Insights Into The Regulation Of Aging

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt vor dem Rat der Fakultät für Mathematik und Informatik

der Friedrich-Schiller-Universität Jena

eingereicht von Emanuel Barth M. Sc.

geboren am 10.04.1990 in Plauen

January 19, 2019

Gutachter

1. Prof. Dr. Manuela (Manja) Marz, Friedrich Schiller University Jena, Germany

2. Prof. Dr. Steve Hoffmann, Fritz Lipmann Institute Jena, Germany

3. Prof. Dr. Peter Dittrich, Friedrich Schiller University Jena, Germany

Tag der Verteidigung: 09.08.2019

"In the fields of observation chance favours only the prepared mind."

— Louis Pasteur

"Death is a disease. Its like any other. And there's a cure and I will find it."

— Tom Creo, The Fountain

Danksagung

Zu aller erst möchte ich meiner grandiosen Doktormutter Manja, danken, nicht nur für die spannenden und ereignisreichen letzten vier Jahre, sondern auch für das immerwährende Vertrauen in mich und meine Arbeit und dass sie es mir ermöglicht hat, mich während meiner Doktorandenzeit völlig frei zu entfalten. Es war ein langer und manchmal auch sehr anstrengender Weg bis hierhin, aber ich bin sehr froh das ich ihn mit dir gehen durfte, Manja!

Ich möchte mich auch von ganzem Herzen bei meinen Eltern bedanken, denn sie haben mich all die Jahre bedingungslos bei all meinen Unternehmungen unterstützt und nie an mir und dem, was ich mache gezweifelt. Auch meinen Geschwistern Theresa, Julian und Benjamin danke ich, denn auch wenn sie vielleicht nicht immer ganz verstanden haben mit welchen biologischen oder mathematischen Problemen ich mich rumschlagen musste, so habe ich trotzdem immer mindestens ein offenes Ohr bei ihnen gefunden. Danke, ihr alle seid die wundervollste Familie, die man sich nur wünschen kann!

Ganz besonderen Dank möchte ich auch meinem Kindergartenfreund Martin aussprechen, für all die tollen und lustigen, langen Jahre, die wir uns nun schon kennen. Auf dass noch viele weitere kommen mögen! Meinen guten Freunden und ehemaligen Kommilitonen Stefanie, Jan und Marco bin ich auch zu viel Dank verpflichtet, denn mit ihnen bin ich nicht nur während unseres gemeinsamen Studiums durch dick und dünn gegangen, sondern wir haben auch zusammen die Abenteuer der Promotion durchgestanden. Genauso muss ich meiner gesamten Arbeits(turm)crew danken. Die letzten Jahre mit euch waren einfach nur fantastisch und nicht ein einziger Tag mit euch zusammen im Büro, auf Konferenz oder bei allerlei Freizeitaktivitäten war

ein verschwendeter Tag. Franzi, Nelly, Max, Sebastian, Kevin, Flo, Daria, Martin, Marie, Daniel, Lisa, Konrad, Celia, Adrian, Bashar, Akash, und Muriel ihr seid die besten Freunde und Kollegen, die man haben kann, danke!

Danken möchte ich auch meinem kleinen Kater Happy, denn er hat ein wunderbares Gespür dafür, wann es Zeit ist sich einmal quer über die Tastatur zulegen und eine Pause von der vielen Schreiberei einzulegen.

Zu guter Letzt gebührt der größte Teil meines Dankes der besten und wundervollsten Frau der ganzen Welt. Danke, Sigrid, für all die Liebe, die du mir schenkst und all das Verständnis, dass du mir entgegenbringst. Danke für die viele Motivation und Unterstützung und all die anderen großen und kleinen Dinge, die du in mein Leben gebracht und um die du mich bereichert hast. Danke für einfach Alles!

Abstract

Aging is doubtlessly one of the most complex and multi-factorial biological processes we have encountered since the beginning of modern life sciences and the systematic study of human and animal biology. Despite many remarkable findings, aging remains an incompletely understood mechanism, causing several severe diseases, such as cardiovascular diseases, neurodegenerative diseases or cancer. It is associated with a progressive loss of cell functions that lead to a decline of tissue functions and finally resulting in death. Uncountable studies were performed over the last five decades to identify possible causes of how and why we age. Nevertheless, there is a still ongoing debate about the true molecular source of aging, giving rise to a variety of competing theories.

Due to its highly complex nature, we have investigated aging from various perspectives, based on the gene expression of different species and tissues. We analyzed a huge set of RNA-Seq transcriptomic data to obtain new insights into the genetic regulation of aging and to identify conserved molecular processes that might be responsible for aging-related disorders. We found that each tissue shows its own distinct pattern of gene expressional changes with age, because they have to respond to different types of stress over time, leading to differing sources of molecular damage and subsequent stress responses. In particular, we could show this for four well-studied aging-related processes: cellular senescence, inflammation, oxidative stress response and circadian rhythms. In addition, we could show that alternative splicing (*i.e.*, the generation of multiple mRNA isoforms from single genes) is in general only slightly affected by aging and probably plays a secondary role in the overall aging process. In contrast, we found microRNAs (very small regulatory RNA molecules) to be important modulators of aging in all investigated species and tissues.

Concluding, the results presented in this thesis describe aging as a stochastic process, leading to an accumulation of different kinds of molecular damage and the respective cellular stress responses. We have identified several genetic factors that could serve as potential diagnostic markers or even therapeutic targets, that could help in the future to slow down the progression of age-associated disorders or preventing them. Nevertheless, the subject of aging remains a challenging research field and many open questions still wait to be answered.

Zusammenfassung

Das Altern ist zweifellos einer der komplexesten und faktorenreichsten biologischen Prozesse, dem wir seit Beginn der modernen Lebenswissenschaften und der systematischen Erforschung der Human- und Tierbiologie begegnet sind. Trotz vieler bemerkenswerter Erkenntnisse bleibt das Altern ein unvollständig verstandener Mechanismus, der mit vielen schweren Krankheiten assoziert ist, wie etwa Herz-Kreislauf-Erkrankungen, neurodegenerative Erkrankungen oder Krebs. Es geht mit einem fortschreitenden Verlust von Zellfunktionen einher, der zu einer Abnahme der Organfunktionen und schließlich zum Tod führt. In den letzten fünf Jahrzehnten wurden unzählige Studien durchgeführt, um mögliche Quellen zu finden, wie und warum wir altern. Auch heutzutage wird noch intensiv über die eigentliche molekulare Ursache des Alterns gestritten, was zu einer Vielzahl konkurrierender Theorien führt. Aufgrund seiner hohen Komplexität haben wir in der vorliegenden Arbeit das Altern aus verschiedenen Perspektiven untersucht, basierend auf der Genexpression verschiedener Spezies und Gewebstypen. Wir haben eine Vielzahl an RNA-Seq Transkriptomdaten analysiert, um neue Einblicke in die genetische Regulation des Alterns zu erhalten und um konservierte molekulare Prozesse zu identifizieren, die möglicherweise für altersbedingte Gebrechen und Krankheiten verantwortlich sind. Wir fanden heraus, dass jedes Gewebe mit dem Alter ein eigenes Muster von Genexpressionsänderungen aufweist, da es im Laufe der Zeit auf verschiedene Arten von Stress reagieren muss, was zu unterschiedlichen Ursachen für molekulare Schäden und nachfolgende Stressreaktionen führt. Insbesondere konnten wir dies für vier gut untersuchte, altersbedingte Prozesse zeigen: zelluläre Seneszenz, Entzündungsreaktionen, oxidative Stressreaktion und zirkadiane Rhythmen. Außerdem konnte gezeigt werden, dass alternatives Spleißen (d. H., die Erzeugung mehrerer mRNA-Isoformen aus einzelnen Genen) im Allgemeinen nur geringfügig vom Altern betroffen ist und wahrscheinlich eine untergeordnete Rolle im gesamten Alterungsprozess spielt. Im Gegensatz dazu haben wir festgestellt, dass microRNAs (sehr kleine regulatorische RNA-Moleküle) in allen untersuchten Spezies und Geweben wichtige Alterungsmodulatoren sind.
Abschließend beschreiben die in dieser Arbeit vorgestellten Ergebnisse das Altern als einen stochastischen Prozess, der zu einer Anhäufung verschiedener Arten von molekularen Schäden und den jeweiligen zellulären Stressreaktionen führt. Wir haben mehrere genetische Faktoren identifiziert, die als potenzielle diagnostische Marker oder sogar als therapeutische Ziele dienen könnten, die in Zukunft dazu beitragen könnten, das Fortschreiten altersbedingter Erkrankungen zu verlangsamen oder zu verhindern. Trotzdem bleibt das Thema Altern weiterhin ein herausforderndes Forschungsfeld und viele offene Fragen warten noch auf ihre Beantwortung.

This thesis is first and foremost based on the following publications:

1. Mario Baumgart$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, Aurora Savino, Marco Groth, Philipp Koch, Andreas Petzold, Ivan Arisi, Matthias Platzer, Manja Marz, Alessandro Cellerino. "A miRNA catalogue and ncRNA annotation of the short-living fish Nothobranchius furzeri". In: *BMC Genomics*, 18 (693), 2017.

2. Akash Srivastava$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, Maria A. Ermolaeva, Madlen Guenther, Christiane Frahm, Manja Marz and Otto W. Witte. "Tissue-specific gene expression changes are associated with aging in mice." In: *Genomics, Proteomics & Bioinformatics* (accepted).

3. **Emanuel Barth**$^{\Omega}$, Mario Baumgart$^{\Omega}$, Luca Dolfi, Marco Groth, Roberto Ripa, Aurora Savino, Matthias Platzer, Manja Marz, Alessandro Cellerino. "Analysis of microRNA expression reveals convergent evolution of the molecular control of diapause in annual fish". *In preparation for Science Ecology.*

4. Mario Baumgart$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, [...], Manja Marz, Alessandro Cellerino. "microRNAs in aging: potentially conserved key regulators of aging". *In preparation for Cell.*

5. **Emanuel Barth**, Sebastian Krautwurst, Manja Marz. "Using gamma distributions to model small RNA-Seq expression data". *In preparation for Bioinformatics.*

6. **Emanuel Barth**$^{\Omega}$, Akash Srivastava$^{\Omega}$, Diane Penndorf$^{\Omega}$, Milan Stojiljkovic, Hubertus Axer, Otto W. Witte, Alexandra Kretz, Manja Marz. "Inter-species and inter-organ comparisons reveal aging-associated changes in clock gene expression". *Submitted to PLOS Computational Biology.*

7. **Emanuel Barth**$^{\Omega}$, Akash Srivastava$^{\Omega}$, Milan Stojiljkovic, Christiane Frahm, Hubertus Axer, Otto W. Witte, Manja Marz. "Conserved Aging-related Signatures of Senescence and Inflammation in different Tissues and Species". *Aging*, vol. 11, 2019.

8. Patricia Sieber$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, Manja Marz. "The landscape of the alternatively spliced transcriptome remains stable during aging across different species and tissues". *Submitted to Giga Science.*

9. Akash Srivastava$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, Maria A. Ermolaeva, Madlen Guenther, Christiane Frahm, Manja Marz and Otto W. Witte. "Inching towards decrypting the enigma of Aging." In: *Neuroscience: Hot Topics*, 2018.

Ω These authors contributed equally to this work.

... and partially based and complemented by:

1. **Emanuel Barth**, Ron Huebler, Aria Baniahmad, Manja Marz, "The Evolution of COP9 signalosome in unicellular and multicellular organisms". In: *Genome Biology and Evolution*, Volume 8, Issue 4, 1 April 2016, Pages 1279–1289.

2. Martin Hoelzer, Verena Kraehling, Fabian Amman, **Emanuel Barth**, [...], Stephan Becker, Manja Marz. "Differential transcriptional responses to Ebola and Marburg virus infection in bat and human cells". In: *Scientific Reports*, Volume 6, Article number: 34589 (2016).

3. Konstantin Riege, Martin Hoelzer, Tilman E. Klassert, **Emanuel Barth**, Julia Braeuer, Maximilian Collatz, Franziska Hufsky, Nelly Mostajo, Magdalena Stock, Bertram Vogel, Hortense Slevogt, Manja Marz. "Massive Effect on LncRNAs in Human Monocytes During Fungal and Bacterial Infections and in Response to Vitamins A and D". In: *Scientific Reports*, Volume 7, Article number: 40598 (2017).

4. Diana Morales$^{\Omega}$, **Emanuel Barth**$^{\Omega}$, Wittaya Chaiwanyen, Ruby N. Gutierrez-Samudio, Stephanie Ospina-Prieto, Manja Marz and Udo R. Markert. "Identification of miRNAs and associated pathways regulated by Leukemia Inhibitory Factor in trophoblastic cell lines". In: *Placenta*, vol. 88, p. 20 – 27, 2019.

5. Laura Graf, Alexej Dick, Franziska Sendker, **Emanuel Barth**, Manja Marz, Oliver Daumke, Georg Kochs. "Effects of allelic variations in the human Myxovirus resistance protein A on its antiviral activity". In: *Journal of Biological Chemistry*, 2018.

6. Kevin Lamkiewicz, **Emanuel Barth**, Manja Marz, Bashar Ibrahim. "Identification of potential microRNAs associated with Herpesvirus family based on bioinformatic analysis". *Submitted to MDPI Viruses.*

7. Kevin Lamkiewicz, Michele Kayser, **Emanuel Barth**, Manja Marz. "ViRAlign: building secondary structure alignments of whole viral genomes". *In preparation.*

8. Florian Mock, Adrian Viewweger, **Emanuel Barth**, Manja Marz. "Viral host prediction with deep learning". *bioRxiv*, p. 541417, 2019.

Ω These authors contributed equally to this work.

Contents

Chapter 1

Introduction

When we come to think about aging, everyone has an intuitive understanding of this concept. This is because we observe it everyday in the people around us and also what it does to ourselves. We experience it as a rather slow but natural process that irreversibly restrains our mobility, health and sometimes also impairs our cognitive abilities, however, always resulting in death.
Everlasting life was always a dream of mankind and with the decoding of the human genome in 2000 and the subsequent revolution of molecular techniques to investigate and manipulate genes and their expression, one was sure to soon find treatments for many (genetic) diseases and finally also something like a "fountain of youth" [1, 2]. These expectations were further impelled by early findings of life-span enhancing genetic programs in invertebrates and other lower organisms. Despite being already known for decades that caloric restrictive diets (a.k.a. *caloric restriction* or *dietary restriction*) have positive effects on the maximal age of a variety of animals, the molecular mechanisms for this effect were just discovered in the early 2000s in the worm *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* [3–5]. A few years later, after there was first evidence that (short-term) caloric restriction slows aging even in primates [6], many believed to finally be on the right track to fully understand the aging process. However, this kind of optimism had to make way for the realization that there is a huge difference between merely knowing the code of life and truly understanding it and that many findings in small short-lived animals could not be so easily transfered to longer-lived species [7] and on the contrary could even be harmful [8]. Making it even worse, in the early years after having the first "complete" version of the human genome the understanding dawned that there are more regulatory levels controlling gene activity than expected, especially in humans and likewise other higher eukaryotes [9]. For example, only a minor fraction of our genomes code for proteins and it was long believed that everything else was just *junk DNA*, *i.e.*, sequences without any function apart from some relatively small promoter regions where proteins could bind specific sites in order to control the activation of certain protein-coding genes [10]. However, it was soon discovered that this assumption was false. More and more so-called *non-coding genes* were identified, not acting as messengers between DNA and proteins but fulfilling enzymatic and regulatory functions as RNA molecules [11, 12]. And still today, we are discovering even more layers of genetic regulation, such as DNA modifications or are

suspecting modulation by genomic elements that we have barely understood, such as repeat regions or pseudogenes.
Concluding, after almost 20 years since the decoding of the human genome it subjectively feels like we have less understanding of how life works than before. And this includes aging too, because it is so deeply entangled with everything we know about life, that every new found or characterized biological process appears to add to its complexity, making it more challenging to understand. Nevertheless, people still dream of a very long and healthy life and a lot of effort is taken into solving aging, because sooner or later everyone is confronted with it.

1.1 What is aging and why do we have to age?

What is aging? That is a really tough question and finding a fitting definition for this phenomenon isn't easy, because we still have only little knowledge about it. In almost all scientific articles or reviews regarding this topic [13–15], aging is described as a complex, multi-factorial process associated with a decline of cellular functions over time. Often it is treated as an illness and described by its associated obvious physiological symptoms, such as impairment of the whole sensory system, deterioration of cognitive functions (loss of memory, rising difficulty to learn new things, ...), frailty (especially in mobility and body coordination), increased susceptibility for diseases or visible physiological changes (wrinkling of the skin, graying and/or loss of hair, ...). In addition, there are many so-called age-related diseases, which mainly effect the elderly and have rising incident with age, as for example most types of cancer, neurodegenerative diseases (*e.g.*, Parkinson's disease, Alzheimer's disease), metabolic disorders (most prominent diabetes) and cardiovascular diseases. All of these are dysfunctions associated with late life and the "advancement of age" is the highest risk-factor for their origin [16]. However, this is more of a recursive definition, that does not really help in uncovering the true molecular systems causing these typical signs of aging. And only in understanding these underlying processes we can completely perceive why we have to age.
In general we can say, that aging is associated with a progressive loss of cell functions that lead to more and more dis-regulated tissue functions and finally to a total system failure, where all vital processes collapse, called death. And we also do know that many different relatively smaller mechanisms play a role in this big process that we call aging (more on that in the following Section). Alternatively, we could say that aging is a function of time, responsible for the decline of cellular function. One answer to the question why we age would be: because our vital system fails to keep homeostasis. And for some reason there was never evolutionary pressure to overcome this "flaw" (if it is one at all). Another interpretation could be that there exists something like a "biological clock", and evolution had arranged it so to restrict the life-span of each generation to make room for the next generation and thus ensuring genetic variability. In Section 1.1.2 pros and cons for both interpretations will be further presented and discussed.

Before we move on, an important distinction between longevity and aging has to be made, because they can be confused as two similar phenomenons. Longevity refers

to the maximal and average life expectancy of species populations [1] and can be measured relative easily. Additionally, there exist some known genetic components, that have a direct influence on the life-span of certain organisms, like the so-called *Sirtuin* genes [8, 16]. In contrast, aging is more difficult to assess, because there are many different factors accounting for it [17]. It appears not so easy to identify universal genetic markers, that can be used to estimate the biological age or get a grasp on this heterogeneous process. This topic will be discussed in more detail in Section 4.4.

In the following sections, I want to give a comprehensive overview about what we already know about the complexity of aging (Section 1.1.1) and what is believed and still discussed of how it is caused and why aging happens at all (Section 1.1.2). After that I will also give a short introduction into a technique called *RNA Sequencing* or in short *RNA-Seq* (Section 1.2), because many of the data analyzed and discussed within this thesis are based on this approach to measure gene activity. In Section 1.3, I will explain the current restrictions in aging research and what problems arise with these constraints and how they could possibly be overcome.

1.1.1 The complexity of aging and how it may be caused

Aging is doubtless one of the most complex biological processes we have encountered since the beginning of modern life sciences and the systematic study of human and animal biology. Despite many remarkable findings, aging remains an incompletely understood mechanism, causing several severe diseases, such as cardiovascular diseases, neurodegenerative diseases or cancer.

Uncountable studies were performed over the last five decades to identify possible causes why we get more and more frail with the passage of time and many remarkable findings were discovered. One of the most early ones was that the restriction of calories has a surprisingly huge impact in the average life-spans of different animals [3–5]. The positive effects of this dietary restriction on longevity were first systematically described in 1934 by McCay and colleagues in rats [18]. Although missing a molecular answer for this observation, many scientists repeated and refined caloric restriction (CR) experiments with confirming results, initiating a new aging research era in the hope to have found a fountain of youth in form of long-term CR [19]. It was not until nearly 70 years later, that first evidence showed that CR is connected to certain signaling pathways responsible for cellular stress regulation and modification of metabolic processes (for a comprehensive review on this topic, see Haigis et al. [20]).

Today their are reasonable doubts that CR could be used as a potential intervention to increase life expectancy in human [7, 8]. With the beginning of the new century, we saw a rapid advancement in biotechnology and subsequently in all life sciences, changing the belief that aging can be explained by rather simple processes or single

[1] see for example the WHO records for humans `http://www.who.int/gho/mortality_burden_disease/life_tables/en/`

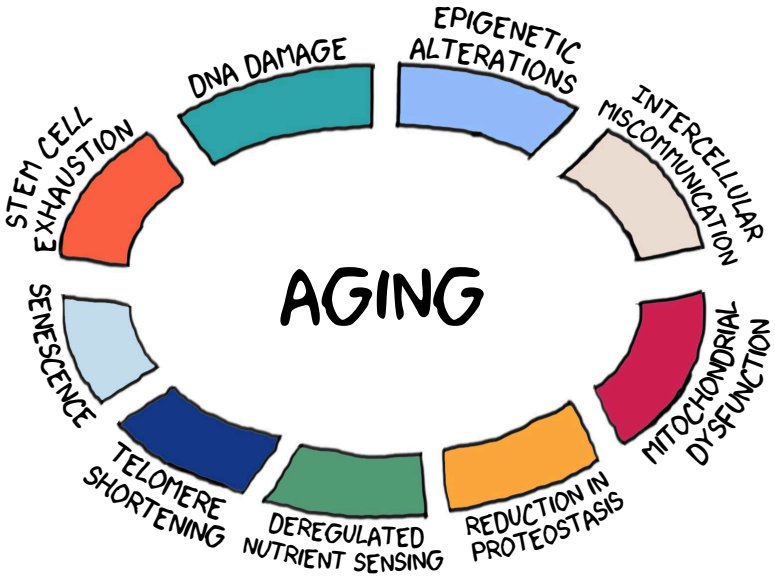


Figure 1.1: **The nine "hallmarks" of aging.**
As adapted from Lopez-Otin et al. [13], they resume these processes to be the main causes for cellular aging.

genes alone and in fact is a multi-factorial process. Lopez-Otin et al. have recently identified nine primary process to be responsible for aging [13] (see Figure 1.1). These "hallmarks" are not independent mechanisms that drive the course of aging, but are highly intertwined and having strong influences on each other. Figure 1.2 is an attempt, based on recent research results, to describe those interactions and relations in more detail, nonetheless, not suggesting to have drawn a complete picture of the current understanding of the aging process.

One of the most reoccurring terms regarding aging and its causes is *cellular stress*. This broad term can be broken down to any event that triggers a specific reaction in order to maintain cell homeostasis. A cell has stress, if it is provoked to activate any means of stress response, such as repair mechanisms due to physical or chemical damage or metabolic pathways in the presence or absence of certain (macro)molecules to sustain itself. Cellular stress was first recognized to be strongly linked to aging in 1995 when it was shown that mild thermal stress could extend the mean life-span of the nematode *Caenorhabditis elegans* [21]. Following research proved many more relations between the stress biology and aging biology (see Epel et al. [22] for a recent review). For that reason, cellular stress is one of the main hubs in Figure 1.2 and many origins were already verified, *e.g.*, oxidative stress due to reactive oxygen species (ROS) [23], accumulation of malignant proteins [24], environmental exposure [25], deregulation of (nutrient) signaling pathways [13, 20] and both the decline in immune activity and the over-activation of inflammatory processes [26, 27].

Another central and well-established factor in the process of aging is cellular senescence, a non-replicative arrest state in the cell's life cycle that prevents the formation of cancer due to abundant damage within the cell. Once a cell has entered

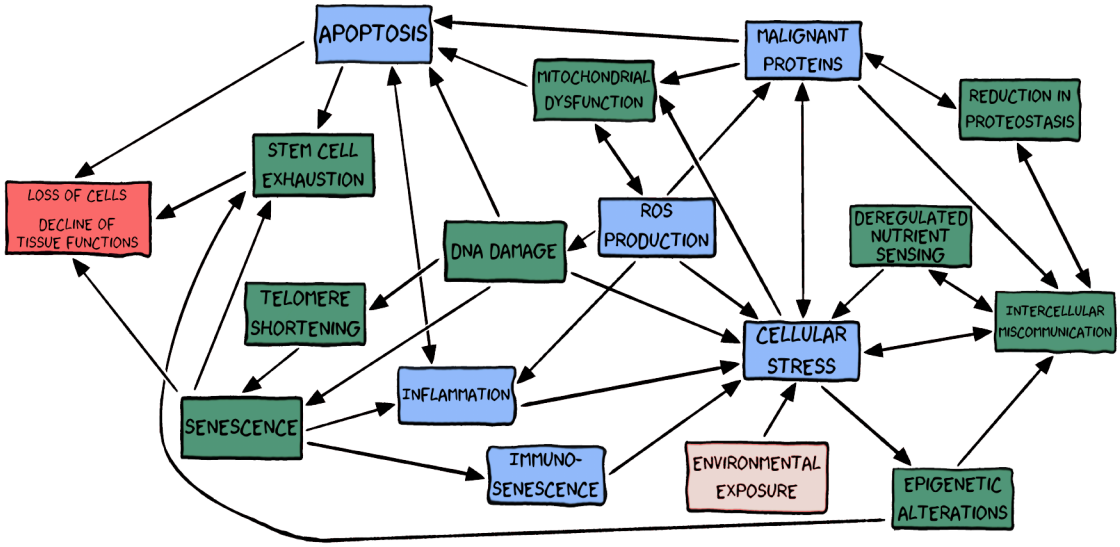


Figure 1.2: **Draft of the aging network.**
Here, as an extension to the nine "hallmarks" from Lopez-Otin et al. [13], the aging process is depicted as a directed graph, towards its ultimate pathologic consequence, the complete loss of cellular function (red). New but minor internal (light blue) and external (cream-colored) processes were added based on the reviewed literature cited in the main text of this section, in order to draw relations between the main mechanics of aging (green). Arrows indicate causal relationships described in different published studies, *e.g.*, stem cell exhaustion is caused by senescence and apoptosis and causes the decline of tissue function. This network is of course not complete and single processes, like "cellular stress" or "environmental exposure", could be further split up into more specific categories. However, for the sake of clarity and comprehensibility this stripped-down version already shows how complex and entangled biological aging is.

senescence it will stop growing and replicating, but continues to sustain itself. The main causes of senescence are the shortening of telomeres [28] and direct damage to the DNA [29]. Inevitably, proliferative somatic and stem cells are lost over time, resulting in a decline of functional and regenerative capacity [30]. As a further consequence, senescent cells develop a pro-inflammantory secretory phenotype, adding to the stress of neighboring cells [29]. Paradoxically, besides the constant activation of the immune system by inflammantion processes, immune activity decreases because of the senescence of immune cells (also known as *immunosenescence*), which leads to a higher vulnerability to infections and yet another source of cellular stress [31, 32].

Next to replicative senescence, apoptosis (programmed cell death) is likewise an immediate cause for the loss of somatic and stem cells. It can be triggered by different events, including hyper-inflammation (as it occurs through senescent cells as described above), the release of the protein cytochrome c from dysfunctional or disrupted mitochondrias, infections, abundant DNA damage and misfolded proteins [33, 34]. However, in the process of apoptosis, certain cellular components are released that can activate inflammatory processes and in a context of already

existing constant inflammation, boosting the cycle of self-harming development [35]. Besides controlling one apoptotic signaling way, mitochondrias are the main power source of our cells by providing chemical energy through respiration. As a natural byproduct, ROS are produced and when being too abundant they can damage DNA as well as proteins [13]. Such dysfunction often happens in times of general cellular stress [24]. The creation of malignant proteins is not only caused by ROS but also inaccurate folding and irregular complex aggregation [36]. These malignant proteins can exhibit genotoxic effects by disturbing cellular communication and sometimes even interfere with the cells normal degradation mechanisms, making them impossible to eliminate and leading to an accumulation of disruptive molecules. This can further contribute to the reduction of proteostasis, ending in the generation of more deleterious proteins [24].
Last but not least, exposure to the environment pose as an additional threat. The huge variety of external factors that can directly damage DNA, proteins and other macromolecules within a living organism, is an extensive source of cellular stress [22, 25].

Summing everything up, there are plenty possible starting points for the decline of individual cells and subsequently whole tissues. Moreover, due to complex interplays and (in parts still unknown) dependencies between all these biological processes, the balance to sustain homeostasis indefinitely appears to be almost impossible.
We see animals aging differently with a huge variance in life-spans, from the short-lived Mayflies ($\sim$24 hours) to the long-lived giant tortoise (>200 years). Even individuals within one species or population show a divers range of life-spans, for example most humans die around 80 years of age, whereas some humans reach the age of 100 or more (the oldest known individual was the French woman Jeanne Calment who died at the age of 122). And inside any individual every tissue shows its own pattern and rate of aging (see Chapter 3). Possibly every cell ages differently and at its own pace and what we can observe is the "average" aging process, that should not be confused with any so-called "general" aging mechanism. This of course makes the whole story even more complex and challenging, but also more interesting. And still a lot of questions remain unanswered with one of the most intriguing being: What is the main driving factor of aging and can it be reversed?

1.1.2 Theories, so many theories!

Because of its incompletely understood nature, there is a decades-old still ongoing debate about the true source of aging, giving rise to a variety of competing theories. There is no doubt in the scientific community that all facets of aging cannot be explained by just one single molecule or genetic pathway (since they are highly connected, as described in the previous section). However, the molecular origin, or in plain words the "starting point" of aging is still highly discussed. The intention of this section is to give an overview of recent but differing arguments on how aging happens on a molecular level, rather than participating in this discussion. Nevertheless, own conclusions will be drawn and justified in the subsequent chapters, when different aspects of aging will be discussed. A final conclusion can be found

in the last chapter of this thesis. For now, I will provide in brief a basis of what are currently the potential driving factors of aging.

In general, all theories about the molecular source of aging can be divided into two classes: The *programmed aging* theories and the *failure accumulation* theories [15]. Since they are in the minority and over the past years less and less evidence could be brought up, I will start with the first group.

Evolutionary genetic theory:
This theory is one of the oldest regarding the reason and the purpose of aging and argues that a species benefits from a constant change in the compilation of its genetic repertoire by letting older individuals die and younger ones take over [37]. Therefore, there has to be a genetic program that allows such a "turnover of generations". However, there was never much evidence that this is true [15]. Nevertheless, pushing this idea was the discovery that in several species there are genes that, if inactive, can prolong the life-span of individuals [38–42]. Consequently, their activation was correlated with the decline of cellular functions and some evolutionary biologists argued that the existence of those genes would not make sense in respect to natural selection, if not for triggering death [43–45].

Telomere loss theory:
Telomeres are the terminal regions at the ends of each chromosome arm, consisting of many relative short nucleotide repeats [46]. They protect the genetic code, because with every genome duplication during a cell's division the chromosomes get shortened at their terminal regions due to the *end replication problem* [47]. Therefore, with every cell division the telomeric regions get shorter (from around 11 kb in human newborns to less than 4 kb on average in centenarians [48, 49]), triggering the cellular senescence replication stop if a critical length is reached. Cells within replicative senescence are unable to further proliferate and their overall function is strongly diminished [50]. During an organisms lifetime, most cells continuously undergo proliferation and cell division and reach the state of senescence at their own pace, which over time leads to an accumulation of senescent cells accompanied by loss of function and integrity of the respective tissues. Some concluded from this observation that telomeres take the role as an internal "countdown" for proliferating somatic cells and restrict division cycles to prevent cancer but also regulate the maximum life-span [51–53]. There exists an enzyme called telomerase, which can renew shortened telomeres, however, it is only active in germ cells and stem cells and some cancer cells, but not in normal somatic cells [54].

Antagonistic pleiotropy theory:
In 1953, Williams [55] argued that evolution would favor genes that have (small) positive effects during the early life, even though they have negative effects during late life [14]. That is because, an early advantage in replication outweighs harmful effects that occur later in an species life history even at the price of aging and death. One of the most prominent examples is the senescence key gene *p53*, which suppresses cancer early in life, but is responsible for the accumulation of senescent

cells later in age [56]. And since most wildlife animals die earlier than their life-span would allow due to accidents, disease, starvation or predation, there was never any evolutionary pressure on these pleiotropic effects [57].

Epigenetic clock theory:
Epigenetic marks are defined as heritable and reversible chemical modifications of histone proteins in order to change chromatin structures or direct changes of the DNA (*e.g.*, through cysteine methylation) [58]. Especially the methylation of DNA is known to be associated with the down-regulation of genes and epigenetic studies are relatively new in the field of aging research [25]. Many independent studies could show that there is a strong correlation between aging and DNA methylation [59–61], leading to new hypotheses that there is an epigenetic program that restricts lifespans by turning down genetic processes with time passing by [62, 63].

All four theories implicitly suggest that the aging process is directly modulated by evolution for every species and that its existence can be explained by natural selection. However, there are strong arguments against such implications. For example, different studies on the mortality of wild animals showed, that in many species almost all individuals die relatively early due to predation, cold or starvation and only few reach an age were factors like accumulation of senescence or epigenetic methylation marks become relevant (for a brief review, see [64]). This implicates that there is less to no evolutionary pressure on any potential genetic program of aging and therefore does not exist or is extremely robust, but not identified yet. Especially the "epigenetic drift" could not be proved to be more than random changes that add up over time, eventually silencing important cellular processes just by chance [65, 66]. Most importantly, none of the theories can explain the non-existence of "cheaters", *i.e.*, individuals that benefit from an deactivated aging program, because the overall suggested evolutionary benefit applies only for the whole group of a species and not for an individual alone [15]. Nevertheless, there are still proven genetic pathways that, when modified, can shorten or prolong life in different animal models significantly, but are mostly counted as longevity instead of aging processes.

Next, I will introduce the main ideas behind the *failure accumulation* theories.

DNA damage theory:
This theory is a generalization of the *free radical theory of aging* [67], based on the fact that cells can enter replicative senescence if their DNA is damaged beyond repair. Due to several internal and external sources of stress, different impairments of the genomic DNA accumulate, like DNA single or double strand breaks, non-canonic mutations or chemical modifications of single bases. Besides having different and complex DNA repair mechanisms, if a cell can not cope with the damage senescence will be triggered to prevent the transformation into a cancer cell. Within the state of senescence a cell fully stops growing and replicating, however, not making place for a new completely functional cell, thus preventing it to participate in sustaining tissue homeostasis. As a consequence, more and more cells enter senescence over time and organ functions decrease to a point where the maintenance of the whole

body collapses.

Mitochondrial theory:
In principal following the same reasoning as the DNA damage theory, the mitochondrial theory additionally identifies the cell's mitochondria not only as the main but also determining source of detrimental reagents for the DNA, namely the *reactive oxygene species* (ROS) [23, 68, 69]. ROS are generated as a byproduct of the mitochondrial electron transport chain, the central energy producing process within cells, and play a natural role in some cell signaling pathways [70]. Mitochondrias hold their own DNA (mtDNA), encoding for integral proteins of the oxygen processing machinery. Since the mtDNA is in close proximity to the side of the metabolic production of ROS, it is extremely susceptible to DNA damage induced by ROS overproduction. However, somatic mutations in mtDNA result in impaired mitochondrial proteins and an increasing dysfunctional electron transport chain, consequently producing even more harmful ROS [13]. As a consequence, with time mitochondrias become more unstable and slowly transmute to internal genotoxic threats, culminating in the activation of replicative senescence or cellular death. Nevertheless, there are conflicting experimental observations in different animals, showing an unaffected or even shortened life-span, after a reduction of oxidative stress by mitochondrias or the complete removal of mtDNA [71–73].

Proteostasis theory:
Sometimes also referred to as *waste accumulation theory*, it is based on the assumption that the main origin of internal cellular stress is caused by misfolded proteins and deleterious protein aggregations [36]. The argumentation is similar to the ones above: over time accumulation of disturbed proteins results in a self-enhancing cycle of stress-induced interference with molecular processes, that in turn lead to an increased formation of harmful proteins triggering senescence. Common examples for such events are neurodegenerative diseases like Alzheimer's or Parkinson's disease and cancer, which are known to be triggered by dysfunctional and constrained protein complexes [74]. Normally, certain proteins are responsible for the correct folding of the majority of newly translated proteins (so-called chaperons or heat shock proteins) or the adequate degradation of old and non-functional proteins (mainly the lysosome). With age, these mechanisms get inaccurate and reduced, possibly due to different internal and external stresses, making cells incapable of controlling the natural protein turnover, resulting in the above described fatal consequences [24].

Inflammation theory and immune theory:
Chapter 5 is dedicated to the topic of immune system changes during aging. There, we will deal with these two overlapping theories in more detail. The central argument of the two theories can be condensed as follows: inflammatory processes and cells of the immune system (especially macrophages) are prone to reach an imbalance in their activity, damaging surrounding tissues and triggering a self-reinforcing process of more self-harming activity [57].

Environment theory:
In contrast to the four aforementioned theories, supporters of the environmental theory (or closely related ones), claim that cellular changes and damages causing aging are mainly driven by exposure to external stresses and outweigh all internal sources of cellular stress [25, 75]. These environmental factors include ultraviolet light, diet, ingested or inhaled toxic chemicals, heavy metals, radiation, heat or cold exposure and psychological stress. Besides from directly damaging DNA, proteins or entire cells the main influence of these factors is reflected by the changes in epigenetic modifications (especially DNA methylation) [76, 77]. This age-dependent epigenetic drift causes a general down-regulation of gene transcription, resulting in intracellular miscommunication and deregulation of important biological processes [78]. Different cohort and twin studies could prove a direct and strong connection between the rate of epigenetic changes and aging, giving the topic of epigenetics much attention in the field of aging in the recent years [79–81].

Stress theory of aging:
It is currently the most accepted one, plainly saying that different causes for aging do not have to be mutually exclusive. Its main statement is, that aging is most likely the combination of the afore mentioned processes acting and interacting in parallel on different functional levels [82–84]. However, the progression of aging is tightly linked to all kinds of stress that an individual experiences through out its life time. The ability to handle internal and external stresses by any means, is proclaimed to determine the rate of aging [22]. This of course makes sense, in light that many aging theories overlap in their argumentation. As discussed in Section 1.1.1, aging is most likely caused by the complex interplay of different molecular sources where cellular stress appears to be one of the central points.

Even today, there is still a big discord in the aging science community, regarding the actual aim of aging research and the proper use of the respective results. Whereas some believe that the aging process could be modified or even reversed to extend human life expectancy drastically, others claim that this can never be achieved (or should not be achieved due to potential negative consequence for our society) and the main goal should be to provide a healthy and frailty-free late life [84–86].

1.2 Bioinformatics and *omics*-analyses in aging research

After we have learned in the previous sections what we currently know or believe and most importantly do not know about aging, let us turn from this theoretical (and in parts philosophical) topic to a more technical one. In this section, I will mainly introduce the field of transcriptomics and modern gene activity measurements, because most of the analyses described within this thesis are based on such expression data. Nevertheless, I will mention other *omics*-technologies and their application and contribution to aging research, without going too much into detail there. The

focus remains on high throughput sequencing technologies of the second generation and the analysis of their respective output.

1.2.1 A very brief introduction to RNA sequencing

It is a well known fact, that cells can alter the expression of their genes, in order to adapt to environmental changes, switch between different life stages (*e.g.*, proliferation, senescence, apoptosis) or react to stress such as infections, starvation or physical injuries [87–89]. A change in active gene expression alters the currently present transcriptome within cells, *i.e.*, the presence and amount of all transcribed RNA molecules. By measuring the changes of RNA transcript quantities, it is possible to infer changes of the activity of specific genes and subsequently draw conclusions about the detailed function of cellular mechanisms as well as the genes involved in these processes [90]. A common scenario is to measure the gene activity of two or more different conditions, say healthy against diseased cells, cells during different stages of the cell cycle, or in our case young against old cells, and then to contrast the obtained gene expressions to ascertain which genes are up- or down-regulated between the investigated conditions. These differentially expressed genes (DEGs) can now help to figure out the cellular program under specific conditions or stresses. This can be used to reconstruct the underlying molecular mechanisms, which finally can be used, for example, for targeted therapeutic interventions or genetic improvements if completely understood [91]. The invention of the so-called next-generation sequencing (NGS) technologies by the end of the 2000s made it possible for the first time to sequence tens of millions of DNA molecules at once in a time and cost efficient way [92]. The term NGS encloses different sequencing technologies from different companies (*e.g.*, *TrueSeq* by Illumina or *IonTorrent* by Thermo Fisher Scientific) that all fulfill in general the same task: rapid high throughput sequencing of DNA, thus enabling big data analysis in life sciences [93, 94]. Since it is easily possible to translate RNA molecules into DNA via reverse transcription [95], it was only a short way to utilize NGS for transcriptome studies and was hence called *RNA sequencing* (RNA-Seq). However, nowadays RNA-Seq can be used for a lot more analyses besides the classical gene expression comparison. The most common ones are identification of novel transcripts or new transcript isoforms, variation analysis of single or short nucleotide mutations in transcribed genomic regions, fusion gene detection or reconstruction of whole transcriptomes [96, 97].

1.2.2 Transcriptomics in a nutshell

In general, an RNA-Seq experiment consists of four major steps: isolation of RNA, library preparation, DNA amplification and sequencing (see Figure 1.3, steps I-IV). The extraction of RNA is mostly done by the *single-step-method* developed by Chomczynski and Sacchi [98], which could demonstrate that after extraction RNA can easily be separated from DNA and other cell compartments within an acidic solution and one centrifugation step. The library preparation step consists of several subordinate steps, that define how closely the isolated RNA will represent the original extracted RNA population. Since the transcriptome consists of many different

classes of coding and non-coding RNAs, it is necessary to further process and filter the extracted RNA, based on the kind of biological question to be answered. In almost all cases ribosomal RNA (rRNA) molecules account for up to 95 % of all RNA molecules within a cell, therefore their concentration has to be depleted in order to increase the measurement accuracy for all other types of RNAs [99]. Messenger RNAs (mRNAs) can be specifically targeted by their poly-adenosine tails either to be enriched or filtered. This enables the possibility to analyze only the protein-coding or only the non-coding part of a transcriptome. Further, RNA molecules can be selected by their size, for example, a common protocol named *small RNA-Seq* filters especially for smaller RNA molecules, such as microRNAs (miRNAs) [100].
After the isolation steps, RNAs must be fragmented and translated into complementary DNA (cDNA) or, depending on the specific NGS technology protocols used, first translated and then fragmented. Fragmentation is important to reduce possible secondary structures, often present in non-coding RNAs, and to fit the cDNAs to the standard length of the chosen NGS method, because the accuracy of the subsequent sequencing is limited to a relative small size of $50-200$ bases per fragment [99]. The resulting cDNA fragments are than amplified, so the abundance of single molecules is high enough for sequencing.
Finally, the created RNA library will be sequenced, resulting in a set of tens of millions sequenced transcript fragments called *reads*. All subsequent analysis steps will be performed on these reads, including, *e.g.*, read mapping, counting and further specific investigations (see Figure 1.3, steps V-VII).

The normally relatively short reads obtained by an RNA-Seq experiment are fragments of transcribed regions of the investigated species' genome, however, from which exact location they originate is not known at that moment. To solve this, all reads (or at least all reads that have a minimum of quality, often measured in probability scores of sequencing errors; the topic of read quality control will not be addressed here and can be, for example consulted, here: [101]) must be aligned back to their reference, that is in most cases a genomic sequence. This process is called *mapping* and several tools have been developed in the past years, to accomplish this task in a reasonable time [102–104]. Despite the pairwise alignment problem being already satisfyingly solved [105], there are still several issues that have to be efficiently dealt with, making the mapping process a non-trivial task. As mentioned above, normally we end up with tens of millions of short reads, possibly containing errors due to sequencing mistakes or mutations in respect to their origin, that have to be mapped back to their reference [106]. Additionally, individual reads can be mapped equally good to multiple locations, *i.e.*, they have optimal alignment scores for more than one position in the reference sequence. Or because of the exon-intron structure of higher eukaryotes' genes, reads have to be split in two or more parts to be able to be mapped correctly. How these problems are solved algorithmically by the recent mapping tools will not be discussed here, I just wanted to point out that the mapping step itself is not easy but crucial for following data investigations [107].

Nevertheless, the abundance of read-to-reference alignments to a specific location is in direct proportion to the measured transcriptional activity of the respective genomic region, which often is annotated to a *sequence feature*. Such features (or

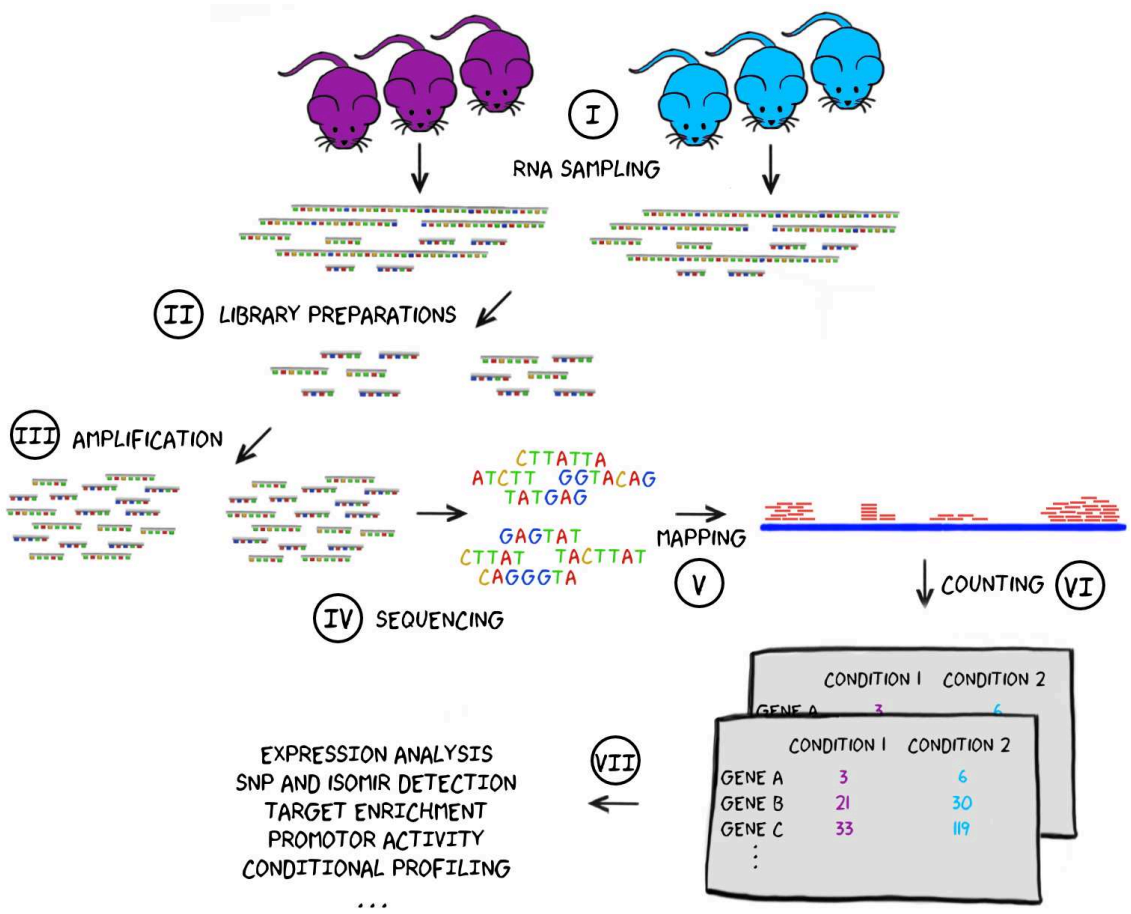


Figure 1.3: **A simplified overview of a typical RNA-Seq experiment and subsequent common analysis steps.**
(I) First, sample material is collected and cells are disrupted to separate the released RNA molecules from all other cell compartments. (II) Various preparation steps can be performed, such as rRNA depletion or size selection, to isolate specific RNA libraries of interest. (III) During the amplification step, all isolated RNA molecules are cut into smaller fragments, are reverse transcribed to cDNA and are duplicated to fit the requirements of the used NGS protocol. (IV) The resulting cDNA fragments are sequenced by an NGS machine, resulting in one or several *fastq* read files. (V) Sequenced reads can be mapped back to their origin by specialized pairwise alignment tools (VI) and after counting, certain genomic features of interest (such as genes) can be assigned an expression count representing its activity. (VII) Based on the obtained mappings and counting tables, further analyses can be performed to extract new biological insights from the sampled organisms under their respective conditions.

genomic features) can refer to genes, transcripts or exons and quantifying the amount of reads mapping to such features is called *read counting*. Again, this is a non-trivial task since many choices have to be made, *e.g.*, minimal overlap of a given read with the feature is required or how to deal with non-uniquely mapped reads, that all can heavily influence the assigned number of reads to a certain feature [108]. What is

obtained at the end are lists of features and their corresponding read number, which can be interpreted as some kind of "activity score", the higher the number the more actively transcribed was the respective genomic feature. Based on these counting lists and on the mapped reads, different analyses can be made to conclude biological insights from the given RNA-Seq libraries, such as differential gene expression analysis, identification of novel genes or transcript isoforms, enrichment analysis, mutation and variation analyses and more. Most of these investigations will be explained and discussed in the following chapters, in particular in Chapter 3, 4 and 6.

1.2.3 *OMICS* and aging

The suffix "-omics" (from the greek neologism -*ομικη*) refers in modern biology to quantitative studies that aim to analyze the totality of certain molecules or biological meaningful features within a sample at a defined time point [109]. As described above in the last section, transcriptomics deals with all transcribed RNA molecules that were current in a given sample during its sampled time point. Consequently, the measurement and analysis of all present proteins (for example measured by mass spectrometry technologies) within a cell or tissue is named *proteomics* or when investigating the entirety of all known genes of a certain species one performs a *genomic* study. These kind of *omics* studies became more and more popular recently, because the according high throughput technologies were just established in the last decade. Besides the three already mentioned *omics* there exist several more, but only two of them, *epigenomics* (the study of all epigenetic modifications) and *interactomics* (the study of all physical interactions between proteins), were relevant for aging research so far.

A number of transcriptomic, proteomic or combined studies revealed many of the possible causes of cellular aging as discussed in Section 1.1.1, and moved the whole field a lot further in recent years (for an excellent collection of these studies, see [110]). This is mostly because aging is a multi-factorial phenomenon that cannot be pinpoint down to single genes, events or pathways, but needs to be viewed from a much broader perspective. With the emergence of the above mentioned high throughput technologies and *omics* studies this was possible for the first time. Still, many open questions have to be answered to understand some of the subtleties but also principals of aging in order to precisely interfere in this process, enabling a longer and healthier life. For that we need more holistic investigations, performed by new and possibly also more advanced *omics* studies. Therefore, biotechnological and algorithmic solutions and advancements remain the key if we want to unravel the big mystery called aging.

1.3 The obstacles of current aging research

As already stressed a lot in the last few sections, biological aging is a highly complex process and a promising way to understand it, is the combination of different modern high throughput data and their analysis. However, before we can accomplish the

task of answering how and why aging happens, we have to be sure to not only ask the right questions but also if we are capable of answering them and perceive those answers with our current technological advancement. There are currently some biotechnological and statistical limitations that one has to face when working in the field of aging and which will be discussed briefly within this section. This is not meant to keep the spirits down or to understate any study on aging (including this thesis), but just to set the presented data and results into the correct frame.

1.3.1 What is "normal" aging?

The title of this subsection already indicates one of the major problems of aging research. Many studies in this area aim to uncover specific parts of molecular processes based on healthy cells or wild type animals to infer information about normal aging, *i.e.*, not influenced by genetic predispositions or other diseases. However, as we have learned from Section 1.1.2 most evidence from research of the last decades currently leads to the conclusion that aging resembles a stochastic process. Organisms, seen as biological living system, accumulate different kinds of damage over time and at some point by chance one of the basic molecular processes, such as DNA integrity or mitochondrial homeostasis, reaches a state of dis-functionality where it starts destabilizing the whole system. As a consequence, other processes are negatively influenced, adding to the already existing damage, subsequently making them also dysfunctional and leading to a kind of self-enhancing damaging process that leads to the total collapse of the biological system. This can also be seen when we recap Figure 1.2 where the complexity of aging processes and their interplays are depicted. Having an imbalance in just one of the shown processes can have substantial influence on the balance of the whole interacting network of different molecular processes with fatal consequences. All this implies, that aging can happen differently from individual to individual, first on the molecular level and later also phenotypically. This can easily be seen on the example of the elderly in our society, were some of them are healthy and fit being in their late 80's, but others already begin to suffer from age-associated afflictions, such as cancer, neurodegenerative diseases or simply physical frailties, starting at the age of 60. There is no indication so far that some kind of common "roadmap" of aging exists, making it more or less impossible to have something being the "normal" aging process. So what actually is investigated in all the different studies on aging are distinct manifestations of aging. And this should be kept in mind, that presumably there is not *the* aging process, making study results difficult to compare. Most likely, there exists not one general solution to the complications that come with aging, instead it has to be dealt with at various frontiers.

1.3.2 The many problems with data on aging

Until now it is still not trivial to find a comprehensively fitting definition for the phenomenon of biological aging (as discussed in Section 1.1) and one reason for that is, that apparently no generic course of aging on the molecular level exists.

Therefore, investigations on aging are often limited to single aspects of this process. This includes specific age-associated mechanisms, particular tissues or species. The problem herein lies in the diversity of aging peculiarities, that seem to be depending (at least in parts) on the aforementioned factors. Brain tissue ages differently compared with liver or skin tissue, since they are exposed to different sources of stress during the lifetime and show highly different regeneration capacities. However, this also depends strongly on the examined species. In most known mammals, the brain shows just little regenerative potential and basically stops growing after maturation, whereas for example some fish, such as *Danio rerio* (zebrafish), still have the ability to regenerate even severe brain injuries during old age [111]. This of course has significant influence on the aging process. Another relatively well-known example are naked mole-rats that can live up to 32 years, thus making them extremely long-lived compared with other closely related rodents of the same size [112]. This is mostlikely due to their low respiratory/metabolic rates, reducing the harmful impact of certain age-associated processes [113]. Accordingly, it is not easy to infer findings from one species to another or from one tissue to another.
But even if we stay within one specific species and tissue, we still have to deal with the versatility of aging processes. Since the progress of aging is of the stochastic persuasion, it is almost impossible to know if individuals, that are used as samples for one certain age group in a classic healthy age group comparison experiment, are good replicates or not (*i.e.*, if aging has effected or damaged the same molecular functions to a comparable extend). If this is not the case, the individual aging processes of every sample appears only as background noise in the respective data and is most likely overshadowed by other processes (for an illustrated example, see Figure 1.4). This problem will be shown and discussed in the subsequent main chapters of this thesis.

Another problem arises when different vital organs are investigated in terms of aging. There are some tissues that are easily accessible and harmless to sample, such as blood or skin. And we can already learn much about aging processes by investigating changes in them at various time points, but only to a certain extend. As it will be shown and discussed later (in particular in Chapter 3), all the different organs show their own pattern of transcriptional shifts over time. Thus, to examine how the distinct and functionally specialized tissues contribute to the overall aging process within one organism, they have to be analyzed individually. To do so, it is inevitable to sacrifice the animal of interest for some organs (*e.g.*, heart, brain, liver, spinal cord, ...). However, this so-called cross-sectional study design brings additional problems. For example, because different individuals are sampled to obtain data on the age-depended regulation or function of one specific organ, it is nearly impossible to track the actual course of molecular changes over time. The reason being again that the accumulation of molecular damage and subsequent dis-functions remain random and can hardly be interpolated between different individuals, because a lot of important information is lost since all animals within the respective study can only be examined once. This leads to an increased stochasticity of the obtained data, making it more difficult to infer viable results (see Figure 1.5).

One more point to consider, especially for longevity studies, is that we do not know how old any individual animal sacrificed at a certain age, would have become. There

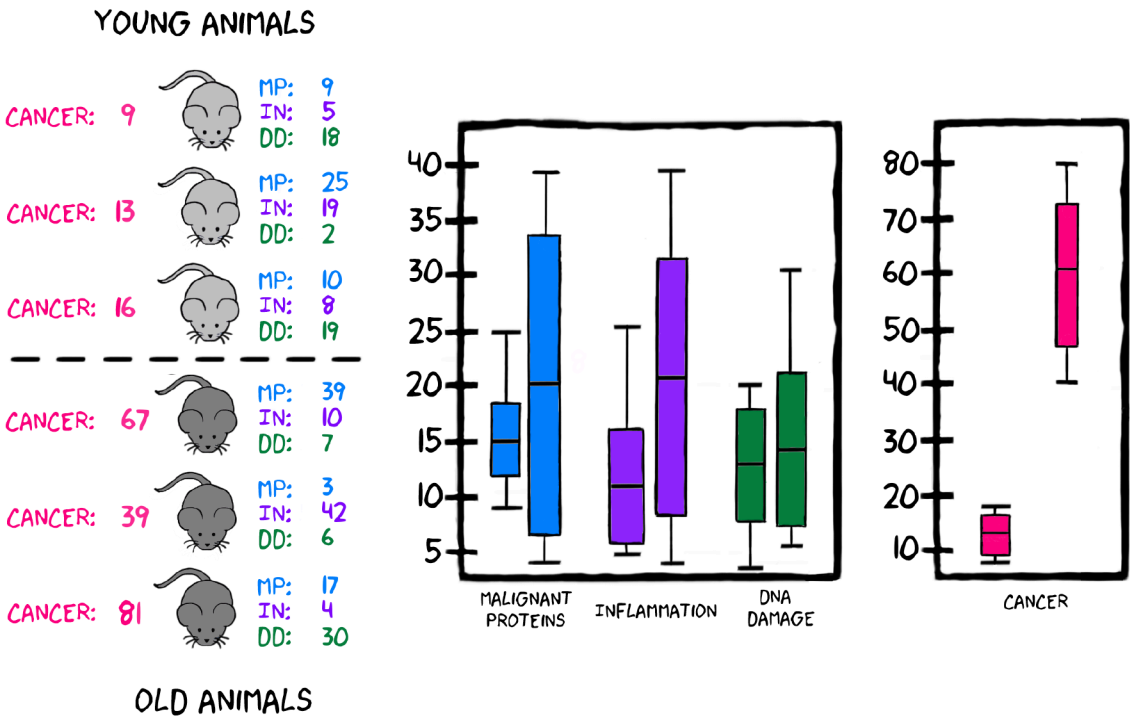


Figure 1.4: **Possible problems with cross-sectional aging data (Part I).** In this simplified example a young and an old mouse group are compared at a certain time point, regarding the activation of three different aging processes and their cancer activity: malignant proteins (MP), inflammation (IN), DNA damage (DD). The higher the measured score, the higher the activity of the respective process. Since aging is a random process, a huge variance is present in the obtained data. This is especially true for the older animals, because just by chance some accumulate more DNA damage through their life time, whereas others have to deal with malignant proteins or hyper-inflammation instead. Therefore, getting a consistent picture of overtime changes within these aging processes is difficult, making it hard to identify this processes to be changed statistically significant. An important factor is also the extremely small sample size, which typically for current expression comparisons, seldomly exceeds more than five samples per group. On the other side, more prominent changes, such as the activity of cancer cells between young an old animals are easily detectable. However, cancer is just a consequence of aging, not a cause of it, overshadowing the already hard to observe underlying process changes.

is some evidence that the life-span of an organism could be modulated by the activity of some genes (namely *Sirtuin genes*). Yet, individuals that express this specific longevity pattern could be sacrificed at an early age and being mixed into one comparative age group with those who did not have these life extending genetic programs in action. If they are now compared to animals that actually have reached the very old ages of their species, the underlying cause can hardly be identified.
Speaking of life-spans, another problem remains: We study aging on animals to achieve an understanding that hopefully provides us the possibilities to healthily extend human life as much as possible. Directly working on human subjects would be for one reason too long-running and for another of course ethically immoral.

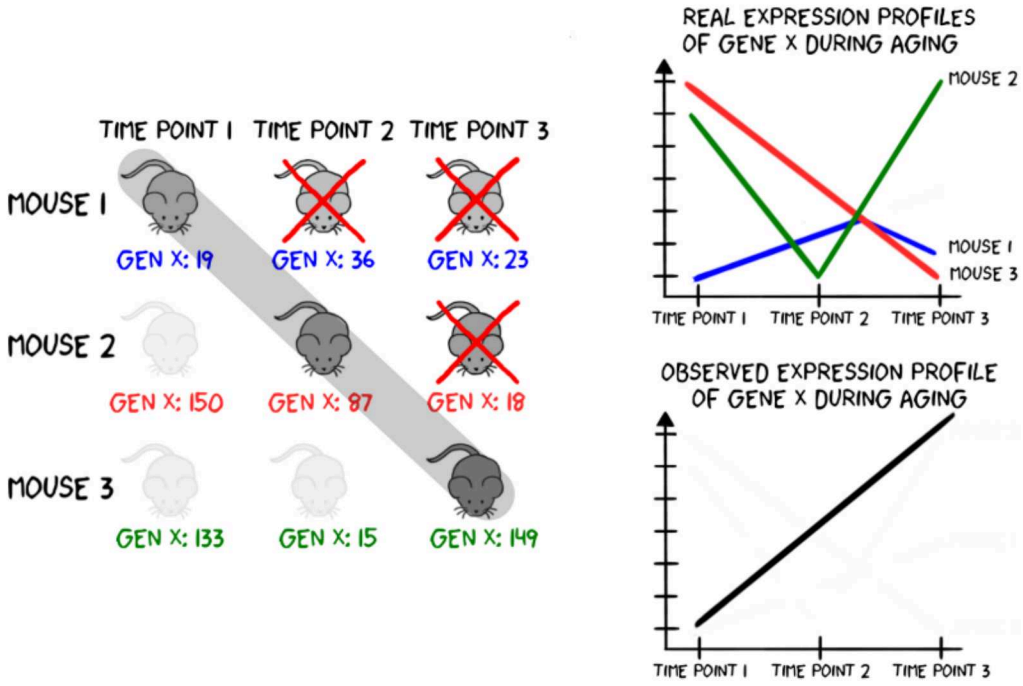


Figure 1.5: **Possible problems with cross-sectional aging data (Part II).** At the moment, when examining age-dependent changes of vital organs in animals like the brain, liver or heart, it is unavoidable to sacrifice the individuals. As displayed in this simplified cross-sectional experiment example, the expression of an arbitrary gene X is measured at three different time points, to estimate its activity changes with age. Three different individuals (or groups of individuals) have to be used to obtain an age expression profile of gene X (the black bottom right function), that appears to indicate some positive relationship of the activation of gene X with aging. However, because none of the animals could be measured at more than one time point, due to being too young or already dead, the true longitudinal expression profiles of gene X were never observed individually (the top right blue, green and red functions) and would have shown that there is no clear correlation of the expression of gene X with aging.

Therefore, working on animals like rodents or fish, that normally life between two and five years is more practical, with the disadvantage that some findings in one species may not be applicable to other species, even to closely related ones (like the example of the naked mole-rat, mentioned earlier). To minimize these risks, we would have to study aging in our closest relatives: primates such as chimpanzee, gorilla or orangutan. They all have wild life expectancies between $30-40$ years and can even reach older ages of 50 or more years in captivity [114, 115]. This fact alone makes them rather unsuitable study objects, let alone the complications that come with strict animal rights and ethics.

To conclude this section, at the momentary scientific frontier of aging research, we are faced with a plethora of limitations when it comes to obtaining and evaluating meaningful data. Some problems could be solved in the future with advanced

biotechnological tools (*e.g.*, less harmful invasive procedures to sample vital organs), whereas others will probably never be solved (unpredictable course of external influences on molecular processes). Nevertheless, despite being a challenging task it is not impossible to gain new and valuable insights from currently available data. And some of them are being presented in the following chapters.

1.4 Contribution and scope of this thesis

Since this thesis is based on many different age-related projects, the topic of aging will be approached from different directions. At first in Chapter 3, we address the topic of tissue-specific changes in the expression of protein coding genes and why it is not easy to identify so-called general biomarkers of aging even within one species. In Chapter 4 we will learn about the regulatory functions of microRNAs, how they can be identified and annotated, which role they have on aging, especially in the short-living killifish *Nothobranchius furzeri* and how we can statistically model the expression of microRNA genes. Within the last three chapters we will concentrate on very specific biological processes and how they influence or are being influenced by aging. First, the relatively well studied phenomenons of age-related senescences and inflammation will be compared among four different species. After that, we will have a look if the gene splicing machinery contributes significantly to the cellular stresses of aging cells or just plays a minor role. And concluding, the course of activity of genes responsible for the maintenance of an individual's circadian rhythm are analyzed and possible implications are discussed.
As it is common in scientific work, the scientific "we" is used through out this thesis instead of the personal pronoun "I", because most of the work presented would not have been possible without the cooperation of my colleagues and collaborative partners.

Chapter 2

Data, Material and Methods

Within this chapter, all data and respective general analyses on which this thesis is based are described. Most of the data presented and analyzed originates from the *JenAge* project, therefore it is explained separately in more detail in the next section (Section 2.1). This mainly includes the experimental part of organ and tissue collection and their experimental processing. This was done by different people and is acknowledged in the respective given citations or stated explicitly. All data analysis, interpretation and visualization parts were performed by myself, if not stated otherwise. In Section 2.2 some general data (pre)processing steps are described, that were applied to all investigated RNA-Seq data sets. Specific analysis are described in detail in the respective chapters.

2.1 The JenAge Data

The Jena Center for Systems Biology of Aging (*JenAge*) is a multidisciplinary research center located in Jena, Germany. It was established in 2009 with a successful grant application at the German Federal Ministry of Education and Research and consisted of ten research groups from four different institutions (`http://www.jenage.de/`). Embedded in this project is the *JenAge Information Center*, a collection of various databases holding aging related biological, demographic and disease specific data as well as corresponding metadata. A major part of the stored biological data are RNA-Seq data of the five species *Homo sapiens*, *Mus musculus*, *Danio rerio*, *Nothobranchius furzeri* and *Caenorhabditis elegans*. For a brief introduction to RNA-Seq data analysis, see Chapter 1.2.2. A number of transcriptome samples was collected and sequenced from different tissues at different ages from the above mentioned species of which a huge part was analyzed within this thesis. The only exception is the *Caenorhabditis elegans* data set, which was not used in any of the further presented chapters. For a detailed overview on the investigated data, see Figure 2.1 and Figure 2.2.

Tissue and organ extractions from all animals, and RNA extraction as well as further experimental processing were realized in frame of the JenAge consortium as published by Irizar et al. [116]. Animal housing and experimental usage was performed in accordance with the ARRIVE (Animal Research: Reporting of In Vivo

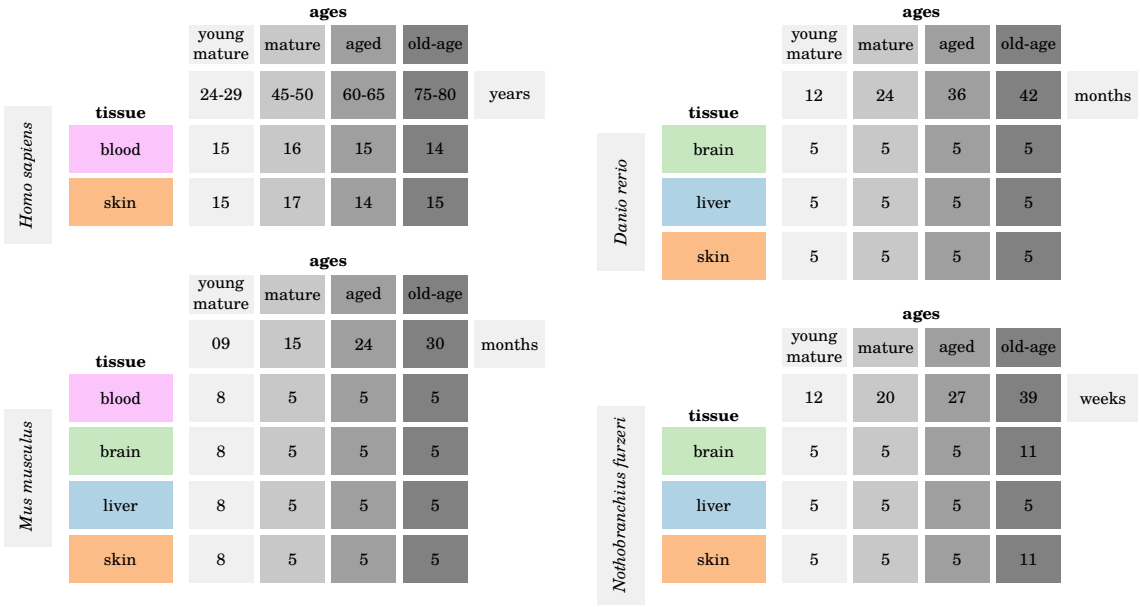


Figure 2.1: **Overview of the analyzed JenAge RNA-Seq data.** Amount of sequenced RNA-Seq libraries for each sampled age, tissue and species.

Experiments) guidelines and approved by the regional animal welfare authorities in the State of Thuringia (Thüringer Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Bad Langensalza, Germany). Inclusion of the four species implied highly different environments, life conditions and basal requirements. Likewise, activity patterns and nutrition periods are greatly different between nocturnal and diurnal species such as mouse and human mammals, and hardly comparable to fish. Moreover, the natural habitat of short-lived Nothobranchius furzeri is characterized by extreme climatic conditions (see 3.2), and daytime instead of seasonal climate.

Tissue sampling

Human tissues

Human tissues comprising blood and skin were derived from identical or different male and female individuals and collected at the Department of Neurology at Jena University Hospital. Donors were separated into the following age categories: 24 – 29 years, 60 – 65 years, and 75 – 79 years (n = 14 – 15 for each). Blood samples were stabilized in PAXgene Blood RNA tubes. A skin area of 3 mm^2 was taken from the ventro-medial, distal part of the upper leg. Human tissue sampling was authorized by the local ethics review committee and implied informed consent by each of the study participants. Donors with strongly pigmented skin, skin diseases or a history of interfering pathologies were excluded.

Murine tissues

Murine tissues included blood, skin, brain and liver from male wild type C56BL/6 mice housed in the animal facility of the Jena University Hospital under controlled

Figure 2.2: **Overview of the analyzed JenAge small RNA-Seq data.** Amount of sequenced small RNA-Seq libraries for each sampled age, tissue and species.

conditions for temperature, humidity and light/dark cycle (14:10 hours, including a daily break of dawn and dusk imitation period), with food and water available ad libitum. As age categories, corresponding to human sampling time points, 9 months, 15 months, 24 months and 30 months were selected (Figure 2.3). Within the same age category, blood and skin (abdominal) samples were collected from the same individuals, whereas liver and brain were extracted from another set of animals (blood, skin, liver, n = 5 for each; brain, n = 8). For tissue collections, mice were deeply anesthetized with an overdose of volatile isoflurane/cervical dislocation, the organs were immediately extracted and frozen in liquid nitrogen until use for RNA-Seq.

Fish tissues

Animal maintenance was performed as described in [117, 118]. All samples from one fish strain were taken at the same daytime (*Nothobranchius furzeri*: 10 a.m.; *Danio rerio*: noon/early afternoon) in a fasted state. For the tissue preparation, the male fish were euthanized with Tricaine mesylate (MS-222) and cooled on crushed ice. Tissues from both, *Danio rerio* (strain TüAB) and *Nothobranchius furzeri* (strain MZM-04/10) included brain, liver and skin (n = 5 for each) and were isolated from 12, 24, 36 and 42 months and 12, 20, 27 and 39 weeks old fish, respectively. Maintenance and experimental use of *Danio rerio* and Nothobranchius furzeri were approved by the local animal welfare authorities in the State of Thuringia (Thüringer Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Bad Langensalza, Germany).

RNA extraction

RNA extraction was performed as described in the following publications: *Homo sapiens* [116], *Mus musculus* [119], *Danio rerio* [120] and *Nothobranchius furzeri* [120, 121].

2.2 General RNA-Seq data processing

All RNA-Seq libraries presented within this thesis are accessible at NCBI's Gene Expression Omnibus (*Homo sapiens*: GSE75337, GSE103232; *Mus musculus*: GSE75192, GSE78130; *Danio rerio*: GSE74244 and *Nothobranchius furzeri*: GSE52462, GSE66712).

Genomes and annotation

The genomes and annotations of *Homo sapiens*, *Mus musculus* and *Danio rerio* used in this thesis were downloaded from `Ensembl` (release version 92) [122]. For *Nothobranchius furzeri*, the currently published assembly version of its genome and the respective annotation was used [121, 123]. Additionally, for the analyses in Chapter 4 the following fish genomes and annotations were downloaded from `Ensembl`: *Oryzias latipes* (HdrR), *Takifugu rubripes* (FUGU5) and *Gasterosteus aculeatus* (BROAD S1).

Total RNA-Seq library processing and mapping

Total RNA-Seq libraries were filtered and quality trimmed with the use of `PRINSEQ` (v0.20.3) [124] by clipping all reads at both sides to achieve a minimum base quality of at least 20 and discarding all reads with a length of less than 15 nt or more than two ambiguous N bases. To monitor the read qualities, we used `FastQC` (v0.11.3; `http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/`). Mapping the quality trimmed RNA-Seq libraries onto the respective current genomes was performed with `TopHat2` (v2.1.1) [125] using the default parameters, allowing spliced reads and mapping of single reads to multiple best fitting locations. Read counting was performed using `featurecounts` (v1.5.3) [108] and reads were normalized by transcripts per million (TPM) [126]:

$$TPM_i = \frac{c_i}{l_i} \cdot \left(\frac{1}{\sum\limits_{j \in N} \frac{c_j}{l_j}} \right) \cdot 10^6$$

where c_i is the raw read count of gene i, l_i is the cumulative exon length of gene i and N is the number of all genes in the given annotation. Genes with an TPM value ≤ 1 in every sample were considered to be not expressed and discarded in all subsequent expression analysis.

Small RNA-Seq library processing and mapping

The RA3 adapter of the TruSeq small RNA preparation kit (5'-TGGAATTCTCGGGTGCCAAGG) was cut from the reads of all small RNA-

Seq libraries. Additionally, `PRINSEQ` (v0.20.3) [124] was used to trim the reads from both sides in order that the read bases had a minimum quality of 20 and minimum lenght of at least 15 bases. The mapping onto the respective genomes was performed with `segemehl` (v0.2.0) [103] using the `-H 1` option, allowing single reads to be mapped to multiple best fitting locations. The visualization of mapped reads was done using `IGV` (v2.0.34) [127].

Identification of differentially expressed genes (DEGs)

The `DESeq2` (v1.10.0) [128] Bioconductor package was used to identify differentially expressed genes (DEGs) between the different ages for of each species and tissue, respectively. The exact age group comparisons are described within the respective chapters individually. Multiple testing adjustment of the resulting p-Values was performed using the Benjamini and Hochberg's FDR approach [129]. Genes with an identified adjusted p-value < 0.05 were considered as differentially expressed. All DEG results, together with the raw and normalized count values can be found in detail at the supplemental data descriptions of the respective chapters.

2.3 Life span comparison

The four species investigated display highly different lifespans. To facilitate comparability, we rendered a normalized lifespan scale to each of them, which was aligned to specific biological stages, *i.e.*, the age of mean maturity, mean survival and 10 % survival age as well as the age of the oldest known individual (see Figure 2.3). Information regarding the different life stages of the four species underlied different sources: for *Homo sapiens* from the *World Health Statistics 2016* of the WHO [130] and [131, 132], for *Mus musculus* from [133–136], for *Danio rerio* from [137–140] and for *Nothobranchius furzeri* from [118, 123, 141, 142].

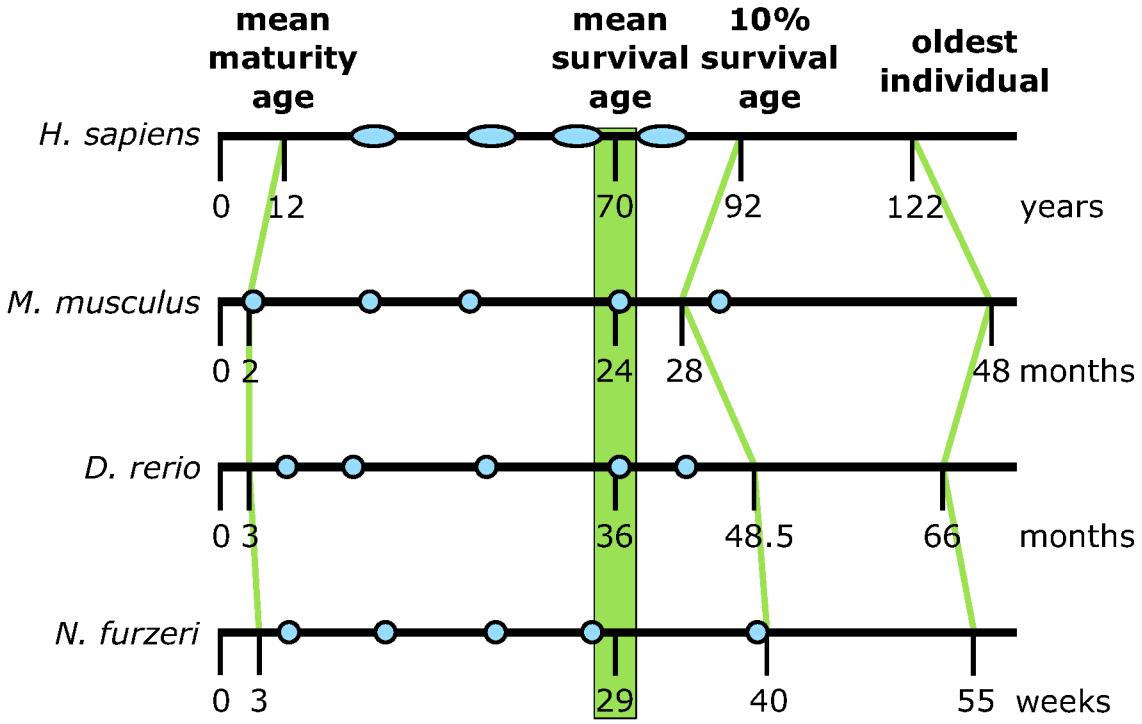


Figure 2.3: **Inter-species normalization of age categories.**
To align the species-specific chronological ages to biological age categories comparable between different species, the total individual's life time, represented by the length of the life time axis, was subdivided by index stages. These were the biological stages of maturation, mean survival age, 10 % survival rate and highest ultimate age reported for an individual belonging to the respective species (see Section 2.3). Time intervals between the resulting intersections were normalized with respect to the mean survival age in a linear fashion. Thus, the sampling time points realized in the study for a certain species (blue dots) matches with the biological age category in all the other species analyzed.

Chapter 3

Tissue-specific Aging Patterns in Mice

This chapter is based on the publications "Tissue-specific gene expression changes are associated with aging in mice" and "Inching towards decrypting the enigma of Aging".[1]

Data used and analyses performed in this chapter

Age comparison setup

The age comparisons were performed according to Figure 3.1. Half of the comparisons can be seen as a linear age-gradient analysis (9 vs. 15 months, 15 vs 24 months and 24 vs 30 months), providing insight into the changes over time in gene expression. The other three comparisons (9 vs. 24 months, 9 vs. 30 months, 15 vs. 30 months) show the more drastic transcriptional changes between adult and old-aged mice. All DEG results can be found in detail at STable 2.

Age profiling of tissue gene expression

Each expressed gene with TPM > 1 was assigned an age profile with respect to its expression behavior over time for each of the four tissues individually (see Figure 3.2). The profiles were determined by analyzing the read fold change between the linear age progression comparisons (9 vs. 15, 15 vs. 24 and 24 vs. 30 months). For each of the four comparisons, every gene was categorized either up/down-regulated (increase/decrease of at least 25 % in fold change) or equally expressed. For the brain, the fold change thresholds were set to 10 %, because the total neuron expression activity was reduced compared to other tissues, because they do not divide further after birth and so the neuronal cell population remains more or less stable, thus expression is more consistent, and hence, a fold change of 10 % can be assumed to have significant effects. Genes were clustered regarding their gene profiles and clusters were analyzed for pathway and GO term enrichment. For details about the gene expression age profiling see STable 3 and SData 1.

[1] The complete supplemental material is available at `https://osf.io/tvrdm/`

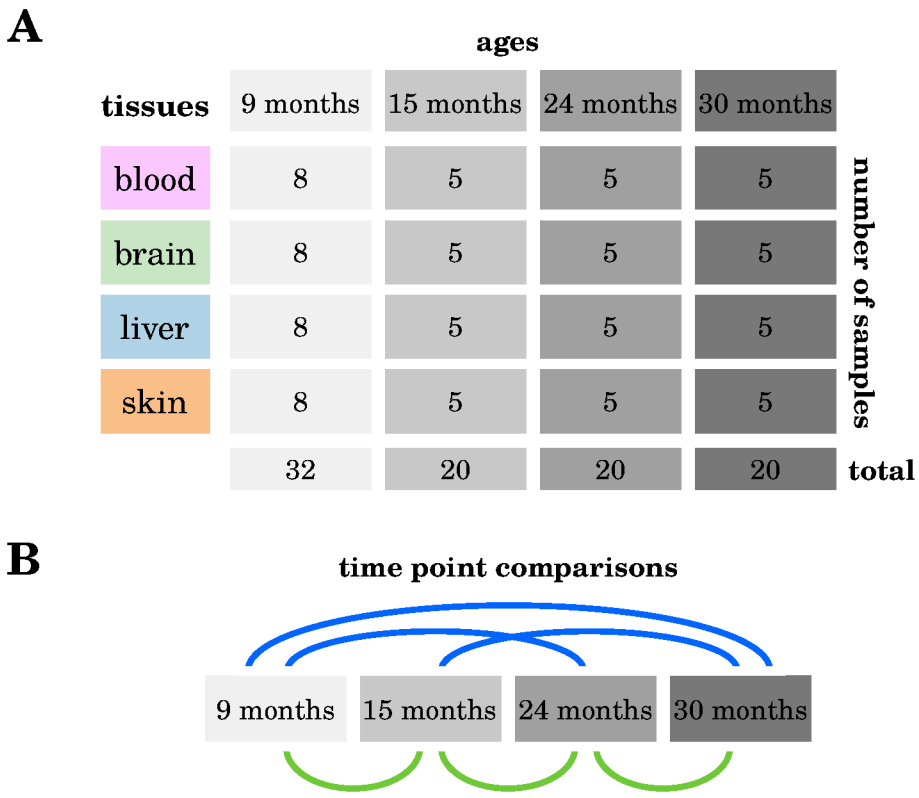


Figure 3.1: **RNA-Seq library overview and age comparisons.**
(A) The number of used samples and RNA-Seq libraries per tissue at every age point. (B) We compared the four age groups for each tissue individually (24 comparisons). Green - short-term comparison; blue - long-term comparison.

Pathway category enrichment

Differentially expressed genes from each time comparison (Figure 3.4) were associated with the KEGG [143] pathways they are involved in, using the functional annotation analyses of DAVID (version 6.8) [144] for each tissue, using *Mus musculus* as a reference and with KEGG pathways being the only option checked-in. All pathways were grouped on the basis of their physiological role and functionality into different categories. Categories that contained at least 5 % of the DEGs of the respective comparison were: metabolism, inflammation/immune response, synapse-related pathways, muscle-related pathways, extra cellular matrix/adhesion/cytoskeleton, signaling pathways and cancer. All remaining categories were grouped into others (Figure 3.5). The enrichment was performed based on up- and down-regulated genes separately, as well as the whole set of DEGs (see SData 3A/B). Additionally, genes that followed a specific expression pattern with aging were separately used for pathway enrichment analysis (see STable 4 and SData 4). Pathways with an FDR adjusted p-value lower than 0.05 were considered significantly regulated and were again further classified into functional groups for each tissue individually.

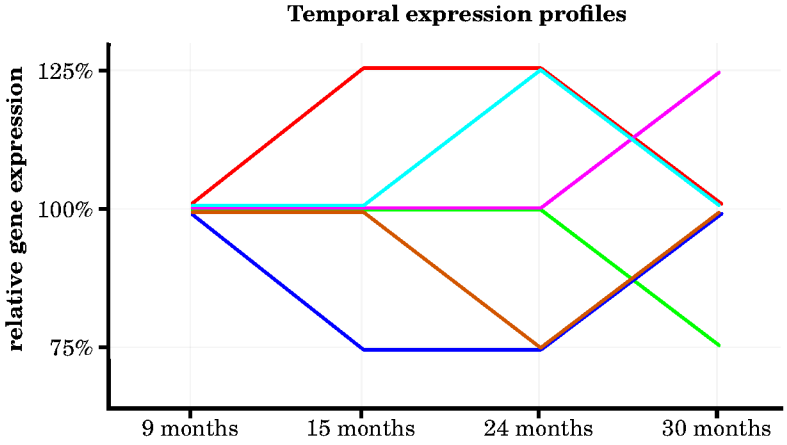


Figure 3.2: **Gene expression profiles of interest.**
All genes that followed the course of one of the depicted expression changes during aging were grouped into separate clusters and analyzed subsequently for an enrichment of functional commonalities.

Validation of expression for common DEGs using qPCR

This experimental part was performed by Madlen Günther.[2]

We validated the expression level of DEGs determined by RNA-Seq by performing qPCR for transcripts using the same methodology described earlier by Sieber *et al.* [145]. We used the same amount of RNA in qPCR as was used for RNA-Seq to reverse-transcribe the cDNA using a RevertAid First Strand cDNA Synthesis Kit from Fermentas (Thermo Fisher Scientific, Waltham, MA, USA). Primers were designed and were diluted to 500 nM for mice. Brilliant III SYBR Green QPCR Mastermix from Agilent Technologies was used to perform the reaction using Qiagen's Rotor-Gene 6000 cycler with the following cycle conditions: 3 minutes of polymerase activation followed by 40 amplification cycles of 95 °C for 10 seconds and 60 °C for 15 seconds. For housekeeping genes, in order to increase the accuracy of our results, we used Hprt1 for every tissue along with Gapdh, Hmbs, Gusb and Actb for brain, blood, skin and liver tissue, respectively. Further, the Ct values were used to calculate the transcript ratios with respect to 9 months age using the Pfaffl equation [146]. A list of the used primers for qPCR can be found in SData 1.

RNA interference treatments

This experimental part was performed by Maria Ermolaeva.[3]

HT115 bacteria containing specific RNAi constructs were grown on lysogeny broth agar plates supplemented with ampicillin and tetracycline. Plates were kept at

[2]Hans-Berger Department of Neurology, University Hospital Jena, Friedrich Schiller University Jena, 07747 Jena, Germany
[3]FLI Leibniz Institute for Age Research, Beutenbergstrasse 11, 07745 Jena, Germany

4 °C. Overnight cultures were grown in lysogeny broth media containing ampicillin. RNAi expression was induced by adding 1 mM isopropylthiogalactoside (IPTG) and incubating the cultures at 37 °C for 20 min before seeding bacteria on NGM agar supplemented with ampicillin and 3 mM IPTG.

Life span analysis

This experimental part was performed by Maria Ermolaeva.[4]

The experiments were carried out at 20 °C under standard conditions. Synchronized L4 larvae were placed on 60 mm dishes containing NGM agar and RNAi bacteria at a density of 70 worms per plate. Worms were transferred to new plates on a daily basis until adulthood at day 6 (AD6) and later transferred to new plates every 3 – 4 days. The number of dead animals was scored daily. The whole experiment was repeated three times. The analysis of the lifespan data, including statistical analysis, was performed using GraphPad Prism software. For more details, see SData 6.

[4]FLI Leibniz Institute for Age Research, Beutenbergstrasse 11, 07745 Jena, Germany

3.1 Aging treats tissues not equally

We already learned from Chapter 1 that aging is a highly complex process characterized by progressive physiological changes. And this process happens throughout all organs and tissues during an individuals life time. However, it has become increasingly clear that aging has a divergent effect on different tissues at both the gene expression and the physiological levels, due to the influence of a variety of intrinsic and extrinsic agents [147, 148]. For example, the role of the liver during aging is intensively discussed and it is suggested that replicative senescence, exposure to toxins such as drugs and free radicals, and diet play important roles in the process of liver aging [149, 150]. In 2015, Lans *et al.* described in *Caenorhabditis elegans* how damage responses are altered uniquely for different tissues because of distinct DNA repair systems [151]. Additionally, several studies have linked the aging process with an increase in inflammation (more on that topic in Chapter 6) and suggest that changes in lymphoid organs also play an important role [152]. However, despite decades of research, it remains incompletely understood how exactly tissues change with age.

One important approach to understanding the mechanisms of aging is to examine changes at the gene expression levels at different ages. An increase in inflammatory and stress response genes along with the accumulation of ubiquitinated proteins with age has been reported by many independent studies based on microarray or RNA-Seq experiments. In particular, antigen processing and presentation, NF-κB signaling, and lipid metabolism, as well as complement and coagulation cascades, were shown to be relevant inflammation pathways during aging [147, 153, 154]. In addition to certain processes, some specific genes have also been associated with aging, such as members of the cathepsin family, apolipoproteins, complement system and members of the STAT protein and Tumor Necrosis Factor receptor superfamily [155–157].

Most of these studies focus on only one tissue (frequently the brain) and include a limited number of time points, often including a time point before maturation. Studies comparing gene expression changes during aging in multiple tissues across different age points are scarce. One exception is the extensive microarray-based study by Jonker *et al.* [147] with five different organs and six age points. This study reports distinct aging signatures in different organs.

Within this chapter, we present a comprehensive study of age-related transcriptional changes in mice in four tissues: brain, blood, liver and skin. We generated a unique set of 92 RNA-Seq transcriptome libraries from mice at four different ages: a young but already mature age of 9 months, an intermediate age of 15 months, an old of 24 months, and the very old age of 30 months. Survival studies have shown that 72 % of the investigated C57BL/6 mice live to an age of 24 months, while only 4 % are long-lived and survive to an age of 30 months [158]. It is important to note that mentioned age groups are in relation to C57BL/6 male mice and may differ for other strains and sex.

Our study revealed tissue-specific and tissue-independent genes and processes that are differentially affected in aging. We identified seven genes that are differentially expressed in all four tissues during aging, along with various subunits of the mi-

tochondrial electron transport chain that are regulated in a similar fashion among the analyzed tissues. While the number of differentially expressed genes and their molecular biological processes were comparable between the skin, brain and blood, the liver shows a much higher level of age-linked gene expression differences, making it an interesting candidate for future research.

3.1.1 PCA shows characteristic regulation at the transcriptome level for every tissue during aging

To get a general overview of the 92 RNA-Seq samples from the four tissues investigated, we performed a principle component analysis (PCA) based on the 750 genes that were most variant with age, see Figure 3.3 and SFigure 1. As expected, the sample replicates of the individual tissues (blood, brain, liver and skin) clustered together. Organs are clearly distinguishable, indicating that organ-specific genes are present among the 750 most variant genes.
Being most separated in the first principle component, the liver showed the largest differences from other organs. In the liver and skin, we observed the highest dispersion of samples in the PCA plot, indicating broader gene expression variability at different time points. The blood and especially the brain replicates clustered more closely together. This is in agreement with the fact that the brain is a tissue that undergoes only minor changes once it is fully developed, despite some new insights which show plasticity in some specific areas [159]. However, a clear distinction of the different ages within the single tissues was hardly possible with this approach (SFigure 1), because age-dependent expression changes were subtle, and no larger set of obvious common aging-linked factors was present in these four organs. Only for the liver could one observe measurable separations of the replicates of the young time point at 9 months and the old time point at 24 months from all other samples.

3.1.2 Predicting important modulators of aging for each tissue

In total, we identified 1329, 579, 5185 and 1237 genes that were differentially expressed in the blood, brain, liver and skin, respectively, among the pairwise age comparisons. Specific numbers of differentially expressed genes (DEGs) identified are given in Figure 3.4 and STable 2.

Aging patterns of gene expression in different organs

Within the brain, consecutive age-wise comparisons showed only minor changes in gene expression, indicating that on the gene expression level, normal aging is a slowly progressive process with subtle changes between close time points. A larger number of genes (400) with significant changes were observed between wider time spans, especially those including the very old mice (9 vs. 30 months and 15 vs. 30 months). This shows the progressive nature of the aging process, which appears to accelerate at later time points. We found a very similar pattern of gene expression changes in blood (with a generally higher number of affected genes) and skin. Skin showed more differentially expressed genes at an earlier time point (9 vs. 24 months), which

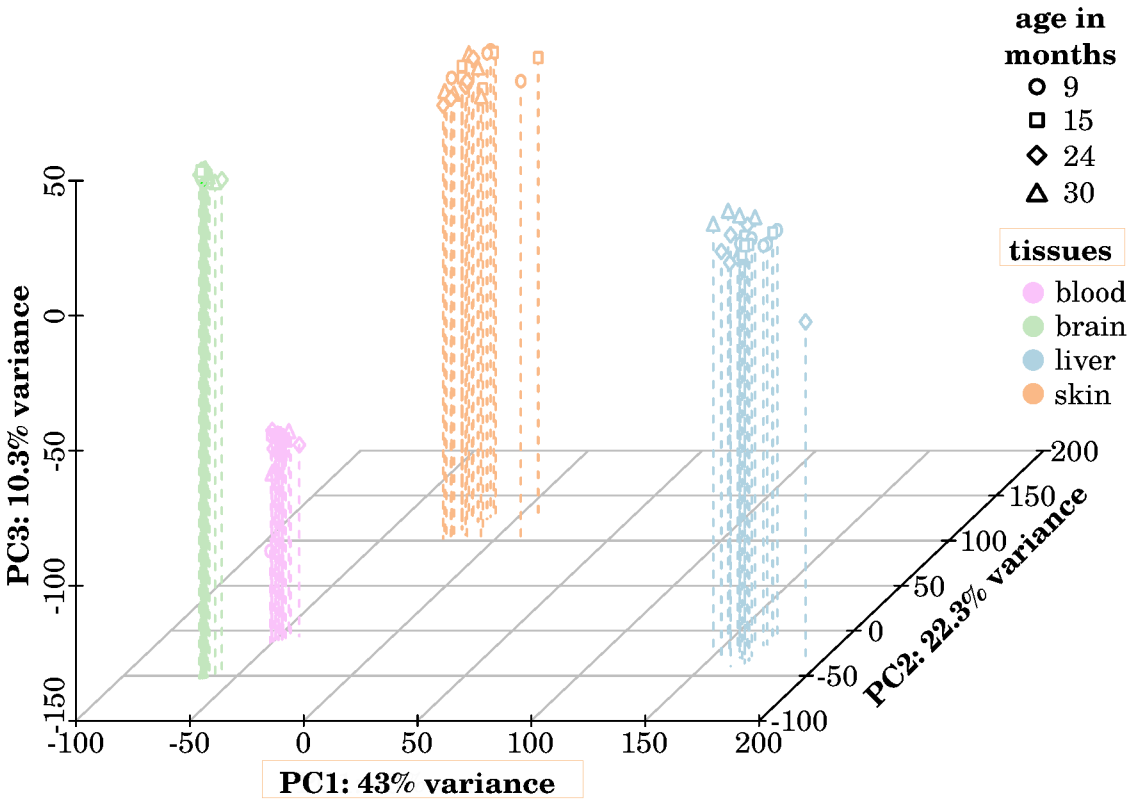
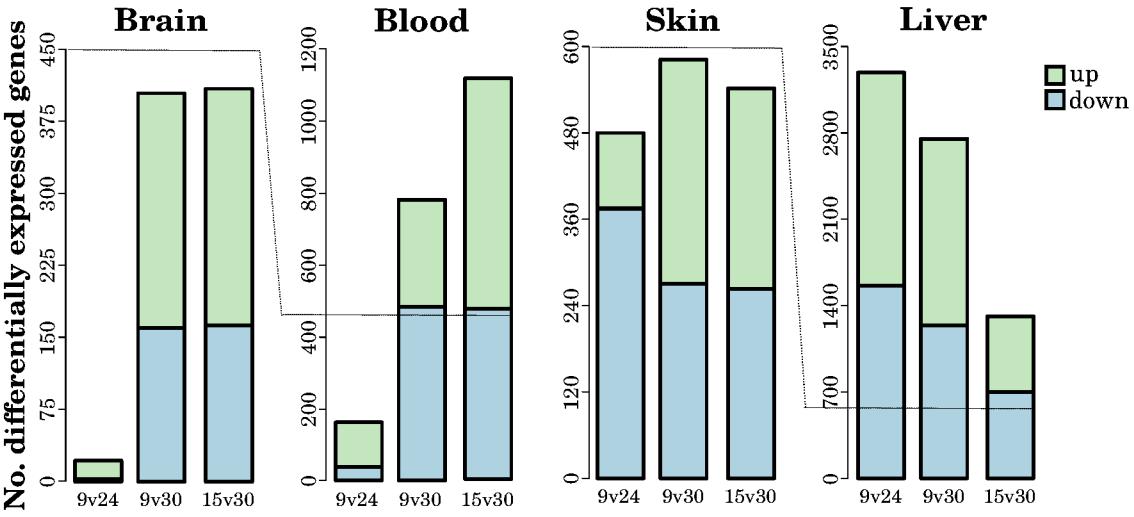


Figure 3.3: **Principal component clustering of RNA-Seq libraries.** The three-dimensional PCA plot of 92 RNA-Seq libraries based on the 750 most variant genes, *i.e.* showing the highest variance in expression between time points in all samples. While all libraries clustered very well together according to their tissues, no clear distinction regarding the different age groups could be made. PCA plots of each tissue and different sets of genes are available at SFigure 1, showing in general always the same clustering.

can be interpreted as a slightly accelerated aging process compared to brain and blood.

The brain shows more signs of inflammaging

In particular, the most significant DEGs in the brain were genes of the complement and immune systems, such as *C4a*, *C4b*, *Tlr2*, *Cst7* and *Ifit3*. This includes brain-specific ones such as *Gfap*, which is known to act during development but is also a marker for activated astrocytes with aging [160]. All these genes have been previously reported to be up-regulated in the brain with age. While proactive immune functions can imply protection from infections, persistent low-grade inflammation can also be deleterious [161]. Thus, inflammaging might be one of the primary causes of aging in this organ. Additionally, we also observed moderate up-regulation of cell cycle regulators *e.g.*, *Cebpa*, which is a common observation in aging neurons [162]. For details, see STable 2.



	Brain			Blood			Skin			Liver		
	9v24	9v30	15v30	9v24	9v30	15v30	9v24	9v30	15v30	9v24	9v30	15v30
up	19	254	258	44	301	621	108	309	279	1667	1493	601
down	2	159	164	17	514	485	374	270	260	1552	1285	695
total	21	413	422	61	815	1106	482	579	539	3219	2778	1296

Figure 3.4: **Differentially expressed genes in mouse tissues during aging.** Number of identified differentially expressed genes (DEGs) within each of the four investigated tissues and between the four different ages. Aging occurs as an accumulation of errors over time, manifesting as a high number of DEGs when comparing 9 vs. 30 months and 15 vs. 30 months in the brain, blood and skin. The liver does not follow this pattern and shows a high rate of DEGs at every time point comparison. Note the different scales of the y-axes of the plots. More details about all time comparisons and identified DEGs can be found in STable 2.

The blood mainly shows markers of senescence due to replicative aging

We observed age-linked down-regulation of the hematopoietic cell growth and survival negative regulator *Rhoh*, which is known to be a strong inhibitor of NF-κB and other general T and B cell migration and differentiation factors such as *Chst3* [163]. In addition, regulators of various cancer pathways appeared to be differentially expressed: *Id3* and *Dusp2*, which hint at a potential cancer suppressing mechanism for very old mice, and *Dpep2*, which is known to play an anti-leukemia role [164, 165]. Thus, dysregulation of cell division and differentiation is highly prominent with aging. In addition, similar to the brain, the immune system was activated with aging in blood cells, and although it was mostly restricted to the up-regulation of immunoglobulins from B cells, we could observe up-regulation of other immune-related genes, such as *Sirpb1-a/b/c*, *Retnlg*, *Mpo* and *Prg2* expressed by macrophages, monocytes and eosinophils, respectively. In comparison with the very late age (30 months), we also found more general inflammation factors, such as *Clec4e*, *Ifitm2*, and *C4b*.

Skin appears to age distinctively due to environmental exposure

A surprisingly large variety of different transcription factors appeared among the topmost differentially expressed genes (DEGs) in the 9 vs. 24 months comparison. We observed down-regulation of many genes associated with the extracellular matrix, such as different members of the collagen protein family, or *Flnc*, which encodes one of three filamins. A reduction of the extracellular matrix is already implicated with aging, due to external factors such as UV radiation and has been examined in a variety of studies [166–168]. Consequently, no clear trend for affected molecular pathways was observed. Only in comparisons to the 30-month-old mice were some common patterns identified. Between 24 months and 30 months of age, we found important skin functions, such as tight junction, focal adhesion and olfactory transduction to be most significantly affected. This underlines the fact that a decline in tissue-specific functionality can be seen with increasing age. When compared to the brain and blood, only a few immune- and inflammation-associated genes are found to be differentially regulated in the skin, indicating that the typical inflammaging signs seem to be less prominent in this tissue.

The liver shows several divergent processes related to aging

In the liver, we identified the highest number of DEGs (5185) related to aging, making it the most dynamic of the investigated tissues. Young adult tissues (9 and 15 months) mainly displayed changes in genes involved in drug metabolism and reactive oxygen species (ROS) management; in particular, these are mainly cytochrome genes. In addition, immune-related genes were highly significantly affected. Notably, inflammatory processes were regulated but showed no clear general activation or deactivation at this age, with half of them being down- (*e.g.*, *C8b* and *Ilr1*) and half being up-regulated (*e.g.*, *Tlr12* and *H2eb1*); see STable 2. With progressing age (9 and 15 months compared to 24 months) more genes related to inflammation were up-regulated, pointing to an increased inflammation process. Additionally, cell cycle and cell adhesion genes such as cyclin D1 and cadherin 1 were very significantly up-regulated in the 15 to 24 months comparison. Mis-regulation of those genes is known to contribute to cancer progression, which is an important cause of death with increasing age [169]. This is consistent with our pathway analysis (see Section 4.1.3), as cancer-promoting pathways increased between 9 and 24 months of age, while genes associated with cancer suppression did not undergo a significant change. When comparing the very old mice with younger mice (9 and 15 months vs. 30 months) we observed an up-regulation of the immune system in the topmost regulated genes, mainly represented by the activation of immunoglobulin genes and interferons. The two most strongly up-regulated genes here were *S100a8* and *S100a9*, which form a heterodimer, belonging to the S100 calcium-binding protein family and are known to play a major role in inflammation and tumorigenesis [170, 171]. When comparing the two oldest time points (24 vs. 30 months), hardly any immune related genes were found among the most significantly altered genes, which is in agreement with the inflammation processes being highly active in old age (24 months) and seeming to not decrease or increase further with ongoing aging (30 months). Instead, the most significant up-regulated DEGs induce cell death (*e.g.*, *Dedd2*) or directly relate to

cancer (*e.g.*, *Brap*) [172]. Interestingly, the autophagy-inducing gene *Atg2a* was up-regulated, although autophagy activity has been described as decreasing across all tissues during aging [173].

The number of DEGs greatly varied between the investigated tissues, being lowest in the brain (579) followed by the skin (1237), blood (1329) and liver (5185). This might be caused not only by aging, but also by the special roles of these organs. The brain is the most protected of all tissues and practically stops growing after maturation, and it also shows only low regenerative capacities in mice. Skin, being the most exposed tissue of an individual, undergoes constant abrasion due to light, friction and other environmental factors. Over time, these external stresses can cause the distinctive aging phenotype of skin. Interestingly, we found changes in the activation of focal adhesion, tight junction and olfactory genes in the old age time point comparison (24 vs. 30 months). This could be explained by loss of function due to permanent exposure to environmental stress. For blood, genes linked to cell division seemed to be the main signature of physiological changes of processes during aging. Blood constantly undergoes proliferation, unlike most tissues, which normally undergo division only when they suffer damage. The liver is metabolically the most active and involves detoxification processes. Thus, the liver is constantly exposed to exogenous factors, undergoing constant regeneration and showing high regenerative potential. However, this extraordinary plasticity likely makes the liver more susceptible to inflammation and cancer development, as indicated by our data on age-linked gene expression changes.

3.1.3 Clustering of KEGG pathways into categories reveals the most prominent processes affected in aging

Analysis of DEGs reveals tissue specific processes with aging

To analyze our data in a broader context, we enriched KEGG pathways using DAVID for each tissue and for specific age comparisons separately (see Figure 3).

Brain

In brain tissue, when comparing 9 and 15 months to 30 months of age, we found 413 and 422 genes to be differentially expressed, respectively. Among these DEGs only 125 and 127, respectively, were found to be annotated and were associated with a known pathway within the KEGG database. For the age comparison of 9 vs. 30 months, half of the affected pathways contributed to inflammation and immune response. Differentially expressed genes (90 %) were higher expressed at 30 months of age. Almost 14 % to signaling pathways, approximately 12.5 % to metabolism, 7 % to cancer and approximately 5.5 % to muscle function. Interestingly, in the 15 vs. 30 months age comparison, we observed an almost 13 % decline in inflammation, while a similar increase was seen in metabolism and signaling pathways. This shows that with progressive aging, there appears to be an increase in chronic inflammation, and hence, a lower number of DEGs and pathways could be

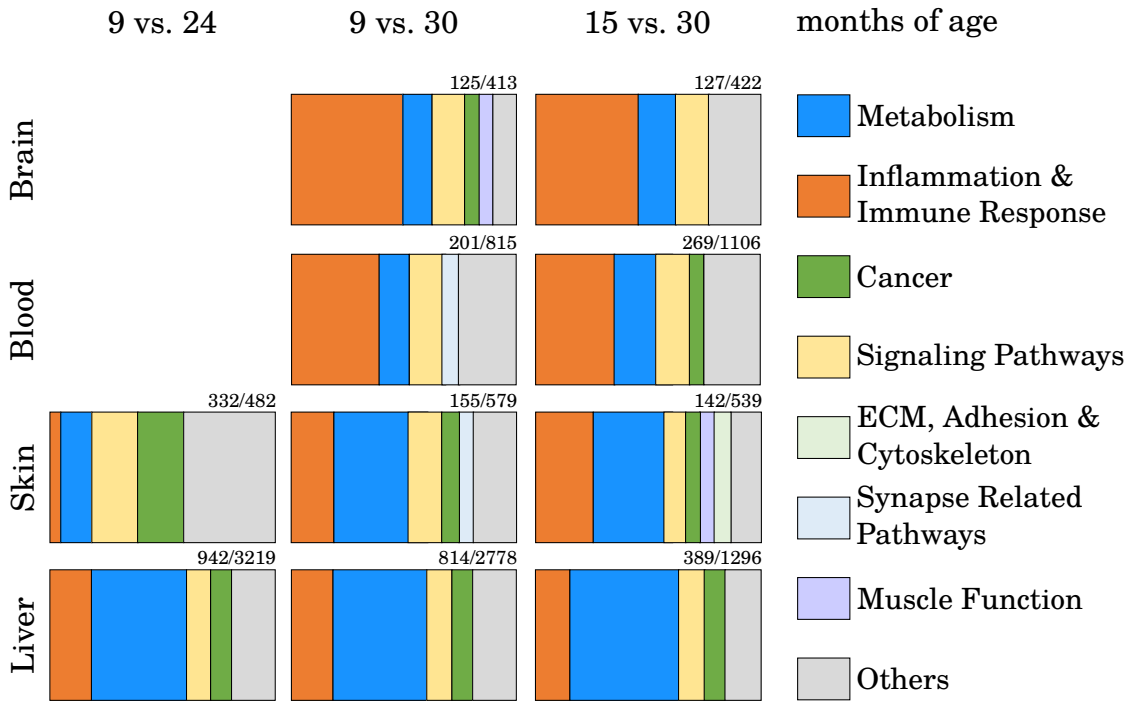


Figure 3.5: **Tissue-specific change of enriched pathways during aging.** Enriched pathway categories based on significantly differentially expressed genes (DEGs) (Figure 3.4) for each tissue and age comparison separately. The number of DEGs for each subplot and the total number of DEGs for this comparison are displayed at the top of each subplot. Any pathway contributing less than 5 % of the genes is categorized as others (grey). The brain and blood show similar patters during aging, while the skin and liver each showed specific aging patterns. Most of the DEGs from the liver were involved in metabolism. Note that not all age comparisons are included here, because they had too few DEGs for a sufficient pathway enrichment analysis. More details can be found in SFigure 3 and SData 3A/B.

annotated to the same category at later time points, while an increase in metabolism- and signaling-related pathways can be due to either dysregulation or an effort to adjust to the increased inflammation. Additionally, the contribution of cancer and muscle function related pathways dropped below 5 %, putting them in the others category. This change indicates that pathways related to cancer are more tightly regulated at 9 months than at 15 months, and therefore, we cannot detect a significant difference when comparing later time points. This can be explained by the accumulation of damage with age [174, 175]. In the remaining DEGs (others category), a relatively high number of DEGs (15) were related to the cell adhesion molecules pathway with a relative low p-value ($2e-6$) for both age comparisons. These factors can be a prognostic for blood-brain barrier integrity.
Prolla [176] has already discussed similar results with respect to brain aging comparing his findings with calorie restricted mice. He reported several up-regulated genes related to stress and immune response, along with an accumulation of ubiquitinated proteins in control mice. However, this effect appears to be slightly reduced

in calorie restricted mice.

Blood
In blood tissue, we identified 815 and 1106 DEGs within the age comparisons of 9 vs. 30 and 15 vs. 30 months respectively, of which 201 and 269 were annotated using DAVID. DEGs contributing to inflammation and immune response still made up the largest group in the significantly altered pathways at both time point comparisons. A closer analysis on the transcript expression revealed, that genes related to inflammation were again higher expressed while genes related to the H2 complex (MHCII) were lower expressed at 30 months. Changes in metabolism-related pathways were on the same order, with only a slight increase to 13 % and 18 % at the two time points. The major difference between brain tissue and blood was apparent with in cancer pathways, which were part of the others category and made up less than 5 % in the 9 vs. 30 months comparison and approximately 6.56 % for the 15 vs. 30 months comparison. At the later age comparison, most of the DEGs comprising cancer pathways were found to be higher expressed at the age of 30 months. This suggests that in blood, the changes related to cancer are more prominent at 15 months and might increase at a later age.

Skin
The skin appears to age in a relatively distinctive manner, possibly due to direct exposure to external environmental factors, which is limited for most of the other tissues. For the 9 vs. 24 months comparison, 482 DEGs were identified. Of these, 332 were known by DAVID and 302 genes belonged to the olfactory transduction pathway with a very low enrichment p-value of $7e-223$. Skin has many different receptors, helping an individual to perceive its immediate environment[177]. All 302 DEGs contributing to olfactory transduction were lower expressed at 24 months of age, probably showing a loss in this specific ability when contrasted to young adult mice (9 months). When comparing the two younger ages (9 months and 15 months) to the long-lived (30 months) mice, we identified approximately the same number of DEGs (579 and 539). Of these, 155 and 142 DEGs were associated with known KEGG pathways. Among both age comparisons, changes related to metabolism and cancer were approximately equal, but there was an increase of 6 % in inflammation- and immune response-related up-regulated genes, from 19.7 % to 25.7 %, with advanced age. This complements the reduction of signaling pathway-related genes from 14.8 % to 9.7 %. The most notable changes here are genes related to muscle function and the extra cellular matrix, adhesion and the cytoskeleton. For these categories we observed an increase to 7 %, showing a dysregulation of the primary functions of the skin, *e.g.*, barrier function or wound healing. It is also complemented with a decreased expression of collagen related genes. Additionally, within the signaling pathway category, we found the PI3K-Akt and the FoxO signaling pathways to be the most significantly changed. Both pathways are strongly involved in the processes of aging and contribute to longevity [13].

Liver
Further, in liver, we have identified 3219, 2778 and 1296 DEGs for the above mentioned age comparisons, for which 942, 814 and 389 were categorized as associated with known pathways, respectively. We observed the same trend in each consecutive age comparison, *i.e.*, a decrease in inflammation and immune processes, while the regulation of metabolic pathways was increasing with age. Major gene expression changes were observed in the comparison of 9 months with 15 months and 15 months with 24 months old mice. In 9 to 15 months, most genes were higher expressed in the older mice, irrespective of their functional category. Furthermore, changes in pathways associated with cancer or signaling were relatively constant. As the age progresses, more and more DEGs fall into mixed functional categories, showing a slow dysregulation with aging. In absolute terms, changes in the metabolic processes were by far the largest in the liver, highlighting the importance of this tissue in the regulation of these functions during aging.

Comparison of tissues
Most of the DEGs from the brain and blood were involved in inflammation and immune response, as we already suspected from their topmost regulated genes. Generally, both tissues show a similar pattern of regulated pathways during aging. In contrast, skin showed a very heterogeneous pattern of regulated pathways across different age time points (see Figure 3.5). It displayed no obvious pattern such as the ones seen in the brain and blood, which confirms our observations on the most significantly regulated DEGs. Again, liver had a unique and specific pathway pattern among the four tissues. In all time comparisons, metabolic pathways in the liver were most strongly regulated and even increased with time. A more detailed subdivision and additional details about these metabolism pathways can be found in SData 3B. However, across all regulated pathways and time comparisons we found cancer-related pathways, which is intuitive as the interplay of cell growth promoting and tumor suppressing genes is ever-present during aging [52].

3.1.4 Temporal expression profiles of tissues reveal a similar regulation of electron transport chain in brain, blood, skin and liver

Next, we clustered all genes according to their temporal expression behavior, for the four tissues separately and analyzed specific clusters, namely, all genes that showed a constant expression at earlier time points but an up- or down-regulation of at least 25 % (blood, liver, skin) or 10 % (brain) at later ages. This mainly included three types of expression profiles: (1) genes that had a relatively constant expression until the age of 24 months and increased or decreased at 30 months, (2) genes that showed constant expression until 15 months of age, with their expression levels rising or dropping at 24 months and at 30 months, and (3) genes with an increased or decreased expression at 15 months and that stayed constant until the age of 30 months, at which time they returned to their initial level of expression (see Figure 3.2). Functional annotation of these genes revealed an enrichment of genes acting in age-related KEGG pathways, such as Alzheimer's disease, Parkin-

son's disease, Huntington's disease, non-alcoholic fatty liver disease and oxidative phosphorylation, in all tissues. A closer look yielded three interesting observations. First, all age-related pathways mentioned above share a number of genes coding for electron transport chain subunits in mitochondria. Second, genes related to these pathways showed a higher expression at 30 months than at 24 months in every tissue but blood, where the age of 24 months marked the highest gene expression. Third, most of these genes were related to mitochondrial dysfunction (for details about these genes and their expression levels, see SData 4 and STable 4). These genes include, ATP-synthetases, NADH Ubiquinone Oxidoreductases (NDUFs) and Cytochrome C oxidases (COXs), which are all part of the respiratory chain. Therefore, they are functionally related to each other, which may explain their similar expression pattern. We conclude that several complexes of the respiratory chain were similarly regulated in all four tissues during aging. A number of studies have already shown that (defective) mitochondria, as a source of ROS, contribute to neurodegenerative disease and to aging in general and that the suppression of complexes I, III and IV of the electron transport chain at a young age can increase overall lifespan[178, 179]. For example, Kowang and Sohal examined the mitochondria isolated from brain, heart, skeletal muscle, liver, and kidney of mice and found that individual complexes of the electron transport chain were decreased in old animals, with potential adversely effect on oxidative phosphorylation [180]. With increasing age, defective mitochondria in post-mitotic cells also increase in size, making it more difficult for them to undergo mitophagy. This could explain the reduced expression of the respiratory complexes in blood compared to the other tissues, because it has a much higher cell turnover rate. In contrast, there are studies showing that down-regulation of ETC-complexes, due to arsenic poisoning, could inhibit aerobic respiration, which can be deleterious, especially for neural tissues. Additionally, defects in SDHD (complex II) can induce tumors [181]. The maintenance of optimal expression levels of the electron transport chain genes is crucial for each tissue, in order to generate sufficient ATP from aerobic transpiration without the excessive production of ROS. Since this is a cross-sectional study, we cannot comment on the exact expression levels in the long-lived mice, leaving the question open whether mice reaching 30 months of age have a more balanced and efficient regulation of the electron transport chain during their younger years.

3.1.5 Tissue independent markers for aging

Processes regulated with aging among all tissues

The striking finding of our pathway enrichment analysis shows every tissue follows its own pattern of aging, driven by tissue-specific processes. However, we also identified processes regulated across all investigated tissues, even if different genes regulate them. In the 9 to 30 months comparison we found the hematopoietic cell lineage pathway to be significantly enriched in all four tissues; however, this is due to an enrichment of inflammation-related genes, such as interleukin receptors (*e.g.*, *Il7r* and *Il4ra*) and histocompatibility complexes (*e.g.*, *H2-be1*). Additionally, several pathways related to immune response and metabolism showed enrichment in three (blood, brain and liver) of the four tissues (see Table 3.1). Both findings support the

concept of inflammaging, as hematopoietic aging is also associated with a change in the microenvironment due to infection and inflammation [182]. Table 3.1 shows all significantly enriched pathways, that were shared by the blood, brain and liver tissues. We determined the phagosomal pathway to be one of the most affected pathways throughout aging, which again is evidence that autophagy might be one of the main driving factors of aging. Interestingly, the cell adhesion molecules pathway seemed also to be highly affected in all three tissues during aging. Cell adhesion molecules play a crucial role in a variety of biological processes related to homeostasis, including immune response, synapse formation and oxidative stress [183, 184]. Very likely, all other shared enriched pathways appeared due to their relation to inflammatory processes or the immune system in general and many of the differentially expressed genes in the investigated tissues are immune or proinflammatory genes. It is conceivable that many of these are a consequence of aging rather than a cause of it. Nevertheless, the role of inflammation and immunity as stressors for age-related diseases and processes is actively being discussed [185, 186]. According to the suggested inflammaging phenomenon (or inflammation hypothesis of aging), the production of proinflammatory factors becomes less useful with age due to the accumulation of frail immune gene variants and an overall loss in specificity and efficiency of the immune system. The result is a constant inflammatory stimulus (or stress), promoting age-related processes/diseases, such as Alzheimer's disease, osteoporosis or diabetes. This reasoning follows the same principal ideas of the antagonistic pleiotropy hypothesis that was proposed in 1957 to explain the contrary effects of senescence and has already been the subject of aging-related discussions [29, 55]. Our data supports these ideas since many of the identified alterations in gene expression are related to the immune response, infectious diseases and/or inflammation. It remains to be elucidated what specific influence these genes might have on the aging of different tissues.

3.1.6 Commonly expressed DEGs in all tissues

Most of the significantly differentially expressed genes were unique to a certain tissue. However, we aimed to focus on potential marker genes for aging in mouse. The overlap of all DEGs is depicted in Figure 3.6. We found 125 genes to be common in at least three examined tissues and seven genes that showed a significant change in all four tissues. The liver showed the highest number of differentially expressed genes (almost 5200 genes), most of which (approximately 80 %) were unique to the liver during aging. Analyses of the skin exposed approximately 65 % unique genes, followed by the blood (47 %) and brain (46 %).
The seven DEGs common to all four tissues were *Vmp1*, *Rap2a*, *Igkv4-62*, *Gm8979*, *S100a6*, *Lcn2*, and *S100a9*, of which the last two were constantly increasing their expression with age (see Figure 3.7. Three of these seven proteins are involved in different parts of the immune system: LCN2 functions in innate immunity [187], whereas IGKV4-62 (immunoglobulin kappa variable 4-62) is part of the adaptive system encoding a kappa light chain for antibodies [188]. The protein S100A9 (also known as MRP14) controls the accumulation of neutrophils and macrophages by binding another of the S100 gene family members (S100A8) and shows a generally

KEGG ID	Pathway name	#KEGG genes	Blood #DEGs	Blood p-Value	Brain #DEGs	Brain p-Value	Liver #DEGs	Liver p-Value	inters. genes
mmu04145	Phagosome	174	24	$3.1e^{-6}$	23	$3.8e^{-10}$	50	$2.3e^{-7}$	TLR2, CYBA, H2-EB1
mmu04514	Cell adhesion molecules	162	15	$1.5e^{-2}$	16	$1.4e^{-5}$	34	$9.6e^{-3}$	ICOSL, H2-EB1
mmu04610	Complement & coagulation cascades	76	11	$2.1e^{-3}$	7	$1.3e^{-2}$	23	$3.2e^{-4}$	C1QA, C1QB
mmu04612	Antigen processing and presentation	82	16	$3.1e^{-6}$	20	$1.0e^{-13}$	37	$1.4e^{-11}$	CD74, H2-EB1
mmu04940	Type I diabetes mellitus	62	9	$6.4e^{-7}$	13	$3.7e^{-8}$	22	$3.5e^{-5}$	H2-EB1
mmu05140	Leishmaniasis	64	15	$6.9e^{-7}$	6	$2.3e^{-2}$	22	$6.0e^{-5}$	TLR2, CYBA, H2-EB1
mmu05142	Chagas disease	103	12	$6.5e^{-3}$	9	$4.6e^{-3}$	25	$4.7e^{-3}$	C1QA, C1QB, TLR2, CCL5
mmu05150	Staphylococcus aureus infection	48	16	$1.4e^{-9}$	8	$1.9e^{-4}$	20	$6.4e^{-6}$	C1QA, C1QB, H2-EB1
mmu05152	Tuberculosis	176	31	$2.0e^{-10}$	14	$5.2e^{-4}$	34	$3.1e^{-2}$	TLR2, CD74, H2-EB1
mmu05164	Influenza A	171	22	$2.6e^{-5}$	12	$4.1e^{-3}$	43	$6.4e^{-5}$	CCL5, H2-EB1
mmu05166	HTLV-I infection	278	21	$2.6e^{-2}$	17	$1.9e^{-3}$	57	$1.2e^{-3}$	H2-EB1
mmu05168	Herpes simplex infection	207	22	$4.1e^{-4}$	27	$9.8e^{-12}$	54	$1.9e^{-6}$	TLR2, CCL5, CD74
mmu05320	Autoimmune thyroid disease	70	9	$1.3e^{-2}$	13	$1.5e^{-7}$	20	$1.9e^{-3}$	H2-EB1
mmu05323	Rheumatoid arthritis	80	15	$1.1e^{-5}$	8	$4.1e^{-3}$	24	$2.6e^{-4}$	TLR2, CCL5, H2-EB1
mmu05330	Allograft rejection	56	8	$1.2e^{-2}$	13	$1.1e^{-8}$	21	$2.2e^{-5}$	H2-EB1
mmu05332	Graft-versus-host disease	52	8	$8.4e^{-3}$	13	$4.4e^{-9}$	21	$6.2e^{-6}$	H2-EB1
mmu05416	Viral myocarditis	79	10	$8.3e^{-9}$	14	$7.6e^{-8}$	29	$6.4e^{-7}$	H2-EB1

Table 3.1: KEGG pathways that were significantly regulated and common in at least three of the investigated tissues. The first two columns give the KEGG identification number and the common name of the respective pathway. For every pathway, the total number of genes and the number of differentially expressed genes within this pathway are given for the blood, brain, and liver, individually. P-values are calculated based on a hypergeometrical approach used by `DAVID 6.8`. The first two pathways are potential marker pathways for aging. No overlap was found with any enriched pathway in skin tissue. Inters. genes – Common DEGs in all three tissues.

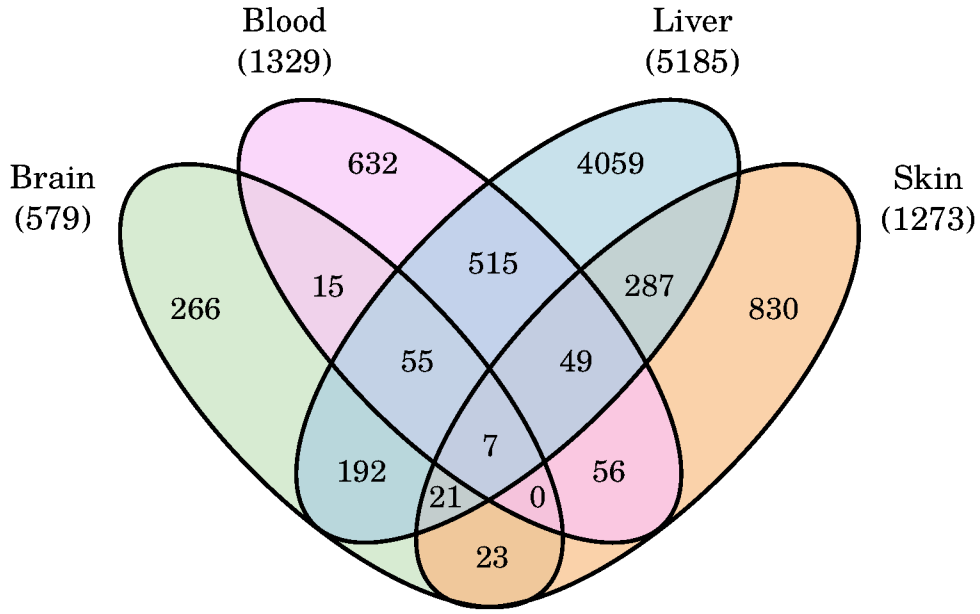


Figure 3.6: **Venn diagram of all significantly differentially expressed genes in the four analyzed tissues.**
Whereas most DEGs were unique to a single tissue, a total of 125 genes were nevertheless common in at least three tissues, and seven were common in all four of them. The total number of DEGs in the individual tissues is written in brackets. Details and gene names can be observed in SFigure 2A and STable 3.

proinflammatory role [189].

The relation of these immune- and inflammation-specific genes to aging might be inversely causal: the change in their transcription might be caused by aging, not necessarily causing aging. Older mice are more vulnerable to infections and diseases, and with increasing age, a progressive activation of the immune system is to be expected. Additionally, LCN2 (Lipocalin-2) has been implicated to play a role in aging disorders in several other studies. For instance, its overexpression in the brain was linked with an increased insulin resistance in advanced age, resulting in obesity [190]. It has also been considered as a marker for multiple sclerosis [191] and was described to be associated with oxidative stress, inflammation and demyelination of neurons, which in turn lead to cognitive impairment [192].
The second member of the S100 gene family that we found to be differentially expressed in every investigated tissue is S100A6, a gene with an uncertain function. It is suggested that S100A6 is involved in cell cycle progression [193] and is a regulator of the S100B-dependent signaling pathway, responsible for a variety of functions in glia cells [194]. Interestingly, in a similar study based on human RNA-Seq data we also identified a member of the S100 gene family (S100P) to have an altered expression in all investigated tissues. VMP1 is known to be a key regulator in the initial steps of the autophagosome formation and is only one of these seven genes that can directly be related to aging [195]. One of its many known and suggested functions (for a comprehensive overview, see review [196]) is to control the half-life

of proteins and cellular organelles. In our data, we observed a small but constant decrease in the expression of *Vmp1* in all investigated tissues over age, except blood, where its expression is slightly increasing. Reduced autophagosomal activity can lead to an accumulation of misfolded or damaged proteins, resulting in apoptosis and neurodegenerative diseases [197]. The role of autophagy in aging has already been discussed intensively in recent years [198], but here, we observed concrete evidence that VMP1 might be a potential common genetic driving factor for aging induced by a reduced autophagy potential. Lastly, almost nothing is known about the two remaining proteins encoded by the genes *Rap2a* and *Gm8979*. Both are known to have the capability to bind and hydrolyze GTP. RAP2A is a member of the Ras protein superfamily and therefore might function in the regulation of cell proliferation, whereas GM8979 is only a predicted protein that has yet to be confirmed by molecular and/or genetic analysis [199].

Age-specific expression pattern validation by qPCR

We performed expression validation by qPCR of the seven abovementioned DEGs for the brain samples first, but unfortunately most of them showed only very low expression (see SFigure 5). However, the expression profile of *Lcn2* seemed to be promising, showing a relatively strong expression with lesser deviation among samples. Thus, we proceeded with the *Lcn2* expression validation among the other tissues. Though we observed the same pattern of expression as in the RNA-Seq profiles, the expression strength was reduced (see Figure 3.8).

3.1.7 Effect of *Lcn2* orthologue knockdown in *Caenorhabditis elegans* using RNAi

Since *Lcn2* showed a consistent expression in all samples and significant up-regulation during aging, we further analyzed its effect on the lifespan in the invertebrate model *Caenorhabditis elegans* using RNAi. *Lcn2* is orthologous to a family of *Caenorhabditis elegans* lipocalin-related proteins, which consists of seven paralogs (*Lpr1,–7*). Only one of the paralogs, *Lpr6*, is expressed in neurons (amphid sensory neurons), suggesting that its function is probably closest to that of mammalian *Lcn2* [200]. We next decided to investigate the impact of *Lpr-6* on longevity, using an RNAi-mediated gene inactivation approach. As shown in Figure 3.9, *Lpr-6* deficient animals show a trend of extended longevity early in life, resulting in a median lifespan extension of one day (from 18 days to 19 days), while maximal longevity is unchanged. This suggests that down-regulation of *Lpr-6* (or *Lnc2*) alone is not sufficient to have a significant influence on the overall life-span, which is plausible considering the complexity of aging. However, its reduced expression still has a positive and reproducible effect on the survival rate of *Caenorhabditis elegans*, even if it is minor.

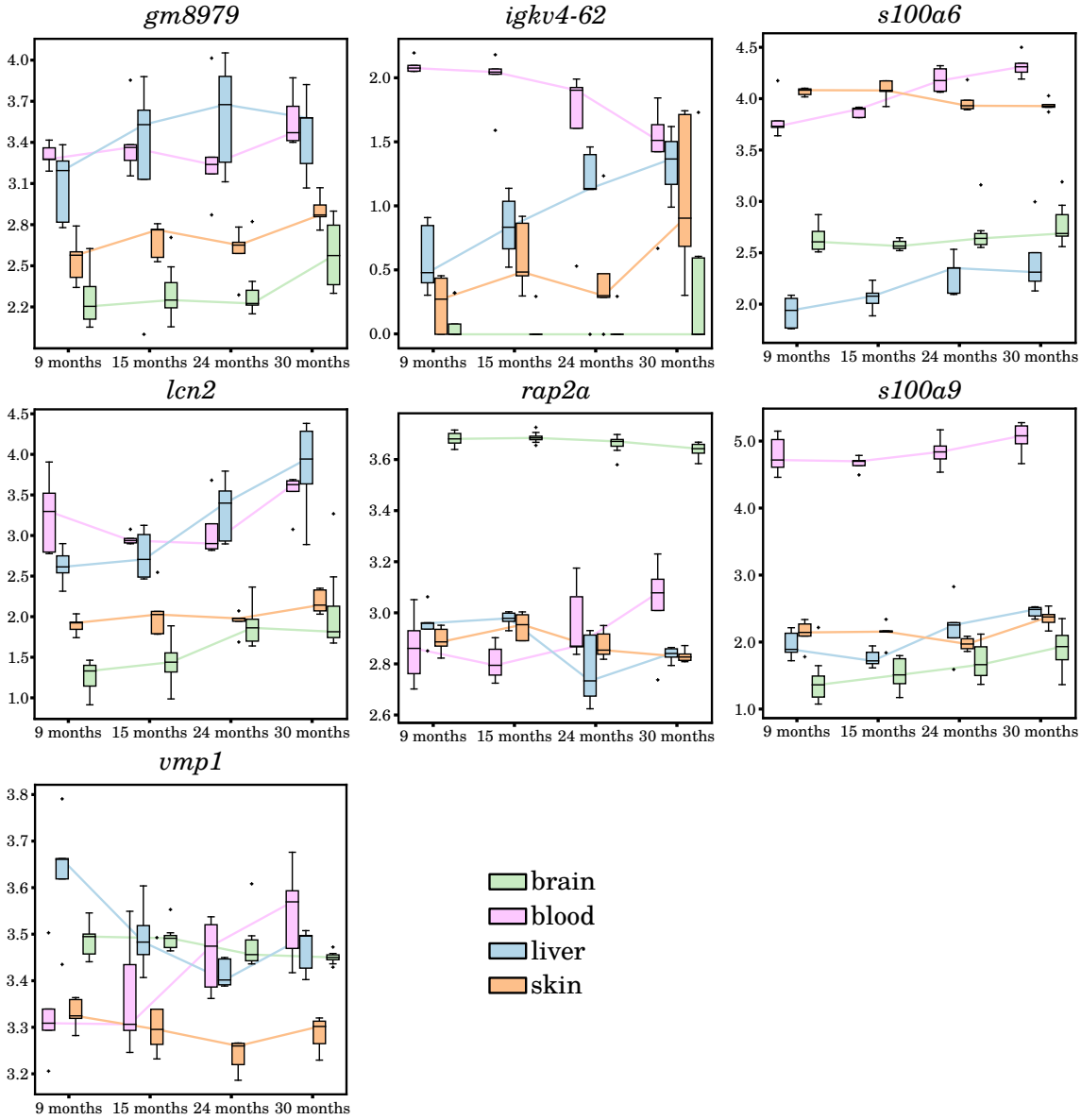


Figure 3.7: **Temporal expression profiles.**
Temporal expression profiles of the seven genes (*Gm8979*, *Igkv4-62*, *Lcn2*, *Rap2a*, *S100a6*, *S100a9* and *Vmp1*) that were differentially expressed in all four tissues during aging. The y-axes provide the expression strength as log_{10} normalized counts of expressed reads of each gene. Small plus signs indicate outliers. With the exceptions of *Lcn2*, *S100a9* and *Gm8979*, none of the genes showed a consistent expression pattern that was shared between the tissues. More details regarding the expression profiles of all seven genes can be found in SFigure 2B and SFigure 2C.

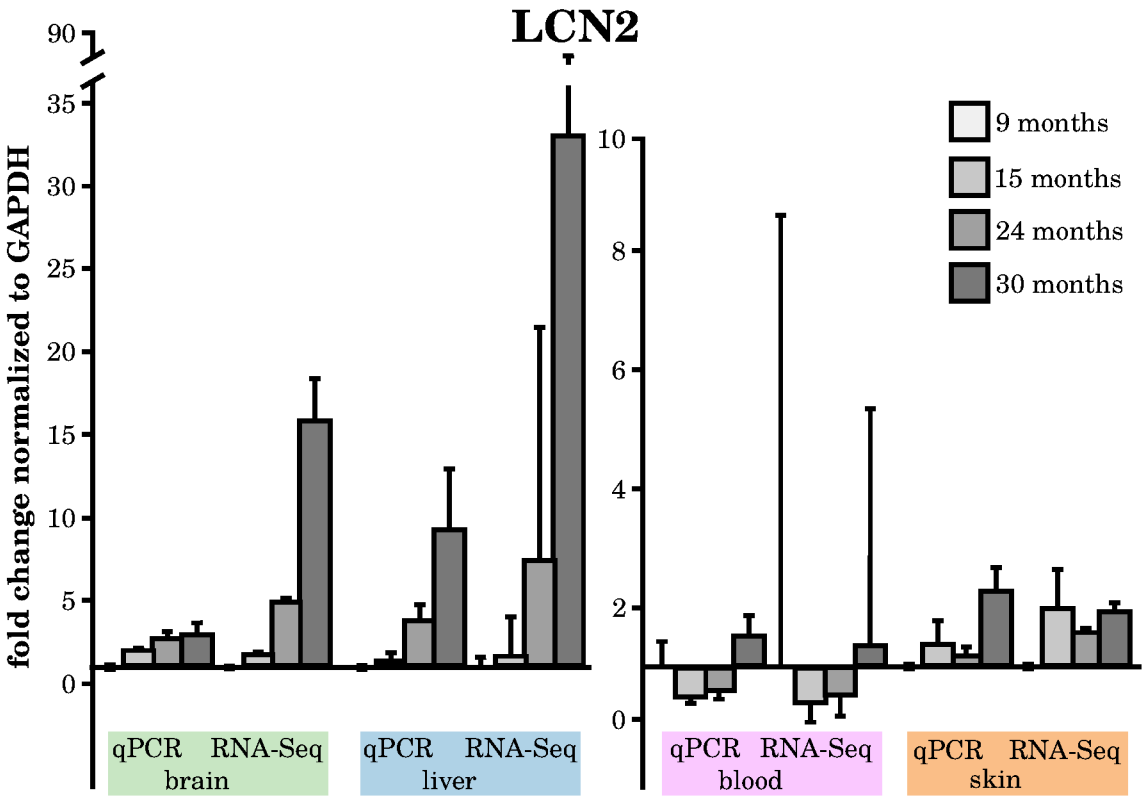


Figure 3.8: **qPCR expression validation of *Lcn2*.**
Expression validation of *Lcn2* by qPCR in comparison to its RNA-Seq profile. While we observed the same pattern of expression of *Lcn2* in both experiments, the strength of expression was reduced in the qPCR validations compared to our RNA-Seq data.

3.2 Why it is not easy to find biomarkers of aging

In this extensive study of 92 male murine RNA-Seq libraries from four organs and several age time points, we showed that organ-specific gene expression changes prevailed over tissue-independent aging-linked gene expression changes (Figure 3.3). This phenomenon indicates that different tissues age differently because of intrinsic and extrinsic factors specific to these tissues. The liver displayed a much higher number of DEGs, suggesting one of two possibilities: This might indicate better adaption to changes over time to coordinate its functions within the aging body, or contrary, it could also be a sign of loss of plasticity with aging. In addition, the liver has to deal with systemic toxins, that become more abundant with aging, causing, among other things cell death and inflammation. Early on, the liver showed signs of inflammation that might later account for dysfunction of hematopoiesis in the liver [201]. On the other hand, anti-ROS-related pathways could work to sustain the normal functioning of the body. With a further increase in age, the liver showed tremendous changes in metabolic pathways, which can be considered a two-way response to aging, *i.e.*, following and resisting certain aging processes. In addition, anti-ROS pathways are important defense mechanisms against the rise of the

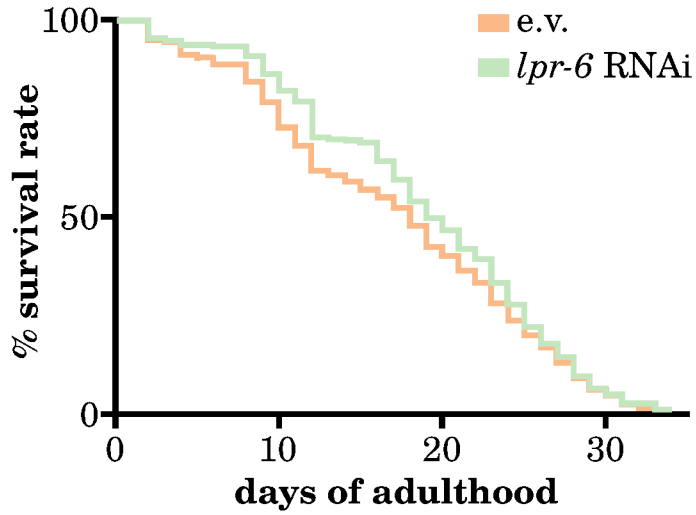


Figure 3.9: ***Caenorhabditis elegans* survival rates with and without *Lpr-6* knockout.**
Survival rates of wild type *Caenorhabditis elegans* (e.v. - empty vector) and *Lpr-6* knockouts. When compared, there seemed to be no significant increase in the total life-span of individuals, but there was a slight protection against early lethality. The survival experiment was repeated three times (n=3).

cancer-related pathways seen when comparing young and old mice [202, 203]. While investigating genes with a distinct pattern of expression during aging, *i.e.*, genes showing more drastic changes at the later time points, we found many genes of the Cytochrome C oxidases (COXs) and NADH Ubiquinone Oxidoreductase (NDUFs) families to be regulated similarly in all four tissues. Members of the NDUF family are part of complex I of the mitochondrial electron transport chain and have been studied prominently with respect to cancer, ROS and lifespan [204]. Being the last enzyme of the mitochondrial respiratory chain, COX plays a key role in senescence, and therefore, its respective subunit genes are eminently studied with respect to aging and cancer and are often used to evaluate the level of oxidative stress [205, 206]. The stability of the COX assembly requires efficient ATP-synthetase activity, and we showed the expression of ATP-synthetase genes that followed the same temporal profile as the COX genes (*e.g.*, *Cox7a2*, *Cox4i1* and *Cox8a*), indicating that an increase in the COX expression level could be a consequence of or an effort to balance disrupted ATP-synthetase [207]. Looking at the general picture of all age affected pathways for the investigated tissues (Figure3.5), we suspect chronic inflammation to be the major or at least the most prominent correlative of aging, in regard to gene expression changes. Nevertheless, the liver appears to oppose the resulting negative changes through metabolic adaptations. The hematopoietic cell lineage pathway was the only pathway shared by all four tissues when comparing adult mice of 9 or 15 months to a highly aged individual of 30 months, because it contains a number of inflammatory genes, which are not specific to blood. This supports the hypothesis of the so-called immunosenescence. It also underlines the fact that several inflammation-related pathways were common in at least three tissues, which

further supports the theory that immunosenescence results in inflammaging [208]. We were able to identify specific DEGs to be common in all or at least three tissues. These genes could be of special interest for further aging research to determine whether they constitute a part of the aging process or results from it. The seven genes common in all investigated tissues are particularly important since they are involved in the immune system, cell cycle, autophagy and calcium signaling. These are processes that are intermittently being discussed as the main cause of aging [13]. Several studies have also shown a change in the cell type population with aging, which may also affect the expression pattern of the whole tissue [209]. Therefore, it would be interesting to study the transcriptome on the single cell level to evaluate the contribution of such cell population changes with aging. Additionally, in this study male mice were chosen to avoid any interference of hormonal cycles of female mice on expression pattern of genes. However, including female mice could be helpful in understanding the process of aging in species per se and also to study sex linked signatures of aging. In a similarly exhaustive study, based on microarray data and histological examination, Jonker et. al. compared five different tissues (liver, kidney, spleen, lung and brain) from mice at five age-points (13, 26, 52, 78, 104 and 130 weeks) [147]. They found several hallmark age-related pathologies across tissues and observed some tissue specific effects, such as an accumulation of lipofuscin in brain and liver, thickening of glomerular membrane in kidney and increased peribronchiolar lymphoid proliferation in lung. They also found *Lilrb4* (Leukocyte Immunoglobulin Like Receptor B4) to be up-regulated in all organs. However, immune-related genes were down-regulated in spleen but up-regulated in kidney and lung which can be justified as they observed reduction in lymphocytolysis in spleen. We conclude that aging is a very heterogeneous process, and any single gene alone can hardly be regarded as a marker of aging for a whole organism. Every tissue has its own specialized function and we showed and justified several tissue specific processes to be regulated during aging. They may overlap in some cases of particular genes or common processes, but except for chronic inflammation and imbalances of the mitochondrial electron transport chain, it is difficult to identify any process or gene to be responsible for cumulative aging. Together with the observation that most of the identified DEGs can only be found performing long-term comparisons this suggests that aging is a more subtle process, manifesting gradually in animals.

Chapter 4

The Role of MicroRNAs in Aging

As mentioned in Chapter 1, with the beginning of the 21st century one revolution chased aftr the next in molecular biology, with one of them being the realization that most parts of eukaryotic genomes were not junk DNA, but encoded for various different controlling mechanisms [9]. One of those regulatory elements were found to be multiple new classes of non-coding RNAs, which besides the commonly known mRNAs, tRNAs and rRNAs seemed to have crucial and determining functions in (almost) all aspects of life [10]. One of those new regulatory non-coding RNA classes were microRNAs (miRNAs). Being discovered already in 1993 as small antisense binding RNA sequences within the nematode *Caenorhabditis elegans* [210], it still took almost 10 more years until they were recognized as an evolutionary conserved class of non-coding RNAs in nearly all domains of life, even including some viruses [211–216]. Since they are key regulators of many different biological pathways, they are also of special interest for the aging process.
In this chapter we want to explore their share on the regulation of age-related processes. At first, we will see how miRNAs can be annotated bioinformatically in newly sequenced and assembled genomes on the example of the short-living fish *Nothobranchius furzeri* (Section 4.2). Next, we investigate their role in the phenomenon of diapause in annual fishes, such as *Nothobranchius furzeri*, and how this is linked to certain aging processes (Section 4.3). Then we introduce a new statistical model, which aims to improve the detection of significant expression changes of miRNAs between two different conditions, like a young and an old time point, or a healthy and a diseased state (Section 4.4). Finally, we will examining conserved aging-related regulatory functions of miRNAs in evolutionarily more distinct species, based on results from our new statistical model (Section 4.5). But before we come to all of that, we will first briefly introduce the biogenesis and molecular function of miRNAs (Section 4.1).

4.1 Small in size, huge in regulation: How miRNAs control biological processes

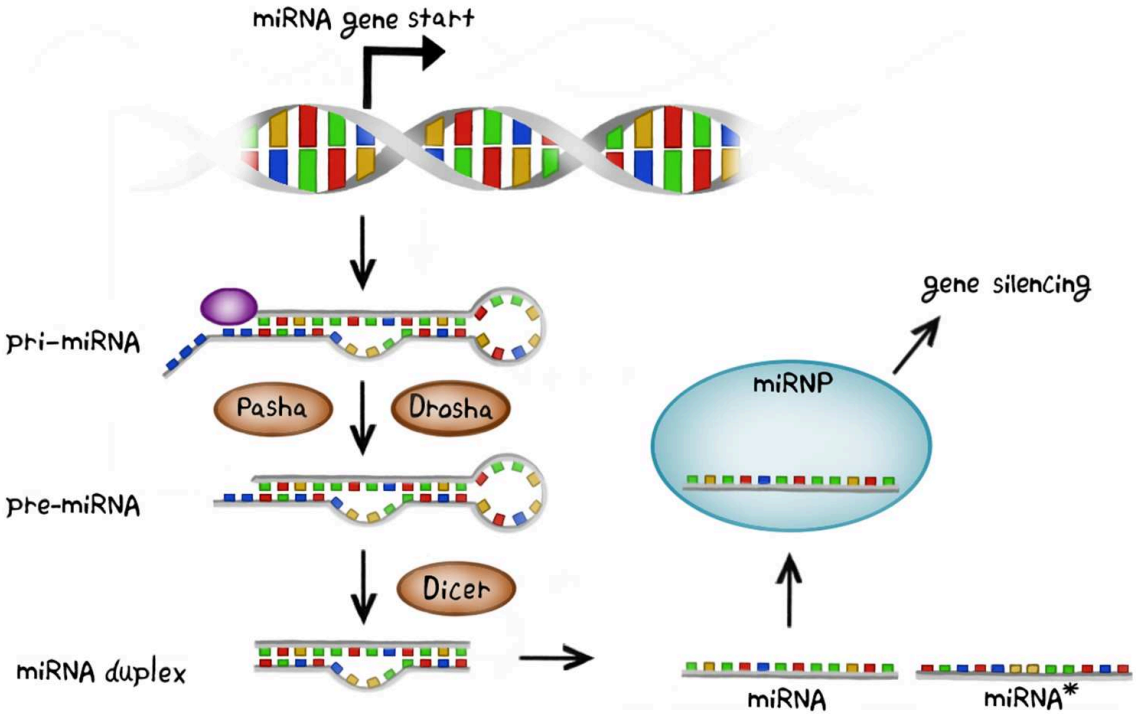


Figure 4.1: **A short overview of the canonical miRNA biogenesis pathway.** Starting with a primary miRNA (pri-miRNA) molecule after transcription, pri-miRNAs get cleaved and further processed into one or more precursor (or premature) miRNAs (pre-miRNAs) by the proteins Pasha and Drosha. After transportation from the nucleus into the cytoplasm by Exportin-5, Dicer cuts pre-miRNA transcripts into double stranded mature miRNA duplexes with an approximate size of 22 nucleotides. Both mature miRNAs are separated from each other and either the more stable one or both are loaded into separate protein complexes, forming the so-called RNA-induced silencing complex (RISC). This ribonucleoprotein complex can now bind target mRNAs, almost always leading to their degradation and translational repression.

Whereas about 40 % of all miRNAs are encoded within introns or exons of other genes and are transcribed alongside their host genes, the majority of miRNA genes consists of an independent promoter and is transcribed by the RNA polymerase II [217–219]. In the latter case a so-called primary miRNA molecule is transcribed, which contains a special 5' end cap (*i.e.*, a modified nucleotide) as well as a polyadenylated tail at the 3' end, similar to mRNA transcripts (see Figure 4.1) [218]. These primary miRNAs can be several hundred nucleotides long and contain up to six miRNA precursors, each with a length of about 70 – 80 nucleotides [218]. The two proteins Drosha and Pasha recognize the double stranded structure of primary miRNAs and process them by cutting them into the precursor-miRNAs, that have a typical hair needle structure with a small unpaired overhang at the 3' end [220].

Such precursor-miRNAs can also be generated skipping the Drosha/Pasha processing mechanisms, if they are directly spliced from their host gene [221]. Either way, precursor-miRNAs are transported from the nucleus into the cell's cytoplasm via the protein Exportin-5, where they are further processed by the Dicer protein [222, 223]. Dicer cuts precursor-miRNAs into about 22 nucleotide long, still double stranded miRNA duplexes by removing the loop of the hair needle structure as well as the overhanging 3' end part of the duplex[224]. In plants this canonical pathway is slightly altered. Instead of two different processing steps, the plant homolog of Dicer (Dicer-like 1) performs both cutting steps, maturing the primary miRNA transcripts directly into miRNA duplexes within the cell's nucleus, before it gets exported by the Exportin-5 homolog named Hasty [225]. Besides this canonical miRNA biogenesis pathways, there exist multiple alternative dicer-independent pathways, generating functional miRNA molecules [226]. As a final step, the miRNA duplex gets unwind and either the more stable or both resulting mature miRNAs get loaded into the so-called RISC complex (RNA-induced silencing complex) [227]. Besides one mature miRNA transcript, the RISC complex is contained of various proteins with many of them belonging to the Argonaute protein family [228]. This ribonucleoprotein complex now acts as a post-transcriptional gene control level by binding mRNAs (usually in their untranslated regions) due to sequence complementarity, downregulating their translational rate either by direct degradation or blocking them from entering ribosomes [228]. These defined binding sequences of miRNAs are also called *seed regions* and usually compromise only seven to eight bases. Since RISC incorporated miRNAs do not have to perfectly base pair with their target mRNAs, one mature miRNA can have multiple mRNA targets, making target prediction analysis rather complicated [229, 230]. But it has also been observed that in some rare cases miRNAs can upregulate the expression of some protein-coding genes by interacting with their promoter regions [231]. In general, miRNA act in a wide variety of biological processes and pathways as repressing regulatory elements via RNA silencing [232].

4.2 Identification and annotation of known and new miRNAs in the killifish *Nothobranchius furzeri*

This chapter's section is based on the publication "A miRNA catalogue and ncRNA annotation of the short-living fish *Nothobranchius furzeri*" [1].

Data used and analyses performed in this section

Analyzed RNA-Seq data

In total, 169 small RNA-Seq libraries from seven different killifish species were analyzed in the study of this section. 157 of them belong to the JenAge dataset (as presented in Chapter 2) including all libraries of *Nothobranchius furzeri* strains *GRZ* and *MZM-0410* at several ages from the three tissues brain, liver and skin.
The remaining additional RNA-Seq libraries obtained from *Aphyosemion striatum*, *Nothobranchius kadleci*, *Nothobranchius rachovii*, *Nothobranchius pienaari*, *Nothobranchius kunthae* and *Nothobranchius korthausae* were used to identify expression patterns at predicted miRNA locations in *Nothobranchius furzeri* and miRBase premature miRNA sequences. For details see, Tab. 4.1, STable 1 and STable 2.

Table 4.1: **SmallRNA-Seq samples from *Nothobranchius* strains generated in this study.** * – unknown; # – number of replicates; + – two weeks post-fertilization plus diapause

Species	Tissue	Age (weeks)	No. library	#
Nothobranchius furzeri MZM	whole embryos	diapause II$^+$	7	7
Nothobranchius furzeri MZM	brain, liver, skin	5, 12, 20, 27, 32, 39	75	4–5
Nothobranchius furzeri GRZ	brain, liver, skin	5, 7, 10, 12,14	75	5
Aphyosemion striatum	brain	*	2	2
Nothobranchius kadleci	brain	*	2	2
Nothobranchius rachovii	brain	*	2	2
Nothobranchius pienaari	brain	*	2	2
Nothobranchius kunthae	brain	*	2	2
Nothobranchius korthausae	brain	*	2	2

[1] The complete supplemental material is available at `https://osf.io/25mxb/`

ncRNA and miRNA annotation

Already characterized and conserved non-coding RNAs were annotated with `GoRAP 2.0`, which is based on the RFAM database, currently holding 2,450 ncRNA families (v12.0) [233]. For an initial prediction of candidate miRNAs, a combination of five tools was used, each of them following a different annotation strategy: `miRDeep*` (v32) [234], `Infernal` (v1.1) [235], `BLAST` (v2.2.30) [236], `GoRAP` (v2.0, unpublished) and `CID-miRNA` (version from April 2015) [237]. A detailed description of the individual searches can be found below.
All results were merged and putative miRNAs overlapping with genes of the recently published *Nothobranchius furzeri* annotation were removed. The expression profiles of the remaining non-redundant candidate miRNA genes were analyzed automatically using `Blockbuster` (v1) [238] and in-house scripts in order to mark candidates that did not exhibit a typical miRNA expression profile (according to [239, 240]). All candidates were additionally manually examined and filtered by carefully checking the features of the potential hairpin secondary structure as well as the precise mapping of reads supporting the predicted precursor miRNA, leading to the final set of miRNA predictions.

miRDeep*

Mappings of 39 MZM brain, 15 GRZ brain, 25 GRZ liver, 28 MZM liver, 3 MZM skin and 7 MZM embryo small RNA-Seq libraries were used on four different fish genomes (*Nothobranchius furzeri*, *Danio rerio*, *Oryzias latipes*, *Takifugu rubripes*) as input for `miRDeep*` (for a detailed list of used libraries, see STable 1). Predictions from all 117 mappings were pooled together in order to obtain a comprehensive representation of the `miRDeep*` results. To each predicted miRNA hairpin sequence, we assigned the average of the `miRDeep*` score computed across the multiple samples were the sequence was found.
The merged non-redundant list of identified miRNA sequences was re-mapped with `BLAT` [241] on the *Nothobranchius furzeri* genome, and only gap-free alignments were accepted. These loci underwent further filtering steps: (i) a hairpin sequence was considered reliable if it showed a `BLAT` hit (one mismatch allowed) in miRBase (release 20) [242] or a miRDeep* score ≥ 7 and (ii) overlapping hairpin loci (*i.e.*, within 100 nt) were discarded, and the sequence with the highest score was kept. Predictions where no hits in miRBase could be obtained were further analyzed based on their secondary structure. Therefore, corresponding sequences were extended by 50 nt on either side and were compared with Rfam using `Infernal`. All predicted loci that had a significant hit to a known miRNA secondary structure or no hit at all were kept, while loci hitting other ncRNAs were discarded.

Infernal

For the `Infernal` search on the *Nothobranchius furzeri* genome, 155 hand-curated pre-miRNA covariance models were used as input [243, 244] and only significant hits with a p-value of $p < 0.005$ were kept.

BLAST
To identify candidates from the most conserved miRNA families, `blastn` was used with all mature and pre-mature miRNA sequences available on miRBase (release 20) [242]. Only non-redundant hits were kept if they spanned the complete sequences of their corresponding input miRNAs to at least 90 % with no gaps allowed. To further reduce false positive hits, a stringent cutoff of $p < 10^{-7}$ was chosen.

CID-miRNA
Being based on a stochastic context-free grammar model to identify possible pre-miRNAs, `CID-miRNA` follows a similar approach as `Infernals` covariance models. The *Nothobranchius furzeri* genome was given as input with the following thresholds: putative miRNAs have a length between 60 bp and 120 bp, and the grammar and structural cutoff were set to the recommended values of -0.609999 and 23, respectively.

miRNA family association

To determine if any of the predicted miRNAs belong to an already known miRNA family, all available hairpin and mature miRNA sequences were downloaded from miRBase. First, predicted pre-mature miRNA sequences were clustered with the known hairpin sequences if they shared at least 90 % identity. Second, for each cluster an alignment was build (using `Mafft` (v7.307) [245]) containing the predicted pre-mature and known miRNA hairpins as well as the known respective mature sequences. A predicted miRNA was only then associated with a given miRNA family if the known mature sequences matched with an identity of 95 % or higher. All predicted miRNAs not fulfilling these criteria or belonging to a cluster which does not contain a single known miRNA hairpin sequence were given new miRNA family names.

miRNA target prediction

To determine putative *Nothobranchius furzeri* mRNA targets of the miRNA candidates the `TargetScan` tool was used [246]. As input the putative miRNA seed regions and the known 3'-UTR sites of all annotated mRNAs of *Nothobranchius furzeri* as well as the ones from *Danio rerio*, *Mus musculus* and *Homo sapiens* were used. The input files and the resulting output can be found in the online supplement.
Enrichment scores of miRNA targets within different published sets of differentially expressed *Nothobranchius furzeri* genes were calculated using the hypergeometric test:

$$p - value = \frac{R!n!(N-R)!(N-n)!}{N!} \sum_{i=r}^{min(n,R)} \frac{1}{i!(R-i)!(n-i)!(N-R-n+i)i},$$

where N is the total amount of known protein coding genes in *Nothobranchius furzeri*, R the amount of differentially expressed genes of one of the given sets, n the number of protein coding genes with predicted miRNA target sites and r the size of differentially expressed genes with predicted miRNA target sites. Enrichment of

Figure 4.2: **The teleost *Nothobranchius furzeri* or turquoise killifish.** Photography of a male mature *Nothobranchius furzeri* (taken from `http://www.jenage.de/assets/images/general/nothobranchius.jpg`).

individual miRNAs, being enriched in any of the gene sets, were calculated similarly, with N being the total amount of protein coding genes with predicted miRNA target sites and n the amount of genes, showing a target site of the respective miRNA. The resulting p-values were adjusted using Benjamini and Hochbergs FDR approach and were considered significant if p was less than 0.05 [129].

4.2.1 *Nothobranchius furzeri* a new model organism for aging research

The annual teleost *Nothobranchius furzeri* is a recent experimental animal model in biomedical research. In the wild, this fish inhabits ephemeral pools in semi-arid *bushveld* of Southern Mozambique. It has adapted to the seasonal drying of its natural environment by producing desiccation-resistant eggs, which can remain dormant in the dry mud for one and maybe more years by entering into diapause. More about the topic of diapause can be found in the following section 3.3.
Due to the very short duration of the rainy season in its habitat, the natural lifespan of these animals is limited to a few months. They represent the vertebrate species with the shortest captive lifespan of only 4 – 12 months and also with the fastest maturation. In addition, they express a series of conserved putative aging markers and are amenable to genetic manipulations, making them an attractive model system for aging research (for a review, see [247, 248])
A striking characteristic of *Nothobranchius furzeri* is the existence of laboratory strains differing in lifespan and expression of aging phenotypes [117, 249]: an extremely short-lived strain (GRZ: median lifespan 3 – 4 months) and several longer-lived strains (*e.g.*, MZM-04/10; median lifespan 7 – 9 months). The molecular basis for this striking difference in aging is unknown. A previous miRNA-Seq study of brain aging that predated genome sequencing and used homology to miRBase to annotate *Nothobranchius furzeri* miRNAs revealed that the two strains have different global patterns of miRNA expression [120].
In this section, a comprehensive microRNA (miRNA) catalogue for *Nothobranchius furzeri* is presented. As already described in the previous section, miRNAs are abundant non-coding RNAs of short length that are produced in a complex biosynthetic pro-

cess starting from longer transcripts and are established as key players in the post-transcriptional regulation of gene expression. MiRNA genes can be hosted within an intron of a protein-coding gene (and their transcriptional regulation follows that of the hosting gene) or can arise from primary transcripts that are regulated independently of any protein-coding RNA. Several miRNAs are grouped in genomic clusters containing mostly two to six individual miRNAs with an intra-miRNA distance of less than 10 kb, which can be co-transcribed. However, unusually large clusters were also found in some species, like the miR-430 cluster in zebrafish, consisting of 57 miRNAs [250–252]. The advantage of this accumulation is unclear. It could be possible that multiple loci are required to increase the copy-number and therefore the expression level of specific miRNAs in particular conditions, like miR-430 in the maternal-zygotic transition in zebrafish (*Danio rerio*) [253].

Furthermore, miRNA genes are grouped into families based on sequence homology and can be defined as a collection of miRNAs that are derived from a common ancestor [233]. On the contrary, miRNA clusters may contain miRNAs belonging to different miRNA families, but are located in relative close proximity to each other. Both the evolutionary conservation of some miRNA families and the innovations leading to appearance of novel miRNAs are well-described. An expansion of the miRNA inventory due to genome duplications in early vertebrates and in ancestral teleosts has already been described [254].

MiRNAs bind target mRNAs, due to sequence complementarity in the seed region (nucleotides 2 – 7), mostly in the 3' untranslated region, thereby silencing expression of the gene product via translational repression and/or transcript degradation. Up to now, several thousands of miRNAs have been predicted and identified in animals, plants and viruses, and one single species can express more than one thousand miRNAs [242]. They frequently represent the central nodes of regulatory networks and may act as "rheostat" to provide stability and fine-tuning to gene expression networks [255, 256].

Before a sequence of the *Nothobranchius furzeri* genome assembly became available [123], it could be shown by use of the *Danio rerio* reference from miRBase that aging in the *Nothobranchius furzeri* brain displays evolutionary conserved miRNA regulation, converging in a regulatory network centered on the antagonistic actions of the oncogenic MYC and tumor-suppressor TP53 [120] and the expression of miR-15a as well as the miR-17/92 cluster, being mainly localized in neurogenetic regions of the adult brain [257]. Two draft genome sequences for *Nothobranchius furzeri* were recently produced [123, 258].

In this section, a comprehensive annotation of the *Nothobranchius furzeri* miRNome based on a combination of Illumina-based small RNA-Seq data, different *in silico* prediction methods on the genome assembly and a final manual curation is now described. Using the newly created miRNA reference, a large dataset of 162 small RNA-Seq libraries was analyzed and tissue-specific miRNA expression of conserved and non-conserved miRNAs in *Nothobranchius furzeri* is reported here. Further the *Nothobranchius furzeri* reference was used to analyze the miRNA expression in other Nothobranchius species and one closely-related non-annual killifish species, which were previously used to analyze positive selection [123] to identify when in the evolutionary history of *Nothobranchius furzeri* non-conserved miRNAs arose.

4.2.2 Prediction, annotation and characterization of miRNAs and other ncRNAs in *Nothobranchius furzeri*

Table 4.2: **Number of annotated ncRNAs.**
Besides housekeeping RNAs, RNA elements controlling metabolism and protein editing were identified. [a]mRNA regulation RNA element, [b]mRNA localization RNA element, [c]editing signal, [d]conserved lncRNA element.

ncRNA class	No.	ncRNA class	No.
18S rRNA	3	7SK	1
5.8S rRNA	11	Vault	12
28S rRNA	6	TPP	1
5S rRNA	28	IRE[a]	3
tRNA	570	CAESAR[a]	1
U1	6	DPB[a]	1
U2	8	Vimentin3[b]	1
U4	8	Antizyme FSE[c]	2
U5	8	U1A PIE[c]	1
U6	6	KRES[c]	26
U11	1	GABA3[c]	6
U12	1	Y RNA	4
U4atac	1	TERC	1
U6atac	1	mascRNA-menRNA[d]	42
RNase P	1	SPYR-IT1[d]	1
RNase MRP	1	MEG8[d]	1
SRP	3		

Annotation of ncRNAs

We could identify more than 750 non-coding RNA (ncRNA) genes in the *Nothobranchius furzeri* genome based on small RNA-Seq reads, including editing signals, RNA elements located in the UTRs of mRNAs either controlling localization or regulation and conserved lncRNA element (see Tab. 4.2, SData 1 and STable 5).
In line with other eukaryotes, we identified multiple gene copies of rRNAs, tRNAs, several major spliceosomal RNAs, signal recognition particle (SRP) RNAs and one copy of a minor spliceosomal RNA set. Further housekeeping RNA genes, such as RNase P, RNase MRP, and the 7SK RNA, are found, as expected, once in the entire genome.
We annotated the widely distributed TPP riboswitch, capable of binding thiamine pyrophosphate and thereby regulating genes that are in charge of the thiamine balance [259, 260]. We could also identify more RNA elements located in the UTRs of mRNAs, being directly involved in the regulation of gene expression (3 copies of *IRE* – controlling iron responsive proteins [261], CAESAR – controlling tissue

growth factor CTGF [262], DPB – controlling DNA polymerase β) [263]), localization of mRNAs (Vimentin3) [264]), DNA replication (four copies of the *Y RNA* gene, and Telomerase RNA TERC) or of unknown function (12 vault RNAs). Additionally, ncRNAs responsible for editing certain mRNAs have also been found (two copies of Antizyme FSE [265], one U1A polyadenylation inhibition element (PIE) [266], 26 Potassium channel RNA editing signals (KRES) [267], and six copies of *GABA3* [268]). Two promising candidate long non-coding RNAs (lncRNAs), SPYR-IT1 and MEG8, were also included in the annotation, even though we were not able to identify all of their exons. Two vague candidates for XIST and MALAT can be viewed in the supplemental material. The MALAT-derived *masc* and *men* RNA gene was clearly detected in 42 copies throughout the genome of *Nothobranchius furzeri*.
Additionally, we detected several copies for the RNA elements *unal2* and *secis* as well as the lncRNA gene *mimt*. However, because of their smooth transition from true copy to pseudogene, we included them in the current annotation as possibly non-functional.

Mapping and miRNA prediction results

For the identification of putative miRNA genes, we used five methods, each following a different prediction approach (`BLAST`, `CID-miRNA`, `Infernal`, `GoRAP`, `miRDeep*`) and `Blockbuster` as verification (see Fig. 4.4 for an example).

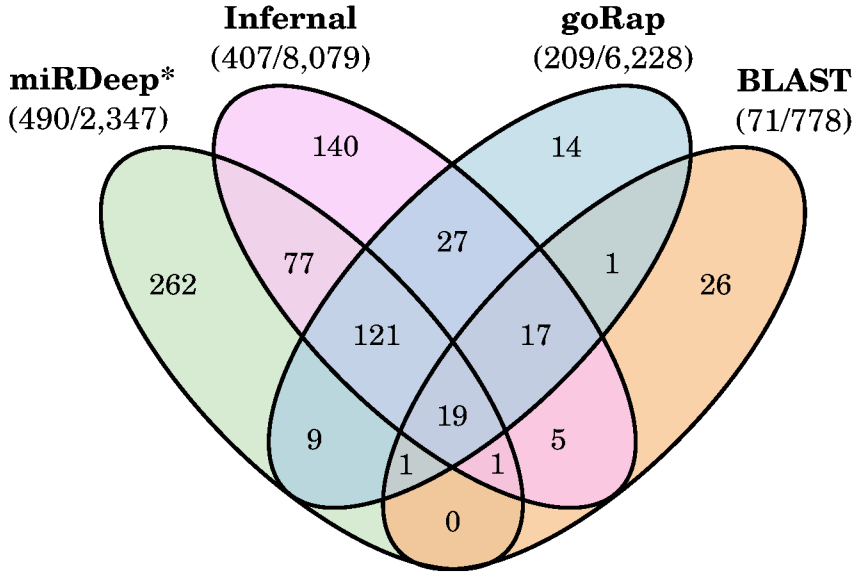


Figure 4.3: **Venn diagram of predicted miRNA genes from four tools `miRDeep*`, `Infernal`, `goRap` and `BLAST`.** Only 2 of the 33 candidates predicted by `CID-miRNA` overlapped with any of the other miRNA candidates. Nevertheless, all 33 candidates were selected as miRNAs after manual inspectations. The total number of miRNA predictions after and before applying any filtering step are shown in brackets for each tool.

The five tools identified 71, 33, 407, 209, 490 miRNA candidates, respectively (Fig. 4.3 shows the variety and the overlap of the different tools). All predictions

were merged, and redundant loci were removed (for details, see the first section of this chapter). Of the remaining 788 candidate miRNAs, 617 (78.3 %) showed expressions and were verified by `Blockbuster` and then manually verified for the correct mapping of the reads on the predicted precursor, with rectangular peaks corresponding to the mature 5p and/or 3p part, separated by a short gap devoid of mapped reads, while cases with more extended mappings were excluded. By this, 34 (4.3 %) candidates were removed, all predicted by `miRDeep*`. Candidates showing no expression in any of the sequenced small RNA-Seq libraries were still kept as putative miRNAs because they were predicted based on conserved and already characterized miRNA genes. In total, we predict a final amount of 754 miRNAs in *Nothobranchius furzeri* by the union of these methods (see SData 2).
Most of the small RNA-Seq reads (up to 88.81%) mapped onto the identified 754 miRNAs. Interestingly, the number of miRNA related reads varies broadly between the tissue samples (see Tab. 4.3). Possibly, this difference correlates with different regenerative capacities of these tissues. Mature brain cells are hardly proliferating, whereas liver cells are constantly renewed [41]. This regeneration might be additionally under the control of certain yet unknown tissue specific miRNAs.
About half of the miRNA annotations are overlapping genes coding for proteins and are therefore intragenic. A minor fraction of reads (see Tab. 4.3) maps to other ncRNAs and proteins. Whereas 333 of the predicted miRNAs can be assigned to one of the known miRBase families, based on sequence identity, 421 miRNAs did not match any known family and can therefore be considered as novel or non-conserved miRNAs (for details see Table 4.4).

The age-dependent expression of the following miRNAs was previously demonstrated by qPCR: tni-miR-15a, tni-miR-101a, tni-miR-101b, dre-miR-145, has-miR-29c-1 (100 % identical to dre-miR-29a), hsa-let-7a-5p, hsa-miR-124a-1, hsa-miR-1-2, ola-miR-21, ola-miR-183-5p and, from cluster dre-miR-17a/18a/19a, and dre-miR-20a (the used primers were Qiagen miScript primer). Expression changes detected by sequencing were validated on an independent set of specimens. All 13 miRNAs showed concordant changes in their expression, of which six reached statistical significance [120]. The expression of the following miRNAs in the brain was confirmed by *in situ* hybridisation using LNA probes (Exiqon): miR-9, miR-124 [269] and miR-15a, miR-20a [257].

4.2.3 Target prediction of the identified miRNA candidates

To get a first insight of the potential regulatory functions of our putative miRNA genes, we performed a target prediction based on the miRNA seed regions and the aligned homologous 3'- UTR mRNA regions of *Nothobranchius furzeri* and *Danio rerio*. Additionally, we repeated this target prediction analysis, including homologous 3'- UTR mRNA regions of *Mus musculus* and *Homo sapiens* to have a more conservative target list for each miRNA candidate, because *in silico* miRNA target predictions tend to have a high number of false positive results [270]. Using only the two fish 3'- UTR alignments, we predicted for 438 of our miRNA candidates potential mRNA targets with a median of 47 putative targets per miRNA. With our

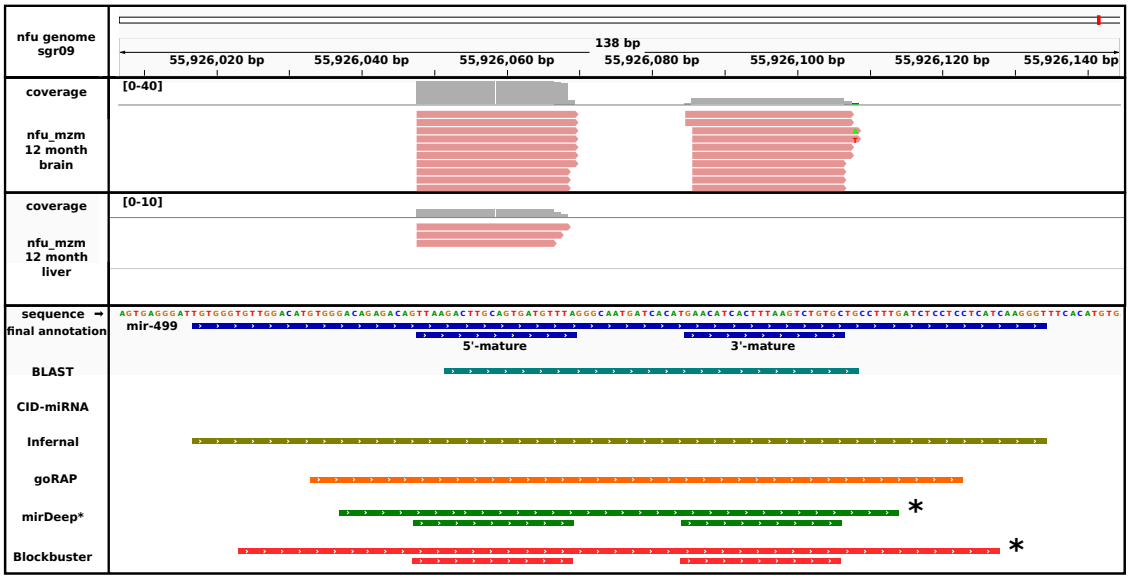


Figure 4.4: **Annotation, expression profiles and prediction comparison for miR-499.** We annotated the pre-miR-499 on sgr09, position 55,926,017–55,926,134 and the two mature miRNAs at 55,926,048–55,926,069 and 55,926,085–55,926,106. The six methods used for miRNA detection are displayed, CID-miRNA was not able to detect this miRNA. Tools working independent of the small RNA-Seq data (`BLAST` (cyan), `Infernal` (blue green) and `goRAP` (orange) vary in their annotation length. The latter two programs are based on covariance models, identifying mostly the complete pre-miRNA. The remaining two programs `miRDeep*` and `Blockbuster` are based on small RNA-Seq data (*) and therefore accurately annotate the mature miRNAs. MiR-499 is expressed weakly within *Nothobranchius furzeri MZM* 12 month liver library and therefore could not be detected by `miRDeep*` and `Blockbuster`. In the *Nothobranchius furzeri MZM* 12 month brain library, miR-499 was expressed high enough to be detected by both programs.

more conservative approach, still 419 miRNA candidates showed targeting potential with a median of 25 putative targets per miRNA.

To further examine these potential targets, we calculated enrichment scores of miRNA binding sites in already known sets of down-regulated genes in the brain of *Nothobranchius furzeri* during aging [271] and in different tissues between young and very old *Nothobranchius furzeri* individuals [123]. In the first study, both clusters, containing genes with decreasing activity during aging, show a significant enrichment of miRNA targets (cluster1: $p = 8.67^{-25}$; cluster5: $p = 1.78^{-5}$). For all three investigated tissues in the second study, we also found a significant enrichment of miRNA target sites within the down-regulated genes (brain: $p = 6.19^{-32}$; liver: $p = 7.72^{-17}$; skin: $p = 1.49^{-9}$).

Additionally, we identified single miRNA candidates, whose targets were enriched in any of the above mentioned gene sets (for details, see online supplement section *miRNA target prediction*). We found *e.g.*, *miR-10*, *miR-29* and *miR-92* showing potential to be significantly involved in the down-regulation of genes in the aging brain of *Nothobranchius furzeri*, like cell cycle regulators (*ccne2* [272], *nek6* [273],

Table 4.3: **Total number of genes known in *Nothobranchius furzeri* and number of ncRNAs covered by small RNA-Seq reads and percental distribution of reads for brain, skin and liver of *Nothobranchius furzeri*.** See STable 1 for more details and Tab. 4.1 for details about the read libraries. Asterisks indicates annotation of protein-coding genes from NCBI. ncRNA – all ncRNAs except miRNA, tRNA and rRNA; none – reads mapping to non-annotated genomic areas.

	Total		Brain		Skin		Liver
miRNA	754	470	88.81 %	438	75.01 %	368	58.10 %
tRNA	570	293	0.35 %	361	1.33 %	361	4.27 %
rRNA	48	37	1.73 %	43	3.61 %	43	6.97 %
ncRNA	250	136	0.09 %	217	0.15 %	204	0.41 %
protein*	22,627	0	6.60 %	0	4.59 %	0	6.35 %
none			2.42 %		15.31 %		23.90 %
Sum	24,125	936	100 %	1059	100 %	976	100 %

cdk13 [274]) or cancer related genes (*mycn* [275], *vav2* [276]), both processes involved in aging. A more detailed analysis of the potential regulatory role of miRNAs in aging follows in Section 4.3 and 4.5.

4.2.4 Effects of tissue and age on global miRNA expression

We used principal component analysis (PCA) to visualize the effects of tissue type and age on the global miRNA expression (see Fig. 4.5). A strong component of tissue-specific expression was detected and samples corresponding to different tissues clustered tightly and widely apart in the plane defined by the first two principal component axes (collectively accounting for 77 % of variance). Remarkably, the third principal component axis (3 % of variance explained) identifies an age-dependent component of miRNA expression that is common to all three tissues with the youngest samples (5 weeks), clearly separated from the rest. A detailed analysis of age- and tissue-dependent miRNA expression, including embryonic development, will be part of a separate publication.

4.2.5 miRNA expression comparison to closely related killifish

To compare and validate the miRNA composition in *Nothobranchius furzeri*, we created for each of the six related killifish species two small RNA-Seq libraries (see Tab. 4). These libraries were mapped simultaneously on all available miRBase (release 21) sequences and our annotated miRNAs of *Nothobranchius furzeri* to observe *Nothobranchius furzeri* miRNA candidates expressed in other killifish and conserved miRNAs possibly missing in *Nothobranchius furzeri* but not in the closely related

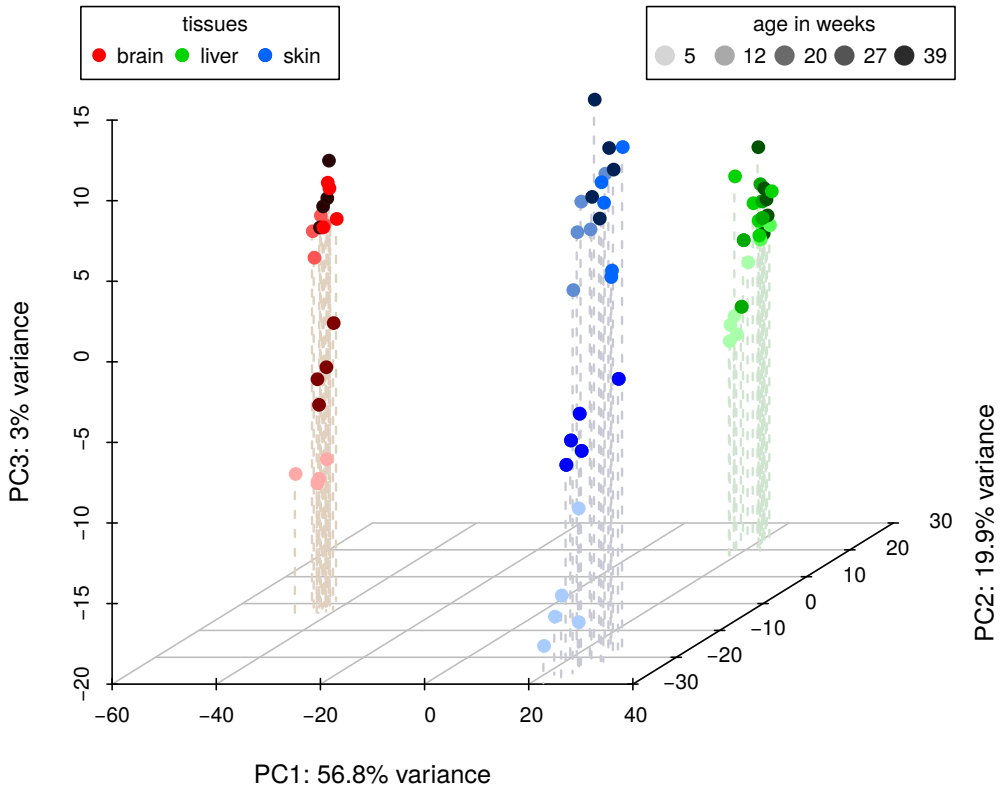


Figure 4.5: **Principal component analysis of the investigated small RNA-Seq samples.** A three-dimensional PCA plot of the *Nothobranchius furzeri MZM* small RNA-Seq libraries of all three tissues (brain – red, liver – green, blue – skin) and all investigated ages (from light to dark: 5, 12, 20, 27, 39 weeks). Whereas the samples cluster well according to their tissue belongings, a distinct separation regarding the ages can only be observed for the youngest samples in each tissue. A PCA plot of the *GRZ* strain, can be found in STable 2.

species.
In total, 546 (93.7 %) of the 583 expressed and 17 (9.9 %) of the 171 non-expressed miRNA candidates in *Nothobranchius furzeri* showed expression in at least one of the related killifish (Fig. 4.7 shows a miRNA not expressed in *Nothobranchius furzeri* but in several of the other killifish). Of these expressed miRNAs, 299 belong to the 421 non-conserved *Nothobranchius furzeri* miRNA genes.
To investigate whether miRNA sequences reflect known phylogenetic relationships, we concatenated the sequences of all expressed miRNAs and constructed a phylogenetic tree. This tree perfectly reflected the evolution of the *Nothobranchius* lineage [277]. It is also interesting that the number of *Nothobranchius furzeri* miRNAs expressed in other killifish species (indicated above the branch in Figure 4.6) is inversely correlated to the evolutionary distance, *i.e.*, this number is higher for killifish the closer they are related to *Nothobranchius furzeri*.
Aphyosemion striatum, *Nothobranchius korthausae*, *Nothobranchius pienaari*, *Nothobranchius rachovii*, *Nothobranchius kunthae* and *Nothobranchius kadleci* showed expression for 352, 428, 488, 473, 496 and 534 miRNAs, respectively.

Most of these expressed miRNAs (> 89 %) are among the 333 conserved miRNAs of *Nothobranchius furzeri* (see Supplement Tab. 3). The composition of expressed miRNAs from the six killifish varies only marginally. The *Nothobranchius* species (except *Nothobranchius furzeri*) had in total 395 expressed miRNAs in common (of which 148 are non-conserved), and *Aphyosemion striatum* expressed 324 of them (of which 116 are non-conserved). These 324 miRNAs represent a core of miRNAs from *Nothobranchiidae*, whose origin predates the emergence of annualism in this clade.

4.2.6 Identified miRNA clusters and gene duplications

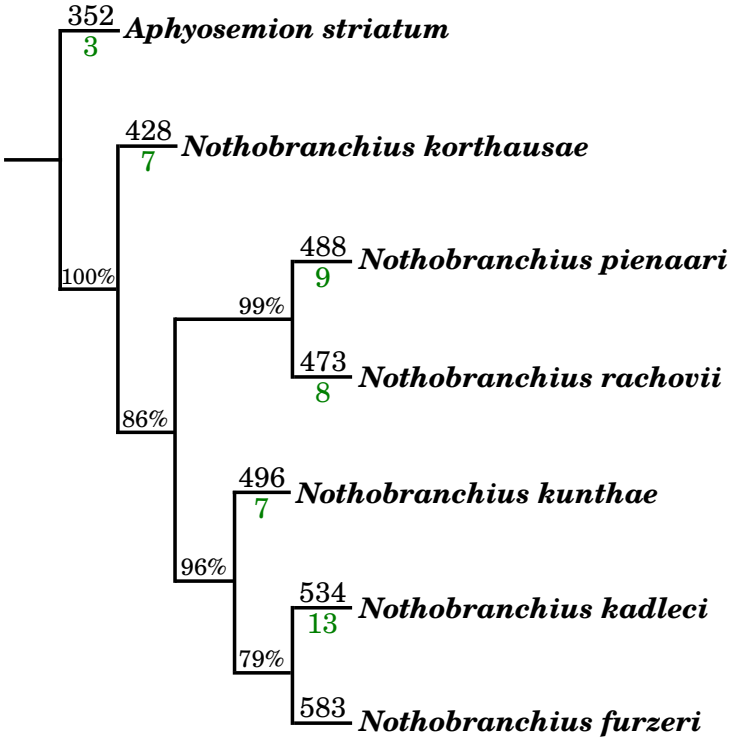


Figure 4.6: **Phylogentic tree based on miRNA expression.** Killifish phylogeny based on the expressed miRNAs calculated via hierarchical clustering using the `R` package `pvclust` [278]. Bootstrap values are given as percentages at the corresponding branches. The amount of identified expressed miRNAs is given next to the species names. The numbers in green indicate the number of these expressed miRNAs, which were annotated but not expressed in any of the sequenced *Nothobranchius furzeri* samples.

MiRNAs are known to often occur in clusters [254]. We define a miRNA cluster to consist of at least two miRNAs, with a maximum distance of 10 kb. Examining the localization and distances of the miRNA genes in the five fish species with assembled genomes, we identified 83, 96, 58, 68 and 59 different clusters in *Nothobranchius furzeri*, *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus* and *Takifugu rubripes*, respectively (see Tab. 4.4, Fig. 4.8-A).

Table 4.4: **Amount of annotated miRNAs, identified miRNA clusters and the number of miRNAs in clusters, as well as known conserved and non-conserved miRNA families in *Nothobranchius furzeri* (Nfu), *Danio rerio* (Dre), *Oryzias latipes* (Ola), *Gasterosteus aculeatus* (Gac) and *Takifugu rubripes* (Fru).** In brackets, the amount of miRNAs associated with the identified miRNA families are given. For detailed lists of miRNA family assignments, see STable 4.

Species	#miRNAs	#miRNA clusters	#miRNAs in clusters	#mirBase families	#unknown families
nfu	754	83	213	94 (333)	383 (421)
dre	765	96	305	99 (307)	302 (458)
ola	366	58	151	104 (269)	55 (97)
gac	504	68	299	102 (413)	63 (91)
fru	337	59	143	99 (278)	51 (59)

In all investigated fish species but *Takifugu rubripes*, the largest cluster is the *mir-430* cluster (containing 7 to 55 miRNAs; see Fig. 4.8-C). This cluster is extremely divergent and evolving relatively quickly in each lineage.

Not only the number of miR-430 copies within each cluster varies greatly but also the number and organization of the members of this miRNA family. Whereas miR-430a and miR-430c can be found in all five fish species, miR-430b and miR-430d seem to occur only in *Danio rerio* and *Oryzias latipes*, respectively. Additionally, no structural similarities or shared repetition patterns can be observed for this miRNA cluster, which is an additional indication of the low purifying selection on this specific gene cluster. However, a clear duplication pattern can be observed for the miR-430 cluster in *Danio rerio* (the order miR-430c/b/a is repeated with only a few exceptions) and *Nothobranchius furzeri* (the order miR-430c/a/a/c/a/a/a is repeated). For *Oryzias latipes* and *Gasterosteus aculeatus*, the order of miR-430 variants appears to be more random, and *Takifugu rubripes* has too few copies to show any repeated pattern.

Fig. 4.8-B depicting the miR-17-92 cluster shows an example of the other extreme: in all five investigated fish species, two perfectly conserved clusters can be found. These represent a duplication of an ancestral cluster present in all vertebrates, and the order of the different members is perfectly conserved. It is known that the miR-17-92 cluster is transcribed polycystronically and acts in oncogenic and tumor suppressor pathways [279, 280]. Furthermore, up to two smaller and lesser conserved clusters, containing at least two miRNAs of the miR-17 or miR-92 family, were identified per fish species, similar to what is known for mammals. Having correctly identified this highly conserved cluster in *Nothobranchius furzeri* is again good evidence for the high quality of its newly assembled genome and completeness of our miRNA catalogue.

Another example for an evolutionarily conserved miRNA cluster is the miR-29 cluster depicted in Fig. 4.8-D. Mir-29 family members are up-regulated during aging in

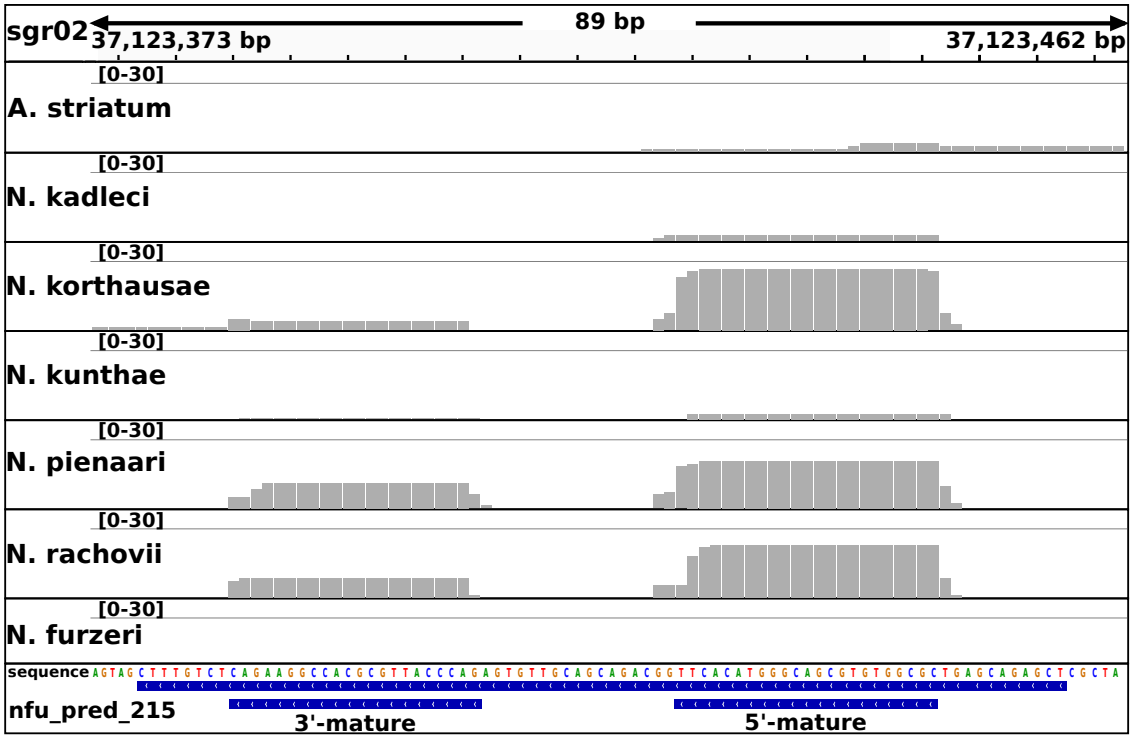


Figure 4.7: **Expression profiles of the predicted miR-215.** Gray bars indicate the amount of aligned reads and therefore coverage at the specific positions. Whereas no expression can be observed for this miRNA in *Nothobranchius furzeri*, clear activation can be seen in *Nothobranchius korthausae*, *Nothobranchius pienaari* and *Nothobranchius rachovii*. *Aphyosemion striatum*, *Nothobranchius kadleci* and *Nothobranchius kunthae* show a weak expression for at least the 5'-mature variant of this miRNA.

a variety of different tissues including muscle, skin, brain and aorta [120, 281–283] and appear to be key regulators of age-dependent gene expression [284]. This cluster consists of miR-29a (which is identical to the mammalian miR-29c) and its variant miR-29b and is duplicated at least once. In some fish species, an additional variant miR-29c is known, which is identical to the miR-29a in mammals, with one nucleotide being different outside the seed region [285]. As from RFAM (version 12.1) and miRBase (release 21), miR-29 genes are mainly identified in vertebrates as well as one *Hemichordata* and one *Arthropoda*, so we can only speculate that the original cluster duplication event arose in the early metazoa lineage. In *Oryzias latipes* and *Takifugu rubripes*, both miR-29 clusters are still present, whereas *Danio rerio* appears to has lost one copy of the miR-29a gene. For *Gasterosteus aculeatus*, we were only able to identify one miR-29 cluster. However, because its genome assembly is incomplete, we assume that the second cluster may not be lost but is missing in the current version of its miRNA annotation. Interestingly, in *Nothobranchius furzeri*, we identified an additional miR-29a/b pair and a fourth single copy of miR-29b. Assuming a complete genome assembly, different scenarios could explain this finding: (1) both original miR-29 clusters were individually duplicated once more, and the fourth miR-29a gene was later lost, (2) one of the two clusters was duplicated as a

whole, whereas in the other only miR-29b was copied or (3) both original clusters were duplicated during the same event, and again one of the miR-29a genes was later lost.
About the same amount of different miRBase miRNA families could be identified for all five fish species, despite their big differences in the number of identified miRNA genes. All miRNA genes not matching any known mirRBase family were clustered based on their sequence identity in order to estimate the amount of miRNA 'families' not covered by the miRBase database (see Tab. 4.4 and Supplement Tab. 4).

4.2.7 Completeness of the *Nothobranchius furzeri* miRNA annotation

This study involved a multitude of small RNA-Seq libraries from several tissues, ages, strains and embryos of *Nothobranchius furzeri* and closely related species. The aim was the characterization of the *Nothobranchius furzeri* miRNome and a detailed annotation in the recently published genome [123]. The inclusion of other killifish species allowed us to analyze the occurrence of novel miRNAs in the group of annual fish. A total of 421 putatively novel miRNAs that have no known homologous genes in other non-killifish species may sound a lot. However, given the fact that between the roughly known 1800 human and 2200 mouse miRNAs only some hundred miRNA families are shared (besides these two mammalian species being relatively closely related), these findings show that there appears to be a comparably large set of species-specific miRNA genes next to an established and evolutionary conserved set of miRNAs. This and the fact that the regulatory function realized of a single miRNA family that does not exist in one species can be compensated by another miRNA, targeting the same mRNAs as the missing one, shows the flexibility of this regulatory mechanism. Since simple single nucleotide mutations can completely change the targets and therefore the functions of an miRNA (for details about biological function of miRNAs refer to the Section 4.1) they are more easily prone to selective pressure but also much more adaptive than protein based regulation like transcription factors.
Due to the fact that we identified roughly the same number of miRNAs in *Nothobranchius furzeri* as known in *Danio rerio* and both fish species share almost equal amounts of miRBase families and unknown miRNA families, we assume that our miRNA catalogue is comparable to the one of the model organism *Danio rerio*.

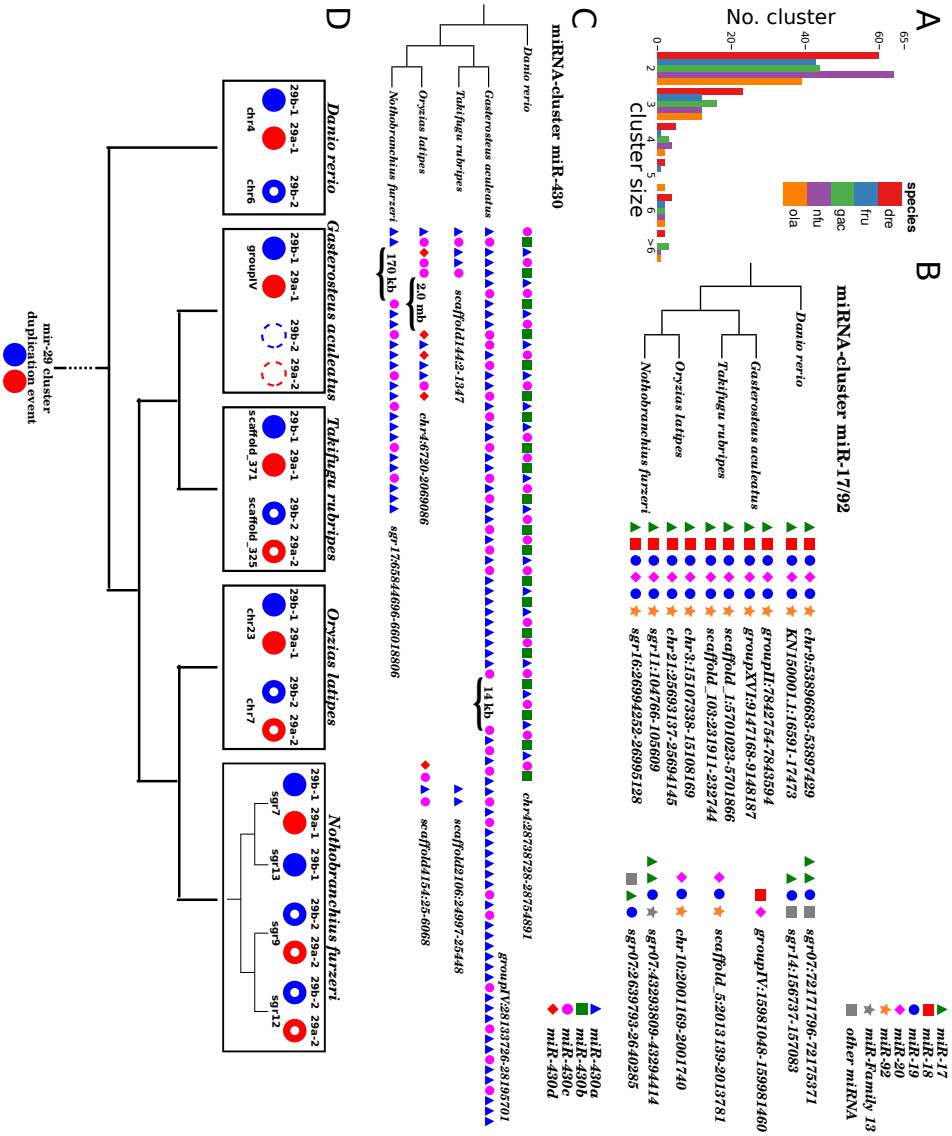


Figure 4.8: **miRNA cluster comparison between fish.** (A) Amount of clusters and their respective sizes with a maximum distance of 10,000 bp between two miRNAs. (nfu - *Nothobranchius furzeri*, dre - *Danio rerio*, ola - *Oryzias latipes*, gac - *Gasterosteus aculeatus*, tru - *Takifugu rubripes*) (B) Structure comparison of the miR-17/92 cluster. Two highly conserved clusters could be identified for each species, as well as some smaller less conserved clusters, containing at least two miRNAs of the miR-17/92 cluster. (C) Structure comparison of the miR-430 cluster. No structural similarity between the different species can be observed. However, *Danio rerio*, *Gasterosteus aculeatus* and *Nothobranchius furzeri* show some distinct but individual repeating pattern. Even though the gene variants miR-430b and miR-430d seem to be unique to *Danio rerio* and *Oryzias latipes*, they can be clearly distinguished, based on sequence alignments. (D) After the ancestral duplication event, the mir-29 cluster is distinguished in the mir-29a/b-1 (filled red and blue dots) and the mir-29a/b-2 cluster (red and blue circles). Whereas for *Danio rerio* the mir-29a-2 gene appears to be lost, we assume that for *Gasterosteus aculeatus* the whole second mir-29 cluster (dashed circles) is only missing, because of the low quality genome sequencing and assembly. In *Nothobranchius furzeri* we observe an additional copy for each of the two clusters, except that the mir-29a/b-1 pair was only partially duplicated or the second mir-29a-1 gene was lost again.

4.3 Diapause regulation in *Nothobranchius furzeri* and its implication to delayed aging

This chapter's section is based on the publication "Analysis of microRNA expression reveals convergent evolution of the molecular control of diapause in annual fishes" [2].

Data used in this section

In total 19 small RNA-Seq libraries from 9 different annual and non-annual killifish were used for the analysis of diapause specific miRNAs. Library generation was performed by Mario Baumgart and Alessandro Cellerino [3], following the same procedure as described in Chapter 2.1. RNA-Seq data analysis was performed as described in Chapter 2.2. The killifish used were *Nothobranchius furzeri*, *Aphyosemion striatum*, *Callopanchax occidentialis*, *Epiplatys dageti*, *Aplocheylus lineatus*, *Nematolebias papilliferous*, *Austrofundulus leoholgnei* and *Rivulus cylindraceus*. For more details regarding these species and RNA-Seq libraries, see Figure 4.9 and SData 1. A list of aging-related differentially expressed protein-coding RNAs of *Nothobanchius furzeri* was obtained from the study of Reichwald *et al.* [123], to correlate them with the identified differentially expressed miRNA genes (see SData 7). For the overlap comparison with aging-related miRNAs, the whole *Nothobranchius furzeri MZM* RNA-Seq data of the JenAge collection was used, including the three tissues brain, liver and skin (see Chapter 2).

4.3.1 Diapause: oversleeping aging

Life-history is a major force in the evolutionary shaping of embryonic development. Extreme examples are seasonal species where embryos undergo diapause, a suspension of development or dormancy state, as it is the case for the larvae of many insect species from temperate climate. Seasonal life cycle has evolved also in a clade of teleost fishes (suborder Aplocheiloidei) known as annual killifishes. Annual fish inhabit ephemeral bodies of water and are adapted to the alternation of wet- and dry-season in Africa and South America. All adult fish die when their habitat dries out and conservation of the species is ensured by desiccation-resistant eggs that enter in diapause and remain encased in the dry mud until the next rainy season. Annual life history is present in both, South American and African killifishes, and is the result of a series of independent events.
Early studies suggested the existence of four events of loss and re-gain of an ancestral annual trait [286, 287] while more recent studies suggest that it repeatedly evolved, at least three times in Africa and three times in South America [288]. In either of the two scenarios, an annual clade has always a sister non-annual clade that is phylogenetically closer than any another annual clade.
The early development of annual fish is conserved in the different lineages [288–290] and is characterized by three points of developmental arrest. The major point of de-

[2] The complete supplemental material is available at `http://www.rna.uni-jena.de/supplements/diapause/`

[3] FLI Leibniz Institute for Age Research, Beutenbergstrasse 11, 07745 Jena, Germany

velopmental arrest occurs after formation of the embryonic axis and organogenesis, at mid-somitogenesis, and is called diapause II (DII). DII is a facultative stage: it can be skipped when embryos are incubated at high temperature [288, 291], but lower temperature, darkness or dehydration (all conditions occurring in natural habitats) induce DII [292, 293]. The duration of DII is highly variable and the embryos can remain in this stage for several months [294] or even years (unpublished observation; personal communication with Alessandro Cellerino). The physiological and molecular mechanisms of diapause were studied in detail in the South-American species, *Austrofundulus limnaeus*. Diapause is characterized by drastic depression of protein synthesis, oxygen consumption and of mitochondrial respiration associated with G1 arrest of the cell-cycle [295, 296]. These basic mechanisms seem to be conserved also in the African annual genus *Nothobranchius* [288, 297].

Studies in the roundworm *Caenorhabditis elegans* have drawn a connection between diapause and aging. This nematode can enter a stage of dormancy called "dauer" when the environmental conditions are unfavorable. Some genetic mutations that influence dauer formation also modulate longevity. In particular, the daf-2 mutation that affects an ortholog of the IGF/insulin receptor, increases lifespan over two-fold. Strikingly, the influence of the IGF/insulin pathway on longevity is conserved also in vertebrates and humans [298, 299]. In addition, the gene expression profile in the dauer larvae stage show high similarities to the expression profile of long-lived adult mutants [300] and in annual fish, the expression of protein-coding RNAs during diapause and aging overlaps with respect to genes controlling cell cycle and mRNA translation [123]. Also small non-coding RNAs are embedded in the genetic network that links diapause and longevity, as exemplified by miR-71. This miRNA is a longevity gene and an aging biomarker in *Caenorhabditis elegans* and is also essential for diapause [301, 302]. These results prompted us to investigate a possible overlap in miRNA regulation between aging and diapause in a vertebrate clade, here on the example of the short-lived fish *Nothobranchius furzeri*. As previously mentioned in the last section, this small annual fish has a captive life-span of a few months and is the vertebrate with the shortest life-span that can be cultured in captivity. For these reasons, *Nothobranchius furzeri* has emerged as model organism of choice for biological investigations into aging [118, 271, 303].

Next generation sequencing (NGS) techniques can be used to identify and at the same time quantify miRNAs in non-model species taking advantage of the small size and extremely high conservation of miRNA sequences [304, 305]. Using this technique, we described an evolutionarily conserved miRNA signature of aging in the brain of *Nothobranchius furzeri* [120]. More recently, based on the available genome sequence and a large database of small RNA sequencing, we published a catalog of both conserved and non-conserved microRNAs in *Nothobranchius furzeri* [121]. A possible function of the non-conserved miRNAs in this species may be to control diapause. Within this section, we characterize this miRNA expression signature of DII of both conserved and non-conserved miRNAs and compare this signature with miRNA regulation observed during aging. To characterize an evolutionarily-conserved expression signature of DII, we compared miRNA expression in embryos at mid-somitogenesis in pairs of annual- and non-annual species from three independent evolutionary lineages of *Aplocheloidei*. This approach is complementary to a recently

published study where small RNA expression was analyzed in embryos of the South-American annual killifish *Austrofundulus limnaeus* at different stages of development and in embryos that skipped diapause and proceeded to direct development [306].

4.3.2 Phylogenetic sampling of the analyzed killifish

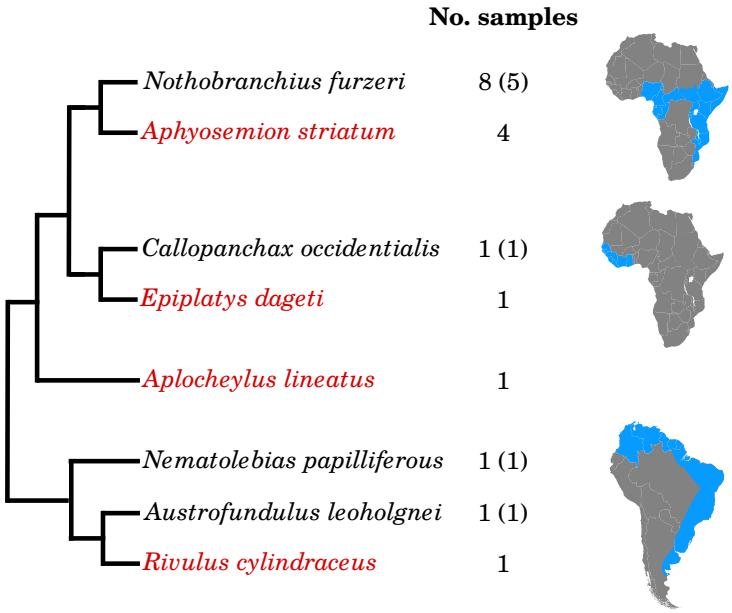


Figure 4.9: **Overview of the sequenced killifish embryos and the respective amount of samples.** Fish marked in red are non-annual species, whereas all others are annual. For the annual species the numbers of samples sequenced during diapause are given in brackets. *Aplocheylus lineatus* originates from India. For detailed information, see STable 1.

Annual killifish are divided into three clades distributed in Africa (one West- and one East- of the Dahomey gap) and in South America (Fig. 4.9). In each of these regions, the clade contains both annual- and non-annual genera. The position of the non-annual genus *Aplocheilus* is debated. Early analysis considered it basal to all annual species [286], but it was later suggested to be nested between African- and South American species [288]. Therefore, *Aplocheloidei* offer a unique opportunity to study parallel evolution of life-history adaptations (see also [290]). We collected eggs in diapause II from annual species or in the corresponding developmental stage (mid-somitogenesis) from non-annual species of all three clades in addition to *Aplocheilus lineatus* and analyzed miRNA expression by small RNA-Seq. From South America, we analyzed miRNA expression in the annual species *Austrofundulus leohoignei* and *Nematolebias papiliferous* and the related non-annual species *Rivulus cylindraceus*. From Africa, West of Dahomey gap, we analyzed the annual species *Callopanchax occidentalis* and the non-annual species *Epiplatys dageti monroviae*. From Africa, East of Dahomey gap, we analyzed the annual species *Nothobranchius furzeri* (5 replicates), and the non-annual species *Aphyosemion striatum* (4 replicates). Finally, we incubated eggs from *Nothobranchius furzeri* (3 replicates)

at higher temperature, a condition known to promote direct developement, *i.e.* diapause skipping [288, 307].

4.3.3 Differences in miRNA expression between annual and non-annual fish

Overall expression activity of conserved and killifish-specific miRNAs

First, we investigated the total number of miRNAs whose expression could be detected in all analyzed killifish species. As depicted in Fig. 4.10, the largest number of expressed conserved and killifish-specific miRNAs could be found in *Nothobranchius furzeri* and the closely related species *Aphyosemion striatum* where multiple replicates were available. A certain bias towards *Nothobranchius furzeri* and the more closely related species was expected, because the killifish miRNA annotation was based on the *Nothobranchius furzeri* reference genome [121]. Nevertheless, we observed the third highest activity in miRNA genes in the more distantly related *Austrofundulus leohoignei* and the lowest number of active miRNAs in *Callopanchax occidentalis*. Still, most of these genes showed only low (10 – 50 reads) or moderate expression (51 – 1000 reads). Taken all samples together, we detected expression activity for roughly 45 % of the miRNAs genes, predicted and described in Section 4.2. Interestingly, some of these miRNA genes being identified in the *Nothobranchius furzeri* genome but did not show active transcription in any of the *Nothobranchius furzeri* samples, were found to be expressed in other species, in particular in those not undergoing diapause.

Expression of miRNAs can separate fish based on their diapause status

Based on the global expression patterns of all miRNAs we have clustered the analyzed samples using principal component analysis (see Figure 4.11). We observed a clear separation not only between the individual species, but also between annual and non-annual embryos. Replicates of *Aphyosemion striatum* and *Nothobranchius furzeri* form defined clusters, whereas the three diapause skipped *Nothobranchius furzeri* embryos are only weakly clustered. Nevertheless, these results demonstrate that samples are divided by their physiological status and not by their phylogenetic relationships, indicating the existence of a convergent transcriptional program associated with diapause across species.

Conserved expression of developmental-related miRNAs

Next we examined the most highly expressed miRNAs of all investigated samples. Interestingly, four conserved miRNA families (miR-10a/b/c, miR-92a, miR-181a and miR-430a/b/c) and two killifish-specific miRNA families (currently named miR-19337 and miR-19344) were always the most highly expressed in all the samples, indicating their special role during the early development of the investigated fishes (see, Fig. 4.12 and STable 1). This is consistent with the report of Romney and Pobrasky [306], who also report these four conserved miRNA families to be the highest-expressed in embryos of *Austrofundulus limnaeus* at mid-somitogenesis. Of

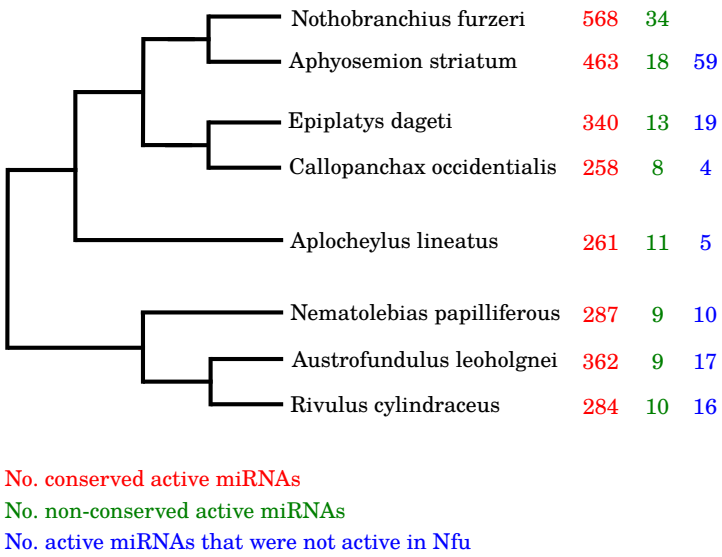


Figure 4.10: **Actively transcribed conserved and killifish-specific miRNAs.** Phylogenetic tree of the investigated killifish species and the number of their actively transcribed miRNA genes (red – miRNAs belonging a known miRNA family; green – miRNAs specific to killifish; blue – miRNAs not being transcribed in any of the *Nothobranchius furzeri* samples).

special interest is the high expression of the two killifish-specific miRNAs miR-19337 and miR-19344, because this could indicate that they potentially control some aspects of killifish embryonic development.
A strong expression of members of the miR-10 family is not unexpected since they are a known regulators of a number of *hox* genes, which are important transcription factors during embryonic development [308]. Indeed, expression of miR-10 strongly increases during somitogenesis in *Austrofundulus limnaeus* [306]. Additionally, it is also known that expression of miR-10 has a positive effect on the ribosomal protein synthesis machinery [309].
Members of the teleost-specific miR-430 family are also key mediators of embryogenesis, regulating multiple early development processes, such as mesendodermal fate specification, brain morphogenesis, silencing of maternal transcripts and development of primordial germ cells [310–313]. Usually miR-430 genes are produced by much larger genomic clusters whose organization differs between teleost species and in the previous section we have described the organization of such clusters (see Figure 4.8). Recent work in *Austrofundulus limnaeus* has shown that expression of miR-430 is downregulated during somitogenesis and is lower in embryos that escape diapause and proceed to direct development, underlining its relevance for the embryogenesis of teleosts [306].
Also miR-181 is necessary for embryonic development, and the deletion of all its paralogs induces embryonic lethality [314, 315]. In addition, it was shown to target another *hox* gene, namely *hox-a11* in mammals, and the homeobox transcription factor Prox1, therefore being an embryonic development regulator too and hinting

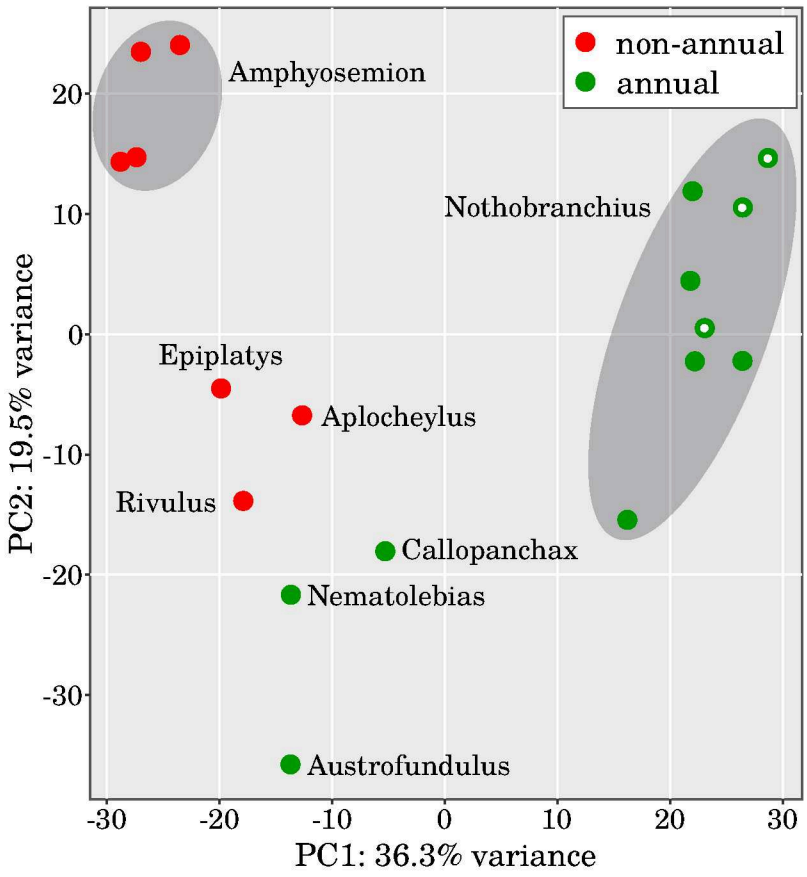


Figure 4.11: **Principal component analysis of the investigated small RNA samples.** PCA of the 18 RNA-Seq libraries based on the 10 most variant genes, *e.g.* showing the highest variance in expression. The libraries generated from annual and non-annual species are relatively well separated and samples from *Nothobranchius furzeri* and *Aphyosemion striatum* cluster together, respectively. A more detailed PCA plot is available, see SFigure 1.

at a potential similar function in fish embryos [316, 317].
Being part of the miR-17/92 cluster, the miR-92 is an important well-studied regulator of the cell cycle [318] and inferring from its strong expression in our samples, it could be especially important during developmental processes.
For both killifish-specific miRNAs, only predicted target mRNAs in *Nothobranchius furzeri* are known [121]. However, TSC2 and FAM83D are both putative targets of miR-19344. TSC2 is of particular interest as it is part of the TSC complex, which is a major upstream regulator of the activity of the mTORC1 complex, representing a central regulatory hub that integrates nutrient sensing and growth-factor signaling to regulate cell growth, protein synthesis and cell proliferation [319]. mTORC1 is also a known major regulator of aging and longevity in multiple organisms [320]. FAM83D, the other potential target of miR-19344, is involved in cell proliferation and motility [321].
Relevant potential targets of miR-19337 are CECR2, being part of a protein complex that regulates neurulation, STARD13B a known inhibitor of cell growth [322] and SRF (Serum response factor), which modulates the expression of many immediate

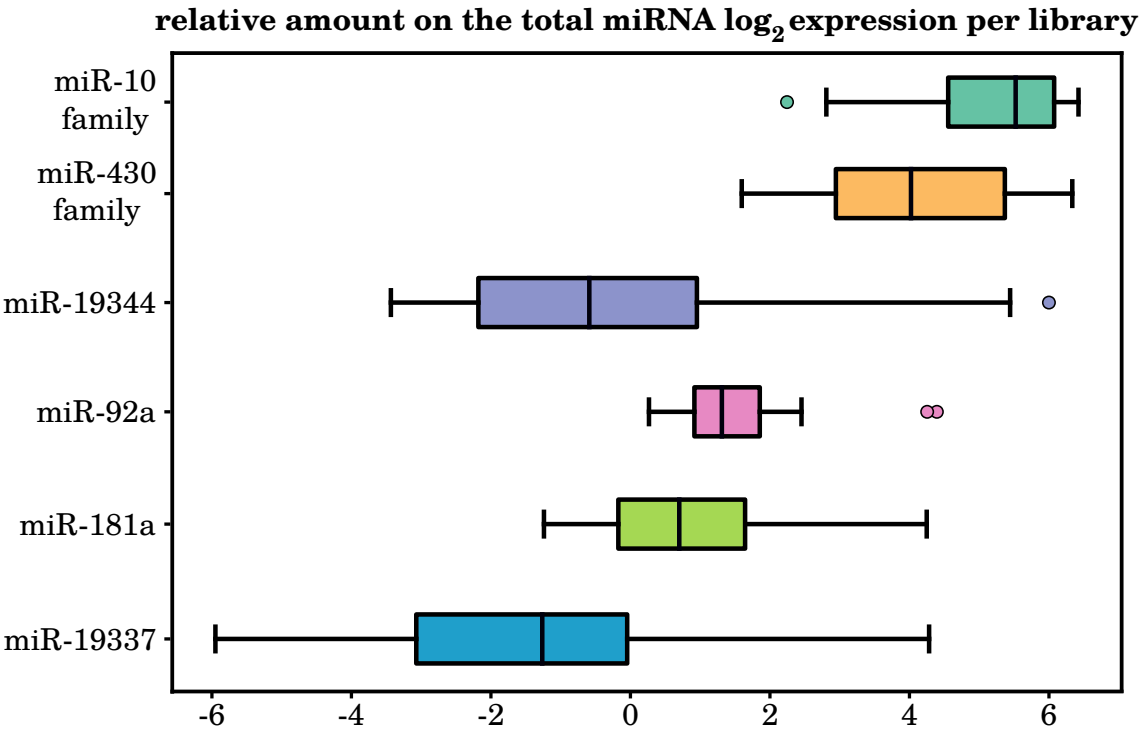


Figure 4.12: **Relative amounts on the total expressed reads within all the sequenced killifish samples of the six most highly expressed miRNA families.** All of these miRNAs are known or predicted to be implicated in the regulation of development processes in embryonic cells. For details on the miRNA expression levels, see STable 1

early genes and therefore is an important key player in embryonic development [323]. From all these examples, we can already observe a strong impact of the developmental stage of the sampled fishes onto the expression of their respective miRNomes (a neologism based on the term transcriptome), transcribing almost exclusively miRNAs which are regulating developmental or associated processes. Especially the dominating abundance of miR-10 and miR-430 genes in almost all samples, indicates the specialized roles of both miRNAs as well as their evolutionarily conserved regulatory importance for teleost embryogenesis.

Differentially expressed miRNAs related to diapause regulation in killifish

To identify differentially expressed mature miRNAs (DEMs) that could be linked to diapause regulation of annual species, we set up three different comparisons of the examined small RNA-Seq samples: (I) the "big annual comparison" contrasting all annual samples from fish that underwent diapause against the non-annual fish derived samples, (II) the "diapause comparison" contrasting the *Nothobranchius furzeri* samples that underwent diapause against those that skipped it and (III) the "small annual comparison" comparing the *Nothobranchius furzeri* diapause samples against the samples of its sister taxon the non-annual *Aphyosemion striatum*. As introduced in Section 4.1, a single pre-miRNA molecule is processed into a small miRNA duplex, which divides into two mature miRNAs after being unwind, named

5 prime (5p) and 3 prime (3p) mature sites, depending on their encoded site within the original pre-miRNA sequence. Both of these mature miRNAs, originating from the same gene that encoded the miRNA precursor, can be incorporated into a RISC complex and subsequently fulfill their silencing function. However, both mature miRNAs from one and the same precursor can target different mRNAs, thus regulating differing biological processes. Because of this, we have examined the mature miRNAs of single miRNA genes separately in the subsequent analyses.
The big annual comparison revealed 242 DEMs (122 upregulated and 120 downregulated in the annual fishes compared with the non-annual ones) belonging to 156 distinct miRNA genes (see SData 1). Some of the most significantly changing miRNAs in the annual comparison were members of the miR-430 family, belonging to the above mentioned highest expressed miRNA genes and being major regulators of developmental processes. Interestingly, we found that the particular miR-430a/c-3p mature miRNAs are is downregulated in all the annual species whereas its mature sibling miR-430a/c-5p is upregulated in the same species. This is striking since the 5p forms are expressed at much lower levels as the corresponding 3p and this is a prominent example of a miRNA isoform switch that we describe more in detail in the next subsection. All the mature miRNAs of miR-430a and miR-430c act in developmental processes, but targeting different mRNAs. Both 5p mature miRNAs are mainly involved in the development of the enteric nervous system but also the brain, having strong impact on the negative regulation of cell differentiation [271]. The 3p mature miRNAs act by regulating different cell signaling pathways, such as the thrombin and Rho protein pathways, but mainly by modulating the negative regulation of cell transcription from RNA polymerase II. In addition to miR-430, we identified other DEMs with important development-related functions, such as miR-19 being part of the miR-17/92 cluster and associated with endothelial cell differentiation in embryonic stem cells [324], miR-200 regulating the epithelial/mesenchymal transition during embryonic development [325] or miR-221 being involved in cell proliferation and angiogenesis [326, 327]. All of these miRNAs seem to be important for a well organized regulation of different developmental processes and may in part be responsible for the carefully orchestrated diapause regulation and embryogenesis in killifish.
The diapause comparison between the *Nothobranchius furzeri* samples in diapause and those that have skipped it, due to high temperature incubation, revealed 17 DEMs (10 up- and 7 downregulated in the diaupausing samples) belonging to 14 distinct miRNA genes (see SData 2). Observing only few differentially regulated mature miRNAs between the diapause and non-diapause embryos of *Nothobranchius furzeri* might indicate a more simple miRNA control layer of diapause at least within this species. Prominent is the upregulation of miR-9-3p and miR-29a-5p during diapause. These miRNAs are both implicated in either regulation of neuronal differentiation and brain development or being involved in cell proliferation processes, making them promising key regulators for diapause [120, 328]. Most interestingly, the killifish-specific miR-19344 shows a 10-fold upregulation within the diapause samples and as discussed already above one of its potential targets is TSC2, regulating the mTORC1 complex, which is a major modulator of cell growth, protein synthesis and cell proliferation [319]. This could designate miR-19344 as the first

killifish-specific miRNA involved in developmental processes and as a marker for diapause. Not much is known about the other DEMs or their absolute change in expression is relatively low: One study suggests that miR-7641 is involved during endothelial differentiation in embryonic stem cells [329] (this miRNA was differentially expressed in the annual comparison too), a recent study links miR-563 to the development of the spine *in vitro* [330] and expression of miR-210 appears to promote angiogenesis by inhibiting *efna3*, a suppressor of blood vessel formation [331]. Finally, within the small diapause comparison we observed a total of 292 DEMs (171 upregulated and 121 downregulated in the annual *Nothobranchius furzeri* samples compared with the non-annual *Aphyosemion striatum*) belonging to 203 distinct miRNA genes (see SData 3). Similar to the big annual comparison, we identified a great number of DEMs with both comparisons having almost all DEMs in common, including the same directional changes in expression (see SData 4). This includes the miR-430, miR-7641-3p, miR-10 and miR-17 mature miRNAs. Still, we identified unique DEMs belonging to 25 and 68 miRNA families in the big and small annual comparison, respectively. Those miRNAs unique to the *Nothobranchius furzeri* and *Aphyosemion striatum* comparison might be in part responsible for the difference in embryogenesis in their clade but not necessarily in the other annual/non-annual killifish clades, because annual life appears to have evolved independently several times [286–288].
In general, the regulatory involvement of miRNAs during the embryogenesis of killifish is extensive, whereas in contrast only few miRNAs seem to be involved in the maintenance of diapause in *Nothobranchius furzeri*, making these miRNAs critical regulators of this dormancy state.

mature miRNA switches: changes in the abundance of specific mature sites with significant consequences

Within our differential expression analysis, we observed some interesting expression patterns of differential miRNA expression. In particular, we found some miRNA genes that showed a "switch" in the expression of their main mature miRNA, *i.e.* not necessarily a change in the total transcribed amount of the pre-miRNA transcripts but in the subsequent processing machinery responsible for the maturing of miRNAs, changing that the relative abundance of the 5p and 3p mature products. In addition to the above described examples of the miR-430 family, we identified two more of such cases in the big (miR-200a and miR-181a) and four in the small annual comparison (miR-128, miR-130c, miR-217 and miR-223), for an example see Figure 4.13 and SFigure 3. We already briefly mentioned the regulatory targets of miR-200a and its potential role during diapause in the previous subsections. Also the miRNAs of the other switch cases show potential regulatory roles in different maturation processes: developmental regulation of T cells and the vascular system by miR-181a [317, 332], miR-217 shows strong expression in late stages of the fetal rat development and an expression disruption showed damaged lung tissue [333] and miR-223 is a well studied modulator in hematopoiesis and osteoclast differentiation [334, 335]. The functions of miR-128 are more linked to the formation and progression of various cancers, suggesting a general role in either cell development or proliferation, and it is known to inhibit telomerase activity [336–338]. Not much

is known about miR-130c, however, its paralog miR-130a appears to be present in hematopoietic progenitor cells but not in mature blood cells [339]. Besides their implications in certain developmental processes, all of these examples demonstrate that miRNA regulation is not just executed by the mere total expression of miRNA genes, but itself is controlled by the selective incorporation of the distinct mature sites into the RISC complex. This is an important observation, suggesting that miRNA expression data has to be modeled mathematically more specific to ensure correct interpretation (more on that topic in Section 3.5 of this chapter).

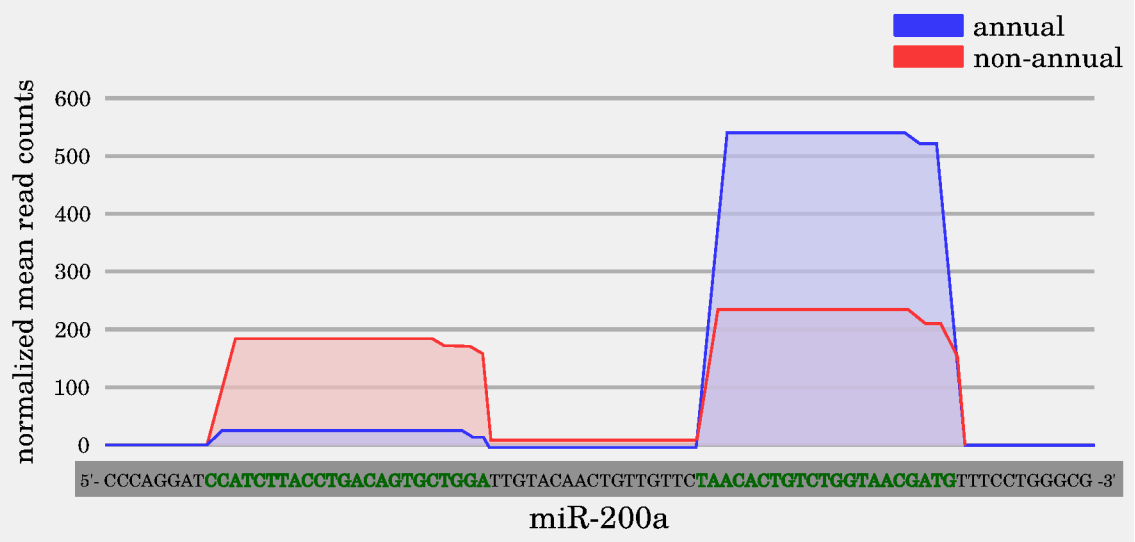


Figure 4.13: **Expression profiles of the miR-200a gene from the small annual comparison.** No significant change in expression can be observed when all reads of the pre-miR-200a annotation are compared among both conditions (p-Value=0.43), whereas a clear difference can be observed for the 5p and 3p mature miRNAs separately (p-Value of $2.0e^{-9}$ and $2.7e^{-4}$). In contrast to the non-annual species were both mature sites are equally expressed, the 3p mature miRNA is increased in expression whereas the 5p mature miRNA appears to be switched off. The green bases indicate the mature sequences. For more examples, see SFigure 3.

Strong correlation between the identified differentially expressed miRNAs and expression levels of developmental-related mRNAs

Next, we examined to which extend the results from the differential expression analysis of the miRNAs can be observed on the transcriptomic level. To do so, already identified aging-related differentially expressed genes [123] were compared with the targets of the identified DEMs. We found that 143 DEMs have 88 unique mRNA targets that also show a significant expression change within the small annual comparison (see SData 7). If we look at targets of some of the DEMs discussed above, we see a clear overlap to the identified differentially expressed mRNAs. Almost all of them are associated with cell developmental or proliferation processes and show opposite expression patterns compared with their targeting miRNAs. This indicates a strong correlation between the expression of the identified DEMs and

the regulation of their target mRNAs. Just as one example, cyclin D1 is one of the targets of the DEM miR-430a/c-3p, promoting G1 progression of the cell cycle [340] and was found differentially expressed.
Additionally, we tested if there was a specific enrichment of mRNA targets from a single DEM and could identify 23 miRNAs were this was the case. Some of the most significant ones were miR-23a-3p, miR-731-5p, miR-17a/b-3p, miR-430a-5p and miR-19b-5p, all of them discussed already above (see SData 7).

The expression of miRNAs involved in embryogenesis and diapause regulation is also modulated in an aging-related manner

We were also interested if there is a connection between the miRNA exercised regulation of early development and diapause processes and possible implications to aging. On the protein-coding RNA level an overlap between the expression of transcripts expressed during diapause and aging coding for cell cycle and translational control, was already observed [123]. These findings prompt that miRNA expression during diapause might also be linked to aging in *Nothobranchius furzeri*.
Indeed, we found not only miR-29a-5p to be similarly upregulated during diapause and aging but also its sibling mature miRNA miR-29a-3p within the aged brain, liver and skin samples of *Nothobranchius furzeri*. The miR-29a gene was already known to be upregulated in highly aged individuals and is also considered a tumor suppressor [120]. Another upregulated miRNA in both conditions was miR-210, which acts in the hypoxia pathway [341]. When active, this miRNA positively regulates the adaptation to a hypoxic environment, helping cells to survive with only low available oxygen [341]. Diapausing killifish embryos show drastic repression of aerobic metabolism [294, 295] and are extremely resistant to hypoxia [292], therefore, under normal oxidative conditions, hypoxia-sensitive genes may be downregulated. Hypoxia is also a known condition in aged tissues, indicating that miR-210 is generally used to respond to this oxidative stress [342]. The expression of miR-222a was observed to be equally and strongly downregulated during diapause and aging. When active, miR-222a promotes the expression of proteins responsible for muscle cell development by silencing the expression of the translational repressor *cpeb3* [343]. However, neither miR-9 nor the killifish-specific miR-19344 could be correlated with aging and seem to be specific for diapause and early developmental regulation. When including the DEMs of the small annual comparison, we observed a quite large overlap of 57 and 40 DEMs being up- and down-regulated in the same way during embyrogenesis and aging, as well as 99 DEMs being modulated in the opposite directions (see SData 6). Closer inspection revealed interesting behavior of some of the already discussed miRNAs. Both 3p mature transcripts of miR-430a/c are downregulated during diapause, but upregulated during aging in the brain. In the aged brain and liver of *Nothobranchius furzeri*, miR-10b-5p is downregulated, whereas it is upregulated in embryonic development. The same is true for miR-92b-3p in liver and skin. Also interesting, miR-181b-3p is down-regulated in both diapause and aging, but miR-181b-5p appears to be up-regulated during diapause but stays down-regulated in the aged liver and skin samples and not significantly

changed in brain during aging.
These exemplarily described behaviors of specific miRNAs already indicate that there is an overlap between aging processes and the regulatory network controlling diapause and embryogenesis in *Nothobranchius furzeri*.

However, to elucidate the precise mechanisms or regulatory responses that act in both conditions, more functional analyses need to be performed. Nevertheless, here we provide one of the first miRNA studies of vertebrate diapause. We could observe that regulation of miRNA expression in different clades of annual fishes shows convergent evolution and the samples of different taxa do not cluster according to their phylogenetic relationships, but according to their physiological status (being in diapause or not). This represents a further example of how annual fish belonging to different clades independently converged on similar adaptations to support embryonic development in their ephemeral habitat [288, 290]. Additionally, we describe different miRNAs to be key modulators not only during fish embryogenesis but also diapause regulation, such as members of the miR-10 and miR-430 family, miR-92, and miR-29a, but also killifish-specific ones, such as miR-19344 and miR-19337. Some of these miRNAs are of special interest, because their implications in specific developmental processes are already further studied in other species. For example, miRNA-430 was shown to act primarily by inhibiting protein synthesis [344]. However, after onset of zygotic transition and in adult cells mRNA degradation appears to be the predominant mechanism [345]. Diapause is characterized by a prominent depression of protein synthesis that is reduced to 10 % of the pre-diapause levels [346]. It is therefore highly likely that both proteins and mRNAs are stabilized during diapause. In this context, miRNAs should act primarily, if not exclusively, by inhibiting protein synthesis. Upon exit of diapause, release of miRNAs would allow immediate onset of translation. This concept is further supported by the observation that diapausing embryos of *Nothobranchius furzeri* show paradoxical up-regulation of genes related to translational elongation [123], suggesting that these embryos are primed for a catch-up process upon exit from diapause.

Studies in the worm *Caenorhabditis elegans* revealed miRNAs that are regulated both during diapause and aging. In particular, miR-71 does not influence diapause entry, but is necessary for survival during diapause in *Caenorhabditis elegans* [347]. Deletion of miR-71 results in shortened life-span and miR-71 abundance increases with age before dropping late in life [348]. The timing of the miR-71 expression drop can be used to predict the longevity of individual worms [301]. Our results show a similar overlap between miRNAs that are regulated during diapause and aging in the annual fish *Nothobranchius furzeri*. Several miRNAs significantly modulated in diapausing *Nothobranchius furzeri* were modulated during aging as well. One of these miRNAs, miR-101a, is particularly interesting. It was shown that miR-101 acts as a highly-connected hub in gene regulatory networks of transcription factors and epigenetic modulators involved in cell cycle progression [349]. Overexpression of miR-101 is also known to induce cell cycle arrest in different cell types [349]. So, high levels of miR-101 in diapausing embryos would contribute to the G1 block that is typical of diapause [296]. On the other hand, we observed downregulation of members of

the miR-17/92 cluster, also known as oncomiR-1. The organization of this cluster is conserved in *Nothobranchius furzeri* and it was observed to be downregulated during aging of the brain [120] and adult neuronal progenitors [257]. OncomiR-1 is a major regulator of cell cycle and overexpressed in several tumors [350]. On the other hand, it is well-established that oncomiR-1 is downregulated during aging of human mitotically-active cells where it targets the cell cycle inhibitor p21 [351]. An overlap in the regulation of cell cycle protein-coding genes between diapause and aging was also demonstrated by RNA-seq analysis in *Nothobranchius furzeri* [123], further supporting the concept that, like in *Caenorhabditis elegans*, aging and diapause are controlled by overlapping genetic pathways. The results of our study serve as a good basis for further research to infer the impact and precise control of miRNA regulation of these overlapping pathways.

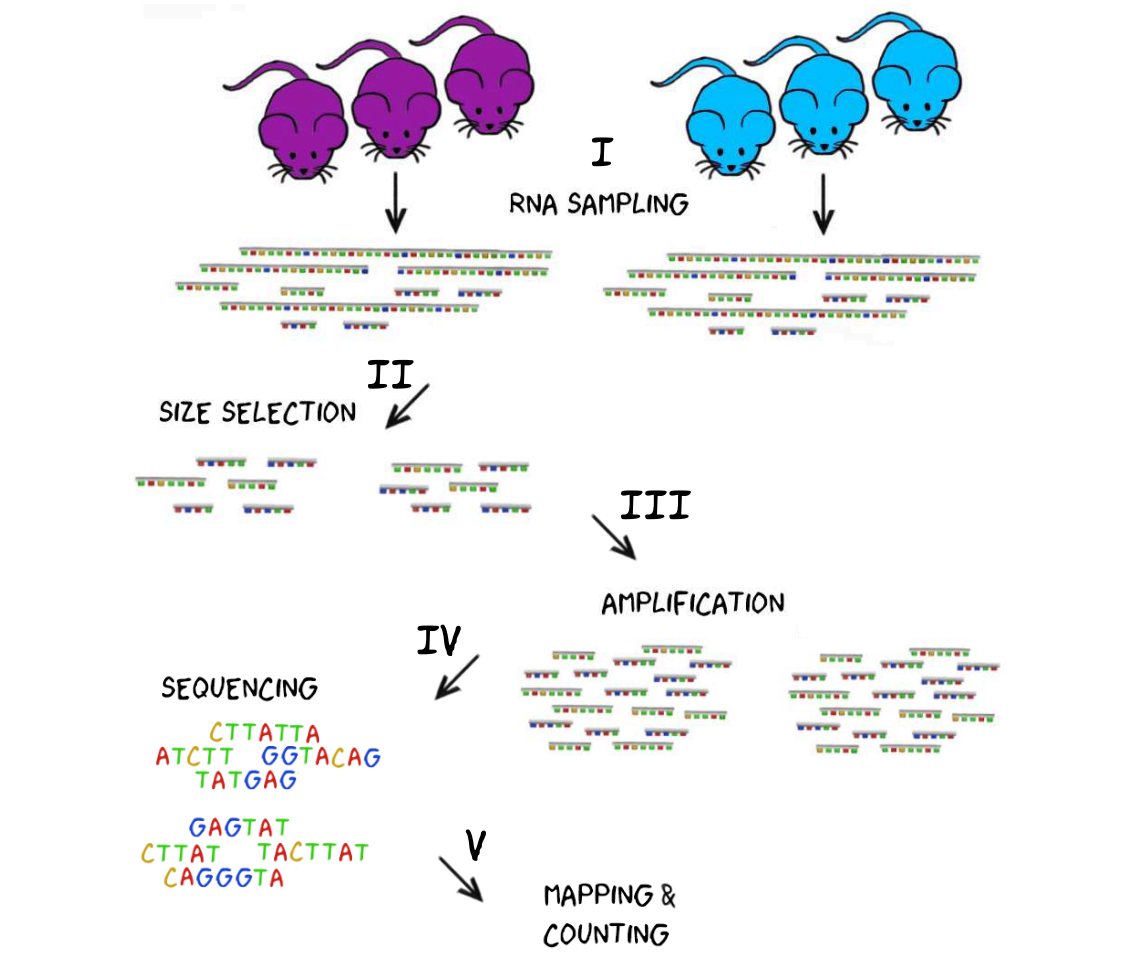
4.4 MeRDE: A new statistical model to infer differential expression of small RNAs from read counts

This chapter is based on the publication "Using Gamma Distributions to Model Small RNA-Seq expression data".[4]

4.4.1 The stochasticity of RNA-Seq expression data

In Section 1.2.2, we have already briefly discussed the topic of transcriptomic RNA-seq data. Because this section dives deeper into the theories and practices that stand behind statistical modeling of such transcription data, we will recapitulate the basics from the corresponding introductory section in greater detail.
An RNA-Seq experiment is used to quantify the RNA transcripts that are present at a certain time point within a cell, providing a quantitative snapshot of the active gene expression [96, 97]. It is performed in several main steps, resulting in a library of RNA molecule sequences, referred to as reads (see Figure 4.14). These read libraries represent the measured repertoire of RNA molecules and every read can (in theory) be assigned to a genomic feature (*e.g.*, gene, exon, intron, ...) based on its mapping to a specific location of the target genome. The summarized number of reads, belonging to a certain genomic feature (from now "gene" will be used synonymously for the term genomic feature) is called *read count* and was found to be in a direct linear relationship to the abundance of the respective gene's transcripts [352]. Thus, RNA-Seq experiments can be condensed into *count tables*, by read library mapping and subsequent assignment and counting of the mapped reads based on a target reference genome and annotation. Each count table then contains the name or ID of each examined gene with the assigned count value. In many cases, these counts are compared gene-wise between different conditions, such as healthy and diseased or young and old cells. Such techniques are known as *differential gene expression* analyses. To assess if a difference in the counts of a given gene between two conditions is significant, *i.e.*, the gene shows a greater or smaller expression than expected due to biological variation, statistical testing is needed. Therefore, replicates for both investigated conditions are required to estimate the natural variation (or *noise*) of gene expression. With large amounts of replicates, significant read count differences could be determined by statistical non-parametric rank based tests, such as the Mann-Whitney U test [353] or the Kruskal-Wallis test [354]. However, the generation of RNA-Seq libraries is still relatively expensive and time-consuming, making it necessary to get along with only a very small number of replicates per condition (as few as three replicates are not uncommon). In the lack of non-parametric approaches, the decision for a certain stochastic distribution is critical for modeling the underlaying biological variation of count data and consequently for the correct identification of expressional differences.
The two most common distributions which are assumed to model RNA-Seq read data best, are the Poisson distribution and the negative binomial distribution, following

[4]The complete supplemental material is available at `https://osf.io/wfjrs/`



	CONDITION 1				CONDITION 2			
	REPLICATE 1	REPLICATE 2	REPLICATE 3	REPLICATE 4	REPLICATE 1	REPLICATE 2	REPLICATE 3	REPLICATE 4
GENE A	10	9	4	9	3	7	10	7
GENE B	78	53	34	51	125	87	177	180
GENE C	64	117	85	106	67	112	95	66
GENE D	1429	1516	1174	1550	629	827	689	737
⋮		. . .				. . .		

Figure 4.14: **Simplified overview of generating read count data from small RNA-Seq expression data.** (I) Sample material is collected from two or more different biological conditions in several replicates each. (II) Besides other sequence library preparation steps (*e.g.*, rRNA degradation), RNA molecules are filtered by size to remove all transcripts with a length of about 200 nt or more. (III) Amplification of molecules by polymerase chain reaction to ensure sufficient transcript abundance for the sequencing step. (IV) Sequencing of the transcripts by the preferred NGS machine to obtain the read sequences. (V) Mapping of the reads to the target genome and counting the reads mapped to certain genomic features, such as miRNA genes.

the arguments that read counts follow multinomial distributions and that they represent discrete data making the use of discrete distributions useful [355–357]. Since Poisson distributions have only a single parameter describing their means and variances, they cannot handle overdispersion (*i.e.*, greater variance than mean) in data very well, which occurs often in biological count data [356, 358]. As a consequence, more elaborate negative binomial models were used, which are capable of modeling mean and variance with two independent parameters, making them less restrictive and more accurate than Poisson-based models [356, 358, 359]. Current state-of-the-art differential expression tools mainly utilize negative binomial distributions to model the noise of count data and calculate significant differences in gene expression [360].
However, within this section we show that for the special case of small RNA-Seq data (sequencing of the expressed miRNA instead of mRNA transcripts) the model of a negative binomial distribution can be replaced by a gamma distribution. We present a new statistical model to accurately infer expression differences within miRNA sequencing data and compare it to other state-of-the-art approaches. We can demonstrate the benefit of tailoring specialized statistical models instead of using more general models for specific classes of genes, such as miRNAs.

Small RNA-Seq data are more likely to follow gamma than negative binomial distribution

In 2016, Qin, Tuschl and Singer [357] suggested that miRNA sequencing data can be better modeled by gamma than Poisson or negative binomial distributions. They argue that most of the stochastic noise in the count data originates from the amplification step of the RNA molecules during the library preparation, rather than random natural variation (step three in Figure 4.14). The amplification step is performed by polymerase chain reaction (PCR) [95] to ensure sufficient DNA molecule abundance for the sequencing step. During each cycle of the PCR every molecule has a certain chance to be copied, which means that after multiple cycles the abundance of each molecule can develop into various different directions, depending on how often it was selected for duplication during the whole PCR [361]. This amplification step resembles a exponential distribution and since single genes are represented by multiple molecules, their total count after PCR is a mixture of exponentially distributed molecule counts, which can be modeled by a gamma distribution [357, 362]. However, the same argumentation cannot be used for normal RNA-Seq data, containing reads from mRNA transcripts or other longer molecules. Whereas in a small RNA-Seq library each sequence read usually represents exactly one transcript, several reads can originate from a single RNA transcript in a mRNA RNA-Seq library due to read size limitations (see Chapter 1.2.1). This results into a bias of selecting reads from longer transcripts more frequently during each PCR cycle, because they occur more often than reads from shorter transcripts with a similar abundance. This results in a multinomial distribution of read counts for every gene [356]
Qin *et al.* could show that their assumption holds true for small RNA-Seq data, based on a small set of technical replicates of miRNA sequencing data [357]. A technical RNA-Seq library replicate is obtained when the same sample is sequenced several times, thus, in contrast to biological replicates, that originate from different

samples of the same condition, cannot reflect natural variation [363]. A certain bias towards random variations of the technical procedure has to be expected in their data.

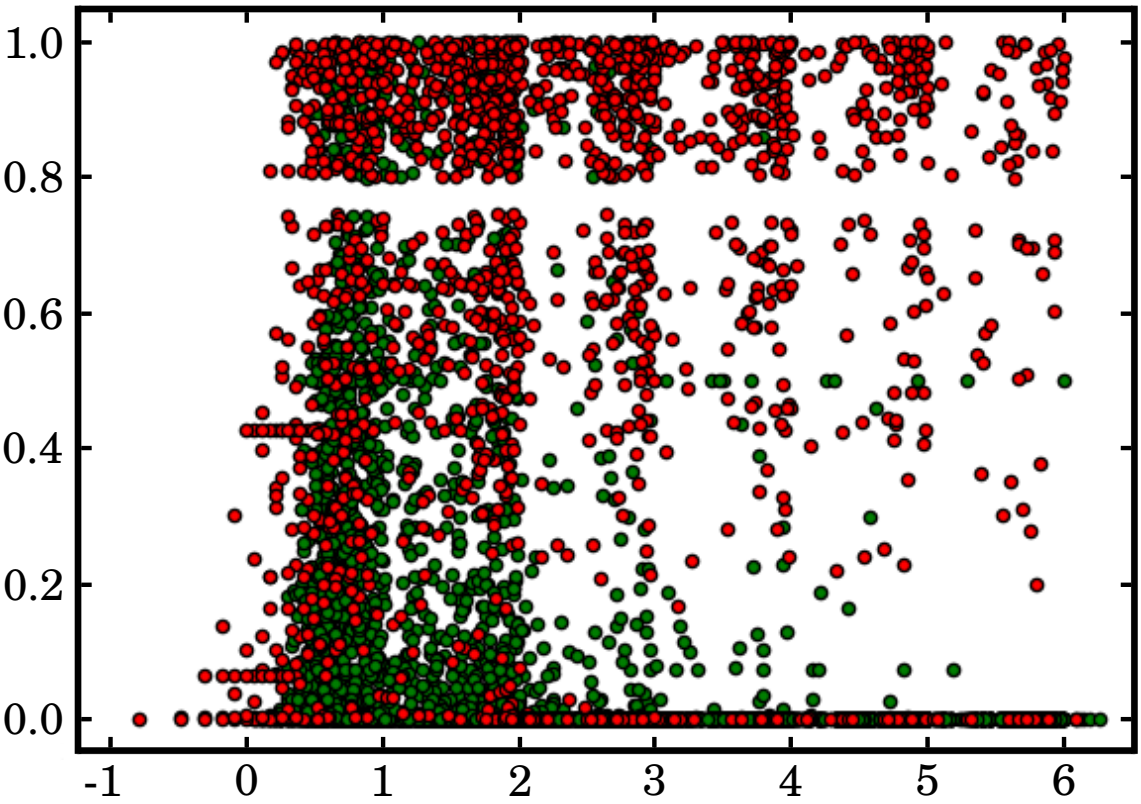


Figure 4.15: **Result of the Kolmogorov-Smirnov test on the miRNA samples of the JenAge RNA-seq dataset.** For every miRNA in all its respective conditions (species, tissue and age sample) the calculated goodness-of-fit p-Value to either follow a gamma distribution (green) or negative binomial distribution (red) was plotted against its log_{10} mean read count. As it can be observed, most miRNAs tend to follow a gamma distribution (for details see SData 1).

To validate the observation of Qin *et al.* we used the Kolmogorov-Smirnov test (KS-test) on the whole JenAge small RNA-Seq dataset, consisting of count data from over 4,100 miRNA genes of around 50 different conditions with four to eight biological replicates each (see Figure 2.2). The KS-test is a non-parametric test to compare a given sample (here count values belonging to biological replicates of single miRNA genes from single conditions) to a reference probability distribution [364]. It expresses the goodness-of-fit of the given sample to the reference distribution, *i.e.*, the likelihood that the sample originates from a population following the reference probability, in form of a p-Value. We calculated for every miRNA and for each of its conditions the goodness-of-fit p-Value under the assumption of either a negative binomial or a gamma distribution (see Figure 4.15). We observed that in almost all cases the examined miRNAs are more likely to follow a gamma distribution than a negative binomial one, independent of their expression strength. This provided

more empirical evidence that the variance in miRNA count data is dominated by the exponential stochasticity of the PCR amplification step during RNA-Seq library preparation. Thus, making gamma distributions useful for modeling miRNA count expression data.

Gene expression strength distributions differ between small and total RNA-Seq data

In addition to examining the read count distributions, we studied the overall distributions of gene expression strengths of both the JenAge miRNA and mRNA sequencing datasets (see Figure 2.1 and Figure 2.2). To do so, we counted the number of genes to which a certain number of reads were assigned in steps of ten for all RNA-Seq libraries (see Figure 4.16). Interestingly, we observed the same negative exponential distribution of relative abundances of gene expression strengths in all small RNA-Seq libraries. At least 70 % of miRNA genes had a read count of less than 10, independent of the respective sample, indicating that the majority of miRNA genes is inactive during any given sampling point. The relative abundances and its variance further decreased with an increase in expression strength, often resulting in just a few miRNA genes whose assigned reads dominated the respective library (an observation that we have already reported in Section 4.3.3). This predictive characteristic of small RNA-Seq data sets was used later to improve the statistical gamma distribution based model of miRNA counts. For the investigated mRNA RNA-Seq libraries we observed a different behavior. The relative abundances showed a greater variance within each gene expression strength category compared with the miRNA sequencing data, resembling a bimodal instead of an unimodal distribution.

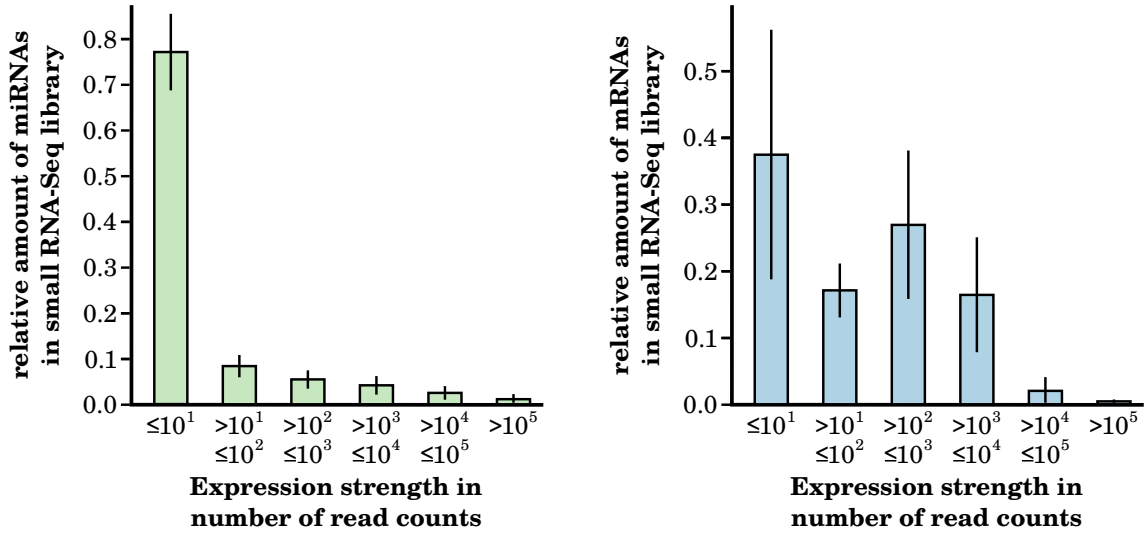


Figure 4.16: **Relative abundances of gene expression strengths of miRNA and mRNA sequencing datasets.** Whereas the abundance of miRNA gene expression strengths follows a negative exponential distribution, mRNA genes show a greater variance within certain expression strength categories (see SData 2).

4.4.2 Using gamma distributions to model small RNA-Seq expression data

In the following subsection, we describe a new statistical model based on gamma distributions to infer significant differences in the expression of miRNA genes between different conditions. The terms *gene read count* and *gene expression strength* are used synonymously.

Model description

First, we assume that for every sample y the amount of reads assigned to a any gene x can be modeled by a gamma distributed random variable G_{xy}:

$$G_{xy} \sim \Gamma(P_{xy}, B_{xy}),$$

with a shape parameter $P_{xy} > 0$ and a scale paramter $B_{xy} > 0$, resulting in non-negative read count values. Thus, the respective probability density function f_{xy} is:

$$f_{xy}(z) = \begin{cases} \frac{B_{xy}^{P_{xy}}}{\Gamma(P_{xy})} \cdot z^{P_{xy}-1} \cdot e^{-B_{xy}z}, & z > 0 \\ 0, & z \leq 0 \end{cases},$$

where $\Gamma(P_{xy})$ is the gamma function, evaluated at P_{xy}. Figure 4.17 displays examples of probability density plots of gamma distributions.

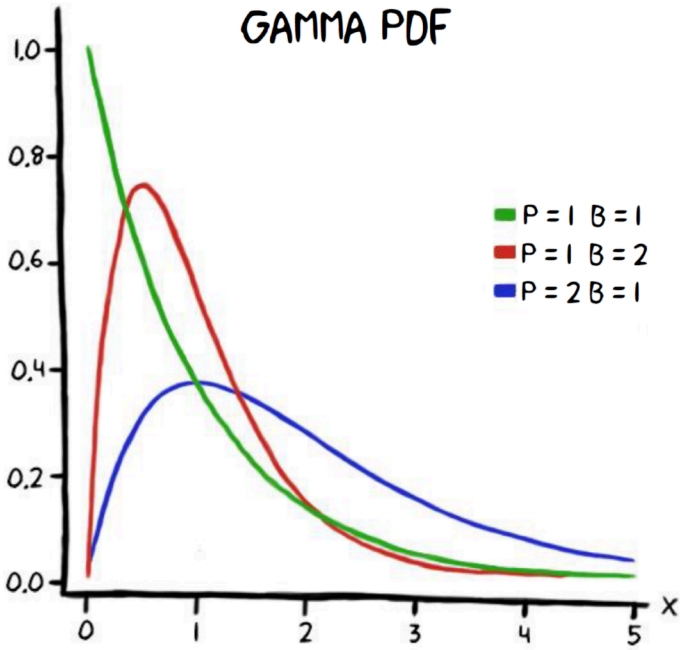


Figure 4.17: **Density of gamma distributions.** Shown are probability density function (PDF) plots of gamma dsitributions for different values of shape P and scale B.

The mean μ_{xy} and variance σ_{xy} of the random variable G_{xy} are defined as:

$$\mu_{xy} = P_{xy} \cdot B_{xy}$$

$$\sigma_{xy} = P_{xy} \cdot B_{xy}^2$$

Based on this, we further assume that the estimated expression strength E_{xy} of gene x in sample y is the product of our modeled gene count and a library-specific size factor S_y:

$$E_{xy} = G_{xy} \cdot S_y$$

When fitting our model to data, we expect this data to be a $x \times y$ count matrix, where $x = 1 \dots n$ is the number of genes, $y = 1 \dots m$ the number of samples and g_{xy} the count value of gene x in sample y. Library size factors are necessary to normalize RNA-Seq libraries in respect to their sequencing depths, or in other words, to equalize the total amount of reads between sequencing libraries that are compared. The calculation of the size factor estimators $\hat{s}_y$ is based on all libraries $y = 1 \dots m$ that are part of a comparison and is performed similarly as suggested by [356]:

$$\hat{s}_y = \underset{x}{median} \frac{g_{xy}}{(\prod_{i=1}^{m} g_{xi})^{1/m}}$$

.

Subsequently, we can estimate the average expression strength $e_{x,c(A)}$ of any gene x of some condition A with $c(A)$ being all sampled read libraries of condition A and $|c(A)| = r \leq m$:

$$e_{x,c(A)} = \frac{\sum_{y \in c(A)} e_{xy}}{r} = \frac{\sum_{y \in c(A)} g_{xy} \cdot \hat{s}_y}{r}.$$

Since we do not know the shape and rate paramters P_{xy} and B_{xy} of our gene expression strengths E_{xy}, we have to estimate them based on the given data. Fortunately, there exist good maximum-likelihood estimators for both parameters (within 1.5 % of the correct value) $\hat{P}_{xy}$ and $\hat{B}_{xy}$, that can be calculated efficiently [365]:

$$\hat{P}_{xy} \approx \frac{3 - t + \sqrt{(t-3)^2 + 24t}}{12t}$$

with

$$t = ln\left(\frac{\sum_{y \in c(A)} g_{xy}}{r}\right) - \left(\frac{\sum_{y \in c(A)} ln(g_{xy})}{r}\right)$$

and

$$\hat{B}_{xy} = \frac{1}{r\hat{P}_{xy}} \cdot \sum_{y \in c(A)} g_{xy}.$$

Building gene expression clusters

With the above introduced random variables E and G as well as both parameter estimators $\hat{P}$ and $\hat{B}$ it is possible to model the gene expression strength of any given miRNA gene from small RNA-Seq data. However, the performances of the shape and scale estimators strongly depend on the available amount of data, *i.e.*, the number of independently measured count values for every gene in each condition. The higher the amount of biological replicates, the better perform $\hat{P}$ and $\hat{B}$. As previously mentioned in Section 4.4.1, the number of biological replicates per investigated condition for RNA-Seq experiments is typically limited, seldom exceeding five or more sequenced samples. To overcome this obstacle, we assume that genes sharing a similar mean expression strength also share a similar dispersion. With this assumption it is possible to cluster genes based on their mean expression strengths and use the available count data of all genes within each cluster respectively to estimate their common shape and scale parameters $\hat{P}$ and $\hat{B}$ (see Figure 4.18).

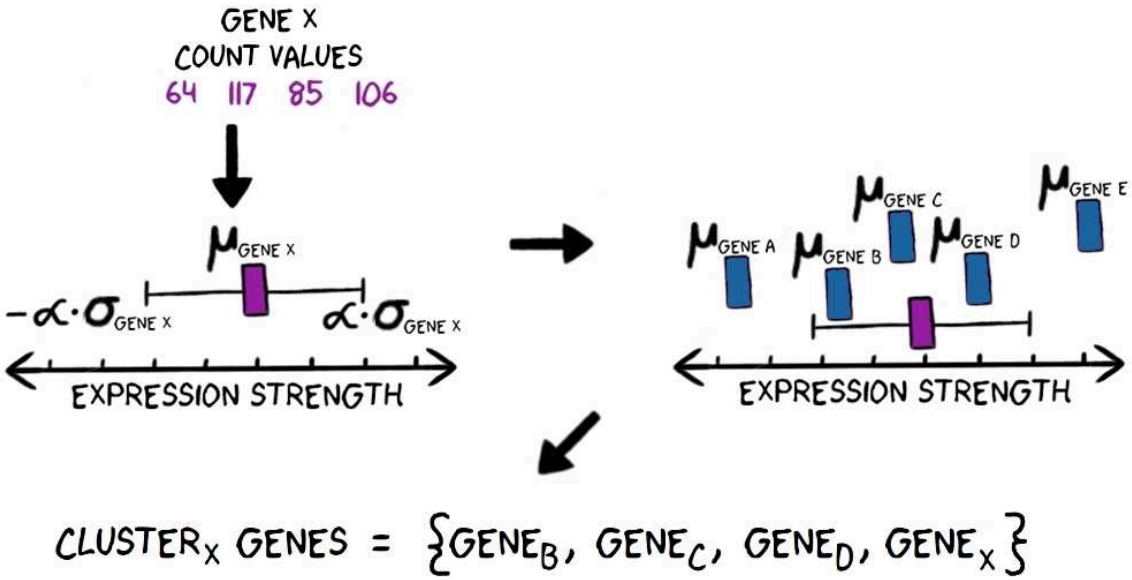


Figure 4.18: **Expression strength cluster approach to pool small RNA-Seq data.** Any gene x spans an expression range interval R_x based on its count mean and standard deviation. All other genes w whose count means are within the interval R_x are assigned to the expression cluster $C(x)$ of gene x. Every expression cluster contains at least its spanning gene.

For every gene x we calculate its expression strength mean μ_x as the empirical mean read count of all replicates of gene x. And we define the *expression range* of every gene x as the interval R_x:

$$R_x = [\mu_x - \alpha\sigma_x, \mu_x + \alpha\sigma_x],$$

with σ_x being the empirical standard deviation of the gene expression of gene x and α an additional scaling factor. At the moment, α can be assumed to be a strictly positive value and its precise function will be discussed later. Again, for every

gene x we can now determine its expression strength cluster $C(x)$, which includes all genes w whose expression mean μ_y lies within R_x. This approach is performed for each condition and their respective replicates separately, allowing to estimate P_x and B_x on the aggregated counts in $C(x)$.

Outlier detection and expression cluster correction

Outliers (*i.e.*, singular extreme values often deriving from experimental errors) within count data represent sources of potentially great bias and have to be eliminated. But they are hard to identify if only few replicates are available, thus, the aggregation of count values for each gene based on the introduced expression clusters can also help in that matter. However, because expression clusters are calculated using mean and standard deviation of gene counts, they can be strongly affected by outliers themselves. In the presence of one or more outlier values in a given gene x its associated expression range R_x is greater than it should be, leading to an inclusion of genes w into the expression cluster $C(x)$ that do not fit to the true mean μ_x. Accordingly, the subsequent estimation of the shape and scale parameters P_x and B_x can be strongly biased towards the outlier values. We can still use the information provided by the expression clusters to detect the presence of outliers within the examined dataset. For that, we developed two criteria that mark gene clusters which are potentially affected by outliers:

1. Single count values of gene x are outside the expression interval spanned by its expression cluster $C(x)$.

2. The number of genes w assigned to expression cluster $C(x)$ is higher than would be expected considering its mean value $\mu_{C(x)}$.

Similarly to the gene expression range R_x, the interval $R_{C(x)}$ spanned by an expression cluster is defined by its mean and standard deviation, but without utilizing an additional scaling parameter α.
If only single outliers are present within an expression cluster $C(x)$ its dispersion is not altered strongly, making it possible to detect them easily (criteria 1). But if multiple outliers are present within $C(x)$, they might have overly increased the cluster expression interval $R_{C(x)}$ by inclusion of too many not fitting genes and are not evident in respect to criteria 1. However, because we know that the abundance of gene expression strengths in small RNA-Seq datasets strictly follow a negative exponential distribution (see Section 4.4.1), we can identify expression clusters that include more genes than would be expected from their mean (criteria 2). Outliers are then removed by the double median absolute deviation (MAD) approach, which is is an extension of the normal MAD approach for skewed distributions [366]. After outliers are removed the gene expression clusters can be corrected and are recalculated for every gene. If a single gene x has too few count values left in any of the investigated conditions, no meaningful assertion can be made about its mean μ_x and it should be removed entirely from the differential expression analysis. However, which exact threshold should be used for an exclusion is highly debatable and should always be chosen depending on the general experimental setup. We would recommend to have at least four valid count values per condition for any gene x.

Hypothesis testing

After the outlier detection and the gene expression cluster correction we suppose that we have good estimators $\hat{P}_x$ and $\hat{B}_x$ for every gene x for two biological conditions I and II. That means we can now model every gene x as a gamma distributed random variable, depending on its condition: $\Gamma(\hat{P}_{x_I}, \hat{B}_{x_I})$ and $\Gamma(\hat{P}_{x_{II}}, \hat{B}_{x_{II}})$. To evaluate if any gene x is differentially expressed, *i.e.*, the alternative hypothesis $\mu_{x_I} \neq \mu_{x_{II}}$ is more likely than the null hypothesis $\mu_{x_I} = \mu_{x_{II}}$, we can calculate the probability that the value μ_{x_I} was drawn from the distribution spanned by $\Gamma(\hat{P}_{x_{II}}, \hat{B}_{x_{II}})$ (and vice versa). In others words, we calculate the likelihood of drawing the mean count value of gene x from one condition given the approximated count distribution for the same gene under the second condition. To obtain the respective p-Value we perform three simple steps (see Figure 4.19):

1. Standardization of the gamma distribution spanned by $\Gamma(\hat{P}_{x_I}, \hat{B}_{x_I})$ by normalizing the scale parameter $\hat{B}_{x_I}$ with σ_{x_I} to obtain $\Gamma(\hat{P}_{x_I}, \frac{\hat{B}_{x_I}}{\sigma_{x_I}})$.

2. Shift of the standardized gamma distribution by β to obtain $\Gamma(\hat{P}_{x_I}, \beta, \frac{\hat{B}_{x_I}}{\sigma_{x_I}})$, because negative values are not well defined for gamma distributions (see Section 4.4.2, Model description).

3. Calculation of the p-Value by:

$$\text{p-Value} = 1 - \int_{-z_{x_{II}}+\beta}^{z_{x_{II}}+\beta} \Gamma(\hat{P}_{x_I}, \beta, \frac{\hat{B}_{x_I}}{\sigma_{x_I}}) dz_{x_{II}}$$

   with the z-Values being

$$z_{x_{II}} = \frac{|\mu_{x_{II}} - \mu_{x_I}|}{\sigma_{x_{II}}(\sqrt{N_{x_{II}}})^{-1}}$$

   and $N_{x_{II}}$ being the number of count values of gene x under condition II.

There is one exception to this procedure of hypothesis testing. If any given gene x has too few count values in its corresponding expression cluster $C(x)$ to calculate good estimators $\hat{P}_x$ and $\hat{B}_x$ for its shape and scale parameter, a different approach is used to calculate the respective p-Value. Due to the negative exponential distribution of the expressions strengths in small RNA-Seq data sets, only few miRNA genes with exceptional strong expression are affected by that. As suggest by Qin *et al.* [357], a cubic root transformation is applied to the count values of the concerning gene x, but instead of a normal t-Test to calculate the p-Value we use Welsch's t-Test, because we cannot assume an equal variance in the count values of both biological conditions or equal sample sizes [367]. Again, defining how many count values are sufficient to estimate $\hat{P}_x$ and $\hat{B}_x$ is debatable, but we recommend to have at least 20 valid count values for any gene x.

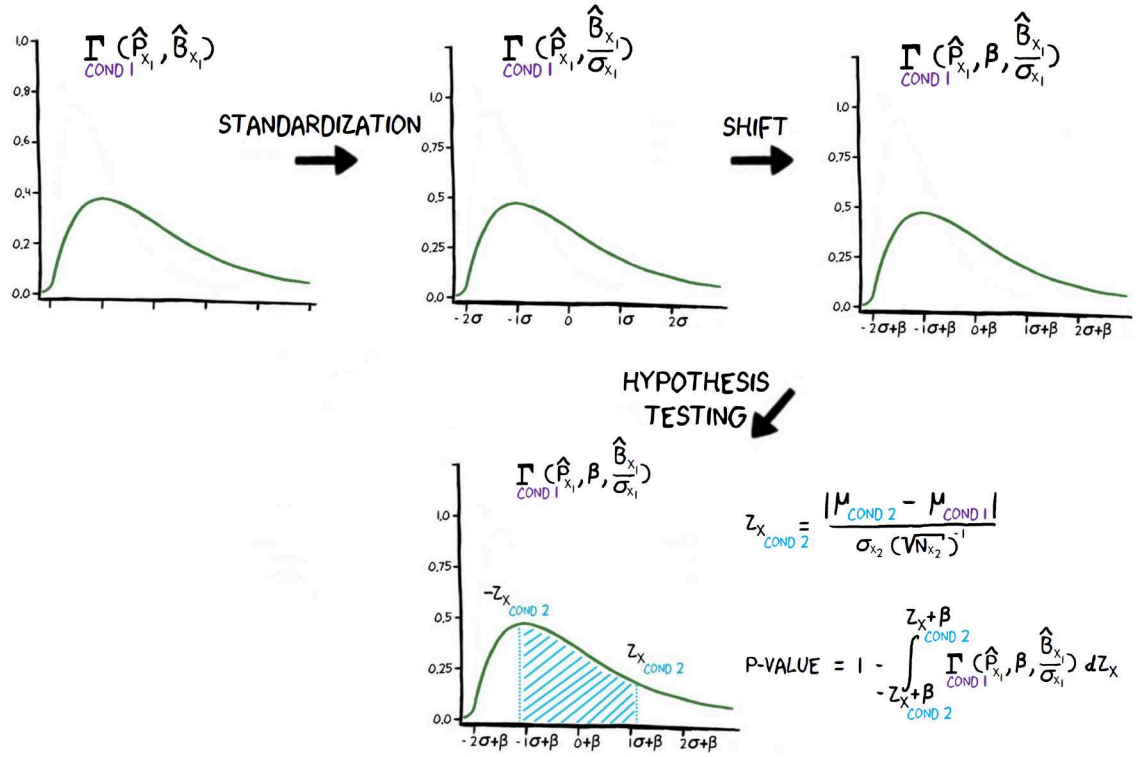


Figure 4.19: **Hypothesis testing using gamma distributions.** To estimate the probability of a significant difference between the measured mean count values of any gene x between two biological conditions, we calculate the p-Value that the null hypothesis is true in three steps: standardization followed by a positive shift of the gamma distribution to obtain $\Gamma(\hat{P}_{x_I}, \beta, \frac{\hat{B}_{x_I}}{\sigma_{x_I}})$, and subsequent calculation of the size of the area under the approximate gamma distribution determined by the respective z-Values.

4.4.3 Comparison to existing negative binomial models

To show that the new gamma distribution based model can compete with existing negative binomial models, we compared it against two state-of-the-art differential expression tools `DESeq2` [128] and `edgeR` [355]. To that end, we generated a several comprehensive artificial datasets of small RNA-Seq count data to have a direct comparison of the tools' true positive and false positive rates. We also compared their general performances on the JenAge small RNA sequencing data.

Implementation of the gamma model

We implemented the statistical model described in Section 4.4.2 in a command line tool written in Python named `MeRDE`, which is an acronym for *microRNA differential expression analysis*. `MeRDE` has two main functions and can therefore be started in two different modes: correct counting of miRNA-derived reads from small RNA-Seq libraries (*count* mode) and estimation of differentially expressed miRNAs between two biological conditions (*analyze* mode). As we have learned in Section 4.1, a

common miRNA gene encodes for two different mature miRNA molecules with potentially differing mRNA targets, regulating varying biological functions. Thus, both mature sites of a miRNA gene have to be treated separately to not only ensure correct counting of the associated reads but also to estimate differential expression correctly (see Figure 4.20). We have already shown that this is relevant for real small RNA-Seq data (see Section 4.3.3, mature miRNA switches).

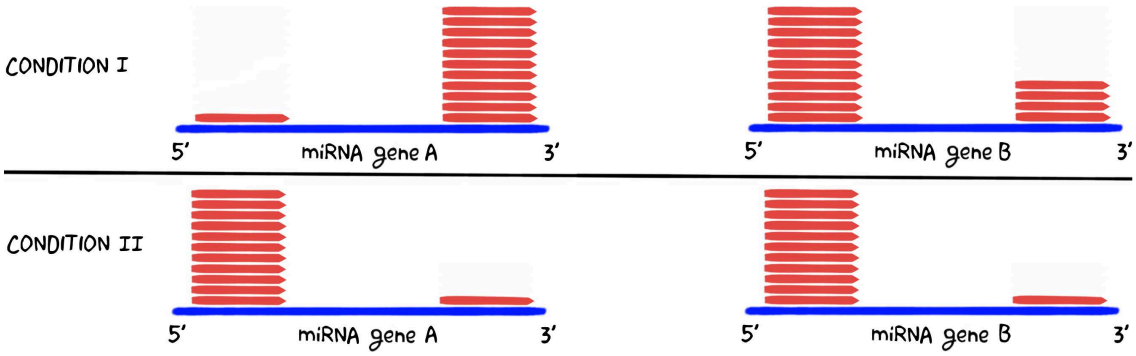


Figure 4.20: **Examples of potential miRNA read counting mistakes.** In this simplified example we have two miRNA genes A and B (blue lines) and corresponding mapped reads (red lines) and want to compare their expression between two different conditions I and II. In both cases we would not detect a significant difference in the total abundance of reads mapped to gene A or gene B between both conditions. However, when considering the mapped reads for the 5p and 3p sites individually, a mature miRNA switch can be detected for gene A as well as a significantly altered expression of the 3p mature miRNA of gene B.

Using the *count* mode, `MeRDE` generates a simple count table in `csv` format for a given `gff` annotation file and any given number of small RNA-Seq datasets in the standard `sam` or `bam` file format. `MeRDE` can automatically split the annotation of any miRNA gene if no mature sites are specified, to ensure the correct counting of reads for the individual mature miRNAs. Additional user adjustable parameters control the size with which a read must overlap an annotated miRNA to be counted, if only uniquely mapped reads should be considered and if the sequencing protocol was strand-specific, making it necessary to also consider the orientation of every mapped read. The full list of available parameters with a detailed description can be obtained using the command `MeRDE.py count --help`.

To start the *analyze* mode of `MeRDE`, one or several counting files in `csv` format are required, which contain all miRNA genes and respective count values that should be tested for differential expression. If the count values are not normalized, `MeRDE` can do so, by calculating the appropriated size factors as described in Section 4.4.2. The user can decide to ignore lowly expressed genes by setting a minimum amount of reads that have to be mapped to any gene in both examined biological conditions. By default, mature miRNAs with less than ten read counts in both conditions (after normalization) are considered not expressed and excluded from the subsequent analysis. Additionally, the user can define if identified outlier values should be kept or

removed from the dataset and what the minimal amount of replicates per gene and condition should be to perform the differential expression analysis. The full list of available parameters with detailed description can be obtained using the command `MeRDE.py analyze --help`.

Artificial datasets

To directly compare the performances of `MeRDE`, `DESeq2` and `edgeR` we created several artificial datasets, mimicking the characteristics of count files derived from real small RNA-Seq libraries. Each of these artificial datasets included 1,000 simulated miRNA genes with randomly associated count values and fulfilled the following criteria:

- For each gene x a set of 'replicated' count values was drawn from a gamma distribution Γ_x for two simulated differing conditions.
- Each of these sets consisted of three up to nine count values, with all simulated genes having the same amount of 'replicates' within one dataset.
- The mean μ_x of the gamma distribution Γ_x of each gene x was drawn from a negative exponential function, describing the commonly observed distribution of expression strengths in small RNA-Seq datasets (see Section 4.4.1 and Figure 4.16).
- The scale parameter (*i.e.*, the dispersion) of the gamma distribution Γ_x of each gene x was set to be either 5, 10 or 20.
- For each gene x every single count value drawn had a chance of 2 % to be an outlier and was either multiplied or divided by a uniformly distributed random float number between 5.0 and 10.0.
- Every gene x had a chance of 5 % to be differentially expressed, meaning that all simulated count values of one condition were multiplied by a uniformly distributed random float number in the range of 1.5 and 3.0.

For each combination of replicate number and gamma scale parameter 1,000 datasets were generated, totaling in 21,000 simulated small RNA-Seq counting files. All three tools were used to identify the differentially expressed genes and we measured their respective mean true positive and false positive rates (see Figure 4.21 left side). We observed a general improvement in the true positive rates of all three tools with an increasing number of replicates per condition, which is expected since all employed statistical models can better estimate their model parameters with more given data. Also not unexpected was a common drop in the respective true positive rates with an increased gamma scale. The higher the gamma scale, the more dispersed were the count values of each gene within a single condition, making it hard to distinguish significant fold changes in the mean gene expression from 'natural' expression variation. With a minimum of four or more replicates per condition `MeRDE` was able to equal or outperform both other tools in terms of identifying differentially expressed genes.

However, to make a fair comparison between the three tools we generated a second collection of 21,000 simulated small RNA-Seq counting files, this time drawing count values from a negative binomial distribution with the respective dispersion parameter varying between 0.5, 0.3 and 0.1 (see Figure 4.21 right side). Still, `MeRDE` could equal and sometimes even outperform the true positive rates of `DESeq2` and `edgeR` five or more replicates per condition. Based on our comprehensive artificial datasets we could show that our gamma distribution based statistical model can compete with state-of-the-art tools on miRNA sequencing data sets.

Real datasets

It is difficult to accurately benchmark the performances of differential gene expression tools, because an agreed-upon test standard dataset based on real RNA-Seq data is still missing. Nevertheless, we wanted to estimate the performance of `MeRDE` to detect differentially expressed genes in real datasets in comparison to `DESeq2` and `edgeR`. Therefore, we employed all three tools to detect differentially expressed miRNA genes (DEMs) within the JenAge small RNA-Seq dataset by performing pairwise comparisons between all ages within each species and its respective tissue samples (see Figure 2.2). In total, 276 comparisons were performed which resulted in 9,980 individually identified differentially expression events (see SData 3). When overlapping the obtained results from the three tools, we observed that only about 889 DEMs were commonly identified by all of them, which is less than 9 % of the total number of estimated DEMS (see Figure 4.22). Interestingly, there were no DEMs exclusively determined by `edgeR`, meaning that it did not contribute to the total number of 9,980 DEMs. In general, `edgeR` was more conservative compared to the other two tools, having estimate only 2,794 DEMs. This is an observation that was already made before and is confirmed with our findings [368]. In contrast, `MeRDE` and `DESeq2` both identified 7,126 and 5,958 DEMs respectively, of which 3,303 DEMs were shared.
When exemplary investigating some of the miRNAs that differed in their differential expression status between `DESeq2` and `MeRDE` more closely, we observed that in many of these cases the negative binomial model of `DESeq2` clearly over- or underestimated the expression variance of those genes and our model was more in line with the empirical evidence (see Figure 4.23). However, this was not true for all cases. We also found some examples were the gamma model of `MeRDE` seemed to over- or underestimate expression variances (see SData 4).
In Section 4.5 we will explore the results of the differential expression analyses of `MeRDE` on a subset of the JenAge miRNA sequencing data in more detail.

Further improvements of the gamma model and MeRDE

The statistical model underlying `MeRDE` as well as certain implementation details are still work in progress and will be improved in the near future. For the model we plan to optimize the calculation for the gene's expression clusters. At the moment, the introduced parameter α that acts as a scaling factor for the expression range R is a implemented as a manually adjustable constant, which is the same for all genes in one analysis. However, the idea behind this scaling factor is to automatically

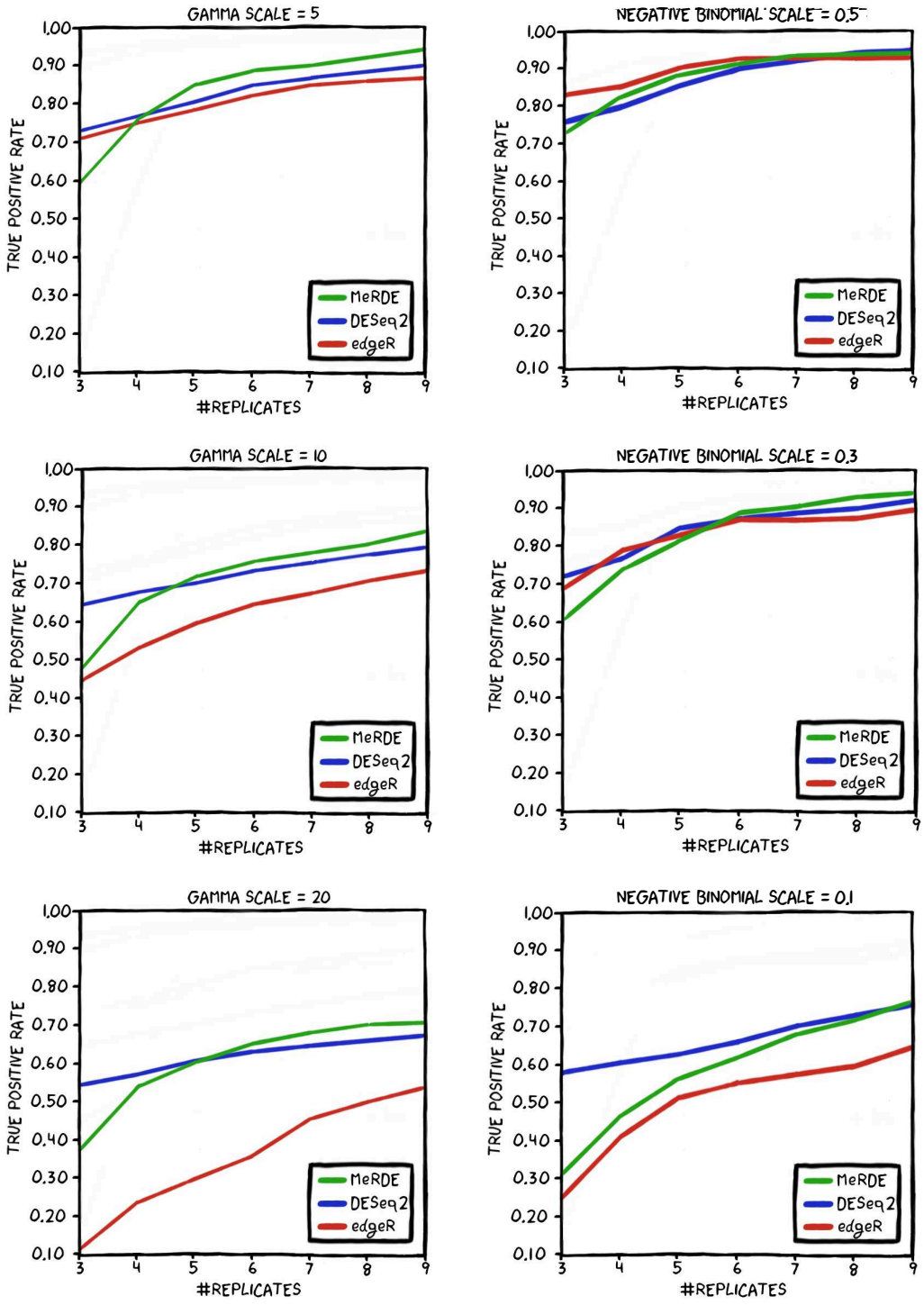


Figure 4.21: **Performance comparison of MeRDE, DESeq2 and edgeR on simulated miRNA sequencing data.** On the left side the comparison results of the gamma distribution based artificial datasets and on the right side the comparison results of the negative binomial distribution based artificial datasets are displayed. The mean rate of correctly identified differentially expressed genes are plotted against the amount of replicated count values per simulated biological condition. With a minimum of four to six replicates per condition `MeRDE` was able to equal or outperform both other tools. All three tools showed mean false positive rates of less than 0.003 during any test condition.

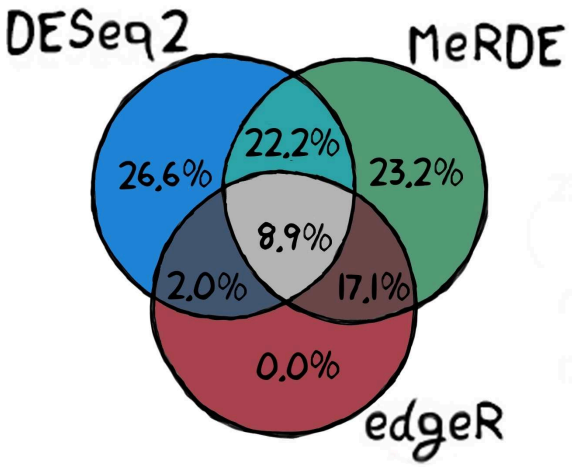


Figure 4.22: **Result comparison of the differential gene expression analyses performed on the JenAge small RNA-Seq datasets.** In sum, `MeRDE`, `DESeq2` and `edgeR` identified 9980 DEMs in the small RNA expression libraries of the JenAge dataset. With only 889 common DEMs, the overlap of all three compared tools was relatively low, which was mostly due to the conservative estimations of `edgeR`. Whereas about 31.1 % of all DEMs were commonly identified by `MeRDE` and `DESeq2`, both tools still had a relative large share of exclusively determined DEMs.

adjust it based on the observed relative variance of every gene, individually. This individual scaling will help to tailor the genes' expression range more specifically to the observed relationship between their mean read count and standard deviation, which is crucial to calculate the correct expression clusters by including only genes that display more similar expression characteristics. Consequently, with the improvement of the gene expression cluster calculation we enhance the estimations of the respective genes' gamma distributions and therefore overall accuracy of differential expression identification.
Nevertheless, we could show on the basis of artificial and real small RNA-Seq datasets that the current gamma distribution model of `MeRDE` is capable of identifying differentially expressed miRNAs with high precision (*i.e.*, displaying a high true positive rate while maintaining a low false positive rate). Furthermore, we could prove that gamma distributed noise in small RNA-Seq data due to the experimental PCR amplification prevails over other potential sources of natural variance (see Figure 4.15). Moreover, we could contribute to further understanding of small RNA sequencing stochasticity by showing that the expression strength abundance distribution of miRNAs strictly follows a negative exponential distribution (see Figure 4.16).
Besides improvements in our statistical model, we also work on a refined and more easy-to-use version of `MeRDE` itself, by setting up a web service where users just have to upload their read or counting data together with a simple description of the experimental design. Alongside the current text-based result file, `MeRDE` is able to generate several diagnostic plots (such as, scatter, volcano, t-SNE, relative variance,

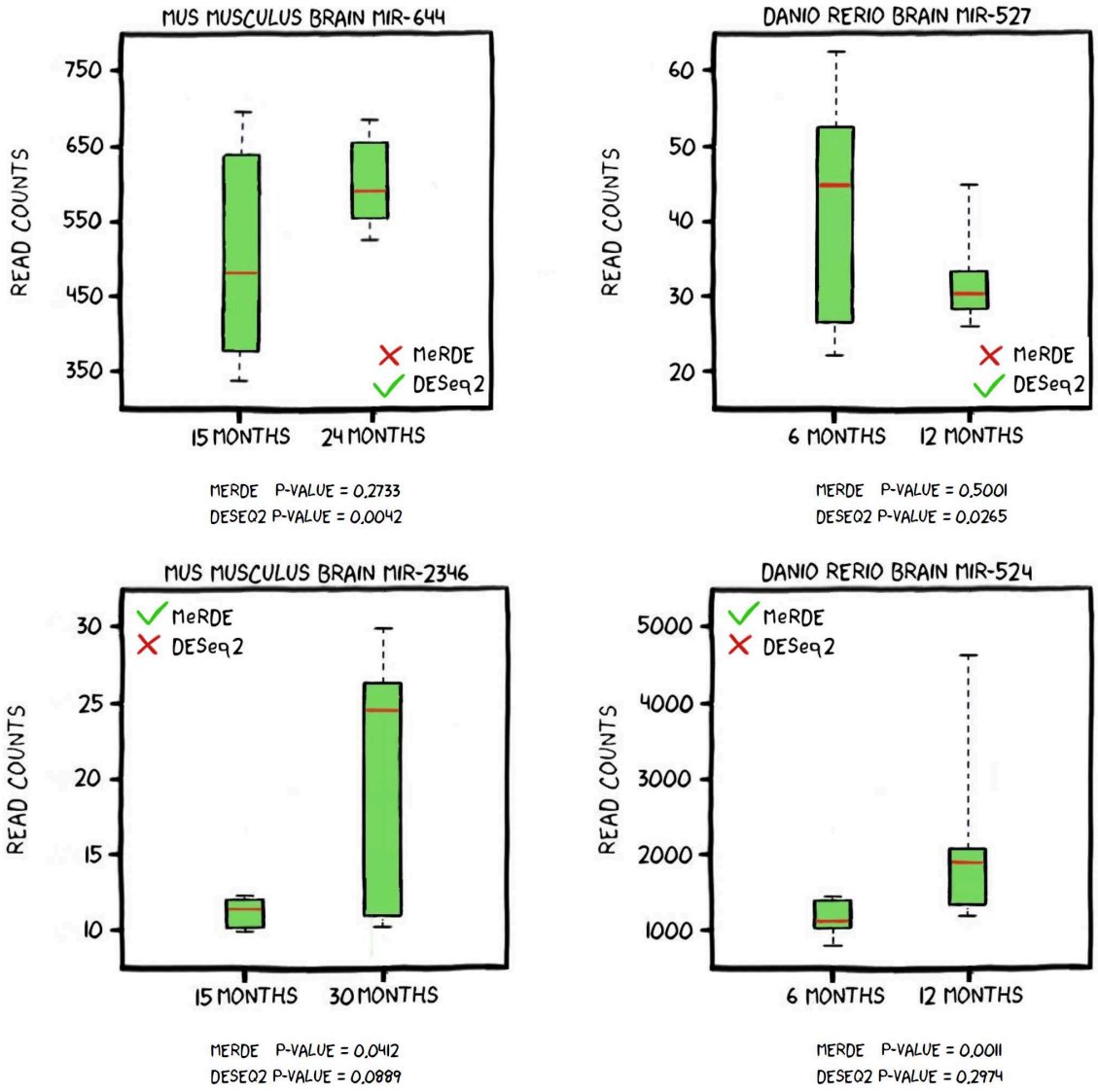


Figure 4.23: **Examples of miRNAs with differing differential expression status between MeRDE and DESeq2.** Boxplots by different time points from four miRNAs that have a `MeRDE`-based p-Value of less than 0.05 and a `DESeq2`-based p-Value of more than 0.05 or vice versa. For more examples, see SData 4.

expression boxplots and p-Value distribution plots) that help in understanding and interpreting the investigated small RNA-Seq data. However, most of these visualization functions are not finally implemented yet and we work on providing an elaborate html-based output that summarizes the results and diagnostics of each analysis.

4.5 The role of miRNAs during aging: A comparison between different species

This chapter is based on the publication manuscript "MicroRNAs in Aging: potentially conserved key regulators of aging".[5]

Data used and analyses performed in this section

Age comparison setup

The age comparisons were performed according to Figure 4.24. For each of the four species *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* we investigated four time points (young mature, mature, aged, old-age) in up to four tissues (blood, brain, liver, skin). All differentially expressed miRNA (DEM) results can be found at SData 1.

A

species	tissues	ages: young mature	mature	aged	old-age		No. samples
H. sapiens	blood, skin	24-29	45-50	60-65	75-80	years	60
M. musculus	blood, brain, skin	09	15	24	30	months	60
D. rerio	brain, skin	12	24	36	42	months	40
N. furzeri	brain, liver, skin	12	20	27	39	weeks	56

B

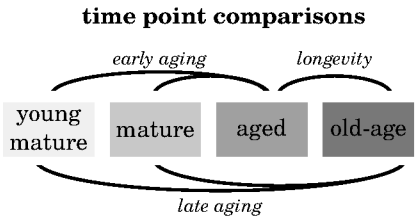


Figure 4.24: **Overview of species, tissues, time points and comparisons of the utilized small RNA-Seq libraries.** (A) Samples of *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* were taken from up to three different tissues (blood, brain, liver, skin) at four different time points. The sampled time points can be assigned to different ages: young mature, mature, aged and old-age. (B) To explore the changes of miRNA expression, we studied three expression comparisons: early aging (young mature and mature vs. aged), late aging (young mature and mature vs. old-age), and longevity (aged vs. old-age). All comparisons were performed individually for each of the studied species and their tissues.

[5]The complete supplemental material is available at `https://osf.io/u86cn/`

Differential expression analysis

Mapping of RNA-Seq libraries was performed as described in Section 2.2. Read counting and estimation of differentially expressed miRNAs was carried out by `MeRDE` with the following parameter setup: `-l 10 -s 20 -m 4 -sigma 2 -n True -e remove`. Parameter descriptions can be found at SData 2.

miRNA size variation and sequence modification analyses

We measured the normalized length of each mapped mature miRNA transcript as the difference in its length with respect to the annotated reference sequence for each sampled age, tissue and species, individually. Size variations between the ages of each tissue and species were compared with the Kruskal-Wallis H-test.
Similarly, single base exchanges of the mapped transcripts sequences compared with the reference genome sequence were counted for each sampled age, tissue and species, individually. We distinguished the sequence modification events based on the exact type of base exchange (*e.g.*, adenosine to cytosine, thymine to guanine, etc.) and compared the relative frequencies of these events between the sampled ages of each tissue and species, using the Kruskal-Wallis H-test.

4.5.1 miRNAs as potential genetic aging factors

We already know that miRNAs are crucial for the regulation of almost all cell functions (see Section 4.1), including aging-related processes, such as senescence (permanent cell cycle arrest), apoptosis (induced cell death), inflammatory response or various other stress responses [369, 370]. Thus, an implicit connection between aging and miRNA-mediated regulation has to exist. However, to which extend miRNA regulatory networks or even single miRNAs influence the progression of aging processes is still subject of recent research. In 2005, lin-4 was discovered to be the first miRNA to explicitly modulate life-span of the nematode *Caenorhabditis elegans* [371]. Later on, more miRNAs were observed to be up- or downregulated with aging in different invertebrates, but also in certain murin tissues or human cell lines [369]. Moreover, the change of miRNA expression could be correlated with the formation or progression of different cancers and other age-associated diseases, such as, neurodegenerative diseases, cardiovascular diseases, diabetes or the phenomenon of inflammaging (see Chapter 5) [370, 372, 373]. Still, due to the already complex nature of aging we are just beginning to understand the precise role of miRNA-mediated control in the interplay between cellular homeostasis and stress response functions. Since many miRNAs are expressed in a tissue-specific manner or are even species-specific, not much information is available if general and evolutionarily conserved miRNA regulators of aging exist. In mammals, a prominent example of such conserved aging-associated miRNAs are members of the mir-17/92 cluster, which were first discovered to be oncogenic miRNAs [374]. It was later suggested that these miRNAs regulate not only the formation of cancer but also modulate organismal life-span, especially miR-19 and miR-92 [374]. The identification of more such conserved miRNA-mediated regulation of aging processes would not only help to understand aging in general, but could help to establish new diagnostic markers of age-associated diseases and potential therapeutic targets [373].
In this section, we examined the small RNA-Seq data from the JenAge dataset to identify evolutionarily conserved miRNAs, that change their expression in an age-related manner. Furthermore, we investigated if and what kind of sequence modifications of mature miRNAs occur during aging. The results presented within this section are part of a bigger study, which involves several biological experiments in the two fish species *Danio rerio* and *Nothobranchius furzeri* to determine and validate the predicted expression and age-dependent change of certain miRNAs. However, because some of these experiments are still in progress, we will present only our bioinformatical predictions, that serve as the basis of the ongoing and future experiments.

4.5.2 Overall expression and modification of miRNAs appears to be less affected by aging

The abundance of active miRNAs is tissue- but not age-dependent

As a starting point for our analyses, we examined the abundance of actively expressed miRNA genes during each time point within the different tissues of all four investigated species (see Figure 4.25). Displaying almost only half as much active

miRNAs, the blood samples of *Homo sapiens* and *Mus musculus* showed a great difference compared with their respective other tissues. In contrast, we observed only little differences in the number of active miRNAs in *Danio rerio* and *Nothobranchius furzeri* irrespective of the tissue type. Also, there was no significant age-dependent change in the abundance of active miRNAs within the individual tissues of any of the investigated species, indicating that at the first glance aging appears to have no pronounced effect on a general decrease or increase in the number of activated miRNA regulatory networks. This observation does not imply that there is no aging-related effect in the expressional activity of certain miRNAs, but suggests that the regulatory level of miRNAs is constantly maintained during life time.

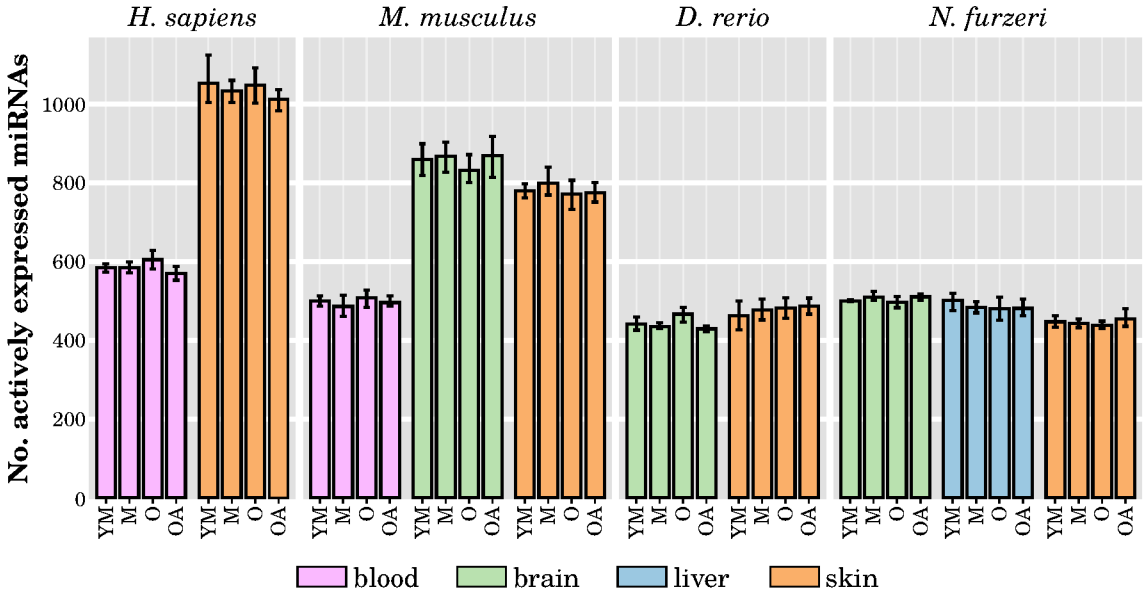


Figure 4.25: **Abundance of actively expressed miRNA genes.** Whereas the number of active miRNAs appears to be unaffected by the sampled age, the tissue type has an observable influence, which is more pronounced in both investigated mammals than the two fishes (YM – young mature, M – mature, A – aged, OA – old-age).

Expression patterns of miRNAs can distinguish tissues and species, but not age

To evaluate if we can discriminate the sampled ages based on the expression patterns of evolutionarily conserved miRNAs, we performed a t-distributed stochastic neighbor embedding (t-SNE) on the read count data of 85 miRNAs that are common to all four investigated species (see Figure 4.26 and SData 3). Whereas all samples formed distinct clusters based on their tissue type and species, we could not observe a clear separation of the different time points within these clusters. In general, the expression of the evolutionarily conserved miRNAs showed to be less variant in the brain samples of *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* compared with all other examined tissues. When applying the t-SNE approach on each species

individually, considering all respective miRNA genes that are expressed during at least one sampled age, we still observe tissue-specific clusters that cannot be further separated into distinct age-specific subclusters (see SData 4). This observation suggests that there seem to be no specific miRNAs in charge of a conserved regulation of age-related processes.

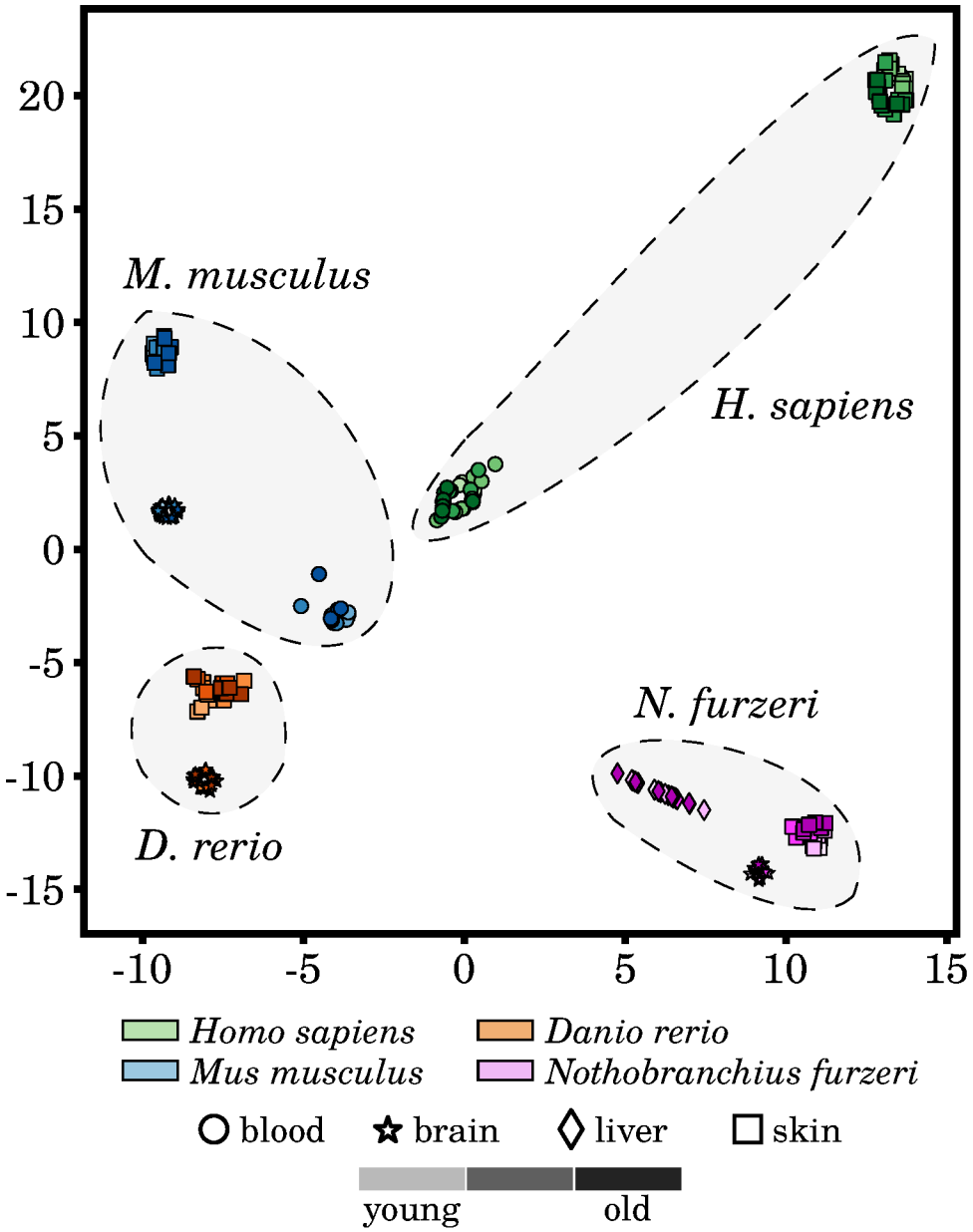


Figure 4.26: **RNA-Seq sample clustering based on the expression patterns of conserved miRNAs.** All small RNA-Seq libraries were clustered by the t-SNE approach based on the expression patterns of 85 miRNA genes, that are evolutionarily conserved in all four investigated species. All sampled tissues form distinctive species-specific and non-overlapping clusters. A further separation of the individual samples in these clusters based on the age could not be observed.

Modifications of miRNA transcripts during their maturation appears no to be age-related

It is estimated that in human at least 16 % of all miRNA transcripts are affected by sequence modification processes during their maturation process within the nucleus [375]. Those modifications include RNA editing, mostly carried out by adenosine deaminase enzymes, catalyzing adenosin to inosin transitions [376]. RNA editing of miRNA transcripts can prevent further maturation to functional miRNA molecules or alter target specificity by changing their seed region, *i.e.*, the mRNA recognition and binding site [377]. But also shortening of mature miRNA molecules is a common modification used to alter miRNA stability and consequently regulate their activity [377].
We investigated if we can identify an age-related change in the size or modification of mature miRNAs in any of the sampled tissues of the four examined species. However, whereas we could observe that the mean length of mature miRNA transcripts was one base shorter than their reference annotation in all samples, no significant shortening with age was identified (see Figure 4.27). The variation in miRNA length was highest in human compared with the other three species, but also no significant differences in size variance was found within or between each species with respect to the sampled ages.
When examining RNA editing events, we found that the majority of base exchanges occur at the first position or last two positions of the miRNA transcripts in respect to their 5' site, independent of species and tissue (see SData 5). Interestingly, taking all investigated samples into account, all four bases were exchanged with about the same frequency, whereas changes to adenine and thymine were observed to happen twice as often as to cytosine or guanine (see SData 6). Base exchanges at the border sites of miRNAs do most likely not change their mRNA target specificity, because their seed regions are located more towards the middle of the mature transcripts. Still, adenosine and thymine form less hydrogen bonds, making the miRNA-mRNA pairing more unstable and as a consequence the RNA silencing effect less effective. These findings indicate that RNA editing of miRNAs more often regulate their inhibition strength than alter their target. However, we could not identify a significant change in the frequency of specific RNA editing events in dependence of the age of the sampled tissues. It appears miRNA modification processes are unaffected by aging.

4.5.3 Differential expression analysis reveals potential miRNA key regulators of aging

In Section 4.4 we introduced `MeRDE`, a new tool to estimate differential expression of miRNAs more accurately, based on a newly developed gamma distribution model. We have calculated differentially expressed miRNAs (DEMs) of the JenAge small RNA sequencing data not only to evaluate its performance, but also to identify miRNA regulators of aging-related processes. We performed three different aging comparisons (early aging, late aging and longevity comparison) between the sampled ages of each tissue and respective species, individually.
In total, only few miRNA genes were found to be differentially expressed with age,

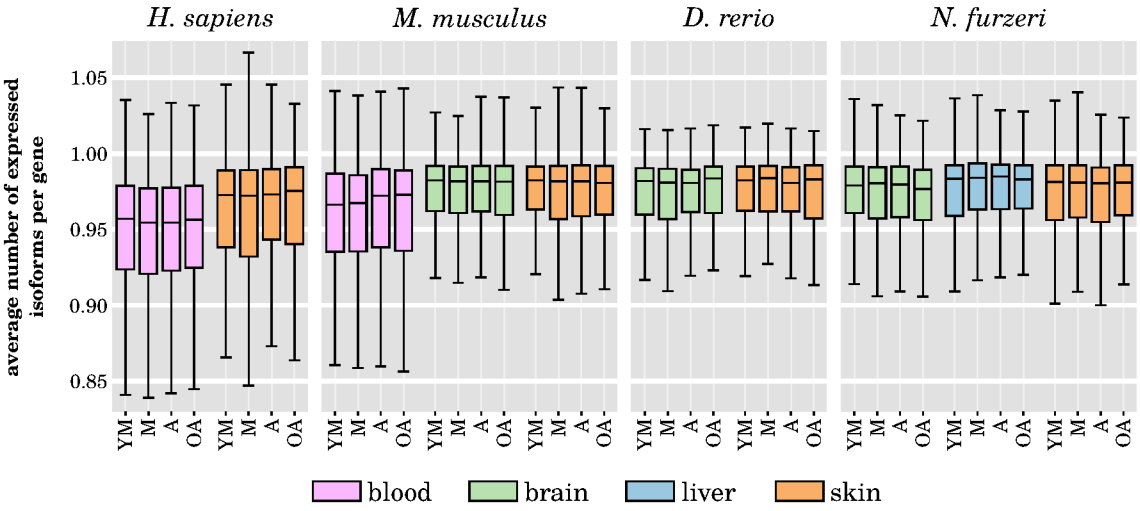


Figure 4.27: **Length variations of mature miRNA transcripts during aging.** Whereas the mean length of mature miRNA transcripts was observed to be one base shorter than their reference annotation in all samples, no significant shortening with age could be identified (YM – young mature, M – mature, A – aged, OA – old-age). For details, see SData 5.

with *Homo sapiens* and *Nothobranchius furzeri* showing the lowest number of DEMs (see Figure 4.28). Still, we observed in almost all cases an increase of DEMs with age. Interestingly, *Danio rerio* displayed the highest amount of DEMs, while having the lowest number of known miRNA genes.

Not unexpectedly, we found only few overlapping DEMs between the tissues of the four investigated species (see SData 7). However, when overlapping all identified DEMs we found 33 to be shared by two, 13 to be shared by three and one to be shared by all four species (see Figure 4.29). The latter commonly shared miRNA is miR-192, an important regulator of the well-studied tumor suppressor p53 [378]. As a transcription factor, p53 modulates the expression of a wide variety of genes, controlling biological functions such as senescence or apoptosis [379]. Usually, p53 activation is triggered by severe DNA damage, either resulting in the accumulation of the cell cycle arrest inducing protein p21 or in the apoptosis pathway activating proteins Noxa and Puma [379]. In several studies it was shown that miR-192 activation has strong tumor suppressing effects in different cell types, by enhancing p53 and consequently p21 expression resulting in the formation of senescent rather than cancerous cells [379–382]. We have already discussed in Chapter 1 (and will extend this discussion further in Chapter 5) that the accumulation of senescent cells is one of the main driving factors of aging. This puts miR-192 already in the position of a potential key regulator of certain aging processes. Moreover, in the presence of oxidative stress, another major source of aging promoting processes, miR-192 was found to be highly expressed, inducing cell death and confirming its important role in further cellular stress responses [383]. But there is even more evidence that suggests miR-192 to be involved in various age-related processes. Some miRNAs were found to be bound by high-density lipoproteins and secreted in the bloodstream, were

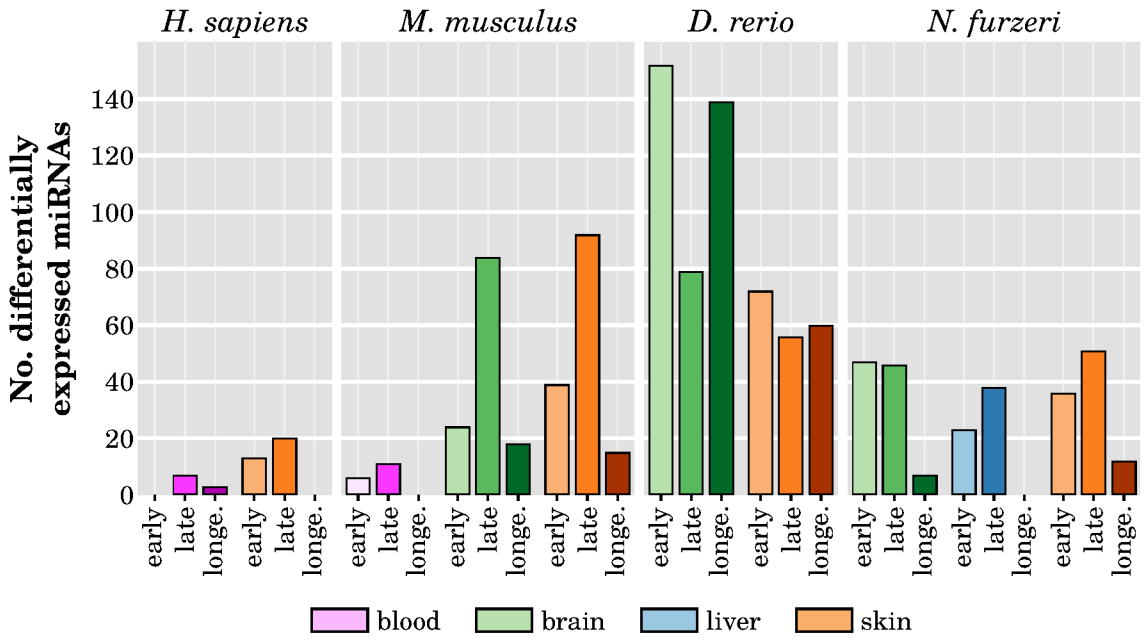


Figure 4.28: **Number of aging-related differentially expressed miRNAs by tissue type and species.** Only few aging-relate DEMs were observed in the JenAge small RNA sequencing dataset, especially in the human and killifish samples. Still, with rising age, the number of DEMs was found to increase, with only few exceptions (early – early aging comparison; late – late aging comparison; longe. – longevity comparison).

they circulate through the body and can function as regulators in distant recipient cells [384]. One of these circulating miRNAs is miR-192 and it was observed to be found in increased abundance during aging-related diseases, such as diabetes [385], degenerative loss of skeletal muscle cells [386] or cancer [384]. Taken together with our observation that miR-192 is differentially expressed in many of the sampled tissues (always upregulated with aging, except within the human skin and murin brain samples) of the four investigated and evolutionarily distinct species, this miRNA may act as a conserved key regulator of vertebrate aging. In addition, it might serve as an easy accessible diagnostic marker or even therapeutic target, because it is a blood circulating miRNA. As discussed in Chapter 1, caloric restriction was found to have prolonging effects on the life-span of various animals and was for some time believed to hold the key in understanding aging. Interestingly, it was observed in mice and rhesus monkeys, that caloric restriction can to some degree reverse the action of miR-192, drawing yet another connection between caloric restriction and aging processes on the molecular level [386, 387].
But besides miR-192, many other of the 46 miRNAs that are differentially expressed during aging in at least two different species are directly implicated with aging or age-related processes (see Table 4.5). They act in a variety of different biological functions, thus resembling the heterogenic nature of aging. Besides having identified miR-192 as a potential key regulator of a multitude of aging-associated processes, there appears to be only few evidence of conserved miRNA regulatory networks that

control aging or at least some aspects of it.

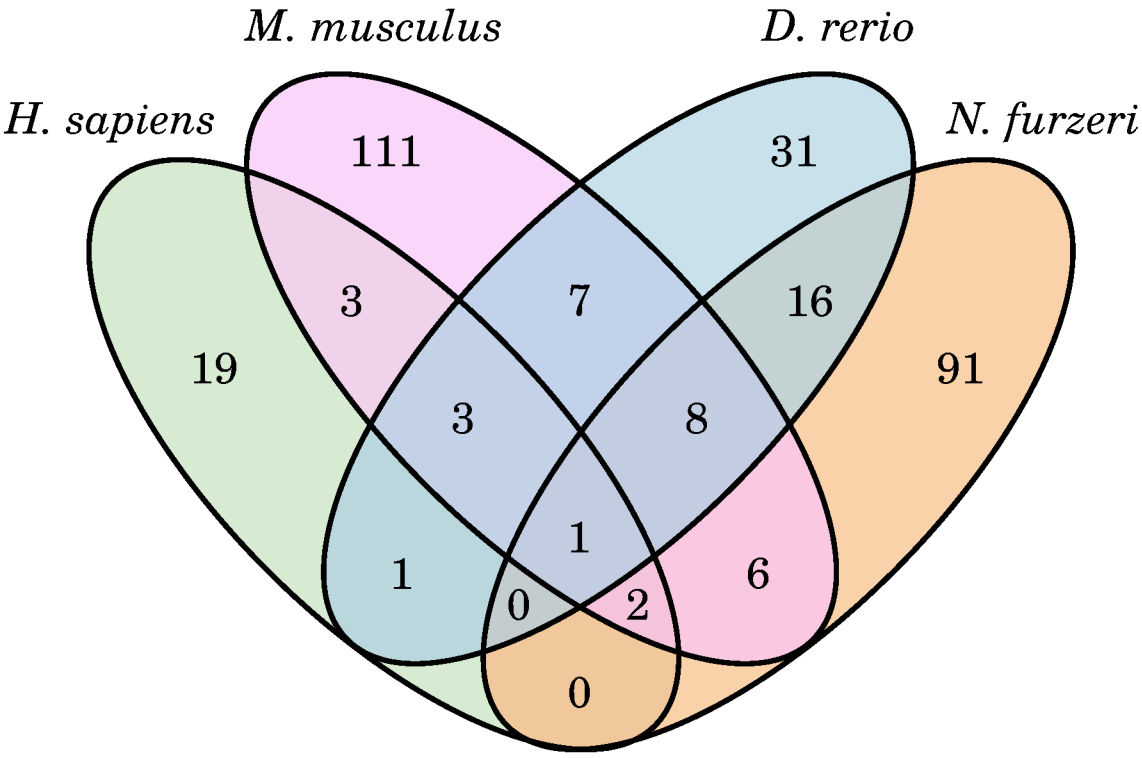


Figure 4.29: **Overlapping differentially expressed miRNAs between the four investigated species.** We found only few overlapping DEMs between the tissues of the four investigated species (see SData 7), but when overlapping all identified DEMs we found 33 to be shared by two, 13 to be shared by three and miR-192 to be shared by all four species.

Table 4.5: **Evolutionary conserved differentially expressed miRNAs with potential regulatory functions in age-related processes.** miRNA names written in bold were already reported to act in an aging-related fashion or in age-associated diseases. Unmarked miRNAs were not directly related to aging yet, but they display potential regulatory function in aging-related processes.

miRNA name	involved biological function or age-associated disease	source
let-7c/f	stem cell aging, cell cycle control, apoptosis	[369]
miR-29b	neurogenerative diseases, senescence, neuronal aging	[369]
miR-31	cell cycle control, cell differentiation	[388]
miR-34a	sirtuin-induced senescence and apoptosis	[369]
miR-92a	senescence, inflammatory stress response	[388]
miR-101a	oxidative response, neuronal aging	[369]
miR-122	IGF-signaling	[389]
miR-125b	inflammation, neuronal aging	[388]
miR-128	apoptosis, neurogenerative diseases	[388]
miR-132	sirtuin-induced inflammation, neurogenerative diseases	[389]
miR-133a	cardiovascular aging, diabetes	[389]
miR-142a	senescence, inflammation	[388]
miR-146a	senescence, inflammation	[369]
miR-181a	senescence, inflammation	[369]
miR-182	apoptosis, IGF-signaling	[369]
miR-200a/b	stem cell aging	[369]
miR-206	neurogenerative diseases, neuronal aging	[389]
miR-216a	vascular aging, senescence	[369]
miR-375	IGF-signaling	[389]
miR-543	sirtuin-controlled stem cell aging	[390, 391]
miR-22	senescence, cardiovascular diseases	[392, 393]
miR-96	apoptosis, cell cycle control	[394]
miR-129	apoptosis, cell cycle control	[395, 396]
miR-184	stem cell maintenance	[397]
miR-190a	diabetes	[398]
miR-203a/b	stem cell maintenance	[399]
miR-499	apoptosis, cardiac diseases	[400]
miR-731	senescence, apoptosis, inflammation	[401]

Chapter 5

Senescence and Inflammaging – Causes or Consequences of Aging?

This chapter is based on the publication "Conserved Aging-related Signatures of Senescence and Inflammation in different Tissues and Species".[1]

Data used and analyses performed in this chapter

Age comparison setup

The age comparisons were performed according to Figure 5.1. For each of the four species *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* we investigated three time points (mature, aged, old-age) in up to four tissues (blood, brain, liver, skin). All differentially expressed gene (DEG) results can be found at SData 1.

Collection of genes of interest and t-SNE clustering

An initial list of 769 genes relevant in the context of senescence and inflammaging was compiled from a detailed literature search (see SData 2). From those 769 genes, we removed all genes that were either missing an ortholog in any of the four species or were not differentially expressed (FDR adjusted p-value < 0.05) at least once in any species and tissue, ending with the final list of 464 genes (see SData 3).
The measured expression of these 464 genes was used to cluster the analyzed RNA-Seq samples using the t-SNE algorithm [402] (see Figure 5.2).

Gene expression coefficient of variation analysis

To estimate the general gene expression variation among the different time points of each species and tissue, we calculated the coefficient of variation for every expressed gene (*i.e.*, a TPM value greater than 1) as the ratio of its standard deviation to its mean. The significance of the variation differences among time points was determined using the two-sided t-test. We repeated this analysis using only the 464 genes of interest (for details, see SData 4).

[1] The complete supplemental material is available at `https://osf.io/kzq5y/`

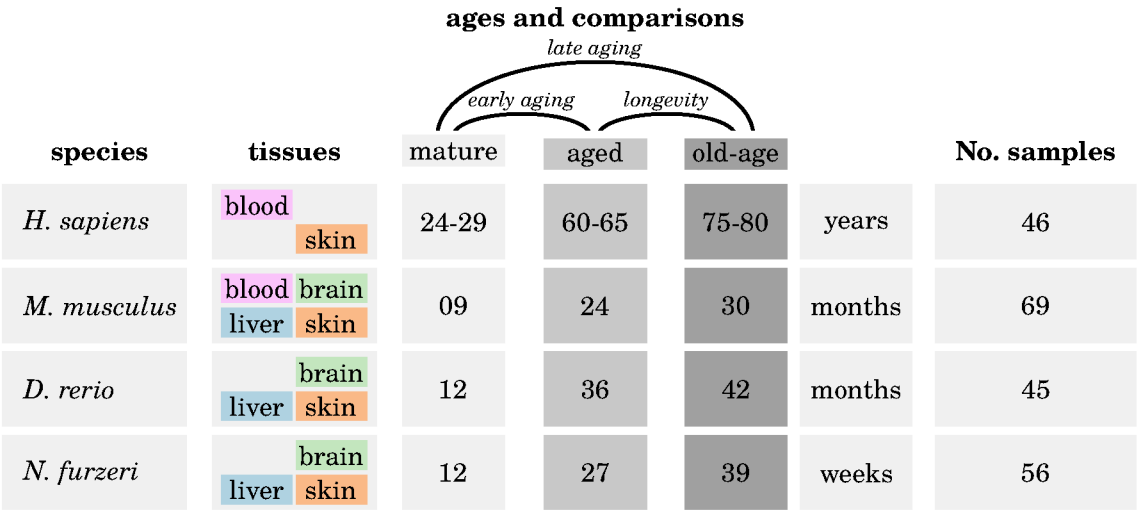


Figure 5.1: **Overview of the analyzed high-throughput transcriptomic data.** For each of the four investigated species, up to four different tissues were sampled at three different ages: one mature (M) time point, one aged (A) time point and one old-aged (OA) time point. Three different comparisons were made to reveal significant differences in expression of genes during early aging (M vs. A), late aging (M vs. OA) and longevity (O vs. OA).

Gene function annotation and expression pattern analysis

The exact biological functions of our preselected gene set of interest were obtained from the functional annotation database `David` (version 6.8) [403] (see SData 5). To identify common signatures of senescence- or inflammation-related processes, we filtered for genes that were differentially expressed at least once within one of the age comparisons and showed the same direction of up- or downregulation by a minimum threshold of 10 % in every other age comparison. This rather low effect-size threshold can be justified, because it was previously shown that expression changes related to aging tend to be more subtle [404]. This process was conducted for every species and tissue individually. Based on this gene subset, we further searched for common expression patterns of single genes among the different tissues and species. When searching for tissue-specific signatures, we applied the same filtering strategy but raised the minimum threshold of expression changes to 25 %.

5.1 How two protective systems slowly fail with age

Because aging is incompletely understood, a decades-old and still ongoing debate exists on the true source of aging, giving rise to a variety of competing theories (see Chapter 1.1.2). Senescence and inflammatory processes are two of the most common themes within these discussions of the molecular driving forces of aging. Cellular senescence is a state in which permanent replication is halted, and thus cells are unable to further proliferate, and their overall function is strongly diminished [50]. Senescence is meant to be a protective mechanism, stopping further proliferation if cells are on the verge to turn into malignant tumor cell due to severe DNA damage, for example because of telomere shortening. During an organism's lifetime, most cells continuously undergo proliferation and cell division and reach the state of senescence at their own pace, when their telomeres have reach a certain shortening threshold. Over time, this process leads to an accumulation of senescent cells accompanied by loss of function and integrity of the respective tissues, which reflects the close connection of senescence with aging [405]. Recently, cellular senescence was described as the "nexus of aging" by Bhatia-Dey *et al.*, suggesting it as the main driver of the aging process [406].

Another systemic process, that is observed in most tissues with age is the increased release of proinflammatory messenger substances. As a consequence, low-grade chronic inflammatory processes slowly but irresistibly begin to damage organs and are viewed as the cause of other age-related chronic diseases, such as Alzheimer's disease, osteoporosis or diabetes[407, 408]. This state of chronic age-dependent inflammation is also suspected as one of the main causes of biological aging and was described as "inflammaging" [409]. Miquel *et al.* proposed an integrative oxidation-inflammation theory of aging, arguing that chronic oxidative stress originating from mitochondria leads to senescence in cells of the regulatory systems, such as the immune system [410]. Unarguably, senescence and inflammation processes are strongly connected and contribute to an organism's aging phenotype as well as its rate of aging. Therefore, understanding how those systemic processes are regulated during aging in different species and tissues could supply many answers related to biological aging itself.

In this chapter, we present a descriptive transcriptomic study on the age-dependent genetic changes of senescence and inflammation in the four evolutionarily distinct species of *Homo sapiens*, *Mus musculus*, *Danio rerio* and the short-lived fish *Nothobranchius furzeri* and up to four different tissues (brain, blood, liver, and skin). Our aim was to identify potential markers of aging across species and tissues by comparing a young mature time point against an aged and old-aged time point. We report 26 different genes that showed consistent upregulation or downregulation towards old age in multiple tissues and discuss their roles in aging. Furthermore, we identified several genes that were similarly regulated during aging among the investigated species in a tissue-specific manner. Additionally, we observed a stricter control of gene expression of aging-related processes in the rather old-aged individ-

uals compared with the normally aged individuals. We conclude that certain of the identified genes that show a conserved age-dependent expression pattern are potentially interesting targets for therapeutic developments designed to achieve healthier aging.

5.1.1 Gene expression discriminates among tissues more strongly than ages

To evaluate the homogeneity of the investigated transcriptomic data, we performed t-SNE clustering of the RNA-Seq libraries based on the measured expression strengths of our 464 preselected senescence- and inflammation-associated genes (see Figure 5.2). We observed a homogeneous clustering of the samples with respect to species and tissue, with almost no segregation of the three different ages within these clusters. This observation was already made and reported in similar studies based on different sets of genes and indicates that aging is a relative subtle process, at least on the transcriptional level [119, 121]. Nevertheless, separations between the youngest and oldest time point can be observed within the skin clusters of all four species but to a much weaker extent in the human samples. This observation might suggest a pronounced and conserved difference in the activity of senescence and inflammation processes during skin aging of evolutionarily distinct species. This observations and the fact that the skin is also relatively easily accessible makes it especially interesting for interspecies comparisons of age-related senescence and inflammation.
Another observation is that although the different tissue samples form species-specific clusters, they are still well separated from each other, *i.e.*, forming tissue-specific clusters, with the exception to the *Homo sapiens* and *Mus musculus* blood samples. This observation hints that the similarities of the inflammaging process among the same tissues of different species prevails over any systematic effect within one individual.

5.1.2 Expression variance is more controlled in long-lived individuals

It is proposed that individuals that have reached a comparatively high age have somehow counteracted the effects of inflammaging by unknown anti-inflammation processes [407, 408]. As one approach to validate this hypothesis, we analyzed how the mean variance in gene expression changes during aging. Interestingly, we found a significantly lower relative standard deviation of the old-age time points compared with the aged time points in almost all species and tissues and even lower than the mature time points in selected cases (see Figure 5.3 and SData 4). This observation indicates that the selected genes associated with inflammation and senescence have a more stable expression, suggesting that both age-driving processes are in fact more controlled in the long-lived individuals. Exceptions are the skin samples of all species other than humans, which show the opposite pattern of a significantly increased variance in gene expression in the old-age time point. Since the skin is the most diverse of the tissues studied in this work (mice have fur and humans have hair, whereas fish have scales) and is the most exposed tissue, a potentially higher

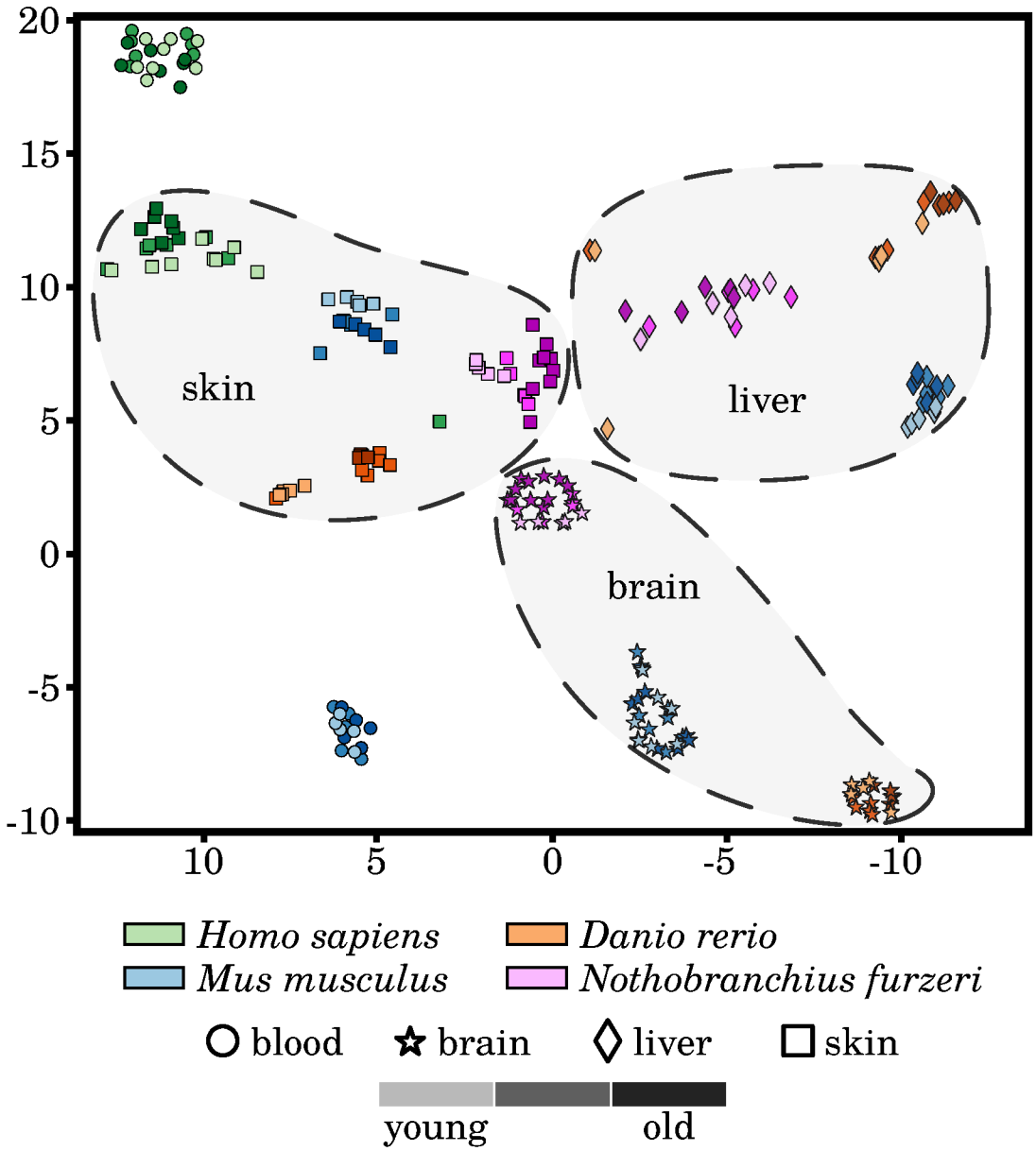


Figure 5.2: **A t-distributed stochastic neighbor embedding (t-SNE) of the analyzed RNA-Seq libraries.** All RNA-Seq samples were clustered based on the expression patterns of the selected senescence and inflammation related genes, utilizing the t-SNE approach. All tissues form distinctive species-specific and non-overlapping clusters with exception to very few single outliers. Additionally, larger species-independent tissue clusters can be observed. However, the three different time points did not generally separate in independent clusters of their own. A weak segregation can only be observed among the mature and old-aged skin samples of all four species.

diversity exists among the factors that affect aging in skin. As a result, skin cells are subjected to more and different stresses, and the rate of senescence is increased compared with other tissues, which also affects the rate of inflammation within the

skin tissues due to the pro-inflammatory secretory phenotype that senescent cells can develop [29]. In contrast, human skin showed a decreased variance in gene expression, indicating less pronounced or at least tighter controlled inflammatory and senescence processes. A possible explanation could be that humans take special healthcare precautions during their lifetime, especially in regard to protecting the skin, thus constantly reducing the amount of stress to this tissue.
The only other observed exception is the brain of *Nothobranchius furzeri*, which displayed a strongly increased gene expression variance at the old-age time point compared with both younger time points. This pattern is similar to that in fish skin and is already reflected in the sample clustering as mentioned above (see Figure 5.2). This observation could indicate a weaker protection of the *Nothobranchius furzeri* brain against inflammaging, but it remains to be further studied.
To verify that the more controlled gene expression observed in long-lived individuals is not biased due to our preselected genes, we repeated the same analysis with all expressed genes in every species and tissue. We confirmed our observation with all tissues showing the same relative variance changes as before, in principal. The overall variance within each age group was decreased, which was expected because many more genes were included in the analysis that show no change in expression during aging.

5.1.3 Oxidative stress response tends to be the prevailing process towards old age

To examine whether certain biological functions predominantly drive the inflammaging processes in a tissue- or species-dependent manner, we analyzed the identified differentially expressed genes (DEGs) with respect to their molecular function. We found that most of the DEGs within the brain and liver samples belong to immune and inflammatory response processes, whereas changes in the expression of senescence-related genes are predominately related to skin aging (see Figure 5.4 and SData 6). Only few DEGs could be identified in blood age comparisons for humans and mice, making it difficult to interpret these results in a meaningful way.

It is a known and common observation of aged skin that it is more susceptible to infections, physical damage and reduced epidermal barrier integrity as well as other age-related deficiencies [411]. Epidermal stem cells maintain the tissue's homeostasis and loss of those stem cells due to premature senescence is the main cause of aging within the skin [412]. Our data confirm this observation, showing mainly senescence-related DEGs within the skin of all four species already in the early and the late aging comparisons. However, whether the source of epidermal stem cell senescence is primarily intrinsic or extrinsic factors is controversially discussed [413, 414]. Apoptosis appears to play a more important role during aging of the liver than in any of the other investigated tissues. This is a confirmed observation, because liver homeostasis is mainly regulated through apoptotic processes and many liver dysfunctions and diseases are related to apoptosis [415, 416].

Most interestingly, we observed that DEGs associated with the oxidative stress re-

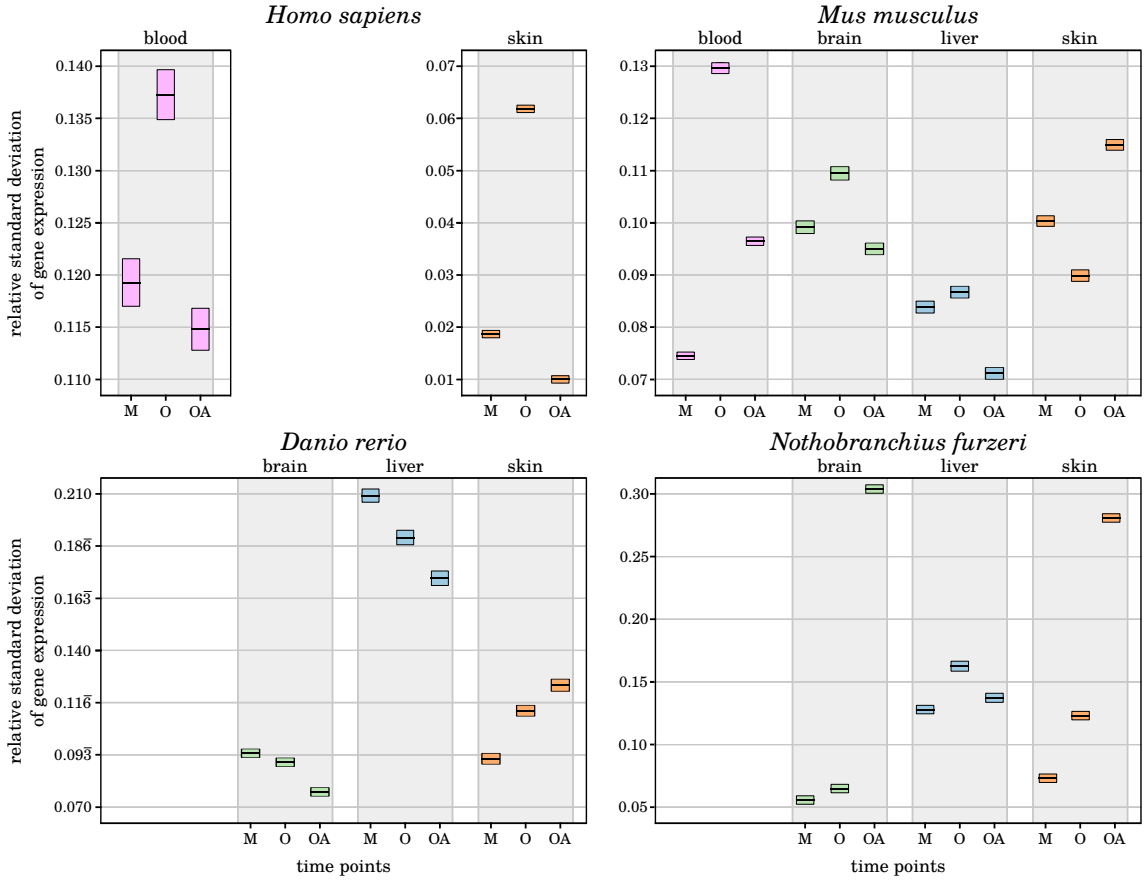


Figure 5.3: **Change of relative standard deviations of gene expression in all four species with age.** For each investigated species and tissue, the measured variance in transcript expression of the preselected senescence- and inflammation-related genes is displayed for every time point (M – mature, O – old, OA – old-aged). The upper and lower bounds of the box plots represent the respective 2.5 % percentiles. All displayed differences in the mean variance of gene expression among the aged and old-aged time points and almost all other age comparisons are significant (p-Value < 0.05) within each tissue and species, individually. A general decrease of variance in gene expression can be observed in the old-age time points, except for the skin samples of *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*. This indicates a tighter control of inflammaging processes in long-lived individuals, reducing negative effects and helped them to reach the high age. For detailed information, see SData 4.

sponse and oxidative processes occur predominantly within the late age and longevity comparisons (*i.e.*, comparisons with the long-lived individuals) of the brain, liver and skin samples. This observation suggests that those individuals that grew older than their species average survival age have either a generally more active or better regulated response mechanism against reactive oxygen species (ROS). Free radicals such as ROS and their great impact on the aging process are much discussed [410, 417]. During respiration of oxygen, free radicals are produced as a harmful byproduct that can oxidize macromolecules such as DNA or proteins and damage them in this

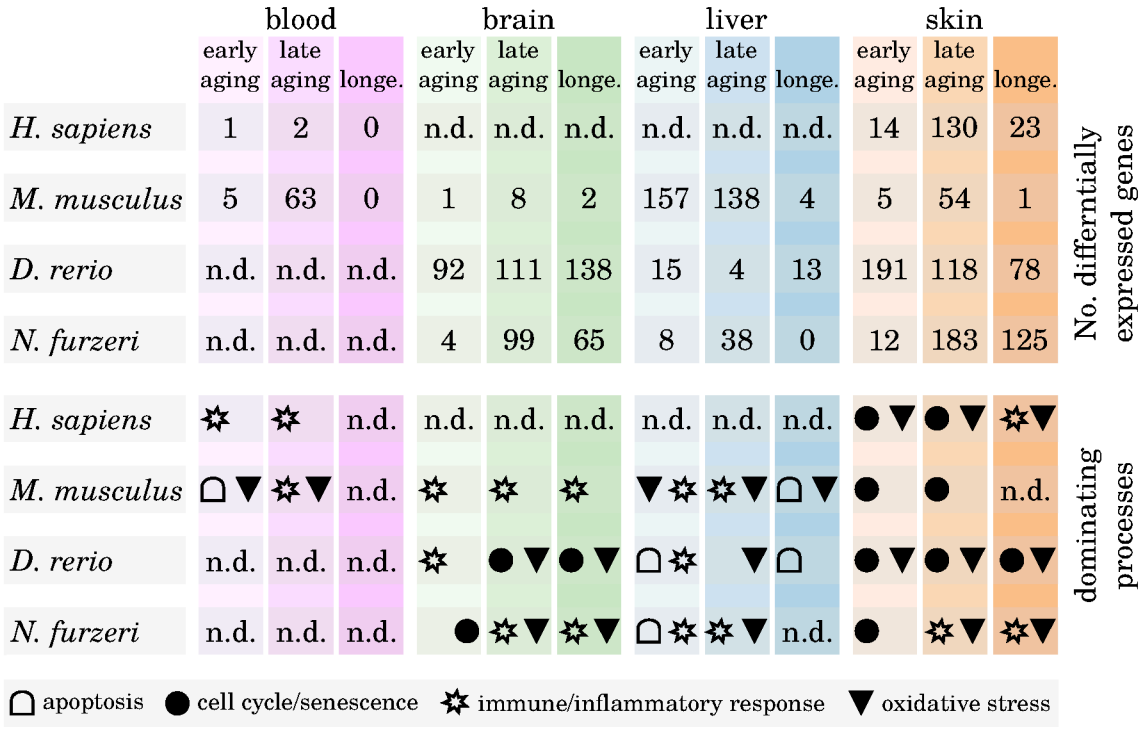


Figure 5.4: **Differentially expressed genes during tissue aging and associated biological processes.** The upper part shows the number of identified differentially expressed genes (DEGs) within all three age comparisons (early aging, late aging, longevity) in every investigated species and tissue. The lower part shows for every comparison the dominant biological processes as determined by the majority of the annotated functions of the respective DEGs. Note that in some comparisons only few DEGs could be identified and hence may only poorly reflect the underlying age-related processes. For detailed information, see SData 6.

process. The main point of aging theories involving oxidative stress is that with age, the rate of harmful ROS accelerates because ROS first target the mitochondria (because they are mainly generated in that location), leading to an accumulation of damage in these cell organelles. The more degenerated the mitochondria become, the higher the rate of released ROS becomes, injuring genomes and membranes of neighboring cells and, as a final after-effect, resulting in their senescence or apoptosis and therefore aging [417]. Different antioxidant mechanisms exist within the cells to prevent the release of free radicals and protect the cells from such deleterious consequences [418]. It was observed that long-lived species appear to have higher protection against ROS than short-lived species [419, 420]. Our data suggest that the same is true for long-lived individuals within one species, confirming the importance of oxidative stress response for longevity.

5.1.4 Conserved aging expression signatures across tissues and species

To identify a more precise signature of the abovementioned observations (Figure 5.4), we analyzed the expression patterns of our preselected gene set in greater detail. We were particularly interested in genes that show a constant increase or decrease towards old age that could also be consistently observed between the investigated tissues and species.
We found 16 genes that were consistently higher expressed and 10 that were consistently lower expressed during aging in at least three of the four species and tissues, totaling 26 genes that show a conserved expression pattern with aging (see 5.5). Certain of these genes have already been reported in the context of aging, and we observed them to be the most conserved genes from our preselected set among species and tissues. Hence, we discuss several of these genes based on their biological function in additional detail. More senescence- and inflammation-related genes were found to have conserved expression patterns but to a lesser extent and can be found in SData 7.

Immune/inflammatory response

The gene *marco* encodes for a scavenger receptor (Marco – macrophage receptor with collagenous structure), which is typically found on macrophages but also other immune cells [421, 422], and acts in the innate immune system by binding and clearing pathogens and initiating an inflammatory response [423, 424]. We observed *marco* to be constantly upregulated (mostly by an increase of two-fold or more) towards old-age in all four species and within the blood, liver and skin. Although Marco does not cause inflammation directly, it is important for the activation of other receptors, such as members of the toll-like family [424]. These receptors subsequently activate NF-κB, inducing an inflammatory response. Interestingly, Marco is associated with Alzheimer's disease, because it is also present on microglia and a decreased response of these cells has been observed after binding amyolid beta peptides [425].
The coagulation factor II receptor (PAR1) is encoded by the gene *f2r*, which similar to *marco* shows persistent upregulation with age in blood, liver or skin within all four species. The major function of PAR1 is mediation between coagulation and inflammation activity, and therefore it has a key role in inflammatory response activation [426, 427]. In addition, PAR1 activity is implicated in aging-associated cardiovascular diseases [428].
Four more genes (*cd40*, *sh2d1b1*, *ptafr*, *adam8*) coding for membranous receptors have a common upregulated signature through several aging tissues and species. These genes have all been shown to promote pro-inflammatory signaling, either directly by binding platelet-activating factors (*ptafr* [429]), activating antigen presenting cells (*cd40* [430]), and acting as a regulator of antigen receptor signal transduction (*sh2d1b1* [431]) or more indirectly by releasing and degrading other cell surface receptors of leukocytes (*adam8* [432]).
Observation of a generally higher expression of ceruloplasmin encoded by the gene *cp* in all species and tissues with age (except *Nothobranchius furzeri* brain) is not unexpected due to its role as an acute-phase protein. These reactants are increas-

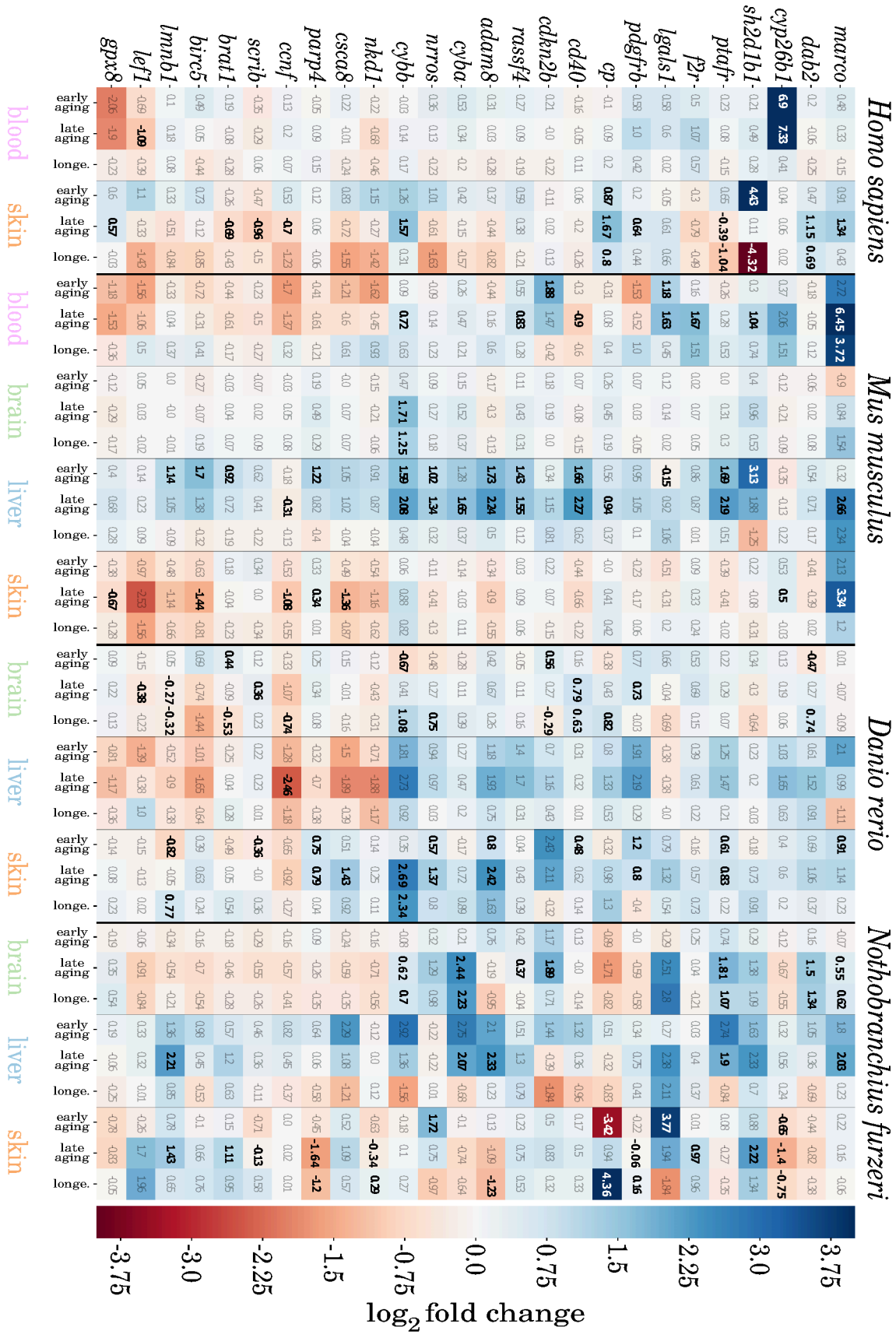


Figure 5.5: **Heatmap representation of potentially conserved senescence- and inflammation-related genes.** Genes are represented as mouse orthologues (a complete list of gene orthologues can be found in SData 8). Numbers indicate log_2 fold changes between two compared ages, where a positive value indicates an upregulation (blue), and a negative value downregulation (red) of the respective gene with aging. All significant changes in gene expression are indicated in bold. For detailed information, see SData 7.

ingly released in the blood plasma during an inflammatory response and support the innate immune response [433]. Thus, increased *cp* expression is most likely a result of the chronic inflammatory nature of aged tissues.
Lef1 is one of the few immune-related genes that showed a decreased expression with age in the blood and skin of human and mice respectively and the brains of both examined fish. The main function of the LEF1 (lymphoid enhancer-binding factor 1) protein is to enhance expression of the T-cell receptor alpha chain [434]. However, a variety of different interaction partners are known, linking LEF1 as an important regulator within the WNT and TGF pathways and implicating a role in apoptosis and cell proliferation of leukocytes [435, 436].

Oxidative stress response

As previously mentioned, ROS and oxidative stress are widely believed to be the driving force behind the cellular aging process [410]. The source of oxidative stress is likely an imbalanced activity of the respiratory chain and antioxidant processes. We observed a constant age-depended upregulation of the genes *cyba*, *cybb*, *cyp26b1* and *nrros* in various combinations of the investigated tissues and species. The first two genes encode for the light and heavy chains of the Cytochrome b-245 protein, a superoxide-producing subunit of the NADPH oxidase [437]. By producing and releasing ROS, phagocytes use Cytochrome b-245, mainly as an antimicrobial strategy during an infection. However, the role of Cytochrome b-245 during aging is still debated [438, 439], and it is associated with several degenerative diseases due to uncontrolled production of ROS [440, 441]. *Cyp26b1* encodes another cytochrome protein, Cytochrome P450, which is involved in many metabolic reactions, especially the oxidation of NADPH [442]. Higher expression of *cyp26b1* during aged time points is suggested to aid in degradation of toxic substances that have accumulated with age [443].
The genes *nrros* and *gpx8* both code for proteins that belong to antioxidant mechanisms controlling the production of ROS. NRROS directly interacts with Cytochrome b-245 and mediates its degradation, limiting the rate of ROS production of the associated NADPH oxidase complex [444]. Similarly, the enzyme GPX8 protects cells from ROS-induced damage by its peroxidase activity by reducing free hydrogen peroxide to water [445]. However, whereas gene expression of *nrros* rises with age, most likely to regulate the increased Cytochrome b-245 activity, the translation of *gpx8* is downregulated in many tissues and species. We observed the opposite regulation during aging only in the human skin and mouse liver samples. Nevertheless, *gpx8* serves as an interesting potential therapeutic target and it was previously shown in the *Caenorhabditis elegans* model that the deletion of several GPx family members is the cause of an accelerated aging process that results in shorter life-span [446].

Senescence and apoptosis

Most of the 26 identified genes with conserved expression patterns in aging, are strongly related to the cell cycle control or apoptotic processes. Both replicative

senescence and apoptosis are immediate causes for the loss of somatic and stem cells with age and can be similarly triggered by chronic inflammation and oxidative stress [33, 34].
We found four of these genes to encode direct inducers of senescence: *ccnf* (Cyclin-F) [447], *scrib* [448]. *cdkn2b* ($p15^{INK4b}$) [449] and *rassf4* encoding an inhibitor of Cyclin D1 [450]. We found all of these genes to be expressed in a senescence-promoting fashion towards aging (*rassf4* and *cdkn2b* upregulated; *scrib* and *ccnf* downregulated) in the examined tissues and species. Additionally, selected genes interacting with cell cycle regulators also showed persistent expression towards old age, namely, *dab2* [451], *lgals1* [452], and *brat1* [453].
The proteins CDCA8 and Survivin encoded by *cdca8* and *birc5*, respectively, are both associated with apoptosis. By inhibiting caspase activity, Survivin is a repressor of apoptosis, whereas CDCA8 directly interacts and stabilizes Survivin [454, 455]. However, both proteins show converse directions of regulation during aging between the same tissues and within individual species, resembling rather weakly conserved expression signatures.

5.1.5 Potential common tissue-specific marker genes of aging

We also focused on senescence- and inflammation-related genes that change their expression during aging consistently and similarly in a tissue-specific manner. These genes should present conserved age-related changes and might be the driving factors of tissue-specific aspects of the inflammaging process, probably closely linked to the specialized function of the respective tissue. In a first comparison of the identified DEGs of the different species, we observed only a few common genes that occur as differentially expressed in all species of a single tissue (see Figure 5.6 and SData 9). However, the blood and liver samples revealed no DEGs to be common in all of the respective species, and two were identified within the brain (*cybb*, *cd68*) and another two within the skin (*Gpx7*, *Tnfrsf25*). *Cd68* is a member of the cell surface receptors of macrophages and other monocytes and is a well-known marker for macrophage activation. The Cytochrome b-245 subunit encoding gene *cybb* and its potentially harmful role during aging were already introduced in the previous section. Tnfrsf25 (a member of the TNF recptor superfamily) stimulates the proliferation of T cells or can initiate apoptotic signaling, leading either to survival or cell death, and is therefore an important regulator of T cell development [456]. *Gpx7* encodes for another glutathione peroxidase and has the same antioxidant function as its above described homolog *gpx8*.
However, none of the four genes exhibits features that serve as a tissue-specific marker because they do not display a consistent up- or downregulation during aging within each investigated species or share the same expression pattern with other tissues. Therefore, we applied a similar but more stringent filtering process to reveal genes displaying the characteristics of potential tissue-specific markers. With this approach, we identified several genes that showed a constant age-related transcriptional increase exclusively in the liver or skin samples of the investigated species.

Liver

Four constantly upregulated genes were identified in the liver samples of *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*: *jag2*, *anax1*, *ralb*, and *sfrb*. The protein encoded by *jag2* (Jagged-2) is a known activating ligand of Notch2, which is mainly involved in many different developmental processes regulating cell fate decisions [457]. Most interestingly, *jag2* overexpression is not only shown to play a critical role in the formation of plasma cell myeloma and progression of other tumors [458, 459] but also induces the secretion of interleukin-6 [460]. Interleukin-6 is a potent stimulator of immune and inflammatory responses and acts in the development and progression of many age-associated diseases, such as Alzheimer's disease [461], atherosclerosis [462], diabetes [463] and various cancers, making it an attractive therapeutic target [464, 465]. Because the liver shows strong signs of inflammation during aging (Figure 5.4), overexpression of *jag2* might be one conserved driving factor of inflammatory processes within hepatic cells and could be of interest as an additional therapeutic target. In contrast, Annexin A1, the protein encoded by *anax1*, has distinct anti-inflammatory and protective properties, by inhibiting NF-κB signal transduction and counterregulating pro-inflammatory signals in a variety of immune cells [466, 467]. The age-related upregulation of Annexin A1 in the liver could be a component of a tissue-specific mechanism attempting to cope with the chronic state of inflammation. SFRP2 (encoded by *sfrp2*) has an oncogenic character and is associated with cancer formation by acting as an inhibitor of the canonical WNT/βcatenin pathway [468, 469]. The last of the four conserved liver-specific upregulated genes, *ralb*, encodes for one of two Ral protein paralogs (Ral-B) and acts as a major modulator of a multitude of cellular processes [470]. Although many functions are shared between the paralogs Ral-A and Ral-B, the latter is more specifically involved in the activation of apoptosis [471], which we observed to occur more predominantly during liver aging (Figure 5.4). Additionally, Ral-B has clinical significance because it promotes tumor progression of several cancers and activation of the innate immune response [470, 472] and hence might contribute to inflammatory stress in the liver.

Skin

If including all four species, we did not observe any gene displaying the characteristics of a potential conserved marker gene for skin aging. Considering the diverse nature of skin among the studied species we decided to examine the skin in two separate groups of, fishes and mammals.
Between human and mouse, only *pecam1* was identified to be consistently upregulated towards old age. The gene encodes the protein Pecam-1, also known as CD31, which is a common immunohistochemistry marker used to evaluate tumor growth, and has a major role in the removal of old neutrophils, which could explain its higher expression with age. However, this gene is already known as a biomarker of inflammatory processes [473].
The two fishes *Danio rerio* and *Nothobranchius furzeri* share 6 genes that showed a persistent increased expression with aging exclusively in their skin and their en-

coded proteins are all involved in mitotic progression of the cell cycle: *ckap2* [474], *cdca8* [455], *aurkc* [475], *cep55* [476], *mis12* [477] and *spc25* [478]. This observation might indicate that fish skin cells are not as affected by senescence as the skin of humans and mice, which is in line with the observation that the skin of zebrafish shows a high regenerative capacity even at higher age [479].

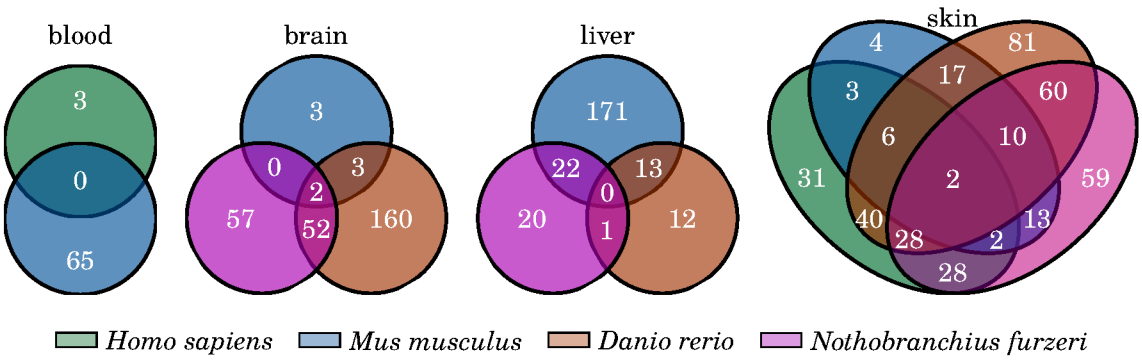


Figure 5.6: **Common inflammation- and senescence-related genes that are significantly changed with age.** Venn diagrams showing the overlap of the identified differentially expressed genes among the four investigated tissues. Only few genes are commonly differentially expressed among all species of any of the four tissue comparisons. For detailed information, see SData 9.

5.2 Age-related stress response is similar in vertebrate species

Within this chapter, we have investigated the inflammation- and senescence-related gene expression during aging in multiple tissues from four evolutionarily distinct species. By analyzing the gene expression profiles, we were able to identify common signatures of aging, *i.e.*, genes that illustrate steady increased or decreased expression with age. We identified 26 genes that shared an age-dependent expression pattern in at least three of the four investigated species and at least three different tissues. These genes represent interesting targets for further study, because they hint at general molecular mechanisms of aging that occur similarly not only within various organs of one organism but also between different species. It is important to note that we focused our study on a specific set of senescence- and inflammation-related genes and other processes might exist that share similar expression changes during aging across distinct species.

Several of the identified genes are not only directly involved in the initiation of an inflammatory response but also play a major role in sustaining the state of inflammation. Modulation of the expression level of certain these genes, such as *marco* and *f2r*, or addressing the respective proteins could be one potential approach to controlling chronic inflammation and thus might result in reduced inflammaging-related cellular stress with aging. In total, six different genes coding for macrophage

cell surface receptors (*marco*, *f2r*, *cd40*, *sh2d1b1*, *ptafr*, *adam8*) have been found to show conserved upregulated expression across different tissues and species. Accumulation and higher activity of macrophages has already been linked to aging and removal of macrophages has shown beneficial effects for age-related disorders, such as neuro-degeneration, or atherosclerosis [480–482].
Another strongly aging-associated intrinsic stress factor is the production of cytotoxic ROS as a byproduct of the respiratory chain. Although phagocytes use Cytochrome b-245 (a heterodimer encoded by the genes *cyba* and *cybb*) to produce ROS to kill microbes, excessive generation of ROS causes premature replicative senescence of cells due to DNA damage [13]. We observed a conserved age-dependent upregulation of selected genes involved in ROS production (*cyba*, *cybb*, *cyp26b1*) together with a similarly upregulated but apparently inefficient (*nrros*) or even downregulated oxidative stress response (*gpx8*). Most of the identified genes with conserved expression patterns were related to senescence and apoptosis and showed an increase in both processes with age in general. As noted by many different theories on aging, each of the mentioned cellular processes is suspected to be the main driving force behind biological aging. However, it has become increasingly apparent that these processes should not be treated as separate and seemingly competing sources of aging, because they are strongly interconnected [13]. During a state of inflammatory stress, excessive production of ROS is enhanced due to dysfunction of the respiratory chain in mitochondria, leading to an accumulation of senescent cells, which can develop a pro-inflammatory secretory phenotype [29]. In addition, inflammation is a known trigger of apoptotic processes and during the process of apoptosis, certain cellular components are released that can further activate inflammatory processes, adding to the overall cellular stress [35]. As a consequence, an environment of chronic inflammatory, oxidative, senescent and apoptotic stress is established, with these processes mutually triggering each other and boosting the cycle of self-harming development [35] over time. Additionally, constant stimulation of cellular stress response mechanisms promotes genetic deregulation, which is reflected in the observed gene expression changes.

Based on the annotated functions of the identified DEGs, we observed that the individual tissues express different aging-related processes more strongly than others. Although the impact on age-related expression changes of immune and inflammatory response processes was stronger in the blood, brain and liver samples of all investigated species, the skin samples displayed greater modulation of cell cycle and senescence-associated genes. This observation is most likely due to the specialized functions of these different organs. Only few tissue-specific and conserved regulated genes could be observed in the liver and skin and none in the blood and brain samples. Nevertheless, the identified genes potentially represent molecular sources for why certain aging-related processes appear more tissue-specific than others.

Additionally, we report that gene transcription in long-lived individuals is generally more controlled compared with average-lived individuals, showing significantly lower variance in gene expression in all tissues and species, except the skin. Maintaining a more stable transcriptional activity, not only for senescence and inflammatory

response processes, appears to have a significant life-prolonging effect [27]. It is also considered that oxidative stress response processes were observed to be more regulated in the old-age comparisons, and a possible reason could be better management of oxidative processes and more efficient antioxidant mechanisms. However, this conclusion remains speculative until further experimental studies can prove this observation.

Overall, in this chapter, we describe previously unknown conserved transcriptional changes across different species and tissues as well as tissue-specific changes with age, supplying a complementary overview of how changes in gene expression relate to processes of aging. In addition, our findings could serve as a basis for new strategies in the development of therapies against aging-related diseases.

Chapter 6

Changes of Alternative Splicing with Age

This chapter is based on the publication "The landscape of the alternatively spliced transcriptome remains stable during aging across different species and tissues".[1]

Data used and analyses performed in this chapter

Age comparison setup

The age comparisons were performed according to Figure 6.1. For each of the four species *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*, we investigated four time points (young mature, mature, aged, old-age) in up to four tissues (blood, brain, liver, skin). All differentially spliced gene (DSG) results can be found in SData 1.

Detection of alternatively spliced genes and differentially expressed spliceosomal genes

We analyzed differential alternative splicing (AS) using `rMATS` (v3.0.8) [483]. This tool has been applied successfully in other publications and showed a high precision in AS identification [484]. Only results with a false discovery rate below 0.05 and an absolute inclusion level difference of AS events per sample (`IncLevelDifference`) above 0.1 were considered [485]. In addition, we manually controlled the results exemplary to ensure that AS takes place as predicted and whether it is already annotated as alternative transcript according to the respective species annotation. Furthermore, we identified the position of each AS event relative to the gene range based on the prediction of `rMATS`.

We selected spliceosomal genes from `Uniprot` [486] and `Ensembl Biomart` [122], because the macromolecular spliceosome not only consists of proteins but also of small non-coding RNAs [487]. Based on these gene lists, we analyzed differentially expressed genes with spliceosomal activity for each species and the respective tissues. Differentially expressed genes were identified using the `DESeq2` (v1.10.0) [128]

[1] The complete supplemental material is available at `https://osf.io/rz6kc/`.

A

species	tissues	ages: young mature	mature	aged	old-age		No. samples
H. sapiens	blood, skin	24-29	45-50	60-65	75-80	years	121
M. musculus	blood, brain, liver, skin	09	15	24	30	months	92
D. rerio	brain, liver, skin	12	24	36	42	months	60
N. furzeri	brain, liver, skin	12	20	27	39	weeks	71

B

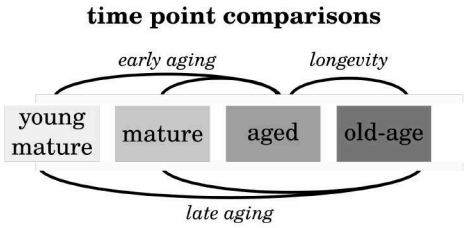


Figure 6.1: **Overview of species, tissues, time points and comparisons of the utilized RNA-Seq libraries.** (A) Samples of *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* were taken from up to four different tissues (blood, brain, liver, skin) at four different time points. The sampled time points can be assigned to different ages: young mature, mature, aged and old-age. (B) To explore the changes of alternative splicing, we studied three expression comparisons: early aging (young mature and mature vs. aged), late aging (young mature and mature vs. old-age), and longevity (aged vs. old-age). All comparisons were performed individually for each of the studied species and their tissues.

Bioconductor package, comparing the four time points of each species and tissue, individually. False discovery rate adjustment of the resulting genes p-values was performed according to Benjamini *et al.* [129].

Quantification of isoform expression

To determine the expression of all annotated isoforms independent of differential expression, we applied the tool `Stringtie` (v1.3.3b) [488] for abundance estimation separately for each sample. We calculated only the abundance of each annotated isoform without predicting new isoforms. Normalizing the abundance of isoforms was executed using transcripts per million (TPM) [126], we considered only those as expressed that had a TPM>1 in at least one sample. This information was used to compute the number of expressed isoforms per gene. For each gene with multiple expressed isoforms, we identified its predominant isoform and changes in the isoform frequencies over time.

Gene set enrichment analysis

Annotated gene functions of the differntially spliced genes of the early aging, late aging and longevity comparisons for each species and tissue, were obtained from the `David` functional annotation database [403].

Protein domain identification

For each gene that showed a switch in its predominant isoform during aging, we investigated the encoded protein domains using `Interproscan` (v5.29-68.0) [489]. For those genes with multiple predominantly expressed isoforms, we identified the corresponding protein domains and compared them among each other, counting the loss or gain of encoded protein domains of a certain gene over time.

6.1 Accumulation of alternative splicing events: A failing regulatory system during aging?

As we have extensively discussed in Chapter 1, aging is unarguably one of the most complex processes of human and animal biology and can be defined as a time-dependent, constant weakening of tissue homeostasis, resulting in various frailties and age-associated diseases. A key part in maintaining correct cellular function and thus healthiness of the organism is proper gene expression and regulation, where alternative splicing (AS) of RNA transcripts plays an important role.
AS is an important co- and post-transcriptional process in eukaryotes that enables altered forms (also referred to as isoforms) of mRNA molecules of a gene, leading to multiple mature transcripts of a single gene with possible different or modified encoded functionality [490]. Thus, AS mainly contributes to the increase of an organism's transcriptomal diversity, allowing to synthesize multiple different proteins from a single protein-coding gene [491]. Initially, AS was considered to be a relatively rare event of genetic regulation, however, the majority of eukaryotic multi-exonic genes undergo AS [492] and AS functions in several important biological processes, partially in a tissue-specific manner [487]. The splicing process is conducted by the macromolecular spliceosome consisting of two spliceosomal complexes, the major and the minor spliceosome, which themselves are assembled from different subunits involving a multitude of proteins and maintaining RNA molecules [492, 493]. Several modes of AS are known, among them exon skipping (ES), intron retention (IR), alternative 5' splice sites (A5S), alternative 3' splice sites (A3S), and mutually exclusive exons (MXE) [490]. While all of these modes may serve the same goal, *i.e.* the alteration of mRNAs by different specific mechanisms, their frequencies seem to vary in a species- and even tissue-specific way [487, 494].
Since AS provides a common level of genetic regulation in almost all biological processes, several age-related diseases are linked to either specific harmful misspliced gene variants or to a general misregulated or dysfunctional splicosomal activity. A well studied example is the Hutchison Gilford progeria syndrome, were a silent point mutation in the *lmna* gene adds a 5' splicing signal, leading to a shortened transcript and subsequently a shorter version of its encoded protein Lamina A [495]. Because

Lamina A is an important structural protein of the cell's nucleus and its shorter version is non-functional, the nuclei of affected cells are misshaped, resulting in impaired chromatin organization and limited cell division [496]. There are many more examples of the implications of AS and age-related diseases, such as Alzheimer's disease [497], Parkinson's disease [498] and different types of cancer [499].
A few recent studies on various species and tissues investigated the change in spliceosomal activity during aging by measuring the transcription of the respective spliceosomal genes, observing an age-dependent decline in their expression [500–503]. As a common assumption, it is argued that the age-dependent decrease in the spliceosomal activity leads to an inaccurate splicing process resulting in an accumulation of non-functional mRNAs. This additional 'noise' in the transcriptome might reflect another source of stress, which aging cells are exposed to, adding to the intrinsic obstacles to maintain homeostasis. Furthermore, Rodriguez *et al.* could directly report an increase in the number of alternatively spliced genes with age in different tissues of mice [504]. Nevertheless, most other studies on AS and its implications to aging are either focused on the age-dependent differential expression of splicing factors or differentially expressed isoforms of specific genes within one species and tissue [500–503, 505]. Only a small number of studies exist, that directly try to assess the general impact of aging on the heterogeneity of a cell's transcriptome. In an RNA-Seq based study, Wood *et al.* reported changes in the usage of certain isoforms within the aging rat brain [154]. Further research is required to understand the age-related change of the isoform landscape and evaluate the true influence between AS and aging.
In this chapter, we investigate and compare the global impact of the various modes of AS on the expressed transcripts in different species and tissues during aging. We analyze a multitude of cross-sectional RNA-Seq data from four different species (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Nothobranchius furzeri*) from up to four different tissues (blood, brain, liver, skin) at four distinct ages. Whereas we identified a great number of differentially spliced genes, with many of them being involved in post-transcriptional mRNA processing pathways, we did not observe a widespread increase of spliced genes with age. Also, the average number of isoforms per gene remained constant within the investigated species and tissues. Additionally, we investigated the switch of isoforms during aging, which showed only little effect on the encoded proteins in respect to their functional domains. Finally, we could confirm a decline in the expression of major and minor spliceosomal genes and selected associated splicing factors in all investigated tissues and species, but to different extends.
We conclude that despite the acknowledged influence of single (mis)spliced isoforms on age-associated processes, AS remains in general stable and likely plays a minor role during normal physiological aging.

6.2 The landscape of the alternatively spliced transcriptome is only minorely affected by aging

6.2.1 Only marginally differences in the general number of expressed isoforms per gene with age

As a starting point, to investigate the general impact of aging on the heterogeneity of the transcriptome, we measured the average number of expressed isoforms per gene for each age group in all examined species and tissues (see Figure 6.2). Whereas clear differences between the tissues within each species have been observed, neither any significant age-dependent change in the average number of isoforms per gene was found, nor does the proportion of genes with multiple expressed isoforms change significantly over time (see SData 2). On average, the actively transcribed genes of both mammalians tend to express two different isoforms in parallel, while both fishes tend to have only one active isoform per gene. However, this finding could be slightly biased since the isoform annotation of protein coding genes is more comprehensive for *Homo sapiens* and *Mus musculus* than for *Danio rerio* and *Nothobranchius furzeri*, with on average about seven, five, three, and four annotated isoforms per protein coding gene, respectively. Nevertheless, there are several genes that have more than two actively transcribed isoforms during any investigated time point (see SData 3). The brain and skin samples of the *Nothobranchius furzeri* old-age time point displayed highly increased variances in the amount of expressed isoforms compared to the other time points, possibly hinting at a more deregulated splicing activity. However, this observation could not be confirmed in general. Thus, as a next step, we identified genes that either constantly increase or decrease the number of expressed isoforms with aging. Only few genes matching these criteria have been observed. *Nothobranchius furzeri* displayed the most of such genes with 14 genes showing a constant increase and 28 genes showing a constant decrease of parallel expressed isoforms (see SData 4). Interestingly, we found many of the genes to be membrane or membrane associated proteins in all investigated species and tissues, showing this particular pattern of age-related rise/decline of isoform numbers. This is most likely due to the regulatory role of AS, which especially determines folding, topology, solubility and, thus, functions of membrane proteins, displaying its decisive task for the respective mRNAs [506]. In general, the specific annotated biological processes of the identified genes were found to be more diverse and could not be further clustered into significantly enriched molecular functions. Nevertheless, these genes still represent interesting targets for further studies, such as *foxm1* and *gas5*, which we will discuss exemplarily.

The forkhead box protein M1 that is encoded by the *foxm1* gene, represents a transcription factor of the FOX family and regulates important biological processes like cell proliferation, cell development but also play a part in longevity [507, 508]. Foxm1 has a key role in cell cycle progression as well as chromosomal segregation and genomic stability and therefore is tightly linked to the age-associated processes of cellular senescence and cancer formation [509, 510]. Three different isoforms are known to be transcribed from *foxm1*, where one acts as a transcriptional repressor (isoform A) and two as activators (isoform B and C), associating the latter two to the

proto-oncogenic character of *foxm1* [510]. We observed a constant age-dependent decrease in the number of expressed isoforms of *foxm1* in the mice skin samples, with all three isoforms being actively transcribed in the first two mature time points then being reduced to two in aged mice and finally just one in the old-age time point. When neglecting the TPM low expression threshold, the same trend could be observed for the human skin samples. All other tissues showed a constant but relatively low expression of just one isoform within all age groups, with exception to the skin samples of *Danio rerio*, expressing the opposite pattern during aging, with just one actively transcribed isoform at the young mature time point and rising to have all three isoforms expressed at the latest old-age time point. This observation might suggest an age-related control of splicing of *foxm1*. For *Nothobranchius furzeri*, however, only one known isoform is currently annotated, so no predication could be made.
The other interesting gene, *gas5*, does not encode a protein but a long non-coding RNA (lncRNA), which was shown to be involved in apoptotic and tumor formation processes [511]. Several different isoforms are known and predicted for *gas5*, of which some host various small nucleolar RNA (snoRNA) sequences [512, 513]. But expression of *gas5* is also strongly linked to growth control and senescence of human T-cells, indicating a potential role in the age-associated phenomenon of immunosenescence [514]. We observed a similar age-dependent decline in the number of transcribed isoforms of *gas5* in the liver and skin samples of *Mus musculus* and *Danio rerio*, showing the lowest number of parallel actively expressed isoforms at the old-age time point. On the opposite, the brain samples of both species and both investigated human tissues show a more diverse pattern in the amount of expressed distinct isoforms, but always with the highest number during old-age. The exact function and role of the different isoforms of this lncRNA during aging remains to be explored, but it serves as an interesting potential target of further research. Again, no homologous gene is currently annotated or known for *Nothobranchius furzeri*.

6.2.2 Changes in the expression of the main isoforms do not disrupt the encoded functional protein domains

Due to the fact that the number of different expressed isoforms and genes do not change significantly with age, we investigated if the mainly expressed isoforms of all genes are switched in an observable age-related fashion (see Fig. 6.3). For this analysis, only single isoforms with a higher expression than other expressed isoforms were taken into account [515].
We observed an aging-dependent switch in the dominantly transcribed isoform for $11-34\,\%$ of all expressed genes in all investigated species and tissues. Interestingly, with about $26\,\%$, the most of all expressed genes switched their main isoform at the aged time point after being unchanged during both mature time points, regardless of tissue and species. The same species- and tissue-independent relative amounts can be observed for genes that switch their main isoform directly after the young mature time point or at the latest old-age time point, with around $16-18\,\%$ of the relevant genes for each of these cases individually. Then we examined these genes with altered predominant isoforms during aging more closely, by determining the

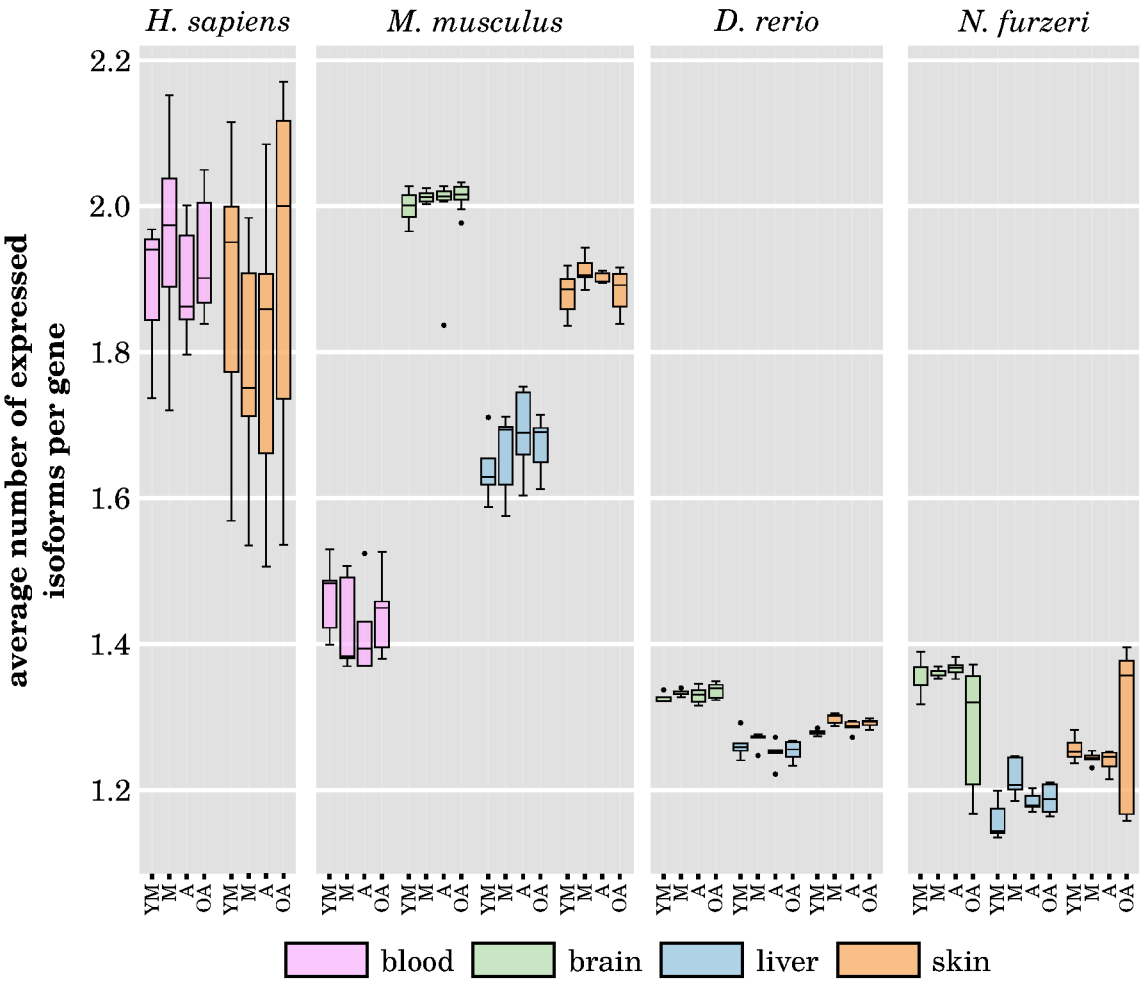


Figure 6.2: **Average number of expressed isoforms per gene.** The average number of expressed isoforms per gene in *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri* for every studied tissue and time point (YM – young mature, M – mature, A – aged, OA – old-age). No significant difference in the number of expressed isoforms per gene can be observed between different ages within each investigated species and tissue. In addition, the variance between the samples of each time point and tissue remains stable, except for the old-age brain and skin sample of *Nothobranchius furzeri*, showing a manifold increased variance. The lower amount of expressed isoforms in the two fish compared with both mammals might be biased due to the less complete isoform annotation. For more details, see SData 2.

amount of encoded functional protein domains in the different dominant isoforms. Functional protein domains (in short functional domains or protein domains) are specific units of amino acid sequences, which are structurally and functionally conserved and determine the precise function of proteins [516]. Surprisingly, in around 80 % of all genes showing an age-dependent isoform switch, the encoded number of protein domains remained unaltered, with the remaining genes showing equally fractions of isoforms with increased or decreased encoded functional domains after being switched (see Figure 6.3). This again holds true for all examined species and

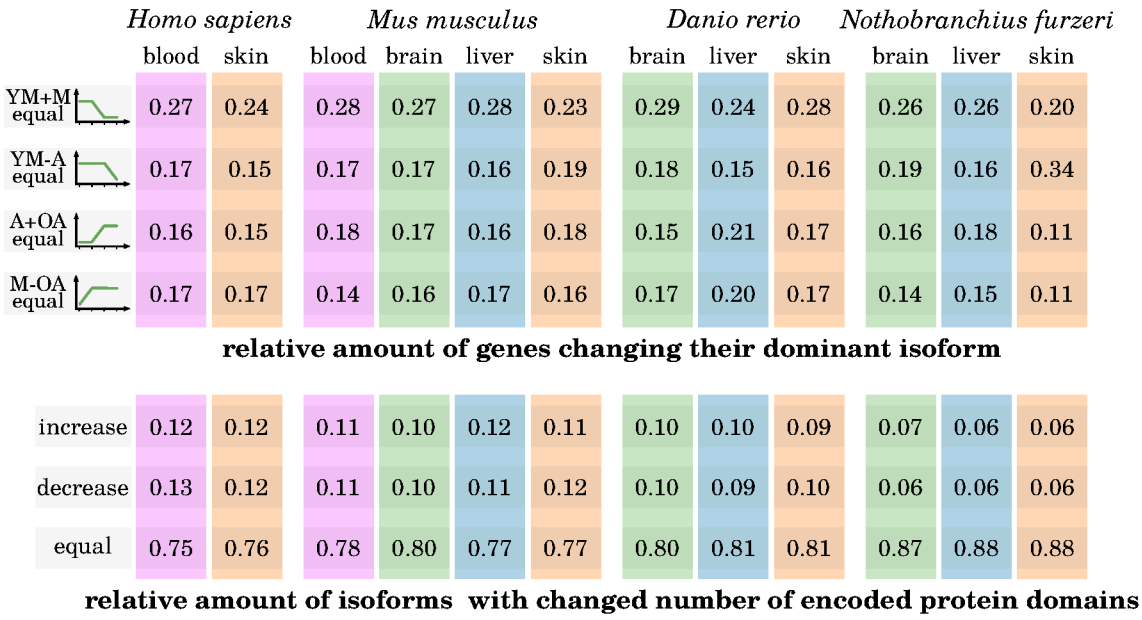


Figure 6.3: **Frequency of isoform switches during aging and impact on the amount of encoded functional domains** In the upper part of the figure, the relative amounts of genes for each species and tissue that switch their main isoform in one of four age-dependent manners are displayed: the dominant isoform switch occurs after the mature time point (YM+M equal), it remains the same until the aged time point and switches during the old-age (YM−A equal), it remains stable from the aged to the old-age time point (A+OA equal), or it switches directly after the the young mature time point (M−OA equal). Generally, age-dependent isoform switching occurs only for a minority genes and takes place more frequently in later time points. The lower part of the figure shows the relative amounts of isoforms, which encode less, more or still the same number of functional protein domains after being switched with age. Only genes with a changing dominant isoform are considered.

tissues. A change in the amount or composition of encoded protein domains within an mRNA has great influence on the topology, localization and function of the translated protein. However, because only a small fraction of switched mRNA isoforms is concerned by such a change in an age-dependent manner, most AS events seem to affect either sites within the mRNAs that encode unstructured protein regions or untranslated regions (UTRs). In fact, it was recently reported that changes in trailing 5' and 3' UTRs as well as intrinsic unstructured regions of mRNAs due to AS are over-represented compared with changes altering encoded protein domains [516]. Still, changes within the 5' or 3' UTRs can have significant effect on the translation of the respective mRNA, because the primary sequence and secondary structure of UTRs regulate the transcript's stability, turnover and efficiency of translation [517]. Yet, since most of the age-dependent isoform switches are subject to non-obvious functional changes, they might influence the effective translational availability rather than their abundance, the real impact of aging on the process of AS is hard to assess. Based on our observations, aging has some influence on the heterogeneity of

the transcriptome within the different tissues and species, affecting the primarily transcribed gene isoforms, but it appears that, for the majority of genes, this influence is more subtle.
Nevertheless, for individual genes, the change in the switched isoform can be more drastic with respect to their encoded function and the proportion of different expressed isoforms of a single gene could have significant impact on several aging-related processes [518]. For example, we found that the mainly expressed isoform of the *bcl2l1* gene switches at the aged time point of the mouse liver samples from encoding is short protein version (Bcl-xS) instead of its long version (Bcl-xL). Whereas Bcl-xL is an apoptosis inhibitor, the short version Bcl-xS is an apoptosis activator, displaying the AS controlled age-dependent transition from repressing to activating apoptotic processes in hepatic cells. Apoptosis is an important mechanism in aging liver tissue to maintain homeostasis and several dysfunctions, and diseases of the liver are related to programmed cell death processes [415, 416].
Also, amongst others, *srsf1* and *srsf6* change their predominantly expressed isoform over time in different murine tissues, and *srsf3* in human skin. These genes encode for SR proteins that are known to regulate certain age-related genes [519]. For example, SRSF1 and SRSF3 have direct impact on the expression of the important tumor protein p53 that regulates the cell cycle, senescence and apoptosis [518, 519]. In addition, the gene *gpr18* in human blood has been identified to change its isoform ratio [500], which is confirmed by our data. Although the isoforms of these genes are not differentially alternatively expressed, the switched expression of their mainly transcribed isoforms can influence the regulation of aging and senescence.

6.2.3 Less than 5 % of transcribed genes are differentially alternatively spliced during aging

To further assess the influence of aging on the splicing process, we investigated the expression change of differentially alternatively spliced genes (DSGs) between three different comparisons: early aging, late aging and longevity (see Figure 6.1). For every species and tissue, we observed quite varying amounts of DSGs and in contrast to previous findings [504], we could not confirm a general increase of DSGs with age (see Figure 6.4A). In their study, Rodriguez *et al.* compared the number of DSGs of murine tissues, but between different ages (4 months to 18 months and 4 months to 28 months) compared with our time points (9/15 months to 24 months and 9/15 months to 30 months), which could explain this discrepancy. In addition, they investigated different tissues (skeletal muscle, thymus, adipose tissue and bone, besides skin tissue) and analyzed microarray data instead of RNA-Seq, adding to the factors that make a direct comparison more difficult. The only similar observation they made was an increase of DSGs in their late aging comparison (4 months to 28 months) compared with their early aging comparison (4 months to 18 months). In general, within our data we observed only few DSGs in the mammalian samples (less than 2 % of the transcribed genes) compared with the fish samples (less than 5 % of the transcribed genes) and highest number of DSGs within the skin tissues, with exception to *Danio rerio*, having more age-dependent spliced genes in brain (311 DSGs) than skin (241 DSGs). This might indicate that the tissue-

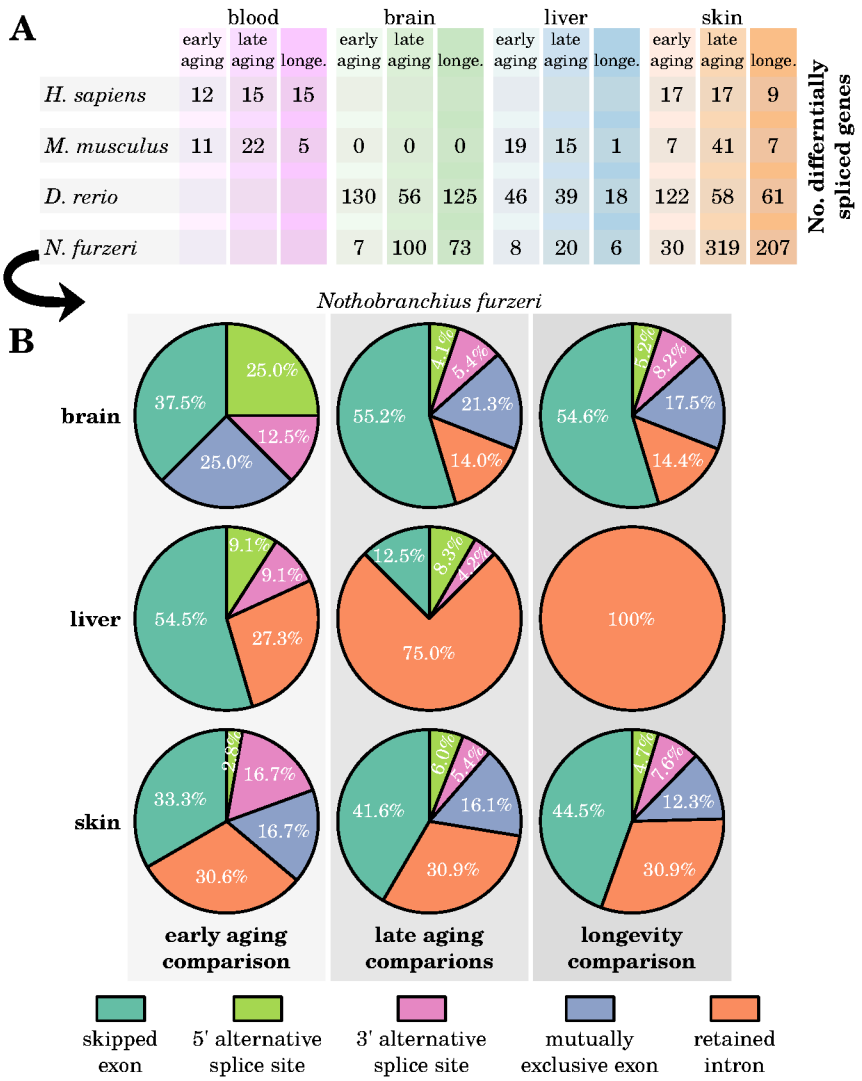


Figure 6.4: **Number of identified differentially spliced genes.** In the upper part of the figure, the amount of differentially spliced genes for the early aging, late aging and longevity (longe.) comparisons for every investigated species and tissue are given. Far less differentially spliced genes could be observed for *Homo sapiens* and *Mus musculus* compared with the fishes *Danio rerio* and *Nothobranchius furzeri*. Also, no general trend of increased gene splicing can be recognized in any of the species or tissues. On the arbitrary example of *Nothobranchius furzeri*, the lower part of the figures depicts the fractions of the exact AS modes (skipped exon, 5'/3' alternative splice site, mutually exclusive exon, retained intron) that contribute to the identified differentially spliced genes. Despite skipped exon being the most abundant AS mode in almost every age comparison, each tissues displays its own signature of splicing events during aging. More detailed information, including the AS modes of the other species can be found in SData 5.

specific regulation of AS processes prevails over any possible common aging effect. When analyzing differential expression of splicing within our age comparisons, we also investigated the occurring AS modes (exon skipping (ES), intron retention (IR), alternative 5' splice sites (A5S), alternative 3' splice sites (A3S), and mutually exclusive exons (MXE)) that contributed to the identified DSGs (see Figure 6.4B). We identified ES as the most prevalent pattern in all species, which is in agreement with previous findings [490]. There are only a few exceptions where one of the other AS modes prevails (see SData 5): In the human blood samples, IR is the main form of the alternatively spliced genes during early aging. Interestingly, A5S and A3S make up more than half of the AS modes in the longevity comparison in human. In the mouse blood samples, a decline from 86 % ES splice events in the early aging comparison to only 20 % in the longevity comparison can be observed, compensated by a rise of MXE splice events to 60 %. The same shift from mainly ES to MXE splice events can be observed in the murine skin samples. In contrast to any of the other AS modes, MXE is a relative complex and controlled splicing process and less likely to happen accidentally due to a deregulated splicing machinery [520], suggesting a stable working spliceosome even with high age. In *Danio rerio*, ES remains the main AS mode within all age comparisons and investigated tissues with a share of 54 % – 79 % of all occurring splice events. The fraction of AS modes does not change considerably in the brain and skin of *Nothobranchius furzeri*, with the notable exception of A5S and A3S, which seem to occur more often during early aging than late aging in the brain samples. Notably, the amount of IR increases during aging, especially in the liver samples of *Nothobranchius furzeri*, although this AS mode is known to be rare in vertebrates [490]. In general, the occurrence of the different AS modes appears to be tissue-specific with no observable distinct drift of specific splicing events during aging, let alone a common tendency between the individual tissues of one of the investigated species. In addition, between the four species, there are only few commonalities, besides ES being the predominant splicing event and the liver samples showing an unexpectedly high rate of DSGs derived from IR splice events. However, given an age-dependent decline in the correct functionality of the splicesome, one could expect a systematic decrease of more controlled splicing events, such as MXE.

Furthermore, we observed that AS tends to either take place in the middle or towards the 3' end of the expressed transcripts, regardless of the splice event, in all investigated tissues and species. In *Nothobranchius furzeri*, even the majority of AS events take place close to the 3' ends of the spliced transcripts (within the last quarter of the transcripts sequences), whereas for the other three species most AS events occurred in the middle of the spliced RNAs (within the second and third quarter of the respective sequences). This is not completely unexpected, because AS mostly changes protein characteristics only slightly [487, 490]. Changes in the primary sequences of transcripts near the 5' end due to AS, resulting in a frameshift, have a higher chance to alter the encoded protein function more significantly, which is commonly believed to be avoided. Again, if there is a general failure of the splicing machinery with increasing age, we would anticipate to see AS events to occur more uniformly distributed in transcribed sequences.

Age-related alternatively spliced genes are tissue- and species-specific

When comparing the observed DSGs between the tissues for each species individually, we found almost no overlap between the tissues (see Figure 6.5 and SData 6), confirming earlier findings of age-related tissue-specific AS regulation [490, 504]. Nevertheless, a few common genes could be identified being altered by AS during aging. For example, during early aging in both the liver and skin of *Mus musculus* the ubiquitously expressed gene *hnrnpa2b1* undergoes AS. It encodes for a member of the heterogeneous nuclear ribonucleoprotein family. Besides being involved in many different cellular functionalities, such as stabilization of telomeres, cell proliferation and splicing of other pre-mRNAs, it is associated with several diseases [125, 521]. *Hnrnpa2b1* has several known isoforms with most of them coding for the same protein, but one non-functional transcript (ENSMUST00000204090) is expressed highly in both tissues. Whether alterations in this genes isoform expression have any influence on further age-related AS remains an interesting open question. Furthermore, it appears to have at least some connection to aging because it was associated with longevity in a previous study [501].
But also many commonly shared DSGs in both fish species have functions in the splicing process or mRNA processing, like *u2af2b*, occurring to be alternatively spliced in all *Danio rerio* investigated tissues during early aging, encoding for an important subunit of the spliceosome [522].

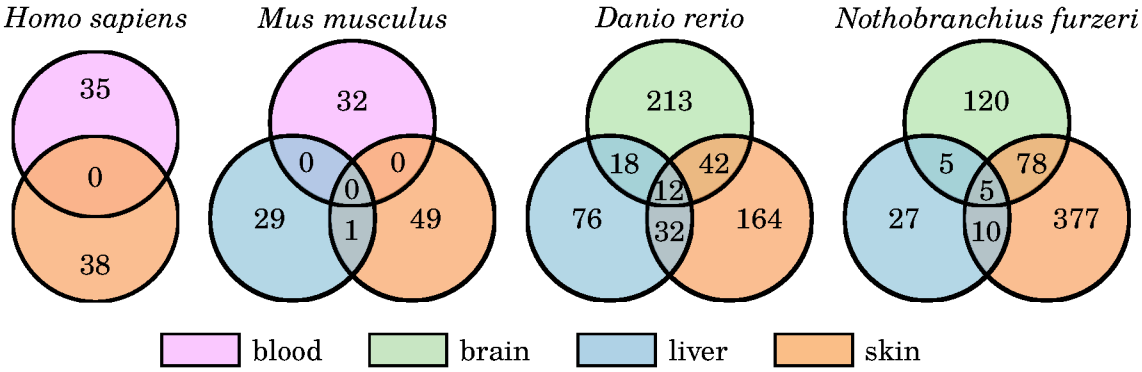


Figure 6.5: **Overlap of DSGs between the different tissues within each investigated species.** The comparison of the age-related DSGs demonstrates the tissue-specificity of the AS processes, showing only little overlap within each species. Note that the different age comparisons of each tissue are collapsed. For further results, see SData 6.

Even fewer overlap of DSGs can be found between the four different species (see SData 6). The only shared alternatively spliced gene in the aging human and mouse blood samples was *rpl13a*, encoding for a ribosomal protein, which was reported to play a special role in specific translational control of inflammation associated genes, next to its normal function as part of the ribosome [523]. This is of particular interest as inflammation processes tend to be deregulated with rising age in vertebrates and being implicated in many age-associated disorders [208, 409].
Most of the few shared DSGs in the skin samples of *Danio rerio* and *Nothobranchius furzeri* were observed to be cell structural genes. However, one more

interesting gene, *tp63*, was found to express different isoforms in both fish species during aging. It encodes for a protein named Tumor protein 63 (or transformation-related protein 63), which is a homolog of the intensively studied tumor suppressor gene p53 [524]. There are two distinct major isoforms, TAp63 and δNp63, with the first regulating apoptotic processes [525] and latter being involved in skin development and stem cell regulation [526]. Interestingly, we see a switch to the TAp63 isoform towards the old-aged individuals. Since it was recently shown that TAp63 prevents premature skin aging by supporting the maintenance of adult stem cells, age-related AS of this gene might have a conserved role in longevity [527].

6.2.4 Alternatively spliced genes heavily function in mRNA processing and transcription regulation during aging

To infer the potential role of the identified aging-dependent DSGs, we systematically examined their annotated functions for the individual tissues and species (see SData 7). Most interestingly, we found that genes directly responsible for splicing, post-transcriptional mRNA processing, transcriptional and translational regulation are differentially spliced in almost all aging comparisons (see Figure 6.6). Most of these DSGs include members of the DEAD box protein family, ribosomal proteins, eukaryotic initiation/elongation factors or splicing factors of the SR family. Similar observations were already made in different mouse tissues and human blood samples [500, 504, 519, 528]. In addition to our observations, it shows that age-related AS of various splicing and transcription components is, to some extend, a conserved aging feature. Since we did not observe a drastic change in the heterogeneity of expressed isoforms with age, we assume that the alterations of these splicing and transcription components do not lead to a generally disturbed spliceosomal or mRNA processing activity. Transcription and splicing are two closely connected processes, wherein several of the splicing factors directly interact with the transcription machinery [529]. Potentially, AS induced age-related changes in some components of the spliceosome might reflect adaptation mechanism to an altered transcription process. However, to understand the consequences of these changes, further research is required on the exact effects for the encoded proteins of the respective DSGs.
In the human blood and skin samples, we additionally observed many age-related DSGs to be involved in the process of nonsense-mediated mRNA decay. Nonsense-mediated mRNA decay (NMD), which is the degradation of mRNA transcripts due to premature termination signals [530], was itself observed to be regulated by AS through aging in the human blood and skin. Whereas the main function of NMD is to discard aberrant mRNAs that encode potentially deleterious proteins, it is also used as a transcriptional regulatory level, controlling stability and availability of mRNAs by AS [531]. Possible implications of an altered NMD mechanism through age-related AS remain again an interesting target of further studies.
Besides these common and partially conserved findings, we observed also many DSGs to function in tissue-specific aging-associated processes. AS of genes involved in inflammatory response were identified mainly in the human and mouse blood and skin samples, most likely being associated with the age-related increase of inflammatory processes in these tissues [208, 409, 532], because AS is extensively used to increase

the isoform diversity of genes of the immune system [533].
Not unexpected, the liver samples of mouse and both fishes display DSGs of various metabolic processes, hinting at possible alterations in digestion or detoxification with age. In the brain samples of *Danio rerio* and *Nothobranchius furzeri*, many DSGs are associated with neuronal receptor regulation and brain development processes. It is a known feature of teleosts, such as both investigated fishes, to maintain neurogenesis even in older ages [271, 534].

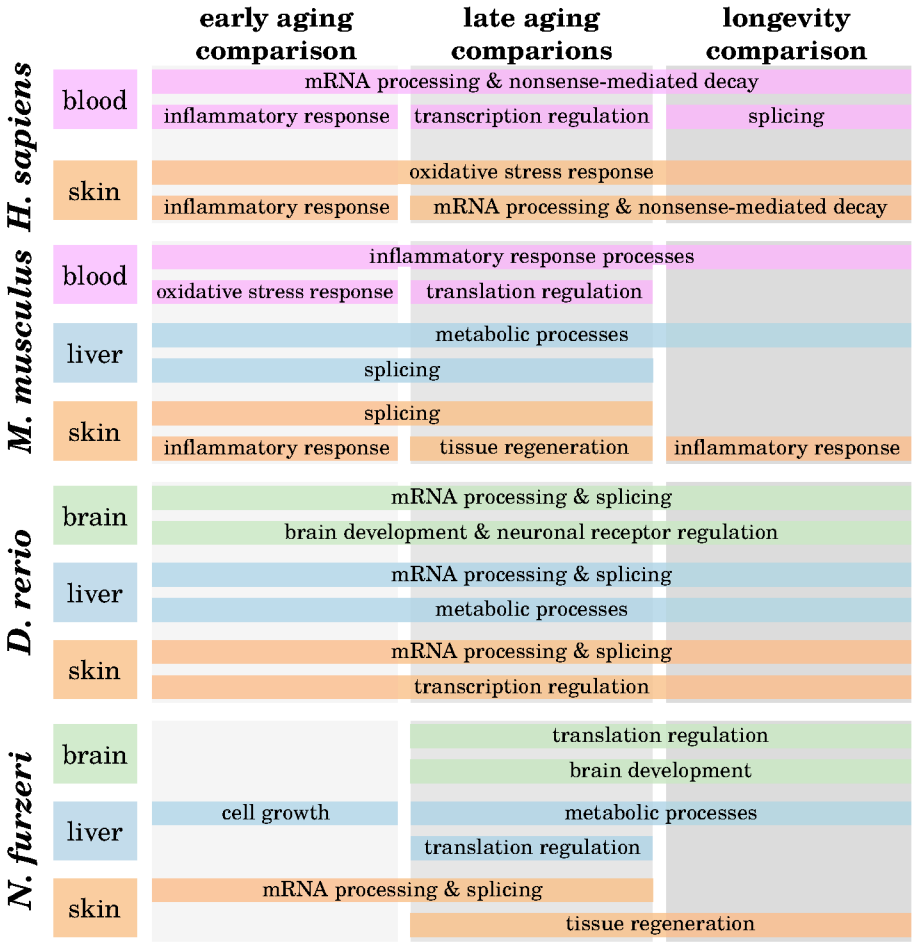


Figure 6.6: **The prevailing biological functions associated with the alternatively spliced genes in the examined age comparisons.** Interestingly, various DSGs involved in splicing, post-transcriptional mRNA processing as well as transcriptional and translation regulation were observed in almost all aging comparisons, regardless of the investigated tissue and species. Amongst others, tissue-specific genes functioning in aging-associated processes were found to be subject to AS, such as tissue regeneration, inflammation or oxidative stress response.

6.2.5 Age-related changes in spliceosomal activity correlates with the number of differentially spliced genes

The spliceosome is a ribonucleoprotein complex, consisting of a high number of components to ensure the correct process of splicing with a high accuracy including several control mechanisms and different modes [487, 492]. As a further step to understand how AS could be affected by aging, we analyzed the expression of the main protein-coding and non-coding RNA genes of both the major and minor spliceosome as well as transcriptional changes of associated splicing factors (see Figure 6.7 and SData 8). The deregulation of the spliceosome or at least some of its components can have severe consequences for the organism and was focus of recent aging-related microarray-based studies [501, 502, 519].
For human and mouse, we observed a stable expression of the main spliceosomal genes during aging with only a few significant changes. In human blood, only the U2 small nuclear RNA is significantly downregulated towards the old-age time point but other RNAs, such as U1 and U4, show the same tendency. In skin, the gene *snrnp70*, encoding for an important subunit of the small nuclear ribonucleoprotein U1, is down-regulated in the late age comparison. Interestingly, the same gene shows an about two-fold upregulation in the mouse liver within the early aging comparison, whereas the small nuclear RNA U6 becomes downregulated in the long-lived individuals. All of the mentioned genes are crucial for the correct assembly and function of the major spliceosome, but their change during aging appears to be little in both mammals. Still, several additional splicing factors show a strong upregulation with age in the human skin samples, such as *nova1*, which is also highly upregulated during early aging in the mouse blood samples. The equally named encoded protein of *nova1* was first thought to be a brain-specific splicing factor [535], but was later discovered to play a key role in the splicing of pancreatic beta cells, too [536]. Here, we observe that it also appears to be active in two more tissues, human skin and murine blood, in an aging-related manner.
However, both investigated fishes present a different picture. Whereas *Nothobranchius furzeri* shows as few significant changes as both mammals, important genes of the major and minor spliceosome are rather up- instead of downregulated during aging. In its skin, we observed an upregulation of the gene *snrnp27*, which is part of the major spliceosomal U4/U6 subunit, and the two genes *snrnp25* and *snrnp48*, which are part of the minor spliceosomal-specific U11/U12 subunit. Additionally, the two spliceosomal associated genes, coding for the ribonucleoproteins SNU13 and SNRPD1 are also more than two-fold increased in their expression. All of these genes show a significant higher activity towards the old-age time point and may indicate a need for an enduring splicing activity. Also, in the brain of *Nothobranchius furzeri*, a significant activation of spliceosomal related genes, like *snrnp40* (ribonucleoprotein U5) or *snrnp48* (ribonucleoprotein U11/U12) was observed with age. However, the transcriptional changes were not as strong compared with the skin.
In contrast to the other examined species, *Danio rerio* displays extensive changes in the expression of its spliceosomal genes and associated splicing factors during aging. It displays a distinctive down-regulation of most of the ribonucleoprotein coding

genes with age. This can especially be seen in the liver and skin samples during the early and late aging comparison. Most interestingly, the longevity comparison between the aged and the old-age animals revealed either no change in expression or a slight upregulation of some ribonucleoproteins, indicating no or only minor further modulations of the spliceosome after a certain age is reached. The brain shows in principal the same pattern as the other tissues, but similar to *Nothobranchius furzeri* to a much weaker extend. If this strong downregulation can be correlated to the comparatively high amount of significantly spliced genes in *Danio rerio*, remains an interesting open question.
In general, we do not identify a significant increase or decrease of expression of the catalytic units of the minor spliceosome (U4atac, U6atac, U11, U12) during aging.

6.3 Alternative splicing as an adaptation to maintain homeostasis in old age

Within this chapter, we have analyzed the impact of aging on AS in four evolutionarily distinct species and up to four different tissues. For each species, we investigated two mature (one young and one middle-aged), an aged and an old-age time point. When comparing these ages with each other, we observed a varying amount of significantly differntially spliced genes (DSGs) for the four species and tissues. Whereas the fishes *Danio rerio* and *Nothobranchius furzeri* showed in total several hundred DSGs, much less were identified in *Homo sapiens* and *Mus musculus*. However, a general increase of AS with age could not be observed, which is in contrast to earlier reports [504]. Most likely, these differing results arise from the different experimental setups. The study of Rodriguez *et al.* [504] was based on microarray data of only murine tissues (skeletal muscle, thymus, adipose tissue, bone and skin) and completely different time points (4 instead of 9 months and 18 instead of 24 months for the mature and aged time points respectively) compared with our data. They observed an increase of AS in the bone, muscle and skin tissues with age, which we could confirm at least for the latter tissue, but to a much lower extend. Taking other tissues and species into consideration, their observation of an age-related accumulation of DSGs cannot be generalized.
Interestingly, even though we identified age-related DSGs to be almost exclusively tissue- and species-specific, many of them function in various post-translation mRNA processing steps, transcriptional and translational control and splicing processes themselves, independent of the investigated species and tissues. This observation indicates AS and other RNA processing mechanisms to be indeed modified with normal physiological aging, and that this feature appears to be conserved to some extend. Other studies observed similar results in human and mice [500, 504, 519, 528]. As a consequence, the specificity of the spliceosome and thus transcription of specific gene isoforms seem to change during aging. Besides mRNA regulatory processes, age-related DSGs are mainly involved in tissue-specific aging-associated processes, such as inflammation in the mammalian blood and skin samples [208, 409, 532], metabolic process in liver and tissue regeneration processes in the fishes brain and skin samples [271, 534]. However, because more aging-related genes are expressed

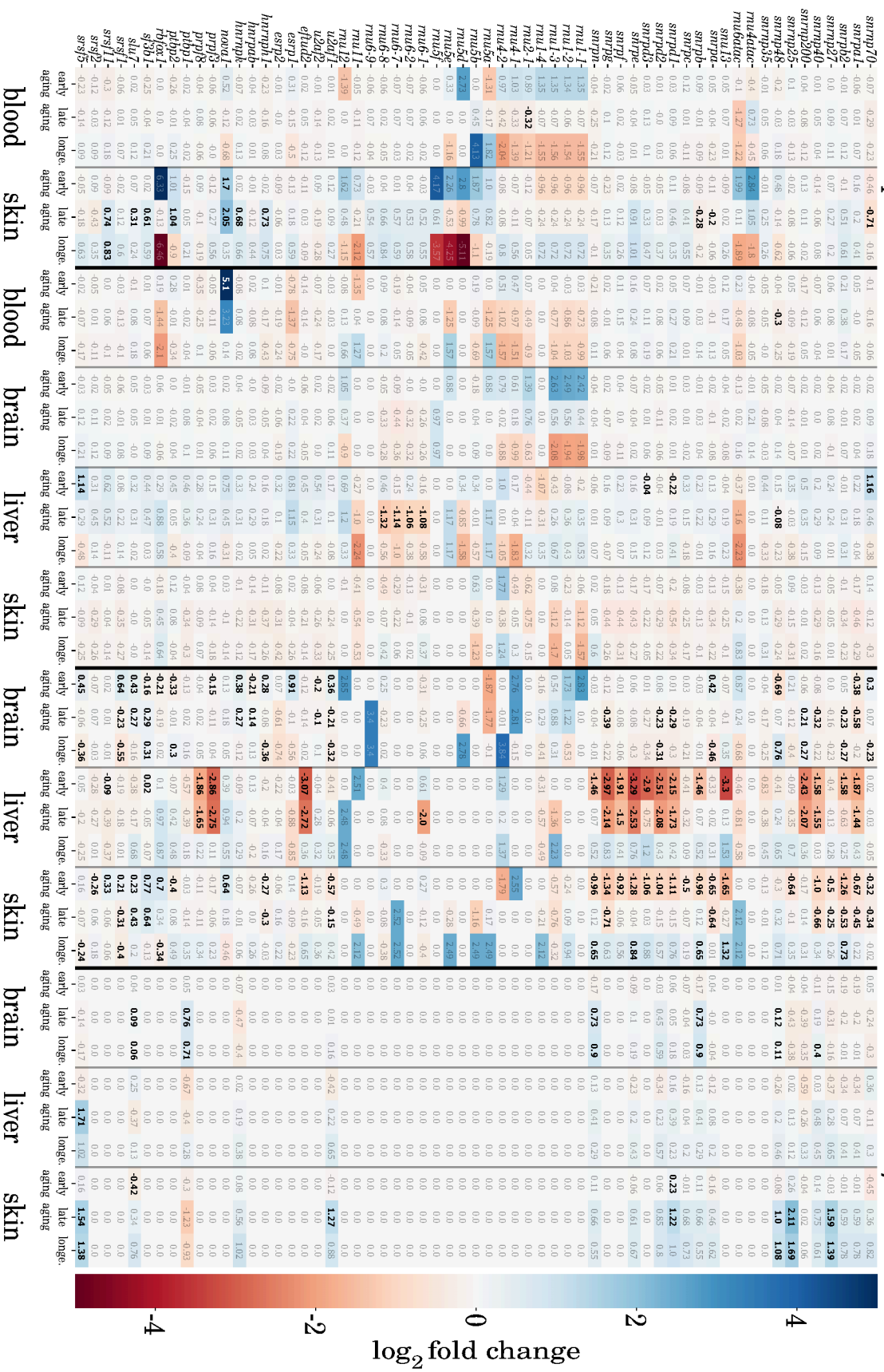


Figure 6.7: **Transcriptional changes of the major and minor spliceosomal genes as well as splicing factor genes during aging.** Numbers indicate log_2 fold changes between two compared ages, were a positive value indicates an upregulation and a negative value a downregulation of the respective gene with aging. Bold numbers mark statistically significant changes.

with increasing age, it is not totally unexpected to observe these genes to be processed by AS.
Other recent microarray-based studies focused on the expression changes of spliceosomal genes and splicing factors, reporting a decline in the expression of several components of the splicing machinery [501, 502, 519], possibly resulting in an increased heterogeneity of isoforms and transcriptional noise [537]. When examining the expression activity of major and minor spliceosomal genes and several splicing factors, we observed only a weak downregulation of these genes during aging in any of the investigated human and mouse samples and even a slight upregulation in *Nothobranchius furzeri*. In contrast, *Danio rerio* displays massive transcriptional modulation of spliceosomal genes and associated splicing factors during aging.
One of the most important findings of our study is that these age-related changes in the splicing process, either by self-regulation through significant changes in the expressed isoforms of various spliceosome components or the up-/downregulation of spliceosomal genes are not reflected in the general landscape of present RNA transcripts. Neither does the number of expressed isoforms per gene change significantly over time, nor are the proteins encoded by genes that switch their predominant isoform during aging affected. Nevertheless, different isoforms expressed by the same gene can still encode exactly the same protein and changes in their untranslated regions can have major impact on their stability and translation efficiency [516]. Further studies based on single cell RNA-sequencing could give deeper insights in the diversity of the transcriptome during aging and its implication in biological processes [538].
In conclusion, while the overall landscape of the aging transcriptome appears to be unaffected by aging-related changes in AS or the spliceosomal activity, single alternatively spliced genes can have crucial influences on certain aging processes but AS shows not to be deregulated in general with increasing age.

Chapter 7

The circadian rhythm might change universally with aging

This chapter is based on the publication "Interspecies and interorgan comparisons reveal aging-associated changes in clock gene expression" [1].

Data used and analyses performed in this chapter

Age comparison setup

According to the time scales displayed in Figure 7.1, we selected three different time points for each species – one mature point (M), one aged point (A) and one old-aged point (OA) – and compared them pairwise. For the individual species these were: 24 – 29, 60 – 65, and 75 – 79 years (*Homo sapiens*); 9, 24 and 30 months (*Mus musculus*); 12, 36 and 42 months (*Danio rerio*); 12, 27 and 39 weeks (*Nothobranchius furzeri*). Details of all differential gene expression results, together with the raw and normalized count values are given in detail in SData 1.

7.1 How aging affects the rhythmicity of genes

We know that aging affects host physiology in different dimensions and disturbances can occur at the cellular and organ levels. However, how they are translated into whole-body aging is still relatively unknown. One important characteristic of higher eukaryotes (and even some fungi and cyanobacteria) is, that they are strongly influenced by daytime, in respect to almost all their functions (in animals, this is best reflected by their sleep-wake cycle). The underlaying mechanisms, which control the expression of almost all genes and therefore controlling most biological processes are called circadian rhythms (CRs) [539]. They enforce a 24 hour periodicity and are self-sustained, but mostly adjust themselves to local environment *zeitgeber*, which are external factors such as light or temperature [540]. Just in 2017, the Nobel Prize in Physiology or Medicine was awarded for the discovery of the molecular basis of CRs on the level of single cells in the fruit fly *Drosophila melanogaster* [336].

[1] The complete supplemental material is available at `http://www.rna.uni-jena.de/supplements/circadian_rhythm_aging/`

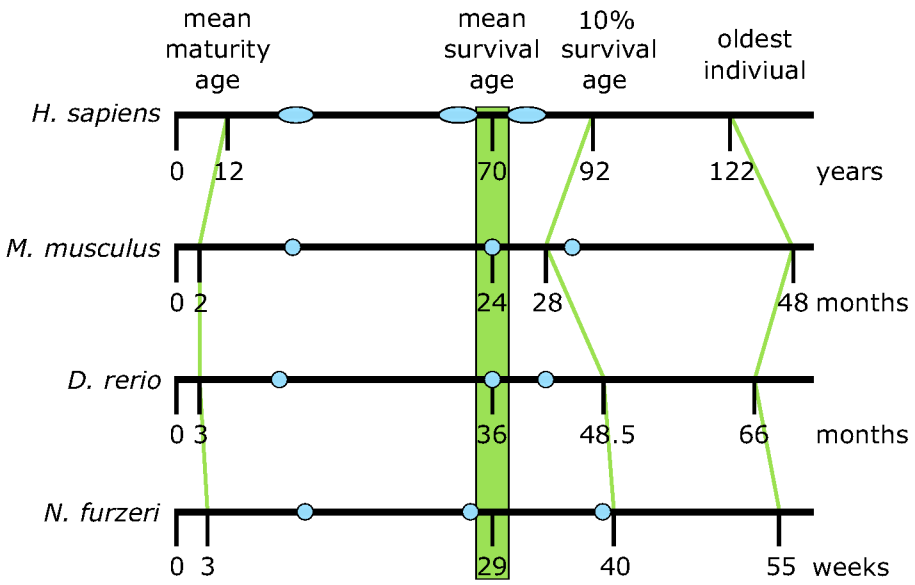


Figure 7.1: **Interspecies normalization of age categories.** This picture is similar to Figure 2.3 but only showing the selected time points for the ages used in the study of this chapter (blue dots). The total individual's life time, represented by the length of the life time axis, was subdivided by index stages. These stages were the biological stages corresponding to maturation, the mean survival age, the 10 % survival rate and the greatest ultimate age reported for an individual belonging to the respective species. The time intervals between the resulting intersections were normalized with respect to the mean survival age in a linear fashion.

In mammals, the suprachiasmatic nucleus (SCN) of the anterior hypothalamus is a superordinate pacemaker that imposes CRs on gene expression in each tissue and cell type and, thus, conveys rhythmicity to all essential processes of physiology and behavior. In parallel with whole-body circadian rhythmicity imprinted by the SCN, circadian oscillations are autonomously regulated at the organ and cellular levels, with a functional clock residing in peripheral tissues and cell types. In comparison with mammals, including humans, CRs in lower vertebrates, such as fish, are less hierarchically organized and do not depend on the SCN [541]. Accordingly, the entrainment of CRs via both environmental and endogenous cues is stronger in fish, where organs such as the skin and liver directly respond to factors such as light exposure and hormonal and nutritive conditions. The functionality of the CRs declines with aging, generally manifesting in lower penetrance, characterized by reduced amplitudes and increased scatter in circadian acrophases, disturbances of cell and tissue synchronization and phase shifts in oscillations during daytime [542, 543]. Likewise, fragmentation and shifts in the periodicity of CRs are prominent in elder mice and humans [543, 544], where they influence rest-activity and sleep-wake period patterns. In addition to causing sleep disorders, disturbances in CRs are associated with pronounced stress responses, impaired DNA repair and cancer and are considered to be an independent risk factor for age-related disorders such as type II diabetes mellitus, Alzheimer's dementia, coronary heart disease and tumors [545–

548]. Furthermore, misalignment of CR has been linked to the loss of proteostasis that is characteristic of many age-related proteinopathies, which manifest in the nervous system in particular [549]. Accordingly, genes governing CR signatures are involved in a plethora of biological processes, at both the single-cell and network levels, influencing cell growth, metabolic homeostasis, DNA repair, autophagy, epigenetic modifications, ER stress and immune functions, among others, all of which play a central role in the initiation and progression of cellular senescence and tissue aging [13]. According to these pleiotropic influences, meta-analyses have recently revealed that at least 40 % of the total protein-encoding genome in mammals is subject to circadian oscillations; among these genes, those associated with age-related disorders are most strongly connected to CR drivers [545]. However, it has not been well studied whether age-dependent variations in CR orchestration are based on different expression levels of core regulatory genes and downstream targets or, rather, are due to a qualitative shift in the set of genes involved. Additionally, it is unresolved how far these changes in CR gene regulation patterns overlap in different tissues and species. An interconnected question is whether a conserved genetic fingerprint of age-related CR modification that is common to different species and various tissues might be perceivable. Moreover, in addition to strong control by the core set of clock genes and transcription factors and their pulsed expression in the brain hypothalamus [550], self-sustained functionality of the CR is well established for the skin and liver [551]. Thus, the verification of such a concept demands the use of species that oppose hierarchical CR regulation to conserved decentralization of the CR through evolution, such as fish. To address these issues, we performed comprehensive transcriptional analyses in different species and subselected them for CR regulatory patterns. Differentially expressed genes (DEGs) were identified as a function of elevated to extremely old ages in comparison to a young maturation grade. The results were compared among *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*, the last of which is a short-lived organism with high dependence on good adaption to ecological niches. The examined tissues included the brain, blood, skin and liver from one or both sexes underlying a cross-sectional study design. Through this comprehensive approach, we performed a longterm interspecies and intertissue comparison of CR regulatory genes for the first time, and we report putative intersections of CR-associated gene alterations with aging-related chronodisruption, pathobiologies and the aging process itself. At the organ level, our results particularly highlight the importance of the skin in CR government in both, mammals and fish and support the notion that decentralized CR regulation might not be unique to fish but may instead be underestimated in higher vertebrates.

7.1.1 Clock-related gene selection

When we initially analyzed the JenAge data in an unbiased, hypothesis-independent mode we identified clock-related genes to be among the most prominent differentially expressed genes (DEGs). Because this observation made us curious, we filtered DEGs bases on a predefined list of clock-related candidate genes (as derived from [545, 552]), supplemented with a manual collection of established clock-related target genes. Thereby, we found significant regulations ranging from a log_2 fold

change of -1.84 up to 5.59 across species, tissues and ages (see Figure 7.3).

7.1.2 Intraspecies homogeneity of tissue samplings

Interage and interspecies comparisons of preselected transcriptome datasets resulted in the identification of 25 DEGs from the total age-matched DEGs that were associated with CR control and showed significant age-dependent regulation in at least one out of the four species in at least one age category and at least one tissue. The intraspecies homogeneity of the RNA-Seq measurements for each tissue and time point analyzed were plotted according to t-SNE, as shown in Figure 7.2. In general, each tissue of the four investigated species formed distinct individual clusters, as illustrated for human skin and blood samples (green symbols) and murine tissue specimens (blue symbols). The only exceptions were the scattered pattern observed for liver samples from both fish species, revealing value interference, and the *Nothobranchius furzeri* brain and skin samples, which overlapped with each other in two separate subclusters (gray symbols). However, because organ and tissue isolation from both fish strains was performed at defined times during the day, this value dispersion appeared to be independent of the sampling strategy. Due to restrictions in sample spectrum acquisition from human individuals, brain and liver were compared only between mice and fish. As a consequence of the limited blood volume in both fish species, nucleated blood cells were analyzed solely in mammals. Skin, as the organ with the highest sensitivity to environmental and exogenous cues influencing the circadian system, was evaluated in all four species. Based on the observed level of intragroup consistency and the intergroup separation of the datasets with respect to the parameters 'species' and 'tissue/organ type', we embarked upon analyses of an age-related signature of the transcriptome output. Since the different age categories applied were not strong enough in their function of segregating RNA-Seq values into independent clusters by t-SNE, the expression profiles were analyzed at the single-transcript level in this case.

7.1.3 Interspecies comparison of age-related CR effects

CR has been intensely studied in mice, but less so in other species. Moreover, a direct interspecies comparison of evolutionarily distant species has not yet been performed. Here, we used four phylogenetically separate species and compared their CR-related transcriptomes as a function of aging. As illustrated by the heat maps in Figure 7.3, the 25 DEGs that we identified comprised the highly conserved core clock regulators *per1 – per3*, *cry2*, *arntl* (*bmal1*), *clock*, *nr1d2* (*rev-erb-β*) and *bhlhe41* (*dec2*) and interacting genes such as *ulk1*, *ntrk2*, *ptgs1* and *ace*. The *arntl* (*bmal1*) and *clock* genes form the positive branch of the circadian feedback loop, acting as transcription factors that control the expression of a plethora of CR-related genes, including per and cry family members. Per1 – 3 and Cry form heterodimers to suppress the Clock/Arntl (Bmal1) complex and, thus, represent the complementary negative arm of the CR loop. Both, *per2* and *bhlhe41* (*dec2*) were found to be regulated in an age-dependent manner in all four species. Moreover, the transcript levels of *per1*, *cry2*, *arntl* (*bmal1*) and *clock* changed in three of the species (mice and the two

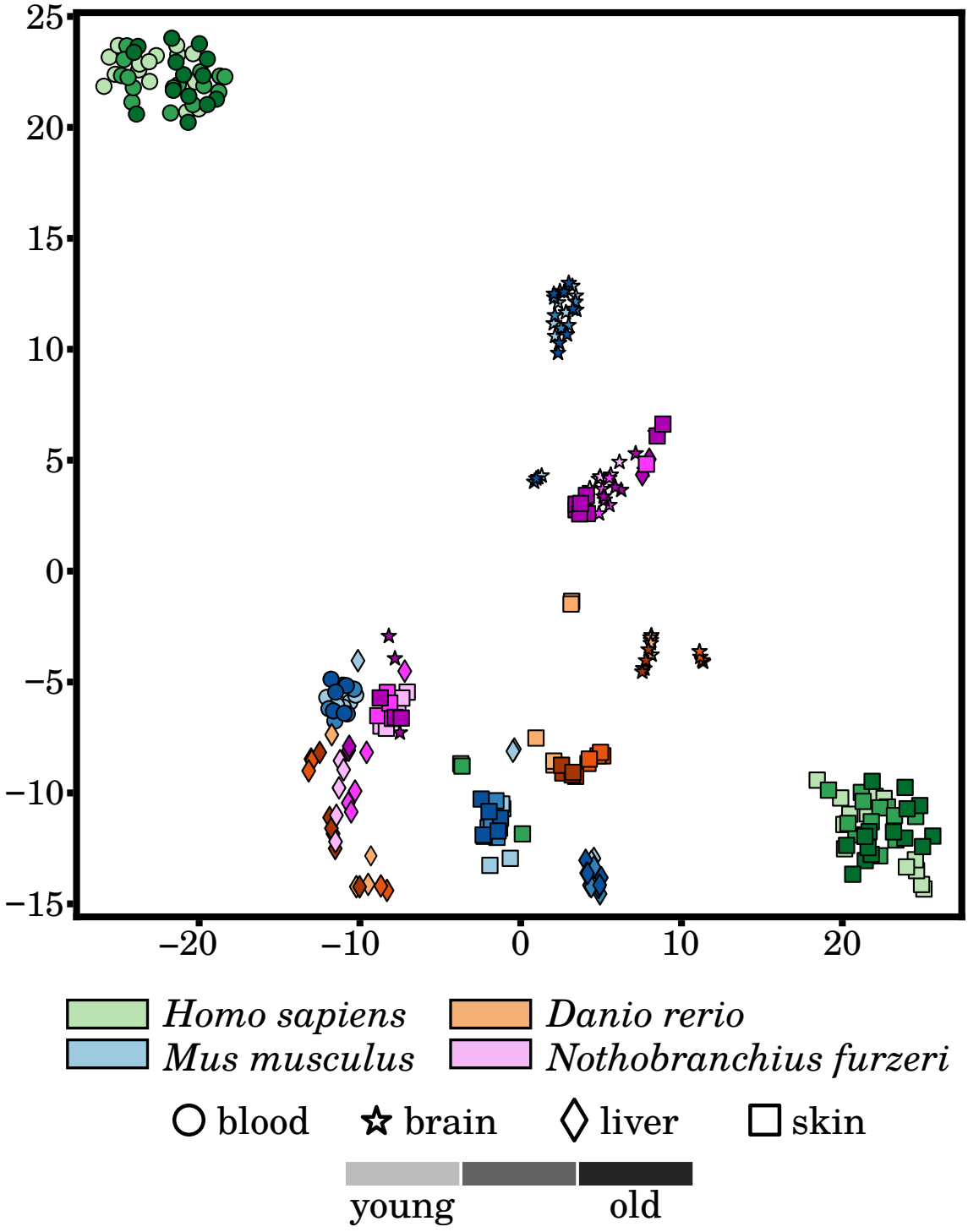


Figure 7.2: **CR-associated gene based t-SNE of the investigate samples.** A t-distributed stochastic neighbor embedding [402] (t-SNE) performed for the parameters 'species', 'tissue' and 'age category' of the RNA-Seq datasets. All RNA-Seq samples displaying the expression of selected CR-associated genes were clustered utilizing the t-SNE approach. In general, all tissues of the four investigated species formed distinctive clusters. As exceptions, the liver samples from both fish species overlapped each other, as well as brain and skin samples of *Nothobranchius furzeri*. However, the different age categories did not segregate in independent clusters. For more details, see SData 2.

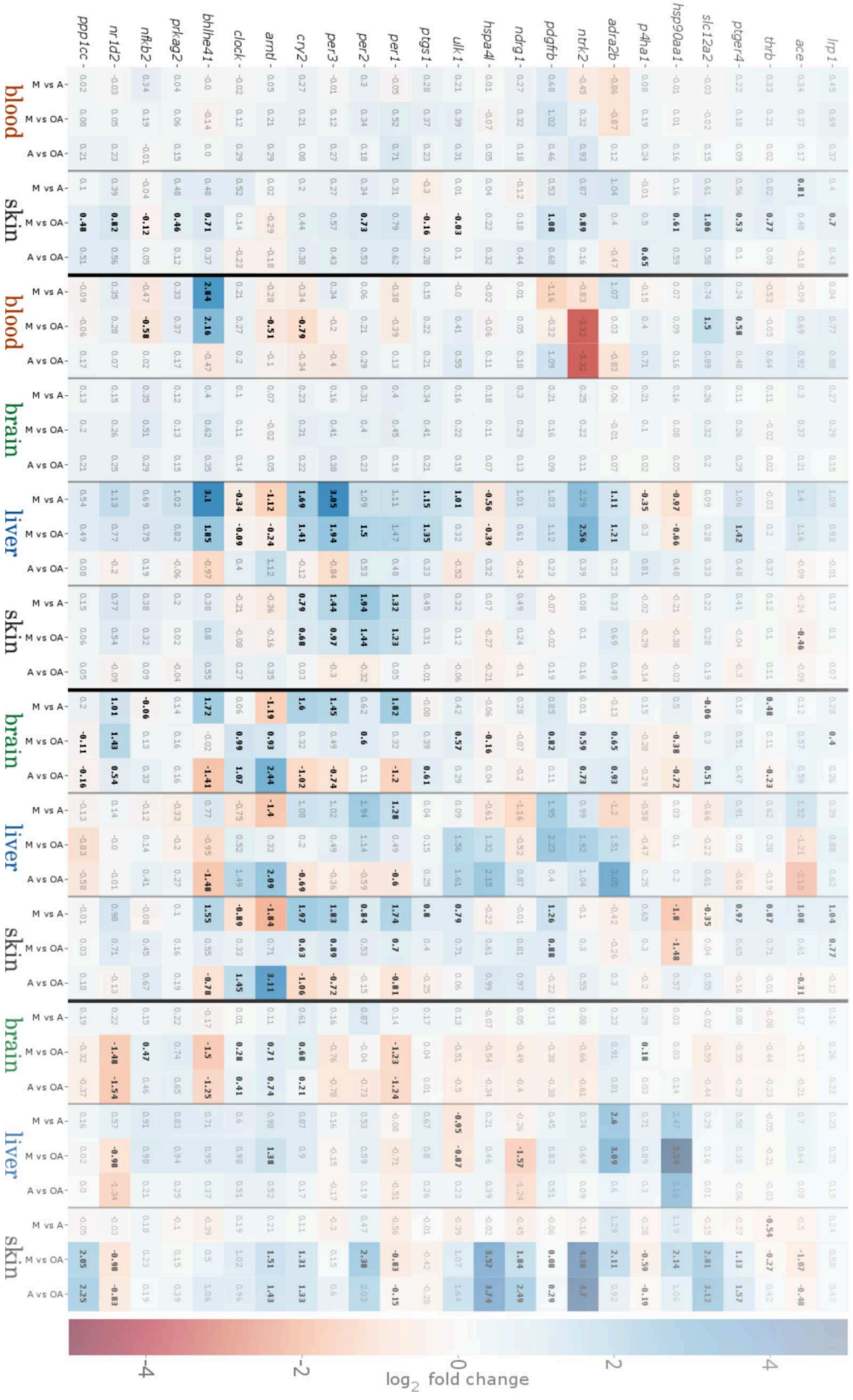


Figure 7.3: **Heatmap representation of CR-related DEGs for the different species, tissues and age categories.** Genes are represented as mouse orthologues (a complete list of gene orthologues can be found in SData 4). Numbers indicate log_2 fold changes between two compared ages, where a positive value indicates an upregulation (blue), and a negative value downregulation (red) of the respective gene with aging. All significant changes in gene expression are indicated in bold.

ray-finned fish). Most of these transcriptional alterations were already detectable in the aging category, A, and persisted up to the highly aged class, OA. The expression of *per2* increased in all of the examined species and organs with aging, whereas the up/downregulation patterns of the other CR-related clock genes were less homogeneous. Although the biological importance of *per2* upregulation is unclear, it may be involved in immune responses and tumor suppression, in addition to its role in CR government [549, 551, 553]. Similarly, *per1* exerts crucial effects on the cell cycle machinery in addition to functioning in CR control. Whereas *per1* overexpression propagates DNA damage response (DDR)-induced apoptosis, its downregulation can enhance tumor growth [554]. It also directly interacts with the DDR checkpoint regulators Atm and Chk2. Since patients with a cancer diagnosis display reduced expression of *per1*, it has been suggested as a target for tumor treatment. Moreover, the transcript levels of the blood pressure-controlling Ace enzyme, an established downstream target of clock regulators, were found to be strongly altered in all four species. Blood pressure deregulation is a frequent symptom in the elderly and a well-established risk factor for many age-related diseases. Considering gene orthologies, these data suggest that the genetic basis of clock gene regulation during aging shares many commonalities across the examined species. This conclusion is in accordance with another recent work comparing CR gene expression between mice and fish [549]. Further evidence of the involvement of CR-associated alterations in the aging process is derived from a mammalian system [553, 555]. In 2015, Chen and colleagues conducted a study that identified DEGs in the prefrontal cortex of elderly, mainly male Caucasians (≤ 60 years) relative to young persons (≤ 40 years). The time of death of each person was normalized to a zeitgeber time, where the time point of sample acquisition allowed conclusions about changes in circadian rhythmicity. The authors identified age-related alterations of rhythmicity in more than 1,000 genes, including core clock components such as *per1* and *per2*. Thus, loss of circadian rhythmicity was identified not only for certain genes, but for a subset of genes that exhibited rhythmic expression exclusively in aged individuals. This set of genes might act as a novel clock compensation mechanism for the disruption of chronobiology during aging [556]. Notably, studies in an elderly Taiwanese population revealed the presence of single nucleotide polymorphisms in several core clock components, including *clock*, which were associated with cognitive aging independent of environmental factors [557]. In mice, numerous studies have analyzed the effects of disruption of different core clock components belonging to both the positive and the negative branches of the circadian feedback loop (*e.g.*, *clock* and *bmal1* (*arntl*) as well as *per2*). Interestingly, independent of endogenous regulatory function (*i.e.*, suppressive or stimulatory), the disruption of core clock components led to lifespan reduction, though to different extents [553]. Likewise, in addition to a shortened lifespan, Bmal1-deficient mice exhibit strong and even premature aging phenotypes, such as age-dependent body and organ weight loss, sarcopenia, cataracts and atrophy of subcutaneous fat [558]. Remarkably, *bmal1* (*arntl*) is the only gene whose disruption is followed by total loss of behavioral rhythmicity [553, 559]. Although mice that are deficient in Clock exhibit a reduced lifespan and develop an aging phenotype, the extent is less pronounced [555]. A more recent study revealed timely restrictions on the effects of Bmal1 deficiency on

aging phenotypes [560]. Yang and colleagues showed that conditional Bmal1 knockout during adulthood ameliorated some age-related phenotypes observed in studies on a conventional murine Bmal1 knockout model, as described above. Moreover, the expression of oscillatory genes was reduced in the conditional knockout model (at least when analyzed in the liver), but overall gene expression remained largely unaltered. Hence, the authors suggested that some of the aging phenotypes that were previously related to disruption of CRs might instead arise from the loss of clock-independent Bmal1 (Arntl) functions [560]. Thus, it appears that the role of distinct clock-regulating factors in actual circadian rhythmicity is not decisive for their effects on aging. Accordingly, we present findings suggesting that clock-related gene regulation occurs during aging, irrespective of the circadian rhythmicity of the genes. We further revealed a lack of a strict activating or suppressing regulation pattern, which has also been observed by others [553]. Taken together, the available findings indicate that the effect of clock regulators on the CRs alone is not sufficient to explain their impact on aging but, rather, suggest more complex interplay of their multidimensional molecular functions and contributions to multiple pathways, better reflecting a complex process such as aging.

7.1.4 How organ-specific CR regulation changes with age

In contrast to the use of a mixture of *clock*-related mRNAs from different tissues and organs, as performed for zebrafish larvae [549], we realized an interorgan comparison. Moreover, in contrast to the study by Yan et al. (2008), who surveyed 14 tissues exclusively among mammalian species (*i.e.*, human, monkey and mouse) from a tissue gene expression atlas [561], we extended the examination of interorgan relationships to four evolutionarily separated species and specified regulation in the context of aging. The organ-specific pattern of CR-associated gene regulation differed between the species, as depicted in the heat map containing the 25 DEGs introduced above (Figure 7.3). Likewise, in human samples, the strongest age-associated impact on CR genes occurred in the skin, whereas blood was devoid of any variation. In mice, the most prominent alterations were present in the liver and, to some extent, the skin, whereas only slight changes were found in blood and none were found in the brain. This pattern was again different from that detected in fish, where the greatest regulation occurred in the skin and brain. The changes in the liver were also less pronounced in fish than in mice. Between the two fish species, *Danio rerio* displayed a higher number of changes, whereas *Nothobranchius furzeri* evidenced stronger gene regulation with aging. Our observations are in line with the hypothesis that a common set of CR genes, including the core regulators *per1/per2/per3*, *cry2*, *bmal1* (*arntl*), *clock* and *bhlhe41* (*dec2*), is intrinsically active in several tissues. In a recent comprehensive study compiling micro-array data from several previous studies, the authors extracted 41 common CR genes from 9,995 genes that oscillated in at least 8 out of 14 tissues in mice [561]. In conclusion, although the transcriptome pattern differed across the species, we identified several CR-related genes that were differentially expressed in at least one tissue within the same species. Although there is apparently a common set of clock genes in many tissues and cell types, as described for the above-mentioned genes herein and complemented by *cry1*, *npas2*,

bhlhe40 (*dec1*), *nr1d1* (*rev-erb-α*), *ror-α* and *rorc* in other studies [545, 561], these genes often differ substantially in their circadian oscillation phases between different tissues [545, 561, 562]. Thus, according to our analysis, it might be more likely that common CR gene expression patterns can be detected within the same tissue across species than that similar age-dependent regulation can be assumed within different organs of a single species.

Brain

The impact of brain-specific CR regulation on the aging process is still not well studied [556, 557]. Recent data from Yan and colleagues illustrated that non-SCN brain tissue compares to other non-SCN tissues in terms of CR phase oscillation [561]. Another interesting finding is that deletion of central clock regulators such as Bmal1 (Arntl) and Clock, in combination with Npas2, in the brain causes severe age-related astrogliosis, whereas deletion of Bmal1 (Arntl) alone results in structural and molecular changes [548]. Here, we assessed brain-specific regulation in mice and fish. Since either the entire brain or a complete hemisphere was assayed in fish and mice, respectively, our data do not reflect age-dependent changes specifically in the hypothalamus. While no regulation was found across the selected age categories in mice, core clock genes were strongly regulated in a heterogeneous pattern in both fish species, with log_2 fold changes ranging from -1.54 to 2.44 when comparing A to OA categories (see Figure 7.3). Likewise, *pdgfrb* and *ulk1* were dominantly changed in *Danio rerio*, but less so in *Nothobranchius furzeri*, among which *ulk1* displays a well-known function in autophagy and longevity pathway regulation [563, 564]. Thus, the CR regulatory input from brain neuronal populations (possibly in addition to the pineal complex and eye photoreceptors) might be much stronger in fish than in mammals or may at least be subjected to stronger age-related perturbances in fish. Consistent with these findings, high resilience of the mammalian hypothalamus to age-related alterations in clock-related transcripts was illustrated in a recent study performed in rhesus macaques. When diurnal expression of SCN regulatory genes was examined in young versus aged rhesus monkeys, no significant changes were observed based on microarray or qRT-PCR techniques. The authors concluded that altered behavior in aged rhesus macaques might originate from CR misalignment in other regulatory, but subordinate, body systems [565]. Moreover, the data indicated that, in spite of decentralized CR regulation, core clock genes in fish are conserved similarly to those in mammals. In support of this observation, several fish orthologues of murine core clock-interacting transcription factors were recently identified as targets of the core clock member *arntl* (*bmal1*) [549] that operates as a key regulator of CR-related gene transcription in mice [566], and was found in the present study to be regulated in mice and both fish.

Blood

Although blood is often utilized for CR studies underlying a longitudinal study design, our cross-sectional approach points to weak representation of age-related CR effects in this compartment compared to other tissues. Likewise, no age-associated alterations were evident in human blood. In mice, only a few genes were differentially

expressed displaying log_2 fold changes between -0.79 and 2.84, with the strongest effect being found for *bhlhe41* (*dec2*). The Bhlhe41 (Dec2) transcription factor is part of the negative feedback branch of the core clock loop that suppresses *clock*/*arntl* (*bmal1*). Bhlhe41 is involved in short sleep manifestation and impaired sleep homeostasis [567] and further participates in pathways relevant to tumor growth and progression [568], which are known to be affected by the aging process [569].

Liver

In the liver, most frequent age-related differences in the expression of CR genes were discovered in mice, whereas alterations in fish were most prominent with log_2 fold changes displaying values between -1.57 and 5.59. Similar to the results in blood, *bhlhe41* (*dec2*) showed the strongest regulation in the liver, followed by *per3*. Deregulation of *per3* has been related to tumor progression and a worse cancer prognosis, *e.g.*, in breast cancer [570]. The pattern of DEG manifestation was similar in aging categories A and OA.

Skin

Skin exhibited significant CR-related gene alterations across all four species showing log_2 fold changes between -1.84 and 4.7. The greatest expression changes were observed in short-lived *Nothobranchius furzeri*, where significant changes were manifest in 17 of the 25 target CR-related genes. A similar pattern was seen in *Danio rerio* featuring significant changes in 16 genes. In human samples, the changes were less prominent but were similarly frequent in number. In mice, the upregulated genes consisted almost exclusively of core clock regulators of the negative branch. Compared to human and fish, the number of DEGs in the mouse skin was lowest and shared commonalities with human only for two genes, *i.e. per2* and *ace*. In contrast, a larger intersection set of genes was found between human and both fish species (Figure 7.4). Interestingly, the expression of *cry2*, the binding partner of the Per group, was increased in all of the species. This pattern reached significance in three species and showed a tendency in human specimens. Cry2 has been described as playing a role in psychiatric disorders, cancer and type II diabetes mellitus [545–548]. Thus, as mentioned above in the context of the brain, the skin might represent one of the CR oscillators subordinate to the SCN that are responsible for the age-related decay in the ability to entrain, synchronize or reset circadian phases. Notably, there was a higher similarity in genes being differentially expressed between human and fish than human and mice (Figure 7.4), though sharing 70 % compared with 85 % with the human genome, respectively.

7.1.5 Regulation changes of CR-associated DEGs

As recently suggested based on mouse transcriptome data, body temperature may represent the key driver that synchronizes the oscillation of HSPs throughout the body [561]. Here, we observed that *hsp90aa1* and *hspa4l* were differentially, but heterogeneously expressed during aging across the four species. Hsp90aa1 is a direct target of Hsf1, a key transcription factor in heat response mediation that is

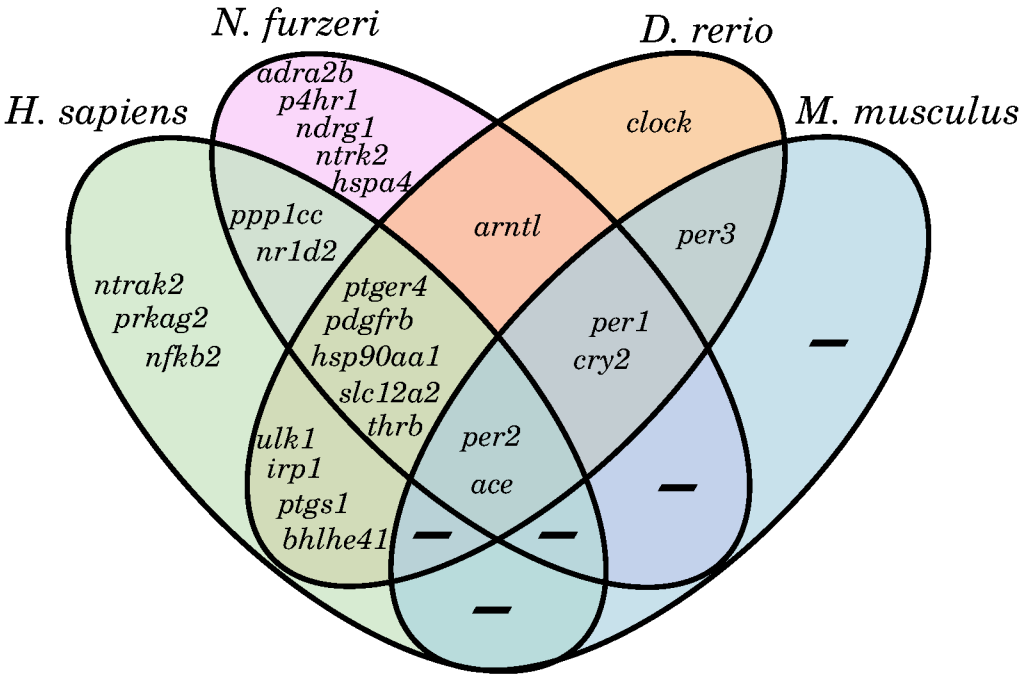


Figure 7.4: **Venn diagram showing the overlap of CR-related DEGs in the skin from the four species investigated.** Across all species, the most prominent regulation of CR-related genes was found in the skin. There was a greater similarity in genes being differentially expressed between humans and fish than humans and mice.

involved in the indirect transmission of heat stimulation to *per1* and *per2*. Similar to other chaperons, Hsp90aa1 has been suggested to impact the aging process by taking part in attenuated primordial stress response. Such a weakened heat shock response is assumed to contribute to loss of proteostasis through reduced clearance of damaged proteins and deregulation of apoptosis. In contrast, sustained mild heat shock responses are part of the hormetic concept propagating anti-aging and pro-longevity effects [571]. Apart, Hspa4l is required for normal spermatogenesis in young mice [572] and found highly expressed in leukemia cells.

7.2 Single circadian rhythm factors are evolutionary conserved in the chronobiology of aging

Multiple efforts have been aimed at mapping physiological and pathological processes to the core clock in single [546, 573] or multiple organs [545] of a single species, or in selected organs underlying a pathophysiological stressor [552]. However, the role of highly conserved clock regulation in physiological aging over time and across multiple species has not yet been addressed. Here, we used whole-genome RNA-Seq to profile the transcriptomes from four different species in a cross-sectional study, in individuals ranging from young mature to old age categories. Within these species, we compared transcriptome elements associated with CR in different or-

gans, including blood, brain, liver and skin. In summary, our analyses illustrate that alterations in CR-related gene expression during aging are a conserved characteristic that is retraceable across evolutionarily different species, including humans. Within the aging period, such alterations were found to be an early characteristic, at least in *Mus musculus* and *Danio rerio*, that persisted into old age classes. Our data also point to the importance of the skin in CR regulation, where we detected the most striking regulations across all four species (Figure 7.4), seemingly due to the light exposure of the skin. These observations suggest that even in organisms with strongly hierarchical CR orchestration by hypothalamic structures, as observed in mice and humans, peripheral subordinate organs such as the skin contribute to alterations in clock gene expression during aging. Additionally, the transcriptional profile governing the decentralized clock system in fish via peripheral tissues and cell types appears to be much closer to mammalian clock gene regulation than expected. Whether CR deregulation is causal or a consequence of aging remains to be explored. This work complements recent studies and provides a comprehensive novel dataset that links CR factors with physiological aging across evolutionarily highly distinct species. It will help to dissect characteristics of chronobiology in healthy aging from those associated with diseases and latent precursor states and contribute to the identification of interventions that improve the well-being of the elderly and prolong a healthy lifespan.

Chapter 8

Conclusion

Preventing aging is the always demanding struggle to keep the balance of a highly complex thermodynamical process that we call *life*. Almost every cell has to face different kinds of stress over time and in order to survive must try to adapt to these new conditions. Such stresses can arise from external or internal factors and different sophisticated molecular strategies have evolved during the history of life to handle cellular stress. Nevertheless, these response mechanisms are not perfect, sometimes affecting each other and cannot always restore the biological integrity of a cell completely. With time more and more 'errors' (*e.g.*, in form of DNA damage, malignant proteins or imbalanced metabolic pathways) accumulate, that further disturb the cell's homeostasis, leading to an impaired response to other stresses. In the end, when too much errors have accumulated the cell either dies or enters a senescent state, loosing the ability to proliferate and replicate. During senescence the cell is still able to sustain itself but has lost almost all its other functions, *i.e.*, if further damaging stress occurs it still will die eventually. At this point, the aging process of an individual cell would end. However, in a multicellular organism with specialized tissues aging manifests in more complex ways. Different cell types have to deal with different kind of stresses and can influence each other strongly in various ways. Thus, due to the complexity of life, aging is a very heterogeneous process and has to be tackled from different angles to understand and maybe prevent it.
Within this thesis we have investigated the topic of aging from many different perspectives to obtain new insights in the underlying genetic programs and to examine if some of these programs are evolutionary conserved. To do so, we have analyzed a huge collection of transcriptome sequencing data from four evolutionarily distinct species (*Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*). From those four species different tissues were sampled (blood, brain, liver and skin) at four different but comparable time points (see Figure 2.3): one young but already mature time point, one middle-aged mature time point, one old time point and finally one long-lived time point. The whole transcriptome dataset was generated in the frame of the *JenAge* project by next generation RNA sequencing experiments during the time of 2009 and 2013.
Within Chapter 1 of this thesis we have introduced the topic of aging comprehensively, discussing what is currently known and hypothesized about how and why we age. We also briefly introduced the idea behind next generation RNA sequencing

and how the resulting data are processed and analyzed, because most of the results presented in this thesis are based on RNA sequencing analysis.
As a starting point, we have examined in Chapter 3 how different tissues from one species are affected by aging. We compared the expressional changes of protein-coding genes in four different mouse organs during aging and found that each of the investigated tissue ages differently due to tissue-specific intrinsic and extrinsic factors. However, besides displaying individual aging patterns in gene expression and stress response, all four investigated tissues showed also some common age-dependent features. For example, we found many genes acting in the mitochondrial electron transport chain to be modulate in their expression significantly with age, which can have direct consequences to cell survival and organismal lifespan [204]. Additionally, we could identify seven genes to be commonly differentially expressed in the four investigated murine organs: *gm8979*, *igkv4-62*, *s100a6*, *lcn2*, *rap2a*, *s100a9* and *vmp1*. These genes are involved in the immune system, cell cycle control, autophagy and calcium signaling, which are all known aging-associated processes. Moreover, we could show in the nematode *Caenorhabditis elegans* that a knockdown of its *lcn2* homolog leads to a protection against early lethality. However, in the end we concluded from our results that any single gene can hardly be regarded as a marker of aging for a whole organism, because aging manifests differently in every tissue.

In Chapter 4, we explored the topic of microRNAs (miRNAs) and their potential role in aging, especially in the short-lived killifish *Nothobranchius furzeri*. We described how these short RNA transcripts are involved in the regulation of almost all cellular functions (Section 4.1) and they are identified and annotated on the example of the newly assembled genome of *Nothobranchius furzeri* (Section 4.2). Besides other classes of non-coding RNAs (ncRNAs), we could annotate over 750 miRNA genes of which about 420 seem to be killifish-specific. We have further described evolutionarily conserved miRNA clusters, such as the miR-430 or the miR-17/92 cluster, and compared them with ones from other fishes. From our findings and expression validation analyses we could assume that our *Nothobranchius furzeri* miRNA catalog is comparable to the one of the model organism *Danio rerio*.
Then, we have investigated the role of miRNAs in the diapause regulation of *Nothobranchius furzeri* and its the connection to aging processes (Section 4.3). Diapause is a kind of dormancy state, which some killifish embryos can undergo if the environmental conditions are not fit for survival. The embryos can remain in diapause for several months or even years, showing no sign of aging until the environmental conditions improve again and they leave their developmental arrest to hatch and develop normally. By comparing the expression of miRNA genes between different annual and non-annual killifish species, we were able to identify several conserved miRNAs (*e.g.*, members of the miR-430 family, miR-29a, miR-200) and two killifish-specific miRNAs (miR-19337 and miR-19344) that are probably in the control of regulating diapause in these fish embryos. In addition, we could also show that some of the diapause-associated miRNAs are also modulated in an aging-dependent manner in *Nothobranchius furzeri*, and that they function in processes related to aging. Thus, we could show that the miRNA regulatory networks involved in diapause and aging

overlap to some extend.
In Section 4.4, we studied the stochasticity of small RNA-Seq data in greater detail, proposing and introducing a new statistical model based on gamma distributions to estimate differential expression of miRNAs more accurately. We could show that expression data of miRNAs are more likely to follow a gamma distribution instead of the typically assumed negative binomial distribution. We have implemented our new statistical model in a tool named `MeRDE` and showed in different comparisons that it can compete with the current-state-of-the-art tools `DESeq2` and `edgeR`. Furthermore, in Section 4.5 we applied `MeRDE` on the JenAge small RNA-Seq dataset and could identify several potentially conserved miRNA regulators of aging-related processes in *Homo sapiens*, *Mus musculus*, *Danio rerio* and *Nothobranchius furzeri*. Among those, the miRNA miR-192 was identified as a possible key regulator of aging in all four species, being heavily involved in the regulatory control of cellular senescence, apoptosis, inflammation and other stress responses as well as the formation and progress of several aging-associated diseases, such as cancer, diabetes and neurodegenerative disease. Since miR-192 is not only expressed in all investigated tissues, but also occurs as a circulating miRNA in the bloodstream of different animals, it could serve as an easy accessible diagnostic marker for aging and even as a potential therapeutic target.

Two of the most suspected drivers of aging are cellular senescence and chronic tissue inflammation (the latter phenomenon also known as *inflammaging*). We examined the progression of both processes during aging in the sampled organs of the four investigated species and identified tissue-specific expression patterns of associated genes (see Chapter 5). For example, the liver and brain appear to be more affected by chronic inflammation and apoptosis, whereas the skin shows strong activation of senescence pathways with age. In addition, we could show that the response to oxidative stress modulates both senescence and inflammation progression. We could identify conserved expression signatures across tissues and species as well as potential tissue-specific markers of aging and concluded from our observation that age-related stress response is executed similarly in vertebrates. Most interestingly, we found the variance in gene expression to be more controlled in long-lived individuals compared with average-lived ones, *i.e.*, suggesting that individuals that reached an exceptionally old age benefited from a more stably executed stress response.

In Chapter 6, we have investigated to which extend the process of alternative splicing (which is the generation of differing mRNA isoforms from single genes) is influenced by aging. Despite previous and contrary observations, we found that the general landscape of the alternatively spliced transcriptomes changes only slightly with age in all examined species and sampled tissues. For example, about 80 % of all genes that showed an age-dependent switch of their mainly expressed isoform, the encoded proteins remained unaltered in respect to its functional domains. Nevertheless, we could observed single (mis)spliced isoforms that are reported to be involved in aging-associated processes. But the general process of alternative splicing remained commonly stable during normal physiological aging.

In the last chapter of this thesis, we dealt with age-related changes of circadian rhythm factors and possible resulting consequences. Almost all biological functions in animals are strongly influenced by daytime (*e.g.*, reflected by the sleep-wake cycle) and the underlying mechanism, which control the expression of most genes, are called circadian rhythms. The functionality of circadian rhythms declines with age, leading to a fragmentation and shifts of their periodicity. As a results, cellular stress responses are impaired and strong disturbances of circadian rhythms are considered to be risk factors of age-associated disorders such as cancer, Alzheimer's dementia and heart diseases.
We observed that age-related changes in circadian rhythm regulation are organ-specific and only few circadian rhythm factors are evolutionarily conserved in the chronobiology of aging. Interestingly, we found that the skin contributes strongly to alterations in clock gene expression during aging in all four investigated species. Additionally, the transcriptional profile governing the decentralized clock system in fish via peripheral tissues and cell types appears to be much closer to mammalian clock gene regulation than expected.

Everything taken together, we have observed that aging manifests in a wide variety of forms, not only in the distinct species but also in the different organs within a single organism. Every tissue has to face different stresses over time due to its specialized function or localization and, thus, has to respond to them in their own specific way. Some of these responses appear to be conserved to some extend, indicating that they play a more dominant role in the formation of aging. Nevertheless, despite all the known aging-related processes and known genetic aging factors, the "starting point" of aging and its progression are still hard to predict. This is most likely because aging resembles a stochastic processes of error accumulation over time, which is also suggest by many of our results. But in spite of its complexity, research on aging has improved our understanding of how different internal and external factors influence cellular processes with time. And doubtlessly it will further help to prevent age-associated diseases and frailties and will enable an extended healthy life-span of humans. With the results presented in this thesis, we have made our contribution to achieve this goal.

Chapter 9

Bibliography

[1] F. S. Collins and V. A. McKusick. Implications of the Human Genome Project for medical science. *JAMA*, 285(5):540–544, 2001.

[2] Francis S. Collins, Michael Morgan, and Aristides Patrinos. The Human Genome Project: lessons from large-scale biology. *Science (New York, N.Y.)*, 300(5617):286–290, 2003.

[3] Matt Kaeberlein, R. Wilson Powers, Kristan K. Steffen, Eric A. Westman, Di Hu, Nick Dang, Emily O. Kerr, Kathryn T. Kirkland, Stanley Fields, and Brian K. Kennedy. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science (New York, N.Y.)*, 310(5751):1193–1196, 2005.

[4] Pankaj Kapahi, Brian M. Zid, Tony Harper, Daniel Koslover, Viveca Sapin, and Seymour Benzer. Regulation of lifespan in Drosophila by modulation of genes in the TOR signaling pathway. *Current biology : CB*, 14(10):885–890, 2004.

[5] Cynthia Kenyon. The plasticity of aging: insights from long-lived mutants. *Cell*, 120(4):449–460, 2005.

[6] Ricki J. Colman, Rozalyn M. Anderson, Sterling C. Johnson, Erik K. Kastman, Kristopher J. Kosmatka, T. Mark Beasley, David B. Allison, Christina Cruzen, Heather A. Simmons, Joseph W. Kemnitz, and Richard Weindruch. Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science (New York, N.Y.)*, 325(5937):201–204, 2009.

[7] Amie J. Dirks and Christiaan Leeuwenburgh. Caloric restriction in humans: potential pitfalls and health concerns. *Mechanisms of ageing and development*, 127(1):1–7, 2006.

[8] L. Robert. Calorie-restriction and longevity. End of a dream, at least for primates? *European Geriatric Medicine*, 4(2):129–132, 2013.

[9] ENCODE Project Consortium et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, 2007.

[10] R. C. Hardison. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends in genetics : TIG*, 16(9):369–372, 2000.

[11] Manolis Kellis, Barbara Wold, Michael P. Snyder, Bradley E. Bernstein, Anshul Kundaje, Georgi K. Marinov, Lucas D. Ward, Ewan Birney, Gregory E. Crawford, Job Dekker, Ian Dunham, Laura L. Elnitski, Peggy J. Farnham, Elise A. Feingold, Mark Gerstein, Morgan C. Giddings, David M. Gilbert, Thomas R. Gingeras, Eric D. Green, Roderic Guigo, Tim Hubbard, Jim Kent, Jason D. Lieb, Richard M. Myers, Michael J. Pazin, Bing Ren, John A. Stamatoyannopoulos, Zhiping Weng, Kevin P. White, and Ross C. Hardison. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17):6131–6138, 2014.

[12] Thomas Derrien, Roderic Guigó, and Rory Johnson. The Long Non-Coding RNAs: A New (P)layer in the Dark Matter. *Frontiers in genetics*, 2:107, 2011.

[13] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–1217, 2013.

[14] T. B. Kirkwood and S. N. Austad. Why do we age? *Nature*, 408(6809):233–238, 2000.

[15] Thomas B. L. Kirkwood. Understanding the odd science of aging. *Cell*, 120(4):437–447, 2005.

[16] L. Robert. Longevity and aging, genetic and post-genetic mechanisms. Which target to choose for postponing and treating age-related diseases. *European Geriatric Medicine*, 3(1):61–66, 2012.

[17] J. Labat-Robert and L. Robert. Longevity and aging. Mechanisms and perspectives. *Pathologie-biologie*, 63(6):272–276, 2015.

[18] Clive M. McCay and Mary F. Crowell. Prolonging the life span. *The Scientific Monthly*, 39(5):405–414, 1934.

[19] Roger B. McDonald and Jon J. Ramsey. Honoring Clive McCay and 75 years of calorie restriction research. *The Journal of nutrition*, 140(7):1205–1210, 2010.

[20] Marcia C. Haigis and Bruce A. Yankner. The aging stress response. *Molecular cell*, 40(2):333–344, 2010.

[21] G. J. Lithgow, T. M. White, S. Melov, and T. E. Johnson. Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proceedings of the National Academy of Sciences of the United States of America*, 92(16):7540–7544, 1995.

[22] Elissa S. Epel and Gordon J. Lithgow. Stress biology and aging mechanisms: toward understanding the deep connection between adaptation to stress and longevity. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 69 Suppl 1:S10–6, 2014.

[23] D. Harman. Aging: a theory based on free radical and radiation chemistry. *Journal of gerontology*, 11(3):298–300, 1956.

[24] Kerstin Nowotny, Tobias Jung, Tilman Grune, and Annika Höhn. Accumulation of modified proteins and aggregate formation in aging. *Experimental gerontology*, 57:122–131, 2014.

[25] Jessica A. Sorrentino, Hanna K. Sanoff, and Norman E. Sharpless. Defining the toxicology of aging. *Trends in molecular medicine*, 20(7):375–384, 2014.

[26] E. G. McGeer and P. L. McGeer. Brain inflammation in Alzheimer disease and the therapeutic implications. *Current pharmaceutical design*, 5(10):821–836, 1999.

[27] C. Franceschi, M. Bonafè, S. Valensin, F. Olivieri, M. de Luca, E. Ottaviani, and G. de Benedictis. Inflamm-aging. An evolutionary perspective on immunosenescence. *Annals of the New York Academy of Sciences*, 908:244–254, 2000.

[28] Sahn-ho Kim Sh, Patrick Kaminker, and Judith Campisi. Telomeres, aging and cancer: in search of a happy ending. *Oncogene*, 21(4):503–511, 2002.

[29] Judith Campisi. Aging, cellular senescence, and cancer. *Annual review of physiology*, 75:685–705, 2013.

[30] Dominick G. A. Burton and Valery Krizhanovsky. Physiological and pathological consequences of cellular senescence. *Cellular and molecular life sciences : CMLS*, 71(22):4373–4386, 2014.

[31] Paul V. Targonski, Robert M. Jacobson, and Gregory A. Poland. Immunosenescence: role and measurement in influenza vaccine response among the elderly. *Vaccine*, 25(16):3066–3069, 2007.

[32] Gregory A. Poland, Inna G. Ovsyannikova, Richard B. Kennedy, Nathaniel D. Lambert, and James L. Kirkland. A systems biology approach to the effect of aging, immunosenescence and vaccine response. *Current opinion in immunology*, 29:62–68, 2014.

[33] X. Liu, C. N. Kim, J. Yang, R. Jemmerson, and X. Wang. Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. *Cell*, 86(1):147–157, 1996.

[34] Ingrid Böhm and Hans Schild. Apoptosis: the complex scenario for a silent cell death. *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging*, 5(1):2–14, 2003.

[35] R. William Vandivier, Peter M. Henson, and Ivor S. Douglas. Burying the dead: the impact of failed apoptotic cell removal (efferocytosis) on chronic inflammatory lung disease. *Chest*, 129(6):1673–1682, 2006.

[36] Evan T. Powers, Richard I. Morimoto, Andrew Dillin, Jeffery W. Kelly, and William E. Balch. Biological and chemical approaches to diseases of proteostasis deficiency. *Annual review of biochemistry*, 78:959–991, 2009.

[37] B. L. Strehler. Understanding aging. *Methods in molecular medicine*, 38:1–19, 2000.

[38] D. B. Friedman and T. E. Johnson. Three mutants that extend both mean and maximum life span of the nematode, Caenorhabditis elegans, define the age-1 gene. *Journal of gerontology*, 43(4):B102–9, 1988.

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

[40] A. Bartke. Mutations prolong life in flies; implications for aging in mammals. *Trends in endocrinology and metabolism: TEM*, 12(6):233–234, 2001.

[41] Shennen A. Mao, Jaime M. Glorioso, and Scott L. Nyberg. Liver regeneration. *Translational research : the journal of laboratory and clinical medicine*, 163(4):352–362, 2014.

[42] Martin Holzenberger, Joëlle Dupont, Bertrand Ducos, Patricia Leneuve, Alain Géloën, Patrick C. Even, Pascale Cervera, and Yves Le Bouc. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*, 421(6919):182–187, 2003.

[43] Justin M. J. Travis. The evolution of programmed death in a spatially structured population. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 59(4):301–305, 2004.

[44] V. P. Skulachev. The programmed death phenomena, aging, and the Samurai law of biology. *Experimental gerontology*, 36(7):995–1024, 2001.

[45] Vladimir P. Skulachev. Programmed death phenomena: from organelle to organism. *Annals of the New York Academy of Sciences*, 959:214–237, 2002.

[46] J. D. Griffith, L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss, and T. de Lange. Mammalian telomeres end in a large duplex loop. *Cell*, 97(4):503–514, 1999.

[47] K. Collins. Mammalian telomeres and telomerase. *Current opinion in cell biology*, 12(3):378–383, 2000.

[48] Koji Okuda, Arlene Bardeguez, Jeffrey P. Gardner, Paulette Rodriguez, Vijaya Ganesh, Masayuki Kimura, Joan Skurnick, Girgis Awad, and Abraham Aviv. Telomere length in the newborn. *Pediatric research*, 52(3):377–381, 2002.

[49] Yasumichi Arai, Carmen M. Martin-Ruiz, Michiyo Takayama, Yukiko Abe, Toru Takebayashi, Shigeo Koyasu, Makoto Suematsu, Nobuyoshi Hirose, and Thomas von Zglinicki. Inflammation, But Not Telomere Length, Predicts Successful Ageing at Extreme Old Age: A Longitudinal Study of Semi-supercentenarians. *EBioMedicine*, 2(10):1549–1558, 2015.

[50] Manuel Collado, Maria A. Blasco, and Manuel Serrano. Cellular senescence in cancer and aging. *Cell*, 130(2):223–233, 2007.

[51] A. M. Olovnikov. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of theoretical biology*, 41(1):181–190, 1973.

[52] Judith Campisi. Aging, tumor suppression and cancer: high wire-act! *Mechanisms of ageing and development*, 126(1):51–58, 2005.

[53] A. Krtolica, S. Parrinello, S. Lockett, P. Y. Desprez, and J. Campisi. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12072–12077, 2001.

[54] J. K. Cowell. Telomeres and telomerase in ageing and cancer. *Age*, 22(2):59–64, 1999.

[55] George C. Williams. Pleiotropy, natural selection, and the evolution of senescence. *Evolution*, 11(4):398–411, 1957.

[56] N. V. Cherdyntseva, P. A. Gervas, N. V. Litvyakov, M. N. Stakcheeva, A. A. Ponomaryeva, A. Yu Dobrodeev, E. V. Denisov, V. A. Belyavskaya, and E. L. Choinzonov. Age-related function of tumor suppressor gene TP53:contribution to cancer risk and progression. *Experimental oncology*, 32(3):205–208, 2010.

[57] Robin Holliday. Aging is no longer an unsolved problem in biology. *Annals of the New York Academy of Sciences*, 1067:1–9, 2006.

[58] Cathérine Dupont, D. Randall Armant, and Carol A. Brenner. Epigenetics: definition, mechanisms and clinical perspective. *Seminars in reproductive medicine*, 27(5):351–357, 2009.

[59] Gregory Hannum, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, Marina Bibikova, Jian-Bing Fan, Yuan Gao, Rob Deconde, Menzies Chen, Indika Rajapakse, Stephen Friend, Trey Ideker, and Kang Zhang. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell*, 49(2):359–367, 2013.

[60] Jordana T. Bell, Pei-Chien Tsai, Tsun-Po Yang, Ruth Pidsley, James Nisbet, Daniel Glass, Massimo Mangino, Guangju Zhai, Feng Zhang, Ana Valdes, So-Youn Shin, Emma L. Dempster, Robin M. Murray, Elin Grundberg, Asa K. Hedman, Alexandra Nica, Kerrin S. Small, Emmanouil T. Dermitzakis, Mark I. McCarthy, Jonathan Mill, Tim D. Spector, and Panos Deloukas. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics*, 8(4):e1002629, 2012.

[61] Carmen M. Koch and Wolfgang Wagner. Epigenetic-aging-signature to determine age in different tissues. *Aging*, 3(10):1018–1027, 2011.

[62] Joseph L. McClay, Karolina A. Aberg, Shaunna L. Clark, Srilaxmi Nerella, Gaurav Kumar, Lin Y. Xie, Alexandra D. Hudson, Aki Harada, Christina M. Hultman, Patrik K. E. Magnusson, Patrick F. Sullivan, and Edwin J. C. G. van den Oord. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Human molecular genetics*, 23(5):1175–1185, 2014.

[63] Jean-Pierre Issa. Aging and epigenetic drift: a vicious cycle. *The Journal of clinical investigation*, 124(1):24–29, 2014.

[64] C. E. Finch and G. Ruvkun. The genetics of aging. *Annual review of genomics and human genetics*, 2:435–462, 2001.

[65] Reid F. Thompson, Gil Atzmon, Ciprian Gheorghe, Hong Qian Liang, Christina Lowes, John M. Greally, and Nir Barzilai. Tissue-specific dysregulation of DNA methylation in aging. *Aging cell*, 9(4):506–518, 2010.

[66] Mario F. Fraga, Esteban Ballestar, Maria F. Paz, Santiago Ropero, Fernando Setien, Maria L. Ballestar, Damia Heine-Suñer, Juan C. Cigudosa, Miguel Urioste, Javier Benitez, Manuel Boix-Chornet, Abel Sanchez-Aguilera, Charlotte Ling, Emma Carlsson, Pernille Poulsen, Allan Vaag, Zarko Stephan, Tim D. Spector, Yue-Zhong Wu, Christoph Plass, and Manel Esteller. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10604–10609, 2005.

[67] T. Finkel and N. J. Holbrook. Oxidants, oxidative stress and the biology of ageing. *Nature*, 408(6809):239–247, 2000.

[68] D. Harman. The biologic clock: the mitochondria? *Journal of the American Geriatrics Society*, 20(4):145–147, 1972.

[69] Denham Harman. The free radical theory of aging. *Antioxidants & redox signaling*, 5(5):557–561, 2003.

[70] T. P. A. Devasagayam, J. C. Tilak, K. K. Boloor, Ketaki S. Sane, Saroj S. Ghaskadbi, and R. D. Lele. Free radicals and antioxidants in human health: current status and future prospects. *The Journal of the Association of Physicians of India*, 52:794–804, 2004.

[71] Samuel E. Schriner, Nancy J. Linford, George M. Martin, Piper Treuting, Charles E. Ogburn, Mary Emond, Pinar E. Coskun, Warren Ladiges, Norman Wolf, Holly van Remmen, Douglas C. Wallace, and Peter S. Rabinovitch. Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science (New York, N.Y.)*, 308(5730):1909–1911, 2005.

[72] Youngmok C. Jang, Viviana I. Pérez, Wook Song, Michael S. Lustgarten, Adam B. Salmon, James Mele, Wenbo Qi, Yuhong Liu, Hanyu Liang, Asish Chaudhuri, Yuji Ikeno, Charles J. Epstein, Holly van Remmen, and Arlan Richardson. Overexpression of Mn superoxide dismutase does not increase life span in mice. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 64(11):1114–1125, 2009.

[73] C. D. Powell, D. E. Quain, and K. A. Smart. The impact of media composition and petite mutation on the longevity of a polyploid brewing yeast strain. *Letters in applied microbiology*, 31(1):46–51, 2000.

[74] Frances M. Platt, Barry Boland, and Aarnoud C. van der Spoel. The cell biology of disease: lysosomal storage disorders: the cellular impact of lysosomal dysfunction. *The Journal of cell biology*, 199(5):723–734, 2012.

[75] Joseph Kochmanski, Luke Montrose, Jaclyn M. Goodrich, and Dana C. Dolinoy. Environmental Deflection: The Impact of Toxicant Exposures on the Aging Epigenome. *Toxicological sciences : an official journal of the Society of Toxicology*, 156(2):325–335, 2017.

[76] Sonia Shah, Allan F. McRae, Riccardo E. Marioni, Sarah E. Harris, Jude Gibson, Anjali K. Henders, Paul Redmond, Simon R. Cox, Alison Pattie, Janie Corley, Lee Murphy, Nicholas G. Martin, Grant W. Montgomery, John M. Starr, Naomi R. Wray, Ian J. Deary, and Peter M. Visscher. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome research*, 24(11):1725–1733, 2014.

[77] Meaghan J. Jones, Sarah J. Goodman, and Michael S. Kobor. DNA methylation and healthy human aging. *Aging cell*, 14(6):924–932, 2015.

[78] Yulia A. Medvedeva, Abdullah M. Khamis, Ivan V. Kulakovskiy, Wail Ba-Alawi, Md Shariful I. Bhuyan, Hideya Kawaji, Timo Lassmann, Matthias Harbers, Alistair R. R. Forrest, and Vladimir B. Bajic. Effects of cytosine methylation on transcription factor binding sites. *BMC genomics*, 15:119, 2014.

[79] David Martino, Yuk Jin Loke, Lavinia Gordon, Miina Ollikainen, Mark N. Cruickshank, Richard Saffery, and Jeffrey M. Craig. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome biology*, 14(5):R42, 2013.

[80] Christopher Faulk, Kevin Liu, Amanda Barks, Jaclyn M. Goodrich, and Dana C. Dolinoy. Longitudinal epigenetic drift in mice perinatally exposed to lead. *Epigenetics*, 9(7):934–941, 2014.

[81] Kathleen M. Gilbert, Sarah J. Blossom, Stephen W. Erickson, Brad Reisfeld, Todd J. Zurlinden, Brannon Broadfoot, Kirk West, Shasha Bai, and Craig A. Cooney. Chronic exposure to water pollutant trichloroethylene increased epigenetic drift in CD4(+) T cells. *Epigenomics*, 8(5):633–649, 2016.

[82] Huiling Xue, Bo Xian, Dong Dong, Kai Xia, Shanshan Zhu, Zhongnan Zhang, Lei Hou, Qingpeng Zhang, Yi Zhang, and Jing-Dong J. Han. A modular network model of aging. *Molecular systems biology*, 3:147, 2007.

[83] Gábor I. Simkó, Dávid Gyurkó, Dániel V. Veres, Tibor Nánási, and Peter Csermely. Network strategies to understand the aging process and help age-related drug design. *Genome medicine*, 1(9):90, 2009.

[84] Matteo Tosato, Valentina Zamboni, Alessandro Ferrini, and Matteo Cesari. The aging process and potential interventions to extend life expectancy. *Clinical interventions in aging*, 2(3):401–412, 2007.

[85] L. Hayflick. How and why we age. *Experimental gerontology*, 33(7-8):639–653, 1998.

[86] L. Hayflick. The future of ageing. *Nature*, 408(6809):267–269, 2000.

[87] Burr Atkinson. *Changes in Eukaryotic Gene Expression in Response to Environmental Stress*. Elsevier Science, Oxford, 5th ed. edition, 1985.

[88] Keith A. Spriggs, Martin Bushell, and Anne E. Willis. Translational regulation of gene expression during conditions of cell stress. *Molecular cell*, 40(2):228–237, 2010.

[89] V. R. Iyer, M. B. Eisen, D. T. Ross, G. Schuler, T. Moore, J. C. Lee, J. M. Trent, L. M. Staudt, J. Hudson, M. S. Boguski, D. Lashkari, D. Shalon, D. Botstein, and P. O. Brown. The transcriptional program in the response of human fibroblasts to serum. *Science (New York, N.Y.)*, 283(5398):83–87, 1999.

[90] Matthew B. Avison. *Measuring gene expression*. The basics. Taylor & Francis, New York and Abingdon [England], 2007.

[91] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):57–63, 2009.

[92] Jay Shendure and Hanlee Ji. Next-generation DNA sequencing. *Nature biotechnology*, 26(10):1135–1145, 2008.

[93] Michael L. Metzker. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1):31–46, 2010.

[94] Sara Goodwin, John D. McPherson, and W. Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature reviews. Genetics*, 17(6):333–351, 2016.

[95] John M. S. Bartlett and David Stirling. A short history of the polymerase chain reaction. *Methods in molecular biology (Clifton, N.J.)*, 226:3–6, 2003.

[96] Fatih Ozsolak and Patrice M. Milos. RNA sequencing: advances, challenges and opportunities. *Nature reviews. Genetics*, 12(2):87–98, 2011.

[97] H. P. J. Buermans and J. T. den Dunnen. Next generation sequencing technology: Advances and applications. *Biochimica et biophysica acta*, 1842(10):1932–1941, 2014.

[98] Piotr Chomczynski and Nicoletta Sacchi. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature protocols*, 1(2):581–585, 2006.

[99] Valerio Costa, Claudia Angelini, Italia de Feis, and Alfredo Ciccodicola. Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of biomedicine & biotechnology*, 2010:853916, 2010.

[100] Colin C. Pritchard, Heather H. Cheng, and Muneesh Tewari. MicroRNA profiling: approaches and considerations. *Nature reviews. Genetics*, 13(5):358–369, 2012.

[101] M. Hoelzer. *The Dark Art of Next-Generation Sequencing – Fundamental approaches for genomics, transcriptomics, and differential gene expression.* PhD thesis, Friedrich Schiller University Jena, 2017.

[102] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, 2009.

[103] Steve Hoffmann, Christian Otto, Stefan Kurtz, Cynthia M. Sharma, Philipp Khaitovich, Jörg Vogel, Peter F. Stadler, and Jörg Hackermüller. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS computational biology*, 5(9):e1000502, 2009.

[104] Cole Trapnell, Lior Pachter, and Steven L. Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9):1105–1111, 2009.

[105] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.

[106] Chandra Shekhar Pareek, Rafal Smoczynski, and Andrzej Tretyn. Sequencing technologies and genome sequencing. *Journal of applied genetics*, 52(4):413–435, 2011.

[107] Hayan Lee and Michael C. Schatz. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. *Bioinformatics (Oxford, England)*, 28(16):2097–2105, 2012.

[108] Yang Liao, Gordon K. Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7):923–930, 2014.

[109] J. Lederberg and A. T. McCray. Ome sweet omics: a genealogical treasury of words. *Scientist*, 15(8), 2001.

[110] Alessandro Cellerino and Alessandro Ori. What have we learned on aging from omics studies? *Seminars in cell & developmental biology*, 70:177–189, 2017.

[111] Volker Kroehne, Dorian Freudenreich, Stefan Hans, Jan Kaslin, and Michael Brand. Regeneration of the adult zebrafish brain from neurogenic radial glia-type progenitors. *Development (Cambridge, England)*, 138(22):4831–4841, 2011.

[112] Rochelle Buffenstein. The naked mole-rat: a new long-living model for human aging research. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 60(11):1369–1377, 2005.

[113] Thomas J. Park, Jane Reznick, Bethany L. Peterson, Gregory Blass, Damir Omerbašić, Nigel C. Bennett, P. Henning J. L. Kuich, Christin Zasada, Brigitte M. Browe, Wiebke Hamann, Daniel T. Applegate, Michael H. Radke, Tetiana Kosten, Heike Lutermann, Victoria Gavaghan, Ole Eigenbrod, Valérie Bégay, Vince G. Amoroso, Vidya Govind, Richard D. Minshall, Ewan St J. Smith, John Larson, Michael Gotthardt, Stefan Kempa, and Gary R. Lewin. Fructose-driven glycolysis supports anoxia resistance in the naked mole-rat. *Science (New York, N.Y.)*, 356(6335):307–311, 2017.

[114] Junaidi Payne and J. Cede Prudente. *Orang-utans: Behaviour, ecology and conservation.* New Holland, London, 1. publ edition, 2008.

[115] Caleb E. Finch. Evolution in health and medicine Sackler colloquium: Evolution of the human lifespan and diseases of aging: roles of infection, inflammation, and nutrition. *Proceedings of the National Academy of Sciences of the United States of America*, 107 Suppl 1:1718–1724, 2010.

[116] Peer Aramillo Irizar, Sascha Schäuble, Daniela Esser, Marco Groth, Christiane Frahm, Steffen Priebe, Mario Baumgart, Nils Hartmann, Shiva Marthandan, Uwe Menzel, Julia Müller, Silvio Schmidt, Volker Ast, Amke Caliebe, Rainer König, Michael Krawczak, Michael Ristow, Stefan Schuster, Alessandro Cellerino, Stephan Diekmann, Christoph Englert, Peter Hemmerich, Jürgen Sühnel, Reinhard Guthke, Otto W. Witte, Matthias Platzer, Eytan Ruppin, and Christoph Kaleta. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nature communications*, 9(1):327, 2018.

[117] Eva Terzibasi, Christel Lefrançois, Paolo Domenici, Nils Hartmann, Michael Graf, and Alessandro Cellerino. Effects of dietary restriction on mortality and age-related phenotypes in the short-lived fish Nothobranchius furzeri. *Aging cell*, 8(2):88–99, 2009.

[118] Eva Terzibasi, Dario Riccardo Valenzano, Mauro Benedetti, Paola Roncaglia, Antonino Cattaneo, Luciano Domenici, and Alessandro Cellerino. Large differences in aging phenotype between strains of the short-lived annual fish Nothobranchius furzeri. *PloS one*, 3(12):e3866, 2008.

[119] A. Srivastava, E. Barth, M. A. Ermolaeva, M. Guenther, C. Frahm, M. Marz, and O. W. Witte. Tissue-specific gene expression changes are associated with aging in mice. *Genomics, Proteomics & Bioinformatics*, pages accepted, publication pending, 2019.

[120] Mario Baumgart, Marco Groth, Steffen Priebe, Jessika Appelt, Reinhard Guthke, Matthias Platzer, and Alessandro Cellerino. Age-dependent regulation of tumor-related microRNAs in the brain of the annual fish Nothobranchius furzeri. *Mechanisms of ageing and development*, 133(5):226–233, 2012.

[121] Mario Baumgart, Emanuel Barth, Aurora Savino, Marco Groth, Philipp Koch, Andreas Petzold, Ivan Arisi, Matthias Platzer, Manja Marz, and Alessandro Cellerino. A miRNA catalogue and ncRNA annotation of the short-living fish Nothobranchius furzeri. *BMC genomics*, 18(1):693, 2017.

[122] Daniel R. Zerbino, Premanand Achuthan, Wasiu Akanni, M. Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, Laurent Gil, Leo Gordon, Leanne Haggerty, Erin Haskell, Thibaut Hourlier, Osagie G. Izuogu, Sophie H. Janacek, Thomas Juettemann, Jimmy Kiang To, Matthew R. Laird, Ilias Lavidas, Zhicheng Liu, Jane E. Loveland, Thomas Maurel, William McLaren, Benjamin Moore, Jonathan Mudge, Daniel N. Murphy, Victoria Newman, Michael Nuhn, Denye Ogeh, Chuang Kee Ong, Anne Parker, Mateus Patricio, Harpreet Singh Riat, Helen Schuilenburg, Dan Sheppard, Helen Sparrow, Kieron Taylor, Anja Thormann, Alessandro Vullo, Brandon Walts, Amonida Zadissa, Adam Frankish, Sarah E. Hunt, Myrto Kostadima, Nicholas Langridge, Fergal J. Martin, Matthieu Muffato, Emily Perry, Magali Ruffier, Dan M. Staines, Stephen J. Trevanion, Bronwen L. Aken, Fiona Cunningham, Andrew Yates, and Paul Flicek. Ensembl 2018. *Nucleic acids research*, 46(D1):D754–D761, 2018.

[123] Kathrin Reichwald, Andreas Petzold, Philipp Koch, Bryan R. Downie, Nils Hartmann, Stefan Pietsch, Mario Baumgart, Domitille Chalopin, Marius Felder, Martin Bens, Arne Sahm, Karol Szafranski, Stefan Taudien, Marco Groth, Ivan Arisi, Anja Weise, Samarth S. Bhatt, Virag Sharma, Johann M. Kraus, Florian Schmid, Steffen Priebe, Thomas Liehr, Matthias Görlach, Manuel E. Than, Michael Hiller, Hans A. Kestler, Jean-Nicolas Volff, Manfred Schartl, Alessandro Cellerino, Christoph Englert, and Matthias Platzer. Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell*, 163(6):1527–1538, 2015.

[124] Robert Schmieder and Robert Edwards. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, 27(6):863–864, 2011.

[125] Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4):R36, 2013.

[126] Günter P. Wagner, Koryu Kin, and Vincent J. Lynch. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences = Theorie in den Biowissenschaften*, 131(4):281–285, 2012.

[127] Helga Thorvaldsdóttir, James T. Robinson, and Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2):178–192, 2013.

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

[129] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[130] World Health Organization. *World Health Statistics 2016: Monitoring Health for the Sustainable Development Goals (SDGs)*. Nonserial Publication. World Health Organization, Geneva, 2016.

[131] Robert V. Kail. *Essentials of human development: A life-span view*. Cengage Learning, Australia, second edition edition, 2017.

[132] Michel Allard, Victor Lèbre, and Jean-Marie Robine. *Jeanne Calment: From Van Gogh's time to ours : 122 extraordinary years*. W.H. Freeman, New York, 1998.

[133] S. Pobojewski. World's oldest mouse reaches milestone birthday. *University of Michigan News Service.*, 19 April 2004. Available from: www.ur.umich.edu/0304/Apr19_04
/26.shtml Cited 19 October 2018.

[134] Roberts Rugh. *The mouse: Its reproduction and development*. Oxford science publications. Oxford Univ. Press, Oxford, [nachdr.] edition, 2008.

[135] Daniel J. Aires, Graham Rockwell, Ting Wang, Jennifer Frontera, Jo Wick, Wenfang Wang, Marija Tonkovic-Capin, Jianghua Lu, Lezi E, Hao Zhu, and Russell H. Swerdlow. Potentiation of dietary restriction-induced lifespan extension by polyphenols. *Biochimica et biophysica acta*, 1822(4):522–526, 2012.

[136] Richard A. Miller, James M. Harper, Robert C. Dysko, Stephen J. Durkee, and Steven N. Austad. Longer life spans and delayed maturation in wild-derived mice. *Experimental biology and medicine (Maywood, N.J.)*, 227(7):500–508, 2002.

[137] Jules Richard Kemadjou Njiwa, Paul Müller, and Roland Klein. Life Cycle Stages and Length of Zebrafish (Danio rerio) Exposed to DDT. *JOURNAL OF HEALTH SCIENCE*, 50(3):220–225, 2004.

[138] Shuji Kishi, Junzo Uchiyama, Anne M. Baughman, Tadateru Goto, Mao C. Lin, and Stephanie B. Tsai. The zebrafish as a vertebrate model of functional aging and very gradual senescence. *Experimental gerontology*, 38(7):777–786, 2003.

[139] Evan T. Keller and Jill M. Murtha. The use of mature zebrafish (Danio rerio) as a model for human aging and disease. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP*, 138(3):335–341, 2004.

[140] Glenn S. Gerhard, Elizabeth J. Kauffman, Xujun Wang, Richard Stewart, Jessica L. Moore, Claudia J. Kasales, Eugene Demidenko, and Keith C. Cheng. Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). *Experimental gerontology*, 37(8-9):1055–1068, 2002.

[141] Stefano Valdesalici and Alessandro Cellerino. Extremely short lifespan in the annual fish Nothobranchius furzeri. *Proceedings. Biological sciences*, 270 Suppl 2:S189–91, 2003.

[142] Milan Vrtílek, Jakub Žák, Martin Pšenička, and Martin Reichard. Extremely rapid maturation of a wild African annual fish. *Current biology : CB*, 28(15):R822–R824, 2018.

[143] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[144] Da Wei Huang, Brad T. Sherman, Qina Tan, Jack R. Collins, W. Gregory Alvord, Jean Roayaei, Robert Stephens, Michael W. Baseler, H. Clifford Lane, and Richard A. Lempicki. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, 8(9):R183, 2007.

[145] Matthias W. Sieber, Madlen Guenther, Nadine Jaenisch, Daniela Albrecht-Eckardt, Matthias Kohl, Otto W. Witte, and Christiane Frahm. Age-specific transcriptional response to stroke. *Neurobiology of aging*, 35(7):1744–1754, 2014.

[146] Stephen A. Bustin. A - Z of quantitative PCR. *Internat. Univ. Line*, 5:89–113, 2004.

[147] Martijs J. Jonker, Joost Pm Melis, Raoul V. Kuiper, Tessa V. van der Hoeven, P. F. K. Wackers, Joke Robinson, Gijsbertus Tj van der Horst, Martijn Et Dollé, Jan Vijg, Timo M. Breit, Jan Hj Hoeijmakers, and Harry van Steeg. Life spanning murine gene expression profiles in relation to chronological and pathological aging in multiple organs. *Aging cell*, 12(5):901–909, 2013.

[148] Ismene Karakasilioti and George A. Garinis. Tissue-specific aging: a tale of functional asymmetry. *Aging*, 6(1):7–8, 2014.

[149] Abhinandana Anantharaju, Axel Feller, and Antonio Chedid. Aging Liver. A review. *Gerontology*, 48(6):343–353, 2002.

[150] Sabrina Gohlke, Michele Mishto, Kathrin Textoris-Taube, Christin Keller, Carolin Giannini, Francesco Vasuri, Elisa Capizzi, Antonia D'Errico-Grigioni, Peter-Michael Kloetzel, and Burkhardt Dahlmann. Molecular alterations in proteasomes of rat liver during aging result in altered proteolytic activities. *Age (Dordrecht, Netherlands)*, 36(1):57–72, 2014.

[151] Hannes Lans and Wim Vermeulen. Tissue specific response to DNA damage: C. elegans as role model. *DNA repair*, 32:141–148, 2015.

[152] Ivan K. Chinn, Clare C. Blackburn, Nancy R. Manley, and Gregory D. Sempowski. Changes in primary lymphoid organs with aging. *Seminars in immunology*, 24(5):309–320, 2012.

[153] Roman M. Stilling, Eva Benito, Michael Gertig, Jonas Barth, Vincenzo Capece, Susanne Burkhardt, Stefan Bonn, and Andre Fischer. De-regulation of gene expression and alternative splicing affects distinct cellular pathways in the aging hippocampus. *Frontiers in cellular neuroscience*, 8:373, 2014.

[154] Shona H. Wood, Thomas Craig, Yang Li, Brian Merry, and João Pedro de Magalhães. Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome. *Age (Dordrecht, Netherlands)*, 35(3):763–776, 2013.

[155] Suzanne E. Hickman, Nathan D. Kingery, Toshiro K. Ohsumi, Mark L. Borowsky, Li-chong Wang, Terry K. Means, and Joseph El Khoury. The microglial sensome revealed by direct RNA sequencing. *Nature neuroscience*, 16(12):1896–1905, 2013.

[156] Patrick M. Loerch, Tao Lu, Kelly A. Dakin, James M. Vann, Adrian Isaacs, Chengiz Geula, Jianbin Wang, Ying Pan, Dana H. Gabuzda, Cheng Li, Tomas A. Prolla, and Bruce A. Yankner. Evolution of the aging brain transcriptome and synaptic regulation. *PloS one*, 3(10):e3329, 2008.

[157] Marie Orre, Willem Kamphuis, Lana M. Osborn, Jeroen Melief, Lieneke Kooijman, Inge Huitinga, Jan Klooster, Koen Bossers, and Elly M. Hol. Acute isolation and transcriptome characterization of cortical astrocytes and microglia from young and aged mice. *Neurobiology of aging*, 35(1):1–14, 2014.

[158] Christiane Frahm, Akash Srivastava, Silvio Schmidt, Jule Mueller, Marco Groth, Madlen Guenther, Yuanyuan Ji, Steffen Priebe, Matthias Platzer, and Otto W. Witte. Transcriptional profiling reveals protective mechanisms in brains of long-lived mice. *Neurobiology of aging*, 52:23–31, 2017.

[159] Anja Urbach, Christoph Redecker, and Otto W. Witte. Induction of neurogenesis in the adult dentate gyrus by cortical spreading depression. *Stroke*, 39(11):3064–3072, 2008.

[160] Bruna Bellaver, Débora Guerini Souza, Diogo Onofre Souza, and André Quincozes-Santos. Hippocampal Astrocyte Cultures from Adult and Aged Rats Reproduce Changes in Glial Functionality Observed in the Aging Brain. *Molecular neurobiology*, 54(4):2969–2985, 2017.

[161] Emily L. Goldberg and Vishwa Deep Dixit. Drivers of age-related inflammation and strategies for healthspan extension. *Immunological reviews*, 265(1):63–74, 2015.

[162] Hyoung-Gon Lee, Gemma Casadesus, Xiongwei Zhu, Rudy J. Castellani, Andrew McShea, George Perry, Robert B. Petersen, Vladan Bajic, and Mark A. Smith. Cell cycle re-entry mediated neurodegeneration and its treatment role in the pathogenesis of Alzheimer's disease. *Neurochemistry international*, 54(2):84–88, 2009.

[163] Xiaoyu Li, Xia Bu, Binfeng Lu, Hava Avraham, Richard A. Flavell, and Bing Lim. The hematopoiesis-specific GTP-binding protein RhoH is GTPase deficient and modulates activities of other Rho GTPases by an inhibitory function. *Molecular and cellular biology*, 22(4):1158–1171, 2002.

[164] Kirsty S. Brown, Derek Blair, Steven D. Reid, Emma K. Nicholson, and Margaret M. Harnett. FcgammaRIIb-mediated negative regulation of BCR signalling is associated with the recruitment of the MAPkinase-phosphatase, Pac-1, and the 3'-inositol phosphatase, PTEN. *Cellular signalling*, 16(1):71–80, 2004.

[165] Xiao-Tong Ma, Bin Xu, Li-Li An, Cheng-Ya Dong, Yong-Min Lin, Yang Shi, and Ke-Fu Wu. Vaccine with beta-defensin 2-transduced leukemic cells activates innate and adaptive immunity to elicit potent antileukemia responses. *Cancer research*, 66(2):1169–1176, 2006.

[166] J. Antoniou, N. M. Goudsouzian, T. F. Heathfield, N. Winterbottom, T. Steffen, A. R. Poole, M. Aebi, and M. Alini. The human lumbar endplate. Evidence of changes in biosynthesis and denaturation of the extracellular matrix with growth, maturation, aging, and degeneration. *Spine*, 21(10):1153–1161, 1996.

[167] Gary J. Fisher, Sewon Kang, James Varani, Zsuzsanna Bata-Csorgo, Yinsheng Wan, Subhash Datta, and John J. Voorhees. Mechanisms of photoaging and chronological skin aging. *Archives of dermatology*, 138(11):1462–1470, 2002.

[168] K. Takeda, A. Gosiewska, and B. Peterkofsky. Similar, but not identical, modulation of expression of extracellular matrix components during in vitro and in vivo aging of human skin fibroblasts. *Journal of cellular physiology*, 153(3):450–459, 1992.

[169] Jordan M. Winter, Angela H. Ting, Felip Vilardell, Eike Gallmeier, Steve B. Baylin, Ralph H. Hruban, Scott E. Kern, and Christine A. Iacobuzio-Donahue. Absence of E-cadherin expression distinguishes noncohesive from cohesive pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(2):412–418, 2008.

[170] Alexander Hermani, Jochen Hess, Barbara de Servi, Senad Medunjanin, Rainer Grobholz, Lutz Trojan, Peter Angel, and Doris Mayer. Calcium-binding proteins S100A8 and S100A9 as novel diagnostic markers in human prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 11(14):5146–5152, 2005.

[171] C. E. Shepherd, J. Goyette, V. Utter, F. Rahimi, Z. Yang, C. L. Geczy, and G. M. Halliday. Inflammatory S100A9 and S100A12 proteins in Alzheimer's disease. *Neurobiology of aging*, 27(11):1554–1563, 2006.

[172] Janet C. Koon and Terrance J. Kubiseski. Developmental arrest of Caenorhabditis elegans BRAP-2 mutant exposed to oxidative stress is dependent on BRC-1. *The Journal of biological chemistry*, 285(18):13437–13443, 2010.

[173] Ana Maria Cuervo. Autophagy and aging: keeping that old broom working. *Trends in genetics : TIG*, 24(12):604–612, 2008.

[174] William C. Burhans and Martin Weinberger. DNA replication stress, genome instability and aging. *Nucleic acids research*, 35(22):7545–7556, 2007.

[175] Alex A. Freitas and João Pedro de Magalhães. A review and appraisal of the DNA damage theory of ageing. *Mutation research*, 728(1-2):12–22, 2011.

[176] Tomas A. Prolla. DNA microarray analysis of the aging brain. *Chemical senses*, 27(3):299–306, 2002.

[177] Anna Christina Sondersorg, Daniela Busse, Jessica Kyereme, Markus Rothermel, Gitta Neufang, Günter Gisselmann, Hanns Hatt, and Heike Conrad. Chemosensory information processing between keratinocytes and trigeminal neurons. *The Journal of biological chemistry*, 289(25):17529–17540, 2014.

[178] M. F. Beal. Aging, energy, and oxidative stress in neurodegenerative diseases. *Annals of neurology*, 38(3):357–366, 1995.

[179] Alexei Terman, Tino Kurz, Marian Navratil, Edgar A. Arriaga, and Ulf T. Brunk. Mitochondrial turnover and aging of long-lived postmitotic cells: the mitochondrial-lysosomal axis theory of aging. *Antioxidants & redox signaling*, 12(4):503–535, 2010.

[180] L. K. Kwong and R. S. Sohal. Age-related changes in activities of mitochondrial electron transport complexes in various tissues of the mouse. *Archives of biochemistry and biophysics*, 373(1):16–22, 2000.

[181] Yan Hong, Fengyuan Piao, Yufeng Zhao, Sheng Li, Yanyan Wang, and Peng Liu. Subchronic exposure to arsenic decreased Sdha expression in the brain of mice. *Neurotoxicology*, 30(4):538–543, 2009.

[182] Larisa V. Kovtonyuk, Kristin Fritsch, Xiaomin Feng, Markus G. Manz, and Hitoshi Takizawa. Inflamm-Aging of Hematopoiesis, Hematopoietic Stem Cells, and the Bone Marrow Microenvironment. *Frontiers in immunology*, 7:502, 2016.

[183] L. C. Rønn, V. Berezin, and E. Bock. The neural cell adhesion molecule in synaptic plasticity and ageing. *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*, 18(2-3):193–199, 2000.

[184] Yani Zou, Kyung Jin Jung, Jung Won Kim, Byung Pal Yu, and Hae Young Chung. Alteration of soluble adhesion molecules during aging and their modulation by calorie restriction. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 18(2):320–322, 2004.

[185] Franz Bauernfeind, Sven Niepmann, Percy A. Knolle, and Veit Hornung. Aging-Associated TNF Production Primes Inflammasome Activation and NLRP3-Related Metabolic Disturbances. *Journal of immunology (Baltimore, Md. : 1950)*, 197(7):2900–2908, 2016.

[186] H. Bruunsgaard, M. Pedersen, and B. K. Pedersen. Aging and proinflammatory cytokines. *Current opinion in hematology*, 8(3):131–136, 2001.

[187] B. S. Nielsen, N. Borregaard, J. R. Bundgaard, S. Timshel, M. Sehested, and L. Kjeldsen. Induction of NGAL synthesis in epithelial cells of human colorectal neoplasia and inflammatory bowel diseases. *Gut*, 38(3):414–420, 1996.

[188] Michael F. Criscitiello and Martin F. Flajnik. Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *European journal of immunology*, 37(10):2683–2694, 2007.

[189] Thomas Vogl, Klaus Tenbrock, Stephan Ludwig, Nadja Leukert, Christina Ehrhardt, Marieke A. D. van Zoelen, Wolfgang Nacken, Dirk Foell, Tom van der Poll, Clemens Sorg, and Johannes Roth. Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nature medicine*, 13(9):1042–1049, 2007.

[190] Ana C. Ferreira, Sandro Dá Mesquita, João C. Sousa, Margarida Correia-Neves, Nuno Sousa, Joana A. Palha, and Fernanda Marques. From the periphery to the brain: Lipocalin-2, a friend or foe? *Progress in neurobiology*, 131:120–136, 2015.

[191] Faiez Al Nimer, Christina Elliott, Joakim Bergman, Mohsen Khademi, Ann M. Dring, Shahin Aeinehband, Tommy Bergenheim, Jeppe Romme Christensen, Finn Sellebjerg, Anders Svenningsson, Christopher Linington, Tomas Olsson, and Fredrik Piehl. Lipocalin-2 is increased in progressive multiple sclerosis and inhibits remyelination. *Neurology(R) neuroimmunology & neuroinflammation*, 3(1):e191, 2016.

[192] Kyoungho Suk. Lipocalin-2 as a therapeutic target for brain injury: An astrocentric perspective. *Progress in neurobiology*, 144:158–172, 2016.

[193] Q. Yang, D. O'Hanlon, C. W. Heizmann, and A. Marks. Demonstration of heterodimer formation between S100B and S100A6 in the yeast two-hybrid system and human melanoma. *Experimental cell research*, 246(2):501–509, 1999.

[194] J. C. Deloulme, N. Assard, G. O. Mbele, C. Mangin, R. Kuwano, and J. Baudier. S100A6 and S100A11 are specific targets of the calcium- and zinc-binding S100B protein in vivo. *The Journal of biological chemistry*, 275(45):35302–35310, 2000.

[195] Maria I. Molejon, Alejandro Ropolo, Andrea Lo Re, Veronica Boggio, and Maria I. Vaccaro. The VMP1-Beclin 1 interaction regulates autophagy induction. *Scientific reports*, 3:1055, 2013.

[196] Ana Maria Cuervo. Autophagy: in sickness and in health. *Trends in cell biology*, 14(2):70–77, 2004.

[197] Taichi Hara, Kenji Nakamura, Makoto Matsui, Akitsugu Yamamoto, Yohko Nakahara, Rika Suzuki-Migishima, Minesuke Yokoyama, Kenji Mishima, Ichiro Saito, Hideyuki Okano, and Noboru Mizushima. Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature*, 441(7095):885–889, 2006.

[198] B. Levine and G. Kroemer. Autophagy in aging, disease and death: the true identity of a cell death impostor. *Cell death and differentiation*, 16(1):1–2, 2009.

[199] Vanessa Nancy, Rob M. F. Wolthuis, Marie-France de Tand, Isabelle Janoueix-Lerosey, Johannes L. Bos, and Jean de Gunzburg. Identification and Characterization of Potential Effector Molecules of the Ras-related GTPase Rap2. *Journal of Biological Chemistry*, 274(13):8737–8745, 1999.

[200] Tamar Hashimshony, Martin Feder, Michal Levin, Brian K. Hall, and Itai Yanai. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature*, 519(7542):219–222, 2015.

[201] Chang H. Kim. Homeostatic and pathogenic extramedullary hematopoiesis. *Journal of blood medicine*, 1:13–19, 2010.

[202] E. Panieri and M. M. Santoro. ROS homeostasis and metabolism: a dangerous liason in cancer cells. *Cell death & disease*, 7(6):e2253, 2016.

[203] Patricia Sancho, David Barneda, and Christopher Heeschen. Hallmarks of cancer stem cell metabolism. *British journal of cancer*, 114(12):1305–1312, 2016.

[204] Yongsoon Kim and Hong Sun. Functional genomic approach to identify novel genes involved in the regulation of oxidative stress resistance and animal lifespan. *Aging cell*, 6(4):489–503, 2007.

[205] E. Cadenas and K. J. Davies. Mitochondrial free radical generation, oxidative stress, and aging. *Free radical biology & medicine*, 29(3-4):222–230, 2000.

[206] S. R. Schwarze, R. Weindruch, and J. M. Aiken. Oxidative stress and aging reduce COX I RNA and cytochrome oxidase activity in Drosophila. *Free radical biology & medicine*, 25(6):740–747, 1998.

[207] Ines Soro-Arnaiz, Qilong Oscar Yang Li, Mar Torres-Capelli, Florinda Meléndez-Rodríguez, Sónia Veiga, Koen Veys, David Sebastian, Ainara Elorza, Daniel Tello, Pablo Hernansanz-Agustín, Sara Cogliati, Jose Maria Moreno-Navarrete, Eduardo Balsa, Esther Fuertes, Eduardo Romanos, Antonio Martínez-Ruiz, Jose Antonio Enriquez, Jose Manuel Fernandez-Real, Antonio Zorzano, Katrien de Bock, and Julián Aragonés. Role of Mitochondrial Complex IV in Age-Dependent Obesity. *Cell reports*, 16(11):2991–3002, 2016.

[208] Claudio Franceschi and Judith Campisi. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 69 Suppl 1:S4–9, 2014.

[209] Lilach Soreq, Jamie Rose, Eyal Soreq, John Hardy, Daniah Trabzuni, Mark R. Cookson, Colin Smith, Mina Ryten, Rickie Patani, and Jernej Ule. Major Shifts in Glial Regional Identity Are a Transcriptional Hallmark of Human Brain Aging. *Cell reports*, 18(2):557–570, 2017.

[210] R. C. Lee, R. L. Feinbaum, and V. Ambros. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854, 1993.

[211] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl. Identification of novel genes coding for small expressed RNAs. *Science (New York, N.Y.)*, 294(5543):853–858, 2001.

[212] Mariana Lagos-Quintana, Reinhard Rauhut, Jutta Meyer, Arndt Borkhardt, and Thomas Tuschl. New microRNAs from mouse and human. *RNA (New York, N.Y.)*, 9(2):175–179, 2003.

[213] A. E. Pasquinelli, B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Müller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89, 2000.

[214] Matthew W. Jones-Rhoades, David P. Bartel, and Bonnie Bartel. MicroRNAS and their regulatory roles in plants. *Annual review of plant biology*, 57:19–53, 2006.

[215] Lee Tuddenham, Jette S. Jung, Béatrice Chane-Woon-Ming, Lars Dölken, and Sébastien Pfeffer. Small RNA deep sequencing identifies microRNAs and other small noncoding RNAs from human herpesvirus 6B. *Journal of virology*, 86(3):1638–1649, 2012.

[216] David P. Bartel. Metazoan MicroRNAs. *Cell*, 173(1):20–51, 2018.

[217] Antony Rodriguez, Sam Griffiths-Jones, Jennifer L. Ashurst, and Allan Bradley. Identification of mammalian microRNA host genes and transcription units. *Genome research*, 14(10A):1902–1910, 2004.

[218] Yoontae Lee, Minju Kim, Jinju Han, Kyu-Hyun Yeom, Sanghyuk Lee, Sung Hee Baek, and V. Narry Kim. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, 23(20):4051–4060, 2004.

[219] Xuefeng Zhou, Jianhua Ruan, Guandong Wang, and Weixiong Zhang. Characterization and identification of microRNA core promoters in four model species. *PLoS computational biology*, 3(3):e37, 2007.

[220] Jinju Han, Yoontae Lee, Kyu-Hyun Yeom, Young-Kook Kim, Hua Jin, and V. Narry Kim. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & development*, 18(24):3016–3027, 2004.

[221] Eugene Berezikov, Wei-Jen Chung, Jason Willis, Edwin Cuppen, and Eric C. Lai. Mammalian mirtron genes. *Molecular cell*, 28(2):328–336, 2007.

[222] Elizabeth P. Murchison and Gregory J. Hannon. miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Current opinion in cell biology*, 16(3):223–229, 2004.

[223] E. Lund and J. E. Dahlberg. Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor symposia on quantitative biology*, 71:59–66, 2006.

[224] Jong-Eun Park, Inha Heo, Yuan Tian, Dhirendra K. Simanshu, Hyeshik Chang, David Jee, Dinshaw J. Patel, and V. Narry Kim. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 475(7355):201–205, 2011.

[225] Christine Lelandais-Brière, Céline Sorin, Marie Declerck, Abdelali Benslimane, Martin Crespi, and Caroline Hartmann. Small RNA diversity in plants and its impact in development. *Current genomics*, 11(1):14–23, 2010.

[226] Jr-Shiuan Yang and Eric C. Lai. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Molecular cell*, 43(6):892–903, 2011.

[227] Dianne S. Schwarz and Phillip D. Zamore. Why do miRNAs live in the miRNP? *Genes & development*, 16(9):1025–1031, 2002.

[228] Ashley J. Pratt and Ian J. MacRae. The RNA-induced silencing complex: a versatile gene-silencing machine. *The Journal of biological chemistry*, 284(27):17897–17901, 2009.

[229] Lee P. Lim, Nelson C. Lau, Philip Garrett-Engele, Andrew Grimson, Janell M. Schelter, John Castle, David P. Bartel, Peter S. Linsley, and Jason M. Johnson. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027):769–773, 2005.

[230] Benjamin P. Lewis, I-hung Shih, Matthew W. Jones-Rhoades, David P. Bartel, and Christopher B. Burge. Prediction of mammalian microRNA targets. *Cell*, 115(7):787–798, 2003.

[231] Robert F. Place, Long-Cheng Li, Deepa Pookot, Emily J. Noonan, and Rajvir Dahiya. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5):1608–1613, 2008.

[232] David P. Bartel. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, 2004.

[233] Eric P. Nawrocki, Sarah W. Burge, Alex Bateman, Jennifer Daub, Ruth Y. Eberhardt, Sean R. Eddy, Evan W. Floden, Paul P. Gardner, Thomas A. Jones, John Tate, and Robert D. Finn. Rfam 12.0: updates to the RNA families database. *Nucleic acids research*, 43(Database issue):D130–7, 2015.

[234] Jiyuan An, John Lai, Melanie L. Lehman, and Colleen C. Nelson. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic acids research*, 41(2):727–737, 2013.

[235] Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics (Oxford, England)*, 29(22):2933–2935, 2013.

[236] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. BLAST+: architecture and applications. *BMC bioinformatics*, 10:421, 2009.

[237] Sonika Tyagi, Candida Vaz, Vipin Gupta, Rohit Bhatia, Sachin Maheshwari, Ashwin Srinivasan, and Alok Bhattacharya. CID-miRNA: a web server for prediction of novel miRNA precursors in human genome. *Biochemical and biophysical research communications*, 372(4):831–834, 2008.

[238] David Langenberger, Clara Bermudez-Santana, Jana Hertel, Steve Hoffmann, Philipp Khaitovich, and Peter F. Stadler. Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics (Oxford, England)*, 25(18):2298–2301, 2009.

[239] D. Langenberger, C. I. Bermudez-Santana, P. F. Stadler, and S. Hoffmann. Identification and classification of small RNAs in transcriptome sequence data. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 80–87, 2010.

[240] Chol-Hee Jung, Martin A. Hansen, Igor V. Makunin, Darren J. Korbie, and John S. Mattick. Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. *BMC genomics*, 11:77, 2010.

[241] W. James Kent. BLAT–the BLAST-like alignment tool. *Genome research*, 12(4):656–664, 2002.

[242] Ana Kozomara and Sam Griffiths-Jones. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research*, 39(Database issue):D152–7, 2011.

[243] Ingo Braasch, Andrew R. Gehrke, Jeramiah J. Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M. Berlin, Michael S. Campbell, Daniel Barrell, Kyle J. Martin, John F. Mulley, Vydianathan Ravi, Alison P. Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Wcisel, Cristian Cañestro, Jason Sydes, Felix E. G. Beaudry, Yi Sun, Jana Hertel, Michael J. Beam, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H. Letaw, Gary W. Litman, Ronda T. Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F. Stadler, Han Wang, John S. Taylor, Quenton Fontenot, Allyse Ferrara, Stephen M. J. Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A. Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T. Amemiya, Byrappa Venkatesh, Peter W. H. Holland, Yann Guiguen, Julien Bobe, Neil H. Shubin, Federica Di Palma, Jessica Alföldi, Kerstin Lindblad-Toh, and John H. Postlethwait. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature genetics*, 48(4):427–437, 2016.

[244] Jana Hertel and Peter F. Stadler. The Expansion of Animal MicroRNA Families Revisited. *Life (Basel, Switzerland)*, 5(1):905–920, 2015.

[245] Kazutaka Katoh and Daron M. Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780, 2013.

[246] Benjamin P. Lewis, Christopher B. Burge, and David P. Bartel. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120(1):15–20, 2005.

[247] Alessandro Cellerino, Dario R. Valenzano, and Martin Reichard. From the bush to the bench: the annual Nothobranchius fishes as a new model system in biology. *Biological reviews of the Cambridge Philosophical Society*, 91(2):511–533, 2016.

[248] Matthias Platzer and Christoph Englert. Nothobranchius furzeri: A Model for Aging Research and More. *Trends in genetics : TIG*, 32(9):543–552, 2016.

[249] Emiliano Di Cicco, Eva Terzibasi Tozzini, Giacomo Rossi, and Alessandro Cellerino. The short-lived annual fish Nothobranchius furzeri shows a typical teleost aging process reinforced by high incidence of age-dependent neoplasias. *Experimental gerontology*, 46(4):249–256, 2011.

[250] Antonio Marco, Maria Ninova, Matthew Ronshaugen, and Sam Griffiths-Jones. Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic acids research*, 41(16):7745–7752, 2013.

[251] Elizabeth J. Thatcher, Jordan Bond, Ima Paydar, and James G. Patton. Genomic organization of zebrafish microRNAs. *BMC genomics*, 9:253, 2008.

[252] Yirong Wang, Junjie Luo, Hong Zhang, and Jian Lu. microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. *Molecular biology and evolution*, 33(9):2232–2247, 2016.

[253] Antonio J. Giraldez, Yuichiro Mishima, Jason Rihel, Russell J. Grocock, Stijn van Dongen, Kunio Inoue, Anton J. Enright, and Alexander F. Schier. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science (New York, N.Y.)*, 312(5770):75–79, 2006.

[254] Jana Hertel, Manuela Lindemeyer, Kristin Missal, Claudia Fried, Andrea Tanzer, Christoph Flamm, Ivo L. Hofacker, and Peter F. Stadler. The expansion of the metazoan microRNA repertoire. *BMC genomics*, 7:25, 2006.

[255] Matteo Osella, Carla Bosia, Davide Corá, and Michele Caselle. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS computational biology*, 7(3):e1001101, 2011.

[256] Velia Siciliano, Immacolata Garzilli, Chiara Fracassi, Stefania Criscuolo, Simona Ventre, and Diego Di Bernardo. MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nature communications*, 4:2364, 2013.

[257] Eva Terzibasi Tozzini, Aurora Savino, Roberto Ripa, Giorgia Battistoni, Mario Baumgart, and Alessandro Cellerino. Regulation of microRNA expression in the neuronal stem cell niches during aging of the short-lived annual fish Nothobranchius furzeri. *Frontiers in cellular neuroscience*, 8:51, 2014.

[258] Dario Riccardo Valenzano, Bérénice A. Benayoun, Param Priya Singh, Elisa Zhang, Paul D. Etter, Chi-Kuo Hu, Mathieu Clément-Ziza, David Willemsen, Rongfeng Cui, Itamar Harel, Ben E. Machado, Muh-Ching Yee, Sabrina C. Sharp, Carlos D. Bustamante, Andreas Beyer, Eric A. Johnson, and Anne Brunet. The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan. *Cell*, 163(6):1539–1554, 2015.

[259] Alexander Serganov, Anna Polonskaia, Anh Tuân Phan, Ronald R. Breaker, and Dinshaw J. Patel. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*, 441(7097):1167–1171, 2006.

[260] Narasimhan Sudarsan, Jeffrey E. Barrick, and Ronald R. Breaker. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA (New York, N.Y.)*, 9(6):644–647, 2003.

[261] N. Aziz and H. N. Munro. Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region. *Proceedings of the National Academy of Sciences of the United States of America*, 84(23):8478–8482, 1987.

[262] S. Kubota, T. Hattori, T. Nakanishi, and M. Takigawa. Involvement of cis-acting repressive element(s) in the 3'-untranslated region of human connective tissue growth factor gene. *FEBS letters*, 450(1-2):84–88, 1999.

[263] Elzbieta Sarnowska, Ewa A. Grzybowska, Krzysztof Sobczak, Ryszard Konopinski, Anna Wilczynska, Maria Szwarc, Tomasz J. Sarnowski, Wlodzimierz J. Krzyzosiak, and Janusz A. Siedlecki. Hairpin structure within the 3'UTR of DNA polymerase beta mRNA acts as a post-transcriptional regulatory element and interacts with Hax-1. *Nucleic acids research*, 35(16):5499–5510, 2007.

[264] G. Bermano, R. K. Shepherd, Z. E. Zehner, and J. E. Hesketh. Perinuclear mRNA localisation by vimentin 3'-untranslated region requires a 100 nucleotide sequence and intermediate filaments. *FEBS letters*, 497(2-3):77–81, 2001.

[265] Ivaylo P. Ivanov, Christine B. Anderson, Raymond F. Gesteland, and John F. Atkins. Identification of a new antizyme mRNA +1 frameshifting stimulatory pseudoknot in a subset of diverse invertebrates and its apparent absence in intermediate species. *Journal of molecular biology*, 339(3):495–504, 2004.

[266] L. Varani, S. I. Gunderson, I. W. Mattaj, L. E. Kay, D. Neuhaus, and G. Varani. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nature structural biology*, 7(4):329–335, 2000.

[267] Tarun Bhalla, Joshua J. C. Rosenthal, Miguel Holmgren, and Robert Reenan. Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nature structural & molecular biology*, 11(10):950–956, 2004.

[268] Johan Ohlson, Jakob Skou Pedersen, David Haussler, and Marie Ohman. Editing modifies the GABA(A) receptor subunit alpha3. *RNA (New York, N.Y.)*, 13(5):698–703, 2007.

[269] Eva Terzibasi Tozzini, Mario Baumgart, Giorgia Battistoni, and Alessandro Cellerino. Adult neurogenesis in the short-lived teleost Nothobranchius furzeri: localization of neurogenic niches, molecular characterization and effects of aging. *Aging cell*, 11(2):241–251, 2012.

[270] Natalia Pinzón, Blaise Li, Laura Martinez, Anna Sergeeva, Jessy Presumey, Florence Apparailly, and Hervé Seitz. microRNA target prediction programs predict many false positives. *Genome research*, 27(2):234–245, 2017.

[271] Mario Baumgart, Marco Groth, Steffen Priebe, Aurora Savino, Giovanna Testa, Andreas Dix, Roberto Ripa, Francesco Spallotta, Carlo Gaetano, Michela Ori, Eva Terzibasi Tozzini, Reinhard Guthke, Matthias Platzer, and Alessandro Cellerino. RNA-seq of the aging brain in the short-lived fish N. furzeri - conserved pathways and novel genes associated with neurogenesis. *Aging cell*, 13(6):965–974, 2014.

[272] J. M. Gudas, M. Payton, S. Thukral, E. Chen, M. Bass, M. O. Robinson, and S. Coats. Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly expressed in human cancers. *Molecular and cellular biology*, 19(1):612–622, 1999.

[273] Min-Young Lee, Hyun-Ju Kim, Myoung-Ae Kim, Hye Jin Jee, Ae Jeong Kim, Yoe-Sik Bae, Joo-In Park, Jay H. Chung, and Jeanho Yun. Nek6 is involved in G2/M phase cell cycle arrest through DNA damage-induced phosphorylation. *Cell cycle (Georgetown, Tex.)*, 7(17):2705–2709, 2008.

[274] F. Marqués, J. L. Moreau, G. Peaucellier, J. C. Lozano, P. Schatt, A. Picard, I. Callebaut, E. Perret, and A. M. Genevière. A new subfamily of high molecular mass CDC2-related kinases with PITAI/VRE motifs. *Biochemical and biophysical research communications*, 279(3):832–837, 2000.

[275] J. M. Cheng, J. L. Hiemstra, S. S. Schneider, A. Naumova, N. K. Cheung, S. L. Cohn, L. Diller, C. Sapienza, and G. M. Brodeur. Preferential amplification of the paternal allele of the N-myc gene in human neuroblastomas. *Nature genetics*, 4(2):191–194, 1993.

[276] YunQing Jiang, Indira Prabakaran, Fei Wan, Nandita Mitra, Dana K. Furstenau, Rupert K. Hung, Siyuan Cao, Paul J. Zhang, Douglas L. Fraker, and Marina A. Guvakova. Vav2 protein overexpression marks and may predict the aggressive subtype of ductal carcinoma in situ. *Biomarker research*, 2:22, 2014.

[277] Alexander Dorn, Zuzana Musilová, Matthias Platzer, Kathrin Reichwald, and Alessandro Cellerino. The strange case of East African annual fishes: aridification correlates with diversification for a savannah aquatic group? *BMC evolutionary biology*, 14:210, 2014.

[278] Ryota Suzuki and Hidetoshi Shimodaira. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics (Oxford, England)*, 22(12):1540–1542, 2006.

[279] Lin He, J. Michael Thomson, Michael T. Hemann, Eva Hernando-Monge, David Mu, Summer Goodson, Scott Powers, Carlos Cordon-Cardo, Scott W. Lowe, Gregory J. Hannon, and Scott M. Hammond. A microRNA polycistron as a potential human oncogene. *Nature*, 435(7043):828–833, 2005.

[280] Andrea Tanzer and Peter F. Stadler. Molecular evolution of a microRNA cluster. *Journal of molecular biology*, 339(2):327–335, 2004.

[281] Ashley M. Fenn, Kristen M. Smith, Amy E. Lovett-Racke, Mireia Guerau-de Arellano, Caroline C. Whitacre, and Jonathan P. Godbout. Increased micro-RNA 29b in the aged brain correlates with the reduction of insulin-like growth factor-1 and fractalkine ligand. *Neurobiology of aging*, 34(12):2748–2758, 2013.

[282] Katie Nolan, Mollie R. Mitchem, Eva M. Jimenez-Mateos, David C. Henshall, Caoimhín G. Concannon, and Jochen H. M. Prehn. Increased expression of microRNA-29a in ALS mice: functional analysis of its inhibition. *Journal of molecular neuroscience : MN*, 53(2):231–241, 2014.

[283] Masaki Takahashi, Akiko Eda, Tatsunobu Fukushima, and Hirohiko Hohjoh. Reduction of type IV collagen by upregulated miR-29 in normal elderly mouse and klotho-deficient, senescence-model mouse. *PloS one*, 7(11):e48974, 2012.

[284] Roberto Ripa, Luca Dolfi, Marco Terrigno, Luca Pandolfini, Aurora Savino, Valeria Arcucci, Marco Groth, Eva Terzibasi Tozzini, Mario Baumgart, and Alessandro Cellerino. MicroRNA miR-29 controls a compensatory response to limit neuronal iron accumulation during adult life and aging. *BMC biology*, 15(1):9, 2017.

[285] Adrian Liston, Aikaterini S. Papadopoulou, Dina Danso-Abeam, and James Dooley. MicroRNA-29 in the adaptive immune system: setting the threshold. *Cellular and molecular life sciences : CMLS*, 69(21):3533–3541, 2012.

[286] W. J. Murphy and G. E. Collier. A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. *Molecular biology and evolution*, 14(8):790–799, 1997.

[287] Tomas Hrbek and Allan Larson. The Evolution of Diapause in the Killifish Family Rivulidae (Atherinomorpha, Cyprinodontiformes): A Molecular Phylogenetic and Biogeographic Perspective. *Evolution; international journal of organic evolution*, 53(4):1200–1216, 1999.

[288] Andrew I. Furness, Kevin Lee, and David N. Reznick. Adaptation in a variable environment: Phenotypic plasticity and bet-hedging during egg diapause and hatching in an annual killifish. *Evolution; international journal of organic evolution*, 69(6):1461–1475, 2015.

[289] J. P. Wourms. The developmental biology of annual fishes. 3. Pre-embryonic and embryonic diapause of variable duration in the eggs of annual fishes. *The Journal of experimental zoology*, 182(3):389–414, 1972.

[290] Luca Dolfi, Roberto Ripa, and Alessandro Cellerino. Transition to annual life history coincides with reduction in cell cycle speed during early cleavage in three independent clades of annual killifish. *EvoDevo*, 5:32, 2014.

[291] Radim Blažek, Matej Polačik, and Martin Reichard. Rapid growth, early maturation and short generation time in African annual fishes. *EvoDevo*, 4(1):24, 2013.

[292] Jason E. Podrabsky and George N. Somero. An inducible 70 kDa-class heat shock protein is constitutively expressed during early development and diapause in the annual killifish Austrofundulus limnaeus. *Cell Stress & Chaperones*, 12(3):199–204, 2007.

[293] J. E. Podrabsky, J. F. Carpenter, and S. C. Hand. Survival of water stress in annual fish embryos: dehydration avoidance and egg envelope amyloid fibers. *American journal of physiology. Regulatory, integrative and comparative physiology*, 280(1):R123–31, 2001.

[294] Podrabsky and Hand. The bioenergetics of embryonic diapause in an annual killifish, austrofundulus limnaeus. *The Journal of experimental biology*, 202 (Pt 19):2567–2580, 1999.

[295] Jeffrey M. Duerr and Jason E. Podrabsky. Mitochondrial physiology of diapausing and developing embryos of the annual killifish Austrofundulus limnaeus: implications for extreme anoxia tolerance. *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology*, 180(7):991–1003, 2010.

[296] Camie L. Meller, Robert Meller, Roger P. Simon, Kristin M. Culpepper, and Jason E. Podrabsky. Cell cycle arrest associated with anoxia-induced quiescence, anoxic preconditioning, and embryonic diapause in embryos of the annual killifish Austrofundulus limnaeus. *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology*, 182(7):909–920, 2012.

[297] P. J. Levels, R. E. Gubbels, and J. M. Denucé. Oxygen consumption during embryonic development of the annual fish Nothobranchius korthausae with special reference to diapause. *Comparative biochemistry and physiology. A, Comparative physiology*, 84(4):767–770, 1986.

[298] Luigi Fontana, Samuel Klein, and John O. Holloszy. Effects of long-term calorie restriction and endurance exercise on glucose tolerance, insulin action, and adipokine production. *Age (Dordrecht, Netherlands)*, 32(1):97–108, 2010.

[299] Cynthia Kenyon. The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1561):9–16, 2011.

[300] Joshua J. McElwee, Eugene Schuster, Eric Blanc, James H. Thomas, and David Gems. Shared transcriptional signature in Caenorhabditis elegans Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. *Journal of Biological Chemistry*, 279(43):44533–44543, 2004.

[301] Zachary Pincus, Thalyana Smith-Vikos, and Frank J. Slack. MicroRNA predictors of longevity in Caenorhabditis elegans. *PLoS genetics*, 7(9):e1002306, 2011.

[302] Konstantinos Boulias and H. Robert Horvitz. The C. elegans microRNA mir-71 acts in neurons to promote germline-mediated longevity through regulation of DAF-16/FOXO. *Cell metabolism*, 15(4):439–450, 2012.

[303] Itamar Harel, Bérénice A. Benayoun, Ben Machado, Param Priya Singh, Chi-Kuo Hu, Matthew F. Pech, Dario Riccardo Valenzano, Elisa Zhang, Sabrina C. Sharp, Steven E. Artandi, and Anne Brunet. A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell*, 160(5):1013–1026, 2015.

[304] Evgeny A. Glazov, Pauline A. Cottee, Wesley C. Barris, Robert J. Moore, Brian P. Dalrymple, and Mark L. Tizard. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome research*, 18(6):957–964, 2008.

[305] Merav Bar, Stacia K. Wyman, Brian R. Fritz, Junlin Qi, Kavita S. Garg, Rachael K. Parkin, Evan M. Kroh, Ausra Bendoraite, Patrick S. Mitchell, Angelique M. Nelson, Walter L. Ruzzo, Carol Ware, Jerald P. Radich, Robert Gentleman, Hannele Ruohola-Baker, and Muneesh Tewari. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem cells (Dayton, Ohio)*, 26(10):2496–2505, 2008.

[306] Amie L. T. Romney and Jason E. Podrabsky. Small noncoding RNA profiles along alternative developmental trajectories in an annual killifish. *Scientific reports*, 8(1):13364, 2018.

[307] Jason E. Podrabsky, Ian D. F. Garrett, and Zachary F. Kohl. Alternative developmental pathways associated with diapause regulated by temperature and maternal influences in embryos of the annual killifish Austrofundulus limnaeus. *The Journal of experimental biology*, 213(Pt 19):3280–3288, 2010.

[308] Joost M. Woltering and Antony J. Durston. MiR-10 represses HoxB1a and HoxB3a in zebrafish. *PloS one*, 3(1):e1396, 2008.

[309] Ulf Andersson Ørom, Finn Cilius Nielsen, and Anders H. Lund. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Molecular cell*, 30(4):460–471, 2008.

[310] Alessandro Rosa, Francesca M. Spagnoli, and Ali H. Brivanlou. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Developmental cell*, 16(4):517–527, 2009.

[311] Saori Tani, Rie Kusakabe, Kiyoshi Naruse, Hiroshi Sakamoto, and Kunio Inoue. Genomic organization and embryonic expression of miR-430 in medaka (Oryzias latipes): insights into the post-transcriptional gene regulation in early development. *Gene*, 449(1-2):41–49, 2010.

[312] Yuichiro Mishima, Antonio J. Giraldez, Yasuaki Takeda, Toshinobu Fujiwara, Hiroshi Sakamoto, Alexander F. Schier, and Kunio Inoue. Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Current biology : CB*, 16(21):2135–2142, 2006.

[313] Antonio J. Giraldez, Ryan M. Cinalli, Margaret E. Glasner, Anton J. Enright, J. Michael Thomson, Scott Baskerville, Scott M. Hammond, David P. Bartel, and Alexander F. Schier. MicroRNAs regulate brain morphogenesis in zebrafish. *Science (New York, N.Y.)*, 308(5723):833–838, 2005.

[314] Yuri Pekarsky, Urmila Santanam, Amelia Cimmino, Alexey Palamarchuk, Alexey Efanov, Vadim Maximov, Stefano Volinia, Hansjuerg Alder, Chang-Gong Liu, Laura Rassenti, George A. Calin, John P. Hagan, Thomas Kipps, and Carlo M. Croce. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. *Cancer research*, 66(24):11590–11593, 2006.

[315] Benjamin Gibert, Céline Delloye-Bourgeois, Charles-Henry Gattolliat, Olivier Meurette, Solen Le Guernevel, Joanna Fombonne, Benjamin Ducarouge, Fabrice Lavial, Frantz Bouhallier, Marion Creveaux, Ana Maria Negulescu, Jean Bénard, Isabelle Janoueix-Lerosey, Annick Harel-Bellan, Olivier Delattre, and Patrick Mehlen. Regulation by miR181 family of the dependence receptor CDON tumor suppressive activity in neuroblastoma. *Journal of the National Cancer Institute*, 106(11), 2014.

[316] Irina Naguibneva, Maya Ameyar-Zazoua, Anna Polesskaya, Slimane Ait-Si-Ali, Reguina Groisman, Mouloud Souidi, Sylvain Cuvellier, and Annick Harel-Bellan. The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nature cell biology*, 8(3):278–284, 2006.

[317] Jan Kazenwadel, Michael Z. Michael, and Natasha L. Harvey. Prox1 expression is negatively regulated by miR-181 in endothelial cells. *Blood*, 116(13):2395–2401, 2010.

[318] Tamar Uziel, Fedor V. Karginov, Suqing Xie, Joel S. Parker, Yong-Dong Wang, Amar Gajjar, Lin He, David Ellison, Richard J. Gilbertson, Gregory Hannon, and Martine F. Roussel. The miR-17˜92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8):2812–2817, 2009.

[319] Ksenia A. Orlova and Peter B. Crino. The tuberous sclerosis complex. *Annals of the New York Academy of Sciences*, 1184:87–105, 2010.

[320] Thomas Weichhart. mTOR as Regulator of Lifespan, Aging, and Cellular Senescence: A Mini-Review. *Gerontology*, 64(2):127–134, 2018.

[321] Zeran Wang, Yueyong Liu, Pengju Zhang, Weiguo Zhang, Weijing Wang, Kenneth Curr, Guangwei Wei, and Jian-Hua Mao. FAM83D promotes cell proliferation and motility by downregulating tumor suppressor gene FBXW7. *Oncotarget*, 4(12):2476–2486, 2013.

[322] Yick-Pang Ching, Chun-Ming Wong, Shing-Fai Chan, Thomas Ho-Yin Leung, David Chi-Heng Ng, Dong-Yan Jin, and Irene Oi-lin Ng. Deleted in liver cancer (DLC) 2 encodes a RhoGAP protein with growth suppressor function and is underexpressed in hepatocellular carcinoma. *Journal of Biological Chemistry*, 278(12):10824–10830, 2003.

[323] S. Dalton, R. Marais, J. Wynne, and R. Treisman. Isolation and characterization of SRF accessory proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 340(1293):325–332, 1993.

[324] Karine Tréguer, Eva-Marie Heinrich, Kisho Ohtani, Angelika Bonauer, and Stefanie Dimmeler. Role of the microRNA-17-92 cluster in the endothelial differentiation of stem cells. *Journal of vascular research*, 49(5):447–460, 2012.

[325] Philip A. Gregory, Andrew G. Bert, Emily L. Paterson, Simon C. Barry, Anna Tsykin, Gelareh Farshid, Mathew A. Vadas, Yeesim Khew-Goodall, and Gregory J. Goodall. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature cell biology*, 10(5):593–601, 2008.

[326] Carmen Urbich, Angelika Kuehbacher, and Stefanie Dimmeler. Role of microRNAs in vascular diseases, inflammation, and angiogenesis. *Cardiovascular research*, 79(4):581–588, 2008.

[327] Stefania Nicoli, Carl-Philipp Knyphausen, Lihua J. Zhu, Abirami Lakshmanan, and Nathan D. Lawson. miR-221 is required for endothelial tip cell behaviors during vascular development. *Developmental cell*, 22(2):418–429, 2012.

[328] M. Shioya, S. Obayashi, H. Tabunoki, K. Arima, Y. Saito, T. Ishida, and J. Satoh. Aberrant microRNA expression in the brains of neurodegenerative diseases: miR-29a decreased in Alzheimer disease brains targets neurone navigator 3. *Neuropathology and applied neurobiology*, 36(4):320–330, 2010.

[329] Jung Ki Yoo, Ho Yong Jung, Chang-Hyun Kim, Woo Sung Son, and Jin Kyeoung Kim. miR-7641 modulates the expression of CXCL1 during endothelial differentiation derived from human embryonic stem cells. *Archives of pharmacal research*, 36(3):353–358, 2013.

[330] H. Zhang, C. Xu, Y. Liu, and W. Yuan. MicroRNA-563 promotes the osteogenic differentiation of posterior longitudinal ligament cells by inhibiting SMURF1. *Zhonghua wai ke za zhi [Chinese journal of surgery]*, 55(3):203–207, 2017.

[331] Shijun Hu, Mei Huang, Zongjin Li, Fangjun Jia, Zhumur Ghosh, Maarten A. Lijkwan, Pasquale Fasanaro, Ning Sun, Xi Wang, Fabio Martelli, Robert C. Robbins, and Joseph C. Wu. MicroRNA-210 as a novel therapy for treatment of ischemic heart disease. *Circulation*, 122(11 Suppl):S124–31, 2010.

[332] Qi-Jing Li, Jacqueline Chau, Peter J. R. Ebert, Giselle Sylvester, Hyeyoung Min, Gwen Liu, Ravi Braich, Muthiah Manoharan, Juergen Soutschek, Petra Skare, Lawrence O. Klein, Mark M. Davis, and Chang-Zheng Chen. miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell*, 129(1):147–161, 2007.

[333] Manoj Bhaskaran, Yang Wang, Honghao Zhang, Tingting Weng, Pradyumna Baviskar, Yujie Guo, Deming Gou, and Lin Liu. MicroRNA-127 modulates fetal lung development. *Physiological genomics*, 37(3):268–278, 2009.

[334] Nadia Felli, Francesca Pedini, Paolo Romania, Mauro Biffoni, Ornella Morsilli, Germana Castelli, Simona Santoro, Simona Chicarella, Antonio Sorrentino, Cesare Peschle, and Giovanna Marziali. MicroRNA 223-dependent expression of LMO2 regulates normal erythropoiesis. *Haematologica*, 94(4):479–486, 2009.

[335] T. Sugatani and K. A. Hruska. MicroRNA-223 is a key factor in osteoclast differentiation. *Journal of cellular biochemistry*, 101(4):996–999, 2007.

[336] Rong-Chi Huang. The discoveries of molecular mechanisms for the circadian rhythm: The 2017 Nobel Prize in Physiology or Medicine. *Biomedical journal*, 41(1):5–8, 2018.

[337] Yihai Lin and Zhangyi Wu. MicroRNA-128 inhibits proliferation and invasion of glioma cells by targeting COX-2. *Gene*, 658:63–69, 2018.

[338] Herlinda Guzman, Katie Sanders, Adam Idica, Aurore Bochnakian, Douglas Jury, Iben Daugaard, Dimitrios G. Zisoulis, and Irene Munk Pedersen. miR-128 inhibits telomerase activity by targeting TERT mRNA. *Oncotarget*, 9(17):13244–13253, 2018.

[339] Giulia Escobar, Davide Moi, Anna Ranghetti, Pinar Ozkal-Baydin, Mario Leonardo Squadrito, Anna Kajaste-Rudnitski, Attilio Bondanza, Bernhard Gentner, Michele de Palma, Roberta Mazzieri, and Luigi Naldini. Genetic engineering of hematopoiesis for targeted IFN-a delivery inhibits breast cancer progression. *Science translational medicine*, 6(217):217ra3, 2014.

[340] V. Baldin, J. Lukas, M. J. Marcote, M. Pagano, and G. Draetta. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes & development*, 7(5):812–821, 1993.

[341] Xin Huang, Quynh-Thu Le, and Amato J. Giaccia. MiR-210–micromanager of the hypoxia pathway. *Trends in molecular medicine*, 16(5):230–237, 2010.

[342] Endashaw Jebessa, Hongjia Ouyang, Bahareldin Ali Abdalla, Zhenhui Li, Auwalu Yusuf Abdullahi, Qingshen Liu, Qinghua Nie, and Xiquan Zhang. Characterization of miRNA and their target gene during chicken embryo skeletal muscle development. *Oncotarget*, 9(25):17309–17324, 2018.

[343] Le Zhang, Philip J. Ebenezer, Kalavathi Dasuri, Sun Ok Fernandez-Kim, Joseph Francis, Nithya Mariappan, Zhanguo Gao, Jianping Ye, Annadora J. Bruce-Keller, and Jeffrey N. Keller. Aging is associated with hypoxia and oxidative stress in adipose tissue: implications for adipose function. *American journal of physiology. Endocrinology and metabolism*, 301(4):E599–607, 2011.

[344] Ariel A. Bazzini, Miler T. Lee, and Antonio J. Giraldez. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science (New York, N.Y.)*, 336(6078):233–237, 2012.

[345] Stephen W. Eichhorn, Huili Guo, Sean E. McGeary, Ricard A. Rodriguez-Mias, Chanseok Shin, Daehyun Baek, Shu-Hao Hsu, Kalpana Ghoshal, Judit Villén, and David P. Bartel. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular cell*, 56(1):104–115, 2014.

[346] J. E. Podrabsky and S. C. Hand. Depression of protein synthesis during diapause in embryos of the annual killifish Austrofundulus limnaeus. *Physiological and biochemical zoology : PBZ*, 73(6):799–808, 2000.

[347] Xiaochang Zhang, Rebecca Zabinsky, Yudong Teng, Mingxue Cui, and Min Han. microRNAs play critical roles in the survival and recovery of Caenorhabditis elegans from starvation-induced L1 diapause. *Proceedings of the National Academy of Sciences of the United States of America*, 108(44):17997–18002, 2011.

[348] Alexandre de Lencastre, Zachary Pincus, Katherine Zhou, Masaomi Kato, Siu Sylvia Lee, and Frank J. Slack. MicroRNAs both promote and antagonize longevity in C. elegans. *Current biology : CB*, 20(24):2159–2168, 2010.

[349] Xiao-Pu He, Yun Shao, Xiao-Lin Li, Wei Xu, Guo-Sheng Chen, Huan-Huan Sun, Hai-Chen Xu, Xian Xu, Dan Tang, Xi-Feng Zheng, Yi-Ping Xue, Guo-Chang Huang, and Wei-Hao Sun. Downregulation of miR-101 in gastric cancer correlates with cyclooxygenase-2 overexpression and tumor growth. *The FEBS journal*, 279(22):4201–4212, 2012.

[350] Virginie Olive, Iris Jiang, and Lin He. mir-17-92, a cluster of miRNAs in the midst of the cancer network. *The international journal of biochemistry & cell biology*, 42(8):1348–1354, 2010.

[351] Matthias Hackl, Stefan Brunner, Klaus Fortschegger, Carina Schreiner, Lucia Micutkova, Christoph Mück, Gerhard T. Laschober, Günter Lepperdinger, Natalie Sampson, Peter Berger, Dietmar Herndler-Brandstetter, Matthias Wieser, Harald Kühnel, Alois Strasser, Mark Rinnerthaler, Michael Breitenbach, Michael Mildner, Leopold Eckhart, Erwin Tschachler, Andrea Trost, Johann W. Bauer, Christine Papak, Zlatko Trajanoski, Marcel Scheideler, Regina Grillari-Voglauer, Beatrix Grubeck-Loebenstein, Pidder Jansen-Dürr, and Johannes Grillari. miR-17, miR-19b, miR-20a, and miR-106a are down-regulated in human aging. *Aging cell*, 9(2):291–296, 2010.

[352] Ali Mortazavi, Brian A. Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7):621–628, 2008.

[353] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[354] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[355] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–140, 2010.

[356] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, 2010.

[357] Li-Xuan Qin, Thomas Tuschl, and Samuel Singer. Empirical insights into the stochasticity of small RNA sequencing. *Scientific reports*, 6:24061, 2016.

[358] Mark D. Robinson and Gordon K. Smyth. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics (Oxford, England)*, 23(21):2881–2887, 2007.

[359] Mark D. Robinson and Gordon K. Smyth. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics (Oxford, England)*, 9(2):321–332, 2008.

[360] Harold Pimentel, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature methods*, 14(7):687–690, 2017.

[361] G. Stolovitzky and G. Cecchi. Efficiency of DNA replication in the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23):12947–12952, 1996.

[362] L. J. Gleser. The gamma distribution as a mixture of exponential distributions. *The American Statistician*, 43(2):115–117, 1989.

[363] Lauren M. McIntyre, Kenneth K. Lopiano, Alison M. Morse, Victor Amin, Ann L. Oberg, Linda J. Young, and Sergey V. Nuzhdin. RNA-seq: technical variability and sampling. *BMC genomics*, 12:293, 2011.

[364] F. J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[365] S. C. Choi and R. Wette. Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics*, 11(4):683–690, 1969.

[366] C. Leys, C. Ley, O. Klein, P. Bernard, and Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

[367] G. D. Ruxton. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4):688–690, 2006.

[368] Fatemeh Seyednasrollah, Asta Laiho, and Laura L. Elo. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in bioinformatics*, 16(1):59–70, 2015.

[369] Thalyana Smith-Vikos and Frank J. Slack. MicroRNAs and their roles in aging. *Journal of cell science*, 125(Pt 1):7–17, 2012.

[370] Fabiola Olivieri, Maria Rita Rippo, Vladia Monsurrò, Stefano Salvioli, Miriam Capri, Antonio Domenico Procopio, and Claudio Franceschi. MicroRNAs linking inflamm-aging, cellular senescence and cancer. *Ageing research reviews*, 12(4):1056–1068, 2013.

[371] Michelle Boehm and Frank Slack. A developmental timing microRNA and its target regulate life span in C. elegans. *Science (New York, N.Y.)*, 310(5756):1954–1957, 2005.

[372] P. H. Reddy, J. Williams, F. Smith, J. S. Bhatti, S. Kumar, M. Vijayan, R. Kandimalla, C. S. Kuruva, R. Wang, M. Manczak, X. Yin, and A. P. Reddy. MicroRNAs, Aging, Cellular Senescence, and Alzheimer's Disease. *Progress in molecular biology and translational science*, 146:127–171, 2017.

[373] Massimiliano Bonafè and Fabiola Olivieri. Circulating microRNAs in aging. *Oncotarget*, 6(3):1340–1341, 2015.

[374] Johannes Grillari, Matthias Hackl, and Regina Grillari-Voglauer. miR-17-92 cluster: ups and downs in cancer and aging. *Biogerontology*, 11(4):501–506, 2010.

[375] Yukio Kawahara, Molly Megraw, Edward Kreider, Hisashi Iizasa, Louis Valente, Artemis G. Hatzigeorgiou, and Kazuko Nishikura. Frequency and fate of microRNA editing in human brain. *Nucleic acids research*, 36(16):5270–5280, 2008.

[376] Julia Winter, Stephanie Jung, Sarina Keller, Richard I. Gregory, and Sven Diederichs. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology*, 11(3):228–234, 2009.

[377] M. Ohman. A-to-I editing challenger or ally to the microRNA process. *Biochimie*, 89(10):1171–1176, 2007.

[378] Flavia Pichiorri, Sung-Suk Suh, Alberto Rocci, Luciana de Luca, Cristian Taccioli, Ramasamy Santhanam, Wenchao Zhou, Don M. Benson, Craig Hofmainster, Hansjuerg Alder, Michela Garofalo, Gianpiero Di Leva, Stefano Volinia, Huey-Jen Lin, Danilo Perrotti, Michael Kuehl, Rami I. Aqeilan, Antonio Palumbo, and Carlo M. Croce. Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development. *Cancer cell*, 18(4):367–381, 2010.

[379] Christian J. Braun, Xin Zhang, Irina Savelyeva, Sonja Wolff, Ute M. Moll, Troels Schepeler, Torben F. Ørntoft, Claus L. Andersen, and Matthias Dobbelstein. p53-Responsive micrornas 192 and 215 are capable of inducing cell cycle arrest. *Cancer research*, 68(24):10094–10104, 2008.

[380] H. W. Z. Khella, M. Bakhet, G. Allo, M. A. S. Jewett, A. H. Girgis, A. Latif, H. Girgis, I. von Both, G. A. Bjarnason, and G. M. Yousef. miR-192, miR-194 and miR-215: a convergent microRNA network suppressing tumor progression in renal cell carcinoma. *Carcinogenesis*, 34(10):2231–2239, 2013.

[381] Yuanzhuo Gu, Xiyang Wei, Yulin Sun, Hongjun Gao, Xin Zheng, Linda L. Wong, Ling Jin, Niya Liu, Brenda Y. Hernandez, Karolina Peplowska, Xiaohang Zhao, Qi-Min Zhan, Xin-Hua Feng, Zhao-You Tang, and Junfang Ji. miR-192-5p silencing by genetic aberrations is a key event in hepatocellular carcinomas with cancer stem cell features. *Cancer research*, 2018.

[382] Volker Ast, Theresa Kordaß, Marcus Oswald, Amol Kolte, David Eisel, Wolfram Osen, Stefan B. Eichmüller, Alexander Berndt, and Rainer König. MiR-192, miR-200c and miR-17 are fibroblast-mediated inhibitors of colorectal cancer invasion. *Oncotarget*, 9(85):35559–35580, 2018.

[383] Paola Fuschi, Matteo Carrara, Christine Voellenkle, Jose Manuel Garcia-Manteiga, Paolo Righini, Biagina Maimone, Elena Sangalli, Francesco Villa, Claudia Specchia, Mario Picozza, Giovanni Nano, Carlo Gaetano, Gaia Spinetti, Annibale A. Puca, Alessandra Magenta, and Fabio Martelli. Central role of the p53 pathway in the noncoding-RNA response to oxidative stress. *Aging*, 9(12):2559–2586, 2017.

[384] Johan Skog, Tom Würdinger, Sjoerd van Rijn, Dimphna H. Meijer, Laura Gainche, Miguel Sena-Esteves, William T. Curry, Bob S. Carter, Anna M. Krichevsky, and Xandra O. Breakefield. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature cell biology*, 10(12):1470–1476, 2008.

[385] Hwa Jin Jung and Yousin Suh. Circulating miRNAs in ageing and ageing-related diseases. *Journal of genetics and genomics = Yi chuan xue bao*, 41(9):465–472, 2014.

[386] Evi M. Mercken, Elisa Majounie, Jinhui Ding, Rong Guo, Jiyoung Kim, Michel Bernier, Julie Mattison, Mark R. Cookson, Myriam Gorospe, Rafael de Cabo, and Kotb Abdelmohsen. Age-associated miRNA alterations in skeletal muscle from rhesus monkeys reversed by caloric restriction. *Aging*, 5(9):692–703, 2013.

[387] Joseph M. Dhahbi, Stephen R. Spindler, Hani Atamna, Amy Yamakawa, Noel Guerrero, Dario Boffelli, Patricia Mote, and David I. K. Martin. Deep sequencing identifies circulating mouse miRNAs that are functionally implicated in manifestations of aging and responsive to calorie restriction. *Aging*, 5(2):130–141, 2013.

[388] Stefanie Dimmeler and Pierluigi Nicotera. MicroRNAs in age-related diseases. *EMBO molecular medicine*, 5(2):180–190, 2013.

[389] Hwa Jin Jung and Yousin Suh. MicroRNA in Aging: From Discovery to Biology. *Current genomics*, 13(7):548–557, 2012.

[390] Seunghee Lee, Kyung-Rok Yu, Young-Sil Ryu, Young Sun Oh, In-Sun Hong, Hyung-Sik Kim, Jin Young Lee, Sunghoon Kim, Kwang-Won Seo, and Kyung-Sun Kang. miR-543 and miR-590-3p regulate human mesenchymal stem cell aging via direct targeting of AIMP3/p18. *Age (Dordrecht, Netherlands)*, 36(6):9724, 2014.

[391] Juan Li, Guoying Dong, Bo Wang, Wei Gao, and Qing Yang. miR-543 promotes gastric cancer cell proliferation by targeting SIRT1. *Biochemical and biophysical research communications*, 469(1):15–21, 2016.

[392] Dan Xu, Fumitaka Takeshita, Yumiko Hino, Saori Fukunaga, Yasusei Kudo, Aya Tamaki, Junko Matsunaga, Ryou-U Takahashi, Takashi Takata, Akira Shimamoto, Takahiro Ochiya, and Hidetoshi Tahara. miR-22 represses cancer progression by inducing cellular senescence. *The Journal of cell biology*, 193(2):409–424, 2011.

[393] Zhan-Peng Huang and Da-Zhi Wang. miR-22 in cardiac remodeling and disease. *Trends in cardiovascular medicine*, 24(7):267–272, 2014.

[394] Irene K. Guttilla and Bruce A. White. Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cancer cells. *Journal of Biological Chemistry*, 284(35):23204–23216, 2009.

[395] Junjie Wu, Jun Qian, Chun Li, Letty Kwok, Feng Cheng, Peijun Liu, Catalina Perdomo, Darrell Kotton, Cyrus Vaziri, Christina Anderlind, Avrum Spira, Wellington V. Cardoso, and Jining Lü. miR-129 regulates cell proliferation by downregulating Cdk6 expression. *Cell cycle (Georgetown, Tex.)*, 9(9):1809–1818, 2010.

[396] M. Karaayvaz, H. Zhai, and J. Ju. miR-129 promotes apoptosis and enhances chemosensitivity to 5-fluorouracil in colorectal cancer. *Cell death & disease*, 4:e659, 2013.

[397] Changmei Liu, Zhao-Qian Teng, Nicholas J. Santistevan, Keith E. Szulwach, Weixiang Guo, Peng Jin, and Xinyu Zhao. Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell stem cell*, 6(5):433–444, 2010.

[398] Paola Mirra, Cecilia Nigro, Immacolata Prevenzano, Teresa Procopio, Alessia Leone, Gregory Alexander Raciti, Francesco Andreozzi, Michele Longo, Francesca Fiory, Francesco Beguinot, and Claudia Miele. The role of miR-190a in methylglyoxal-induced insulin resistance in endothelial cells. *Biochimica et biophysica acta. Molecular basis of disease*, 1863(2):440–449, 2017.

[399] Ulrich Wellner, Jörg Schubert, Ulrike C. Burk, Otto Schmalhofer, Feng Zhu, Annika Sonntag, Bettina Waldvogel, Corinne Vannier, Douglas Darling, Axel Zur Hausen, Valerie G. Brunton, Jennifer Morton, Owen Sansom, Julia Schüler, Marc P. Stemmler, Christoph Herzberger, Ulrich Hopt, Tobias Keck, Simone Brabletz, and Thomas Brabletz. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nature cell biology*, 11(12):1487–1495, 2009.

[400] Jian-Xun Wang, Jian-Qin Jiao, Qian Li, Bo Long, Kun Wang, Jin-Ping Liu, Yan-Rui Li, and Pei-Feng Li. miR-499 regulates mitochondrial dynamics by targeting calcineurin and dynamin-related protein-1. *Nature medicine*, 17(1):71–78, 2011.

[401] Bao-Cun Zhang, Ze-Jun Zhou, and Li Sun. pol-miR-731, a teleost miRNA upregulated by megalocytivirus, negatively regulates virus-induced type I interferon response, apoptosis, and cell cycle arrest. *Scientific reports*, 6:28354, 2016.

[402] L. van der Maaten and Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[403] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1):44–57, 2009.

[404] João Pedro de Magalhães, Caleb E. Finch, and Georges Janssens. Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing research reviews*, 9(3):315–323, 2010.

[405] Jessie C. Jeyapalan and John M. Sedivy. Cellular senescence and organismal aging. *Mechanisms of ageing and development*, 129(7-8):467–474, 2008.

[406] Naina Bhatia-Dey, Riya R. Kanherkar, Susan E. Stair, Evgeny O. Makarev, and Antonei B. Csoka. Cellular Senescence as the Causal Nexus of Aging. *Frontiers in genetics*, 7:13, 2016.

[407] Brian Giunta, Francisco Fernandez, William V. Nikolic, Demian Obregon, Elona Rrapo, Terrence Town, and Jun Tan. Inflammaging as a prodrome to Alzheimer's disease. *Journal of neuroinflammation*, 5:51, 2008.

[408] Sonya Vasto and Calogero Caruso. Immunity & Ageing: a new journal looking at ageing from an immunological point of view. *Immunity & ageing : I & A*, 1(1):1, 2004.

[409] Claudio Franceschi, Miriam Capri, Daniela Monti, Sergio Giunta, Fabiola Olivieri, Federica Sevini, Maria Panagiota Panourgia, Laura Invidia, Laura Celani, Maria Scurti, Elisa Cevenini, Gastone C. Castellani, and Stefano Salvioli. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mechanisms of ageing and development*, 128(1):92–105, 2007.

[410] Mónica de La Fuente and Jaime Miquel. An update of the oxidation-inflammation theory of aging: the involvement of the immune system in oxi-inflamm-aging. *Current pharmaceutical design*, 15(26):3003–3026, 2009.

[411] Jerry L. McCullough and Kristen M. Kelly. Prevention and treatment of skin aging. *Annals of the New York Academy of Sciences*, 1067:323–331, 2006.

[412] Sigrun Ressler, Jirina Bartkova, Harald Niederegger, Jiri Bartek, Karin Scharffetter-Kochanek, Pidder Jansen-Dürr, and Meinhard Wlaschek. p16INK4A is a robust in vivo biomarker of cellular aging in human skin. *Aging cell*, 5(5):379–389, 2006.

[413] Matthew M. Stern and Jackie R. Bickenbach. Epidermal stem cells are resistant to cellular aging. *Aging cell*, 6(4):439–452, 2007.

[414] Adam Giangreco, Mei Qin, John E. Pintar, and Fiona M. Watt. Epidermal stem cells are retained in vivo throughout skin aging. *Aging cell*, 7(2):250–259, 2008.

[415] M. E. Guicciardi and G. J. Gores. Apoptosis: a mechanism of acute and chronic liver injury. *Gut*, 54(7):1024–1033, 2005.

[416] Maria Eugenia Guicciardi, Harmeet Malhi, Justin L. Mott, and Gregory J. Gores. Apoptosis and necrosis in the liver. *Comprehensive Physiology*, 3(2):977–1010, 2013.

[417] J. Miquel. An update on the oxygen stress-mitochondrial mutation theory of aging: genetic and evolutionary implications. *Experimental gerontology*, 33(1-2):113–126, 1998.

[418] Helmut Sies, Carsten Berndt, and Dean P. Jones. Oxidative Stress. *Annual review of biochemistry*, 86:715–748, 2017.

[419] Gustavo Barja. Free radicals and aging. *Trends in neurosciences*, 27(10):595–600, 2004.

[420] Reinald Pamplona and Gustavo Barja. Aging Rate, Mitochondrial Free Radical Production, and Constitutive Sensitivity to Lipid Peroxidation: Insights From Comparative Studies. In Thomas von Zglinicki, editor, *Aging at the molecular level*, Biology of aging and its modulation, pages 47–64. Springer, Dordrecht and London, 2011.

[421] Annette Plüddemann, Claudine Neyen, and Siamon Gordon. Macrophage scavenger receptors and host-derived ligands. *Methods (San Diego, Calif.)*, 43(3):207–217, 2007.

[422] Dawn M. E. Bowdish and Siamon Gordon. Conserved domains of the class A scavenger receptors: evolution and function. *Immunological reviews*, 227(1):19–31, 2009.

[423] Subhankar Mukhopadhyay, Audrey Varin, Yunying Chen, Baoying Liu, Karl Tryggvason, and Siamon Gordon. SR-A/MARCO-mediated ligand delivery enhances intracellular TLR and NLR function, but ligand scavenging from cell surface limits TLR4 response to pathogens. *Blood*, 117(4):1319–1328, 2011.

[424] Kyle E. Novakowski, Angela Huynh, SeongJun Han, Michael G. Dorrington, Charles Yin, Zhongyuan Tu, Peter Pelka, Peter Whyte, Alba Guarné, Kaori Sakamoto, and Dawn M. E. Bowdish. A naturally occurring transcript variant of MARCO reveals the SRCR domain is critical for function. *Immunology and cell biology*, 94(7):646–655, 2016.

[425] Yang Yu and Richard D. Ye. Microglial Ab receptors in Alzheimer's disease. *Cellular and molecular neurobiology*, 35(1):71–83, 2015.

[426] S. R. Coughlin, T. K. Vu, D. T. Hung, and V. I. Wheaton. Characterization of a functional thrombin receptor. Issues and opportunities. *The Journal of clinical investigation*, 89(2):351–355, 1992.

[427] He Wu, Zhiyi Zhang, Ying Li, Ruibo Zhao, Heng Li, Yuejia Song, Jiping Qi, and Jian Wang. Time course of upregulation of inflammatory mediators in the hemorrhagic brain in rats: correlation with brain edema. *Neurochemistry international*, 57(3):248–253, 2010.

[428] Andrew J. Leger, Lidija Covic, and Athan Kuliopulos. Protease-activated receptors in cardiovascular diseases. *Circulation*, 114(10):1070–1077, 2006.

[429] Satoshi Ishii, Takahide Nagase, and Takao Shimizu. Platelet-activating factor receptor. *Prostaglandins & other lipid mediators*, 68-69:599–609, 2002.

[430] Antonios Chatzigeorgiou, Maria Lyberi, Gregorios Chatzilymperis, Adrianos Nezos, and Elli Kamper. CD40/CD40L signaling and its implication in health and disease. *BioFactors (Oxford, England)*, 35(6):474–483, 2009.

[431] M. Morra, J. Lu, F. Poy, M. Martin, J. Sayos, S. Calpe, C. Gullo, D. Howie, S. Rietdijk, A. Thompson, A. J. Coyle, C. Denny, M. B. Yaffe, P. Engel, M. J. Eck, and C. Terhorst. Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells. *The EMBO journal*, 20(21):5840–5852, 2001.

[432] S. Yamamoto, Y. Higuchi, K. Yoshiyama, E. Shimizu, M. Kataoka, N. Hijiya, and K. Matsuura. ADAM family proteins in the immune system. *Immunology today*, 20(6):278–284, 1999.

[433] Nathan E. Hellman and Jonathan D. Gitlin. Ceruloplasmin metabolism and function. *Annual review of nutrition*, 22:439–458, 2002.

[434] A. Milatovich, A. Travis, R. Grosschedl, and U. Francke. Gene for lymphoid enhancer-binding factor 1 (LEF1) mapped to human chromosome 4 (q23-q25) and mouse chromosome 3 near Egf. *Genomics*, 11(4):1040–1048, 1991.

[435] Rajesh Kumar Gandhirajan, Peter Anton Staib, Katharina Minke, Iris Gehrke, Günther Plickert, Axel Schlösser, Esther Katharina Schmitt, Michael Hallek, and Karl-Anton Kreuzer. Small molecule inhibitors of Wnt/beta-catenin/lef-1 signaling induces apoptosis in chronic lymphocytic leukemia cells in vitro and in vivo. *Neoplasia (New York, N.Y.)*, 12(4):326–335, 2010.

[436] E. Labbé, A. Letamendia, and L. Attisano. Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8358–8363, 2000.

[437] Cécile Hervé, Thierry Tonon, Jonas Collén, Erwan Corre, and Catherine Boyen. NADPH oxidases in Eukaryotes: red algae provide new hints! *Current genetics*, 49(3):190–204, 2006.

[438] Sanghamitra Sahoo, Daniel N. Meijles, and Patrick J. Pagano. NADPH oxidases: key modulators in aging and age-related cardiovascular diseases? *Clinical science (London, England : 1979)*, 130(5):317–335, 2016.

[439] Laura Baciou, Rawand Masoud, Hager Souabni, Xavier Serfaty, Gilda Karimi, Tania Bizouarn, and Chantal Houée Levin. Phagocyte NADPH oxidase, oxidative stress and lipids: Anti- or pro ageing? *Mechanisms of ageing and development*, 172:30–34, 2018.

[440] Karen Bedard and Karl-Heinz Krause. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiological reviews*, 87(1):245–313, 2007.

[441] Marina S. Hernandes, Joana C. D'Avila, Silvia C. Trevelin, Patricia A. Reis, Erika R. Kinjo, Lucia R. Lopes, Hugo C. Castro-Faria-Neto, Fernando Q. Cunha, Luiz R. G. Britto, and Fernando A. Bozza. The role of Nox2-derived ROS in the development of cognitive impairment after sepsis. *Journal of neuroinflammation*, 11:36, 2014.

[442] S. Harayama, M. Kok, and E. L. Neidle. Functional and evolutionary relationships among diverse oxygenases. *Annual review of microbiology*, 46:565–601, 1992.

[443] Jacob M. Zahn, Rebecca Sonu, Hannes Vogel, Emily Crane, Krystyna Mazan-Mamczarz, Ralph Rabkin, Ronald W. Davis, Kevin G. Becker, Art B. Owen, and Stuart K. Kim. Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS genetics*, 2(7):e115, 2006.

[444] Rajkumar Noubade, Kit Wong, Naruhisa Ota, Sascha Rutz, Celine Eidenschenk, Patricia A. Valdez, Jiabing Ding, Ivan Peng, Andrew Sebrell, Patrick Caplazi, Jason DeVoss, Robert H. Soriano, Tao Sai, Rongze Lu, Zora Modrusan, Jason Hackney, and Wenjun Ouyang. NRROS negatively regulates reactive oxygen species during host defence and autoimmunity. *Nature*, 509(7499):235–239, 2014.

[445] Stefano Toppo, Stefano Vanin, Valentina Bosello, and Silvio C. E. Tosatto. Evolutionary and structural insights into the multifaceted glutathione peroxidase (Gpx) superfamily. *Antioxidants & redox signaling*, 10(9):1501–1514, 2008.

[446] Taro Sakamoto, Kana Maebayashi, Yasuhito Nakagawa, and Hirotaka Imai. Deletion of the four phospholipid hydroperoxide glutathione peroxidase genes accelerates aging in Caenorhabditis elegans. *Genes to cells : devoted to molecular & cellular mechanisms*, 19(10):778–792, 2014.

[447] Zhiwei Wang, Pengda Liu, Hiroyuki Inuzuka, and Wenyi Wei. Roles of F-box proteins in cancer. *Nature reviews. Cancer*, 14(4):233–247, 2014.

[448] L. E. Dow, J. S. Kauffman, J. Caddy, K. Zarbalis, A. S. Peterson, S. M. Jane, S. M. Russell, and P. O. Humbert. The tumour-suppressor Scribble dictates cell polarity during directed epithelial migration: regulation of Rho GTPase recruitment to the leading edge. *Oncogene*, 26(16):2272–2282, 2007.

[449] G. J. Hannon and D. Beach. p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest. *Nature*, 371(6494):257–261, 1994.

[450] Minglei Zhang, Dapeng Wang, Tongtong Zhu, and Ruofeng Yin. RASSF4 Overexpression Inhibits the Proliferation, Invasion, EMT, and Wnt Signaling Pathway in Osteosarcoma Cells. *Oncology research*, 25(1):83–91, 2017.

[451] Junqi He, Jianguo Xu, Xiang-Xi Xu, and Randy A. Hall. Cell cycle-dependent phosphorylation of Disabled-2 by cdc2. *Oncogene*, 22(29):4524–4530, 2003.

[452] Laurent Gauthier, Benjamin Rossi, Florence Roux, Elise Termine, and Claudine Schiff. Galectin-1 is a stromal cell ligand of the pre-B cell receptor (BCR) implicated in synapse formation between pre-B and stromal cells and in pre-BCR triggering. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):13014–13019, 2002.

[453] Eui Young So and Toru Ouchi. BRAT1 deficiency causes increased glucose metabolism and mitochondrial malfunction. *BMC cancer*, 14:548, 2014.

[454] N. K. Sah, Z. Khan, G. J. Khan, and P. S. Bisen. Structural, functional and therapeutic biology of survivin. *Cancer letters*, 244(2):164–171, 2006.

[455] Reto Gassmann, Ana Carvalho, Alexander J. Henzing, Sandrine Ruchaud, Damien F. Hudson, Reiko Honda, Erich A. Nigg, Dietlind L. Gerloff, and William C. Earnshaw. Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. *The Journal of cell biology*, 166(2):179–191, 2004.

[456] Christopher W. Borysenko, William F. Furey, and Harry C. Blair. Comparative modeling of TNFRSF25 (DR3) predicts receptor destabilization by a mutation linked to rheumatoid arthritis. *Biochemical and biophysical research communications*, 328(3):794–799, 2005.

[457] K. Shimizu, S. Chiba, N. Hosoya, K. Kumano, T. Saito, M. Kurokawa, Y. Kanda, Y. Hamada, and H. Hirai. Binding of Delta1, Jagged1, and Jagged2 to Notch2 rapidly induces cleavage, nuclear translocation, and hyperphosphorylation of Notch2. *Molecular and cellular biology*, 20(18):6913–6922, 2000.

[458] Alexander Pietras, Kristoffer von Stedingk, David Lindgren, Sven Påhlman, and Håkan Axelson. JAG2 induction in hypoxic tumor cells alters Notch signaling and enhances endothelial cell tube formation. *Molecular cancer research : MCR*, 9(5):626–636, 2011.

[459] David Chiron, Sophie Maïga, Géraldine Descamps, Philippe Moreau, Steven Le Gouill, Séverine Marionneau, Thibauld Ouiller, Jérôme Moreaux, Bernard Klein, Régis Bataille, Martine Amiot, and Catherine Pellat-Deceunynck. Critical role of the NOTCH ligand JAG2 in self-renewal of myeloma cells. *Blood cells, molecules & diseases*, 48(4):247–253, 2012.

[460] Christiane Houde, Yulin Li, Lynda Song, Kevin Barton, Qing Zhang, John Godwin, Sucha Nand, Amir Toor, Serhan Alkan, N. Veronique Smadja, Hervé Avet-Loiseau, Carmen S. Lima, Lucio Miele, and Lionel J. Coignet. Overexpression of the NOTCH ligand JAG2 in malignant plasma cells from multiple myeloma patients and cell lines. *Blood*, 104(12):3697–3704, 2004.

[461] Walter Swardfager, Krista Lanctôt, Lana Rothenburg, Amy Wong, Jaclyn Cappell, and Nathan Herrmann. A meta-analysis of cytokines in Alzheimer's disease. *Biological psychiatry*, 68(10):930–941, 2010.

[462] Andrzej Dubiński and Zygmunt Zdrojewicz. Rola interleukiny 6 w wystapieniu i progresji miazdzycy. *Polski merkuriusz lekarski : organ Polskiego Towarzystwa Lekarskiego*, 22(130):291–294, 2007.

[463] Ole P. Kristiansen and Thomas Mandrup-Poulsen. Interleukin-6 and diabetes: the good, the bad, or the indifferent? *Diabetes*, 54 Suppl 2:S114–24, 2005.

[464] Beverly E. Barton. Interleukin-6 and new strategies for the treatment of cancer, hyperproliferative diseases and paraneoplastic syndromes. *Expert opinion on therapeutic targets*, 9(4):737–752, 2005.

[465] Josef S. Smolen and Ravinder N. Maini. Interleukin-6: a new therapeutic target. *Arthritis research & therapy*, 8 Suppl 2:S5, 2006.

[466] Mauro Perretti and Fulvio D'Acquisto. Annexin A1 and glucocorticoids as effectors of the resolution of inflammation. *Nature reviews. Immunology*, 9(1):62–70, 2009.

[467] Zhiquan Zhang, Liqun Huang, Wenping Zhao, and Basil Rigas. Annexin 1 induced by anti-inflammatory drugs binds to NF-kappaB and inhibits its activation: anticancer effects in vitro and in vivo. *Cancer research*, 70(6):2379–2388, 2010.

[468] Huensuk Kim, Seungyeul Yoo, Ruoji Zhou, An Xu, Jeffrey M. Bernitz, Ye Yuan, Andreia M. Gomes, Michael G. Daniel, Jie Su, Elizabeth G. Demicco, Jun Zhu, Kateri A. Moore, Dung-Fang Lee, Ihor R. Lemischka, and Christoph Schaniel. Oncogenic role of SFRP2 in p53-mutant osteosarcoma development via autocrine and paracrine mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 115(47):E11128–E11137, 2018.

[469] Jie Ren, Fangfang Jian, Hong Jiang, Yuhao Sun, Sijian Pan, Changwei Gu, Xiao Chen, Weiqing Wang, Guang Ning, Liuguan Bian, and Qingfang Sun. Decreased expression of SFRP2 promotes development of the pituitary corticotroph adenoma by upregulating Wnt signaling. *International journal of oncology*, 52(6):1934–1946, 2018.

[470] Michal Simicek, Sam Lievens, Mathias Laga, Dmytro Guzenko, Vasily N. Aushev, Peter Kalev, Maria Francesca Baietti, Sergei V. Strelkov, Kris Gevaert, Jan Tavernier, and Anna A. Sablina. The deubiquitylase USP33 discriminates between RALB functions in autophagy and innate immune response. *Nature cell biology*, 15(10):1220–1230, 2013.

[471] Awet Tecleab, Xiaolei Zhang, and Said M. Sebti. Ral GTPase down-regulation stabilizes and reactivates p53 to inhibit malignant transformation. *Journal of Biological Chemistry*, 289(45):31296–31309, 2014.

[472] David F. Kashatus. Ral GTPases in tumorigenesis: emerging from the shadows. *Experimental cell research*, 319(15):2337–2342, 2013.

[473] Jesús Villar, Haibo Zhang, and Arthur S. Slutsky. Lung Repair and Regeneration in ARDS: Role of PECAM1 and Wnt Signaling. *Chest*, 2018.

[474] L. Maouche-Chrétien, N. Deleu, C. Badoual, P. Fraissignes, R. Berger, P. Gaulard, P. H. Roméo, and K. Leroy-Viard. Identification of a novel cDNA, encoding a cytoskeletal associated protein, differentially expressed in diffuse large B cell lymphomas. *Oncogene*, 17(10):1245–1251, 1998.

[475] Scott D. Slattery, Michael A. Mancini, Bill R. Brinkley, and Rebecca M. Hall. Aurora-C kinase supports mitotic progression in the absence of Aurora-B. *Cell cycle (Georgetown, Tex.)*, 8(18):2984–2994, 2009.

[476] Armando van der Horst, Jacinta Simmons, and Kum Kum Khanna. Cep55 stabilization is required for normal execution of cytokinesis. *Cell cycle (Georgetown, Tex.)*, 8(22):3742–3749, 2009.

[477] Chikashi Obuse, Osamu Iwasaki, Tomomi Kiyomitsu, Gohta Goshima, Yusuke Toyoda, and Mitsuhiro Yanagida. A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nature cell biology*, 6(11):1135–1141, 2004.

[478] Mark L. McCleland, Marko J. Kallio, Gregory A. Barrett-Wilt, Cortney A. Kestner, Jeffrey Shabanowitz, Donald F. Hunt, Gary J. Gorbsky, and P. Todd Stukenberg. The vertebrate Ndc80 complex contains Spc24 and Spc25 homologs, which are required to establish and maintain kinetochore-microtubule attachment. *Current biology : CB*, 14(2):131–137, 2004.

[479] Troy C. Lund, Tiffany J. Glass, Jakub Tolar, and Bruce R. Blazar. Expression of telomerase and telomere length are unaffected by either age or limb regeneration in Danio rerio. *PloS one*, 4(11):e7688, 2009.

[480] Francesco Prattichizzo, Massimiliano Bonafè, Fabiola Olivieri, and Claudio Franceschi. Senescence associated macrophages and "macroph-aging": are they pieces of the same puzzle? *Aging (Albany NY)*, 8(12):3159–3160, 2016.

[481] Bennett G. Childs, Darren J. Baker, Tobias Wijshake, Cheryl A. Conover, Judith Campisi, and Jan M. van Deursen. Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science (New York, N.Y.)*, 354(6311):472–477, 2016.

[482] Elizabeth E. Spangenberg, Rafael J. Lee, Allison R. Najafi, Rachel A. Rice, Monica R. P. Elmore, Mathew Blurton-Jones, Brian L. West, and Kim N. Green. Eliminating microglia in Alzheimer's mice prevents neuronal loss without modulating amyloid-b pathology. *Brain : a journal of neurology*, 139(Pt 4):1265–1281, 2016.

[483] Shihao Shen, Juw Won Park, Zhi-xiang Lu, Lan Lin, Michael D. Henry, Ying Nian Wu, Qing Zhou, and Yi Xing. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, 111(51):E5593–601, 2014.

[484] Joan E. Hooper. A survey of software for genome-wide discovery of differential splicing in RNA-Seq data. *Human genomics*, 8:3, 2014.

[485] Wei Liao, Gwen Jordaan, Phillipp Nham, Ryan T. Phan, Matteo Pelegrini, and Sanjai Sharma. Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens. *BMC cancer*, 15:714, 2015.

[486] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic acids research*, 46(5):2699, 2018.

[487] Olga Kelemen, Paolo Convertini, Zhaiyi Zhang, Yuan Wen, Manli Shen, Marina Falaleeva, and Stefan Stamm. Function of alternative splicing. *Gene*, 514(1):1–30, 2013.

[488] Mihaela Pertea, Geo M. Pertea, Corina M. Antonescu, Tsung-Cheng Chang, Joshua T. Mendell, and Steven L. Salzberg. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*, 33(3):290–295, 2015.

[489] Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F. Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9):1236–1240, 2014.

[490] Hadas Keren, Galit Lev-Maor, and Gil Ast. Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews. Genetics*, 11(5):345–355, 2010.

[491] Mo Chen and James L. Manley. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews. Molecular cell biology*, 10(11):741–754, 2009.

[492] Cindy L. Will and Reinhard Lührmann. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, 3(7), 2011.

[493] A. Gregory Matera and Zefeng Wang. A day in the life of the spliceosome. *Nature reviews. Molecular cell biology*, 15(2):108–121, 2014.

[494] Yoseph Barash, John A. Calarco, Weijun Gao, Qun Pan, Xinchen Wang, Ofer Shai, Benjamin J. Blencowe, and Brendan J. Frey. Deciphering the splicing code. *Nature*, 465(7294):53–59, 2010.

[495] Annachiara de Sandre-Giovannoli, Rafaëlle Bernard, Pierre Cau, Claire Navarro, Jeanne Amiel, Irène Boccaccio, Stanislas Lyonnet, Colin L. Stewart, Arnold Munnich, Martine Le Merrer, and Nicolas Lévy. Lamin a truncation in Hutchinson-Gilford progeria. *Science (New York, N.Y.)*, 300(5628):2055, 2003.

[496] Kan Cao, John J. Graziotto, Cecilia D. Blair, Joseph R. Mazzulli, Michael R. Erdos, Dimitri Krainc, and Francis S. Collins. Rapamycin reverses cellular phenotypes and enhances mutant protein clearance in Hutchinson-Gilford progeria syndrome cells. *Science translational medicine*, 3(89):89ra58, 2011.

[497] Shufen Chen, Kirk Townsend, Terry E. Goldberg, Peter Davies, and Concepcion Conejero-Goldberg. MAPT isoforms: differential transcriptional profiles related to 3R and 4R splice variants. *Journal of Alzheimer's disease : JAD*, 22(4):1313–1329, 2010.

[498] Soraya Scuderi, Valentina La Cognata, Filippo Drago, Sebastiano Cavallaro, and Velia D'Agata. Alternative splicing generates different parkin protein isoforms: evidences in human, rat, and mouse brain. *BioMed research international*, 2014, 2014.

[499] A. Sveen, S. Kilpinen, A. Ruusulehto, R. A. Lothe, and R. I. Skotheim. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*, 35(19):2413–2427, 2016.

[500] Lorna W. Harries, Dena Hernandez, William Henley, Andrew R. Wood, Alice C. Holly, Rachel M. Bradley-Smith, Hanieh Yaghootkar, Ambarish Dutta, Anna Murray, Timothy M. Frayling, Jack M. Guralnik, Stefania Bandinelli, Andrew Singleton, Luigi Ferrucci, and David Melzer. Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging cell*, 10(5):868–878, 2011.

[501] Benjamin P. Lee, Luke C. Pilling, Florence Emond, Kevin Flurkey, David E. Harrison, Rong Yuan, Luanne L. Peters, George A. Kuchel, Luigi Ferrucci, David Melzer, and Lorna W. Harries. Changes in the expression of splicing factor transcripts and variations in alternative splicing are associated with lifespan in mice and humans. *Aging cell*, 15(5):903–913, 2016.

[502] Alice C. Holly, David Melzer, Luke C. Pilling, Alexander C. Fellows, Toshiko Tanaka, Luigi Ferrucci, and Lorna W. Harries. Changes in splicing factor expression are associated with advancing age in man. *Mechanisms of ageing and development*, 134(9):356–366, 2013.

[503] Pavel Mazin, Jieyi Xiong, Xiling Liu, Zheng Yan, Xiaoyu Zhang, Mingshuang Li, Liu He, Mehmet Somel, Yuan Yuan, Yi-Ping Phoebe Chen, Na Li, Yuhui Hu, Ning Fu, Zhibin Ning, Rong Zeng, Hongyi Yang, Wei Chen, Mikhail Gelfand, and Philipp Khaitovich. Widespread splicing changes in human brain development and aging. *Molecular systems biology*, 9:633, 2013.

[504] Sofía A. Rodríguez, Diana Grochová, Tomás McKenna, Bhavesh Borate, Niraj S. Trivedi, Michael R. Erdos, and Maria Eriksson. Global genome splicing analysis reveals an increased number of alternatively spliced genes with aging. *Aging cell*, 15(2):267–278, 2016.

[505] James R. Tollervey, Zhen Wang, Tibor Hortobágyi, Joshua T. Witten, Kathi Zarnack, Melis Kayikci, Tyson A. Clark, Anthony C. Schweitzer, Gregor Rot, Tomaž Curk, Blaž Zupan, Boris Rogelj, Christopher E. Shaw, and Jernej Ule. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome research*, 21(10):1572–1582, 2011.

[506] Kathleen F. Mittendorf, Catherine L. Deatherage, Melanie D. Ohi, and Charles R. Sanders. Tailoring of membrane proteins by alternative splicing of pre-mRNA. *Biochemistry*, 51(28):5541–5556, 2012.

[507] H. Ye, T. F. Kelly, U. Samadani, L. Lim, S. Rubio, D. G. Overdier, K. A. Roebuck, and R. H. Costa. Hepatocyte nuclear factor 3/fork head homolog 11 is expressed in proliferating epithelial and mesenchymal cells of embryonic and adult tissues. *Molecular and cellular biology*, 17(3):1626–1641, 1997.

[508] Armando van der Horst and Boudewijn M. T. Burgering. Stressing the role of FoxO proteins in lifespan and disease. *Nature reviews. Molecular cell biology*, 8(6):440–450, 2007.

[509] Jamila Laoukili, Matthijs R. H. Kooistra, Alexandra Brás, Jos Kauw, Ron M. Kerkhoven, Ashby Morrison, Hans Clevers, and René H. Medema. FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nature cell biology*, 7(2):126–136, 2005.

[510] Inken Wierstra and Jürgen Alves. FOXM1, a typical proliferation-associated transcription factor. *Biological chemistry*, 388(12):1257–1274, 2007.

[511] M. Mourtada-Maarabouni, M. R. Pickard, V. L. Hedge, F. Farzaneh, and G. T. Williams. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*, 28(2):195–208, 2009.

[512] C. M. Smith and J. A. Steitz. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Molecular and cellular biology*, 18(12):6897–6909, 1998.

[513] G. Raho, V. Barone, D. Rossi, L. Philipson, and V. Sorrentino. The gas 5 gene shows four alternative splicing patterns without coding for a protein. *Gene*, 256(1-2):13–17, 2000.

[514] Mirna Mourtada-Maarabouni, Vanessa L. Hedge, Lucy Kirkham, Farzin Farzaneh, and Gwyn T. Williams. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *Journal of cell science*, 121(Pt 7):939–946, 2008.

[515] Iakes Ezkurdia, Jose Manuel Rodriguez, Enrique Carrillo-de Santa Pau, Jesús Vázquez, Alfonso Valencia, and Michael L. Tress. Most highly expressed protein-coding genes have a single dominant isoform. *Journal of proteome research*, 14(4):1880–1887, 2015.

[516] Sara Light and Arne Elofsson. The impact of splicing on protein domain architecture. *Current opinion in structural biology*, 23(3):451–458, 2013.

[517] A. W. van der Velden and A. A. Thomas. The role of the 5' untranslated region of an mRNA in translation regulation during development. *The international journal of biochemistry & cell biology*, 31(1):87–106, 1999.

[518] Huan Li, Ziyue Wang, Tianyi Ma, Gang Wei, and Ting Ni. Alternative splicing in aging and age-related diseases. *Translational Medicine of Aging*, 1:32–40, 2017.

[519] Mathieu Deschênes and Benoit Chabot. The emerging role of alternative splicing in senescence and aging. *Aging cell*, 16(5):918–933, 2017.

[520] Klas Hatje, Raza-Ur Rahman, Ramon O. Vidal, Dominic Simm, Björn Hammesfahr, Vikas Bansal, Ashish Rajput, Michel Edwar Mickael, Ting Sun, Stefan Bonn, and Martin Kollmar. The landscape of human mutually exclusive splicing. *Molecular systems biology*, 13(12):959, 2017.

[521] Claudio R. Alarcón, Hani Goodarzi, Hyeseung Lee, Xuhang Liu, Saeed Tavazoie, and Sohail F. Tavazoie. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell*, 162(6):1299–1308, 2015.

[522] P. D. Zamore, J. G. Patton, and M. R. Green. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, 355(6361):609–614, 1992.

[523] Rupak Mukhopadhyay, Partho Sarothi Ray, Abul Arif, Anna K. Brady, Michael Kinter, and Paul L. Fox. DAPK-ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Molecular cell*, 32(3):371–382, 2008.

[524] A. Yang, M. Kaghad, Y. Wang, E. Gillett, M. D. Fleming, V. Dötsch, N. C. Andrews, D. Caput, and F. McKeon. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Molecular cell*, 2(3):305–316, 1998.

[525] Gregor B. Deutsch, Elisabeth M. Zielonka, Daniel Coutandin, Tobias A. Weber, Birgit Schäfer, Jens Hannewald, Laura M. Luh, Florian G. Durst, Mohamed Ibrahim, Jan Hoffmann, Frank H. Niesen, Aycan Sentürk, Hana Kunkel, Bernd Brutschy, Enrico Schleiff, Stefan Knapp, Amparo Acker-Palmer, Manuel Grez, Frank McKeon, and Volker Dötsch. DNA damage in oocytes induces a switch of the quality control factor TAp63a from dimer to tetramer. *Cell*, 144(4):566–576, 2011.

[526] Christopher P. Crum and Frank D. McKeon. p63 in epithelial survival, germ cell surveillance, and neoplasia. *Annual review of pathology*, 5:349–371, 2010.

[527] Xiaohua Su, Maryline Paris, Young Jin Gi, Kenneth Y. Tsai, Min Soon Cho, Yu-Li Lin, Jeffrey A. Biernaskie, Satrajit Sinha, Carol Prives, Larysa H. Pevny, Freda D. Miller, and Elsa R. Flores. TAp63 prevents premature aging by promoting adult stem cell maintenance. *Cell stem cell*, 5(1):64–75, 2009.

[528] Lucinda K. Southworth, Art B. Owen, and Stuart K. Kim. Aging mice show a decreasing correlation of gene expression within genetic modules. *PLoS genetics*, 5(12):e1000776, 2009.

[529] Y. W. Fong and Q. Zhou. Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414(6866):929–933, 2001.

[530] Indrani Rebbapragada and Jens Lykke-Andersen. Execution of nonsense-mediated mRNA decay: what defines a substrate? *Current opinion in cell biology*, 21(3):394–402, 2009.

[531] Saiprasad G. Palusa and Anireddy S. N. Reddy. Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *The New phytologist*, 185(1):83–89, 2010.

[532] Marcello Pinti, Victor Appay, Judith Campisi, Daniela Frasca, Tamas Fülöp, Delphine Sauce, Anis Larbi, Birgit Weinberger, and Andrea Cossarizza. Aging of the immune system: Focus on inflammation and vaccination. *European journal of immunology*, 46(10):2286–2301, 2016.

[533] Mehmet Yabas, Hannah Elliott, and Gerard F. Hoyne. The Role of Alternative Splicing in the Control of Immune Homeostasis and Cellular Differentiation. *International journal of molecular sciences*, 17(1), 2015.

[534] Caghan Kizil, Jan Kaslin, Volker Kroehne, and Michael Brand. Adult neurogenesis and brain regeneration in zebrafish. *Developmental neurobiology*, 72(3):429–461, 2012.

[535] R. J. Buckanovich, Y. Y. Yang, and R. B. Darnell. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(3):1114–1122, 1996.

[536] Olatz Villate, Jean-Valery Turatsinze, Loriana G. Mascali, Fabio A. Grieco, Tatiane C. Nogueira, Daniel A. Cunha, Tarlliza R. Nardelli, Michael Sammeth, Vishal A. Salunkhe, Jonathan L. S. Esguerra, Lena Eliasson, Lorella Marselli, Piero Marchetti, and Decio L. Eizirik. Nova1 is a master regulator of alternative splicing in pancreatic beta cells. *Nucleic acids research*, 42(18):11818–11830, 2014.

[537] Christopher D. Wiley, James M. Flynn, Christapher Morrissey, Ronald Lebofsky, Joe Shuga, Xiao Dong, Marc A. Unger, Jan Vijg, Simon Melov, and Judith Campisi. Analysis of individual cells identifies cell-to-cell variability following induction of cellular senescence. *Aging cell*, 16(5):1043–1050, 2017.

[538] Celia Pilar Martinez-Jimenez, Nils Eling, Hung-Chang Chen, Catalina A. Vallejos, Aleksandra A. Kolodziejczyk, Frances Connor, Lovorka Stojic, Timothy F. Rayner, Michael J. T. Stubbington, Sarah A. Teichmann, Maike de La Roche, John C. Marioni, and Duncan T. Odom. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science (New York, N.Y.)*, 355(6332):1433–1436, 2017.

[539] M. H. Vitaterna, J. S. Takahashi, and F. W. Turek. Overview of circadian rhythms. *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, 25(2):85–93, 2001.

[540] Joseph Bass. Circadian topology of metabolism. *Nature*, 491(7424):348–356, 2012.

[541] Esther Isorna, Nuria de Pedro, Ana I. Valenciano, Ángel L. Alonso-Gómez, and María J. Delgado. Interplay between the endocrine and circadian systems in fishes. *The Journal of endocrinology*, 232(3):R141–R159, 2017.

[542] Kun Hu, Eus J. W. van Someren, Steven A. Shea, and Frank A. J. L. Scheer. Reduction of scale invariance of activity fluctuations with aging and Alzheimer's disease: Involvement of the circadian pacemaker. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8):2490–2494, 2009.

[543] Sahar Farajnia, Stephan Michel, Tom Deboer, Henk Tjebbe vanderLeest, Thijs Houben, Jos H. T. Rohling, Ashna Ramkisoensing, Roman Yasenkov, and Johanna H. Meijer. Evidence for neuronal desynchrony in the aged suprachiasmatic nucleus clock. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(17):5891–5899, 2012.

[544] W. Witting, I. H. Kwa, P. Eikelenboom, M. Mirmiran, and D. F. Swaab. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biological psychiatry*, 27(6):563–572, 1990.

[545] Ray Zhang, Nicholas F. Lahens, Heather I. Ballance, Michael E. Hughes, and John B. Hogenesch. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45):16219–16224, 2014.

[546] Biliana Marcheva, Kathryn Moynihan Ramsey, Ethan D. Buhr, Yumiko Kobayashi, Hong Su, Caroline H. Ko, Ganka Ivanova, Chiaki Omura, Shelley Mo, Martha H. Vitaterna, James P. Lopez, Louis H. Philipson, Christopher A. Bradfield, Seth D. Crosby, Lellean JeBailey, Xiaozhong Wang, Joseph S. Takahashi, and Joseph Bass. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature*, 466(7306):627–631, 2010.

[547] Anne M. Curtis and Garret A. Fitzgerald. Central and peripheral clocks in cardiovascular and metabolic function. *Annals of medicine*, 38(8):552–559, 2006.

[548] Erik S. Musiek, Miranda M. Lim, Guangrui Yang, Adam Q. Bauer, Laura Qi, Yool Lee, Jee Hoon Roh, Xilma Ortiz-Gonzalez, Joshua T. Dearborn, Joseph P. Culver, Erik D. Herzog, John B. Hogenesch, David F. Wozniak, Krikor Dikranian, Benoit I. Giasson, David R. Weaver, David M. Holtzman, and Garret A. Fitzgerald. Circadian clock proteins regulate neuronal redox homeostasis and neurodegeneration. *The Journal of clinical investigation*, 123(12):5389–5400, 2013.

[549] Ying Li, Guang Li, Haifang Wang, Jiulin Du, and Jun Yan. Analysis of a gene regulatory cascade mediating circadian rhythm in zebrafish. *PLoS computational biology*, 9(2):e1002940, 2013.

[550] Bruno C. R. Ramos, Maria Nathália C. M. Moraes, Maristela O. Poletini, Leonardo H. R. G. Lima, and Ana Maria L. Castrucci. From blue light to clock genes in zebrafish ZEM-2S cells. *PloS one*, 9(9):e106252, 2014.

[551] Jianguo Liu, Gautam Malkani, Guantham Mankani, Xiaoyan Shi, Mark Meyer, Susana Cunningham-Runddles, Xiaojing Ma, and Zhong Sheng Sun. The circadian clock Period 2 gene regulates gamma interferon production of NK cells in host response to lipopolysaccharide-induced endotoxic shock. *Infection and immunity*, 74(8):4750–4756, 2006.

[552] Hiroaki Hojo, Sora Enya, Miki Arai, Yutaka Suzuki, Takashi Nojiri, Kenji Kangawa, Shinsuke Koyama, and Shinpei Kawaoka. Remote reprogramming of hepatic circadian transcriptome by breast cancer. *Oncotarget*, 8(21):34128–34140, 2017.

[553] Elizabeth A. Yu and David R. Weaver. Disrupting the circadian clock: gene-specific effects on aging, cancer, and other phenotypes. *Aging*, 3(5):479–493, 2011.

[554] Sigal Gery, Naoki Komatsu, Lilit Baldjyan, Andrew Yu, Danielle Koo, and H. Phillip Koeffler. The circadian gene per1 plays an important role in cell growth and DNA damage control in human cancer cells. *Molecular cell*, 22(3):375–382, 2006.

[555] Yulia V. Dubrovsky, William E. Samsa, and Roman V. Kondratov. Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. *Aging*, 2(12):936–944, 2010.

[556] Cho-Yi Chen, Ryan W. Logan, Tianzhou Ma, David A. Lewis, George C. Tseng, Etienne Sibille, and Colleen A. McClung. Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 113(1):206–211, 2016.

[557] Eugene Lin, Po-Hsiu Kuo, Yu-Li Liu, Albert C. Yang, Chung-Feng Kao, and Shih-Jen Tsai. Effects of circadian clock genes and environmental factors on cognitive aging in old adults in a Taiwanese population. *Oncotarget*, 8(15):24088–24098, 2017.

[558] Roman V. Kondratov, Anna A. Kondratova, Victoria Y. Gorbacheva, Olena V. Vykhovanets, and Marina P. Antoch. Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. *Genes & development*, 20(14):1868–1873, 2006.

[559] M. K. Bunger, L. D. Wilsbacher, S. M. Moran, C. Clendenin, L. A. Radcliffe, J. B. Hogenesch, M. C. Simon, J. S. Takahashi, and C. A. Bradfield. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell*, 103(7):1009–1017, 2000.

[560] Guangrui Yang, Lihong Chen, Gregory R. Grant, Georgios Paschos, Wen-Liang Song, Erik S. Musiek, Vivian Lee, Sarah C. McLoughlin, Tilo Grosser, George Cotsarelis, and Garret A. Fitzgerald. Timing of expression of the core clock gene Bmal1 influences its effects on aging and survival. *Science translational medicine*, 8(324):324ra16, 2016.

[561] Jun Yan, Haifang Wang, Yuting Liu, and Chunxuan Shao. Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLoS computational biology*, 4(10):e1000193, 2008.

[562] Daniel Pouly, Sébastien Chenaux, Virginie Martin, Maja Babis, Rafael Koch, Emi Nagoshi, Vladimir L. Katanaev, Frédéric Gachon, and Olivier Staub. USP2-45 Is a Circadian Clock Output Effector Regulating Calcium Absorption at the Post-Translational Level. *PloS one*, 11(1):e0145155, 2016.

[563] Ryan C. Russell, Ye Tian, Haixin Yuan, Hyun Woo Park, Yu-Yun Chang, Joungmok Kim, Haerin Kim, Thomas P. Neufeld, Andrew Dillin, and Kun-Liang Guan. ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. *Nature cell biology*, 15(7):741–750, 2013.

[564] Willemijn M. Passtoors, Marian Beekman, Joris Deelen, Ruud van der Breggen, Andrea B. Maier, Bruno Guigas, Evelyna Derhovanessian, Diana van Heemst, Anton J. M. de Craen, David A. Gunn, Graham Pawelec, and Pieternella E. Slagboom. Gene expression analysis of mTOR pathway: association with human longevity. *Aging cell*, 12(1):24–31, 2013.

[565] Dominique H. Eghlidi, Selva L. Luna, Donald I. Brown, Vasilios T. Garyfallou, Steven G. Kohama, and Henryk F. Urbanski. Gene expression profiling of the SCN in young and old rhesus macaques. *Journal of molecular endocrinology*, 61(2):57–67, 2018.

[566] Guillaume Rey, François Cesbron, Jacques Rougemont, Hans Reinke, Michael Brunner, and Felix Naef. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS biology*, 9(2):e1000595, 2011.

[567] Renata Pellegrino, Ibrahim Halil Kavakli, Namni Goel, Christopher J. Cardinale, David F. Dinges, Samuel T. Kuna, Greg Maislin, Hans P. A. van Dongen, Sergio Tufik, John B. Hogenesch, Hakon Hakonarson, and Allan I. Pack. A novel BHLHE41 variant is associated with short sleep and resistance to sleep deprivation in humans. *Sleep*, 37(8):1327–1336, 2014.

[568] Fuyuki Sato, Ujjal K. Bhawal, Tomohiro Yoshimura, and Yasuteru Muragaki. DEC1 and DEC2 Crosstalk between Circadian Rhythm and Tumor Progression. *Journal of Cancer*, 7(2):153–159, 2016.

[569] Jan R. Aunan, William C. Cho, and Kjetil Søreide. The Biology of Aging and Cancer: A Brief Overview of Shared and Divergent Molecular Hallmarks. *Aging and disease*, 8(5):628–642, 2017.

[570] Cristina Cadenas, Leonie van de Sandt, Karolina Edlund, Miriam Lohr, Birte Hellwig, Rosemarie Marchan, Marcus Schmidt, Jörg Rahnenführer, Henrik Oster, and Jan G. Hengstler. Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell cycle (Georgetown, Tex.)*, 13(20):3282–3291, 2014.

[571] P. Verbeke, J. Fonager, B. F. Clark, and S. I. Rattan. Heat shock response and ageing: mechanisms and applications. *Cell biology international*, 25(9):845–857, 2001.

[572] Torsten Held, Amal Z. Barakat, Belal A. Mohamed, Ilona Paprotta, Andreas Meinhardt, Wolfgang Engel, and Ibrahim M. Adham. Heat-shock protein HSPA4 is required for progression of spermatogenesis. *Reproduction (Cambridge, England)*, 142(1):133–144, 2011.

[573] Kai-Florian Storch, Ovidiu Lipan, Igor Leykin, N. Viswanathan, Fred C. Davis, Wing H. Wong, and Charles J. Weitz. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83, 2002.

List of Figures

List of Tables

Curriculum vitae

Education

since 01/2015	**Ph.D. Student** Friedrich Schiller University Jena RNA Bioinformatics and High Throughput Analysis PhD thesis: "Insights Into The Regulation Of Aging"
2012 – 2014	**Master of Science in Bioinformatics** Friedrich Schiller University Jena Master thesis: "Comparison of *de novo* transcriptome assemblers"
2009 – 2012	**Bachelor of Science in Bioinformatics** Friedrich Schiller University Jena Bachelor thesis: "Homology search of small non-coding RNAs in *Bacillus subtilis*"
2008	**Diploma qualifying for university admission** Diesterweggymnasium Plauen

Conferences and workshops

02/2018	33^{th} TBI Winterseminar *Bled, Slovenia*	**Talk**
10/2017	15^{th} Herbstseminar der Bioinformatik *Doubice, Czech Republic*	**Talk**
04/2017	1^{st} HACKEN (Workshop) *Jena, Germany*	**Coordinator**
03/2017	1^{st} EVBC Meeting *Jena, Germany*	**Coordinator**
05/2017	Central German Meeting on Bioinformatics 2017 *Leipzig, Germany*	**Poster**
02/2017	32^{th} TBI Winterseminar *Bled, Slovenia*	**Talk**
10/2016	14^{th} Herbstseminar der Bioinformatik *Doubice, Czech Republic*	**Talk**
02/2016	31^{th} TBI Winterseminar *Bled, Slovenia*	**Talk**
01/2016	ZAJ Symposium *Jena, Germany*	**Poster**
10/2015	13^{th} Herbstseminar der Bioinformatik *Doubice, Czech Republic*	**Talk**
08/2015	Central German Meeting on Bioinformatics 2015 *Halle, Germany*	**Poster**
02/2015	30^{th} TBI Winterseminar *Bled, Slovenia*	**Talk**
11/2014	"Fight against Ebola – *in silico*" (Workshop) *Jena, Germany*	**Attendance**
10/2014	12^{th} Herbstseminar der Bioinformatik *Doubice, Czech Republic*	**Talk**

Ehrenwörtliche Erklärung

Hiermit erkläre ich

- dass mir die Promotionsordnung der Fakultät bekannt ist,
- dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigene Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe,
- dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe.

Bei der Auswahl und Auswertung des Materials sowie Herstellung des Manuskripts haben mich folgende Personen unterstützt:

Manuela Marz

Ich habe weder die gleiche, noch eine ähnliche oder eine andere Arbeit an einer Hochschule als Dissertation eingereicht.

Jena,

Emanuel Barth