bioconductor**. Nature Methods, 2015 Feb; 12(2):115-121.

Husnjak K, Elsasser S, Zhang N, Chen X, Randles L, Shi Y, Hofmann K, Walters KJ, Finley D, Dikic I. **Proteasome subunit Rpn13 is a novel ubiquitin receptor**. Nature 2008 May 22; 453(7194):481-8

Janssens GE, Meinema AC, Gonzalez J, Wolters JC, Schmidt A, Guryev V, Bischoff R, Wit EC, Veenhoff LM, Heinemann M. **Protein biogenesis machinery is a driver of replicative aging in yeast**. Elife, 2015 Dec 1; 4:e08527.

Jubb, RA. **A new *Nothobranchius* (Pisces, Cyprinodontidae) from Southeastern Rhodesia**. Journal of the American Killifish Association, 1971; 8:12–19.

Keshishian H, Addona T, Burgess M, Mani DR, Shi X, Kuhn E, Sabatine MS, Gerszten RE, Carr SA. **Quantification of Cardiovascular Biomarkers in Patient Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution.** Molecular & Cellular Proteomics, 2009 Oct 1; 8(10):2339-49.

Kempermann G. **Seven principles in the regulation of adult neurogenesis**. Eur J Neurosc. 2011 Mar; 33(6):1018-24.

Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP. **Structure of the human 80S ribosome**. Nature, 2015 Apr 30; 520(7549):640-5.

Kirkwood TB. **Evolution of ageing**. Nature, 1977; 270:301–304.

Kirschner J, Weber D, Neuschi C, Franke A, Bottger M, Zielke L, Powalsky E, Groth M, Shagin D, Petzold D, Hartmann N, Englert C, Brockmann GA, Platzer M, Cellerino A, Reichwald K. **Mapping of quantitative trait loci controlling lifespan in the short-lived fish Nothobranchius furzeri – a new vertebrate model for age research.** Aging Cell, 2012 Apr; 11(2):252-261

Knierim E, Lucke B, Schwarz MJ, Schuelke M, Seelow D. S**ystematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing**. PloS One, 2011 Nov 11; 6(11): e28240

Knowles PJ, Vendruscolo M, Dobson CM. **The amyloid state and its association with protein misfolding diseases.** Nature Reviews Molecular Cell Biology 2014; 15: 384-396

Kumsta C, Chang JT, Schmalz J, Hansen M. **Hormetic heat stress and HSF-1 induce autophagy to imrpove survival and proteostasis in *C.elegans***. Nature Communications, 2017; 8(14337)

Lahens NF, Kavakli IH, Zhang R, Hayer K, Black MB, Dueck H, Pizarro A, Kim J, Irizarry R, Thomas RS, Grant GR, Hogenesch JB. **IVT-seq reveals extreme bias in RNA sequencing**. Genome Biology, 2014 Jun 30; 15(6):R86.

Langfelder P, Horvath S. **WGCNA: an R package for weighted correlation network analysis**. BMC Bioinformatics, 2008; 9(559).

Lee B, Lee MJ, Park S, Oh D, Elsasser S, Chen P, Gartner C, Dimova N, Hanna J, Gygi SP, Wilson SM, King RW, Finley D. **Enhancement of Proteasome Activity by a Small-Molecule Inhibitor of Usp14**. Nature. 2010 Sep 9; 467(7312):179-184.

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. **Comprehensive comparative analysis of strand-specific RNA sequencing methods**. Nature Methods, 2010 Sep; 7(9):709-15.

Liu C, Wang X, Feng W, Li G. **Differential expression of aging biomarkers at different life stages of the annual fish *Nothobranchius guentheri***. Biogerontology, 2012 Aug; 13(5):501-10

Lo CC, Chain PSG. **Rapid evaluation and quality control of next generation sequencing data with FaQCs**. BMC Bioinformatics, 2014 Nov 19; 15(1):366.

Loman NJ, Pallen MJ. **Twenty years of bacterial genome sequencing**. Nature Reviews Microbiology, 2015 Nov 9; 13:787-94.

Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ. **High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity**. Nature Reviews Microbiology, 2012 Aug 6; 10:599-606.

Love MI, Huber W, Anders S. **Moderated estimation of fold changes and dispersion for RNA-seq data with DESeq2.** Genome Biology, 2014; 15(12):550.

Lucas-Sanchez A, Almaida-Pagan PF, Mendiola P, de Costa J. **Nothobranchius as a model for aging studies. A review**. Aging Dis, 2014 Aug; 5(4):281-291

Lucas-Sanchez A, Almaida-Pagan PF, Madrid JA, de Costa J, Mendiola P. **Age-realted changes in fatty acid profile and locomotor activity rhytms in *Nothobranchius korthausae***. Experimental Gerontology, 2011 Dec; 46(12):970-978.

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. **GAGE: generally applicable gene set enrichment for pathway analysis.** BMC Bioinformatics, 2009; 10(161).

Mallick P, Kuster B. **Proteomics: a pragmatic perspective**. Nature Biotechnology, 2010 Jul; 28(7):695-709.

Marthandan S, Bamgart M, Priebe S, Groth M, Schaer J, Kaether C, Guthke R, Cellerino A, Platzer M, Diekmann S, Hemmerich P. **Conserved Senescence associated genes and pathways in Primary Human Fibroblasts detected by RNA-seq**. PLoS One, 2016; 11(5):e0154531.

Mazuze, FM. **Analysis of adoption of orange-fleshed sweetpotatoes: the case study of Gaza province in Mozambique**. Institute of Agricultural Research of Mozambique, Maputo, 2007; Research Report No. 4E:35 pp.

McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. **MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.** Anal Chem, 2014 Jul 15; 86(14):7150-8.

McShane E, Sin C, Zauber H, Wells JN, Donnelly N, Wang X, Hou J, Chen W, Storchova Z, Marsh JA, Valleriani A, Selbach M. **Kinetic analysis of Protein Stability Reveals Age-Dependent Degradation.** Cell, 2016 Oct 20; 167(3):803-815.

Medawar PB. **An Unsolved Problem of Biology.** HK Lewis, London, 1952.

Miller JR, Koren S, Sutton G. **Assembly algorithms for next-generation sequencing data**. Genomics 2010; 95:315-27.

Morley JF, Brignull HR, Weyers JJ, Morimoto RI. **The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caernohabitis elegans***. PNAS 2002 Aug 16; 99(16): 10417-10422

Morozova O, Hirst M, Marra MA. **Applications of new sequencing technologies for transcriptome analysis**. Annu Rev Genetics Hum Genet, 2009; 10:135-51.

Mu TW, Ong DST, Wang YJ, Balch WE, Yates JR, Segatori L, Kelly JW. **Proteostasis regulators and Pharmacologic Chaperones Synergize to Correct Protein Misfolding Diseases**. Cell, 2008 Sep 5; 134(5):769-781.

Murphy WJ, Collier GE. **A molecular phylogeny for aplocheiloid fishes (Aterinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism**. Mol BiolEvol. 1997 Aug; 14(8):790-9.

Naef V, Monticelli S, Corsinovi D, Mazzetto M, Cellerino A, Ori M. **The age-regulated zinc finger factor ZNF367 is a new modulator of neuroblast proliferation during embryonic neurogenesis**. Scientific Reports, 2018 Aug 7; 8(11836).

Nagalashkmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. **The transcriptional landscape of the yeast genome defined by RNA sequencing**. Science, 2008 Jun 6; 320(5881):1344-9.

Naiman RJ, Gerking SD, Ratcliff TD. **Thermal Environment of a Death Valley Pupfish**. Copeia, 1973 May 22; 1973(2):366-369

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. **Resolution of ray-finned fish phylogeny and timing of diversification**. PNAS, 2012 Aug 21; 109(34):13698-703.

Ng'oma E, Groth M, Ripa R, Platzer M, Cellerino A. **Transcriptome profiling of natural dichromatismin the annual fishes *Nothobranchius furzeri* and *Nothobranchius kadleci***. BMC Genomics, 2014 Sep 2; 15(1):754.

Niccoli T, Partridge L. **Ageing as a risk factor for disease**. Current Biology 2012 Sep 11; 22(17): R471-52

Norton WT, Aquino DA, Hozumi I, Chiu FC, Brosnan CF. **Quantitative aspects of reactive gliosis: a review.** Neurochemical Research, 1992; 17:877–885.

Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. **Accurate quantitation of protein expression and site-specific phosphorylation**. PNAS, 1999.

Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU, Vabulas RM. **Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions.** Cell 2011 Jan 7; 144(1):67-78.

Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM. **Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database**. Proteomics, 2005 Aug; 5(13):3226-43.

Omodei D, Fontana L. **Calorie restriction and prevention of age-associated chronic disease.** FEBS Lett. 2011 Jun 6; 585(11):1537-42.

Onken B, Driscoll M. **Metformin induces a dietary restriction-like state and the oxidative stress response to extend C.elegans healtspan via AMPK, LKB1, and SKN-1**. PloS One, 2010 Jan 18; 5(1): e8758

Ori A, Iskar M, Buczak K, Kastritis P, Parca L, Andres-Pons A, Singer S, Bork P, Beck M. **Spatiotemporal variation of mammalian protein complex stoichiometries**. Genome Biology, 2016; 17(47).

Ori A, Banterle N, Iskar M, Andres-Pons A, Excher C, Khanh Bui H, Sparks L, Solis-Mezarino V, RInner O, Bork P, Lemke EA, Beck M. **Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines**. Mol Syst Biol, 2013; 9:648.

Ozsolak F, Milos PM. **RNA sequencing: advances, challenges and opportunities**. Nat Rev Genet, 2011 Feb; 12(2):87-98

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, Mclnerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. **Direct RNA sequencing**. Nature Letters, 2009 Sep 23; 461: 814-818.

Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR. **High-density miniaturized thermal shift assays as a general strategy for drug discovery**. J Biomol Screen, 2001 Dec; 6(6):429-40.

Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I. **The impact of amplification on differential expression analyses by RNA-seq**. Scientific Reports, 2016 May 9; 6(25533).

Petzold A, Reichwald K, Groth M, Taudien S, Hartmann N, Priebe S, Shagin D, Englert C, Platzer M. **The transcript catalogue of the short-lived fish Nothobranchius furzeri provides insights into age-dependent changes of mRNA levels**. BMC Genomics, 2013 Mar 16; 14:185.

Polačik M, Blažek R, Řežucha R, Vrtílek M, Terzibasi Tozzini E, Reichard M. **Alternative intrapopulation life-history strategies and their trade-offs in an African annual fish**. Journal of Evolutionary Biology, 2014; 27:854–865.

Polacik M, Reichard M. **Diet overlap among three sympatric African annual Killifish species *Nothobranchius* spp. From Mozambique**. J Fish Biol, 2010 Aug; 77(3):754-68.

Podrabsky JE, Somero GN. **Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus***. J Exp Biol, 2004 Jun; 207(Pt 13):2237-54.

Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE. **Biological and chemical approaches to diseases of proteostasis deficiency.** Annu Rev Biochem. 2009; 78:959-91

Ramskold D, Wang ET, Burge CB, Sandberg R. **An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data**. PloS One, Computational Biology, 2009 Dec; 5(12):e1000598.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. **Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data**. Genome Biology, 2013; 14(9):R95

Reichard M, Polacik M, Blažek R, Vrtílek M. **Female bias in the adult sex ratio of African annual fishes: interspecific differences, seasonal trends and environmental predictors**. Evolutionary Ecology, 2014; 28:1105–1120.

Reichwald K, Petzold A, Koch P, Downie BR, Hartmann N, Pietsch S, Baumgart M, Chalopin D, Felder M, Bens M, Sahm A, Szafranski K, Taudien S, Groth M, Arisi I, Weise A, Bhatt SS, Sharma V, Kraus JM, Schmid F, Priebe S, Liehr T, Görlach M, Than ME, Hiller M, Kestler HA, Volff JN, Schartl M, Cellerino A, Englert C, Platzer M. **Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish**. Cell, 2015 Dec 3; 163(6):1527-38.

Reichwald K, Lauber C, Nanda I, Kirschner J, Hartmann N, Schories S, Gaussmann U, Taudien S, Schilhabel MB, Szafranski K, Glockner G, Schmid M, Cellerino A, Scharti M, Englert C, Platzer M. **High tandem repeat content in the genome of the short-lived annual fish Nothobranchius furzeri: a new vertebrate model for aging research**. Genome Biology, 2009 Feb 11; 10(2):R16.

Reis-Rodrigues P, Czerwienec G, Peters TW, Evani US, Alavez S, Gaman EA, Vantipalli M, Mooney SD, Gibson BW, Lithgow GJ, Hughes RE. **Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan**. Aging Cell, 2012Feb; 11(1):120-7.

Rhoads TW, Burhans MS, Chen VB, Hutchins PD, Rush MJP, Clark JP, Stark JL, Mcllwain SJ, Eghbalnia HR, Pavelec DM, Ong IM, Denu JM, Markley JL, Coon JJ, Colman RJ, Anderson RM. **Caloric Restriction engages hepatic RNA Processing in Rhesus Monkeys**. Cell Metabolism, 2018 Mar 6; 27: 677-88.

Ripa R, Dolfi L, Terrigno M, Pandolfini L, Savino A, Arcucci V, Groth M, Terzibasi Tozzini E, Baumgart M, Cellerino A. **MicroRNA miR-29 controls a compensatory response to limit neuronal iron accumulation during adult life and aging**. BMC Biology, 2017 Feb 13; 15:9.

Robida-Stubbs S, Glover-Cutter K, Lamming DW, Mizunuma M, Narasimhan SD, Neumman-Haefelin E, Sabatini DM, Blackwell TK. **TOR signalling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO.** Cell Metab. 2012 May 2; 15(5): 713-24

Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. **Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution**. Science, 2019 Mar 29; 363(6434):1463-67.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. **Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents**. Mol Cell Proteomics, 2004 Dec; 3(12):1154-69.

Sahm A, Bens M, Platzer M, Cellerino A. **Parallel evolution of genes controlling mitonuclear balance in short-lived annual fishes**. Aging Cell, 2017 Jun; 16(3):488-96.

Sakthianandeswaren A, Parsons MJ, Mouradov D, Sieber OM. **MACROD2 deletions cause impaired PARP1 activity and chromosome instability in colorectarl cancer**. Oncotarget, 2018 Sep 4; 9(69):33056-58.

Sarkar S, Ravikumar B, Floto RA, Rubinsztein DC. **Rapamycin and mTOR-independent autophagy inducers ameliorate toxicity of polyglutamine-expandend huntingtin and related proteinopathies**. Cell Death Differ. 2009 Jan; 16(1):46-56.

Savitzky MM, Reinhard FB, Franken H, Werner T, Savitski MF, Eberhard D, Martinez Molina D, Jafari R, Dovega RB, Klaeger S, Kuster B, Nordlund P, Bantscheff M, Drewes G1. **Tracking cancer drugs in living cells by thermal profiling of the proteome**. Science, 2014 Oct 3; 346(6205):1255784.

Senisterra GA, Markin E, Yamazaki K, Hui R, Vedadi M, Awrey DE. **Screening for ligands using a generic and high-throughput light-scattering-based assay**. Journal of biomolecular screening, 2006 Dec; 11(8):940-48.

Shanker S, Paulson A, Edenberg HJ, Peak A, Perera A, Alekseyev YO, Beckloff N, Bivens NJ, Donnelly R, Gillapsy AF, Grove D, Gu W, Jafari N, Kerley-Hamilton JS, Lyons RH, Tepper C, Nicolet

CM. **Evaluation of Commercially Available RNA Amplification Kits for RNA Sequencing Using Very Low Input Amounts of Total RNA**. J Biomol Tech, 2015 Apr; 26(1):4-18.

Sharp Z.D.,Richardson A. **Aging can cancer: can mTOR inhibitors kill two birds with one drug?** Target. Oncol. 2011; 6:41-51

Shen D, Coleman J, Chan E, Nicholson TP, Dai L, Sheppard PW, Patton WF. **Novel cell- and tissue-based assays for detecting misfolded and aggregated protein accumulation within aggresomes and inclusion bodies**. Cell biochemistry and biophysics, 2011; 60(3):173-185.

Shendure J, Ji H. **Next-generation DNA sequencing**. Nature Biotechnology, 2008 Oct; 26(10):1135-45.

Shi Y, Chen X, Elsasser S, Stocks BB, Tian G, Lee BH, Shi Y, Zhang N, de Poot SA, Tuebing F, Sun S, Vannoy J, Tarasov SG, Engen JR, Finley D, Walters KJ. **Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome.** Science 2016 Feb 19; 351(6275). pii:aaad9421.

Shim SM, Lee WJ, Kim Y, Chang JW, Song S, Jung YK. **Role of S5b/PSMD5 in proteasome inhibition caused by TNF-alpha/NFkB in higher eukaryotes**. Cell Reports, 2012 Sep 27; 2(3):603-15.

Silva FG. **The aging kidney: a review-part I**. International Urology and Nephrology 37, 2005:pp. 185–205.

Silva FG. **The aging kidney: a review-part II**. International Urology and Nephrology 37, 2005:pp. 419–432.

Sittler A, Lurz R, Lueder G, Priller J, Lehrach H, Hayer-Hartl MK, Hartl FU, Wanker EE. **Geldanamycin activates a heat shock response and inhibits huntingtin aggregation in a cell culture model for Huntogton's disease**. Hum Mol Genet. 2001 Jun 1; 10(12):1307-15.

Stefanatos R, Sanz A. **The role of mitochondrial ROS in the aging brain**. FEBS Letters, 2017.

Steffen KK, Dillin A. **A ribosomal perspective on Proteostasis and Aging**. Cell Metabolism, 2016 Jun 14; 23(6):1004-12.

Su Z, Labai PP, Li S, Kreil DP, Mason CE, Shi L. **A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium**. Nature Biotechnology, 2014; 32:903-14.

Sung MK, Porras-Yakunski TR, Reitsima JM, Huber FM, Sweredoski MJ, Hoelz A, Hess S, Deshaies RJ. **A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins**. eLife, 2016; 5:e19105.

Supek F, Bosnjak M, Skunca N, Smuc T. **REVIGO summarizes and visualizes long lists of gene ontology terms**. PLoS One, 2011; 6(7):e21800.

Tabrez SS, Sharma RD, Jain V, Siddiqui AA, Mukhopadhyay A. **Differential alternative splicing coupled to nonsense-mediated decay of mRNA ensures dietary restriction-induced longevity**. Nature Communications, 2017 Aug 21; 8(1):306.

Tanaka K. **The proteasome: overview of structure and functions**. Proc Jn Acad Ser B Phys Biol Sci. 2009; 85(1):12-36

Taylor DS, Turner BJ, Davis WP, Chapman BB. **A novel terrestrial fish habitat inside emergent logs**. Am Nat. 2008 Feb 17; 171(2):263-6

Terzibasi E, Lefrancois C, Domenici P, Hartmann N, Graf M, Cellerino A. **Effects of dietary restriction on mortality and age-related phenotypes in the short-lived fish *Nothobranchius furzeri***. Aging Cell, 2009 Mar 26; 8(2):88-99.

Terzibasi E, Valenzano DR, Benedetti M, Roncaglia P, Cattaneo A, Domenici L, Cellerino A. **Large differences in aging phenotype between strains of the short-lived annual fish *Nothobranchius furzeri***. PloS One, 2008; 3(12):e3866.

Terzibasi Tozzini E, Dorn A, Ng'oma E, Polacik M, Blazek R, Reichwald K, Petzold A, Watters B, Reichard M, Cellerino A. **Parallel evolution of senescence in annual fishes in response to extrinsic mortality**. BMC Evolutionary Biology, 2013; 13(77)

Tian X, Seluanov A, Gorbunova V. **Molecular mechanisms determining lifespan in short- and long-lived species**. Trends Endocrinol Met, 2017 Oct; 28(10):72-734.

Ting L, Rad R, Gygi SP, Haas W. **MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics**. Nature Methods, 2011 Oct 2; 8(11):937-40.

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Jonstone R, Mohammed AK, Hamon C. **Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS**. Anal Chem, 2003 Apr 15; 75(8):1895-904.

Tomko RJ, Hochstrasser M. **Incorporation of the Rpn12 Subunit couples completion of Proteasome Regulatory Particle lid Assembly to Lid-Base joining**. Molecular Cell, 2011 Dec 23; 44(6):907-17.

Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR 3rd, Hetzer MW. **Identification of long-lived proteins reveals exceptional stability of essential cellular structures**. Cell, 2013 Aug 29; 154(5):971-982.

Tozzini ET, Baumgart M, Battistoni G, Cellerino A. **Adult neurogenesis in the short-lived teleost *Nothobranchius furzeri*: localization of neurogenic niches, molecular characterization and effects of aging**. Aging Cell, 2012 Apr; 11(2):241-51.

Tsaytler P, Harding HP, Ron D, Bertolotti A. **Selective inhibition of a regulatory subunit of protein phosphatase 1 restores proteostasis**. Science, 2011 Apr 1; 332(6025):91-4.

Tyers M, Mann M. **From genomics to proteomics**. Nature, 2003 Mar 13; 422(6928):193-7.

Valdesalici S, Cellerino A. **Extremely short lifespan in the annual fish *Nothobranchius furzeri***. Proceedings of the Royal Society B, 2003 Nov 7; 270(2)

Valenzano DR, Benayoun BA, Priya Singh P, Zhang E, Etter PD, Hu CK, Clement-Ziza M, Willemsen D, Cui R, Harel I, Machado BE, Yee MC, Sharp SC, Bustamante CD, Beyer A, Johnson EA, Brunet A. **The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan**. Cell, 2015 Dec; 163(6):1539-1554

Valenzano DR, Kirschner J, Kamber RA, Zhang E, Weber D, Cellerino A, Englert C, Platzer M, Reichwald K, Brunet A. **Mapping Loci Associated With Tail Color and Sex Determination in the Short-Lived Fish Nothobranchius furzeri**. *Genetics, 2009 Dec; 183(4):1385-95.*

Valenzano DR, Terzibasi E, Cattaneo A, Domenici L, Cellerino A. **Temperature affects longevity and age-related locomotor and cognitive decay in the short-lived fish *Nothobranchius furzeri***. Aging Cell, 2006; 5, 275–278.

Valenzano DR, Terzibasi E, Genade T, Cattaneo A, Domenici L, Cellerino A. **Resveratrol prolongs lifespan and retards the onset of age-related markers in a short-lived vertebrate**. Curr Biol, 2006 Feb 7; 16(3):296-300.

Van Verk MC, Hickman R, Pieterse CM, Van Wees SC. **RNA-Seq: revelation of the messengers**. Trends Plant Sci, 2013 Apr; 18(4):175-9.

Vandesompele J, De Preteer K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. Genome Biology, 2002 Jun 18; 3:r0034.1.

Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD, Nordlund P, Sundstrom M, Weigelt J, Edwards AM. **Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination**. Proc Natl Acad Sci USA, 2006 Oct 24; 103(43):15835-40.

Vilchez D, Boyer L, Morantte I, Lutz M, Meckwirth C, Joyce D, Sencer B, Page L, Masliah E, Berggren WT, Gage FH, Dillin A. **Increased proteasome activity in human embryonic stem cells is regulated by PSMD11**. Nature, 2012 Sep 13; 489(7415):304-8.

Vrtilek M, Zak J, Psenicka M, Reichard M. **Extremely rapid maturation of a wild African annual fish**. Current Biology, 2018 Aug 6; 28(15):PR822-R824.

Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, Ciryam P, Morimoto RI, Dobson CM, Vendruscolo M, Mann M, Hartl FU. **Widespread Prtoeome Remodelling and Aggregation in Aging *C.elegans***. Cell, 2015 May 7; 161(4):919-32.

Wang Z, Gerstein M, Snyder M. **RNA-Seq: a revolutionary tool for transcriptomics**. Nat Rev Genet, 2009 Jan; 10(1):57-63.

Wei YN, Hu HY, Xie GC, Fu N, Ning ZB, Zeng R, Khaitovich P. **Transcript and protein expression decoupling reveals RNA binding proteins and miRNAs as potential modulators of human aging**. Genome Biology, 2015; 16(41).

Wildekamp HR. **A World of Killies: Atlas of the Oviparous *Cyprinodontiform* Fishes of the World**. American Killifish Association , Elyria, 2004; Volume 4.

Wildekamp HR, Haas R. **Redescription of Nothobranchius microlepis, description of two new species from northern Kenya and southern Somalia, and note on the status of Paranothobranchius (Cyprinodontiformes: Aplocheilidae)**. Ichthyological Explorations in Freshwaters, 1992 Feb; 1-16.

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF. **From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.** Biotechnology (NY), 1996 Jan; 14(1):61-5.

Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL. **Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it**. Biotechnology Genet Eng Rev, 1996; 13:19-50.

Williams GC. **Pleiotropy, natural selection, and the evolution of senescence**. Evolution, 1957; 11:398–411.

Woodhead AD. **Aging changes in the heart of a poeciliid fish, the guppy *Poecilia reticula***. Experimental Gerontology, 1984;19:383–391.

Wojcik C, DeMartino GN. **Analysis of *Drosophila* 26 S Proteasome using RNA Interference**. The Journal of biological Chemistry, 2002 Feb 22; 277(8):6188-97.

Yang Z, Klionsky DJ. **Mammalian autophagy: core molecular machinery and signalling regulation**. Curr Opin Cell Biol. 2010 Apr; 22(2):124-31

Yates JR. **The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis**. J Am Chem Soc. 2013 Feb 16; 135(5):1629-40.

Yoshino A, Saitoh M, Oka A, Okumura A, Kubota M, Saito Y, Takanashi J, Hirose S, Yamagata T, Yamanouchi H, Mizuguchi M. **Epidiemiology of acute encephalopathy in Japan, with emphasis on the association of viruses and syndromes.** Brain Dev. 2012 May; 34(5):337-43.

Yu X, Li G. **Effects of resveratrol on longevity, cognitive ability and aging-related histological markers in the annual fish *Nothobranchius guentheri***. Experimental Gerontology, 2012; 47:940–949.

Zhang B, Kirov S, Snoddy J. **WebGestalt: an integrated system for exploring gene sets in various biological contexts**. Nucleic Acids Research, 2005 Jul 1; 33(Web Server Issue):W741-8.

Zecca L, Youdim MB, Riederer P, Connor JR, Crichton RR. **Iron, brain ageing and neurodegenerative disorders**. Nat Rev Neurosci, 2004 Nov; 5(11):863-73.

Zoncu R, Bar-Peled L, Efeyan A, Wang S, Sancak Y, Sabatini DM. **mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase**. Science, 2011Nov 4; 334(6056):678-83.