Age-dependent processes and molecular pathways of the fungal ageing model *Podospora anserina*: A bioinformatics approach

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

Vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Oliver Philipp
aus Siegen

Frankfurt (2017)
(D30)

Vom Fachbereich Informatik und Mathematik der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Andreas Bernig

Gutachter: Prof. Dr. Ina Koch und Prof. Dr. Heinz D. Osiewacz

Datum der Disputation: 14.11.2017

"Aging is mostly the failure to repair."
Gregory Benford

Zusammenfassung

Biologisches Altern ist ein degenerativer und unumkehrbarer Prozess, der schlussendlich zum Tod des Organismus führt. Der komplexe Vorgang des Alterns wird sowohl von Umweltbedingungen als auch von genetischen Merkmalen beeinflusst und unterliegt darüber hinaus verschiedenen stochastischen Faktoren. Obwohl sich bereits viele Theorien innerhalb der letzten Jahrzehnte etabliert haben, kann keine dieser Theorien die komplexen Mechanismen des Alterns vollständig erklären.

Generell führen sowohl biologische Prozesse als auch Umweltfaktoren zu molekularen Schäden und zur Anhäufung von beeinträchtigten, zellulären Komponenten. Dem gegenüber stehen zelluläre Überwachungsmechanismen und -systeme, die Reparatur, Remodellierung oder Degradation von beschädigten bzw. beeinträchtigten Komponenten umfassen. Nichtsdestotrotz sind diese Mechanismen ab einem bestimmten Zeitpunkt nicht mehr länger wirksam, da die zunehmende Menge der molekularen Schäden nicht mehr effizient beseitigt werden können oder weil die Reparatur- und Überwachungsmechanismen selbst durch diverse schädigende Effekte beeinträchtigt werden. Der Organismus verfällt und stirbt schließlich. Um das komplexe Wechselspiel von Verfall und Aufrechterhaltung zu untersuchen und zu verstehen, bedarf es holistischer und systembiologischer Analysen. Deshalb wurde der Prozess, der zum Altern im pilzlichen Modellorganismus *Podospora anserina* führt, mit Hilfe von bioinformatischen Methoden untersucht. Im Vergleich zu vielen anderen Alterungsmodellen zeichnet sich *P. anserina* insbesondere durch seine kurze Lebensspanne, seine gute Zugänglichkeit für molekulare und genetische Manipulationen als auch durch seine geringere, biochemische Komplexität aus.

Um einen generellen Überblick über die verschiedenen Prozesse zu erhalten, die während des Alterns von *P. anserina* eine wichtige Rolle spielen, wurde zunächst eine umfassende Untersuchung durchgeführt die zum Ziel hatte, Gene aufzufinden die alternsrelevant reguliert und exprimiert werden. Diese Untersuchung basierte auf den Daten einer alternsabhängigen Transkriptomanalyse. Die hier angewandten, umfassenden Analysen wiesen auf verschiedene alternsrelevante Stoffwechselwege hin und zeigten auf, dass insbesondere die Autophagie eine wesentliche Rolle im Alterungsprozess spielt. So zeigte sich, dass die Expression von autophagie-assoziierten Genen im Verlauf des Alterns ansteigt.

Um anschließend die Autophagie zu charakterisieren und die mit ihr assoziierten Komponenten und deren Interaktionen untereinander zu untersuchen wurde

PATH2PPI, ein neues bioinformatisches Softwarepaket, entwickelt und implementiert. PATH2PPI ermöglicht die Vorhersage von Protein-Protein Interaktionsnetzwerken von bestimmten Stoffwechselwegen, basierend auf dem Vergleich von Sequenz-Homologien. Das Paket wurde verwendet, um das Autophagie-Interaktionsnetzwerk von *P. anserina* zu erstellen.

Anschließend wurde das vorhergesagte Netzwerk um experimentelle Daten erweitert. Diese Daten umfassen zum einen die Ergebnisse aus der Transkriptomanalyse und zum anderen Protein-Protein-Interkationen, die mit Hilfe einer Hefe-Zwei-Hybrid Analyse experimentell ermittelt wurden. Um die biologische Signifikanz des erstellten Autophagie-Netzwerks zu untermauern, wurden dessen topologischen Eigenschaften untersucht und mit Hilfe verschiedener mathematischer und statistischer Methoden mit den Eigenschaften von zusätzlich generierten Zufallsnetzwerken verglichen. Weiterhin konnten, basierend auf dieser topologischen und funktionalen Analyse, die für die Autophagie wichtigsten Proteine ermittelt und funktionale Module identifiziert werden, die den verschiedenen Phasen der Autophagie entsprechen. Aufgrund der integrierten Transkriptomdaten wurde das Autophagie-Netzwerk mit den Prozessen des Alterns in Beziehung gebracht. So konnten Proteine identifiziert werden, deren Gene im Verlauf des Alterns kontinuierlich hoch- bzw. herunter reguliert werden und es konnte erstmalig gezeigt werden, dass autophagie-assoziierte Gene im Verlauf des Alterns signifikant häufig co-exprimiert werden.

Insgesamt stellt das vorgestellte biologische Netzwerk eine systembiologische Sicht auf Autophagie dar und ermöglicht weitere darauf aufbauende Studien die darauf abzielen, die Zusammenhänge von Autophagie und Altern näher zu untersuchen. Darüber hinaus bietet es Ansätze für weitere Untersuchungen zu potentiellen Methoden der Intervention in den Alterungsprozess und damit zur Verlängerung der gesunden Lebenspanne sowohl von *P. anserina* als auch von anderen eukaryotischen Organismen, insbesondere des Menschen.

Abstract

Biological ageing is a degenerative and irreversible process, ultimately leading to death of the organism. The process is complex and under the control of genetic, environmental and stochastic traits. Although many theories have been established during the last decades, none of these are able to fully describe the complex mechanisms, which lead to ageing. Generally, biological processes and environmental factors lead to molecular damage and an accumulation of impaired cellular components. In contrast, counteracting surveillance systems are effective, including repair, remodelling and degradation of damaged or impaired components, respectively. Nevertheless, at some point these systems are no longer effective, either because the increasing amount of molecular damages can not longer be removed efficiently or because the repairing and removing mechanisms themselves become affected by impairing effects. The organism finally declines and dies. To investigate and to understand these counteracting mechanisms and the complex interplay of decline and maintenance, holistic and systems biological investigations are required. Hence, the processes which lead to ageing in the fungal model organism *Podospora anserina*, had been analysed using different advanced bioinformatics methods. In contrast to many other ageing models, *P. anserina* exhibits a short lifespan, a less biochemical complexity and it provides a good accessibility for genetic manipulations.

To achieve a general overview on the different biochemical processes, which are affected during ageing in *P. anserina*, an initial comprehensive investigation was applied, which aimed to reveal genes significantly regulated and expressed in an age-dependent manner. This investigation was based on an age-dependent transcriptome analysis. Sophisticated and comprehensive analyses revealed different age-related pathways and indicated that especially autophagy may play a crucial role during ageing. For example, it was found that the expression of autophagy-associated genes increases in the course of ageing.

Subsequently, to investigate and to characterise the autophagy pathway, its associated single components and their interactions, PATH2PPI, a new bioinformatics approach, was developed. PATH2PPI enables the prediction of protein-protein interaction networks of particular pathways by means of a homology comparison approach and was applied to construct the protein-protein interaction network of autophagy in *P. anserina*.

The predicted network was extended by experimental data, comprising the transcriptome data as well as newly generated protein-protein interaction data achieved from a yeast two-hybrid analysis. Using different mathematical and statistical methods the topological properties of the constructed network had been compared with those of randomly generated networks to approve its biological significance. In addition, based on this topological and functional analysis, the most important proteins were determined and functional modules were identified, which correspond to the different sub-pathways of autophagy. Due to the integrated transcriptome data the autophagy network could be linked to the ageing process. For example, different proteins had been identified, which genes are continuously up- or down-regulated during ageing and it was shown for the first time that autophagy-associated genes are significantly often co-expressed during ageing.

The presented biological network provides a systems biological view on autophagy and enables further studies, which aim to analyse the relationship of autophagy and ageing. Furthermore, it allows the investigation of potential methods for intervention into the ageing process and to extend the healthy lifespan of *P. anserina* as well as of other eukaryotic organisms, in particular humans.

Scientific contributions

Peer-reviewed publications

First authorships

2013 1. Dez. 20, 2013. "A Genome-Wide Longitudinal Transcriptome Analysis of the aging Model *Podospora anserina*". In: *PLoS One* **8**(12), e83109

2016 2. Feb. 04, 2016. "PATH2PPI: an R package to predict protein-protein interaction networks for a set of proteins". In: *Bioinformatics* **32**(9), 1427-1429

2017 3. Mar. 27, 2017. "The autophagy interaction network of the aging model *Podospora anserina*". In: *BMC Bioinformatics* **18**(196)

Co-authorships

2016 1. Feb. 16, 2016 (with Carina R Guevara et al.). "Global protein oxidation profiling suggests efficient mitochondrial proteome homeostasis during aging". In: *Molecular & Cellular Proteomics* **15**(5), 1692-1709

Poster

2012 1. Nov. 23, 2012. "Age-dependent transciptome analysis of genes associated with mitochondria and the respiratory chain in *Podospora anserina*". In: Jahrestagung der Deutschen Gesellschaft für Alternsforschung. Hofheim (Taunus).

2013 2. Sep. 10, 2013. "A longitudinal transcriptome analysis of a fungal aging model indicates that autophagy compensates age-dependent proteasomal impairments". In: German Conference on Bioinformatics. Göttingen.

3. Sep. 29, 2013. "An age-dependent transcriptome analysis of *Podospora anserina* identifies autophagy as a potential mechanism to compensate age-related proteasomal impairments". In: XI International Fungal Biology Conference. Karlsruhe.

2014 4. Sep. 28, 2014. "The autophagy interaction network of the fungal aging model *Podospora anserina*". In: German Conference on Bioinformatics. Bielefeld.

Conference talks

2012 1. Sep. 19, 2012. "Co-expressed genes of the energy metabolism during ageing of *Podospora anserina*". In: German Conference on Bioinformatics, Systems Biology of ageing (Workshop). Göttingen.

Reviewer activities

2013 1. Aug. 2, 2013, "Active Reviewer". In: *The Pharmacogenomics Journal*.

2015 2. Apr. 28, 2015, "Active Reviewer". In: *BMC Bioinformatics*.

2016 3. Jan. 4, 2016, "Active Reviewer". In: *BMC Bioinformatics*.

Contents

List of illustrations

Figures

Tables

List of Abbreviations

ATG	autophagy-related gene
BLAST	Basic Local Alignment Search Tool
BP	Biological process
CC	Cellular component
cDNA	complementary DNA
COG	Clusters of orthologous groups
DNA	Deoxyribonucleic acid
GFP	Green fluorescent protein
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
MF	Molecular function
MS	Mass spectrometry
mRNA	messenger RNA
NGS	Next generation sequencing
ORF	Open reading frame
Pcc	Pearson correlation coefficient
PN	Petri net
PPI	Protein-protein interaction
QC	quality control
qRT-PCR	Quantitative Real-time PCR
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SAGE	Serial Analysis of Gene Expression
TOR	Target of Rapamycin
tpm	tags per million
UPS	Ubiquitin-proteasome system

Chapter 1

Introduction

1.1 Motivation

Most eukaryotic organisms are affected by the phenomenon of biological ageing, which leads to an increase of cellular impairments, the general decline of a biological system, to severe diseases and finally to death (Kirkwood and Austad, 2000). Ageing is a complex process, which is influenced by several environmental as well as cellular conditions (Bowen and Atwood, 2004; Dillin et al., 2014).

The scientific field of ageing research aims to generally identify and understand the processes, which lead to the symptoms of ageing and which occur during ageing amongst the different affected organisms. One major objective of this scientific field is a comprehensive understanding of ageing which may allow the investigation of potential methods of intervention into the ageing process, to reduce age-related impairments and to extend the healthy lifespan.

During the last decades, many theories have been established, which describe different parts of the underlying biological processes. These theories comprise genetic, proteomic as well as biochemical aspects of ageing (Rattan, 2006; Jin, 2010). Nevertheless, to date none of these theories can sufficiently explain the entire process of ageing and many aspects are still unknown and require further investigations.

Such investigations strongly dependent on experiments in model organisms, since studies on humans can only be based on comparative and correlative data due to ethical constraints. Ideally, model organisms in ageing research have a short lifespan and are accessible for genetical and biochemical manipulations. Already some decades ago model organisms have been used to unravel the underlying mechanisms of ageing (Rizet, 1953). During these early days in ageing research, the prevailing opinion was that ageing can be attributed to some genes. Hence, many early theories and studies aimed to identify and to modify single targets in such model organisms in order to affect ageing. Nevertheless, the increasing amount of different genes and biological pathways which have been identified to have an impact on the lifespan of the different organisms, evinced that ageing is affected by many different

complex mechanisms and molecular pathways. It became obvious, that these reductionistic approaches, i.e. the modification of single target genes, will not be able to sufficiently explain ageing. System-oriented perspectives are required (Vaiserman, 2014). Furthermore, it was found, that indeed, each biological system has specific and non-conserved processes of ageing, but basal mechanisms exist which are evolutionary conserved amongst organisms (Martin et al., 1996; Partridge and Gems, 2002).

Investigations and systems-biologics approaches which focus on these conserved age-dependent processes will lead to new insights and may help to unravel the complex mechanisms of age-dependent processes. Such investigations benefit from model organisms which exhibit the same fundamental biochemical processes as higher eukaryotes, e.g. humans, but provide the advantages of a less complex biological system.

The filamentous fungus *Podospora anserina* is such a model organism (Scheckhuber and Osiewacz, 2008). It exhibits different age-related characteristics in development and physiology and has many biochemical processes in common with most eukaryotic organisms. Most advantageously and in contrast to other ageing models, it has a well-defined limited lifespan of about a few weeks, until it finally dies. For many years several studies have been published, which successfully used this short-lived ageing model to investigate the various processes affecting organismal ageing (Fischer et al., 2012; Knuppertz and Osiewacz, 2016). For example, previous works have demonstrated the impact of reactive oxygen species and the role of cellular respiration as well as copper homoeostasis and mitochondrial dynamics during ageing of *P. anserina*. Furthermore, some of these investigated processes were first identified in *P. anserina* and subsequently found to play a general role in ageing of other organisms. A first extensive bioinformatics analysis was applied on *P. anserina*, which especially focused on the relationship of ageing with the energy metabolism and the components, which are associated with the respiratory chain (Philipp, 2012). More specifically, it was measured and analysed which genes of the energy metabolism and the respiratory chain are significantly regulated and expressed in the course of ageing.

Notwithstanding, further comprehensive and in particular, unbiased investigations and analyses, i.e with no presumptions, are required to combine and extend the current knowledge, to compile existing data and to reveal new aspects in ageing research. Such analyses can largely profit from theoretical studies and advanced bioinformatics approaches in combination with experimentally accessible model systems like *P. anserina*.

1.2 Project aims

The general objective of this work was to investigate and to characterise age-dependent processes in the fungal ageing model *P. anserina* using advanced bioinformatics approaches. Basically, the present work can be divided into three major parts (see figure 1.1):

1. **Data mining**: Initially, the data of two comprehensive experimental approaches had been analysed. One of these approaches aimed to reveal the effects of reactive oxygen species (ROS) on the mitochondrial proteome of *P. anserina* in the course of ageing. A new biochemical workflow and a particular mass spectrometry approach (iTRAQ) was applied to achieve a comprehensive set of data about modified and/or oxidised proteins (Guevara et al., 2016). To analyse these data, the implementation of a new bioinformatics pipeline was required. The second experimental approach was an age-dependent transcriptome analysis, where the data had already been partially analysed with a special focus on genes associated with the energy metabolism and the respiratory chain (Philipp, 2012). In contrast, in the present work the entire transcriptome data were analysed to investigate biological pathways which are significantly affected by or associated with ageing (Philipp et al., 2013).

2. **Method development and implementation**: The latter study revealed several genes associated with different biological pathways which appeared to be age-related. To identify the proteins and their interactions which are associated with these pathways, a new bioinformatics approach and software tool, Path2PPI, was developed, implemented and introduced (Philipp et al., 2016). It enables the prediction of protein-protein interaction networks applying a homology comparison approach. Path2PPI contributed to the Bioconductor project and its reliability and advantages were demonstrated.

3. **Applying the approach and pathway characterisation**: One of the most promising findings of the transcriptome study was the evidence that the autophagy pathway is significantly involved in ageing processes in *P. anserina*. Hence, based on the findings of the transcriptome study and the subsequently developed prediction approach, during the third part of this work it was attempted to identify and characterise the autophagy pathway in *P. anserina* (Philipp et al., 2017). Path2PPI was applied to construct a biological network comprising the single components of autophagy and their interactions. In addition, newly generated experimental data as well as the transcriptome data were integrated into the network. By means of different mathematical approaches it was shown that the network exhibits biological meaningful properties.

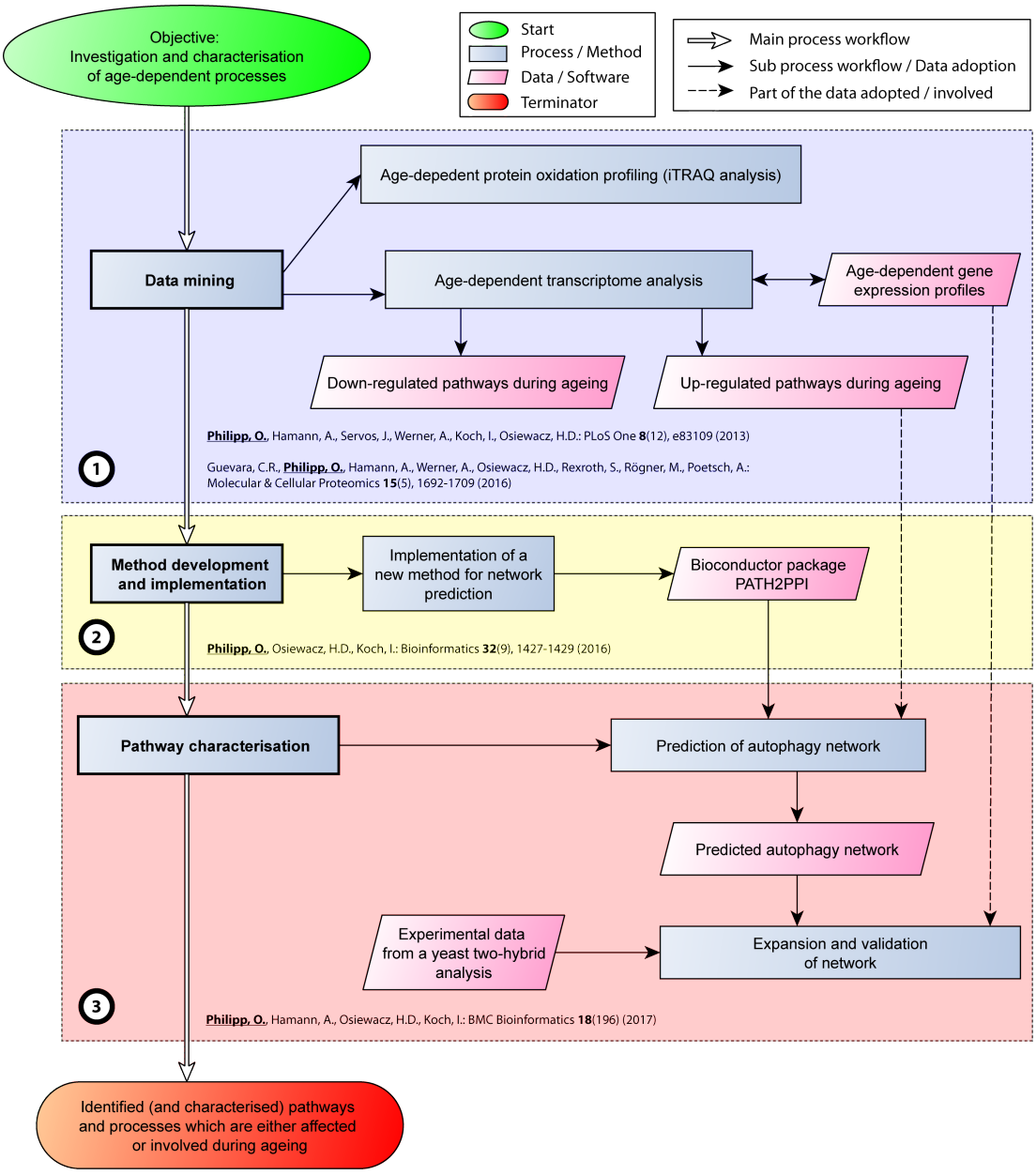


Figure 1.1: Flowchart depicting general workflow and bioinformatics pipeline. The present work can be divided into three consecutive main project parts (indicated by differently coloured backgrounds), where each part consists of different sub-processes. Each main part resulted in several findings, conclusions and achieved data sets which influenced subsequent parts. The approaches and findings of each part where published in peer reviewed journals (see references at the bottom of each coloured background box).

Chapter 2

Biological Background

This chapter gives a brief overview on the biological background of this work and the current state of research. It comprises a general description of the ageing process and the different research fields within this scientific area. Furthermore, it gives a short introduction on the fungal model organism *P. anserina*. Subsequently the biological background of the different methods of investigation which have been performed during this work are provided.

Additional and more detailed information about the different topics can be found within the accordingly referenced sources.

2.1 Biological ageing

Due to the complexity of ageing and the diversity of the biological systems which are affected, it is even a challenging task to find a clear definition for this process. According to Kirkwood and Austad (2000) ageing is accompanied by the continuously decline of the physiological functionality. The organisms fertility decreases, diseases occur with increasing frequency (morbidity), and finally mortality increases. On the one hand, in ageing research, theories exists which aim to describe the evolutionary background or reasoning of ageing, respectively (Kirkwood, 2005; Rose et al., 2008). These theories try to describe why ageing has evolved during evolution and for example what are the advantages for a population if its individuals age and die. On the other hand, other theories intend to understand and investigate the underlying biological mechanisms which lead to ageing. The present work deals with the latter aspect and aims to identify biological components, processes and pathways which lead to ageing or are affected by ageing in *P. anserina* as well as in other organisms.

It is clear that ageing is affected by and has an influence on different biological levels, e.g. the genetic or proteomic level. Furthermore, different biochemical mechanisms, may interact or counteract during ageing. Although various age-related theories have been proposed, these underlying complex mechanisms and their interactions still have to be unravelled. Nevertheless, as reviewed in Jin (2010), the

current concepts which aim to describe the ageing process fall into two main categories:

- First, the theories which state that ageing is due to programmed effects and follows a biological schedule. That means some of these mechanisms may already be encoded in the genome and changes in gene expression will alter the physiological constitution of the organism as well as the efficiency of potential, counteracting repairing and defense mechanisms. For example, one of these theories which describe programmed factors states that the length of the telomeres is associated with ageing. It is suggested that these regions at the end of the chromosomes represent a "molecular clock" while they become shorter with every division of the cell (Mather et al., 2011).

- Second, the damage or error theories, which postulate that, e.g. cellular, genetic or physiological depletions, impairments and damages accumulate in the course of ageing, that finally lead to the death of the organism. For example, one popular theory is the free radical theory of ageing which proposes that damage induced by ROS accumulate during ageing (Harman, 1956; Pole et al., 2016). Since the majority of ROS are inevitable side products of the energy metabolism, basically all organisms are affected by the phenomenon of ageing which are dependent on an energy transduction cascade, e.g. the oxidative phosphorylation.

As aforementioned, most organisms have evolved complex cellular quality control systems comprising different counteracting repairing and defense mechanisms which are able to either repair or to degrade impaired components. For example, in early studies it has been found that calorie restriction can extend the lifespan in a multiplicity of different model organisms (Masoro, 2002). Even if the underlying life extending effect has not yet been clarified, in subsequent investigations it was found that nutrient depletion induces different stress-related biochemical processes and pathways (see also section 2.3).

Notwithstanding, at a certain time point these counteracting mechanisms are no longer, or much less, effective and the organism declines and finally dies. Hence, a more sophisticated understanding of the various quality assurance mechanisms may be as important as investigations into the processes which lead to ageing. Even if ageing persists to be inevitable and irreversible, investigations into counteracting mechanisms may allow the intervention into the ageing process and to prolong the healthy lifespan of an organism.

2.2 *Podospora anserina*: A fungal ageing model

The ascomycete *P. anserina* has already been used 60 years ago as a model organism in experimental gerontology (Rizet, 1953) and has established as ageing model until today. Generally, this is due to two major advantages.

First and in contrast to many other ageing organisms, *P. anserina* exhibits a relatively short lifespan of only a few weeks. Phenotypical changes during this ageing process of *P. anserina* are clearly visible and can easily be monitored. Like all other filamentous fungi *P. anserina* consists of many branching hyphae (*sing.* hypha) which build the mycelium of the fungi (figure 2.1a). During growth, basically only the tips of the hyphae divide and expand this cellular network. Finally, at the end of the lifespan, the apical cells stop dividing and the phenotypical characteristics become visible. In contrast to a juvenile or middle-aged individual (figure 2.1b left), the senescent organism shows a darker pigmentation and its mycelium becomes more rough and irregular (figure 2.1b right). Subsequently to this senescence phase in the life cycle of *P. anserina*, the apical cells finally die and burst (Osiewacz, 2002a).

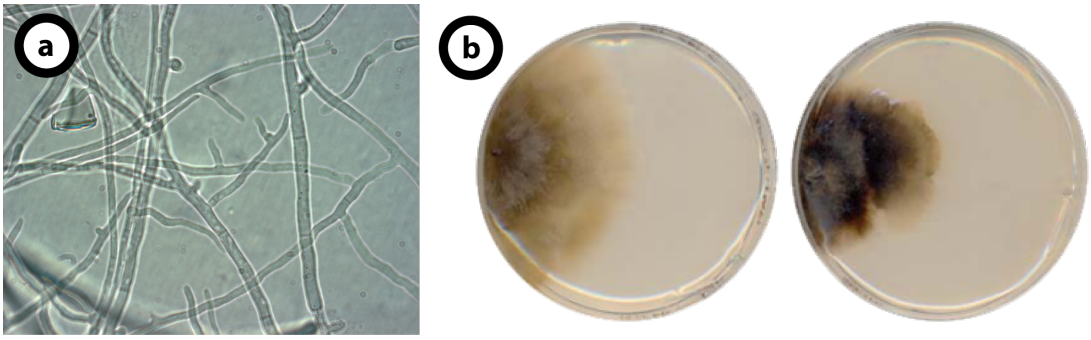


Figure 2.1: The mycelium of *P. anserina*. a) A closeup of the branching hyphae which build the mycelium of *P. anserina*. b) Shown are a middle-aged (left) and a senescent individual of *P. anserina* grown on solid medium (Scheckhuber and Osiewacz, 2008). In contrast to the middle-aged, the senescent mycelium exhibits a dark pigmentation and a rough shape.

The second major advantage of *P. anserina* as a model organism in ageing research is the good accessibility for cultivation and experimental manipulation. For many years, different aspects and theories of ageing have been experimentally investigated in this ageing model to unravel the network of biological pathways affecting organismal ageing and lifespan control. For example, previous works have demonstrated the impact of DNA stability, the generation and scavenging of ROS, respiration, copper homeostasis, proteolysis, mitochondrial dynamics, programmed cell death and autophagy in lifespan control (reviewed in: Hamann et al. (2008); Figge et al. (2012); Fischer et al. (2012); Osiewacz and Hamann (1997); Osiewacz and Borghouts (2000)). Overall, there is a strong mitochondrial etiology of ageing in *P. anserina* (Osiewacz, 2002b). During ageing massive changes in mitochondrial morphology and ultra-structure occur and, ultimately, lead to programmed cell death and the death of the individual (Brust et al., 2010; Scheckhuber et al., 2007; Daum

et al., 2013). Some of these pathways, like age-related mtDNA reorganisations, mitochondrial dynamics, and copper homeostasis were first identified in *P. anserina* and subsequently found to play a general role in the ageing of other organisms.

Furthermore, the entire genome of *P. anserina* has been sequenced and is available for comprehensive bioinformatics studies and analyses (Espagne et al., 2008). Its genome consists of $\approx$ 10.000 genes that is significantly larger than that of the yeast *Saccharomyces cerevisiae* ($\approx$ 6.000 genes) but about half of that of humans ($\approx$ 21.000 genes). Thus, compared to mammals, the coding potential already suggests a reduced complexity of functional units. One example for such a difference is programmed cell death (apoptosis). While 12 different caspases are active in humans, yeast contains one single metacaspase (YCA) and *P. anserina* the two metacaspases PaMCA1 and PaMCA2 (Strobel and Osiewacz, 2013).

Summarising, the particular strengths of the *P. anserina* system in experimental gerontology are: (1) a short lifespan, (2) simple phenotypic monitoring of the ageing process, (3) the ease of accessibility to genetic manipulation, (4) the availability of specific mutants affecting various pathways, (5) the availability of the complete genomic sequence, (6) an easy way to genetically generate double-, triple- and higher-order mutants and (7) the fact that individual pathways are controlled by fewer components than in many other systems.

2.3 Quality assurance and autophagy

As described in section 2.1 various organisms have evolved a complex system of different quality control (QC) pathways which impact on development and degeneration has been demonstrated (for some reviews see: Tatsuta and Langer (2008); Fischer et al. (2012); Korolchuk et al. (2010)). From these investigations, a cross-talk between individual pathways appears to be active.

One of these QC systems is the autophagy pathway (Greek: *auto*-, "self" and *phagein*, "to eat"). For many years it was supposed that autophagy is exclusively induced as a cellular recycling mechanism during starvation in order to ensure the proper functionality of the system. In the last decade, autophagy has evolved as a major pathway which is active in recycling not only during starvation but in molecular QC as well.

During macroautophagy, the main pathway of autophagy, molecules or organelles become enclosed by membranes and the resultant autophagosomes are subsequently delivered to lysosomes (in animals) or the vacuole (in plants and fungi) where they became enzymatically degraded (see Figure 2.2). The building blocks of the degraded components (e.g. amino acids) are subsequently reused to generate new functional components. The main molecular processes occurring during autophagy are very similar and strongly conserved among organisms. The core machinery encoded by the "autophagy-related genes" (ATG) is effective in the induction of

autophagy, vesicle-nucleation and -expansion, fusion with the lysosome or the vacuole and the final degradation of the regarding components (Kundu and Thompson, 2008; He and Klionsky, 2009; Knuppertz and Osiewacz, 2016).

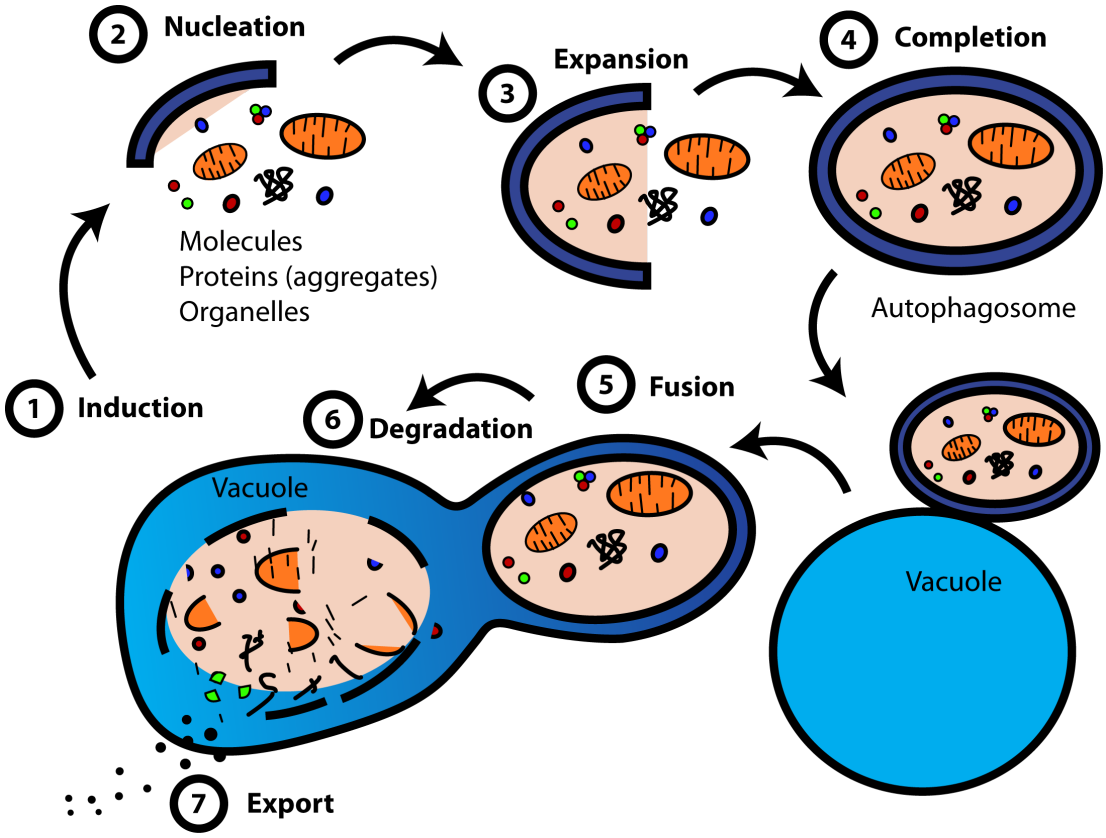


Figure 2.2: Scheme of the degradation process during macroautophagy. The degradation process of autophagy can be divided into seven distinct steps. (1) After induction either by nutrient depletion (non-selective autophagy) or by recognition of cellular components which were accordingly tagged (selective autophagy), (2) the nucleation of the autophagosome begins. (3) During expansion the components became encapsulated by the expanding autophagosome and are fully encased after (4) completion of autophagsome developing. (5) The autophagsome fuses with the vacuole (lysosome in mammals) and its content is released into the vacuole where it became enzymatically degraded into the single building blocks. (7) Finally, during export, the degraded, single building blocks are secreted back into the cytosol and are available for further biochemical and metabolic processes.

In different organisms it has been shown that there is evidence for a link of autophagy with the ageing process. For example, an impaired autophagy system shortens lifespan in mice (Pyo et al., 2013) or is responsible for various age-dependent diseases like Alzheimers, Huntingtons or Parkinsons disease (Nassif and Hetz, 2012). At the same time, an increase of autophagy by treatment with rapamycin or by overexpression of specific autophagy-related genes was shown to extend lifespan in various organisms like yeast, fruit flies or mice (Alvers et al., 2009; Bjedov et al., 2010; Pyo et al., 2013).

In *P. anserina* the first indication for a role of autophagy in ageing was obtained during this work by means of a genome-wide longitudinal transcriptome analysis

(Philipp et al., 2013). Subsequently, it was experimentally validated that autophagy is strongly related with ageing in *P. anserina* (Knuppertz et al., 2014). In that study it was found that autophagy is induced during ageing and that an impairment of autophagy shortens the lifespan of *P. anserina*. These findings evinced that autophagy is a longevity-assurance mechanism in *P. anserina*. The identification of autophagy as a longevity assurance mechanism and evidence for interactions of autophagy with other QC pathways in *P. anserina* are consistent with relationships proposed, but not mechanistically elucidated, in other systems (Lilienbaum, 2013; Löw, 2011; Park and Cuervo, 2013).

Autophagy is primarily negatively regulated by the TOR kinase (Kundu and Thompson, 2008), in contrast to the protein- and ribosome biogenesis which is positively regulated by TOR. For example, if autophagy is activated by an inactivation of TOR, the expression of ribosomal genes is inactivated and vice versa. This relationship was also found in the transcriptome data analysed in this work (Philipp et al., 2013). Furthermore, a compensatory relationship of the proteasome system and autophagy, as suggested by this data, was previously reported in other systems (Lilienbaum, 2013).

The ubiquitin-proteasome system (UPS) as part of the QC network is responsible for the degradation of impaired, misfolded or aggregated proteins. In different studies in which the UPS was found to be impaired due to experimental manipulations or various diseases, a compensation of this impairment by autophagy was observed (Pandey et al., 2007; Schreiber and Peter, 2014). Some details about the underlying "cross-talk" have been discussed but not finally experimentally demonstrated and proven in detail (Korolchuk et al., 2010).

To understand the role and the relevance of autophagy in the ageing process and how it is associated with other QC pathways it is necessary to analyse autophagy on a systems biology level rather than focusing on single components. For instance, experiments which aim to reveal how the modification of single components, like the depletion of a particular protein within a pathway, effects the system may lead to imprecise or even false conclusions, as other processes and components may have an impact on the system as well. Hence, it is crucial to unravel interactions and the dynamics of the individual components that are part of the complete QC system.

A mathematical model, which summarises in an abstract manner the most important components and interactions, provides valuable insights into autophagy and its dynamic role in ageing for further experimental evaluation and modulation. Already in Figge et al. (2012) a probabilistic modelling approach was applied to provide a mathematical model which enabled to simulate the dynamics of mitochondrial quality control during ageing in *P. anserina*. For instance, the proposed model suggested that the reduction of mitochondrial fusion-fission cycles during ageing are due to a systemic adaptation to extend the lifespan.

In contrast, the mathematical model, proposed in this work, is a network based

model of the autophagy pathway, which allows to identify, by means of various algorithmic approaches, "bottle neck" situations or concurrent processes, where a genetic manipulation of the corresponding components can influence the rate of autophagy.

2.4 From gene expression to protein interactions: Different impact on biological systems

The flow of information in a living cell is organised in a cascading manner. Generally, the genetic information (the entity of all genes) is encoded in the genome. During gene expression the genes become transcribed to RNA and finally translated to proteins. Each of these processes is regulated by a complex system of various biochemical mechanisms adapting to the prevailing environmental conditions. Furthermore, each single entity of these biological levels, i.e., genes, RNA molecules or proteins, are not independent but rather interact with each other. For example genes can become activated or deactivated, and if expressed, one single gene or its subsequent product can induce or inhibit further gene expression. The same is true on the proteomic level. Several proteins can interact with each other in different ways and induce, activate or inhibit down-stream proteins and processes (Jones and Thornton, 1996; Braun and Gingras, 2012). Such interacting and counteracting sets of proteins that are associated with or involved in similar processes are defined as biological pathways.

Basically, the present work aimed to investigate ageing on two different biological levels and to reveal the corresponding single components, interactions and important biological pathways:

- On the level of gene expression, it should be investigated which genes are regulated and expressed in an age-dependent manner. It was of interest to identify the genes which exhibit a similar degree of expression during ageing indicating a common underlying regulatory program. Furthermore, since genes are associated with one or more biological pathways, appropriate bioinformatics analyses were performed to reveal which pathways are associated with ageing, at least on a genetic level.

- Proteins are the major functional units of the cell which perform a broad range of various biochemical tasks within a biological pathway. To gain a more comprehensive picture of a pathway it is of particular importance to know the involved proteins and how they are interacting. Hence, the present work investigated ageing on the proteomic level as well and aimed to reveal the different protein-protein interactions (PPI) of age-related pathways. Different types of PPI exist and can be distinguished depending on the particular

biochemical process. For example, two or more proteins can build a larger protein complex which comprise different specialised subunits. Further examples of typical PPIs can be found during signal transduction (Krauss, 2006) or during ubiquitination where particular target proteins are tagged with ubiquitine (Pickart, 2001).

Obviously the aforementioned biological levels are interrelated and must not be considered separately. Hence, the age-related investigations which have been done during this work on different biological levels are built on each other. The present thesis aims to summarise the results and to bring them into an holistic context.

2.5 Next generation sequencing

Next generation sequencing (NGS) methods enable the reading of a large amount of sequence information of different types of DNA as well as RNA molecules with a reasonable effort (for a current review see van Dijk et al. (2014)). Most advantageously, these methods allow different approaches of comprehensive gene expression analyses. Basically, by means of gene expression analyses it is possible to measure which genes are activated and / or regulated under particular conditions. For example, in this work the age-dependent transcriptomes of *P. anserina* had been analysed where the different age-stages represent the different conditions.

Generally, it is possible to distinguish two types of gene expression analyses, the array-based and the tag- or fragment-based approaches, respectively. Array-based approaches are based on DNA microarrays which contains as much as possible or required predefined cDNA sequences which are complementary to the sequence-molecules in the corresponding samples. If a molecule binds or hybridises to such a complementary sequence on the chip it can be detected or counted, respectively (Schulze and Downward, 2001; Allison et al., 2006; Guindalini and Pellegrino, 2016). The major advantages of array-based approaches are the relatively low costs and the reproducibility. Nevertheless, since the sequences have to be predefined and the space on a microarray is limited, the most obvious disadvantage is that low abundant or even unknown molecules may be missed and are not detectable. Indeed, modern chip technologies partly overcame these drawbacks. Notwithstanding, tag-based methods do not have such problems. In contrast, the sequence-molecules in each sample become biochemically fragmented first. Next, the fragments, called sequence-tags, become sequenced by a high-throughput approach, subsequently counted and, if applicable, identified and mapped to a reference genome. The latter enables the detection of (known) genes which are expressed under particular conditions.

Amongst other things, in this work gene expression data generated by a tag-based approach was analysed. The applied method will be briefly described in the following section.

2.5.1 SuperSAGE analysis

Serial Analysis of Gene Expression (SAGE) is a tag-based approach to detect and identify mRNA-molecules within a sample in a qualitative as well as quantitative manner (Velculescu et al., 1995). Basically, SAGE produces short but representative sequence-fragments of each mRNA-molecule in a sample; the sequence-tags. Subsequently, each tag gets sequenced and aligned to the reference genome, i.e., the longer the sequence-tags the more accurate and unambiguous becomes the mapping. Hence, the quality of the different SAGE approaches depends on the length of the produced tags.

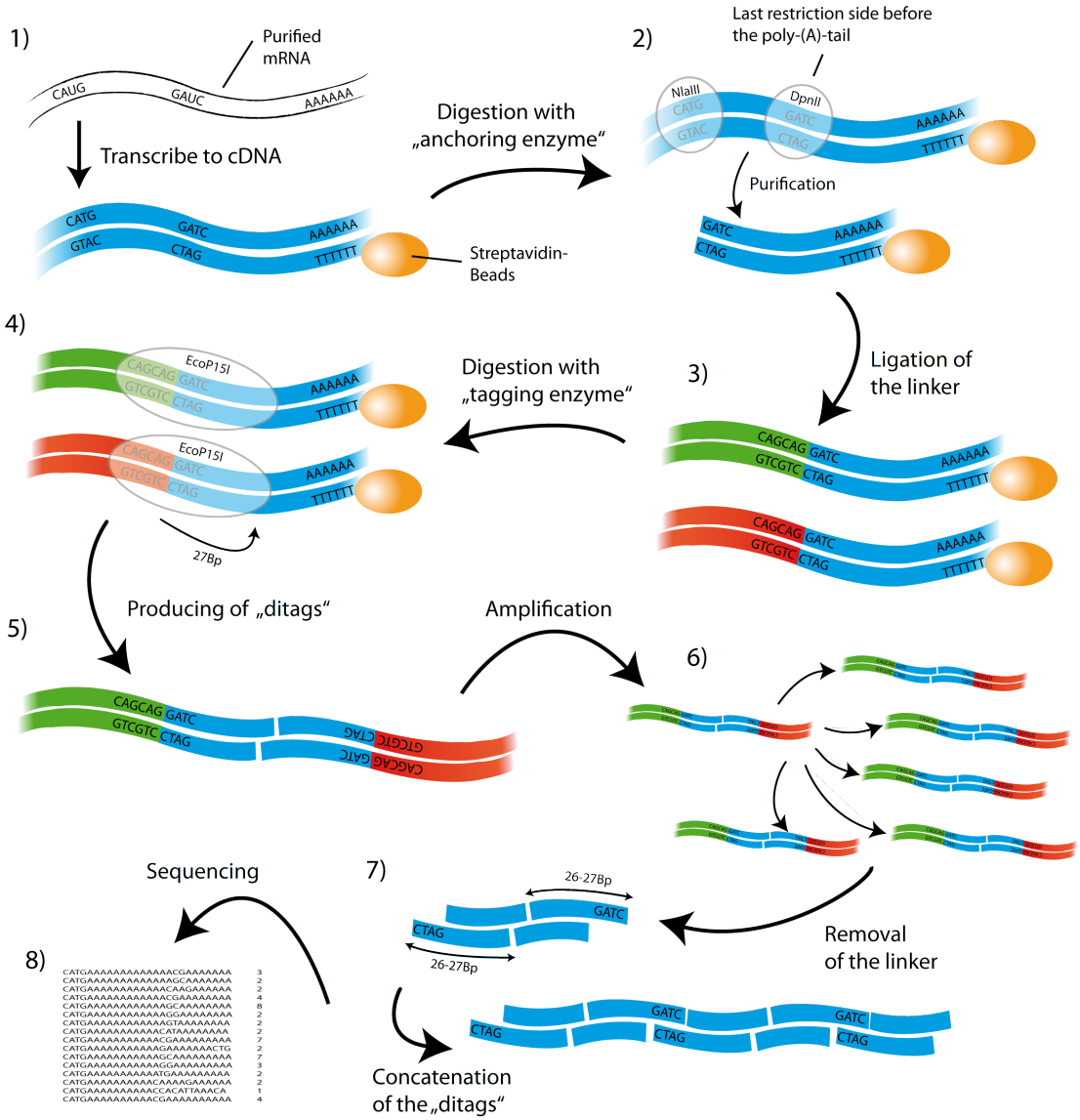


Figure 2.3: Workflow of the SuperSAGE analysis (modified from *Oliver Philipp, 2012, Diploma thesis*). The single mRNA-transcripts become digested by enzymes in order to achieve short but representative sequence tags. These tags finally get sequenced by a high-throughput approach (e.g. illumina sequencing). The details of the workflow are described in the text.

In contrast to former SAGE approaches SuperSAGE produces sequence-tags of 26 bases length which leads to a nearly unambiguous assignment to the reference

genome (Matsumura et al., 2003; Molina et al., 2008). The biochemical workflow of the SuperSAGE approach works as followed (see figure 2.3):

1. The purified mRNA molecules become transcribed into a double stranded cDNA and attached to a carrier material (Streptavidin beads).
2. The cDNA molecules become fragmented by two restriction enzymes which especially digest high frequent base sequences. After purification only the beads with the attached cDNA fragments remain.
3. Specific cDNA-linker, are ligated to the cDNA fragments. These linker contain different base sequences each (green and red parts in Figure 2.3) which act as primer sequences for the amplification process in step six.
4. The new ligated cDNA-molecules are now digested with the "tagging enzyme" which cuts the cDNA at a distance of 27 base pairs from the recognition sequence in the previously attached linker.
5. "Ditags" are produced by combining each of the digested cDNA fragments.
6. An amplification of the "ditags" ensures a sufficient amount of RNA data.
7. The linker is removed and a concatenation of the "ditags" is applied.
8. The "ditags" become sequenced by a high-throughput approach. Each tag is mapped to the reference genome in order to obtain absolute abundances of each detected mRNA molecule.

Subsequently, an appropriate bioinformatics pipeline is required to analyse the achieved data (see next chapter).

Chapter 3

Bioinformatical Background

A vast amount of available as well as newly generated biological data makes it necessary to apply advanced computer-aided methods to analyse these data. Various bioinformatics approaches have been developed and introduced during the last years dedicated to a diverse landscape of biological data. Nevertheless, specific biological problems require appropriate bioinformatics pipelines consisting of and combining different methods on the one hand. On the other hand, based on the corresponding biological problem, new bioinformatics methods have to be developed.

During this work, several types of data on different biological levels accrued, e.g. transcriptomic and proteomic data. This chapter briefly describes the different bioinformatics approaches, tools and methods which had been applied and developed in the course of this work. Beyond these descriptions further information is available in the corresponding single publications.

3.1 Gene expression analysis

3.1.1 Data preparation for the SuperSAGE analysis

A SuperSAGE analysis was applied to achieve gene expression data of *P. anserina* (see also section 2.5.1). As a result of a SuperSAGE analysis, for each sample or condition of the study, respectively, a comprehensive data set was generated consisting of the absolute abundances of each mRNA molecule. The entity of all mRNA molecules or sequence tags of a sample is defined as the library, and the total amount of detectable molecules as the library size. For example, in this work the transcriptome data of seven age-stages had been analysed resulting in seven distinct tag libraries of different sizes. The different sizes are due to biological or experimental variations. To compare the gene expression, i.e. the changes in the abundances of the corresponding mRNA molecules, the different libraries had to be normalised. For this purpose, each library had been normalised to one million tags, i.e. the unity of the expression strength of each gene on each age-stage is defined as

"tags per million" (tpm):

$$tpm = \frac{n * 10^6}{N}, \tag{3.1}$$

where n is the absolute abundance of the mRNA molecule and N the total amount of the corresponding library.

3.1.2 Statistical approaches to analyse gene expression data

Based on the design of the study and the initial question, different statistical and bioinformatics approaches can by applied to the gene expression data. Basically, two major study designs can be distinguished: First, the comparative analyses, where the differences of two (or more) conditions are compared directly, e.g. to investigate the influence of a particular drug or a gene mutation to an organism's gene expression. That means, the transcriptome of an untreated (or wild type) sample is compared with the transcriptome of the treated (or mutated) sample (see figure 3.1 left). Second, time-series or longitudinal studies, where the changes in gene expression are compared over a range of different conditions or time points (see figure 3.1 right).

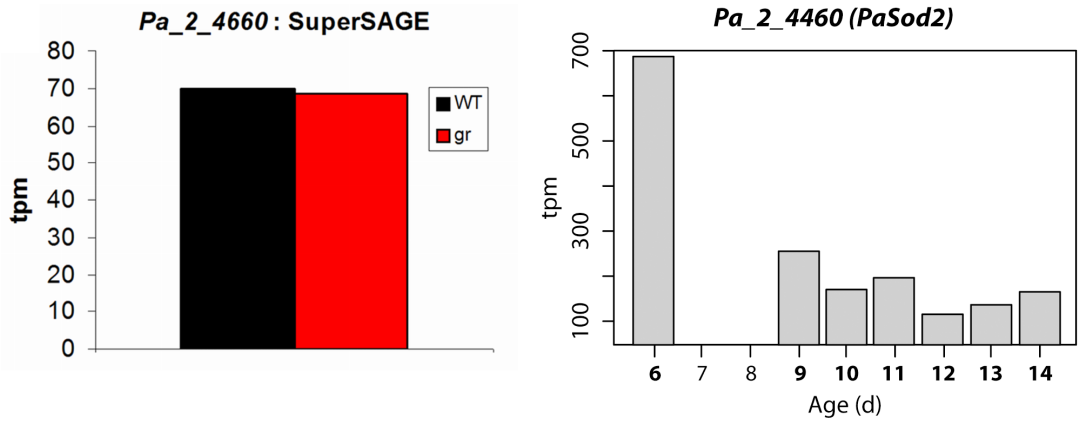


Figure 3.1: Example of a comparative and a longitudinal transcriptome analyses. The left histogram gives the expression strength of the gene coding for the superoxid dismutase 2 (*PaSod2*) in the wild type strain of *P. anserina* compared with the expression strength of the same gene in the grisea mutant (modified from Servos et al. (2012)). In contrast, the right picture shows the expression levels of *PaSod2* at different (more than two) conditions. In this case the conditions are the subsequent age stages in the wildtype of *P. anserina* (modified from Philipp et al. (2013)).

Normalisation

At the very beginning of both approaches, the data has to be normalised to take experimental and biological fluctuation into account on the one hand (see section 3.1.1). On the other hand normalisation can be necessary due to particular metrices, e.g. the Euclidean, which may otherwise distort the data and the resulting conclusions. Furthermore, for longitudinal studies, typically not the difference in

abundances are of major importance but the process or trend of the gene expression, represented as expression profiles (see figure 3.2).

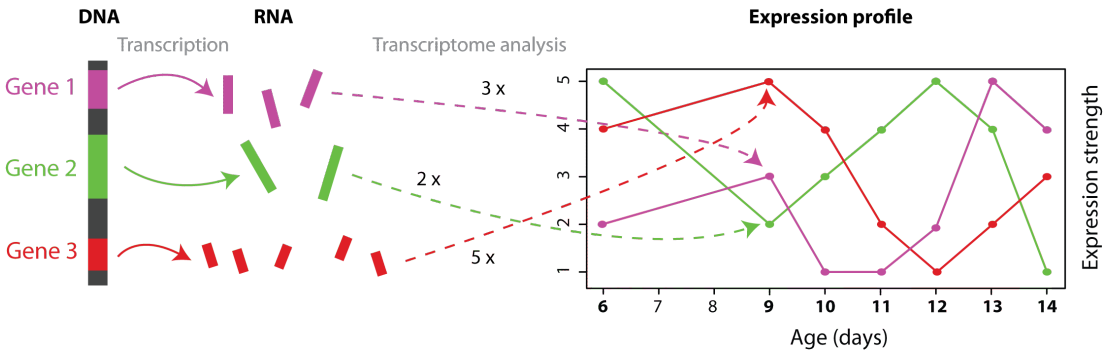


Figure 3.2: From gene expression to expression profiles. During gene expression the single genes are transcribed into the corresponding RNA molecules. A transcriptome analysis detects and counts the amount of each RNA molecule on a particular time point. The amount of each transcribed gene on each day (or condition) can be plotted into a graph. The result is a set of expression profiles, where each profile represents the change of gene expression of a particular gene during ageing.

A normalisation of these profiles and the corresponding data points to a default expression strength of 0 and a standard deviation of 1 enables the direct comparison of the gene regulation of two distinct genes independently from their actual expression strength:

$$S = \frac{X - \frac{1}{n}\sum_{i=1}^{n} x_i}{\sigma_X} \tag{3.2}$$

with the variance

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 \tag{3.3}$$

and the standard deviation

$$\sigma_X = \sqrt{Var(X)}, \tag{3.4}$$

where $X = \{x_1, x_2, ..., x_n\}$ represents the expression profile, n the number of conditions, here, the age stages, and x_i the count of the corresponding transcript in sample i.

Identification of age-related expression profiles and pathways

To reveal significant changes in gene expression and the corresponding genes, different statistical approaches exist. Either basics like a t-test or more sophisticated and appropriate analyses can be performed (Lovén et al., 2012; Conesa et al., 2016). For example, for the data analysed in this work, an approach was applied which is based on Bayesian statistics and computes p-values for each pair of conditions or age stages respectively (Audic and Claverie, 1997). Nevertheless, rather than comparing expression levels of genes in two conditions, in a time series analysis it is of major

interest to identify genes and the corresponding expression profiles, which exhibit striking and/or age-related patterns. Multivariate as well as correlative approaches are feasible to reveal such expression profiles.

During this work two major approaches had been applied to reveal age-dependent expression profiles. First, a cluster analysis was performed (Kumar and Futschik, 2007) which assorts expression profiles into groups or clusters with similar patterns, respectively. Hence, this unbiased approach only provides a pre-classification of the different expression profiles. Subsequently, age-related profile groups have to be identified manually, e.g. by searching for profile groups, where the expression strength decreases or increases in late or early age stages. Second, and in contrast to the former unbiased clustering approach, each profile had been correlated with time, using the Pearson correlation coefficient. This method only reveals genes, where the expression profiles decrease or increase continuously in the course of ageing.

Basically, clustering and correlation analyses of time series-data, only led to different groups of expression profiles, each with similar patterns. To reveal putative age-related biological pathways within these groups, it is necessary to identify significant amounts of genes which are associated with the same pathways or functions.

Different biological classification repositories exist, which aim to classify genes and their products by function, localization in the cell or their involvement in particular biological pathways. For example, the KEGG database (Kanehisa et al., 2014) provides information about different pathways and the associated genes or proteins, respectively. The COG classification system (Tatusov et al., 2000) gives a rather rough allocation than the KEGG database of proteins to different biological functions. The most comprehensive and prominent classification approach is the gene ontology (GO) classification system (The Gene Ontology Consortium, 2015) which aims to classify and describe genes and their products, independently from species and organism, by means of three different ontologies, firstly biological process (BP), secondly molecular function (MF) and thirdly, cellular component (CC). A gene or its product, respectively, can be described by one or more assigned GO terms, where each belongs to one of the three ontologies. Most importantly, each of the ontologies are hierarchically structured, i.e., the level of detail increases with decreasing hierarchical level. For example, GO terms which describe general and global processes are localised on the very top level of the "GO-tree". Often, a GO term is a parent of further child terms which correspond to a more precise description of the corresponding functionality or process. Summarizing, each term which is assigned to a gene or protein, usually represents a leaf within a GO tree with many other terms (see figure 3.3).

A GO enrichment analysis is a mathematical approach, which aims to consider each gene of a specific group, e.g. a cluster, with all assigned GO terms as well as each underlying GO tree, to statistically reveal whether a biological process or function is significantly over-represented (enriched) within such a group. In this

work, the GO classification system was applied and a GO analysis was performed to reveal biological processes or even pathways which are particularly enriched in age-dependent groups of expression profiles.

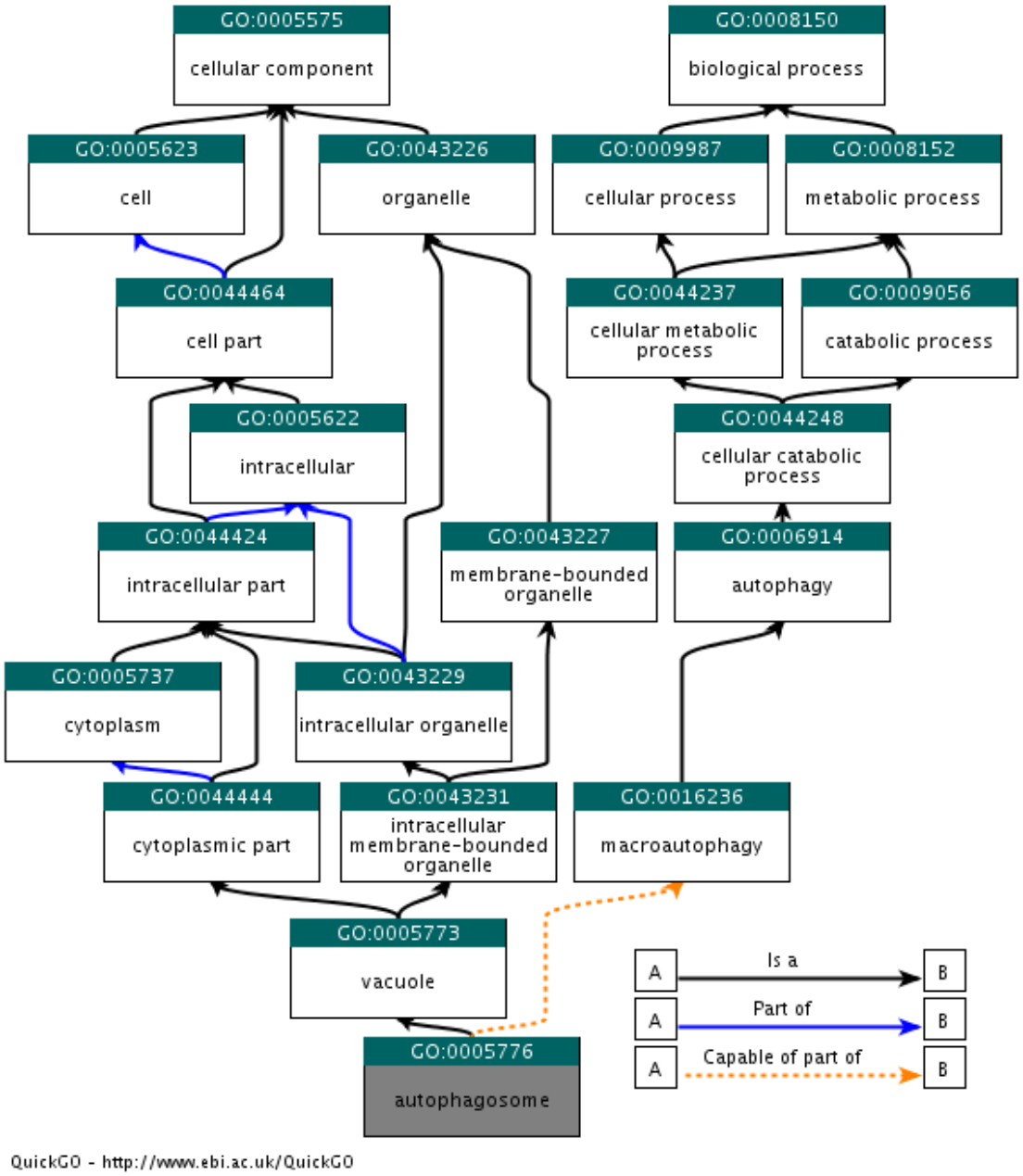


Figure 3.3: Example of the hierarchical structure of a Gene Ontology-tree (Binns et al., 2009). The GO term "autophagosome", which may have been assigned to a protein located within the autophagosomes, is a child leaf of different ancestor vertices. The more the tree reaches the root vertex the more the detail of description decreases. Each node in an ontology tree consists of an unique GO identifier and a descriptive GO term. The directed edges specify the type of relation of two nodes or GO terms, respectively.

3.2 Protein-protein interaction (PPI) networks

Protein-protein interaction (PPI) networks are graph-theoretical interpretations of proteins and their interactions, ideally within a biological pathway or process (see Figure 3.4). Each of the nodes in a PPI network corresponds to a particular protein and each edge indicates that the corresponding two proteins are interacting or are related with each other, respectively. Since the second main part of this work was about PPI network prediction, reconstruction and analysis, this section gives a brief overview on different aspects of the bioinformatics background of PPI networks.

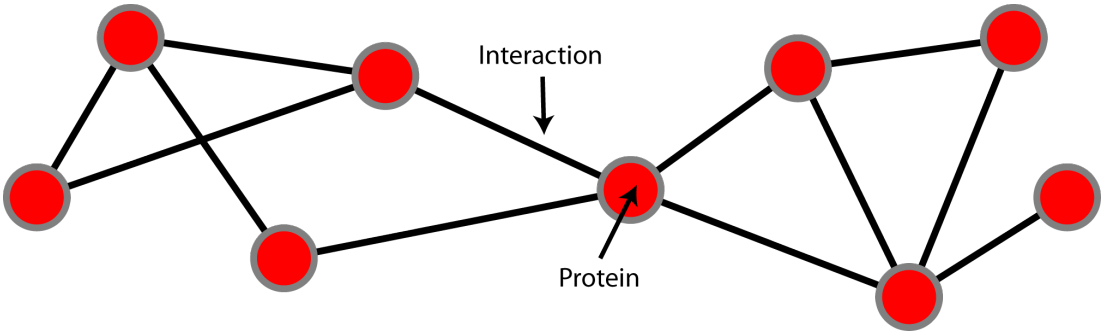


Figure 3.4: Representation and example of a PPI network. Each node represents a protein, and the edges indicate which proteins are interacting with each other.

3.2.1 Definition and representation

A PPI network defines which proteins are interacting with each other. Depending on the requirements and preconditions, different definitions for PPI networks exist. Generally, a PPI network is an undirected graph $PPI = (V, E)$ consisting of

- a set of vertices V, where each vertex $v \in V$ corresponds to a protein $v \in P$ with P as the set of proteins
- and a set of edges E, where each edge $e = \{v, u\} \in E$ indicates that the corresponding proteins v and u are interacting.

For example, if a protein v builds a complex with the two proteins u and t then all three vertices are connected by the edges $\{v, u\}$, $\{v, t\}$ and $\{u, t\}$. Further optional properties are possible depending on the requirements:

- To specify the relation or the interaction of two proteins v and u a PPI network can be a weighted graph with a scoring function $w(e)$, which defines a weight for each edge. For example, if an interaction between the two proteins v and u has been revealed by a prediction approach, then $w(v, u)$ gives the probability score for the predicted edge.
- To distinguish edge types, e.g. different types of interactions, a PPI graph can be defined as triple $PPI = (V, E, f)$ where each edge is coloured according to the colouring function f, i.e. a discrete value is assigned to each edge in PPI.

- In addition, if interactions are directed and the required information is available, then a PPI network consists of directed edges with ordered pairs of vertices (v, u), e.g. if protein v activates a protein u.

- In contrast to simple graphs, PPI networks can contain loops $\{v, v\}$ which represent self interacting proteins.

During this work different types of PPI networks had been taken into account and developed. If applicable, the corresponding definitions and specifications are described in detail in the respective location.

3.2.2 Databases and interaction repositories

Many experimental and biochemical methods exist to investigate protein-protein interactions or to reveal which proteins are related with each other (Xing et al., 2016). They differ in quality, sensitivity, effort and costs. In addition, different approaches had been established to predict PPI data (see next section). All methods have in common that they produce valuable data which may be of interest for further studies and investigations. Therefore, a broad range of different databases and repositories exist, which comprise PPI data of different sources deduced from various organisms. Generally, the current publicly available PPI data repositories differ in some major properties which are briefly described in the following accompanied by some examples. Nevertheless, many of the existing PPI databases are not exclusively limited to one of these categories but provide aspects of the others as well.

- **Species-specific** PPI data repositories provide protein interaction data of one particular organism each. Generally, these type of databases are only available for some of the most studied and established model organisms, where a broad scientific community and a comprehensive amount of data exist, e.g. for yeast (Cherry et al., 2012), fruit fly (Yu et al., 2008) or human (Prasad et al., 2009).

- In contrast, **Species-independent** PPI databases provide protein interaction data for various species and organisms. Even if for a single species only a small range of PPI is available, these databases enable the investigation and search of interactions in related organisms to extend the limited information and the transfer of knowledge amongst species. For example, the KEGG database provides information on biological pathways, which are similar or evolutionary conserved throughout different organisms (Kanehisa et al., 2014). Also the BioGRID database contains a vast amount of manually curated interaction data for a broad range of distinct organisms (Chatr-Aryamontri et al., 2015).

- Some PPI data repositories provide **predicted PPI data**. The prediction is based on data mining as well as on appropriate statistical and mathematical

approaches, which predict new interactions by transferring and combining information from different sources, e.g. experimental findings in other species (see section 3.2.3). These databases enable to obtain PPI data even for organisms, where only little experimental information and corresponding studies are available. In contrast, the major drawbacks of these databases are that they are limited to a predefined set of organisms and that they represent only a snap-shot of the latest available data. The quality and reliability strongly depends on the assumption and settings which had been made by the database provider. The most established of these prediction databases is the STRING database which provides, additionally to experimental data, predicted PPI data, based on different sources and methods, e.g. text mining, correlation studies and biological experiments (Franceschini et al., 2013).

- **Meta databases** collect and combine the available data of other PPI databases to provide a central and integrative PPI data repository. Basically, these meta databases only use experimental rather than predicted data and combine the advantages of the different species-specific and species-independent databases. For example, iRefIndex is a meta database which provides protein interaction data of the major PPI databases, e.g. BIND, BioGRID and HPRD (Razick, 2008).

3.2.3 Prediction of protein-protein interaction networks

For many organisms, only little or even nothing is known about biological pathways and protein-protein interactions. Since biological experiments are laborious and costly and mostly are feasible for only some selected proteins, computer-aided PPI prediction approaches can be applied with much less effort to reveal interacting proteins in a particular organism. Several prediction approaches and algorithms exist which take different sources into account. For example, if two proteins exhibit similar functionality or a biochemical structure, then a particular algorithm assumes that both proteins may interact. If two distinct genes show similar expression and are probably co-regulated, then this may indicate that the corresponding proteins interact as well. Other approaches are based on simple text-mining, i.e., a broad range of databases and publications are screened for protein names and identifiers. Proteins which are mentioned in the same publication or which occur in the same study may also interact under particular conditions. Some prediction approaches combine different methods to minimise the rate of false positives. Nevertheless, each of the several approaches are accompanied by different advantages and disadvantages (for a current review see Rao (2014)). The major disadvantage of the most approaches is that they require a sufficient amount of previous knowledge and species-specific information, which is often not available.

In contrast, for many organisms, the entire genomes had been sequenced or are

planned to be sequenced, since highly effective and relatively inexpensive NGS approaches are available nowadays (see also section 2.5). Along with the sequenced genomes the corresponding protein sequences become available too. Different bioinformatics approaches exist which allow the identification of the open reading frames (ORF) of such sequenced genomes and the translation of the nucleic acid to the corresponding amino acid sequences (Reeves et al., 2009). That means, in general, the sequences of most genes and proteins are known even if neither their functions nor the associated processes are yet characterised.

Sequence-based prediction approaches benefit from this broad range of available sequence information. They are based on two assumptions. Firstly, a similar protein sequence corresponds to a similar function or association with the same biological process in a related or homologous organism, respectively. Secondly, if two proteins have similar sequences to two proteins in a related organism, and the latter two proteins are known to interact, then both homologous proteins may interact as well (see Figure 3.5).

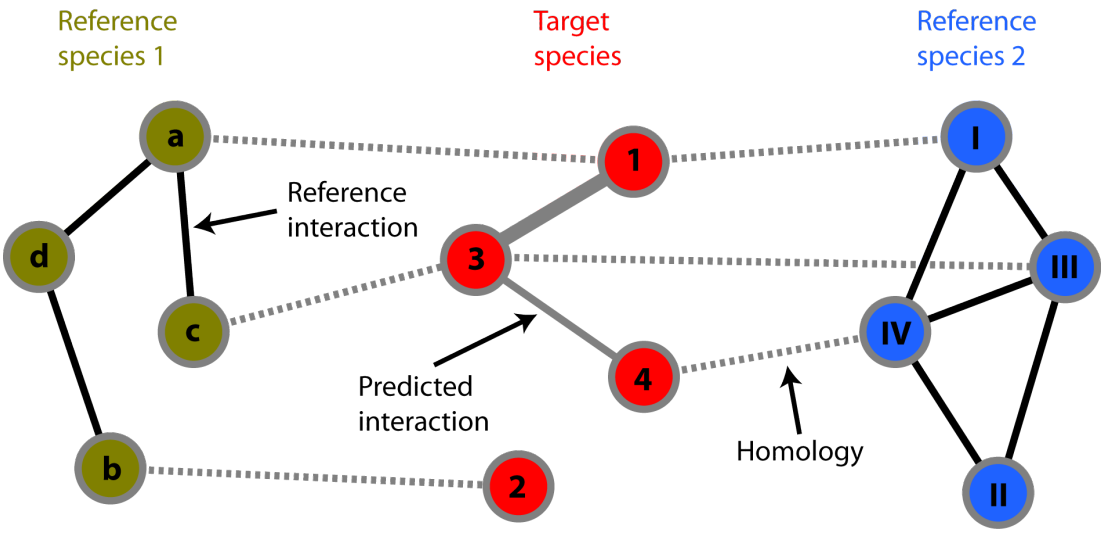


Figure 3.5: Homology-based prediction of PPI networks. Homology-based prediction of PPI networks relies on the assumption that, if two proteins are interacting in a reference species, then two homologous proteins (grey dotted edges) in the target species may interact as well (grey solid edges). PATH2PPI also considers that the probability of an interaction in the target species increases if the corresponding interaction was found in more than one reference species, indicated by a thick grey edge between node 1 and 3 of the target species PPI network.

During the last few years different algorithms and tools have been introduced to perform homology-based PPI prediction. Nevertheless, none of them appeared to be fully satisfactory and suitable for the requirements which arose during this work. Hence, a new PPI prediction approach, PATH2PPI, was developed and published in the course of this work enabling the prediction of PPI networks in *P. anserina* as well as in other organisms. The tool overcomes different deficiencies of other currently available prediction tools and approaches.

For example, many sequence-based tools required a predefined set of proteins

and/or interactions to subsequently evaluate whether these proteins may interact or the proposed interactions are reliable (Murakami and Mizuguchi, 2014; Chen et al., 2009). Since one major objective of many studies is to gain knowledge about a certain pathway in a target species for which nearly nothing is known, it is not clear which pairs of proteins to choose *a priori*. Moreover, it is not comfortable to check various protein pairs consecutively whether they interact or not and to subsequently assemble the putative interaction network manually, using these different tested protein pairs. In addition, most of the currently available prediction tools exhibit relatively "closed" architectures. These tools are deposited on servers, i.e. the source code is hidden and does not allow the customization of the implemented algorithm to own demands. Another drawback of different prediction tools is that they are focusing on some model organism and do not provide the possibility to predict interactions for many other, less studied organisms (Wiles et al., 2010). In many of the prediction tools it is a challenging task to achieve the underlying information which led to the predicted and proposed interactions, i.e. to find out which species or reference interactions had been taken into account.

PATH2PPI was developed to overcome these and other drawbacks of currently available prediction solutions. It was implemented with an open-source architecture, i.e. it is not a "black box" and provides the possibility to add additional functionalities. PATH2PPI does not require presumptions about the pathway in the target species and can be applied to every fully sequenced organism. Furthermore, each prediction can be attributed to the reference interaction (see section 4.3).

Chapter 4

Publications

This chapter contains all manuscripts published in the course of this work. For every publication the authors' contributions are described, and the relative amount of contribution is estimated for Oliver Philipp.

Basically, the first two papers aimed to achieve a general overview on age-dependent processes and the involved components on two different biological levels. While, the first publication (**1**) deals with transcriptome data deduced from entire cell extracts, the second publication (**2**) focused on proteome data deduced from isolated mitochondria of *P. anserina*. Each of both studies comprised data which had been gained from samples of *P. anserina* strains of different and consecutive age stages. The general matter of publication (**1**) was an unbiased investigation of all genes that are involved in age-related processes and are either responsible for ageing and decline of the organisms or are involved in age-dependent quality assurance mechanisms. Furthermore, due to the holistic and systems biological approach of that study it was possible to reveal several biological age-dependent pathways where the associated genes are regulated in an age-dependent manner. Publication (**2**) especially aimed to reveal whether and how ROS influence and damage the proteome of *P. anserina* in the course of ageing. Since ROS generally are generated in the mitochondria, the study focused on the proteomes of these organelles.

The third publication (**3**) introduces a new software package which enables the prediction of protein-protein interaction networks of particular biological pathways. The motivation of implementing the software package is based on findings of publication (**1**). It was implemented to gain deeper knowledge about particular pathways.

Finally, the fourth publication (**4**) used the software package introduced in publication (**3**) and aims to construct a comprehensive biological network of the autophagy pathway. The latter was the most promising age-dependent pathway found in publication (**1**).

Manuscripts (**1**), (**3**) and (**4**) are first author publications where (**1**) has been published in PLoS One (Impact Factor (IF): 3.54), the second (**3**) in Bioinformatics (IF: 5.76) and the third (**4**) in BMC Bioinformatics (IF: 2.43). The co-authorship publication (**2**) has been published in Molecular & Cellular Proteomics (IF: 5.91).

4.1 A Genome-Wide Longitudinal Transcriptome Analysis of the aging Model *Podospora anserina*

Publication status: Published

Summary

Advanced next generation sequencing (NGS) techniques enable the analysis of a vast amount of biological data from different sources at once. Since ageing is also accompanied by differential gene expression, such an NGS technique was applied to study the genetic context of ageing in *P. anserina*. By means of a SuperSAGE analysis (performed by GenXPro, Frankfurt) the transcriptomes of three individuals on seven consecutive days in the lifetime of *P. anserina* were captured. This longitudinal analysis resulted in more than 10.000 expression profiles which had been analysed to reveal age-dependent gene expression and biological pathways which play essential roles during ageing.

Initially, a general statistical pipeline was applied, comprising filter, normalisation and data fitting approaches to enhance the quality of the transcriptome data. A qRT-PCR analysis, experimentally validated the reliability of the data. Subsequently, the processed data was analysed by three different types of statistical and bioinformatics analyses. Firstly, a fuzzy cluster analysis was performed to group similar expression profiles into eight (overlapping) groups of distinct profile patterns. Secondly, the focus was set to genes which seemed to be continuously up- or down-regulated during ageing. Thirdly, the data set was reduced and only the first and the last day of measurement was considered in order to reveal significant changes in gene expression when comparing only young with old individuals.

It was found that different pathways were affected during ageing and several biological processes were identified which were particularly regulated in an age-dependent manner. For example, genes associated with the ribosomes and protein biogenesis are down-regulated during ageing. Furthermore, genes involved in protein degradation via the ubiquitin-proteasome system (UPS) were continuously down-regulated as well. In contrast, genes associated with another cellular quality assurance system, the autophagy pathway, seemed to be up-regulated during ageing. Genes associated with the energy metabolism, especially exhibit significant differences in gene expression when only comparing very young with very old individuals.

Additionally, to the bioinformatics and statistical analyses, the entire raw as well as the processed data has been published and deposited in the Euopean Bioinformatics Institute's ArrayExpress public data repository.

Authors' contribution

Authors: Oliver Philipp (OP), Andrea Hamann (AH), Jörg Servos (JS), Alexandra Werner (AW), Ina Koch (IK) and Heinz D. Osiewacz (HDO)

Percentage of contribution for OP:

Conception/Design:	50 %
Experimental work:	0 %
Implementation:	70 %
Analysis/Evaluation:	70 %
Manuscript:	35 %

OP performed advanced statistical analysis on semi-raw data (after preprocessing by GenXPro). OP developed pipeline for data filtering, normalisation and fitting. Additionally, OP planned, designed and performed the entire bioinformatics analysis pipeline and applied each described approach to analyse the transcriptome data and to draw the biological conclusions. JS applied homology comparison to achieve GO universe. AH and AW performed biological experiments. HDO conceived and designed the experiment. HDO, OP, AH, IK wrote the manuscript. HDO and IK supervised the study.

OPEN ACCESS Freely available online

PLOS ONE

A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserine*

Oliver Philipp[1,2], Andrea Hamann[1], Jörg Servos[1], Alexandra Werner[1], Ina Koch[2], Heinz D. Osiewacz[1]*

1 Molecular Developmental Biology, Institute of Molecular Biosciences, Faculty for Biosciences & Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe University, Frankfurt am Main, Germany, **2** Molecular Bioinformatics, Institute of Computer Science, Faculty of Computer Science and Mathematics & Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe University, Frankfurt am Main, Germany

Abstract

Aging of biological systems is controlled by various processes which have a potential impact on gene expression. Here we report a genome-wide transcriptome analysis of the fungal aging model *Podospora anserina*. Total RNA of three individuals of defined age were pooled and analyzed by SuperSAGE (serial analysis of gene expression). A bioinformatics analysis identified different molecular pathways to be affected during aging. While the abundance of transcripts linked to ribosomes and to the proteasome quality control system were found to decrease during aging, those associated with autophagy increase, suggesting that autophagy may act as a compensatory quality control pathway. Transcript profiles associated with the energy metabolism including mitochondrial functions were identified to fluctuate during aging. Comparison of wild-type transcripts, which are continuously down-regulated during aging, with those down-regulated in the long-lived, copper-uptake mutant grisea, validated the relevance of age-related changes in cellular copper metabolism. Overall, we (i) present a unique age-related data set of a longitudinal study of the experimental aging model *P. anserina* which represents a reference resource for future investigations in a variety of organisms, (ii) suggest autophagy to be a key quality control pathway that becomes active once other pathways fail, and (iii) present testable predictions for subsequent experimental investigations.

Citation: Philipp O, Hamann A, Servos J, Werner A, Koch I, et al. (2013) A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserine*. PLoS ONE 8(12): e83109. doi:10.1371/journal.pone.0083109

Editor: Gustavo Henrique Goldman, Universidade de Sao Paulo, Brazil

Received August 14, 2013; **Accepted** November 8, 2013; **Published** December 20, 2013

Funding: This work was supported by a grant by the Bundesministerium für Bildung und Forschung (BMBF; GerontoMitoSys 0315584A) to HDO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Osiewacz@bio.uni-frankfurt.de

Introduction

Biological aging is a complex process leading to physiological impairments, the degeneration of cellular and organ functions, the development of disease and finally death of the system [1–3]. The underlying molecular mechanisms are multifactorial and only partially defined. It is clear that aging is accompanied by changes in gene expression. The available data, however, are mainly derived from single gene analyses and the comparison of only a few age stages (e.g., young vs. old). The comprehensive and systematic analyses of changes over the lifetime of individuals can identify new key pathways and regulatory circuits involved in aging and lifespan control and can open the field for the development of strategies to intervene into aging and age-related diseases (e.g., cancer, dementia, Parkinson's disease, cardiovascular impairments). Nowadays, the availability of efficient high-throughput techniques makes such studies possible, in particular when the study is performed with experimentally accessible short-lived systems.

Podospora anserina is such a system [4–6]. In contrast to most filamentous fungi this ascomycete is characterized by a well-defined aging process that is under the control of genetic and environmental traits. After germination of an ascospore, a mycelium develops which grows at the periphery until it reaches a phase where the growth rate first decreases until it comes to a complete growth stop [7]. Finally, the hyphal tips burst and die. This process occurs under nutrient-replete growth conditions and thus clearly differs from those described as 'aging' in fungi grown under nutrient starvation [8] and as 'chronological aging' in the yeast *Saccharomyces cerevisiae* [9,10]. In *P. anserina*, various molecular pathways have been identified to be involved in the control of aging and development [11–13]. The lifespan of the fungus is short (typically two to four weeks for wild-type strain 's') and depends on the growth medium and on cultivation conditions [5,14]. The vegetation body of *P. anserina* is simply consisting of branched filamentous cellsforming a mycelium. For sexual reproduction specialized organs, protoperithecia and spermogonia, are formed in dikaryotic as well as in monokaryotic strains. *P. anserina* is accessible to experimentation [4,5]. Biomolecules like DNA, RNA or proteins as well as whole mitochondria can be isolated and analyzed from individuals of well-defined age [5]. The complete genome of *P. anserina* is sequenced and consists of about 36 MBp coding for more than 10,600 putative proteins [15,16]. *P. anserina* can be genetically manipulated by classical genetic approaches and by genetic engineering [5,17,18].

Here we describe a genome-wide transcriptome profiling of three *P. anserina* individuals from which total RNA was isolated after 6, 9, 10, 11, 12, 13 and 14 days of cultivation. Quantitative transcript profiles were generated by serial analysis of gene expression (SuperSAGE) and analyzed by bioinformatical and statistical approaches [19–21]. Previously we used SuperSAGE

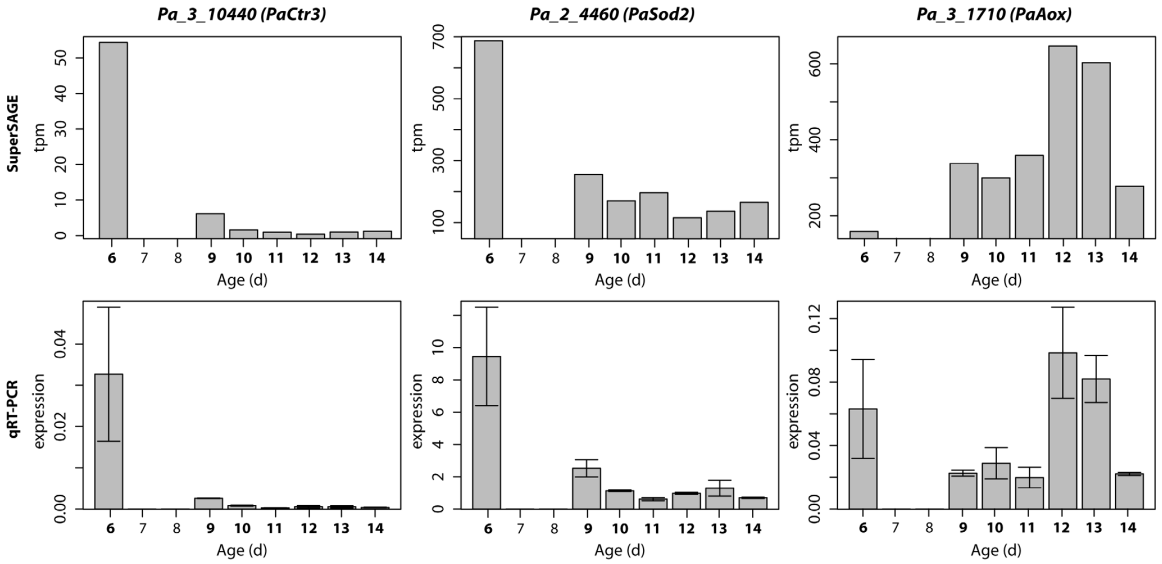


Figure 1. Transcript analysis of Pa_3_10440 (PaCtr3), Pa_2_4460 (PaSod2) and Pa_3_1710 (PaAox). Upper row: RNA of the three biological replicates (individuals) was isolated, pooled and analyzed by SuperSAGE. Gene expression was quantified as "tags per million" (tpm). Lower row: The same RNA was used for qRT-PCR analysis. RNA samples from each individual were individually analyzed for relative gene expression. Error bars represent the standard error. In both analyses, x-axes indicate the age of the individuals at which total RNA was isolated.
doi:10.1371/journal.pone.0083109.g001

successfully to characterize the transcriptome of a specific long-lived mutant of *P. anserina* and compared it to the transcriptome of the wild type. Validation by qRT-PCR demonstrated the reliability of this method [22]. The data of the current longitudinal study, in which RNA was isolated from the same fungal individuals after a defined period of growth and subjected to a genome-wide SuperSAGE analyses, identified autophagy as a quality control pathway up-regulated late in the life of *P. anserina* at a time when transcripts, encoding components of other pathways (e.g., proteasome), are down-regulated.

Materials and Methods

Podospora anserina Strains and Cultivation

For all experiments, three independent monokaryotic spore isolates (mating type minus) of the wild-type strain 's' [7] were used. Cultivation was essentially performed as described previously [23]. Briefly, single ascospores were germinated for 2 days on germination medium. Pieces of mycelium of this two day old culture were either directly transferred to a fresh PASM [24] plate overlaid with a cellophane sheet or, in order to generate strains of older age, to solid PASM medium and incubated under permanent light at 27°C. After 5, 6, 7, 8, 9, and 10 days, respectively, pieces from the growth front of the latter cultures were transferred to a fresh PASM plate (overlaid with a cellophane sheet). After two days of growth, the mycelium of the developed culture was transferred from the cellophane to liquid CM medium [25] and incubated for additional 2 days at 27°C under light and agitation. This last incubation step leads to the formation of enough mycelium (biomass) that, free of agar, can be easily harvested for the isolation of RNA. Following this regime, mycelium grown for a defined period of time (different age stages) of 6, 9, 10, 11, 12, 13, and 14 days, respectively, was available for isolation of total RNA. To make sure that all three isolates have a similar aging behaviour, the lifespan as period of linear growth on solid PASM medium was recorded. All isolates had a lifespan of 14 days, thus the oldest age stage (14 days) represents a senescent culture.

Isolation of Total RNA

Total RNA was isolated using a CsCl density gradient as described previously [22].

Quantitative Real-time PCR

Quantitative Real-time PCR (qRT-PCR) was performed as described in [22]. Primer sequences can be found in Table S1.

SuperSAGE Analysis

A SuperSAGE analysis [19] was performed for each of the seven samples consisting of the pooled RNA of three genetically identical individuals as described above. Sequence tag identification and annotation, and basal statistics were performed by GenXPro (Frankfurt) as described in [22]. The raw data have been deposited in the European Bioinformatics Institute's ArrayExpress public data repository (http://www.ebi.ac.uk/arrayexpress/) [26] and are available under accession number E-MTAB-2016.

Data Preparation and Filtering

Absolute tag counts were corrected for experimental biases and normalized to tags per million (tpm) in order to compare the different samples (see: Text S1). The latter is necessary because each library (absolute number of all tag molecules per sample) can differ from each other due to experimental and biological fluctuations. Furthermore, in order to compare the n different expression profiles they have been standardized to a mean expression strength of zero and a standard deviation of one:

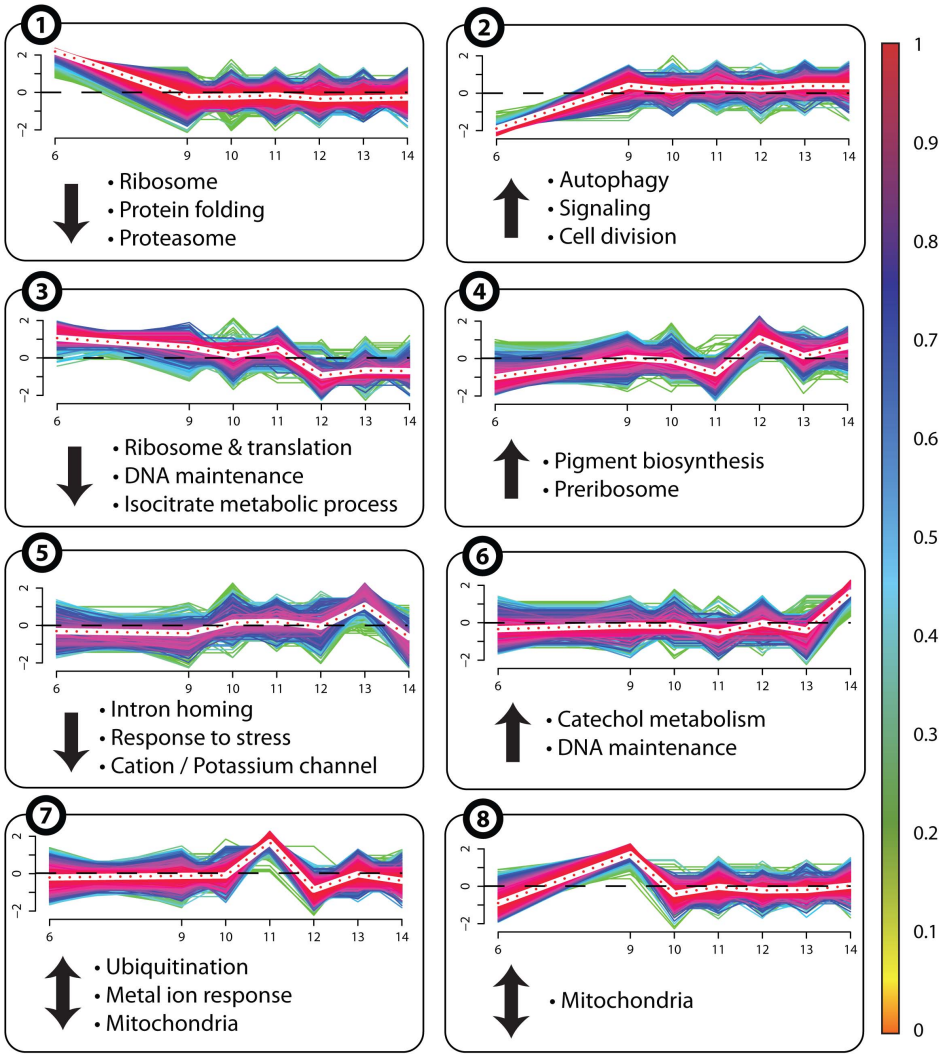


Figure 2. Fuzzy clustering analysis. All of the 7,467 "smoothed" expression profiles of the "cluster profile library" are distributed to eight fuzzy clusters. The x-axis represents the seven time points at which RNA was isolated from the three investigated *P. anserina* individuals. The y-axis indicates the relative expression strength. The color gradient corresponds to the membership value (see color bar on the right). A membership value of 1 means a perfect fitting to the corresponding cluster core (deep red). Arrows indicate the general tendency of the corresponding cluster. Below each cluster, a selection of categories to which significantly enriched GO terms (p-value ≤0.01) for particular functions were associated are listed (see description in the text).
doi:10.1371/journal.pone.0083109.g002

$$S = \frac{X - \frac{1}{n}\sum_{i=1}^{n} x}{\sigma_X},$$

with X as the expression profile and σ_X as its standard deviation. If analyses consider profiles in a Euclidean metric the regarding profiles were standardized, as well (e.g., for cluster analysis – see below). Expression profiles with more than 25% missing values were removed.

Gene Ontology Analysis

At the time our transcriptome data became available, the genes in the genome of *P. anserina* were only poorly assigned to GO terms [21]. To extend the list of assignments for further analyses, we performed a BLAST [27] search of all 10,635 putative *P. anserina* genes and the proteins in the UniProt [28] data library adopting the information of other better annotated species. Subsequently, we assigned all GO terms of each hit with an E-value of ≤1e-20 to the corresponding gene of *P. anserina*. The complete annotation library is provided in Table S2.

The GO enrichment analysis was performed using the programming language R [29] by means of the R package

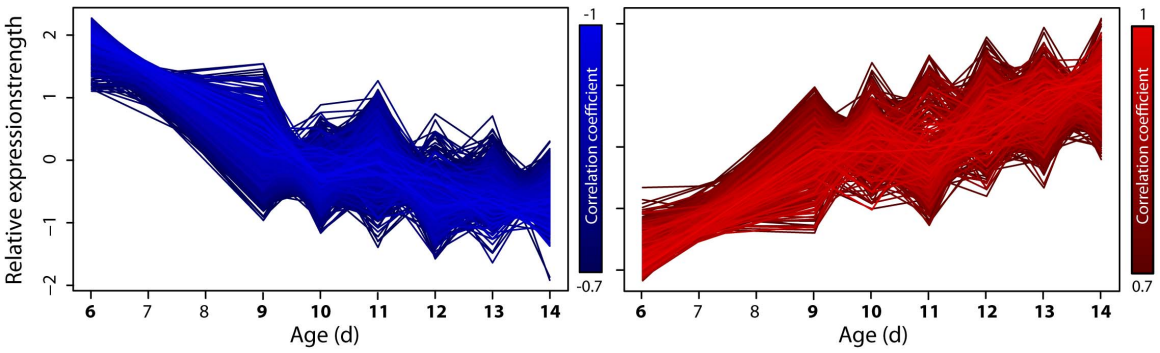


Figure 3. Age-related profiles of continuously down- or up-regulated genes. The Pearson correlation coefficient for each profile was determined. It indicates whether a gene was down- (≤ -0.7) or up-regulated (≥ 0.7) during aging. The identified groups of down- (left) and up-regulated (right) expression profiles are depicted. The brighter (weaker) the color of a profile, the higher (lower) is the correlation coefficient (see color bars). Days on the x-axis in bold identify the time points, at which RNA samples were isolated. 1,202 genes with a decreasing (left) and 418 genes with an increasing tendency (right) were identified.
doi:10.1371/journal.pone.0083109.g003

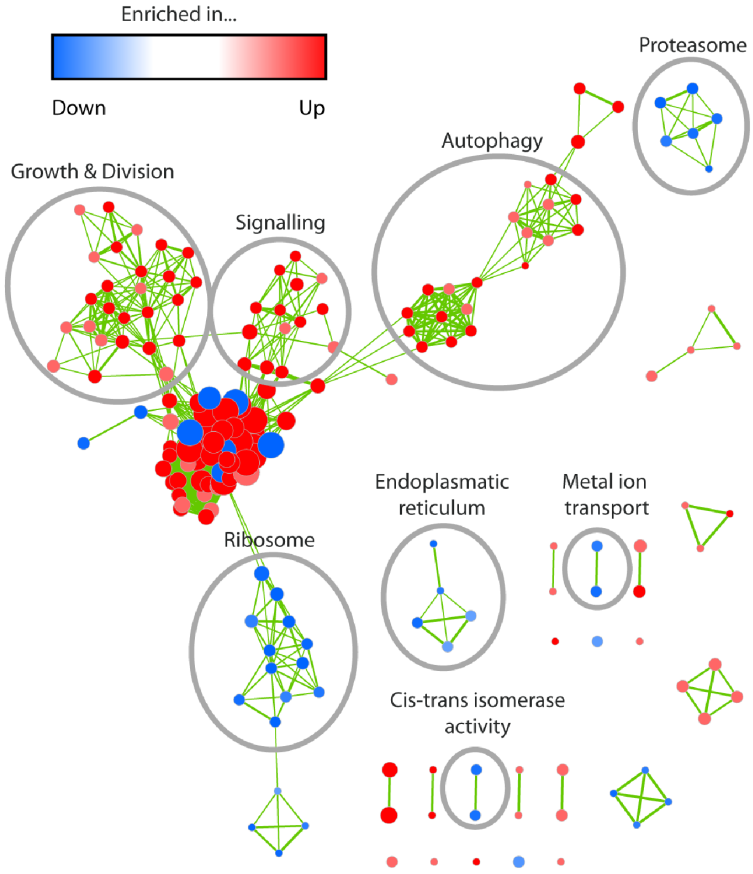


Figure 4. GO term enrichment map for the library of down- and up-regulated genes. Each node represents one GO term. The thickness of the edges represents the number of genes shared by two terms. The node color corresponds to the group and the degree of significance for enrichment (see color bar). Node sizes indicate the number of genes within the corresponding GO term. GO terms are grouped by their similarity degree defined by the number of common genes or related terminology. Striking groups were manually circled and labelled. Map regions with a high density of thick-sized nodes (center of figure) represent GO terms, describing very general processes with rather low information.
doi:10.1371/journal.pone.0083109.g004

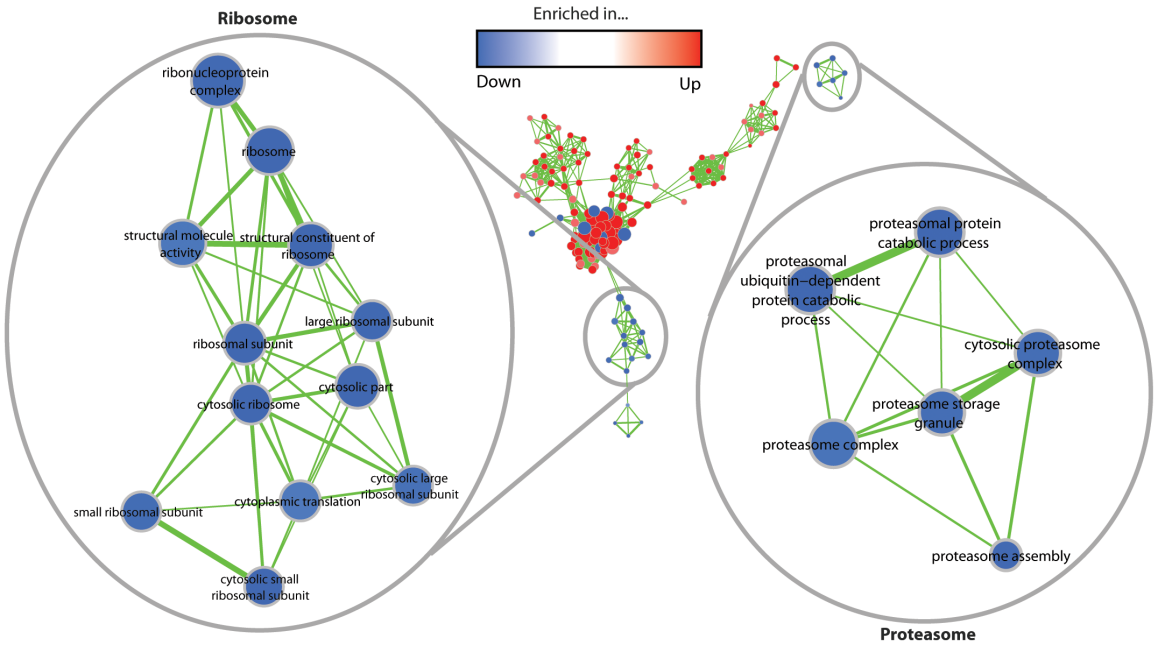


Figure 5. GO term enrichment map for the "ribosome" and the "proteasome" categories in the set of down-regulated genes. As in Figure 4, each node represents one GO term and node colors correspond to the group and the degree of significance. Edge thickness indicates the amount of genes shared by two terms and the node sizes are proportional to the number of genes assigned to the corresponding term.
doi:10.1371/journal.pone.0083109.g005

GOstats [30] and an one-sided hyper geometric test (test for over-representation). GO terms were assumed to be statistically significantly enriched if the probability (p)-value was ≤ 0.01.

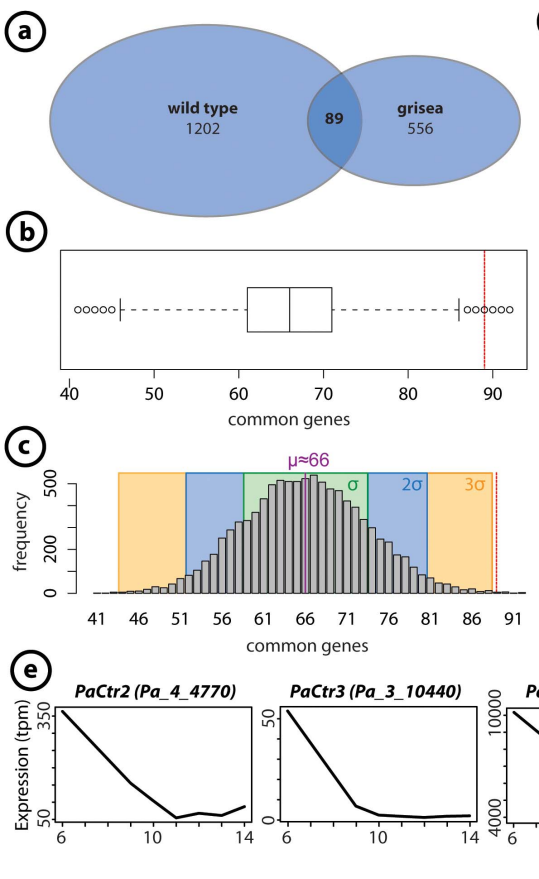
To identify and visualize possible links between individual GO terms deduced from the GO enrichment analyses, in selected cases, enrichment maps were generated [31] using the Cytoscape plugin [32]. Such a map is a graph-based representation which overcomes the redundancy problem in a GO enrichment analysis. In order to use the enrichment map plugin, a dedicated R script, providing the necessary interface, was developed. Basically, the enrichment analysis results preliminary generated by the GOstats package were adjusted and exported to a file containing the obligatory columns: GO identifiers, GO terms, enrichment p-values and phenotype (up- or down-regulated indicated by 1 or 0, respectively). Furthermore, a gmt-file was generated that contains all GO assignments, consisting of three columns: GO identifiers, GO descriptions, and all assigned gene accession numbers with the corresponding GO identifier. The R script is available upon request. In Cytoscape the default values for the enrichment map plugin were kept, only the p-value cutoff was set to 0.01, and a Jaccard coefficient of 0.25 as overlap coefficient was applied.

Identification of Expression Profiles with Continuous Expression Tendencies

The gene expression profiles were searched for up- and down-regulated genes. For this purpose, each profile, i.e., the transcript abundance for each of the seven time points, was correlated with time (days 6, 9, 10, 11, 12, 13, 14) using the Pearson correlation coefficient. The threshold was set to 0.7. This threshold is accepted to indicate a strong (positive or negative) linear correlation of two variables or expression data, respectively [33,34]. Thus, with a coefficient of ≤ -0.7 the corresponding expression profiles show a decrease and profiles with a coefficient of ≥ 0.7 an increase during aging. We only considered profiles further if the p-value for differential expression of day 6 (first day in measurement) to day 14 (last day) is $\leq$1e-10.

Significance Smoothing

The clustering algorithm we performed used a Euclidean metric. To compare the different patterns, it was necessary to standardize the expression profiles. A major drawback of standardization is that various information regarding variances and dispersion get lost. To avoid as much as possible an over-fitting and over-interpretation and to get rid of low abundant and/or biased expression profiles, we applied a "significance smoothing" to the gene expression data. This means, for each gene and for each pair of age stages (day 6 to day 9, day 9 to day 10, day 11 to day 12, day 12 to day 13, day 13 to day 14) the p-value indicates whether the corresponding gene is differentially expressed. If this p-value is less or equal than 1e-3, we assumed a significant differential expression for the corresponding gene during the two age stages and kept the original expression strength. If the p-value is above that threshold, we assumed no significant differential gene expression and correct the corresponding expression value for this age stage, using the mean expression value of the previous and current non-significant expression values (Figure S1 in File S1). After this smoothing step only significant expression patterns remained. Those patterns, which were very likely due to experimental or biological fluctuations, were corrected and smoothed. This led to expression profiles with no expression changes, which were removed from the dataset.



d

GO ID	GO term	Ont.	P-value
GO:0035434	copper ion transmembrane transport	BP	1,60E-05
GO:0006825	copper ion transport	BP	1,54E-04
GO:0000041	transition metal ion transport	BP	1,87E-04
GO:0030001	metal ion transport	BP	2,48E-04
GO:0006826	iron ion transport	BP	3,07E-04
GO:0015891	siderophore transport	BP	4,45E-04
GO:0015688	iron chelate transport	BP	4,45E-04
GO:0015677	copper ion import	BP	6,65E-04
GO:0033214	iron assimilation by chelation and transport	BP	9,27E-04
GO:0019439	aromatic compound catabolic process	BP	1,64E-03
GO:0006725	cellular aromatic compound metabolic process	BP	2,62E-03
GO:0033212	iron assimilation	BP	3,90E-03
GO:0034220	ion transmembrane transport	BP	5,00E-03
GO:0006879	cellular iron ion homeostasis	BP	5,40E-03
GO:0009698	phenylpropanoid metabolic process	BP	5,75E-03
GO:0006085	acetyl-CoA biosynthetic process	BP	6,44E-03
GO:0019551	glutamate catabolic process to 2-oxoglutarate	BP	6,80E-03
GO:0055072	iron ion homeostasis	BP	6,95E-03
GO:0006573	valine metabolic process	BP	7,17E-03
GO:0006098	pentose-phosphate shunt	BP	9,57E-03
GO:0000293	ferric-chelate reductase activity	MF	5,87E-05
GO:0016723	oxidoreductase activity, oxidizing metal ions, NAD or NADP as acceptor	MF	1,38E-04
GO:0016491	oxidoreductase activity	MF	1,50E-04
GO:0016722	oxidoreductase activity, oxidizing metal ions	MF	3,81E-04
GO:0008442	3-hydroxyisobutyrate dehydrogenase activity	MF	9,54E-04
GO:0004806	triglyceride lipase activity	MF	3,01E-03
GO:0004616	phosphogluconate dehydrogenase (decarboxylating) activity	MF	4,78E-03
GO:0009055	electron carrier activity	MF	5,34E-03
GO:0004352	glutamate dehydrogenase (NAD+) activity	MF	8,13E-03
GO:0003868	4-hydroxyphenylpyruvate dioxygenase activity	MF	8,13E-03

Figure 6. Comparison of down-regulated genes in the grisea mutant and the wild type. (a) Common down-regulated genes within the age-dependent transcriptome of the wild-type 's' and the long-lived mutant grisea of *P. anserina*. 556 genes of the grisea transcriptome were identified to be down-regulated with a factor of at least 3 and a p-value of ≤0.01 for differential expression [22]. The 1,202 genes, identified as being down-regulated during aging in this study, have 89 genes in common. (b) 10,000 simulated experiments with two sets of randomly picked gene names, each set of the same size as the two sets (wild type, grisea), were applied. The box plot shows that the 89 common genes were extreme outliers. (c) The results of the random experiments are normally distributed with a mean of μ≈66.05 and a standard deviation of σ≈7.39. Our determined value of 89 is greater than three times sigma, indicating that the corresponding genes do not occur randomly. (d) All enriched GO terms in the set of the 89 common genes with a p-value ≤0.01 are depicted. Copper associated terms can be found on the very top. (e) Four genes already identified as putative target genes of transcription factor GRISEA are found within these 89 common genes. Due to its irregular expression profile, the gene, encoding the putative copper transporter PaCTR1, is not among the 89 common genes.
doi:10.1371/journal.pone.0083109.g006

Fuzzy Cluster Analysis

We performed a fuzzy cluster analysis to identify significant and striking age-dependent expression patterns generated during the lifespan of *P. anserina*. In contrast to hard clustering approaches, where disjunctive clusters are generated, fuzzy clustering leads to overlapping clusters. In this case, each profile belongs to each cluster with a given degree of membership in the interval [0; 1] where 1 indicates a total fit of the expression profile to the corresponding cluster core. Because gene expression itself and the experimental conditions as well can vary, this method is suitable as it avoids too stringent selection criteria. In our analysis we used the fuzzy clustering algorithm fuzzy c-means (FCM) [35] based on the modified R package "Mfuzz" [20]. The optimal number of clusters, which had to be defined *a priori*, was determined by means of the Xie-Beni index which computes a value for a fuzzy cluster result [36]. The smaller this index, the better is the partitioning. A repeated approach revealed an optimal size of eight clusters (Figure S2 in File S1).

Results and Discussion

Overview, Data Preparation and Validation

From single gene analyses it is known that aging of *P. anserina* is correlated with differential gene expression. In the corresponding analyses, transcript levels of selected genes (e.g., *PaDnm1*, *PaCtr3*, *PaSod2*, *PaMt1*) were found to differ in juvenile and senescent cultures [37–39]. Recently, we reported a differential genome-wide transcriptome analysis of juvenile strains of the long-lived mutant grisea and the wild-type 's' [22]. This analysis allowed the assignment of 9,700 transcripts.

Age-related transcriptomes. In order to obtain a systematic view about age-related gene expression of this aging model, we

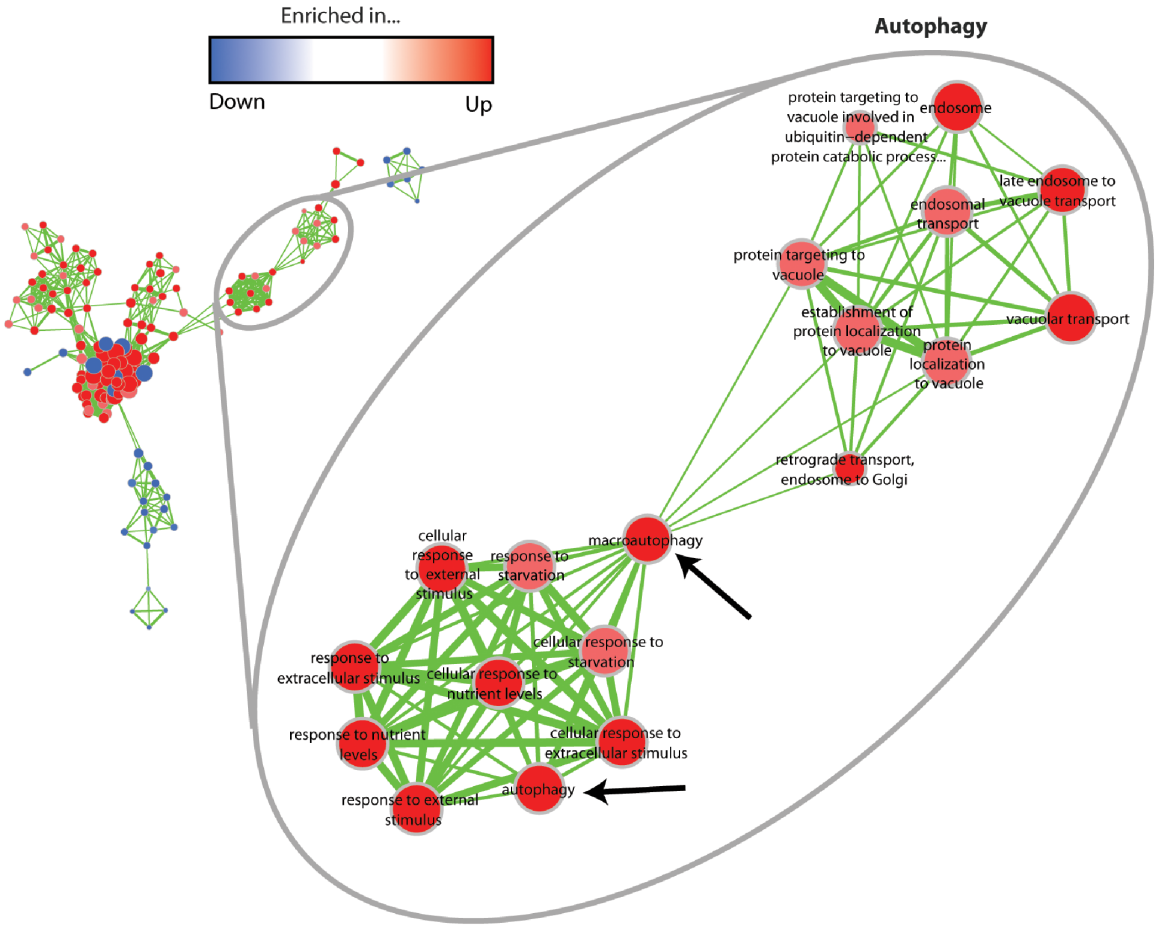


Figure 7. GO term enrichment map for the ''autophagy'' category in the up-regulated gene set. Analogous to Figure 4, each node represents one GO term and node colors correspond to the group and the degree of significance. Edge thickness indicates the amount of genes shared by two terms, and node sizes are proportional to the number of genes assigned to the corresponding term.
doi:10.1371/journal.pone.0083109.g007

performed a genome-wide longitudinal expression profile study of the wild-type strain 's' [5]. In this analysis, total RNA was isolated from three independent individuals (biological replicates) after 6, 9, 10, 11, 12, 13 and 14 days of growth, pooled and subjected to a SuperSAGE analysis. Overall, more than 85 million single tag molecules were sequenced (Table S3) leading to 646,000 distinct different tags. A BLAST search led to the assignment of about 150,000 sequence tags to more than 10,200 of the predicted 10,635 genes encoded in the genome of *P. anserina*. After applying some filtering routines (e.g., removal of incomplete profiles) a "basic expression profile library" of 10,059 genes remained which covers about 95% of all genes in the genome of *P. anserina*. The processed expression data are compiled in Table S4.

Data validation. To validate the results obtained in the SuperSAGE analysis, qRT-PCR experiments were performed with three selected genes (*PaCtr3*, *PaSod2*, *PaAox*) of *P. anserina* (Figure 1). For *PaCtr3* and *PaSod2* it is known from previous work that their abundance changes from juvenile to senescent [38]. For *PaAox*, a northern blot analysis of the juvenile and the senescent stage revealed that the transcript was shown to decline during aging [40]. However, with the additional samples covering more age stages now, it becomes clear that the trend has a peak at days 12 and 13 (pre-senescent stage) and then declines in the oldest age stage (senescent, day 14). This trend has not been seen before when RNA samples from only two age stages juvenile and senescent were studied [40]. The increased expression of *PaAox* in the pre-senescent phase may be explained as a compensatory attempt to rescue the breakdown of cytochrome c oxidase dependent respiration.

The RNA samples of the current study were the same in both, the qRT-PCR and the SuperSAGE analysis. However, in contrast to the SuperSAGE study, the three samples (biological replicates) of each age stage were not pooled for the qRT-PCR, but analyzed separately, allowing the determination of a standard error among the different biological replicates. Comparing the highest and lowest transcript levels in each profile, at least 3-fold differences in RNA abundance were found for each gene. Moreover, the transcript levels determined by the two different techniques were comparable (Figure 1). This conclusion is also in agreement with another recent transcriptome analysis [22].

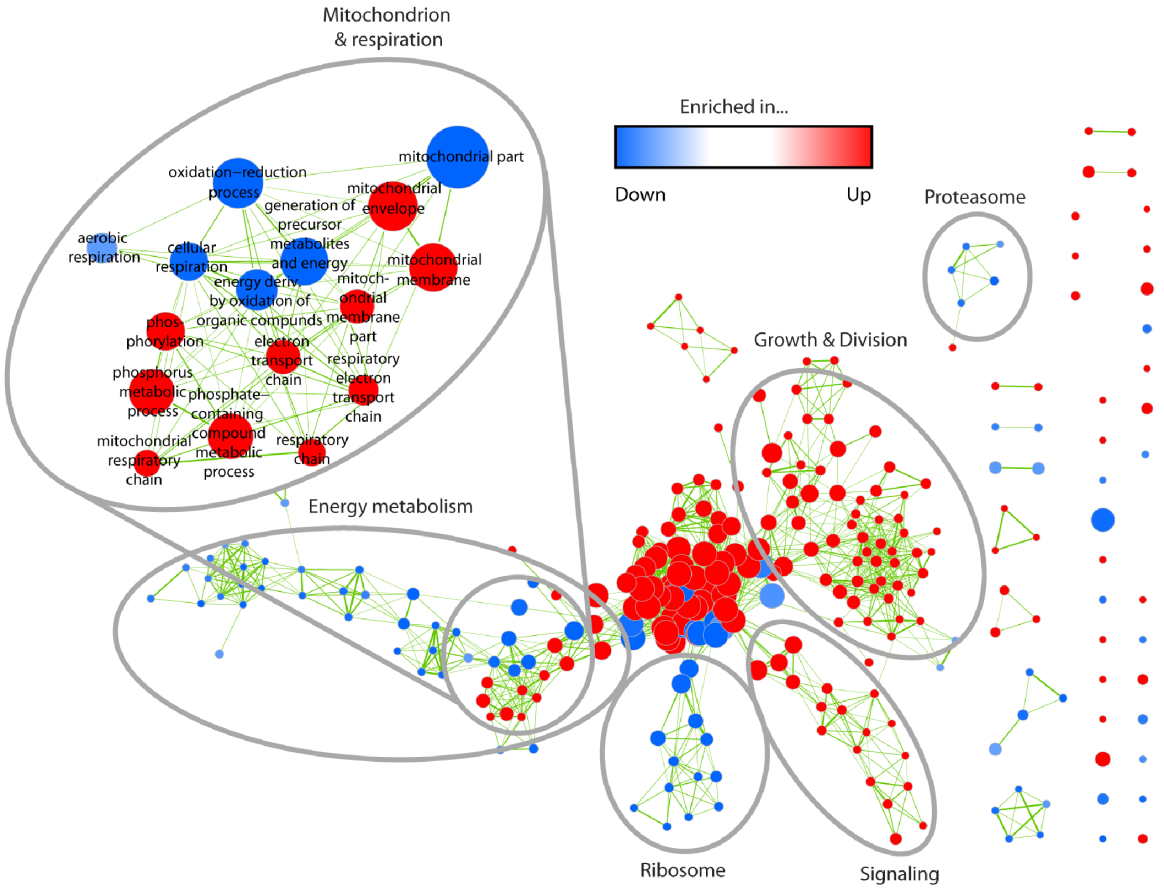


Figure 8. Transcriptome comparison of young (day 6) and old individuals (day 14). The figure depicts the enrichment map for the corresponding GO enrichment analysis. Analogous to Figure 4, each node represents one GO term and node colors correspond to the group and the degree of significance. Edge thickness indicates the amount of genes shared by two terms and node sizes are proportional to the number of genes assigned to the corresponding term. Striking groups were subsequently circled and labelled.
doi:10.1371/journal.pone.0083109.g008

Gene Ontology (GO) term assignment. Using the *P. anserina* genome data base, we first performed an extended homology search to identify putative homologs encoded by all of the 10,635 putative *P. anserina* genes. We used this information to assign 2,334 additional proteins with an E-value of ≤1e-20. Overall, 7,609 (72%) of the 10,635 gene products are now assigned to GO terms (Table S2), while 2,451 of the genes in the basic expression profile library remain to code for putative products of unknown function.

Significance smoothing. In order to analyze the 10,059 expression profiles for age-dependent expression patterns and/or to investigate co-regulated genes associated with similar pathways and processes, we performed a cluster analysis using a Euclidean metric. For this purpose, we first standardized the 10,059 generated profiles to a mean of zero and a standard deviation of 1 (see: Methods). A major drawback of standardization approaches is that information regarding variances and dispersion become lost. For example, if a gene's expression changes from 1 tag per million (tpm) to 3 tpm after standardization it seems to have the same variation as a gene's expression change of 100 tpm to 300 tpm, even if these small expression changes (1 tpm to 3 tpm) are probably due to experimental noises. Hence, to avoid an over-fitting and over-interpretation, we first applied a "significance smoothing" to the 10,059 expression profiles and ended up with a "cluster profile library" of 7,467 reliable expression patterns which exhibit significant (age-related) differential expression.

Fuzzy Cluster Analysis

In order to analyze the 7,467 profiles in the "cluster profile library" for age-dependent expression patterns and/or to investigate co-regulated genes associated with similar pathways and processes, we performed a fuzzy cluster analysis. We used the fuzzy c-means algorithm (FCM). For the results see Table S5. The optimal size of clusters was statistically determined to eight (see: Methods).

Figure 2 depicts the eight fuzzy clusters and their corresponding expression patterns, including all of the age-dependent profiles of the *P. anserina* "cluster profile library". In contrast to "hard" clustering approaches each profile fits to each "fuzzy" cluster with a specific membership degree. The thick white profile represents the core profile of each cluster and corresponds to the weighted mean of all profiles assigned. The color gradient used for the

profiles indicates the congruence with the cluster core, i.e., a red profile displays a perfect fit to the corresponding cluster core profile (see color bar).

The eight main patterns, can be assigned to following profile types: (i) profiles showing a general decrease of transcripts (cluster 1, 3, and 5), (ii) profiles with a general transcript increase (cluster 2, 4, and 6), and (iii) profiles with a strong transcript increase earlier or later in the life cycle which subsequently falls back to the basic level (clusters 7 and 8).

In order to assign characteristic pathways, processes and components to the different clusters, we performed a GO enrichment analysis. The results of the complete analysis are provided in the supporting data (Table S6). In Figure 2, below each cluster, a selection of categories related to GO terms with particular interest for aging and lifespan control is indicated. The categories were manually generated by compiling entries from the GO universe. For instance, the terms "ribosome", "small ribosomal subunit", "cytosolic small ribosomal subunit", "structural constituent of ribosome" and some more were grouped into the category "ribosome".

Clusters 1 and 3 with down-regulated transcripts were found to contain enriched genes coding for products associated with protein folding (e.g., "peptidyl-prolyl cis-trans isomerase activity"), the "proteasome", the "ribosome", "DNA maintenance", and the "isocitrate metabolic process".

The main profile of cluster 5 describes a weak increase of expression till day 13 followed by a rapid decrease below the basal level. In this cluster, GO terms related to categories like "response to stress" and "cation/potassium channels" were found to be significantly enriched. While "response to stress" is a well characterized type of age-related processes basically nothing is known about the impact of "cation/potassium channels" in aging of *P. anserina*. In contrast, an important role of "intron homing", another GO term identified to be enriched in cluster 5, has been well investigated before. "Intron homing" was reported to be involved in the age-related reorganization of the mitochondrial DNA (mtDNA) as it occurs during aging of cultures [41–44]. The underlying mechanism is related to the mobility of the first cytochrome c oxidase intron [45], its release from and its reintegration into "homing sites" in the mtDNA and the subsequent recombination between intron repetition sequences [1,44]. As a consequence, large parts of the mtDNA are deleted in senescent cultures leading to deficiencies in remodeling of affected proteins encoded by the mt genome.

Overall, stress response pathways are identified as being down-regulated during aging of *P. anserina*. This category includes the GO term "isocitrate metabolic process", the role of which in regulating cellular defence against oxidative stress has been described before [46]. The other categories include DNA and protein quality control pathways like "telomere maintenance", "recombinational repair", "protein refolding", protein degradation via the "proteasome", and the replacement of affected proteins via the expression of the corresponding genes.

In cluster 2, in which transcripts increase in abundance at early life stages, enriched GO terms like "signal transduction", "regulation", "regulation of cell communication" or "intracellular protein kinase cascade" were identified and compiled in the category "signaling". This category makes clear that during aging, there is a strong need for cellular readjustment and remodeling. This conclusion is in concordance with the down-regulation of important components (e.g., the proteasome) and the accumulation of impairments during aging [47–49]. A second enriched category of GO terms is associated with "cell division" and suggests an increased need for the expression of genes coding for products of the developmental machinery. Two additional GO terms, "autophagy" and "macroautophagy", which are enriched in cluster 2, attracted our special interest. We combined these terms in the category "autophagy". The up-regulation of genes in this category appears to be of special relevance in a situation where the proteasome as a major quality control component is impaired (see below).

In the second cluster with an up-regulated tendency (cluster 4), GO terms are found that are related to "pigment biogenesis". In this category, GO terms, like "carotenoid metabolic process" and "pigment metabolic process", were found to be enriched. Strikingly, in an earlier study [50] it was shown that an over-expression of carotenoid associated genes leads to increased lifespan of *P. anserina*, probably by the ROS scavenging activity of this class of pigments. Furthermore, we found enriched GO terms related to other pigment metabolic processes. This is in concordance with an age-related increase in pigmentation of *P. anserina* cultures. Strikingly, although the category "ribosome" was found in clusters with a decrease of transcripts, we identified GO terms which can be assigned to the category "pre-ribosome" consisting of terms like "rRNA processing" and "pre-ribosome, large subunit precursor". Since early processes of the ribosomal assembly occur in the nucleus before these pre-ribosomal structures are exported to the cytosol for final maturation [51], it is reasonable to assume that the early processes in ribosome assembly are not affected. The observed increase in expression of genes, encoding pre-ribosomal proteins, may be an unsuccessful cellular rescue attempt.

In the third cluster with an up-regulated tendency (cluster 6), the GO term "SWI/SNF complex" was found to be enriched. This protein complex is part of the nucleosome remodeling complex and is implicated in various processes like gene expression, nuclear organization, centromere function, and chromosomal stability [52]. It is thus part of a category defined as "DNA maintenance". Significantly, in contrast to the corresponding GO terms identified in cluster 3, transcripts linked to this term were found to increase during aging. A second category of GO terms, termed "catechol metabolism", with particular relevance to aging was found to be significantly enriched by GO terms like "catechol catabolic process, ortho-cleavage", and "catechol-containing compound, catabolic process". Previously, work on *P. anserina* unravelled a role of such polyphenols in aging. In particular, it was found that the o-methyltransferase PaMTH1 as a longevity assurance factor, which increases in abundance during aging of *P. anserina* [53–55], appears to be involved in methylation of polyphenols with vicinal hydroxyl groups which, in the presence of Cu2+ and Fe3+, are prone to the generation of the highly reactive hydroxyl radical with its potential to damage all kinds of molecules [56–58]. The enzyme was shown to methylate various flavonoids (e.g., quercetin, myricetin) *in vitro*. Over-expression of the corresponding gene in *P. anserina* resulted in a reduction of protein damage and a lifespan extension [59]. The transcriptome data of the current study provide a first direct evidence for an age-related increase of enzymes associated with catechol in *P. anserina* (Pa_1_20370 and Pa_7_9460– "Put. hydroxyquinol 1,2-dioxygenase"), providing first evidence about *in vivo* substrates of PaMTH1. Significantly, this metabolism and the activity of the mammalian PaMTH1 homolog catechol-o-methyl-transferase (COMT) play a key role in a variety of human disorders including the age-associated Parkinson disease [60,61].

In the clusters 7 and 8 transcript abundance increases transiently at different age stages and subsequently falls back to base line levels. These are the only clusters, which contain significantly enriched GO terms associated with "mitochondria"

identified in the cluster analysis. Transcript levels increase from day 10 to 11 (cluster 7) and day 6 to 9 (cluster 8). At this period of time in the life cycle of the fungus, an optimal function of mitochondria to generate enough adenosine triphosphate (ATP) appears to be important for the generation of sexual reproduction structures (protoperithecia and spermogonia). In cluster 7, GO terms linked to the categories "ubiquitination" and "metal ion response" are enriched. Changes in metal ions concentrations in cellular compartments have been indeed demonstrated to occur during aging. In particular, an increase of cytoplasmic copper in earlier age stages has been observed during aging in *P. anserina* and human fibroblasts [38,62]. These data are consistent with the observation of increases in the expression of controlling metal homeostasis (see below) in cluster 7. Finally, an increase in "ubiquitination" may be required for the induction of protein quality control pathways to control cellular proteostasis. The corresponding pathways in *P. anserina* have not been investigated yet.

Continuously Up- or Down-regulated Expression Profiles

The cluster analysis was performed as a first rough and non-biased evaluation of all age-related expression profiles ("cluster profile library") prepared for our bioinformatical analyses. Next, in order to strengthen the clues revealed by this analysis, we reduced the complexity of the data set to profiles exhibiting a continuously up- or down-regulation of gene expression (see: methods). We identified 537 genes displaying an increase and 1,766 genes with a decrease in abundance. In order to reduce the risk that the observed profiles result from experimental or biological fluctuation, we subsequently considered only those profiles with p-values less or equal to 1e-10 for differential expression between day 6 and 14 (first and last day in the measurement). The corresponding data libraries contain 418 genes with an up-regulation (right site of Figure 3), and 1,202 genes with down-regulated age-related trend (left site of Figure 3, see also Table S7).

As in the cluster analysis, we performed a GO enrichment analysis using the two age-specific expression profile groups (Table S8). The reduction of the complexity of data allowed us to construct intuitive enrichment maps using the method introduced by [31] (see: Methods). Figure 4 represents an example for such a map. In this map each significantly enriched GO term (p-value $\leq$0.01) from both profile groups is depicted as a single node. The number of different genes in one node is indicated by the node size. Nodes that are sharing genes are connected by edges. The thickness of each edge corresponds to the number of genes shared by two nodes. GO terms enriched within the group of down-regulated genes are indicated in blue and those enriched in the up-regulated set in red. Nodes associated with similar categories of pathways and processes were manually circled. GO terms describing very general processes like "cellular metabolic process", "intracellular" or "protein binding", although they contain a larger number of genes, are not further investigated because they are linked to rather unspecific processes.

As indicated in the cluster analysis before, the GO enrichment analysis revealed a significant enrichment of terms or genes associated with the ribosome and the proteasome (Figure 5). Most significantly, as revealed by a BLAST search, the GO term "proteasome assembly" contains the genes *Pa_2_8950* and *Pa_1_12250*, which encode two putative homologs of the human proteasome subunits beta 2 and 5, respectively. These subunits are also known to decrease in abundance in late passages of human WI38 fibroblasts along with beta subunits 1, leading to a reduced proteasome activity. Significantly, the over-expression of the genes, encoding either the beta subunits 1 or 5, can rescue proteasome activity [63].

Also an enrichment of components linked to the ribosome was found in the down-regulated clusters (Figure 5). These data are consistent with experimental findings reporting a decrease of ribosome-related transcripts during replicative aging in *Saccharomyces cerevisiae* [64]. The decrease may result from the deactivation of the "target of rapamycin" (TOR) kinase, a protein complex which is highly conserved from yeast to mammals and has also been studied in *P. anserina* [65]. This kinase is the central regulator of the TOR pathway [66], which is known to positively regulate the transcription of genes associated with ribosome biogenesis [67].

Among the GO terms which attracted our special interest were the terms "metal ion transport" and "cis-trans isomerases" (Figure 4). In the first category, the gene encoding the high affinity copper transporter CTR3 was found, which is regulated by the copper-sensing and -binding transcription factor GRISEA [68,69]. From earlier work we know that transcription of *PaCtr3* strongly decreases during aging [70] as the result of an age-related increase in cytoplasmic copper and the repression of transcription factor GRISEA [62]. Apart from *PaCtr3*, genes encoding the two additional copper transporters, PaCTR1 and PaCTR2, the genes *PaMth1* encoding an o-methyltransferase, and *PaFre1*, which codes for a putative ferric reductase, were suggested as potential target genes of GRISEA [22]. In order to validate this conclusion and to identify additional putative genes regulated by transcription factor GRISEA, we compared the 1,202 down-regulated genes from the current age-related study with the 556 down-regulated genes in the grisea mutant (Figure 6a). We found 89 genes to overlap in the two datasets (Table S9). In order to test, whether the number of overlapping genes is higher than expected by chance, we randomly draw from the pool of the 10,059 available genes of the wild-type transcriptome 1,202 genes and subsequently from the 9,704 transcripts of the 'grisea' genome 556 samples. This procedure was performed 10,000 times. The results are compiled in the box plot shown in Figure 6b. The red line indicates that the 89 identified genes lay beyond the upper whisker of the box plot and are extreme outliers. As expected, the results of this random experiment are normally distributed (Figure 6c) with a mean of $\mu \approx 66.05$ and a standard deviation of $\sigma \approx 7.39$. Regarding our observation of 89 common genes, $89 \geq 3\sigma$ holds, indicating that the identification of 89 common genes is very unlikely to occur by chance.

Next, we applied a GO analysis for this set of 89 genes. The most significantly enriched GO terms are associated with the copper metabolism (Figure 6d). While *PaCtr2*, *PaCtr3*, *PaFre1*, and *PaMth1* were found within the overlapping genes, *PaCtr1* was not found in this group (Figure 6e). Comparing the age-related transcript profiles of the five genes revealed that the profile of *PaCtr1* differs strongly from those of the four other genes. After a decrease of transcripts in the middle of the life cycle there is an increase late in life, indicating that the regulation of this gene may not exclusively be controlled by transcription factor GRISEA and cytoplasmic copper. Also the profile of PaMth1 with an increase of transcripts may be influenced by additional regulatory circuits which remain to be elucidated.

In our analysis, the down-regulation of genes, coding for cis-trans isomerases, was rather surprising, since earlier experiments indicated an increase during aging of cyclophilin D (PaCYPD), a mitochondrial enzyme of this group [54]. Moreover, this regulator of the mitochondrial permeability transition pore was identified to induce a cell death program at the end of the life of *P. anserina* [71]. A close look to the individual genes grouped in the list of GO

terms revealed that the gene coding for PaCYPD was not included. The transcription profile of this gene shows that at early life stages the transcript exhibits a fluctuating profile with a strong increase in expression at the last age stages (days 13 and 14).

One significantly enriched group of GO terms within the up-regulated gene set is related to "signaling". Here, various components, including kinases and phosphatases and known components of signal transduction pathways, were identified (Table S8). As the GO terms on "regulation" are also significantly enriched, it becomes clear that during aging there is a strong need to re-adjust the cellular metabolism. This need is in concordance with the down-regulation of important components (e.g., the proteasome) and the accumulation of impairments during aging.

While the proteasome as a component of cellular quality control systems appears to be down-regulated during aging, another cellular quality control pathway, macroautophagy (autophagy), becomes induced (Figure 7). This is indicated by the up-regulation of genes encoding homologs of ATG proteins, signaling components and proteins involved in ubiquitination. Specifically, we found that the abundance of the transcript of an ATG13 homolog, which is essentially for autophagy initiation, increases during aging. Apart from the transcripts coding for an ATG13 homolog, transcripts of ATG6 and VPS15 were identified in the group of up-regulated transcripts coding for putative components of the autophagosome. VPS15 is a serine/threonine kinase, which regulates the activity of VPS34. Both proteins act together in the class III phosphatidylinositol 3-kinase (PI3K) complex necessary for the nucleation and assembly of the initial phagophore membrane and have been shown to be essential in the filamentous fungus *Sordaria macrospora* [72]. On first glance, it is surprising that only *PaVps15,* but not *PaVps34,* is continuously up-regulated during aging. However, from the transcript data it is not clear, which one of the transcripts is limiting and how efficient the corresponding mRNAs are translated. Nevertheless, despite the general difficulties to conclude on the activity of specific proteins, an interesting new aspect arose from the transcriptome analysis.

Autophagy is negatively regulated by the TOR kinase which phosphorylates ATG13. Phosphorylated ATG13 is unable to initiate autophagy via the interaction with ATG1 to form the pre-autophagosomal structure [73]. Both, the transcriptional down-regulation of ribosome genes as well as the up-regulation of macroautophagy transcripts suggests that TOR becomes inactivated during aging of *P. anserina.* It seems that this inactivation acts as a compensatory mechanism to delay deleterious effects of the age-dependent damage accumulation, since TOR inhibition leads to lifespan extension [74–76]. However, obviously, this process is not able to prevent the organismal death finally, especially since in *P. anserina* during aging the mtDNA becomes grossly rearranged, leading to a complete loss of the remodeling capacity of mitochondria-encoded proteins [77]. Autophagy has been reported to occur in different filamentous fungi and is important for fungal development, sexual and asexual reproduction and pathogenicity (reviewed in [78,79]). In a recent study, it has been described that autophagy promotes survival in 'aging' submerged cultures of *Aspergillus niger* [8]. It should be stressed that this kind of 'aging' is different from 'organismic aging' as a process that occurs under nutrient replete conditions. The decline of cellular functions during organismic aging is studied in various aging models (e.g., *P. anserina*, *Caenorhabditis elegans*, *Drosophila melanogaster*, mammals) and is not the result of starvation. Under nutrient starvation autophagy acts mainly as a recycling process to ensure cellular homeostasis even under conditions where the constant supply of nutrients is disturbed. Beside this important role of starvation-induced autophagy, autophagy has been demonstrated to have another important function. It serves as part of a complex quality control system involved in the removal of functionally impaired molecules and organelles (reviewed in: [80–82]). This latter function is of key importance to slow down aging of systems and thus has an important impact on lifespan.

The induction of autophagy as a major quality control pathway under conditions the function of the ubiquitin proteasome system (UPS) is impaired, is in agreement with earlier suggestions that autophagy is a compensatory mechanism, allowing the cell to remove UPS substrates. This conclusion is supported by investigations, using mouse cell lines, in which autophagy was induced by rapamycin and the proteasome inhibited by lactacystin. Under these conditions, lactacystin-induced apoptosis and levels of ubiquitinated protein aggregates were reduced, suggesting that autophagy induction acts as a compensatory pathway [83]. Similar conclusions were drawn in *Drosophila melanogaster* in which autophagy was induced as a result of mutations affecting the function of the proteasome [84]. In another study, Gamerdinger et al. [85] found two members of the BAG family, BAG1 and BAG3, to be key modulators of the proteasomal and autophagic pathways, respectively. While expression of *Bag1* becomes down-regulated during cellular and brain aging, *Bag3* expression is up-regulation in aged cells suggesting a molecular switch between autophagic and proteasomal degradation. In our study, we provide first evidence that autophagy becomes induced during normal organismal aging when the proteasome appears to be functionally impaired. Since ubiquitination is relevant for both, the degradation via the UPS system and for autophagy [86], it is not surprising that the GO term "ubiquitination" is not down-regulated like the GO term "proteasome".

Significantly Differential Expression between a Young and an Old Individual

In a final comparative approach we reduced the transcriptome data sets by comparing only transcript levels in samples of the youngest (6 days) and oldest (14 day) individuals. Subsequently, the data were filtered for all p-values for differential expression of ≤1e-10 leading to 3,976 significantly differentially expressed genes. This gene set contains 2,037 up-regulated and 1,939 down-regulated genes (Table S10). The subsequent GO analysis (Table S11) shows GO terms which were not identified in the same clarity in the former analyses. In particular, genes associated with the energy metabolism, and especially the mitochondrial respiratory chain and the citric acid cycle are significantly enriched within both, the up- and the down-regulated group (Figure 8). The terms in the category "energy metabolism" specifically contains the GO terms "tricarboxylic acid cycle" and "pyruvate metabolic process" and a sub-category of "mitochondrion & respiration" with GO terms, like "mitochondrial part", "respiratory chain", or "cellular respiration". Mitochondria and the energy metabolism are well-known to play a key role in aging of *P. anserina* and other organisms [87–90]. Here, it is of particular interest that only the comparative analyses of two age stages, young vs. old, led to the identification of this correlation which appears to vanish in the longitudinal transcription analysis when all age stages are evaluated. This effect may be due to the fact that the energy metabolism machinery is differentially regulated over the life cycle with fluctuations in gene expression at certain times, according to the energetic demands during development. Bioinformatical processing of these changes, e.g., by applying 'fuzzy' algorithms, may lead to the loss of important information. Applying different bioinformatics approaches in parallel may help to minimize the risk to miss important clues and conclusions.

Conclusions

In this study, we report the results of a genome-wide longitudinal transcriptome analysis of the aging model *P. anserina*. Such an analysis is unique and cannot easily be performed in many other organisms, in particular in long-lived species. In our study, quantitative changes in the transcriptome were followed from an early age to a time-point shortly before the corresponding individuals die. The data represent a unique reference data resource that is valuable also for studies with other organisms. It becomes clear that the expression of genes over the lifetime of an organism does not just follow a simple up- or down-regulation. Many genes are differentially regulated at different times. A reliable comparison of data sets is therefore only possible if the correct samples from a multiple data set are analyzed. Failure in doing so may lead to contradicting results as they are often found in the literature.

The bioinformatical analysis of the transcriptome data in our study led to different relevant conclusions. For instance, we found that after having generated a complete data library it is helpful to reduce the complexity of this library to not miss important information about the processes studied. As a key conclusion of our study, we identified, for the first time, the induction of genes involved in autophagy during normal organismal aging. The induction of this quality control pathway was found to occur after the function of the proteasome and the ribosome is impaired, suggesting that autophagy induction may act as a compensatory pathway in situations when other quality control pathways fail. This is in concordance with earlier findings (reviewed in: [91]). Certainly, the conclusions from our analysis require a careful validation by next generation experiments. In *P. anserina*, such experiments can be performed, specifically by switching from high-throughput analysis to a specific genetic analysis. Experiments to demonstrate the significance of autophagy for aging and lifespan control of *P. anserina* and to elucidate the mechanisms triggering the observed responses have been raised. They will be published in a separate paper.

Supporting Information

Table S1 Summary of primer sequences used in qRT-PCR analysis of selected genes.
(XLS)

Table S2 GO term annotation library of *P. anserine*.
(XLS)

Table S3 Counted tags for each day.
(XLS)

Table S4 Result of the transcriptome analysis.
(XLS)

Table S5 Fuzzy cluster analysis.
(XLS)

Table S6 GO enrichment of transcript profiles in the individual clusters.
(XLS)

Table S7 Compilation of continuously down- and up-regulated genes.
(XLS)

Table S8 GO enrichment analysis of the continuously up- and down-regulated *P. anserina* genes.
(XLS)

Table S9 Putative target genes of transcription factor GRISEA.
(XLS)

Table S10 Up- or down-regulated genes in RNA preparations from 6-day and 14-day old cultures.
(XLS)

Table S11 GO enrichment of up- and down-regulated genes in RNA samples from 6-day and 14-day old cultures.
(XLS)

Text S1 Tag-library correction and normalization to tags per million.
(DOCX)

File S1 Figure S1. Samples for the smoothing approach. Figure S2. Computed Xie-Beni indices for several cluster result.
(PDF)

Acknowledgments

We thank Dr. Nadine Schöne (Goethe University Frankfurt am Main) very much for valuable suggestions for statistical analysis of the data.

Author Contributions

Conceived and designed the experiments: HDO. Performed the experiments: OP AH AW. Analyzed the data: OP. Contributed reagents/materials/analysis tools: AW. Wrote the paper: HDO OP AH IK. Performed homology comparison: JS. Supervised experiment: HDO IK.

References

1. Osiewacz HD (1997) Genetic regulation of aging. J Mol Med 75: 715–727.
2. Kirkwood TB, Austad SN (2000) Why do we age? Nature 408: 233–238.
3. Kenyon CJ (2010) The genetics of ageing. Nature 464: 504–512.
4. Esser K (1974) *Podospora anserina*. In: King RC, editors. Handbook of Genetics. New York: Plenum Press. 531–551.
5. Osiewacz HD, Hamann A, Zintel S (2013) Assessing organismal aging in the filamentous fungus *Podospora anserina*. Methods Mol Biol 965: 439–462.
6. Scheckhuber CQ, Osiewacz HD (2008) *Podospora anserina*: a model organism to study mechanisms of healthy ageing. Mol Genet Genomics 280: 365–374.
7. Rizet G (1953) Sur l'impossibilité d'obtenir la multiplication végétative ininterrompue et illimitée de l'ascomycète *Podospora anserina*. C R Acad Sci Paris 237: 838–855.
8. Nitsche BM, Burggraaf-van Welzen AM, Lamers G, Meyer V, Ram AF (2013) Autophagy promotes survival in aging submerged cultures of the filamentous fungus *Aspergillus niger*. Appl Microbiol Biotechnol 97: 8205–8218.
9. Fabrizio P, Longo VD (2003) The chronological life span of *Saccharomyces cerevisiae*. Aging Cell 2: 73–81.
10. Longo VD, Shadel GS, Kaeberlein M, Kennedy B (2012) Replicative and chronological aging in *Saccharomyces cerevisiae*. Cell Metab 16: 18–31.
11. Osiewacz HD, Kimpel E (1999) Mitochondrial-nuclear interactions and lifespan control in fungi. Exp Gerontol 34: 901–909.
12. Osiewacz HD (2002) Aging in fungi: role of mitochondria in *Podospora anserina*. Mech Ageing Dev 123: 755–764.
13. Osiewacz HD (2011) Mitochondrial quality control in aging and lifespan control of the fungal aging model *Podospora anserina*. Biochem Soc Trans 39: 1488–1492.
14. Osiewacz HD (1992) Genetic control of aging in the ascomycete *Podospora anserina*. Berlin; Heidelberg: Springer. 164 p.
15. Espagne E, Lespinet O, Malagnac F, Da SC, Jaillon O, et al. (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol 9: R77.
16. Osiewacz HD, Clairmont A, Huth M (1990) Electrophoretic karyotype of the ascomycete *Podospora anserina*. Curr Genet 18: 481–483.
17. Osiewacz HD, Skaletz A, Esser K (1991) Integrative transformation of the ascomycete *Podospora anserina*: identification of the mating-type locus on chromosome VII of electrophoretically separated chromosomes. Appl Microbiol Biotechnol 35: 38–45.
18. Osiewacz HD (1994) A versatile shuttle cosmid vector for the efficient construction of genomic libraries and for the cloning of fungal genes. Curr Genet 26: 87–90.
19. Molina C, Rotter B, Horres R, Udupa SM, Besser B, et al. (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9: 553.

20. Kumar L, Futschik M (2007) Mfuzz: A software package for soft clustering of microarray data. Bioinformation 2: 5–7.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.
22. Servos J, Hamann A, Grimm C, Osiewacz HD (2012) A differential genome-wide transcriptome analysis: impact of cellular copper on complex biological processes like aging and development. PLoS ONE 7: e49292.
23. Rexroth S, Poetsch A, Rögner M, Hamann A, Werner A, et al. (2012) Reactive oxygen species target specific tryptophan site in the mitochondrial ATP synthase. Biochim Biophys Acta 1817: 381–387.
24. Brust D, Hamann A, Osiewacz HD (2010) Deletion of *PaAif2* and *PaAmid2*, two genes encoding mitochondrial AIF-like oxidoreductases of *Podospora anserina*, leads to increased stress tolerance and lifespan extension. Curr Genet 55: 225–235.
25. Borghouts C, Kerschner S, Osiewacz HD (2000) Copper-dependence of mitochondrial DNA rearrangements in *Podospora anserina*. Curr Genet 37: 268–275.
26. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, et al. (2013) ArrayExpress update - trends in database growth and links to data analysis tools. Nucleic Acids Res 41: D987–D990.
27. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
28. The UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40: D71–D75.
29. R Core Team (2013) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
30. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23: 257–258.
31. Merico D, Isserlin R, Stueker O, Emili A, Bader DB (2010) Enrichment Map: A network-based method for gene-set enrichment visualization and interpretation. PLoS ONE 5: e13984.
32. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2012) A travel guide to Cytoscape plugins. Nat Methods 9: 1069–1076.
33. Lin WD, Liao YY, Yang TJ, Pan CY, Buckhout TJ, et al. (2011) Coexpression-based clustering of Arabidopsis root genes predicts functional modules in early phosphate deficiency signaling. Plant Physiol 155: 1383–1402.
34. Wang S, Yin Y, Ma Q, Tang X, Hao D, et al. (2012) Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. BMC Plant Biol 12: 138.
35. Nikhil RP, Bezdek JC, Hathaway RJ (1996) Sequential competitive learning and the fuzzy c-means clustering algorithms. Neural Networks 9: 787–796.
36. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. TPAMI 13: 841–847.
37. Scheckhuber CQ, Erjavec N, Tinazli A, Hamann A, Nyström T, et al. (2007) Reducing mitochondrial fission results in increased life span and fitness of two fungal ageing models. Nat Cell Biol 9: 99–105.
38. Borghouts C, Scheckhuber CQ, Stephan O, Osiewacz HD (2002) Copper homeostasis and aging in the fungal model system *Podospora anserina*: differential expression of *PaCtr3* encoding a copper transporter. Int J Biochem Cell Biol 34: 1355–1371.
39. Averbeck NB, Borghouts C, Hamann A, Specke V, Osiewacz HD (2001) Molecular control of copper homeostasis in filamentous fungi: increased expression of a metallothionein gene during aging of *Podospora anserina*. Mol Gen Genet 264: 604–612.
40. Borghouts C, Werner A, Elthon T, Osiewacz HD (2001) Copper-modulated gene expression and senescence in the filamentous fungus *Podospora anserina*. Mol Cell Biol 21: 390–399.
41. Stahl U, Lemke PA, Tudzynski P, Kück U, Esser K (1978) Evidence for plasmid like DNA in a filamentous fungus, the ascomycete *Podospora anserina*. Mol Gen Genet 162: 341–343.
42. Esser K, Tudzynski P, Stahl U, Kück U (1980) A model to explain senescence in the filamentous fungus *Podospora anserina* by the action of plasmid like DNA. Mol Gen Genet 178: 213–216.
43. Esser K, Kück U, Stahl U, Tudzynski P (1981) Mitochondrial DNA and senescence in *Podospora anserina*. Curr Genet 4: 83.
44. Sellem CH, Lecellier G, Belcour L (1993) Transposition of a group II intron. Nature 366: 176–178.
45. Osiewacz HD, Esser K (1984) The mitochondrial plasmid of *Podospora anserina*: A mobile intron of a mitochondrial gene. Curr Genet 8: 299–305.
46. Kil IS, Huh TL, Lee YS, Lee YM, Park JW (2006) Regulation of replicative senescence by NADP+ -dependent isocitrate dehydrogenase. Free Radic Biol Med 40: 110–119.
47. Osiewacz HD, Hermanns J (1992) The role of mitochondrial DNA rearrangements in aging and human diseases. Aging (Milano) 4: 273–286.
48. Osiewacz HD, Hermanns J, Marcou D, Triffi M, Esser K (1989) Mitochondrial DNA rearrangements are correlated with a delayed amplification of the mobile intron (plDNA) in a long-lived mutant of *Podospora anserina*. Mutat Res 219: 9–15.
49. Osiewacz HD, Hamann A (1997) DNA reorganization and biological aging. A review. Biochemistry (Mosc) 62: 1275–1284.
50. Scheckhuber CQ, Mack SJ, Strobel I, Ricciardi F, Gispert S, et al. (2010) Modulation of the glyoxalase system in the aging model *Podospora anserina*: effects on growth and lifespan. Aging (Albany NY) 2: 969–980.
51. Nissan TA, Baßler J, Petfalski E, Tollervey D, Hurt E (2002) 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. EMBO J 21: 5539–5547.
52. Euskirchen GM, Auerbach RK, Davidov E, Gianoulis TA, Zhong G, et al. (2011) Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. PLoS Genet 7.
53. Averbeck NB, Jensen ON, Mann M, Schägger H, Osiewacz HD (2000) Identification and characterization of PaMTH1, a putative *o*-methyltransferase accumulating during senescence of *Podospora anserina* cultures. Curr Genet 37: 200–208.
54. Groebe K, Krause F, Kunstmann B, Unterluggauer H, Reifschneider NH, et al. (2007) Differential proteomic profiling of mitochondria from *Podospora anserina*, rat and human reveals distinct patterns of age-related oxidative changes. Exp Gerontol 42: 887–898.
55. Kunstmann B, Osiewacz HD (2009) The S-adenosylmethionine dependent O-methyltransferase PaMTH1: a longevity assurance factor protecting *Podospora anserina* against oxidative stress. Aging (Albany NY) 1: 328–334.
56. Zhu BT, Ezell EL, Liehr JG (1994) Catechol-O-methyltransferase-catalyzed rapid O-methylation of mutagenic flavonoids. Metabolic inactivation as a possible reason for their lack of carcinogenicity in vivo. J Biol Chem 269: 292–299.
57. Nappi AJ, Vass E (1998) Hydroxyl radical formation via iron-mediated Fenton chemistry is inhibited by methylated catechols. Biochim Biophys Acta 1425: 159–167.
58. Knab B, Osiewacz HD (2010) Methylation of polyphenols with vicinal hydroxyl groups: A protection pathway increasing organismal lifespan. Cell Cycle 9: 3387–3388.
59. Kunstmann B, Osiewacz HD (2008) Over-expression of an S-adenosylmethionine-dependent methyltransferase leads to an extended lifespan of *Podospora anserina* without impairments in vital functions. Aging Cell 7: 651–662.
60. Eriksen JL, Wszolek Z, Petrucelli L (2005) Molecular pathogenesis of Parkinson disease. Arch Neurol 62: 353–357.
61. Ding Q, Dimayuga E, Keller JN (2006) Proteasome regulation of oxidative stress in aging and age-related diseases of the CNS. Antioxid Redox Signal 8: 163–172.
62. Scheckhuber CQ, Grief J, Boilan E, Luce K, Debacq-Chainiaux F, et al. (2009) Age-related cellular copper dynamics in the fungal ageing model *Podospora anserina* and in ageing human fibroblasts. PLoS One 4: e4919.
63. Chondrogianni N, Stratford FL, Trougakos IP, Friguet B, Rivett AJ, et al. (2003) Central role of the proteasome in senescence and survival of human fibroblasts: induction of a senescence-like phenotype upon its inhibition and resistance to stress upon its activation. J Biol Chem 278: 28026–28037.
64. Yiu G, McCord A, Wise A, Jindal R, Hardee J, et al. (2008) Pathways change in expression during replicative aging in *Saccharomyces cerevisiae*. J Gerontol A Biol Sci Med Sci 63: 21–34.
65. Pinan-Lucarré B, Iraqui I, Clavé C (2006) *Podospora anserina* target of rapamycin. Curr Genet 50: 23–31.
66. Martin DE, Soulard A, Hall MN (2004) TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. Cell 119: 969–979.
67. Reiter A, Steinbauer R, Philippi A, Gerber J, Tschochner H, et al. (2011) Reduction in ribosomal protein synthesis is sufficient to explain major effects on ribosome production after short-term TOR inactivation in *Saccharomyces cerevisiae*. Mol Cell Biol 31: 803–817.
68. Osiewacz HD, Nuber U (1996) GRISEA, a putative copper-activated transcription factor from *Podospora anserina* involved in differentiation and senescence. Mol Gen Genet 252: 115–124.
69. Borghouts C, Osiewacz HD (1998) GRISEA, a copper-modulated transcription factor from *Podospora anserina* involved in senescence and morphogenesis, is an ortholog of MAC1 in *Saccharomyces cerevisiae*. Mol Gen Genet 260: 492–502.
70. Borghouts C, Scheckhuber CQ, Werner A, Osiewacz HD (2002) Respiration, copper availability and SOD activity in *P. anserina* strains with different lifespan. Biogerontology 3: 143–153.
71. Brust D, Daum B, Breunig C, Hamann A, Kühlbrandt W, et al. (2010) Cyclophilin D links programmed cell death and organismal aging in *Podospora anserina*. Aging Cell 9: 761–775.
72. Voigt O, Herzog B, Jakobshagen A, Pöggeler S (2013) Autophagic kinases SmVPS34 and SmVPS15 are required for viability in the filamentous ascomycete *Sordaria macrospora*. Microbiol Res.
73. Kundu M, Thompson CB (2008) Autophagy: basic principles and relevance to disease. Annu Rev Pathol 3: 427–455.
74. Powers RW III, Kaeberlein M, Caldwell SD, Kennedy BK, Fields S (2006) Extension of chronological life span in yeast by decreased TOR pathway signaling. Genes Dev 20: 174–184.
75. Jia K, Chen D, Riddle DL (2004) The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. Development 131: 3897–3906.
76. Kapahi P, Zid BM, Harper T, Koslover D, Sapin V, et al. (2004) Regulation of lifespan in Drosophila by modulation of genes in the TOR signaling pathway. Curr Biol 14: 885–890.
77. Kück U, Osiewacz HD, Schmidt U, Kappelhoff B, Schulte E, et al. (1985) The onset of senescence is affected by DNA rearrangements of a discontinuous mitochondrial gene in *Podospora anserina*. Curr Genet 9: 373–382.

78. Pollack JK, Harris SD, Marten MR (2009) Autophagy in filamentous fungi. Fungal Genet Biol 46: 1–8.
79. Voigt O, Pöggeler S (2013) Self-eating to grow and kill: autophagy in filamentous ascomycetes. Appl Microbiol Biotechnol.
80. Tatsuta T, Langer T (2008) Quality control of mitochondria: Protection against neurodegeneration and ageing. EMBO J 27: 306–314.
81. Weber TA, Reichert AS (2010) Impaired quality control of mitochondria: aging from a new perspective. Exp Gerontol 45: 503–511.
82. Fischer F, Hamann A, Osiewacz HD (2012) Mitochondrial quality control: An integrated network of pathways. Trends Biochem Sci 37: 284–292.
83. Pan T, Kondo S, Zhu W, Xie W, Jankovic J, et al. (2008) Neuroprotection of rapamycin in lactacystin-induced neurodegeneration via autophagy enhancement. Neurobiol Dis 32: 16–25.
84. Pandey UB, Nie Z, Batlevi Y, McCray BA, Ritson GP, et al. (2007) HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. Nature 447: 859–863.
85. Gamerdinger M, Hajieva P, Kaya AM, Wolfrum U, Hartl FU, et al. (2009) Protein quality control during aging involves recruitment of the macroautophagy pathway by BAG3. EMBO J 28: 889–901.
86. Lilienbaum A (2013) Relationship between the proteasomal system and autophagy. Int J Biochem Mol Biol 4: 1–26.
87. Harman D (1972) The biologic clock: the mitochondria? J Am Geriatr Soc 20: 145–147.
88. Osiewacz HD, Stumpferl SW (2001) Metabolism and aging in the filamentous fungus *Podospora anserina*. Arch Gerontol Geriatr 32: 185–197.
89. Osiewacz HD (2002) Mitochondrial functions and aging. Gene 286: 65–71.
90. Wallace DC (2001) A mitochondrial paradigm for degenerative diseases and ageing. Novartis Found Symp 235: 247–263.
91. Korolchuk VI, Menzies FM, Rubinsztein DC (2010) Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. FEBS Lett 584: 1393–1398.

PLOS | ONE
TENTH ANNIVERSARY

Correction: A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*

Oliver Philipp, Andrea Hamann, Jörg Servos, Alexandra Werner, Ina Koch, Heinz D. Osiewacz

Published: December 31, 2013 • http://dx.doi.org/10.1371/annotation/03280dea-66ce-4ba6-8ac5-f985f51dea37

The species name "*Podospora anserina*" is spelled incorrectly in the article title. The correct title is: "A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*." The correct Citation is: Philipp O, Hamann A, Servos J, Werner A, Koch I, et al. (2013) A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*. PLoS ONE 8(12): e83109. doi:10.1371/journal.pone.0083109

Citation: Philipp O, Hamann A, Servos J, Werner A, Koch I, Osiewacz HD (2013) Correction: A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*. PLoS ONE 8(12): 10.1371/annotation/03280dea-66ce-4ba6-8ac5-f985f51dea37. doi:10.1371/annotation/03280dea-66ce-4ba6-8ac5-f985f51dea37

Published: December 31, 2013

Competing interests: No competing interests declared.

OPEN ACCESS Freely available online

PLOS ONE

Correction

Correction: A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*

The *PLOS ONE* Staff

In the Funding Statement, an organization providing funding to the last author (HDO) is incorrectly omitted. The Funding Statement should read:

"This work was supported by a grant by the Bundesministerium für Bildung und Forschung (BMBF; GerontoMitoSys 0315584A) and the LOEWE excellence initiative of the State of Hesse within the framework of the Cluster for Integrative Fungal Research (IPF) to HDO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

Reference

1. Philipp O, Hamann A, Servos J, Werner A, Koch I, et al. (2013) A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*. PLoS ONE 8(12): e83109. doi:10.1371/journal.pone.0083109

Citation: The *PLOS ONE* Staff (2014) Correction: A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*. PLoS ONE 9(3): e91590. doi:10.1371/journal.pone.0091590

Published March 25, 2014

4.2 Global protein oxidation profiling suggests efficient mitochondrial proteome homeostasis during aging

Publication status: Published

Summary

The free-radical theory of ageing postulates that ROS induce cellular damage which accumulates during ageing. These cellular impairments may lead to a continuous decline of the organism and finally to death. The impact of ROS on ageing and lifespan was demonstrated in *P. anserina* (Gredilla et al., 2006; Scheckhuber et al., 2007; Kunstmann and Osiewacz, 2008; Wiemer and Osiewacz, 2014) as well as in other organisms (Beal, 1995; Groebe et al., 2007). Basically, the majority of ROS are generated during the process of oxidative phosphorylation in the respiratory chain which is located in the inner membrane of the mitochondria. Obviously, mostly mitochondrial components are affected by oxidative stress. In contrast, different cellular mechanisms exist which act as ROS scavenging systems or as quality control instances which are able to repair or degrade damaged components. Nevertheless, the study aimed to reveal whether and how ROS is affecting and impairing the mitochondrial proteome during ageing in *P. anserina*. Additionally, it was of interest to investigate whether the ratio of damaged and undamaged proteins substantially changed during ageing.

For this purpose the mitochondrial proteomes of six distinct individuals on four different time points in the lifespan of *P. anserina* were extracted. Subsequently, an appropriate, experimental workflow was applied comprising chemical labeling (iTRAQ) of proteins or peptides, a mass spectrometry (MS) approach to measure the single protein ratios and a new statistical pipeline to evaluate and interpret the MS data. The workflow enabled the identification of 22 different types of amino acid related oxidations or modifications respectively, and the relative quantification of modified and unmodified proteins for which the corresponding abundance profiles are provided. The comprehensive statistical pipeline was developed to distinguish these different types of oxidations and to consider the experiment-related, varying quality of the data. It comprises a new mathematical approach to measure the quality of the single data points as well as an appropriate method to detect significant changes in protein ratio.

Summarising, in the study 2,341 proteins could be quantified for which 746 proteins exhibited modified as well as unmodified variants, enabling to compare the degree of oxidation during ageing. Generally, it was found that the majority of the detected proteins show a positive correlation in abundance of modified to unmodified variants in the course of ageing, suggesting that either the impairing effect of ROS

is not as extensive as expected or the various ROS scavenging and quality control mechanisms are highly effective even in older individuals of *P. anserina*.

Authors' contribution

Authors: Carina Ramallo Guevara (CRG), Oliver Philipp (OP), Andrea Hamann (AH), Alexandra Werner (AW), Heinz D. Osiewacz (HDO), Sascha Rexroth (SR), Matthias Rögner (MR) and Ansgar Poetsch (AP)

Percentage of contribution for OP:

Conception/Design:	0 %
Experimental work:	0 %
Implementation:	100 %
Analysis/Evaluation:	25 %
Manuscript:	15 %

OP conceived and developed a bioinformatics and statistical pipeline for data analysis. In particular, OP implemented a web-based species-specific database and platform which comprised experimental data as well as the analysis pipeline. HDO, AP, SR and MR conceived and designed the study. AH and AW acquired isolated mitochondria. CRG acquired proteome data. AH, HDO, AP, OP and CRG analysed and interpreted the data. CRG and OP drafted the manuscript. AH, HDO, AP, OP, CRG, MR and SR critically revised the manuscript and provided intellectual content.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

Research

Global Protein Oxidation Profiling Suggests Efficient Mitochondrial Proteome Homeostasis During Aging*

Carina Ramallo Guevara‡, Oliver Philipp§¶, Andrea Hamann§, Alexandra Werner§, Heinz D. Osiewacz§, Sascha Rexroth‡, Matthias Rögner‡, and Ansgar Poetsch‡‖

The free radical theory of aging is based on the idea that reactive oxygen species (ROS) may lead to the accumulation of age-related protein oxidation. Because the majority of cellular ROS is generated at the respiratory electron transport chain, this study focuses on the mitochondrial proteome of the aging model *Podospora anserina* as target for ROS-induced damage. To ensure the detection of even low abundant modified peptides, separation by long gradient nLC-ESI-MS/MS and an appropriate statistical workflow for iTRAQ quantification was developed. Artificial protein oxidation was minimized by establishing gel-free sample preparation in the presence of reducing and iron-chelating agents. This first large scale, oxidative modification-centric study for *P. anserina* allowed the comprehensive quantification of 22 different oxidative amino acid modifications, and notably the quantitative comparison of oxidized and nonoxidized protein species. In total 2341 proteins were quantified. For 746 both protein species (unmodified and oxidatively modified) were detected and the modification sites determined. The data revealed that methionine residues are preferably oxidized. Further prominent identified modifications in decreasing order of occurrence were carbonylation as well as formation of N-formylkynurenine and pyrrolidinone. Interestingly, for the majority of proteins a positive correlation of changes in protein amount and oxidative damage were noticed, and a general decrease in protein amounts at late age. However, it was discovered that few proteins changed in oxidative damage in accordance with former reports. Our data suggest that *P. anserina* is efficiently capable to counteract ROS-induced protein damage during aging as long as protein *de novo* synthesis is functioning, ultimately leading to an overall constant relationship between damaged and undamaged protein species. These findings contradict a massive increase in protein oxidation during aging and rather suggest a protein damage homeostasis mechanism even at late age. ***Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.055616, 1692–1709, 2016.***

Reactive oxygen species (ROS)[1] are highly reactive intermediates leading to oxidative damage of virtually all biomolecules (1, 2). Mitochondria are known as the main source of endogenous ROS, mainly generated as by-products of oxidative phosphorylation (OXPHOS) at complexes I and III of the respiratory electron transport chain (3, 4). Consequently, mitochondria are inevitably the most prominent target of ROS-induced damage (5). The accumulation of oxidatively damaged macromolecules, particularly proteins, has critical consequences, for example on mitochondrial structure and activity of the respiratory chain (6–8). Accordingly, ROS are involved in several diseases (9, 10), and the free radical theory of aging postulates that the cumulative ROS-induced damage plays a causative role in aging (11, 12). From the chemical and analytical point of view, oxidative damage of proteins is complex and leads to a variety of products, with the accumulation of irreversible oxidative protein modifications contributing to the development of disease (13, 14) and to aging (12, 15).

Although ROS inflicted damage has frequently been reported for higher organisms and humans (16, 17), their long lifespan and laborious molecular manipulation has drawn attention toward alternative model systems such as *Podospora*

From the ‡Plant Biochemistry, Faculty of Biology & Biotechnology, Ruhr University Bochum, Bochum-44801, Germany; §Molecular Developmental Biology, Faculty of Biosciences and Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe University, Frankfurt am Main-60438, Germany; ¶Molecular Bioinformatics, Faculty of Computer Science and Mathematics and Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe University, Frankfurt am Main-60325, Germany

Received September 16, 2016, and in revised form, February 5, 2016

Published, MCP Papers in Press, February 16, 2016, DOI 10.1074/mcp.M115.055616

Author contributions: C.R., A.H., H.O., S.R., M.R., and A.P. designed research; C.R. and A.W. performed research; O.P. contributed new reagents or analytic tools; C.R., O.P., A.H., H.O., and A.P. analyzed data; C.R., O.P., A.H., H.O., and A.P. wrote the paper.

[1] The abbreviations used are: ROS, reactive oxygen species; FA, formic acid; FASP, Filter-aided sample preparation; FDR, false discovery rate; Fwhm, full width at half maximum; HCD, higher-energy collisional dissociation; IAA, Iodoacetamide; iTRAQ, isobaric tags for relative and absolute quantification; *P. anserina*, *Podospora anserina*; PSM, peptide spectrum match; PTM, posttranslational modification; SRM, selected reaction monitoring; TEAB, triethylammonium bicarbonate.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

anserina. This fungus is a well-established model system in aging research because of its senescence syndrome and a short lifespan of ~25 days (18–21). Importantly, for *P. anserina*, it has been shown that in mitochondria the generation of ROS at the electron transport chain increases with age: For instance, the age-related accumulation of endogenous hydrogen peroxide is indicated by increased secretion of this ROS from old individuals. Deletion of a gene encoding for a mitochondrial fission factor leads to a strong increase in the healthy lifespan and goes along with a delay in hydrogen peroxide release in comparison to the wild type (22). The mutation of a nuclear gene, *Grisea*, coding for a transcription factor that is involved in the control of high affinity copper uptake leads to a switch from a copper-dependent standard respiration to an iron-dependent alternative respiration. This switch, because of a loss of respiratory complex III, a major generator of superoxide anion, leads to a decreased generation of this ROS (23). Although carbonylation of proteins visualized by the commonly used Western blot technique did not identify a prominent age-related change of carbonylated proteins in *P. anserina* (24), a strong decrease of carbonylated proteins was found in a strain in which a gene was overexpressed that encodes a mitochondrial methyltransferase which protects against ROS generation. The healthy lifespan of the corresponding transgenic strain was increased by 115% compared with that of the wild type (25). So far, a proteomic view to understand the effects of ROS-induced protein damage, such as carbonylation, in molecular detail during aging is missing.

Because of its high resolution power of (modified) proteins, 2D-electrophoresis has successfully been applied for the identification and quantification of oxidative protein modifications on the proteome level (26). Carbonylated proteins were detected with fluorophore-labeling, differential ProteoTope radioactive quantification (17, 27), and the immunochemical detection technique (25, 28–30). Despite their popularity, antibody and 2D gel-based methods for identification and quantification of oxidative protein damage have well-known limitations, such as under-representation of certain protein categories, limited dynamic range and comigration of proteins (31–33). In particular, these detection techniques allow only the analysis of one specific protein modification at a time. Furthermore, observed oxidative protein modifications by gel electrophoresis have to be interpreted with caution, because proteins can undergo artificial oxidation in polyacrylamide gels (34, 35). Because oxidation results in a specific mass shift, it can be precisely pinpointed with tandem mass spectrometry of intact proteins or their proteolytic digest (36). Hence, gel-free mass spectrometry analyses with previous enrichment step of low-abundant oxidized peptides have been developed (37). However, they are limited to only a few selected protein oxidations (38, 39).

In order to perform a large-scale, unbiased temporal analysis of prominent ROS-induced protein oxidations, we developed a gel-free quantitative proteomic workflow using chemical labeling of peptides with iTRAQ reagents (40, 41). This enables parallel quantification of protein species from mitochondria at different age stages; also, a beforehand data analysis with a novel statistical framework allows an interpretation and comparison of temporal trends of both oxidized and nonoxidized protein species to verify proteome homeostasis during increasing ROS exposure with age.

EXPERIMENTAL DESIGN AND STATISTICAL RATIONALE—

Growth of P. anserina and Isolation of Crude Mitochondria—The cultivation of the wild-type strain s and the isolation of crude mitochondrial fractions from *P. anserina* were performed as previously described (42). Cultures of six individuals serving as biological replicates were harvested at four different age stages (6 days, 9 days, 13 days and 16 days) resulting in a total of 24 mitochondrial samples.

Experimental Workflow for In-filter Protein Digestion and iTRAQ Labeling—The mitochondrial samples were processed according to the in-filter protein digestion (FASP II) procedure described by (43) with minor modifications. FASP combines the advantages of in-gel and in-solution digestion for mass spectrometry-based proteomics and enables the solubilization of crude mitochondria with SDS and detergent removal prior to LC-MS analysis. Because we focus on the mitochondrial proteome a solubilization step is inevitable for high protein coverage. Fifty micrograms of crude mitochondria were denatured and solubilized in SDT-lysis buffer by sonication for 3 min and subsequent shaking at RT for 30 min. Next, samples were mixed with 8 M urea and prepared using the FASP II protocol. All solutions for FASP II included additionally 1 mM DTT and 1 mM EDTA to avoid artificial oxidation during sample processing. In addition, triethylammonium bicarbonate (TEAB) was used as a tertiary amine buffer instead of ammonium bicarbonate. After centrifugation the resulting peptides were acidified with 50% (v/v) ACN in 0.5% (v/v) TFA and dried using a SpeedVac. Digested and carbamidomethylated samples were labeled with iTRAQ 4plex tags according to the manufacturer's protocol. Deviant from manufacturer's protocol, peptide samples were labeled with half the amount of iTRAQ reagent after tryptic digestion for multiplexing all age stages from one of each biological replicate. Briefly, each iTRAQ reagent was dissolved in 70 μl ethanol and the aliquots of iTRAQ 114, 115, 116, and 117 were combined with the peptide mixtures of mitochondria from four different age stages (6, 9, 13 and 16 days), respectively. Consequently, for each experiment 35 μl of reagent solution was used for 50 μg of sample, which was resuspended in 25 μl iTRAQ dissolution buffer (0.5 M TEAB). After 1.5 h of incubation at RT, labeling reaction was stopped with 10 mM glycine. The differently-labeled samples of all 4 age stages from one individual were pooled and dried again in a SpeedVac. Prior to nLC-MS analysis all samples were desalted by solid-phase extraction using Spec C18AR pipette tips according to the manufacturer's protocol. To address any potential label bias, labels for all age stages of one biological replicate were swapped, too. The assignment of label tags to samples is summarized in supplemental Table S1.

Nano-LC-MS/MS Analysis—All desalted samples were resuspended in 2% ACN in 0.1% FA (1 μg/μl) by sonication for 10 min prior to one-dimensional nLC-ESI-MS/MS analysis. Measurements were performed on a LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA) coupled to a nanoACQUITY gradient UPLC pump system combined with an autosampler (all Waters, Milford, MA). The nanoACQUITY UPLC system was equipped with a reversed phase UPLC HSS T3 column (1.8 μm, 75 μm x 250 mm, Waters) and a PicoTip Emitter (Silica TipTM, 10 μm i.d, New Objective, Woburn, MA) as a nanospray source. Four microliters of the sample were loaded directly onto the analytical column using the

TABLE I

All examined single amino acid modifications induced by oxidative damage. Molecular formula and monoisotopic mass shift of the difference between the native amino acid and the oxidized product are given. Based on protein modifications for mass spectrometry (www.unimod.org)

Amino acid	Product	Molecular formula	Monoisotopic mass shift
Arg	Glutamic semialdehyde	−5H −1C −3N +1O	−43.05343
Arg	+14 Da[a]	−2H +1O	+13.97927
Asp	3-hydroxyaspartic acid	+1O	+15.99492
Cys	Sulfinic acid	+2O	+31.98983
Cys	S-nitrosylation[a]	H(-1) N O	+28.990164
His	2-oxohistidine	+1O	+15.99492
His	4-hydroxynonenal[a]	H(16) C(9) O(2)	+156.115030
Leu	+14 Da[a]	−2H +1O	+13.97927
Leu	3-hydroxyleucine	+1O	+15.99492
Met	Methionine sulfoxide	+1O	+15.99492
Met	Methionine sulfone	+2O	+31.98983
Phe	Hydroxyphenylalanine	+1O	+15.99492
Pro	Pyrrolidinone	−2H −1C −1O	−30.01057
Pro	Pyroglutamic acid	−2H +1O	+13.97927
Pro	Hydroxyproline	+1O	+15.99492
Thr	2-amino-3-ketobutyric acid	−2H	−2.01560
Trp	Kynurenine	−1C +1O	+3.99490
Trp	Hydroxytryptophan	+1O	+15.99492
Trp	Hydroxykynurenine	−1C +2O	+19.98983
Trp	N-formylkynurenine	+2O	+31.98983
Trp	6-nitrotryptophan	−1H 1N +2O	+44.98508
Phe	2-nitrophenylalanine	−1H 1N +2O	+44.98508

[a] Precise structure of these products is unknown.

nanoACQUITY autosampler with 99% buffer A (0.1% FA) and 1% buffer B (0.1% FA in ACN) for 60 min. Peptides were eluted by a multiple step gradient of buffer A and B at a flow rate of 300 nL/min in 270 min. Subsequently the concentration of buffer B was increased to 3% within 20 min, followed by a linear gradient to 30% buffer B in 225 min (60–80 min: 3% buffer B; 80–200 min: 13% buffer B; 200–260 min: 20% buffer B; 260–305 min: 30% buffer B). To elute all peptides from the column, the concentration of buffer B was raised to 99% in 20 min and kept constant for 5 min before the column was re-equilibrated at 1% buffer B within 40 min. The column oven was set to 45 °C and the heated desolvation capillary was set to 275 °C. Data-dependent acquisition on the LTQ Orbitrap Velos was operated via instrument method files of Xcalibur (Rev. 2.1.0) in positive ion mode (91) at a spray voltage of 1.3–1.8 kV. The full MS scan was performed in the Orbitrap in a range of 400–1400 *m/z* at a resolution of 60.000 using polysiloxane (*m/z* 445.120024) for internal lock mass calibration (Olsen *et al.*, 2005). The ten most intense precursors per cycle were isolated and fragmented consecutively with CID and HCD. For peptide identification, CID was performed with relative collision energy of 35%, an isolation width of 1.5 Th and an activation time of 10 ms before analysis at normal scan rate and mass range in the linear ion trap. For reporter ion quantification, HCD fragmentation spectra were acquired with normalized collision energy of 65%, an isolation width of 1.2 Th and an activation time of 0.1 ms at a resolution of 7500. Dynamic exclusion was enabled with a repeat count of four and a 90 s exclusion duration window. Unassigned charge states, singly and more than triply charged ions were rejected from MS/MS. For each biological replicate three technical replicates were performed with following MS-settings in Xcalibur: 1. Top10, 2. Bottom10, and 3. Top10 with exclusion list which contained both the retention times and the m/z of previously identified peptides to be excluded for MS/MS. Furthermore, Top10 method meant that each of the ten most intense peaks in a full scan were fragmented in the order of highest to lowest intensity, whereas fragmentation order was from lowest to highest in the Bottom 10 method.

Protein Identification and Quantification—MS data search was performed against a *P. anserina* protein database (version 6.32) from http://podospora.igmors.u-psud.fr/, containing 10612 sequences (44), using the SEQUEST algorithm embedded in Proteome-Discoverer 1.3.0.339 (Thermo Electron © 2008–2011). All acquired raw-files from three technical replicates of one biological replicate were combined in one data search, for which following search parameters were applied: (1) fully tryptic as enzyme specificity, (2) a maximum of two missed cleavages, (3) precursor ion mass tolerance of 5 ppm, (4) fragment ions mass tolerance of 1 Da, (5) Carbamidomethylation of Cys and (6) iTRAQ4plex(N-term) as fixed as well as (7) iTRAQ4plex(K,Y) as variable modifications. The S/N threshold of Peak Filters in the Orbitrap was set to 3. Of note, database searches for suspected protein modifications are limited in the Proteome-Discoverer 1.3.0.339 to four. Therefore, database searches had to be sequentially conducted considering different oxidative amino acid modifications as fixed modification to allow detection of all possible posttranslational modifications (PTMs). All searched oxidative modifications are specified in Table I. Protein quantification is performed with the Reporter Ions Quantifier tool embedded in the Proteome-Discoverer software. Settings were kept at the default values: (1) integration window tolerance of 20 ppm, (2) integration method of most confident centroid, (3) mass analyzer is the Orbitrap with (4) MS^2 order, and (5) HCD as activation type. Reporter based quantification is normalized by the protein ratio median. The factor normalizes all peptide ratios by the median protein ratio. Additionally, to minimize unwanted quantification of co-isolated peptides the allowed relative isolation interference was set to $< 20\%$ of precursor signal intensity. For the determination of the false discovery rate (FDR) a decoy database search was performed with the percolator validation in Proteome Discoverer. The q-value is the minimal FDR at which the identification is considered correct and was set to 1%. The q-values are estimated using the distribution of scores from the decoy database search (45). For data analysis the mass spec format-(msf)-files were filtered with peptide confidence “high” and one peptide per

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

protein with peptide rank 1. Protein or peptide grouping were disabled to achieve the highest number of protein identifications. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository (46) with the dataset identifier PXD001023.

Calculation of Age Dependent iTRAQ Protein Ratios Between the Age Stages—To compare all mitochondrial proteomes from different ages to each other we calculated the following iTRAQ protein ratios for each biological replicate: (1) 114/115 (6/9 days), (2) 114/116 (6/13 days), (3) 114/117 (6/16 days). The iTRAQ protein ratios were calculated as the median over all distinct peptide ratios belonging to a protein. For protein quantification in six biological replicates we performed a multiconsensus report in which all calculated protein ratios were displayed separately for each biological replicate, in order to survey a possibly variance between biological replicates. Consequently, we obtained three different iTRAQ ratios per protein (i, ii, iii) for each biological replicate, resulting in (3 × 6) iTRAQ protein ratios in total, also here referred to as replicate ratios. In this study we carried out several database searches depending on different oxidatively amino acid modifications as well as nonoxidized peptides leading to a total of 23 multiconsensus reports. For detailed information about each protein quantified in multiconsensus reports from six biological replicates and for each of examined modifications all tables are deposited in the ProteomeXchange repository with the dataset identifier PXD001023 as well as all MS/MS spectra of identified and quantified peptides.

Protein Ratios and Computation of a Quality Score for Each Age Stage—To calculate a final protein ratio for each of the three age stages, all individual biological replicate ratios were combined into one iTRAQ protein ratio per age ratio (i, ii, iii) using again the median. Unfortunately, not in all experiments a value for every of the six possible replicate ratios could be measured for each available protein and each age stage. Even if a ratio is available for all of the six biological replicates they often exhibit an experimental variation in strength and direction of their signs (up- and down-regulated).

Hence, because of the need to estimate the actual quality of each single measurement, we computed the quality score q for each protein ratio, enabling us to evaluate the reliability of the computed final protein ratio. The quality score is a value within the interval 0;1 where 1 indicates a highly reliable experimental measurement and 0 an unreliable protein ratio because of strong experimental variation. It is important to highlight that we do not use this quality score in order to test for significance of the expression at one time point in contrast to another. We needed this scoring function only to decide whether the measurement of a single protein ratio under a specific condition is reliable; because our data exhibit the drawbacks described in results sections (see also supplemental Fig. S2).

The quality score bases upon three assumptions and was computed as described in the following. Per protein P and experiment E, we obtain a set $N_{P,E} = \{x_1, x_2, \ldots, x_6\}$ of 6 results. A result x_i is either a real number for the ratio of abundances of the protein P or x_i set to "not a number" (NA) if no peptide of the protein has been detected in the *i*th replicate of experiment E. Obviously, a result x_i = NA or $x_i = 1$ produces no valuable information on the up- or down-regulation of the protein. We eliminate these results and get a subset $M_{P,E} \leq N_{P,E}$ of *m* positive real numbers that are either larger than one (up-regulated) or smaller than one (down regulated). The assignment of a protein P to be up/down regulated in experiment E relies on the set of replicates $M_{P,E}$. An unambiguous assignment is impossible if half of the results in $M_{P,E}$ indicate up-regulation and the other half down-regulation. The number of values that identifies a unique type of regulation is relevant for reliability assignment. The same is true for the variance of real values in $M_{P,E}$. To account for the reliability and unambiguousness of our assignment of a protein P to be up or down-regulated we compute a score based on three features of $M_{P,E}$:

(1) Number of replicates in which a specific protein is detected. In the best case we obtain one ratio for each of the six possible replicates per protein that would be rated with 1 and in the worst case - no detected replicate ratio - with 0:

$$q_n = \frac{m}{n} \quad \text{(Eq. 1)}$$

For example: Of the six possible replicate ratios we detect four: $\frac{4}{6} \approx 0.66$ (see also supplemental Fig. S2*a*).

(2) Quotient of up- to down-regulated replicates. In the best case each detected replicate exhibits the same tendency. That is all replicates are either up- or down-regulated which should lead to a value of 1. In the worst case there are as many up-regulated as down-regulated replicates, which should be rated with the worst value of 0:

$$r_{ratio} = \frac{\max(d, u)}{m} \geq 0.5 \quad \text{(Eq. 2)}$$

where d is the number of down-regulated and u the number of up-regulated replicates in M. Hence, in the worst case we now get a value of 0.5 (instead of 0), because there are as many up- as down-regulated replicates. Next, we "stretched" the interval 0.5;1 to the required interval 0;1:

$$q_r = \frac{r_{ratio} - 0.5}{0.5} \quad \text{(Eq. 3)}$$

(see also supplemental Fig. S2*b*)

(3) The variance of the replicates ratios. Next, we have to evaluate the variances of the replicates. In the best case there is no variance in M and all ratios exhibit exactly the same values which should be rated with 1. For this purpose we use the coefficient of variance $v_c = \frac{\sigma(M)}{\overline{M}}$ which results in 0 if there is no variance among the ratios and it does not have an upper limit. To map this one-sided open range to the interval (0, 1) we computed the quality of variance with

$$q_v = e^{-v_c} \quad \text{(Eq. 4)}$$

That is if the ratios show no variance we calculate 1 and the greater the variance the more q_v approaches asymptotically 0:

$$\lim_{v_c \to \infty} q_v = 0 \quad \text{(Eq. 5)}$$

Finally, we combined all values to one quality score using the geometric mean—

$$q = \sqrt[3]{q_n \times q_v \times q_r} \quad \text{(Eq. 6)}$$

which ensures that q is 0 if one of the three values is 0.

We assume a protein ratio to be reliable if $q \geq 0.6$ holds. This semi-arbitrary threshold bases upon the fact that two of the three subvalues (q_n and q_r) relied on discrete values (*n* and *m*). A threshold of 0.6 for both subvalues ensures that more than half of the possible replicates are available, and more than twice as many replicates for a protein ratio are up- than down-regulated and *vice versa*. Furthermore, because of the fact that the coefficient of variation is expressed as percentage and is subsequently stretched to the interval [0, 1], a threshold of 0.6 ensures that the protein ratios do not vary more than 50%. A subsequent manual checkup of the q-values confirmed that 0.6 indicates a reliable protein ratio (see examples in supplemental Table S2). For example, if there are 3 replicates available out of the 6

possible replicates this results in $q_n = \frac{3}{6} = 0.5$ whereas four possible (more than half as many) replicates lead to $q_n = \frac{4}{6} = 0.\overline{6}$.

Criteria for Significant Up- and Down-regulation—The Reporter Ions Quantifier tool exported the replicate ratios always compared with the lowest mass iTRAQ reagent for each biological replicate, which was in our case the youngest sample: (i), (ii), (iii). Conversely, for a more intuitive interpretation and especially to compare the protein ratios easier with previous studies, we determined the inverses for all ratios; that is subsequently the following protein ratios are considered as abundance changes of oxidized or nonoxidized species from an older in comparison to a younger individual (iv) 115/114 (9/6 days), (v) 116/114 (13/6 days), (vi) 117/114 (16/6 days).

To decide whether a protein (modified- or unmodified) was up- or down-regulated on a certain day in respect to the reference day 6 we had to choose a reliable threshold. Unfortunately, on average the fold changes were very close to 1, that is we had to take into account that changes of protein abundances in our experiments normally are small.

Furthermore, regarding the drawbacks of our measurements described in results sections (see statistical approach: quality score) it was not reasonable to use a conventional parametric significance test to check whether the abundance of a protein (modified or unmodified) is high or low. For example, a two sided *t* test presumes a normal distribution and an appropriate sample size. It also just quantitatively takes into account the variances or the means, respectively, of the single (actually available) values and of each local distribution but not the qualitative properties like same regulative direction. We take into account the latter in combination with the quality score by using an appropriate approach which considers the complete data and all distributions as one.

We computed specific thresholds based on the distribution of our data. For this purpose, we first computed for all ratios the respective logarithmic ratios (to the base of 2) to obtain symmetric behavior. Subsequently, we calculated the quartiles for all of the logarithmic ratios and defined the 25% and the 75% quartile as our lower- and upper-threshold, respectively. That is 50% of all ratios lay within this interquartile range and all others beyond these borders (see supplemental Table S3 and are stronger down- or up-regulated than 50% of all other ratios. The supplemental Table S3 indicates that these 50% of all ratios are within the interval [0.9, 1.54]. Hence, a decrease of more than 10% and an increase of more than 54% in protein abundance were assumed to be significant.

Comparison of Modified and Unmodified Protein Profiles—With our experimental approach it is possible to detect and quantify both oxidized and nonoxidized species for each protein via iTRAQ in a single analysis, which largely eliminates technical variance. Therefore, comparing the abundance changes of modified and unmodified species at different age stages is possible, and we recorded the trend of changes on the peptide levels in protein oxidation and protein amount over the age stages in order to determine age-related protein damage. By this means, the actual relative change in oxidation degree for a certain protein during aging based on the determined ratio of oxidized to nonoxidized species could be calculated.

To test whether abundance changes of oxidized and nonoxidized protein species are similar, we compared these profiles and decided whether they are equal by means of the two-sided two-sample Kolmogorow-Smirnoff-Test, which tests for differences between two one-dimensional distributions (47). The *n* replicate-ratios for profile 1 and the *m* replicate-ratios for profile 2 were used to compute the empirical distribution functions $F_{1,n}$ and $F_{2,m}$. Hence, the two-sided hypothesis can be formulated as

$$H_0{:}F_1(x) = F_2(x) \quad \text{(Eq. 7)}$$

(similar or indifferent distributions or profiles, respectively) and

$$H_1{:}F_1(x) \neq F_2(x) \quad \text{(Eq. 8)}$$

(different distributions or profiles, respectively).

Subsequently, based on the D-statistic

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad \text{(Eq. 9)}$$

a *p* value has been computed as described in (48) and a significance threshold of 0.05 was applied, meaning that if the *p* value was less than or equal to 0.05 the null-hypothesis was rejected and we assumed that two profiles were significantly different. To control for multiple hypothesis testing we applied the false discovery rate method introduced by Benjamini and Hochberg (49).

Confirmation of Both Oxidized and Nonoxidized Species for Selected Proteins Via SRM MS—For selected proteins, additional MS experiments were performed to confirm the data obtained by iTRAQ quantification.

Relative quantification of the protein abundance ratios using crude AQUA peptides was performed using a triple quadrupole mass spectrometer (TSQ Vantage, Thermo Fisher Scientific) in SRM mode. Twenty-four suitable peptides, 3 per protein, were selected from the iTRAQ data set based on sequence uniqueness, length between 5–20 amino acids, missed tryptic cleavage sites, and amino acid oxidation predisposition. Corresponding crude heavy peptides (nontagged SpikeTides L) were synthesized from JPT Peptide Technologies containing C-terminal lysine or arginine stable isotopes to induce mass shifts of 8 or 10 Da per peptide (see on Passel PASS00738). Four transitions per peptide were selected based on maximum signal intensities observed during nLC-iTRAQ MS/MS and collision energy for each peptide was generated *in-silico* using Skyline software (version 2.6.0.7176; MacCoss Lab, University of Washington; (50) (see on Passel PASS00738). The four age stages of the six independent biological replicates were prepared in the same manner as described in experimental section above (Experimental workflow for in-filter protein digestion and iTRAQ labeling), however without iTRAQ labeling step. One hundred femtomole or one picomole of each heavy peptide was spiked into the tryptic digested samples prior to the desalting procedure (see on Passel PASS00738). SRM measurements were performed as described previously by (51) with slight modifications. Samples (5-μl injection) were loaded onto the column with 2% buffer B. Peptides were eluted from the column with a multistep gradient of buffer A and buffer B which was established as follows: for 0–5 min: 2% buffer B; for 5–10 min: 2–5% buffer B; for 10–71 min: 5–35% buffer B; for 71–77 min: 35–85% buffer B; and for 77–105 min: 2% buffer B. The SRM instrument method consisted of one SRM scan event over a 105 min run-time whereby 206 transitions were measured with a 10 s cycle time (0.05 ms scan time per transition). Fixed parameters were 0.2 fwhm Q1 for precursor ions and 0.7 fwhm Q3 resolution for product ions.

For targeted quantification of the corresponding proteins with their oxidative modification sites, the six independent iTRAQ-labeled samples (see supplemental Table S1) were remeasured on the TSQ Vantage in SRM-mode. The selected oxidized peptides, precursor ions, SRM transitions and collision energies used for this analysis are deposited to the PeptideAtlas (PASS00738). Collision energy for each peptide was generated again by the skyline software and further experimentally refined for the iTRAQ-labeled peptides according to signal-to-noise measurements during SRM trials. The SRM instrument method consisted of one SRM scan event over 105 min runtime, whereby for all SRMs a scan width of 0.01 m/z and a scan time of 0.02 s was set, and fixed parameters were 0.2 fwhm Q1 and 0.7 fwhm Q3 resolution for peptide fragments and 0.3 fwhm Q3 resolution

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

for iTRAQ reporter ions. Each biological replicate was analyzed three times.

All SRM-data have been deposited to the PeptideAtlas SRM Experiment Library (PASSEL) and are accessible via the website http://www.peptideatlas.org/PASS/PASS00738 (52).

SRM Data Analysis—For SRM data normalization of nonoxidized protein species the total protein amount of each sample was determined by precursor ion quantification (53). For this normalization, all 24 samples from SRM experiment were measured on a LTQ Orbitrap XL mass spectrometer, and for each sample the total protein area from all identified protein areas was summed to generate a normalization factor. More details are available in supplemental methods. Data analysis for age-related protein abundance changes was carried out by using Skyline 3.1.0.7382 software (50). All SRM-data were manually inspected to ensure correct peak identification whereby not accurately identified peptides based on selected transitions were excluded from the data set. Further, two samples were excluded from the further data analysis, based on peculiarities of the samples exhibiting shifting of retention times for all peptides and a lower peptide and protein identification rate compared across all samples. The ratios between the peak areas of each light and heavy peptide were calculated using Skyline and exported to Excel for further statistical analysis. To account for differences in protein amount across the samples, the peptide peak ratios of the different samples were normalized based on the total protein area of each corresponding sample. As described above, the equal age ratios for each peptide of one of each biological replicate were calculated: (i), (ii), (iii). Subsequently we determined the inverses for all age ratios: (vi), (v), (vi) and logarithmised to base 2. The final protein age ratios were calculated as the median over all logarithmic distinct peptide ratios belonging to a protein. Additionally, the mean and the standard deviation were calculated in the same manner (supplemental Table S8).

Data analysis for age-related changes in protein oxidation of the corresponding proteins was performed by using Skyline 2.6.0.7176 software (50). Again all SRM-data were manually inspected to ensure correct peak identification and not accurately identified peptides by transition peaks were excluded from data set. Calculated areas of the iTRAQ reporter ions from each biological and technical replicate were summarized for each protein in an excel worksheet, and the same iTRAQ ratios per age ratio were computed: (i), (ii), (iii). All iTRAQ age ratios were normalized on the factors from respective biological replicate which were obtained during data analysis for unmodified protein species via Proteome Discoverer 1.3, and were therefore most reliable for normalization. The inverses for all iTRAQ age ratios were determined: (vi), (v), (vi) followed by the logarithm to base 2. The median was computed over logarithmic iTRAQ age ratios from technical and biological replicates for final protein age ratios. Furthermore the mean and the standard deviation over logarithmic iTRAQ age ratios from technical and biological replicates were determined (supplemental Table S9).

Estimation of Modification Site Occupancy—Peak areas of proteins and of modified and unmodified peptides were obtained from Proteome Discoverer 1.4.1.14. Fractional intensity of each of the four iTRAQ labeled samples was calculated from respective peptide peak area and iTRAQ reporter ratios so that all 4 labeled samples sum up to the peptide peak area. For this purpose, the normalized (setting protein median ratio between each sample to one) relative intensity of iTRAQ reporter ratios was used. For each peptide of each age state and biological replicate the median fractional peptide intensity was calculated from first all quantified peptide spectrum matches and second from all three technical replicates. Then fractional peptide peak areas were normalized between biological replicates by division with the corresponding protein peak area in each biological replicate. Site occupancies in the range between 0 (no modification) and 1 (completely modified) were calculated under the simplified assumption that a) total peak area of a peptide is the sum of the peak areas of the one modified and the one unmodified peak area, and b) that the "flyability" (*i.e.* ionization efficiency) of modified and unmodified peptide is identical based on the formulas given in (54). Occupancy calculations were realized by using an in-house Perl script.

RESULTS

Optimal Experimental Workflow for Comprehensive Data Acquisition of Oxidative Protein Modifications and Global Proteome Analysis—The technical goal of this study was to develop an in-solution sample processing workflow (Fig. 1*A*) using iTRAQ labeling to enable the simultaneous, unbiased identification and quantification of oxidized and nonoxidized proteins from mitochondria of *P. anserina* to comprehensively analyze aging-related modified proteins. To avoid artificial protein modifications from SDS-PAGE, we adapted the filter-aided protein digestion (FASP) prior to iTRAQ labeling. Labeling efficiency was determined by comparison of database searches conducted with or without iTRAQ modification as parameter for each biological replicate (see supplemental Table S4). Based on these results, we can conclude that an average labeling efficiency of 92% was achieved in our experiments. Furthermore, one experiment was repeated with a label swap in order to verify any potential labeling bias, and we found that this switch did not significantly affect iTRAQ protein ratios; an eventual label bias can be neglected for regulated proteins (see supplemental Table S5).

iTRAQ labeling enables pooling the four age stages of one biological replicate and allows for simultaneous relative quantification of oxidized and nonoxidized protein species within a single LC-MS/MS run. This reduces experimental variance; however, it also leads to higher sample complexity. Moreover, a critical factor of reporter ion quantification is precursor mixing (55), leading to compression of observed iTRAQ protein ratios. This analytical challenge could be properly addressed by achieving high separation power with a 25 cm column and an optimized four-hour linear solvent gradient, resulting in over 1000 quantified proteins. Additionally, each experiment was analyzed on a LTQ Orbitrap Velos with CID-HCD coupled method in three technical replicates (Top10, Bottom10, Exclusion list-Top10) to account for the expected low amount of modified (oxidized) protein species. Combining all acquired raw-files from three technical replicates of one experiment in a single database search increased ~3-fold the number of quantified iTRAQ-labeled peptides. Mainly, the Bottom10 method enlarged the detection of unique peptides of proteins by an average of 655 (Fig. 1*B*). To verify the robustness of our experimental setup, technical variance and biological variance were estimated from three biological replicates. Therefore, we calculated the mean standard deviation of measured protein ratios related to their age stage for technical and biological replicates and confirmed that the mean standard deviation from technical replicates is smaller than the mean standard deviation from biological replicates (sup-

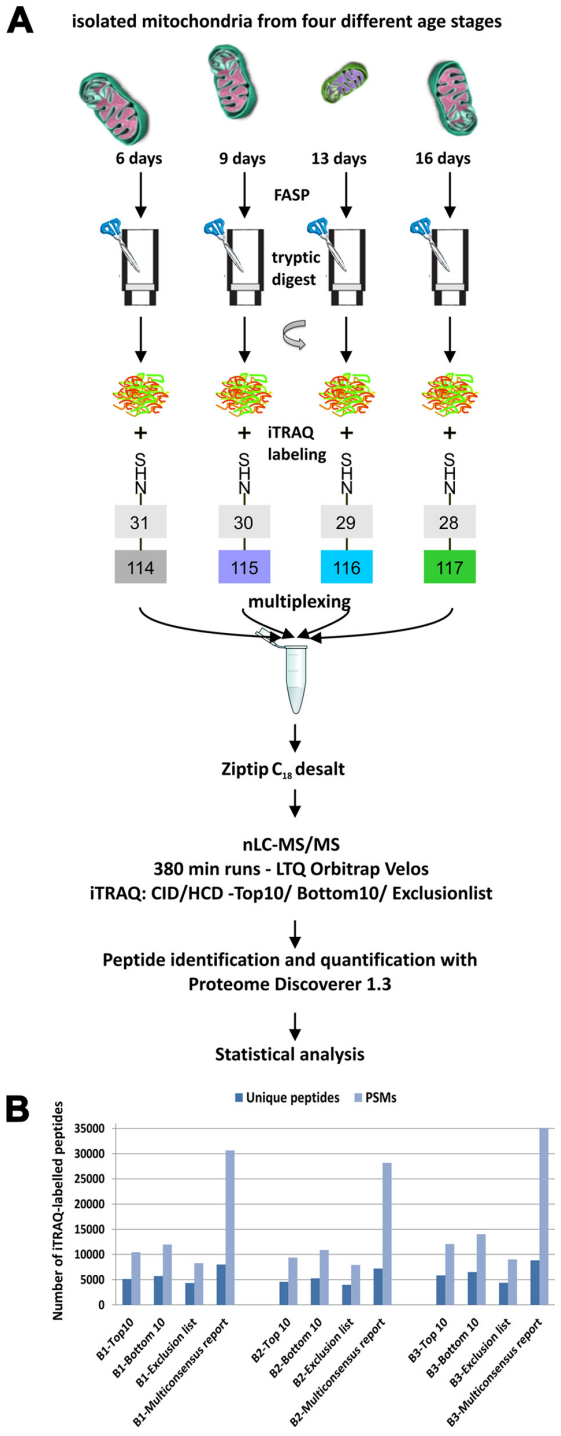


FIG. 1. **The proteomic strategy for global quantification of oxidative protein damage.** *A*, Experimental workflow for iTRAQ-nLC-MS/MS. Whole mitochondria from four different age stages were prepared gel-free using the FASP protein digestion method. Peptides were labeled with iTRAQ and combined to analyze all age stages of one biological replicate in one single 380 min nLC MS/MS run. All samples were acquired on a LTQ Orbitrap Velos instrument with CID/HCD method in three technical replicates (Top10, Bottom10, Exclusion list-Top10). Peptides were identified and quantified with Proteome Discoverer 1.3. *B*, Number of unique peptide sequences as well as total number of identified peptide spectral matches (PSMs) quantified with each of the three MS-methods, shown for three biological replicates ($n = 3$). Higher numbers of quantified iTRAQ-labeled peptides were gained by merging of technical replicates in one database search. Notably, the Bottom10 method enlarged the detection of unique peptides of proteins by an average of 655. Detailed list of iTRAQ protein ratios, their averages and standard deviation between technical replicates can be found in supplemental Table S4.

plemental Fig. S1). In detail, we detected a mean standard deviation of 0.08 for technical variance and 0.10 for biological variance in day 9, 0.07 and 0.11 in day 13 and finally 0.09 and 0.15 in the oldest age stage (day 16). A complete overview of measured iTRAQ protein ratios from representative experiments, as well as mean values and standard deviations between technical and biological replicates can be found in supplemental Table S6.

General Survey of Global iTRAQ-based Proteome Analysis—Samples of six independent *P. anserina* individuals were analyzed according to the defined experimental design (Fig. 1*A*). The optimized proteomic strategy allowed an identification and quantification of in total 2341 proteins, whereof for a number of 746 both protein species (unmodified and oxidatively modified) were detected (Fig. 2*A*). No unmodified peptides could be quantified for 333 proteins. In these cases, only the oxidized protein form is listed. Overall, the iTRAQ proteomic analysis revealed remarkably moderate changes in protein oxidation and total protein abundance. Fig. 2*B* provides a compilation of quantified protein oxidation sites before statistical analysis of temporal changes and reveals a high level of mono-oxidation in amino acid residues. The data show that methionine is the main oxidation site within the detected proteins, which is reasonable because of the easy oxidation of sulfur-containing amino acid side chains. Other prominent identified modifications in decreasing order of occurrence are mono-oxidation of several amino acid residues, then carbonylation as well as formation of N-formylkynurenine and pyrrolidinone.

Statistical Approach and Time Course of Oxidative Protein Damage—In order to determine age-related protein damage, the ratios of iTRAQ-labeled peptides from four age stages were used to calculate the trend in protein oxidation and amount. Unfortunately, advanced bioinformatics tools for a comprehensive, comparative analysis of oxidized protein species and their nonoxidized counterparts in complex samples are currently missing, which made it necessary to develop a new, customized statistical approach. First, computational analysis and evaluation of the measured iTRAQ data had to be established: a quality score q for each protein ratio from six experiments was calculated that accounted for the deviation

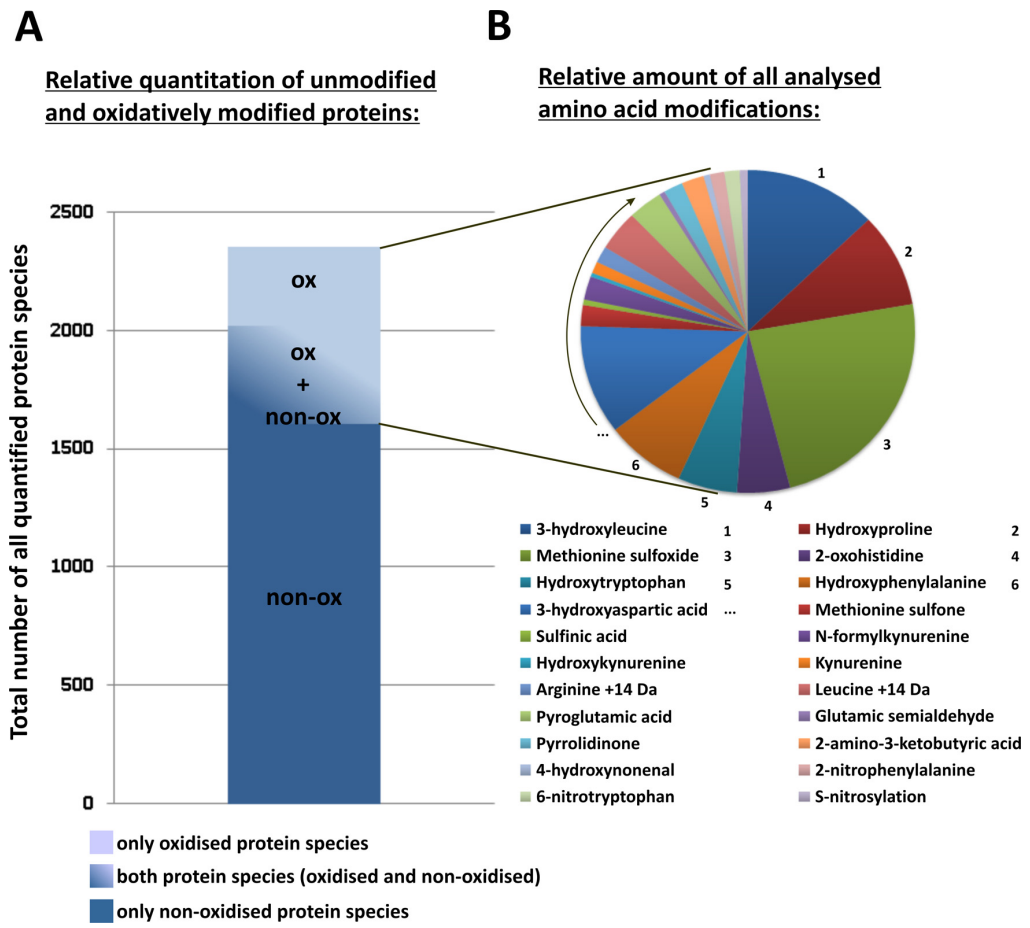


FIG. 2. **General survey of global iTRAQ-based proteome analysis (n = 6).** *A*, Total number of all quantified protein species in nonoxidized and oxidized form. In total, 2341 proteins were quantified of which 32% of proteins carried an oxidative amino acid modification (*B*). Relative amount of all analyzed amino acid modifications. Database searches using Proteome Discoverer 1.3 were sequentially conducted to identify 22 different oxidative amino acid modifications.

of our data from normal distribution and missing data values, which enabled to evaluate the reliability of the computed final protein ratio for each age stage. To clarify the need for such a quality score we illustrate the relative frequencies of measurements with missing replicate values and different regulative direction using two barplots (see supplemental Fig. S2). Overall, only a third of all protein ratios are computed using six single values deduced from six available replicates (supplemental Fig. S2*a* - rightmost bar). In addition, only half of all protein ratios exhibit a distinct regulatory direction when going to the level of the corresponding single replicate ratios. That is all available replicates for one protein ratio are indicating an up- or down-regulation (supplemental Fig. S2*b* - rightmost bar). Secondly, a specific threshold based on the box plot distribution of our data was computed to decide whether a protein (modified or unmodified) was significantly up- or down-regulated on a certain day in respect to the reference day 6 (supplemental Table S3). After application of these defined statistical criteria on our proteome data, still 20% of oxidized proteins remained suitable for further analysis of their regulation.

Fig. 3*A* illustrates the distribution of modified proteins before and after the statistical analysis and exposes the total number of proteins with significantly altering abundance during aging, sorted by the oxidative modification. In the end, 18 types of protein oxidation remained after applying statistical criteria with methionine oxidation still being the most prominent, and were subjected to further data analysis. As depicted in Fig. 3*B*, these significant, differentially oxidized proteins were plotted against their calculated abundance ratio for each age stage to gain a complete overview of oxidation trends during the course of aging: concomitantly with increasing age, there is an increase in the number of proteins with significant abundance and oxidation changes. However, there exists no age-related uniform trend of changes in protein oxidation

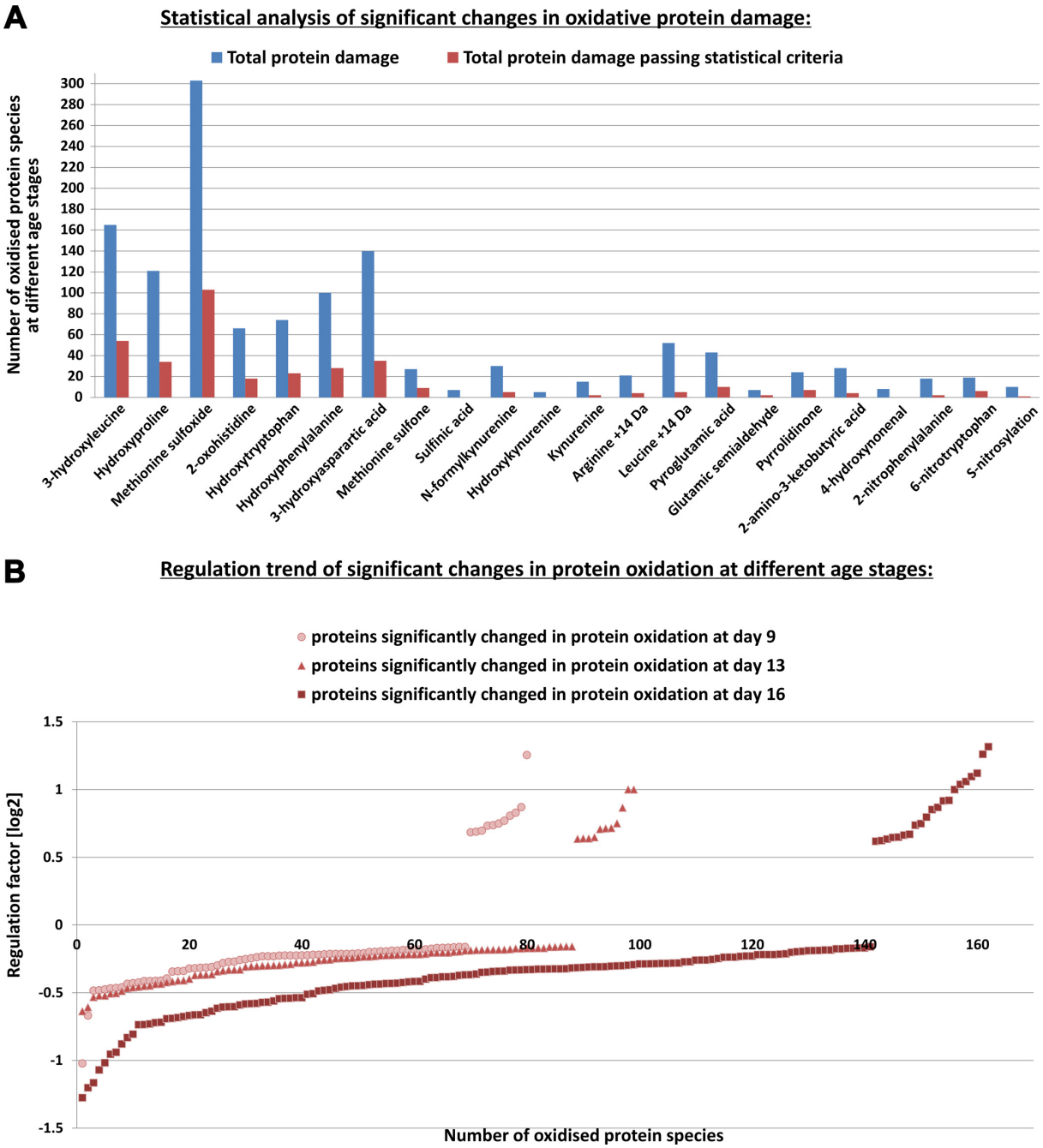


FIG. 3. **Statistical analysis of significant changes in oxidative protein damage (n = 6).** *A*, Protein modifications between different age stages passing defined statistical criteria are shown in red. Criteria are: 1. Number of quantified biological replicates ≥3, 2. Consistent up- or down-regulation of protein damage, 3. Variance of protein damage between biological replicates. In total 20% of protein modifications passed statistical criteria and were subsequently used for further data-analysis. *B*, Regulation trend of significant changes in protein oxidation at different age stages. Most proteins show a decrease in oxidative damage during the course of aging. Few proteins exhibit small increases in abundance of protein oxidation of which most were quantified in the oldest age stages.

because individual proteins display an increase as well as a decrease in protein damage with higher age.

Age-related Changes in Protein Damage Based on Only Oxidized Protein Species—Usually proteomic studies target only the modified (here oxidized) proteins in complex biological samples for subsequent identification and quantification. Even though our data affords quantification of both the oxidized and cognate nonoxidized peptide species (see the following sections), we investigated first only the oxidized protein species. By this means, we are able to compare our results with previous works on oxidative protein modification in *P. anserina*, and to analyze those proteins, where only quantification of the oxidized peptides was possible. supplemental Fig. S3 shows the proteins with the highest significant age-dependent changes in oxidation level, ignoring their unmodified counterparts (if detected). In accordance with the statistical analysis mentioned in the previous chapter, only oxidized protein species with at least 0.66 (log2[-0.6])-fold decrease or 1.5 (log2[0.6])-fold increase in amounts normalized to day 6 had to be considered and belonged to 8 diverse biological processes based on the clusters of orthologous groups of proteins (COGs from www.jgi.doe.gov) (56, 57). Below, we describe some proteins in more detail because of their protective role against aging.

Three chaperones (Pa_1_6520, Pa_3_9400, and Pa_6_2570) were found with increased protein damage of about 1.5 (log2[0.6]) to 2 (log2[1])-fold during aging. One of them, the mitochondrial protein SSC1 (Pa_6_2570) belongs to the 70-kDa heat shock protein family, functioning as molecular chaperone to protect cells against the adverse effects of stress. For five proteins only oxidized peptides could be detected, suggesting they are prone to ROS-induced oxidation. These were a putative oxygenase (Pa_4_520), the putative cytochrome P450 (Pa_1_9520) and a putative transporter protein (Pa_4_6490), which all showed an overall decrease in protein damage. Increase of oxidation during aging was observed for the putative methyltransferase (Pa_5_11950). Additionally, the putative polyketide synthase (Pa_1_11870) showed a sudden 1.5 (log2[0.6])-fold increase in protein damage from day 13 to 16.

Changes in Abundance and Oxidation Positively Correlate for the Majority of Proteins During Aging in P. anserina—Comparison of trends between modified and unmodified protein species were statistically evaluated by correlation analyses. Thus, we compared the trends of all oxidized protein species passing the statistical criteria with their unmodified counterpart and verified with the two-sided two sample Kolmogorow-Smirnoff test, whether for a given protein the trends for abundance changes of modified (oxidized) and nonmodified peptides positively correlate or not during aging (supplemental Table S7). After computational analysis of all statistically relevant proteins, only 17% of these proteins exhibit a significant difference (p value $\leq$ 0.05) between the trend profiles of both protein species (see next section), whereas 69% exhibit a positive correlation. In addition, we controlled for multiple hypothesis testing by applying the false discovery rate method by Benjamini and Hochberg with a p value $\leq$ 0.05. Under these even stricter criteria, only ~6% exhibit significantly different trend profiles (see supplemental Table S7 with details for all examined modifications). These findings affirm our conclusion that the majority of proteins revealed a consistent ratio of oxidized to nonoxidized species during aging, and that it is mandatory to compare changes in protein oxidation with changes in protein abundance for correct interpretation of the aging process. Accordingly, we uncovered that changes in oxidative damage in those proteins listed in supplemental Fig. S3 correlate with their respective abundance changes for all proteins where an unmodified counterpart was found.

Proteins with a Changing Ratio of Oxidized to Nonoxidized Protein Species During Aging—Our findings revealed that no global increase in protein damage occurs in mitochondria from *P. anserina* in the course of aging. In this section, we will present the few proteins with uncorrelated trend profiles for abundance changes of protein ratios based on oxidized and cognate nonoxidized peptides. In particular, proteins prone to increasing ROS-induced damage during aging should be important for decaying mitochondrial function.

Forty-three proteins displayed a significant difference in trend profiles ($p \leq 0.05$) determined by the two-sided two sample Kolmogorow-Smirnoff test and are listed in supplemental Fig. S4, and after additional Benjamini and Hochberg test still 14 of them remained. Those with p values ≤ 0.05 for both tests are highlighted in supplemental Fig. S4 (red asterisk) and marked with the symbol [*] throughout the text. Notably, half of these proteins are participating in energy metabolism, as depicted in supplemental Fig. S4. Among these proteins, most exhibited an overall decreasing oxidation degree with aging. However, for some proteins first an increase and then a decline at older age stages and conversely was observed, indicating a more complex process during aging (supplemental Fig. S4). Frequently several probable amino acid oxidations were discovered for a certain protein, often with similar trends, exemplified by the putative electron transfer flavoprotein beta-subunit 6 (Pa_6_1650). Still, one must bear in mind that determination of the exact localization of the oxidized amino acid residue within a peptide sequence was very challenging and often not possible, because often a complete series of fragment ions is required for unequivocal assignment. Therefore, we presented and quantified all possible (according to database search results) amino acid modifications for a given peptide to monitor the degree of protein oxidation in supplemental Fig. S4.

Of particular interest is the putative mitochondrial ATP synthase gamma chain (Pa_1_9450) [*] that revealed a continuous increase in oxidative damage up to a 2.4 (log2[1.2])-fold change, whereas abundance of the nonoxidized protein species was unaltered during aging. This protein subunit belongs

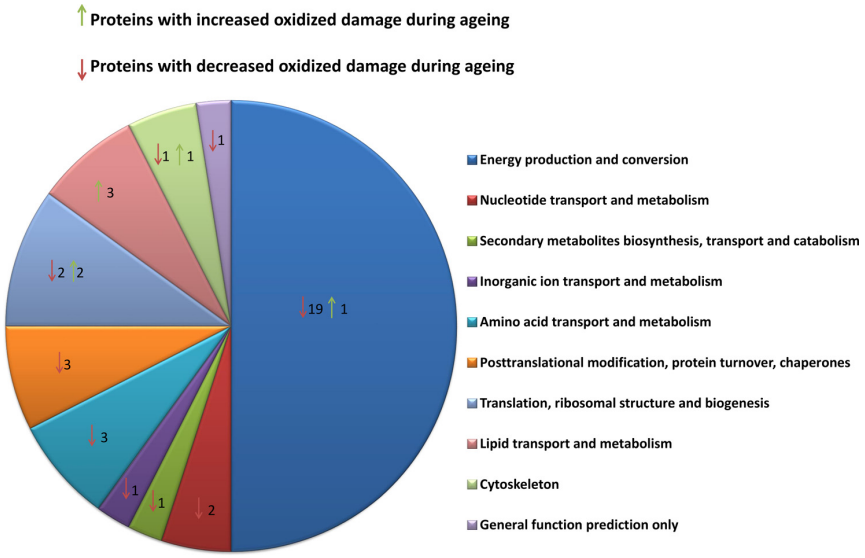


FIG. 4. **Functional classification of proteins according to the COG database that reveal age-dependent differences between altered protein abundance and their respective oxidative damage.** Only proteins with a p value < 0.05 were considered. Arrows indicate the overall increase or decrease of changes in protein oxidation degree from young to old (16 days) mitochondria.

to the complex V of respiratory chain and is the only protein implicated in energy metabolism that exhibited a trend of increasing oxidation. In contrast, an overall decrease of protein oxidation is confirmed for the mitochondrial OXPHOS system, because protein subunits of complex I, II, IV and V showed a decline of about 20% at different time points. According to our results complex I and V could be oxidative hotspots, because several subunits of one complex are damaged by oxidation. However, the abundance change in oxidation degree decreases for almost all subunits of each complex with age or correlates with their respective protein amount. Also, the plasma membrane ATPase type III (Pa_3_6820 [*]) showed a significant decline in protein oxidation. Moreover, three proteins (Pa_1_4920, Pa_6_5750 and Pa_6_2570 [*]) that are involved in cellular stress response, such as chaperones, revealed a slight decrease in oxidized protein species. RNA helicase mss116 (Pa_6_3510 [*]) contributes to the translation mechanism and exhibited a remarkable increase of about 2 (log2[1])-fold in oxidized protein species from day 6 to day 9. Moreover, two proteins (Pa_2_1640 and Pa_4_9350) related to lipid metabolism showed a marked increase in oxidation degree during aging. We also detected several cytoplasmic proteins - probably because of isolation of crude mitochondria- which evidently changed their oxidation degree during aging. Representative proteins with a possible link to aging are the putative woronin body major protein HEX1 (Pa_1_17000 [*]) and the putative eukaryotic translation initiation factor 5A-2 (Pa_3_4460) with an overall decreasing oxidation degree and further a putative GTP-binding protein (Pa_1_1540) with a remarkable increase in oxidation of aspartic acid residues during aging.

Confirmation of Age-related Changes in Abundance and Oxidation for Selected Proteins by SRM Analysis—To validate the most notable age-related changes in protein oxidation and total protein abundance, a targeted proteomic approach for relative quantification was used. Consequently, age-related protein abundance changes were verified by the use of selected AQUA peptides, whereby for verification of age-related changes in protein oxidation the iTRAQ-labeled samples were re-measured in SRM modus on a TSQ mass spectrometer. SRM data analysis for nonoxidized and oxidized species of selected proteins are presented in supplemental Table S8 and S9, respectively. To facilitate data interpretation the temporal trends of changes in protein oxidation or protein abundance from analyzed peptides obtained either on an Orbitrap Velos or a TSQ instrument are shown in direct comparison (Fig. 5). Fig. 5 illustrates that the application of targeted quantification by SRM MS provided a good means of confirmation since for all analyzed proteins the trend profiles for changes in abundance were highly similar between iTRAQ-based and AQUA-based relative quantification, even though the absolute log2 ratios are not always equal during aging. For changes in protein oxidation, five proteins exposed similar results for iTRAQ ratios obtained on Orbitrap Velos *versus* TSQ, whereas just for four protein results did not agree: the putative laccase (Pa_6_7880), a copper-containing oxidase enzyme, the cytochrome *c* oxidase (Pa_6_5480), the putative SAM-dependent o-methyltransferase PaMTH1 (Pa_2_7880) and the putative ATP-dependent RNA helicase mss116 (Pa_6_3510). Thus these proteins weren't any more credible for a statement about the processes during biological aging. Nevertheless, it was possible to confirm an increasing oxida-

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

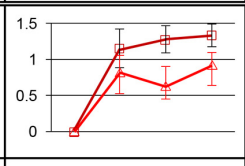
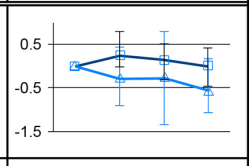
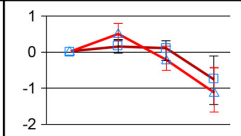
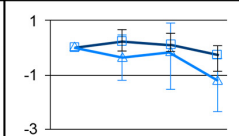
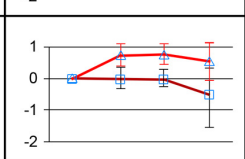
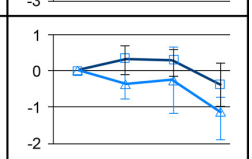
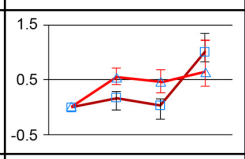
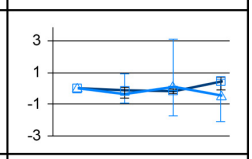
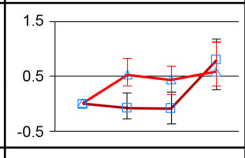
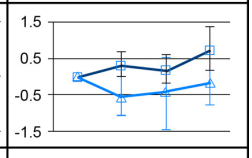
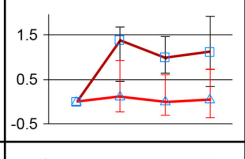
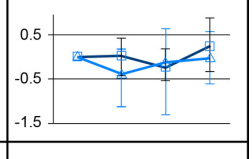
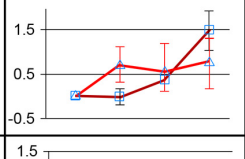
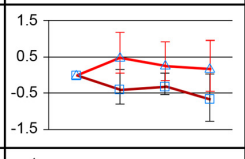
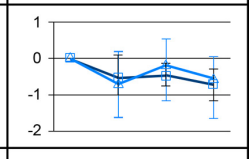
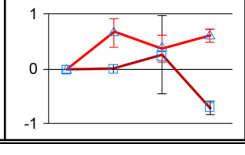
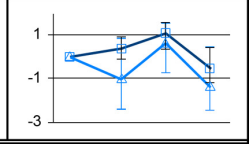
Accession	Description	Legend TSQ: iTRAQ TSQ: AQUA Velos (HCCD): iTRAQ Velos (HCD): iTRAQ Peptide sequence	change oxidative damage log2 ratio 6/6 log2 ratio 9/6 log2 ratio 13/6 log2 ratio 16/6	change protein abundance log2 ratio 6/6 log2 ratio 9/6 log2 ratio 13/6 log2 ratio 16/6
Pa_1_9450	Putative mitochondrial ATP synthase gamma chain precursor	VYQSAETKPL(+10)EGEGKK; NIVLNFSGVGK, YQILFNR		
Pa_mito_nad5	NADH-ubiquinone oxidoreductase chain 5	YITNL1DLGGQM(+10)TK GSIELFGPFGLEK, GLVNFSK		
Pa_6_5480	Cytochrome c oxidase polypeptide 5 encoded by the cox5 gene	GNWAELT(+O)IAEK, EWQEAT(+O)NEYLK; EWQEATNEYLK, AQNSDPLTGISSEGYK		
Pa_3_9400	Putative mitochondrial GrpE protein homolog precursor	AL(+O)TMVPEEK; DLVDSVDNFDR, EHNTVFHTQOK		
Pa_6_2570	Putative mitochondrial pre-cursor of heat shock protein SSC1	DIOEVIL(+O)VGGMTR; TTPSVVAFAEDGER, DAGQIAGLNVLR, ADSVLTDTEK		
Pa_6_3510	Putative ATP-dependent RNA helicase mss116	YIDP(-2H-1C-1O)TNFEAQTVK; TLAFLVPILQK, AIVISPTR		
Pa_5_11950	Putative methyltransferase	L(+O)VVAL(+O)NGL(+O)DHL(+O)K		
Pa_6_7880	Putative laccase precursor	GW(+O)QDPLSEPSLYDTMCADEPYDK; W(+O)MMHEAPFR, EAIEESCSW(+O)R; NWYDDLEIGHEGEGR, WFTPDK		
Pa_2_7880	Putative SAM-dependent O-methyltranferase encoded by the PaMTH1 gene	LQGSW(+O)MIFTAR; AEIVTLEYSPK, TLEGEFDLIEVDANK, DGYAGYVK		

FIG. 5. **Confirmation of iTRAQ data results by SRM quantification of both oxidized and nonoxidized species for selected proteins.** Additional SRM experiments were performed on a TSQ mass spectrometer and directly compared against the iTRAQ data obtained on an

tion level for the chaperone SSC1 with high age (day 16) although log2 ratios differed between the Orbitrap Velos and the TSQ. Concerning this matter the same applies for the putative methyltransferase (Pa_5_11950).

Of particular interest is the confirmation of the changes in protein abundance and oxidation for the ATP synthase gamma chain (Pa_1_9450), further substantiating the age-related protein damage. Furthermore the NADH-ubiquinone oxidoreductase (Pa_mito_nad5, complex I) and the chaperone protein GrpE were confirmed in their age-related changes in protein abundance and oxidation.

Modification Site Occupancy Indicates Low Protein Oxidation Level—Fractional modification site occupancies were estimated for the mono-oxidation of methionine and leucine, and carbonylation of proline which leads to the product pyroglutamic acid (supplemental Table S10). For most sites, occupancies only varied slightly between the different age stages. The distribution of leucine and methionine occupancies peaked around 0.2 (supplemental Fig. S5), whereas for the few estimated proline carbonylations, the occupancy was mostly less than 0.1. Hence, for the majority of sites, the nonoxidized form was in excess. Nevertheless, for some peptides and proteins only the oxidized form could be identified, which may be because of high occupancy, yet its missing unmodified counterpart precluded occupancy calculation. In this respect, it was unfortunately not possible to calculate the occupancy for the leucine oxidation site on the gamma subunit, because the unmodified peptide was not identified. However, data shows that some sites of the gamma and other ATP synthase subunits exhibit relatively high occupancy; in more detail occupancy for another site on the gamma subunit was 0.58, for two sites on the alpha subunit it was 0.48 and 0.63, respectively, and for two sites on the beta subunit it was 0.48 and 0.38. Altogether, these data suggest that some regions of the holoenzyme are hotspots for ROS. For the two other proteins with changes unequivocally confirmed by SRM, *i.e.* the NADH-ubiquinone oxidoreductase (Pa_mito_nad5, complex I) and the chaperone protein GrpE, mean site occupancies were 0.41 and 0.14, respectively.

DISCUSSION

A prominent theory of aging research explains aging as the result of ROS-induced accumulation of random molecular damage to biomolecules, where the majority of ROS are generated in mitochondria. Therefore oxidative modifications of mitochondrial proteins should contribute to their functional decay with age (12, 58). Accordingly, protein damage has been assumed to be a causative factor or at least a key contributor to degeneration and death, also in humans (59). In *P. anserina* different studies demonstrate that oxidative stress increases during aging and that interfering with the generation or scavenging of ROS has a strong impact on aging and lifespan control (22, 23, 25).

Despite this and other evidence, the detailed role of oxidative stress in aging is only insufficiently explained at the molecular level, such as the proteome, because of difficulties in tracking *in vivo* oxidants and oxidized biomolecules (60). As discussed in the following sections, these technical challenges were overcome by a newly developed workflow consisting of large-scale, untargeted protein modification analysis and statistical evaluation of results to scrutinize the role of ROS-induced protein damage for aging.

Assessment of the Untargeted iTRAQ-based Proteomic Approach to Monitor ROS-induced Protein Oxidation in P. anserina—In the present work, a large dataset for age-dependent protein abundance changes of oxidatively modified and unmodified protein species was obtained. A clear advantage of the used untargeted proteomic approach is no *a priori* limitation to a particular protein modification. However, previous untargeted proteome analyses have concentrated almost exclusively on the identification of post-translational modifications, rather than on their quantification (61, 62). Based on these previous qualitative studies it is well-known that an extremely diverse mixture of protein oxidation products occurs during aging, which we aimed to analyze quantitatively, too.

The applied workflow enabled identification and quantification of in total 2341 proteins, whereof both species (unmodified and oxidatively modified) were detected in 746 proteins, and no oxidized counterpart peptides could be quantified for 1273 unmodified proteins. The underlying reasons for the latter are probably both biological and technical. For instance, a previous determination of the oxidation degree in ATP synthase from *P. anserina* (42) suggests a low amount of oxidized species for many proteins. Moreover, the oxidized peptide is more hydrophilic than its corresponding nonmodified form, leading to differences in elution and consequently different interferences from co-eluting background ions during RP chromatography. Despite these challenges, our study stands well against previous quantitative analyses, even when pre-enrichment of modified peptides was done. Compared with (37), who identified and quantified over 200 carbonylated proteins of crudely enriched rat skeletal muscle mitochondria by virtue of pre-enrichment, we were able to quantify almost 100 carbonylated proteins without using any enrichment pro-

Orbitrap Velos. Consequently, age-related protein abundance changes were checked by selected AQUA peptides whereby for confirmation of age-related changes in protein oxidation the iTRAQ-labeled samples were remeasured in a SRM modus on a TSQ mass spectrometer. The temporal trends of changes in protein oxidation or protein abundance from analyzed peptides obtained either on an Orbitrap Velos or a TSQ instrument are shown in direct comparison. Error bars display the standard deviation over six biological replicates per age-stage. The SRM approach confirmed trend profiles of protein abundance change, and trend profiles for protein oxidation were similar for more than half of all proteins.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

cedure for low-abundant oxidized peptides. Admittedly, targeted quantitative enrichment of thiol groups with OxICAT achieved better coverage of cysteine oxidation status, for example 400 different protein thiols in *Saccharomyces cerevisiae* (63) and thiol redox status for 140 proteins in *Caenorhabditis elegans* lifespan (64, 65) were obtained. Even though powerful technologies to target large-scale methionine oxidation sites exist (66), they have not been applied to mitochondrial aging. Unsurprisingly, our untargeted modification approach does not achieve the same qualitative and quantitative proteome coverage for a single oxidative modification as recent targeted studies and therefore we might have missed information pertaining to oxidized peptides which are critical for the proper function of a protein (*e.g.* active sites); but instead it has enabled tracing a much wider spectrum of modifications, amounting to more than 2000 identified protein species which allows the investigation of a global ROS-induced protein damage to verify the ROS theory of aging. The suitability of our iTRAQ-based shotgun proteomics workflow was validated by confirming selected peptides values with SRM analysis. The sometimes observed result differences between the two methods could probably be best explained by missing data points in the shotgun approach.

Can Previous Studies on Oxidative Damage and Protein Abundance Changes in Mitochondria During Aging be Extended?—Obviously the central biological question we wanted to address was whether our results are in line and extend former studies on the role of ROS in aging. Although many studies focused on the relationship between ROS-induced protein damage and aging, an age-related increase of bulk oxidized mitochondrial proteins could not be observed. Instead only individual proteins seem to be modified, such as proteins of the OXPHOS or aconitase (17, 24, 37). Indeed, an immunochemical analysis of carbonylated proteins failed to detect significant age-dependent difference in bulk carbonylation of mitochondrial proteins; nevertheless, a remarkable increase of proteins containing carboxymethylated lysine residues during aging in *P. anserina* could be shown (24). In the present study we were able to quantify 18 different protein oxidations concurrently over the age stages in *P. anserina*. Unfortunately, the used software (Proteome Discoverer 1.3) did not allow searches for oxidative modifications on lysine residues since already modified by the iTRAQ reagent. Overall, the iTRAQ proteomic analysis revealed remarkably moderate changes in protein oxidation and total protein abundance. It is known that this quantitative approach suffers to some extent from compression of iTRAQ protein ratios in complex samples but the direction of change (differential up-, down-, or unregulated) is itself unaffected (67). In agreement with our work other comparative approaches on *P. anserina*, such as transcriptome analysis, 2D-DIGE analysis or label free protein quantification, revealed the same moderate effects on the mitochondrial proteome during aging (68–70), suggesting a compensatory response to ROS-induced protein damage during aging of *P. anserina*. Further, a recent SRM analysis on the mitochondrial proteome of *P. anserina PaSod3* over-expression mutant revealed that the protein changes in the mutant are much more remarkable than in the aging wild type (70). Our work extends previous studies which showed only minor changes in the protein abundance involved in oxidative phosphorylation, energy metabolism, stress response and protein quality control (68–70). Even though the made assumptions for occupancy calculation demand cautious result interpretation, it seems fair to say that overall a low site oxidation level could be maintained at all investigated age stages of *P. anserina*. This result again indicates a proteome homeostasis mechanism. It might further indicate that sites with a low or "decorative" level of oxidation do not impair protein function. On the contrary, the relatively high occupancy of several modification sites in the ATP synthase together with the observed significant increase of oxidation for its gamma subunit, hints toward functional impairment at late age. Furthermore, it can be concluded from the wealth of obtained data that protein abundance and oxidation during aging do not follow a simple trend, but instead show different and varying kinetics, although the overall changes are quite subtle, and the composition of the mitochondrial proteome of *P. anserina* can be considered as generally stable even when reaching late age.

Besides the important global assessment of mitochondrial protein damage, novel findings for individual proteins related to aging were obtained: (17) summarized examples of age-related proteins (ATP-synthase F1 complex, reticulocalbin) with redundant isoforms in mitochondria from three different species, *P. anserina*, rat and human. The study revealed that in the ATP-synthase F1 complex the underlying posttranslational modifications are associated with distinct domains of ATP synthase subunits; whereas other parts of the protein remain unaffected. Furthermore, (37) approve the susceptibility of the OXPHOS machinery to carbonylation. In agreement with these observations our iTRAQ-quantification showed that the OXPHOS complexes are certainly targets of ROS, including several oxidized protein forms. In addition, our data revealed a remarkable increase of leucine oxidation within the ATP synthase gamma chain with age that could be also confirmed by the SRM validation (Fig. 5) and may therefore be a hotspot to ROS-induced damage during aging. However, regarding the complete proteome level, we end up with a different biological interpretation, namely protein damage homeostasis, because correlations between the amount of modified and unmodified peptide species were found. Indeed, no pronounced increase of global protein damage at late age could be observed; instead our data rather indicate a general reduction in protein oxidation and abundance mostly from day 13 to day 16. This does not mean that an absolute increase of oxidized protein species could not negatively affect cellular and organismal physiology even though the fractional amount does not change, but we hypothesize that as long as the

organismal scavenger systems and the protein *de novo* synthesis are properly functioning, the occurring protein damage can be largely, if not completely, compensated, which will be discussed in following sections.

No Massive Increase in Protein Oxidation Degree During Aging in P. anserina—

How Could ROS-induced Protein Damage be Compensated?—Extensive correlative evidence has been collected over decades that corroborate the oxidative stress theory of aging including the discovery that aging and many age-related diseases are accompanied by considerable cellular oxidative damage (28, 39, 71–73) (74, 75). Despite these and many other reports supporting an increased ROS level, recently a series of conflicting results suggested a more complex interplay between endogenous oxidants, antioxidants, protein quality control and lifespan (76, 77). For instance, the comprehensive proteomic analysis of different mouse tissues revealed only minor age-related abundance changes and rather suggested a functional protein homeostasis up to a relatively late age (78). Notably, a redox proteomics study performed with the OxICat technique revealed a ROS burst as an early event in *C. elegans* life, but not during aging (65). Also, the previous SRM analysis of mitochondria from *P. anserina* showed that the examined oxidative modification in the ATP synthase subunit alpha remained rather unaffected by aging (42). These recent findings suggest that ROS level is well-controlled during aging and functionally regulates essential physiological processes in the cell (79, 80). For this purpose, pathways for special ROS signaling seem to exist which regulate also ROS homeostasis for alleviating the toxicity of ROS (2). For instance, our data reveal that methionine residues are the main oxidation site within the detected proteins, which could be removed, if repair mechanisms are sufficiently early activated. It is known that methionine residues constitute an important antioxidant defense mechanism because surface exposed methionine residues create an extremely high concentration of reactant, available as an efficient oxidant scavenger. Reduction back to methionine by methionine sulfoxide reductases allow the antioxidant system to be restored (81). Here, two different classes of methionine sulfoxide reductases, MsrA and MsrB, play a pivotal role (82); accordingly, *msrA* overexpression in *Drosophila melanogaster* and *S. cerevisiae* extended the lifespan and increased the resistance against oxidative stress, respectively. Methionine oxidation followed by reduction may act as a constant sink for reactive oxygen species, which primarily protects amino acids in close vicinity against ROS (83, 84). We identified unmodified peptide species of both putative methionine sulfoxide reductases (Pa_1_1390 & Pa_4_7040) and found that the protein amount of Pa_1_1390 remained stable during aging, whereas the putative methionine sulfoxide reductase Pa_4_7040 decreased slightly - interestingly in earlier age stages (day 9 and 13). This is supported by a recent transcriptome analysis revealing a decrease of the corresponding gene (Pa_4_7040) from day 6 to 14 with a factor of 0.67 (69) and indicating insufficient capacity of the Msr proteins to compensate oxidative damage alone. Indeed, concomitant increase of oxidized and nonoxidized forms of chaperons (GrpE, Hsp70 (SCC1) and Hsp60) suggests that *P. anserina* elicits a stress response for additional compensation of accumulated protein damage during aging.

Proteome Homeostasis Collapse Occurs at a Very Late Age—The here observed decrease of many mitochondrial proteins can be best explained by an age-dependent decrease in protein synthesis and an increase in protein removal by various proteolytic processes: Apparently the protein synthesis breaks down after day 13 and the organism cannot compensate the removal of ROS-damaged proteins anymore. Several studies revealed age-related increases in mtDNA mutations that contribute to physiological decline in aging and degenerative diseases (85–87) and are mirrored in senescent *P. anserina* (88, 89). These mtDNA mutations lead to deficiencies in remodelling of affected oxidized proteins that are encoded by the mtDNA, such as proteins of the respiratory chain that require replacement by *de novo* synthesis (21). This conclusion agrees with the observed general down-regulation of important nuclear as well as mtDNA-encoded mitochondrial proteins from day 13 on. Concerning the postulated increase of proteolytic processes with age, the recent genome-wide transcriptome analysis of *P. anserina* suggests that autophagy is a compensatory mechanism during aging when all other pathways failed to restore the proceeding accumulation of damaged biomolecules (69). A recent analysis showing that autophagy is up-regulated in old *P. anserina*, as monitored by the autophagy-dependent degradation of a PaSOD1::GFP reporter protein (90) underpins this assumption: autophagy evidently becomes induced during normal organismal aging when the proteasome and transcription/translation appear to be functionally impaired (69). Importantly, loss of autophagy significantly shortens lifespan (90). However, the exact role of autophagy in the degradation of oxidatively damaged mitochondrial proteins of *P. anserina* remains to be determined.

CONCLUSION

In summary, this study shows the potential of untargeted iTRAQ-shotgun proteomics and adequate statistics to monitor simultaneously many different oxidative modifications. Our approach allowed to conclude that: (1) no global increase in protein oxidation occurs in the mitochondria of *P. anserina* during aging, (2) not only protein amounts remain rather constant, but also the ratio of damaged to undamaged protein, and (3) ROS-induced protein damage is efficiently compensated by protein *de novo* synthesis until late age, when in *P. anserina* autophagy is activated. Hence, efficient proteome homeostasis is the key to maintain mitochondrial function during increasing ROS exposure with age.

Acknowledgments—

* This study was funded by the German Federal Ministry for Education and Research (BMBF) through the *GerontoMitoSys* project (FKZ0315584) to HDO and MR. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (46) via the PRIDE partner repository with the dataset identifier PXD001023.

S This article contains supplemental material.

‖ To whom correspondence should be addressed: Ruhr-University Bochum, Universitätsstr. 150, D-44780 Bochum, Germany. Tel.: +49 (0) 234 3228419; Fax: +49 (0) 234 3214322; E-mail: Ansgar.Poetsch@rub.de.

Email addresses: Carina Ramallo Guevara carina.ramalloguevara@rub.de; Oliver Philipp o.philipp@bioinformatik.uni-frankfurt.de; Andrea Hamann a.hamann@bio.uni-frankfurt.de; Alexandra Werner alexandra.werner@bio.uni-frankfurt.de; Heinz D. Osiewacz osiewacz@bio.uni-frankfurt.de; Sascha Rexroth sascha.rexroth@rub.de; Matthias Rögner matthias.roegner@rub.de; Ansgar Poetsch* ansgar.poetsch@rub.de.

CONFLICT OF INTEREST: The authors declare that there are no conflicts of interest.

REFERENCES

1. Halliwell, B., and Gutteridge, J. (1990) The antioxidants of human extracellular fluids. *Arch. Biochem. Biophys.* **280,** 1–8
2. D'Autreaux, B., and Toledano, M. B. (2007) ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat. Rev. Mol. Cell Biol.* **8,** 813–824
3. Lanciano, P., Khalfaoui-Hassani, B., Selamoglu, N., Ghelli, A., Rugolo, M., and Daldal, F. (2013) Molecular mechanisms of superoxide production by complex III: a bacterial versus human mitochondrial comparative case study. *Biochim. Biophys. Acta* **1827,** 1332–1339
4. Dröse, S., and Brandt, U. (2012) Molecular mechanisms of superoxide production by the mitochondrial respiratory chain. *Adv. Exp. Med. Biol.* **748,** 145–169
5. Calabrese, V., Scapagnini, G., Giuffrida Stella, A. M., Bates, T. E., and Clark, J. B. (2001) Mitochondrial involvement in brain function and dysfunction: relevance to aging, neurodegenerative disorders and longevity. *Neurochem. Res.* **26,** 739–764
6. Lenaz, G., Bovina, C., D'Aurelio, M., Fato, R., Formiggini, G., Genova, M. L., Giuliano, G., Pich, M. M., Paolucci, U. G. O., and Castelli, G. P. (2002) Role of mitochondria in oxidative stress and aging. *Ann. N.Y. Acad. Sci.* **959,** 199–213
7. Cui, Z. J., Han, Z. Q., and Li, Z. Y. (2011) Modulating protein activity and cellular function by methionine residue oxidation. *Amino Acids* **43,** 505–517
8. Wang, C. H., Wu, S. B., Wu, Y. T., and Wei, Y. H. (2013) Oxidative stress response elicited by mitochondrial dysfunction: implication in the pathophysiology of aging. *Exp. Biol. Med. (Maywood)* **238,** 450–460
9. Jenner, P. (2003) Oxidative stress in Parkinson's disease. *Ann. Neurol.* **53 Suppl 3,** S26–36; discussion S36–38
10. Lin, M. T., and Beal, M. F. (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443,** 787–795
11. Harman, D. (1956) Aging: a theory based on free radical and radiation chemistry. *J. Gerontol.* **11,** 298–300
12. Harman, D. (1972) The biologic clock: the mitochondria? *J. Am. Geriatr. Soc.* **20,** 145–147
13. Levine, R. L. (2002) Carbonyl modified proteins in cellular regulation, aging, and disease. *Free Radic. Biol. Med.* **32,** 790–796
14. Cui, H., Kong, Y., and Zhang, H. (2012) Oxidative stress, mitochondrial dysfunction, and aging. *J. Signal Transduct.* **2012**, 646354
15. Harman, D. (1992) Free radical theory of aging. *Mutat. Res.* **275,** 257–266
16. Beal, M. F. (1995) Aging, energy, and oxidative stress in neurodegenerative diseases. *Ann. Neurol.* **38,** 357–366
17. Groebe, K., Krause, F., Kunstmann, B., Unterluggauer, H., Reifschneider, N. H., Scheckhuber, C. Q., Sastri, C., Stegmann, W., Wozny, W., Schwall, G. P., Poznanović, S., Dencher, N. A., Jansen-Dürr, P., Osiewacz, H. D., and Schrattenholz, A. (2007) Differential proteomic profiling of mitochondria from *Podospora anserina*, rat and human reveals distinct patterns of age-related oxidative changes. *Exp. Gerontol.* **42,** 887–898
18. Osiewacz, H. D. (2002) Aging in fungi: role of mitochondria in *Podospora anserina*. *Mech. Ageing Dev.* **123,** 755–764
19. Osiewacz, H. D. (2002) Mitochondrial functions and aging. *Gene* **286,** 65–71
20. Lorin, S., Dufour, E., and Sainsard-Chanet, A. (2006) Mitochondrial metabolism and aging in the filamentous fungus *Podospora anserina*. *Biochim. Biophys. Acta* **1757,** 604–610
21. Osiewacz, H. D., Hamann, A., and Zintel, S. (2013) Assessing organismal aging in the filamentous fungus *Podospora anserina*. *Methods Mol. Biol.* **965,** 439–462
22. Scheckhuber, C. Q., Erjavec, N., Tinazli, A., Hamann, A., Nyström, T., and Osiewacz, H. D. (2007) Reducing mitochondrial fission results in increased life span and fitness of two fungal ageing models. *Nat. Cell Biol.* **9,** 99–105
23. Gredilla, R., Grief, J., and Osiewacz, H. D. (2006) Mitochondrial free radical generation and lifespan control in the fungal aging model *Podospora anserina*. *Exp. Gerontol.* **41,** 439–447
24. Luce, K., and Osiewacz, H. D. (2009) Increasing organismal healthspan by enhancing mitochondrial protein quality control. *Nat. Cell Biol.* **11,** 852–858
25. Kunstmann, B., and Osiewacz, H. D. (2008) Over-expression of an S-adenosylmethionine-dependent methyltransferase leads to an extended lifespan of *Podospora anserina* without impairments in vital functions. *Aging Cell* **7,** 651–662
26. Bakala, H., Ladouce, R., Baraibar, M. A., and Friguet, B. (2013) Differential expression and glycative damage affect specific mitochondrial proteins with aging in rat liver. *Biochim. Biophys. Acta* **1832,** 2057–2067
27. Chaudhuri, A. R., de Waal, E. M., Pierce, A., Van Remmen, H., Ward, W. F., and Richardson, A. (2006) Detection of protein carbonyls in aging liver tissue: a fluorescence-based proteomic approach. *Mech. Ageing Dev.* **127,** 849–861
28. Aksenov, M. Y., Aksenova, M. V., Butterfield, D. A., Geddes, J. W., and Markesbery, W. R. (2001) Protein oxidation in the brain in Alzheimer's disease. *Neuroscience* **103,** 373–383
29. Kunstmann, B., and Osiewacz, H. D. (2009) The S-adenosylmethionine dependent O-methyltransferase PaMTH1: a longevity assurance factor protecting *Podospora anserina* against oxidative stress. *Aging* **1,** 328–334
30. Surco-Laos, F., Cabello, J., Gómez-Orte, E., González-Manzano, S., González-Paramás, A. M., Santos-Buelga, C., and Dueñas, M. (2011) Effects of O-methylated metabolites of quercetin on oxidative stress, thermotolerance, lifespan and bioavailability on *Caenorhabditis elegans*. *Food Funct.* **2,** 445–456
31. Sheehan, D., McDonagh, B., and Bárcena, J. A. (2010) Redox proteomics. *Expert Rev. Proteomics* **7,** 1–4
32. Törnvall, U. (2010) Pinpointing oxidative modifications in proteins—recent advances in analytical methods. *Anal. Methods* **2,** 1638–1650
33. Fedorova, M., Bollineni, R. C., and Hoffmann, R. (2014) Protein carbonylation as a major hallmark of oxidative damage: update of analytical strategies: protein carbonylation. *Mass Spectrom. Rev.* **33,** 79–97
34. Froelich, J. M., and Reid, G. E. (2008) The origin and control of ex vivo oxidative peptide modifications prior to mass spectrometry analysis. *Proteomics* **8,** 1334–1345
35. Perdivara, I., Deterding, L. J., Przybylski, M., and Tomer, K. B. (2010) Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification? *J. Am. Soc. Mass Spectrom.* **21,** 1114–1117
36. Møller, I. M., Rogowska-Wrzesinska, A., and Rao, R. S. P. (2011) Protein carbonylation and metal-catalyzed protein oxidation in a cellular perspective. *J. Proteomics* **74,** 2228–2242
37. Meany, D. L., Xie, H., Thompson, L. V., Arriaga, E. A., and Griffin, T. J. (2007) Identification of carbonylated proteins from enriched rat skeletal muscle mitochondria using affinity chromatography-stable isotope labeling and tandem mass spectrometry. *Proteomics* **7,** 1150–1163
38. Mann, M., and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21,** 255–261
39. Baraibar, M. A., Ladouce, R., and Friguet, B. (2013) Proteomic quantification and identification of carbonylated proteins upon oxidative stress and during cellular aging. *J. Proteomics* **92,** 63–70
40. Ross, P. L. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell.*

Proteomics **3,** 1154–1169
41. Evans, C., Noirel, J., Ow, S. Y., Salim, M., Pereira-Medrano, A. G., Couto, N., Pandhal, J., Smith, D., Pham, T. K., Karunakaran, E., Zou, X., Biggs, C. A., and Wright, P. C. (2012) An insight into iTRAQ: where do we stand now? *Anal. Bioanal. Chem.* **404,** 1011–1027
42. Rexroth, S., Poetsch, A., Rögner, M., Hamann, A., Werner, A., Osiewacz, H. D., Schäfer, E. R., Seelert, H., and Dencher, N. A. (2012) Reactive oxygen species target specific tryptophan site in the mitochondrial ATP synthase. *Biochim. Biophys. Acta* **1817,** 381–387
43. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
44. Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B. M., Couloux, A., Aury, J.-M., Ségurens, B., and Poulain, J. (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.* **9,** R77
45. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7,** 40–44
46. Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.-A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H.-J., Albar, J. P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226
47. Marsaglia, G., Tsang, W. W., and Wang, J. (2003) Evaluating Kolmogorov's Distribution. *J. Stat Softw* **8,** 1–4
48. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical Recipes in C: the art of scientific computing, Second Edition*, Cambridge Univ. Press, New York
49. Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. B* **57,** 289–300
50. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968
51. Wolff, D., ter Veld, F., Kohler, T., and Poetsch, A. (2013) Combined application of targeted and untargeted proteomics identifies distinct metabolic alterations in the tetraacetylphytosphingosine (TAPS) producing yeast Wickerhamomyces ciferrii. *J. Proteomics* **82,** 274–287
52. Farrah, T., Deutsch, E. W., Kreisberg, R., Sun, Z., Campbell, D. S., Mendoza, L., Kusebauch, U., Brusniak, M. Y., Huttenhain, R., Schiess, R., Selevsek, N., Aebersold, R., and Moritz, R. L. (2012) PASSEL: the PeptideAtlas SRMexperiment library. *Proteomics* **12,** 1170–1175
53. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5,** 144–156
54. Eidhammer, I., Barsnes, H., Eide, G. E., and Martens, L. (2013) Quantification of Post-Translational Modifications. in *Computational and Statistical Methods for Protein Quantification by Mass Spectrometry*. John Wiley & Sons, pp 244–253
55. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8,** 937–940
56. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**
57. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* **278,** 631–637
58. Muller, F. (2000) The nature and mechanism of superoxide production by the electron transport chain: Its relevance to aging. *AGE* **23,** 227–253
59. Devasagayam, T., Tilak, J. C., Boloor, K. K., Sane, K., Ghaskadbi, S., and Lele, R. (2004) Free radicals and antioxidants in human health: current status and future prospects. *Japi* **52,** 794–804
60. Muller, F. L., Lustgarten, M. S., Jang, Y., Richardson, A., and Van Remmen, H. (2007) Trends in oxidative aging theories. *Free Radic. Biol. Med.* **43,** 477–503
61. Tveen-Jensen, K., Reis, A., Mouls, L., Pitt, A. R., and Spickett, C. M. (2013) Reporter ion-based mass spectrometry approaches for the detection of non-enzymatic protein modifications in biological samples. *J. Proteomics* **92,** 71–79
62. McClintock, C. S., Parks, J. M., Bern, M., GhattyVenkataKrishna, P. K., and Hettich, R. L. (2013) Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional LC–MS/MS data. *J. Proteome Res.* **12,** 3307–3316
63. Brandes, N., Reichmann, D., Tienson, H., Leichert, L. I., and Jakob, U. (2011) Using quantitative redox proteomics to dissect the yeast redoxome. *J. Biol. Chem.* **286,** 41893–41903
64. Kumsta, C., Thamsen, M., and Jakob, U. (2011) Effects of oxidative stress on behavior, physiology, and the redox thiol proteome of *Caenorhabditis elegans*. *Antioxid Redox Signal* **14,** 1023–1037
65. Knoefler, D., Thamsen, M., Koniczek, M., Niemuth, Nicholas J., Diederich, A.-K., and Jakob, U. (2012) Quantitative *in vivo* redox sensors uncover oxidative stress as an early event in life. *Mol. Cell* **47,** 767–776
66. Ghesquiere, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M., Schymkowitz, J., Rousseau, F., Vandekerckhove, J., and Gevaert, K. (2011) Redox proteomics of protein-bound methionine oxidation. *Mol. Cell. Proteomics* **10,** M110 006866
67. Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009) iTRAQ underestimation in simple and complex mixtures: "The good, the bad and the ugly". *J. Proteome Res.* **8,** 5347–5355
68. Chimi, M. A., Dröse, S., Wittig, I., Heide, H., Steger, M., Werner, A., Hamann, A., Osiewacz, H. D., and Brandt, U. (2013) Age-related changes in the mitochondrial proteome of the fungus *Podospora anserina* analyzed by 2D-DIGE and LC-MS/MS. *J. Proteomics* **91,** 358–374
69. Philipp, O., Hamann, A., Servos, J., Werner, A., Koch, I., and Osiewacz, H. D. (2013) A Genome-Wide Longitudinal Transcriptome Analysis of the Aging Model *Podospora anserina*. *PLoS ONE* **8,** e83109
70. Plohnke, N., Hamann, A., Poetsch, A., Osiewacz, H. D., Rögner, M., and Rexroth, S. (2014) Proteomic analysis of mitochondria from senescent *Podospora anserina* casts new light on ROS dependent aging mechanisms. *Exp. Gerontol.* **56,** 13–25
71. Finkel, T., and Holbrook, N. J. (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* **408,** 239–247
72. Beal, M. F. (2002) Oxidatively modified proteins in aging and disease. *Free Radic. Biol. Med.* **32,** 797–803
73. Lionaki, E., and Tavernarakis, N. (2013) Oxidative stress and mitochondrial protein quality control in aging. *J. Proteomics* **92,** 181–194
74. Bokov, A., Chaudhuri, A., and Richardson, A. (2004) The role of oxidative damage and stress in aging. *Mech. Ageing Dev.* **125,** 811–826
75. Dencher, N. A., Frenzel, M., Reifschneider, N. H., Sugawa, M., and Krause, F. (2007) Proteome alterations in rat mitochondria caused by aging. *Ann. N.Y. Acad. Sci.* **1100,** 291–298
76. Lapointe, J., and Hekimi, S. (2010) When a theory of aging ages badly. *Cell. Mol. Life Sci.* **67,** 1–8
77. Doonan, R., McElwee, J. J., Matthijssens, F., Walker, G. A., Houthoofd, K., Back, P., Matscheski, A., Vanfleteren, J. R., and Gems, D. (2008) Against the oxidative damage theory of aging: superoxide dismutases protect against oxidative stress but have little or no effect on life span in *Caenorhabditis elegans*. *Genes Dev.* **22,** 3236–3241
78. Walther, D. M., and Mann, M. (2011) Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging. *Mol. Cell. Proteomics* **10,** M110 004523
79. Rhee, S. G. (2006) Cell signaling. H_2O_2, a necessary evil for cell signaling. *Science* **312,** 1882–1883
80. Held, J. M., and Gibson, B. W. (2012) Regulatory control or oxidative damage? Proteomic approaches to interrogate the role of cysteine oxidation status in biological processes. *Mol. Cell. Proteomics* **11,** R111 013037
81. Levine, R. L., Berlett, B. S., Moskovitz, J., Mosoni, L., and Stadtman, E. R. (1999) Methionine residues may protect proteins from critical oxidative damage. *Mech. Ageing Dev.* **107,** 323–332
82. Weissbach, H., Resnick, L., and Brot, N. (2005) Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage. *Biochim. Biophys. Acta* **1703,** 203–212
83. Levine, R. L., Moskovitz, J., and Stadtman, E. R. (2000) Oxidation of methionine in proteins: roles in antioxidant defense and cellular regulation. *IUBMB Life* **50,** 301–307

84. Friguet, B., Bulteau, A.-L., and Petropoulos, I. (2008) Mitochondrial protein quality control: implications in ageing. *Biotechnol J.* **3,** 757–764
85. Chomyn, A., and Attardi, G. (2003) MtDNA mutations in aging and apoptosis. *Biochem. Biophys. Res. Commun.* **304,** 519–529
86. Kujoth, G. C. (2005) Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309,** 481–484
87. Wallace, D. C. (2010) Mitochondrial DNA mutations in disease and aging. *Environ. Mol. Mutagen.* **51,** 440–450
88. Osiewacz, H. D., and Hermanns, J. (1992) The role of mitochondrial DNA rearrangements in aging and human diseases. *Aging* **4,** 273–286
89. Osiewacz, H. D. (1997) Genetic regulation of aging. *J. Mol. Med.* **75**, 715–727
90. Knuppertz, L., Hamann, A., Pampaloni, F., Stelzer, E., and Osiewacz, H. D. (2014) Identification of autophagy as a longevity-assurance mechanism in the aging model *Podospora anserina. Autophagy* **10,** 822–834
91. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4,** 2010–2021

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

4.3 Path2PPI: an R package to predict protein-protein interaction networks for a set of proteins

Publication status: Published

Summary

Various databases and repositories exist which comprise a vast amount of different protein interactions. Some of these databases consist of experimental data (Razick, 2008; Prasad et al., 2009; Kanehisa et al., 2014), i.e., protein interactions revealed by different type of biological experiments, for example affinity chromatography, yeast two-hybrid screen or Western Blot. Other databases mostly provide predicted interactions (Franceschini et al., 2013; Wiles et al., 2010). The predictions can be based on different approaches like homology comparison, supervised learning or rely on particular biological data, e.g. functional annotation, co-expression and / or text-mining data (Deng, 2013; Rao, 2014). Nevertheless, most of these databases have in common that they only provide comprehensive data for some already well established model organisms or rely on sets of *a priori* available biological data. To date, no satisfactory solution existed to gain knowledge about proteins and their interaction of particular pathways for organisms like *P. anserina* for which only less PPI data is available.

The R Package PATH2PPI was developed to close this gap. It was implemented and deposited at the Bioconductor platform to provide a software tool which enables the prediction of proteins and their interactions for any fully sequenced organism. Typically, these proteins belong to a particular pathway of interest and are based on information available about that pathway from other better studied reference organisms. The package relies on the iRefIndex database (Razick, 2008) which provides interaction data from various biological experiments about seven of the most established and studied model organisms, human, mouse, rat, yeast, *E.coli*, *C.elegans* and *D.melanogaster* which serve as reference organisms. PATH2PPI automatically transfer interaction data by means of the evolutionary similarity of the target species with the reference species or the corresponding proteins, respectively. Summarising, the major features of PATH2PPI are:

- definition of the reference species (up to seven)
- works with each fully sequenced target organism
- automatic search for all relevant reference interactions based on protein sets (support of different protein identifiers)

- scoring of predicted interactions based on degree of homology and the number of reference species
- retrieve information about each underlying reference interaction
- various graphical representations of the predicted PPI network.

Although PATH2PPI features an intuitive usability it comprises a detailed user manual and a comprehensive tutorial (see appendix) which helps to construct the putative PPI network of interest. Since PATH2PPI provides different graphical representations and output formats it enables the performing of subsequent network topological as well as experimental analyses on the predicted network.

Authors' contribution

Authors: Oliver Philipp (OP), Heinz D. Osiewacz (HDO) and Ina Koch (IK)

Percentage of contribution for OP:

Conception/Design:	80 %
Experimental work:	Not applicable
Implementation:	100 %
Analysis/Evaluation:	90 %
Manuscript:	70 %

OP conceived, designed and developed the method. OP implemented the package, wrote the manual and the tutorial of the package and developed the sample data and network. OP, IK and HDO wrote the manuscript. IK and HDO supervised the study.

Bioinformatics Advance Access published February 4, 2016

Bioinformatics, 2016, 1–3
doi: 10.1093/bioinformatics/btv765
Advance Access Publication Date: 5 January 2016
Applications Note

Systems biology

Path2PPI: an R package to predict protein–protein interaction networks for a set of proteins

Oliver Philipp[1,2], Heinz D. Osiewacz[2] and Ina Koch[1,*]

[1]Molecular Bioinformatics, Institute of Computer Science, Faculty of Computer Science and Mathematics & Cluster of Excellence 'Macromolecular Complexes', Goethe-University, Frankfurt am Main, Germany and [2]Molecular Developmental Biology, Institute of Molecular Biosciences, Faculty for Biosciences & Cluster of Excellence 'Macromolecular Complexes', Goethe-University, Frankfurt am Main, Germany

*To whom correspondence should be addressed.
Associate Editor: Ziv Bar-Joseph

Received on 19 August 2015; revised on 18 November 2015; accepted on 25 December 2015

Abstract

Summary: We introduce PATH2PPI, a new R package to identify protein–protein interaction (PPI) networks for fully sequenced organisms for which nearly none PPI are known. PATH2PPI predicts PPI networks based on sets of proteins from well-established model organisms, providing an intuitive visualization and usability. It can be used to combine and transfer information of a certain pathway or biological process from several reference organisms to one target organism.
Availability and implementation: PATH2PPI is an open-source tool implemented in R. It can be obtained from the Bioconductor project: http://bioconductor.org/packages/Path2PPI/
Contact: ina.koch@bioinformatik.uni-frankfurt.de
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Plenty of databases exist which contain protein–protein interaction (PPI) data for various organisms (e.g. Chatr-aryamontri *et al.*, 2015; Franceschini *et al.*, 2013). For some well-established model organisms, species-specific data repositories are available (e.g. Güldener *et al.*, 2006; Prasad *et al.*, 2009) providing also PPI data. In contrast, for the majority of organisms such a comprehensive amount of PPI data is not available. Therefore, different approaches have been developed to predict PPIs. Some of these approaches aim to deduce new interactions from known PPIs by means of homology-based mapping based on sequence similarity. Other methods apply supervised learning to filter and score the predicted interactions, using additional biological data, e.g. functional annotation, co-expression and / or text-mining data (Yu *et al.*, 2010; for a recent review see Rao *et al.*, 2014). Approaches that predict PPIs based on sequence data or network topology information often provide only precomputed data (Franceschini *et al.*, 2013; Pesch and Zimmer, 2013) or predict interactions only for a set of predefined organisms (Deng *et al.*, 2013; Wiles *et al.*, 2010). Furthermore, most of the methods do not supply information about the underlying reference interactions. In the majority, only the scores are provided to validate a predicted interaction. Often it is necessary to easily access the entire underlying information about the predicted interactions since the interpretation and experimental validation is one of the most important steps after the prediction.

As we were interested in aging processes and interaction networks of age-related pathways in the fungal model organism *Podospora anserina* (Osiewacz *et al.*, 2013; Philipp *et al.*, 2013), we found only a few data repositories for interaction data. For example, the *KEGG* database (Kanehisa *et al.*, 2014) provides small subnetworks of some selected, mainly metabolic, pathways. Recently, also the *STRING* database (e.g. Franceschini *et al.*, 2013) involves some predicted interactions for *P.anserina*. Nevertheless, there was no satisfactory solution and no easy and fast way to directly gain knowledge about proteins and their interactions of certain biological processes, which are well established in some model organisms, but nearly unknown in the target organism. Homology-based tools which theoretically enable to predict or transfer interactions between species are mostly implemented for a set of predefined organisms or

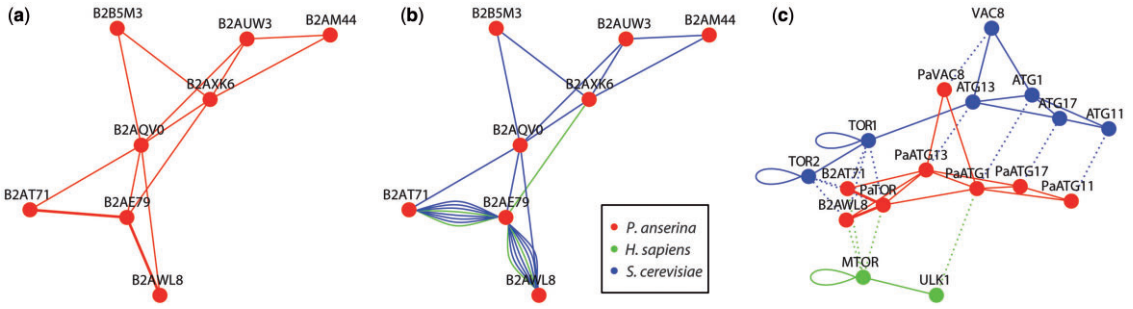


Fig. 1. The predicted PPI network of autophagy induction in *P.anserina* based on the corresponding PPIs in human and yeast. (**a**) The predicted PPI network (*normal* view). The edge thickness corresponds to the scores and the number of reference species showing the interaction. (**b**) *Detailed* view of the PPI network. Each edge is specifically colored (see legend), indicating in which reference species this interaction occurs. Multiple edges represent multiple findings of an interaction. That means, that e.g. the interaction of the proteins 'B2AT71' and 'B2AE79' was found six times in yeast (blue edges) and two times in human (green edges). (**c**) *Hybrid* network representation of the predicted PPI: The relevant parts of the PPI networks of the reference species are included together with the predicted PPI network of the target species. Interactions are depicted as solid lines in the respective color (see legend). Homologous relations of the reference proteins to those of the target species are drawn as dotted edges in the respective color of the target node

require pairs of proteins in the target organism to decide whether they may interact (Chen *et al.*, 2009; Murakami and Mizuguchi, 2014). Here, we report the implementation of PATH2PPI which helps finding proteins and interactions of certain pathways or biological processes in each fully sequenced organism without the need for predefinition of putative proteins and interactions.

2 Features

Using PATH2PPI, the user can choose up to seven of the most established model organisms (human, mouse, rat, yeast, *E.coli*, *C.elegans* and *D.melanogaster*). Based on sets of proteins from these reference species PATH2PPI uses the interaction repository *iRefIndex* (Razick *et al.*, 2008) to find the corresponding relevant interactions. We implemented a more flexible and comfortable search engine than provided by the *iRefR* package (Mora and Donaldson, 2011). Additionally, PATH2PPI requires results of *NCBI BLAST+* (Camacho *et al.*, 2009) searches of all reference species against the target species. Based on these data, PATH2PPI computes new interactions in the target species and scores them. The score is based on the degree of homology and the number of reference species which show the corresponding interaction. A major advantage of PATH2PPI is the easy access to the underlying reference interactions, i.e. all information provided by *iRefIndex*, e.g. source database, interaction type and reference publication. Based on the *igraph* package (Csardi and Nepusz, 2006) the computed PPI can directly be visualized in R (see Fig. 1).

3 Implementation

PATH2PPI can be obtained from the Bioconductor project (Huber *et al.*, 2015). It contains a comprehensive tutorial and for the case study, data files necessary to predict interactions of the induction step of autophagy in *P.anserina* by means of the corresponding PPIs in human and yeast. There are three types of visualization methods available, the *normal*, *detailed* and *hybrid* (Fig 1a–c). Additionally, detailed information about each interaction can be obtained. Results are provided as data frame or as *igraph* objects, enabling for subsequent analyses in R or in advanced analysis tools like *Cytoscape* (Cline *et al.*, 2007). Through the *S4* class architecture PATH2PPI can be easily extended by further prediction and validation algorithms. The example depicted in Figure 1, the prediction algorithm and all features of PATH2PPI are described in detail in the tutorial, see the supplement and the corresponding Bioconductor web site.

4 Conclusion

We introduced a new R package to predict PPI networks based on sets of proteins which may belong to a specific biological pathway, providing an intuitive visualization and usability. We implemented PATH2PPI to reveal putative proteins and interactions for a pathway or a biological process in organisms for which nearly none PPI information is available. The results can serve as starting points for further network modeling studies and experimental validations.

Acknowledgement

We thank Heiko Giese, Hendrik Schäfer and Tim Schäfer for testing and the reviewers for the valuable comments.

Funding

This work was partly funded by the LOEWE program of the State of Hesse (Germany), the 'Bundesministerium für Bildung und Forschung' (BMBF) and by the Prof. Dr. Dieter Platt Stiftung.

Conflict of Interest: none declared.

References

Camacho,C. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.

Chatr-aryamontri,A. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D4708.

Chen,C.C. *et al.* (2009) PPIsearch: a web server for searching homologous protein–protein interactions across multiple species. *Nucleic Acids Res.*, **37**, W369–W375.

Cline,M.S. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.

Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *Int. J.*, **1695**, 1–9.

Deng,Y. *et al.* (2013) ppiPre: predicting protein–protein interactions by combining heterogeneous features. *BMC Syst. Biol.*, **7**, S8.

Franceschini,A. *et al.* (2013) STRING v9. 1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.

Güldener,U. *et al.* (2006) MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res.*, **34**, D436–D441.

Huber,W. *et al.* (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods*, **12**, 115–121.

Kanehisa,M. *et al.* (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.

Mora,A. and Donaldson,I.M. (2011) iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. *BMC Bioinf.*, **12**, 455.

Murakami,Y. and Mizuguchi,K. (2014) Homology-based prediction of interactions between proteins using averaged one-dependence estimators. *BMC Bioinf.*, **15**, 213.

Osiewacz,H.D. *et al.* (2013) Assessing organismal aging in the filamentous fungus *Podospora anserina*. *Methods Mol. Biol.*, **965**, 439–462.

Pesch,R. and Zimmer,R. (2013) Complementing the eukaryotic protein interactome. *PLoS ONE*, **8**, e66635.

Philipp,O. *et al.* (2013) A genome-wide longitudinal transcriptome analysis of the aging model *Podospora anserina*. *PLoS ONE*, **8**, e83109.

Prasad,T.K. *et al.* (2009) Human protein reference database 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.

Rao,V.S. *et al.* (2014) Protein–protein interaction detection: Methods and analysis. *Int. J. Proteomics*, **2014**, 147648.

Razick,S. *et al.* (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinf.*, **9**, 405.

Wiles,A.M. *et al.* (2010) Building and analyzing protein interactome networks by cross-species comparisons. *BMC Syst. Biol.*, **4**, 36.

Yu,J. *et al.* (2010) Simple sequence-based kernels do not predict protein–protein interactions. *Bioinformatics*, **26**, 2610–2614.

4.4 The autophagy interaction network of the ageing model *Podospora anserina*

Publication status: Published

Summary

In the former transcriptome study it was found that autophagy may play an essential role during ageing since the expression of autophagy-associated genes significantly increased during the lifespan of *P. anserina*. A subsequent study by Knuppertz et al. (2014) proved that autophagy is a longevity assurance mechanism.

To reveal the single components of autophagy, i.e., the involved proteins and how they are interacting, the study aimed to construct and analyse the protein-protein interaction network of *P. anserina*. However, prior to the date of the study, little was known about PPIs in *P. anserina*. Various data repositories which contain PPI data only comprise protein interactions for some well-established model organisms and only minor information for many other species, as for *P. anserina*.

Hence, in the present study, the information available about autophagy in yeast and humans had been transferred to *P. anserina* by means of the currently released Bioconductor package PATH2PPI which uses a prediction approach based on sequence homology. Additionally, the predicted network was expanded and complemented by own experimental data: Firstly, by protein interaction data achieved by a yeast two-hybrid analysis and secondly, by gene expression data from the former longitudinal transcriptome analysis. The final network consists of 89 proteins and 186 interactions.

Subsequently, different bioinformatics and statistical approaches were applied to the constructed network in order to mathematically prove that it is not random but exhibited biological meaningful properties. It was found that the network consists of seven densely connected protein modules which partly correspond to the seven known sub-processes of autophagy. In addition, proteins had been identified which represent "major player" or hub nodes within the constructed network. The subsequent mapping of the gene expression data to the PPI network revealed that autophagy-associated genes are significantly often co-expressed during ageing. Furthermore, as shown by the transcriptome analysis, it was confirmed that expression of genes which are coding for proteins involved in autophagy are continuously up-regulated during ageing.

Summarizing, the study provides a comprehensive age-dependent biological network of the autophagy pathway in *P. anserina* comprising predicted as well as experimental data. It enables the implementation of further studies which aim to analyse autophagy and ageing in *P. anserina* as well as in other organisms.

Authors' contribution

Authors: Oliver Philipp (OP), Andrea Hamann (AH), Heinz D. Osiewacz (HDO) and Ina Koch (IK)

Percentage of contribution for OP:

Conception/Design:	80 %
Experimental work:	0 %
Implementation:	100 %
Analysis/Evaluation:	70 %
Manuscript:	60 %

OP conceived and designed the study under the supervision of IK and HDO. OP performed the bioinformatics analyses, developed the statistical pipeline and analysed the data. OP, AH, IK and HDO wrote the manuscript. AH performed biological experiments and wrote the biological part in the methods section of the manuscript. IK and HDO proofread and revised the manuscript. All authors read and approved the final version of the manuscript.

Philipp *et al. BMC Bioinformatics* (2017) 18:196
DOI 10.1186/s12859-017-1603-2

BMC Bioinformatics

RESEARCH ARTICLE Open Access

The autophagy interaction network of the aging model *Podospora anserina*

Oliver Philipp[1,2], Andrea Hamann[2], Heinz D. Osiewacz[2] and Ina Koch[1*]

Abstract

Background: Autophagy is a conserved molecular pathway involved in the degradation and recycling of cellular components. It is active either as response to starvation or molecular damage. Evidence is emerging that autophagy plays a key role in the degradation of damaged cellular components and thereby affects aging and lifespan control. In earlier studies, it was found that autophagy in the aging model *Podospora anserina* acts as a longevity assurance mechanism. However, only little is known about the individual components controlling autophagy in this aging model. Here, we report a biochemical and bioinformatics study to detect the protein-protein interaction (PPI) network of *P. anserina* combining experimental and theoretical methods.

Results: We constructed the PPI network of autophagy in *P. anserina* based on the corresponding networks of yeast and human. We integrated PaATG8 interaction partners identified in an own yeast two-hybrid analysis using ATG8 of *P. anserina* as bait. Additionally, we included age-dependent transcriptome data. The resulting network consists of 89 proteins involved in 186 interactions. We applied bioinformatics approaches to analyze the network topology and to prove that the network is not random, but exhibits biologically meaningful properties. We identified hub proteins which play an essential role in the network as well as seven putative sub-pathways, and interactions which are likely to be evolutionary conserved amongst species. We confirmed that autophagy-associated genes are significantly often up-regulated and co-expressed during aging of *P. anserina*.

Conclusions: With the present study, we provide a comprehensive biological network of the autophagy pathway in *P. anserina* comprising PPI and gene expression data. It is based on computational prediction as well as experimental data. We identified sub-pathways, important hub proteins, and evolutionary conserved interactions. The network clearly illustrates the relation of autophagy to aging processes and enables further specific studies to understand autophagy and aging in *P. anserina* as well as in other systems.

Keywords: Autophagy, Aging, *Podospora anserina*, Protein-protein interaction, ATG8, Network analysis, Topological properties

Background

In the last decade, different forms of autophagy were detected as major pathways active in recycling during starvation and in molecular quality control (QC). During macroautophagy, hereafter termed autophagy, molecules, organelles, or whole bacteria become enclosed by membranes. The resulting autophagosomes are subsequently delivered to lysosomes, in animals, or the vacuole, in plants and fungi, where they are enzymatically degraded. The building blocks of the degraded components, e.g. amino acids, are reused to generate new functional components. The main molecular processes are conserved among organisms. A core machinery encoded by "autophagy-related genes" (ATG) controls the induction of autophagy, vesicle nucleation and expansion, fusion of autophagosomes with lysosomes or vacuoles, and the final degradation of the corresponding components [14, 21].

In various organisms, evidence for a link of autophagy with aging processes has been demonstrated [18]. For example, an impaired autophagy system shortens the lifespan of mice [36] or is involved in the development of various age-dependent diseases such as Alzheimer's,

*Correspondence: ina.koch@bioinformatik.uni-frankfurt.de
[1]Molecular Bioinformatics, Institute of Computer Science, Faculty of Computer Science and Mathematics and Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe-University Frankfurt am Main, Robert-Mayer-Str. 11-15, 60325 Frankfurt am Main, Germany
Full list of author information is available at the end of the article

Huntington's, or Parkinson's disease [27]. At the same time, an increase of autophagy by treatment with rapamycin or by overexpression of specific autophagy-related genes leads to an extension of the lifespan in several organsisms. For example, in yeast treatment with 10, 20, and 40 *nM* rapamycin extends chronological lifespan. This effect is dependent on the presence of ATG1 and ATG7 [2]. In fruit flies, 50, 200 and 400 μM rapamycin increases median lifespan by up to about 20% in females and to a lesser extent also in males [4]. Moderate overexpression of *Atg5* results in enhanced autophagy and extends median lifespan by 17% in mice [36].

In the fungus *Podospora anserina*, first evidence for a role of autophagy in aging was obtained from a genome-wide longitudinal transcriptome analysis [34]. In this fungal aging model, growth of peripheral hyphae is limited. After about 25 days wild-type strain 's' stops growth and dies. The lifespan of this fungus is defined as the period of growth and can be measured in days or in centimeters [30]. In the longitudinal study, the transcriptome of *P. anserina* was captured at seven consecutive age points and analyzed. The approximately 10,000 expression profiles were filtered to yield age-dependent profiles. Two groups of expression profiles with similar patterns were of particular interest: 1,202 continuously down- and 418 up-regulated profiles. A gene ontology (GO) analysis evinced that genes involved in autophagy were significantly up-regulated (p-value=$3.52e{-}04$) while those of the ubiquitin proteasome system (p-value=$9.90e-04$) were down-regulated (the details of the analyses are described in Philipp et al. (2013) [34]).

A subsequent analysis revealed that autophagy is a longevity assurance mechanism in *P. anserina* [19]. Ablation of essential autophagy machinery components, such as ATG1, leads to a shortened lifespan of the corresponding strains. Moreover, it was found that aging of the strains leads to an increase of autophagosomes and magnified autophagy activity.

The identification of autophagy as a longevity assurance mechanism in *P. anserina* is consistent with findings proposed, but not mechanistically elucidated, in other systems [22, 23, 32]. Nevertheless, only little is known about the regulatory network of autophagy as a QC mechanism effective during aging. A systematic analysis of the involved components of the autophagy machinery during aging is a promising approach to uncover new information about these branches of the network relevant for aging and lifespan control.

Although *P. anserina* is a well-established aging model for which a large body of data about pathways, affecting aging and lifespan, has been generated [30, 41], there is only little known about the single proteins and interactions involved in autophagy in this species. For example, the KEGG database [16] gives a small sub-network of autophagy in *P. anserina*. Recently, the STRING database [11] maintains some predicted interactions for *P. anserina* but with a very limited possibility to track the source of the provided interactions.

To overcome this incompleteness of PPI data, we applied the software tool PATH2PPI [33] to predict the PPI network of autophagy in *P. anserina* based on information about autophagy from the two reference systems yeast and human. Additionally, we performed a yeast two-hybrid analysis to identify interaction partners of the *P. anserina* ATG8 protein, a central component of autophagy, which is homologous to the ATG8 protein in yeast and the LC3/GABARAP protein family in human. Combining the findings of the PPI prediction approach with those of the yeast two-hybrid analysis and the expression profiles of the age-dependent transcriptome analysis, we were able to generate a comprehensive putative PPI network of autophagy in *P. anserina*, which illustrates the relation between autophagy and aging.

Methods

Yeast two-hybrid analysis

Using PaATG8 (UniProt Q8J282) as a bait, a yeast two-hybrid analysis was performed by Dualsystems Biotech AG (Zurich, Switzerland). The bait construct for yeast two-hybrid screening was generated by subcloning the cDNA, encoding amino acids 1 to 115 of PaATG8 (= Pa_3_5250) into the vector pLexA-DIR (Additional file 1). To prevent conjugation to substrates, not the full-length protein sequence (121 amino acids) was used, but a truncated version, lacking the last six codons (encoding Gly-Gly-Phe-Glu-Thr-Ala). The bait construct was transformed into the NMY32 yeast strain (MATa his3200 trp1-901 leu2-3,112 (lexAop)8-ADE2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ GAL4) using standard procedures [12]. Correct expression of the bait was verified by western blot analyses of cell extracts using a mouse monoclonal antibody directed against the LexA domain. The absence of self-activation was verified by co-transformation of the bait together with a control prey and selection on minimal medium, lacking the amino acids tryptophan, leucine, and histidine (selective medium). For the yeast two-hybrid screen, the bait was co-transformed together with a cDNA library into NMY32. To obtain a cDNA library, total RNA of three *P. anserina* wild-type isolates (strain 's') was isolated from liquid cultures treated with 500 ng/ml rapamycin 3.5 h before harvest. pGAD-HA was used as prey vector (Additional file 2).

$5 * 10^6$ transformants were screened, yielding 96 transformants that grew on selective medium and were positively tested for β-galactosidase activity, using a quantitative β-galactosidase assay. From 70 of these clones successful plasmid rescue and sequencing from a 5' junction

sequencing primer was possible. Six clones were discarded because the coding sequence was not in frame. Two additional clones were discarded since they encode transcription factors and therefore, obviously represent false positives. The remaining 62 clones were assigned to 21 different proteins (see Additional file 3).

Databases and data repositories

In the UniProt database, we searched for proteins tagged with the key words "autophagy" or "macroautophagy" and additionally declared to be "reviewed" to get the high-quality, manually annotated entries based on experimental results [39].

To get all known interactions, we used the latest IREFINDEX MITAB file (release 14) for human (taxonomy id 9606) and yeast (taxonomy id 559292) [37]. The sets of relevant reference proteins and the IREFINDEX files were assigned to PATH2PPI. This software package uses IREFINDEX for retrieval of interaction data since it is a meta database for protein-protein interactions which combines interactions from different databases and arrange them with many additional information. This information is provided in a well compiled format and can easily be parsed and handled by computational analysis pipelines and methods.

For the protein or gene identifiers of yeast and human we used the accession numbers provided by the UniProt database. In contrast, each *P. anserina* accession number or identifier, respectively, was adopted from the *P. anserina* genome database [10].

Homology search and settings for PATH2PPI

PATH2PPI requires the homologous relations for each of the previously gathered proteins. We performed a BLAST search, using the BLAST toolkit (release 2.2.29) with default settings and an E-value cutoff of $1e-4$ [5].

We applied PATH2PPI, using default settings except the definition of the homology range, which is defined by a lower and an upper E-value bound, i.e. in this context we interpreted and defined the BLAST E-value as a degree of "similarity" (see Philipp et al. (2016) [33]). If a BLAST E-value for two proteins is greater than or equal to the upper bound, h_u, the score equals 0. An E-value which is less than or equal to the lower bound, h_l, will lead to the top score of 1. Thus, each E-value between these bounds will be scored according to the range $[0, 1]$. The existence of only one homolog to a protein with an E-value of at most $1e-5$ will also lead to the top score of 1. To find appropriate parameters for these two bounds, we repeatedly run the prediction algorithm, using different values for h_u and h_l. We started with an E-value of $1e-20$ and decrease it stepwise by $1e-20$ until we reach $1e-200$:

$h_u = [1e-20, 1e-40, 1e-60, \ldots, 1e-160]$ and

$h_l = [h_u * 1e-40, h_u * 1e-60, h_u * 1e-80, \ldots, 1e-200]$.

In our case, it is not feasible to use default evaluation approaches of prediction classifiers, like ROC curves or a precision and recall approach, since only sparse information is available. For example, a false-positively predicted interaction could occur due to the lack of information. To choose appropriate values for the homology range and to evaluate the different results, we compared each network with the very small autophagy network of *P. anserina* of the KEGG database [16] (see Additional file 4). This small, manually curated network consists of 15 proteins and 12 interactions, which are based on a mapping of *P. anserina* proteins and the corresponding interactions to the basal autophagy pathway. This basal pathway is strongly conserved throughout most eukaryotic organisms. We covered as much as possible of these known interactions and proteins.

Statistical significance

In biological research, the experimental data are often incomplete and of different quality and quantity. To ensure that the set of data we consider is large enough for biological analysis, we have to prove that the model could not be randomly formed, but exhibits "real-world" properties.

Furthermore, the transfer and compilation of interactions from different sources to one PPI network must rely on well-defined rules and assumptions. The applied approach must not produce a random network. To mathematically prove the predicted network for biological consistency and to decide whether it differs from networks generated by chance, we computed several topological features [28]. We compared these features with the median value of thousand networks of randomly chosen interaction partners of the same proteins. We took two sets of randomly generated networks. The first set (*set 1*) consists of networks with unchanged node degree, i.e. the number of adjacent edges was the same, only the corresponding interacting partners were randomly changed. The second set (*set 2*) consists of randomly chosen interaction partners, only preserving the total number of edges. Which one of the two sets we considered depended on the topological feature to be compared. Unless otherwise stated, for the computation of the single topological features, we applied the functions of the IGRAPH-package [9] using default settings. We computed the topological features for the predicted and for each random network to show the differences between them. For this comparison approach, we considered the predicted proteins and interactions resulted from PATH2PPI, but not the proteins from the yeast two-hybrid analysis.

We considered the following topological features: the diameter, modules and modularity, the transitivity or clustering coefficient, the node degree and the node

betweenness (hereinafter called degree or betweenness, respectively). The comparison of the different topology values of the predicted network with those of the random networks requires appropriate statistical tests. Hence, to compute empirical p-values for each of the considered topologies, we applied appropriate statistical tests for the corresponding topological feature (see detailed descriptions below and Additional file 5).

Diameter

The diameter of a network is defined as the longest path of all shortest paths between each two nodes. "Real-world" networks, such as metabolic or protein interaction networks, exhibit a small diameter, having a "small world" architecture, indicating that any node can be reached via a relatively short path from another node [44]. We compared the diameter of the predicted network with the median diameter of the randomly rewired networks (*set 1*). Since the set of diameters of the random networks represents a discrete distribution, we computed the relative frequency of each diameter which can be interpreted as an empirical p-value.

Modules and modularity

Modules are subgraphs, in which the connections within them are much denser than between them. In a biological network, modules often correspond to different functional sub-pathways. The modularity represents a quality measure for network partitioning. For module detection we applied a random walk approach as described by [35]. The computation of the modularity is based on the method of Clauset et al. (2004) [8]. To compute an empirical p-value for the modularity of the predicted network, we first applied the Shapiro-Wilk test on the modularities of the random networks to check whether they are normally distributed. Subsequently, we used the normal distribution with the mean and standard deviation of the random modularities (*set 1*) to compute the p-value of the predicted network's modularity.

The transitivity or clustering coefficient

The transitivity or clustering coefficient indicates how dense the nodes of a network are connected. More precisely, the global transitivity is defined as the ratio of the triangles and the connected triples in the network. If n is the number of triangles and t the number of triplets, then the clustering coefficient c is computed by $c = \frac{3n}{t}$ [3]. To compute the p-value for the predicted network's cluster coefficient, we applied the same test procedure as for the modularity.

Node degree and node betweenness

To reveal major hubs in the network, we computed the node degrees and the node betweennesses. Generally, PPI networks exhibit the scale-free property, i.e. they consist of some nodes with a high degree and many nodes with a low degree, and the degree distribution follows a power law.

The betweenness property is an indicator for the importance of a node, i.e. it can reveal how strong a node influences a network. The higher a node's betweenness the more increases its importance. It indicates the "traffic load" on one node under the assumption that the flow of information follows the shortest path.

Each network exhibits a certain distribution of degrees and betweennesses. To compare the cumulative distribution of the random networks (*set 2*) with the distribution of the predicted network, we applied the Kolmogorov-Smirnov test, which computes a p-value for the probablity that two samples were drawn from the same distribution. A significantly low p-value will indicate that the random network's betweennesses and degree distributions, are significantly different from those of the predicted network.

Statistical test for co-expression and significant regulation

We applied two different approaches to integrate the information available from the corresponding age-dependent expression profiles of the longitudinal transcriptome analysis [34].

First, we computed the Pearson correlation coefficient (Pcc) for each pair of expression profiles of all genes. A pair of profiles was assumed to be co-expressed, if the Pcc is greater than or equal to a threshold of 0.9. Two profiles were assumed to be expressed in opposite directions, if the Pcc is less than or equal to -0.9. We divided these Pcc values into four groups to achieve the contingency table for Fisher's exact test, first, a group of all co-expressed pairs of genes, second, a group which comprises all other co-expressed pairs, third, a group of all autophagy gene pairs which are not co-expressed, and fourth, a group of pairs which were neither co-expressed nor involved in autophagy. To compute expectation values or the number of expected co-expressed genes in the group of autophagy-associated genes, we applied the hypergeometric distribution function.

Second, we computed the Pcc of each expression profile in correlation with age to reveal which genes are up- or down-regulated in the course of aging. According to the statistical approach for co-expression, we applied the Fisher's test to prove that the number of up-regulated transcripts of the predicted PPI network is higher than expected by chance. For this purpose, we divided the 10,059 expression profiles into four sets to get the contingency table, first, a group of up-regulated transcripts in the PPI network, second, all other transcripts, third, all other up-regulated profiles, and fourth, the remaining profiles.

The entire gene expression data from the age-dependent transcriptome analysis can be found in Philipp et al. (2013) [34] or is available at the European Bioinformatics Institute's ArrayExpress public data repository (http://www.ebi.ac.uk/arrayexpress/) with the accession number E-MTAB-2016.

Results

To achieve the autophagy PPI network of *P. anserina* we applied a comprehensive bioinformatics pipeline, which includes different data sets, sources and analysis approaches (see flowchart in Fig. 1).

Initial data gathering and homology comparison

We applied the software package PATH2PPI [33] which requires three data sets for both reference species. First, PATH2PPI requires the proteins which are associated with autophagy in the corresponding species. In the UniProt database, we identified 259 proteins in human and 108 proteins in yeast directly or indirectly associated with autophagy. Second, PATH2PPI uses the IREFINDEX repository to find all known interactions between these proteins. We identified 487 interactions in human for which both proteins were in the initial protein list. Analogously, we found 313 interactions for yeast (see Additional file 6). Third, PATH2PPI needs to know the homologous relationships between each protein of *P. anserina* and each protein of human and yeast. We applied a comprehensive BLAST search, using the proteomes of *P. anserina*, yeast, and human. We identified 4686 *P. anserina* proteins with one or more human homologs and 4995 *P. anserina* proteins with yeast homologs, each with an E-value less than or equal to $1e-4$.

Parameter fitting for PPI network prediction

To determine appropriate initial parameters for the construction of the PPI network, we first run the prediction algorithm several times with different values for the lower and upper bound of the homology range. Subsequently, we compared the predicted PPI networks with the "small" autophagy PPI network of *P. anserina* in the KEGG database, which consists of 15 proteins and

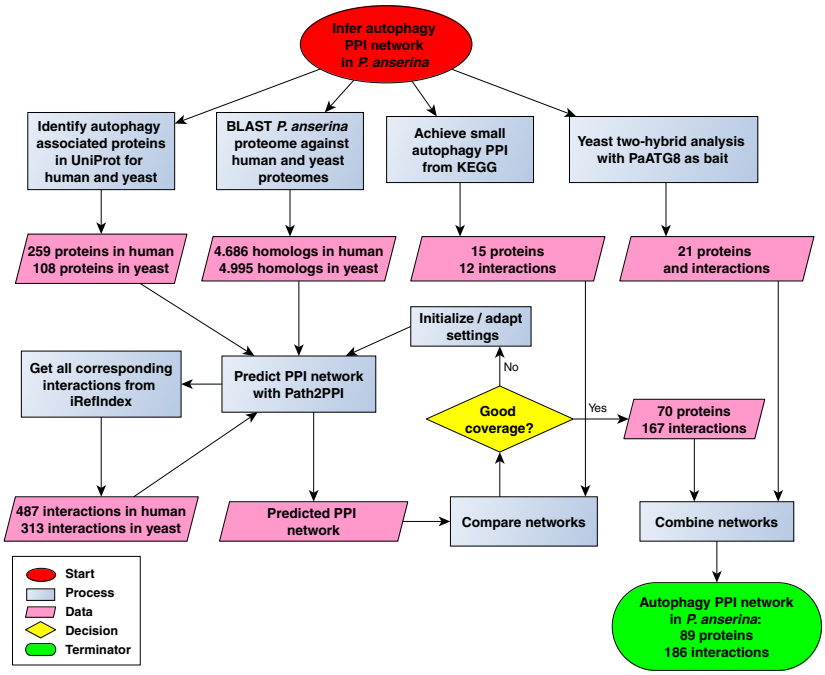


Fig. 1 Flowchart of the analysis pipeline which was applied to infer the autophagy PPI network in *P. anserina*. At the beginning of the pipeline, each autophagy-associated protein from yeast and human was identified. In parallel, a comprehensive BLAST search was applied, using the proteomes of *P. anserina*, yeast and human. Based on the initial protein set, the software package PATH2PPI was applied to find each interaction in the IREFINDEX database which is associated with autophagy in yeast and human. Subsequently, PATH2PPI used the sets of reference proteins, interactions and BLAST results to predict the putative autophagy PPI in *P. anserina*. To achieve the best possible prediction parameters, the prediction approach was repeated several times and each result was compared with the small, manually curated autophagy PPI network from KEGG. The best predicted PPI network was combined with the experimental results of a yeast two-hybrid analysis. Finally, the pipeline led to the autophagy PPI network in *P. anserina* which consists of 89 proteins and 186 interactions

12 interactions. The best possible result was a coverage of 14 proteins (93%) and 9 (75%) interactions (see Additional file 7). Since we wanted the best possible prediction with a minimum of strictness and a maximum of possible discrimination between the different E-values, we applied a range of $[1e-200, 1e-20]$ for the autophagy prediction approach.

Comparison of autophagy-associated interactions in human, yeast, and *P. anserina*

Figure 2 depicts the homology-based, hybrid PPI network, which consists of all proteins of the human's and yeast's PPI networks considered for the autophagy PPI network of *P. anserina*, with 70 proteins and 167 potential interactions (see also Additional file 8). It is based on 39 proteins and 75 interactions of the human PPI network and 50 proteins and 114 interactions of the yeast PPI network. Most of the reference interactions (see Additional file 9: Table S4) are characterized by the terms "physical interactions" (61%) or "direct interaction" (27%), in contrast to the remaining PPIs characterized by "association" (2%), "colocalization" (<2%), or "covalent binding" (<1%). Less than 10% of the reference interactions were not characterized by IREFINDEX (not shown). Additionally, each homologous relation is depicted as dotted edge. Overall, 57 homologous relationships of *P. anserina* proteins to human and 65 to yeast were detected.

The result from the initial parameter fitting approach was a first proof of concept. We compared the predicted PPI network with the "small" autophagy PPI network of *P. anserina* (see Additional file 4) and found that 14 of the 15 proteins and 9 of the 12 interactions were correctly predicted. The protein Pa_1_7190 (putative ATG10, as referred by the KEGG database) was not found in the predicted network, since no homolog exists neither in yeast nor in human. The assignment of the protein Pa_1_7190 to ATG10 by KEGG was most probably based on a homologous protein in a closely related fungus, for which the ATG10 protein had already been investigated and characterized. Two of the three not predicted interactions are

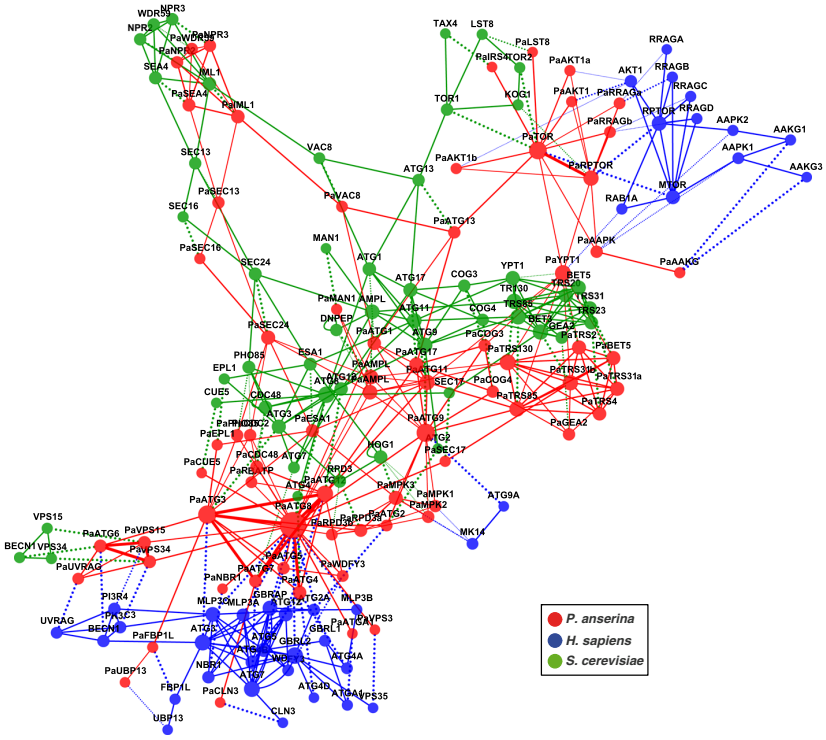


Fig. 2 Reference and predicted PPI networks in relation with the inferred homology network of autophagy in *P. anserina*. The figure illustrates the relation of each reference network with the predicted network. This hybrid network is a combination of different network types: it comprises both reference PPI networks, the predicted PPI network and the homology network which combines each of the PPI networks. All proteins from human (*blue nodes*) and yeast (*green nodes*) and their corresponding interactions used for the *P. anserina* PPI network (*red nodes and edges*) are shown. Homologous relationships between two proteins are depicted as *dotted lines*

based on Pa_1_7190. The third one is the interaction of Pa_2_1770 (putative PaATG7) with Pa_5_5430 (putative PaATG4). In the predicted PPI network, both proteins are not directly interacting, but are connected via one of the three proteins, Pa_3_5250 (PaATG8), Pa_1_20610 (PaATG3), or Pa_4_7460 (PaATG12). The missing direct link between PaATG7 and PaATG4 was probably due to the very strict criteria and settings we applied for the network prediction approach. This ensured that we avoid as much false positives as possible but could also miss potential true positives.

Yeast two-hybrid analysis

The yeast two-hybrid analysis, using PaATG8 as a bait, revealed 21 putative interacting partners of PaATG8. Additional file 10: Figure S1 (red nodes) depicts the corresponding interactions and all homologous proteins from yeast and human with an E-value of at least $1e-20$. Interestingly, we found interactions in the reference species of ATG8 with ATG4 and of ATG8 with NBR1. These two interactions were the most promising ones provided by the yeast two-hybrid analysis because independent clones were found seven and 26 times, respectively (see numbers at the red edges in Additional file 10: Figure S1).

The autophagy PPI network

We combined the two PPI networks, the predicted and the one deduced from the yeast two-hybrid analysis, to one autophagy PPI network of *P. anserina* (Fig. 3). The PPI network consists of 89 proteins participating in 188 interactions (see also Additional file 9). Additionally, we integrated the age-dependent expression profiles of the corresponding transcripts from the former transcriptome study to indicate which of the autophagy-related genes are down- or up-regulated during aging [34].

Many of the predicted proteins were unknown or uncharacterized for *P. anserina* and no trivial protein names were assigned. Hence, if applicable, we consider

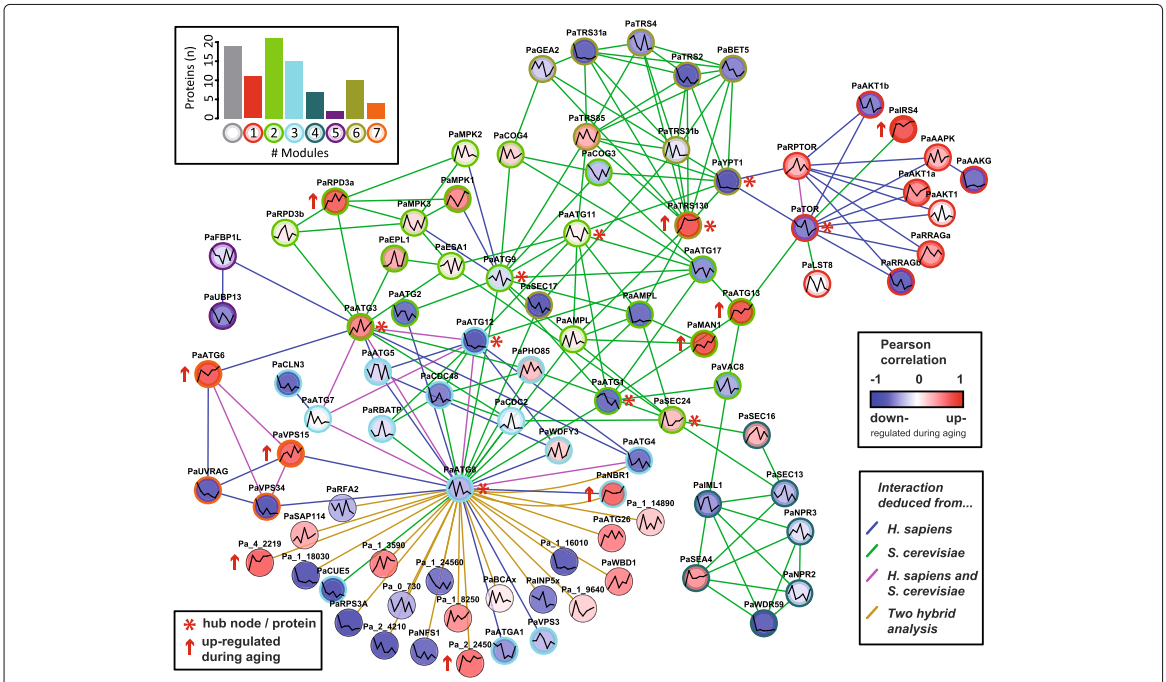


Fig. 3 Predicted autophagy PPI network of *P. anserina* with age-dependent expression profiles. The colors of the solid lines indicate the species that exhibits the interaction (*green edges* for yeast and *blue* for human). If an interaction was found in both species, the edge is colored *purple*. In addition, if an interaction was found in the yeast two-hybrid analysis, an edge is colored *yellow*. For example, an interaction of PaATG8 and PaATG4 was found in yeast and human (*purple edge*) and additionally in the yeast two-hybrid analysis (*yellow*). Hence, this interaction is represented by two edges (*purple* and *yellow*), since it was found in all three PPI data sources. The nodes are color-coded according to the Pearson correlation coefficient of the corresponding age-dependent expression profiles. The more the expression of a transcript was increased during aging the more red is the node. The opposite direction is colored *blue*. The *red arrows* indicate transcripts which were continuously up-regulated during aging according to [34]. Red stars highlight the hub nodes, see Table 1. The number of proteins per module - from module 1 (*red*) to module 7 (*orange*) - are depicted in the bar plot in the upper left corner. The *grey bar* corresponds to proteins which were not assigned to a module, but subsequently added. Nodes, which were grouped to one of the seven functional modules, were discriminated by similar border colors

the protein names of the most homologous proteins from yeast or human, respectively, otherwise we used the protein identifiers provided by the *P. anserina* genome database [10].

The global topology of the PPI network indicates biological significance

Biological networks differ from randomly generated networks by well-defined topological features, such as the reachability of the single nodes, the degree of cross-linking and grouping, and the general size of the network. To substantiate the biological significance of the elaborated network, we computed different topology features. These features were compared with those of the randomly generated networks (see Table 1 and Additional file 5). The first topological feature was the network's **diameter**. The median diameter of the random networks was 6 and of the predicted network 8 (p-value=0.007). As one may assume that biological networks have smaller diameters due to the need of increasing the network's efficiency, e.g. transition time or the flow of information, the result seems to be counterintuitive. Nevertheless, biological networks have diameters greater than expected, since a higher degree of modularization of a network increases its diameter [43].

We compared the **modularities** of the random networks and the predicted network. We detected seven modules (Fig. 3, node border colors) for the predicted network. The proteins detected by the yeast two-hybrid analysis, which we added later, were not considered and therefore are not contained in the modules. Based on these seven modules, we yield a modularity of about 0.57 (p-value=$1e-13$) for the predicted network which is significantly larger than the median modularity of 0.32 for the random networks.

We considered the **transitivity** or the **clustering coefficient**, respectively. The transitivity of the predicted network was about 0.35 (p-value=$3e-42$), in contrast to the random networks' median of transitivities of about 0.12. This is in accordance with the observation that small-world networks exhibit higher clustering coefficients or transitivities, respectively [44].

Subsequently, we considered the nodes' **degree** distributions and the nodes' **betweennesses** (see Additional file 11). While the betweenness distribution of the predicted network significantly differs from the betweenness distribution of the random networks (p-value=$1.17e-08$), the significance test of the degree distribution did not resulted in such a significant p-value (0.09). Most probably this was due to one major drawback of the degree distributions. In contrast to the betweennesses, the degree distributions consists of only a few different discrete values, i.e. 13 for the predicted network and 17 for all random networks. Hence, the Kolmogorov-Smirnoff test was not able to significantly distinguish both distributions. To overcome this drawback it was more feasible to visually compare the courses of the random and the real degree distributions. As expected, the distributions of the node degrees of the randomly chosen interaction partners were bell-shaped (Additional file 10: Figure S2a), in contrast to the degree distribution of the predicted PPI network which rather follows a power-law distribution depicted in the log-log-scaled plot with the fitted regression line (Additional file 10: Figure S2b-bottom). The power-law distribution indicates that there are many nodes with a low degree and some nodes with a high degree (scale-free property). These hub nodes indicate that the corresponding proteins are probably major players within the pathway (see also Table 2). We combined the node degrees with the node betweennesses to get the top ten of the most important proteins in the network (Table 2). In addition, for each of these ten proteins, we used the *Saccharomyces* Genome Database [7] to achieve information about the effects on yeast strains where the corresponding homologous gene had been knocked out.

We describe these major players in more detail in the "Discussion" section.

Table 1 Results of the topological analysis

Topology	Pred. net. (median)	Rand. net.	Test statistic	p-value
Diameter	8	6	Relative frequencies	0.007
Modularity	0.57	0.32	Shapiro-Wilk + Norm.dist	$1.22e-13$
Transitivity	0.35	0.12	Shapiro-Wilk + Norm.dist.	$2.74e-42$
Betweenness	*NA*	*NA*	Kolmogorov-Smirnov	$1.17e-08$
Degrees	*NA*	*NA*	Kolmogorov-Smirnov	0.09

Different test procedures were applied according to the corresponding topological feature (see "Methods" section and Additional file 5). The topological values of the predicted network and the median values of the random networks are listed. In addition, the corresponding p-values are shown, indicating the significance of the difference between the random and the predicted network. The computation of a single topological value for transitivity and betweenness was not applicable, since each node of each network has one betweenness and one degree. Here, only the distributions had been compared using the Kolmogorov-Smirnoff test

Table 2 The ten most important proteins of the predicted PPI network

Module no. (color code)	Protein identifier	Alias (no. of interactions)	Ranks degree / between.	Total rank	Effects in knockout yeast strains
3 (turquois)	Pa_3_5250	PaATG8 (22)	1 / 1	2	ne, aa, dv
2 (green)	Pa_5_5550	PaATG9 (13)	2 / 4	6	ne, aa, dv
1 (red)	Pa_4_9630	PaTOR (11)	4 / 2	6	TOR1: ne; TOR2: e
2 (green)	Pa_1_20610	PaATG3 (13)	3 / 6	9	ne, aa, dv
6 (ocher)	Pa_3_7690	PaYPT1 (9)	9 / 3	12	e
2 (green)	Pa_1_14210	PaATG11 (9)	8 / 8	16	ne, ma
3 (turquois)	Pa_4_7460	PaATG12 (10)	5 / 12	17	ne, aa, dv
6 (ocher)	Pa_5_4470	PaTRS130 (10)	6 / 13	19	e
2 (green)	Pa_5_5670	PaSEC24 (7)	16 / 9	25	e
2 (green)	Pa_7_10890	PaATG1 (6)	22 / 5	27	ne, aa, dv

The first column gives the module and its color according to Fig. 3. The column "Ranks" gives the ranks based on the degree and the betweenness. We summed up both values to a total rank and sort the table in descending order. The last column gives the most important effects on the yeast strain where the corresponding homologous gene had been knocked out [7]. Abbreviations: ne=non-essential, aa=autophagy absent, dv=decreased viability, e=essential, ma=mitophagy absent. For PaTOR, two homologs exist, TOR1 and TOR2. The protein identifiers were adopted from the *P. anserina* genome database [10]

Protein coding genes of the autophagy PPI network are significantly often co-expressed during aging

We were interested in how the coding genes of the predicted proteins are connected on the gene regulatory level. Hence, we considered the former longitudinal transcriptome analysis [34] and computed the Pearson correlation coefficients of each pair of the corresponding age-dependent expression profiles. Applying a threshold of 0.9 for the Pcc, we found 165 co-expressed gene pairs among the autophagy-related genes (see Additional file 12). Gene pairs, which show similar expression profiles, are probably co-regulated during aging. We computed an expectation value of about 85 co-expressed gene pairs and a p-value of $1.008e-14$, i.e. exhibiting 165 co-expressed gene pairs is very unlikely to occur by chance. Subsequently, we applied a threshold of -0.9 to search for autophagy-related gene pairs with mirrored expression profiles. The corresponding genes were probably contrarily co-expressed during aging. We found 54 pairs to be co-expressed in an opposite manner with an expectation value of 25 pairs and a p-value of $5.79e-07$. Nevertheless, when comparing the co-expressed gene pairs with the PPI network, we did not find a significant overlap, i.e. if two proteins are interacting, then the corresponding gene pairs are not necessarily co-expressed. This finding suggests that autophagy is indeed strongly age-dependently regulated, but the regulatory program on the genetic level is different to that on the protein level.

Further, we tested whether the genes, coding for the proteins of the predicted PPI network, were more frequently up-regulated during aging than expected by chance as found in the former transcriptome study. We computed a p-value of 0.003728 indicating a significant number of up-regulated genes for a given confidence level of 0.99.

Overall, the additional consideration of the transcriptome data associated with the proteins of the autophagy PPI network led to two major conclusions. First, we found a significant number of co-expressed gene pairs and groups, suggesting a strong, age-dependent regulatory relationship of the predicted proteins on the gene expression level. Second, we confirmed that genes associated with autophagy are significantly up-regulated during aging. Figure 3 may appear contradictory to this result as it seems that there are more down-regulated than up-regulated expression profiles. In fact, in the transcriptome study we found three times more decreasing than increasing expression profiles. Hence, within the autophagy pathway there are significantly more increasing expression profiles than expected by chance given the ratio of down- to up-regulated profiles.

These findings substantiated that investigations on the autophagy pathway and its relation to aging should include both protein as well as gene expression data even if both biological levels are not comparable.

Discussion

Evidence for a biological network

The presented PPI network of autophagy in *P. anserina* is based on a computational prediction approach combined with additional interactions deduced from the yeast two-hybrid analysis. We applied well-established, mathematical approaches to verify that the network was not generated by chance. We considered topological features which are known to be valid for biological networks.

One of these features is the larger network diameter in contrast to random networks, which is due to the higher degree of modularization. The second feature is the cluster coefficient, which is higher than its random

counterpart. A high cluster coefficient indicates that some highly connected subgraphs exist. Often in a biological network, subgraphs can be interpreted as functional sub-pathways, for example as a metabolic or signaling pathway.

The autophagy pathway can be manually divided into the seven disjunctive sub-pathways: 1. induction and initiation, 2. targeting of substrates, 3. autophagosome nucleation, 4. autophagosome expansion and completion, 5. autophagosomal fusion with the vacuole, 6. the breakdown of the autophagosomes and the substrates, and 7. export of the secreted components.

Strikingly, also the independent module detection approach proposed seven modules of highly connected autophagy protein groups. The presence of important marker proteins led to the following alignment: one module which comprises the TOR kinase (red module in Fig. 3) corresponded to the initiation processes. Another module referred to the induction process, involving the putative proteins PaATG1, PaATG13, PaATG11, PaVAC8, PaATG17, and other proteins necessary for autophagy induction after signaling by the TOR pathway (green module in Fig. 3).

A third module covers the autophagosome nucleation and / or elongation step, comprising the proteins PaATG5, PaATG7, PaATG8, and PaATG12 (turquois module in Fig. 3).

We considered the degree distribution and the betweenness, two topological features which provide information about "major players" in the network. The degree distribution of the predicted network reveals that there are many nodes with a small number of neighbors, but a few nodes with much more interacting partners. This is in accordance with the scale-free property of "real-world" biochemical networks.

Hub nodes and evolutionary conserved interactions

The interactions of the yeast two-hybrid analysis were not included in the topological analysis. In contrast to the prediction approach, where we considered all possibilities of interactions, the yeast two-hybrid analysis aimed to reveal interaction partners of only PaATG8 and was not statistically independent. Table 2 lists the ten most important proteins in the PPI network based on their degrees and betweennesses. In addition, for these ten proteins we included the phenotype information of the corresponding knockout yeast strains which is provided by the *Saccharomyces* Genome Database [7]. Since, for most yeast proteins the corresponding knockout strains are available, we were interested in how a loss of the homologous proteins affects the organism. Interestingly, eight of these proteins are either essential or a loss leads to a decreased viability and, more importantly, to an absence of autophagy. The two exceptions are, first, the ATG11 yeast mutant which exhibits a loss of mitophagy and second, PaTOR. Since, for PaTOR two homologs exist in yeast, TOR1 and TOR2, we included the information of both phenotypes. We found that in contrast to *Tor1*, *Tor2* is an essential gene. These findings reinforced the assumption that a high node degree generally correlates to the node's importance for the pathway.

On the very top of the list of the ten most important proteins, we found the protein **PaATG8**, which is involved in most interactions (highest degree) and also in many of the shortest paths (highest betweenness). The finding of PaATG8 to be one of the major hub proteins in autophagy corresponds to its known relevance for the autophagy pathway in other organisms. The autophagy-related protein PaATG8 is homologous to the yeast ATG8 protein and the LC3 protein in mammals. The ATG8 and LC3 proteins are known to be essential for formation and expansion of the autophagosome membrane during autophagy and to be furthermore crucial for other autophagy-related processes, such as cargo delivery into autophagosomes or for selective targeting of cytosolic components for lysosomal and / or vacuolar degradation [38]. Due to its ubiquitin-like function, the ATG8 protein is part of a larger protein complex which binds to the phospholipid membrane of the autophagosome. In addition, it binds different cargo receptors during selective autophagy and mediates the transport of the autophagosomes to the lysosome or the vacuole, respectively. Due to this known relevance for autophagy, we initially applied the yeast two-hybrid analysis to reveal putative interaction partners of PaATG8 in *P. anserina*, before we started the autophagy interaction study. For most of the proteins identified in the yeast two-hybrid analysis, we found only a weak evidence for interaction with PaATG8, because most of them were identified only in one or two clones (see Additional file 10: Figure S1). These proteins may represent experimental artefacts and unspecific binding issues. They have not been predicted by the bioinformatics approach, indicating that they have not been found and probably do not exist in other species. In contrast, the two most promising interacting partners, **PaNBR1** and **PaATG4**, for which multiple independent clones (26 and 7, respectively) were selected in the yeast two-hybrid screen, were also found in the predicted PPI network.

In mammals, the **NBR1** protein contains a LIR-motif (LC3-interaction region) enabling the proteins to bind to LC3 (aPaATG8 homolog). At the C terminus of PaNBR1, a putative LIR-motif was found with the LIR consensus sequence [DE]-[DE]-[DE]-[WFY]-X-X-[LIV] [15] probably responsible for interaction with PaATG8. Kraft et al. (2010) [20] describe that NBR1 mediates autophagosomal engulfment of misfolded protein aggregates which were too large to get degraded by the 26S proteasome. Interestingly, the number of transcripts of the *P. anserina*

PaNBR1 increases during aging by the factor of 3.1. This could be due to the increasing number of misfolded and larger protein aggregates during aging. In the same transcriptome study we found that the expression of proteasomal genes significantly decreases during aging. We hypothesize that in *P. anserina* during aging, the increased number of *PaNbr1* transcripts also leads to an increased number of PaNBR1 proteins to prevent an accumulation of protein aggregates due to the declined function of the proteasome system.

The protein **PaATG4** is another PaATG8-interacting protein identified by the prediction and the yeast two-hybrid analysis. In yeast and other organisms, ATG4 together with **ATG7** and **ATG3** is active in the conjugation of phosphatidylethanolamine with ATG8, enabling the association of ATG8 with the pre-autophagosomal structure. Remarkably, all three interactions, ATG8 with ATG3, ATG4, and ATG7, were found in both reference species (see purple edges in Fig. 3), indicating an evolutionary conservation of this conjugation system already proposed by Kikuma and Kitamoto (2011) [17] for *Aspergillus oryzae*. In this filamentous fungus, it was found that the deletion of *AoAtg4* leads to impaired autophagy. The same was observed in *Sordaria macrospora*, another filamentous fungus, after transcriptional down-regulation of the ATG7 homolog [29]. The third protein is the E2 enzyme **PaATG3** whose homologs are very important for autophagy in other organisms [1, 25]. It is on the very top of the list in Table 2, indicating its importance for the predicted autophagy PPI network. Summarizing, we suggest that PaATG8 and its interactions with the conjugation system, consisting of **PaATG7**, **PaATG4**, and **PaATG3**, play a crucial role in the autophagy process in *P. anserina* as well.

The ubiquitin-like protein **PaATG12** is the fourth protein which is involved in putative evolutionary conserved interactions (purple edges in Fig. 3). It is associated with PaATG8 and the proteins forming the conjugation system. The yeast homolog, ATG12p, is activated by **ATG7** and finally linked to **ATG5** to build the ATG12p conjugation system, which is essential for autophagosome formation [29]. Meijer et al. (2007) [24] described that genes, coding for the corresponding proteins or interacting partners, respectively, are also conserved in filamentous fungi.

Three additional proteins and interactions were predicted to be evolutionary conserved, since they are based on proteins and interactions in human and in yeast. These are the two vacuolar protein-sorting enzymes, **PaVPS15** and **PaVPS34**, and the autophagy-related protein, **PaATG6**. In *S. macrospora*, it was not possible to generate knockout strains for these genes, suggesting that the corresponding proteins are essential for viability in this fungus [40]. Since both are very homologous to their counterparts, PaVPS15 and PaVPS34 (with a BLAST E-value of 0 each), it is likely that they have similar effects on viability in *P. anserina*. The evolutionary conservation and importance of ATG6 and corresponding homologs (like the mammalian Beclin 1) for autophagy was already described for various other organisms. Even its relation to aging processes and lifespan control has already been discussed [6, 13, 26].

The second and third most important proteins in the predicted network, **PaATG9** and **PaTOR**, were equally important. The reason is that we considered the autophagy pathway, but not all the other biological processes which are affected by the TOR kinase. Nevertheless, for autophagy in *P. anserina*, PaATG9 seems to play an essential role. For other systems, it is known that ATG9 is activated by **ATG1** which is also in the list of the ten most important proteins. The activated ATG9 is required for recruitment of ATG8 to the pre-autophagosomal structure and therefore essential for proper efficiency of autophagy [31].

Transcriptional relationship of aging and autophagy

Although the current study is mainly focused on the PPI network of autophagy in *P. anserina*, we also considered the related transcripts and corresponding expression profiles in the course of aging due to two considerations. First, we were interested in a holistic and systems biology view of autophagy and its relationship to aging. Since we hold a comprehensive set of gene expression data, we included the data into the network. Second, we presumed that there is a considerable relationship of transcription, protein level, and activity of autophagy, because autophagy-associated gene expression significantly increases during aging [34], and the subsequent experimental validation identified an age-related increase of autophagy in later stages of the life cycle, before the system finally breaks down [19].

Initially, we were interested in the general expressional behavior and analyzed whether the genes, coding for the predicted proteins, are more often co-expressed than expected by chance. We found 165 co-expressed gene pairs with a p-value of $1.008e-14$, indicating a very significant and strong co-expression within the set of autophagy-related genes. That are approximately twice as many co-expressed gene pairs as expected in a random network. To the best of our knowledge, it is the first time that a significant degree of co-expression during aging of autophagy-associated genes is demonstrated, supporting again the relevance of the autophagy pathway for aging processes in *P. anserina*. This approach revealed that there is a significant number of co-expression within the genes, which correspond to the predicted proteins and it was experimentally validated that the predicted components of the network are strongly related with each other. By means of the topological analysis, we verified

that the predicted interactions exhibit properties which are characteristic for biological networks. Next, we analyzed the age-dependent expressional tendency of each single gene. At a first glance, it seems that the majority of transcripts, coding for the proteins of the autophagy PPI network, were down-regulated during aging what would be in contrast to the former transcriptome study [34]. Nevertheless in that study, there were essentially more down-regulated than up-regulated transcripts, which was probably due to a general and continuous decline of the organism during aging. Hence, it was of great importance to statistically prove whether the number of up-regulated transcripts exceeds the expectation. Indeed, the p-value of 0.0037 indicated that the amount of ten transcripts, showing a continuously up-regulation during aging, was more than expected by chance, which is in accordance with the transcriptome study.

Since up-regulation of autophagy seems to act as an important mechanism and rescue effort during aging in *P. anserina*, we wanted to identify the single up-regulated components of autophagy. In addition to the hub proteins, we analyzed and discussed the ten proteins, for which the corresponding transcripts exhibit continuously up-regulated expression profiles during aging. These proteins are listed in Table 3 and marked with red arrows in Fig. 3. In the following, we refer **p**roteins with **u**p-**r**egulated expression profiles of the corresponding genes as **PUR** proteins. Interestingly, only one of the PUR proteins is also a hub protein, PaTRS130. In yeast, the homologous protein, Trs130, is part of the TRAPP II complex (Transport protein particle II), which mediates the trafficking of autophagy proteins from the Golgi to pre-autophagosome structures. It was found that autophagy was impaired in a trs130 temperature-sensitive mutant, suggesting an important role for the early autophagy induction processes [45].

Table 3 Proteins with up-regulated expression profiles (PUR proteins)

Protein	Alias	Module no.	Pearson correlation	Ratio
Pa_0_1500	PaATG13	2	0.87	1.74
Pa_1_10350	PaMAN1	2	0.87	2.10
Pa_5_4470	PaTRS130	6	0.87	3.94
Pa_7_7880	PaIRS4	1	0.86	2.79
Pa_1_11020	PaATG6	7	0.84	3.48
Pa_1_5330	PaRPD3a	2	0.81	1.65
Pa_4_2219	Pa_4_2219	TH	0.79	4.96
Pa_2_12420	PaNBR1	3	0.78	3.13
Pa_6_1830	PaVPS15	7	0.77	2.40
Pa_2_2450	Pa_2_2450	TH	0.72	3.38

For ten proteins, the corresponding transcripts were up-regulated during aging (see expression profiles in Fig. 3 of the nodes marked by red arrows). The higher the Pearson correlation coefficient the straighter increases the related expression profile or the number of gene products, respectively. TH indicates that the protein was not assigned to any module, since it was only found in the yeast two-hybrid analysis. The last column gives the ratio of the transcript abundance from the first to the last day of measurement. The protein identifiers were adopted from the *P. anserina* genome database

The other PUR proteins are not important hub nodes with a greater number of interactions or crossing shortest paths. This can be explained by the assumption that gene expression is not or only weakly related with the actual relevance and importance of the corresponding protein. Because we suppose a considerable relationship of transcription, protein level, and activity of autophagy, we also assume that these PUR proteins may play important regulatory roles for down-streamed sub-pathways. For example, we previously supposed that the increase of the transcripts of PaNBR1 is caused by an increased number of misfolded and larger protein aggregates. Another example is the PaATG13 protein, which is involved at the very beginning of the autophagy process when it becomes activated or deactivated by the TOR kinase. Nevertheless, the two most promising PUR proteins are PaATG6 and PaVPS15, since both are the only ones which are involved in evolutionary conserved interactions. Furthermore, PaATG6 homologs were already described to be relevant for aging processes in other organisms. It will be of interest to see whether the corresponding transcripts show an increased abundance during aging in other systems.

Summarizing, the provided PPI network of autophagy in *P. anserina* is based on experimental data and mainly on a prediction approach which relies on sequence homology. Since there are known and reasonable drawbacks of such prediction methods [42], it is important to carefully analyze the results by appropriate mathematical approaches. Finally, only biological experiments allow the validation of the drawn conclusions. Nevertheless, we substantiate the reliability of the predicted network using two main approaches. First, we applied different analyses which revealed topological features that are significant for biological networks. Since these topological features were probably already effective for the origin networks of yeast and human, they indicate that the transfer of the interactions relies on well-defined rules and assumptions and thus, no random network was developed. Second and in addition to the first assumption, the experimentally achieved expression data indicated that the predicted components are strongly related and with high probability belong to the same biological pathway. Hence, we provide a statistically validated biological network, which also was partly substantiated by experimental data and which indeed has to be finally confirmed by subsequent experiments. The network can help to plan further studies which aim to investigate the relationship of aging and autophagy in *P. anserina*.

Conclusions

In this study, we developed and investigated the putative protein-protein interaction network of autophagy in the aging model *P. anserina*. The study is based on the PPI networks of the reference organisms, human and yeast, and own experimental data. Using well-defined and established network analysis approaches, we demonstrate that the present PPI network is not random, but a "real-world" one. For expansion of the predicted network, we integrated the data of a yeast two-hybrid analysis. Additionally, we considered age-dependent gene expression data from a former transcriptome study to expand and combine the current knowledge of autophagy in *P. anserina* to one single unifying biological network.

We identified hub proteins that act as major players, e.g. PaATG8, PaATG9, and PaATG3, and seven functional modules, representing sub-pathways within autophagy. We confirmed that genes associated with autophagy are significantly more frequently up-regulated during aging than expected by chance. In general, autophagy-associated genes are much more co-expressed and, therefore, probably more co-regulated than randomly chosen gene groups.

In contrast to other model organisms, only little was known about PPIs of autophagy in *P. anserina*. Only the basal components of autophagy in *P. anserina* had been investigated until now. The present study aimed to collect, combine, and compile already existing data from *P. anserina* as well as from other organisms to achieve a new and general overview on autophagy in *P. anserina* which has not been available before. Moreover, since *P. anserina* is a well established aging model the provided network allows further investigations to gain a deeper knowledge on autophagy during aging of *P. anserina* as well as of other organisms.

Additional files

Additional file 1: Nucleotide sequence of the vector pLexA-DIR. The sequence is provided as text-based FASTA file. (FASTA 6 kb)

Additional file 2: Nucleotide sequence of the vector pGAD-HA. The sequence is provided as text-based FASTA file. (FASTA 8 kb)

Additional file 3: Yeast two-hybrid analysis. The results of the yeast two-hybrid analysis including the accession ID, the number of clones, the trivial names, and the putative function. For each prey clone of the non-singletons the position of the first amino acid is given in a separate sheet. In addition, the file contains each edge and node of the hybrid network, which was deduced from the interactions of the yeast two-hybrid analysis and the corresponding homologous proteins from yeast and human (see Additional file 10: Figure S1). Two edge types can be distinguished. First, edges indicating an interaction, where the source and the target node correspond to proteins in the same species. Second, edges indicating a homologous relation, where the source protein belongs to *P. anserina* and the target protein to yeast or human. UniProt accession numbers were used as identifiers for proteins of yeast and human. The identifiers for *P. anserina* proteins were adopted from the *P. anserina* genome database [10]. Species were distinguished by their NCBI taxonomy IDs: *Homo sapiens* (9606), *Podospora anserina* (5145) and *Saccharomyces cerevisiae S288c* (559292). (XLS 88 kb)

Additional file 4: *P. anserina* autophagy pathway from KEGG. List of all proteins and interactions for the autophagy pathway in *P. anserina* provided by the KEGG database. The identifiers for *P. anserina* proteins are either KEGG accession numbers (PODANSxxx) or were adopted from the *P. anserina* genome database (Pa_xx_xx). (XLS 37 kb)

Additional file 5: Results and statistic of topological and random network analysis. The file contains all topological values of each randomly generated network. In addition, the results of the comparison of the predicted network with the random networks are provided. (XLS 1896 kb)

Additional file 6: Lists of reference proteins and interactions. The file contains each protein and interaction from both reference species, which initially had been used for network prediction. UniProt accession numbers were used as identifiers for proteins of yeast and human. (XLS 133 kb)

Additional file 7: Parameter fitting. The results of the parameter fitting approach are given, including the lower and upper bounds of the E-values, their logarithms to the base 10, the differences between the logarithmic upper and lower bounds as well as the number of proteins and interactions in the reference KEGG network and in the predicted network. The parameters we have applied for the prediction approach in PATH2PPI are highlighted red. (XLS 43 kb)

Additional file 8: Hybrid PPI network. This file gives the underlying data of Fig. 2. In contrast to Additional file 1, this file contains all interactions, proteins and homologous relations of the autophagy hybrid PPI network, comprising both reference PPI networks, the predicted PPI network and the homology network which connects all three PPI networks. As for Additional file 1, two edge types can be distinguished. First, edges indicating an interaction, where the source and the target node correspond to proteins in the same species. Second, edges indicating a homologous relation where the source protein belongs to *P. anserina* and the target protein to yeast or human. UniProt accession numbers were used as identifiers for proteins of yeast and human. The identifiers for *P. anserina* proteins were adopted from the *P. anserina* genome database. Species were distinguished by their NCBI taxonomy IDs: *Homo sapiens* (9606), *Podospora anserina* (5145) and *Saccharomyces cerevisiae S288c* (559292). (XLS 105 kb)

Additional file 9: Autophagy PPI network of P. anserina. This files contains the tabular representation of Fig. 3 and provides all underlying data of the constructed PPI network of autophagy. The network consists of all interactions from the prediction approach and the yeast two-hybrid analysis. It includes the ID of the interaction, the source ID, the target ID, and the score. This file also contains further information about the corresponding reference interactions. For each predicted interaction, the reference interaction from yeast and / or human with the corresponding publication, author, and interaction type is listed. In addition, the file contains the expression data which are depicted as expression profiles in Fig. 3. UniProt accession numbers were used as identifiers for proteins of yeast and human. The identifiers for *P. anserina* proteins were adopted from the *P. anserina* genome database. Species were distinguished by their NCBI taxonomy IDs: *Homo sapiens* (9606), *Podospora anserina* (5145) and *Saccharomyces cerevisiae S288c* (559292). (XLS 271 kb)

Additional file 10: Additional figures. The figures concern the graphical representation of the results of the yeast two-hybrid analysis and node degree distributions of the random network and of the predicted network. (PDF 243 kb)

Additional file 11: Node degrees and betweennesses. The degrees and betweennesses for each node of the predicted autophagy PPI network, including the IDs, the alias names, the node degrees, the betweenesses, the rank of the degrees, the rank of the betweenesses, the rank of the sum and the corresponding module number. The identifiers for *P. anserina* proteins were adopted from the *P. anserina* genome database. (XLS 43 kb)

Additional file 12: Co-expressed genes. All co-expressed gene pairs deduced from the predicted PPI network and the former longitudinal transcriptome analysis, including the IDs of the genes and the Pearson correlation coefficient. The identifiers for *P. anserina* proteins were adopted from the *P. anserina* genome database. (XLS 48 kb)

Abbreviations
ATG: Autophagy-related genes; BLAST: Basic local alignment search tool; cDNA: Complementary deoxyribonucleic acid; GO: Gene ontology; Pcc: Pearson correlation coefficient; PPI: Protein-protein interaction; PUR: Proteins with up-regulated expression profiles of the corresponding genes; QC: Quality control

Acknowledgements
We thank the Cluster of Excellence 'Macromolecular Complexes' for support.

Funding
This work was partly supported by the Goethe University via the Frankfurt Autophagy Network (FAN) program and the Deutsche Forschungsgemeinschaft (OS75/17-1; SFB1177), the 'Bundesministerium für Bildung und Forschung' (BMBF) and the LOEWE program Ubiquitin Networks (Ub-Net) of the State of Hesse (Germany). These funding bodies were not directly involved in design of the study or collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' contributions
Conception/Design: OP, IK, HDO. Generated/Analyzed the data: OP. Wrote the manuscript: OP, AH, IK, HDO. Performed biological experiments: AH. Supervised the study: IK, HDO. All authors read and approved the final version of the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Molecular Bioinformatics, Institute of Computer Science, Faculty of Computer Science and Mathematics and Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe-University Frankfurt am Main, Robert-Mayer-Str. 11-15, 60325 Frankfurt am Main, Germany. [2]Molecular Developmental Biology, Institute of Molecular Biosciences, Faculty for Biosciences & Cluster of Excellence 'Macromolecular Complexes', Johann Wolfgang Goethe-University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.

Received: 3 November 2016 Accepted: 15 March 2017
Published online: 27 March 2017

References
1. Altman BJ, Jacobs SR, Mason EF, Michalek RD, MacIntyre AN, Coloff JL, Ilkayeva O, Jia W, He YW, Rathmell JC. Autophagy is essential to suppress cell stress and to allow BCR-Abl-mediated leukemogenesis. Oncogene. 2011;30(16):1855–67.
2. Alvers AL, Wood MS, Hu D, Kaywell AC, Dunn WA Jr, Aris JP. Autophagy is required for extension of yeast chronological life span by rapamycin. Autophagy. 2009;5(6):847–9.
3. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci USA. 2004;101(11):3747–752.
4. Bjedov I, Toivonen JM, Kerr F, Slack C, Jacobson J, Foley A, Partridge L. Mechanisms of life span extension by rapamycin in the fruit fly *Drosophila melanogaster*. Cell Metab. 2010;11(1):35–46.
5. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. Blast+: architecture and applications. BMC Bioinforma. 2009;10(1):421.
6. Cao Y, Klionsky DJ. Physiological functions of Atg6/Beclin 1: a unique autophagy-related protein. Cell Res. 2007;17(10):839–49.
7. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40(D1):D700–D5.
8. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys Rev E. 2004;70(6):066111.
9. Csardi G, Nepusz T. The Igraph software package for complex network research. InterJournal Compl Syst. 2006;1695(5):1–9.
10. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury JM, Ségurens B, Poulain J, et al. The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol. 2008;9(5):R77.
11. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, et al. String v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41(D1):808–15.
12. Gietz RD, Woods RA. Genetic transformation of yeast. Biotechniques. 2001;30(4):816–31.
13. Hansen M, Chandra A, Mitic LL, Onken B, Driscoll M, Kenyon C. A role for autophagy in the extension of lifespan by dietary restriction in *C. elegans*. PLoS Genet. 2008;4(2):e24.
14. He C, Klionsky DJ. Regulation mechanisms and signaling pathways of autophagy. Annu Rev Genet. 2009;43:67–93.
15. Johansen T, Lamark T. Selective autophagy mediated by autophagic adapter proteins. Autophagy. 2011;7(3):279–96.
16. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(D1):199–205.
17. Kikuma T, Kitamoto K. Analysis of autophagy in *Aspergillus oryzae* by disruption of Aoatg13, Aoatg4, and Aoatg15 genes. FEMS Microbiol Lett. 2011;316(1):61–9.
18. Knuppertz L, Osiewacz HD. Orchestrating the network of molecular pathways affecting aging: Role of nonselective autophagy and mitophagy. Mech Ageing Dev. 2016;153:30–40.
19. Knuppertz L, Hamann A, Pampaloni F, Stelzer E, Osiewacz HD. Identification of autophagy as a longevity-assurance mechanism in the aging model *Podospora anserina*. Autophagy. 2014;10(5):822–34.
20. Kraft C, Peter M, Hofmann K. Selective autophagy: ubiquitin-mediated recognition and beyond. Nat Cell Biol. 2010;12(9):836–41.
21. Kundu M, Thompson CB. Autophagy: basic principles and relevance to disease. Annu Rev Pathol. 2008;3:427–55.
22. Lilienbaum A. Relationship between the proteasomal system and autophagy. Int J Biochem. 2013;4(1):1–26.
23. Löw P. The role of ubiquitin–proteasome system in ageing. Gen Comp Endocrinol. 2011;172(1):39–43.
24. Meijer WH, van de Klei IJ, Veenhuis M, Kiel JA. ATG genes involved in non-selective autophagy are conserved from yeast to man, but the selective Cvt and pexophagy pathways also require organism-specific genes. Autophagy. 2007;3(2):106–16.
25. Metlagel Z, Otomo C, Takaesu G, Otomo T. Structural basis of ATG3 recognition by the autophagic ubiquitin-like protein ATG12. Proc Natl Acad Sci USA. 2013;110(47):18844–9.
26. Morselli E, Galluzzi L, Kepp O, Criollo A, Maiuri MC, Tavernarakis N, Madeo F, Kroemer G. Autophagy mediates pharmacological lifespan extension by spermidine and resveratrol. Aging. 2009;1(12):961–70.
27. Nassif M, Hetz C. Autophagy impairment: a crossroad between neurodegeneration and tauopathies. BMC Biol. 2012;10(1):78.
28. Newman M. Networks: an Introduction. Oxford: Oxford University Press; 2010.
29. Nolting N, Bernhards Y, Pöggeler S. SmATG7 is required for viability in the homothallic ascomycete *Sordaria macrospora*. Fungal Genet and Biol. 2009;46(8):531–42.
30. Osiewacz HD, Hamann A, Zintel S. Assessing organismal aging in the filamentous fungus *Podospora anserina*. Methods Mol Biol. 2013;965:439–62.

31. Papinski D, Schuschnig M, Reiter W, Wilhelm L, Barnes CA, Maiolica A, Hansmann I, Pfaffenwimmer T, Kijanska M, Stoffel I, et al. Early steps in autophagy depend on direct phosphorylation of Atg9 by the Atg1 kinase. Mol Cell. 2014;53(3):471–83.
32. Park C, Cuervo AM. Selective autophagy: talking with the UPS. Cell Biochem Biophys. 2013;67(1):3–13.
33. Philipp O, Osiewacz HD, Koch I. PATH2PPI: an R package to predict protein-protein interaction networks for a set of proteins. Bioinformatics. 2016;32(9):1427–9.
34. Philipp O, Hamann A, Servos J, Werner A, Koch I, Osiewacz HD. A genome-wide longitudinal transcriptome analysis of the aging model *Podospora anserina*. PLoS One. 2013;8(12):e83109.
35. Pons P, Latapy M. Computing communities in large networks using random walks. J Graph Algorithms Appl. 2006;10(2):191–218.
36. Pyo JO, Yoo SM, Ahn HH, Nah J, Hong SH, Kam TI, Jung S, Jung YK. Overexpression of Atg5 in mice activates autophagy and extends lifespan. Nat Commun. 2013;4:2300.
37. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinforma. 2008;9(1):405.
38. Shpilka T, Weidberg H, Pietrokovski S, Elazar Z. Atg8: an autophagy-related ubiquitin-like protein family. Genome Biol. 2011;12(7):226.
39. The UniProt Consortium. Activities at the universal protein resource (UniProt). Nucleic Acids Res. 2014;42(D1):191–8.
40. Voigt O, Herzog B, Jakobshagen A, Pöggeler S. Autophagic kinases SmVPS34 and SmVPS15 are required for viability in the filamentous ascomycete *Sordaria macrospora*. Microbiol Res. 2014;169(2):128–38.
41. Wiemer M, Grimm C, Osiewacz HD. Molecular control of fungal senescence and longevity. In: The Mycota I, Growth, differentiation and sexuality. 3rd edn. Switzerland: Springer; 2016. p. 155–81.
42. Yu J, Guo M, Needham C, Huang Y, Cai L, Westhead D. Simple sequence-based kernels do not predict protein-protein interactions. Bioinformatics. 2010;26(20):2610–614.
43. Zhang Z, Zhang J. A big world inside small-world networks. PLoS One. 2009;4(5):5686.
44. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev. 2007;21(9):1010–24.
45. Zou S, Chen Y, Liu Y, Segev N, Yu S, Liu Y, Min G, Ye M, Zeng Y, Zhu X, et al. Trs130 participates in autophagy through GTPases Ypt31/32 in *Saccharomyces cerevisiae*. Traffic. 2013;14(2):233–46.

Chapter 5

Discussion and conclusion

The topic of this work generally was an unbiased analysis of all genes which are age-dependently regulated. A bioinformatics pipeline was developed to reveal these age-related genes and the corresponding pathways. This initial, open-ended approach led to different findings that had already been presented and discussed in the first publication. The most promising conclusion of that study was the identification of the autophagy pathway to be of major importance for the ageing process. Based on this finding, the objectives of publications (**3**) and (**4**) were to analyse and characterise the autophagy pathway in *P. anserina* and its relationship with ageing. In comparison, the objective of publication (**2**) was to analyse the impact of ROS on the mitochondrial proteome, which required a new appropriate bioinformatics pipeline as well. Nevertheless, since no or only a minor increased amount of oxidised proteins could be recognised during ageing, it had been discussed that this probably was due to a complex quality assurance system active in the course of ageing. The autophagy pathway is one major process within this system.

Since the biological relevances and conclusions had already been discussed in each publication, this chapter focuses on the overall impact of the single publications for ageing research in *P. anserina* as well as in other organisms. Furthermore, it discusses the central theme of this work and how the single publications are related with each other.

5.1 A Genome-Wide Longitudinal Transcriptome Analysis of the aging Model *Podospora anserina*

The study provides a comprehensive age-dependent transcriptome analysis in a short-lived fungal organism. Such studies cannot be performed in other long-lived organisms or only with a lot of effort and difficulties. For example, such a study in mammals would take years of investigation. In addition, it would be difficult to

define constant environmental conditions or to specify which of the different tissue types should be taken into account. *P. anserina* provides the possibility to investigate age-dependent processes on the gene regulatory level within a few weeks. It possesses most of the basal molecular pathways and processes as other eukaryotic systems, enabling the transfer of information and knowledge to other organisms. The entire set of gene expression data is available either as raw or as statistically processed data analysed by means of a new bioinformatics pipeline. It can serve as a comprehensive reference for further age-dependent studies either in *P. anserina* or other organisms which are not accessible for such types of studies.

Considering the advantage of transferring knowledge amongst organisms due to the similar basal processes, another major conclusion could be drawn from the study. Three of these basal processes which are strongly conserved amongst organisms are the protein biogenesis via the ribosomes, the UPS system and the autophagy pathway. These three processes arose to be of particular interest during the study. This was firstly due to their general impact on the organism and secondly genes associated with these pathways showed significant age-dependent expression profiles. Genes associated with the ribosomes or the UPS system, exhibit continuously decreasing expression profiles. In contrast, genes associated with autophagy were significantly up-regulated during ageing. Furthermore, former studies suggested and partly showed that these three processes are known to be regulated by the TOR kinase in an antagonistic and opposite manner (Pandey et al., 2007; Reiter et al., 2011). The results of the study supported these assumptions on the gene regulatory level.

Most importantly, during the recent years it was found that autophagy is not only a cellular recycling pathway but a quality assurance process as well (reviewed in: Tatsuta and Langer (2008); Fischer et al. (2012); Knuppertz and Osiewacz (2016)). It is probably effective during ageing or is involved in ageing processes, respectively (Weber and Reichert, 2010; Nitsche et al., 2013). Moreover, autophagy seemed to play an essential role for longevity assurance purposes. Indeed, this aspect of autophagy was proved in a subsequent study on *P. anserina* (Knuppertz et al., 2014).

Based on the latter study and the transcriptome analysis, the following questions arose:

- How is autophagy associated with ageing?
- Which are the single components of autophagy?
- How are these components related to each other in *P. anserina*?

This study prepared and paved the way for the following studies, investigations and papers comprised by this PhD thesis.

Since its publication in December 2013, the article achieved about 3000 views, over 450 downloads and 15 citations . The corresponding entire raw as well as the processed transcriptome data set deposited at the European Bioinformatics Institutes ArrayExpress (accession number E-MTAB-2016) has been viewed about 650 times substantiating the impact of the study and its significance for other investigations on ageing research.

5.2 Global protein oxidation profiling suggests efficient mitochondrial proteome homeostasis during aging

The study overcame the experimental and technical difficulties of qualitative as well as quantitative *in vivo* measuring of oxidised proteins. The impact of ROS on the mitochondrial proteome during ageing was investigated by means of a novel experimental and statistical workflow which comprises a newly developed bioinformatics pipeline.

Since the iTRAQ-based proteome analysis led to a large amount of modified as well as unmodified proteins at once, it also produced a large but also a noisy set of data which had to be statistically evaluated. For the bioinformatics part of the study it was necessary to take different characteristics of the data into account, e.g. the partial incompleteness and ambiguity of the single protein data which was due to biological noise and experimental artifacts. The initial data set consisted of over 220,000 peptide sequences already annotated to the corresponding 2341 distinct proteins. Each detected peptide exhibited a particular amino acid modification and an abundance ratio of one of the three days in contrast to the first day of measurement. First of all, this data had been deposited in a MySQL database and the statistical pipeline was implemented in a web-based analysis tool. The pipeline had to compile each of these single data points to achieve final protein ratios for each possible modification on each of the three different time points. At this stage of the pipeline it was important to develop and implement a new measurement, the quality score, which indicates the reliability of these single, compiled data points. That quality score considered that some of the data was incomplete and exhibited variances as well as ambiguity in the direction of regulation.

In addition, the bioinformatics pipeline consists of particular approaches which enabled the finding of significant protein ratios and the comparison of unmodified with modified protein abundances. The latter approach aimed to reveal proteins which may be particularly affected and impaired by ROS during ageing.

Summarizing, the study presented an innovative experimental workflow on the one hand. On the other hand, it provided a new developed bioinformatics pipeline and statistical approaches which can be adapted to other similar iTRAQ-based

studies. Here, this workflow, which was applied to the age-dependent proteomes, revealed that ROS either do not impair proteins as expected or that rather different quality assurance mechanisms are effective, which can compensate ROS-damages even in later age stages, in *P. anserina*. Autophagy may be one of these mechanisms, which would be in concordance with the findings of former studies. For example, it was found that autophagy-associated gene-expression and the number of autophagosomes increase during ageing. Hence, it is conceivable that indeed ROS may lead to an increased amount of impaired proteins but that an increased autophagy or mitophagy rate, may compensate that effect by degradation of these damaged protein species or mitochondria, respectively.

5.3 Path2PPI: an R package to predict protein-protein interaction networks for a set of proteins

The R package PATH2PPI had been developed to overcome the drawbacks of existent PPI data repositories and prediction solutions (see sections 3.2.3 and 4.3). For example, the former transcriptome study revealed different pathways which seem to be affected or even may play essential roles in the ageing process of *P. anserina*. It was of particular interest to get detailed information about the involved proteins of these pathways but current PPI repositories did not contain sufficient data for *P. anserina*. PATH2PPI is a software tool, which enables the prediction and construction of PPI networks in organisms, for which only little is known about particular pathways and biological processes. It transfers information available about the pathways and the corresponding interactions from one or more of the seven most established model organisms.

In the tutorial of the package (see appendix) the initial step of the autophagy pathway was predicted, which is based on autophagy induction in yeast and humans. The small predicted network was compared with the network provided by the KEGG database (Kanehisa et al., 2014), which generally comprises manually curated interactions. PATH2PPI is able to transfer interaction data amongst organisms; all of these known interactions of autophagy induction in *P. anserina* had been predicted correctly. Furthermore, PATH2PPI contributed to the Bioconductor project (Huber et al., 2015), which is a well-established platform for various bioinformatics tools and algorithms and represents one of the major reasons why the programming language R has been widely adopted for bioinformatics studies. Since its participation in the Bioconductor project on August 2015, PATH2PPI has been downloaded over 2000 times.

Summarising, the contribution to the Bioconductor project and the number of downloads demonstrate that PATH2PPI is an advantageous and reliable tool and

provides a useful and new approach to support studies which aim to analyse PPI data. By using of PATH2PPI, a study was performed to reveal the putative interaction network of autophagy in *P. anserina*, which has not been possible before without great effort and/or many expensive biological experiments.

Since PATH2PPI works species-independent and only relies on protein sequences, it may not only help to predict further pathways and the corresponding interactions in *P. anserina* but also for many other fully sequenced organisms, which lack reliable PPI data repositories.

5.4 The autophagy interaction network of the ageing model *Podospora anserina*

In the former transcriptome study it was found that autophagy may play an important role during ageing in *P. anserina*. A subsequent study proved the finding and confirmed that not only the expression of autophagy-associated genes increases during ageing but the actual rate of autophagy as well (Knuppertz et al., 2014). To investigate how autophagy is associated with ageing and which components are involved in autophagy, this study aimed to reveal the associated proteins and their interactions.

Basically, the objective of the study was to construct and analyse the interaction network of autophagy in *P. anserina*. It is the first comprehensive PPI study which is based on PATH2PPI. Therefore, the different topological and statistical analyses which had been applied on the predicted network, did not only confirm that it is biological meaningful but also proved the reliability of PATH2PPI and the corresponding prediction approach. For example, it was shown that the network's diameter, modularity and transitivity are significantly higher than expected in a random network. Furthermore, the betweenness and the node degree distributions noticeably differ from the distributions of random networks.

Nevertheless, it is important to carefully gather and prepare the initial original data, since the underlying reference data are as important as the applied prediction method. If only a few or even compromised reference data are available, the method will produce insufficient results. Therefore, the initial autophagy PPI network predicted by PATH2PPI relies on the corresponding PPI networks of yeast and humans, which are based on known and reviewed proteins and their interactions associated with autophagy. Both reference species were chosen deliberately. Yeast is the best established and studied eukaryotic model organism and provides a broad scientific community and many data repositories. The human autophagy system was concerned due to its obvious relevance for understanding the ageing processes in humans. The possibility of combining both reference networks enables the extension of the underlying data, the increase of the resulting information and the

retrieval of putative evolutionary conserved interactions. For example, PATH2PPI suggests nine interactions to be active in the autophagy system of yeast, human and *P. anserina*.

To improve the quality of the predicted PPI network, additional experimental data was subsequently included. Firstly, the PPI data achieved from a yeast two-hybrid analysis were integrated and secondly, the gene expression data from the former transcriptome study was mapped to the final PPI network. This led to a comprehensive age-dependent biological network of autophagy in *P. anserina*. In the study it was discussed that autophagy and ageing are related on the gene expression as well as on the protein activity and interaction level. Hence, since the constructed network combines different biological levels, it provides a comprehensive view on autophagy, produced valuable results and enables the drawing of different conclusions about its relationship with ageing. The most important of these results and conclusions were, firstly, the detected modules, which correspond to the different sub-pathways of autophagy. Secondly, ten of the major proteins or key players in the network were introduced. It will be of interest to analyse how a manipulation (e.g. deletion or enhancement) of these hub proteins or the corresponding genes will impact the system. Thirdly, it was confirmed that autophagy-associated genes are significantly often up-regulated and moreover, much more frequently co-expressed during ageing than expected by chance. The latter findings suggest a strong relationship of ageing, autophagy and gene expression. Additional results and conclusions had been discussed or were drawn in the study allowing to establish different presumptions and the conceiving of further studies, which aim to analyse the autophagy system and its relationship with ageing in *P. anserina* as well as in other organisms (see chapter 6).

Chapter 6

Outlook

The proposed network of autophagy in *P. anserina* provides an excellent basis for further studies which aim to investigate how autophagy is related to ageing in *P. anserina*. It is most appropriate and valuable for subsequent network modelling approaches. Nevertheless, even at this initial or early phase of network modelling, questions arose which could be investigated by specific biological experiments to achieve new insights into autophagy.

6.1 Potential biological experiments and validation of the authophagy network

- The reliability of the ten most important proteins in the PPI network was substantiated by different statistical approaches. In addition, the effects on the corresponding knockout yeast strains were taken into account. The latter showed that each of these genes or proteins, respectively, is either essential or a loss led to the absence of autophagy in yeast. Nevertheless, similar knockout experiments in *P. anserina* will confirm the reliability of the proposed PPI network and the detected hub proteins. Moreover, as each of the hub proteins are assigned to specific modules which correspond to the different sub-pathways of autophagy, it will be of interest whether a particular knockout leads to a complete absence of autophagy or only to a depletion of sub-pathways which are down-stream to the deleted gene's sub-pathway. Such an analysis would substantiate the modularity approach during network modelling. Using reporters for measuring selected autophagy pathways (e.g. PaSOD3mut::GFP for mitophagy, PaSOD1::GFP for general autophagy) allows the dissection of the affected pathways.

- Ten genes, which are associated with the autophagy pathway, were identified to be up-regulated during ageing. Since this finding is based on a high-throughput approach, as described in the transcriptome study (Philipp et al., 2013), the

up-regulation should be validated by additional, specific experiments, e.g. by a qRT-PCR analysis. During the study, such a validation had already been done for some selected genes to generally substantiate the reliability of the transcriptome analysis.

- Only one of the genes which code for the ten most important proteins, appeared to be up-regulated during ageing (*PaTrs130*). This lead to the question whether autophagy components are limiting during ageing. If so, up-regulation of the most important proteins should increase autophagy which can be measured using appropriate reporter proteins. If autophagy up-regulation is possible, the physiological consequences should be investigated, i.e. what is the impact on lifespan and stress resistance. For instance, the hub protein PaATG12 was computationally assigned to the module which seemed to be associated with autophagosome formation. Since the level of gene expression does not necessarily correlate with the protein amount, a potential experiment will reveal whether an over-expression of *PaAtg12* leads to an increased amount of PaATG12 as well, and more importantly to an increased amount of autophagosomes or pre-autophagosomal structures. This may partly increase the performance and efficiency of the degradation process during autophagy.

- Nine of the significantly up-regulated genes do not correspond to the most important proteins in the network. Since the transcriptome study revealed that, in general, there are considerably more down-regulated than up-regulated genes during ageing, the finding of a significant number of up-regulated genes, indicate the importance of these genes for autophagy as well. To investigate the impact of these genes on autophagy, the corresponding deletion strains should be investigated. Especially their contribution to different forms of autophagy (e.g selective vs. general autophagy) needs to be analysed. Such analyses are only feasible if the particular gene is not essential, i.e. a loss is not lethal. Table 6.1 lists these ten up-regulated genes and the most important effects on the yeast strain where the corresponding homologous gene was knocked out. For three of the genes no homologue in yeast was found and one gene is essential in yeast. Hence, the remaining six genes represent a good starting point for such analyses. Remarkably, a deletion of *Rpd3* led to increased autophagy in the corresponding knockout yeast strain. A deletion of the *P. anserina* homologue *PaRpd3a* may probably led to an increase of autophagy in *P. anserina* as well. Vice versa, this would indicate that the observed, age-dependent increase of *PaRpd3a* may even counteract to the rate of autophagy.

- It was found that PaNBR1 interacts with PaATG8 and the corresponding gene *PaNbr1* is up-regulated during ageing. Nevertheless, this does not necessarily mean that the protein amount or its activity, respectively, increases during ageing as well. In Kraft et al. (2010) it was described that NBR1 is involved in the

Table 6.1: Up-regulated autophagy genes compared with yeast homologues. For ten proteins of the PPI autophagy network, the corresponding transcripts were up-regulated during aging (Philipp et al., 2017). The higher the Pearson correlation coefficient the straighter increases the related expression profile or the number of gene products, respectively. The ratio gives the difference of transcript abundance from the first to the last day of measurement. In addition, the last column compiles the most important effects on the yeast strain where the corresponding homologous gene was knocked out (Cherry et al., 2012). Abbreviations: ne=non-essential, aa=autophagy absent, dv=decreased viability, dvg=decreased vegetative growth, e=essential, ai=autophagy increased. NA (not applicable) indicates that no homologue exists.

Podospara anserina gene	Pearson cor. coef.	Ratio	Yeast homologue	Effects in knockout yeast strains
PaAtg13 (Pa_0_1500)	0.87	1.74	*Atg13* (YPR185W)	ne, aa, dv
PaMan1 (Pa_1_10350)	0.87	2.10	*Ams1* (YGL156W)	ne, dvg
PaTrs130 (Pa_5_4470)	0.87	3.94	*Trs130* (YMR218C)	e
PaIrs4 (Pa_7_7880)	0.86	2.79	*Irs4* (YKR019C)	ne, dvg
PaAtg6 (Pa_1_11020)	0.84	3.48	*Vps30* (YPL120W)	ne, aa, dv
PaRpd3a (Pa_1_5330)	0.81	1.65	*Rpd3* (YNL330C)	ne, ai
Pa_4_2219	0.79	4.96	NA	NA
PaNbr1 (Pa_2_12420)	0.78	3.13	NA	NA
PaVps15 (Pa_6_1830)	0.77	2.40	*Vps15* (YBR097W)	ne, aa
Pa_2_2450	0.72	3.38	NA	NA

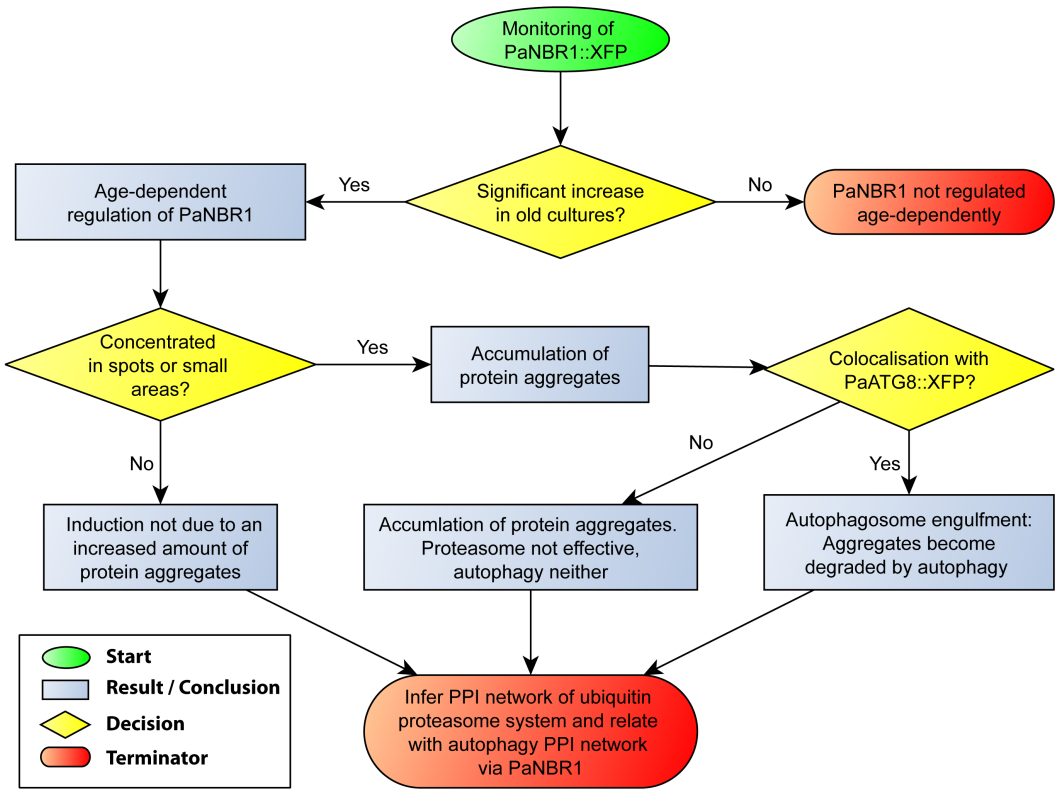


Figure 6.1: Flowchart of possible PaNBR1 monitoring experiments. Labelling of PaNBR1 with a fluorescing protein (e.g. GFP, here indicated by ::XFP) or another protein tag (e.g. His/FLAG tag) allows the monitoring of PaNBR1 in old and young *P. anserina* strains. An increase of PaNBR1 in older cultures will indicate that not only transcript but also protein level is up-regulated during ageing. Fluorescence microscopy of a fluorescent PaNBR1 variant (e.g. PaNBR1::GFP) additionally will allow the evaluation of the localisation of the protein. An accumulation of larger protein aggregates is indicated, if the majority of PaNBR1 is localised or concentrated in spots or small areas within the hyphae. A colocalisation of these spots with PaATG8 substantiates the assumption, that larger protein aggregates accumulate during ageing and became degraded by autophagy rather than by the ubiquitin-proteasome system. Subsequently, an inferred PPI network of the ubiquitin-proteasome system could be combined with the autophagy PPI network via PaNBR1 to reveal how both QC systems are related with each other.

degradation of protein aggregates via autophagosome formation if the aggregates are too large for degradation by the 26S proteasome. The age-dependent transcriptome study revealed that the expression of genes coding for the proteasome decreases during ageing (Philipp et al., 2013). Hence, a verification that the number of PaNBR1 proteins increases during ageing can indicate that degradation via the ubiquitin-proteasome system is impaired and/or that the accumulation of large protein aggregates increases during ageing as well. The amount of PaNBR1 labelled with an appropriate tag (e.g. His/FLAG) can be measured in total protein extracts by western blot analyses. Moreover, labelling with GFP or other fluorescence proteins will allow the localisation of PaNBR1 in the hyphae during ageing. In addition, parallel monitoring of

colocalisation of PaNBR1 with a labelled PaATG8 protein, which is important for initiation of autophagosome formation (see Knuppertz et al. (2014)) may reveal a further important role of PaNBR1 during ageing, as depicted in the flowchart in figure 6.1. Depending on the result of such an analysis a subsequent study could infer the PPI network of the ubiquitin-proteasome system and combine it with the proposed PPI network of autophagy, either via a linkage at PaNBR1 or at another putatively conjugating protein (see next section).

6.2 Bioinformatics analysis and network expansion

- During this work an age-dependent relation of the ubiquitin-proteasome system with the autophagy pathway was shown. The transcriptome study revealed that the expression of a significant number of proteasomal genes decreases while autophagy associated gene expression increases during ageing. These data suggested that autophagy may compensate proteasomal impairments and vice versa. Such a compensatory relationship was also proposed by former studies (Lilienbaum, 2013). Hence, a subsequent work can infer the PPI network representing the assembly and degradation process of the ubiquitin-proteasome system and combine it with the PPI network of the autophagy pathway. As formerly described evidence suggest that PaNBR1 as well as PaATG8 can probably be two major linking proteins of both degradation pathways.

- The module detection approach assigned the predicted proteins to seven disjunct modules and it was shown that these modules partly correspond to the different sub-pathways of autophagy. For instance, the proteins PaATG1, PaATG13, PaATG11, PaVAC8, PaATG17 and others are known to be involved in the induction process. These proteins were assigned to the same module. Since the modules were only superficially related to the different pathways, a subsequent study may build on this initial approach, and could aim to verify the modules and the assignment to pathways in a more elaborative manner. This could be accomplished by gathering additional knowledge from other species and studies, where the autophagy proteins already had been associated with a particular sub-pathway. A subsequent statistical analysis, will then investigate whether the assignment is reliable. Such a statistical analysis can test whether the number of assigned proteins to a module is due to a random distribution or whether this number exceeds the expected number of a random approach indicating a biological meaningful assignment.

- The study on the autophagy network of *P. anserina* revealed that there are a significant number of co-expressed gene pairs indicating a strong regulatory relationship of autophagy-associated genes during ageing. Nevertheless, it was also found that the co-expression network which consists of these co-expressed gene pairs only weakly covers the predicted interactions of the proposed PPI network. This indicated that transcriptional regulation strongly differs from translational regulation. However, a challenging but promising task will be the compilation and conjugation of both networks. While the PPI network primarily consists of qualitative data, the co-expression network also consists of qualitative data but is accompanied by quantitative transcriptome data as well. Such a compilation step can be achieved by quantitative network modelling approaches as described in the next section.

6.3 Mathematical and quantitative modelling

Most importantly, the proposed autophagy network allows further network-modelling and the implementation of different pathway analysis approaches. It consists of a sufficient amount of reliable qualitative data, enabling the quantitative modelling of the autophagy pathway in *P. anserina*. Due to the deficient data at the beginning of this work such analyses had not been possible before. Basically, quantitative mathematical models allow the researcher to make predictions about systems behaviour, comprising stoichiometry and stochastic reactions within a biological pathway. For example, among the existing methods, which allow for modelling of qualitative as well as quantitative data, the Petri Net (PN) formalism is very suitable (Koch et al., 2010). It provides sound analysis methods for systems with concurring processes, which can occur at the same time (synchronised) or not (non-synchronised). In particular, biochemical systems exhibit such a behavior of concurrent processes, usually containing many alternative pathways like autophagy which consist of different sub processes.

Based on the findings of the predicted biological network of autophagy a subsequent PN modelling approach may reveal further regulatory properties of this pathway. As an example for such a possible study, the PN model of the autophagy induction during starvation is depicted in figure 6.2 which has been developed during this work for further subsequent investigations.

While the schematic view is only inadequately able to visualise the progressive process of a signalling cascade, the PN, deduced from data of the PPI prediction approach, enables the differentiated and gradually description of the processes which occur during induction. The qualitative model can now be extended by further processes such as the nucleation and the expansion of the autophagosome, the subsequent fusion with the vacuole and finally the degradation of the autophagosomal content. Furthermore, as described above, in a later modelling step, quantitative

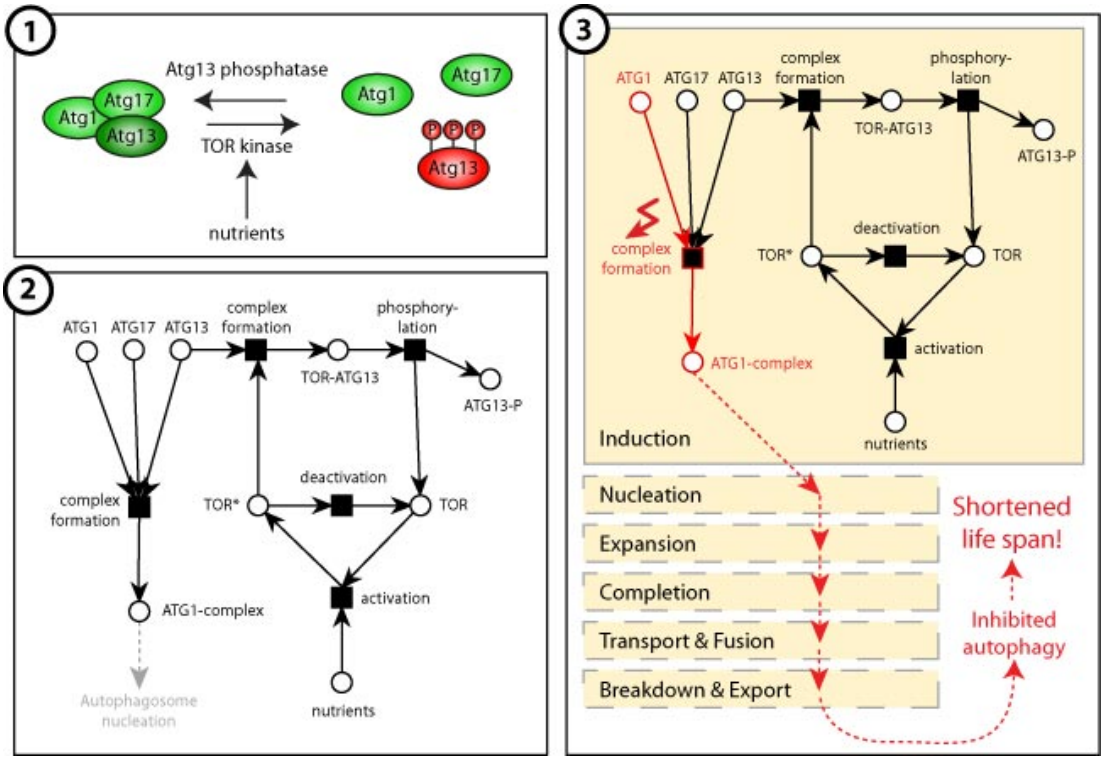


Figure 6.2: Example for a PN model of autophagy induction during starvation. (1) Schematic view of the induction of autophagy and the (2) PN model which provides more detailed information. Nutrient availability activates the TOR kinase, which phosphorylates ATG13. Hyper-phosphorylated ATG13 cannot interact with ATG1 and ATG17 to build the ATG1 complex, which is important for autophagosome nucleation. Hence, autophagy is inhibited. Under nutrient-depleted conditions, TOR is deactivated, ATG13 is dephosphorylated and available for complex formation with ATG1 and ATG17. The next step of the autophagy process - autophagosome nucleation can begin. (3) An example for a simulation scenario in a PN model. If *Atg1* is deleted, the transition (reaction) complex formation and all subsequent autophagy processes are inactivated, as well, i.e., autophagy is inhibited which leads to a shortened lifespan (see Knuppertz et al. (2014)).

properties can be added if experimental data is available. For example, a great advantage of the transcriptome analysis is the possibility to start the network model on a gene regulatory level, since not only qualitative but quantitative transcript abundances are available. Overall, a major advantage of the PN formalism is that the level of abstraction can be iteratively adapted to the current problem and the coverage of available data. Hence, a progressive switch from a primarily qualitative model to a subsequent quantitative validation is possible.

Summarising, a subsequent implementation of a PN model could generate new testable predictions, e.g. for a "cross-talk" of different QC pathways during ageing, to identify mechanisms and components which can influence the autophagy process to extend the healthy lifespan.

Chapter 7

Summary

The present PhD thesis about the bioinformatics analysis of age-dependent pathways in *P. anserina* led to the following conclusions and results:

1. The transcriptome analysis revealed various genes and the corresponding pathways or processes, which are down-regulated in the course of ageing. The two most significant of these sets of genes are those, which are involved in protein biogenesis via the ribosomes and the genes, which are associated with protein degradation via the proteasome system.

2. In contrast, the most significant and promising up-regulated process on the genetic level appears to be the autophagy pathway.

3. The transcriptome data suggest an underlying mechanism, which is effective during ageing and forces the aforementioned pathways to behave in a counteracting or antagonistic manner.

4. A new bioinformatics and statistical analysis pipeline was developed to analyse iTRAQ data and the effect of oxidisation and modification of proteins during ageing. Since only minor oxidative changes had been identified it was suggested that a potential quality assurance mechanism is effective during ageing in *P. anserina*, e.g. the autophagy pathway.

5. A new method was developed and implemented which enables a reliable prediction of PPI networks for pathways and organisms, for which no or only a little data were available.

6. A comprehensive biological network of the autophagy pathway in *P. anserina* was developed, comprising predicted as well as experimental data.

7. The network was tested for mathematical, topological and biological properties substantiating its reliability and that the results did not arise due to random effects. Different sub-pathways were discovered and the ten most important key proteins in the autophagy pathway were identified.

8. The network and several data sets generated during this work provide a valuable basis for further pathway modelling approaches.

Generally, the achieved data during this work led to the conclusion that the autophagy pathway is of major importance for understanding the ageing process. Evidence suggests that gene expression and regulation is as important as analyses, which aim to describe the biochemical process of autophagy based on the involved proteins and their interactions. Moreover, autophagy appears to be a promising target for intervention into ageing to extend the healthy lifespan of *P. anserina* as well as of other organisms like humans.

Chapter 8

Bibliography

Allison, D.B., Cui, X., Page, G.P., Sabripour, M.: Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics **7**(1), 55–65 (2006)

Alvers, A.L., Wood, M.S., Hu, D., Kaywell, A.C., Dunn, W.A. Jr, Aris, J.P.: Autophagy is required for extension of yeast chronological life span by rapamycin. Autophagy **5**(6), 847–849 (2009)

Audic, S., Claverie, J.-M.: The significance of digital gene expression profiles. Genome Research **7**(10), 986–995 (1997)

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

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'donovan, C., Apweiler, R.: QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics **25**(22), 3045–3046 (2009)

Bjedov, I., Toivonen, J.M., Kerr, F., Slack, C., Jacobson, J., Foley, A., Partridge, L.: Mechanisms of life span extension by rapamycin in the fruit fly *Drosophila melanogaster*. Cell Metabolism **11**(1), 35–46 (2010)

Bowen, R.L., Atwood, C.S.: Living and dying for sex. A theory of aging based on the modulation of cell cycle signaling by reproductive hormones. Gerontology **50**(5), 265–290 (2004)

Braun, P., Gingras, A.-C.: History of protein–protein interactions: From egg-white to complex networks. Proteomics **12**(10), 1478–1498 (2012)

Brust, D., Daum, B., Breunig, C., Hamann, A., Kühlbrandt, W., Osiewacz, H.D.: Cyclophilin D links programmed cell death and organismal aging in *Podospora anserina*. Aging Cell **9**(5), 761–775 (2010)

Chatr-Aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., *et al.*: The BioGRID interaction database: 2015 update. Nucleic Acids Research **43**(D1), 470–478 (2015)

Chen, C.-C., Lin, C.-Y., Lo, Y.-S., Yang, J.-M.: PPISearch: a web server for searching homologous protein–protein interactions across multiple species. Nucleic Acids Research **37**(2), 369–375 (2009)

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., *et al.*: *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Research **40**(D1), 700–705 (2012)

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al.: A survey of best practices for RNA-seq data analysis. Genome Biology **17**(13) (2016)

Daum, B., Walter, A., Horst, A., Osiewacz, H.D., Kühlbrandt, W.: Age-dependent dissociation of atp synthase dimers and loss of inner-membrane cristae in mitochondria. Proceedings of the National Academy of Sciences **110**(38), 15301–15306 (2013)

Deng, Y.*et al.*: ppiPre: predicting protein-protein interactions by combining heterogeneous features. BMC Systems Biology **7**(Suppl 2), 8 (2013)

Dillin, A., Gottschling, D.E., Nyström, T.: The good and the bad of being connected: the integrons of aging. Current Opinion in Cell Biology **26**, 107–112 (2014)

Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B.M., Couloux, A., *et al.*: The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biology **9**(5), 77–17722 (2008)

Figge, M.T., Reichert, A.S., Meyer-Hermann, M., Osiewacz, H.D.: Deceleration of fusion–fission cycles improves mitochondrial quality control during aging. PLoS Computational Biology **8**(6), 1002576 (2012)

Fischer, F., Hamann, A., Osiewacz, H.D.: Mitochondrial quality control: an integrated network of pathways. Trends in Biochemical Sciences **37**(7), 284–292 (2012)

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L.: STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research **41**(D1), 808–815 (2013)

Gredilla, R., Grief, J., Osiewacz, H.D.: Mitochondrial free radical generation and lifespan control in the fungal aging model *Podospora anserina*. Experimental Gerontology **41**(4), 439–447 (2006)

Groebe, K., Krause, F., Kunstmann, B., Unterluggauer, H., Reifschneider, N.H., Scheckhuber, C.Q., Sastri, C., Stegmann, W., Wozny, W., Schwall, G.P., *et al.*: Differential proteomic profiling of mitochondria from podospora anserina, rat and human reveals distinct patterns of age-related oxidative changes. Experimental Gerontology **42**(9), 887–898 (2007)

Guevara, C.R., Philipp, O., Hamann, A., Werner, A., Osiewacz, H.D., Rexroth, S., Rögner, M., Poetsch, A.: Global protein oxidation profiling suggests efficient mitochondrial proteome homeostasis during aging. Molecular & Cellular Proteomics **15**(5), 1692–1709 (2016)

Guindalini, C., Pellegrino, R.: Gene expression studies using microarrays. In: Rodent Model as Tools in Ethical Biomedical Research, pp. 203–216. Springer, (2016)

Hamann, A., Brust, D., Osiewacz, H.D.: Apoptosis pathways in fungal growth, development and ageing. Trends in Microbiology **16**(6), 276–283 (2008)

Harman, D.: Aging: a theory based on free radical and radiation chemistry. The Journal of Gerontology **1**, 298–300 (1956)

He, C., Klionsky, D.J.: Regulation mechanisms and signaling pathways of autophagy. Annual Review of Genetics **43**, 67–93 (2009)

Huber, W., Carey, J., V., Gentleman, R., Anders, S., Carlson, M., Carvalho, S., B., Bravo, *et al.*: Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods **12**(2), 115–121 (2015)

Jin, K.: Modern biological theories of aging. Aging and Disease **1**(2), 72 (2010)

Jones, S., Thornton, J.M.: Principles of protein-protein interactions. Proceedings of the National Academy of Sciences **93**(1), 13–20 (1996)

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M.: Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research **42**(D1), 199–205 (2014)

Kirkwood, T.B.: Understanding the odd science of aging. Cell **120**(4), 437–447 (2005)

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

Knuppertz, L., Osiewacz, H.D.: Orchestrating the network of molecular pathways affecting aging: role of nonselective autophagy and mitophagy. Mechanisms of Ageing and Development **153**, 30–40 (2016)

Knuppertz, L., Hamann, A., Pampaloni, F., Stelzer, E., Osiewacz, H.D.: Identification of autophagy as a longevity-assurance mechanism in the aging model *Podospora anserina*. Autophagy **10**(5), 822–834 (2014)

Koch, I., Reisig, W., Schreiber, F.: Modeling in Systems Biology: the Petri Net Approach vol. 16. Springer Science & Business Media, (2010)

Korolchuk, V.I., Menzies, F.M., Rubinsztein, D.C.: Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. FEBS Letters **584**(7), 1393–1398 (2010)

Kraft, C., Peter, M., Hofmann, K.: Selective autophagy: ubiquitin-mediated recognition and beyond. Nature Cell Biology **12**(9), 836–841 (2010)

Krauss, G.: Biochemistry of Signal Transduction and Regulation. John Wiley & Sons, (2006)

Kumar, L., Futschik, M.E.: Mfuzz: a software package for soft clustering of microarray data. Bioinformation **2**(1), 5–7 (2007)

Kundu, M., Thompson, C.B.: Autophagy: basic principles and relevance to disease. Annual Review of Pathology **3**, 427–455 (2008)

Kunstmann, B., Osiewacz, H.D.: Over-expression of an S-adenosylmethionine-dependent methyltransferase leads to an extended lifespan of *Podospora anserina* without impairments in vital functions. Aging Cell **7**(5), 651–662 (2008)

Lilienbaum, A.: Relationship between the proteasomal system and autophagy. International Journal of Biochemistry and Molecular Biology **4**(1), 1–26 (2013)

Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., Young, R.A.: Revisiting global gene expression analysis. Cell **151**(3), 476–482 (2012)

Löw, P.: The role of ubiquitin–proteasome system in ageing. General and Comparative Endocrinology **172**(1), 39–43 (2011)

Martin, G.M., Austad, S.N., Johnson, T.E.: Genetic analysis of ageing: role of oxidative damage and environmental stresses. Nature Genetics **13**(1), 25–34 (1996)

Masoro, E.J.: Caloric Restriction: a Key to Understanding and Modulating Aging vol. 1. Elsevier, (2002)

Mather, K.A., Jorm, A.F., Parslow, R.A., Christensen, H.: Is telomere length a biomarker of aging? A review. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **66**(2), 202–213 (2011)

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Krüger, D.H., Terauchi, R.: Gene expression analysis of plant host–pathogen interactions by SuperSAGE. Proceedings of the National Academy of Sciences **100**(26), 15718–15723 (2003)

Molina, C., Rotter, B., Horres, R., Udupa, S.M., Besser, B., Bellarmino, L., Baum, M., Matsumura, H., Terauchi, R., Kahl, G., et al.: SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics **9**(553) (2008)

Murakami, Y., Mizuguchi, K.: Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators. BMC Bioinformatics **15**(213) (2014)

Nassif, M., Hetz, C.: Autophagy impairment: a crossroad between neurodegeneration and tauopathies. BMC Biology **10**(78) (2012)

Nitsche, B.M., Burggraaf-van Welzen, A.-M., Lamers, G., Meyer, V., Ram, A.F.: Autophagy promotes survival in aging submerged cultures of the filamentous fungus *Aspergillus niger*. Applied Microbiology and Biotechnology **97**(18), 8205–8218 (2013)

Osiewacz, H., Hamann, A.: DNA reorganization and biological aging. A review. Biochemistry **62**(11), 1275–1284 (1997)

Osiewacz, H.D.: Genes, mitochondria and aging in filamentous fungi. Ageing Research Reviews **1**(3), 425–442 (2002a)

Osiewacz, H.D.: Mitochondrial functions and aging. Gene **286**(1), 65–71 (2002b)

Osiewacz, H.D., Borghouts, C.: Mitochondrial oxidative stress and aging in the filamentous fungus *Podospora anserina*. Annals of the New York Academy of Sciences **908**(1), 31–39 (2000)

Pandey, U.B., Nie, Z., Batlevi, Y., McCray, B.A., Ritson, G.P., Nedelsky, N.B., Schwartz, S.L., DiProspero, N.A., Knight, M.A., Schuldiner, O., *et al.*: HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. Nature **447**(7146), 860–864 (2007)

Park, C., Cuervo, A.M.: Selective autophagy: talking with the UPS. Cell Biochemistry and Biophysics **67**(1), 3–13 (2013)

Partridge, L., Gems, D.: Mechanisms of aging: public or private? Nature Reviews Genetics **3**(3), 165–175 (2002)

Philipp, O.: Analyse des altersabhängigen Transkriptoms im Ascomyceten *Podospora anserina* unter besonderer Berücksichtigung von Genen der Atmungskette. Diploma thesis, Goethe Universität Frankfurt am Main, Frankfurt am Main, Germany (2012)

Philipp, O., Osiewacz, H.D., Koch, I.: Path2PPI: an R package to predict protein–protein interaction networks for a set of proteins. Bioinformatics **32**(9), 1427–1429 (2016)

Philipp, O., Hamann, A., Servos, J., Werner, A., Koch, I., Osiewacz, H.D.: A genome-wide longitudinal transcriptome analysis of the aging model *Podospora anserina*. PLoS One **8**(12), 83109 (2013)

Philipp, O., Hamann, A., Osiewacz, H.D., Koch, I.: The autophagy interaction network of the aging model *Podospora anserina*. BMC Bioinformatics **18**(196) (2017)

Pickart, C.M.: Mechanisms underlying ubiquitination. Annual Review of Biochemistry **70**(1), 503–533 (2001)

Pole, A., Dimri, M., Dimri, G.P.: Oxidative stress, cellular senescence and ageing. AIMS Molecular Science **3**(3), 300–324 (2016)

Prasad, T.K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., *et al.*: Human protein reference database - 2009 update. Nucleic Acids Research **37**(suppl 1), 767–772 (2009)

Pyo, J.-O., Yoo, S.-M., Ahn, H.-H., Nah, J., Hong, S.-H., Kam, T.-I., Jung, S., Jung, Y.-K.: Overexpression of Atg5 in mice activates autophagy and extends lifespan. Nature Communications **4**(2300) (2013)

Rao, V.S.*et al.*: Protein-protein interaction detection: Methods and analysis. International Journal of Proteomics **2014** (2014)

Rattan, S.I.: Theories of biological aging: genes, proteins, and free radicals. Free Radical Research **40**(12), 1230–1238 (2006)

Razick, S.*et al.*: iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics **9**(405) (2008)

Reeves, G.A., Talavera, D., Thornton, J.M.: Genome and proteome annotation: organization, interpretation and integration. Journal of The Royal Society Interface **6**(31), 129–147 (2009)

Reiter, A., Steinbauer, R., Philippi, A., Gerber, J., Tschochner, H., Milkereit, P., Griesenbeck, J.: Reduction in ribosomal protein synthesis is sufficient to explain major effects on ribosome production after short-term TOR inactivation in *Saccharomyces cerevisiae*. Molecular and Cellular Biology **31**(4), 803–817 (2011)

Rizet, G.: Sur la longevite des phenomen des souches de *Podospora anserina*. Comptes Rendus de l'Acadmie des Sciences **237**, 1106–1109 (1953)

Rose, M.R., Burke, M.K., Shahrestani, P., Mueller, L.D.: Evolution of ageing since darwin. Journal of Genetics **87**(4), 363–371 (2008)

Scheckhuber, C.Q., Osiewacz, H.D.: *Podospora anserina*: a model organism to study mechanisms of healthy ageing. Molecular Genetics and Genomics **280**(5), 365–374 (2008)

Scheckhuber, C., Erjavec, N., Tinazli, A., Hamann, A., Nyström, T., Osiewacz, H.: Reducing mitochondrial fission results in increased life span and fitness of two fungal ageing models. Nature Cell Biology **9**(1), 99–105 (2007)

Schreiber, A., Peter, M.: Substrate recognition in selective autophagy and the ubiquitin–proteasome system. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1843**(1), 163–181 (2014)

Schulze, A., Downward, J.: Navigating gene expression using microarrays - a technology review. Nature Cell Biology **3**(8), 190–195 (2001)

Servos, J., Hamann, A., Grimm, C., Osiewacz, H.D.: A differential genome-wide transcriptome analysis: impact of cellular copper on complex biological processes like aging and development. PLoS One **7**(11), 49292 (2012)

Strobel, I., Osiewacz, H.D.: Poly (ADP-ribose) polymerase is a substrate recognized by two metacaspases of *Podospora anserina*. Eukaryotic Cell **12**(6), 900–912 (2013)

Tatsuta, T., Langer, T.: Quality control of mitochondria: protection against neurodegeneration and ageing. The EMBO Journal **27**(2), 306–314 (2008)

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V.: The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research **28**(1), 33–36 (2000)

The Gene Ontology Consortium: Gene ontology consortium: going forward. Nucleic Acids Research **43**(D1), 1049–1056 (2015)

Vaiserman, A.M.: Aging-modulating treatments: from reductionism to a system-oriented perspective. Frontiers in Genetics **5**(446) (2014)

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C.: Ten years of next-generation sequencing technology. Trends in Genetics **30**(9), 418–426 (2014)

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W.: Serial analysis of gene expression. Science **270**(5235), 484 (1995)

Weber, T.A., Reichert, A.S.: Impaired quality control of mitochondria: aging from a new perspective. Experimental Gerontology **45**(7), 503–511 (2010)

Wiemer, M., Osiewacz, H.D.: Effect of paraquat-induced oxidative stress on gene expression and aging of the filamentous ascomycete *Podospora anserina*. Microbial Cell **1**(7), 225–240 (2014)

Wiles, A.M., Doderer, M., Ruan, J., Gu, T.-T., Ravi, D., Blackman, B., Bishop, A.J.: Building and analyzing protein interactome networks by cross-species comparisons. BMC Systems Biology **4**(36) (2010)

Xing, S., Wallmeroth, N., Berendzen, K.W., Grefen, C.: Techniques for the analysis of protein-protein interactions in vivo. Plant Physiology **171**(2), 727–758 (2016)

Yu, J., Pacifico, S., Liu, G., Finley, R.L.: DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics **9**(461) (2008)

Chapter 9

Appendix

Path2PPI - Tutorial, example and the algorithm

Path2PPI - Tutorial, example and the algorithm

Prediction of autophagy induction in *Podospora anserina*

Contents

1 Introduction

Prediction of protein-protein interaction (PPI) networks is an important approach to gain knowledge about protein interactions in model organisms where only a small number of PPI information is available. Current PPI databases, providing predicted interaction data, lack many organisms or contain less reproducible information about the predicted interactions. Currently available prediction approaches are mainly based on biological data (functional annotation, co-expression etc.) which often are not available for many "less established" organisms, for example, where only sequence data is available. In addition, it is of major interest to get knowledge about a certain pathway in such a "less-studied" organism. To overcome these drawbacks **Path2PPI** can be used to predict proteins and interactions of a certain pathway of interest in a target organism by using and combining the PPIs of other well established model organisms.

To do so, it needs a list of proteins of interest from each reference species and the result files produced by the local NCBI BLAST (Camacho et al., 2009) tool (see next chapter). The relevant interactions based on the users' protein lists are automatically extracted from the corresponding *iRefIndex* files (Razick et al., 2008).

2 Preparation of the data

In this tutorial, we make use of the test data set provided with the package. This data set consists of all data files necessary to predict the interactions of the induction step of autophagy in *Podospora anserina* by means of the corresponding PPIs in human and yeast. Hence, we first load the "autophagy induction" test data set:

```
data(ai) #Load test data set
ls() #"ai" contains six data objects
```

```
## [1] "human.ai.irefindex"   "human.ai.proteins"    "pa2human.ai.homologs"
## [4] "pa2yeast.ai.homologs" "yeast.ai.irefindex"   "yeast.ai.proteins"
```

As stated by `ls()` the test data set contains six data objects (three for each of the two reference species human and yeast). First, the algorithm requires a list of proteins which define the corresponding pathway for each reference species, defined in *"human.ai.proteins"* and *"yeast.ai.proteins"* (see section 2.1). Second, the algorithm requires the data frames which contain the interactions of each reference species defined in *"human.ai.irefindex"* and *"yeast.ai.irefindex"* (described in more detail in section 2.1). Third, the algorithm needs to know the homologous relations between the target species with each reference species. These relations are defined in the data frames *"pa2human.ai.homologs"* and *"pa2yeast.ai.homologs"* (we describe this in more detail in section 2.2).

If you want to use **Path2PPI** for your own demands, you have to generate and prepare the necessary data files.

2.1 Proteins and interactions of pathways of interest

We list the proteins which are associated with a specific pathway of interest in a character vector for each reference species. To give you an example for such lists, we take a brief look into the loaded data set. Among others, we found the two named character vectors *"human.ai.proteins"* and *"yeast.ai.proteins"* which consist of the corresponding proteins for yeast and human, our two reference species:

```
human.ai.proteins
```

```
##  P42345  O75385  Q8IYT8  Q6PHR2  O75143
##  "MTOR"  "ULK1"  "ULK2"  "ULK3" "ATG13"
```

```
yeast.ai.proteins
```

```
##  P35169  P32600  P53104  Q06410  Q12527  Q06628  P39968
##  "TOR1"  "TOR2"  "ATG1" "ATG17" "ATG11" "ATG13"  "VAC8"
```

In this example, the values are the trivial names of the proteins and the names are the actual protein identifiers. **Path2PPI** also accepts simple character vectors where the values are the protein identifiers, if the trivial names of the proteins are not available. For example, this simple character vector, only consisting of the protein identifiers, would be also a valid protein list:

```
## [1] "P42345" "O75385" "Q8IYT8" "Q6PHR2" "O75143"
```

The major advantage of using a named character vector with the trivial names, is that these names will be shown in the plots allowing for a more comfortable interpretation. You can use various accession formats for the protein identifiers which are supported by *iRefIndex* (e.g. UniProt, SwissProt, Ensembl). However, we urgently recommend to use UniProt identifiers, since those are the most established ones.

Use the default R functions to load your own protein lists of interest into R (e.g. `read.table`).

These proteins of interest are applied to find relevant interactions in the corresponding species *iRefIndex* file. *iRefIndex* tables are available for the seven most established model organisms and can be found here: http://irefindex.org/wiki/index.php?title=iRefIndex. You can also use the corresponding *iRefR*-package to directly archive the *iRefIndex* data frames from this page. Unfortunately, the package is not updated as frequently as the web page and it may be that you do not get the latest release of a corresponding file.

In the "autophagy induction" test data set, only a very small part of the *iRefIndex* files for yeast and human are provided which contain the relevant interactions necessary for this tutorial. The complete files are much larger. The data frames *"human.ai.irefindex"* and *"yeast.ai.irefindex"* in the test data set contain these corresponding *iRefIndex* parts. See:

```
str(human.ai.irefindex)
str(yeast.ai.irefindex)
```

2.2 Get homology files using NCBI BLAST+

You also need the result files produced by the BLAST+ toolkit provided by the NCBI web page: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download. The test data set already includes the necessary results of the BLAST searches of the proteoms of *P. anserina* against the proteoms of yeast and human (*"pa2human.ai.homologs"* and *"pa2yeast.ai.homologs"*):

```
head(pa2yeast.ai.homologs)
```

```
##          V1     V2    V3  V4  V5 V6  V7  V8   V9  V10   V11   V12
## 534  B2AX00 P53104 25.66 304 165 12  15 268   21  313 1e-15  79.0
## 960  B2AXW7 P53104 28.52 291 154 10 510 751   30  315 7e-24 106.0
## 1278 B2AUR5 P35169 26.74 288 154 10 602 837 2078 2360 2e-11  65.9
## 1279 B2AUR5 P32600 25.09 275 152  9 610 835 2090 2359 1e-09  60.5
## 1555 B2B7B1 P53104 25.99 177  99  7 807 963  149  313 5e-09  59.3
## 2469 B2ASL9 P53104 26.42 352 171 13  20 314   24  344 2e-19  87.8
```

```
head(pa2human.ai.homologs)
```

```
##          V1     V2    V3  V4  V5 V6  V7  V8 V9 V10   V11   V12
## 2123 B2AX00 Q6PHR2 33.09 269 146 15  11 268 13 258 5e-24 106.0
## 2177 B2AX00 O75385 30.30 231 131  9  24 239 22 237 3e-22 103.0
## 2213 B2AX00 Q8IYT8 29.24 236 142  8  24 247 15 237 4e-21  99.4
## 4588 B2AXW7 Q6PHR2 32.96 267 159 12 499 754  6 263 9e-30 125.0
## 4649 B2AXW7 Q8IYT8 29.17 216 136  6 509 714 14 222 2e-23 107.0
## 4658 B2AXW7 O75385 29.95 217 133  9 509 714 21 229 5e-23 106.0
```

The second column (V2) contains the protein identifiers of the corresponding reference species to which the protein of the target species (here: *P. anserina*) in the first column (V1) is homologous. Keep in mind that these protein identifiers are equal to those we used in the protein lists described above.

If you are unfamiliar with this toolkit, we refer to the BLAST+ user manual or to the broadly available tutorials in the web.

Nevertheless, we want to give you a very short description on how to use this toolkit. We assume that you already have loaded and installed the NCBI BLAST+ toolkit and you also have each proteome file in FASTA format of each species (target and reference species). You first have to create the databases for each reference species, here, for human and yeast:

```
makeblastdb -in human.fa -input_type fasta -dbtype prot -out human_proteins
-title human_proteins
makeblastdb -in yeast.fa -input_type fasta -dbtype prot -out yeast_proteins
-title yeast_proteins
```

Subsequently, you can start the comprehensive BLAST searches using the FASTA file of your target species (here: *P. anserina*):

```
blastp -query panserina.fa -db human_proteins -out human_panserina.out -evalue
0.0001 -outfmt 6
blastp -query panserina.fa -db yeast_proteins -out yeast_panserina.out -evalue
0.0001 -outfmt 6
```

Please, make sure that you use as the output format the tab delimited list indicated by the parameter `-outfmt 6`. The two species-specific homology files which are now generated, can be imported into your R session, using the function `read.table`, and subsequently used as data frames for the **Path2PPI**-package.

3 Predict PPI in target species

After the necessary data sets are generated or loaded, respectively, we can start with the prediction.

3.1 The Path2PPI object

An object of the class *Path2PPI* represents the major instance which is responsible for storing and managing of each data set and for each computation and prediction step. Hence, we first have to create a new instance of the class *Path2PPI* with the corresponding information:

```
ppi <- Path2PPI("Autophagy induction", "Podospora anserina", "5145")
```

The arguments are the title of the pathway we want to predict, the taxonomy name of the target species ("Podospora anserina") and its corresponding taxonomy id ("5145").

3.2 Add reference species

This new instance does not contain any reference species or a predicted PPI, yet:

```
ppi
```

```
## Autophagy induction in Podospora anserina (5145)
## -------------------------------------------------
## No reference species yet.
## -------------------------------------------------
## No predicted PPI yet.
```

To add the reference species, for which we have collected the necessary data, we make use of the method `addReference`.

```
ppi <- addReference(ppi, "Homo sapiens", "9606", human.ai.proteins,
                    human.ai.irefindex, pa2human.ai.homologs)
```

```
## Search for all relevant interactions:
## 0%--25%--50%--75%--100%
## Remove irrelevant homologs.
```

Besides the taxonomy name and the taxonomy identifier, this method requires the list, containing the proteins of the pathway of interest, the corresponding *iRefIndex*-data frame or the file name of the corresponding *iRefIndex* file, and the species specific homology data set generated by the NCBI BLAST+ toolkit. This method searches for all relevant interactions in the *iRefIndex* data frame. There are different and often ambiguous protein identifiers defined in an *iRefIndex* file and the "major" identifiers are not necessarily those defined in the corresponding "major" columns "uidA" and "uidB". Furthermore, *iRefIndex* also contains complexes. Hence, this method applies an advanced search algorithm to find automatically relevant interactions associated with the pathway or the proteins of interest, respectively. You do not have to predefine the identifiers' types (UniProt, Swissprot, Ensembl etc.), since these types are often assigned ambiguously. The algorithm searches for each identifier in 10 columns where any type of identifier or accession number is defined, for example, "uidA", "altA", "OriginalReferenceA", "FinalReferenceA", "aliasA", "uidB", "altB", "OriginalReferenceB", "FinalReferenceB" and "aliasB". Additionally, it searches for each complex to which one or more of the predefined proteins are associated. Subsequently, each homologous relationship which is not relevant for the previously found interactions is declined.

In the same manner we add yeast to our Path2PPI-instance:

```
ppi <- addReference(ppi, "Saccharomyces cerevisiae (S288c)", "559292",
                    yeast.ai.proteins, yeast.ai.irefindex,
                    pa2yeast.ai.homologs)
```

```
## Search for all relevant interactions:
## 0%--25%--50%--75%--100%
## Remove irrelevant homologs.
```

In this tutorial, we want to predict the PPIs in *P. anserina* based on these two reference species. You can use other and/or more reference species for your demands.

Now, we can get all processed information about the added reference species using the method *showReferences*:

```
showReferences(ppi)
```

```
## Homo sapiens (TaxId: 9606)
## ---------------------------
## 5 proteins (0 not used)
## 894 interactions:
## - 6 interactions have both interactors in protein list.
## - 349 interactions have at least one interactor in protein list.
## - 660 interactions in 102 protein complexes.
##
##
## Saccharomyces cerevisiae (S288c) (TaxId: 559292)
## ------------------------------------------------
```

```
## 7 proteins (0 not used)
## 2910 interactions:
## - 15 interactions have both interactors in protein list.
## - 834 interactions have at least one interactor in protein list.
## - 2207 interactions in 102 protein complexes.
```

If we want to know which interactions have been found or which interactions are associated with the proteins of interest in a specific reference species (e.g. human), we can use the method as follows:

```
interactions <- showReferences(ppi, species="9606",
                               returnValue="interactions")
head(interactions)
```

```
##       ref      A.db                 A.accession A.in.prot.list      B.db
## 1  287217   complex qx1eWqPyfshUfC/6x17AYjcT/3w          FALSE  replaced
## 7  287217   complex qx1eWqPyfshUfC/6x17AYjcT/3w          FALSE uniprotkb
## 13 287217   complex qx1eWqPyfshUfC/6x17AYjcT/3w          FALSE uniprotkb
## 19 436141 uniprotkb                  A0A090N900          FALSE  replaced
## 32 502959 uniprotkb                      P62942          FALSE  replaced
## 45 408315 uniprotkb                      Q8N122          FALSE  replaced
##    B.accession B.in.prot.list
## 1        P42345           TRUE
## 7   A0A0A0MR05          FALSE
## 13       Q6R327          FALSE
## 19       P42345           TRUE
## 32       P42345           TRUE
## 45       P42345           TRUE
```

For more information about the method we refer to the corresponding manual page (`?showReferences`).

3.3 Predict PPI

After we added all reference species and all necessary data, we can start with the prediction. To predict the PPI network in the target species we use the method `predictPPI`:

```
ppi <- predictPPI(ppi,h.range=c(1e-60,1e-20))
```

```
## Begin with Homo sapiens
## 6 interactions processed. These lead to 5 interactions in target species.
## ------------------------------
## Begin with Saccharomyces cerevisiae (S288c)
## 15 interactions processed. These lead to 22 interactions in target species.
## ------------------------------
## Combine results to one single PPI.
## A total of 13 putative interactions were predicted in target species.
```

This method uses different arguments to influence the prediction approach and to define the output of the PPI network. For a detailed description of the various arguments we refer to the corresponding manual (`?predictPPI`). Here, we only use the argument `h.range` where the first value corresponds to the lower bound and the second value to the upper bound of the homology range. That means that each E-value which is equal or less the lower bound will be scored with 1, and each E-value which is equal or larger than the upper bound will be scored with 0 (see appendix for a detailed description).

According to the reports generated by this method two species specific PPI networks led to a PPI network in the target species with 13 interactions. To achieve further information about the former prediction step, we just type:

```
ppi #show(ppi)
```

```
## Autophagy induction in Podospora anserina (5145)
## ------------------------------------------------
## 2 reference species: 9606, 559292
## ------------------------------------------------
## Number of predicted proteins: 8
## Number of predicted interactions: 13
## Predicted PPI based on 2 reference species:
## 9606 (2 interactions and 4 homologous relations)
## 559292 (11 interactions and 11 homologous relations)
## ------------------------------------------------
## Settings:
## Homology threshold: 1e-05
## Homology range: [1e-60,1e-20]
## Interactions threshold: 0.7
## Consider complexes: FALSE
```

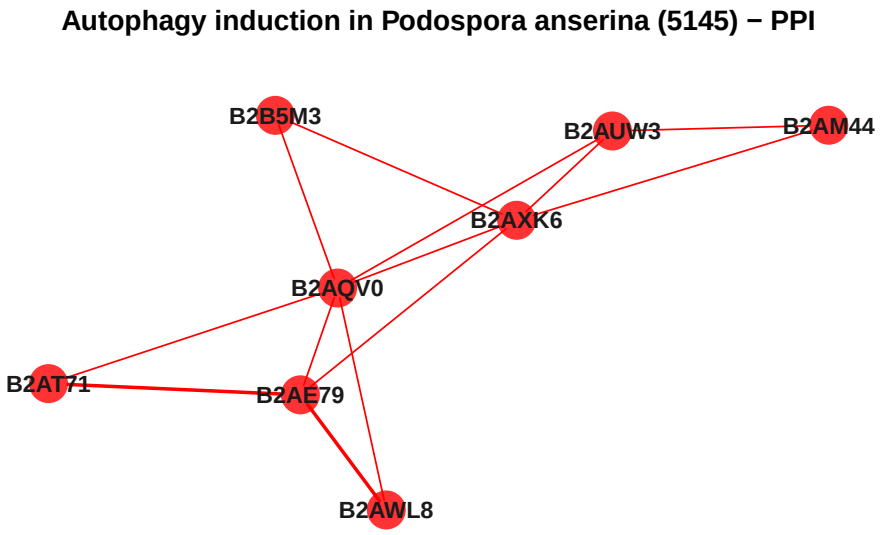
4 Results of the prediction

After we predicted the PPI network of the "autophagy induction" pathway in *P. anserina* we now want to know how this network looks like. And we want to know which proteins and interactions actually are associated with this pathway in our target species.

4.1 Plotting the results

To get a graphical representation of the predicted PPI network, *Path2PPI* provides three different plotting types. First, to get only the predicted PPI, we use the plot function of the Path2PPI-object, which is based on the *igraph* plotting function (Csardi and Nepusz, 2006):

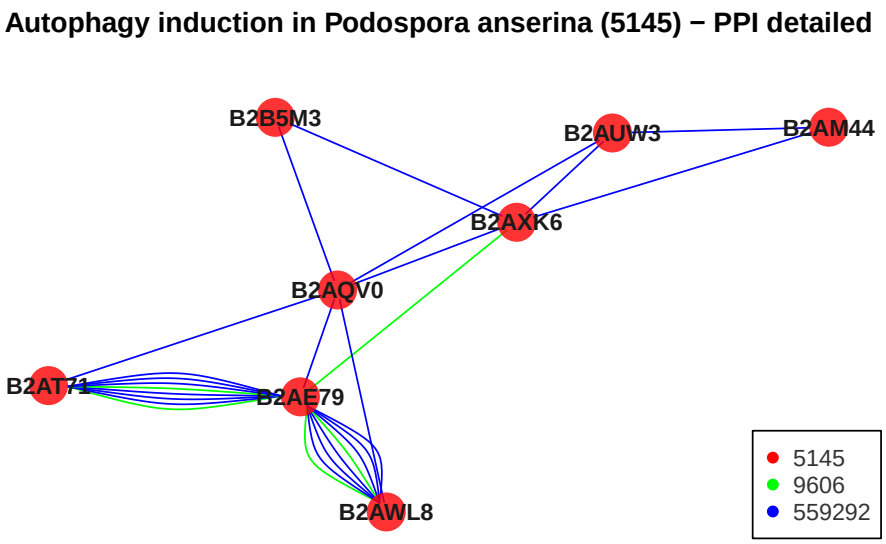
```
set.seed(12) #Set random seed
coordinates <- plot(ppi, return.coordinates=TRUE)
```



There are various arguments provided with this method (see `?plot.Path2PPI`). Here, we initially use the `return.coordinates` argument since we want to save the coordinates of the vertices for the next plotting approach.

In the second approach, we want to know from which reference species the different predicted interactions originated. We assign the previously computed coordinates to the plotting function since we want to compare both networks:

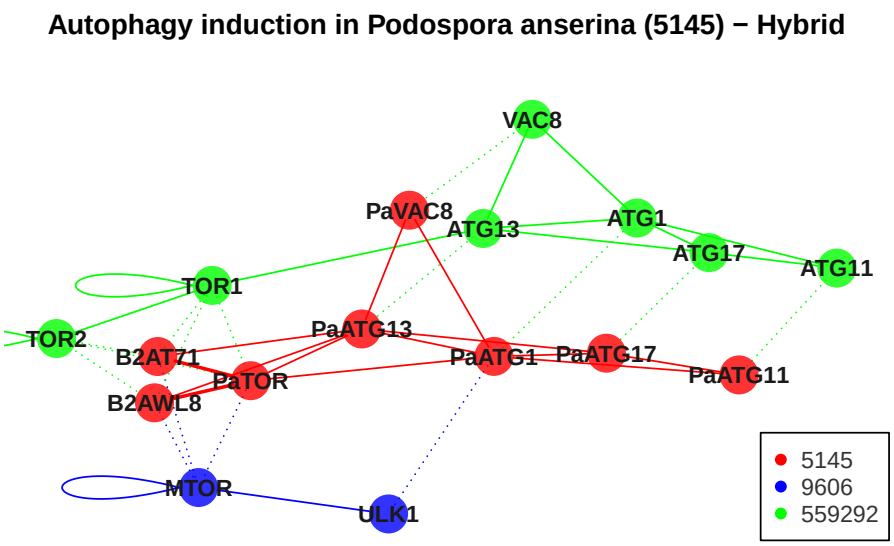
```
plot(ppi,multiple.edges=TRUE,vertices.coordinates=coordinates)
```



The different colors of the edges correspond to the species, see the taxonomy identifiers in the legend: 5154 for *P. anserina*, 9606 for human, and 559292 for yeast) from which the interaction was deduced. For example, we can see that the edge between the proteins "B2AWL" and "B2AE79" in the upper network is thicker than the others. This indicates that the interaction was found in more than one reference species. In the second plot, we see that this interaction is based on six interactions found in yeast and two interactions found in human.

Next, we want to plot the so-called *hybrid* PPI network, where we additionally can see the underlying reference interactions or the underlying reference PPI networks, respectively, and each homologous relationship. We also want to set the vertex labels, since we know the trivial names of the target species proteins. You can set the label for each protein of each species. Additionally, we want to change the species colors:

```
set.seed(40)
target.labels<-c("B2AE79"="PaTOR","B2AXK6"="PaATG1",
                 "B2AUW3"="PaATG17","B2AM44"="PaATG11",
                 "B2AQV0"="PaATG13","B2B5M3"="PaVAC8")
species.colors <- c("5145"="red","9606"="blue","559292"="green")
plot(ppi,type="hybrid",species.colors=species.colors,
     protein.labels=target.labels)
```



The dotted edges correspond to an homologous relationship between a protein of the target species and a reference species. Only those proteins and interactions of the reference species are shown which actually led to an interaction in the target species.

4.2 Get detailed information about each interaction

After we know how the predicted PPI network of this pathway may look like in the target species we want to know more about the predicted interactions. For this purpose we make use of the method `showInteraction` which requires the protein identifiers of the interaction:

```
showInteraction(ppi,interaction=c("B2AT71","B2AE79"))
```

```
## Score: 1.989472
## 8 reference interactions: 559292 (6) 9606 (2)
```

For further details about the underlying reference interactions we can use the additional argument `mode`:

```
showInteraction(ppi,interaction=c("B2AT71","B2AE79"),mode="detailed",
                verbose=FALSE)
```

```
##   source.id target.id score h.scoreA h.scoreB r.species r.species.s
## 1    B2AT71    B2AE79 0.855     0.71     1.00      9606      P42345
## 2    B2AT71    B2AE79 0.820     0.64     1.00    559292      P35169
## 3    B2AT71    B2AE79 0.825     0.65     1.00    559292      P32600
## 4    B2AT71    B2AE79 0.820     0.64     1.00    559292      P35169
```

```
## 5    B2AE79    B2AT71 0.855     1.00     0.71      9606      P42345
## 6    B2AE79    B2AT71 0.825     1.00     0.65    559292      P35169
## 7    B2AE79    B2AT71 0.825     1.00     0.65    559292      P32600
## 8    B2AE79    B2AT71 0.820     1.00     0.64    559292      P35169
##   r.species.t pos.edges used.edges    ref
## 1      P42345         6          4 742389
## 2      P32600         6          4 522660
## 3      P32600         6          4 858136
## 4      P35169         6          4 105672
## 5      P42345         6          4 742389
## 6      P32600         6          4 522660
## 7      P32600         6          4 858136
## 8      P35169         6          4 105672
```

This data frame contains each single predicted interaction of the current interaction and all corresponding reference interactions. For the interaction of the proteins in columns one and two, the third column gives the prediction score. The fourth and the fifth columns show the homology score of the source protein (A) or the target protein (B), respectively, in the target species to its equivalent (column seven and eight) in the reference species (column six). The column "pos.edges" (possible edges) indicates how many interactions could be derived from this interaction in the reference species, in contrast, to the number of interactions in "used.edges" which actually were adopted. The last column gives the identifier of this interaction in the *iRefIndex* data set.

To get the corresponding *iRefIndex* entries for these reference interactions we can use this method as follows:

```
ref.interaction <- showInteraction(ppi,interaction=c("B2AT71","B2AE79"),
                                   mode="references.detailed",verbose=FALSE)
```

The data frame now stored in the variable `ref.interaction` is part of the *iRefIndex* table which contains all information about the current reference interactions. We can use this data frame with default R functionality to search for specific information of the reference interactions. For example, if we want to know from which study (author and publication) and from which database the interaction of the proteins "P42345" and "P42345" with the interaction identifier "742389" (first row) was adopted, we use:

```
ref.interaction[ref.interaction$irigid=="742389",
                c("author","pmids","sourcedb")]
```

```
##                   author
## 9096                   -
## 143905 Thedieck K (2007)
## 143974 Takahara T (2006)
## 143979    Urano J (2007)
## 387372                 -
## 387373                 -
## 586846       remy-2001-1
##                                                                    pmids
## 9096                                                     pubmed:15467718
## 143905                                                   pubmed:18030348
## 143974                                                   pubmed:16870609
## 143979                                                   pubmed:17360675
## 387372 pubmed:10702316|pubmed:11438723|pubmed:12030785|pubmed:16870609
## 387373 pubmed:10702316|pubmed:11438723|pubmed:12030785|pubmed:16870609
## 586846                                                   pubmed:11438723
```

```
##                sourcedb
## 9096      MI:0462(bind)
## 143905 MI:0463(biogrid)
## 143974 MI:0463(biogrid)
## 143979 MI:0463(biogrid)
## 387372    MI:0468(hprd)
## 387373    MI:0468(hprd)
## 586846  MI:0469(intact)
```

In this manner you are able to search for each entry provided by the *iRefIndex* table.

4.3 Export results

Now, we want to work with the predicted PPI network and do further analyses either directly in R or using an advanced network analysis tool like Cytoscape (Cline et al., 2007). To do so, we can use the method `getPPI`, which gives us the edge list of the PPI network:

```
my.ppi <- getPPI(ppi)
my.ppi
```

```
##    source.id target.id    score   r.species
## 1     B2AT71    B2AE79 1.989472 559292,9606
## 2     B2AE79    B2AWL8 1.969948 559292,9606
## 3     B2AQV0    B2B5M3 1.000000      559292
## 4     B2AQV0    B2AXK6 1.000000      559292
## 5     B2AM44    B2AXK6 1.000000      559292
## 6     B2AM44    B2AUW3 1.000000      559292
## 7     B2AQV0    B2AUW3 1.000000      559292
## 8     B2AE79    B2AQV0 1.000000      559292
## 9     B2AXK6    B2AUW3 1.000000      559292
## 10    B2AXK6    B2B5M3 1.000000      559292
## 11    B2AXK6    B2AE79 0.820000        9606
## 12    B2AT71    B2AQV0 0.820000      559292
## 13    B2AWL8    B2AQV0 0.780000      559292
```

If you want this PPI network with each single (not the combined one) predicted interaction, use the additional argument `raw=TRUE`.

We also want the edge list of the hybrid network which includes the PPIs of the reference species. For this purpose we use the method `getHybridNetwork`:

```
my.hybrid <- getHybridNetwork(ppi)
my.hybrid
```

```
##     source.id target.id t.species.id    score        type
## 1      B2AT71    P42345         9606 0.710000    homology
## 2      B2AT71    P35169       559292 0.640000    homology
## 3      B2AT71    P32600       559292 0.650000    homology
## 5      B2AE79    P42345         9606 1.000000    homology
## 6      B2AE79    P35169       559292 1.000000    homology
## 7      B2AE79    P32600       559292 1.000000    homology
## 13     B2AWL8    P42345         9606 0.510000    homology
```

```
## 14     B2AWL8    P35169       559292 0.560000    homology
## 15     B2AWL8    P32600       559292 0.610000    homology
## 17     B2AXK6    O75385         9606 0.640000    homology
## 18     B2AQV0    Q06628       559292 1.000000    homology
## 20     B2AM44    Q12527       559292 1.000000    homology
## 26     B2AXK6    P53104       559292 1.000000    homology
## 45     B2B5M3    P39968       559292 1.000000    homology
## 48     B2AUW3    Q06410       559292 1.000000    homology
## 11     P42345    P42345         9606 1.000000 interaction
## 24     P35169    P32600       559292 1.000000 interaction
## 310    P32600    P32600       559292 1.000000 interaction
## 4      P35169    P35169       559292 1.000000 interaction
## 171    O75385    P42345         9606 1.000000 interaction
## 181    Q06628    P39968       559292 1.000000 interaction
## 19     Q06628    P53104       559292 1.000000 interaction
## 201    Q12527    P53104       559292 1.000000 interaction
## 21     Q12527    Q06410       559292 1.000000 interaction
## 22     Q06628    Q06410       559292 1.000000 interaction
## 23     P35169    Q06628       559292 1.000000 interaction
## 261    P53104    Q06410       559292 1.000000 interaction
## 27     P53104    P39968       559292 1.000000 interaction
## 29     B2AT71    B2AE79         5145 1.989472 interaction
## 30     B2AE79    B2AWL8         5145 1.969948 interaction
## 31     B2AQV0    B2B5M3         5145 1.000000 interaction
## 32     B2AQV0    B2AXK6         5145 1.000000 interaction
## 33     B2AM44    B2AXK6         5145 1.000000 interaction
## 34     B2AM44    B2AUW3         5145 1.000000 interaction
## 35     B2AQV0    B2AUW3         5145 1.000000 interaction
## 36     B2AE79    B2AQV0         5145 1.000000 interaction
## 37     B2AXK6    B2AUW3         5145 1.000000 interaction
## 38     B2AXK6    B2B5M3         5145 1.000000 interaction
## 39     B2AXK6    B2AE79         5145 0.820000 interaction
## 40     B2AT71    B2AQV0         5145 0.820000 interaction
## 41     B2AWL8    B2AQV0         5145 0.780000 interaction
```

If you want to work with an *igraph* object instead, use the additional argument `igraph=TRUE` in both methods. To export the edge lists, use default R functions like `write.table`.

5 References

Camacho, C. et al. (2009). BLAST+: architecture and applications. BMC Bioinformatics, 10(1), 421.

Cline, M. S. et al. (2007). Integration of biological networks and gene expression data using Cytoscape. Nature Protocols, 2(10), 2366-2382.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems, 1695(5), 1-9

Razick, S. et al. (2008). iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics, 9(1), 405.

Kanehisa, M. et al. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research, 42(D1), D199-D205.

6 Session info

```
sessionInfo()
```

```
## R version 3.2.1 (2015-06-18)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 8 x64 (build 9200)
##
## locale:
## [1] LC_COLLATE=German_Germany.1252  LC_CTYPE=German_Germany.1252
## [3] LC_MONETARY=German_Germany.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] Path2PPI_1.1.1 igraph_1.0.1
##
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5    formatR_1.2.1   tools_3.2.1     htmltools_0.2.6
##  [5] yaml_2.1.13     stringi_1.0-1   rmarkdown_0.8.1 knitr_1.11
##  [9] stringr_1.0.0   digest_0.6.8    evaluate_0.8
```

7 Appendix

7.1 Biological evidence of the predicted PPI network

The example of autophagy induction was choosen, since it is well known and described for many eukaryotic organisms, including yeast, human and even *P. anserina*. Representatively and briefly described for yeast, the induction step works as follows:

Generally, nutrient availability activates the TOR-kinase, which phosphorylates ATG13. Hyper-phosphorylated ATG13 cannot interact with ATG1 and ATG17 to build the ATG1 complex (including ATG11 and VAC8) , which is important for autophagosome nucleation. Hence, autophagy is inhibited. In contrast, under nutrient-depleted conditions, TOR is deactivated, ATG13 is dephosphorylated and available for complex formation with ATG1 and ATG17. The next step of the autophagy process - autophagosome nucleation - can begin.

As reference proteins we used those provided by the KEGG database (Kanehisa et al., 2014) for yeast (seven) and human (five). We compared the predicted PPI with that provided by the KEGG database for *P. anserina*.

We found that two of the five proteins of human, ULK1 and TOR, and all of the seven proteins of yeast have been taken into account from the algorithm. This is due to the closer evolutionary distance of the both fungi and the corresponding proteins. The alogrithm predicts all of the proteins and interactions provided by the KEGG database for the induction step of autophagy in *P. anserina*. Additionally, it also includes two additional proteins, which are not characterized yet. This is propably due to the fact that these two proteins are putative serine/threonine kinases with similar motifs and sequences like the TOR kinase.

Summarizing, this example shows that **Path2PPI** is able to transfer interaction data amongst organisms. The resulting PPI networks can serve as starting points for further analyses and PPI studies.

7.2 The prediction algorithm

Aim: We aim to predict an interaction network in a target organism based on the interaction networks of well established model organisms. The networks are based on a predefined set of proteins which may belong to a certain pathway. We consider the degree of homologies and the number of reference species.

Requirements: The major initial requirements for the algorithm are the PPIs for each reference species, the BLAST results of the target species (*P. anserina*) against each reference species, and an E-value range defined by the upper bound h_u and the lower bound h_l. The range is required to map the BLAST E-value interval $[h_l, h_u]$ to a homology score in the interval $[1, 0]$ where 0 is the worst and 1 the best value. We will exemplarily describe the algorithm based on the two reference species, human and yeast.

7.2.1 Computing preliminary reference species-specific PPIs

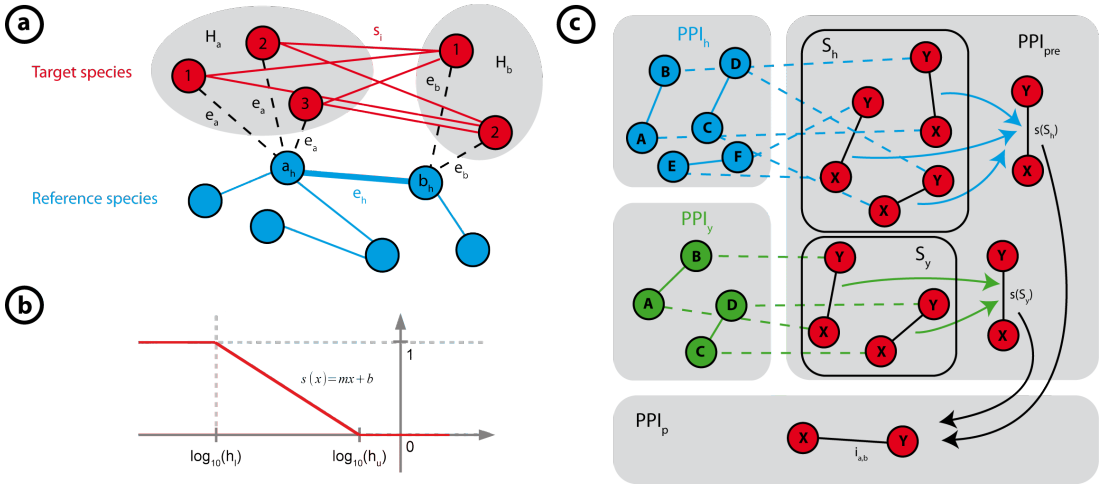


Figure 1. Steps of the prediction algorithm a) It searches for each interaction (blue edges) in each reference species (blue) the corresponding homologs (dotted edges) in the target species (red). Each valid homologous protein from the first set, H_a, will now be connected with each homologous protein from the other set, H_b. The deduced interactions are scored, using the function described below. b) The linear function to map the BLAST E-values to the interval $[0, 1]$. c) Combining the redundantly predicted interactions to one final PPI network, PPI_p, of the target species.

At the beginning, we have the (undirected) PPI graphs of human $PPI_h = (V_h, E_h)$ and yeast $PPI_y = (V_y, E_y)$ (figure 1a, the blue graph corresponds to one reference species), a set of unassigned nodes V_p which represents all *P. anserina* proteins (figure 1a, red nodes) and an empty preliminary PPI graph of *P. anserina* $PPI_{pre} = (V_p, E_p) = \emptyset$. The homology comparing approach leads to a directed homology graph

$H = (V_H, E_H)$ where

$V_H \subseteq (V_h \cup V_y \cup V_p)$ and

$E_H \subseteq \{(a, b) \mid a \in V_p, b \in (V_y \cup V_h)\}$.

Each edge in H is weighted by the E-values (EV) of the BLAST search.

Before starting the algorithm we have to set three major parameters. First, the parameter e_{thresh}, which defines a threshold for the E-values. Homologous relations which exceed that theshold are immeditely declined and will not be taken into account anymore. The second and third parameters are the upper and the lower bound for an E-value range $[h_l, h_u]$, where $h_l < h_u \leq e_{thresh}$ which will be mapped to the interval $[1, 0]$ by the scoring function s where 1 is the best score and 0 the worst (figure 1b). Since generally, homology-based network inference has to consider that that more homologous proteins exist, we aim to distinguish and weight

these relations by different scores provided by the scoring function. If only one homologous relationship exists, the scoring function will lead to the best score of 1, as well. This exception is due to the unambigous and desired situation if only one protein in the target species is homologous to the corresponding protein in the reference species.

Each interaction of each reference species is handled one by one and contributes to the prediction and scoring with an equal weight of 1, i.e. we do not distinguish between the underlying experimental methods. Furthermore, it is assumed that interactions and the reference species are stochastically independent, i.e. the occurance of one interaction does not influence the probality of the occurence of another interaction. We start with the first reference species, here, PPI_h:

For each $e_h = \{a_h, b_h\} \in E_h$ (figure 1a, vertices a_h and b_h):

1. Get $H_a \subseteq H$ with each $e_a = \{(a_{ea}, b_{ea}) \mid a_{ea}, b_{ea} \in V_H, b_{ea} = a_h\}$ and $H_b \subseteq H$ with each $e_b = \{(a_{eb}, b_{eb}) \mid a_{eb}, b_{eb} \in V, b_{eb} = b_h\}$ (figure 1a, dotted edges pointing to set H_a or set H_b)
2. *Predefinition: Given the graph H and $V(H)$ gives the set of vertices in H and $E(H)$ the set of edges in H then $\mid (V(H) \mid$ gives the number of vertices in H and $\mid (E(H) \mid$ the number of edges in H.*

   If $V(H_a) = \emptyset$ or $V(H_b) = \emptyset$, decline e_h, otherwise compute the homology score $s_h \rightarrow [0, 1]$ for each edge $e_a \in H_a$ and each edge $e_b \in H_b$ based on the single E-values and the cardinalities of H_a and H_b:

   If $\mid (E(H_a) \mid = 1$ then $s_h(e_a) = 1$ else $s_h(e_a) = \mid mlog_{10}(EV(e_a)) + b \mid$,

   if $\mid (E(H_b) \mid = 1$ then $s_h(e_b) = 1$ else $s_h(e_b) = \mid mlog_{10}(EV(e_b)) + b \mid$

   with $m = \frac{1}{log_{10}(h_l) - log_{10}(h_u)}$ and $b = -(mlog_{10}(h_u))$ (figure 1b).
3. Decline edges which are now scored with 0, i.e. set $H_a = H_a \setminus \{(e_a \in H_a), s_h(e_a) = 0\}$ and $H_b = H_b \setminus \{(e_b \in H_b), s_h(e_b) = 0\}$.
4. Add each remaining vertex $\{v \in V(\{H_a \cup H_b\}) \wedge v \in V_p\}$ from target species to PPI_{pre} and connect each node deduced from set H_a with each node deduced from set H_b (figure 1a, red edges). Compute a score s_i for each of these new predicted edges (interactions) using the arithmetic mean:

   $s_i = \frac{s_h(e_a) + s_h(e_b)}{2}$.

Decline edges where $s_i \leq s_{thresh}$ with s_{thresh} as a predefined threshold.

Repeat step 1-4 with PPI_y and each other reference species.

7.2.2 Combining PPIs deduced from each reference species

Now, the algorithm has to combine all predicted and probably redundant interactions in PPI_{pre} to one combined PPI of *P. anserina*, $PPI_p = (V_p, E_p)$. In particular, it has to consider interactions which were suggested by both reference species PPI networks (figure 1c):

For each $e_i = \{a_{ei}, b_{ei}\} \in E(PPI_{pre})$:

1. Get $I_{a,b} \subseteq PPI_{pre}$ with each $e_j = \{a_{ej}, b_{ej} \mid a_{ej}, b_{ej} \in E(PPI_{pre}), ((a_{ej} = a_{ei}) \wedge (b_{ej} = b_{ei})) \vee ((a_{ej} = b_{ei}) \wedge (b_{ej} = a_{ei}))\}$, i.e. further interactions with the same interacting partners of e_i (each edge X-Y in figure 1c).
2. Divide $I_{a,b}$ into subsets $S_h \in I_{a,b}$ and $S_y \in I_{a,b}$, depending on the reference species where the corresponding interactions have been found (black bordered areas in figure 1c).

3. Combine each score $s_i(s_h \in S_h)$ to one intra species score $s(S_h)$ and each score $s_i(s_y \in S_y)$ to one intra species score $s(S_y)$. Since a higher number of interactions increases the probability that $I_{a,b}$ exists in the target species, and the predicted interaction should be rated at least with the highest score, we sum up each score of each interaction in S_h or S_y, respectively, using a recursive function where s_k is the kth score in $s_i(s_y \in S_y)$. Additionally, $s_i(s_y \in S_y)$ has been sorted in a decreasing order:

   $s(s_k) = s_k$ if $k = 1$,

   $f(s_k) = f(s_{k-1}) + (1 - f(s_k - 1)) * s_k$ otherwise.

Hence, the intra species score is at least as high as the biggest sub score. For example, if we have three interactions deduced from one reference species, predicting all the same interaction in the target species with the scores 0.7, 0.9 and 0.6. Then, the final species score is:
$s(0.6, 0.7, 0.9) = 0.9 + (1 - 0.9) * 0.7 + (1 - (0.9 + (1 - 0.9) * 0.7)) * 0.6 = 0.988$.

4. Combine the two intra species scores to one common inter-species or final-score $i_{a,b}$ (figure 1c lower area):
   $i_{a,b} = \frac{s(S_h)+s(S_y)}{m} + m - 1$, where m is the number of reference species for which the interaction $I_{a,b} \subseteq PPI_{pre}$ was found (i.e. $m = 1$, if only in yeast or human and $m = 2$, if in both).
5. Finally, add the new edge (interaction) $I_{a,b}$ to PPI_p with the score $i_{a,b}$.