Introduction

As the aging population grows worldwide, identifying factors that promote prolonged life- and health-span is an urgent public health priority [1]. One fundamental goal of aging research is to deepen the understanding of factors that contribute to healthy aging and delay onset of functional decline and disability. This is reflected in the mission statement of the National Institute on Aging (NIA) Intramural Research Program (IRP), National Institutes of Health (NIH):

The central focus of our research is understanding age-related changes in physiology and the ability to adapt to environmental stress. This understanding is then applied to developing insight about the pathophysiology of age-related diseases. The NIA IRP seeks to understand the changes associated with healthy aging and to define the criteria for evaluating when changes should be considered pathologic and require treatment. Thus, in addition to studying common age-related diseases, such as Alzheimer's Disease, Parkinson’s Disease, stroke, atherosclerosis, osteoarthritis, diabetes and cancer, we also explore the determinants of healthy aging as possible targets for interventions aimed at improving health and quality of life in the older population at large [2].

In our research group, epidemiologic studies are one method by which we examine biological, environmental and behavioral factors that contribute...
to health span. Two longitudinal cohort studies of aging drive the majority of epidemiological research in our research unit. The first, the Baltimore Longitudinal Study of Aging (BLSA) is a unique cohort study that began in 1958 to understand the process of normal aging [3]. Currently, over 1300 men and women aged 21 to over 100 years participate in the BLSA and are assessed every 1 to 4 years depending on their age. At each visit, participants come to the clinical research unit of the NIA IRP for 3 to 4 days to perform a battery of tests designed to measure biological, behavioral, environmental and molecular changes (https://www.blsa.nih.gov/) [3]. The second cohort is the InCHIANTI study, an epidemiological study of aging that aims to understand the factors that lead to mobility problems in old age (http://inchiantistudy.net) [4]. InCHIANTI began in 1998 with 1453 subjects recruited from two sites in the Tuscany region of Italy (Greve in Chianti and Bagno a Ripoli). To date, four follow up visits approximately 3 years apart have been conducted. Similar to the BLSA, the InCHIANTI study collects a vast number of clinical, biological, behavioral and molecular data at each visit. Both the BLSA and InCHIANTI provide ideal platforms to not only describe the changes that occur with aging but also to identify factors that contribute to these changes.

In recent years, development of omics technology has enabled the collection of molecular data such as genetics, epigenetics (methylation), gene expression, metabolomics and proteomics. Such data provides an opportunity to analyze large numbers of biomarkers in relation to aging and the potential to uncover some of the important molecular mechanisms underlying the aging process (Figure 1). Here, I review the research on aging conducted using ‘omics’ data in the BLSA and InCHIANTI studies with an emphasis on genetics but also more recent endeavors focused on methylation, gene expression and proteomics data.

1. Genetic association studies in aging

1.1 Genetics and Longevity

Longevity is a trait that aggregates in families, where offspring of parents with a longer lifespan tend to live longer than those whose parents are short lived [5-7]. In addition, children of longer lived parents have fewer disease risk factors and lower risk of cardiovascular disease, cancer and cognitive decline in midlife [5, 8-13]. These observations suggest

![Diagram of genomics, epigenomics, transcriptomics, and proteomics](image)

**Figure 1** Aging biomarker discovery using omics data.

Comprehensive assessment of molecular biomarkers allows large scale measurement of biomarkers at different levels. The different levels of omics data are shown here, starting with genomics where single nucleotide polymorphisms or other changes in DNA sequence are measured. Post transcriptional changes such as DNA methylation are important controls of gene expression. Transcriptomics is the comprehensive measurement of RNA in different tissues types. Finally, in proteomics, the goal is to measure as many proteins within the target tissue as possible. These omics data can be correlated with aging phenotypes to identify important molecular pathways that change with age.
that genetics contributes to longevity as well as other important age-related phenotypes. Family based studies can be used to estimate the heritability, or the proportion of the phenotypic variability that can be explained by genetic factors. Using this design, many aging traits have been shown to have varying levels of heritability [14]. For example, longevity has low to moderate heritability ranging from 15-25% [14-17]. Over the last few decades, rapid advancements in genotyping technology have allowed high throughput genotyping in population studies which has enabled large scale genetic studies of complex traits. International efforts to identify, catalog, and understand patterns of human genetic variation such as the International HapMap project [18] and the 1000 genome project consortium [19] have built a reference genome that allows millions of unmeasured single nucleotide polymorphisms (SNPs) to be imputed. Using both genotyped and imputed data, genome-wide association studies (GWAS) of complex diseases has become one of the standard methods to investigate the genetic architecture of human conditions including aging. We have participated in many international projects aimed at identifying genetic loci of aging (described below). Overall, these efforts have revealed the complexity of human aging and highlight the challenges in studying the genetics of longevity and healthy aging.

One of our initial projects in genetics of aging was the GWAS meta-analysis of longevity as part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [20]. In this study, participants were drawn from 11 cohorts of European descent where 6036 long-long lived individuals were compared with 3757 control subjects. Longevity was defined as survival beyond the age of 90 years and controls were subjects who died between 55 to 80 years of age. In the initial meta-analysis of ~2.5 million genotyped and imputed SNPs, no variants reached the genome-wide significance threshold of \( p \leq 5 \times 10^{-8} \); however 7 loci in or near GRIK2, CADM2, RGS7, SOX6, MBOAT1, PFKM, and LIMCHI genes showed suggestive associations at the \( p \leq 1 \times 10^{-5} \). A look-up of these seven loci along with a SNP in the FOXO3 gene that nearly reached the threshold \( p=8.56 \times 10^{-5} \) was conducted in other studies of aging with similar phenotypes [21]. Although none of the associations reached the Bonferroni corrected threshold \( p<0.0006 \), the FOXO3 SNP showed a consistent direction of association with marginal significance \( p=0.02 \). In addition to the GWAS approach, candidate loci look-up of previously identified loci from a GWAS of centenarians [22] and linkage analysis of longevity sib-pairs [23] provided a replication of the APOE locus (OR=1.20; \( P=4.8 \times 10^{-6} \)).

Several promising novel genetic loci have been identified through GWAS of longevity, however, the majority of these have not been replicated across studies [21, 24-28]. The two loci that have had the most consistent associations with longevity are the APOE and FOXO3. The APOE locus with three isoforms e2, e3 and e4 defined by two SNPs (rs429358 and rs7412) is the most widely studied. This locus was first described in association with longevity in a candidate gene association study of 338 centenarians from France where the e4 allele was found less frequently in centenarian compared to control adults aged 20 to 70 years [29]. Similar associations were subsequently described in other centenarian and nonagenarians [14]. This e4 isoform is also a risk factor for other aging traits including dyslipidemia, cardiovascular disease, and cognitive impairment. The two SNPs that define the APOE isoforms are not tagged well in the genotyping chip. Therefore, it is unclear whether the APOE locus identified through GWAS is the same signal. Whether there is one or multiple loci within the APOE gene region, this genetic region appears to be important in aging.

1.2 Genetics and Parental longevity

Another strong proxy phenotype of longevity and health span is parental longevity. In the Health and Retirement study (HRS), a representative multiethnic sample of Americans aged 51 to 61 years [30], with every decade of maternal and paternal survival beyond 65 years of age was associated with a decline of all-cause mortality by 19% and 14% respectively. In addition, offspring with one or both long-lived parents had a lower incidence of cancer (HR=0.76, \( P=0.001 \)), diabetes (HR=0.89, \( P=0.002 \)), heart disease (HR=0.88, \( P<0.001 \)) and stroke (HR=0.86, \( P=0.002 \)). Subjects with two long-lived parents had a 40% slower rate of cognitive decline and were less likely to be diagnosed with memory disorder compared to those with no long-lived parents [9]. These associations between parental longevity and favorable health outcomes have been reported in other cohorts [8, 10-13]. One advantage of using parental lifespan as the phenotype for a genetics study is the increased power attained by larger sample sizes as data on parental age of death is more likely to be available than data on the age of death of the participant, particularly for younger or middle age cohorts. There have been a few GWAS studies of parental lifespan with interesting observations [31-33]. Two notable studies in the UK biobank simultaneously reported a genome-wide significant signal in the
nicotine receptor locus (CHRNA3) [31, 32]. This locus was previously linked with smoking behavior and lung cancer. The allele associated with greater smoking was also linked with younger parental age of death and this effect was sex-dependent where the association was stronger or specific to paternal age of death. In addition to the CHRNA3 locus, the APOE locus was also associated with parental lifespan. An interesting observation was made where assessment of genetic risk scores for common age-related traits showed that offspring of long lived parents had an enrichment of protective alleles for coronary artery disease, systolic blood pressure, body mass index, total cholesterol, triglycerides, type 1 diabetes, inflammatory bowel disease and Alzheimer’s disease [31].

1.3. Genetics and Muscle function

In addition to longevity, other age-related phenotypes have been evaluated in GWAS studies including muscle function using grip strength as a proxy measure. Epidemiological studies have shown that grip strength in mid-life is a strong predictor of survival as well as other important measures of aging such as mobility disability[34, 35]. Although decline in muscle strength with age is determined by multiple factors such as physical activity and hormonal changes with age, family studies have shown that genetic factors also contribute to grip strength variability with heritability estimates ranging from 40 to 65% [36-40]. We therefore sought to identify genetic loci associated with grip strength through a GWAS in 27,581 individuals of European descent from 14 cohorts in the CHARGE consortium [41]. All cohorts measured grip strength using a hand dynamometer. Two genome-wide significant associations were discovered on chromosomes 8 (rs752045, p=3.09x10^-8) and 10 (rs3121278, p=2.68x10^-8). In addition to these two loci, 3 other SNPs that had suggestive associations on chromosomes 7, 8, and 11 were tested in an independent replication sample of 6393 individuals from three cohort studies. The SNP rs752045 on chromosome 8 was consistently associated with grip strength in the replication samples (P_{meta-analysis}=5.2x10^-10). This variant is in an intergenic region of chromosome 8, where the closest gene is over 500kb away. While functional annotation suggests that this region may be important for myotube differentiation and muscle repair, the exact function underlying this association is unclear. Further studies are needed to confirm this association as well as to understand the biology underlying the signal.

2. DNA Methylation and chronological age and aging

Epigenetics describes post-transcriptional modification of DNA and RNA packaging that influences gene expression without changing DNA sequence [42]. As epigenetics is modifiable, it is one of the ways in which the genome can respond to the environment. One such epigenetic marker, DNA methylation, is quickly becoming a promising biomarkers of aging. It has been shown that there is an overall decrease in genomic DNA methylation with age but an increase in DNA methylation variability [43, 44]. DNA methylation occurs on the cytosine of the DNA and is frequently found on CpG dinucleotides [45]. These DNA methylation patterns can be measured using an array or through bisulfite sequencing methods.

There is growing interest in using DNA methylation based age prediction score referred to as the “epigenetic clock” or “epigenetic age” to study the relationship between epigenetics and aging [46-49]. Epigenetic age can predict chronological age using as few as three CpG with high accuracy with mean average deviation of less than 5 years [46, 47]. The epigenetic clock developed by Horvath using 353 CpG sites predicted chronological age with an overall correlation of 0.96 with an error of 3.6 years. Most interestingly, this association was robust across several tissues including peripheral blood mononuclear cells, whole blood, brain, breast, liver and adipose tissue [49]. Although the epigenetic clock is highly correlated with age, methylation age appears to be an important risk factor for aging traits independent of chronological age. Our group and others have shown that those with higher epigenetic age than their chronological age, or accelerated methylation age, had increased risk of all-cause mortality [50-53]. Accelerated methylation aging was also associated with other aging conditions such as frailty [54], cancer [55], and physical and cognitive function [56]. Taken together, these results suggest accelerated aging based on DNA methylation captures an aspect of biological aging that has a measurable effect of aging phenotypes.

3. Gene expression

Transcriptomics is the comprehensive assessment of RNA transcripts in cells. Similar to DNA methylation, gene expression is dynamic and can change over time. It is estimated that there are ~19,000 protein coding human genes [57]. Gene expression profiles can be measured in different tissues using gene expression arrays or through RNA-seq. There are multiple reports of age-related expression differences in tissues
including the brain, skin, adipose and kidney. There appears to be little consistency in gene expression changes across different tissues which would suggest that tissue specificity may be important depending on the phenotype of interest. Ideally, investigating gene expression changes in all tissues would provide a comprehensive assessment of transcriptomic changes with aging, but in epidemiological studies, the most widely available data is gene expression in whole blood.

The InCHIANTI study was among the most comprehensive transcriptomics studies of age conducted in whole blood which involved 14,983 individuals from 13 cohorts [58]. In this study, approximately half of the genes (n=11,908) in the human genome were assessed for differential expression with age. There were 1495 age-associated genes of which 897 were negatively and 600 were positively associated with age. Many of the genes identified represent known aging pathways including immune and mitochondrial related pathways. In addition, novel pathways not previously been described in human aging such as glycosaminoglycan degradation and actin remodeling were highlighted. Using all the genes that were measured, a transcriptomic age predictor was developed that was highly correlated with chronological age. Similar to the DNA methylation clock, the deviation between the transcriptomic age and chronological age (mean difference was 7.8 years) was used as a measure of biological aging. Transcriptomic age was significantly but modestly correlated with two epigenetic age measures with correlations ranging from 0.1 to 0.33 suggesting that biological aging measured by DNA methylation and transcription show some consistency but may reflect different biological phenomenon. In addition, higher transcriptomic age was associated with higher systolic blood pressure, waist-hip ratio and higher prevalence of smoking. Thus, similar to the epigenetic clock, transcriptomic age correlates with age-related risk factors.

4. Proteomics

In proteomics research, various assays including immunoassays and mass spectrometry are used to create a comprehensive catalog of proteins in various tissues [59]. While the number of proteins that can be measured have steadily increased, performing discovery proteomics in serum, plasma, urine and in tissues such as skin, skeletal muscle and adipose tissue remains a challenge. Some of the main obstacles are the large numbers of proteins present, the extremely wide range of concentration, and the presence of highly abundant proteins that mask those with low concentration [60-62]. One unique multiplex method for relative protein quantification of up to ~1300 proteins utilizes single-stranded DNA oligonucleotides that can bind to protein with high affinity and specificity [63]. This technology utilizes DNA microarrays and allows high throughput measurements in different tissues. Recently, this methodology was used in two studies to evaluate the association of protein levels with chronological age. In the first study, 800 proteins were assessed in cerebrospinal fluid of 90 cognitive normal adults aged 21 to 85 years. Eighty-one proteins were associated with age and the most significant gene ontologies enriched in these proteins were those involved in inflammation and response to injury [64]. In a second study, 1129 proteins were assessed in plasma samples from 202 females from the Twins UK cohort [65]. Thirteen proteins were found to be associated with age, of which 10 were confirmed in an independent replication sample of 677 individuals. Preliminary analysis in BLSA confirmed these finding (unpublished results). The most significant protein, chordin-like protein 1 was also found to be significantly associated with higher birthweight and lower Framingham 10-year cardiovascular risk score. Other notable proteins associated with age in plasma include insulin-like growth factor-binding protein 6 and matrix metalloproteinase-12 both of which have been implicated in the aging process [65]. These studies provide promising results and support the use of proteomics in discovery studies of important molecular pathways of aging.

Conclusions

In this rapidly expanding field of omics, comprehensive molecular data are being collected in a large number of studies. One of the next exciting new chapters of genetics is the ability to capture greater detail of the genome through whole genome sequencing. This will improve the ability to accurately capture rare variants that may explain portions of the “missing heritability” or the portion of genetic variance that is not explained by common variants (with minor allele frequencies >5%) [66]. Methods to identify personal variants that are unique to an individual are being developed and may have important implications in understanding genetics of longevity. Aging studies using DNA methylation and gene expression data clearly show that there are epigenetic and transcriptomic signatures of aging that identify individuals at increased risk for premature aging outcomes. It is clear that one major lesson learned from the early omics studies is that no single
set of molecular biomarkers can be used to fully explain the process of aging. The future challenge will be to implement methods where the different omics data can be examined simultaneously. This multi-omics approach to build complex networks will likely improve our understanding of the aging process and contribute to the development of effective methods to prolong both life and health span in humans.

References
46. Bocklandt S, Lin W, Sehl ME, Sanchez FJ.