

From *Hydra* to humans
- Insights into molecular mechanisms of
aging and longevity -

Dissertation

in fulfillment of the requirements for the degree “Dr. rer. nat.”
of the Faculty of Mathematics and Natural Sciences
at Kiel University

submitted by

Guillermo Gonzalo Torres Estupiñan

Kiel, May 2019

First examiner: Prof. Dr. Thomas Roeder

Second examiner: Prof. Dr. Almut Nebel

Date of oral examination: **12.07.2019**

Approved for print on:

Approved by:

Prof. Dr. Frank Kempken, Dean

“Aging seems to be the only available way to live a long life.”

Daniel-Francois-Esprit Auber, French composer 1782-1871

Table of contents

Table of contents	I
List of tables	III
List of figures.....	IV
1. General introduction	1
1.1. Relevance of the present work.....	1
1.2. Aging and longevity in model organisms	2
1.2.1. The potentially immortal freshwater polyp <i>Hydra vulgaris</i>	3
1.3. Human longevity.....	5
1.3.1. The genetics of human longevity	5
1.4. The human gut microbiota in aging and longevity.....	10
1.4.1. The human gut microbiota.....	10
1.4.2. The human gut microbiota composition during development and in adulthood	11
1.4.3. Changes in the human gut microbiota composition in the elderly.....	13
1.4.4. Signatures of longevity in the human gut microbiome	13
1.5. Aims of the thesis	15
1.5.1. Specific aims.....	16
References	18
2. Chapter I: Boundary maintenance in the ancestral metazoan <i>Hydra</i> depends on histone acetylation.....	29
Graphical abstract	30
Abstract	32
2.1. Introduction	33
2.2. Results	34
2.2.1. Comparative transcriptomics identifies compartment-specific genes	34
2.2.2. Functional annotations outline the compartment-specific gene expression program.....	35
2.2.3. Metabolism, cell proliferation and adhesion pathways contribute to the boundary between head and body	36
2.2.4. Inhibition of histone deacetylases (HDAC) breaks down head/body column boundary	37
2.2.5. Cell cycle control and histone deacetylation	39
2.2.6. Wnt signaling and histone deacetylation.....	40
2.3. Discussion	40
2.3.1. Principles of boundary formation in the early emerging metazoan <i>Hydra</i> .	40
2.3.2. An evolutionary perspective on HDAC and boundary formation: parallels between <i>Hydra</i> and patterning in other animals and plants.....	43
2.4. Materials and methods	44
Acknowledgements	49
Funding	49
Author contributions	50
References	51

3. Chapter II: Exome-wide association study identifies <i>FN3KRP</i> and <i>PGP</i> as new candidate longevity genes in humans	69
Abstract	71
3.1. Introduction	72
3.2. Results	73
3.2.1. Single-variant association analysis reveals a longevity association of rs1046896 in <i>FN3KRP</i>	73
3.2.2. The longevity-association of rs1063192 (<i>CDKN2B</i>) replicates with borderline significance in an independent cohort	74
3.2.3. Gene-based analysis reveals <i>PGP</i> as a potential new longevity locus and strengthens the <i>FN3KRP</i> association	75
3.2.4. SNV-SNV interaction analysis reveals a functional longevity framework.....	75
3.3. Discussion	76
3.4. Conclusion	79
3.5. Materials and methods	79
Acknowledgments	84
Conflicts of interests.....	84
Funding	84
References	86
Tables	93
Figure legends	96
4. Chapter III: Human gut microbiome in healthy aging and longevity	101
4.1. Introduction	103
4.2. Results	104
4.2.1. Physiological parameters remain relatively stable during aging in healthy individuals	104
4.2.2. Four enterotype-like clusters stratify the GM and the global microbial diversity tends to increase with age, but appears to be lower in LLI	104
4.2.3. Enterotypes explain 19.8% of the bacterial structure variability	105
4.2.4. Microbial changes associated with aging.....	106
4.3. Discussion	107
4.4. Material and methods.....	111
References	114
Tables	121
Figure legends	122
5. Conclusion and outlook	131
References	137
6. Summary	139
7. Zusammenfassung	141
8. Declaration.....	143
9. Curriculum vitae.....	145
10. Acknowledgements	147

List of tables

Table 3.1. Association statistics for the 11 longevity-associated SNVs identified by the single-variant association approach in the whole German study population	93
Table 3.2. Single-variant replication and meta-analysis statistics for candidate SNVs in the French and Danish populations.....	94
Table 3.3. Association statistics for the 16 longevity-associated genes identified by the gene-based association approach in the whole German study population	95
Table 4.1. Environmental factors that significantly contribute to shaping of the microbial structure of the study population.....	121

List of figures

Figure 1.1. World life expectancy estimates compared, 1800 - 2001	1
Figure 1.2. The freshwater polyp <i>Hydra</i> in its completely differentiated form.....	4
Figure 1.3. Proportion of individuals reaching extreme ages either by surviving, delaying or escaping morbidity in three age groups	6
Figure 1.4. Gut microbiome changes from infancy to adulthood and factors influencing the microbial composition at the single stages.....	11
Figure 1.5. Proposed changes in the gut microbiota composition during aging and in extreme longevity.....	14
Figure 1.6. Overview of the three research projects	15
Figure 2.1. General approach followed to investigate <i>Hydra</i> boundaries.....	59
Figure 2.2. Identification of region-specific genes by comparative transcriptomics...	60
Figure 2.3. Summary of unbiased RNAseq approach for the identification of compartment-specific functional groups of genes	61
Figure 2.4. Changes in the transcriptional landscape upon TSA treatment	62
Figure 2.5. A higher expression of cell cycle genes is identified in head tissue when <i>Hydra</i> is treated with TSA	64
Figure 2.6. TSA treatment prevents the cells in the head expressing organizer genes.....	65
Figure 2.7. Histone deacetylation model.....	67
Figure 3.1. Workflow of the association analyses in the case-control study of longevity using the Illumina HumanExome BeadChip	97
Figure 3.2. Manhattan plot summarizing the findings from the single-variant analysis	97
Figure 3.3. Functional annotation map generated by ClueGo	97
Figure 4.1. Age distribution of the individuals enrolled in the study	122
Figure 4.2. Relationships among anthropometric measures, dietary parameters and the age of the individuals.....	122
Figure 4.3. Gut bacterial diversity of healthy Germans across a broad age range.....	122
Figure 4.4. Microbial stratification and enterotype-specific alpha-diversity	122
Figure 4.5. Inter-individual GM variability explained by microbial stratification, dietary and anthropometrics parameters	122
Figure 4.6. Taxa whose abundances change during aging	122
Figure 4.7. Differentially abundant taxa in LLI compared with younger subjects.....	122
Figure 5.1. Overview of the three research projects and their respective outcomes....	132

1.

General introduction

1.1. Relevance of the present work

Population aging and the global burden of late-life diseases

Only one century, i.e. from 1850 to 1950, was needed for humans to double their life expectancy [1] (**Figure 1.1**). Much of this dramatic increase in life expectancy can be explained by lower early life mortality after improvements in the quality of water, food, hygiene, housing, and medical care [2]. In this millennium (from 2000 to 2016), life expectancy has only risen 4.76 ± 1.12 years so far (life expectancy nowadays is ~ 77 years in developed countries; see <http://apps.who.int/gho/data/view.main.SDG2016LEXREGv?lang=en>; accessed 03.2019).

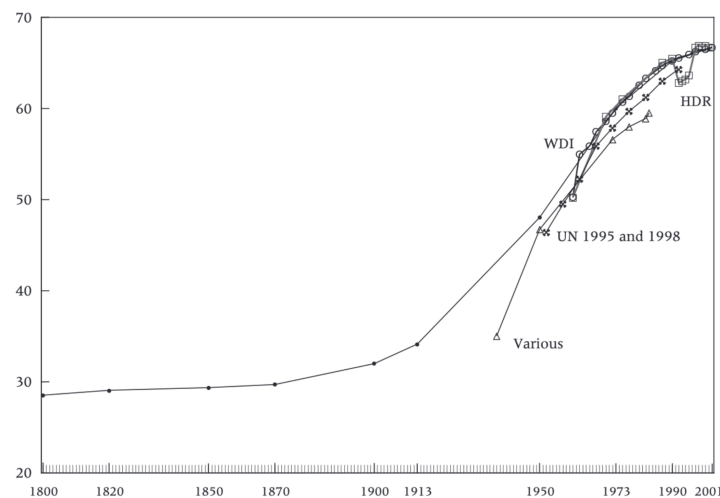


Figure 1.1. World life expectancy estimates compared, 1800 - 2001 (modified from [3]). HDR = UNDP, Human Development Report, various years; WDI = World Bank, World Development Indicators 2004. SOURCES: World Bank 2004; United Nations Development Programme (various years); United Nations 1995, 1998; Siampos 1989, who on p. 424 reproduces without reference what may be a League of Nations or UN estimate for 1930; United Nations 1975a: 1; 1975b, I: 175; and 1988

However, living longer does not necessarily mean that the years gained are characterized by health and reproductivity [4]. Infectious diseases do no longer represent leading causes of death

in developed populations, but they have been replaced by chronic degenerative diseases like cancer and cardiovascular diseases that occur later in life (a phenomenon known as "epidemiologic transition" [5]. While the shift from early mortality by infectious diseases to long lives with chronic diseases was triggered, as aforementioned, by improved lifestyle (hygiene) and medical conditions (vaccines, antibiotics), curing degenerative diseases nowadays is still hindered by knowledge gaps concerning the mechanisms of aging [6, 7]. It is well accepted that aging, the progressive accumulation of molecular damage and consequential loss of biological functioning, is the most important risk factor for chronic diseases and drives both morbidity and mortality [8].

The proportion of elderly is increasing and an alarming 16-20% of the late life is spent in morbidity, partly because medical treatments allow diseased people to survive longer with their illness [9, 10]. Late-life diseases become a global burden with vast economic implications [7]. If or how long the increase in life expectancy will continue or if there is a limit to human lifespan is lively debated among scientists [6, 11]. Some researchers even foresee a reversal, i.e. an eventual decrease in life expectancy [12].

One focus of modern aging research is on how to compress the morbid phase at the end of life (see e.g. [13]). Both aging and longevity have a genetic contribution, and both are also vastly influenced by environmental factors [14, 15]. Obviously, long-lived individuals (LLI) are also aged and show the common signs of aging (e.g. wrinkles, grey hair). The likelihood of a long lifespan, i.e. of becoming long-lived, is limited by individual aging processes. It becomes clear that aging and longevity are different phenotypes, but highly connected. In the past, both studies using (genetically modifiable) model organisms and research on LLI have helped shed light on genes and pathways responsible for or contributing to (healthy) aging and longevity.

1.2. Aging and longevity in model organisms

Aging can be defined as a progressive functional decline of an organism with a gradual deterioration of physiological function, including a decrease in fecundity [16, 17]. Sometimes, "aging" and "senescence" are used synonymously; however, the latter term rather refers to processes in a single cell [18].

The aging process, particularly in metazoans, appears to derive from intrinsic cellular mechanisms acting in parallel with changes in tissues. Scientists hypothesize that bone marrow

stem cells are crucial in the aging process due to their high self-renewal capacity [19]. Others, however, state that aging is controlled by a specific tissue or organ (e.g., the hypothalamus) rather than single cell populations [20–22]. Yet, despite the complexity of the aging processes, there is no scientific basis for the notion that aging is an inevitable process that cannot be stopped or postponed [23]. Many aging and longevity-associated pathways are evolutionarily highly conserved. Therefore, studies in model organisms, including yeast, nematodes, fruit flies, and mice, help shed light on those processes. Despite the yet not fully resolved aging mechanisms, it has been demonstrated in model organisms that lifespan can be extended to some degree. For example, mutations in genes associated with the insulin/insulin-like signaling pathway (such as *daf-2*, *daf-12*, *daf-16*, or PI3K_{CS}) in the nematode *Caenorhabditis elegans* (*C. elegans*) extended the lifespan of the organisms nearly 10-fold [24]. In mice, over-expression of the *klotho* gene which acts as a circulating hormone extended lifespan by ~30% [25], and disruption of the growth hormone receptor extended lifespan by more than 40% [26]. Furthermore, mice with mutations in *Prop1* (a transcription factor that regulates *Pit1*, a gene related to the insulin/insulin-like signaling pathway) lived twice as long as control mice that did not carry the mutation [27].

Caloric restriction (CR), i.e. reduced energy intake without malnutrition, appears to be the only lifestyle intervention so far that has led to increased lifespan in various lower model organisms, e.g. in *C. elegans* and in the fruit fly *Drosophila melanogaster* (*D. melanogaster*). Studies have also shown beneficial effects of CR on survival and health in mammals like mice and rhesus monkeys [28–30], and very recently, CR led to improved health measures (e.g. reduction of oxidative damage) in healthy non-obese humans [31].

1.2.1. The potentially immortal freshwater polyp *Hydra vulgaris*

Along the evolutionary tree, diverse forms of aging appear which are reflected in wide differences in lifespan. For example, some species of reptiles, amphibians, fishes and birds appear to be able to delay age-related pathologies, maybe even the aging process itself [32]. Others, like the freshwater polyp *Hydra* (**Figure 1.2**), display negligible senescence or even escape the aging process by coupling vegetative propagation to sexual reproduction [33, 34].

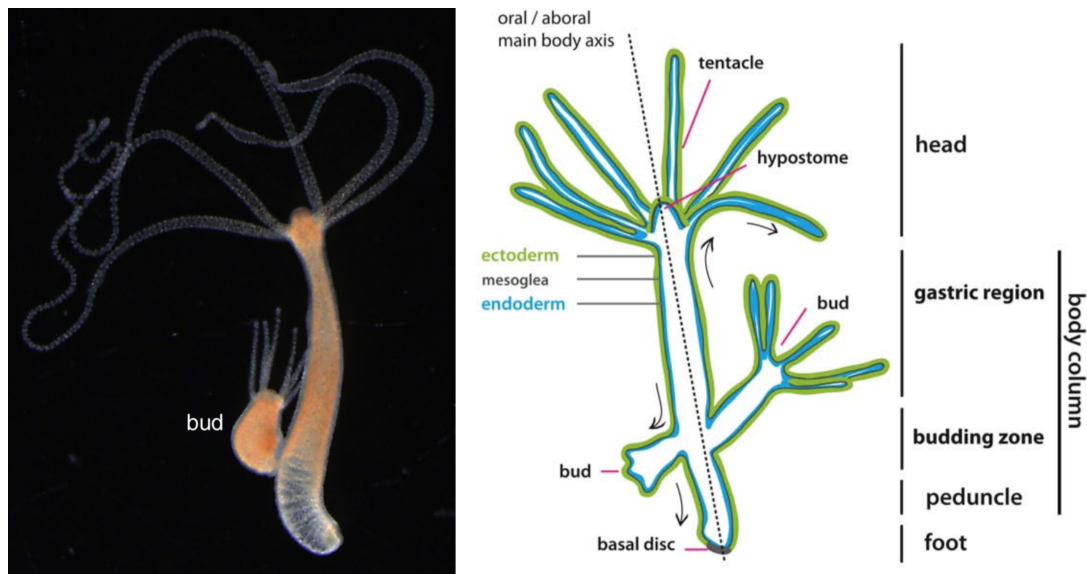


Figure 1.2. The freshwater polyp *Hydra* in its completely differentiated form (modified from [35]).

Hydra is considered to be potentially immortal. Impressively, when gametogenesis is induced in *Hydra* (e.g., *H. oligactis* or *Pelmatohydra robusta*), aging patterns seem to emerge which lead to the death of the organism within 100 to 120 days [34]. Gametogenesis in *Hydra* was shown to be associated with loss of interstitial stem cells and their somatic derivatives [34]. This pattern resembles the humans aging process which is related to the reduction of the capacity of self-renewal due to loss of somatic stem cells [36, 37]. Therefore, *Hydra*'s immortality was initially thought to be due to its asexual mode of reproduction. However, nowadays, it is widely accepted that the self-renewal capacity of the stem cell lineages contributes significantly to *Hydra*'s immortality, and one of the genes correlated with this capacity is the forkhead box O (FoxO) gene [38]. FoxO is an evolutionarily conserved transcription factor with a well-known role in cellular responses to environmental and physiological stress [39], and has been validated as a central regulator of lifespan across species [40]. In *H. vulgaris*, FoxO correlated with stem cell proliferation, budding processes and continuous self-renewal capacity [38]. Additionally, FoxO was shown to be crucial for the functional maintenance of the innate immune system and immune peptides production [35, 41]. Experimental evidence from studies in flies, worms and mice indicated that FoxO extends lifespan by stimulating cell survival, stem cell control, and tissue homeostasis [39, 40]. Strikingly, specific mutations in the human *FOXO3* gene predispose to human longevity [42–44].

In addition to *Hydra*'s immortality status, its potential as a model organism for aging and

longevity studies is remarkable. Studies focusing on the identification of orthologous genes among species showed that *Hydra* shares at least 6,071 genes with humans [45]. Impressively, of 259 human aging genes, 207 (80%) were shared by *Hydra*. In contrast, some of the human aging genes shared by *Hydra* were missing or poorly conserved in *D. melanogaster* and *C. elegans* (e.g., *MDM2* or TGF β inhibitor noggin) [46]. This evidence, along with the unique experimental techniques (e.g., genomics, transcriptomics, and transgenics), published in previous studies, highlights *Hydra* as a powerful model for mechanistic aging studies [46, 47].

Studying aging processes in other animals and comparing the way different species age or escape aging, helps scientists to gather knowledge about human aging and may eventually yield clues on how to delay it. In summary, research in animal models and humans has shown that genetic and environmental factors affect both life- and healthspan and both can potentially be prolonged.

1.3. Human longevity

Longevity could be defined as the period of time an organism is expected to live under ideal circumstances. Evolutionary and genetic research indicates that longevity has been a by-product of evolution, and results from the integration of many genetic and environmental factors which furthermore interact with each other; therefore, longevity can be considered as a complex trait [16, 17, 48, 49].

1.3.1. The genetics of human longevity

Hundreds of genes have been identified to be capable of influencing longevity [50]; most of them can be allocated to one or more of the following pathways i) protein homeostasis, ii) insulin/insulin-like growth factor signaling (IIS), iii) mitochondrial metabolism, iv) sirtuin signaling, v) chemosensory function or vi) pathways associated with CR [51, 52]. This aging gene pool was predominantly defined with the help of model organisms (e.g., *Saccharomyces cerevisiae* (*S. cerevisiae*), *C. elegans* and *D. melanogaster*) using experimental approaches such as random mutagenesis [53] or overexpression screenings [54, 55]. However, aforementioned techniques are suitable for models that are relatively simple and inexpensive to maintain in the laboratory (e.g., *Hydra*), but not for long-lived models and/or those less flexible to targeted genetics (e.g., humans). In humans, genetic mapping techniques, either by studying a secondary phenotype that correlates with longevity (e.g., for humans: Hutchinson–Gilford progeria, mandibuloacral dysplasia, Werner’s syndrome) [15], or the investigation of a genetically

diverse population with a natural variation in lifespan allow scientists to identify longevity-associated loci [14].

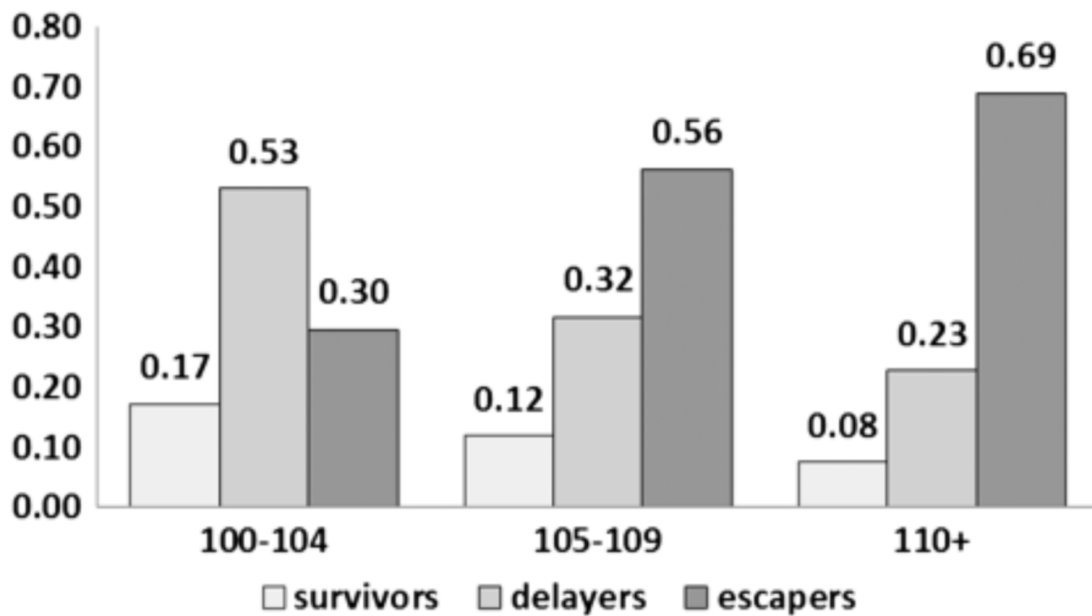


Figure 1.3. Proportion of individuals reaching extreme ages either by surviving, delaying or escaping morbidity in three age groups: centenarians (100-104years), semisupercentenarians (105-109 years) and supercentenarians (110+ years) [56]. “Survivors” refers to individuals with at least one disease with an onset of prior to age 80 years; “delayers”, at least one disease with an onset between age 80 and 90 years; escapers, at least one disease with an onset after age 100 years [56].

Epidemiological and demographical analyses indicate that natural variation in human longevity is diverse [57, 58]. While some studies showed that centenarians escape major diseases [59], other reports indicated that centenarians often suffer from multi-morbidity [57, 60, 61]. Hence, it has been suggested that humans can reach extreme ages either by surviving, delaying or escaping morbidity (**Figure 1.3**) [56, 58]. This indicates that there are multiple ways to achieve exceptional longevity which has implications on both the way how contrasting populations (case-control cohorts) are defined and on the approaches used to identify genetic variants that affect this trait, e.g., longitudinal, cross-sectional, genetic linkage, and association-based studies. Additionally, the genetic contribution to longevity was shown to increase with age; therefore, individuals with 100 years of age or older are likely to be the most informative subjects for investigating the genetics of human longevity [14].

Longitudinal studies in humans would represent the gold standard in human longevity research [62, 63]; however, securing the budget and retention of the participants within a long-term study are two major problems generally encountered with a longitudinal study design. Longitudinal longevity studies specifically are further hampered by the fact that the proportion of individuals that will reach exceptional longevity is too small to conduct statistically meaningful analyses [64]. Alternatively, genetic linkage (a family-based approach) and population-based association analyses have been largely employed in human longevity research [64].

Linkage analysis is a technique which is based on the fact that genes or DNA segments that are physically close to each other tend to be inherited together. For any trait with a genetic contribution, this enables the identification of chromosomal regions that co-segregate among affected family members using polymorphic markers that are evenly distributed throughout the genome [65]. In longevity research, linkage studies conducted on twins yielded the estimates for the longevity heritability in humans which was found to be rather modest, i.e. between 20% and 30% [66–68]. Additionally, with relative large sample sizes, at least ten longevity loci (3p22, 4q22-25, 8q23, 9q31-34, 12q24, 14q11, 17q12-q21, 17q22, 18q23-24, and 19p13) were identified by applying this study design [69–71].

Mutations in the genome can contribute to an increased or reduced disease or phenotype risk in a population. Genome-wide association studies (GWAS) help to identify mutations related to the phenotype of interest. The most commonly used type is the case-control study design, in which frequencies of genetic markers (e.g., single nucleotide polymorphisms (SNPs)) are compared between contrasting populations, e.g., healthy (controls) and sick (cases) individuals, to identify SNPs that significantly differ in frequency. Longevity GWAS can be divided into two major types depending on the evaluated phenotype. In the first (traditional) type, the genotype of each individual is determined and correlated with the phenotype of the individual itself. The second type focuses on parental longevity and evaluates the genotype of each individual correlated with the age or phenotype of the parents.

Since longevity-associated variants have barely reached genome-wide significance (GWS; P -value $< 5 \times 10E-08$), longevity is thought to depend on small-effect alleles [72–75]. Yashin et al. [76] genotyped 1,173 individuals and analyzed 169 putative SNPs for an association with longevity (P -value $< 1 \times 10E-06$). They demonstrated a significant and substantial joint influence of small-effect alleles on human lifespan. This finding has two major implications:

First, a relaxation of the GWS threshold in longevity is needed to identify the right trade-off between false-positive and biologically relevant signals; in other words, statistics needs to be adapted to the biological problem. As discussed later, this would also contribute to solving the so-called “missing-heritability problem” [77]. Second, complex traits like longevity might exhibit a non-additive (nonlinear) joint genetic influence (epistasis). Therefore, it has been proposed to analyze small-effect alleles, the so-called omnigenic model [78]. In this model it is taken into account that association signals tend to be spread across the genome and may include genes without an obvious connection to the trait under study. The omnigenic model differentiates between core and peripheral genes and accounts for the effect size of the genetic variants. The core genes are those which harbor mutations with strong effects (normally reaching GWS) and a biologically interpretable role in the trait studied. The peripheral genes include all genes with apparently no relevant effects on the trait (fail reaching GWS). Longevity likely is a result of many processes involving core genes such as *FOXO3* [42, 43, 79] and apolipoprotein E (*APOE*) [72, 80, 81] as well as other loci with only a suggestive association with longevity (P -value $< 5 \times 10E-06$).

The first longevity association dates back to the 1990s. Back then, the $\epsilon 4$ allele in the apolipoprotein E (*APOE*) gene was reported to be negatively associated with human longevity [80]. This finding derived from a candidate gene association study (CGAS), in which associations of variants in pre-defined loci with a trait, e.g., longevity, are investigated. *APOE* plays an important role in lipid metabolism in both peripheral tissues and the brain. The deleterious $\epsilon 4$ allele is associated with an increased risk and an earlier onset of various chronic degenerative diseases, like cardiovascular and Alzheimer’s disease [72, 80, 81]. Up to now, the negative longevity association of *APOE* $\epsilon 4$ has been confirmed in various populations.

In 2008, the *FOXO3*-longevity association was published; first reported in Japanese-Americans [42], and later confirmed in other populations including Germans [43, 82, 83]. The protein *FOXO3* is a key player in the IIS pathway [39, 52]. Activation of IIS correlates inversely with the expression of *FOXO3* and its targets, as upon IIS activation *FOXO3* is phosphorylated and excluded from the nucleus. In contrast, oxidative stress activates *FOXO3* by inducing its dephosphorylation and translocation to the nucleus where *FOXO3*, in its function as a transcription factor, enhances the transcription of stress-response genes, which aid in cell protection and maintenance. Increased *FOXO3* expression led to lifespan extension in *Hydra*, *C. elegans* and *D. melanogaster* [52, 84]. Many non-coding *FOXO3*-variants have been shown

to be associated with human longevity [74]. Some of them have already been analyzed in more depth with regard to functional implications and/or phenotypic associations. For example, the longevity-associated G allele of rs2802292 was reported to be associated with a higher peripheral and hepatic insulin sensitivity and increased *FOXO3* expression in skeletal muscle [85]. The minor allele (G) of rs12212067 was associated with an anti-inflammatory role [86] which was reflected in lower "inflammaging" in LLI [87]. In a recent study, the longevity-associated minor alleles of rs12206094 (T) and rs4946935 (A) both showed considerable enhancer activities in luciferase activity assays that were reversed by IGF-1 treatment. This may indicate an allele-specific response to nutritional stress. In the same report, the longevity alleles of both SNPs were shown to be associated with increased *FOXO3* expression in various tissues [44].

In order to extend the list of loci associated with human longevity, several large-scale meta-analyses have been conducted. One of these included 14 different studies with individuals of European descent [81]. They identified a novel locus, chromosome 5q33.3 (rs2149954, P -value $< 5 \times 10E-08$), to be associated with longevity. The T-allele of this single-nucleotide variant (SNV) had previously been reported to be associated with low blood pressure in middle age and to reduce the cardiovascular mortality risk (independent of the effects on blood pressure) [88, 89]. A meta-analysis on parental lifespan led to the discovery of two more regions: *HLA-DQA1/DRB1* (rs34831921) and *LPA* (rs55730499) [86]. Furthermore, in that study, previous longevity-associated loci were validated, i.e. *CHRNA3/5* (rs8042849), *CDKN2A/B* (rs1333049, rs4977756) and *SH2B3* (rs3184504) [90]. A disease-informed GWAS, in which knowledge from large studies of age-related diseases was integrated to narrow down the potential gene pool of longevity, identified genetic loci associated with exceptional human longevity including *APOE*, *CDKN2B/ANRIL*, *ABO* (tags the O blood group), and *SH2B3/ATXN2* [91]. This study also substantiated the genetic overlap between longevity and age-related diseases [92].

Despite the long time that has passed since the initial findings and the suggestive variants that have been discovered, up to now, only a handful variants have remained with a confirmed longevity association across populations: Next to *APOE* and *FOXO3*, this includes the locus on chromosome 5 (chromosome 5q33.3), and the variants in *CDKN2B*.

Despite the power of GWAS for the identification of new susceptibility loci, in longevity, the meta-analyses barely yielded novel genetic findings or overlap between studies. This

phenomenon has been partially attributed to inconsistent phenotype definitions across studies (especially with regard to the controls) [14, 93]. In addition, the individual longevity variants showed only weak effects which cannot reach GWS. Another surprising outcome from GWAS on complex traits was that GWAS-detected trait-associated SNVs explained only a small proportion of the total heritability (on average 5-10%). Furthermore, over 90% of the identified variants are located in non-coding regions, e.g., intronic or intergenic regions [14, 94–96]; this fact makes functional interpretations of GWAS findings highly challenging.

Hitherto, GWAS efforts have been mostly focused on discovering common genetic variants based on the common disease-common variant (CD/CV) hypothesis. However, from an evolutionary perspective, it is more likely that rare mutations have larger effects on common diseases [97]. Joint influences of large-effect (rare) and weak-effect (common) variants might explain the genetic susceptibility to both common complex diseases/phenotypes [98] and late-onset diseases. For example, the type 2 diabetes-associated loci *MTNR1B* was shown to harbor both rare and common variants [99]. Additionally, low-density lipoprotein levels were reported to be associated with common and rare variants in *APOE* and *NPC1L1* [100, 101]. Likewise, variation in blood pressure was linked to rare and common variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* [102]. Also in Alzheimer’s disease, common and rare SNV-burden was detected in genes such as *TREM2*, *CLU*, *SORL1*, *ABCA7*, *APP*, *PLD3*, *EPHA1*, *CR1*, and *BINI* [103]. Recently, a whole-exome sequencing of 100 long-lived individuals (≥ 98 years) was performed to identify rare variants with large effects associated with extreme longevity [104]. Unfortunately, due to the limited sample size, no SNV reached statistical significance; however, the authors suggested *LYST*, *MDN1*, and *RBMXL1*, which showed nominal significance, to be genes that harbor an increased burden of rare variants.

The evidence summarized above indicates that as a complex trait, longevity is very likely to be controlled by rare variants at many loci [105]; therefore extending the spectrum to rare susceptibility variants may help explain the missing heritability of longevity.

1.4. The human gut microbiota in aging and longevity

1.4.1. The human gut microbiota

The gut microbiome (GM) is the complex collection of microorganisms that inhabit our gastrointestinal (GI) tract; according to initial estimates, approximately 10^{13} to 10^{14} microorganisms reside inside the GI tract, exceeding the number of human cells in the body by

a factor ~10. Bacterial genes were reported to outnumber human genes even by a factor 100 [106]. However, in a more recent report, the stated values are much lower, but still impressive (~10¹³ microorganisms, ratio human:microbial genes 1:1) [107]. In a balanced healthy state, the relationship between the human host and his gut microbes is symbiotic; therefore, humans are sometimes considered "meta-" or "superorganisms" [108, 109]. Indeed, the microbiota contributes to human health in multiple ways. As one of the key functions, it plays a pivotal role in the development of the immune system and in balancing pro- and anti-inflammatory responses. The metabolic capacity perfectly complements the human enzymatic repertoire, e.g. by facilitating energy harvest from otherwise indigestible fibers. Furthermore, it helps maintain the integrity of the gut barrier and protects the host from colonization by pathogenic microorganisms [109].

1.4.2. The human gut microbiota composition during development and in adulthood

Aging is accompanied by changes in the gut microbiota composition. The colonization of the gut starts immediately after birth (or even already *in utero*), but microbial communities fluctuate particularly during the first three years of life until, after then, the structures stabilize to an adult-like status [110].

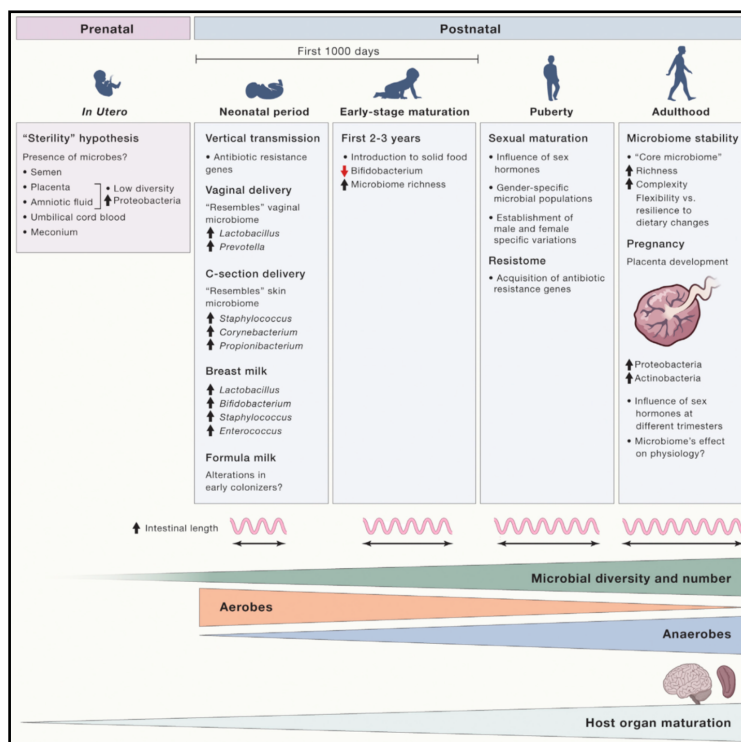


Figure 1.4. Gut microbiome changes from infancy to adulthood and factors influencing the microbial composition at the single stages ([111]).

Since the infant gut is initially an aerobic environment, facultative aerobe bacteria like *Escherichia* and *Enterococcus*, which derive from the mother and the environment, settle first. They induce a gradual change to more anaerobic conditions within a few days, which in turn allows a colonization by *Firmicutes* like *Clostridia*, *Bacteroidetes*, and *Bifidobacteria* [112]. During this process, the biodiversity of the GM increases in complexity and reaches a stable ecology structure at approximately three years of age [113]. Compared with the adult stage, childhood GM is still lower in microbial diversity, and characterized by both a low abundance of *Bacteroidetes* and higher abundances of *Firmicutes* and *Actinobacteria*. Species from the genus *Bifidobacterium*, the *Lactobacillus* group, and from the family *Coriobacteraceae* are also more abundant than in adulthood [114–116]. From a functional perspective, the childhood microbiota is characterized by an enrichment of genes that support growth, development, and production of vitamin B12 (124). In adulthood, the GM composition remains relatively stable. In healthy individuals, the dominating phyla are *Firmicutes* (families *Lachnospiraceae* and *Ruminococcaceae*), *Bacteroidetes* (*Bacteroidaceae*, *Prevotellaceae*, and *Rikenellaceae*), and *Actinobacteria* (*Bifidobacteriaceae* and *Corinobacteriaceae*). The families *Lachnospiraceae* (*Eubacterium rectale* group, 10 - 45%), *Ruminococcaceae* (*Clostridium leptum* group, 16 - 27%), and *Bacteroidaceae/Prevotellaceae* (12 - 60%) together account for 60 - 90% of all the faecal bacteria [113, 117–119]. Functionally, the adult GM is enriched in genes associated with dietary utilization [115].

Noteworthy, although only a few phyla dominate the adult human gastrointestinal tract, at the species level, these are composed of ~1,800 genera and ~16,000 phylotypes. Furthermore, every person carries specific and dynamic subsets of ~160 species [109, 120]. This highlights the complexity of any study, in which the GM composition and/or changes in GM composition in health and/or disease are targeted. The microbial diversity, i.e. the number of different microbial species in a community, generally serves as an indicator of a "healthy gut", owing this role to its association with productivity, functioning, and stability [121, 122]. Both a reduced microbial diversity and perturbations of the microbial balance, so-called dysbiosis, in the gut have been associated with various adverse health outcomes, including obesity, type 2 diabetes mellitus (T2D), inflammatory bowel disease, and other inflammatory and infectious diseases [108, 123, 124]. Also the changes in the microbial composition that take place during aging may affect host health, although it has remained controversial so far whether this remodelling is a cause (dysbiosis) or rather an adaptation to the aging processes [123].

However, maintaining both a high microbial diversity and a balanced gut microbial composition appear to be crucial in healthy aging and longevity.

1.4.3. Changes in the human gut microbiota composition in the elderly

The gastrointestinal tract as well as dietary patterns undergo substantial changes with aging, shifting GM gradually over time [125–127]. In the elderly, the GM diversity is reduced compared with younger adults, while the inter-individual variation is increased [126]. Part of the GM instability in old age has been associated with stress, antibiotics, lifestyle and diet [128], while some alterations have been linked to age-related diseases such as obesity, T2D, and cardiovascular diseases (CVD) [129–133]. In addition, shifts in the dominant species occur; the amount of beneficial microorganisms declines and both facultative anaerobic bacteria and subdominant bacterial species increase, such as members of *Bacteroidetes* and *Proteobacteria* over *Firmicutes* (*Clostridium* cluster XIV, *Fecalibacterium prausnitzii*, and *Bifidobacterium*) [126, 134–136]. Susceptibility to pathogenic infections has been associated with depletion of *Bifidobacterium* in older adults [137]. The proliferation of pathobionts (resident microbes with pathogenic potential) together with the lower microbial diversity is hypothesized to contribute to the pro-inflammatory status in the elderly and to be a risk factor for chronic health conditions [125, 127, 138]. Moreover, a lack of microbiota-derived metabolites (i.e. short-chain fatty acids, SSFA) due to the reduction in beneficial bacteria (*Clostridium* cluster XIV, *Fecalibacterium prausnitzii*, and *Bifidobacterium*, etc.) has been associated with aging-related disorders such as irregular bowel transit, reduced appetite, frailty, weight loss, cognitive decline, hypertension, vitamin D deficiency, diabetes, arthritis, and sarcopenia [132, 139, 140]. Altogether, these findings substantiate the hypothesis that maintaining GM homeostasis is crucial for healthy aging and for achieving longevity [141–144]. Therefore, restoration of GM homeostasis in elderly at risk is likely to become a main challenge in the future, especially in view of the ongoing population aging.

1.4.4. Signatures of longevity in the human gut microbiome

The most profound changes in GM have been seen in centenarians. One of the unequivocal age-related affliction is "inflammaging" (chronic low-grade inflammatory status [87]). In the gut, inflammation boosts aerobiosis and the production of reactive oxygen species, turning the gut into a disadvantageous environment for the strict anaerobic bacteria (e.g. *Firmicutes*) while uplifting the colonization of facultative aerobes (e.g. *Enterobacteraceae*, *Enterococcaceae*, *Staphylococcaceae*). These facultative microorganisms can compromise the immune

homeostasis in favor of a pro-inflammatory profile creating a vicious inflammatory cycle (as it has been demonstrated for mice [145]), which in turn may contribute to both frailty and the progression of diseases. Healthy LLI do not follow a linear pattern that continues to nourish the inflammatory cycle. Centenarians have been shown to exhibit a noticeable adaptive remodeling of the GM [135, 146, 147].

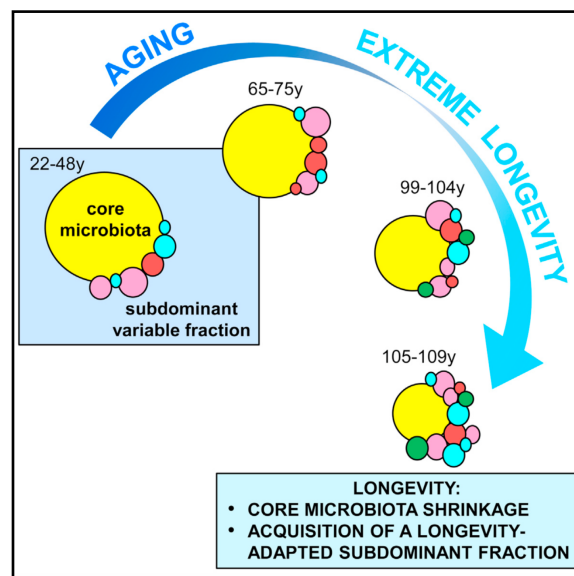


Figure 1.5. Proposed changes in the gut microbiota composition during aging and in extreme longevity ([146]).

Among the multiple characteristics of the centenarians' GM, one of the most prominent is the high microbial diversity (**Figure 1.5**) [136, 146, 147]. The increase in diversity entails a rearrangement of the core microbiota; for example, in centenarians from Italy, a reduction in the occurrence of the *Clostridium* cluster XIVa, an increment in *Bacillus* species and compositional changes in the *Clostridium* cluster IV [146] have been observed. The GM was also characterized by an increase in pathogenic species, such as *Proteobacteria*, *Fusobacterium*, *Bacillus*, *Staphylococcus*, *Corynebacterium*, and from members of the families *Ruminococcaceae*, *Christensenellaceae* and *Micrococcaceae* [135]. The proliferation of pathobionts like *Proteobacteria*, which may contribute to inflammaging, and a reduction in butyrate-producing bacteria, such as *Fecalibacterium*, *Roseburia*, *Coprococcus*, *Blautia* and *Eubacterium* have been reported in LLI from Italy, Japan and China [136, 146, 147]. This evidence has led to the hypothesis that the centenarians' GM displays a particular signature embedded in a microbial community shaped by lifestyle, nutrition, geographical/population or social factors as well as genetics, which still remains to be determined [148].

1.5. Aims of the thesis

We need a better understanding of the mechanisms underlying aging; only this knowledge will enable us to eventually intervene in aging-related biological processes, influence or manipulate them for the better and thereby extend the healthy lifespan of humans. As aging is a highly complex phenotype, the present work aims at contributing to the growing body of evidence about aging mechanisms on three very different levels (**Figure 1.6**):

1. Basic research in the immortal fresh water polyp *Hydra vulgaris* (Chapter I)
2. Genetic analyses focusing on the heritable component of human longevity, a healthy aging phenotype (Chapter II)
3. Investigations of the human intestinal microbiome and its role in healthy aging (Chapter III)

On all these levels, the work requires the handling, management and analysis of diverse big data sets (microarray and next-generation sequencing transcriptomics (Chapter I), exome-wide SNP arrays (Chapter II), microbiome 16S rRNA sequencing data along with host genetic and phenotypic information (Chapter III)) and the application of cutting-edge statistical and computational methods.

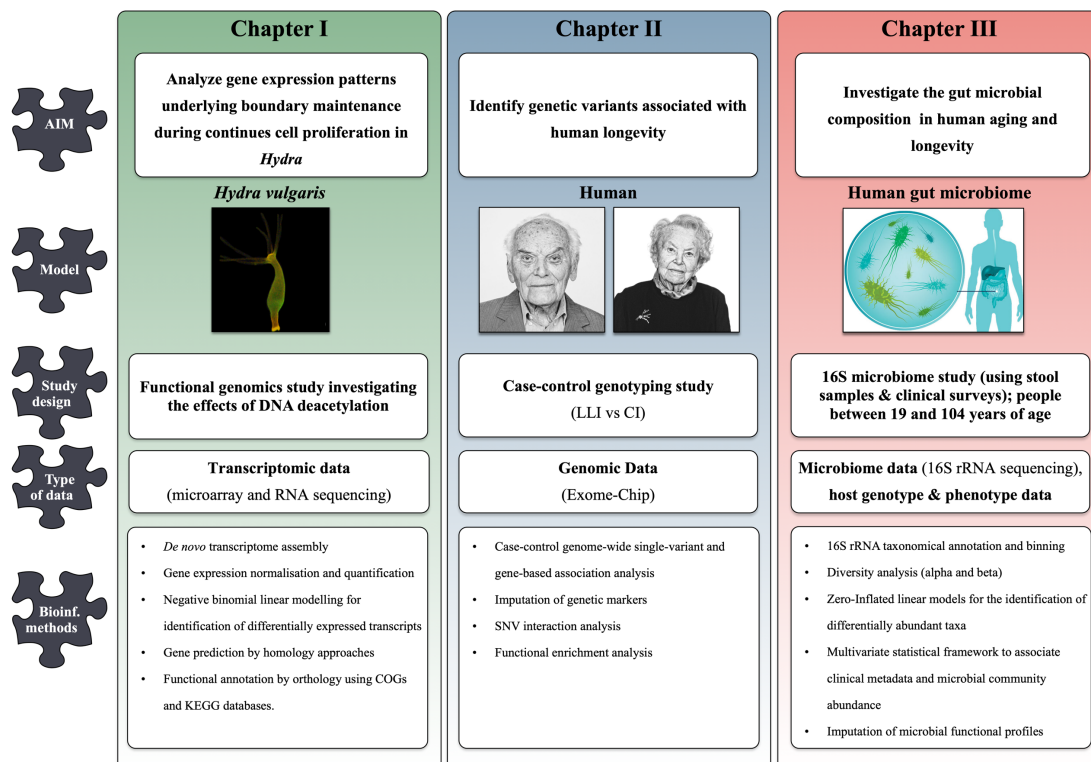


Figure 1.6. Overview of the three research projects.

CI, control individuals; COG, cluster of orthologous groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; LLI, long-lived individuals; SNV, single nucleotide variant.

1.5.1. Specific aims

Chapter I

The *Hydra* chapter comprises fundamental research on the genetically driven regulation of development and aging. Many genes and pathways are evolutionarily conserved and, consequently, a large number of orthologs of human genes have been found in the *Hydra* genome; FoxO is only one prominent example. Therefore, basic research in *Hydra* might contribute in the long-term to unravel evolutionary signatures of aging regulation, which may help to understand human aging.

Specifically, molecular mechanisms underlying *Hydra* body compartmentalization, i.e. boundary formation and maintenance under constant proliferation, and the genetic contribution to these processes have remained largely unknown so far. Therefore, the aims of *Chapter I* are to:

1. define the gene expression profile of the body parts of *Hydra* (foot, body, head and tentacles) using RNAseq techniques;
2. subsequently determine by differential gene expression analyses the degree to which gene expression changes underlie the compartmentalization;
3. identify which patterns in gene expression are altered when *Hydra*'s epigenome is disrupted employing a histone acetylation inhibitor.

Eventually, the work aims to discover the regulatory mechanism(s) underlying the regionalization processes as well as to investigate a possible involvement of FoxO.

Chapter II

Despite extensive research efforts, there is still a huge knowledge gap about the genes that specifically contribute to human longevity – so far, only a handful of SNVs have shown consistent longevity-associations across population-cohorts. It appears plausible that apart from common susceptibility variants also (relatively) rare SNVs with large effect sizes that reside in coding regions may at least partly explain the missing longevity heritability. Previous studies focused almost exclusively on intronic or regulatory SNVs with limited success. In this subproject, exonic variants are investigated. This is the first study of this kind in Germans. An

exome-wide chip-based SNV case-control association analysis is conducted in a large cohort comprising more than 1,200 LLI including 599 centenarians. This project also involves the replication of results in Danish and French cohorts and an in-depth analysis of the genotyping data with bioinformatic methods.

Chapter III

The intestinal microbial composition changes during aging. However, these changes appear to depend on dietary patterns and/or geographic region. No systematic large-scale study over a broad age range, including LLI, has been conducted in Germans. Therefore, the aims of this subproject are to:

1. identify age-related changes in the microbiome composition in healthy German individuals covering the age range between 19 and 104 years;
2. determine to what extent host genetics (associated with gut microbial composition and metabolic syndromes) as well as environmental factors influence aging;
3. investigate whether there are any specific microbial patterns in German LLI.

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; 296:1029-1031.
2. Fogel RW, Costa DL. A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography* 1997; 34:49-66.
3. Riley JC. Estimates of regional and global life expectancy, 1800–2001. *Population and development review* 2005; 31:537-543.
4. Crimmins EM, Beltrán-Sánchez H. Mortality and morbidity trends: is there compression of morbidity. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 2011; 66:75-86.
5. Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. 1971. *Milbank Q.* 2005; 83:731-757.
6. Barbi E, Lagona F, Marsili M, Vaupel JW, Wachter KW. The plateau of human mortality: Demography of longevity pioneers. *Science* 2018; 360:1459-1461.
7. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature* 2018; 561:45-56.
8. Kennedy BK, Berger SL, Brunet A, et al. Geroscience: linking aging to chronic disease. *Cell* 2014; 159:709-713.
9. Jagger C, Gillies C, Moscone F, Cambois E, Van Oyen H, Nusselder W, Robine JM, EHLEIS team. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* 2008; 372:2124-2131.
10. Crimmins EM. Lifespan and healthspan: past, present, and promise. *Gerontologist* 2015; 55:901-911.
11. Newman SJ. Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biol.* 2018; 16:e2006776.
12. Olshansky SJ, Passaro DJ, Hershow RC, Layden J, Carnes BA, Brody J, Hayflick L, Butler RN, Allison DB, Ludwig DS. A potential decline in life expectancy in the United States in the 21st century. *N. Engl. J. Med.* 2005; 352:1138-1145.
13. Hubert HB, Bloch DA, Oehlert JW, Fries JF. Lifestyle habits and compression of morbidity. *J. Gerontol. A Biol. Sci. Med. Sci.* 2002; 57:M347-51.
14. Giuliani C, Garagnani P, Franceschi C. Genetics of human longevity within an eco-evolutionary nature-nurture framework. *Circ. Res.* 2018; 123:745-772.
15. Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, Monti D, Capri M, Salvioli S. The continuum of aging and age-related diseases: common mechanisms but different rates. *Front. Med.* 2018; 5:61.
16. Partridge L, Mangel M. Messages from mortality: the evolution of death rates in the old. *Trends Ecol. Evol.* 1999; 14:438-442.
17. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell* 2013; 153:1194-1217.
18. De Magalhães JP, Passos JF. Stress, cell senescence and organismal ageing. *Mech. Ageing Dev.* 2018; 170:2-9.
19. Geiger H, Van Zant G. The aging of lympho-hematopoietic stem cells. *Nat. Immunol.* 2002; 3:329-333.

20. Mattson MP, Duan W, Maswood N. How does the brain control lifespan. *Ageing Res. Rev.* 2002; 1:155-165.
21. Zhang G, Li J, Purkayastha S, Tang Y, Zhang H, Yin Y, Li B, Liu G, Cai D. Hypothalamic programming of systemic ageing involving IKK- β , NF- κ B and GnRH. *Nature* 2013; 497:211-216.
22. Rera M, Azizi MJ, Walker DW. Organ-specific mediation of lifespan extension: more than a gut feeling. *Ageing Res. Rev.* 2013; 12:436-444.
23. De Magalhães JP. The scientific quest for lasting youth: prospects for curing aging. *Rejuvenation Res.* 2014; 17:458-467.
24. Ayyadevara S, Alla R, Thaden JJ, Shmookler Reis RJ. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging Cell* 2008; 7:13-22.
25. Kurosu H, Yamamoto M, Clark JD, et al. Suppression of aging in mice by the hormone Klotho. *Science* 2005; 309:1829-1833.
26. Coschigano KT, Holland AN, Riders ME, List EO, Flyvbjerg A, Kopchick JJ. Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. *Endocrinology* 2003; 144:3799-3810.
27. Brown-Borg HM, Borg KE. Dwarf mice and the ageing process. *Nature* 1996; 384:33.
28. Mattison JA, Roth GS, Beasley TM, et al. Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. *Nature* 2012; 489:318-321.
29. Colman RJ, Beasley TM, Kemnitz JW, Johnson SC, Weindruch R, Anderson RM. Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nat. Commun.* 2014; 5:3557.
30. Mattison JA, Colman RJ, Beasley TM, Allison DB, Kemnitz JW, Roth GS, Ingram DK, Weindruch R, de Cabo R, Anderson RM. Caloric restriction improves health and survival of rhesus monkeys. *Nat. Commun.* 2017; 8:14063.
31. Redman L, Smith SR, Burton JH, Martin CK, Il'yasova D, Ravussin E. Metabolic slowing and reduced oxidative damage with sustained caloric restriction support the rate of living and oxidative damage theories of aging. *Cell Metabol.* 2018; 1-11.
32. De Magalhães JP. Species selection in comparative studies of aging and antiaging research. In: Conn PM, ed. *Handbook of Models for Human Aging*. Burlington, MA, 9-20; Elsevier Academic Press, 2006.
33. Martínez DE. Mortality patterns suggest lack of senescence in hydra. *Exp. Gerontol.* 1998; 33:217-225.
34. Yoshida K, Fujisawa T, Hwang JS, Ieko K, Gojobori T. Degeneration after sexual differentiation in hydra and its relevance to the evolution of aging. *Gene* 2006; 385:64-70.
35. Mortzfeld BM, Taubenheim J, Fraune S, Klimovich AV, Bosch TCG. Stem cell transcription factor FoxO controls microbiome resilience in *Hydra*. *Front. Microbiol.* 2018; 9:629.
36. Van Zant G, Liang Y. The role of stem cells in aging. *Exp. Hematol.* 2003; 31:659-672.
37. Jung Y, Brack AS. Cellular mechanisms of somatic stem cell aging. eds. *Current Topics in Developmental Biology: Stem Cells in Development and Disease*. Elsevier, 2014: 405-438.
38. Bridge D, Theofiles AG, Holler RL, Marcinkevicius E, Steele RE, Martínez DE. FoxO

and stress responses in the cnidarian *Hydra vulgaris*. PLoS One. 2010; 5:e11686.

39. Morris BJ, Willcox DC, Donlon TA, Willcox BJ. FOXO3: a major gene for human longevity - a mini-review. *Gerontology* 2015; 61:515-525.
40. Webb AE, Kundaje A, Brunet A. Characterization of the direct targets of FOXO transcription factors throughout evolution. *Aging Cell* 2016; 15:673-685.
41. Boehm A-M, Khalturin K, Anton-Erxleben F, Hemmrich G, Klostermeier UC, Lopez-Quintero JA, Oberg H-H, Puchert M, Rosenstiel P, Wittlieb J, Bosch TCG. FoxO is a critical regulator of stem cell maintenance in immortal *Hydra*. *Proc. Natl Acad. Sci. U S A* 2012; 109:19697-19702.
42. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, Masaki KH, Willcox DC, Rodriguez B, Curb JD. FOXO3A genotype is strongly associated with human longevity. *Proc. Natl Acad. Sci. U S A* 2008; 105:13987-13992.
43. Flachsbart F, Caliebe A, Kleindorp R, Blanché H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl Acad. Sci. U S A*. 2009; 106:2700-2705.
44. Flachsbart F, Dose J, Gentschew L, et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat. Commun.* 2017; 8:2063.
45. Wenger Y, Galliot B. Punctuated emergences of genetic and phenotypic innovations in eumetazoan, bilaterian, euteleostome, and hominidae ancestors. *Genome Biol. Evol.* 2013; 5:1949-1968.
46. Tomczyk S, Fischer K, Austad S, Galliot B. *Hydra*, a powerful model for aging studies. *Invertebr. Reprod. Dev.* 2015; 59:11-16.
47. Nebel A, Bosch TC. Evolution of human longevity: lessons from *Hydra*. *Aging (Albany NY)* 2012; 4:730-731.
48. De Magalhães JP. Is mammalian aging genetically controlled. *Biogerontology* 2003; 4:119-120.
49. Seluanov A, Gladyshev VN, Vijg J, Gorbunova V. Mechanisms of cancer resistance in long-lived mammals. *Nat. Rev. Cancer* 2018; 18:433-441.
50. Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhães JP. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* 2013; 41:D1027-33.
51. Fontana L, Partridge L, Longo VD. Extending healthy life span—from yeast to humans. *Science* 2010; 328:321-326.
52. Kenyon CJ. The genetics of ageing. *Nature* 2010; 464:504-512.
53. De Castro E, Hegi de Castro S, Johnson TE. Isolation of long-lived mutants in *Caenorhabditis elegans* using selection for resistance to juglone. *Free Radic. Biol. Med.* 2004; 37:139-145.
54. Landis GN, Bhole D, Tower J. A search for doxycycline-dependent mutations that increase *Drosophila melanogaster* life span identifies the VhaSFD, Sugar baby, filamin, fwd and Cctl genes. *Genome Biol.* 2003; 4:R8.
55. Paik D, Jang YG, Lee YE, Lee YN, Yamamoto R, Gee HY, Yoo S, Bae E, Min KJ, Tatar M, Park JJ. Misexpression screen delineates novel genes controlling *Drosophila* lifespan. *Mech. Ageing Dev.* 2012; 133:234-245.

56. Andersen SL, Sebastiani P, Dworkis DA, Feldman L, Perls TT. Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. *J. Gerontol. A Biol. Sci. Med. Sci.* 2012; 67:395-405.
57. Andersen-Ranberg K, Schroll M, Jeune B. Healthy centenarians do not exist, but autonomous centenarians do: a population-based study of morbidity among Danish centenarians. *J. Am. Geriatr. Soc.* 2001; 49:900-908.
58. Evert J, Lawler E, Bogan H, Perls T. Morbidity profiles of centenarians: survivors, delayers, and escapers. *J. Gerontol. A Biol. Sci. Med. Sci.* 2003; 58:232-237.
59. Hitt R, Young-Xu Y, Silver M, Perls T. Centenarians: the older you get, the healthier you have been. *Lancet* 1999; 354:652.
60. Berzlanovich AM, Keil W, Waldhoer T, Sim E, Fasching P, Fazyen-Dorner B. Do centenarians die healthy? An autopsy study. *J. Gerontol. A Biol. Sci. Med. Sci.* 2005; 60:862-865.
61. Aramillo Irizar P, Schäuble S, Esser D, et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat. Commun.* 2018; 9:327.
62. Ferrucci L. The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future. *J. Gerontol. A Biol. Sci. Med. Sci.* 2008; 63:1416-1419.
63. Elliott P, Peakman TC, UK B. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* 2008; 37:234-244.
64. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat. Rev. Genet.* 2006; 7:436-448.
65. Pulst SM. Genetic linkage analysis. *Arch. Neurol.* 1999; 56:667-672.
66. McGue M, Vaupel JW, Holm N, Harvald B. Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *J. Gerontol.* 1993; 48:B237-44.
67. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum. Genet.* 1996; 97:319-323.
68. Mitchell BD, Hsueh WC, King TM, Pollin TI, Sorkin J, Agarwala R, Schäffer AA, Shuldiner AR. Heritability of life span in the Old Order Amish. *Am. J. Med. Genet.* 2001; 102:346-352.
69. Boyden SE, Kunkel LM. High-density genomewide linkage analysis of exceptional human longevity identifies multiple novel loci. *PLoS One.* 2010; 5:e12432.
70. Kerber RA, O'Brien E, Boucher KM, Smith KR, Cawthon RM. A genome-wide study replicates linkage of 3p22-24 to extreme longevity in humans and identifies possible additional loci. *PLoS One.* 2012; 7:e34746.
71. Beekman M, Blanché H, Perola M et al. Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. *Aging Cell* 2013; 12:184-193.
72. Nebel A, Kleindorp R, Caliebe A, et al. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* 2011; 132:324-330.
73. Newman AB, Walter S, Lunetta KL, et al. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging

Research in Genomic Epidemiology Consortium. *J. Gerontol. A Biol. Sci. Med. Sci.* 2010; 65:478-487.

74. Broer L, Buchman AS, Deelen J, et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. A Biol. Sci. Med. Sci.* 2015; 70:110-118.

75. Deelen J, Evans DS, Arking DE, Nygaard M, et al. A meta-analysis of genome-wide association studies of longevity identifies ApoE ϵ 4 and ϵ 2 associations. *Nat. Comm.* 2019; 10:3669

76. Yashin AI, Wu D, Arbeevev KG, Ukraintseva SV. Joint influence of small-effect genetic variants on human longevity. *Aging (Albany NY)* 2010; 2:612-620.

77. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* 2016; 99:139-153.

78. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 2017; 169:1177-1186.

79. Soerensen M, Dato S, Christensen K, McGue M, Stevnsner T, Bohr VA, Christiansen L. Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. *Aging Cell* 2010; 9:1010-1017.

80. Schächter F, Faure-Delanef L, Guénot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D. Genetic associations with human longevity at the APOE and ACE loci. *Nat. Genet.* 1994; 6:29-32.

81. Deelen J, Beekman M, Uh HW, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* 2014; 23:4420-4432.

82. Anselmi CV, Malovini A, Roncarati R, Novelli V, Villa F, Condorelli G, Bellazzi R, Puca AA. Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. *Rejuvenation Res.* 2009; 12:95-104.

83. Li Y, Wang W-J, Cao H, et al. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Hum. Mol. Genet.* 2009; 18:4897-4904.

84. Martins R, Lithgow GJ, Link W. Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. *Aging Cell* 2016; 15:196-207.

85. Banasik K, Ribbel-Madsen R, Gjesing AP, Wegner L, Andersson A, Poulsen P, Borglykke A, Witte DR, Pedersen O, Hansen T, Vaag A. The FOXO3A rs2802292 G-allele associates with improved peripheral and hepatic insulin sensitivity and increased skeletal muscle-FOXO3A mRNA expression in twins. *J. Clin. Endocrinol. Metab.* 2011; 96:E119-24.

86. Lee JC, Espéli M, Anderson CA, et al. Human SNP links differential outcomes in inflammatory and infectious disease to a FOXO3-regulated pathway. *Cell* 2013; 155:57-69.

87. Franceschi C, Capri M, Monti D, Giunta S, Olivieri F, Sevini F, Panourgia MP, Invidia L, Celani L, Scurti M, Cevenini E, Castellani GC, Salvioli S. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech. Ageing Dev.* 2007; 128:92-105.

88. Molander L, Lövheim H, Norman T, Nordström P, Gustafson Y. Lower systolic blood pressure is associated with greater mortality in people aged 85 and older. *J. Am. Geriatr. Soc.* 2008; 56:1853-1859.

89. Oates DJ, Berlowitz DR, Glickman ME, Silliman RA, Borzecki AM. Blood pressure and survival in the oldest old. *J. Am. Geriatr. Soc.* 2007; 55:383-388.

90. Joshi PK, Pirastu N, Kentistou KA, et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* 2017; 8:910.
91. Fortney K, Dobriban E, Garagnani P, Pirazzini C, Monti D, Mari D, Atzmon G, Barzilai N, Franceschi C, Owen AB, Kim SK. Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. *PLoS Genet.* 2015; 11:e1005728.
92. Jeck WR, Siebold AP, Sharpless NE. Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell* 2012; 11:727-731.
93. Sebastiani P, Bae H, Gurinovich A, Soerensen M, Puca A, Perls TT. Limitations and risks of meta-analyses of longevity studies. *Mech. Ageing Dev.* 2017; 165:139-146.
94. Maher B. Personal genomes: the case of the missing heritability. *Nature* 2008; 456:18-21.
95. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. U S A* 2009; 106:9362-9367.
96. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461:747-753.
97. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000; 156:297-304.
98. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* 2009; 19:212-219.
99. Bonnefond A, Clément N, Fawcett K et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* 2012; 44:297-301.
100. Benn M, Stene MC, Nordestgaard BG, Jensen GB, Steffensen R, Tybjaerg-Hansen A. Common and rare alleles in apolipoprotein B contribute to plasma levels of low-density lipoprotein cholesterol in the general population. *J. Clin. Endocrinol. Metab.* 2008; 93:1038-1045.
101. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc. Natl Acad. Sci. U S A* 2006; 103:1810-1815.
102. Ji W, Foo JN, O’Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.* 2008; 40:592-599.
103. Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer’s disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* 2016; 15:857-868.
104. Nygaard HB, Erson-Omay EZ, Wu X, Kent BA, Bernales CQ, Evans DM, Farrer MJ, Vilariño-Güell C, Strittmatter SM. Whole exome sequencing of an exceptional longevity cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* 2018 [Epub ahead of print].
105. Sebastiani P, Solovieff N, Dewan AT, et al. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; 7:e29848.
106. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. *Science* 2006; 312:1355-1359.
107. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria

cells in the Body. *PLoS Biol.* 2016; 14:e1002533.

108. Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem. J.* 2017; 474:1823-1836.
109. Ottaviani E, Ventura N, Mandrioli M, Candela M, Franchini A, Franceschi C. Gut microbiota as a candidate for lifespan extension: an ecological/evolutionary perspective targeted on living organisms as metaorganisms. *Biogerontology* 2011; 12:599-609.
110. Nagpal R, Mainali R, Ahmadi S, Wang S, Singh R, Kavanagh K, Kitzman DW, Kushugulova A, Marotta F, Yadav H. Gut microbiome and aging: physiological and mechanistic insights. *Nutr. Healthy Aging* 2018; 4:267-285.
111. Kundu P, Blacher E, Elinav E, Pettersson S. Our gut microbiome: the evolving inner self. *Cell* 2017; 171:1481-1493.
112. Houghteling PD, Walker WA. Why is initial bacterial colonization of the intestine important to infants' and children's health. *J. Pediatr. Gastroenterol. Nutr.* 2015; 60:294-307.
113. Yatsunenkov T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature* 2012; 486:222-227.
114. Agans R, Rigsbee L, Kenche H, Michail S, Khamis HJ, Paliy O. Distal gut microbiota of adolescent children is different from that of adults. *FEMS Microbiol. Ecol.* 2011; 77:404-412.
115. Hollister EB, Riehle K, Luna RA, et al. Structure and function of the healthy pre-adolescent pediatric gut microbiome. *Microbiome* 2015; 3:36.
116. Ringel-Kulka T, Cheng J, Ringel Y, Salojärvi J, Carroll I, Palva A, de Vos WM, Satokari R. Intestinal microbiota in healthy U.S. young children and adults—a high throughput microarray analysis. *PLoS One* 2013; 8:e64315.
117. Sghir A, Gramet G, Suau A, Rochet V, Pochart P, Dore J. Quantification of bacterial groups within human fecal flora by oligonucleotide probe hybridization. *Appl. Environ. Microbiol.* 2000; 66:2263-2266.
118. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. *Science* 2005; 308:1635-1638.
119. Maukonen J, Saarela M. Human gut microbiota: does diet matter. *Proc. Nutr. Soc.* 2015; 74:23-36.
120. Rajilić-Stojanović M, Smidt H, de Vos WM. Diversity of the human gastrointestinal tract microbiota revisited. *Environ. Microbiol.* 2007; 9:2125-2136.
121. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. *BMJ* 2018; 361:k2179.
122. Reese AT, Dunn RR. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. *MBio* 2018; 9.
123. Saraswati S, Sitaraman R. Aging and the human gut microbiota—from correlation to causality. *Front. Microbiol.* 2014; 5:764.
124. Harris VC, Haak BW, Boele van Hensbroek M, Wiersinga WJ. The intestinal microbiome in infectious diseases: the clinical relevance of a rapidly emerging field. *Open Forum Infect. Dis.* 2017; 4:ofx144.
125. O'Toole PW, Claesson MJ. Gut microbiota: changes throughout the lifespan from

infancy to elderly. *Internat. Dairy J.* 2010; 20:281-291.

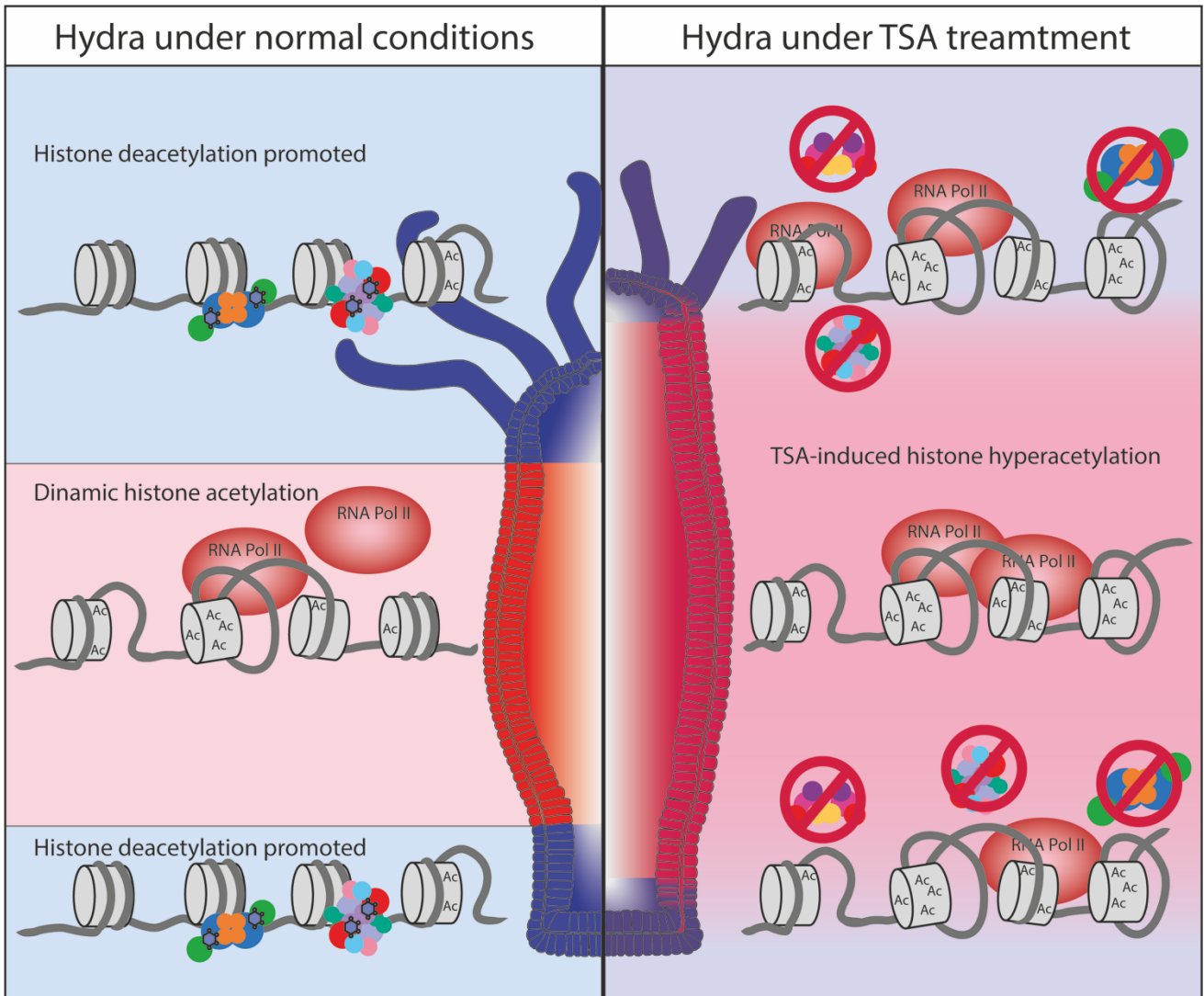
126. Claesson MJ, Cusack S, O'Sullivan O, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl Acad. Sci. U S A.* 2011; 108 Suppl 1:4586-4591.
127. Claesson MJ, Jeffery IB, Conde S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* 2012; 488:178-184.
128. Zhang YJ, Li S, Gan RY, Zhou T, Xu DP, Li HB. Impacts of gut bacteria on human health and diseases. *Int. J. Mol. Sci.* 2015; 16:7493-7519.
129. Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH, Jakobsen M. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* 2010; 5:e9085.
130. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012; 490:55.
131. Damms-Machado A, Mitra S, Schollenberger AE, Kramer KM, Meile T, Königsrainer A, Huson DH, Bischoff SC. Effects of surgical and dietary weight loss therapy for obesity on gut microbiota composition and nutrient absorption. *Biomed. Res. Int.* 2015; 2015:806248.
132. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006; 444:1027-1031.
133. Tang WH, Kitai T, Hazen SL. Gut microbiota in cardiovascular health and disease. *Circ. Res.* 2017; 120:1183-1196.
134. Enck P, Zimmermann K, Rusch K, Schwartz A, Klosterhalfen S, Frick JS. The effects of ageing on the colonic bacterial microflora in adults. *Z Gastroenterol.* 2009; 47:653-658.
135. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, Nikkila J, Monti D, Satokari R, Franceschi C, Brigidi P, De Vos W. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* 2010; 5:e10667.
136. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao JZ, Abe F, Osawa R. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol.* 2016; 16:90.
137. Leung K, Thuret S. Gut Microbiota: A Modulator of brain plasticity and cognitive function in ageing. *Healthcare (Basel)* 2015; 3:898-916.
138. Jackson MA, Jackson M, Jeffery IB, Beaumont M, Bell JT, Clark AG, Ley RE, O'Toole PW, Spector TD, Steves CJ. Signatures of early frailty in the gut microbiota. *Genome Med.* 2016; 8:8.
139. Rampelli S, Candela M, Turrone S, Biagi E, Collino S, Franceschi C, O'Toole PW, Brigidi P. Functional metagenomic profiling of intestinal microbiome in extreme ageing. *Ageing (Albany NY)* 2013; 5:902-912.
140. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilán CG, Salazar N. Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 2016; 7:185.
141. Collino S, Montoliu I, Martin FP, Scherer M, Mari D, Salvioli S, Bucci L, Ostan R, Monti D, Biagi E, Brigidi P, Franceschi C, Rezzi S. Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. *PLoS One* 2013; 8:e56564.

- 142.** Martin-Montalvo A, Mercken EM, Mitchell SJ, et al. Metformin improves healthspan and lifespan in mice. *Nat. Commun.* 2013; 4:2192.
- 143.** Biagi E, Candela M, Turrone S, Garagnani P, Franceschi C, Brigidi P. Ageing and gut microbes: perspectives for health maintenance and longevity. *Pharmacol. Res.* 2013; 69:11-20.
- 144.** Han B, Sivaramakrishnan P, Lin CJ, Neve IAA, He J, Tay LWR, Sowa JN, Sizovs A, Du G, Wang J, Herman C, Wang MC. Microbial genetic composition tunes host longevity. *Cell* 2017; 169:1249-1262.e13.
- 145.** Thevaranjan N, Puchta A, Schulz C, et al. Age-associated microbial dysbiosis promotes intestinal permeability, systemic inflammation, and macrophage dysfunction. *Cell Host Microbe* 2017; 21:455-466.e4.
- 146.** Biagi E, Franceschi C, Rampelli S, Severgnini M, Ostan R, Turrone S, Consolandi C, Quercia S, Scurti M, Monti D, Capri M, Brigidi P, Candela M. Gut microbiota and extreme longevity. *Curr. Biol.* 2016; 26:1480-1485.
- 147.** Kong F, Hua Y, Zeng B, Ning R, Li Y, Zhao J. Gut microbiota signatures of longevity. *Curr. Biol.* 2016; 26:R832-R833.
- 148.** Santoro A, Ostan R, Candela M, Biagi E, Brigidi P, Capri M, Franceschi C. Gut microbiota changes in the extreme decades of human life: a focus on centenarians. *Cell Mol. Life Sci.* 2018; 75:129-14

2.

**Chapter I: Boundary maintenance in the
ancestral metazoan *Hydra* depends on
histone acetylation**

Graphical abstract



This manuscript was submitted to *Developmental Cell* (status: major revision)

Boundary maintenance in the ancestral metazoan *Hydra* depends on histone acetylation

Javier A. López-Quintero^{1†}, Guillermo G. Torres^{2†}, Rafik Neme^{3,4}, Thomas C.G. Bosch^{1*}

¹ Zoological Institute, Kiel University, Germany

² Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Straße 12 24105 Kiel, Germany

³ Max-Planck Institute for Evolutionary Biology, 24306 Plön, Germany.

⁴ Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 1212 Amsterdam Avenue, New York, NY 10027, USA

† these authors contributed equally to this work

* **Correspondence**

tbosch@zoologie.uni-kiel.de

Abstract

Despite being a defining feature of many animal taxa, much remains to be understood about boundary formation during development. Axial patterning of *Hydra*, member of the ancient phylum Cnidaria which diverged prior to the bilaterian radiation, involves a steady state of production and loss of tissue and is dependent on an organizer located in the upper part of the head. We show that the sharp boundary which separates tissue in the body column from head and foot tissue depends on histone acetylation. Histone deacetylation decomposes the boundary by affecting numerous developmental genes including Wnt components and prevents stem cells from entering the position dependent differentiation program. Overall, our results suggest that reversible histone acetylation is an ancient regulatory mechanism for partitioning the body axis into domains with a specific identity, which has been present in the common ancestor of cnidarians and bilaterians, at least 600 million years ago.

2.1. Introduction

Boundary formation during animal development is important to establish tissue with regional identity along the body axis (Batlle and Wilkinson, 2012; Cayuso et al., 2015). Although destabilization of boundaries by cell intercalation or physical disruption and dispersal during morphogenesis directly interferes with normal development (Dahmann et al., 2011a), the molecular principles by which boundaries are maintained throughout the life of an organism have remained undiscovered.

In *Hydra* (**Figure 2.1 A**), cell proliferation takes place continuously along the single apical-basal body axis but is restricted to the gastric region of the animal (**Figure 2.1 B**) (Campbell and David, 1974; David and Campbell, 1972). The division of stem cells in this region causes a constant displacement of differentiated cells towards extremities (head and foot) and buds (Campbell, 1967). Due to these tissue dynamics, patterning signals are constantly active in the adults, which not only allows polyps to proliferate by budding continuously but also is constantly challenging the formation of stable boundaries. Such boundaries are those who strictly separate the stem cell populations in the body column from the committed or terminally differentiated cells in the head and foot compartments (**Figure 2.1 B-C**).

The morphologically invisible boundary which separates body column and head tissue correlates with distinct expression domains of both transcription factors and signalling as well as effector molecules (**Figure 2.1, D-H**) (Boehm et al., 2012; Endl et al., 1999b; Weinziger et al., 1994; for review see Böttger and Hassel, 2012). Transcriptional regulators potentially involved in establishing regional identity along the body axis include CnOtx, HyAlx, Budhead and Prdl-a (Gauchat et al., 1998; Martinez et al., 1997; Smith et al., 1999, 2000). Genes demarcating the basal end of the tentacles and thus the boundary between head and body column tissue include HyDsh, HvWnt8 and Hmfz2 (Minobe et al., 2000; Philipp et al., 2009). *Hydra* specific effector gene *ks1* (Endl et al., 1999b) is strongly expressed in ectodermal epithelial cells immediately after they leave the body column compartment and before they terminally differentiate to become head specific “battery” epithelial cells (**Figure 2.1 A and D**).

The molecular mechanisms maintaining stable boundaries in the constantly proliferating *Hydra* tissue remain largely unknown. It is also not known to what extent changes in gene expression underlie the regionalization, nor do we understand how these changes are regulated. The

mechanistic nature of the control system that specifically sets the head-body column boundary along the body axis is therefore still unknown. Towards this goal, we here combined genetic, cell biological, molecular and computational approaches. We find that the molecular circuitry at the sharp boundary which separates tissue in the body column from head tissue deploys numerous developmental genes and depends on histone acetylation. Histone deacetylation decomposes the boundary and transforms head tissue into body column-like tissue with constant mitotic activity. Overall, our results suggest a model in which partitioning of the *Hydra* body axis into domains with specific identity involves epigenetic programming mediated by histone acetylation.

2.2. Results

2.2.1. Comparative transcriptomics identifies compartment specific genes

To uncover the encoded genetic logic underlying the maintenance of region-specific compartments in adult *Hydra*, we profiled the transcriptome from six defined regions along the body column (**Figure 2.1 C**). The assembly of 256702 contigs led to the identification of 29358 transcripts. As a quality threshold and to minimize variability, we concentrated on 14127 genes corresponding to transcripts which appeared in at least three libraries with at least two counts per million reads. Three independent biological replicates for each region showed high reproducibility (**Figure 2.2 A**). Principle component analysis (PCA) revealed that component 1 and 2 (PC1, PC2) were sufficient for separating the samples into four defined groups: tentacle, head, foot, and all the samples from the three body parts (B1, B2 and Bb) (**Figure 2.2 A**). The first component (PC1) distinguished body column expressed genes from head and tentacle specific transcripts and accounted for 19 % of the variability. The largest differences in this dataset were between the tentacles and the body region (**Figure 2.2 A**). These two regions exhibited 2199 differentially regulated genes. The 9 percent of them were upregulated, and 6.5 percent downregulated (**Supplementary figure 2.1**). When comparing head and body column tissue, from 1022 differentially expressed genes 4.5% were up and 2.6% downregulated. Moreover, from 967 genes differentially expressed in tentacle and foot tissue, 4% were up and 3% downregulated suggesting a quite distinct transcriptome profile for these two compartments. The tentacle and head programs were relatively similar to each other with less than 1% up or downregulated genes from a total of 137 genes.

2.2.2. Functional annotations outline the compartment specific gene expression program

One goal of the study was to delineate a region-specific gene expression profile along *Hydra*'s body column to understand the mechanistic underpinnings. As a first step, we used the differentially expressed genes that were expressed at least 2-fold higher in region-specific samples and plotted them in the PCA correlation circle following their annotation in KEGG pathways (<http://www.genome.ad.jp/kegg/>) (**Figure 2.2 B**). Genes contributing to body column tissue (right side of the circle) are mostly associated with terms such as protein processing, pyrimidine metabolism, ribosome and spliceosome pathways. Terms enriched in the tentacle and head samples (left side of the circle) are associated with cell adhesion, ECM-receptor interaction, calcium signalling, focal adhesion and the MAPK pathway (**Figure 2.2 B**). Terms enriched in foot samples (upper side of the circle) are associated with cell adhesion, MAPK (FGFR) and Wnt signalling pathway (**Figure 2.2 B**). Terms enriched in head tissue (lower part of the circle pathway) are associated with other components of the Wnt signalling pathway as well as oxidative phosphorylation and phagosome formation (**Figure 2.2 B**).

To gain further insight into the extent of gene expression change between the different regions along the body column, we computed the transcriptome divergence between adjacent tissue regions (**Figure 2.2 C**). We included all differentially expressed genes (adjusted P-value < 0.05) with at least two-fold change between the two compartments to uncover the set of genes responsible for their identity. This process resulted in 1310 genes that were further classified based on the Gene Ontology Biology Process terms (<http://www.geneontology.org/>) and plotted according to their expression (counts per million) in each region of the body column (**Figure 2.2 C**). Interestingly, for each body region, we also identified a high number of genes with an unknown Gene Ontology annotation, which we considered as “orphans” (**Figure 2.2 C**). Orphan genes have been suggested to be involved in species-specific processes and to have effector functions (Khalturin et al., 2009; Neme and Tautz, 2013; Tautz and Domazet-Lošo, 2011).

To identify genes that behave similarly and might operate in networks, we next cluster all differentially expressed genes (2497 with adjusted P-value < 0.05) (**Figure 2.2 D**). The clusters were sorted by their expression trend along the body axis and grouped according to particular expression patterns (upregulated in one or more compartments). Four BCs were considered for further analysis (**Figure 2.2 D**): genes upregulated in tentacle and head tissue (dark blue in **Figure 2.2 D**); genes upregulated in foot tissue (yellow in **Figure 2.2 D**); genes upregulated in

both extremities, i.e. tentacle head and foot (light blue in **Figure 2.2 D**); and genes specifically upregulated in the body column (red in **Figure 2.2 D**).

2.2.3. Metabolism, cell proliferation and adhesion pathways contribute to the boundary between head and body

In a next step, we characterized the differentially expressed genes of each of the four “BCs” by identifying KEGG pathway terms enriched in these gene sets (**Figure 2.3**). We considered only pathways where more than 10 % of the corresponding members could be identified in our data set.

When analyzing differentially expressed genes enriched in body column tissue, we observed three distinct KEGG categories (**Figure 2.3**). The first one includes terms associated with DNA replication, cell cycle regulation, mismatch repair and nucleotide excision pathway (**Figure 2.3, Supplementary figure 2.2**). This is consistent with the well-known proliferation processes that occur in this region (Campbell, 1967; Bosch et al., 2010). The second category of terms enriched in the body column samples are genes involved in metabolism. We found in particular, terms associated with glutathione metabolism, synthesis of purine and pyrimidine, cysteine and methionine production, oxidative phosphorylation (OxPhos) and One-carbon (1C) pool by folate to be enriched in the *Hydra* body column. The third category of terms enriched in the body column samples concerns processing and degrading of proteins. Components of both the ubiquitin proteolysis and proteasome pathway are differentially expressed with some of them upregulated in the terminally differentiated tissue.

KEGG pathways defining terminally differentiated *Hydra* tissue involve the Ca⁺⁺/Phosphatidylinositol system as well as cell-adhesion/cytoskeleton-associated processes (**Figure 2.3**). With regard to the Ca⁺⁺/Phosphatidylinositol system, essential components of the phosphatidylinositol signal transduction cascade including Phospholipase C (PLC), protein kinase C (PKC) and calmodulin are transcriptionally upregulated in differentiated *Hydra* tissue (**Figure 2.3 and Supplementary figure 2.2**). Terminally differentiated tissue was also found to be characterized by the expression of key components of the inositolphosphate-metabolism pathway such as IP3K kinase, PI4K kinase, PI(3)P5K kinase and IP5 phosphatase (**Supplementary table 2.2**).

2.2.4. Inhibition of histone deacetylases (HDAC) breaks down head/body column boundary

We generated transgenic animals in which the body-head boundary can be visualized through the expression of reporter genes, as a system for positional dependent gene expression. Embryo-microinjection of a construct *Ks1::GFP /Actin::dsRED* (see supplement) yielded polyps which expressed eGFP in ectodermal cells in the head and dsRED in the whole body as a marker of a positive transgenic cell (**Figure 2.1 A; Figure 2.4 A**). Exposure of these polyps to low doses (2- 10nM) of a potent and specific inhibitor of histone deacetylase (Trichostatin A, TSA) resulted in animals which progressively lost terminally differentiated tentacle and head tissue (**Figure 2.4 A-C**). Tissue in the body column apparently was not affected. This led us to address the role of histone acetylation in the establishment and maintenance of the head/body column boundary.

Further, we detected the complete repertoire of histone deacetylases (HDAC) including transcripts and gene models corresponding to each of the HDAC groups described in *Homo sapiens* (HDAC1/2, 3, 4/5/7/9, 6, 8, 11) in the genome of *Hydra magnipapillata* (Chapman et al., 2010) and in the transcriptome of *Hydra vulgaris* AEP (Hemmrich et al., 2012) (**Supplementary figure 2.3 A**). Moreover, when transplanting TSA-treated *Hydra* tissue to untreated tissue followed by immunohistochemical analysis, we discovered an accumulation of acetylated histone 3 (H3ac) in the nuclei in TSA treated tissue (**Supplementary figure 2.3 B** Anti-H3ac +TSA) in comparison to untreated control tissue (Supplement 4, B, Anti-H3ac DMSO) indicating that TSA suppresses the activity of *Hydra's* HDAC leading to an increase in histone acetylation.

To link the morphological changes caused by TSA treatment (**Figure 2.4 A-C**) to the gene expression profiles, we used a microarray analysis. A PCA was employed to inspect the expression profiles of the head and upper body tissue (Head and Body 1, see **Figure 2.1 B**) when the polyps were exposed to 2.5 or 10nM of TSA (**Figure 2.4 D**). TSA responsive expression patterns were highly consistent in three independent replicates and clustered together according to the origin of the tissue and the TSA concentration. Interestingly, the PC1, which explains 44.2% of the variability in the dataset, clearly depict the transition of the expression profile between head and body under TSA treatment (**Figure 2.4 D**). TSA responsive functional classes contributing to this shift include genes involved in calcium signalling, pyrimidine metabolism, focal adhesion, MAPK signalling (FGF), regulation of actin

cytoskeleton, and the ligands of Wnt signalling pathway (**Figure 2.4 E**). In addition, genes involved in cell cycle, purine metabolism, pyrimidine metabolism, RNA degradation, RNA polymerase, spliceosome and ubiquitin-mediated proteolysis characterized the TSA treated head samples (**Figure 2.4 E**). Apparently, these functional processes are a distinguishing factor between head to body column tissue. The PC2 explains 24.3% of the variability and noticeably discriminates among TSA concentrations. Among the TSA sensitive genes, we highlight those involved in arginine and proline metabolism, mitochondrial genes, oxidative phosphorylation, purine metabolism and tryptophan metabolism as well as apoptosis, cell cycle, MAPK signalling (FGF/FGFR), and peroxisome and pyrimidine metabolism (**Figure 2.4 E**).

To investigate the extent to which TSA treatment changes the expression profile along the body axis, we next identified the differentially expressed transcripts and classified them into clusters (**Figure 2.4 F and G and Supplementary figure 2.4**). Interestingly, plotting the genes differentially expressed in TSA treated head tissue positioned their expression within the body column clusters or nearest to body column gene expression fashion (**Figure 2.4 F and G and Supplementary figure 2.4 A and B**). This analysis revealed, that inhibition of *Hydra*'s HDAC strongly affected the expression profile of cells in terminally differentiated head tissue and causes them to adapt a more body-like expression mode.

Having demonstrated that HDAC is involved in shaping the gene expression profile along the *Hydra* body axis, we next asked which pathways are involved. Thus, based on KEGG annotation of the differentially expressed genes, we found that TSA treatment in head tissue induces the expression of genes involved in cell cycle, DNA replication, spliceosome, TCA cycle, ubiquitin-mediated proteolysis and oxidative phosphorylation (e.g. cluster 1, see **Figure 2.4 F**). This analysis also showed that TSA treatment causes cells in the head region to under-express genes (e.g. cluster 6, 7 and 10, see **Figure 2.4 G and Supplementary figure 2.4 A and B**) involved in MAPK, insulin, hedgehog and calcium signaling as well as the WNT signaling pathway. The fact that several components of the Wnt signalling pathway are prominently expressed in the head (Hobmayer et al., 2000; Nakamura et al., 2011; Petersen et al., 2015), and that body column tissue is characterized by the expression of cell cycle genes (Buzgariu et al., 2018) raised the hypothesis that histone deacetylation establishes and maintains the boundary between head and body column tissue. We therefore, targeted genes involved in cell cycle and stem cell proliferation as well as genes from the Wnt signaling pathway and examined the effects of TSA treatment on their expression.

2.2.5. Cell cycle control and histone deacetylation

KEGG analysis revealed numerous genes involved in cell cycle control to be differentially expressed upon TSA treatment (**Figure 2.5 A**). The majority of the transcripts belonging to this pathway are predominantly expressed in the body region which corresponds with the localization of the three stem cell lineages in *Hydra* (Red boxes in **Figure 2.5 A**). Examples include cyclins A, B, E and H, CDK1 and CDK4, 6, E2F4, 5, HDACs, and the APC/C complex. While only few cell cycle genes are up-regulated in untreated head tissue (blue boxes in **Figure 2.5 A**), a number of them is sensitive to TSA treatment (orange boxes in **Figure 2.5 A**). Intriguingly, genes encoding Cyclin A, B and E and CDK1 as well as components of the APC complex and ubiquitin ligase Cdh1 were part of this altered gene expression profile.

Direct visualization of cells in S-phase along the body column by using 5-bromo-2'-deoxyuridine (BrdU) labelling demonstrated an increased number of S-phase cells in the body column of treated polyps (**Figure 2.5 B-D**). Intriguingly, inhibition of HDAC appeared to allow cells localized in terminally differentiated head tissue to replicate their DNA (**Figure 2.5 B-D**). We also noted that with increasing concentrations of TSA the tentacle length became much shorter and that S-phase cells can be monitored even in the most terminal tissue (**Figure 2.5 D**). Overall, our analysis strongly indicated that inhibition of HDAC affects cell cycle genes along the body column and effectively expands the zone of proliferation to include terminally differentiated tissue.

To address possible consequences of this expanded zone of proliferation, we monitored the expression of transcription factor FoxO along the body column using a transgenic line which expresses GFP under the control of the FoxO promoter (**Figure 2.5 E**). We previously have shown that *Hydra* FoxO is a key regulator of stem cell proliferation (Boehm et al., 2012). Consistent with our observation of the impact of TSA cell proliferation (**Figure 2.5 A-D**), FoxO transgenic polyps show FoxO expression along the whole animal even ectopically in the most terminal tissue regions (**Figure 2.5 F-H**). Interestingly, treatment with increasing concentrations of TSA not only expanded the zone of activity of FoxO; the TSA-induced high and ectopic level of FoxO activity also resulted in polyps with multiple body axes (**Figure 2.5 F-H**) indicating interference of FoxO with the endogenous patterning system.

2.2.6. Wnt signaling and histone deacetylation

Patterning in *Hydra* is under control of the Wnt pathway (Gee et al., 2010; Gufler et al., 2018; Hobmayer et al., 2000; Iachetta et al., 2018; Lengfeld et al., 2009; Nakamura et al., 2011; Petersen et al., 2015). To investigate the impact of histone deacetylation on the Wnt signalling pathway, we used KEGG analysis to ask which of the TSA-dependent gene clusters (**Figure 2.4**) can be classified as genes involved in the Wnt pathway. **Figure 2.6 A** reveals that many components of the Wnt pathway in *Hydra* are differentially expressed between head and body column tissue. Seven Wnt encoding genes, two Frizzled transmembrane receptors, Dishevelled, Axin, β -Catenin and GSK-3 are restricted to or upregulated in the head region (see **Figure 2.6 A**, blue boxes). Other transcripts including two other transmembrane Frizzled proteins, a casein kinase II (CK2), a protein phosphatase (PP2A) and the transcription factors cMyc and cJun are more abundant in the body region (**Figure 2.6, A** Red boxes). Intriguingly, and consistent with our previous findings of TSA-treated head tissue to become more body column like (**Figure 2.4 F and G**), our data reveal that the head-specific expression of two Wnt encoding genes (Wnt1 and Wnt5) are downregulated upon HDAC inhibition (**Figure 2.6 A**, light blue boxes). **Figure 2.6 B** summarizes our current understanding of how histone hyperacetylation affects *Hydra* by extending the stem cell containing body region and reducing the differentiated compartments.

2.3. Discussion

2.3.1. Principles of boundary formation in the early emerging metazoan *Hydra*

Decades of studies have established the basic features of patterning in *Hydra* (Bode, 2011; Gierer, 2012; Koinuma et al., 2000; Meinhardt, 2012; Watanabe et al., 2014). The single body axis is established during embryogenesis and maintained in adult polyps by local Wnt activity (Gee et al., 2010; Gufler et al., 2018; Hobmayer et al., 2000; Nakamura et al., 2011). Transcription factors including c-myc (Ambrosone et al., 2012; Hartl et al., 2014) and FoxO (Boehm et al., 2012; Mortzfeld and Bosch, 2017) contribute to maintaining continuous stem cell proliferation and differentiation in the body column. Here we present a multidimensional gene expression dataset from carefully selected regions along the *Hydra* body axis which provides a foundational gene expression resource encompassing central aspects of patterning in *Hydra*. It supports earlier findings of a spatial (i.e. region-specific) organization of gene expression in adult polyps (**Figure 2.1**) and reveals that numerous not conserved “orphan” genes along with conserved regulators and pathways contribute to the patterning process (**Figure 2.2**). The sensitivity of the expression of many of these genes to experimentally induced

changes in histone acetylation (**Figure 2.4**) reveals that boundaries between compartments are formed and maintained by mechanisms that depend on reversible histone acetylation and deacetylation.

Our results suggest that specific genes are in charge of the boundary maintenance along the body axis. Among these genes, we found members of the cell cycle pathways such as cyclins, cyclins-dependent-kinases (CDKs), proteins forming complexes responsible for DNA replication and the transcription factors c-myc and FoxO (**Figure 2.3, Figure 2.5 A, and Supplementary figure 2.2**). The body column was also characterized by the upregulation of metabolic pathways mainly related to energy supply and protein processing. In addition, many enzymes are upregulated in the body column for the production of purines and pyrimidines, OxPhos, glutathione, cysteine and methionine metabolism (**Figure 2.2 B, Figure 2.3, Figure 2.4 E**).

Interestingly, many of these pathways are known to be important in maintaining the stemness of pluripotent stem cells in different organisms (Shiraki et al., 2014; Shyh-Chang et al., 2013). The OxPhos pathway, for example, is used by mitochondria in totipotent mammalian stem cells and differentiating embryonic stem cells (ESCs) for oxidation of pyruvate (Shyh-Chang et al., 2013). Moreover, cell growth of undifferentiated human ESCs and induced pluripotent stem cells (iPSCs) is dependent on large amounts of methionine; and deprivation of this amino acid result in cell growth inhibition and cell cycle arrest (Shiraki et al., 2014). A particularly interesting class of metabolic terms enriched in the body column was associated with S-Adenosyl L-methionine (SAM) production (**Supplementary figure 2.2**). Since SAM is considered a universal methyl donor (Ducker and Rabinowitz, 2017), it may play a major role in epigenetics because it is the principal methyl substrate used by methyltransferases to methylate metabolites, RNA, DNA, and proteins including histones (Mentch and Locasale, 2016). After the methyl group from SAM is transferred to an acceptor substrate, S-adenosylhomocysteine (SAH) is produced. In turn, SAH is hydrolyzed to adenine and homocysteine, and to complete the cycle, the homocysteine can be re-methylated via methionine synthase to produce again methionine.

At the same time, genes encoding phosphatidylinositol- and related phosphoinositides-modifying enzymes are expressed in terminally differentiated head tissue (**Figure 2.3**). One regulatory step later, enzymes up or downstream the conventional phosphatidylinositol

signaling pathway including PKC and PLC, diacylglycerol kinase and calmodulins are also differentially upregulated in head tissue (**Figure 2.3**). These findings support earlier studies in *Hydra* (Hassel and Bieller, 1996; Müller, 1990, 1995). They reported severe defects in patterning processes when, e.g., PKC is chemically inhibited. Interestingly, these pathways were previously associated with trophoblast differentiation (Kent et al., 2010), keratinocyte differentiation (Haase et al., 1997), and in the regulation of growth and differentiation of hematopoietic cells (Michell et al., 1990) and neurons (Loss et al., 2013).

Besides, many genes involved in the architecture of the actin cytoskeleton including F actin, myosin, integrins, Tallins (**Figure 2.3, Supplementary figure 2.2**); calcium signalling pathways, focal and cell adhesion are upregulated in differentiated head tissue in comparison with the body column. Head tissue is also characterized by the expression of genes encoding extracellular components including collagens, thrombospondin, and ligands such as fibroblast growth factors. Another group of genes differentially expressed in head tissue encodes the anaphase-promoting complex/cyclosome (APC/C) which is a multifunctional ubiquitin-protein ligase that targets various substrates for proteolysis inside and outside of the cell cycle (Alfieri et al., 2017).

The complex expression patterns are controlled and maintained by histone deacetylases. Prolonged inhibition of histone deacetylase by TSA results in progressive loss of terminally differentiated head tissue (**Figure 2.4 C**). Ectopic upregulation of cell cycle genes (**Figure 2.4 F, Figure 2.5 A**) and the sudden appearance of actively replicating cells in the most apical region of the body axis (**Figure 2.5 B-D**) indicates that repression of histone deacetylases prevents the correct onset of the differentiation program at the head/body column boundary. HDAC repression also affects both the cell-cycle regulatory complex APC/C and genes APC 1,8,10 which normally are only expressed in the body column as well as genes encoding components of the SCF complex, which mediates ubiquitination of proteins and has been associated with the degradation of cell cycle proteins. These repressions also remove the confinement of FoxO expression to the body column region (**Figure 2.5 F-H**).

Previous studies have provided evidence that at least some of the HDACs including HDAC 1,2 and 3 are regulated by phosphatidylinositides (Millard et al., 2013a, 2017). Our dataset shows that many genes involved in the phosphatidylinositol pathway and inositol phosphate metabolism are upregulated in tissue from the apical (head) or basal (foot) region (**Figure 2.3**

and **Figure 2.7 C**). The strikingly strong presence of phosphatidylinositides in differentiated (head and foot) tissue, therefore, could cause the region-specific full activation of HDAC (see also Watson et al., 2016).

To summarize, whereas in *Hydra* a common organizer is used to initiate and to maintain the region-specific expression of both conserved as well as un-annotated genes, the different and complex patterns of gene expression at the head/body boundary are controlled by histone deacetylases (HDACs) and thus by epigenetic regulators which regulate transcription through chromatin modification without directly binding response elements on DNA. We conclude (**Figure 2.7 A-D**) that reversible histone acetylation and deacetylation appears as key component maintaining the sharp boundary between the stem cell (body column) and terminally differentiated (head) compartment in *Hydra*.

2.3.2. An evolutionary perspective on HDAC and boundary formation: parallels between *Hydra* and patterning in other animals and plants

Besides *Hydra*, there are other invertebrates and also vertebrates in which an interference (mutation, repression) with HDACs activity affects patterning and embryonic development (de Ruijter et al., 2003). For example, exposure of *Drosophila* to Trichostatin A causes developmental defects (Ikegami et al., 1993; Nemer, 1998). Mutations in the *Drosophila* hdac-1 homolog Rpd3 (**Figure 2.7 E**) are lethal and lead to a paired-rule segmentation phenotype with disturbed parasegmental boundary formation (Mannervik and Levine, 1999; Mottus et al., 2000; reviewed by Dahmann et al., 2011b; Umetsu and Dahmann, 2010). Moreover, treatment of sea urchin embryos with 10ng/ml TSA induces hyperacetylation of histones in cells of blastula stages. Development of the treated embryos is arrested at an early gastrula stage, indicating that a developmental window in the midblastula stage is sensitive to the normal cycle of histone acetylation and deacetylation (Ikegami et al., 1993). Interfering with HDAC activity also affects vertebrate development. Homozygous hdac-1 zebrafish knock-out mutants show numerous developmental abnormalities in the heart and neural epithelial tissue (Pillai et al., 2004). Similarly, treatment of *Xenopus* frog embryos with TSA leads to developmental defects in the head and tail region accompanied by embryonic lethality (Almouzni et al., 1994). In mice, hdac-1 knockout led to severe proliferation defects, growth retardation and abnormal head and allantois development followed by embryonic lethality before E10.5 (Lagger et al., 2002).

Interestingly, as in continuously proliferating *Hydra*, the remarkable ability of some plants to

retain populations of dividing undifferentiated cells (meristems) also depends on histone acetylation. It is widely known that angiosperms, root and shoot growth is maintained and regulated through the activity of the apical meristems (Shishkova et al., 2008). ‘*Potentially, the plant axis can grow indefinitely in length through the activity of its apical meristem and in width through the activity of the vascular cambium* (cited from Sinnott, 1960). In the *Arabidopsis* root (**Figure 2.7 F**), stem cells localized in the apex of the root meristem are characterized by relatively weak and dynamic histone-DNA interactions. However, the constant supply of new cells from the root meristem towards the differentiation zone results in a progressively stronger bond of chromatin to histones (Rosa et al., 2014). In fact, histone hyperacetylation by means of TSA leads to an increase in meristem size (schematically depicted in **Figure 2.7 F**) and in the expression of the meristem marker RHD6 in cells from the differentiation zone, together with an overall decrease of histone stability (Rosa et al., 2014). These results reveal a role for acetylation and histone stability in determining meristem competency and root development (Rosa et al., 2014).

The observations as portrayed in **Figure 2.7 D to F** for *Hydra*, *Drosophila* and *Arabidopsis* indicate an evolutionary conserved and therefore the important role of histone acetylation in regulating both differentiation and cell cycle progression in a spatially controlled framework and for partitioning the body axis into domains with a specific identity. Continuously proliferating *Hydra* polyps, therefore, may present a unique opportunity to decipher the molecular mechanisms regulating boundary formation beyond the spatial and temporal restrictions imposed by most animal embryos.

2.4. Materials and methods

Animals and culture conditions

Experiments were carried out using *Hydra magnipapillata* and *Hydra vulgaris* strain AEP. All animals were cultured identically under constant environmental conditions that include: culture medium, *Artemia salina* as food, and temperature according to standard procedures at 18°C (Lenhoff and Dubois Brown, 1970).

Generation of transgenic *H. vulgaris* strain AEP

For reporter of Ks1 (GenBank, AM161049.1) and FoxO (GenBank JX118843), transgenic polyps were generated, expressing GFP under the control of Ks1 and FoxO promoter respectively. Therefore, the promoter region of Ks1 and FoxO (aprox. 1.1 kb. for Ks1 and 1.5kb

for FoxO from the 5' flanking region) were cloned into ligAF vector (Wittlieb et al., 2006) in front of *enhanced green fluorescent protein (eGFP)* and followed by actin terminator. A second cassette was included with actin-promoter driving the expression of *red fluorescent protein (dsRED)* and finalized through actin terminator (figure 5, E). The resulting vector was injected into *H. vulgaris* (AEP) embryos as previously described (Wittlieb et al., 2006). Founder polyps showed stable dsRED expression in a group of ectodermal cells and were expanded further by clonal propagation.

In situ hybridization

Gene expression analysis was performed by whole-mount in situ hybridization as described previously (Khalturin et al., 2007). The gene expression pattern was detected with digoxigenin (DIG)-labelled RNA probes. Anti-sense RNA probes were designed to recognize specifically the sequence of the following genes: Ks1 (GenBank, AM161049.1), foxO (GenBank JX118843), and at metazome server: HMG3b3 (Hma2.227348) Aristaless 1 (Hma2.219326), KLF13 (Hma2.209274). DIG-labelled sense probes were used as a control. Signal was obtained using anti-DIG antibodies conjugated to alkaline phosphatase (1:2000, Roche Diagnostics) and NBT/BCIP staining solution (Roche).

Immunohistochemistry

Immunohistochemical detection of acetylated histone 3 in whole mount *Hydra* preparations was performed following standard procedures (Engel et al., 2002) using polyclonal rabbit anti-Histone H3 (acetyl K9 + K14 + K18 + K23 + K27) antibody – CHIP Grade (1:2000) and Alexa 488 anti-rabbit antibody (1:1000). Phalloidin and TO-PRO staining was done as described previously (Anton-Erxleben et al., 2009). Confocal laser-scanning microscopy was done using TCS SP1 laser-scanning confocal microscope (Leica).

BrdU labelling and detection

For analysis of cell with active replication in the whole polyp, animals were exposed for 3h to 5mmol l⁻¹ of 5-bromo-2'-deoxyuridine (BrdU) (Holstein et al., 1991). Detection of incorporated BrdU was carried out according to Holstein et al. (1991) using monoclonal anti-BrdU antibodies (1:100, Roche), alkaline phosphatase-conjugated sheep-anti-mouse secondary antibodies (1:5000, Millepore ®) and NBT/BCIP staining solution (Roche).

Microscopy analysis

Fluorescent images of transgenic polyps were taken on a Zeiss Axioscope fluorescence microscope equipped with an Axiocam (Zeiss) digital camera. Laser-scanning confocal data of transgenic cells were acquired by using LeicaTCSSP1 CLS microscope. The *in situ* hybridization preparations were analysed using Zeiss Axioscope microscope.

Phylogenetic analysis

For the calculation of the HDAC tree, an amino acid alignment of 7 HDAC sequences from *Hydra magnipapillata*, 5 HDAC sequences from *Hydra vulgaris* AEP transcriptome, 11 HDACs from humans and one homologue from *Saccharomyces cerevisiae* HDA1 were used. The sequences were aligned using Clustal W with the standard parameters for multiple alignments (Thompson et al., 1994). All ambiguous positions were removed for each sequence pair, giving a total of positions in the final data set of 1462. The statistical method used to make the reconstruction was Neighbour-joining and the bootstrap values were calculated based on 1000 replicates, this was implemented by using MEGA 6 (Tamura et al., 2013) (with p-distance model).

Chemical interference

For the chemical interference experiments the polyps were incubated in 5 or 10nM of TSA or DMSO (control polyps) for 24, 48 or 72 hours. *Hydra* medium was supplemented with a stock solution of TSA (1000nM in DMSO) to a final concentration of 5 or 10nM. TSA was obtained from Sigma-Aldrich (catalog number T8552). The control polyps were kept in *Hydra* medium with the corresponding amount of DMSO for each concentration.

***Hydra* axial transcriptome, assembly and annotation**

Total RNA was isolated using TRIzol™ reagent (Invitrogen™) according to manufacturer's protocol and followed by a purification protocol using PureLink™ RNA mini Kit columns from Ambion®. For the RNAseq experiments total RNA was isolated from 6 *Hydra* -body sections (Tentacles, Head, Body 1, Body 2, Body bud and Foot, Figure 1,C) using TRIzol™ reagent (Invitrogen™) according to manufacturer's protocol and combine with a purification protocol using PureLink™ RNA mini Kit columns from Ambion®. For them we pooled from 150-200 sections of each polyp part. RNA quality was measured with the Agilent RNA Nano Kit, for the individual pools. Samples with RIN values above 7.5 were used for sequencing. Library preparation was done using the Illumina TruSeq stranded Total RNA preparation,

following manufacturers' instructions. Sequencing was done in Illumina MiSeq sequencer. Transcripts were assembled using Trinity v2.1.1 (Haas et al., 2013). Trinity was run with `-min_kmer_cov 2` and `-min_contig_length 200` to reduce memory requirements and discard contigs shorter than 200 bp. In all other respects, default parameters were used. In total, 58.258.591 raw paired reads were processed. Poor quality reads shorter than 25bp were removed. The first ten nucleotides of each read and adapter contamination were also removed prior further analysis. Poor quality nucleotides (average base quality over a four bp sliding window < 5) were trimmed. The remaining 52.631.072 paired reads were used for de novo transcriptome assembly. The assembled transcriptome was assessed for gene completeness using the BUSCO (Benchmarking Universal Single-Copy Orthologs) library (Simão et al., 2015).

After Trinity assembly, annotation was performed using Trinotate pipeline v3.0.1 (<https://trinotate.github.io/>). Gene open reading frames (ORFs) were predicted using TransDecoder v5.0.0 (<http://transdecoder.github.io>). Only the longest predicted ORFs that were at least 100 amino acids long were retained. Obtained ORFs were blasted against the SwissProt- UniProt database (v06.2017) with an E-value cut-off of 10^{-5} using BlastP (Altschul et al., 1990). Additionally, the software Hmmer v.3.1b2 was used for protein domain identification (Finn et al., 2015), Tmhmm v.2.0c for prediction of transmembrane helices in proteins (Krogh et al., 2001) and SignalP v.4.1 to predict signal peptide cleavage sites (Petersen et al., 2011).

Differentially expressed genes along *Hydra* body parts

Differentially expressed (DE) genes among the three parts of *Hydra*'s body plan (foot, body, and head) were identified using the R-package DESeq2 (Love et al., 2014). RNAseq libraries from each body part were mapped to the de novo *Hydra* transcriptome and the transcripts were quantified using RSEM (Li and Dewey, 2011). Transcripts with more than two counts per million in at least three libraries were kept for further analysis. Correction for batch effects in the samples was applied through a surrogate variable analysis using the sva R-package (Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Storey JD, 2017). Negative binomial generalized linear models in DESeq2 were used to determine the differential expression of the transcripts among the body parts. Transcripts with a significance cut-off of adjusted P-value < 0.05 were selected for comparisons. The P-value was adjusted according to the false discovery rate (FDR) (Benjamini and Hochberg, 1995).

Cluster and GO enrichment analysis

The expression patterns of the DE transcripts were analysed with a hierarchically clustering tree. A regularized log transformation (Love et al., 2014) of the expression of DE transcripts and the Euclidean distance was used to construct the tree. The clusters of similar expression patterns were calculated cutting the tree at 50% of the maximum height.

Microarray hybridization and analysis

Custom Agilent Gene Expression Microarray was designed using the *H.magniapillata* gene models available at Metazome (<https://metazome.jgi.doe.gov/pz/portal.html>). For the microarray analysis of gene expression, total RNA was isolated from head and the first part of the body (Body 1, Figure 1, C) of polyps treated with 2,5 TSA or 10nM TSA or the control polyps in DMSO using TRIzol™ reagent (Invitrogen™) according to manufacturer's protocol and combine with a purification protocol using PureLink™ RNA mini Kit columns from Ambion®. About 400ng of total RNA were labelled with Cy3 using the one-colour Quick-Amp Labeling Kit (Agilent Technologies). Labelled cRNA samples (n= 24 in total) were hybridized to Agilent Gene Expression Microarray 4x44k slides for 17h at 65°C and treated according to the Agilent protocol. The microarrays were scanned using the Agilent High Resolution G2565CA Microarray Scanner System. Agilent single-channel microarrays were processed using an in-house R-script which integrates the package limma (Smyth, 2005) for quality control, normalization and gene differential expression assessment. For each microarray, we applied a background correction (Ritchie et al., 2007) using normexp-saddle (Ritchie et al., 2006; Silver et al., 2009). For replicated probes on the same array, we replaced the individual value with the respective average. For array comparability, we performed quantile-normalization (Bolstad et al., 2003). Differentially expressed genes were identified by a linear modeling approach and the empirical Bayes statistics (Smyth, 2004).

Quality of the assembly - supplementary results

The quality of the *Hydra* transcriptome was made based on statistics such as the ExN50 value and the number of genes longer than 1kb. Transcriptome completeness regarding gene content was assessed as a complementary approach. The transcriptome was tested for the presence or absence of a list of conserved orthologous genes. The BUSCO (Benchmarking Universal Single-Copy Orthologs) library of Metazoa orthologous genes was used (Simão et al., 2015). This library represents a collection of 843 well-annotated and conserved single-copy metazoan orthologs.

In total, 730 (86.6%) complete BUSCO hits were obtained, and duplicate hits to 143 (17%) genes. Another 45 (5.3%) were fragmented and 68 (8%) were missing. In theory, duplicates represent gene duplication and/or mechanisms such as alternative splicing in the sample. Indeed, duplicates were high in our total assembly; therefore, this could be explained by the expression of paralogs and/or isoforms in the different *Hydra*'s body plan (foot, body, and head). However, the high number of complete genes provides an important validation of the depth and completeness of the assembly.

D. mela Ref C:98% [D:6.4%], F:0.6%, M:0.3%, n:2 675

D. mela busco C:99% [D:3.7%], F:0.2%, M:0.0%, n:2 675

C. eleg Ref C:85% [D:6.9%], F:2.8%, M:11%, n:843

C. eleg busco C:90% [D:11%], F:1.7%, M:7.5%, n:843

H. sapi Ref C:89% [D:1.5%], F:6.0%, M:4.5%, n:3 023

H. sapi busco C:99% [D:1.7%], F:0.0%, M:0.0%, n:3 023

L. giga Ref C:89% [D:2.3%], F:4.3%, M:5.8%, n:843

L. giga busco C:90% [D:13%], F:7.8%, M:2.1%, n:843

A. nidu Ref C:98% [D:1.8%], F:0.9%, M:0.2%, n:1 438

A. nidu busco C:95% [D:7.3%], F:3.8%, M:0.9%, n:1 438

Acknowledgements

We thank Jörg Wittlieb for technical assistance. Konstantin Khalturin for help with the transcriptome analysis. Anna-Marei Boehm for the *Hydra* transgenic line Ks1- GFP, and Alexander Klimovich for the help with Confocal Microscopy Images. Further we appreciate the expertise provided by the sequencing facility at the Institute of Clinical Molecular Biology (IKMB) in Kiel, Germany.

Funding

The work was supported in part by grants from the Deutsche Forschungsgemeinschaft (DFG) and the CRC 1182 („Origin and Function of Metaorganisms“). J.A.L.-Q acknowledges support from the DAAD. R.N. was supported by a fellowship of the IMPRS for Evolutionary Biology, fund from the European Research Council (NewGenes – 322564) and the Pew Latin American Fellowship from Pew Charitable Trust. T.C.G.B. appreciates support from the Canadian Institute for Advanced Research (CIFAR) and is grateful for a Fellow award from the

Wissenschaftskolleg in Berlin. The position of G.G.T. was funded by the DFG Cluster of Excellence “Inflammation of Interfaces” (EXC 306).

Author contributions

J.A.L-Q, T.C.G.B., designed experiments and wrote the manuscript. J.A.L-Q, performed experiments. J.A.L-Q, G.G.T, R.N. analysed data. G.G.T, R.N. analysed bioinformatic data.

References

- Alfieri, C., Zhang, S., and Barford, D. (2017). Visualizing the complex functions and mechanisms of the anaphase promoting complex/cyclosome (APC/C). *Open Biol.* 7, 170204.
- Almouzni, G., Khochbin, S., Dimitrov, S., and Wolffe, A.P. (1994). Histone acetylation influences both gene expression and development of *Xenopus laevis*. *Dev. Biol.* 165, 654–669.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Ambrosone, A., Marchesano, V., Tino, A., Hobmayer, B., and Tortiglione, C. (2012). Hymc1 downregulation promotes stem cell proliferation in *Hydra vulgaris*. *PLoS One* 7, e30660.
- Anton-Erxleben, F., Thomas, A., Wittlieb, J., Fraune, S., and Bosch, T.C.G. (2009). Plasticity of epithelial cell shape in response to upstream signals: a whole-organism study using transgenic *Hydra*. *Zoology (Jena)* 112, 185–194.
- Battle, E., and Wilkinson, D.G. (2012). Molecular mechanisms of cell segregation and boundary formation in development and tumorigenesis. *Cold Spring Harb. Perspect. Biol.* 4, a008227.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.
- Bode, H. (2011). Axis formation in *Hydra*. *Annu. Rev. Genet.* 45, 105–117.
- Boehm, A.-M., Khalturin, K., Anton-Erxleben, F., Hemmrich, G., Klostermeier, U.C., Lopez-Quintero, J.A., Oberg, H.H., Puchert, M., Rosenstiel, P., Wittlieb, J., et al. (2012). FoxO is a critical regulator of stem cell maintenance in immortal *Hydra*. *Proc. Natl Acad. Sci. U S A* 109, 19697–19702.
- Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193.
- Bosch, T.C.G., Anton-Erxleben, F., Hemmrich, G., and Khalturin, K. (2010). The *Hydra* polyp: nothing but an active stem cell community. *Dev. Growth Differ.* 52, 15–25.
- Böttger, A., and Hassel, M. (2012). *Hydra*, a model system to trace the emergence of boundaries in developing eumetazoans. *Int. J. Dev. Biol.* 56, 583–591.
- Buzgariu, W., Wenger, Y., Tcaciuc, N., Catunda-Lemos, A.-P., and Galliot, B. (2018). Impact of cycling cells and cell cycle regulation on *Hydra* regeneration. *Dev. Biol.* 433, 240–253.
- Campbell, R.D. (1967). Tissue dynamics of steady state growth in *Hydra littoralis*. *Dev. Biol.*

- 15, 487–502.
- Campbell, R.D., and David, C.N. (1974). Cell cycle kinetics and development of *Hydra attenuata*. *J. Cell Sci.* 16.
- Cayuso, J., Xu, Q., and Wilkinson, D.G. (2015). Mechanisms of boundary formation by Eph receptor and ephrin signaling. *Dev. Biol.* 401, 122–131.
- Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., et al. (2010). The dynamic genome of *Hydra*. *Nature* 464, 592–596.
- Dahmann, C., Oates, A.C., and Brand, M. (2011). Boundary formation and maintenance in tissue development. *Nat. Rev. Genet.* 12, 43–55.
- David, C.N., and Campbell, R.D. (1972). Cell cycle kinetics and development of *Hydra attenuata*. *J. Cell Sci.* 11, 557–568.
- Ducker, G.S., and Rabinowitz, J.D. (2017). One-carbon metabolism in health and disease.
- Endl, I., Lohmann, J.U., Bosch, T.C. (1999). Head-specific gene expression in *Hydra*: Complexity of DNA- protein interactions at the promoter of *ks1* is inversely correlated to the head activation potential. *Proc. Natl Acad. Sci. U S A* 96, 1445–1450.
- Engel, U., Ozbek, S., Streitwolf-Engel, R., Petri, B., Lottspeich, F., Holstein, T.W., Ozbek, S., and Engel, R. (2002). Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in *Hydra*. *J. Cell Sci.* 115, 3923–3934.
- Eskeland, R., Freyer, E., Leeb, M., Wutz, A., and Bickmore, W.A. (2010). Histone acetylation and the maintenance of chromatin compaction by polycomb repressive complexes. *Cold Spring Harb. Symp. Quant. Biol.* 75, 71–78.
- Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–W38.
- Gauchat, D., Kreger, S., Holstein, T., and Galliot, B. (1998). *prdl-a*, a gene marker for *Hydra* apical differentiation related to triploblastic paired-like head-specific genes. *Development* 125.
- Gee, L., Hartig, J., Law, L., Wittlieb, J., Khalturin, K., Bosch, T.C.G., and Bode, H.R. (2010). β -catenin plays a central role in setting up the head organizer in *Hydra*. *Dev. Biol.* 340, 116–124.
- Gierer, A. (2012). The *Hydra* model - a model for what? *Int. J. Dev. Biol.* 56, 437–445.
- Gufler, S., Artes, B., Bielen, H., Krainer, I., Eder, M.-K., Falschlunger, J., Bollmann, A., Ostermann, T., Valovka, T., Hartl, M., et al. (2018). β -Catenin acts in a position-independent

- regeneration response in the simple eumetazoan *Hydra*. *Dev. Biol.* 433, 310–323.
- Haase, I., Liesegang, C., Binting, S., Henz, B.M., and Rosenbach, T. (1997). Phospholipase C-mediated signaling is altered during HaCaT cell proliferation and differentiation. *J. Invest. Dermatol.* 108, 748–752.
- Hartl, M., Glasauer, S., Valovka, T., Breuker, K., Hobmayer, B., and Bister, K. (2014). *Hydra* myc2, a unique pre-bilaterian member of the myc gene family, is activated in cell proliferation and gametogenesis. *Biol. Open* 3, 397–407.
- Hassel, M., and Bieller, A. (1996). Stepwise transfer from high to low lithium concentrations increases the head-forming potential in *Hydra vulgaris* and possibly activates the PI cycle. *Dev. Biol.* 177, 439–448.
- Hemmerich, G., Khalturin, K., Boehm, A.M., Puchert, M., Anton-Erxleben, F., Wittlieb, J., Klostermeier, U.C., Rosenstiel, P., Oberg, H.-H., Domazet-Loso, T., et al. (2012). Molecular signatures of the three stem cell lineages in *Hydra* and the emergence of stem cell function at the base of multicellularity. *Mol. Biol. Evol.* 29, 3267–3280.
- Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C.M., von Laue, C.C., Snyder, P., Rothbacher, U., and Holstein, T.W. (2000). WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*. *Nature* 407, 186–189.
- Holstein, T.W., Hobmayer, E., and David, C.N. (1991). Pattern of epithelial cell cycling in *Hydra*. *Dev. Biol.* 148, 602–611.
- Iachetta, R., Ambrosone, A., Klimovich, A., Wittlieb, J., Onorato, G., Candeo, A., D'andrea, C., Intartaglia, D., Scotti, N., Blasio, M., et al. (2018). Real time dynamics of β -catenin expression during *Hydra* development, regeneration and Wnt signalling activation. *Int. J. Dev. Biol.* 62, 311–318.
- Ikegami, S., Ooe, Y., Shimizu, T., Kasahara, T., Tsuruta, T., Kijima, M., Yoshida, M., and Beppu, T. (1993). Accumulation of multiacetylated forms of histones by trichostatin A and its developmental consequences in early starfish embryos. *Roux's Arch. Dev. Biol.* 202, 144–151.
- Ikeuchi, M., Iwase, A., and Sugimoto, K. (2015). Control of plant cell differentiation by histone modification and DNA methylation. *Curr. Opin. Plant Biol.* 28, 60–67.
- Kent, L.N., Konno, T., and Soares, M.J. (2010). Phosphatidylinositol 3 kinase modulation of trophoblast cell differentiation. *BMC Dev. Biol.* 10.
- Khalturin, K., Anton-Erxleben, F., Milde, S., Plötz, C., Wittlieb, J., Hemmerich, G., and Bosch, T.C.G. (2007). Transgenic stem cells in *Hydra* reveal an early evolutionary origin for key elements controlling self-renewal and differentiation. *Dev. Biol.* 309, 32–44.

- Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T.C.G. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* *25*, 404–413.
- Koinuma, S., Umesono, Y., Watanabe, K., and Agata, K. (2000). Planaria FoxA (HNF3) homologue is specifically expressed in the pharynx-forming cells. *Gene* *259*, 171–176.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* *305*, 567–580.
- Lagger, G., O’Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., Schuettengruber, B., Hauser, C., Brunmeir, R., Jenuwein, T., et al. (2002). Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J.* *21*, 2672–2681.
- Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Storey JD, Z.Y. and T.L. (2017). sva: surrogate variable analysis. R package version 3.26.0.
- Lengfeld, T., Watanabe, H., Simakov, O., Lindgens, D., Gee, L., Law, L., Schmidt, H.A., Özbek, S., Bode, H., and Holstein, T.W. (2009). Multiple Wnts are involved in *Hydra* organizer formation and regeneration. *Dev. Biol.* *330*, 186–199.
- Lenhoff, H.M., and Dubois Brown, R. (1970). Mass culture of *Hydra*: an improved method and its application to other aquatic invertebrates. *Lab. Anim.* *4*, 139–154.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* *12*, 323.
- Loss, O., Wu, C.T., Riccio, A., and Saiardi, A. (2013). Modulation of inositol polyphosphate levels regulates neuronal differentiation. *24*.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Mannervik, M., and Levine, M. (1999). The Rpd3 histone deacetylase is required for segmentation of the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 6797–6801.
- Martinez, D.E., Dirksen, M.L., Bode, P.M., Jamrich, M., Steele, R.E., and Bode, H.R. (1997). Budhead, a fork head/HNF-3 homologue, is expressed during axis formation and head specification in *Hydra*. *Dev. Biol.* *192*, 523–536.
- Meinhardt, H. (2012). Modeling pattern formation in *Hydra*: a route to understanding essential steps in development. *Int. J. Dev. Biol.* *56*, 447–462.
- Mentch, S.J., and Locasale, J.W. (2016). One-carbon metabolism and epigenetics: understanding the specificity. *Ann. N.Y. Acad. Sci* *1363*, 91–98.

- Michell, R.H., Conroy, L.A., Finney, M., French, P.J., Brown, G., Creba, J.A., Bunce, C.M., and Lord, J.M. (1990). Inositol Lipids and Phosphates in the Regulation of the Growth and Differentiation of Haemopoietic and other Cells. *Philos. Trans. R. Soc. London B Biol. Sci.* 327.
- Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C. V., Fairall, L., and Schwabe, J.W.R. (2013a). Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol. Cell* 51, 57–67.
- Millard, C.J., Watson, P.J., Fairall, L., and Schwabe, J.W.R. (2017). Targeting class I histone deacetylases in a “complex” environment. *Trends Pharmacol. Sci.* 38, 363–377.
- Millard, C.J.J., Watson, P.J.J., Celardo, I., Gordiyenko, Y., Cowley, S.M.M., Robinson, C.V. V, Fairall, L., and Schwabe, J.W.R.W.R. (2013b). Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol. Cell* 51, 57–67.
- Minobe, S., Fei, K., Yan, L., Sarras Jr., M.P., and Werle, M.J. (2000). Identification and characterization of the epithelial polarity receptor “Frizzled” in *Hydra vulgaris*. *Dev. Genes Evol.* 210, 258–262.
- Morao, A.K., Bouyer, D., and Roudier, F. (2016). Emerging concepts in chromatin-level regulation of plant cell differentiation: timing, counting, sensing and maintaining. *Curr. Opin. Plant Biol.* 34, 27–34.
- Mortzfeld, B.M., and Bosch, T.C. (2017). Eco-Aging: stem cells and microbes are controlled by aging antagonist FoxO. *Curr. Opin. Microbiol.* 38, 181–187.
- Mottus, R., Sobel, R.E., and Grigliatti, T.A. (2000). Mutational analysis of a histone deacetylase in *Drosophila melanogaster*: missense mutations suppress gene silencing associated with position effect variegation. *Genetics* 154, 657–668.
- Müller, W.A. (1990). Ectopic head and foot formation in *Hydra*: Diacylglycerol-induced increase in positional value and assistance of the head in foot formation. *Differentiation* 42, 131–143.
- Müller, W.A. (1995). Competition for factors and cellular resources as a principle of pattern formation in *Hydra*: I. Increase of the potentials for head and bud formation and rescue of the regeneration-deficient mutant reg-16 by treatment with diacylglycerol and arachidonic acid. *Dev. Biol.* 167, 159–174.
- Nakamura, Y., Tsiairis, C.D., Ozbek, S., and Holstein, T.W. (2011). Autoregulatory and repressive inputs localize *Hydra* Wnt3 to the head organizer. *Proc. Natl. Acad. Sci.* 108, 9137–9142.
- Neme, R., and Tautz, D. (2013). Phylogenetic patterns of emergence of new genes support a

- model of frequent de novo evolution. *BMC Genomics* *14*, 117.
- Nemer, M. (1998). Histone deacetylase mRNA temporally and spatially regulated in its expression in sea urchin embryos. *Dev. Growth Differ.* *40*, 583–590.
- Perrimon, N., Lanjuin, A., Arnold, C., and Noll, E. (1996). Zygotic lethal mutations with maternal effect phenotypes in *Drosophila melanogaster*. II. Loci on the second and third chromosomes identified by P-element-induced mutations. *Genetics* *144*, 1681–1692.
- Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* *8*, 785–786.
- Philipp, I., Aufschnaiter, R., Ozbek, S., Pontasch, S., Jenewein, M., Watanabe, H., Rentzsch, F., Holstein, T.W., and Hobmayer, B. (2009). Wnt/B-Catenin and noncanonical Wnt signaling interact in tissue evagination in the simple eumetazoan *Hydra*. *Proc. Natl. Acad. Sci.* *106*, 4290–4295.
- Pillai, R., Coverdale, L.E., Dubey, G., and Martin, C.C. (2004). Histone deacetylase 1 (HDAC-1) required for the normal formation of craniofacial cartilage and pectoral fins of the zebrafish. *Dev. Dyn.* *231*, 647–654.
- Ritchie, M.E., Diyagama, D., Neilson, J., van Laar, R., Dobrovic, A., Holloway, A., and Smyth, G.K. (2006). Empirical array quality weights in the analysis of microarray data. *BMC Bioinformatics* *7*, 261.
- Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., and Smyth, G.K. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics* *23*, 2700–2707.
- Rosa, S., Ntoukakis, V., Ohmido, N., Pendle, A., Abranches, R., and Shaw, P. (2014). Cell differentiation and development in *Arabidopsis* are associated with changes in histone dynamics at the single-cell level. *Plant Cell* *26*, 4821–4833.
- de Ruijter, A.J.M., van Gennip, A.H., Caron, H.N., Kemp, S., and van Kuilenburg, A.B.P. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* *370*, 737–749.
- Shiraki, N., Shiraki, Y., Tsuyama, T., Obata, F., Miura, M., Nagae, G., Aburatani, H., Kume, K., Endo, F., and Kume, S. (2014). Methionine Metabolism Regulates Maintenance and Differentiation of Human Pluripotent Stem Cells.
- Shishkova, S., Rost, T.L., and Dubrovsky, J.G. (2008). Determinate root growth and meristem maintenance in angiosperms. *Ann. Bot.* *101*, 319–340.
- Shyh-Chang, N., Daley, G.Q., and Cantley, L.C. (2013). Stem cell metabolism in tissue development and aging. *Development* *140*, 2535–2547.

- Silver, J.D., Ritchie, M.E., and Smyth, G.K. (2009). Microarray background correction: maximum likelihood estimation for the normal-exponential convolution. *Biostatistics* *10*, 352–363.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* *31*, 3210–3212.
- Sinnott, E.W. (Edmund W. (1960). *Plant morphogenesis* (New York: McGraw-Hill).
- Smith, K.M., Gee, L., Blitz, I.L., and Bode, H.R. (1999). CnOtx, a Member of the Otx Gene Family, Has a Role in Cell Movement in *Hydra*. *Dev. Biol.* 392–404.
- Smith, K.M., Gee, L., and Bode, H.R. (2000). HyAlx, an aristaless-related gene, is involved in tentacle formation in *Hydra*. *Development* *127*, 4743–4752.
- Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* *3*, 1–25.
- Smyth, G.K. (2005). limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, (New York: Springer-Verlag), pp. 397–420.
- Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* *12*, 599–606.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* *30*, 2725–2729.
- Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nat. Rev. Genet.* *12*, 692.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* *22*, 4673–4680.
- Umetsu, D., and Dahmann, C. (2010). Compartment boundaries. *Fly (Austin)*. *4*, 241–245.
- Watanabe, H., Schmidt, H.A., Kuhn, A., Höger, S.K., Kocagöz, Y., Laumann-Lipp, N., Özbek, S., and Holstein, T.W. (2014). Nodal signalling determines biradial asymmetry in *Hydra*. *Nature* *515*, 112–115.
- Watson, P.J., Millard, C.J., Riley, A.M., Robertson, N.S., Wright, L.C., Godage, H.Y., Cowley, S.M., Jamieson, A.G., Potter, B.V.L., and Schwabe, J.W.R. (2016). Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nat. Commun.* *7*, 11262.
- Weinziger, R., Salgado, L., David, C., and Bosch, T. (1994). Ks1, an epithelial cell-specific

gene, responds to early signals of head formation in *Hydra*. *Development* *120*, 2511–2517.

Wittlieb, J., Khalturin, K., Lohmann, J.U., Anton-Erxleben, F., and Bosch, T.C.G. (2006). Transgenic *Hydra* allow in vivo tracking of individual stem cells during morphogenesis. *Proc. Natl Acad. Sci. U S A* *103*, 6208–6211.

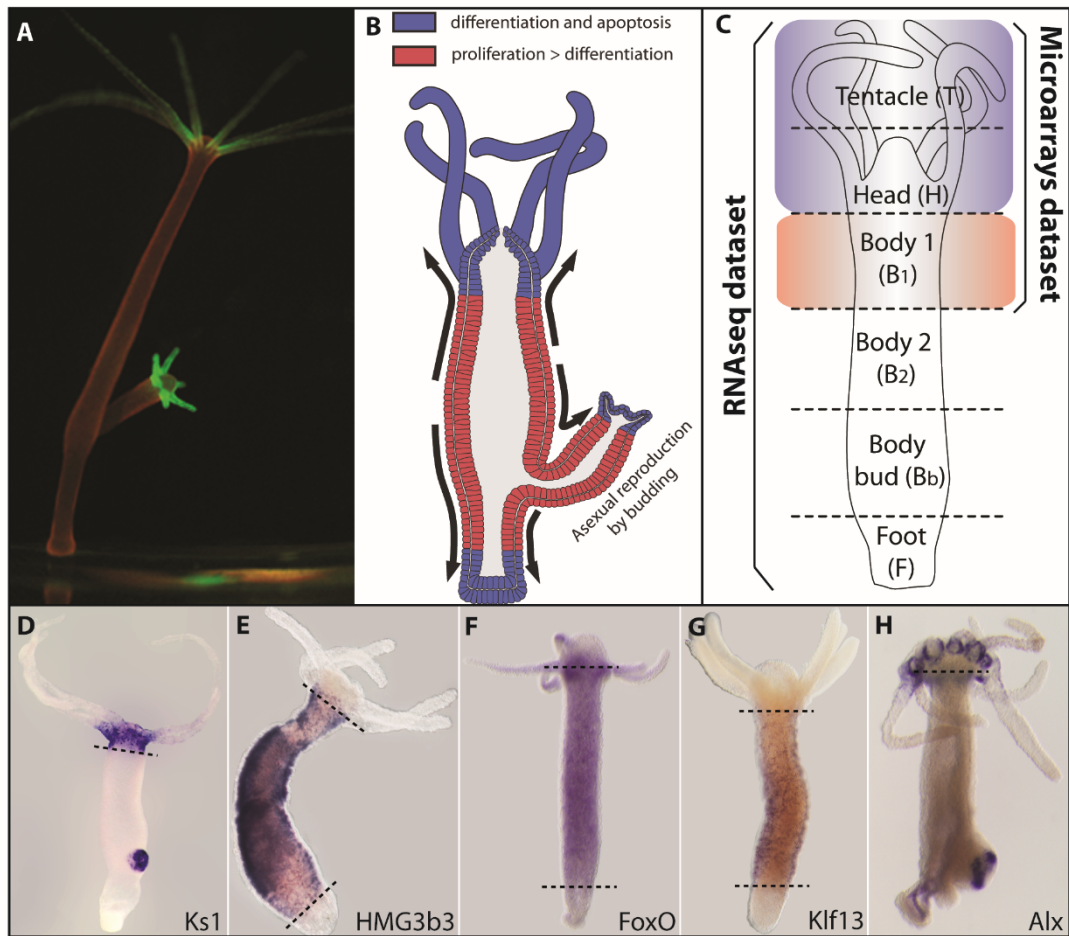


Figure 2.1. General approach followed to investigate *Hydra* boundaries. A. Transgenic *Hydra* used to visualize the head boundary. B. Regionalization of *Hydra* body. The arrows show the constant passive movements of the cells due to constant proliferation of stem cells in the body column. C. Collection of datasets used in this study. D-E. Examples of some genes in *Hydra* showing a position restricted expression. The dashed lines indicate the concerned boundary.

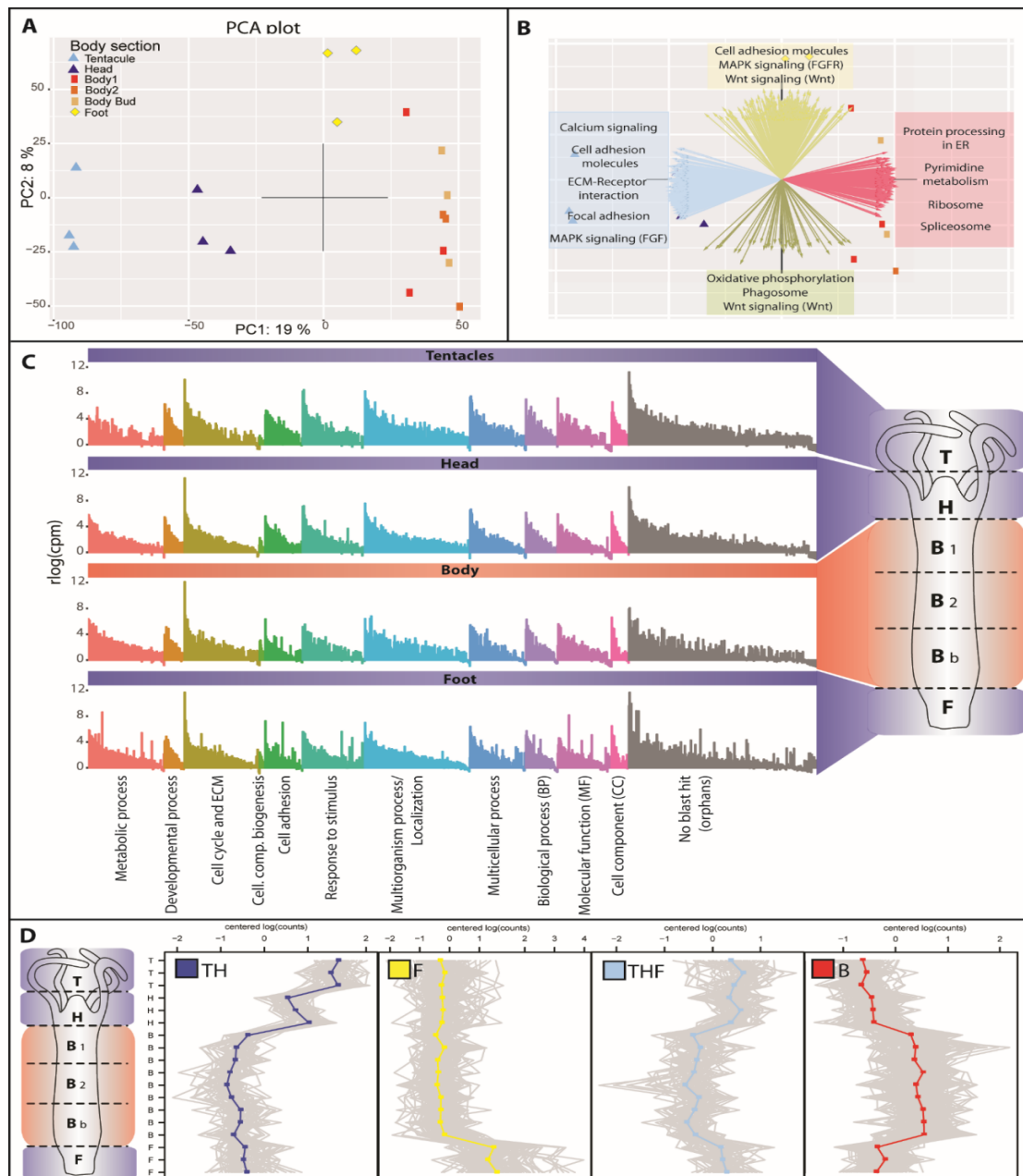


Figure 2.2. Identification of region-specific genes by comparative transcriptomics. A. RNA sequenced samples from *Hydra* body parts plotted in a principal component analysis. B. Correlation circle. The higher contributors to the variation from the PCA were plotted and annotated, the functional processes are shown in the direction where they contribute from the center of the PCA. C. Graphical representation of DEGs annotated by Gene Ontologies where the differential expression was at least of 2 fold among the parts. D. DEGs organized in clusters according to their expression in the different part of the body. Here we show 4 models of the clusters we selected for the analysis. TH: Genes expressed mostly in the Tentacle- Head region. F: Genes expressed specifically in the foot region. THF: Common genes expressed in differentiated tissue (tentacles, head and foot), and B: Genes overexpressed in the body column.

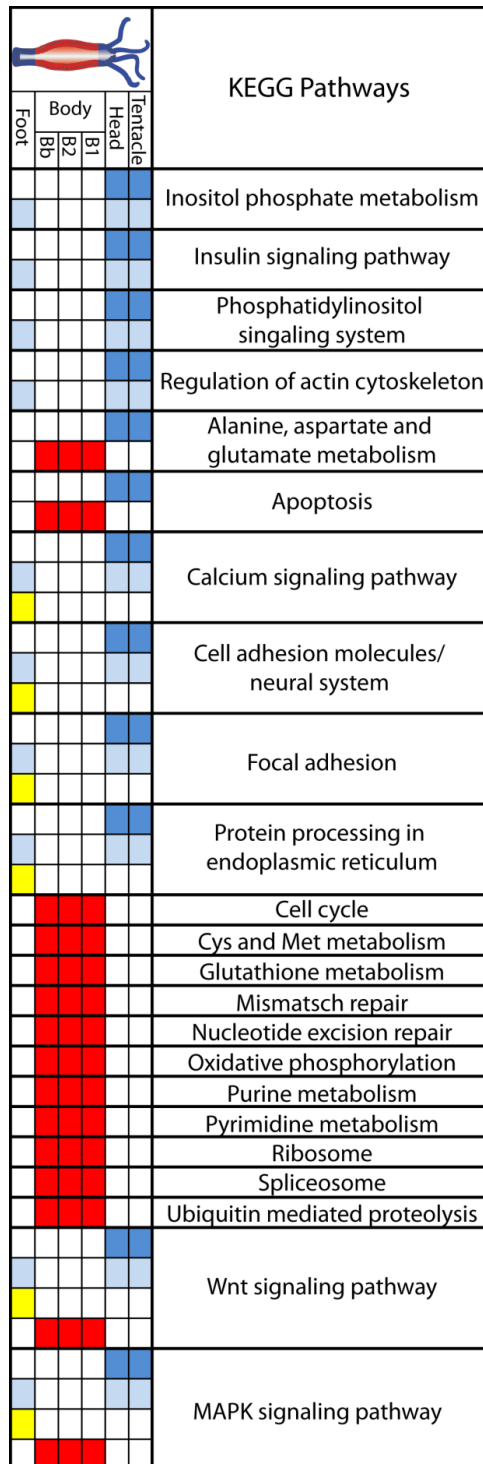


Figure 2.3. Summary of the unbiased RNAseq approach to identify compartment specific functional groups of genes. Group of clusters sharing similar expression pattern (Figure 2D) were annotated using KEGG pathways to identify the function and biological processes dominating in each body part. The colors correlate with the group of clusters from Figure 2,D. Notice that members a single pathway can be expressed in a different fashion according to the position in the *Hydra* body axis.

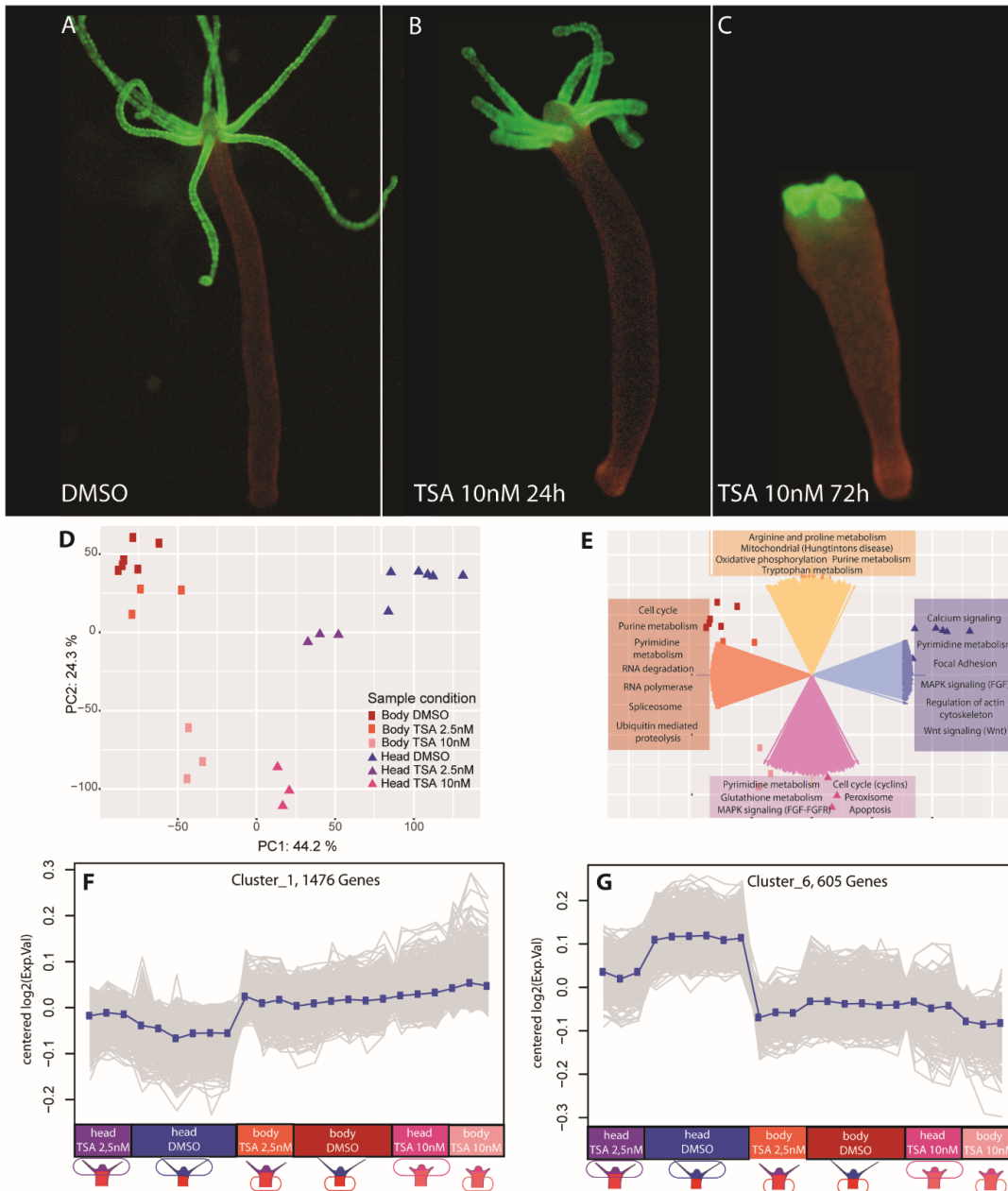


Figure 2.4. Changes in the transcriptional landscape upon TSA treatment. D: PCA from all the samples in the microarrays assay E. Correlation circle. The higher contributors to the variation from the PCA were plotted and annotated, the functional processes are shown in the direction where they contribute from the center of the PCA. The organization of the genes in clusters according to their expression in the different treatments gave rise to 2 main clusters. F: cluster 1, where genes that in the control (head DMSO) are downregulated in the head, after TSA treatment are upregulated like in the body. G. cluster 6, here the genes that normally are upregulated in the head, after TSA treatment show downregulation to a level as they are expressed in the body.

Figure 2.5. A higher expression of cell cycle genes is identified in head tissue when *Hydra* is treated with TSA. A. The genes from cluster 1 were annotated by blast and analyzed in the KEGG pathway database, here we show the cell cycle pathway. Gene expression pattern along the body axis and their change upon TSA treatment is shown in different colored boxes according to the legend below. The boxes without color correspond to genes that were either not giving a Blast hit or that were falling out in the signal cleaning process. Notice that most of the red colored boxes which correspond to genes expressed in the body compartment are also with a skin color, meaning that many genes with restricted expression to the body, are also sensible to TSA and change their expression pattern: they are also upregulated in the head upon TSA treatment. B-D: TSA treatment leads to BrdU labelled s-phase cells in head tissue in differentiated cells. Whole mount BrdU labeling with a lapse of 4 days incubation in DMSO (B), TSA 5nM (C) or TSA 10nM (D), with one time feeding and followed by 2 hours incubation in BrdU. The arrows show positive BrdU nuclei. Notice that in TSA treatments these nuclei are increased in the head and appear in the tentacles where they never did in the control. E-H: TSA allows cell to express the stem cell specific transcription factor FoxO. E: reporter construct used to check the expression pattern in living animals. The expression of foxO is concentrated in the body region of the ectodermal cells (GFP positive region). F: DMSO, G: transgenic animals kept 4 days in 5nM TSA, notice that ectodermal cells from head and foot are GFP positive. H: transgenic animals kept 10 days in 5nM TSA, notice that the detachment of the bud is disturbed.

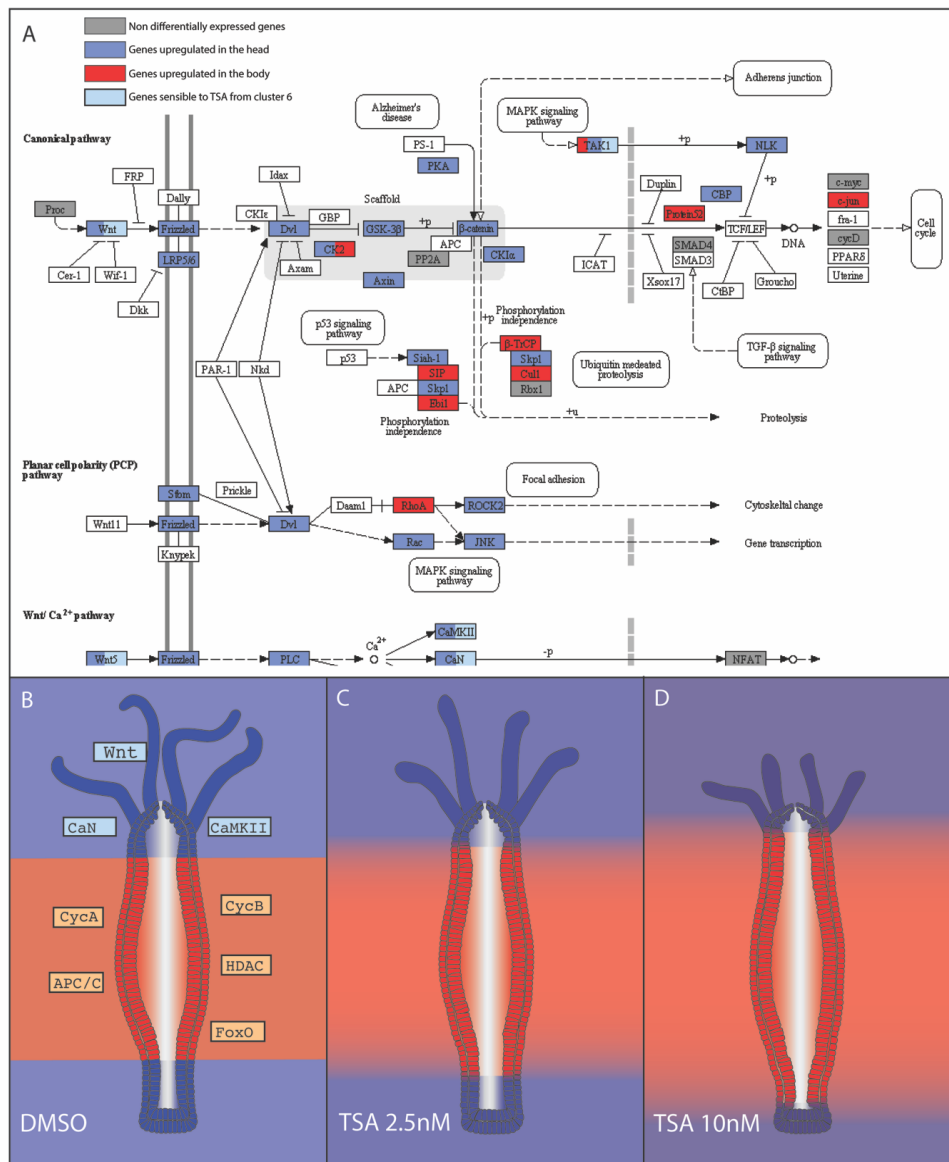


Figure 2.6. TSA treatment prevents the cells in the head expressing organizer genes. A: Here we show some genes belonging to cluster 6. The genes were annotated by blast and organized in KEGG categories. Their comparative expression according to the localization in the body axis is shown in different colored boxes according to the legend above. The boxes without color correspond to genes that were either not giving a Blast hit or that were falling out in the signal cleaning process. Notice that the Wnt proteins which are expressed specifically in the head (Blue), are also sensible to TSA been downregulated (see also figure 5, D). B-D: Model of action of a persistent histone hyperacetylation state in *Hydra*. Transition of a normal pattern into an extended stem cell compartment and a concurrent reduction of the differentiated compartments (most notorious in the tentacles). B: *Hydra* in normal conditions or in DMSO. C: After 2-3 days incubation in TSA 2.5nM. D: After 2-3 days incubation in TSA 10nM.

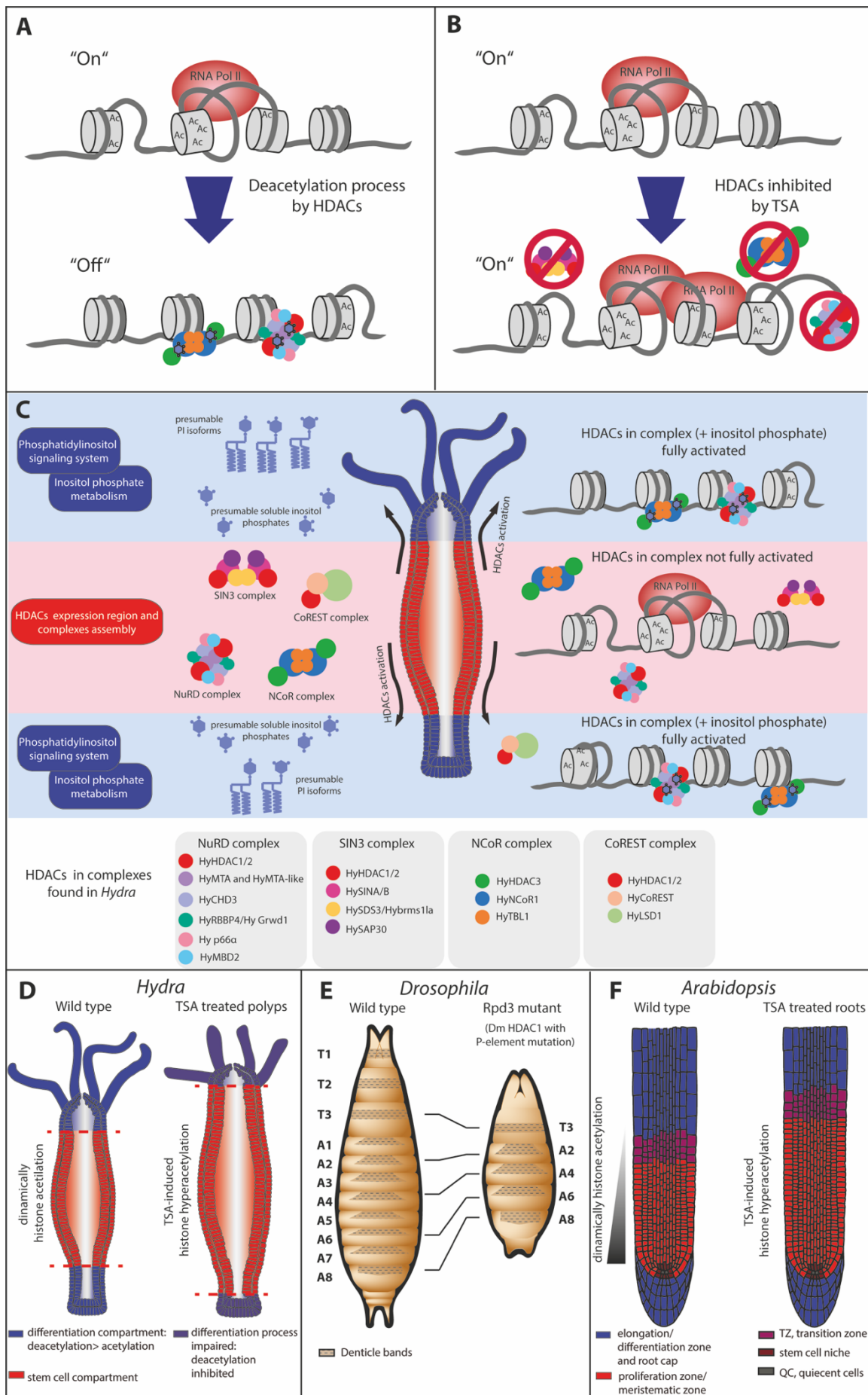


Figure 2.7. Histone deacetylation model: **A.** Histone deacetylation by HDACs (Millard et al., 2017, 2013b) leads conventionally to the inhibition of gene expression (Struhl, 1998) and the first compaction step of chromatin (Eskeland et al., 2010). **B.** TSA inhibits HDACs and leads to a histone hyperacetylation state, thereby many genes cannot be silenced. **C.** In *Hydra*, increased expression of members of the phosphatidylinositol signaling system and inositol phosphate metabolism pathways in differentiated tissue (head and foot) results in numerous phosphatidylinositols (PI) and soluble inositol phosphate isoforms which cause fully activation of HDACs in complex (see Watson et al., 2016). **D.** Exposure of *Hydra* polyps to low dose of TSA hampers the differentiation process, interferes with the maintenance of the boundary and results in a transition of the whole body into a stem cell compartment. **E.** *Drosophila* embryos of mutant Rpd3 (HDAC 1/2 in *Drosophila melanogaster*) germline clones display a drastically disturbed segmentation phenotype (Mannervik and Levine, 1999; Perrimon et al., 1996). **F.** A similar effect of TSA is seen on the Root Apical Meristem (RAM) from *Arabidopsis thaliana*. HDACs inhibition causes an increase in meristem size without increase in cell division which has been interpreted as a delay of differentiation (Rosa et al., 2014), suggesting that controlling the level of histone acetylation is fundamental for the timely transition into differentiation (Ikeuchi et al., 2015). Acetylation representation modified from (Morao et al., 2016).

3.

**Chapter II: Exome-wide association study
identifies *FN3KRP* and *PGP* as new
candidate longevity genes in humans**

Exome-wide association study identifies *FN3KRP* and *PGP* as new candidate longevity genes in humans

Guillermo G. Torres¹, Janina Dose¹, Marianne Nygaard^{2,3}, H el ene Blanch e⁴, Chantalat Sophie⁵, Wolfgang Lieb⁶, Lene Christiansen^{2,7}, Jean-Fran ois Deleuze^{4,5}, Kaare Christensen^{2,3,8}, Stefan Schreiber¹, Andre Franke¹, Friederike Flachsbart¹, Almut Nebel¹

¹ Institute of Clinical Molecular Biology, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany; g.torres@ikmb.uni-kiel.de, j.dose@ikmb.uni-kiel.de, f.flachsbart@mucosa.de, a.franke@mucosa.de

² The Danish Twin Registry and The Danish Aging Research Center, Department of Public Health, University of Southern Denmark, J.B. Winsloews Vej 9B, 5000 Odense C, Denmark; mnygaard@health.sdu.dk

³ Department of Clinical Genetics, Odense University Hospital, J.B. Winsloews Vej 4, 5000 Odense C, Denmark

⁴ Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 27 Rue Juliette Dodu, 75010 Paris, France; helene.blanche@fjd-ceph.org

⁵ Centre National de Recherche en G enomique Humaine CNRGH-CEA, 91000 Evry, France; deleuze@cng.fr, chantalat@cng.fr

⁶ Institute of Epidemiology and Biobank Popgen, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Niemannsweg 11, 24105 Kiel, Germany; w.lieb@epidemiology.uni-kiel.de

⁷ Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Tagensvej 20, 2200 Copenhagen N, Denmark; lene.christiansen.02@regionh.dk

⁸ Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Kl overv anget 47, 5000 Odense C, Denmark; kchristensen@health.sdu.dk

Corresponding author:

Almut Nebel, PhD

Institute of Clinical Molecular Biology

Kiel University

Rosalind-Franklin-Str. 12

24105 Kiel, Germany

phone: +49 (0)431 500 15155, fax: +49 (0)431 500 15168

email: a.nebel@mucosa.de

Abstract

Despite enormous research efforts, the genetic component of longevity has remained largely elusive. Since the investigation of common variants, mainly located in intronic regulatory regions in the genome, has yielded only little new information on the longevity heritability, we performed a chip-based exome-wide association study focusing on both common and rare variants in our German longevity cohort comprising more than 1,200 long-lived individuals, including 599 centenarians. In a single-variant analysis, we found rs1046896 in the gene fructosamine-3-kinase-related-protein (*FN3KRP*) to be significantly associated with longevity. An eQTL database search revealed the rs1046896 longevity allele to be associated with higher gene expression in various human tissues. A gene-based analysis, in which potential collective effects of common and rare variants were considered, yielded the gene phosphoglycolate phosphatase (*PGP*) as another potential longevity locus. With both the single-variant and the gene-based analysis, we validated previously reported longevity associations, i.e. for *CDKN2B* (rs1063192) and *OTOL1*, respectively. Replication of our results in French and Danish longevity collections, was successful for rs1063192 in the French; the longevity-association was further substantiated by a meta-analysis.

3.1. Introduction

Human longevity is a complex phenotype influenced by both genetic and environmental factors, which furthermore interact, e.g., via epigenetic changes [1]. Heritability estimates for longevity range from 12% [2] to 30% [3]; however, in the oldest-old, the genetic influence is even higher (up to 48%) [4]. Candidate studies have yielded many potential longevity associations, but most of them could not be replicated in independent investigations. To date, variation in only four loci has been confirmed to impact longevity across populations: *APOE* [5, 6], *FOXO3* [7, 8], the 5q33.3 locus [9, 10] and *CDKN2B* [11–13]. The first three reached genome-wide significance in large genome-wide association studies (GWAS) [9, 14]. However, the identified variants together only explain a small proportion of the longevity heritability. Thus, novel approaches are needed to identify additional loci involved in the phenotype.

Longevity studies have mostly been conducted following the common disease/common variant hypothesis, which is based on the assumption that the likelihood of becoming long-lived depends on a small number of single nucleotide variants (SNVs) that occur at high frequency in all populations. Yet, it has been estimated that common variants explain only 5 - 10% of the heritability of complex traits [15–17]. Rare variants may contribute a considerable fraction to the heritability of complex traits, i.e. according to initial estimates, half the heritability explained by common SNVs [18]. Therefore, rare variants may substantially regulate longevity, but they are often not covered by genome-wide genotyping arrays or imputations [14]. Common variants often reside in regulatory/intronic regions [19], and GWAS SNPs for complex traits are enriched in functionally important intronic regions [20]. However, rare variants in or near coding regions are thought to have larger effects on the trait of interest than rare or common regulatory variants [19]. To date, only in a few longevity studies, the coding regions of the genome have been targeted or covered, e.g. by the application of exome-based genotyping [21] or whole-genome [22] or exome-sequencing [23–27]. These approaches yielded important results, not only in terms of new longevity variants (e.g. mutations in *CLEC3B* and *HLA-DQB1* [21, 26]); they also provided novel analysis methods that could extend the discovery spectrum of loci and genes influencing complex traits (e.g. analysis for an accumulation and/or combination of multiple variants) [11, 28].

Here, we performed an exome-wide association study (EWAS) in a large German longevity cohort comprising more than 1,200 centenarians and nonagenarians. We used the Illumina Infinium HumanExome BeadChip that covers both rare and common variants. In a single-

variant analysis, we found rs1046896, located in the gene fructosamine-3-kinase-related-protein (*FN3KRP*), to be significantly associated with longevity. An eQTL database search revealed the rs1046896 longevity allele to be associated with higher gene expression in various human tissues. We furthermore performed a gene-based analysis to consider potential collective effects of individual common and rare variants. This analysis yielded, next to *FN3KRP*, the gene phosphoglycolate phosphatase (*PGP*) as a potential longevity locus. With both the single-variant and the gene-based analysis, we validated previously reported longevity associations, i.e. for *CDKN2B* (rs1063192) and *OTOL1*, respectively. Replication of our results in French and Danish longevity collections was successful for rs1063192 in the French; the longevity-association was further substantiated by a meta-analysis. By considering SNV-SNV interactions, we were able to define a functional context highlighting the relevance of Tie2, PI3K/AKT and mTOR signaling, as well as calcium mobilization in healthy aging and longevity.

3.2. Results

3.2.1. Single-variant association analysis reveals a longevity association of rs1046896 in *FN3KRP*

To identify new common and rare genetic variants with moderate to high penetrance associated with longevity, a chip-based EWAS was conducted using the Illumina Infinium HumanExome BeadChip (**Figure 3.1**). This array covers rare and common variants in a ratio of approximately rare:common = 8:1 [29]. In total, 1,248 German LLI and 6,941 younger controls were included in the study. The genotyping analysis was performed based on 62,488 SNVs, and 1,212 LLI and 6,762 younger controls, which had remained after QC (**Supplementary figure 1, Supplementary figure 2**). The single-variant association approach yielded 11 candidate SNVs ($P < 1 \times 10E-04$) (**Table 3.1**, and **Supplementary figure 3**). Among the associated variants, we identified rs2075650, rs4420638, and rs769449 located in the *TOMM40/APOE/APOC1* region, a region that is well known to be negatively associated with longevity [5, 9, 11, 30]. Apart from these SNPs, the best association signal was obtained for rs1046896 in the gene *FN3KRP* (minor allele frequency (MAF) = 0.32, $P = 7.40 \times 10E-07$; **Table 3.1**), which passes the exome-wide association significance threshold of $1 \times 10E-06$ for common variants [31]. In addition, we observed a longevity association for rs1063192 in the *CDKN2B-AS1* region ($2.99 \times 10E-05$; **Table 3.1**), which has previously been reported in the context of human longevity [11].

The effects of the candidate variants were investigated for independency of the *TOMM40/APOE/APOC1* locus by a conditional association test. The results of the conditional analysis confirmed the associations of rs1046896 (*FN3KRP*), rs55882518 (*NOTCH3*), rs1063192 (*CDKN2B-ASI* region), rs1319846 (*TMEM131L*) and rs1790706 (*DSC2*) with longevity (**Table 3.1**). Additionally, an association analysis with the centenarian subpopulation (n=599 individuals ≥ 100 years) was performed. Although the centenarian subset comprised substantially fewer individuals, the association analysis yielded higher ORs for rs1046896 (*FN3KRP*), rs55882518 (*NOTCH3*), 1063192 (*CDKN2B-ASI* region) and rs184214819 (*SPZ1*) (**Supplementary table 2**). This effect has been reported before [32], and is consistent with the greater genetic influence with increasing age.

Since we had noticed that our best-associated new candidate longevity SNV rs1046896 was located in a 3'UTR region and might therefore have structural implications for the mRNA stability [33]. (i.e. might affect gene expression in an allele-dependent manner), we investigated whether this SNV or SNVs in high LD ($r^2 > 0.8$ based on HapMap-CEU individuals; 1,000 Genomes phase 3 [34, 35]) influences local or distant gene expression. Using the publicly available databases Blood eQTL browser (<https://genenetwork.nl/bloodeqtlbrowser/>) [36] and the GTEx eQTL database (<https://www.gtexportal.org/home/>, accessed April 5 2019), we observed significant cis-eQTL associations of rs1046896 (and high-LD SNVs) with its vicinity genes, especially with the expression of *FN3KRP* and *FN3K* in several tissues (e.g. brain regions, testis, pancreas; **Supplementary table 3**). For all reported tissues, *FN3KRP* gene expression was higher in the presence of the rs1046896 longevity allele C (major allele) (**Supplementary table 3, Supplementary figure 4**).

3.2.2. The longevity-association of rs1063192 (*CDKN2B*) replicates with borderline significance in an independent cohort

We aimed for replication of the association results in two independent cohorts. Sample sizes of the Danish and French cohorts (Danish: 1,002 LLI, 737 younger controls and French: 1,264 LLI and 1,830 younger controls) limited the replication approach to the common variants rs1046896 (*FN3KRP*), rs1063192 (*CDKN2B*), and rs1319846 (*TMEM131L*). The association of rs1063192 (*CDKN2B*) reached borderline significance in the French ($P = 0.056$, OR = 1.14; **Table 3.2**), but not in the Danish ($P = 1.00$, OR = 1.04; **Table 3.2**). The other SNPs did not replicate, either in the French or in the Danish (**Table 3.2**).

In a meta-analysis, we observed large inconsistency (I^2) of the genetic effects across the three studies ($I^2 > 75\%$) for rs1046896 (*FN3KRP*) and rs1319846 (*TMEM131L*), but moderate I^2 ($25 > I^2 < 75\%$) for rs1063192 (*CDKN2B*) (**Table 3.2**). The between-population heterogeneity of the genetic effects for rs1063192 (*CDKN2B*) was estimated as 40.72% (**Table 3.2**). To overcome a potential heterogeneity bias, we used the random effects summary odd ratio (OR(R)). The strongest evidence for an association with longevity was observed for rs1063192 (*CDKN2B*) (OR(R) = 1.14, $P = 0.00174$; **Table 3.2**). The association for this SNV (or others in high LD like rs4977756) had been already replicated previously by others [11–13], however, in the Danish sample, we could not detect any association with longevity. Both the slightly smaller sample size (1,003 cases compared with 1,248 cases in the German and 1,264 cases in the French cohort) and the older age of the Danish (< 71 years) compared with the German (< 60 years) or French (< 62 years) controls might have contributed to the missing replication.

3.2.3. Gene-based analysis reveals *PGP* as a potential new longevity locus and strengthens the *FN3KRP* association

In addition to the single-variant association approach, we assessed the cumulative effects of common and rare variants within one genomic region. In the gene-based association test, 13 genes (apart from *TOMM40/APOE/APOC1*) were identified with an enriched burden of rare and common variants ($P < 1 \times 10E-04$). However, only *PGP* survived Bonferroni correction in both the SKATO and the burden test (**Table 3.3**). Of the 13 genes, four (*FN3KRP*, *GRN*, *SKORI*, *SPZI*) had already been observed in the single-variant association analysis. With the gene-based analysis, we validated the previously reported association of *OTOLI* with human longevity (rs1425609; [37]). In the present work, *OTOLI* reached a P -value $< 1.61 \times 10E-06$ in the burden test. An association analysis using the centenarian subset only, yielded significant associations ($P < 3 \times 10E-06$) for seven genes (apart from *APOE*) that survived Bonferroni correction (**Supplementary table 4**). Of these seven genes, *FN3KRP* and *HMHA1* overlapped with the gene-based analysis using the whole German study population. The remaining five genes (*TMEM14A*, *IRAK1BP1*, *ACPP*, *PLXNB1*, and *GARI*) were identified in the centenarian subpopulation only (**Supplementary table 4**).

3.2.4. SNV-SNV interaction analysis reveals a functional longevity framework

To extend the understanding of the complex biology of longevity, we performed a SNV-SNV interaction analysis. A total of 52,310 SNVs (all SNVs from the array apart from those in linkage disequilibrium (LD) and those on the gonosomes) were used to test all possible

interactions with SNPsyn [38] without *a priori* hypothesis. For 86 SNVs, we found 148 significant interacting pairs with positive synergy, i.e. each SNV conveys non-redundant information (**Supplementary table 5**). The genes annotated to the 86 SNVs and those annotated to the significantly associated SNVs from the association analyses (in total, 90 core genes) were uploaded to GeneMANIA [39] and ClueGo [40] to predict their functional implications. Based on the genome-wide map of human genetic interactions published by Lin *et al.* [41], GeneMANIA confirmed most of the interactions generated with SNPsyn (**Supplementary figure 4**). Additionally, GeneMANIA predicted 20 genes to be functionally related to the 90 core genes (**Supplementary table 6**). From the whole set of 110 genes, a functional map of longevity was generated using ClueGo that comprised seven functional groups (RNA modification; regulation of cholesterol transport; role of LAT2 on calcium mobilization; insulin receptor pathway; IGF1R signaling cascade; Tie2 signaling, **Figure 3.1** and **Supplementary table 7**).

3.3. Discussion

In our large German longevity cohort comprising more than 1,200 LLI (incl. ~600 centenarians), we screened 62,488 common and rare exonic SNVs for an association with longevity. We discovered rs1046896 in the 3'UTR region of the gene fructosamine-3-kinase-related-protein (*FN3KRP*) as a novel longevity-associated variant that reached exome-wide significance. Additionally, we replicated the longevity-association of rs1063192 in *CDKN2B-ASI* [11] as well as rs2075650, rs4420638, and rs769449 in the *TOMM40/APOE/APOC1* region [9, 11]; the latter three are well known for being both negatively associated with longevity and positively associated with Alzheimer's disease [6, 42].

Our meta-analysis, which included data from our German cohort as well as from French and Danish sample collections, revealed very high between-population heterogeneity for rs1046896 in *FN3KRP*. This might indicate that the identified SNV is not the causal variant and shows different LD patterns with the causal polymorphism in the three populations included in the meta-analysis [43]. However, based on the eQTL prediction analysis (**Supplementary table 3**, **Supplementary figure 4**), rs1046896 and SNVs in high LD appear to influence the stability of *FN3KRP* and thereby regulate its expression. In future studies, a fine-mapping of the *FN3KRP* gene region is needed to identify the causal variant.

FN3KRP belongs to a gene family with an important role in the reversal of the non-enzymatic

glycation of proteins [44]. Glycation adversely affects protein function, which leads to arterial stiffening [45] in particular and to an accumulation of damaged proteins. This in turn promotes aging [44] and the development of age-related diseases [46]. Both FN3RKP and fructosamine-3-kinase (FN3K), which show 65% sequence similarity, seem to protect proteins from non-enzymatic glycation and to stop the formation of certain advanced glycation end products [44, 47]. Strikingly, rs1046896-T, which showed a lower allele frequency in the LLI in our study (OR_{cond.} = 0.77; OR_{cond._centenarian} = 0.70), was identified as a risk locus for glycated hemoglobin (HbA1c), a critical non-enzymatic glycation product used to monitor and diagnose diabetes [48]. Our findings, together with the substantial role of FN3KRP in cell maintenance and viability, support it as a promising candidate which may facilitate healthy aging and longevity by preventing physiological deterioration.

Interestingly, rs1063192-G in the *CDKN2B* region, which was enriched in the LLI of our sample and whose longevity-association was further supported by our meta-analysis, was reported to be a protective variant in glaucoma, a classical age-related disease [49], which is also characterized by an increased burden of advanced glycation end products [46].

Apart from rs1046896 (*FN3KRP*), rs1063192 (*CDKN2B-AS1*), and rs2075650, rs4420638, and rs769449 (*TOMM40/APOE/APOC1*), six additional SNVs showed an association with longevity ($P < 2 \times 10^{-4}$, **Table 3.1**). We found exome-wide significant (EWS) associations only for the three variants in the *TOMM40/APOE/APOC1* region and rs1046896 (*FN3KRP*). However, variants without EWS (or genome-wide significance) may still reflect true association signals with biological relevance. It was shown that eight of the ten top CHARGE SNVs without GWS from a meta-analysis of GWAS [30], correspond to mouse lifespan quantitative trait loci (QTL) [50]. Of the 11 SNVs, for which we report a longevity association, four represented rare variants (MAF < 0.05%, **Table 3.1**). In a recent study, in which the exomes of 100 LLI were sequenced, no rare protein-altering SNV was observed to be enriched in the genomes of the LLI compared to younger controls [27]. However, our results show that rare coding variants may be drivers of longevity.

Studies on other complex traits have shown that gene-based tests, in which all common and rare SNVs within a gene locus are considered jointly, can be more powerful than single-variant association approaches. When we examined the cumulative effect of common and rare variants within each gene, we discovered *PGP* (cumulative effect of three variants: two rare and one

common) as another novel longevity-associated locus ($P = 8.90 \times 10E-07$). Via its function as glycerol-3-phosphate (Gro3P) phosphatase, PGP controls the levels of Gro3P, which is an important metabolite formed during glycolysis [51]. The availability of Gro3P is crucial for the regulation of both glucose and fat metabolism and eventually determines the generation of signaling and regulatory molecules, which further affect many biological processes (e.g., insulin secretion and sensitivity, inflammation, fat synthesis and storage, (cancer) cell proliferation) [52]. Therefore, PGP may aid in the detoxification of excess nutrient/fuel supplies, thereby preventing metabolic stress with its associated pathologic conditions, e.g. type 2 diabetes, cardiovascular diseases and metabolic syndrome [52–54], and eventually facilitating longevity.

In the gene-based test, apart from PGP and *TOMM40/APOE/APOC1*, we identified 17 additional potential longevity genes ($P < 1 \times 10E-04$), including otolin 1 (*OTOL1*). One intronic SNV in *OTOL1*, rs1425609, has already been associated with longevity in a single-variant analysis in a previous study [37]. The fact that the gene-based test led to the replication of a longevity-association, whose identification was based on a single variant, not only validated the gene-based approach, but also points towards the added value of analyzing both common and rare SNVs jointly for explaining the missing longevity heritability. Next to *OTOL1*, the 17 genes included the ribosomal protein S6 kinase B1 (*RPS6KBI*), a downstream effector of the nutrient-responsive mTOR (mechanistic target of rapamycin kinase), which was shown to promote longevity in yeast, worms, and flies [55]. Furthermore, *TMEM14A*, *IRAK1BPI*, *ACPP*, *PLXNBI*, and *GARI*, all involved in cell proliferation and cell growth [56–60], also reached nominal statistical significance. Noteworthy, the gene *FN3KRP* reached a P -value of $P = 3.08 \times 10E-06$ and survived multiple testing correction in the centenarians subset (**Supplementary table 4**). This supports the *FN3KRP*-longevity association to be a true-positive signal. An accumulation of common and rare variants has already been observed in genes associated with other complex traits like type 2 diabetes [61] and Alzheimer's disease [42]. With respect to extreme aging, the genes *LYST*, *MDN1* and *RBMXLI* were found to harbor an increased burden of rare coding variants in LLI versus younger controls; however, with nominal significance only [27]. To our knowledge, joint effects of common and rare variants on human longevity have not been investigated yet in a cohort of comparable size like ours. In a recent study, sets of variants with low and rare frequency ($MAF < 0.05$) were analysed for an association with longevity in East Asians (530 nonagenarians/centenarians). More than 100 genes reached nominal significance in that study; however, EWS was not met by any of the genes [21] and we

could not identify an overlap with our results.

SNV-SNV interaction analysis predicted genetic interactions between 86 genes that, together with the longevity-associated genes annotated to the significant SNPs from the association analyses in the present work, may depict the molecular and physiological pathways underlying human longevity. The functional network of the longevity-associated genes identified in this study tended to form hubs, which were determined rather by co-expression and genetic interaction data than by co-localization or physical interaction (**Supplementary figure 4**). Overall, our functional network supported previous reports on pathways that likely control longevity, i.e. the mTOR and insulin/IGF1 signaling pathways [55, 62]. Tie2 signaling has been suggested to be indispensable for neovascularization and protection of age-related macular degeneration and glaucoma [63, 64], which is in concordance with our findings for rs 1063192 and *FN3KRP*. Telomere maintenance has also been suggested to be crucial for attaining longevity [62]. In our functional network, the genes *GARI* and *NAT10*, both involved in rRNA modification, were reported to be telomerase-associated proteins, and particularly over-expression of *NAT10* was shown to induce telomere shortening [65].

3.4. Conclusion

With our study, we contribute to the genetic framework of longevity with two new potential candidate genes, *FN3KRP* and *PGP*, which were identified by single-variant and gene-based analyses, respectively. The two genes influence longevity likely by their role in metabolic processes, i.e. the reverse glycation of proteins (*FN3KRP*) and control of Gro3P levels (*PGP*). However, with respect to *FN3KRP*, the variant that we report here (rs1046896) is unlikely to be the causal variant, considering the high between-population heterogeneity values in the meta-analysis of the German, Danish and French association results. Future fine-mapping studies are warranted to identify the true functional variant in high LD with rs1046896. With the combination of analysis methods and, in particular, the investigation of cumulative effects of common and rare variants within one genetic region, we are a step closer to accounting for the missing heritability of human longevity.

3.5. Materials and methods

German, Danish and French study populations

The German sample comprised 1,248 German LLI (male/female ratio: approximately 1/3; age range: 94 - 110 years; mean age: 99) as described previously [6, 8]. Briefly, the control sample

contained 6,941 younger individuals (age < 60 years) from a German population-based collection recruited with the help of the biobank popgen; in detail, 2,905 adults from the KORA study (Collaborative Health Research in the Region of Augsburg) [66], 2,360 individuals from HCDEBONN, 998 Germans from the popgen biorepository [67], 473 participants from HCDEMICK, and 205 additional individuals from HCGCON. All study participants provided a written informed consent prior to enrolment in the study. Approval for the project was obtained from the Ethics Committee of the Medical Faculty of Kiel University.

For the Danish data set, the 1,003 cases (male/female ratio 1/3, age range: 91 - 101 years) consisted of participants drawn from seven nation-wide surveys collected at the University of Southern Denmark; the Study of Danish Old Sibs (DOS), the 1905 Birth Cohort Study, the 1910 Birth Cohort Study, the 1911-12 Birth Cohort Study, the 1915 Birth Cohort Study, the Longitudinal Study of Danish Centenarians (LSDC), and the Longitudinal Study of Ageing Danish Twins (LSADT). Briefly, DOS was initiated in 2004 and included families in which at least two siblings were ≥ 90 years of age at intake. The LSDC and 1905, 1910, and 1915 Birth Cohort Studies were prospective follow-up studies initiated in 1995, 1998, 2010, and 2010, when participants were 100, 92-93, 100, and 95 years of age, respectively [68]. The 1911-1912 Birth Cohort Study consisted of individuals reaching the age of 100 years in the period from May 2011 to July 2012 [69] and LSADT was initiated in 1995 and includes Danish twins ≥ 70 years of age [70]. From DOS and LSADT, one individual from each sib-ship or twin pair was randomly selected among participants that had reached an age of ≥ 91 years for DOS, and ≥ 90 years for LSADT. From the 1905 and 1915 Birth Cohort Studies, participants were selected among individuals reaching an age of minimum 96 years. The 1,189 controls (age < 71 years) consisted of individuals recruited by the Danish Twin Registry (DTR) as part of the study of Middle-Aged Danish Twins (MADT). MADT was initiated in 1998 and included 4,314 twins randomly chosen from each of the birth years 1931-1952 [71]. Surviving participants were revisited from 2008 to 2011 [72], where the blood samples used for DNA extraction were collected. Signed informed consents were obtained from all participants. Collection and use of biological material and survey information were approved by the Regional Committees on Health Research Ethics for Southern Denmark, and the study was approved by the Danish Data Protection Agency.

The French data set comprised 1,264 individuals (male/female ratio $\sim 1/4.5$, mean age 102.4 years; age-range 91-115+ years) [73]. French siblings were recruited when at least two siblings

fulfilled the age criterium of 90 years or older; in such a case the oldest sibling was selected for the study. All subjects signed a written informed consent form in accordance with the local review board. The French controls consisted of 1,830 subjects (age < 62 years) selected in a population-based sample of French subjects that had participated in the Supplementation in Vitamins and Mineral Antioxidants (SU.VI.MAX) study [74].

Exomechip genotype calling and quality control

Study samples were genotyped on the BeadChip Illumina HumanExome-12v1.1 (N = 244,770 SNVs) or BeadChip Illumina HumanExome-12v1.2 (N = 247,870 SNVs) (Illumina Inc., San Diego, USA). Genotype calling was performed using the GenomeStudio software and the GenTrain v2.0 clustering algorithm (Illumina Inc., San Diego, USA). To improve genotype calling, zCall software v3 [75] and call rate > 95% was used. In order to avoid false positive associations, an independent quality control (QC) was performed for samples and SNVs. QC was done with the software Plink 1.9 [76]. In total, 224 samples were removed because they had failed one or more of the following inclusion criteria: discordant sex information, missing genotype < 3%, heterozygosity rate greater or lower than $\pm 4sd$ from the mean, and no relatedness of individuals. Relatedness was estimated using identity by descent metric (IBD) [77]. For related individuals (IBD > 0.1875; halfway between the expected IBD for third- and second-degree relative [77], only one individual was included in the analysis. SNVs were excluded if the missing rate was too high (> 3%) or, for common SNVs in the control sample, if they deviated from Hardy-Weinberg equilibrium ($P < 0.0001$). Due to the lack of sufficient statistical power, SNVs with extremely low minor allele frequency (MAF < 0.003, equivalent to a minor allele count (MAC) ≥ 21) were removed. Power calculations were carried out with the

the	Gas	Power	Calculator
-----	-----	-------	------------

 (http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html). For the single-variant analysis, the Bonferroni-corrected P -value threshold was based on the number of markers tested: $P < 8 \times 10E-07$ (significance threshold $0.05/62,488$ number of markers tested). Concordantly, the gene-based Bonferroni-corrected P -value threshold was based on the number of gene sets tested: $P < 3.3 \times 10E-06$ (significance threshold of $0.05/14,790$ gene sets).

Prior to further analysis, a confounder correction was conducted based on gender, linkage disequilibrium (LD) and population stratification. Population substructure was evaluated with the principal component analysis (PCA) using a common set of independent markers (N = 16,782 SNVs). The principal components (PC) were calculated using PLINK 1.9 [76], and the

first five PCs were selected to identify potential confounders for the correction. Additionally, stratification outliers were identified based on the local outlier factor ($LOF > 1.7$) [78] and excluded to mitigate population stratification.

Association analysis in the German sample

Single-variant association analysis was performed using the logistic regression test implemented in PLINK1.9 [76], assuming an additive genetic model. Candidate longevity SNVs with a discovery P -value $< 1 \times 10E-04$ were selected for replication. This threshold is less stringent than $1 \times 10E-06$ which was proposed for EWAS [31]. Relaxation in the selecting threshold allows the identification of longevity SNVs which usually have small effects on the phenotype [79]. To identify additional association signals and to test for independency of the newly identified SNVs from the effects of the known longevity-associated locus *TOMM40/APOE/APOC1* [9, 11, 30], a conditional association test was performed considering the SNVs rs2075650, rs4420638, and rs769449. The conditional test was done using the logistic regression test in PLINK1.9 [76]. The genomic inflation was estimated based on the P -values of the association. The Sanger imputation service (<http://www.sanger.ac.uk/science/tools/sanger-imputation-service>) and the 1000 Genomes phase I v.3 reference panel was used to enrich the common SNV pool.

In addition to the single-variant association testing, a gene-based analysis was performed. Potential cumulative effects of rare and common variants with longevity were tested using both burden and non-burden (i.e. SKAT) approaches from the algorithm RC-SKAT [80]. Both approaches were considered because burden tests were shown to perform better if multiple variants in the regions are causal and influence the phenotype in the same direction [81], while non-burden tests, like SKAT [82], are more advantageous if SNVs in the region interact or show opposing directions of effect [83]. From the 244,770 SNVs on the HumanExome-v12 BeadChip, a gene-set file was prepared using the R-package biomaRt [84]. The overall effect of rare and common variants in a gene was evaluated based on the adaptive sum test [80] in combination with either burden or SKAT.

Synergistic effects among SNVs were assessed using SNPsyn [38]. SNPsyn uses an information-theoretic approach for a synergistic interaction analysis. A synergy between a pair of SNVs is measured as the difference between the information of the SNV-SNV phenotype and the sum of the information encoded by the two individual SNVs [38]. Shared information

(depicted in the information score) between SNV pairs can be either positive (synergistic SNVs) or negative. If negative, the two SNVs carry redundant information; an effect that is frequently observed among highly correlated SNVs. The significances of the information scores were calculated using permuting genotype data across samples 500 times. The scores were corrected for multiple testing using false discovery rate (FDR) [85]. SNV interactions were considered significant if the corrected P -value < 0.001 .

The list of core genes (annotated to SNVs from both the SNV-SNV interaction analysis and association analyses) was used together with GeneMANIA Cytoscape plugin (under default parameters) [39] to enrich the functional context of the gene-associated candidates. GeneMANIA uses a network weighting approach to determine how genes in a gene list are connected to one another or for determining which types of functional or genomic data are the most useful to collect for finding more genes like those in the query list [86]. Later, functional annotation was performed in the context of the Gene Ontology (GO) terms [87], KEGG [88] and REACTOME [89] pathways using ClueGo Cytoscape plugin [40]. ClueGo creates first a binary gene-term matrix using gene's annotations, similar as described by Huang *et al.* [90]. GO terms and pathways with less than three annotated genes were filtered out. A chance-corrected measure of co-occurrence between pair of terms (kappa statistics) was used to determine the association strength between terms [91]. Thus, ClueGo creates a network that represents the terms as nodes linked based the kappa score level (> 0.3). Functional groups in the network were created by iterative merging and those with a corrected (FDR) P -value < 0.05 were selected for the representation.

Genotyping in the replication cohorts

DNA was extracted from dried blood spot cards using either the DNA Mini or Micro Kits (Qiagen, Hilden, Germany) or the Extract-N-Amp™ Blood PCR Kit (Sigma-Aldrich, St. Louis, MO, USA) followed by amplification using the GenomePlex Complete Whole Genome Amplification (WGA) Kit (Sigma Aldrich, St. Louis, MO, USA), or from whole blood using a manual [92] or a semi-automatic (Autopure, Qiagen, Hilden, Germany) salting out method.

Danish LLI were genotyped using the Illumina HumanOmniExpress Array (Illumina Inc., San Diego, CA, USA). Pre-imputation quality control included filtering of SNPs on genotype call rate $< 95\%$, HWE $P < 10^{-4}$, and MAF $< 1\%$, and individuals on sample call rate $< 95\%$, relatedness and gender mismatch. After imputation to the 1000 Genomes phase I v.3 reference

panel, genotype probabilities were converted to hard-called genotypes in Plink (using a cut-off of 90%) [76]. For the Danish controls, data of the SNPs rs1063192, rs1046896 and rs13119846 was extracted from quality controlled genotype data created using the Illumina Infinium PsychArray (Illumina Inc., San Diego, CA, USA).

French individuals were genotyped by TaqMan (Thermo Fisher Inc., Waltham, Massachusetts, USA) on a 7900HT Fast Real-time PCR System (Thermo Fisher Scientific Inc., Waltham, USA). Association analysis was performed using logistic regression accounting for gender as covariate in the software Stata [93] and Plink [76].

Meta-analysis was performed with Plink [76]. Fixed and random effect models were used to estimate the effect size. Cochran's Q statistics and I² metric were calculated to measure between-study heterogeneity.

Acknowledgements

Genotyping of the Danish controls was conducted by the SNP&SEQ Technology Platform, Science for Life Laboratory, Uppsala, Sweden (<http://snpseq.medsci.uu.se/genotyping/snp-services/>).

Conflict of interest

The authors declare that they have no competing interests.

Funding

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 22167-390884018. The biobank PopGen and the PopGen 2.0 network are financed by the BMBF (grant 01EY1103). The CEPH centenarian cohort, maintained at the CEPH Biobank, Paris, France [BIORESOURCES], was supported by the “Ministère de l'Enseignement supérieur et de la Recherche”. The SU.VI.MAX cohort was funded by the French Institut National de la Santé et de la Recherche Médicale, the Institut National de la Recherche Agronomique, the Université Paris 13. The genotyping of French samples was supported by the Commissariat à l'Energie Atomique-Centre National de Recherche en Génomique Humaine. Danish replication study was supported by The National Program for Research Infrastructure 2007 (grant no. 09-063256), the Danish Agency for Science Technology and Innovation, the Velux Foundation, the US National

Institute of Health (P01 AG08761), the Danish Agency for Science, Technology and Innovation/The Danish Council for Independent Research (grant no. 11-107308), The Danish Interdisciplinary Research Council, the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679, the INTERREG 4 A programme Syddanmark-Schleswig-K.E.R.N. (by EU funds from the European Regional Development Fund), the CERA Foundation (Lyon), the AXA Research Fund, Paris, and The Health Foundation (Helsefonden), Copenhagen, Denmark.

References

1. Morris BJ, Willcox BJ, Donlon TA. Genetic and epigenetic regulation of human aging and longevity. *Biochim. Biophys. Acta Mol. Basis Dis.* 2019; 1865:1718-1744.
2. Ruby JG, Wright KM, Rand KA, et al. Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics* 2018; 210:1109-1124.
3. vB Hjelmberg J, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Kaprio J, Pedersen NL, Christensen K. Genetic influence on human lifespan and longevity. *Hum. Genet.* 2006; 119:312-321.
4. Brooks-Wilson AR. Genetics of healthy aging and longevity. *Hum. Genet.* 2013; 132:1323-1338.
5. Schächter F, Faure-Delanef L, Guénot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D. Genetic associations with human longevity at the APOE and ACE loci. *Nat. Genet.* 1994; 6:29-32.
6. Nebel A, Kleindorp R, Caliebe A, et al. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* 2011; 132:324-330.
7. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, Masaki KH, Willcox DC, Rodriguez B, Curb JD. FOXO3A genotype is strongly associated with human longevity. *Proc. Natl Acad. Sci. U S A* 2008; 105:13987-13992.
8. Flachsbarth F, Caliebe A, Kleindorp R, Blanché H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl Acad. Sci. U S A* 2009; 106:2700-2705.
9. Deelen J, Beekman M, Uh HW, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* 2014; 23:4420-4432.
10. Zeng Y, Nie C, Min J et al. Novel loci and pathways significantly associated with longevity. *Sci. Rep.* 2016; 6:21243.
11. Sebastiani P, Solovieff N, Dewan AT, et al. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; 7:e29848.
12. Fortney K, Dobriban E, Garagnani P, Pirazzini C, Monti D, Mari D, Atzmon G, Barzilai N, Franceschi C, Owen AB, Kim SK. Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. *PLoS Genet.* 2015; 11:e1005728.
13. Joshi PK, Pirastu N, Kentistou KA, et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* 2017; 8:910.

14. Broer L, Buchman AS, Deelen J et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. A Biol. Sci. Med. Sci.* 2015; 70:110-118.
15. Maher B. Personal genomes: the case of the missing heritability. *Nature* 2008; 456:18-21.
16. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461:747-753.
17. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. U S A* 2009; 106:9362-9367.
18. Speed D, Cai N, Ucleb C, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 2017; 49:986-992.
19. Bombá L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* 2017; 18:77.
20. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 2015; 24:R102-10.
21. Tanisawa K, Arai Y, Hirose N, et al. Exome-wide association study identifies clec3b missense variant p.s106g as being associated with extreme longevity in East Asian populations. *J. Gerontol. A Biol. Sci. Med. Sci.* 2017; 72:309-318.
22. Gierman HJ, Fortney K, Roach JC, Coles NS, Li H, Glusman G, Markov GJ, Smith JD, Hood L, Coles LS, Kim SK. Whole-genome sequencing of the world's oldest people. *PLoS One* 2014; 9:e112430.
23. Cash TP, Pita G, Domínguez O, et al. Exome sequencing of three cases of familial exceptional longevity. *Aging Cell* 2014; 13:1087-1090.
24. Tindale LC, Zeng A, Bretherick KL, Leach S, Thiessen N, Brooks-Wilson AR. Burden of common complex disease variants in the exomes of two healthy centenarian brothers. *Gerontology* 2015; 62:58-62.
25. Akhtarkhavari T, Joghataei MT, Fattahi Z, Akbari MR, Larti F, Najmabadi H, Kahrizi K. Genetic investigation of an Iranian supercentenarian by whole exome sequencing. *Arch. Iran. Med.* 2015; 18:688-697.
26. Yang F, Sun L, Zhu X, et al. Identification of new genetic variants of HLA-DQB1 associated with human longevity and lipid homeostasis-a cross-sectional study in a Chinese population. *Aging (Albany NY)* 2017; 9:2316-2333.
27. Nygaard HB, Erson-Omay EZ, Wu X, Kent BA, Bernales CQ, Evans DM, Farrer MJ, Vilariño-Güell C, Strittmatter SM. Whole exome sequencing of an exceptional longevity

cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* 2018 [Epub ahead of print].

28. Dato S, Soerensen M, De Rango F, Rose G, Christensen K, Christiansen L, Passarino G. The genetic component of human longevity: New insights from the analysis of pathway-based SNP-SNP interactions. *Aging Cell* 2018; e12755.
29. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: The CHARGE Consortium. *PLoS One* 2013; 8:e68095.
30. Newman AB, Walter S, Lunetta KL, et al. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: The cohorts for heart and aging research in genomic epidemiology consortium. *J. Gerontol. A Biol. Sci. Med. Sci.* 2010; 65:478-487.
31. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* 2016; 24:1202-1205.
32. Flachsbart F, Dose J, Gentschew L, et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat. Commun.* 2017; 8:2063.
33. Arnold M, Ellwanger DC, Hartsperger ML, Pfeufer A, Stümpflen V. Cis-acting polymorphisms affect complex traits through modifications of microRNA regulation pathways. *PLoS One* 2012; 7:e36694.
34. 1000 GPC, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491:56-65.
35. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015; 31:3555-3557.
36. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 2013; 45:1238-1243.
37. Walter S, Atzmon G, Demerath EW, et al. A genome-wide association study of aging. *Neurobiol Aging* 2011; 32:2109.e15-28.
38. Curk T, Rot G, Zupan B. SNPsyn: detection and exploration of SNP-SNP interactions. *Nucleic Acids Res.* 2011; 39:W444-9.
39. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* 2010; 26:2927-2928.
40. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009; 25:1091-1093.

41. Lin A, Wang RT, Ahn S, Park CC, Smith DJ. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res.* 2010; 20:1122-1132.
42. Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* 2016; 15:857-868.
43. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* 2007; 2:e841.
44. Szwergold BS, Bunker RD, Loomes KM. The physiological substrates of fructosamine-3-kinase-related-protein (FN3KRP) are intermediates of nonenzymatic reactions between biological amines and ketose sugars (fructation products). *Med. Hypotheses* 2011; 77:739-744.
45. Sell DR, Monnier VM. Molecular basis of arterial stiffening: role of glycation - a mini-review. *Gerontology* 2012; 58:227-237.
46. Sadowska-Bartosz I, Bartosz G. Effect of glycation inhibitors on aging and age-related diseases. *Mech. Ageing Dev.* 2016; 160:1-18.
47. Conner JR, Beisswenger PJ, Szwergold BS. The expression of the genes for fructosamine-3-kinase and fructosamine-3-kinase-related protein appears to be constitutive and unaffected by environmental signals. *Biochem. Biophys. Res. Commun.* 2004; 323:932-936.
48. Soranzo N. Genetic determinants of variability in glycated hemoglobin (HbA(1c)) in humans: review of recent progress and prospects for use in diabetes care. *Curr. Diab. Rep.* 2011; 11:562-569.
49. Hu Z, He C. CDKN2B gene rs1063192 polymorphism decreases the risk of glaucoma. *Oncotarget* 2017; 8:21167-21176.
50. Murabito JM, Yuan R, Lunetta KL. The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. *J. Gerontol. A Biol. Sci. Med. Sci.* 2012; 67:470-479.
51. Mugabo Y, Zhao S, Seifried A, et al. Identification of a mammalian glycerol-3-phosphate phosphatase: Role in metabolism and signaling in pancreatic β -cells and hepatocytes. *Proc. Natl Acad. Sci. U S A* 2016; 113:E430-9.
52. Possik E, Madiraju SRM, Prentki M. Glycerol-3-phosphate phosphatase/PGP: Role in intermediary metabolism and target for cardiometabolic diseases. *Biochimie* 2017; 143:18-28.
53. Prentki M, Madiraju SR. Glycerolipid metabolism and signaling in health and disease. *Endocr. Rev* 2008; 29:647-676.
54. Collard F, Baldin F, Gerin I, et al. A conserved phosphatase destroys toxic glycolytic side products in mammals and yeast. *Nat. Chem. Biol.* 2016; 12:601-607.
55. Kaeberlein M, Kennedy BK. Hot topics in aging research: protein translation and TOR

signaling, 2010. *Aging Cell* 2011; 10:185-190.

56. Zhang Q, Chen X, Zhang X, Zhan J, Chen J. Knockdown of TMEM14A expression by RNAi inhibits the proliferation and invasion of human ovarian cancer cells. *Biosci. Rep.* 2016; 36:e00298.
57. Conner JR, Smirnova II, Moseman AP, Poltorak A. IRAK1BP1 inhibits inflammation by promoting nuclear translocation of NF-kappaB p50. *Proc. Natl Acad. Sci. U S A* 2010; 107:11477-11482.
58. Staley LA, Ebbert MT, Bunker D, Bailey M, Alzheimer's DNI, Ridge PG, Goate AM, Kauwe JS. Variants in ACPP are associated with cerebrospinal fluid prostatic acid phosphatase levels. *BMC Genomics* 2016; 17 Suppl 3:439.
59. Liu B, Gu X, Huang T, Luan Y, Ding X. Identification of TMPRSS2-ERG mechanisms in prostate cancer invasiveness: involvement of MMP-9 and plexin B1. *Oncol. Rep.* 2017; 37:201-208.
60. Lin P, Mobasher ME, Hakakian Y, Kakarla V, Naseem AF, Ziai H, Alawi F. Differential requirements for H/ACA ribonucleoprotein components in cell proliferation and response to DNA damage. *Histochem. Cell Biol.* 2015; 144:543-558.
61. Bonnefond A, Clément N, Fawcett K, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* 2012; 44:297-301.
62. Deelen J, Uh HW, Monajemi R, et al. Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *Age (Dordr)* 2013; 35:235-249.
63. Kim J, Park DY, Bae H, et al. Impaired angiopoietin/Tie2 signaling compromises Schlemm's canal integrity and induces glaucoma. *J. Clin. Invest.* 2017; 127:3877-3896.
64. Kim J, Park JR, Choi J, et al. Tie2 activation promotes choriocapillary regeneration for alleviating neovascular age-related macular degeneration. *Sci. Adv.* 2019; 5:eaau6732.
65. Fu D, Collins K. Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol. Cell* 2007; 28:773-785.
66. Holle R, Happich M, Löwel H, Wichmann HE, MONICA/KORA SG. KORA - a research platform for population based health research. *Gesundheitswesen* 2005; 67 Suppl 1:S19-25.
67. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* 2006; 9:55-61.
68. Rasmussen SH, Andersen-Ranberg K, Thinggaard M, Jeune B, Skytthe A, Christiansen

L, Vaupel JW, McGue M, Christensen K. Cohort Profile: the 1895, 1905, 1910 and 1915 Danish Birth Cohort Studies - secular trends in the health and functioning of the very old. *Int. J. Epidemiol.* 2017; 46:1746-1746j.

69. Robine JM, Cheung SL, Saito Y, Jeune B, Parker MG, Herrmann FR. Centenarians today: new insights on selection from the 5-COOP study. *Curr. Gerontol. Geriatr. Res.* 2010; 2010:120354.

70. Skytthe A, Kyvik K, Holm NV, Vaupel JW, Christensen K. The Danish Twin Registry: 127 birth cohorts of twins. *Twin Res. Hum. Genet.* 2002; 5:352-357.

71. Gaist D, Bathum L, Skytthe A, Jensen TK, McGue M, Vaupel JW, Christensen K. Strength and anthropometric measures in identical and fraternal twins: no evidence of masculinization of females with male co-twins. *Epidemiology* 2000; 11:340-343.

72. Skytthe A, Christiansen L, Kyvik KO, et al. The Danish twin registry: linking surveys, national registers, and biological information. *Twin Res. Hum. Genet.* 2013; 16:104-111.

73. Blanché H, Cabanne L, Sahbatou M, Thomas G. A study of French centenarians: are ACE and APOE associated with longevity. *C R Acad Sci III* 2001; 324:129-135.

74. Hercberg S, Galan P, Preziosi P, Rousset AM, Arnaud J, Richard MJ, Malvy D, Paul-Dauphin A, Briançon S, Favier A. Background and rationale behind the SU.VI.MAX Study, a prevention trial using nutritional doses of a combination of antioxidant vitamins and minerals to reduce cardiovascular diseases and cancers. *SUPPLEMENTATION EN VITAMINES ET MINÉRAUX ANTIOXYDANTS STUDY. Int. J. Vitam. Nutr. Res.* 1998; 68:3-20.

75. Goldstein JI, Crenshaw A, Carey J, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 2012; 28:2543-2545.

76. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81:559-575.

77. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat. Protoc.* 2010; 5:1564-1573.

78. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: identifying density-based local outliers. *ACM sigmod record* 2000; 29: 93-104.

79. Yashin AI, Wu D, Arbeevev KG, Ukraintseva SV. Joint influence of small-effect genetic variants on human longevity. *Aging (Albany NY)* 2010; 2:612-620.

80. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am. J. Hum. Genet.* 2013; 92:841-853.

81. Basu S, Pan W. Comparison of statistical tests for disease association with rare variants.

Genet. Epidemiol. 2011; 35:606-619.

- 82.** Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 2011; 89:82-93.
- 83.** Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* 2012; 91:224-237.
- 84.** Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005; 21:3439-3440.
- 85.** Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 2001; 29:1165-1188.
- 86.** Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010; 38:W214-20.
- 87.** Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000; 25:25-29.
- 88.** Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 2002; 30:42-46.
- 89.** Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2018; 46:D649-D655.
- 90.** Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 2007; 8:R183.
- 91.** Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J. Clin. Epidemiol.* 1993; 46:423-429.
- 92.** Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988; 16:1215.
- 93.** StataCorp. Stata Statistical Software: Release 15. College Station, TX: StaCorp LLC, 2017.

Tables

Table 3.1. Association statistics for the 11 longevity-associated SNVs identified by the single-variant association approach in the whole German study population.

SNV	Gene	Chr	MAF ^a			Basic association test		Conditional analysis	
			LLI	C	MA	OR ^b	<i>P</i> ^d	OR ^b	<i>P</i> ^d
						[95% C.I.] ^c		[95% C.I.] ^c	
rs769449	<i>APOE</i>	19	0.056	0.109	A	0.48 [0.40 - 0.58]	7.77E-15	-	-
rs4420638	<i>APOC1</i>	19	0.109	0.169	G	0.60 [0.52 - 0.69]	3.55E-13	-	-
rs2075650	<i>TOMM40</i>	19	0.109	0.147	G	0.70 [0.61 - 0.80]	3.51E-07	-	-
rs1046896	<i>FN3KRP</i>	17	0.276	0.324	T	0.78 [0.71 - 0.86]	7.40E-07	0.77 [0.70 - 0.85]	2.32E-07
rs55882518	<i>NOTCH3</i>	19	0.013	0.005	T	2.69 [1.73 - 4.18]	1.07E-05	2.77 [1.78 - 4.30]	6.23E-06
rs13119846	<i>TMEM131L</i>	4	0.486	0.438	C	1.22 [1.11 - 1.33]	1.73E-05	1.21 [1.11 - 1.32]	2.45E-05
rs1063192	<i>CDKN2B</i>	9	0.482	0.439	G	1.21 [1.11 - 1.32]	2.99E-05	1.22 [1.12 - 1.33]	1.08E-05
rs184214819	<i>SPZ1</i>	5	0.009	0.003	A	3.01 [1.77 - 5.11]	4.72E-05	2.79 [1.64 - 4.74]	1.52E-04
rs200956599	<i>SKOR1</i>	15	0.014	0.006	T	2.34 [1.55 - 3.54]	5.24E-05	2.29 [1.52 - 3.47]	8.65E-05
rs63750412	<i>GRN</i>	17	0.008	0.002	T	3.57 [1.93 - 6.61]	5.18E-05	3.34 [1.80 - 6.18]	1.29E-04
rs1790706	<i>DSC2</i>	18	0.159	0.189	A	0.79 [0.70 - 0.89]	9.59E-05	0.78 [0.69 - 0.88]	3.35E-05

APOC1, apolipoprotein C1; APOE, apolipoprotein E; CDKN2B, cyclin dependent kinase inhibitor 2B; DSC2, desmocollin 2; FN3KRP, fructosamine 3 kinase related protein; GRN, granulin precursor; NOTCH3, notch 3; SKOR1, SKI family transcriptional corepressor 1; SPZ1, spermatogenic leucine zipper 1; TMEM131L, transmembrane 131 like; TOMM40, translocase of outer mitochondrial membrane 40

Long lived individuals (LLI); controls (C); chromosome (Chr)

^aMinor allele frequency, MAF; the definition of the minor allele (MA) is based on controls

^bOdds ratio for longevity, OR; based on the MA in controls

^c95% confidence interval, 95% C.I.; C.I. for the OR

^dAllelic *P*-values, calculated with chi - squared test with one degree of freedom

Table 3.2. Single-variant replication and meta-analysis statistics for candidate SNVs in the French and Danish populations.

SNV	Gene	Chr	MAF ^a			Basic association test		Conditional analysis	
			LLI	C	MA	OR ^b	<i>P</i> ^d	OR ^b	<i>P</i> ^d
						[95% C.I.] ^c		[95% C.I.] ^c	
rs769449	<i>APOE</i>	19	0.056	0.109	A	0.48 [0.40 - 0.58]	7.77E-15	-	-
rs4420638	<i>APOC1</i>	19	0.109	0.169	G	0.60 [0.52 - 0.69]	3.55E-13	-	-
rs2075650	<i>TOMM40</i>	19	0.109	0.147	G	0.70 [0.61 - 0.80]	3.51E-07	-	-
rs1046896	<i>FN3KRP</i>	17	0.276	0.324	T	0.78 [0.71 - 0.86]	7.40E-07	0.77 [0.70 - 0.85]	2.32E-07
rs55882518	<i>NOTCH3</i>	19	0.013	0.005	T	2.69 [1.73 - 4.18]	1.07E-05	2.77 [1.78 - 4.30]	6.23E-06
rs13119846	<i>TMEM131L</i>	4	0.486	0.438	C	1.22 [1.11 - 1.33]	1.73E-05	1.21 [1.11 - 1.32]	2.45E-05
rs1063192	<i>CDKN2B</i>	9	0.482	0.439	G	1.21 [1.11 - 1.32]	2.99E-05	1.22 [1.12 - 1.33]	1.08E-05
rs184214819	<i>SPZI</i>	5	0.009	0.003	A	3.01 [1.77 - 5.11]	4.72E-05	2.79 [1.64 - 4.74]	1.52E-04
rs200956599	<i>SKOR1</i>	15	0.014	0.006	T	2.34 [1.55 - 3.54]	5.24E-05	2.29 [1.52 - 3.47]	8.65E-05
rs63750412	<i>GRN</i>	17	0.008	0.002	T	3.57 [1.93 - 6.61]	5.18E-05	3.34 [1.80 - 6.18]	1.29E-04
rs1790706	<i>DSC2</i>	18	0.159	0.189	A	0.79 [0.70 - 0.89]	9.59E-05	0.78 [0.69 - 0.88]	3.35E-05

C, younger controls; CDKN2B, cyclin dependent kinase inhibitor 2B; FN3KRP, fructosamine 3 kinase related protein; LLI, long-lived individuals; TMEM131L, transmembrane 131 like. Listed are rs-numbers, annotated gene name, chromosome, allele frequencies in cases and controls, the minor allele, odds ratios with 95% confidence intervals and allelic *P*-values for each study population. The effective size of the German population was 1,248 LLI and 6,762 younger controls; for the French 1,270 LLI and 1,824 younger controls, and for the Danish 1,002 LLI and 737 younger controls.

^aMinor allele frequency, MAF; the definition of the minor allele (MA) is based on controls

^bOdds ratio for longevity, OR; based on the MA in controls

^c95% confidence interval, 95% C.I.; C.I. for the OR

^dAllelic *P*-values, calculated with chi-squared test with one degree of freedom; P_{corr}, corrected *P*-value using false discovery rate (corrected for three tests in the French and Danish study populations and for 62,488 tests in the German sample)

^eRandom-effects meta-analysis *P* value; ^fRandom-effects OR estimate; ^g*P*-value for Cochrane's Q statistic, Q; ^hI² heterogeneity index (0-100)

Table 3.3. Association statistics for the 16 longevity-associated genes identified by the gene-based association approach in the whole German study population.

Gene	Chr	P_skato	P_burden	SNVs				No. rs
				All	Tested	Rare	Common	
<i>APOE</i>	19	3,25E-15	3,25E-15	2	2	1	1	rs769449, rs769452
<i>APOC1</i>	19	2,59E-11	1,77E-02	3	3	0	3	rs439401, rs445925, rs4420638
<i>TOMM40</i>	19	1,35E-06	4,34E-02	3	3	1	2	rs157580, rs2075650, rs142412517
<i>PGP</i>	16	2,50E-06	8,90E-07	3	3	2	1	rs200526199, rs116977380, rs200615324
<i>OTOL1</i>	3	2,56E-04	7,11E-06	5	5	3	2	rs199791179, rs149127996, rs199985412, rs3921595, rs202021352
<i>FN3KRP</i>	17	9,19E-06	3,75E-02	5	5	3	2	rs138953335, rs61743692, rs144986629, rs142718764, rs1046896
<i>SETD9</i>	5	1,19E-05	6,95E-05	6	6	4	2	rs2257505, rs149334074, rs40497, rs141692637, rs150526244, rs146260337
<i>RPS6KB1</i>	17	1,56E-05	1,56E-05	2	2	1	1	rs201316437, rs1051424
<i>GRN</i>	17	1,59E-05	8,28E-03	4	4	4	0	rs63750723, rs63750043, rs63750541, rs63750412
<i>PSG7</i>	19	2,64E-05	4,19E-04	2	2	2	0	rs199532805, rs112354282
<i>SKOR1</i>	15	3,38E-05	9,20E-05	2	2	2	0	rs200956599, rs143419968
<i>HNF4G</i>	8	3,95E-05	2,17E-02	4	4	3	1	rs2943549, rs201625743, rs138897994, rs148532560
<i>ASB17</i>	1	4,42E-05	4,22E-05	3	3	1	2	rs149522654, rs11161887, rs3795251
<i>SPZ1</i>	5	8,09E-05	2,26E-04	5	5	2	3	rs1862136, rs139471643, rs184214819, rs200249535, rs35337118
<i>BFSP1</i>	20	1,25E-03	6,00E-05	6	6	5	1	rs145703098, rs140116733, rs6080719, rs147718368, rs143865632, rs142092768
<i>HMHA1</i>	19	1,54E-03	3,49E-05	8	8	5	3	rs1801284, rs2074442, rs36084354, rs142614852, rs150294461, rs139988914, rs61734935, rs139251906

APOE, apolipoprotein E; APOC1, apolipoprotein C1; ASB17, ankyrin repeat and SOCS box containing 17; BFSP1, beaded filament structural protein 1; FN3KRP, fructosamine 3 kinase related protein; GRN, granulin; HMHA1, Rho GTPase activating protein 45; HNF4G, hepatocyte nuclear factor 4 gamma; OTOL1, otolin 1; PGP, phosphoglycolate phosphatase; PSG7, pregnancy specific beta-1-glycoprotein 7; RPS6KB1, ribosomal protein S6 kinase B1; SETD9, SET domain containing 9; SKOR1, SKI family transcriptional corepressor 1; SPZ1, spermatogenic leucine

Figure legends

Figure 3.1. Workflow of the association analyses in the case-control study of longevity using the Illumina HumanExome BeadChip.

Figure 3.2. Manhattan plot summarizing the findings from the single-variant analysis. The inner plot represents the basic association results, the outer plot the association results after conditioning for the longevity-associated locus *TOMM40/APOE/APOC1*. The y-axis shows the $-\log(P\text{-value})$, while the red line depicts the P -value threshold ($1 \times 10E0-4$). The SNVs with $P < 1 \times 10E0-4$ are highlighted as red dots.

Figure 3.3. Functional annotation map generated by ClueGo. The nodes represent the GO terms. The size of the nodes indicate the enrichment significance of the respective term (big node = high significance).

Figure 3.1

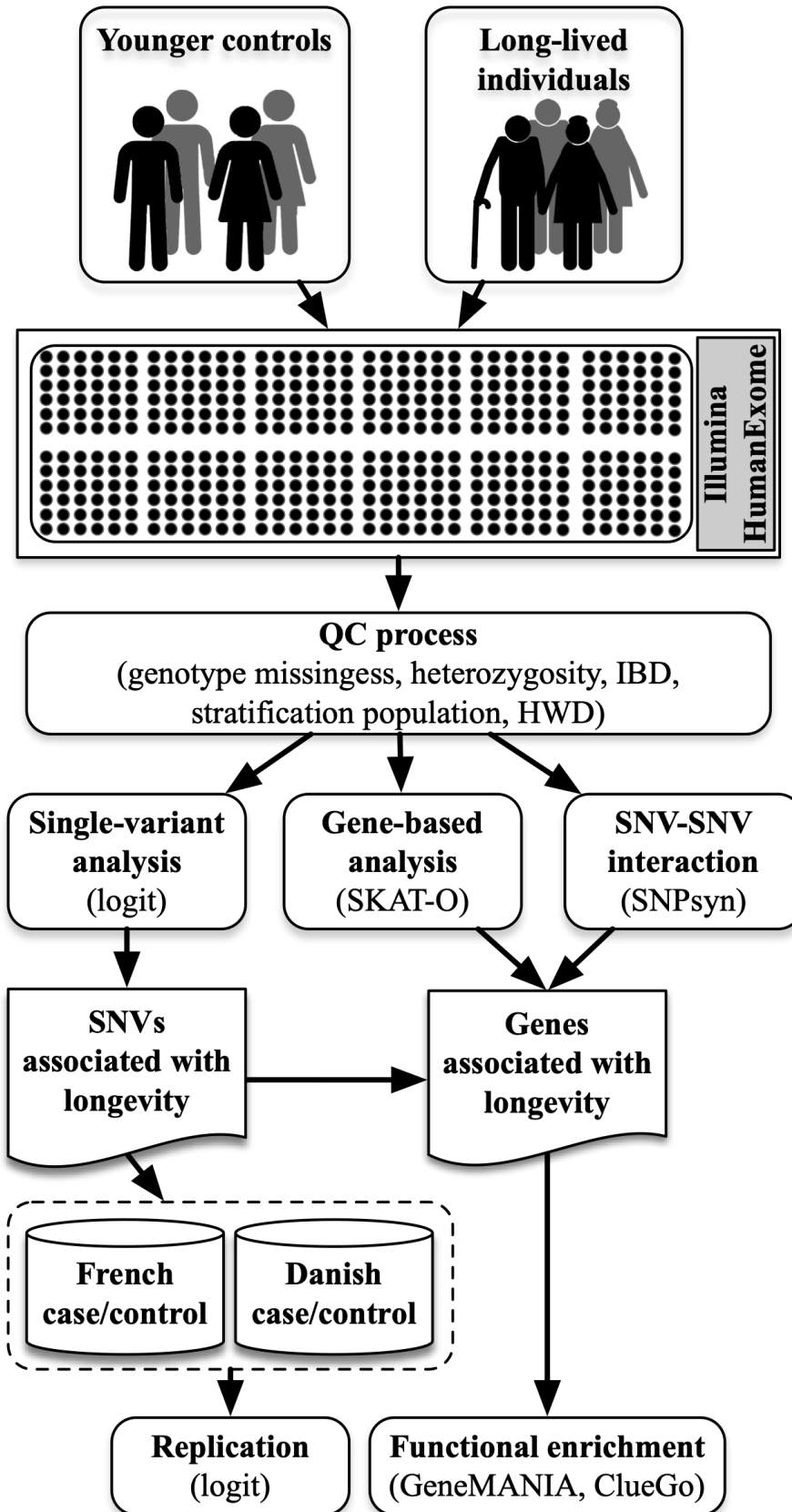


Figure 3.2

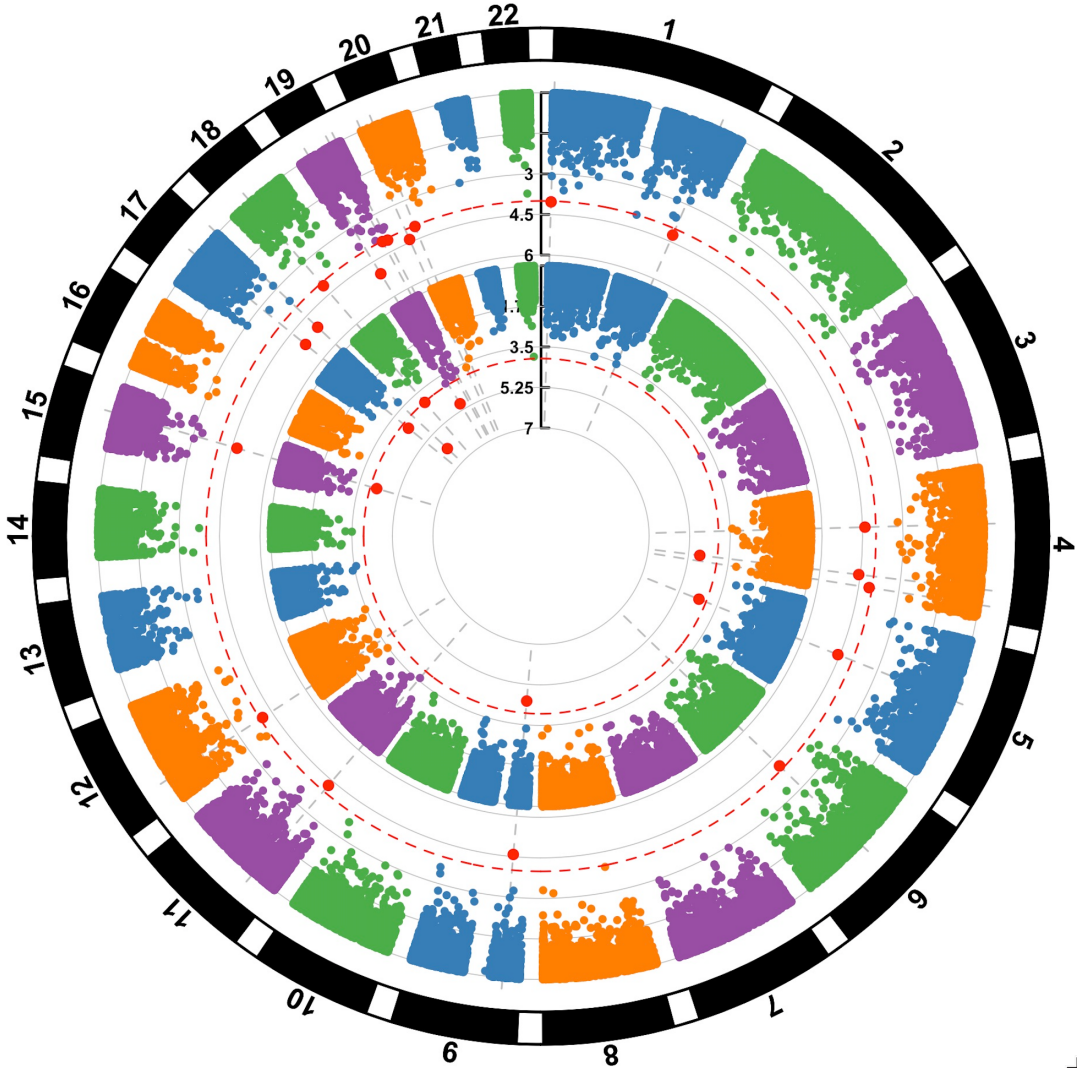
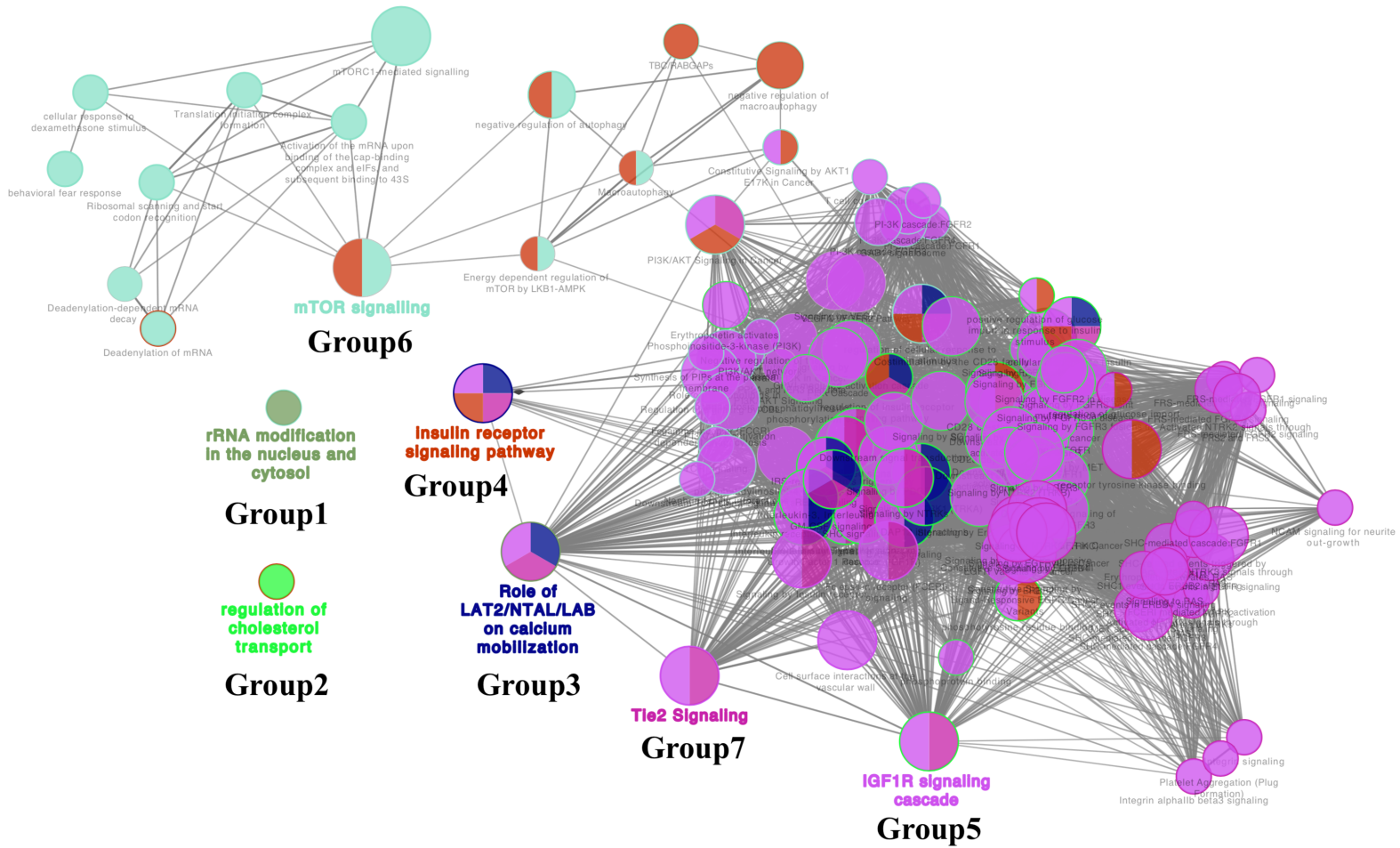


Figure 3.3.



4.

**Chapter III: Human gut microbiome in
healthy aging and longevity**

This manuscript was prepared as hybrid of a manuscript and dissertation chapter.

Gut microbiome in aging and longevity

Guillermo G. Torres¹, Janina Dose¹, Femke-Anouska Heinsen¹, Philipp Bergmann², Matthias Laudes², Wolfgang Lieb³, Stefan Schreiber¹, Andre Franke¹, Friederike Flachsbart¹, Almut Nebel¹

¹ Institute of Clinical Molecular Biology, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; g.torres@ikmb.uni-kiel.de, j.dose@ikmb.uni-kiel.de, f.flachsbart@mucosa.de, f.heinsen@ikmb.uni-kiel.de, a.franke@mucosa.de

² Department of Internal Medicine I, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; philipp.bergmann@uksh.de, matthias.laudes@uksh.de

³ Institute of Epidemiology and Biobank Popgen, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; w.lieb@epidemiology.uni-kiel.de

Corresponding author:

Almut Nebel, PhD

Institute of Clinical Molecular Biology

Kiel University

Rosalind-Franklin-Str. 12

24105 Kiel, Germany

phone: +49 (0)431 500 15155, fax: +49 (0)431 500 15168

email: a.nebel@mucosa.de

4.1. Introduction

Humans experience several changes during the aging process, some of them obvious, but others subtle, but still correlated with health maintenance or disease development, like changes in the gut microbiome (GM) composition. GM is a complex collection of microorganisms inhabiting our gastrointestinal tract. Immediately after birth, the gut is colonized, first by facultative anaerobes, followed by several successive remodelling steps which end in the colonization with strict anaerobes [1]. After the first settlement, the biodiversity of the GM increases and reaches a stable ecology structure at around an age of three years old [2]. In adulthood, the GM remains relatively stable and fecal bacteria in healthy adults are dominated by the phyla *Firmicutes* (families *Lachnospiraceae* and *Ruminococcaceae*), *Bacteroidetes* (families *Bacteroidaceae*, *Prevotellaceae*, and *Rikenellaceae*), and *Actinobacteria* (families *Bifidobacteriaceae* and *Coriobacteriaceae*) [2–5].

As soon as the GM has reached its biodiversity peak, the composition remains relatively stable for long periods of time [6]. However, GM appears to be very sensitive to changes in dietary patterns [5, 7], drugs or medical treatment [8, 9]. Furthermore, human genetics, ethnicity, and physiological shifts which occur during aging were shown to be relevant factors that shape the structure of the GM [10–13]. Interestingly, when humans reach extreme ages (older than 90 years of age), certain patterns of the taxonomical composition of the GM apparently change, which results in a longevity-specific GM profile [14–16].

Despite extensive efforts to understand GM structure and composition, it remains unclear to which degree GM changes are attributed to ethnicity, diet, genetics, gender, and aging, respectively. For Germans, no systematic large-scale study on GM changes during aging in individuals covering a broad age range, including long-lived individuals (LLI, >90 years of age), has been conducted so far. In this study, we performed a human stool metagenomic analysis based on 16S rRNA in 1,301 healthy individuals of German ancestry (age range 19 to 104 years) to i) identify age-related changes in the microbiome composition, ii) determine to which extent host genetics (associated with gut microbial composition and metabolic conditions) as well as environmental factors affect the gut microbial composition during aging, and iii) investigate whether there are any specific microbial patterns that characterize German LLI.

4.2. Results

To investigate the characteristics of the gut microbiome in Germans during aging, we performed a large-scale 16S rRNA-based microbiome study in 1,301 healthy German individuals which covered the broad age range of 19 to 104 years of age. The individuals derived from three different sample collections, abbreviated as BSP/SPC, FOC, and AGE, respectively, with the latter cohort comprising the oldest (90+ years) participants. The characteristics of the different study populations are shown in **Figure 4.1** and **Supplementary table 1**.

4.2.1. Physiological parameters remain relatively stable during aging in healthy individuals

Because the symbiotic microorganisms in the gut are susceptible to both external and host physiological factors, we first explored how selected physiological parameters of the study individuals change during aging and if anthropometric and dietary parameters contribute to explaining the morphophysiological changes. We first analyzed medical records of the study participants. Parameters with values for at least 70% of the individuals ($n(\text{parameters})=47$; **Supplementary table 3**) were selected for the analysis. Principal component analysis (PCA) showed that both gender (**Figure 4.2 A**) and age groups (groups covering 30 years each) led to anthropometric profiles (**Figure 4.2 B, C**). Stratification by age was mainly explained by anthropometric parameters, while the inter-individual variability was largely due to dietary habits (**Figure 4.2 B, C**). To identify which of the anthropometric parameters correlated with age, we used a linear regression model independently for each gender. The model revealed a subtle difference between males and females. Along the age spectrum, both males and females significantly ($P < 0.05$) increased their waist-to-hip ratio (WHR), fiber consumption and the total energy intake. On the contrary, we observed a reduction of plant protein and long chain fatty acid intake as well as water consumption (**Figure 4.2 C - E**). In addition, in females, blood glucose concentrations tended to increase along the age spectrum.

4.2.2. Four enterotype-like clusters stratify the GM and the global microbial diversity tends to increase with age, but appears to be lower in LLI

To investigate potential alterations in the gut bacterial ecology caused by aging, we analyzed 16S rRNA sequencing libraries of all 1,301 individuals. The assembly and binning of 50,924,857 sequencing reads led to the identification of 1,255 OTUs (**Supplementary table 3**). Subsequent taxonomical annotation revealed that more than $87.9\% \pm 15.8\%$ of all bacteria

belonged to the three phyla *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*, respectively, with the latter one representing the least abundant (in our samples $9.9\% \pm 11.6\%$ of the total bacterial load; **Figure 4.3**). The Shannon index as a measure of alpha-diversity (i.e. the taxonomical composition) was not significantly different between the study cohorts (5.1 ± 0.37 for BSP/SPC, 4.92 ± 0.39 for FOC, and 4.7 ± 0.77 for AGE).

To explore the gut microbiome stratification, we used a *de novo* enterotype identification approach. We identified four clusters (**Figure 4.4 A, B**); three were consistent with the already known *Prevotella* (ET_P), *Bacteroides* (ET_B) and *Firmicutes* (ET_F) enterotype, and the newly discovered fourth enterotype-like structure was driven by *Escherichia shigella* (ET_Esh) (**Figure 4.4 A**). When we compared the results of our enterotype prediction approach with the enterotypes generated with the enterotype classifier (<http://enterotypes.org>); 62.53% of our predicted enterotypes were concordant with the classifier output, 10% were swap-classified (ET_B as ET_F and *vice versa*) and 13.8 % represented a miss-classified enterotype (ET_Esh; ET_Esh was supposed to be classified as unknown (XO) as this enterotype is not recognized by the enterotype classifier; but instead was classified as either ET_B or ET_F) (**Figure 4.4 C**). We also observed that the proportion of the enterotypes differed between the three cohorts, but was not affected by age (**Figure 4.4 D**). The microbial diversity within the enterotypes was not statistically different either between males and females or between the cohorts (**Figure 4.4 E**); however, for all enterotypes, the GM tended to increase with aging. Strikingly, the diversity tended to decline in nonagenarians/centenarians, independently of the enterotype, and the lower diversity reached statistical significance for ET_Esh (**Figure 4.4 E**).

4.2.3. Enterotypes explain 19.8% of the bacterial structure variability

We further investigated changes in the beta-diversity (i.e. the comparison of the taxonomical composition between populations) in relation to the microbiome stratification, individual genetics, morphometric (**Supplementary table 4**) and dietary parameters by redundancy analysis (RDA; with $P < 0.001$). The results showed that enterotype, gender, age, WHR, BMI, intestinal inflammatory status (estimated from fecal calprotectin measurements), triglycerids in blood, dietary parameters (such as protein consumption, fibre, plant protein, alcohol intake), and the genetic variants rs958798-T (SNV annotated to *KCNC4*) and rs6666120-G (SNV annotated to *ACTL8*) explained 20.65% of the variability in bacterial abundance (**Figure 4.5**). Impressively, only the enterotypes accounted already for 19.8% of the variability (**Figure 4.5 B**). When we considered the bacterial community structure of each enterotype separately,

anthropometric parameters, genetics, diet and blood values explained 4.9 % of the total variability (**Table 4.1**). Strikingly, age alone explained not more than 1.33% of the total GM variability (**Table 4.1**).

Next, we tested if there were any differences in the bacterial diversity between the enterotypes. Using a zero-inflated Log-normal mixture modeling [17], we observed 32 differentially abundant genera between the enterotypes ($P < 0.05$) (**Supplementary table 5**). We investigated the functional implications of the 32 genera with Tax4Fun [18] and, subsequently, determined the distinctive metabolic components using MetaPath thresholds [19]. This approach yielded 624 significant reactions ($P < 0.01$) and 69 sub-pathways ($P < 0.05$), which were differentially abundant between the four enterotypes (**Supplementary figure 1, Supplementary table 5, 6**). The enriched pathways comprised: for ET_P, those related to folate biosynthesis, vitamin B6/B3 metabolism, beta-Alanine metabolism, lipopolysaccharides biosynthesis; in ET_B, butanoate, synthesis and degradation of ketone bodies, vitamin B3 metabolism; in ET_F, pathways related to biosynthesis of ansamycins, thiamine metabolism, porphyrin and chlorophyll metabolism, and in ET_Sh, pathways related to tyrosine metabolism.

4.2.4. Microbial changes associated with aging

One major goal of our study was to identify microbial changes that occur during aging. We used the software MaAslin2 [20] to model the effects of age on microbial occurrences and included the anthropometric, genetic, and dietary parameters as covariates. The abundance of 43 genera correlated significantly with age (**Figure 4.6, Supplementary table 7**). Positive correlations were seen for bacterial communities such as *Desulfovibrio*, *Oxalobacter*, some butyrate-producing bacteria (*Roseburia*, *Butyrivibrio*, and *Anaerotruncus*), and families that may influence host metabolism, in particular *Coriobacteriaceae*, *Erysipelotrichaceae*, and *Victivallaceae*. Interestingly, the families *Enterobacteraceae* and *Peptostreptococcaceae* and the class *Gammaproteobacteria*, which were negatively correlated with age in our analyses, were previously shown to be associated with health conditions like colorectal cancer, diarrhea, and Crohn's disease [21] [22]. Noteworthy, with increasing age, we observed a significant increase in the occurrence of unknown bacteria ($P = 8.97E-03$, $P_{adj.} = 0.097$, **Figure 4.6**).

We paid special attention to differences in the microbial abundance in healthy LLI (>90 years of age) compared with younger study participants. MaAslin2 identified 45 genera with differential occurrence between LLI and either elderly (75 - 80 years old) or younger adults

(<75 years old) (**Figure 4.7, Supplementary table 8**). In the LLI, we observed a higher abundance of so-called "pathobionts" such as *Clostridiales*, *Enterococcus*, *Enterobacteraceae*, *Klebsiella*, *Lactobacillus*, *Prevotella*, *Pseudomonas* and *Streptococcus*. In contrast, beneficial butyrate producers such as *Eubacterium*, *Faecalibacterium*, *Roseburia* and *Coprococcus* showed a lower abundance in the LLI. Furthermore, the LLI exhibited elevated levels of *Eggerthella*, *Bifidobacterium*, *Anaerotruncus*, *Eggerthella*, *fusobacterium*, and *Pseudomonas* (**Figure 4.7**).

4.3. Discussion

We investigated the role of aging on the microbial composition in healthy individuals from 19 to 104 years old. We identified moderate but significant changes that occur during aging, and a specific pattern in nonagenarians and centenarians. Interestingly, the BSP/SPC and FOC cohorts exhibited different enterotype proportions (**Figure 4.4 D**). The correction for enterotypes reduced the confounding bias and increased the comparability of findings from the different cohorts.

Fecal bacteria in healthy adults are dominated by the phyla *Firmicutes* (families *Lachnospiraceae*, *Ruminococcaceae*, and *Veillonellaceae*), *Bacteroidetes* (families *Bacteroidaceae*, *Prevotellaceae*, *Porphyromonadaceae*, and *Rikenellaceae*), and *Actinobacteria* (families *Bifidobacteriaceae* and *Coriobacteriaceae*) [2–5]. As expected, our results were in line with this pattern, apart from the third dominant phylum which, in our case, was *Proteobacteria* (family *Enterobacteraceae*) (**Figure 4.4**). For the latter phylum, we observed a continuous increase in the abundance of the genera *Desulfovibrio*, *Oxalobacter*, *Klebsiella*, and *Hafnia* with aging. This was independent of both the enterotype class and anthropometric parameters of the individuals. Both *Klebsiella* and *Hafnia* were shown to be able to secrete small molecules known as siderophores, which enable bacteria to compete with the host for iron [23, 24]. This competition can disturb the iron pool of the mucosal cells, which may in turn provoke the activation of innate and adaptive immune responses including proliferation, differentiation and secretion of inflammatory cells and mediators, respectively [25, 26]. In mice, members of the phylum *Proteobacteria*, such as *Desulfovibrio*, have been previously suggested to contribute to the chronic low-grade inflammatory state observed with advanced age [27]. Therefore, it is tempting to speculate that *Desulfovibrio*, *Klebsiella* and *Hafnia* are active players of the human GM that contribute to the pro-inflammatory status of elderly people.

Both the gastrointestinal tract as well as dietary habits of humans are known to undergo substantial changes with aging, which lead to a gradual shifting of the GM over time [28–30]. In our study, dietary patterns substantially explained the inter-individual variation observed along the age spectrum and between males and females (**Figure 4.2**): Examining the taxa profile more closely, we identified *Parabacteroides* (family *Porphyromonadaceae*), *Alistipes* (family *Rikenellaceae*), and *Dialister* (family *Veillonellaceae*) as the most abundant genera (**Figure 4.4**). These taxa have been previously associated with normal GM in adulthood and they may be preventive for obesity [31–33]. *Dialister* has been related to diets rich in fiber and whole grains, and to improvements in physiological measures of the host like reduced plasma interleukin-6 (IL-6) and reduced postprandial glucose levels [34]. Therefore, *Parabacteroides*, *Alistipes* and *Dialister* may exert health-promoting effects during aging, especially conferring resistance to low-grade inflammation (the so-called "inflammaging") [35]. Therapeutic measures to increase these beneficial bacteria in the GM could facilitate healthy aging.

Strikingly, the enterotype-like clusters explained most of the inter-individual variation (19.8 %). It has been suggested previously that enterotypes differ in terms of functional and ecological properties. For example, ET_P or *Prevotella* was shown to be enriched in individuals with non-Western and/or fibre-rich diets [36, 37]. *Prevotella* possess particular hydrolases specialized in plant-fiber degradation and a reduced lipolytic and proteolytic fermentation potential [38, 39]. In contrast, ET_B was reported to be associated with diets rich in animal protein and saturated fats and, additionally, *Bacteroidetes* were shown to exhibit an increased carbohydrate metabolism [38–41]. Accordingly, our functional predictions on the basis of the identified enterotypes supported the lipolytic, proteolytic and saccharolytic metabolism of ET_B (**Supplementary figure 1, Supplementary table 5**). For ET_P, we predicted an increment in fatty acid, folate, cofactor and vitamin biosynthesis, as reported previously [42]. Apart from ET_B and ET_P, we observed an *Enterobacteraceae* enterotype, (ET_Esch) with *Escherichia-Shigella* as its main driver. Despite members of the family *Enterobacteraceae*, including *E. coli*, *Klebsiella spp* and *Proteus spp*, were recognized to preferentially settle in inflamed environments [43–45], the *Enterobacteraceae* enterotype has not been associated as a cause of a pro-inflammatory status so far [46, 47]. Moreover, our results did not show any correlation of ET_Esch either with any of the available pro-inflammatory markers or with age or gender. Furthermore, despite our functional predictions of ET_Esch showed enrichment in pro-inflammatory pathways (**Supplementary figure 1, Supplementary table 5**), we did not

observe any evidence for a correlation of ET_Esch with a pro-inflammatory status or age-related diseases or dysbiosis.

From a taxonomical point of view, ET_Esch was enriched in *Acinetobacter*, *Hungatella*, *Coprobacillus*, *Enterococcus*, *Klebsiella*, *Enterobacteriaceae* and *Escherichia-Shigella*. Interestingly, we found *Acinetobacter* to be positively correlated with BMI and *Hungatella* with female gender. Furthermore, abundance of *Escherichia-Shigella* was inversely correlated with water intake and the G-allele of rs4973961 (in the gene Unc-51 Like Kinase 4 (*ULK4*)) (**Supplementary figure 2, Supplementary table 7**). Rs4973961-G has already been associated with microbial taxonomies and functional units previously [12]; specifically, a negative correlation with both thiamine biosynthesis and presence of *Bifidobacteriales* (some of its members are opportunistic pathogens [48]) has been reported [12]. Here, we observed rs4973961-G to be negatively correlated with opportunistic pathogens such as *Haemophilus*, *Veillonella* and *Escherichia-Shigella*, but to be positively correlated with *Christensenella*. The latter taxon has been linked to health-promoting effects and was shown to be strongly influenced by the genetic makeup of the host [49]. Therefore, our results might indicate a pro-inflammatory microbial signature in the gut of the study participants carrying this ET, although, in our healthy individuals, this was not yet reflected in the values of the collected inflammatory markers like calprotectin. However, in a diseased state, ET_Esch could influence disease progression [43–45]. Furthermore, with the association between ET_Esch and rs4973961, we showed that the likelihood of carrying ET_Esch could be determined by host genetic factors.

Aging entails significant changes in body constitution. These changes include quantitative and qualitative progressive loss of skeletal muscle mass together with a body fat redistribution and an increase in low-density lipoprotein levels (LDL), triglycerides and pro-inflammatory cytokines (i.e. IL-6), independent of any disease development [50–52]. In the individuals included in this study, which were all healthy Germans, BMI, WHR, GGT, HDL, triglycerides, calprotectin, creatinine and blood glucose levels correlated with aging (**Figure 4.2**). Additionally, BMI, WHR, triglycerides, calprotectin, and glucose in blood explained changes in some specific taxa during aging (**Supplementary figure 2, Supplementary table 7**). However, the GM profile of the healthy agers in our study showed a tendency towards a higher diversity and several taxa with health-promoting effects such as butyrate-producers and regulators of metabolic processes (**Figure 4.6**). With advancing age, we observed a continuous increase in butyrate-producing bacteria (*Roseburia*, *Butyrivibrio*, and *Anaerotruncus*); furthermore, in

Oxalobacter, which potentially contribute to protection from calcium oxalate kidney stones [53, 54], the family *Coriobacteriaceae*, which plays an important role in fatty liver disease and obesity [55], and *Erysipelotrichaceae*, which were shown to be positively associated with changes in liver fat content in humans and host lipid metabolism [56–58]. Moreover, we observed a reduction in *Peptostreptococcus* and *Gamaproteobacteria*, which have been associated with colorectal cancer and behavioral disorders, respectively [59]. Interestingly, we observed that during aging, the abundance of *Bacteroidales S24-7* tended to increase, in contrast to a decrease in *Coprobacter*. Such a pattern has been previously observed in Chinese individuals on a high-fat diet [60]; therefore, these bacteria might contribute to the increase in BMI and triglycerides observed in our individuals during aging.

Physiological shifts in the gut environment (e.g. higher levels of pro-inflammatory cytokines, thinning of the mucosal layer in the gut) during aging allow the colonization of opportunistic pathogens [28–30]. In the present study, we identified *Cloacibacillus*, which has been previously reported to be associated with intestinal infections [61]. Furthermore, we found *Holdemanella*, previously shown to be associated with an unhealthy serum lipid profile (low HDL cholesterol and/or high triglycerides or high LDL cholesterol) in obese women [62] as well as *Fusobacterium*, *Blautia* and *Enterobacteraceae*, which are all typical inhabitants of pro-inflammatory environments [26, 63].

It has been previously stated that in LLI, the GM exhibits an adaptive remodeling; in an extent that would even justify to consider centenarians as a separate population [14, 15, 64]. Our results are in line with this hypothesis and substantiate in large part the microbial profile seen in nonagenarians/centenarians from Japan [16], China [65], and Italy [66, 67]. However, we observed a few differences from earlier reports, i.e. German LLI exhibited a higher abundance of *Ordobacter* and *Parabacteroides*, which resembles Chinese centenarians [65]. Furthermore, German LLI also displayed depletion in *Lachnospiraceae* and *Akkermansia*, which has been previously observed in Japanese nonagenarians/centenarians [16]. Some other differences that we observed are likely explained by the differences in the dietary patterns between populations of different cultures. For example, we noticed an elevated abundance of *Halomonas*, *Lactobacillus* and *Shewanella*, which altogether have been associated with ingestion of fermented foods [68]. We also detected an overall reduced bacterial diversity in the LLI compared with younger study participants (**Figure 4.7**) and, specifically, a reduced abundance of *Coriobacteriaceae*, *Eubacterium* and *Oxalobacter*, while *Lachnoclostridium* abundance was

increased. This pattern partially resembles the signature of antibiotic treatment [8]. Although we excluded individuals who had reported antibiotic treatment, we cannot rule out misreporting. Moreover, in the LLI, we found an elevated abundance of *Eggerthella*, which has been linked to cardiac medication [9].

4.4. Material and methods

Study populations

The study included participants from three different sample collections from Schleswig-Holstein (Germany), abbreviated as BSP/SPC, FOC, and AGE. The BSP/SPC cohort comprised 601 individuals, 1,115 individuals derived from the FOC cohort, and 22 individuals were recruited for the AGE cohort (**Supplementary table 1**). The study cohorts were recruited independently from each other with the help of the biobank Popgen [69]. Popgen also provided the corresponding phenotype and genotype information [69]. Fecal samples were collected by the participants at home using standard fecal tubes. Samples were kept at room temperature and either shipped or brought to the collection center by the participants themselves (within 24 hours after collection). Samples were stored at -80°C until further analysis. All study participants provided a signed written informed consent prior to enrollment in the study. Approval for the study was obtained from the Ethics Committee of the Medical Faculty of Kiel University.

Genotyping data

Samples of the BSP/SPC and FOC cohorts were genotyped on different genotyping arrays. BSP/SPC samples were genotyped using the following chips: Affymetrix 6.0, Affymetrix Axiom, Illumina 550k, custom Illumina ImmunoChip and Illumina MetaboChip arrays, with sample sizes ranging from 678 to 1,218 before quality control and variant coverage of 196,524 to 934,968 variants. The FOC samples were genotyped on the custom Illumina ImmunoChip (Illumina Inc., San Diego, USA) and the Omni Express Exome (Illumina Inc., San Diego, USA), with 1,024 and 1,713 samples before quality control, respectively, and a variant coverage of 195,732 to 964,193 variants. For each cohort, genotype data from every array underwent quality control separately and, subsequently, data were merged and imputed. For our study, we included 325 single-nucleotide variants which had been previously reported to be associated with the gut microbiome [11, 12, 49, 70, 71] and 11 SNVs associated with longevity (annotated to the genes *FOXO3* [72] and *APOE* [73], respectively) (**Supplementary table 2**).

DNA extraction and 16S rRNA sequencing

Bacterial genomic DNA was extracted using the QIAamp DNA Stool Mini kit from Qiagen on a QIAcube system (Promega Corporation, Madison, USA). The V1–V2 region of the 16S rRNA gene was sequenced on the MiSeq platform using the 27F–338R primer pair and dual MID indexing (8 nt each on the forward and reverse primers) as described previously [74]. Sequencing was performed with the MiSeq Reagent Kit v2. After sequencing, MiSeq fastq files were obtained from base calls for paired-end reads and both indices (I1 and I2) using the Bcl2fastq module in CASAVA 1.8.2 (Illumina Inc., San Diego, USA). Stringent demultiplexing was carried out by allowing no mismatches in none of the index sequences.

Processing of sequencing dataset

Paired-end reads were processed with Mothur v 1.36.5 [75]. To minimize effects of sequencing errors, sequences that contained more than one undetermined nucleotide, had more than six homopolymers, or showed a Q-score average < 25 in a window of 50 pb were trimmed. Trimmed sequences with a length < 250 pb were eliminated. Spurious sequences were reduced in two steps. First, a pre-clustering step for noise reduction was performed as described previously [76], and second, a chimeras' elimination using Vsearch algorithm [77]. High-quality trimmed reads were used to calculate an uncorrected distance matrix. This matrix was used to cluster the sequences into OTUs of 97% similarity with the optclust algorithm [78]. Taxonomical annotation of the OTUs was done using the consensus taxonomy of the sequences, which were identified with the Bayesian-RDP classifier [79] and the SILVA reference database release 123 [80]. Prior to statistical analysis, the OTU-count table was pre-processed to reduce spurious and chimeric OTUs which met the core measurable microbiome threshold (CMMt) [81]. CMMt was calculated according to pairwise combinations of technical replicates of 49 different samples. For each sample, CMMt was defined as the minimum number of reads for which correlation between technical replicates was higher than 80%. Therefore, only OTUs were retained, whose count values were above the average CMMt and which were present in at least 2% (size of the smallest cohort (LLI)) of the libraries.

Statistical analysis

Morphophysiology across samples was explored by multiple factor analysis. Missing values were imputed using a multivariate approach [82]. OTU counts were normalized by the total-sum scaling method [17]. The R packages vegan [83] and metagenomeSeq [17] were used to calculate the alpha- and beta-diversity on the basis of the normalized counts. Principal

Coordinate Analysis (PCoA) based on Bray-Curtis dissimilarity [83] was used to assess differences in the community composition across samples. Enterotype-like microbiomes were identified using a *de novo* clustering and reference classification approach. Clusters were generated using partition around medoids over a Jensen-Shannon dissimilarity matrix. The number of optimal clusters was calculated with the silhouette and gap statistic method with a 500 times Monte Carlo re-sampling [84]. Additionally, the enterotypes were imputed using the enterotype classifier (<http://enterotypes.org>), which is limited to the identification of three enterotypes [7]. Differentially abundant genera between enterotypes were identified employing the Log-Normal zero-inflated mixture model [17]. Differentially abundant taxa during aging were analyzed using the multivariate statistical framework MaAsLin2 [20] which performs boosted, additive, general linear models between metadata and microbial abundance data.

Functional prediction

Prediction of functional profiles was performed on the basis of SILVA-labeled OTU abundances using Tax4Fun [18]. Differentially abundant pathways and reactions were identified by MetaPath [19]. Enriched pathways were determined based on two *P*-values, *p*abund and *p*struct [19], and enriched reactions were selected based on *p*abund only.

References

1. Fanaro S, Chierici R, Guerrini P, Vigi V. Intestinal microflora in early infancy: composition and development. *Acta Paediatr. Suppl.* 2003; 91:48-55.
2. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature* 2012; 486:222-227.
3. Sghir A, Gramet G, Suau A, Rochet V, Pochart P, Dore J. Quantification of bacterial groups within human fecal flora by oligonucleotide probe hybridization. *Appl. Environ. Microbiol.* 2000; 66:2263-2266.
4. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. *Science* 2005; 308:1635-1638.
5. Maukonen J, Saarela M. Human gut microbiota: does diet matter. *Proc. Nutr. Soc.* 2015; 74:23-36.
6. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. The long-term stability of the human gut microbiota. *Science* 2013; 341:1237439.
7. Costea PI, Hildebrand F, Arumugam M, et al. Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* 2018; 3:8-16.
8. Raymond F, Ouameur AA, Déraspe M, et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J.* 2016; 10:707-720.
9. Koppel N, Bisanz JE, Pandelia ME, Turnbaugh PJ, Balskus EP. Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *Elife* 2018; 7.
10. Brooks AW, Priya S, Blekhman R, Bordenstein SR. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol.* 2018; 16:e2006842.
11. Wang J, Thingholm LB, Skiecevičienė J, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* 2016; 48:1396-1406.
12. Bonder MJ, Kurilshikov A, Tigchelaar EF, et al. The effect of host genetics on the gut microbiome. *Nat. Genet.* 2016; 48:1407-1412.
13. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 2018; 555:210-215.
14. Kong F, Hua Y, Zeng B, Ning R, Li Y, Zhao J. Gut microbiota signatures of longevity. *Curr. Biol.* 2016; 26:R832-R833.

15. Biagi E, Franceschi C, Rampelli S, Severgnini M, Ostan R, Turrone S, Consolandi C, Quercia S, Scurti M, Monti D, Capri M, Brigidi P, Candela M. Gut microbiota and extreme longevity. *Curr. Biol.* 2016; 26:1480-1485.
16. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao JZ, Abe F, Osawa R. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol.* 2016; 16:90.
17. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* 2013; 10:1200-1202.
18. ABhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 2015; 31:2882-2884.
19. Liu B, Pop M. MetaPath: identifying differentially abundant metabolic pathways in metagenomic datasets. *BMC Proc.* 2011; 5 Suppl 2:S9.
20. Himel M, L. TT, J. ML, Gholamali R, George W, N. PJ, Siyuan M, Boyu R, Emma S, Ayshwarya S, A. FE, Corrada BH, Curtis H. Multivariable association in population-scale meta'omic surveys. 2019 [In submission].
21. Rizzatti G, Lopetuso LR, Gibiino G, Binda C, Gasbarrini A. Proteobacteria: a common factor in human diseases. *Biomed. Res. Int.* 2017; 2017:9351507.
22. Sutin AR, Terracciano A, Deiana B, Naitza S, Ferrucci L, Uda M, Schlessinger D, Costa PT. High neuroticism and low conscientiousness are associated with interleukin-6. *Psychol. Med.* 2010; 40:1485-1493.
23. Ellermann M, Arthur JC. Siderophore-mediated iron acquisition and modulation of host-bacterial interactions. *Free Radic. Biol. Med.* 2017; 105:68-78.
24. Podschun R, Fischer A, Ullmann U. Characterisation of *Hafnia alvei* isolates from human clinical extra-intestinal specimens: haemagglutinins, serum resistance and siderophore synthesis. *J. Med. Microbiol.* 2001; 50:208-214.
25. Wang L, Cherayil BJ. Ironing out the wrinkles in host defense: interactions between iron homeostasis and innate immunity. *J. Innate Immun.* 2009; 1:455-464.
26. Atarashi K, Suda W, Luo C, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science* 2017; 358:359-365.
27. Fransen F, van Beek AA, Borghuis T, Aidy SE, Hugenholtz F, van der Gaast-de Jongh C, Savelkoul HFJ, De Jonge MI, Boekschoten MV, Smidt H, Faas MM, de Vos P. Aged gut microbiota contributes to systemical inflammaging after transfer to germ-free mice. *Front. Immunol.* 2017; 8:1385.
28. O'Toole PW, Claesson MJ. Gut microbiota: changes throughout the lifespan from

infancy to elderly. *International Dairy Journal* 2010; 20:281-291.

29. Claesson MJ, Cusack S, O'Sullivan O, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl Acad. Sci. U S A* 2011; 108 Suppl 1:4586-4591.
30. Claesson MJ, Jeffery IB, Conde S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* 2012; 488:178-184.
31. Del Chierico F, Abbatini F, Russo A, et al. Gut microbiota markers in obese adolescent and adult patients: age-dependent differential patterns. *Front. Microbiol.* 2018; 9:1210.
32. Wu TR, Lin CS, Chang CJ, Lin TL, Martel J, Ko YF, Ojcius DM, Lu CC, Young JD, Lai HC. Gut commensal *Parabacteroides goldsteinii* plays a predominant role in the anti-obesity effects of polysaccharides isolated from *Hirsutella sinensis*. *Gut* 2019; 68:248-262.
33. Louis S, Tappu RM, Damms-Machado A, Huson DH, Bischoff SC. Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. *PLoS One* 2016; 11:e0149564.
34. Martínez I, Lattimer JM, Hubach KL, Case JA, Yang J, Weber CG, Louk JA, Rose DJ, Kyureghian G, Peterson DA, Haub MD, Walter J. Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J.* 2013; 7:269-280.
35. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* 2018; 14:576-590.
36. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. U S A* 2010; 107:14691-14696.
37. Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, North ACFRB, Coutinho PM, Henrissat B, Nelson KE. Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microb. Ecol.* 2010; 60:721-729.
38. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014; 505:559-563.
39. Vieira-Silva S, Falony G, Darzi Y, Lima-Mendez G, Garcia Yunta R, Okuda S, Vandeputte D, Valles-Colomer M, Hildebrand F, Chaffron S, Raes J. Species-function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* 2016; 1:16088.
40. Wu GD, Chen J, Hoffmann C, et al. Linking long-term dietary patterns with gut

microbial enterotypes. *Science* 2011; 334:105-108.

41. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 2009; 37:D233-8.
42. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature* 2011; 473:174-180.
43. Zeng MY, Inohara N, Nuñez G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* 2017; 10:18-26.
44. Garrett WS, Gallini CA, Yatsunenkov T, Michaud M, DuBois A, Delaney ML, Punit S, Karlsson M, Bry L, Glickman JN, Gordon JI, Onderdonk AB, Glimcher LH. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* 2010; 8:292-300.
45. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* 2007; 2:119-129.
46. Liang C, Tseng HC, Chen HM, Wang WC, Chiu CM, Chang JY, Lu KY, Weng SL, Chang TH, Chang CH, Weng CT, Wang HM, Huang HD. Diversity and enterotype in gut bacterial community of adults in Taiwan. *BMC Genomics* 2017; 18:932.
47. Castaño-Rodríguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell HM, Kaakoush NO. Gut microbiome analysis identifies potential etiological factors in acute gastroenteritis. *Infect. Immun.* 2018; 86.
48. Zhang G, Gao B, Adeolu M, Khadka B, Gupta RS. Phylogenomic analyses and comparative studies on genomes of the Bifidobacteriales: identification of molecular signatures specific for the order Bifidobacteriales and its different subclades. *Front. Microbiol.* 2016; 7:978.
49. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. Human genetics shape the gut microbiome. *Cell* 2014; 159:789-799.
50. Santanasto AJ, Goodpaster BH, Kritchevsky SB, Miljkovic I, Satterfield S, Schwartz AV, Cummings SR, Boudreau RM, Harris TB, Newman AB. Body composition remodeling and mortality: the Health Aging and Body Composition Study. *J. Gerontol. A Biol. Sci. Med. Sci.* 2017; 72:513-519.
51. Baumgartner RN. Body composition in healthy aging. *Ann. N Y Acad. Sci.* 2000; 904:437-448.

52. Pantsulaia I, Iobadze M, Kikodze N, Pantsulaia N, Chikovani T. lipid profile and cytokines interactions during successful aging. *Georgian Med. News* 2015; 46-51.
53. Barnett C, Nazzal L, Goldfarb DS, Blaser MJ. The presence of oxalobacter formigenes in the microbiome of healthy young adults. *J. Urol.* 2016; 195:499-506.
54. PeBenito A, Nazzal L, Wang C, Li H, Jay M, Noya-Alarcon O, Contreras M, Lander O, Leach J, Dominguez-Bello MG, Blaser MJ. Comparative prevalence of Oxalobacter formigenes in three human populations. *Sci. Rep.* 2019; 9:574.
55. Le Roy T, Llopis M, Lepage P, Bruneau A, Rabot S, Bevilacqua C, Martin P, Philippe C, Walker F, Bado A, Perlemuter G, Cassard-Doulier AM, Gérard P. Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. *Gut* 2013; 62:1787-1794.
56. Kwok AH, Li Y, Jiang J, Jiang P, Leung FC. Complete genome assembly and characterization of an outbreak strain of the causative agent of swine erysipelas--*Erysipelothrix rhusiopathiae* SY1027. *BMC Microbiol.* 2014; 14:176.
57. Palm NW, de Zoete MR, Cullen TW, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* 2014; 158:1000-1010.
58. Martínez I, Perdicaro DJ, Brown AW, Hammons S, Carden TJ, Carr TP, Eskridge KM, Walter J. Diet-induced alterations of host cholesterol metabolism are likely to affect the gut microbiota composition in hamsters. *Appl. Environ. Microbiol.* 2013; 79:516-524.
59. Kim HN, Yun Y, Ryu S, Chang Y, Kwon MJ, Cho J, Shin H, Kim HL. Correlation between gut microbiota and personality in adults: a cross-sectional study. *Brain Behav. Immun.* 2018; 69:374-385.
60. Qian L, Gao R, Hong L, Pan C, Li H, Huang J, Qin H. Association analysis of dietary habits with gut microbiota of a native Chinese community. *Exp. Ther. Med.* 2018; 16:856-866.
61. Domingo MC, Yansouni C, Gaudreau C, Lamothe F, Lévesque S, Tremblay C, Garceau R. *Cloacibacillus* sp., a potential human pathogen associated with bacteremia in Quebec and New Brunswick. *J. Clin. Microbiol.* 2015; 53:3380-3383.
62. Brahe LK, Le Chatelier E, Prifti E, Pons N, Kennedy S, Hansen T, Pedersen O, Astrup A, Ehrlich SD, Larsen LH. Specific gut microbiota features and metabolic markers in postmenopausal women with obesity. *Nutr. Diabetes* 2015; 5:e159.
63. Yang Y, Weng W, Peng J et al. *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating toll-like receptor 4 signaling to nuclear factor- κ B, and up-regulating expression of MicroRNA-21. *Gastroenterology* 2017; 152:851-866.e24.
64. Biagi E, Candela M, Turrone S, Garagnani P, Franceschi C, Brigidi P. Ageing and gut

microbes: perspectives for health maintenance and longevity. *Pharmacol. Res.* 2013; 69:11-20.

65. Wang F, Yu T, Huang G, et al. Gut microbiota community and its assembly associated with age and diet in Chinese centenarians. *J. Microbiol. Biotechnol.* 2015; 25:1195-1204.
66. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, Nikkila J, Monti D, Satokari R, Franceschi C, Brigidi P, De Vos W. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* 2010; 5:e10667.
67. Biagi E, Rampelli S, Turroni S, Quercia S, Candela M, Brigidi P. The gut microbiota of centenarians: signatures of longevity in the gut microbiota profile. *Mech. Ageing Dev.* 2017; 165:180-184.
68. Angelakis E, Yasir M, Bachar D, Azhar EI, Lagier JC, Bibi F, Jiman-Fatani AA, Alawi M, Bakarman MA, Robert C, Raoult D. Gut microbiome and dietary patterns in different Saudi populations and monkeys. *Sci.Rep.* 2016; 6:32191.
69. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* 2006; 9:55-61.
70. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D, Clark AG. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* 2015; 16:191.
71. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* 2016; 48:1413.
72. Flachsbart F, Dose J, Gentschew L, et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat. Commun.* 2017; 8:2063.
73. Nebel A, Kleindorp R, Caliebe A, et al. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* 2011; 132:324-330.
74. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 2013; 79:5112-5120.
75. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 2009; 75:7537-7541.
76. Huse SM, Welch DM, Morrison HG, Sogin ML. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 2010; 12:1889-1898.

77. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *Peer J*. 2016; 4:e2584.
78. Westcott SL, Schloss PD. OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. *mSphere*. 2017; 2.
79. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 2007; 73:5261-5267.
80. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 2013; 41:D590-D596.
81. Benson AK, Kelly SA, Legge R, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. U S A* 2010; 107:18933-18938.
82. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. *J. Stat. Softw.* 2016; 70.
83. Oksanen JO, Blanchet GFB, Friendly MF, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. *vegan: Community Ecology Package*. 2017 [Available at: <http://cran.r-project.org/>].
84. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J. Royal Stat. Soc. Ser. B* 2001; 63:411-423.

Tables

Table 4.1. Environmental factors that significantly contribute to shaping of the microbial structure of the study population.

Environmental factor	Explanatory variable	Explained variability							
		ET_Ech		ET_P		ET_F		ET_B	
		By variable (%)	Total (%)	By variable (%)	Total (%)	By variable (%)	Total (%)	By variable (%)	Total (%)
Anthropometrics	Age	0.92		0.32		1.33		0.41	
	WHR	1.61		2.1		0.24		1.44	
	BMI	0.85							
Blood measurements	Triglycerides					0.33		0.48	
	Glucose							0.11	
	Calprotectin			0.83					
Dietary	Alcohol	0.52	4.58	1.1			2.7	0.31	3.9
	Long-Chain fatty acids							0.16	
	Fibre							0.98	
	Total energy			0.55	4.9				
Genetic	HIVEP3_rs12563071_T	0.68							
	ACTL8_rs6666120_G					0.54			
	KCNC4_rs958798_T					0.21			

BMI, Body Mass Index; ET, enterotype [Ech, *Escherichia shigella*; P, *Prevotella*; F, *Firmicutes*; B, *Bacteroides*]; WHR, Waist-to-hip ratio

Figure legends

Figure 4.1. Age distribution of the individuals enrolled in the study. The age distribution is shown separately for males and females.

Figure 4.2. Relationships among anthropometric measures, dietary parameters and the age of the individuals. A - B. Principal component analysis showing how dietary and anthropometric parameters shape the inter-individual variability of the 1,301 healthy individuals. **C - E.** Linear regression between anthropometric parameters and age. **C.** Prediction model was explained with determination coefficient of 0.395 for males and 0.306 for females. Determination coefficient (R^2) was adjusted for the number of prediction variables in the model. **D.** Residual plot of the prediction models.

Figure 4.3. Gut bacterial diversity of healthy Germans across a broad age range. Relative abundances of the ten most abundant phyla, orders, families and genera for each age group.

Figure 4.4. Microbial stratification and enterotype-specific alpha diversity. A. Principal coordinate analysis (PCoA) based on Jensen-Shannon dissimilarity matrix on the basis of the GM profiles of the subjects. **B.** Optimal number of clusters estimated by the gap statistic method with a 500 times Monte Carlo re-sampling. **C.** Enterotype *de novo* classification in comparison with the enterotypes obtained with the enterotype classifier (<http://enterotypes.org>). **D.** Enterotype distribution in relation to the age sets, population-cohorts and gender. **E.** GM diversity based on the Shannon diversity index in relation to the enterotypes, population-cohorts and gender. **F.** GM diversity based on the Shannon diversity index in relation to the enterotypes, population-cohorts and age sets.

Figure 4.5. Inter-individual GM variability explained by microbial stratification, dietary and anthropometrics parameters. A. Redundancy analysis (RDA) biplot representing the relationship between the explanatory variables and the relative abundance of bacterial communities. The biplot map explains 15.9% of the total variability. **B.** Contribution plot for the explanatory variables; depicted variables explain 21.6% of the total variability.

Figure 4.6. Taxa whose abundances change during aging. Forty-three genera whose occurrence is associated with age.

Figure 4.7. Differentially abundant taxa in LLI compared with younger subjects. Twenty-two genera with differential abundance between LLI and younger subjects.

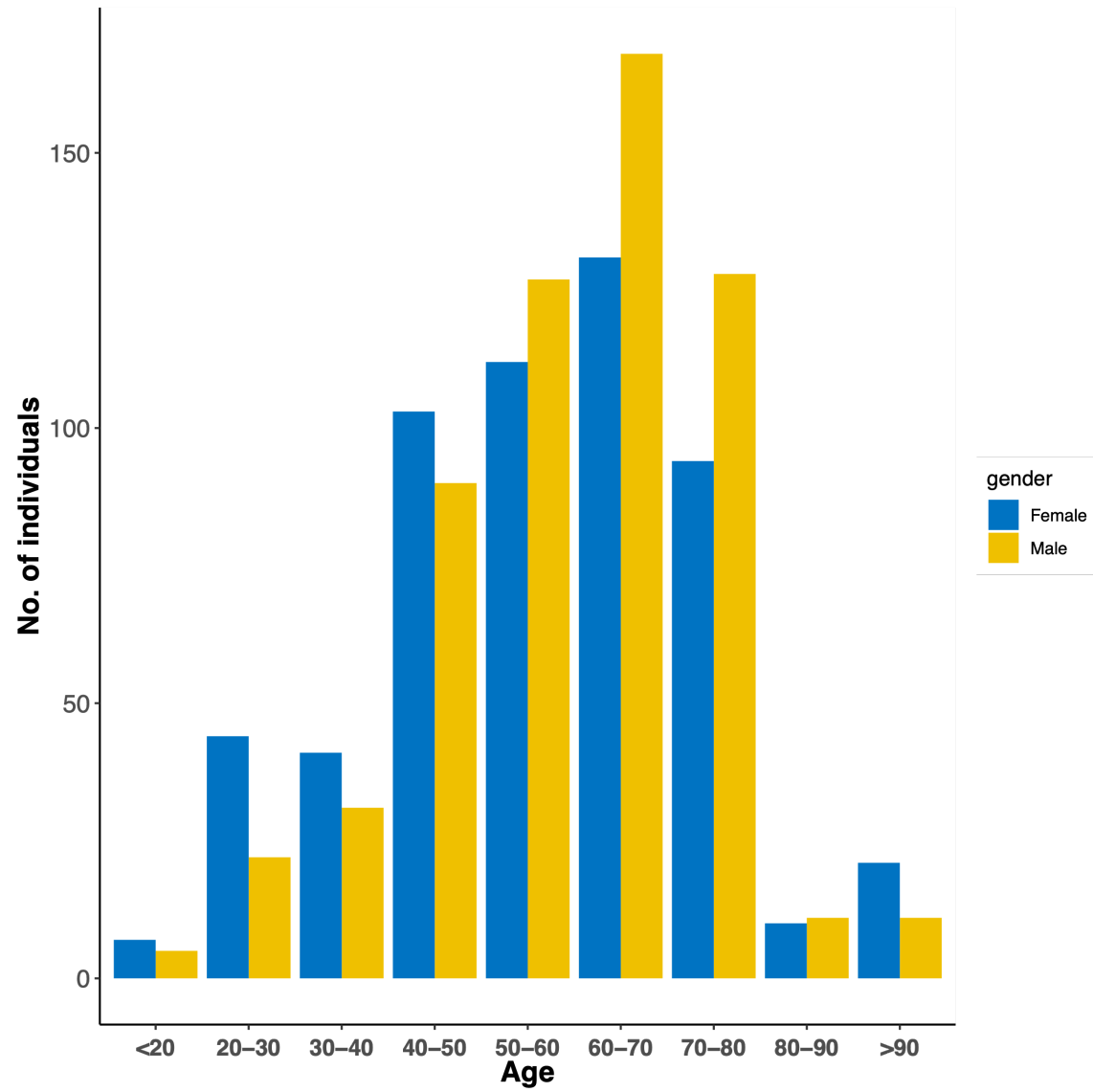


Figure 4.1

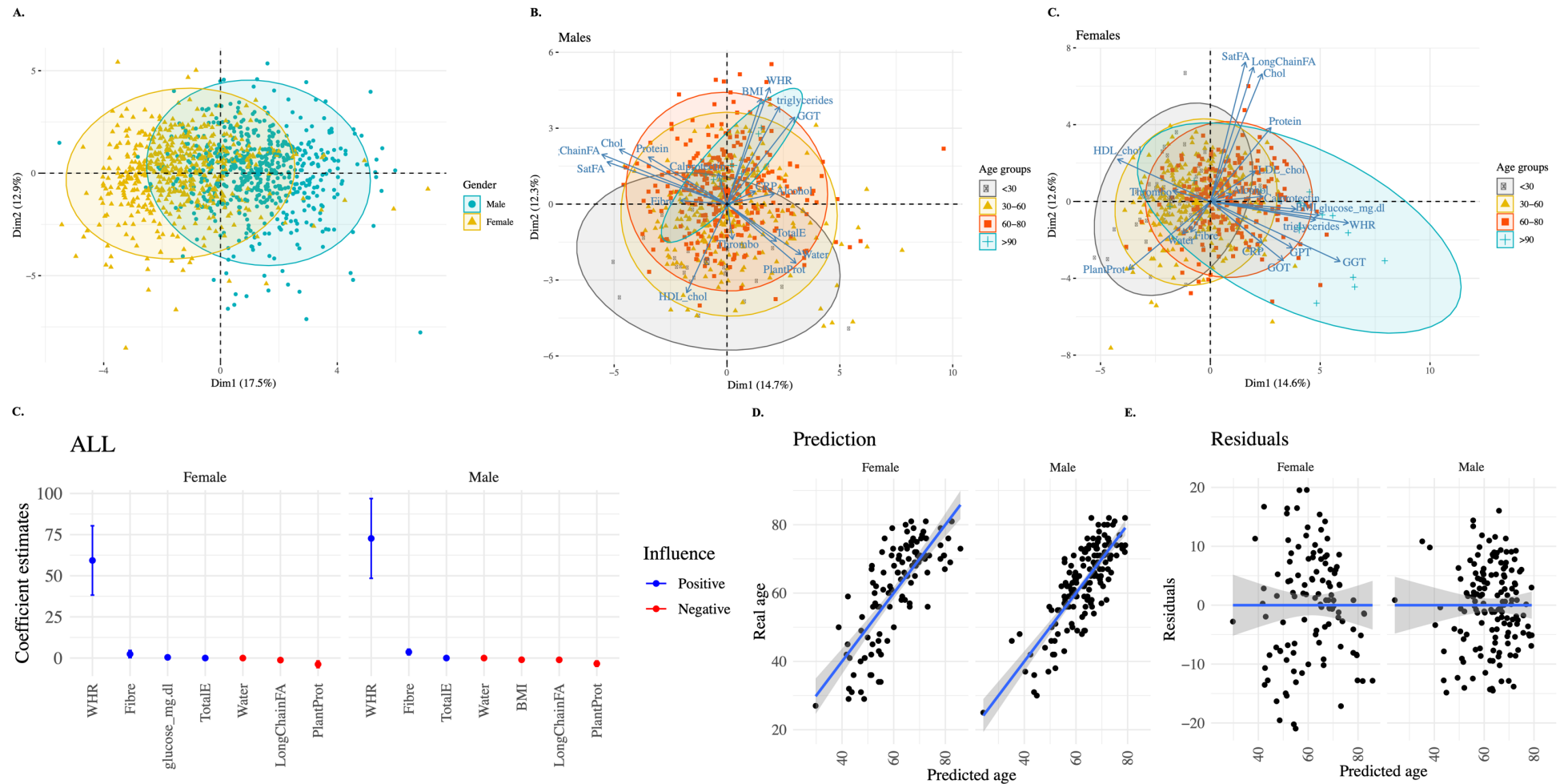


Figure 4.2

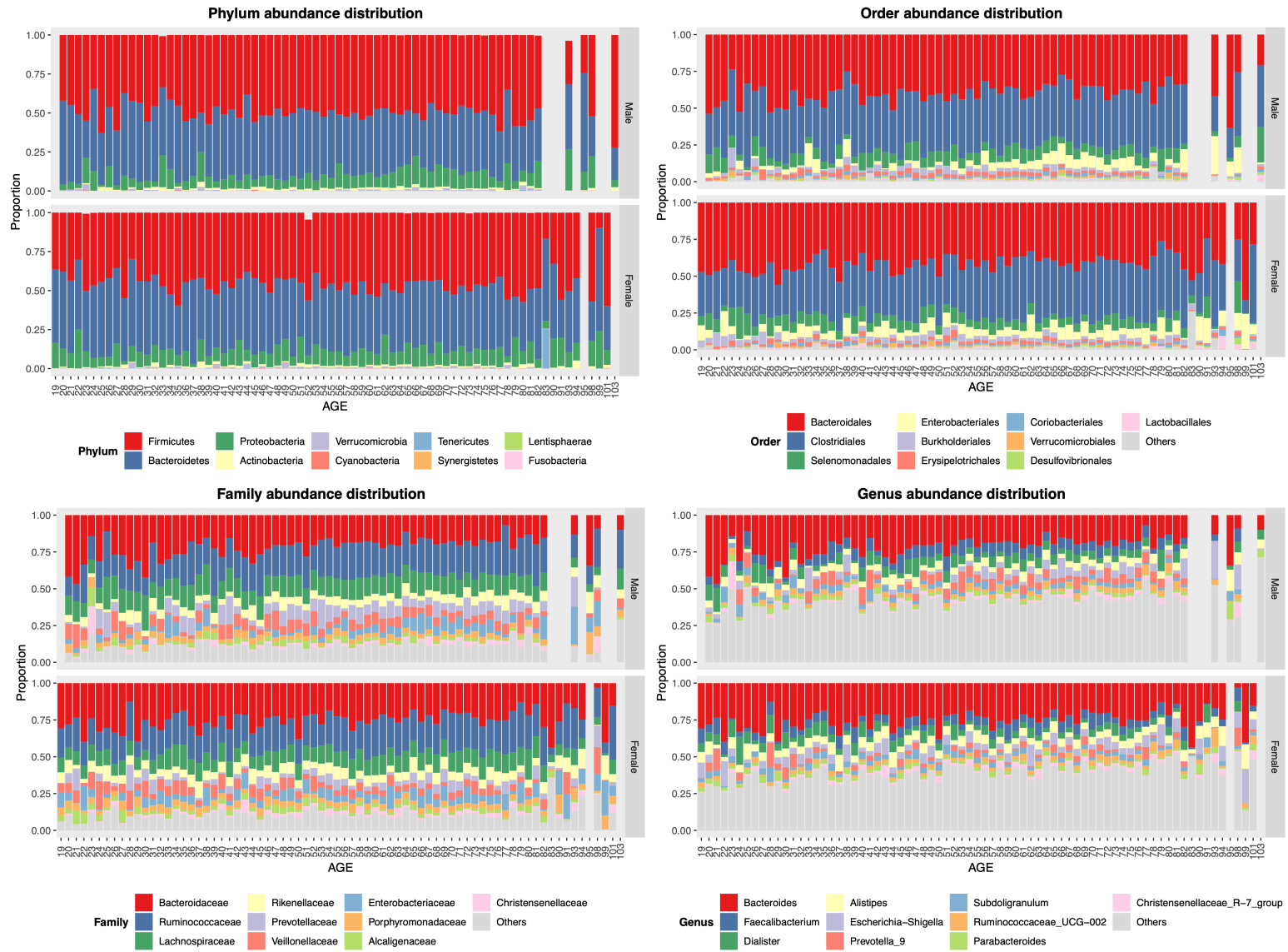


Figure 4.3

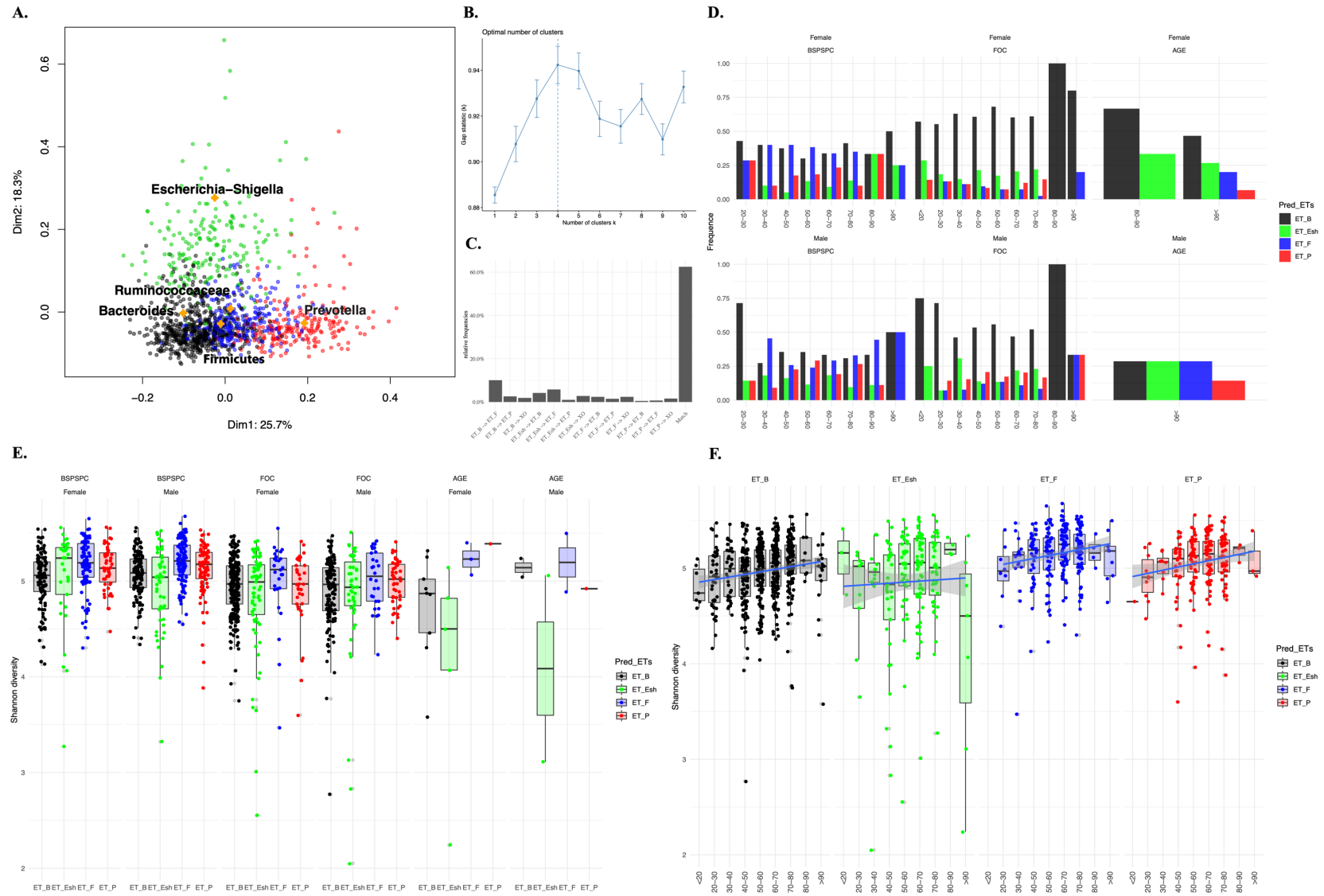
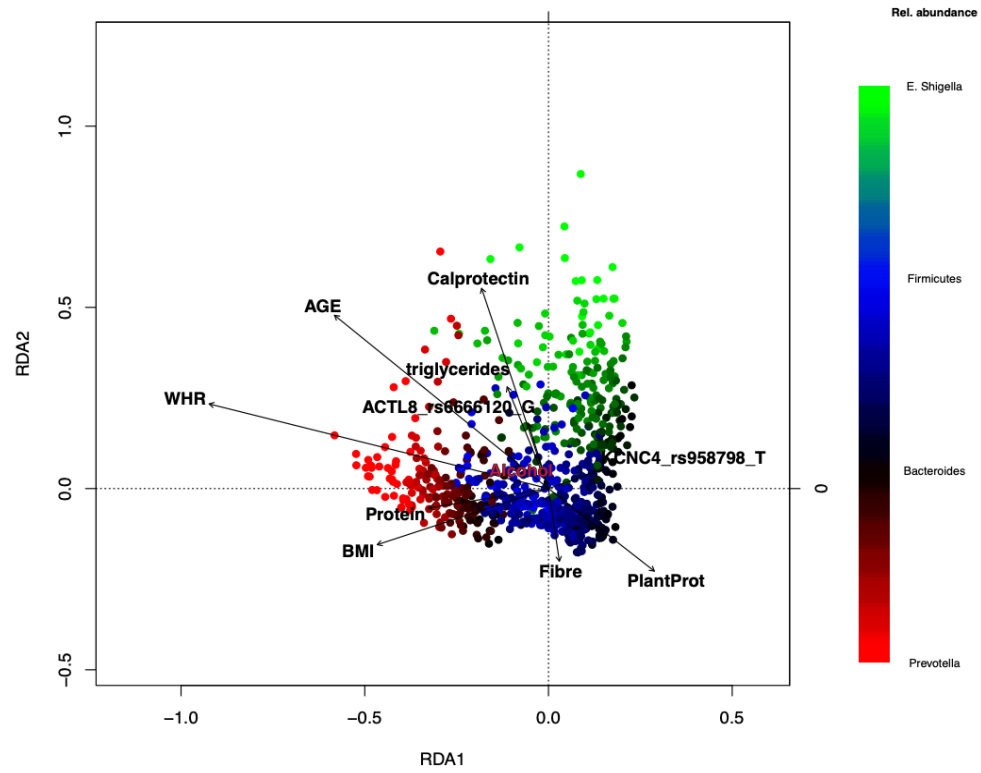
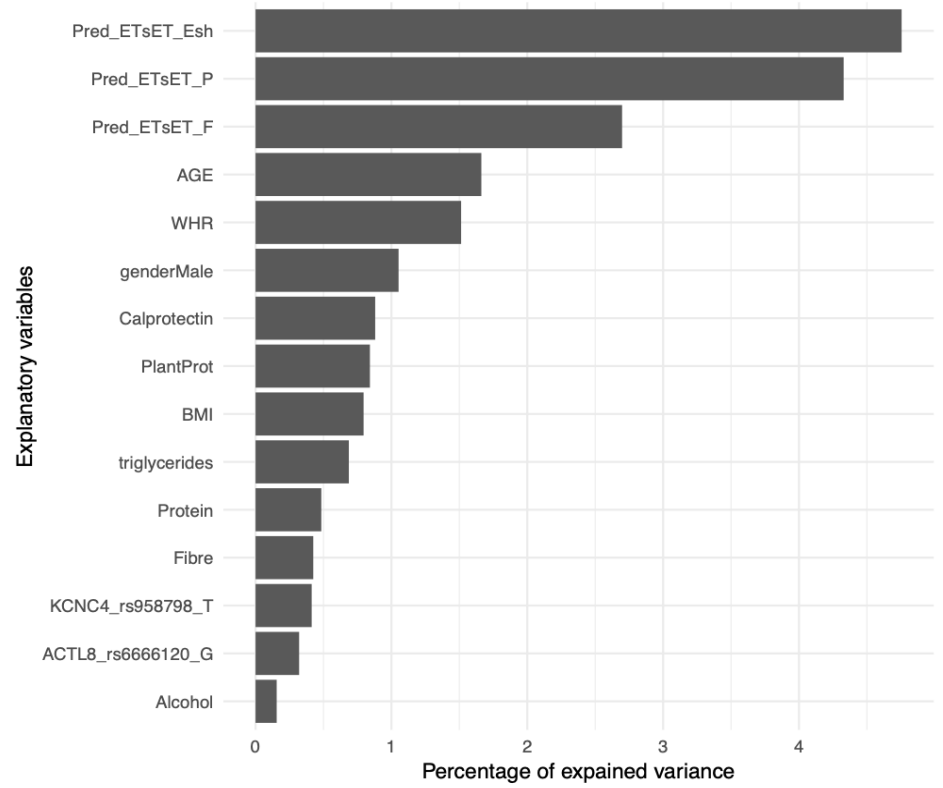


Figure 4.4

A.**B.****Figure 4.5**

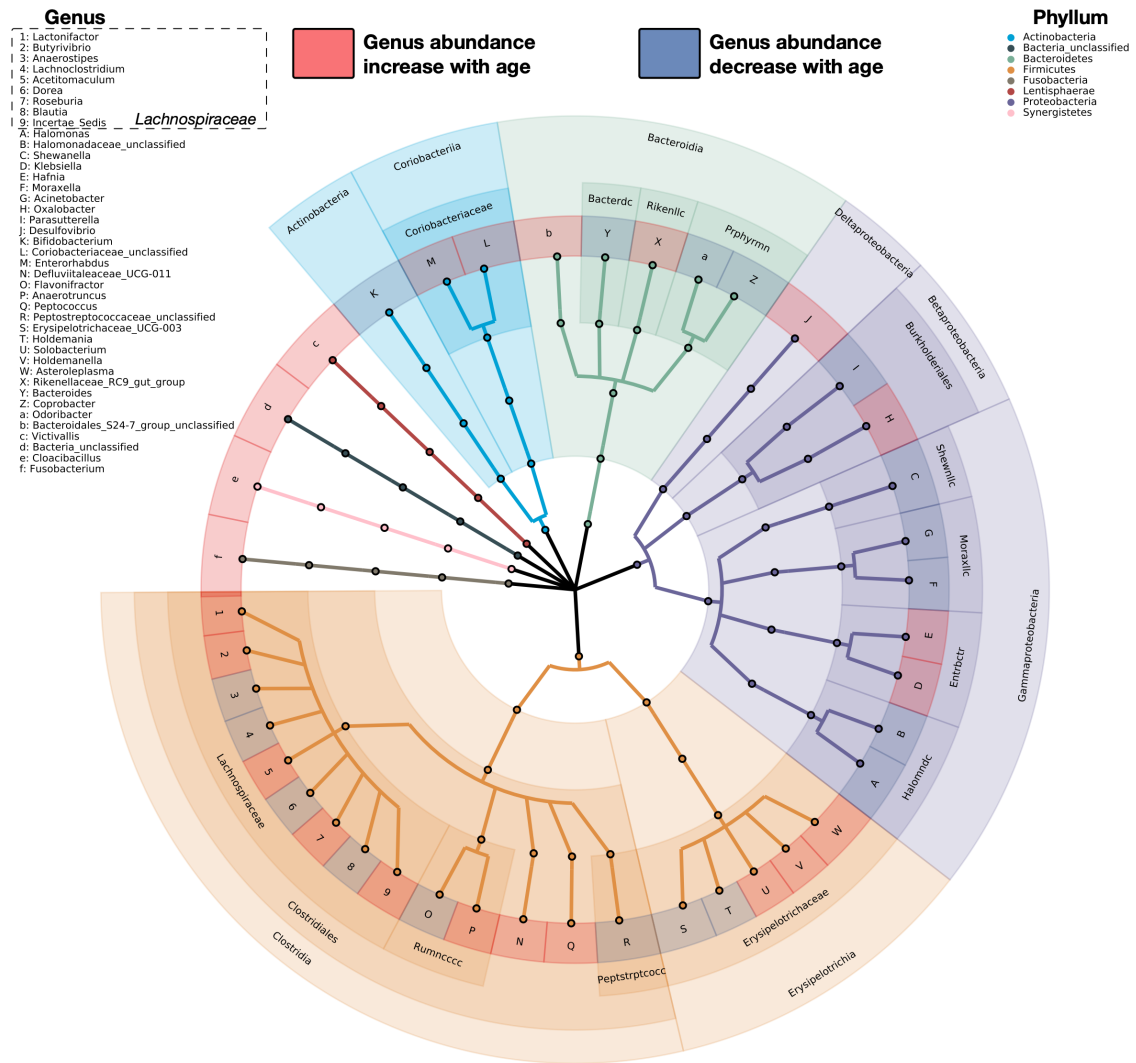


Figure 4.6

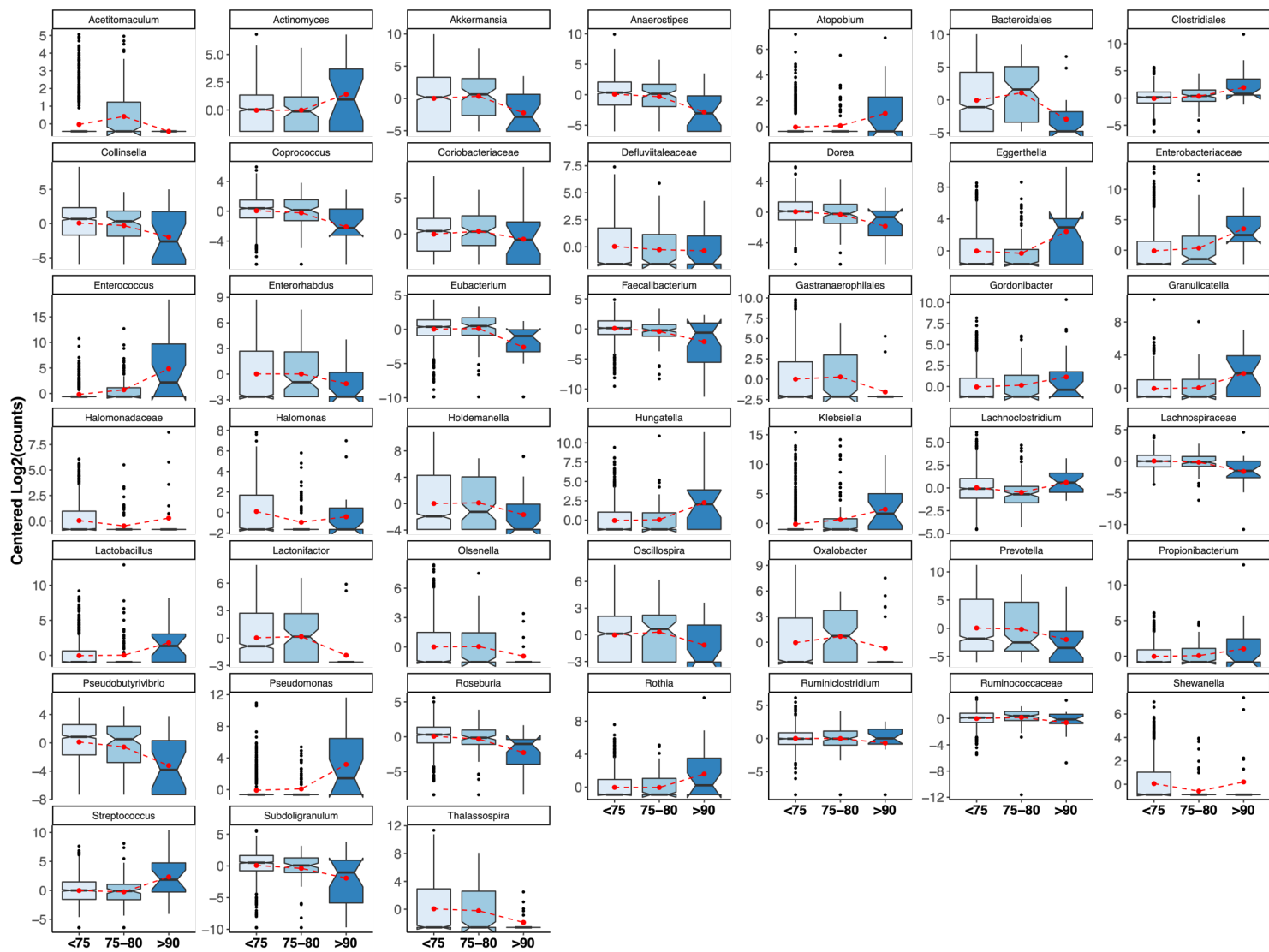


Figure 4.7

5.

Conclusion and outlook

“Aging seems to be the only available way to live a long life.”

Daniel-Francois-Esprit Auber, French composer 1782-1871

Every human being on earth faces aging and, so far, no one can escape the aging process. Only a few individuals achieve the very special phenotype "longevity". While LLI also do age and show clear signs of aging (wrinkles, grey hair, etc.), they seem to grow old in a more healthy manner than average-lived people. Disentangling the tight link between aging and longevity is difficult and has been occupying researchers all over the world for decades. The work presented in this thesis has clearly shown that we are still far from a comprehensive understanding of the highly complex mechanisms influencing (healthy) aging and longevity. Thus, only a broad view of these phenomena will eventually enable us to identify the underlying regulatory pathways.

Figure 5.1 summarizes the aims, methods and results of the three chapters of the present work. Chapter I comprised basic research on aging and development in the relatively simple model organism *Hydra*. Fundamental aging mechanisms and involved genes overlap widely between different species like *Hydra*, *Drosophila* or mice and humans. Thus, research in animals can help overcome experimental and ethical limitations that are often encountered in studies involving humans. Chapter III (gut microbiome in aging) once more highlights the complexity of aging that is exacerbated if we consider human genome–microbiome–environment interactions. Another challenge in aging studies is to unravel causes of aging from its consequences. This challenge was met in chapter II (genetics of human longevity), where the identified SNVs are constitutive and must therefore underlie the longevity phenotype instead of being a consequence of it.

Each chapter is a stand-alone project; at different levels, however, they all revolved around the central topic of how to influence or manipulate the aging process to achieve healthy longevity. Every chapter also revealed either limitations of the respective study design or results that cannot be explained with our current level of understanding.



Figure 5.1. Overview of the three research projects and their respective outcomes.

CI, control individuals; COG, cluster of orthologous groups; GM, gut microbiome; HDAC, histone deacetylase; KEGG, Kyoto Encyclopedia of Genes and Genomes; LLI, long-lived individuals; SNV, single nucleotide variant.

Intriguingly, in the three chapters, FOXO has not shown up as a common denominator. The important role of this gene in longevity and lifespan regulation is undisputed. Not surprisingly, FOXO was recently suggested as a "critical rate-of-aging-regulator" [1].

In *Hydra*, FoxO drives developmental processes and stem cell renewal [1]. In the *Hydra* subproject (chapter I) FoxO expression was detected in all the *in situ* GFP-experiments and also

the differential FoxO expression between body and head was observed. However, for unknown reasons, this differential FoxO expression was not confirmed either in the microarray or the RNA sequencing data. Furthermore, the epigenetic modifications in terms of DNA deacetylation apparently left FoxO expression patterns unaffected; at least, they did not change to a measurable extent. These results were completely unexpected. It could well be that FoxO expression in *Hydra* is restricted to one specific tissue layer (most likely the epidermal layer (ectoderm)) and that by performing a transversal cut through the *Hydra* tissue during the experiment, the expression was diluted to a level below the detection limit. Single-cell transcriptomics or layer-specific analyses could help to answer this question in the future.

FoxO was recently shown to fundamentally affect both the microbiome composition and resilience in *Hydra* [2]. *Drosophila* exhibited an age-related increase in FoxO activity, which correlated with a higher bacterial load [3]. Congruently, *Drosophila* with impaired FoxO signaling were found to be more prone to intestinal infections [4]. In the nematode *C. elegans*, longevity was linked to *Bacillus subtilis* biofilm formation, driven by Daf16, the worm orthologue of human *FOXO3* [5, 6]. Furthermore, the pro-longevity effect of metformin, a potent *FOXO3*-inducing drug [7], is known to be indirectly mediated by its influence on bacteria, particularly via inhibiting bacterial folate and methionine metabolism [8]. Taking these observations into account, it is surprising that in the microbiome subproject (chapter III) no influence of *FOXO3* variants on the intestinal microbiome was detected. Studies in model organisms are generally highly standardized in terms of environment, genetic background, and diet, while in humans a potential influence of *FOXO3* variants on the microbiome – if present – was likely masked by other factors (e.g. diet, genetic background, etc.).

FOXO3 variants were consistently reported to be associated with human longevity. In chapter II, no *FOXO3* variants were significantly associated with longevity. However, this was anticipated as coding variants in *FOXO3* had previously been shown to be unlikely key determinants - the *FOXO3*-longevity-association rather seems to be driven by intronic SNVs [9]. Therefore, exonic *FOXO3* variants covered by the exome chip were *a priori* not expected to yield statistically significant longevity associations.

It is intriguing that despite extensive research efforts, all currently known longevity variants together explain not more than approximately 1 or 2% of the estimated total heritability [10]. This may point towards an analytical bias that has left many genetic factors contributing to human longevity undiscovered so far, in agreement with the research hypothesis of chapter II.

However, another possibility is that the longevity heritability is highly overestimated, as claimed in a recent paper [10]. The latter option is currently the subject of controversial discussion in the longevity community. The work on chapter II also revealed limitations of the commonly applied approaches. The analysis of joint effects of rare and common variants within one genetic region added substantial value to the initially performed traditional single-variant association analysis. This highlights the importance of being open to new methods and/or combinations of existing ones as they can help to discover new or validate previous findings. Besides, it might be time to rethink longevity association studies - hypothesis-free GWAS prove successful in discovering genetic variants with huge effect sizes or associated with highly prevalent diseases/phenotypes, but they are obviously unrewarding in longevity research. Hypothesis-driven, potentially evolutionary-informed, approaches may be more promising. Furthermore, up to date, there is no consensus about an/the ideal design for longevity studies in humans, if there is any at all. As a result, published studies are often hard, if not impossible, to compare. Longevity is such a rare phenotype that the recruitment of LLI, especially centenarians, semi-super-centenarians and super-centenarians is very challenging. The selection of proper control individuals is not less problematic when taking into account birth cohort effects [11, 12]; there are generally no controls available from the same birth cohort the LLI belong to. Cut-offs for cases and controls appear highly variable and sometimes even specific to one study. Both hard age cut-offs or age ranges (90+ yrs, 95+ yrs, 98+ yrs, 100+ yrs and 50-65 yrs, <65 yrs, <70 yrs, <85 yrs for cases and controls, respectively) and percentile thresholds (e.g. >90th/ 99th percentile and <60th percentile for cases and controls, respectively) are chosen apparently almost arbitrarily. In view of these challenges, studies with centenarian offspring and parental lifespan, respectively, appear promising. However, these studies come with the drawback of genotype and phenotype “dilution”.

Environmental factors affect longevity to a much greater extent than genetics, probably even more than the currently estimated ~70-80%, if in fact the lower heritability of the phenotype proves to be correct. Noteworthy, in contrast to the unchangeable genetic architecture, environmental factors are generally modifiable.

The gut microbiome is increasingly gaining attention in aging and longevity research. Maintaining a healthy gut microbial architecture appears to be of utmost importance for staying healthy while growing old. In chapter III, the microbial composition remained mostly stable in healthy individuals during aging, which supports previous findings [13]. Strikingly, age-related

diseases often coincide with dysbiosis [14, 15]. The results obtained in chapter III highlighted the relevance of environmental factors as determinants of the gut microbial composition. The thesis has shown that with only 1% in magnitude, the effects of host genetics on the microbial structure in the intestine were remarkably low, while all collected environmental factors combined contribute 5% when the microbial population stratification was removed. This is in concordance with results obtained by others [16]. The gut microbiota can be regarded as a major target of intervention strategies to maintain or restore a stable healthy gut flora. Indeed, first attempts to restore a healthy gut microbial structure, e.g. by treatment with pre- and probiotics or fecal transplantation yielded promising results [16]. Infants may deserve special attention as the importance of environmental factors might be greatest in this age group when the microbial architecture has not stabilized yet. Therefore, early factors (e.g. delivery, breast milk, animal or formula milk, hygiene) likely have long-term consequences. Noteworthy, the example of the metformin-treated *C. elegans* mentioned above supports the notion that the individual microbial metabolic potential can affect the therapeutic response to drugs. This might eventually enable us to develop personalized approaches to fight aging and age-related diseases and to achieve longevity.

Generally, there is low consistency among microbiome studies. Some of the observed differences may be, at least partly, the result of cultural differences in lifestyle and diet. However, also methodological issues may contribute to discrepant observations, for instance, differences in the sampling procedure, DNA extraction protocol, or handling of sequencing errors, to name but a few. To avoid false conclusions, comparisons should ideally be restricted to within one study. Furthermore, in future aging-microbiome-studies, it might be advisable to stratify elderly participants according to their health status to avoid too much variation in the data. This approach would require bigger population sizes to retain statistical power.

One challenge that remains is to distinguish between effects due to the aging process per se and those caused by changes in diet and lifestyle associated with aging. The thesis contributed to this with deep bioinformatic analyses; however, the complexity of the data and the fact that the study participants derived from three different sample collections, for which different (though overlapping) phenotypes were collected, made this work particularly challenging. Large-scale studies with comprehensive phenotyping and categorization of participants by health status might be one way to address the current challenges of microbiome research. To summarize, gut microbes are clearly associated with both health and disease and the environment impacts

considerably on both the gut microbial composition and the likelihood of becoming long-lived. With this in mind, the identification of the key environmental drivers of these two phenotypes, gut microbial architecture and longevity, should be a focus in future studies.

References

1. Boehm A-M, Khalturin K, Anton-Erxleben F, Hemmrich G, Klostermeier UC, Lopez-Quintero JA, et. al. FoxO is a critical regulator of stem cell maintenance in immortal Hydra. *Proc. Natl Acad. Sci. U S A* 2012; 109:19697-19702.
2. Mortzfeld BM, Taubenheim J, Fraune S, Klimovich AV, Bosch TCG. Stem cell transcription factor FoxO controls microbiome resilience in Hydra. *Front. Microbiol.* 2018; 9:629.
3. Rera M, Clark RI, Walker DW. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. *Proc. Natl Acad. Sci. U S A.* 2012; 109:21528-21533.
4. Fink C, Hoffmann J, Knop M, Li Y, Isermann K, Roeder T. Intestinal FoxO signaling is required to survive oral infection in Drosophila. *Mucosal Immunol.* 2016; 9:927-936.
5. Lee SS, Kennedy S, Tolonen AC, Ruvkun G. DAF-16 target genes that control *C. elegans* life-span and metabolism. *Science* 2003; 300:644-647.
6. Webb AE, Kundaje A, Brunet A. Characterization of the direct targets of FOXO transcription factors throughout evolution. *Aging Cell* 2016; 15:673-685.
7. Sato A, Sunayama J, Okada M, Watanabe E, Seino S, Shibuya K, et. al. Glioma-initiating cell elimination by metformin activation of FOXO3 via AMPK. *Stem Cells Transl. Med.* 2012; 1:811-824.
8. Cabreiro F, Au C, Leung KY, Vergara-Irigaray N, Cochemé HM, Noori T, et. al. Metformin retards aging in *C. elegans* by altering microbial folate and methionine metabolism. *Cell* 2013; 153:228-239.
9. Flachsbart F, Dose J, Gentschew L et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat. Commun.* 2017; 8:2063.
10. Ruby JG, Wright KM, Rand KA et al. Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics* 2018; 210:1109-1124.
11. Cleries R, Martínez JM, Valls J, Pareja L, Esteban L, Gispert R, Moreno V, Ribes J, Borràs JM. Life expectancy and age-period-cohort effects: analysis and projections of mortality in Spain between 1977 and 2016. *Public Health* 2009; 123:156-162.
12. Giuliani C, Garagnani P, Franceschi C. Genetics of human longevity within an eco-evolutionary nature-nurture framework. *Circ. Res.* 2018; 123:745-772.
13. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. The long-term stability of the human gut microbiota. *Science* 2013; 341:1237439.
14. Zhang YJ, Li S, Gan RY, Zhou T, Xu DP, Li HB. Impacts of gut bacteria on human health and diseases. *Int. J. Mol. Sci.* 2015; 16:7493-7519.
15. Tang WH, Kitai T, Hazen SL. Gut microbiota in cardiovascular health and disease. *Circ. Res.* 2017; 120:1183-1196.
16. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 2018; 555:210-215.

6. Summary

Human aging is characterized by progressive functional decline that coincides with both increased morbidity and mortality. Aging affects every human being and only few individuals achieve longevity, a very special phenotype marked by extraordinary healthy aging. This thesis consists of three chapters; each one is devoted to a separate project that contributes to the growing body of knowledge about aging and longevity. The work required the compilation, management and analysis of diverse big data sets and the application of cutting-edge statistical and computational methods.

Chapter 1 - A functional genomics study was conducted in the potentially immortal freshwater polyp *Hydra* using body part-specific microarray and RNA sequencing data. The results revealed gene expression patterns that allow boundary maintenance during *Hydra*'s continuous cell proliferation and tissue self-renewal. Furthermore, this study provided evidence for de-acetylation as a key mechanism underlying compartmentalization. Surprisingly, FoxO, which is known to substantially drive developmental processes and stem cell renewal in *Hydra*, did not seem to be affected by the acetylation status.

Chapter 2 - Long-lived individuals (LLI, >95 years of age) epitomize the healthy aging phenotype and are thought to carry beneficial genetic variants that predispose to human longevity. Despite extensive research efforts, only few of these genetic factors in LLI have been identified so far. In contrast to previous investigations which mainly focused on intronic variants, a genome-wide exome-based case-control study was performed. DNA samples of more than 1,200 German LLI, including 599 centenarians (≥ 100 years), and about 6,900 younger controls were used for single-variant and gene-based association analyses that yielded two new candidate longevity genes, fructosamine 3 kinase related protein (*FN3KRP*) and phosphoglycolate phosphatase (*PGP*). FN3KRP functions in the deglycation of proteins to restore their function, while PGP via controlling glycerol-3-phosphate levels affects both glucose and fat metabolism. Given the biological functions of the genes, their longevity-associations appear very plausible.

Chapter 3 - In recent years, the intestinal microbiome (GM) has increasingly gained attention in aging and longevity research. A 16S rRNA microbiome study was conducted using 1301 stool samples of healthy individuals (age range: 19 - 104 years) that were drawn from three cohorts. The aim was to investigate potential associations among GM composition, host genetics and environmental factors during aging. The GM composition changed with age, showing an increase of opportunistic pathogens that may generate an inflammatory environment in the gut. Age explained only ~1% of the inter-individual variation, whereas anthropometric measures, genetic background and dietary patterns together explained 20%. Strikingly, clear GM population stratification in terms of four enterotype-like clusters was observed, which were predominantly associated with dietary patterns. The correction for these clusters was shown to increase the comparability of findings from the different cohorts. In addition, the LLI showed a specific gut microbial pattern, which is in line with previously published reports.

The present work shows that a thorough bioinformatics expertise helps to address the complexity of the two phenotypes aging and longevity. One highlight of the thesis is the discovery of two new candidate longevity loci that, in view of the limited output of previous study approaches, enlarge the existing database.

7.

Zusammenfassung

Altern geht beim Menschen mit fortschreitender Funktionalitätsabnahme des Organismus und erhöhter Morbidität und Mortalität einher. Jeder Mensch altert, jedoch nur wenige Individuen erreichen *Langlebigkeit* - ein besonderer Phänotyp, der durch eine außergewöhnlich gesunde Form des Alterns gekennzeichnet ist. Die vorliegende Doktorarbeit besteht aus drei Kapiteln, die jeweils einem separaten Projekt gewidmet sind. Alle drei Projekte tragen zur Erweiterung des Kenntnisstandes zu *Altern* und *Langlebigkeit* des Menschen bei. Die geleistete Forschungsarbeit erforderte Zusammenstellung, Management und Analyse diverser großer Datensätze sowie den Einsatz modernster statistischer und bioinformatischer Methoden.

Kapitel 1 - Am potenziell unsterblichen Süßwasserpolyt *Hydra* wurde eine funktionell-genomische Studie unter Verwendung von körperbereich-spezifischen Microarray- und RNA-Sequenzierungsdaten durchgeführt. Die Ergebnisse zeigten Genexpressionsmuster, die die Aufrechterhaltung abgegrenzter Körperbereiche während der kontinuierlichen Zellproliferation und Selbsterneuerung der Gewebe von *Hydra* ermöglichen. Darüber hinaus lieferte die Studie Belege, dass Deacetylierungsreaktionen eine Schlüsselfunktion bei der Kompartimentierung zukommt. Überraschenderweise schien FoxO, ein Gen, das bekanntermaßen Entwicklungsprozesse und Stammzellerneuerung in *Hydra* wesentlich vorantreibt, vom Acetylierungsstatus nicht betroffen zu sein.

Kapitel 2 - Langlebige Individuen (engl. *long-lived individuals*, LLI; >95 Jahre) verkörpern den Phänotyp des gesunden Alterns. Es wird vermutet, dass sie genetische Varianten tragen, die für Langlebigkeit beim Menschen prädisponieren. Trotz umfangreicher Forschung konnten bisher nur wenige dieser genetischen Faktoren in LLI identifiziert werden. Im Gegensatz zu früheren Untersuchungen, die sich hauptsächlich auf intronische Varianten konzentrierten, wurde im Rahmen dieser Promotion eine genomweite exombasierte Fall-Kontroll-Studie durchgeführt. DNA-Proben von mehr als 1200 deutschen LLI, darunter 599 Hundertjährige (≥ 100 Jahre), und etwa 6900 jüngeren Kontrollen wurden für einzelvarianten- und genbasierte

Assoziationsanalysen verwendet. So konnten zwei neue Kandidatengene für Langlebigkeit identifiziert werden, Fructosamin-3-Kinase-related protein (*FN3KP*) und Phosphoglykolat-Phosphatase (*PGP*). *FN3KP* wirkt der Glykierung von Proteinen entgegen, während *PGP* über die Kontrolle des Glycerin-3-Phosphatspiegels sowohl den Glukose- als auch den Fettstoffwechsel beeinflusst. Angesichts dieser biologischen Funktionen erscheinen die Langlebigkeitsassoziationen der beiden Gene sehr plausibel.

Kapitel 3 - Das Darmmikrobiom (engl. *gut microbiome*, GM) hat in den letzten Jahren zunehmend an Bedeutung in der Alterungs- und Langlebigkeitsforschung gewonnen. Im Rahmen der Promotion wurde eine 16S rRNA-Mikrobiomstudie mit 1301 Stuhlproben von gesunden Personen (Altersbereich: 19-104 Jahre) aus drei verschiedenen Kohorten durchgeführt. Untersucht wurden potentielle Assoziationen zwischen GM-Zusammensetzung, Wirtsgenetik und Umweltfaktoren während des Alterns. Die GM-Zusammensetzung änderte sich mit zunehmendem Alter, insbesondere in Richtung einer Zunahme opportunistisch pathogener Bakterien, die ein entzündliches Umfeld im Darm erzeugen können. Das Alter *per se* erklärte lediglich ~1% der interindividuellen Variation, während anthropometrische Messungen, genetischer Hintergrund und Ernährungsmuster der Probanden zusammen 20% erklärten. Auffällig war die deutliche Stratifizierung der GM-Population in vier Enterotypen-ähnliche Cluster, die mit verschiedenen Ernährungsmustern assoziiert waren. Erst die Korrektur für diese Cluster ermöglichte die Vergleichbarkeit der Ergebnisse aus den verschiedenen Kohorten. Darüber hinaus zeigten LLI ein spezifisches GM-Muster, das mit zuvor veröffentlichten Berichten größtenteils übereinstimmt.

Die vorliegende Doktorarbeit zeigt auf, wie mittels bioinformatischer Expertise und Methodenentwicklung die Komplexität der Phänotypen *Altern* und *Langlebigkeit* angegangen werden kann. Angesichts des bisher begrenzten Forschungsstandes ist insbesondere die Entdeckung der neuen Kandidaten-Loci, die die Datengrundlage zur Langlebigkeit erweitern, eine zentrale Forschungsleistung.

8. Declaration

Herewith, I confirm that the submitted thesis is completely the result of my own work. Apart from the advice of my supervisors, all sources and cooperation partners are listed within the thesis. This thesis has not been submitted elsewhere. It has been carried out in strict accordance with the rules of Good Scientific Practice of the *Deutsche Forschungsgesellschaft*. In addition, no academic degree has ever been withdrawn.

Kiel,

.....

Guillermo Gonzalo Torres Estupiñan

9.

Curriculum vitae

Name: Guillermo Gonzalo Torres Estupiñan

Address: Holtenuer Str. 99, 24105 Kiel

Date of birth: 27 May 1985

Place of birth: San Juan de Pasto, Nariño, Colombia

Citizenship: Colombian

Marital status: Single

Basic education:

1993 - 2000 Elementary and high school education

Colegio Champagnat Pasto

2000 - 2001

High school education (Academic degree)

Colegio José Artemio Mendoza Carvajal

Higher education:

2003 - 2009

Undergraduate degree in Biology

National University of Colombia

2010 - 2013

Master of Science, Major Genetics

Institute of Biotechnology

National University of Colombia

2015 - Today

Ph.D. student

Institute of Clinical Molecular Biology

Christian-Albrechts-University, Kiel

10.

Acknowledgements

I will always ascribe this work to myself. But the persons really responsible, as they are for most of the best things in my life, are my parents. He, who has never stopped encouraging my curiosity towards the wonders of nature and who infected my childhood with ideas of "wise-people" with PhDs. She, who has put her motherly selfishness aside to support a dream and who has always balanced my eccentricities. Thank you, Libardo G. Torres, Flor Alba Estupiñan, and Mercedes Benavidez.

The journey of my PhD, which is coming to an end, started in 2015. Many thanks to Prof. Dr. Thomas Roeder for agreeing to be my first supervisor and for supporting me in this quest. Thanks also to Dr. Friederike Flachsbart who has ignited the thrill of the PhD. I will always be extremely grateful to her for providing me with the opportunity to perform my PhD in Germany, surrounded by the beautiful and windy atmosphere of Kiel.

My highest appreciation to my advisor Prof. Almut Nebel for holding my feet to the fire and for forcing me to make my writing as good as possible. She provided me with all the favorable conditions that made this work feasible. I will always cherish the confidence that she entrusted to me as well as the advices that she kindly shared with me.

I extend my gratitude to the Cluster of Excellence 'Inflammation at Interfaces' for financing my position and this work, likewise to all members of the PopGen biobank for the recruitment of probands and sample collection.

I would like to express my utmost gratitude to Prof. Dr. Thomas Bosh who has always kindly and constructively supported me and my research initiatives. Also, thanks to all the members of the Bosch Lab, in particular to Javier Lopez who I have known since I was an undergraduate and with whom I embarked in the great adventure of Chapter I. My most affable thoughts towards him.

I want to express sincere gratitude to Prof. Andre Franke who spent part of his time helping me

overcome several drawbacks of this work, also for inviting me to the soccer game on Tuesdays that I enjoy so much.

I had the pleasure to meet Prof. Dr. Christoph Kaleta and his research group. Many thanks to them for their constructive insights and valuable feedbacks.

I would like to extend my acknowledgements to Prof. Javier Buldu and Dr. Johann Martinez, members of the Laboratory of Biological Networks (in Spain) for their guidance in the fantastic world of the complex networks and for the memorable time in Spain.

I wish to thank all the members of the Human Longevity group, the Ancient DNA group, and Genetics & Bioinformatics group from the Institute of Clinical Molecular Biology (IKMB). In these groups, I have been privileged to work with an as talented, insightful and generous group of colleagues as I can imagine. Many of these colleagues played a role in the development of these research projects which enriched my work and my life. With warm thanks to Julian Susat, Lena Möbus and Alex Flache for their friendship and for showing me the most amiable face of Germany.

Endless thanks to Janina Dose. She stood by me and shared with me both the good and the awkward moments. Thank you also for all the passionate scientific discussions that we had, for enduring my wordiness and for editing my sleep-deprived writing into a form that has been at least minimally acceptable. Thank you so much for your love and care.

It is a pleasure also to thank my friends. Some from a distance accompanied me on this long journey and always put a smile in my face. Many thanks to Camilo, Esteban, Teresa, Alejandro, Bibiana and Milena (Mike).