HHS Public Access

Author manuscript

*Transl Med Aging*. Author manuscript; available in PMC 2020 November 24.

Published in final edited form as:

*Transl Med Aging*. 2019 ; 3: 132–143. doi:10.1016/j.tma.2019.11.004.

# Computational tools for geroscience

**Joseph C.P. Kruempel**[a,1], **Marshall B. Howington**[b,1], **Scott F. Leiser**[a,c,*]

[a]Molecular & Integrative Physiology Department, University of Michigan, Ann Arbor, MI, 48109, USA

[b]Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI, 48109, USA

[c]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA

## Abstract

The rapid progress of the past three decades has led the geroscience field near a point where human interventions in aging are plausible. Advances across scientific areas, such as high throughput "-omics" approaches, have led to an exponentially increasing quantity of data available for biogerontologists. To best translate the lifespan and healthspan extending interventions discovered by basic scientists into preventative medicine, it is imperative that the current data are comprehensively utilized to generate testable hypotheses about translational interventions. Building a translational pipeline for geroscience will require both systematic efforts to identify interventions that extend healthspan across taxa and diagnostics that can identify patients who may benefit from interventions prior to the onset of an age-related morbidity. Databases and computational tools that organize and analyze both the wealth of information available on basic biogerontology research and clinical data on aging populations will be critical in developing such a pipeline. Here, we review the current landscape of databases and computational resources available for translational aging research. We discuss key platforms and tools available for aging research, with a focus on how each tool can be used in concert with hypothesis driven experiments to move closer to human interventions in aging.

### Keywords

Biogerontology; Geroscience; Databases; Bioinformatics; Preventative medicine

## 1. Introduction

Aging is the primary risk factor for mortality and many common pathologies, including cardiovascular disease, stroke, dementia, and many cancers [1,2]. The rapidly growing



*Corresponding author. Molecular & Integrative Physiology Department, University of Michigan, Ann Arbor, MI, 48109, USA. leiser@umich.edu (S.F. Leiser).

[1]Denotes equal contribution.

Conflict of interest

The authors declare there is no conflict of interest.

Declaration of competing interest

None.

population of older individuals drives increases in the burden of age-related diseases worldwide [2]. Thus, developing interventions that maintain the overall health of aging populations is a critical challenge for healthcare in the 21st century.

Research into the biological mechanisms of aging has advanced rapidly over the last thirty years. These advances gained traction with the seminal discovery that mutations affecting the insulin and insulin-like growth factor signaling pathway allow the nematode C. elegans to live twice as long as wild-type animals [3-6]. The geroscience field has since discovered numerous genetic interventions that extend the lifespan of invertebrate models and a handful that extend lifespan in mice [7,8]. Furthermore, a variety of pharmacological interventions that extend lifespan in invertebrates have been identified, along with several that increase the longevity of healthy mice [9-12]. Ongoing clinical studies are testing a mechanistic target of rapamycin (mTOR) inhibitor to treat the age-related decline in immune function in humans [13]. Together, these data suggest that the tools developed by basic biogerontology research are on the verge of translation to the clinic.

Building a translational pipeline for interventions that target the aging process will require a systematic effort from stakeholders in many disciplines, including 1) basic scientists, who discover genes and interventions that promote longevity and ameliorate age-related pathologies in animal models, 2) bioinformaticians, who identify aging-related molecular changes and biomarkers that can predict age-related pathology, and 3) clinicians, who identify patients that will benefit from preventative interventions and perform clinical trials (Fig. 1).

This multidisciplinary effort will require computational tools to organize and analyze vast amounts data, make them accessible to researchers in different fields, and generate new insights from large datasets. Existing resources that incorporate aging research from basic science to clinical applications can form the foundation for translating biogerontology into clinically useful preventative medicine. Here, we review the available and emerging tools within the geroscience field, many of which may be useful for translational approaches. In Section 2, we review databases that compile information on aging-related interventions, molecular changes, pathologies, and the biological and clinical characteristics of aging populations. In Section 3, we address the application of computational methods in geroscience by reviewing tools for analyzing survival data and discussing machine learning and its applications in discovering new aging genes, interventions, and biomarkers.

## 2. Databases for aging research

A major challenge in translating geroscience is to collect and organize the many published observations of interventions that affect aging in animal models. Research groups have identified numerous genetic alterations and pharmacological agents that can modify lifespan, predominantly using short-lived invertebrate aging models, such as *C. elegans, D. melanogaster*, and *S. cerevisiae*. Transcription profiling technology has identified thousands of genes whose expression levels are altered by aging or by interventions, such as dietary restriction, that extend lifespan. Longitudinal studies in mammalian models, predominantly mice, have characterized genetic contributions to longevity and aging-related pathologies

and have identified diets and drugs that increase both lifespan and healthspan. Biobanks and cohort studies are integrating biological data, such as high-density genotyping, with electronic health records to uncover the origins of age-related disease in the clinic. Building a translational pipeline for biogerontology will likely require extensive use of many or all of these data types. Collecting these data from numerous primary publications and performing secondary analysis is prohibitively time consuming for most research groups. Several online databases have begun aggregating data on aging-related molecular and phenotypic changes, interventions, and clinical outcomes, enabling research groups to better integrate their work with relevant data from basic, pre-clinical, and clinical aging research. We focus our discussion on the most established of these databases, including the Human Aging Genomic Resources and Digital Aging Atlas databases (containing data on aging-related genes, molecular and phenotypic changes, and interventions); the Mouse Phenome Database (containing data on pre-clinical aging studies in mice); and centralized databases, such as the UK Biobank and the NIH GERA/RPGEH cohort that connect genetic and biological data on large real-world human populations to clinical information from health records (Fig. 2).

## 2.1. Human Aging Genomic Resources (HAGR) databases

The HAGR are a collection of databases and search tools that were developed by the research group of Dr. Joao Pedro De Magalhaes and are hosted at https://genomics.senescence.info/. Initially launched in 2005, HAGR comprises six databases and complementary tools containing curated information on aging-related genes, phenotypes, and interventions in both humans and numerous animal models, all of which are available to the public without registration [14-17].

**2.1.1. GenAge—**One of the most broadly applicable resources in HAGR is GenAge. GenAge includes two curated, searchable datasets: “GenAge—human genes” and “GenAge—model organisms” [8]. Since human mutations that directly affect longevity are not known, the 307 genes included in “GenAge—human genes” are primarily chosen based on homology to genes that affect longevity in other animal models. For example, growth hormone receptor (GHR) is included on the basis that GHR mutants affect longevity in mouse models [18]. Three additional genes are included because loss-of-function mutations at these loci cause progeroid conditions. “GenAge—human genes” has been used to conduct a meta-analysis of the systems biology characteristics of aging genes that concluded aging genes are often hubs of gene-gene interaction networks [19].

“GenAge—model organisms” is a dataset that includes hundreds of genes that affect aging in animals. Most of the data come from invertebrate models, but mouse, golden hamster, and zebrafish genes are also represented. These genes are included based on reports of a genetic manipulation (mutation, knockdown, or overexpression) causing a change in lifespan. Each database entry includes published data indicating why the gene was included, the magnitude and direction of the associated lifespan change, and homologous genes in other organisms. This provides a tool for researchers to identify previously published data on a gene of interest. The “GenAge—model systems” database could be used for systematic efforts to repeat lifespan data and to identify genes that cause similar lifespan changes when

manipulated in different animal models, thereby identifying genes that regulate aging through conserved mechanisms from invertebrates to mammals.

GenAge also includes a "Genes Commonly Altered During Aging" dataset. This dataset includes 73 genes with orthologues that are differentially regulated during aging in human, mouse, and rat models. They were identified based on a meta-analysis of available microarray data, with each gene cluster being filed under the name of the human orthologue [20]. While it is somewhat more out of date than other GenAge resources, being based on a 2009 paper, this dataset can be a useful tool for identifying genes and biological processes that are robustly associated with aging or that can predict aging-related outcomes. However, the set of robustly changed genes included will be too small for many types of computational analysis.

**2.1.2. GenDR—**The GenDR database is a collection of genes that are involved in dietary restriction (DR). GenDR includes two searchable datasets: a dataset labeled “Gene Manipulations” cataloging “DR—essential genes” that interact genetically with DR, and a dataset labeled “Gene Expression” cataloging genes that are transcriptionally regulated by DR in mammals [16,21,22]. The "Gene Manipulations" dataset includes 214 genes from *Schizosaccharomyces pombe, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster*, and *Mus musculus* that are essential for lifespan extension by DR [21]. "GeneDR: Gene expression" is a separate dataset of 173 genes that are consistently differentially expressed during DR in multiple mammalian species, based on an analysis of available micro-array data [22].

**2.1.3. DrugAge—**HAGR also includes DrugAge, a curated database of published pharmacological interventions that have been tested for effects on lifespan [23]. This database includes 70 compounds that have been tested in mice, hundreds of compounds that have been tested in conventional invertebrate aging models, and many compounds that have been tested in non-conventional aging models, such as rotifers and crickets. Each compound entry includes the dosage and genetic background, as well as a graphical representation of each reported test of the compound and the percent change in lifespan, regardless of whether lifespan was increased, decreased, or not affected. Like GenAge, this database may ultimately help identify compounds that reproducibly extend lifespan in multiple invertebrate species. It may also identify conserved target proteins as prime candidates for translation into mammalian models of aging and disease. Additionally, meta-analysis of these data can potentially identify shared features or target protein networks associated with longevity enhancing drugs, leading to predictions of new lifespan extending therapies [24].

**2.1.4. The Longevity Map—**The Longevity Map is a curated database of genes, gene variants, and chromosomal locations that have been examined for an association with longevity in genome-wide association studies (GWAS), including negative results [25]. In addition to curating the large volume of data on genes associated with human longevity, the Longevity Map can be used in combination with other databases to identify genes that influence longevity, are differentially regulated with aging, or that may be involved in the pathogenesis of age-related diseases [17].

**2.1.5. AnAge—**AnAge is an integrative database of longevity and life history data on over 4000 species [26]. It includes data on maximum species longevity as well as other mortality parameters and life history traits, such as body size, metabolic rate, and development schedules. It also includes a list of species with negligible senescence. An analysis of the life history data available in AnAge was able to conclude that metabolic rate does not correlate with longevity in eutherians or birds, after correction for body size and phylogeny, but that there is a positive correlation between longer developmental time and longer adult lifespan after correction for body size [27]. AnAge is now the most used resource in HAGR, based on the HAGR team's publicly available tracking of site usage and citations, and it enables comparative biology studies that use genome sequence data or primary cell lines to explore the vast differences in longevity observed across taxa [17,28].

**2.1.6. CellAge—**Senescence, the irreversible halting of cell division, may be a contributor to human aging and disease [29,30]. CellAge is a curated database of genes that influence senescence in human cells, based on gene manipulation experiments reported in the literature [17]. In addition to improving understanding of cellular senescence, CellAge might be a useful tool for the development of senolytic drugs that aim to promote clearance of senescent cells to prolong lifespan or prevent age-associated pathologies.

**2.1.7. HAGR analysis tools and applications—**The HAGR databases are a valuable tool to research a gene or phenotype of interest. In addition, these databases can enable meta-analysis that generate new testable hypotheses. HAGR includes several tools that enable analysis of the data from multiple databases. Published analyses have begun using the data available in HAGR to identify new longevity regulating genes and drugs.

Of particular interest to translational geroscience, HAGR includes the Aging-Related Disease Genes (ARDs) tool [17]. The tool includes a database of genes that are associated with 20 different aging-related diseases based on an analysis of available GWAS data. While these genes can be downloaded in isolation, the tool also enables ARDs to be searched and cross-referenced with other databases in HAGR, including GenAge and the Longevity Map. This enables straightforward identification of genes that, for example, are likely to both influence development of aging-related diseases and that are differentially expressed during aging, allowing the formation of new hypotheses about how age-related changes lead to dysfunction.

For users with some coding experience, HAGR offers a Perl toolkit called Ageing Research Computational Tools (ARCT) [14,15,17]). ARCT consists of eight modules that allow for data-mining of the HAGR databases, interaction with public databases, such as GenBank, RefSeq, PubMed, etc., and prediction of protein-protein interactions or phylogenetic relationships similar to those found in HAGR databases.

Analyses of the data in HAGR have been used to identify new genes and drugs that regulate longevity. One study used the "DR—essential genes dataset" and publicly available gene-gene interaction databases to identify genes that were predicted to be involved in DR on the basis of interactions with known DR—essential genes [21]. Testing a subset of these

candidates identified several new genes that are required for lifespan extension by DR in *S. cerevisiae* [21].

Another analysis used DrugAge data from *C. elegans* and machine learning to create a model that predicts novel longevity extending compounds based on the chemical and protein interaction characteristics of known longevity extending compounds [24]. They applied this model to over 6000 compounds from the Drug Gene Interaction Database and identified 20 compounds with a high probability of expending lifespan [24]. While many of the identified compounds, such as temsirolimus, a pro-drug of rapamycin, and valspodar, an experimental chemosensitizer drug, represent exciting potential lifespan extending drugs, their results have not yet been confirmed through wet lab experiments.

Together, the resources available in HAGR represent a powerful centralized tool for a spectrum of geroscience applications, ranging from quickly identifying previous work on a gene or compound of interest to performing detailed meta-analyses. Existing studies provide a roadmap for how these resources can be applied to identify new genes and drugs that can regulate the aging process. In the future, larger scale studies could use the data available in HAGR to systematically test genetic and pharmacological manipulations that have been reported to extend lifespan for reproducibility across species, thereby identifying interventions for translation into vertebrate models and the clinic.

### 2.2. The Digital Aging Atlas (DAA)

Another de Magalhães group resource that complements HAGR is the Digital Aging Atlas (http://ageing-map.org/). The Digital Aging Atlas (DAA) is a collection of changes that occur during aging, ranging from gene expression changes to physiological and pathological changes [31]. The available phenotypic data focuses on human aging and is supplemented by gene expression changes collected from mice. The data available in the DAA can be searched by keyword or by choosing a phenotype or affected tissue from a list. Information is arranged hierarchically so that the entry for an age-associated change often links to the entries for related changes. For example, the entry for “wrinkles” links to “decreased amounts of collagen IV and VII, and abnormal elastin in dermis”, while the entry for IGF1 links to “morning IGF-1 decline”. This organization makes the DAA an excellent tool for researchers to understand the importance of an age-related change of interest at different organizational levels, from the molecular to the physiological. The DAA also has built-in analysis tools that allow any numerical datum present in the DAA (mostly gene expression changes) to be compared against any other, facilitating meta-analysis of the data to identify new age-related correlations.

### 2.3. GeneWeaver databases

Numerous researchers have used -omics approaches, such as RNA-seq, micro-array studies and GWAS, to inquire about the genes and gene expression changes responsible for changes in longevity and age-related disease phenotypes. Meta-analysis of these data can be useful in identifying genes/gene networks that regulate aging or age-related phenotypes and biomarkers for age-related diseases. However, analysis of these data from the primary

literature is time consuming and requires a degree of bioinformatic expertise that is beyond the skill-set of many wet-lab biologists and clinicians.

GeneWeaver (https://geneweaver.org/) is an online repository of -omics data and a suite of analysis tools that was created to bring together -omics datasets and facilitate meta-analysis [32-35]. It contains almost 200,000 public gene sets incorporated from databases, including the Kyoto Encyclopedia of Genes and Genomes (KEGG), Molecular Signatures Database (MSigDB), the drug-related gene database of the Neuroscience Information Framework (NIF), and Online Mendelian Inheritance of Man (OMIM), among many others. Types of GeneWeaver Data that are returned when searching for "senescence OR aging OR longevity" include aging-related (1655 datasets), Gene Ontology (GO) term-based (71), Mammalian and Human Phenotype (63 and 109, respectively), Quantitative Trait Loci (409), Medical Subject Headings (231), gene expression (169), and drug-related gene sets (253) [34]. GeneWeaver is freely accessible and allows for the uploading of one's own datasets (with the option to make public or private) to facilitate meta-analysis.

**2.3.1. GeneWeaver analysis tools—**Among the GeneWeaver tools available for use are the Combine, Jaccard Similarity, GeneSet Graph, STRING, and Ingenuity Pathway Analysis (IGA) tools. A more in-depth application of these tools and of the use of GeneWeaver in aging-related studies can be found in a recent publication [34]. Briefly, the Combine tool returns a single gene set from multiple input gene sets that counts the number of sets in which each gene is found. This can be useful for rapidly identifying genes that may be up- or down-regulated in related datasets, such as cellular senescence and functional decline. The Jaccard Similarity tool uses pairwise comparison of two gene sets to generate a Venn diagram that shows the overlaps of genes between the two sets. The degree of overlap between the two sets is summarized in the Jaccard Similarity coefficient, where 1.0 is perfect overlap and 0.0 is no overlap.

The GeneSet Graph tool, similar to the Combine tool, identifies common genes among multiple input gene sets, but instead of returning a single gene set, GeneSet Graph returns a partitioned set of genes and gene sets. For example, the GeneSet Graph tool has been used to identify the three most highly connected genes from 73 aging-related gene sets across six different species [34]. This tool could be used to rapidly identify genes that are common in many different gene sets and establish orthogonal relationships across species that were previously unknown.

The STRING and IGA tools focus more on fitting input data into existing protein interaction models and molecular pathways, respectively. STRING determines if a group of proteins is likely to have a higher degree of interaction than would be expected from a similar size set of proteins sampled randomly from the genome. For example, Bubier, et al. identified 10 genes common to functional decline and senescence and determined that six of those genes interact in the MAP kinase pathway. Lastly, the IGA tool utilizes built-in algorithms to map input genes to existing molecular pathways. This could be useful, for example, in determining if a newly discovered longevity gene fits into pre-existing longevity pathways, such as the DR or insulin/insulin-like growth factor signaling pathways, or if it is sufficient to establish a novel longevity pathway.

One possible application of GeneWeaver will be to identify mechanisms of lifespan extension that are common across species. GeneWeaver has been used to identify cross-species gene homologues that are transcriptionally regulated by DR and by lifespan extending drugs in mouse and *Drosophila* [34]. GeneWeaver has also been used to identify a novel *C. elegans* lifespan regulator, *Cd63*, on the basis that it is present in a large number of aging-associated gene sets across multiple species [34]. In the future, GeneWeaver might be useful to identify genes whose transcriptional upregulation serves as a biomarker, to identify drugs that extend lifespan across species, and to identify the conserved mechanisms by which interventions extend lifespan across taxa.

Together, the database and tools afforded by GeneWeaver currently represent a user-friendly and pragmatic union to help facilitate meta-analysis. Several other publications exist that utilize GeneWeaver in ways less focused on aging, but also serve to illustrate the usefulness of the platform [32,33,35].

## 2.4. Model organism databases

**2.4.1. Invertebrate model databases—**Both HAGR and GeneWeaver provide sources of information and meta-analysis tools that enable new hypotheses to be formed from existing data. These analyses can be especially useful for researchers using invertebrate model systems, who can, for example, perform mechanistic studies on orthologues of genes that are implicated in mammalian aging or identify pharmacological interventions from DrugAge that have conserved effects on longevity in multiple species. In addition to HAGR and GeneWeaver, organism specific databases, such as the Saccharomyces Genome Database (https://www.yeastgenome.org/), WormBase (https://wormbase.org/#012-34-5), FlyMine (https://www.flymine.org/), and FlyBase (https://flybase.org/) can facilitate design of these studies [36-39]. These databases focus on species or taxa of interest (e.g. WormBase includes information for multiple nematodes species, while FlyMine includes data for multiple insect species) and include numerous types of data, such as gene and protein sequences, genome annotations from the ModEncode project, expression patterns, transcriptome data, proteome data, summaries of publications, and lists of investigators. The data can be accessed in multiple ways including simple searches for a gene or phenotype of interest, as well as more complex queries [40,41].

One critical aspect of study design in invertebrate model systems is pairing invertebrate genes with mammalian genes that are likely functional orthologues. Numerous existing tools are able to identify possible orthologues of a gene of interest based on conservation in nucleotide or protein sequence; however, it can be challenging to determine which genes from a lengthy list of homologous sequences may have a conserved function in humans and a distantly related invertebrate species. WORMHOLE is a recently developed webtool, hosted at http://wormhole.jax.org/, that aims to overcome this challenge by improving identification of gene orthologues with the least sequence divergence from a gene of interest, called least-diverged-orthologues or LDO's. WORMHOLE uses 17 independently developed orthologue identification algorithms to identify possible LDOs, and then applies a second algorithm, developed using known LDO's and machine learning, to identify LDO's based on patterns in the initial set of predictions [42]. Despite its name, WORMHOLE

currently accepts *Saccharomyces cerevisiae, Drosophila Melanogaster, Danio Rerio*, and *Mus Musculus* genes, in addition to *C. elegans* and human genes, as either inputs or outputs. Together, these tools facilitate the design of studies to identify genes and drugs that regulate longevity and healthspan across taxa. Interventions that extend lifespan in multiple invertebrate species may represent promising targets for translation into mammalian pre-clinical models.

**2.4.2. The Mouse Phenome Database—**Efforts to develop interventions that extend longevity and healthspan into mouse models present challenges that are distinct from simpler model systems. These challenges include identifying relevant genetic models to study longevity or particular age-related diseases, as well as deciding on the most important health parameters to measure and developing a detailed protocol, well in advance of starting a lengthy and expensive long-term study with a limited number of animals.

The Mouse Phenome Database (MPD) was developed by the Jackson laboratory to help overcome these challenges and facilitate meta-analysis by collecting, integrating, and providing tools to analyze mouse studies. The MPD includes extensive information on experimental protocols, genetic sequences, and phenotype data, including studies of longevity and other aging relevant endpoints [43-45]. MPD accepts data from any verifiable mouse strain or population, such as inbred, mutant, and transgenic strains and UMHET-3 mice used in the National Institute on Aging Intervention Testing Program (ITP). The MPD currently houses data on over 1700 mouse strains [45].

Data available in the MPD include individual animal phenotypes of specific mouse strains as well as animals subjected to various interventions. The data are searchable by keyword. For example, searching the keyword “aging” as of this writing returns 53 “phenotype strain survey measures”, 11 “QTL phenotype measures”, 36 “phenotyping protocols”, and 2 “collaborating centers”. Phenotyping projects are displayed in an easily navigable table with information, including the investigator, a detailed study protocol, associated references, a graphical display of data, statistics, and individual animal values.

MPD users can create an account (called MyMPD) to which they save project data to perform reproducible secondary analysis [45]. MPD has several tools available for secondary analysis. These include a correlation finder that enables users to easily compare values from different studies with the same strain and to identify correlations between phenotypes, such as strain lifespan. This allows users to form new hypotheses about strain traits that might underlie differences in survival. For example, MyMPD can be used to identify a (slight negative) correlation between BMI and lifespan in inbred strains for which both parameters have been measured in independent studies.

MyMPD also features a strain selection tool that enables users to rank the mouse strains used in a set of studies by their adherence to user-defined criteria, such as lifespan, body weight, food intake, etc. A large portion of the aging data available for secondary analysis in MyMPD originate from a series of phenotyping studies on inbred mouse strains conducted by the Nathan Shock Center at the Jackson Laboratories [46-48]. A published analysis used the MPD strain selection tool to identify a series of strains considered good candidates for

studies of compounds that might extend lifespan or improve age-related health parameters [45]. Their criteria included “high alanine aminotransferase (ALT), low thyroxine (T4), low B cell percentage, high neutrophil percentage, high albumin: creatinine ratio, low gait score, low BMI, and short life span” [45].

The MPD data and the strain selection tool could be highly useful for pre-clinical evaluation of candidate lifespan and healthspan extending molecules, including compounds discovered in high-throughput studies on invertebrate models. The molecular targets of a compound of interest can be used, along with other tools, such as HAGR and GeneWeaver, to identify healthspan and disease phenotypes that a drug may alter. The MPD database can then be used to identify appropriate mouse models to test the hypothesis that a drug will affect survival or another pathologically relevant endpoint in mammals. Since MPD contains information about the point in mouse lifespan when age-related changes are observed, it may also be used to identify relevant timepoints for intervention studies. This approach could allow the study of healthspan phenotypes in pre-existing aged mouse populations without the time and expense of performing full survival studies in mice.

One barrier to the MPD reaching its full potential as a tool for pre-clinical translational aging studies is the relative paucity of data available in the database on aging mice. Currently, most of the longitudinal data available in the database comes from a handful of studies, including those performed at the Jackson Labs and by the Intervention Testing Program. There is a need for investigators with existing longitudinal data in mice to submit to the MPD (or a similar platform) to facilitate meta-analysis.

### 2.5. Clinical databases on aging populations

Translating interventions that affect the aging process into the clinic will likely require a detailed understanding of how aging contributes to disease progression in human patients and the ability to identify patients who will benefit from preventative medicine. Many cohort studies have been performed to examine clinical characteristics of aging populations and have been reviewed elsewhere [49-55]. A recent trend is the establishment of large-scale studies that pair detailed information on patient health with extensive biological resources, such as high-density genotyping and biobanked samples. Among many other applications, these studies will enable collection of biological data on human aging that can inform clinical practice. Several studies that exemplify trends are included in Table 1 [56-63]. A few resources that provide exceptionally centralized and detailed information connecting biology and health outcomes in aging cohorts are described in more detail below.

**2.5.1. GERA/RPGEH—**The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort was assembled as a partnership between the US National Institute on Aging and the Kaiser Permanente health insurance network. A detailed description of the study cohort and the procedure for requesting access to the data are hosted at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v3.p3. The study enabled genome-wide single nucleotide polymorphism (SNP) genotyping of over 100,000 individuals who are members of the Kaiser Permanente Medical Care Plan, Northern California Region (KPNC) and participants in its Research Program on Genes Environment

and Health (RPGEH) [64]. Participants ranged in age from 18 to over 100, with an average age of 63 at the time of the RPGEH survey (2007) [64]. These participants provided saliva samples, filled out a detailed questionnaire, and provided broad consent for the use of their data in biomedical research. Data on the diagnosed health conditions of individual participants are available from electronic medical records. Investigators who wish to access the data can apply to RPGEH Access Review Committee through a dedicated web portal. SNP profiling data, along with survey and health record data for 78,479 patients who gave additional consent, are deposited in the NIH database of genotypes and phenotypes (dpGaP). Institutional investigators can apply for access to these data by submitting a Data Access Request to NIH.

The data collected through the GERA project is being used to better understand the biological basis of human aging and to provide insights into disease. For example, salivary telomere length was measured for over 100,000 patients in the RPGEH cohort [65]. A published abstract from the RPGEH research group reports a prospective association between very short telomere length and increased likelihood of mortality, although the quantitative data from this study are not available for analysis [66]. GWAS using the GERA cohort have identified loci associated with conditions, including blood pressure, drug response, schizophrenia, intraocular pressure, type 2 diabetes, asthma, and hernia [67-73].

**2.5.2. The UK biobank—**Biobanks have been established in many countries, including Iceland, UK, Sweden, Denmark, Latvia, Estonia, Canada, South Korea, Japan, Singapore, China, Taiwan, and in the US at the Mayo Clinic [51]. These repositories are aimed at collecting biological and clinical data on representative populations, with the goal of identifying biological and epidemiological factors that contribute to health outcomes in aging populations [51]. The UK biobank is unique among these because of its 1) high number of patients recruited, 2) variety of biological, genetic, and phenotypic data available on participants, 3) comprehensiveness of available clinical data on health outcomes (obtained through patient medical records filed with its national health service), and 4) open access policy that allows data to be shared widely with researchers inside and outside of the UK [74].

The UK biobank is a prospective cohort study including nearly 500,000 participants from the United Kingdom. The participants, aged 40 to 69 when they were recruited between 2006 and 2010, answered questions regarding their demography, lifestyle, and health-related factors. They also completed a range of physical measures, provided blood, saliva, and urine samples, and provided informed consent for follow-up through linkage to their electronic health records [74]. A large subset of recruitment visits also included eye and hearing assessments, as well as electrocardiograph and arterial stiffness tests. The majority of participants underwent high-density SNP genotyping and health-related biomarker data, including lipids for vascular disease, sex hormone levels for cancer, Hb1AC for diabetes, rheumatoid factor for arthritis, and markers for liver and kidney function. Available tissue samples permit multiple additional assessments, such as metabolomics and proteomics, to be performed for each participant. Additional data, including repeat assessments, objective measurements of physical activity (using a tri-axial accelerometer) and a multi-modal imaging study (including MRI, x-ray and ultrasound) are being collected for subsets of

participants. The UK Biobank data are “available to all bona fide researchers for all types of health-related research that is in the public interest” and can be accessed by application to its resource access management system.

Data in the UK biobank have been used to identify health parameters that best predict all-cause mortality, for GWAS studies examining human longevity and for studies on diverse health topics, including obesity and diabetes risk, frailty and the risk of multi-morbidity and mortality, cognitive function, depression, bipolar disorder, and chronotype [75,77-82]. This wide range of studies demonstrates the utility of this dataset for identifying genetic predictors of age-associated disease among other potential uses. To maximize the global impact of these studies, it will be crucial to determine whether important discoveries made using the UK biobank data can be re-capitulated in other populations by continuing to develop biobank resources around the world.

#### 2.5.3. Comprehensive Assessment of the long-term Effects of Reducing Intake of Energy (CALERIE)—

CALERIE (Comprehensive Assessment of the Long-term Effects of Reducing Intake of Energy) was a two-year study conducted at Duke University, in which 218 participants, ranging in age from 21 to 51, were randomized onto control or 25% calorie restricted diets [83]. CALERIE represents the first controlled human trial of an intervention that extends lifespan and delays or prevents multiple age-associated pathologies in animal models. Participants in CALERIE answered demographic and health-related questions, completed adverse event diaries while on the protocol, and received nutritional and psychological support in transitioning to a calorie restricted diet. Participants underwent a physical and were phenotyped for a variety of biomarkers, including lipids, markers of immune function, inflammation, antibody response to vaccines, growth factors, and serum insulin. Subsets of participants also donated blood, urine, or muscle biopsy samples, with a small group providing all three. The data are available on an open access basis with an application to the CALERIE External Science Committee. Access to biological samples can also be obtained through an application to the Science Committee. Detailed study protocols and online application for access to data or samples are hosted at https://calerie.duke.edu/.

Samples and data from the CALERIE study have been used for secondary analysis, including studies of parameters that are difficult to assess in animal models, such as the effect of calorie restriction on subjective psychological well-being [84]. In the future, these data could be used to identify biological changes, such as differences in the proteome or metabolome, that are associated with calorie restriction and that might serve as biomarkers for evaluating pharmacological interventions that mimic some aspects of calorie restriction.

### 2.6. Summary of aging research databases

Databases that collect information on age-related biological changes and interventions that affect them will be valuable tools for discovering new interventions that increase longevity and healthspan. Collections of transcriptomic and other -omics data on long-lived species and strains can be used to identify conserved molecular changes that are associated with long life and can, in turn, serve as biomarkers to identify new drugs that alter the aging process. Collections of data on genetic changes and drugs that promote increased longevity

can be used to enable systematic efforts to identify the genetic pathways and interventions that improve health during aging across taxa. Databases of preclinical studies in aged mammalian models will help to identify clinically relevant phenotypes that are altered by an aging-modifying intervention.

Databases that collect extensive biological and clinical data on aging populations will be essential resources for translating interventions that alter some aspects of biological aging to the clinic. For a pharmacological intervention that extends longevity to be useful for humans, it will be necessary to perform clinical trials demonstrating that it prevents or slows progression of an age-associated disease in humans. Trials to assess impacts of a candidate drug on multiple morbidities in large populations will be difficult and expensive. Thus, it would be ideal to identify small populations of patients who are at high risk of progressing to an age-related disease state, and then perform trials of a compound that slows aging to determine whether it prevents disease progression. Centralized collections of biological data on aging populations may be used to develop biomarkers that will identify patients who may benefit from trials of aging modifying drugs.

Developing biomarkers that accurately predict development of age-associated diseases will likely require use of multiple types of complex data, such as high-density genotyping, sequencing, DNA methylomics, proteomics, and metabolomics. Likewise, systematic efforts to identify conserved mechanisms of lifespan extension and interventions that modify them will require analysis of complex data, including transcriptomics and gene interaction networks. Thus, Section 3 will discuss emerging developments in computational tools for analyzing aging-related datasets.

## 3. Computational tools for aging research

Many aging studies initially focus on identifying genes or interventions that alter survival. A variety of different tools and statistical approaches can be used to assess differences in survival, taking into account complicated interpretations due to the varied demographics of aging populations. Furthermore, new computational methods, such as machine learning, are emerging as critical tools for identifying new aging-related genes and interventions, as well as creating biomarkers of “biological age” that may prove useful for predicting onset of age-associated diseases and developing preventative interventions. In Section 3, we discuss available tools for analyzing survival data and introduce machine learning and its applications in aging research.

### 3.1. Tools for analyzing survival data

Identifying interventions that affect longevity is a major component of geroscience research. Robust statistical analysis of survival data is crucial for accurate interpretation and reporting of data. There are various methods for conducting, analyzing, and reporting survival analysis, creating challenges for interpreting reported data. Here we will discuss three widely used tools that offer standardized analysis of lifespan data.

**3.1.1. Prism—**Prism is a proprietary statistical analysis and graphing program developed by the company GraphPad that enables generation of publication quality visually appealing

plots from a variety of types of data. It is available for purchase at https://www.graphpad.com/scientific-software/prism/. Of interest to aging research, Prism can plot survival data as a Kaplan-Meier curve in a variety of visual formats. Prism can also test the hypothesis that two survival curves differ using the standard log-rank test (which calculates a test statistic based on the cumulative difference between two survival curves vs. the null hypothesis, that there is no difference in the probability of death at any point, while making no assumptions about the survival distribution) or the Wilcoxon-Gehan test (functioning similarly to the log-rank test but giving higher weight to deaths at earlier timepoints) [85]. Prism is widely used in *C. elegans, Drosophila*, and mouse aging and aging-related diseases research [86-88].

**3.1.2. Online application for survival analysis (OASIS)—**OASIS is an online tool for statistical analysis of lifespan data, developed by the research group of Sanguk Kim and hosted at https://sbi.postech.ac.kr/oasis2/. It is free to use and runs in a web browser. Survival data can be entered by copy-pasting from a simply formatted excel file into a text box and selecting "analysis options" in a graphical user interface, making it straightforward to adopt regardless of programming or statistical expertise [89,90]. In OASIS, survival data can be plotted as survival curves or cumulative log-hazard plots and can be subjected to a variety of statistical tests. In addition to the standard log-rank test, these include tests (e.g. Kolmogorov-Smirnov, Neyman's Smooth, and Chow tests) that examine differences in the shape of survival curves, tests (e.g. Boschloo's test and a modified Mann-Whitney test) that examine differences in maximum lifespan, and tests (e.g. the survival time *F*-test and the partial slopes rank-sum test) that examine differences in the variance of survival time between populations [89,90]. In addition to testing differences between survival curves, OASIS provides Cox-proportional hazards regression that are suitable for analyzing the effects of risk factors (i.e. sex, obesity, calorie intake) on survival. OASIS also contains tools for plotting and comparing values in multiple groups factored by condition and time (ANOVA). This is useful for experiments observing the interaction between treatment, time, and physiological parameters other than survival, such as motility or other markers of healthspan. Overall, OASIS provides a tool offering a good cross-section between ease-of-use and availability of advanced statistics, albeit with less ability to create high-quality graphs available in other applications. OASIS is now widely used, with over 200 Google Scholar citations at the time of this writing, including several high-impact papers [91-93].

**3.1.3. R—**Another useful set of tools for analyzing survival data are the survival, flexsurv, and survminer packages in R [94-96]. R is a coding language used by the open source data visualization software RStudio (https://www.r-project.org/). While it does require a measure of familiarity with coding and converting survival datasets to an R-specific data frame format, the packages in R allow for rapid generation of Kaplan-Meier survival curves as well as statistical analysis using the log-rank test. Similar to analysis using OASIS, the packages in R can also use Cox-proportional hazard regression to determine the effects of risk factors on survival [95]. Importantly, the Cox-regression feature in R can be used to determine the significance of statistical interactions among more than two treatments. For example, this can be used to infer an epistatic interaction between a long-lived mutant and a suppressor mutation that reduces the degree of lifespan extension relative to what is observed in WT. R

is primarily used for data analysis and visualization; therefore, many of these packages have functions that can determine and analyze more advanced parameters of survival data, including generating and fitting standard parametric survival models and merging and manipulating datasets (Jackson, 2016). Importantly, R has a wealth of resources and tutorials available online to assist with utilizing its platform for survival analysis [97,98]. The R survival package has been used extensively in *C. elegans, Drosophila*, and mouse aging and aging-related diseases research [99-101].

**3.1.4. summary of survival tools—**Tools for standardized analysis of survival data, such as those available in Prism, OASIS, and R, can facilitate accurate assessment of survival data in the literature. Additional tools might further facilitate robust standardized analysis and meta-analysis of survival data. A centralized database, including primary survival data from multiple labs, along with meta-data (such as when the experiments were conducted, visible phenotypes, and annotation of why animals were censored), and built-in statistical tools, might facilitate both standardization of lifespan protocols and meta-analysis of many lifespan experiments. Furthermore, easy-to-use online applications that enable testing for significant interactions between more than two survival curves could facilitate more robust interpretation of studies in the genetics of aging.

## 3.2. Machine learning in aging research

New computational tools are necessary to best make use of the high volume of data collected in both model organism and clinical aging studies. In particular, computational tools will be useful for finding patterns in basic research on aging-related genes and interventions that will enable prediction of new genes, gene networks, and interventions that regulate the aging process. Likewise, in the clinical realm, computational tools will be crucial to identify patterns in large datasets that can predict the development of age-related diseases and identify patients who may benefit from preventative interventions translated from basic and pre-clinical geroscience. Ongoing developments in machine learning will be useful for both of these applications.

**3.2.1. What is machine learning?—**Over the past decade, machine learning has emerged as a powerful tool to analyze large datasets. Although use of machine learning algorithms by non-experts is still some time away, it is useful to be aware of the translational applications in literature. A comprehensive review of the concepts and applications of supervised machine learning in aging research was recently published [102]. In brief, machine learning is the automated process of establishing relationships and correlations between variables of input data to generate a model that can be used to evaluate the variables of unknown data. There are two main types of machine learning algorithms: unsupervised and supervised. Unsupervised machine learning relies on the algorithm to establish *de novo* relationships between the variables of the dataset with no further input from the user. Supervised machine learning, however, relies on the user to provide annotations with the data through which the algorithm can base its model. While there are a tremendous number of applications for both unsupervised and supervised machine learning, the latter is most commonly used in aging research. In supervised machine learning, the algorithm used for training the model (e.g. k-nearest neighbor, random forest, support vector machine, etc.) is

supplied with a training dataset. From this dataset, the algorithm establishes correlations among the user-provided annotations. The main types of supervised machine learning are classification (predicting discrete variables) and regression (predicting continuous variables) problems. Examples of classification applications include classifying genes as pro- and anti-longevity; while examples of regression problems include establishing biological age based on methylation sites. Additional applications of machine learning in aging research are covered in the following sections.

**3.2.2. Applications for machine learning in predicting aging related genes and drugs—**Machine learning has proven to be a useful tool when applied to aging research paradigms. Examples of approaches include 1) classifying DNA repair genes into aging- or non-aging-related, 2) classifying genes in *C. elegans* and other model organisms as pro- or anti-longevity 3) classifying aging-related mouse proteins, 4) predicting aging genes in humans and *C. elegans*, 5) identifying biomarkers of aging in humans, and 6) establishing aging- and mortality-related gene expression profiles in humans [103-113]. These applications use a wide-array of input data types to generate models, including protein-protein interactions, gene expression levels and profiles, GO terms, KEGG pathway features, and DNA methylation profiles. While these publications have revealed interesting insights into aging research, there are several challenges to maximizing the utility of machine learning for translational geroscience: 1) machine learning algorithms are complex and not yet as user-friendly as other analysis tools; 2) machine learning models skew toward previous research, limiting the ability of machine learning to find novel aging genes/drugs [102]; and 3) much of the machine learning literature has not been confirmed with wet lab experiments. The utility of machine learning might be improved by encouraging collaborations between computational and experimental research groups to validate results generated by machine learning and further refine models with new data.

**3.2.3. Applications of machine learning in estimating biological age and predicting health outcomes—**One application of machine learning relevant to translational medicine of aging is the development of tools designed to determine a patient's health status and to predict their probability of developing age-related diseases based on biological inputs. Toward this end, multiple research groups have developed "epigenetic clocks" that aim to predict a patient's age based on age-related changes in genome-wide DNA methylation [105,114,115].

The published epigenetic clocks use genome-wide DNA methylation arrays to determine the methylation status of millions of CpG sites throughout the human genome [116]. They then apply a model, constructed using supervised machine learning, to predict the age of the DNA source based on the methylation status of individual CpG sites. While the methylation status of individual CpG sites included in the age estimator is often only weakly correlated with age, age estimators using hundreds of CpG sites can predict chronological age with relatively high accuracy. Different groups have developed epigenetic clocks using distinct tissues as DNA sources. Some epigenetic clocks use a single tissue, such as saliva or blood; whereas, other clocks incorporate data from multiple tissue types and are designed to provide accurate estimates regardless of the DNA source used [105,114-116]. A widely

studied multi-tissue epigenetic age estimator, developed by the Horvath group, is available on an open access basis to anyone with methylation data from a supported CpG array (https://dnamage.genetics.ucla.edu/home).

Existing epigenetic age estimators have shown promise as predictors of health status. Patients whose age estimate from an epigenetic clock is higher than their chronological age are also at higher risk for age-related adverse health outcomes, including neuropathology, Parkinson's disease, reduced physical and cognitive fitness, and certain types of cancer, as well as all-cause mortality, while low epigenetic age relative to chronological age is associated with centenarian status [116-123]. Levine et al. recently reported development of a tool called "DNAm PhenoAge" that predicts a weighted average of clinical characteristics (chronological age, creatine, glucose, and c-reactive protein levels, etc.) rather than age alone [124]. This clock outperforms prior epigenetic clocks at predicting mortality, cardiovascular disease, and other measures of multi-morbidity [116,124].

**3.2.4. Summary of machine learning in aging research—**In the future, machine learning may be used with training data from biobanks of large populations to create clocks that use DNA methylation and other biological predictors, such as genome-wide SNP genotyping and urine metabolite levels, to generate models that accurately predict a patient's risk of developing age-related morbidities. These tools could be used to initiate clinical trials using candidate age-delaying compounds from invertebrate and mouse preclinical studies to delay or prevent progression of age-related disease. Machine learning-based clocks might also be used to evaluate whether an intervention is actually modifying epigenetic readouts of aging or progression to a disease state. These tools might be useful to more rapidly identify lifespan extending compounds in mouse studies or to provide surrogate endpoints for clinical trials of compounds that aim to delay progression of age-related diseases.

## 4. Conclusion

We have outlined a series of existing and developing computational tools that will aid investigators in developing new hypotheses and interventions for translational geroscience. In brief, databases of genetic and pharmacological interventions that affect aging in invertebrate models will enable organized efforts to reproduce existing data and identify interventions that affect lifespan and healthspan across taxa, while machine learning software can be trained on these datasets to generate new hypotheses about candidate genes and compounds to test for aging phenotypes. Databases of aging-related phenotype data in mice will enable investigators to identify appropriate strains to test potential interventions for effects on progression of age-associated health phenotypes. Clinical databases, coupling individual level data with high-density genotyping, epigenetic, proteomic, and metabolomic studies, will inform computational models that can identify patients who may benefit from preventative interventions translated from rodent aging models.

Significant challenges remain in developing existing databases and computational tools into a platform for a translational biogerontology pipeline. First, meta-analysis of aging-related data will be more feasible if data can be easily made available to the community via centralized platforms. Expecting open access to data has become the norm in academic

biomedical research. For example, investigators funded by the U.S. National Institutes of Health are required to provide a data-sharing plan, which can include uploading data to public databases, when applying for major funding. Likewise, some major journals, such as Aging Cell, require authors to submit data from -omic studies to public databases. However, *centralized* general purpose platforms for archiving aging-related biological data, such as lifespan and healthspan observations in animal models, have not been developed: the closest programs are the National Archive of Computerized Data on Aging (NACDA) (which focuses on social science research), along with the AMP-AD Knowledge Portal, the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS), which focus on Alzheimer's disease, and the model organism-specific databases described above [125-130]. In the age of big data, it would make sense for funding institutions to make a long-term investment in centralized data repositories where aging-related primary data, including lifespan and healthspan data in multiple organisms, could be deposited in an easily adopted format and used for meta-analysis. Data in this repository could be linked to existing and developing databases and tools. Centralized data archives would encourage investigators to make data that would otherwise be archived in individual labs readily available to the community, aiding development of new computational resources and new discoveries from secondary analysis. It would also be useful to develop more centralized resources for organizing the increasing volume of clinical aging data from biobanks and large genotyping studies, though we note that such efforts must make ethical considerations and the informed consent of participants to data sharing their top priority.

Secondly, developing the most useful possible computational tools will require diverse expertise not typically found in single research groups. Collaborations between groups with computational expertise, wet lab scientists, and clinicians will be necessary to develop computational tools that can be applied to real-world problems and then refined using new data. The numerous studies that have associated DNA methylation-based calculations of biological age with clinical outcomes are good examples of how collaborative efforts between computational and clinical researchers can be fruitful.

Translating interventions that modify the aging process in animal models into preventative medicine applications will be a major challenge in biomedicine over the coming decades. A research program that uses database and computational resources to enable systematic efforts in translating anti-aging interventions from invertebrate models to the clinic will require significant investments in developing and applying resources. However, this program may enable us to realize the goal of developing preventative interventions that extend healthy life in humans.

## Acknowledgements

The authors wish to thank Dr. Albert Chen, Dr. Scott Pletcher, and members of the Leiser lab for helpful discussions of the manuscript. JK and SFL are supported by the Glenn Foundation for Medical Research, MH is supported by NIH T-32-GM007315, SFL is also supported by NIH R21AG059117 and NIH R01AG058717.

## References

[1]. Niccoli T, Partridge L, Ageing as a risk factor for disease, Curr. Biol 22 (17) (2012) R741–R752. [PubMed: 22975005]

[2]. Chang AY, et al., Measuring population ageing: an analysis of the global Burden of disease study 2017, The Lancet Public Health 4 (3) (2019) e159–e167. [PubMed: 30851869]

[3]. Kenyon C, et al., A C. Elegans mutant that lives twice as long as wild type, Nature 366 (6454) (1993) 461. [PubMed: 8247153]

[4]. Johnson TE, Increased life-span of age-1 mutants in Caenorhabditis elegans and lower Gompertz rate of aging, Science 249 (4971) (1990) 908–912. [PubMed: 2392681]

[5]. Klass MR, A method for the isolation of longevity mutants in the nematode Caenorhabditis elegans and initial results, Mech. Ageing Dev 22 (3–4) (1983) 279–286. [PubMed: 6632998]

[6]. Kimura KD, et al., daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans, Science 277 (5328) (1997) 942–946. [PubMed: 9252323]

[7]. Kenyon CJ, The genetics of ageing, Nature 464 (7288) (2010) 504. [PubMed: 20336132]

[8]. de Magalhaes JP, Toussaint O, GenAge: a genomic and proteomic network map of human ageing, FEBS Lett. 571 (1–3) (2004) 243–247. [PubMed: 15280050]

[9]. Harrison DE, et al., Acarbose, 17-α-estradiol, and nordihydroguaiaretic acid extend mouse lifespan preferentially in males, Aging Cell 13 (2) (2014) 273–282. [PubMed: 24245565]

[10]. Harrison DE, et al., Rapamycin fed late in life extends lifespan in genetically heterogeneous mice, Nature 460 (7253) (2009) 392. [PubMed: 19587680]

[11]. Miller RA, et al., Rapamycin, but not resveratrol or simvastatin, extends life span of genetically heterogeneous mice, J. Gerontol.: Series A 66 (2) (2011) 191–201.

[12]. Strong R, et al., Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice, Aging Cell 7 (5) (2008) 641–650. [PubMed: 18631321]

[13]. Mannick JB, et al., mTOR inhibition improves immune function in the elderly, Sci. Transl. Med 6 (268) (2014), 268ra179–268ra179.

[14]. De Magalhães JP, Costa J, Toussaint O, HAGR: the human ageing genomic resources. Nucleic acids research 33 (suppl_1) (2005) D537–D543. [PubMed: 15608256]

[15]. De Magalhães JP, et al., The Human Ageing Genomic Resources: online databases and tools for biogerontologists, Aging Cell 8 (1) (2009) 65–72. [PubMed: 18986374]

[16]. Tacutu R, et al., Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing, Nucleic Acids Res. 41 (D1) (2012) D1027–D1033. [PubMed: 23193293]

[17]. Tacutu R, et al., Human ageing genomic resources: new and updated databases, Nucleic Acids Res. 46 (D1) (2017) D1083–D1090.

[18]. Coschigano KT, et al., 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 144 (9) (2003) 3799–3810. [PubMed: 12933651]

[19]. Zhang Q, et al., Systems-level analysis of human aging genes shed new light on mechanisms of aging, Hum. Mol. Genet 25 (14) (2016) 2934–2947. [PubMed: 27179790]

[20]. De Magalhães JP, Curado J, Church GM, Meta-analysis of age-related gene expression profiles identifies common signatures of aging, Bioinformatics 25 (7) (2009) 875–881. [PubMed: 19189975]

[21]. Wuttke D, et al., Dissecting the gene network of dietary restriction to identify evolutionarily conserved pathways and new functional genes, PLoS Genet. 8 (8) (2012) e1002834. [PubMed: 22912585]

[22]. Plank M, et al., A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms, Mol. Biosyst 8 (4) (2012) 1339–1349. [PubMed: 22327899]

[23]. Barardo D, et al., The DrugAge database of aging-related drugs, Aging Cell 16 (3) (2017) 594–597. [PubMed: 28299908]

[24]. Barardo DG, et al., Machine learning for predicting lifespan-extending chemical compounds, Aging (Albany NY) 9 (7) (2017) 1721. [PubMed: 28783712]

[25]. Budovsky A, et al., LongevityMap: a database of human genetic variants associated with longevity, Trends Genet. 29 (10) (2013) 559–560. [PubMed: 23998809]

[26]. De Magalhaes J, Costa J, A database of vertebrate longevity records and their relation to other life-history traits, J. Evol. Biol 22 (8) (2009) 1770–1774. [PubMed: 19522730]

[27]. Magalhães J.P.d., Costa J, Church GM, An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts, J. Gerontol. Ser. A Biol. Med. Sci 62 (2) (2007) 149–160.

[28]. HAGR-papers and relevant publications, Available from: https://genomics.senescence.info/papers.html#high, 2019.

[29]. Campisi J, di Fagagna F.d.A., Cellular senescence: when bad things happen to good cells, Nat. Rev. Mol. Cell Biol 8 (9) (2007) 729. [PubMed: 17667954]

[30]. de Magalhães JP, Passos JF, Stress, cell senescence and organismal ageing, Mech. Ageing Dev 170 (2018) 2–9. [PubMed: 28688962]

[31]. Craig T, et al., The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource, Nucleic Acids Res. 43 (D1) (2014) D873–D878. [PubMed: 25232097]

[32]. Baker E, et al., GeneWeaver: data driven alignment of cross-species genomics in biology and disease, Nucleic Acids Res. 44 (D1) (2015) D555–D559. [PubMed: 26656951]

[33]. Baker EJ, et al., GeneWeaver: a web-based system for integrative functional genomics, Nucleic Acids Res. 40 (D1) (2011) D1067–D1076. [PubMed: 22080549]

[34]. Bubier JA, et al., Integration of heterogeneous functional genomics data in gerontology research to find genes and pathway underlying aging across species, PLoS One 14 (4) (2019) e0214523. [PubMed: 30978202]

[35]. Jay JJ, Chesler EJ, Performing integrative functional genomics analysis in GeneWeaver. org, in: Gene Function Analysis, Springer, 2014, pp. 13–29.

[36]. Lee RYN, et al., WormBase 2017: molting into a new stage, Nucleic Acids Res. 46 (D1) (2017) D869–D874.

[37]. Cherry JM, et al., SGD: Saccharomyces genome database, Nucleic Acids Res. 26 (1) (1998) 73–79. [PubMed: 9399804]

[38]. Lyne R, et al., FlyMine: an integrated database for Drosophila and Anopheles genomics, Genome Biol. 8 (7) (2007) p. R129. [PubMed: 17615057]

[39]. Drysdale R, F. Consortium, FlyBase, in: Drosophila, Springer, 2008, pp. 45–59.

[40]. Gerstein MB, et al., Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project, Science 330 (6012) (2010) 1775–1787. [PubMed: 21177976]

[41]. Roy S, et al., Identification of functional elements and regulatory circuits by Drosophila modENCODE, Science 330 (6012) (2010) 1787–1797. [PubMed: 21177974]

[42]. Sutphin GL, et al., WORMHOLE: novel least diverged ortholog prediction through machine learning. PLoS computational biology 12 (11) (2016) e1005182. [PubMed: 27812085]

[43]. Bogue MA, et al., Mouse phenome database (MPD), Nucleic Acids Res. 35 (suppl_1) (2006) D643–D649. [PubMed: 17151079]

[44]. Bogue MA, et al., Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. Nucleic acids research 46 (D1) (2017) D843–D850.

[45]. Bogue MA, et al., Accessing data resources in the mouse phenome database for genetic analysis of murine life span and health span, J. Gerontol. Ser. A Biol. Med. Sci 71 (2) (2014) 170–177.

[46]. Yuan R, et al., Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels, Aging Cell 8 (3) (2009) 277–287. [PubMed: 19627267]

[47]. Yuan R, et al., Genetic coregulation of age of female sexual maturation and lifespan through circulating IGF1 among inbred mouse strains. Proceedings of the National Academy of Sciences 109 (21) (2012) 8224–8229.

[48]. Leduc MS, et al., Identification of genetic determinants of IGF-1 levels and longevity among mouse inbred strains, Aging Cell 9 (5) (2010) 823–836. [PubMed: 20735370]

[49]. Kingston A, Jagger C, Review of methodologies of cohort studies of older people, Age Ageing 47 (2) (2017) 215–219.

[50]. Erten-Lyons D, et al., Review of selected databases of longitudinal aging studies, Alzheimer's Dementia 8 (6) (2012) 584–589.

[51]. De Souza YG, Greenspan JS, Biobanking past, present and future: responsibilities and benefits. AIDS (London, England) 27 (3) (2013) 303.

[52]. Visscher PM, et al., 10 years of GWAS discovery: biology, function, and translation, Am. J. Hum. Genet 101 (1) (2017) 5–22. [PubMed: 28686856]

[53]. Denny JC, Bastarache L, Roden DM, Phenome-wide association studies as a tool to advance precision medicine, Annu. Rev. Genom. Hum. Genet 17 (2016) 353–373.

[54]. Ribeiro CE, et al., A revision and analysis of the comprehensiveness of the main longitudinal studies of human aging for data mining research, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov 7 (3) (2017) e1202.

[55]. Kaiser A, A review of longitudinal datasets on ageing, J. Popul. Ageing 6 (1–2) (2013) 5–27.

[56]. Avlund K, et al., Copenhagen Aging and Midlife Biobank (CAMB) an Introduction, Sage Publications Sage CA, Los Angeles, CA, 2014.

[57]. Lund R, et al., Cohort Profile: the Copenhagen aging and midlife biobank (CAMB), Int. J. Epidemiol 45 (4) (2015) 1044–1053. [PubMed: 26210613]

[58]. Fan C-T, Lin J-C, Lee C-H, Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island, Pharmacogenomics 9 (2) (2008) 235–246. [PubMed: 18370851]

[59]. Kowal P, et al., Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE), Int. J. Epidemiol 41 (6) (2012) 1639–1649. [PubMed: 23283715]

[60]. He W, Muenchrath MN, Kowal PR, Shades of Gray: a Cross-Country Study of Health and Well-Being of the Older Populations in SAGE Countries, 2007-2010, US Department of Commerce, Economics and Statistics Administration, US, 2012.

[61]. Olson JE, et al., The Mayo Clinic Biobank: a building block for individualized medicine, in: Mayo Clinic Proceedings, Elsevier, 2013.

[62]. Correa-De-Araujo R, A wealth of shared data, specimens for aging research 1-30-2019 8-1-2019], Available from: https://www.nia.nih.gov/research/blog/2019/01/wealth-shared-data-specimens-aging-research.

[63]. Biobank Sverige, Sweden Welcome to Biobank Sweden, 7-04-2019 6-10-2019]; Available from: https://biobanksverige.se/english/research/, 2019.

[64]. Kvale MN, et al., Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort, Genetics 200 (4) (2015) 1051–1060. [PubMed: 26092718]

[65]. Lapham K, et al., Automated assay of telomere length measurement and informatics for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort, Genetics 200 (4) (2015) 1061–1072. [PubMed: 26092717]

[66]. Schaefer C, et al., B4-3: demographic and behavioral influences on telomere length and relationship with all-cause mortality: early results from the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), Clin. Med. Res 11 (3) (2013), 146–146.

[67]. Dahlin A, et al., Large-scale genome-wide association study of asthma in the Kaiser Permanente Northern California's genetic epidemiology research on adult health and aging (KPNC-GERA) cohort, in B21 Genetics, Genomics, And Gene Expression In Asthma And Copd B21, American Thoracic Society, 2016 A7938–A7938.

[68]. Hoffmann TJ, et al., Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation, Nat. Genet 49 (1) (2017) 54. [PubMed: 27841878]

[69]. Wen C, et al., Genome-wide association study identifies ABCG2 (BCRP) as an allopurinol transporter and a determinant of drug response, Clin. Pharmacol. Therapeut 97 (5) (2015) 518–525.

[70]. Loh P-R, et al., Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis, Nat. Genet 47 (12) (2015) 1385. [PubMed: 26523775]

[71]. Choquet H, et al., A large multi-ethnic genome-wide association study identifies novel genetic loci for intraocular pressure, Nat. Commun 8 (1) (2017) 2108. [PubMed: 29235454]

[72]. Cook JP, Morris AP, Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility, Eur. J. Hum. Genet 24 (8) (2016) 1175. [PubMed: 27189021]

[73]. Jorgenson E, et al., A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia, Nat. Commun 6 (2015) 10130. [PubMed: 26686553]

[74]. Bycroft C, et al., The UK Biobank resource with deep phenotyping and genomic data, Nature 562 (7726) (2018) 203. [PubMed: 30305743]

[75]. Ganna A, Ingelsson E, 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study, The Lancet 386 (9993) (2015) 533–540.

[77]. Ntuk UE, et al., Ethnic-specific obesity cutoffs for diabetes risk: cross-sectional study of 490,288 UK biobank participants, Diabetes Care 37 (9) (2014) 2500–2507. [PubMed: 24974975]

[78]. Lane JM, et al., Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank, Nat. Commun 7 (2016) 10889. [PubMed: 26955885]

[79]. Davies G, et al., Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151), Mol. Psychiatry 21 (6) (2016) 758. [PubMed: 27046643]

[80]. Pilling LC, et al., Human longevity: 25 genetic loci associated in 389,166 UK biobank participants, Aging (Albany NY) 9 (12) (2017) 2504. [PubMed: 29227965]

[81]. Smith DJ, et al., Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants, PLoS One 8 (11) (2013) e75362. [PubMed: 24282498]

[82]. Hanlon P, et al., Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants, The Lancet Public Health 3 (7) (2018) e323–e332. [PubMed: 29908859]

[83]. Ravussin E, et al., A 2-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of health span and longevity, J. Gerontol.: Series A 70 (9) (2015) 1097–1104.

[84]. Martin CK, et al., Effect of calorie restriction on mood, quality of life, sleep, and sexual function in healthy nonobese adults: the CALERIE 2 randomized clinical trial, JAMA Int. Med 176 (6) (2016) 743–752.

[85]. Prism 8 user guide, Available from: https://www.graphpad.com/guides/prism/8/user-guide/index.htm, 2019.

[86]. Chen ATY, et al., Effects of C aenorhabditis elegans sgk-1 mutations on lifespan, stress resistance, and DAF-16/F ox O regulation, Aging Cell 12 (5) (2013) 932–940. [PubMed: 23786484]

[87]. Mitchell SJ, et al., Nicotinamide improves aspects of healthspan, but not lifespan, in mice, Cell Metabol. 27 (3) (2018) 667–676, e4.

[88]. Simonsen A, et al., Promoting basal levels of autophagy in the nervous system enhances longevity and oxidant resistance in adult Drosophila, Autophagy 4 (2) (2008) 176–184. [PubMed: 18059160]

[89]. Han SK, et al., OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research, Oncotarget 7 (35) (2016) 56147. [PubMed: 27528229]

[90]. Yang J-S, et al., OASIS: online application for the survival analysis of lifespan assays performed in aging research, PLoS One 6 (8) (2011) e23525. [PubMed: 21858155]

[91]. Tepper RG, et al., PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity, Cell 154 (3) (2013) 676–690. [PubMed: 23911329]

[92]. Lim JS, et al., Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer, Nature 545 (7654) (2017) 360. [PubMed: 28489825]

[93]. Hahm J-H, et al., C. elegans maximum velocity correlates with healthspan and is maintained in worms with an insulin receptor mutation, Nat. Commun 6 (2015) 8919. [PubMed: 26586186]

[94]. Jackson CH, flexsurv: a platform for parametric survival modeling, R. J. Stat. Softw (2016) 70.

[95]. Therneau TM, Lumley T, Package ‘survival’ Survival analysis Published on CRAN, 2014.

[96]. Kassambara A, et al., Package ‘survminer’ (2017). Available at: https://cran.r-project.org/web/packages/survminer/index.html.

[97]. Rickert J, Survival analysis with R.

[98]. Schütte D, Survival analysis in R for beginners, Available from: https://www.datacamp.com/community/tutorials/survival-analysis-R, 4 26 2018.

[99]. Mair W, et al., Lifespan extension induced by AMPK and calcineurin is mediated by CRTC-1 and CREB, Nature 470 (7334) (2011) 404. [PubMed: 21331044]

[100]. Whitaker R, et al., Increased expression of Drosophila Sir 2 extends life span in a dose-dependent manner, Aging (Albany NY) 5 (9) (2013) 682. [PubMed: 24036492]

[101]. Bos PD, et al., Genes that mediate breast cancer metastasis to the brain, Nature 459 (7249) (2009) 1005. [PubMed: 19421193]

[102]. Fabris F, De Magalhães JP, Freitas AA, A review of supervised machine learning applied to ageing research, Biogerontology 18 (2) (2017) 171–188. [PubMed: 28265788]

[103]. Kerber RA, O'Brien E, Cawthon RM, Gene expression profiles associated with aging and mortality in humans, Aging Cell 8 (3) (2009) 239–250. [PubMed: 19245677]

[104]. Nakamura E, Miyao K, A method for identifying biomarkers of aging and constructing an index of biological age in humans, J. Gerontol. Ser. A Biol. Med. Sci 62 (10) (2007) 1096–1105.

[105]. Horvath S, DNA methylation age of human tissues and cell types, Genome Biol. 14 (10) (2013) 3156.

[106]. Putin E, et al., Deep biomarkers of human aging: application of deep neural networks to biomarker development, Aging (Albany NY) 8 (5) (2016) 1021. [PubMed: 27191382]

[107]. Li Y-H, Zhang G-G, Guo Z, Computational prediction of aging genes in human, in: 2010 International Conference on Biomedical Engineering and Computer Science, IEEE, 2010.

[108]. Li Y-H, Dong M-Q, Guo Z, Systematic analysis and prediction of longevity genes in Caenorhabditis elegans, Mech. Ageing Dev 131 (11–12) (2010) 700–709. [PubMed: 20934447]

[109]. Wan C, Freitas A, Prediction of the pro-longevity or anti-longevity effect of Caenorhabditis Elegans genes based on Bayesian classification methods, in: 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE, 2013.

[110]. Wan C, Freitas AA, De Magalhães JP, Predicting the pro-longevity or anti-longevity effect of model organism genes with new hierarchical feature selection methods, IEEE ACM Trans. Comput. Biol. Bioinform 12 (2) (2015) 262–275.

[111]. Feng K, et al., Topological anaylysis and prediction of aging genes in Mus musculus, in: 2012 International Conference on Systems and Informatics (ICSAI2012), IEEE, 2012.

[112]. Freitas AA, Vasieva O, de Magalhães JP, A data mining approach for classifying DNA repair genes into ageing-related or non-ageing-related, BMC Genomics 12 (1) (2011) 27. [PubMed: 21226956]

[113]. Kerepesi C, et al., Prediction and characterization of human ageing-related proteins by using machine learning, Sci. Rep 8 (1) (2018) 4094. [PubMed: 29511309]

[114]. Bocklandt S, et al., Epigenetic predictor of age, PLoS One 6 (6) (2011) e14821. [PubMed: 21731603]

[115]. Hannum G, et al., Genome-wide methylation profiles reveal quantitative views of human aging rates, Mol. Cell 49 (2) (2013) 359–367. [PubMed: 23177740]

[116]. Horvath S, Raj K, DNA methylation-based biomarkers and the epigenetic clock theory of ageing, Nat. Rev. Genet (2018) 1.

[117]. Marioni RE, et al., DNA methylation age of blood predicts all-cause mortality in later life, Genome Biol. 16 (1) (2015) 25. [PubMed: 25633388]

[118]. Marioni RE, et al., The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936, Int. J. Epidemiol 44 (4) (2015) 1388–1396. [PubMed: 25617346]

[119]. Levine ME, et al., Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning, Aging (Albany NY) 7 (12) (2015) 1198. [PubMed: 26684672]

[120]. Horvath S, Ritz BR, Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients, Aging (Albany NY) 7 (12) (2015) 1130. [PubMed: 26655927]

[121]. Horvath S, et al., Decreased epigenetic age of PBMCs from Italian semisupercentenarians and their offspring, Aging (Albany NY) 7 (12) (2015) 1159. [PubMed: 26678252]

[122]. Levine ME, et al., DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative, Aging (Albany NY) 7 (9) (2015) 690. [PubMed: 26411804]

[123]. Ambatipudi S, et al., DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. European Journal of Cancer 75 (2017) 299–307. [PubMed: 28259012]

[124]. Levine ME, et al., An epigenetic biomarker of aging for lifespan and healthspan, Aging (Albany NY) 10 (4) (2018) 573. [PubMed: 29676998]

[125]. NIH data sharing policy and implementation guidance.

[126]. NIH Data Sharing Repositories (2019). Available from: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html.

[127]. NIH grants & funding: frequently asked questions, data sharing.

[128]. Wang L-S, et al., Nia genetics of Alzheimer'S disease data storage site (Niagads): 2014 update, Alzheimer's Dementia: The Journal of the Alzheimer's Association 10 (4) (2014) P634–P635.

[129]. Hodes RJ, Buckholtz N, Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing, Taylor & Francis, 2016.

[130]. Aging Cell: Author Guidelines, Available from, https://onlinelibrary.wiley.com/page/journal/14749726/homepage/forauthors.html, 2019.

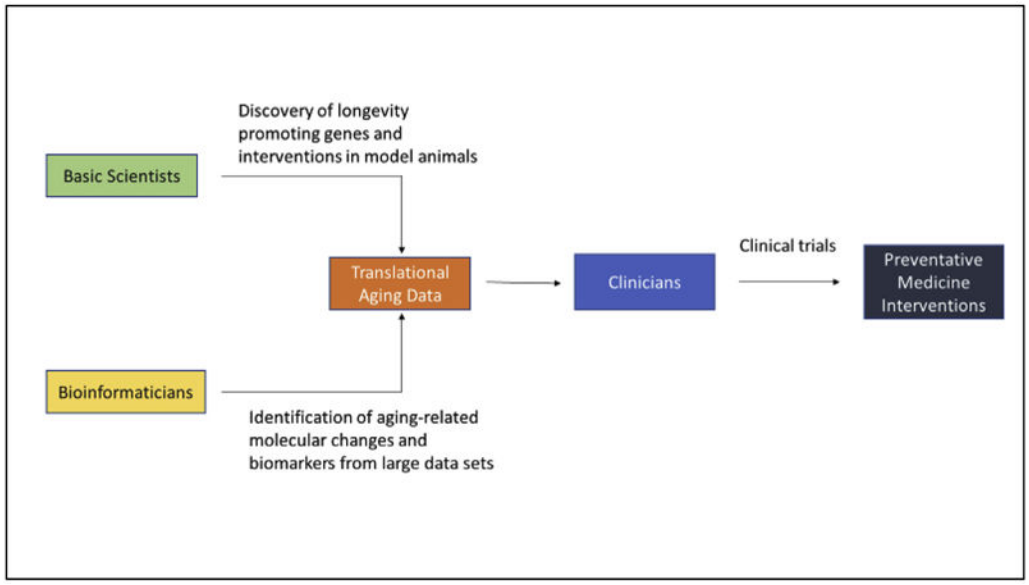


**Fig. 1.**
Translational pipeline for aging interventions.

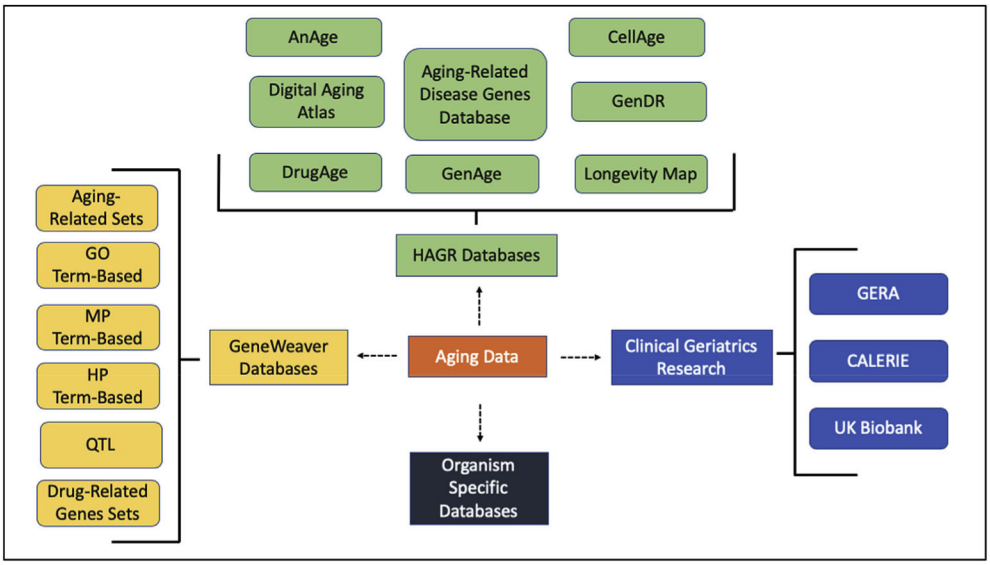


**Fig. 2.**
Online databases used in aging research.

**Table 1**

Databases and resources collecting data on basic and clinical geroscience.

| Name | Organism(s) | Type of data |
|---|---|---|
| AnAge (HAGR) | Many | Species maximum lifespan and other life history characteristics (sexual maturity, weight etc.). Currently contains 4244 species. |
| DrugAge (HAGR) | Many | Drugs that have been tested for effects on lifespan, dose used, whether a significant effect was observed, and percent lifespan change in numerous species. Currently contains 1821 drugs. |
| Aging-Related Disease Genes (ARD) Database (HAGR) | Human and aging models (*M. Musculus*, *C. elegans*, *D. Melanogaster*, *S. Cerevisiae*) | Gene sets associated with age-related diseases (derived from the Genetic Association database). Currently contains 769 genes. |
| CellAge (HAGR) | Human and Mouse | Curated database of gene manipulations affecting cellular senescence. Currently contains 279 genes. |
| Longevity Map (HAGR) | Human | Curated database of loci associated with longevity from GWAS studies. Currently contains 550 variants. |
| GenDR (HAGR) | Mouse and aging models | Database of gene manipulations that are essential for the lifespan phenotypes of dietary restriction in model systems, and gene expression changes caused by dietary restriction. Currently includes 214 manipulations and 173 expression changes. |
| GenAge (HAGR) | Human and aging models | Database of genes associated with aging in humans and genetic interventions that alter aging in animal models. Includes 307 human genes and 2152 model organism genes. |
| Digital Aging Atlas | Human | Database of gene expression, physiological, and pathological changes that occur during aging. Currently contains 8176 records of aging-related changes and genes, including data from humans and mice. |
| GeneWeaver | Human, Mouse, and aging model animals | Curated database of gene datasets based on Gene Ontology (GO) terms, Mammalian and Human Phenotype, Quantitative Trait Loci, Medical Subject Headings, gene expression, and drug-related gene sets. |
| Mouse Phenome Database | Mouse | Longevity and other phenotype data on inbred mouse strains available from Jackson Laboratory. |
| Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort | Human | Genotyping data, summary data from questionnaires, health records, saliva samples for large aging population in California, U.S. Includes over 100,000 participants |
| CALERIE | Human | Health records, survey data, tissue samples for healthy humans undergoing calorie restriction and controls. Includes 218 participants |
| UK BioBank | Human | Health records, tissue samples, genotyping data for large UK population (with MRI imaging study coming soon). Includes 500,000 participants |
| WHO Study on Global AGEing and Adult Health (SAGE) | Human | Cross-sectional health survey data and limited biological data (blood spots) for aging cohorts in China, Ghana, India, Mexico, Russian Federation and South Africa, with a total sample size of over 40,000. |
| Mayo Clinic Biobank | Human | Questionnaire and health record data, blood and other samples from over 21,000 patients at the Mayo Clinic in Rochester, Minnesota. |
| BioBank Sweden | Human | Group governing biobanks in Sweden with over 150 million samples, health register data. |
| Copenhagen Aging and Midlife Biobank (CAMB) | Human | Questionnaire data and blood samples from over 5000 participants, focused on aging and middle-aged population, linked to Danish national health and social registers. |
| NIH Aging Research Biobank | Human | Samples and genetic databases collected by previous National Institutes on Aging funded studies. |
| Taiwan Biobank | Human | Samples and health record data from Taiwanese population with 200,000 participants. |