DISSERTATION

IDENTIFYING NOVEL MOLECULAR MECHANISMS OF HEALTHSPAN USING MULTI-OMICS

Submitted by

Meghan Elizabeth Smith

Department of Health and Exercise Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Tom LaRocca

Karyn Hamilton
Josiane Broussard
Nicole Ehrhart

ABSTRACT

IDENTIFYING NOVEL MOLECULAR MECHANISMS OF HEALTHSPAN USING MULTI-OMICS

An important goal in research on aging is to extend healthspan, the period of life spent healthy and disease-free. Next-generation sequencing and other emerging bioinformatics technologies (e.g., RNA-seq/transcriptomics, epigenetic profiling, and proteomics) have made it possible to broadly profile potential molecular mediators of aging, and perhaps identify therapeutic targets. The studies in this dissertation focus on using transcriptomics and complementary “multi-omics” strategies to characterize novel cellular mechanisms of aging, and to determine their relevance to systemic/functional health in humans. With the guidance of my mentoring team, I completed three studies in which I identified novel mediators of healthspan-related exercise training responsiveness, age-related inflammation, and cognitive/motor function decline in middle-aged and older adults. One particularly novel focus among these studies was the role of non-coding repetitive RNAs (derived from transposable elements) in healthspan. Transposable elements have been linked to known mechanisms of aging, and this topic is reviewed at the start of this dissertation to provide perspective on their role in the context of research on aging biology. Collectively, my findings represent new ideas for targetable genes and proteins that may influence human healthspan.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my mentor, Dr. Tom LaRocca, for his guidance and commitment throughout this degree. Words are not enough to describe how thankful I am that he took a chance and welcomed me into the HBL. Thank you for seeing us as human beings, and for the countless hours of whiteboard madness and editing over the years. When people ask me what I want to be when I grow up, I tell them I want to be Tom.

Thank you to my committee members, Drs. Karyn Hamilton, Josie Broussard, and Nicole Ehrhart for their time, constructive feedback, and inspiration as scientific role models. Thank you to my support system in Fort Collins and at home. I would not be writing this today if not for the friends I've made in HES who have laughed, struggled, and persevered alongside me for the last 5 years. Thank you to my mom and dad, Tara and Sinclair, my brother Justin, and extended family for making sure I feel their love and support from afar every day. Lastly, thank you to my almost-husband Andrew Keenan for supporting me in every way possible and making sure I eat dinner every night, and to our dogs Pie and Maple for their unconditional love and reminders to get off the computer and get outside.

TABLE OF CONTENTS

TRANSPOSABLE ELEMENTS AS A NOVEL HALLMARK OF AGING: A REVIEW

Introduction

An important goal in aging research is to extend healthspan, the period of time spent free of disease. Next-generation sequencing and bioinformatics technologies (e.g., RNA-seq/transcriptomics) are powerful ways to identify novel mechanisms of aging and therapeutic targets[1]. However, most transcriptome studies of aging have focused on identifying genes that drive aging/age-related disease, while transposable elements (TEs), which make up a large portion (~45%) of the genome[2], have generally been neglected[3]. Evidence suggests that the activation and accumulation of TE transcripts may be related to key mechanisms of aging because: 1) TE transcripts increase with age across the lifespan and can be used to predict age[4]; 2) healthy aging interventions (e.g., calorie restriction, rapamycin, exercise) reduce TE transcripts in pre-clinical models and humans[5]; and 3) TE transcript activation/accumulation is associated with key hallmarks of aging and age-related disease[6].

TEs include retrotransposons and DNA transposons, which represent most of our non-coding repetitive genomic sequences. Retrotransposons can propagate via a 'copy and paste' mechanism with an RNA intermediate, which is reverse transcribed into complementary DNA (cDNA) that can re-integrate at different locations in the genome[7,8]. Retrotransposons include long-interspersed nuclear elements (LINEs), short-interspersed nuclear elements (SINEs), and long terminal repeats (LTRs). In contrast, DNA transposons mobilize through a DNA intermediate via a 'cut and paste' mechanism. In the mammalian genome, most retrotransposons are inactive because they no longer contain the machinery for transcription and translocation. However, in humans, a subset (~0.1%) of LINEs (e.g., L1) contain two open reading frames encoding for an RNA chaperone and reverse transcriptase enzyme, which allows them to remain active, propagate and re-integrate in the genome[6,8]. SINEs, the other main type of retrotransposon, do not encode for the proteins necessary for autonomous

retrotransposition and must use LINE machinery to self-replicate. Like a subset of LINEs, some SINEs (e.g., *Alu* elements) are also active in the human genome as indicated by *de novo* insertions [9]. Similar to LINEs and SINEs, LTR retrotransposons (e.g., endogenous retroviruses) also copy via RNA intermediates and cDNA[10]. Finally, unlike LINEs, SINEs and LTR retrotransposons, DNA transposons are thought to be largely inactive in the human genome[11].

While most retrotransposons (and transposons) are epigenetically silenced at the genome level, recent reports show that certain types of retrotransposons are transcribed with aging as a result of chromatin dysregulation or other epigenetic mechanisms[12-16], and that activation/accumulation of their transcripts may drive other cellular mechanisms of aging[4,17-20]. However, this is an emerging topic, and the many possible connections among TEs and mediators of aging have not been discussed in detail in the current literature. Therefore, the purpose of this review is to highlight the most recent evidence supporting the connections between TE dysregulation and the ten established hallmarks of aging (**Figure 1.1**): genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, and inflammation[21,22]. We also discuss potential interventions and therapies to target TEs and mitigate their interactions with these hallmarks.

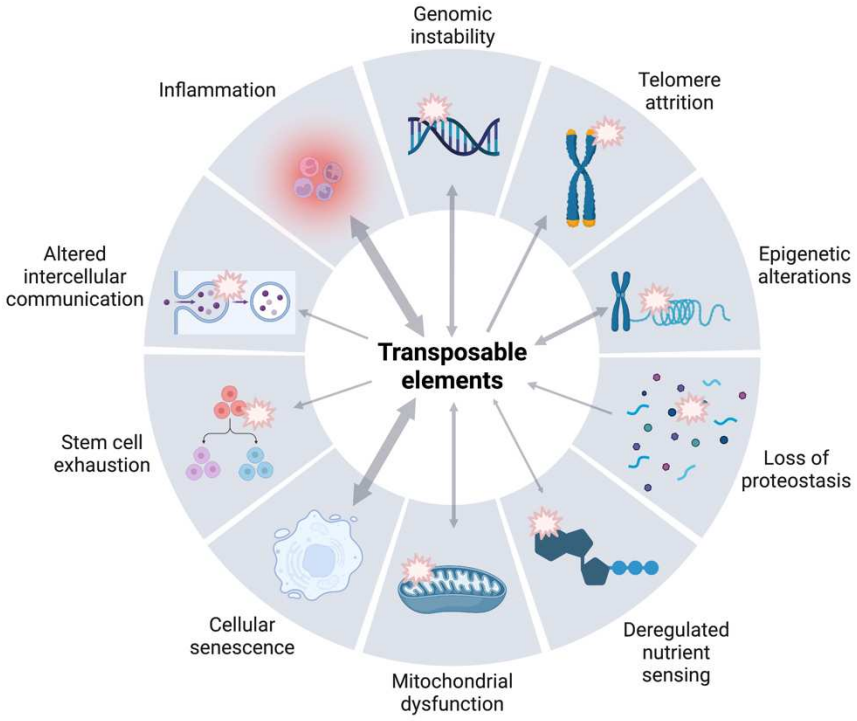


Figure 1.1. TE transcript accumulation and the hallmarks of aging. Age-related TE transcript accumulation/dysregulation relates to the ten hallmarks of aging, which includes genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, and inflammation. Arrows indicate directional associations, and the thickness of each arrow represents how much evidence supports the connection between TE dysregulation and each hallmark.

TEs and each hallmark of aging

1. Epigenetic alterations

Epigenetic alterations are often associated with TE dysregulation with aging. Although changes to the epigenome during aging are complex and multifactorial, changes to global chromatin structure (including the loss of histones) and reduced DNA methylation patterns appear to be most directly related to TE activation with aging[22]. Importantly, the propagation/re-insertion of TEs into the genome might also have a direct influence on the epigenome by altering the formation of heterochromatin[23] and causing the activation of DNA repair processes, which can reduce DNA methylation[24].

1.1. Loss of heterochromatin: Dysregulation of chromatin contributes to the transcription of TEs[27], but this has mainly been demonstrated in simple eukaryotic organisms including yeast and flies[28]. Briefly, chromatin includes euchromatin (loosely-packed regions of the genome that are transcriptionally active) and heterochromatin (regions of the genome that are tightly packed and transcriptionally repressed)[12]. The 'heterochromatin loss model of aging' suggests that decondensation of heterochromatin is closely associated with aging/age-related disease because its structure breaks down over time, partly due to a reduction in histone maintenance and synthesis, and this leads to loss of transcriptional control and subsequent cellular dysfunction[29,30]. Interestingly, the overexpression of histone proteins (H3 and H4) results in reduced retrotransposition in yeast[31] and heterochromatin is enriched in regions of the genome containing retrotransposons[32]. Therefore, one consequence of the loss of heterochromatin and global histone reductions during aging might be the transcription/activation of retrotransposons that are normally repressed[33]. In support of this general idea, data show that heterochromatin formation/maintenance increases lifespan in pre-clinical models[34], and that reductions in heterochromatin maintenance accelerate aging in human cells, in part by inducing senescence[35]. Still, as noted above, it remains to be determined whether the loss of heterochromatin is directly associated with the activation of TEs in mammalian tissues.

1.2. DNA hypomethylation: DNA methylation patterns can contribute to either the activation or repression of TEs, and this is dependent on developmental stage[36]. For example, during embryogenesis and development, reduced DNA methylation is important to activate TEs, thus contributing to developmental innovations and even genome evolution (e.g., because TE insertions can have a direct influence on neighboring genes)[37]. Reduced DNA methylation in regions of the genome containing autonomous (active) TEs enables them to propagate and multiply (ensuring their survival in the genome with multiple copies), but this expansion has

deleterious consequences (e.g., cell death) in the context of aging[38]. As a result, during aging, increased DNA methylation is critically important to suppress TEs that might otherwise become active and contribute to disease[36]. Even so, DNA *hypomethylation* is a well-characterized feature of cellular aging[39], and this age-associated change is the foundation of many "biological clocks" that estimate biological age (vs. chronological age)[40,41]. This age-related hypomethylation is likely central to TE dysregulation, as epigenetic studies have generally found global, age-dependent DNA hypomethylation in regions of the genome enriched in TEs[42], and aging in humans is associated with decreased methylation of *Alu* and LINE1 elements[43]. Additionally, retrotransposon activity is increased in DNA methylation-deficient mouse models (*Dnmt3L* and *Miwi2* mutants), and retrotransposition in these animals impairs genomic integrity by reducing chromatin stability and inducing double stranded DNA breaks[44]. These events may be somewhat self-perpetuating, in that propagation/re-insertion of TEs also has a direct influence on the epigenome by altering the formation of heterochromatin[23], and the insertion of TEs into the genome activates DNA repair processes, which can reduce DNA methylation[24]. Together, these data support the idea that epigenetic changes during aging, particularly the loss of heterochromatin and global reductions in histones and DNA hypomethylation, activate TEs that would normally be suppressed.

2. Genome instability and telomere attrition

2.1. Genome instability: A common mechanism of aging is the accumulation of DNA damage (e.g., due to UV radiation, replication errors, or reactive oxygen species)[21]. DNA lesions, which often manifest as double-stranded DNA breaks, occur often and randomly throughout the genome over the lifespan, and the ability to repair this damage decreases with aging[45]. Interestingly, it has been shown that that TE insertions into the genome can cause double-stranded DNA breaks (therefore inducing genome instability)[23]. These lesions initiate a DNA damage response (DDR) that temporarily pauses the cell cycle for DNA repair mechanisms,

and chronic DDR activation can lead to a permanent halting or change in cell fate such as apoptosis (programmed cell death) or senescence (also a hallmark of aging)[46]. Recent evidence suggests that epigenetic dysregulation and genomic instability are mainly due to DNA damage, and that the combination of these factors can increase derepression of TEs, including LINE1 in humans and IAP (a type of active LTR) in mice[24]. These and other data indicate that DNA damage may be a direct cause of TE dysregulation. For example, the increased mobility of some active human TEs (LINE-1) is associated with reduced DNA repair machinery[47], and it has also been shown that repressing certain TEs results in enhanced genome stability. In fact, targeting *Alu* elements (a type of highly mobile SINE that account for ~11% of the human genome) with small interfering RNA (siRNA) leads to improved resistance to DNA damaging agents, increased cell proliferation (reduced senescence), and enhanced genome stability[48,49]. Taken together, these data suggest the relationship between genome instability and TE activation is bidirectional in that: (1) TE activation/insertions can have a direct impact on genome instability (e.g., by inducing double stranded DNA breaks); and (2) genome instability can contribute to the activation of TEs (e.g., through initiating the DDR or senescence).

2.2. Telomere attrition: Telomeres are repetitive portions of the genome that protect the ends of chromosomes during cell replication[21]. The shortening of telomeres during aging contributes to genomic material loss and can lead to replicative senescence or apoptosis[50]. In contrast to the adverse effects of TEs on other hallmarks of aging, most of the existing research on this topic suggests that certain TEs are essential for maintaining telomere structure/length. Studies in *Drosophila* suggest that telomeres are maintained by three families of active TEs (HeT-a, TART, and Tahre)[51]. Also, in eukaryotic cells, telomeres are supported and elongated by a specialized reverse transcriptase enzyme, telomerase, which adds protective DNA to the ends of telomeres and has been evolutionarily linked to TEs (i.e., perhaps due to its reverse transcriptase activity)[52]. However, recent studies completed in yeast have found that erosion of telomeres is

associated with direct activation of *Ty1* retrotransposons (a type of LTR retrotransposon)[55], and in human cells, LINE1 activation and insertions have been identified at dysfunctional telomeres[56]. These findings suggest that the relationship between TEs and telomeres is complex and likely differs based on TE type, and more research is needed to determine if there is a link between telomere *shortening* and TEs.

3. Altered intercellular communication and chronic inflammation

TE transcript dysregulation and accumulation has been linked with altered intercellular communication including altered neuronal transmission, metabolic control (i.e., blood glucose and lipid levels), and impaired immune response in tumor microenvironments[57,58]. The precise roles of TEs in altered intercellular communication (e.g., hormones, signaling proteins, ligand receptor binding) are still not known. However, the most important link between TEs and age-related changes in intercellular communication involves cytokines and chemokines, which play a central role in cellular signaling, immune activation, and chronic low-grade inflammation[59]. Several studies point to inflammation as a "unifying" hallmark of aging ('inflammaging') and one of the main contributors to age-related disease[60]. Perhaps the strongest evidence for TE involvement in low-grade inflammation (including the activation of cytokines/chemokines) involves the fact that retrotransposons, including LINEs, SINEs, and LTRs, can be reverse transcribed, resulting in the formation of cDNA. During aging, this cDNA can escape the nucleus (perhaps due to decreased nuclear envelope integrity) and be recognized as a pathogen-associated molecular pattern that is detected by cyclic GMP-AMP synthase (cGAS)[61]. This cascade of events can activate the cGAS-STING pathway, an important signaling axis for immune responses/inflammation[62]. cGAS-STING initiates a type-1 interferon response, contributing to the expression of inflammatory cytokines that communicate the threat to surrounding cells, and these events have been associated with various age-related diseases, including cancers[62] (**Figure 1.2**).

In addition to the generation of cytoplasmic cDNA, more recent but limited evidence suggests that TE accumulation may contribute to chronic, low-grade inflammation via TE-derived double stranded RNA (dsRNA) that can also be recognized as a foreign entity and trigger inflammation/immune responses. For example, evidence suggests that transcripts from TEs can fold and base-pair ("intra-strand") to form dsRNA in the cytoplasm. Similarly, nearby transcripts that are complementary may pair up to form dsRNA ("inter-strand")[63]. These dsRNAs can present a virus-like stimulus that triggers cellular inflammatory and immune response signaling pathways via sensors such as Melanoma Differentiation Associated Protein 5 (MDA5), Retinoic Acid Inducible Gene 1 (RIG-1), and Protein-Kinase RNA activated (PKR)[64,65] (**Figure 1.2**). These events may represent additional mechanisms by which non-autonomous TEs (e.g., DNA transposons) can contribute to chronic, low-grade inflammation during aging.

Emerging evidence supports a role for cDNA, dsRNA and other TE-derived molecules in aging and disease. For example, in pre-clinical models and humans cells, elevated TE activity contributes to chronic inflammation and interferon responses[66], and this is reportedly due to increased expression of the LINE element L1 retrotransposon, which stimulates an inflammatory interferon response in senescent cells via cDNA, and during normal healthy (mouse) aging[67]. Similarly, endogenous retroviruses (ERVs, a major type of retrotransposon) have been shown to activate cytokines, chemokines, and downstream inflammatory pathways[68]. The activation of other retrotransposons like HERVK (a human-specific type of ERV) during aging also results in the production of endogenous, virus-like particles that cause inflammatory signaling and senescence in other cells[70]. Interestingly, the order of these events may be reversed in some cases, as a recent study showed that stimulation of Toll-Like Receptors 4 and 5 (TLR-4 and TLR-5, respectively) with pathogen-associated proteins altered the expression profile of TEs including LINEs and human endogenous retroviruses (HERVs) [69]. Finally, recent data show that pathogenic tau (a main hallmark of Alzheimer's disease) drives the activation of TEs that may

contribute to increased levels of TE-derived dsRNA and neuroinflammation, a major mechanism of brain aging and neurodegeneration[72].

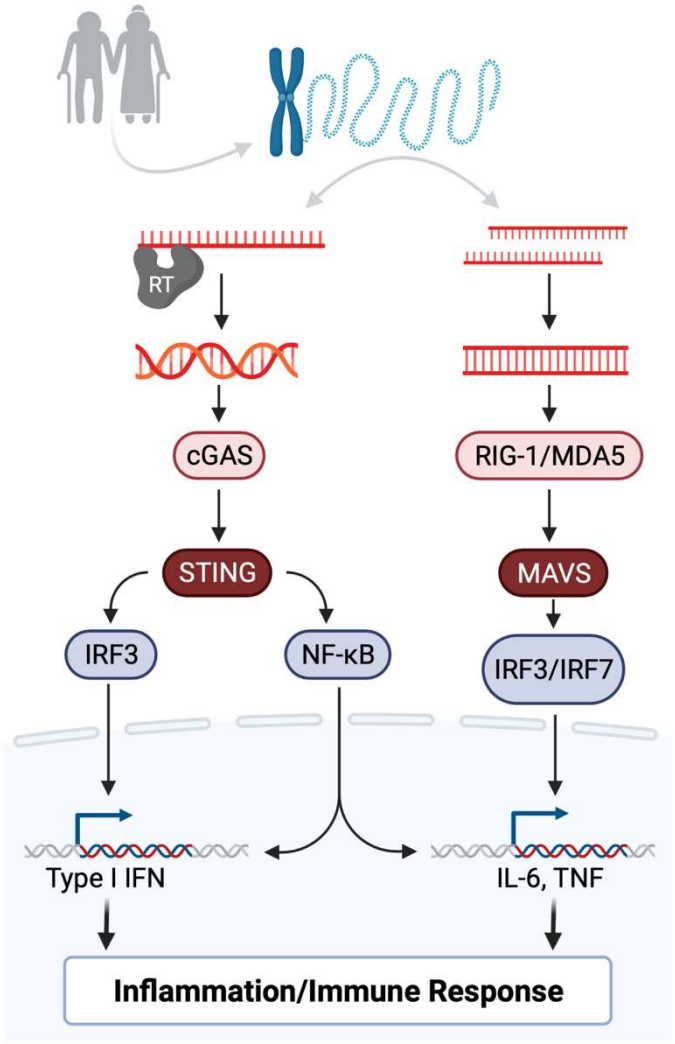


Figure 1.2. Mechanisms by which TE transcript accumulation triggers inflammatory signaling. Aging and age-related changes to the epigenome are associated with TE transcript accumulation. TE transcripts can be reverse transcribed into cDNA by reverse transcriptase (RT), and cDNA trigger the cGAS-STING inflammatory pathway. Additionally, palindromic TE transcripts can fold and base pair, or two complementary transcripts can pair up to form dsRNA, stimulating dsRNA sensors such as RIG-1/MDA5. These scenarios ultimately lead to type 1 interferon and inflammatory responses.

4. Cellular senescence

Cellular senescence is initiated by acute and chronic cell damage, and senescent cells accumulate with aging in multiple tissues [73,74]. Senescence contributes to cellular events that

may be important during aging/disease, including morphological changes, proliferation arrest, telomere dysfunction, inflammation/immune activation, and DNA damage—all of which result in senescent cells adopting a pro-inflammatory profile known as the “senescence-associated secretory phenotype” (SASP)[75]. . Most evidence points towards the fact that cellular senescence and the SASP can lead to an increase in TEs (e.g., LINE1 activity) and activate a type-1 interferon (IFN-I) response and inflammation in human cells[67]. However, this may be a bidirectional process, as chronic inflammation caused by TE-derived dsRNA or cDNA can also induce the SASP phenotype and increase the expression of pro-inflammatory cytokines and chemokines (e.g., IL-6, IL-8)[76]. Most recently, senescence in human cells was associated with increased levels of HERVK transcripts (a type of active human endogenous retrovirus), and an accumulation of HERVK-derived virus-like particles that cause inflammation/immune activation and senescence in other cells[70]. Interestingly, among all of these observations, one consistent change that occurs with cellular senescence is marked decondensation of satellite-specific heterochromatin. Satellites are not considered to be TEs per se, but they are a type of structural repetitive element that undergoes similar age-related increases in expression[4]. Decondensation of satellite heterochromatin is known as senescence-associated distension of satellites (SADS), and this phenomenon suggests that redistribution of condensed/decondensed heterochromatin is an early event in senescence[77]. As such, the current data indicate that: (1) senescence and the SASP contribute to the direct activation of TEs and inflammation/immune activation[6]; but (2) active TEs may also contribute to senescence and the SASP (e.g., by activating interferons and inflammation).

5. Deregulated nutrient sensing and impaired proteostasis

5.1. Deregulated nutrient sensing: Nutrient sensing pathways are highly conserved in evolution and involve hormones and proteins including Insulin/IGF, PI3K/AKT, AMPK, MTOR, and

Sirtuins [78]. These pathways respond to nutrients/stressors and modulate the activity of many transcription factors to regulate broad cellular processes including autophagy (recycling of damaged cellular components), inflammation, mitochondrial biogenesis, and longevity[79] . Recent evidence shows that several important proteins that mediate nutrient sensing also affect TEs, and this process is likely multifactorial (e.g., through chromatin maintenance or influencing DNA repair mechanisms). For example, sirtuins are key metabolic regulators that have been linked to healthspan and lifespan[80]. Mammalian sirtuins largely function as histone deacetylases (HDACs) that play a large role in regulating glucose and lipid metabolism, DNA repair/stability, stress responses and inflammation[81], and they are activated by calorie restriction and exercise (two healthy aging interventions). Consistent with these observations and the idea that TEs modulate healthspan, SIRT1, SIRT6, and SIRT7 have all been shown to play an important role in maintaining chromatin structure and DNA repair mechanisms, and to repress L1 retrotransposon activity and reduce downstream IFN-I and inflammation[80,82]. Similarly, AMP-activated protein kinase (AMPK) is part of a classic nutrient sensing pathway that mediates autophagy and other energy homeostasis processes during times of insufficient intracellular energy, and it is a direct regulator of epigenetic enzymes like HDACs and DNA methyltransferases (DNMTs). Increased AMPK activity is associated with reduced L1 mRNA expression[85], but AMPK expression declines with aging[83,84]. These emerging data suggest that TE expression may be linked to the activity of several key nutrient sensing proteins that have roles in healthspan and lifespan[86], but future studies are needed to understand the order of these events.

5.2. Loss of proteostasis: Aging is associated with impaired chaperone protein expression and protein homeostasis, known as proteostasis—coordinated synthesis and degradation of cellular proteins and organelles via autophagy and the proteasome—which leads to an accumulation of intra- and extracellular protein aggregates[87]. The links between proteostasis and TEs are still

under investigation. However, recent evidence suggests that impaired proteostasis during aging could underlie TE transcript accumulation (i.e., downstream increased transcription). For example, autophagy has been shown to be a requirement for the degradation of retrotransposon RNAs[129], and the proteasome appears to limit the genomic re-integration of some TEs[130]. Although not proteostasis per se, age-related declines also occur in the expression and function of related nucleic acid binding proteins, which help fold and stabilize RNA/DNA and can influence TE/dsRNA expression. As an example, the activity/function of the RNA-binding protein TAR DNA-Binding Protein 43 (TDP-43) is affected during aging (perhaps due impaired proteostasis[87]), and reduced TDP-43 function is associated with the accumulation of TE transcripts, dsRNA, and inflammation[88]. Finally, in addition to overall proteostasis and direct RNA/DNA binding proteins, age-related changes in cellular quality control proteins that are associated with TE control/degradation and disrupt dsRNA/cDNA (e.g., exonucleases) may become impaired with aging. For example, dsDNA fragments or cytoplasmic DNA from TEs that have escaped the nucleus can be degraded by the Three Prime Repair Exonuclease 1 (TREX1) exonuclease, which would prevent cGAS activation and the subsequent interferon and inflammation responses, and TREX1 activity may decline with age[62]. Overall, these data indicate that impaired proteostasis with aging may influence TEs via multiple pathways, particularly those affecting proteins meant to target/degrade cDNA, dsRNA, or TE transcripts.

6. Mitochondrial dysfunction

With aging, mitochondrial function deteriorates partly due to increases in reactive oxygen species and reduced mitochondrial proteostasis[91]. Interestingly, another hallmark of mitochondrial aging is the permeabilization of the mitochondrial membrane which can lead mitochondrial DNA (mtDNA) leakage into the cytosol and trigger the cGAS-STING pathway, as described above[66]. Mitochondrial dsRNAs, derived from aberrant transcription of mtDNA, can also travel through the permeable membrane and set off dsRNA immune sensors such as

mitochondrial antiviral signaling protein (MAVS), which is located on the outer membrane of the mitochondria, and this can initiate a proinflammatory response[92]. These events involve mtDNA/RNA rather than TEs per se, but they reflect the fact that mitochondrial dsRNAs can activate similar pro-inflammatory pathways as TE-derived cDNA/dsRNAs. Recent evidence, however, also shows that TEs can interact with mitochondria and contribute to their dysfunction during aging, perhaps via some of these same pathways[13]. For example, the transfection of *Alu* elements into retinal pigment epithelial cells causes mitochondrial dysfunction by increasing reactive oxygen species (ROS) and activating the inflammasome and apoptosis, in part via MAVS-related pathways[93]. Interestingly, *Alu* elements also appear to be enriched in genes that are involved in mitochondrial function[94], and it has been hypothesized that *Alu*-mediated mitochondrial dysfunction (perhaps via *Alu* effects on gene expression/function) may contribute to neurodegenerative disease[94]. Together, these limited data suggest that TEs might contribute to mitochondrial dysfunction and dysregulation by several mechanisms including the production of ROS.

7. Stem cell exhaustion

Aging is associated with reduced cell renewal and organ tissue/repair[22]. Interestingly, active TE insertions (e.g., of LINE-1) play an important role in the development of induced pluripotent stem cells (that can differentiate into any cell in the body)[95], and the partial derepression of TEs may be important for embryonic development and gametogenesis[96]. These results provide insight into mechanisms by which TEs affect stem cells in the context of development, but the links between stem cell exhaustion and TEs in aging/disease are not well understood. Evidence shows that mechanisms preserving somatic cell integrity are important for stem cell maintenance during aging, and several genomic regulatory mechanisms (e.g., Piwi and RNA interference) are present to help preserve heterochromatin maintenance and suppress retrotransposon activation in advanced age[98]. Limited evidence suggests that the derepression

of retrotransposons combined with reduced activity of Piwi may contribute to stem cell exhaustion in advanced age by causing germline stem cell loss and activating immune signaling[99]. As such, there may be divergent roles of TEs in stem cell function/regulation during development and aging, and more research is necessary to determine whether TE activation contributes to age-related stem cell exhaustion.

The role of TEs in age-related disease

The role of TEs in age-related diseases has been thoroughly reviewed elsewhere[18,100], particularly in the context of brain aging and neurodegeneration[101-104]. So far, TEs have been associated with macular degeneration[105], frontotemporal degeneration/amyotrophic lateral sclerosis[106-108], Alzheimer's disease (AD)[72,109] and age-related cognitive decline[110]. Emerging AD research suggests that pathogenic tau protein aggregation, a well-known characteristic of AD, drives heterochromatin decondensation and TE/retrotransposition activation[109].

In addition to brain aging/neurodegeneration, TEs are involved in cancer, but the mechanisms of their involvement are complex. On one hand, LINE-1 transcripts may contribute to inflammation and cell proliferation that precedes cancer growth[6,111]. On the other hand, some studies suggest that the interferon responses stimulated by TE activation may actually increase tumor cell death and surveillance, which would be protective against cancers[112,113]. Similar observations have been made with regards to other diseases of aging[6,114], and the relationship between TEs and cancer and likely depends on cell type and TE type. Thus, while derepression of TEs clearly occurs with age and in some age-related diseases, the specific mechanisms connecting TEs to disease pathology are still being elucidated.

Targeting TEs to improve healthspan

In addition to the growing evidence linking TE dysregulation with aging and disease, recent studies support several emerging therapies for *targeting* age-related TE dysregulation to reduce age-related adverse cellular events and diseases.

Nucleoside reverse-transcriptase inhibitors (NRTIs): NRTIs may protect against TE accumulation by preventing TE transcripts from being reverse transcribed into cDNA, which can either re-integrate into the genome, or escape the nucleus and cause cellular inflammation, as described above. Several reports show that NRTIs attenuate proinflammatory pathways associated with inflammaging (e.g., cGAS-STING), suggesting that these compounds may have the potential to increase healthspan. These studies have shown that NRTI supplementation reduces inflammation and, in some model organisms, increases health/lifespan. For example, in one study, the NRTI 3TC reduced LINE-1 copy number in the genome, and this was associated with reduced inflammation and senescence/SASP[67]. In other studies, the NRTI 3TC was shown to improve health and motor function progeria mice[115], increase lifespan in flies[28], and improve cognitive function in a mouse model of accelerated aging by reducing inflammation/immune responses[116]. Finally, recent studies from our lab and others' showed that 3TC improves cognitive function/memory, protects neuronal health/function, and reduces inflammation/immune response in old wild-type mice[70,110]. It is important to note, however, that several studies have shown adverse metabolic/mitochondria-related side effects of 3TC and other NRTIs[117]. As such, future work should examine the role of 3TC and other NRTIs on mitochondria and the metabolome[118], as this could provide insight on the therapeutic potential for NRTIs in the context of aging/disease.

Calorie restriction (CR): CR is defined as a total caloric reduction by 10-40% without malnutrition, and it has been shown to improve healthspan and lifespan in many pre-clinical

models and humans[119]. We recently reported that long-term 20% CR reduces global levels of TE transcripts in mice[5], and others have shown that CR reduces retrotransposition in aging mouse tissues[16]. Growing evidence also suggests that CR might reduce TE transcripts by activating the nutrient sensing pathways and remodeling chromatin during aging[120]. As described above, this chromatin remodeling might have a direct impact on TE activation via genome wide alterations in DNA methylation[121]. While more research should be completed in this area (e.g., evaluating short- vs. long-term and/or different degrees of CR) and CR is not a "TE-specific" therapy, these emerging results suggest that CR, which is a robust treatment to prevent disease in pre-clinical models, might partly provide its beneficial effects by acting on TEs.

Exercise: Like CR, exercise is a robust intervention or lifestyle approach to promote healthy aging[122]. Our lab demonstrated that habitual exercise is associated with reductions in TE transcripts in peripheral blood mononuclear cells (PBMCs) in older humans. These reductions were also related to higher maximal aerobic capacity (VO_2max, a predictor of healthspan)[5]. Another study demonstrated that resistance training downregulates skeletal muscle LINE-1 activity[123]. The mechanistic links between exercise and genome maintenance/TE activation are unclear. However, like CR, exercise is associated with improved methylation profiles and heterochromatin maintenance[124], and these effects are most likely mediated by the same nutrient sensors (e.g., sirtuins and AMPK) described above.

DNA methyltransferase (DNMT) activators: DNMTs are responsible for methylating genes and TEs[38], and their activity decreases with age[125,126]. Recent studies in mice have demonstrated that DNMTs are critical to promote DNA methylation and that the loss of DNMT function (via *Dnmt1* genetic knockdown) increases inflammation and reduces health and cognitive function[127]. Studies like this have not investigated TEs specifically, but their findings are

consistent with the idea that DNMT activators could help restore/maintain TE methylation with aging (and therefore reduce downstream inflammation/immune activation). Additional work in model organisms (and eventually humans or at least human cells) is needed to determine whether DNMT activators are a safe intervention for improving healthspan/lifespan, and the extent to which TE repression may be involved.

Anti-aging compounds: Several anti-aging compounds that extend health/lifespan are supported by strong evidence. These include rapamycin, acarbose, 17 alpha-estradiol, and Protandim[128]. Our lab has shown that both short-term and long-term supplementation with these compounds reduces global levels of TE transcripts in the liver (a key metabolic organ), and that some of these effects are associated with gene signatures consistent with improved DNA repair and reduced inflammation/immune activation[5]. Importantly, these compounds have also been shown to directly modulate the hallmarks of aging. Together, these observations: 1) further support the idea that TEs interact with key mechanisms of aging; and 2) suggest a common or conserved pathway that links TE activation with these aging drivers. Determining what these common pathways may be is a key direction for future research and could be accomplished using experimental designs that directly target TE activation (via reverse transcriptase inhibition or DNMT inhibitors, etc.) and interrogate each hallmark of aging simultaneously.

Summary

In summary, TE activation/accumulation is linked with many hallmarks of aging. The strongest and most recent evidence suggests strong links among TE activation/accumulation, epigenetic alterations (chromatin dysregulation and alterations in DNA methylation), senescence/SASP, and chronic inflammation. Current research is focusing on the mechanisms by which TE accumulation/activation affects the other hallmarks of aging, but these mechanisms remain less clear, especially for “newer” hallmarks of aging not discussed here (e.g., dysbiosis).

Interventions that reduce global levels of TE transcripts and improve healthspan include NRTIs, non-pharmacological approaches like CR and exercise, and anti-aging compounds. Identifying additional interventions that target TEs might have therapeutic promise for improving health in older age.

REFERENCES

1 Wahl, D. & LaRocca, T. J. Transcriptomic Effects of Healthspan-Promoting Dietary Interventions: Current Evidence and Future Directions. *Front Nutr* **8**, 712129 (2021). https://doi.org:10.3389/fnut.2021.712129

2 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001). https://doi.org:10.1038/35057062

3 Gloss, B. S. & Dinger, M. E. Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med* **50**, 1-8 (2018). https://doi.org:10.1038/s12276-018-0087-0

4 LaRocca, T. J., Cavalier, A. N. & Wahl, D. Repetitive elements as a transcriptomic marker of aging: Evidence in multiple datasets and models. *Aging Cell* **19**, e13167 (2020). https://doi.org:10.1111/acel.13167

5 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R. & LaRocca, T. J. Healthy Aging Interventions Reduce Repetitive Element Transcripts. *J Gerontol A Biol Sci Med Sci* **76**, 805-810 (2021). https://doi.org:10.1093/gerona/glaa302

6 Gorbunova, V. *et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43-53 (2021). https://doi.org:10.1038/s41586-021-03542-y

7 Fukuda, S. *et al.* Cytoplasmic synthesis of endogenous Alu complementary DNA via reverse transcription and implications in age-related macular degeneration. *Proceedings of the National Academy of Sciences* **118**, e2022751118 (2021). https://doi.org:doi:10.1073/pnas.2022751118

8 Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biology* **19**, 199 (2018). https://doi.org:10.1186/s13059-018-1577-z

9 Belancio, V. P., Hedges, D. J. & Deininger, P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res* **18**, 343-358 (2008). https://doi.org:10.1101/gr.5558208

10 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009). https://doi.org:10.1038/nrg2640

11 Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331-368 (2007). https://doi.org:10.1146/annurev.genet.40.110405.090448

12 Wood, J. G. & Helfand, S. L. Chromatin structure and transposable elements in organismal aging. *Front Genet* **4**, 274 (2013). https://doi.org:10.3389/fgene.2013.00274

13 Bravo, J. I., Nozownik, S., Danthi, P. S. & Benayoun, B. A. Transposable elements, circular RNAs and mitochondrial transcription in age-related genomic regulation. *Development* **147**, dev175786 (2020). https://doi.org:10.1242/dev.175786

14 Cardelli, M. The epigenetic alterations of endogenous retroelements in aging. *Mech Ageing Dev* **174**, 30-46 (2018). https://doi.org:10.1016/j.mad.2018.02.002

15 De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247-256 (2013). https://doi.org:10.1111/acel.12047

16 De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5**, 867-883 (2013). https://doi.org:10.18632/aging.100621

17 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R. & LaRocca, T. J. Healthy aging interventions reduce repetitive element transcripts. *J Gerontol A Biol Sci Med Sci* (2020). https://doi.org:10.1093/gerona/glaa302

18 Andrenacci, D., Cavaliere, V. & Lattanzi, G. The role of transposable elements activity in aging and their possible involvement in laminopathic diseases. *Ageing Research Reviews* **57**, 100995 (2020). https://doi.org:https://doi.org/10.1016/j.arr.2019.100995
19 Chen, H., Zheng, X., Xiao, D. & Zheng, Y. Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. *Aging Cell* **15**, 542-552 (2016). https://doi.org:10.1111/acel.12465
20 Gorbunova, V. *et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43-53 (2021). https://doi.org:10.1038/s41586-021-03542-y
21 López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-1217 (2013). https://doi.org:10.1016/j.cell.s2013.05.039
22 López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243-278 (2023). https://doi.org:10.1016/j.cell.2022.11.001
23 Hsu, P. S. *et al.* More than causing (epi)genomic instability: emerging physiological implications of transposable element modulation. *J Biomed Sci* **28**, 58 (2021). https://doi.org:10.1186/s12929-021-00754-2
24 Yang, J. H. *et al.* Loss of epigenetic information as a cause of mammalian aging. *Cell* **186**, 305-326.e327 (2023). https://doi.org:10.1016/j.cell.2022.12.027
25 Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**, 1419-1428 (2009). https://doi.org:10.1101/gr.091678.109
26 Choi, J. Y. & Lee, Y. C. G. Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genet* **16**, e1008872 (2020). https://doi.org:10.1371/journal.pgen.1008872
27 Lee, J.-H., Kim, E. W., Croteau, D. L. & Bohr, V. A. Heterochromatin: an epigenetic point of view in aging. *Experimental & Molecular Medicine* **52**, 1466-1474 (2020). https://doi.org:10.1038/s12276-020-00497-4
28 Wood, J. G. *et al.* Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. *Proc Natl Acad Sci U S A* **113**, 11277-11282 (2016). https://doi.org:10.1073/pnas.1604621113
29 Pal, S. & Tyler, J. K. Epigenetics and aging. *Science Advances* **2**, e1600584 (2016). https://doi.org:doi:10.1126/sciadv.1600584
30 Yi, S. J. & Kim, K. New Insights into the Role of Histone Changes in Aging. *Int J Mol Sci* **21** (2020). https://doi.org:10.3390/ijms21218241
31 Hu, Z. *et al.* Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev* **28**, 396-408 (2014). https://doi.org:10.1101/gad.233221.113
32 Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471-476 (2004). https://doi.org:10.1038/nature02651
33 Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**, 615-627 (2011). https://doi.org:10.1038/nrg3030
34 Larson, K. *et al.* Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet* **8**, e1002473 (2012). https://doi.org:10.1371/journal.pgen.1002473
35 Zhang, W. *et al.* Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* **348**, 1160-1163 (2015). https://doi.org:10.1126/science.aaa1356
36 Jansz, N. DNA methylation dynamics at transposable elements in mammals. *Essays Biochem* **63**, 677-689 (2019). https://doi.org:10.1042/ebc20190039

37 Etchegaray, E., Naville, M., Volff, J.-N. & Haftek-Terreau, Z. Transposable element-derived sequences in vertebrate development. *Mobile DNA* **12**, 1 (2021). https://doi.org:10.1186/s13100-020-00229-5
38 Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc Natl Acad Sci U S A* **117**, 19359-19366 (2020). https://doi.org:10.1073/pnas.1921719117
39 Jin, B. & Robertson, K. D. DNA methyltransferases, DNA damage repair, and cancer. *Adv Exp Med Biol* **754**, 3-29 (2013). https://doi.org:10.1007/978-1-4419-9967-2_1
40 Lowe, D., Horvath, S. & Raj, K. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* **7** (2016).
41 Pellegrini, C. *et al.* A Meta-Analysis of Brain DNA Methylation Across Sex, Age, and Alzheimer's Disease Points for Accelerated Epigenetic Aging in Neurodegeneration. *Frontiers in Aging Neuroscience* **13** (2021). https://doi.org:10.3389/fnagi.2021.639428
42 Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature Genetics* **45**, 836-841 (2013). https://doi.org:10.1038/ng.2649
43 Bollati, V. *et al.* Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* **130**, 234-239 (2009). https://doi.org:10.1016/j.mad.2008.12.003
44 Zamudio, N. *et al.* DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev* **29**, 1256-1270 (2015). https://doi.org:10.1101/gad.257840.114
45 Yousefzadeh, M. J. *et al.* An aged immune system drives senescence and ageing of solid organs. *Nature* **594**, 100-105 (2021). https://doi.org:10.1038/s41586-021-03547-7
46 Xiong, H. *et al.* DNA damage response and GATA4 signaling in cellular senescence and aging-related pathology. *Front Aging Neurosci* **14**, 933015 (2022). https://doi.org:10.3389/fnagi.2022.933015
47 Farkash, E. A. & Luning Prak, E. T. DNA damage and L1 retrotransposition. *J Biomed Biotechnol* **2006**, 37285 (2006). https://doi.org:10.1155/jbb/2006/37285
48 Deininger, P. Alu elements: know the SINEs. *Genome Biology* **12**, 236 (2011). https://doi.org:10.1186/gb-2011-12-12-236
49 Patchsung, M. *et al.* Alu siRNA to increase Alu element methylation and prevent DNA damage. *Epigenomics* **10**, 175-185 (2018). https://doi.org:10.2217/epi-2017-0096
50 Adwan-Shekhidem, H. & Atzmon, G. in *Epigenetics of Aging and Longevity* Vol. 4 (eds Alexey Moskalev & Alexander M. Vaiserman) 119-136 (Academic Press, 2018).
51 McGurk, M. P., Dion-Côté, A.-M. & Barbash, D. A. Rapid evolution at the Drosophila telomere: transposable element dynamics at an intrinsically unstable locus. *Genetics* **217**, iyaa027 (2021). https://doi.org:10.1093/genetics/iyaa027
52 Kordyukova, M., Olovnikov, I. & Kalmykova, A. Transposon control mechanisms in telomere biology. *Curr Opin Genet Dev* **49**, 56-62 (2018). https://doi.org:10.1016/j.gde.2018.03.002
53 Pardue, M.-L. & DeBaryshe, P. G. in *Mobile DNA II* 870-887 (2007).
54 McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792-801 (1984). https://doi.org:10.1126/science.15739260
55 Scholes, D. T., Kenny, A. E., Gamache, E. R., Mou, Z. & Curcio, M. J. Activation of a LTR-retrotransposon by telomere erosion. *Proc Natl Acad Sci U S A* **100**, 15736-15741 (2003). https://doi.org:10.1073/pnas.2136609100
56 Morrish, T. A. *et al.* Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446**, 208-212 (2007). https://doi.org:10.1038/nature05560

57 Zhang, X., Zhang, R. & Yu, J. New Understanding of the Relevant Role of LINE-1 Retrotransposition in Human Disease and Immune Modulation. *Frontiers in Cell and Developmental Biology* **8** (2020). https://doi.org:10.3389/fcell.2020.00657
58 Pastuzyn, E. D. *et al.* The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* **172**, 275-288.e218 (2018). https://doi.org:https://doi.org/10.1016/j.cell.2017.12.024
59 Francis, K. & Palsson, B. O. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc Natl Acad Sci U S A* **94**, 12258-12262 (1997). https://doi.org:10.1073/pnas.94.23.12258
60 Ferrucci, L. & Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol* **15**, 505-522 (2018). https://doi.org:10.1038/s41569-018-0064-2
61 Miller, K. N. *et al.* Cytoplasmic DNA: sources, sensing, and role in aging and disease. *Cell* **184**, 5506-5526 (2021). https://doi.org:https://doi.org/10.1016/j.cell.2021.09.034
62 Decout, A., Katz, J. D., Venkatraman, S. & Ablasser, A. The cGAS–STING pathway as a therapeutic target in inflammatory diseases. *Nature Reviews Immunology* **21**, 548-569 (2021). https://doi.org:10.1038/s41577-021-00524-z
63 Sadeq, S., Al-Hashimi, S., Cusack, C. M. & Werner, A. Endogenous Double-Stranded RNA. *Noncoding RNA* **7** (2021). https://doi.org:10.3390/ncrna7010015
64 Lässig, C. & Hopfner, K.-P. Discrimination of cytosolic self and non-self RNA by RIG-I-like receptors. *Journal of Biological Chemistry* **292**, 9000-9009 (2017). https://doi.org:https://doi.org/10.1074/jbc.R117.788398
65 Chen, Y. G. & Hur, S. Cellular origins of dsRNA, their recognition and consequences. *Nature Reviews Molecular Cell Biology* (2021). https://doi.org:10.1038/s41580-021-00430-1
66 Cao, W. IFN-Aging: Coupling Aging With Interferon Response. *Frontiers in Aging* **3** (2022). https://doi.org:10.3389/fragi.2022.870489
67 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78 (2019). https://doi.org:10.1038/s41586-018-0784-9
68 Dhillon, P. *et al.* Increased levels of endogenous retroviruses trigger fibroinflammation and play a role in kidney disease development. *Nat Commun* **14**, 559 (2023). https://doi.org:10.1038/s41467-023-36212-w
69 Dopkins, N. *et al.* Endogenous Reverse Transcriptase Inhibition Attenuates TLR5-Mediated Inflammation. *mBio* **14**, e03280-03222 (2023). https://doi.org:doi:10.1128/mbio.03280-22
70 Liu, X. *et al.* Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell* **186**, 287-304.e226 (2023). https://doi.org:10.1016/j.cell.2022.12.017
71 Rostami, M. R. & Bradic, M. The derepression of transposable elements in lung cells is associated with the inflammatory response and gene activation in idiopathic pulmonary fibrosis. *Mobile DNA* **12**, 14 (2021). https://doi.org:10.1186/s13100-021-00241-3
72 Ochoa, E. *et al.* Pathogenic tau-induced transposable element-derived dsRNA drives neuroinflammation. *Sci Adv* **9**, eabq5423 (2023). https://doi.org:10.1126/sciadv.abq5423
73 Van Meter, M. *et al.* SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nature Communications* **5**, 5011 (2014). https://doi.org:10.1038/ncomms6011
74 Kumari, R. & Jat, P. Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Frontiers in Cell and Developmental Biology* **9** (2021). https://doi.org:10.3389/fcell.2021.645593

75 Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nature Reviews Molecular Cell Biology* **22**, 75-95 (2021). https://doi.org:10.1038/s41580-020-00314-w
76 Schmitz, C. R. R., Maurmann, R. M., Guma, F. T. C. R., Bauer, M. E. & Barbé-Tuana, F. M. cGAS-STING pathway as a potential trigger of immunosenescence and inflammaging. *Frontiers in Immunology* **14** (2023). https://doi.org:10.3389/fimmu.2023.1132653
77 Swanson, E. C., Manning, B., Zhang, H. & Lawrence, J. B. Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence. *J Cell Biol* **203**, 929-942 (2013). https://doi.org:10.1083/jcb.201306073
78 Chang, H. C. & Guarente, L. SIRT1 and other sirtuins in metabolism. *Trends Endocrinol Metab* **25**, 138-145 (2014). https://doi.org:10.1016/j.tem.2013.12.001
79 Wahl, D. *et al.* Aging, lifestyle and dementia. *Neurobiol Dis* **130**, 104481 (2019). https://doi.org:10.1016/j.nbd.2019.104481
80 Zhao, L. *et al.* Sirtuins and their Biological Relevance in Aging and Age-Related Diseases. *Aging Dis* **11**, 927-945 (2020). https://doi.org:10.14336/ad.2019.0820
81 Carafa, V. *et al.* Sirtuin functions and modulation: from chemistry to the clinic. *Clinical Epigenetics* **8**, 61 (2016). https://doi.org:10.1186/s13148-016-0224-3
82 Taniguchi, R. *et al.* SIRT1 stabilizes extrachromosomal gene amplification and contributes to repeat-induced gene silencing. *Journal of Biological Chemistry* **296**, 100356 (2021). https://doi.org:https://doi.org/10.1016/j.jbc.2021.100356
83 Ge, Y., Zhou, M., Chen, C., Wu, X. & Wang, X. Role of AMPK mediated pathways in autophagy and aging. *Biochimie* **195**, 100-113 (2022). https://doi.org:https://doi.org/10.1016/j.biochi.2021.11.008
84 Zhang, W., Qu, J., Liu, G.-H. & Belmonte, J. C. I. The ageing epigenome and its rejuvenation. *Nature Reviews Molecular Cell Biology* **21**, 137-150 (2020). https://doi.org:10.1038/s41580-019-0204-5
85 Romero, M. A. *et al.* Five months of voluntary wheel running downregulates skeletal muscle LINE-1 gene expression in rats. *Am J Physiol Cell Physiol* **317**, C1313-c1323 (2019). https://doi.org:10.1152/ajpcell.00301.2019
86 Hunt, N. J. *et al.* Targeting the liver in dementia and cognitive impairment: Dietary macronutrients and diabetic therapeutics. *Adv Drug Deliv Rev* **190**, 114537 (2022). https://doi.org:10.1016/j.addr.2022.114537
87 Cascella, R., Fani, G., Bigi, A., Chiti, F. & Cecchi, C. Partial Failure of Proteostasis Systems Counteracting TDP-43 Aggregates in Neurodegenerative Diseases. *Int J Mol Sci* **20** (2019). https://doi.org:10.3390/ijms20153685
88 LaRocca, T. J., Mariani, A., Watkins, L. R. & Link, C. D. TDP-43 knockdown causes innate immune activation via protein kinase R in astrocytes. *Neurobiol Dis* **132**, 104514 (2019). https://doi.org:10.1016/j.nbd.2019.104514
89 Terry, D. M. & Devine, S. E. Aberrantly High Levels of Somatic LINE-1 Expression and Retrotransposition in Human Neurological Disorders. *Frontiers in Genetics* **10** (2020). https://doi.org:10.3389/fgene.2019.01244
90 Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* **134**, 587-598 (2008). https://doi.org:10.1016/j.cell.2008.06.032
91 Chistiakov, D. A., Sobenin, I. A., Revin, V. V., Orekhov, A. N. & Bobryshev, Y. V. Mitochondrial aging and age-related dysfunction of mitochondria. *Biomed Res Int* **2014**, 238463 (2014). https://doi.org:10.1155/2014/238463
92 Chen, Y., Parker, W. D., Chen, H. & Yang, K. Aberrant mitochondrial RNA in the role of aging and aging associated diseases. *Medical Hypotheses* **85**, 178-182 (2015). https://doi.org:https://doi.org/10.1016/j.mehy.2015.04.022

93 Tarallo, V. *et al.* DICER1 Loss and Alu RNA Induce Age-Related Macular Degeneration via the NLRP3 Inflammasome and MyD88. *Cell* **149**, 847-859 (2012). https://doi.org:https://doi.org/10.1016/j.cell.2012.03.036
94 Larsen, P. A. *et al.* The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. *Alzheimers Dement* **13**, 828-838 (2017). https://doi.org:10.1016/j.jalz.2017.01.017
95 Gerdes, P. *et al.* Retrotransposon instability dominates the acquired mutation landscape of mouse induced pluripotent stem cells. *Nature Communications* **13**, 7470 (2022). https://doi.org:10.1038/s41467-022-35180-x
96 Gruhn, W. H. *et al.* Epigenetic resetting in the human germ line entails histone modification remodeling. *Science Advances* **9**, eade1257 https://doi.org:10.1126/sciadv.ade1257
97 Jin, L. *et al.* Transposable elements activation triggers necroptosis in mouse embryonic stem cells. *Cell Death Dis* **14**, 184 (2023). https://doi.org:10.1038/s41419-023-05705-3
98 Sousa-Victor, P. *et al.* Piwi Is Required to Limit Exhaustion of Aging Somatic Stem Cells. *Cell Rep* **20**, 2527-2537 (2017). https://doi.org:10.1016/j.celrep.2017.08.059
99 Lin, K.-Y. *et al.* Piwi reduction in the aged niche eliminates germline stem cells via Toll-GSK3 signaling. *Nature Communications* **11**, 3147 (2020). https://doi.org:10.1038/s41467-020-16858-6
100 Mosaddeghi, P., Farahmandnejad, M. & Zarshenas, M. M. The role of transposable elements in aging and cancer. *Biogerontology* (2023). https://doi.org:10.1007/s10522-023-10028-z
101 Jönsson, M. E., Garza, R., Johansson, P. A. & Jakobsson, J. Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders. *Trends in Genetics* **36**, 610-623 (2020). https://doi.org:https://doi.org/10.1016/j.tig.2020.05.004
102 Reilly, M. T., Faulkner, G. J., Dubnau, J., Ponomarev, I. & Gage, F. H. The role of transposable elements in health and diseases of the central nervous system. *J Neurosci* **33**, 17577-17586 (2013). https://doi.org:10.1523/jneurosci.3369-13.2013
103 Tam, O. H., Ostrow, L. W. & Gale Hammell, M. Diseases of the nERVous system: retrotransposon activity in neurodegenerative disease. *Mob DNA* **10**, 32 (2019). https://doi.org:10.1186/s13100-019-0176-1
104 Terry, D. M. & Devine, S. E. Aberrantly High Levels of Somatic LINE-1 Expression and Retrotransposition in Human Neurological Disorders. *Front Genet* **10**, 1244 (2019). https://doi.org:10.3389/fgene.2019.01244
105 Tarallo, V. *et al.* DICER1 loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. *Cell* **149**, 847-859 (2012). https://doi.org:10.1016/j.cell.2012.03.036
106 Liu, E. Y. *et al.* Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep* **27**, 1409-1421.e1406 (2019). https://doi.org:10.1016/j.celrep.2019.04.003
107 Prudencio, M. *et al.* Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. *Hum Mol Genet* **26**, 3421-3431 (2017). https://doi.org:10.1093/hmg/ddx233
108 Tam, O. H. *et al.* Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Rep* **29**, 1164-1177.e1165 (2019). https://doi.org:10.1016/j.celrep.2019.09.066
109 Guo, C. *et al.* Tau Activates Transposable Elements in Alzheimer's Disease. *Cell Rep* **23**, 2874-2880 (2018). https://doi.org:10.1016/j.celrep.2018.05.004
110 Wahl, D. *et al.* The reverse transcriptase inhibitor 3TC protects against age-related cognitive dysfunction. *Aging Cell*, e13798 (2023). https://doi.org:10.1111/acel.13798

111 Kitkumthorn, N. & Mutirangura, A. Long interspersed nuclear element-1 hypomethylation in cancer: biology and clinical applications. *Clin Epigenetics* **2**, 315-330 (2011). https://doi.org:10.1007/s13148-011-0032-8
112 Roulois, D. *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973 (2015). https://doi.org:10.1016/j.cell.2015.07.056
113 Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986 (2015). https://doi.org:10.1016/j.cell.2015.07.011
114 Copley, K. E. & Shorter, J. Repetitive elements in aging and neurodegeneration. *Trends Genet* **39**, 381-400 (2023). https://doi.org:10.1016/j.tig.2023.02.008
115 Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab* **29**, 871-885.e875 (2019). https://doi.org:10.1016/j.cmet.2019.02.014
116 Li, M. *et al.* Lamivudine improves cognitive decline in SAMP8 mice: Integrating in vivo pharmacological evaluation and network pharmacology. *J Cell Mol Med* **25**, 8490-8503 (2021). https://doi.org:10.1111/jcmm.16811
117 Lin, H. *et al.* Zidovudine-Mediated Autophagy Inhibition Enhances Mitochondrial Toxicity in Muscle Cells. *Antimicrob Agents Chemother* **63** (2019). https://doi.org:10.1128/aac.01443-18
118 Note, R. *et al.* Mitochondrial and metabolic effects of nucleoside reverse transcriptase inhibitors (NRTIs) in mice receiving one of five single- and three dual-NRTI treatments. *Antimicrob Agents Chemother* **47**, 3384-3392 (2003). https://doi.org:10.1128/aac.47.11.3384-3392.2003
119 Mitchell, S. J. *et al.* Effects of Sex, Strain, and Energy Intake on Hallmarks of Aging in Mice. *Cell Metab* **23**, 1093-1112 (2016). https://doi.org:10.1016/j.cmet.2016.05.027
120 Green, C. D. *et al.* Impact of Dietary Interventions on Noncoding RNA Networks and mRNAs Encoding Chromatin-Related Factors. *Cell Rep* **18**, 2957-2968 (2017). https://doi.org:10.1016/j.celrep.2017.03.001
121 Kim, C. H. *et al.* Short-term calorie restriction ameliorates genomewide, age-related alterations in DNA methylation. *Aging Cell* **15**, 1074-1081 (2016). https://doi.org:10.1111/acel.12513
122 Wahl, D., Cavalier, A. N. & LaRocca, T. J. Novel Strategies for Healthy Brain Aging. *Exerc Sport Sci Rev* **49**, 115-125 (2021). https://doi.org:10.1249/jes.0000000000000242
123 Romero, M. A. *et al.* Acute and chronic resistance training downregulates select LINE-1 retrotransposon activity markers in human skeletal muscle. *Am J Physiol Cell Physiol* **314**, C379-c388 (2018). https://doi.org:10.1152/ajpcell.00192.2017
124 da Silva Rodrigues, G. *et al.* Exercise training modifies the whole blood DNA methylation profile in middle-aged and older women. *J Appl Physiol (1985)* **134**, 610-621 (2023). https://doi.org:10.1152/japplphysiol.00237.2022
125 Cui, D. & Xu, X. DNA Methyltransferases, DNA Methylation, and Age-Associated Cognitive Function. *Int J Mol Sci* **19** (2018). https://doi.org:10.3390/ijms19051315
126 Oliveira, A. M., Hemstedt, T. J. & Bading, H. Rescue of aging-associated decline in Dnmt3a2 expression restores cognitive abilities. *Nat Neurosci* **15**, 1111-1113 (2012). https://doi.org:10.1038/nn.3151
127 Liu, L. *et al.* Insufficient DNA methylation affects healthy aging and promotes age-related health problems. *Clinical Epigenetics* **2**, 349-360 (2011). https://doi.org:10.1007/s13148-011-0042-6
128 Wahl, D., Anderson, R. M. & Le Couteur, D. G. Antiaging Therapies, Cognitive Impairment, and Dementia. *J Gerontol A Biol Sci Med Sci* **75**, 1643-1652 (2020). https://doi.org:10.1093/gerona/glz135

129 Guo H, Chitiprolu M, Gagnon D, Meng L, Perez-Iratxeta C, Lagace D, Gibbings D. Autophagy supports genomic stability by degrading retrotransposon RNA. Nat Commun. 2014 Nov 4;5:5276. doi: 10.1038/ncomms6276. PMID: 25366815.
130 Black HH, Hanson JL, Roberts JE, Leslie SN, Campodonico W, Ebmeier CC, Holling GA, Tay JW, Matthews AM, Ung E, Lau CI, Whiteley AM. UBQLN2 restrains the domesticated retrotransposon PEG10 to maintain neuronal health in ALS. Elife. 2023 Mar 23;12:e79452. doi: 10.7554/eLife.79452. PMID: 36951542; PMCID: PMC10076021.

CHAPTER 2: NOVEL WHOLE BLOOD TRANSCRIPTOMIC PREDICTORS OF CHANGES IN MAXIMAL AEROBIC CAPACITY IN RESPONSE TO ENDURANCE EXERCISE TRAINING IN HEALTHY WOMEN

Introduction

Maximal aerobic capacity ($\dot{V}O_2max$) is one of the strongest independent predictors of morbidity and mortality[1], and aerobic (endurance) exercise training is the primary strategy for increasing $\dot{V}O_2max$[2]. However, improvements in maximal aerobic capacity ($\Delta\dot{V}O_2max$) in response to endurance exercise training vary widely among individuals exposed to the same standardized training program[3-5]. Despite extensive interest in this phenomenon, the specific biological mechanisms associated with this inter-individual variability remain largely unknown. As a result, there is significant interest in using "omics" approaches to identify novel biological mechanisms and potential predictors of $\Delta\dot{V}O_2max$[6,7].

Ideally, omics signatures of $\Delta\dot{V}O_2max$ would provide insight on mechanisms of inter-individual variability and be both reproducible across studies and detectable in easily collected samples. For example, recent data demonstrate that plasma proteome signatures related to muscle physiology differ in subjects who experience greater vs. lesser $\Delta\dot{V}O_2max$ in response to training, and that these signatures can be used to predict $\Delta\dot{V}O_2max$ in multiple datasets[8]. Similar insight has not been achieved at the upstream gene expression level, in part because transcriptome studies of $\Delta\dot{V}O_2max$ have been somewhat limited in scope and based on microarrays[3,9,10]. However, newer and more comprehensive transcriptomics techniques could help identify novel gene expression signatures that provide insight in this context. Here, we report on a pilot transcriptome study to identify gene expression signatures of $\Delta\dot{V}O_2max$ using comprehensive high-throughput sequencing. We performed RNA-seq on easily accessible whole blood samples, which have the potential to reflect systemic inter-individual differences in biology[11], obtained from a randomized controlled trial of aerobic exercise training in which we also measured $\Delta\dot{V}O_2max$ in response to lower vs. higher training volume and intensity. The

goals of our analyses were to determine if differences in baseline gene expression: 1) are associated with greater vs. lesser $\Delta\dot{V}O_2$max in response to exercise training; 2) are modulated by exercise training volume and intensity; 3) provide insight on potential mechanisms of inter-individual differences in $\Delta\dot{V}O_2$max; and 4) have the potential to predict $\dot{V}O_2$max responses to exercise training.

Methods

Exercise intervention trial

Samples for this study were obtained from a randomized clinical trial (NCT02032628, clinicaltrials.gov) designed to evaluate the protective effects of different durations and intensities of exercise training on breast cancer risk factors, described in detail elsewhere[12,13]. The study was approved by the Institutional Review Board at the University of Colorado Boulder and performed in accordance with the Declaration of Helsinki. Briefly, healthy (body mass index [BMI] < 29.0 kg/m^2 and absence of metabolic or cardiovascular disease), sedentary (<60 min/week aerobic exercise) women aged 30-45 years visited the laboratory after an overnight fast for baseline venous blood draws (collected in PaxGene RNA tubes, BD Biosciences) and then $\dot{V}O_2$max measurements via incremental treadmill testing (Balke protocol). Participants were randomized to one of four aerobic exercise training (treadmill walking or running) groups for a 16-week supervised intervention: 1) higher volume/higher intensity (40 min at 75-85% $\dot{V}O_2$max); 2) higher volume/lower intensity (40 min at 55-65% $\dot{V}O_2$max); 3) lower volume/higher intensity (20 min at 75-85% $\dot{V}O_2$max); and 4) lower volume/lower intensity (20 min at 55-65% $\dot{V}O_2$max). Intensity prescriptions were based on current guidelines and calculated using the heart rate (HR) reserve method to estimate %$\dot{V}O_2$max[14,15], and HR monitors were used to ensure maintenance of HR within ±5% of target intensity. Participants visited the laboratory to complete four supervised training sessions per week, and attendance and target intensity

adherence were documented. Blood sampling and $\dot{V}O_2$max testing were repeated 48 hours after the final exercise training session to avoid acute effects on the transcriptome. For the present analyses, we selected 15 participants from the larger study with robust improvements in $\dot{V}O_2$max (RR, $\Delta\dot{V}O_2$max ≥ 15%) and 15 with minimal/no change or a modest decrease in $\dot{V}O_2$max (NR, $\Delta\dot{V}O_2$max ≤ 5%), with both groups including subjects from each of the four exercise training conditions. Sample size for this pilot study was chosen based on costs and reports from our lab and others demonstrating gene expression differences in peripheral blood cells with n ≤ 10/group[9,10,16]. Group composition was matched for phenotypic characteristics including age, BMI, systolic and diastolic blood pressure, training-induced weight change, training adherence, voluntary effort (maximal rate of perceived exertion [RPE]) and objective measures of effort (maximal respiratory exchange ratio [RER]) in $\dot{V}O_2$max tests.

RNA extraction, sequencing and bioinformatics analyses

Poly(A) RNA-seq was performed as previously described[16]. Briefly, RNA was recovered from PaxGene blood collection tubes using PaxGene-specific isolation kits (Qiagen), and RNA libraries were generated using Illumina TruSeq kits and sequenced on an Illumina NovaSeq 6000 platform to produce >40M 151-bp, paired-end reads per sample. Differential gene expression was analyzed using standard techniques as previously reported[16]. Reads were trimmed and quality filtered with *fastp*[17] then aligned to the human genome (hg38 *Homo sapiens*) using the STAR aligner[18] with --outFilters adjusted to maximize unique mapping and reads mapping to globins were removed prior to downstream analyses. Differentially expressed genes were then identified using DESeq2[19] for R. Gene ontology (GO) analyses were performed with g:Profiler[20], and the most specific GO modules were identified as terminal GO terms in the output produced using the GO.db R program. These GO modules were collapsed to identify key processes using REVIGO[21] prior to process mapping/clustering with the Enrichment Map program in Cytosape[22], and hub names were simplified manually for clarity.

Statistical analyses

Significantly differentially expressed genes (DEGs) were identified using a false discovery cutoff (FDR) of 0.05 and GO enrichment statistics were extracted from g:Profiler output. Group differences in RR/NR gene sets were analyzed by Mann-Whitney t-test vs. mean Log2 Fold-Difference for all transcripts within the same group using GraphPad Prism. JMP software was used to calculate simple Pearson correlations among the top 5000 DEGs in data from this and the second study analyzed (described below), and for regression analyses of gene/transcript predictors of $\Delta\dot{V}O_2$max. For data from the present study, an unbiased LASSO regression model was used to avoid overfitting with all DEGs, a normal response distribution and AICc validation. The same predictor genes identified in this analysis were entered into a forward stepwise regression model with JMP default p-value thresholds of 0.25 to enter and 0.10 to leave in secondary dataset analyses.

Study data

Transcriptome data for this study are available on the Gene Expression Omnibus (GEO) website under accession number GSE206505, and DEGs from all analyses are presented in **Supplementary Data** (URL: https://figshare.com/articles/dataset/Supplementary_materials_5_25_23/23202659, DOI: https://doi.org/10.6084/m9.figshare.23202659). Previously published RNA-seq data on gene expression changes in muscle with exercise training are available under GEO accession number GSE97084, and related $\Delta\dot{V}O_2$max data (shared by authors M.M.R. and K.S.N.) are available in Supplementary Data.

Results

As described above, in the parent study, participants were randomly assigned to one of four training groups for a supervised, 16-week aerobic exercise intervention (**Figure 2.1A**). 219

women enrolled in the study, and ~55% completed all post-testing (**Supplementary Data**). Details on the overall study, including attrition analyses showing no substantial differences among participants have been reported on elsewhere[12,13]. At baseline, there were no significant differences in $\dot{V}O_2$max among the four training groups. $\dot{V}O_2$max increased with training in all groups in an exercise intensity-dependent manner, consistent with previous reports[2,4]. However, within each condition and the overall group, we observed individual training-induced $\Delta\dot{V}O_2$max values that ranged from robust increases to little/no change or modest decreases (**Figure 2.1B**), consistent with other published data[3,9,10,23].

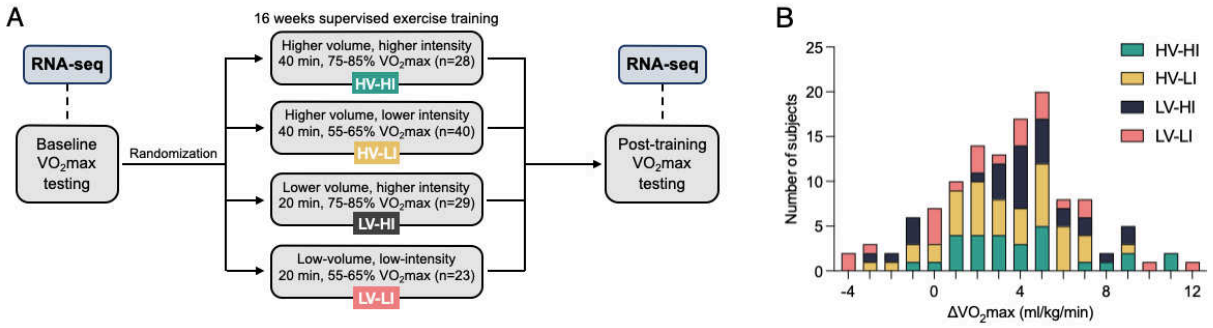


Figure 2.1. Aerobic exercise training dose-response study design, and inter-individual variability in $\Delta\dot{V}O_2$max with exercise training. (A) Subjects completed baseline $\dot{V}O_2$max tests and blood sampling and were randomized to training groups. Following 16 weeks of supervised exercise training, post-training $\dot{V}O_2$max testing and blood sampling were performed. RNA-seq analyses were performed on whole blood samples collected prior to baseline $\dot{V}O_2$max tests and 48 hours after post-training tests. **(B)** Aerobic exercise training-induced $\Delta\dot{V}O_2$max in all 120 subjects who completed the study. HV-HI, high-volume/high-intensity; HV-LI, high-volume/low-intensity; LV-HI, low-volume/high-intensity; LV-LI, low-volume/low-intensity.

Differences in $\Delta\dot{V}O_2$max in the overall study were not related to baseline $\dot{V}O_2$max or other clinical variables. Therefore, to gain insight into the biological mechanisms underlying this inter-individual variability, we selected a subset of subjects including 15 participants with robust $\Delta\dot{V}O_2$max responses (RR, $\Delta\dot{V}O_2$max ≥ 15%) and 15 with minimal/no response (NR, $\Delta\dot{V}O_2$max ≤ 5%) for further analyses. RR and NR subgroups had similar clinical characteristics and age (36 and 39 years, respectively) at baseline (**Table 2.1**). However, the RR group demonstrated a mean $\Delta\dot{V}O_2$max of +19.1% in response to training, whereas $\dot{V}O_2$max did not change in the NR

group (mean $\Delta\dot{V}O_2max$ = +0.003%). The lack of increase in $\dot{V}O_2max$ with training in the NR group was observed despite the group having ~13% lower baseline $\dot{V}O_2max$ compared with the RR group, which, per the law of initial values, should have exerted the effect of increasing $\dot{V}O_2max$ to a greater extent with training. In other words, subjects with lower baseline $\dot{V}O_2max$ would have been expected to experience greater improvements than those with higher baseline values, as reported elsewhere[24]. Importantly, $\Delta\dot{V}O_2max$ in these groups was only modestly and positively related to baseline $\dot{V}O_2max$ (R^2 = 0.19), and group differences were not due to: different training programs (subjects were distributed similarly across all four training groups); differences in adherence, as all subjects completed >85% of all prescribed training sessions (average of 89.1% and 86.3% for RR and NR, respectively); differences in voluntary effort during treadmill testing (as indicated by maximal HR, RPE and RER); or differential changes in clinical or physiological markers with training (e.g., BMI, resting blood pressure), all of which were similar before and after training among groups.

Table 2.1. Effects of 16-week aerobic exercise intervention in all RR and NR.

	Responders		**Non-Responders**	
Age (years)	36 ± 5		39 ± 4	
	Pre	**Post**	**Pre**	**Post**
BMI (kg/m^2)	26.1 ± 3.4	25.8 ± 3.1	28.1 ± 5.5	27.7 ± 5.5
Systolic BP (mmHg)	115 ± 7	115 ± 7	116 ± 7	117 ± 10
Diastolic BP (mmHg)	72 ± 7	72 ± 8	72 ± 7	74 ± 7
$\dot{V}O_2max$ (ml/kg/min)	32.7 ± 4.8	38.0 ± 5.9*	28.4 ± 4.4†	28.8 ± 3.8†
Max HR (bpm)	184 ± 8	181 ± 9*	178 ± 12	175 ± 13*
Max RPE (1-10 scale)	8.9 ± 1.5	7.9 ± 1.6	9.0 ± 1.3	8.4 ± 1.5
Max RER ($VCO_2/\dot{V}O_2$)	1.1 ± 0.0	1.1 ± 0.1	1.1 ± 0.3	1.2 ± 0.3

Values are means ± SD.
*P <0.05 vs. Pre within same group.
†P <0.05 vs. RR within same condition.

Despite RR and NR subjects being matched for age, BMI, and resting blood pressure, we found significant transcriptome differences at baseline (i.e., prior to training, **Figure 2.2A**), with >700 DEGs having higher expression and >1400 DEGs have lower expression in RR vs. NR (**Supplementary Data**). To confirm that these differences were specific to $\Delta\dot{V}O_2max$ (and

not simply due to gene expression variability among all subjects or library preparation artifacts), we compared transcriptomes in subjects with higher vs. lower baseline $\dot{V}O_2$max (top vs. bottom 50% of all subjects), and we found few differences. In RR vs. NR at baseline, GO analyses showed that DEGs were associated with numerous biological processes relevant to exercise physiology. The most specific of these cellular processes included chromatin/epigenetic processes and innate immune/inflammatory signaling (increased), as well as numerous signatures of redox and mitochondrial function (decreased) (**Figure 2.2B**). Increased DEGs in RR vs. NR were also associated with additional processes related to epigenetics and gene expression (e.g., histone modifications), whereas decreased DEGs were associated with multiple signatures of protein translation and metabolism. Enrichment mapping of these GO terms together revealed one cluster of immune/inflammation signaling linked with most processes related to the control of gene expression, as well as multiple mitochondria- and translation-related processes that were largely separate from this cluster (**Figure 2.2C**). These findings could suggest that inflammatory gene expression in the RR phenotype may be due to epigenetic events (e.g., differences in histone or DNA modifications that can result in changes in gene transcription)[25], and perhaps that separate events underlie reduced transcription and translation of mitochondria-related genes in RR. In any case, consistent with the idea that these gene expression differences we observed may be biologically relevant, we also found that DEGs in RR vs. NR at baseline were dose-dependently modulated by exercise training in all subjects (**Figure 2.2D**). That is, with higher training intensity and volume, the significantly “up” DEGs in RR at baseline were more reduced, whereas the significantly “down” DEGs in RR at baseline were more increased.

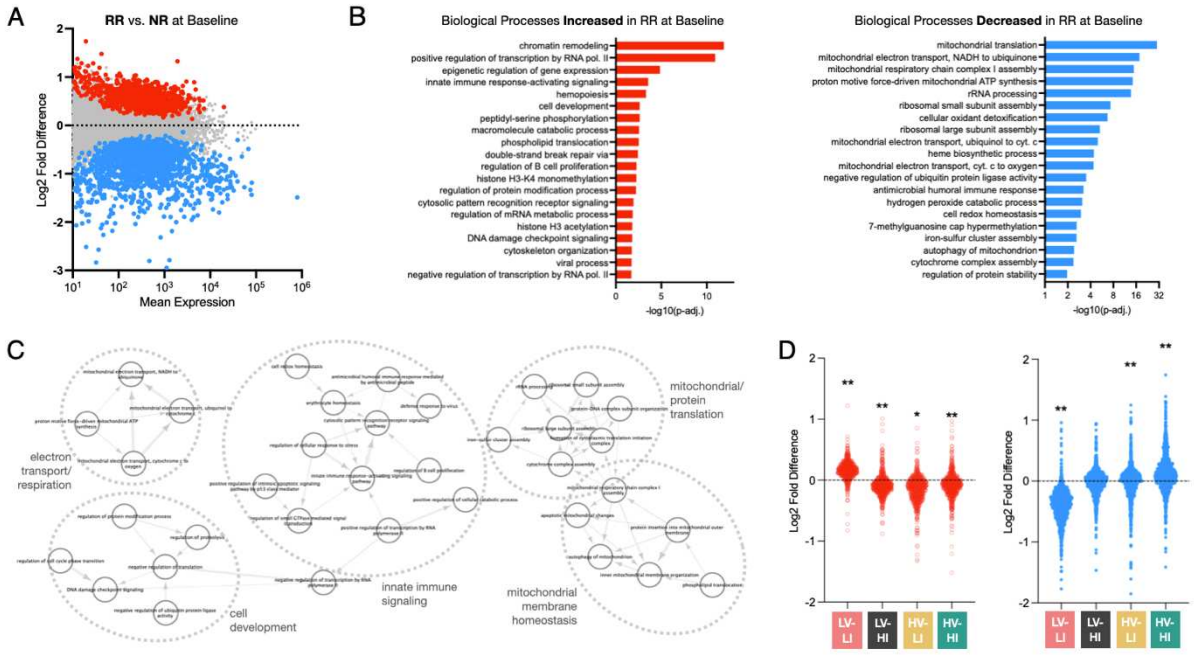


Figure 2.2. Transcriptomic differences between RR and NR before training. (A) MA plot showing relative gene/transcript expression detected by RNA-seq on whole blood samples in RR vs. NR before exercise interventions (red points, significantly greater; blue points significantly lower; FDR<0.05). **(B)** The most specific biological processes related to increased and decreased transcripts before training (i.e., those enriched or depleted in RR vs. NR). **(C)** Simplified enrichment map of all GO/gene expression. **(D)** Effects of the different aerobic exercise training interventions on genes/transcripts significantly increased (left) and decreased (right) in RR vs. NR at baseline. *p<0.05, **p<0.01 vs. average Log2Fold change within each training group.

Finally, to determine which gene expression patterns in RR vs. NR might be reproducibly associated with $\Delta\dot{V}O_2max$, including in different tissues, we compared our findings to RNA-seq data from a previously published study of skeletal muscle transcriptome responses to exercise training[26]. In the original study, the authors did not report on RR vs. NR phenotypes; however, gene expression changes with exercise training included differences in transcripts related to inflammation, mitochondria and protein translation[26]. In data shared by the authors, we also found that aerobic training was associated with a range of $\Delta\dot{V}O_2max$ responses that were not related to baseline $\dot{V}O_2max$ (**Supplementary Data**). Therefore, we asked which of the most increased/decreased genes in RR vs. NR in the present study were related in either direction to

$\Delta\dot{V}O_2$max in this second, published dataset. Overall, we found a greater number of individual genes/transcripts that were negatively vs. positively related to $\Delta\dot{V}O_2$max in both datasets (**Figure 2.3A**). Moreover, whereas genes/transcripts positively related to $\Delta\dot{V}O_2$max in both datasets were weakly associated with generic GO biological processes, transcripts negatively related to $\Delta\dot{V}O_2$max strongly reflected mitochondrial function and protein translation (**Figure 2.3B**). Additionally, in multiple regression analyses (LASSO penalized regression model to avoid overfitting[3,8]), we were able to identify a regression model based on 14 transcripts that strongly predicted $\Delta\dot{V}O_2$max in the present study (R^2 = 0.89, **Figure 2.3C and Supplementary Data**), and only 4 of the transcripts in this model were necessary to predict $\Delta\dot{V}O_2$max in the second dataset using stepwise regression (R^2 = 0.65, **Figure 2.3C**). Interestingly, genes/transcripts that contributed to both prediction models included NEB, a sarcomere contractile protein linked with mitochondrial function[27], and TNFRSF10B, which has been implicated in cell death-associated mitochondrial dysfunction[28]. Taken together, these findings suggest that mitochondria/protein translation-related gene expression may be an important systemic mediator of $\dot{V}O_2$max responses to training, and that this may even be reflected in the circulating transcriptome.

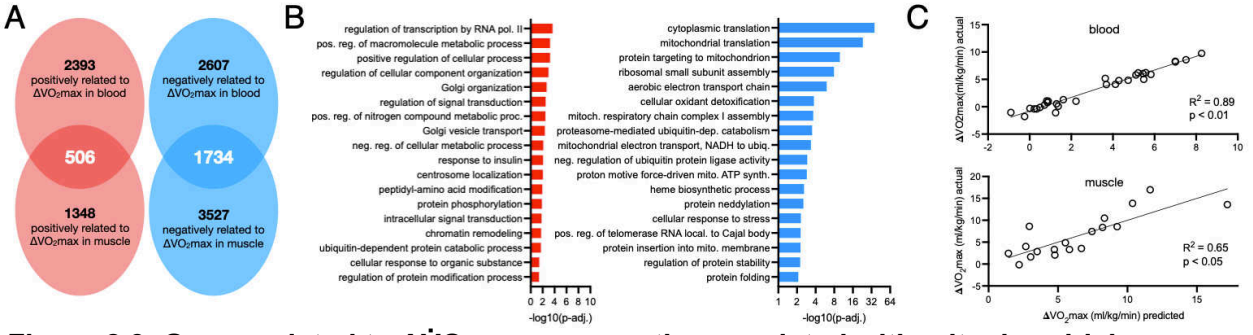


Figure 2.3. Genes related to $\Delta\dot{V}O_2$max are mostly associated with mitochondrial function/translation and can be used to predict $\dot{V}O_2$max training responses. **(A)** Genes/transcripts among the top 5000 DEGs in RR vs. NR that were positively and negatively related to $\Delta\dot{V}O_2$max in both the present study and a second dataset (muscle biopsies from young and older adults, n=18, males and females)[26]. **(B)** The most specific biological processes reflected by transcripts positively and negatively related to $\Delta\dot{V}O_2$max in both datasets (in panel A overlap regions). **(C)** Predicted vs. actual change in $\dot{V}O_2$max with exercise based on a regression model derived from genes differentially expressed in RR vs. NR at baseline in blood (present study, top), and a subset of the same genes/transcripts used to predict $\Delta\dot{V}O_2$max in the second dataset on muscle (bottom).

Discussion

$\dot{V}O_2$max is a powerful predictor of morbidity and mortality risk, and a strong indicator of overall functional capacity[1]. As such, understanding inter-individual variability in exercise training-induced $\Delta\dot{V}O_2$max is an important biomedical research goal[3,7]. In this pilot study, leveraging samples from an exercise intervention trial with a dose-response paradigm[12,13], we present an RNA-seq analysis of $\Delta\dot{V}O_2$max transcriptome signatures in whole blood. Our key findings are that: 1) gene expression signatures at baseline are different in subjects who respond to training with marked vs. little/no increase in $\dot{V}O_2$max; 2) these genes/transcripts are associated with biological processes relevant to cardiorespiratory fitness, most notably mitochondrial/protein translation-related gene expression; 3) greater exercise dose (duration and intensity) modulates $\Delta\dot{V}O_2$max-related transcripts to a greater degree; and 4) these pre-training transcriptome differences can be used to predict $\Delta\dot{V}O_2$max among individuals. Our findings contribute to the evolving literature on this clinically relevant topic.

Early evidence of inter-individual variability in training-induced $\Delta\dot{V}O_2$max came from small exercise intervention trials[4,29], and findings were later confirmed in larger multi-center trials[3,5] that led to genome-wide association studies (GWAS) of $\Delta\dot{V}O_2$max predictors[23]. However, genes/variants linked with $\Delta\dot{V}O_2$max in these reports were not always clearly related to physiological determinants of cardiorespiratory fitness or reproducible in other settings[3,23]. As a result, in part, recent work has focused on “omics” approaches that may provide more direct insight into the biology of $\Delta\dot{V}O_2$max variability[3,7]. For example, Robbins et al.[8] recently identified plasma proteome signatures associated with $\dot{V}O_2$max and exercise training-induced $\Delta\dot{V}O_2$max, and several specific proteins were related to muscle physiology and predictive of $\Delta\dot{V}O_2$max in other datasets. Most work at the upstream, gene expression level has not been as definitive, in part because early transcriptome studies of $\Delta\dot{V}O_2$max were small and based on transcript-limited microarrays[3,9,10]. As such, our findings may be an important extension on the current literature. We used RNA-seq to profile the entire transcriptome in blood samples from a clinical

trial of aerobic exercise training in healthy, middle-aged, previously sedentary women, a historically under-studied group in exercise-related omics research. Also, in contrast to previous studies using one standardized dose of exercise, our supervised intervention trial included two different training volumes and intensities (four groups, fully crossed). We were able to leverage this trial design to identify novel transcriptomic signatures of $\Delta\dot{V}O_2$max. In fact, we found several thousand genes/transcripts that were differentially expressed before exercise training in $\Delta\dot{V}O_2$max responders. These transcripts were associated with biological processes including inflammation and mitochondrial function (both relevant to cardiorespiratory fitness[30]) and modulated by aerobic exercise training in a dose-dependent fashion. Interestingly, most adverse processes like inflammation/immune signaling and DNA damage responses (which could reflect exercise-associated repair processes) were enriched in RR vs. NR at baseline, whereas mitochondria/oxidative metabolism-related processes (which play a key role in aerobic fitness) and protein translation were mostly decreased. Reduced expression of mitochondria/protein translation-related genes was also the dominant pre-training transcriptome signature of $\Delta\dot{V}O_2$max, both in our study and a second RNA-seq dataset based on muscle. The implications of these findings are unclear, but one possible interpretation is that gene expression may be more dynamic and responsive to interventions in RR. It also is possible that exercise can induce greater increases in mitochondria and translation-related transcripts when they are expressed at lower levels before training. This idea would be consistent with many other studies documenting the importance of increased mitochondrial function and protein translation in the physiological adaptations to exercise[30,31], including the secondary dataset we analyzed here[26]. Future studies are needed to confirm these transcriptomic differences, including in different subject populations (e.g., males, older adults, etc.) and in response to different training approaches.

Interestingly, we found no gene expression signatures associated with baseline $\dot{V}O_2$max, and aside from the differences observed in $\Delta\dot{V}O_2$max-related transcripts, the exercise interventions in our study were associated with mostly modest changes in gene expression

(which is consistent with previous reports). These observations further underscore the transcriptomic differences we observed in RR vs. NR at baseline and suggest that inflammation- and especially mitochondria/protein translation-related gene expression in blood may be central to inter-individual variability in training-induced $\Delta\dot{V}O_2$max. In this context, one limitation of our study is that the parent clinical trial did not include biomarker sampling (e.g., for circulating cytokines, myokines or metabolites related to mitochondria and protein synthesis), which could be a promising direction for future work.

A long-term goal of omics-based investigations of inter-individual variability in $\Delta\dot{V}O_2$max is to identify reproducible predictors of exercise training responsiveness. These could be useful for individualizing exercise prescriptions (i.e., "precision exercise medicine")[7], especially given that baseline $\dot{V}O_2$max does not always predict responses to training, and/or in settings where exercise testing is not safe or feasible. Circulating predictors could be particularly helpful in this context, as they are easy to access and have the potential to reflect signals from multiple tissues. Here, we found that pre-training gene expression signatures in whole blood from previously sedentary females could be used to predict training-induced $\Delta\dot{V}O_2$max more accurately than baseline $\dot{V}O_2$max measurements (R^2 = 0.89 vs. 0.19). A subset of the predictor genes/transcripts we found could also be used to predict $\Delta\dot{V}O_2$max in a second RNA-seq dataset based on muscle that included both males and females of different ages. Although most RNA in such samples would be expected to originate from white blood cells, these results suggest that with comprehensive RNA-seq analyses of the whole blood transcriptome, it may be possible to detect transcripts that reflect systemic differences in the biological processes contributing to exercise training responsiveness. In this context, futures studies could benefit from: 1) including RNA-seq on blood and other tissues from the same subjects to identify tissue-specific transcripts; and/or 2) specifically examining the contribution of cellular vs. circulating RNAs (e.g., in exosomes, which are thought to mediate inter-organ communication, or by using dedicated cell-free RNA blood collection tubes). Future work could also investigate upstream

mediators of transcriptomic differences in RR vs. NR, perhaps using multi-omics approaches to also profile the epigenome, especially given our observation that epigenetics-related transcriptome modules were altered in RR vs. NR. In these efforts, demonstrating reproducibility will be critical, especially for transcriptome-based $\Delta\dot{V}O_2max$ predictors, as RNA-seq studies involve many methods-related variables that may differ among laboratories and the technique itself is not quantitative *per se*. However, progress in this area would contribute significantly to our understanding of the mechanisms underlying inter-individual variability in the response to exercise training, and potentially to our ability to develop personalized exercise prescriptions for improving health and reducing disease risk.

REFERENCES

1 Ross, R. *et al.* Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* **134**, e653-e699 (2016). https://doi.org:10.1161/CIR.0000000000000461

2 Milanovic, Z., Sporis, G. & Weston, M. Effectiveness of High-Intensity Interval Training (HIT) and Continuous Endurance Training for VO2max Improvements: A Systematic Review and Meta-Analysis of Controlled Trials. *Sports Med* **45**, 1469-1481 (2015). https://doi.org:10.1007/s40279-015-0365-0

3 Sarzynski, M. A., Ghosh, S. & Bouchard, C. Genomic and transcriptomic predictors of response levels to endurance exercise training. *J Physiol* **595**, 2931-2939 (2017). https://doi.org:10.1113/JP272559

4 Lortie, G. *et al.* Responses of maximal aerobic power and capacity to aerobic training. *International journal of sports medicine* **5**, 232-236 (1984).

5 Bouchard, C. *et al.* Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol (1985)* **87**, 1003-1008 (1999). https://doi.org:10.1152/jappl.1999.87.3.1003

6 Sarzynski, M. A. & Bouchard, C. World-class athletic performance and genetic endowment. *Nat Metab* **2**, 796-798 (2020). https://doi.org:10.1038/s42255-020-0233-6

7 Ross, R. *et al.* Precision exercise medicine: understanding exercise response variability. *Br J Sports Med* **53**, 1141-1153 (2019). https://doi.org:10.1136/bjsports-2018-100328

8 Robbins, J. M. *et al.* Human plasma proteomic profiles indicative of cardiorespiratory fitness. *Nat Metab* **3**, 786-797 (2021). https://doi.org:10.1038/s42255-021-00400-z

9 Dias, R. G. *et al.* PBMCs express a transcriptome signature predictor of oxygen uptake responsiveness to endurance exercise training in men. *Physiol Genomics* **47**, 13-23 (2015). https://doi.org:10.1152/physiolgenomics.00072.2014

10 Timmons, J. A. *et al.* Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J Appl Physiol (1985)* **108**, 1487-1496 (2010). https://doi.org:10.1152/japplphysiol.01295.2009

11 Schmidt, M. *et al.* The Human Blood Transcriptome in a Large Population Cohort and Its Relation to Aging and Health. *Front Big Data* **3**, 548873 (2020). https://doi.org:10.3389/fdata.2020.548873

12 Gillman, A. S., Stevens, C. J. & Bryan, A. D. Women's exercise identity increases after a 16-week exercise RCT and is linked to behavior maintenance at follow-up. *Psychol Sport Exerc* **54** (2021). https://doi.org:10.1016/j.psychsport.2021.101888

13 Stevens, C. J., Gillman, A. S., Giordano, G. R. & Bryan, A. D. Modeling longitudinal variation in affective response to exercise across a 16-week randomized control trial (RCT). *Health Psychol* **40**, 928-939 (2021). https://doi.org:10.1037/hea0001023

14 Campbell, W. W. *et al.* High-Intensity Interval Training for Cardiometabolic Disease Prevention. *Med Sci Sports Exerc* **51**, 1220-1226 (2019). https://doi.org:10.1249/MSS.0000000000001934

15 Garber, C. E. *et al.* American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* **43**, 1334-1359 (2011). https://doi.org:10.1249/MSS.0b013e318213fefb

16 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R. & LaRocca, T. J. Healthy aging interventions reduce repetitive element transcripts. *J Gerontol A Biol Sci Med Sci* (2020). https://doi.org:10.1093/gerona/glaa302
17 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560
18 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
19 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8
20 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019). https://doi.org:10.1093/nar/gkz369
21 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011). https://doi.org:10.1371/journal.pone.0021800
22 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010). https://doi.org:10.1371/journal.pone.0013984
23 Williams, C. J. *et al.* Genes to predict VO2max trainability: a systematic review. *BMC Genomics* **18**, 831 (2017). https://doi.org:10.1186/s12864-017-4192-6
24 Skinner, J. S., Jaskolski, A., Saskolska, A., Krasnoff, J., Gagnon, J., Leon, A.S., Rao, D.C., Wilmore, J.H., Bouchard, C. Age, sex, race, initial fitness, and response to training: the HERITAGE Family Study. *J Appl Physool* **90**, 1770-1776 (2001).
25 Miller, J. L. & Grant, P. A. The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcell Biochem* **61**, 289-317 (2013). https://doi.org:10.1007/978-94-007-4525-4_13
26 Robinson, M. M. *et al.* Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans. *Cell Metab* **25**, 581-592 (2017). https://doi.org:10.1016/j.cmet.2017.02.009
27 Guillen Sacoto, M. J., Chapman, K. A., Heath, D., Seprish, M. B. & Zand, D. J. An uncommon clinical presentation of relapsing dilated cardiomyopathy with identification of sequence variations in MYNPC3, KCNH2 and mitochondrial tRNA cysteine. *Mol Genet Metab Rep* **3**, 47-54 (2015). https://doi.org:10.1016/j.ymgmr.2015.03.007
28 Han, K. W. W., Po, W. W., Sohn, U. D. & Kim, H. J. Benzyl Isothiocyanate Induces Apoptosis via Reactive Oxygen Species-Initiated Mitochondrial Dysfunction and DR4 and DR5 Death Receptor Activation in Gastric Adenocarcinoma Cells. *Biomolecules* **9** (2019). https://doi.org:10.3390/biom9120839
29 Bouchard, C. in *Health risk estimation, risk reduction and health promotion. Proceedings of the 18th annual meeting of the Society of Prospective Medicine.* 463-476 (Canadian Public Health Association Ottawa, Ontario, Canada).
30 Hawley, J. A., Hargreaves, M., Joyner, M. J. & Zierath, J. R. Integrative biology of exercise. *Cell* **159**, 738-749 (2014). https://doi.org:10.1016/j.cell.2014.10.029
31 Egan, B. & Zierath, J. R. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. *Cell Metab* **17**, 162-184 (2013). https://doi.org:10.1016/j.cmet.2012.12.012

CHAPTER 3: REPETITIVE ELEMENT TRANSCRIPT ACCUMULATION IS RELATED TO INFLAMMAGING IN HUMANS

Introduction

Advancing age is the primary risk factor for most chronic disorders including metabolic, cardiovascular, and neurodegenerative diseases, and chronic, low-grade inflammation is a central mechanism of aging that drives the development of these diseases[1]. This "inflammaging" contributes to impairments in multiple domains of physiological function (e.g., cognitive and motor performance) that predict the risk of disease. As a result, determining the underlying causes of age-related inflammation is a major goal of current geroscience[2]. Advances in next-generation sequencing have provided important insight on transcriptomic and epigenetic changes upstream of cellular inflammation and physiological dysfunction. However, most studies have focused on protein-coding genes, which make up only ~2% of the human genome[3,4].

One particularly overlooked fraction of the human genome is composed of non-coding, repetitive sequences known as repetitive elements (RE). RE account for ~60% of the human genome and can be divided into sub-types including transposons, retrotransposons, structural sequences like satellites, and others[4,5]. These repetitive sequences are typically epigenetically suppressed, but growing evidence indicates that RE transcripts accumulate with age[6-8] and contribute to several processes that are hallmarks of aging[9-11], including pro-inflammatory signaling. For example, others have shown that inhibiting retrotransposons, such as LINE-1, reduces inflammation with aging in cells and model organisms[10]. We also recently reported that RE transcript expression is sensitive to healthspan modifiers such as habitual exercise (which reduces inflammation and is associated with lower RE transcript levels in humans) and high-fat diet (a pro-inflammatory stress that increases RE transcripts in mice)[6]. Furthermore, we found that RE transcript levels were reduced by rapamycin, calorie restriction and acarbose (lifespan-

extending interventions), and that genes/transcripts reduced in parallel with RE in response to these interventions were largely related to inflammation[6].

Despite evidence linking RE dysregulation, aging, and inflammation in cells and model organisms, clinical data in support of this idea are lacking. There also is limited information on specific pathways by which RE transcripts may contribute to inflammaging in humans, although two likely mechanisms[5] include: 1) some RE transcripts (particularly those from retrotransposons) can be reverse-transcribed, leading to cytoplasmic complementary DNA (cDNA), a virus-associated molecular pattern that can activate innate immune sensors and subsequent inflammatory signaling[12]; and/or 2) palindromic or partially self-complementary RE transcripts can form endogenous double-stranded RNA (dsRNA), another virus-associated molecular pattern that activates cellular inflammatory responses via dsRNA sensors[13]. It also is possible that both of these processes occur during aging, and that a combination of endogenous, RE-derived cDNA and dsRNA eventually exceeds a cellular "self-tolerance" threshold, leading to inflammatory signaling[5,14].

Finally, upstream of RE transcript-induced inflammation, the exact mechanisms of age-related RE dysregulation are also unclear. RE are typically located in epigenetically suppressed regions of the genome, characterized by DNA methylation and heterochromatinization that inhibit transcription. Age-related DNA hypomethylation and loss of heterochromatin are established hallmarks of aging, and it is possible that these changes underlie RE transcript accumulation with aging[15-18]. However, again, whether these events are associated with RE transcript dysregulation and inflammaging in humans is unknown. Here, we present new transcriptomic and clinical data showing that RE transcript levels are associated with age-related inflammation in humans and related to the expression of multiple cDNA and dsRNA innate immune sensors. We also present new RNA-seq data from an exercise intervention study and show that this healthspan-extending intervention decreases RE transcript levels, and

we use these findings as a framework for examining epigenetic changes that may contribute to age-related RE transcript dysregulation.

Methods

Human subjects and samples

Samples for this study were obtained from several previously conducted clinical trials[19-26]. All study procedures conformed to the Declaration of Helsinki and were approved by the Institutional Review Board of the University of Colorado Boulder. Written informed consent was obtained from all subjects.

Older adult samples: RNA-seq was performed on peripheral blood mononuclear cells (PBMC) collected from 90 middle-aged and older adults (aged 45 to 79 years) prior to the start of any intervention studies. All subjects were healthy as assessed by medical history and physical examination, absence of chronic or age-related disease, and body mass index (BMI) less than 29 kg/m^2. Subjects were also sedentary as defined by self-reported exercise less than 2 d/week and less than 30 min/day. Basic clinical measurements (e.g., blood pressure) were performed using standard techniques, and cognitive and motor/physical functions were assessed using the NIH Toolbox testing batteries as previously described[27,28]. Maximal oxygen consumption (VO_2max) was assessed during treadmill exercise (Balke protocol) as previously described[29]. Blood samples were obtained from fasted subjects via venipuncture for measurements of standard markers of health (e.g., cholesterol) using standard procedures and inflammation (e.g., serum IL-6) via high-sensitivity ELISA. PBMC were isolated by traditional Ficoll gradient centrifugation, and RNA-seq and gene/RE expression analyses were performed using standard methods described below.

Exercise training intervention samples: RNA-seq and whole-genome bisulfite sequencing (WGBS, to profile methylation) were also performed on PBMC collected as above from healthy, sedentary, young adults (29 ± 1 years, n=10, reference controls), and from healthy, sedentary, older adults (72 ± 3 years, n=13) before (pre) and after (post) 16 weeks of supervised aerobic interval training (treadmill walking/running at different intensities, 3 days per week)[25,26].

Sequencing and bioinformatics analyses

RNA sequencing and analyses: RNA isolation and transcriptome analyses were performed as previously reported[6,7,30]. RNA was isolated from snap-frozen PBMC using Trizol (Thermo) and a spin column kit (Direct-Zol, Zymo Research) that included a DNase I treatment to remove genomic DNA. To capture as many genes and RE transcripts as possible, and because samples had been stored for several years (RNA integrity numbers 5-8), RNA libraries were generated using Takara Pico total RNA (low-input ribo-depletion) kits. All libraries were sequenced on an Illumina NovaSeq 6000 platform to produce >40 M 151-bp paired-end fastq reads per sample. RE transcripts were analyzed with both the TEtranscripts[31] and RepEnrich2[32] programs to confirm similar findings with different analysis pipelines[7,30], but primary data are from TEtranscripts, which produced fewer TE counts on average. Briefly, RNA-seq reads were trimmed and quality filtered with *fastp*[33], then aligned to the human genome (hg38 *Homo sapiens*) using STAR followed by gene and RE count generation with TEtranscripts[31]. To compensate for RNA quality due to due to long-term storage, the following STAR parameters were used during alignments to maximize uniquely mapping reads and generate multi-mapping counts as suggested in the TEtranscripts manual: --outFilterScoreMinOverLread 0.3 --outFilterMatchNminOverLread 0.3 --winAnchorMultimapNmax 200 --outFilterMultimapNmax 100. Gene counts from STAR and TE transcript counts (generated in Tetranscripts, which

counts multi-mapping reads against a human repeat-masker annotation file with all RE) were analyzed together for differential expression in R using DESeq2 software[34]. Normalized counts extracted from DESeq2 were used for all statistical analyses. Gene ontology (GO) analyses of differentially expressed genes were performed using rank-ordered lists and the g:Profiler algorithm[35], and the most specific GO modules (biological processes) were identified as terminal GO processes using GO.db for R[36]. To identify gene and RE transcripts that correlated with markers of health and inflammation, weighted gene correlation network analysis (WGCNA) [37] was performed using normalized transcript counts for all samples and a minimum module size of 30, followed by GO analyses on modules correlating with markers of interest at $p<0.05$ as described above.

Whole-genome bisulfite sequencing and analysis (WGBS): DNA isolation from PBMC was performed using a spin column kit (QuickDNA, Zymo). For WGBS, genomic DNA was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research) and sequencing libraries were prepared with Zymo-seq WGBS kits and sequenced (151bp paired-end fastq reads) at >20x coverage on an Illumina NovaSeq instrument. WGBS reads were mapped using abismal[38], and the methylation analysis pipeline MethPipe[39] was used to remove duplicate reads, calculate methylation levels at individual sites, estimate bisulfite conversion rates, and identify hypo- and hyper-methylated regions of the genome (i.e., differentially methylated regions) according to the MethPipe manual. The resulting differentially methylated region lists (bed files) were then intersected with a gtf/bed file including annotated repetitive DNA sequences (obtained from repeatmasker.org) to identify the RE located in these regions.

Differential chromatin accessibility analyses: Differential chromatin accessibility was analyzed in a previously published RNA-seq and transposase-accessible chromatin with sequencing (ATAC-seq) dataset on PBMC from 15 young and 23 older adults[40]. TEtranscripts

was used to identify differentially expressed genes and RE in this dataset as described above (in RNA-seq data), and the existing ATAC-seq data were analyzed using Genrich[41] to identify peaks of significant chromatin enrichment. The resulting bed files (listing chromatin-accessible regions) were intersected with a gtf/bed file including annotated repetitive DNA sequences (obtained from repeatmasker.org) to identify RE located in these regions.

Statistical analyses

Differential expression of genes and RE transcripts was quantified using DESeq2[6,7,30], and significant differences were determined using the default Wald test. Simple linear regressions, normality, and *t* tests, as well as WGCNA module heat mapping were performed and constructed using GraphPad Prism software. RE comparisons by type were performed using Shapiro-Wilk normality testing followed by Mann-Whitney test versus the mean Log2 fold-change for all transcripts with >10 mean counts (to account for any global transcriptome, library preparation, and/or normalization artifacts). For overlap comparisons of RE transcript expression and methylation or chromatin accessibility, the most differentially expressed RE in older vs. young subjects, and post vs. pre-exercise training subjects were identified as those with Log2 fold-change greater than one standard deviation above the mean Log2 fold-change for all RE.

RNA-seq datasets and availability

Transcriptome and genomic/methylation data from this study have been deposited in the NCBI Gene Expression Omnibus website (GEO) and are accessible under accession number GSE[…]. Existing RNA- and ATAC-seq data that were re-analyzed here can be found in the European Genome-phenome Archive under accession number EGAS00001002605. Lists of differentially expressed genes, differentially methylated RE, differentially chromatin-accessible

RE, and GO terms supporting the analyses presented here are available in the Supporting Information section of this article.

Results

RE transcript levels are related to systemic inflammation in healthy older adults

We recently reported that RE transcripts increase with aging in human peripheral blood mononuclear cells (PBMC)[6]. To determine if this age-related increase is associated with "inflammaging," we performed total RNA-seq on PBMC samples from 90 healthy, older sedentary adults (ages 63 ± 7 years) (**Table 3.1**). Anthropometric measurements and clinical biomarkers including serum interleukin-6 (IL-6) and C-reactive protein (CRP), which are "gold-standard" markers of inflammation[42], were collected from these same participants as part of several larger studies[19-24]. Overall subject characteristics (**Table 3.1**) were consistent with what others have reported for older, healthy sedentary adults[22,43].

Table 3.1. Subject characteristics for 90 healthy, older adults.

	Males (n = 45)	**Females (n = 45)**
Age (years)	63.2 ± 7.2	64.4 ± 7.4
Body mass index (kg/m^2)	23.7 ± 3.0	25.9 ± 4.4*
Education (years)	20.8 ± 1.3	21.6 ± 1.5
Resting systolic blood pressure (mmHg)	116.3 ± 14.1	125.8 ± 11.3*
Resting diastolic blood pressure (mmHg)	70.0 ± 6.5	74.7 ± 7.5*
Resting heart rate (bpm)	57.7 ± 7.6	55.7 ± 7.1
Total cholesterol (mg/dL)	182.8 ± 27.1	173.9 ± 33.1
HDL (mg/dL)	66.0 ± 19.8	49.8 ± 12.4*
LDL (mg/dL)	101.7 ± 22.4	107.8 ± 28.1
Oxidized LDL (mg/dL)	51.6 ± 32.7	54.8 ± 23.5
Triglycerides (mg/dL)	76.6 ± 25.4	96.1 ± 58.0*
Glucose (mg/dL)	84.5 ± 6.6	87.9 ± 6.7
Insulin (pmol/L)	9.1 ± 6.0	8.3 ± 2.4
C-reactive protein (mg/dL)	1.1 ± 1.3	1.2 ± 1.1
Interleukin 6 (pg/mL)	1.0 ± 0.6	1.0 ± 0.5

Values are reported as mean ± SD.
*$p < 0.05$ compared to Males.

First, to confirm our previous findings in this dataset, we selected RNA samples from the 10 oldest (average age 76 years) and 10 youngest (middle-aged, 51 years, matched for sex) subjects in the larger group and analyzed differentially expressed genes and RE transcripts (**Figure 3.1A**). Consistent with our previous findings[6,7], we found an age-related increase in all major types of RE transcripts including LINEs, SINEs, LTRs, DNA transposons, and satellites (**Figure 3.1B**), reflecting "global" RE transcript dysregulation and accumulation with aging.

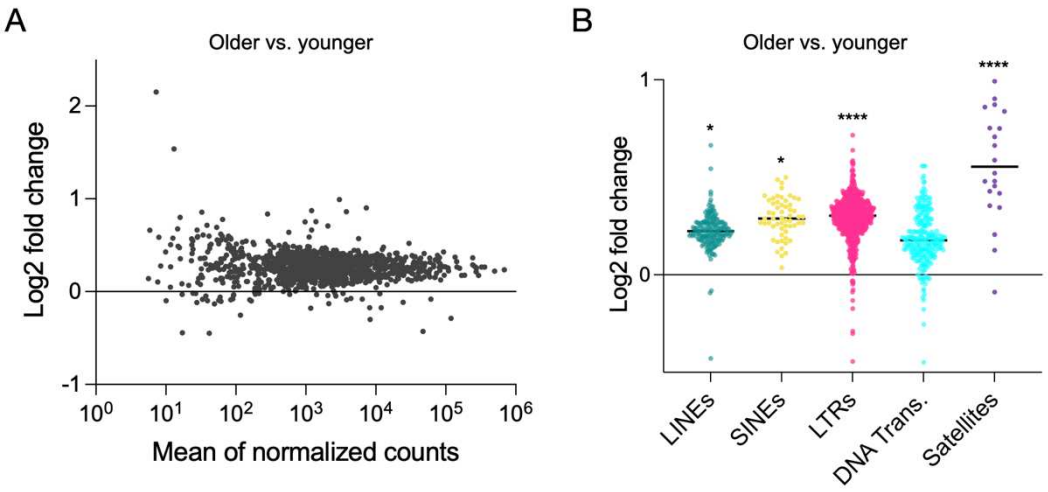


Figure 3.1. RE transcripts accumulate with age, consistent with previous reports. A) MA plot of differentially expressed RE in the ten oldest adults (average age 76 years) vs. ten youngest adults (average age 51 years) in the cohort. **B)** Transcript levels (Log2 fold change) for each major type of RE in the same oldest vs. youngest adults. *p<0.05; ****p<0.0001

To identify gene/RE transcripts related to markers of function and health in this group (including markers of inflammation) in an unbiased fashion, we performed a weighted gene correlation network analysis (WGCNA)[37] using both gene and RE transcripts (**Figure 3.2A**). This approach clusters transcripts into groups (y-axis) referred to as "modules" based on their correlation with physiological traits of interest (x-axis). We found that several gene/RE modules in our WGCNA analysis correlated with clinically important physiological parameters (e.g., module D, which included genes/biological processes like chemokine signaling and coagulation and was positively related to oxidized LDL cholesterol but negatively related to cognitive

function), demonstrating the validity of this network analysis approach. However, most important for the present investigation, we found one module of genes/RE transcripts that correlated positively with serum IL-6, a key clinical marker of age-related inflammation (**Figure 3.2A**, module F) ($r = 0.69$, $p < 0.0001$). Gene Ontology (GO) analysis of the top transcripts contributing to the correlation between module F and IL-6 yielded biological processes related to cell signaling and cellular responses to stress (**Figure 3.2B**).

Next, we examined the genes/transcripts in module F that contributed most to the correlation with IL-6 (**Figure 3.2C**). We found several RE among these transcripts, including DNA transposon transcripts like *UCON69* and *MER123*, the long terminal repeat (LTR) retrotransposon *LTR13A*, and satellite transcripts like *HSAT6* and *SATR2*. Interestingly, *MER123* is a palindromic RE with the potential to form endogenous dsRNA, whereas LTR transcripts may form cDNA via reverse transcription[44]. *LTR13A* is a human endogenous retrovirus K (HERV-K) in the HML-2 group, which is considered to be the most active group of HERVs. In fact, recently published data show that active HERV-Ks directly promote inflammatory signaling and tissue aging/premature senescence[45]. Satellite transcripts are also a known marker of cellular senescence[46], which can be induced by various types of immune/inflammatory signaling. Furthermore, consistent with the idea that these RE transcripts may stimulate inflammation, interferon induced protein 144 (*IFI44*), which has been implicated in the response to viral infections and retrotransposon activation[47], was also a top contributor to the module F correlation with serum IL-6. Other genes that contributed to this module/trait correlation included the antiviral response factor Mov10 RISC complex RNA helicase (*MOV10*, also implicated in cDNA sensing)[48] and histone family member X (*H2AFX*), which is activated in response to DNA breaks (i.e., due to RE retrotransposition/other events)[49] . We also found a positive correlation between total RE transcript counts and serum IL-6 ($r = 0.39$, $p < 0.0001$) (**Figure 3.2D**), in line with our WGCNA data, and this was not driven by any particular type of RE, as transcripts from DNA transposons, LINEs, SINEs, LTRs and satellites were similarly

related to IL-6 levels (**Supplementary Table 3.5**). Together, these findings are consistent with the idea that cellular innate immune activation may link RE transcript accumulation with inflammation during aging in humans.

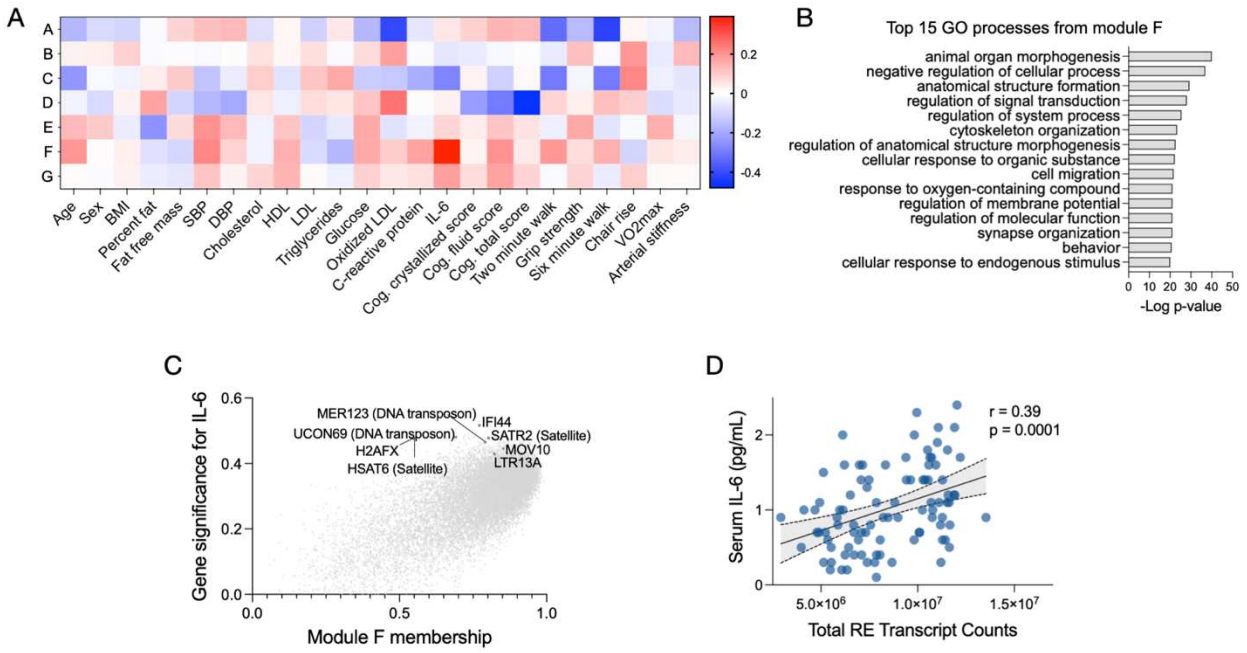


Figure 3.2. Age-related RE transcript expression is correlated with serum IL-6. A) Weighted gene correlation network analysis (WGCNA) of all normalized gene expression counts (including RE transcripts) and clinical traits in 90 healthy older adults. **B)** Biological processes (GO terms) associated with genes in module F, which correlated with IL-6. **C)** Contribution of genes in module F to its correlation with IL-6. Top RE and genes related to inflammation and DNA damage are labeled. **D)** Correlation of total normalized RE transcript counts with serum IL-6 in the entire sample of 90 older adults. Grey bands represent 95% confidence intervals.

Potential mechanisms by which RE transcript accumulation may lead to inflammatory signaling

The mechanisms by which RE transcripts contribute to systemic inflammation could involve: 1) RE-derived cDNA, which can bind to cGAS and trigger the cGAS-STING inflammatory signaling pathway[12,50]; or 2) RE-derived dsRNA, which can bind to dsRNA sensors[13,51] that also trigger inflammatory signaling (**Figure 3.3A)**. Therefore, we looked for correlations among total RE transcript counts and the expression of genes encoding several key mediators in these inflammatory pathways, including cGAS, STING, dsRNA sensors (PKR,

MDA5, OAS, TLR3 and NLRP1) and other mediators of innate immune activation/inflammation such as IRF3, IL-1β and IL-6. (**Figure 3.3B-K**). We found that total RE transcript count was positively related to expression of the IL6 gene (consistent with our findings above), as well as the cytoplasmic dsRNA sensors MDA5, PKR, OAS, NLRP1, the endosomal dsRNA sensor TLR3, the cDNA sensor STING, and inflammatory cytokine IL-1β. Interestingly, RE transcript levels were not related to cGAS (which is upstream of STING) or IRF-3, which transduces transcriptional responses to dsRNA sensor activation. Still, these data demonstrate that higher RE transcript levels are associated with transcriptomic increases in cellular mediators of immune and inflammatory signaling in humans, possibly suggesting a role for both RE-derived cDNA and dsRNA in inflammaging.

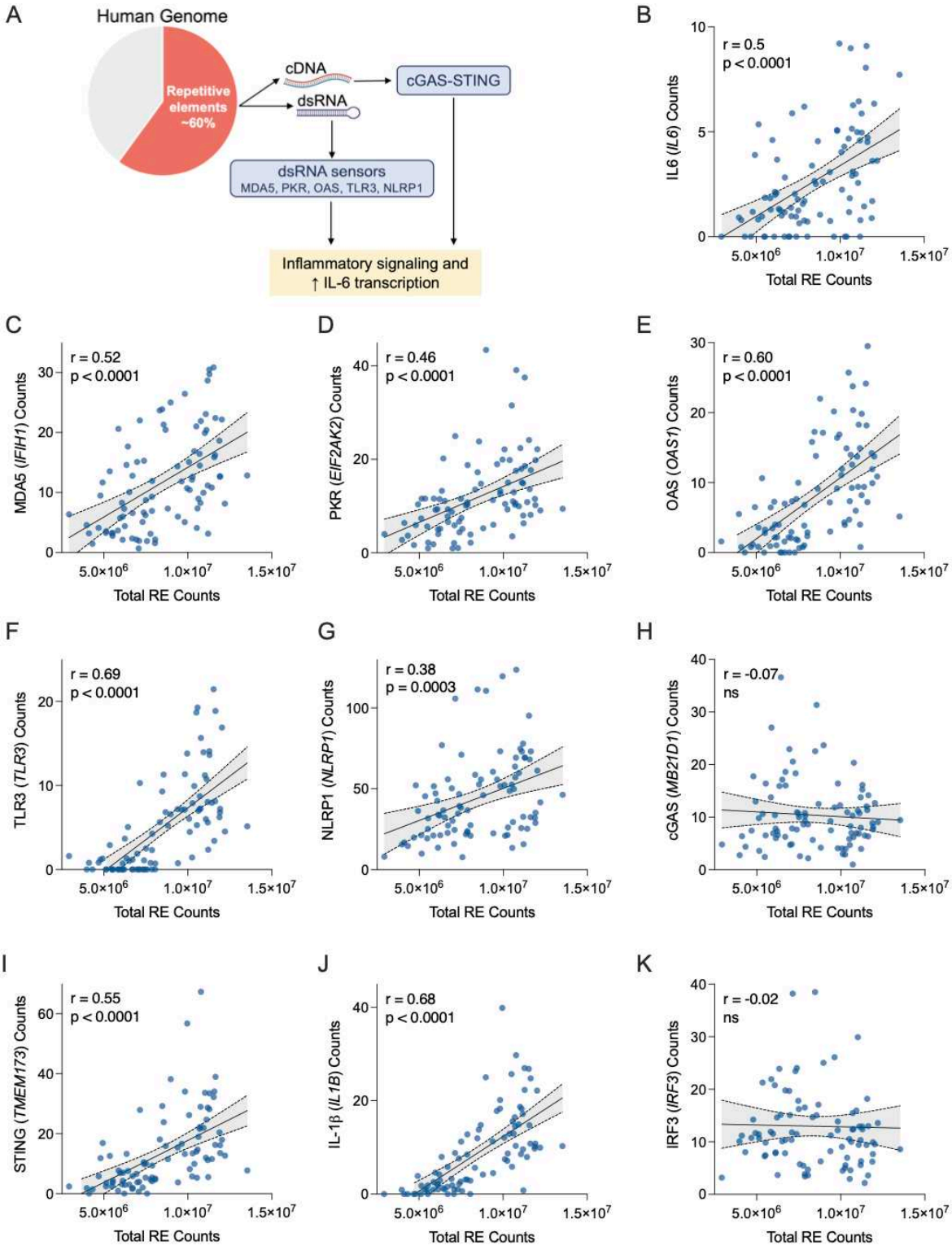


Figure 3.3. RE transcripts are related to inflammatory signaling, potentially mediated by formation of cDNA or dsRNA. A) Schematic of potential mechanisms by which RE transcripts may initiate inflammatory signaling, including formation of complementary DNA (cDNA) and double-stranded RNA (dsRNA). **B-K)** Correlations among total RE transcript counts and

normalized counts for the genes that encode IL-6, MDA5, PKR, OAS, TLR3, NLRP1, cGAS, STING, IL-1β, and IRF3. Grey bands represent 95% confidence intervals.

Age-related RE transcript increases are reversed by exercise

Exercise training increases healthspan, delays age-related physiological dysfunction and is often used as a framework for examining the mechanisms underlying age-related biological changes (i.e., to determine if and how they are reversed by exercise[52-55]). Therefore, to determine if RE transcript levels are altered by exercise, we studied PBMC from young sedentary adults (29 ± 1 years, n=10, reference controls) and older (72 ± 3 years, n=13) adults before (pre) and after (post) 16 weeks of supervised aerobic exercise training (part of a separate study[25,26]) (**Table 3.2**).

Table 3.2. Subject characteristics for 13 healthy, young adults and 10 healthy, older adults pre and post a 16-week supervised exercise training intervention.

	Young	Older Pre	Older Post
Age (years)	28.8 ± 0.86	72.2 ± 2.6**	--
Sex	8 M / 5 F	6 M / 3 F	--
Body mass index (kg/m^2)	26.8 ± 1.4	24.7 ± 1.2	24.9 ± 1.3
VO_2max (ml/kg/min)	34.8 ± 2.8	24.7 ± 1.9*	25.3 ± 1.4
HR_{max} (bpm)	189 ± 3	148 ± 8**	143 ± 8
RER_{max} (VCO_2/VO_2)	1.1 ± 0	1.1 ± 0	1.2 ± 0
RPE_{max} (6 – 20 scale)	17.9 ± 0.4	17.5 ± 0.5	15.4 ± 0.6$^{\#}$

RER: respiratory exchange ratio
RPE: rate of perceived exertion
Values are reported as mean ± SEM
* $p < 0.05$ compared to Young
** $p < 0.0001$ compared to Young
$^{\#}p < 0.05$ compared to Older Pre

In this second dataset/subject group, we again found greater RE transcript levels in older adults (older pre-training vs. young individuals), confirming previous reports and our findings above (**Figure 3.4A**). GO analysis of the genes with higher expression in older pre vs. young subjects reflected increases in immune response-associated processes (**Figure 3.4B**), while genes/transcripts with lower expression were related to processes including cytoplasmic translation, which could be consistent with activation of dsRNA sensors like PKR that inhibit

translation[51,56] (**Figure 3.4C**). Importantly, when we compared older participants post-exercise training vs. pre-exercise training, we found that global RE transcript expression was decreased by the intervention (**Figure 3.4D**). GO analyses also showed that transcripts increased post- vs. pre-exercise training in older adults were related to immune cell responses, as well as metabolism and mitochondrial ATP synthesis (**Figure 3.4E**), whereas transcripts that decreased with exercise training were associated with biological processes related to transcription and neuron/cell differentiation, among others (**Figure 3.4F**). These are the first data to show decreased RE transcript expression in humans after an exercise training intervention. They agree with our previous cross-sectional analysis of habitual exercisers compared to sedentary individuals[6], and they suggest that RE transcript accumulation may be relevant to health/healthspan.

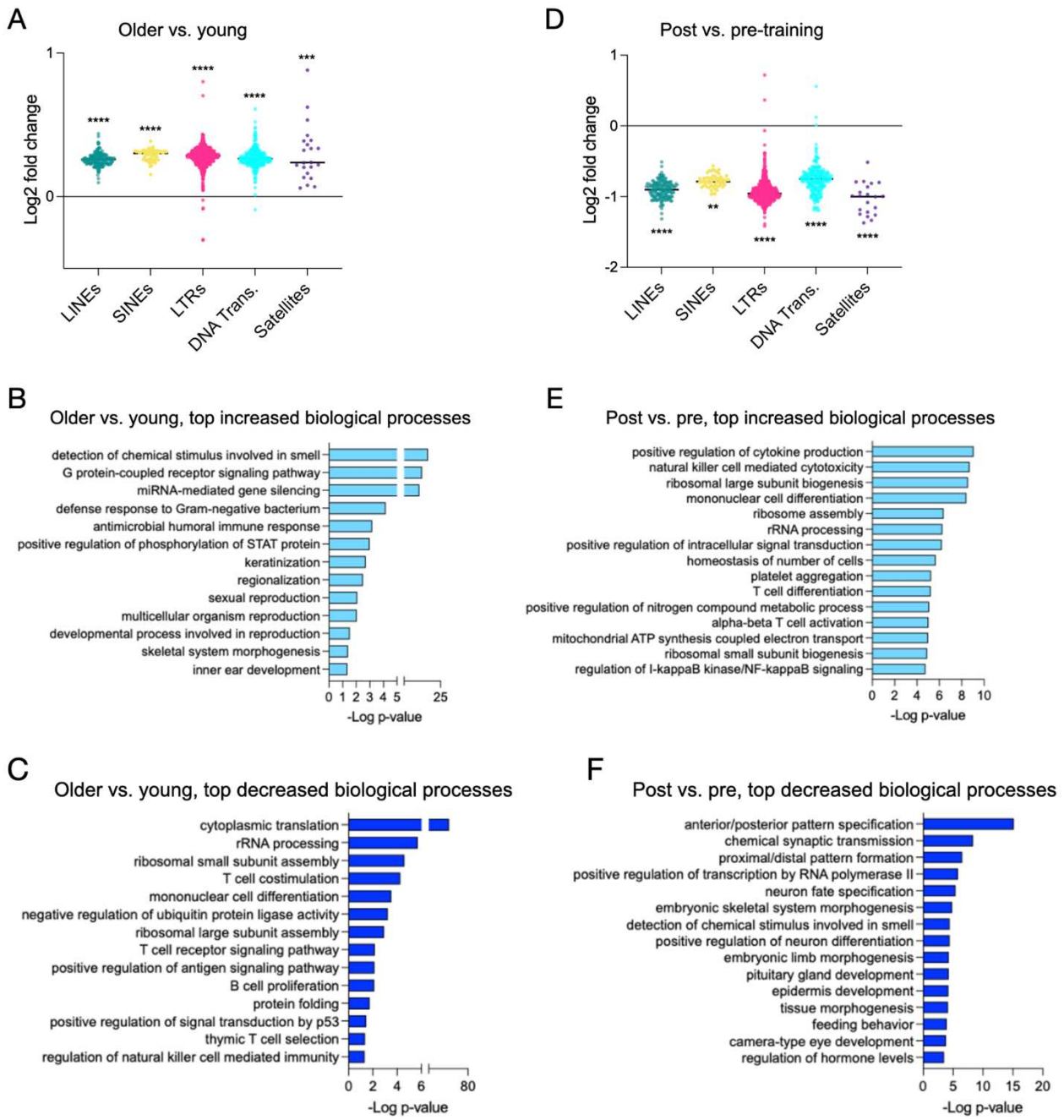


Figure 3.4. Exercise training attenuates the age-related increase in RE transcript expression. A) Transcript levels (Log2 fold change) for each major category of RE in older vs. young adults before the older adults underwent aerobic exercise training (pre-training). **B)** Biological processes related to genes with higher transcript expression in older adults pre-training compared to young. **C)** Biological processes related to genes with lower transcript expression in older adults pre-training compared to young. **D)** Transcript levels for each major category of RE in older adults after 12 weeks of exercise training vs. pre-training. **E)** Biological processes related to genes with higher transcript expression in older adults after training vs. pre-training. **F)** Biological processes related to genes with lower transcript expression in older adults post-training vs. pre-training. **p<0.01; ***p<0.001; ****p<0.0001.

Age- and exercise-related RE transcript changes are associated with differential methylation

Age-related changes in DNA methylation occur across the genome, with intergenic regions (where many RE reside) generally experiencing declines in methylation, or hypomethylation[5,15]. To determine if hypomethylation might underlie age-related RE dysregulation in our subjects/samples, we performed whole-genome bisulfite sequencing (WGBS) on duplicate PBMC samples from the same 13 young sedentary and 10 older participants pre- and post-exercise training (**Table 3.2**). We found that ~80% of RE that were increased in terms of expression in older vs. young subjects were also located in regions of the genome that were differentially hypomethylated in older subjects (**Figure 3.5A**). However, we also found that ~85% of RE with decreased expression in older adults post- vs. pre-exercise training were located in regions of the genome that were differentially hypermethylated (**Figure 3.5B**). Notably, there was significant overlap among the RE that were in the most hypomethylated/increased with age (older vs. young) and most hypermethylated/decreased with exercise (older post vs. pre-training). In fact, >200 RE that were increased in terms of expression and located in hypomethylated genome regions with aging were reduced in expression and traceable to hypermethylated genome regions with exercise (**Figure 3.5C**). Among the RE identified in this analysis, ~76% were LTRs, followed by 11% DNA transposons, 9% LINEs, 2.5% SINEs, and 0.8% satellites (**Figure 3.5D**). These results suggest that key age-related RE transcript increases that are reduced by exercise may be explained by changes in DNA methylation.

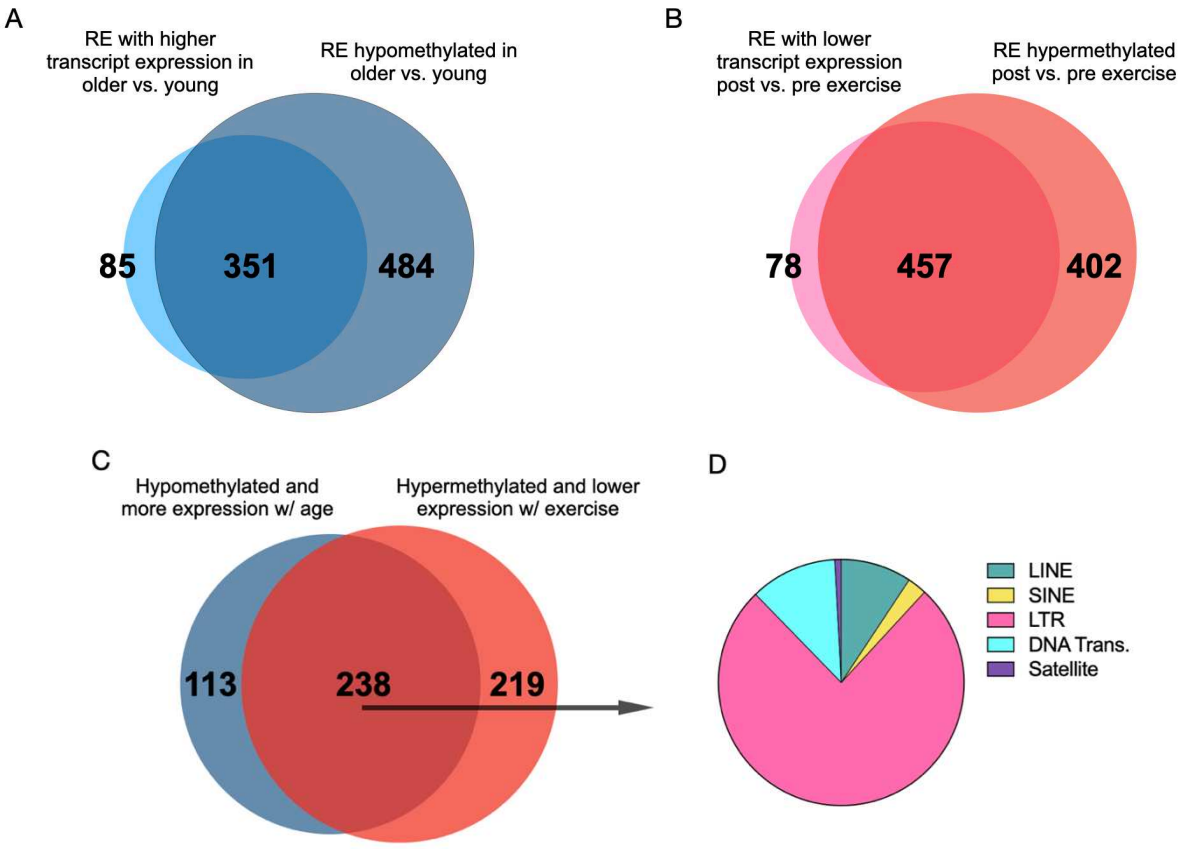


Figure 3.5. RE that increase with age are also hypomethylated, and exercise attenuates age-related hypomethylation of RE. A) Venn diagram showing overlap of RE transcripts increased in older vs. young subjects with RE located in differentially hypomethylated regions of the genome in the same subjects. **B)** Venn diagram showing overlap of RE transcripts reduced in older adults post-exercise training vs. before exercise training with RE located in differentially hypermethylated regions of the genome in the same subjects. **C)** Venn diagram showing intersection of the increased and hypomethylated RE (overlap from panel A) and decreased and hypermethylated RE (overlap from panel B). **D)** Pie chart depicting RE by type in panel C overlap.

To examine the specific RE modified most by aging and exercise, we identified the top RE of each major type that were the most differentially expressed and methylated (examples in **Table 3.3**). Interestingly, many LINE elements that were modulated by age and exercise were in the LINE-1 family, which accounts for ~17% of the human genome and includes some of the only autonomously active retroelements[57], some of which have been shown to induce type-1 interferon/inflammatory signaling via cDNA[10,58]. All of the most differentially expressed and methylated SINEs were Alu elements, which are mobile retrotransposons that require LINE-1

element machinery for their amplification[57,59]. Alu elements are also known to form dsRNA and stimulate interferon signaling mediated by MDA5[14,60]. Collectively, these data suggest that changes in methylation (i.e., hypomethylation) may be an important, upstream mechanism of RE transcript accumulation with aging, especially because these effects are at least partly reversed by exercise training, a healthspan-promoting intervention.

Table 3.3. Most differentially expressed and methylated RE by type.

	Older vs. young **Top 5 increased/hypomethylated RE**	**Post vs. pre-training** **Top 5 decreased/hypermethylated RE**
LINEs	L1ME3Cz:L1:LINE L1ME3A:L1:LINE L4_B_Mam:RTE-X:LINE L1MC1:L1:LINE L1MCb:L1:LINE	L1M1:L1:LINE L1PREC2:L1:LINE L1M3c:L1:LINE L1ME5:L1:LINE L1M3f:L1:LINE
SINEs	AluYk2:Alu:SINE AluYk4:Alu:SINE AluYc:Alu:SINE AluSx:Alu:SINE AluSx1:Alu:SINE	AluYa5:Alu:SINE AluYk4:Alu:SINE AluYe6:Alu:SINE AluYk2:Alu:SINE AluYi6_4d:Alu:SINE
LTRs	MamGyp-int:Gypsy:LTR PRIMA41-int:ERV1:LTR MER57A-int:ERV1:LTR MER4A1:ERV1:LTR ERV24B_Prim-int:ERV1:LTR	MER31-int:ERV1:LTR MLT1J:ERVL-MaLR:LTR HERVE_a-int:ERV1:LTR MER61C:ERV1:LTR ERV24B_Prim-int:ERV1:LTR
DNA transposons	Zaphod2:hAT-Tip100:DNA MamRep488:hAT-Tip100:DNA Arthur1B:hAT-Tip100:DNA Arthur1:hAT-Tip100:DNA Tigger1:TcMar-Tigger:DNA	BLACKJACK:hAT-Blackjack:DNA Arthur1B:hAT-Tip100:DNA Arthur1:hAT-Tip100:DNA MamRep38:hAT:DNA Arthur1C:hAT-Tip100:DNA
Satellites	SST1:centr:Satellite ACRO1:acro:Satellite	REP522:telo:Satellite TAR1:telo:Satellite SST1:centr:Satellite ACRO1:acro:Satellite GSAT:centr:Satellite

Potential contribution of chromatin changes to age-related RE dysregulation

Finally, in addition to hypomethylation, the age-related loss of heterochromatin could be an additional upstream mechanism of RE dysregulation. We were unable to study this in our frozen PBMC samples. However, to provide relevant insight on this possibility, we analyzed RE

in RNA-seq and ATAC-seq (chromatin accessibility) data from a previously published study of aging based on PBMC samples from 23 healthy young and 15 healthy older adults[40]. Similar to our findings in the present study (i.e., Figures 1 and 4A), we found that RE transcript levels of each major type were significantly higher in older compared to young adults in this dataset (**Figure 3.6A**). Additionally, we found that ~30% of these RE could be traced to ATAC-accessible regions of the genome (**Figure 3.6B**), suggesting that fewer RE transcripts that increase with aging may be chromatin-accessible than hypomethylated (i.e., as shown in Figure 5A). In contrast to our results suggesting that methylation changes are most prevalent in LTRs, the chromatin-accessible RE in this dataset were mostly LINEs (**Figure 3.5C**). Several RE that were the most increased with age in this dataset also overlapped with those that were most differentially expressed/methylated in the present study (**Table 3.4**). Additionally, we found that 38 RE transcripts that increased with aging could be traced to genomic regions that were both hypomethylated (in our data) and chromatin-accessible in this secondary dataset (**Supplementary Table 3.6**), and the majority of these were LINEs. However, taken together, these analyses (albeit in different datasets) suggest that DNA hypomethylation could play a more significant role in the age-related RE transcript increases we and others have observed.

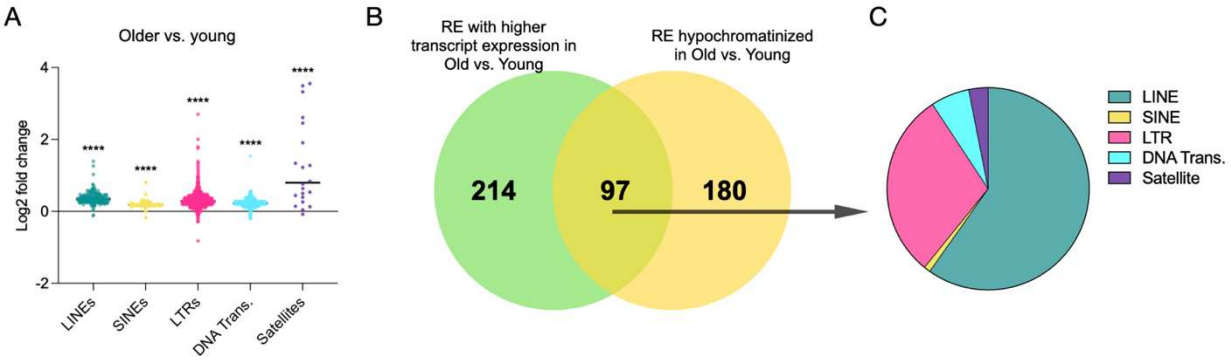


Figure 3.6. RE that are chromatin-accessible with age contribute to RE transcript accumulation. A) Differential RE expression in older vs. young adults from a separate dataset generated by Ucar and colleagues[40]. **B)** Venn diagram of RE that have increased expression in older vs. young and RE that are chromatin-accessible in older vs. young. **C)** Pie chart depicting RE by type in the panel B overlap. ****p<0.0001

Table 3.4. Top 5 RE from each major category that are the most highly expressed and chromatin-accessible with age in data from Ucar et al.

	Older vs. young (Ucar et al.)
LINEs	L1MA1:L1:LINE L1PBa:L1:LINE L1MEb:RTE-X:LINE L1M4a2:L1:LINE L1MEa:L1:LINE
SINEs	AluYa8:Alu:SINE
LTRs	LTR1E:ERV1:LTR LTR7:ERV1:LTR MER73:ERV1:LTR HERVIP10FH-int:ERV1:LTR LTR9:ERV1:LTR
DNA transposons	Charlie17b:hAT-Charlie:DNA Arthur1:hAT-Tip100:DNA MamRep4096:hAT-Tip100:DNA Charlie25:hAT-Charlie:DNA MER6A:TcMar-Tigger:DNA
Satellites	ACRO1:acro:Satellite SST1:centr:Satellite ALR/Alpha:centr:Satellite

Discussion

Chronic inflammation is a major hallmark of aging and a key mechanism of age-related physiological dysfunction and disease, but the upstream causes of inflammaging are incompletely understood. Here, we present transcriptome and clinical data supporting a role for RE transcript accumulation in inflammaging in humans, consistent with growing evidence linking RE transcripts with age-related inflammation in model organisms[5,10,58,61]. We examined the relationship between RE transcript expression and gold standard, clinical markers of systemic inflammation in aging humans, and we characterize epigenetic changes that may underlie these events. Our key findings were that the systemic inflammatory marker serum IL-6 is positively related to RE transcript levels in a large group of sedentary, middle-aged to older adults, and that RE transcript levels in these subjects are also related to expression of innate immune sensors that may detect RE-derived dsRNA and/or cDNA. We also used samples from a

separate exercise intervention to show that most RE transcripts that increase with aging likely originate from regions of the genome that are hypomethylated, and that exercise may reverse these events. Together, these findings are important, as they provide translational evidence that RE transcripts may stimulate age-related inflammation and suggest that these events may be relevant to healthspan (as they are reversed by exercise).

"Inflammaging" was first described to emphasize the close relationship between inflammation, advancing age and age-related disease. Many reports demonstrate that basal levels of immune activation and inflammatory signaling increase with age, a phenomenon known as "sterile inflammation", and that these events contribute to and increase the risk for multiple diseases of aging[10,61]. Although many adverse processes could contribute to inflammaging (e.g., senescence, mitochondrial dysfunction), its upstream causes are incompletely understood. Importantly, growing data suggest that RE transcripts may accumulate progressively with aging, and this accumulation could contribute to inflammaging via several mechanisms. For example, LINE-1 retrotransposons have been experimentally linked to cDNA-mediated interferon signaling and inflammation, and reverse transcriptase inhibitors (which can prevent the generation of RE-derived cDNA) have been used to inhibit this phenomenon in aged mice[10,58]. In addition, repression of retrotransposon expression in *drosophila* has been shown to reduce double-stranded DNA breaks, another potential upstream mechanism of inflammaging in response to DNA damage. In the same study, DNA damage was shown to be a direct result of natural, age-related retrotransposon derepression and subsequent retrotransposition[17]. Consistent with these reports in pre-clinical models and model organisms, here we took advantage of PBMC stored from multiple clinical studies and used transcriptomics to show that RE transcripts are broadly increased with aging in humans, even in older vs. middle-aged subjects, similar to our previously published data[6,7]. We also showed that this age-related accumulation of RE transcripts is associated with increased expression of immune and inflammatory signaling mediators in humans. Importantly, in this context the use of PBMC RNA

for this study can be considered both a strength and limitation, as PBMCs are largely immune cells and are expected to express a strong inflammatory and immune activation signal. Additional studies are needed to profile age-related inflammation and RE expression in other relevant tissues such as muscle, liver, and brain. In a best-case scenario, a comprehensive study would include different tissue samples from the same participants to determine the origin of RE transcripts and inflammatory signaling.

Our findings support previous studies suggesting that RE transcripts can initiate retrotransposon-associated inflammatory signaling via cGAS-STING[12], which senses cDNA in the cytoplasm. While we found no correlation between total RE transcript counts and cGAS (*MB21D1*) expression, we did find a significant positive correlation between RE and STING (*TMEM173*) transcript counts. We also found strong correlations among RE transcript counts and multiple dsRNA sensors, which is consistent with growing evidence of a role for dsRNA in age-related diseases[13]. Specifically, in our data, RE transcripts were positively related to expression of the cytoplasmic dsRNA sensors OAS (*OAS1*), MDA5 (*IFIH1*), PKR (*EIF2AK2*) and NLRP1 (*NLRP1*), as well as TLR3, a dsRNA sensor located in the endosome. This latter finding could suggest that RE-derived dsRNA has the potential to enter extracellular vesicles, which are important mediators of inter-organ communication and innate immune signaling[62,63]. In fact, recent studies have found that select retrotransposon-derived dsRNAs are packaged into extracellular vesicles and involved in interferon signaling[62,63], which could represent a mechanism by which age-related inflammation spreads. Again, a key limitation of our findings is their correlational nature, as this was a retrospective study of stored samples, so we cannot draw conclusions as to cause and effect or the specific signaling events connecting RE transcripts and inflammation. As such, future studies are needed to identify which specific RE transcripts may be most important in these events, potentially via RNA immunoprecipitation sequencing of key cDNA/dsRNA sensors, as this could provide insight on therapeutic targets. If exercise or other interventions that inhibit RE transcript activity (e.g., reverse transcriptase

inhibitors) were included in such studies, it is possible to perform cDNA/dsRNA immunoprecipitation sequencing using RNA from pre- and post-intervention. This would provide more causal/mechanistic insight in humans. Another mechanistic approach could include culturing PBMCs or fibroblasts from the same subjects to measure inflammatory signaling after knockdown of specific RE using targeted siRNA treatment. These approaches could help determine which specific RE transcripts may be responsible for initiating cDNA/dsRNA induced inflammatory signaling.

Age-related RE transcript accumulation has been previously demonstrated, and our laboratory has reported that common “anti-aging” interventions may reduce RE transcripts. However, there are no published data on RE expression in response to exercise, perhaps the best intervention for increasing healthspan, in humans. Therefore, in addition to our analyses of RE transcripts in a larger group of sedentary older adults, we also performed RNA-seq on samples collected from young healthy adults and older healthy adults before and after a 12-week exercise intervention[25,26]. In these subjects, we confirmed increased RE transcript levels in older adults (pre-exercise intervention) compared to young adults, and we found that RE transcript levels were reduced with exercise in the same older adults. This finding is in line with our previous data showing lower RE transcript expression in older habitually exercising adults compared to age- and sex-matched adults who are non-exercisers[6], but these are the first data to demonstrate this relationship in the same individuals before and after an exercise intervention. These results add further support to the idea that RE transcripts modulate healthspan[64,65]. One limitation of our sample is that the older adults did not improve objective measures of fitness such as VO_2max. This may be due to age (72 years) or the fact that the exercise intervention was low to moderate intensity. However, there was a significant decrease in perceived effort during the exercise test from pre to post training, indicating that the participants felt better at maximal effort after training. Additionally, the samples we had access to for these analyses did not have associated measurements of inflammation (e.g., serum IL-6),

but many previous reports have documented the anti-inflammatory effects of exercise[55,66,67]. In the same samples, biological processes associated with age-related transcript increases included mostly generic processes, consistent with frequently reported age-related gene ontology modules[3]. However, several decreased transcriptome/biological processes in older vs. young adults were associated with cytoplasmic translation, which could be related to the fact that a key cellular response to dsRNA is to reduce translation (a protective response to stress and/or viral infections) [51,68]. Other reduced transcriptome signatures with aging were related to immune cell and p53 signaling, which could reflect immunosenescence, an established consequence and mediator of chronic, low-grade inflammation[69]. Interestingly, senescence and p53 signaling have both been linked with dysregulation of RE. How exactly these observations relate to RE transcript increases with aging in humans requires further investigation, perhaps using pharmacological approaches (e.g., senolytics, reverse transcriptase inhibitors) in human cell culture models to test cause and effect. In any case, in the same older adults, we found that exercise broadly reduced RE transcripts and increased transcriptome signatures related to cytokine/natural killer cell function, mitochondrial ATP synthesis and NF-κB signaling, among others (i.e., somewhat the opposite of age-related differences). These data could reflect the hormetic stress response stimulated by exercise (increases in mitochondrial function and protective stress-associated NF-κβ signaling), as others have reported[70,71]. Future studies, perhaps in pre-clinical models, are needed to determine if these events are related to RE transcript reductions.

Importantly, despite the growing evidence of a role for RE transcript increases in aging, the upstream mechanisms of age-related RE dysregulation are incompletely understood. Two key epigenetic mechanisms for repressing gene and RE transcription are DNA methylation and heterochromatin maintenance[15,72,73]. Although age-related changes in these processes are complex, aging is generally associated with reduced DNA methylation (hypomethylation) and a

relaxation of heterochromatin, both of which can lead to de-repression of genes and RE[73,74]. To determine if methylation differences might explain RE transcript changes with aging and exercise in the present study, we performed WGBS to identify differentially methylated genome regions and intersected our results with RNA-seq data on the same samples. Using this approach, we found that most RE transcripts that increased with age were also associated with RE sequences in hypomethylated regions of the genome. Interestingly, the majority of these RE were LTRs, which have the potential to form both cDNA and dsRNA[5 75]. Furthermore, in the same older adults post- vs. pre-exercise training, we found that most RE transcripts that were lowered by the intervention were also associated with differentially hypermethylated regions of the genome, and again, the majority of these were LTRs. Moreover, most of the RE that were hypomethylated and increased in terms of transcript levels with aging were also hypermethylated and decreased with exercise.

To determine which specific/individual RE were the most differentially expressed and methylated, we identified RE with the highest relative expression and greatest number of occurrences in hypo/hypermethylated genome regions. We found that the most affected RE were primarily LINE elements, consistent with literature linking LINE-1 RE to aging and inflammation[8,10,76,77]. LINE-1 is the major human retrotransposable element type that is capable of autonomous retrotransposition, and it is therefore considered one of the only “active” RE (along with HERVKs). We also found multiple SINEs that were modulated by age and exercise, and these were primarily Alu elements, which can hijack LINE-1 machinery to propagate in the human genome[57]. Finally, in this same analysis we found two strongly modulated satellite RE (REP522 and TAR1), which are both located in telomeres. This observation could be related to reports that exercise encourages genomic stability of and increases telomere length, although this possibility requires further investigation. In any case, taken together, the strong overlaps we observed among RE hypomethylation/transcript increases with aging and

hypermethylation/transcript decreases with exercise suggest that age-related RE dysregulation is preceded by loss of methylation at in RE-enriched regions of the genome.

One limitation of our study was that we used PBMC samples that had been stored long-term. We used library preparation kits designed for low-input degraded RNA to address this, but also as a result, we were unable to directly investigate the role of heterochromatin in RE transcript de-repression. Still, to address this possibility, we mined a published dataset[40] which included RNA-seq and ATAC-seq data on PBMC from young and older adults. We found higher expression of RE transcripts in older vs. young adults in these data, consistent with the data generated on our RNA samples and other reports. Of the RE with higher transcript expression, ~30% could be traced to significantly chromatin-accessible regions of the genome (fewer than those that could be traced to hypomethylated regions in our data). Interestingly though, we found that most (~60%) of these chromatin-accessible RE were LINEs, in contrast to our WGBS results showing that methylation-accessible RE were mostly LTRs. Although these analyses were performed on different samples/data, they could suggest a role for both hypomethylation and chromatin relaxation as upstream mechanisms of RE derepression in humans. Given that these cross-sectional results suggest that RE types may be differentially affected by methylation/chromatin changes, future studies would benefit from performing RNA-seq, WGBS, and ATAC-seq on the same samples (although this would be a costly endeavor).

In summary, determining the underlying mechanisms of inflammaging in humans is an important goal of research aiming to extend healthspan. These are the first data to link age-related RE dysregulation with increased systemic inflammation and to show the effects of a healthspan-extending intervention on RE transcript expression in humans. Future studies, perhaps using *ex vivo* human cell culture or pre-clinical models, are needed to directly examine the mechanisms linking these events (e.g., RE-derived cDNA and/or dsRNA inflammatory signaling), and to determine if these events may be targetable, as this could reflect an important direction for research on healthspan-promoting therapeutics. Ultimately, the observations made

here are based on retrospective analyses of samples and data derived from different studies. Future studies should examine the relationships between aging, exercise, epigenetic alterations, and RE transcript expression within the same people/samples. Ideally, blood samples would be collected using PaxGene vacutainer tubes that instantly stabilize DNA and RNA at baseline in adults from young to older adults (including middle-age) with a wide variety of exercise habits from sedentary to extremely active. A subset of sedentary subjects could be included in a supervised exercise training intervention with post-intervention blood sampling to determine the acute effects of training. Subjects could also be included in a longitudinal arm of the study in which blood samples and activity levels are collected every 3-5 years to determine the effect of chronic habits/lifestyle on RE transcript expression and other inflammatory markers of interest.

Supplementary Table 3.5

R and p-values for correlation between each type of RE and serum IL-6		
	r value	p value
LINEs	0.42	<0.0001
SINEs	0.41	<0.0001
LTRs	0.40	<0.0001
DNA Trans.	0.37	0.0003
Satellites	0.38	0.0002

Supplementary Table 3.6

38 common elements in Figure 5A (up and hypomethylated with age) and Figure 6B (up and chromatin-accessible with age)		
Arthur1	hAT-Tip100	DNA
BLACKJACK	hAT-Blackjack	DNA
Charlie25	hAT-Charlie	DNA
L1MC1	L1	LINE
L1ME1	L1	LINE
L1MC2	L1	LINE
L1MA7	L1	LINE
L1MA8	L1	LINE
L1MA6	L1	LINE
L1MA9	L1	LINE
L1MC3	L1	LINE
L1MB1	L1	LINE
L1MA4	L1	LINE

L1MCc	L1	LINE
L1M6	L1	LINE
L1M3b	L1	LINE
L1MA10	L1	LINE
LTR78	ERV1	LTR
THE1B	ERVL-MaLR	LTR
LTR7	ERV1	LTR
HERV16-int	ERVL	LTR
MER73	ERVL	LTR
HERVL-int	ERVL	LTR
LTR16C	ERVL	LTR
MER65-int	ERV1	LTR
MamGypsy2-I	Gypsy	LTR
LTR48B	ERV1	LTR
LTR16D	ERVL	LTR
LTR8B	ERV1	LTR
THE1A	ERVL-MaLR	LTR
LTR78B	ERV1	LTR
LTR81C	Gypsy	LTR
LTR8A	ERV1	LTR
MLT1K-int	ERVL-MaLR	LTR
MLT1H1-int	ERVL-MaLR	LTR
SST1	centr	Satellite
ACRO1	acro	Satellite
AluYa8	Alu	SINE

REFERENCES

1 Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Curr Biol* **22**, R741-752 (2012). https://doi.org:10.1016/j.cub.2012.07.024

2 Santoro, A. *et al.* Inflammaging, hormesis and the rationale for anti-aging strategies. *Ageing Res Rev* **64**, 101142 (2020). https://doi.org:10.1016/j.arr.2020.101142

3 Shokhirev, M. N. & Johnson, A. A. Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell* **20**, e13280 (2021). https://doi.org:10.1111/acel.13280

4 Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol* **19**, 199 (2018). https://doi.org:10.1186/s13059-018-1577-z

5 Gorbunova, V. *et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43-53 (2021). https://doi.org:10.1038/s41586-021-03542-y

6 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R. & LaRocca, T. J. Healthy aging interventions reduce repetitive element transcripts. *J Gerontol A Biol Sci Med Sci* (2020). https://doi.org:10.1093/gerona/glaa302

7 LaRocca, T. J., Cavalier, A. N. & Wahl, D. Repetitive elements as a transcriptomic marker of aging: Evidence in multiple datasets and models. *Aging Cell* **19**, e13167 (2020). https://doi.org:10.1111/acel.13167

8 De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247-256 (2013). https://doi.org:10.1111/acel.12047

9 Kreiling, J. A. *et al.* in *Human Retrotransposons in Health and Disease* (ed Gael Cristofari) 297-321 (Springer International Publishing, 2017).

10 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78 (2019). https://doi.org:10.1038/s41586-018-0784-9

11 Goodier, J. L. Restricting retrotransposons: a review. *Mob DNA* **7**, 16 (2016). https://doi.org:10.1186/s13100-016-0070-z

12 Motwani, M., Pesiridis, S. & Fitzgerald, K. A. DNA sensing by the cGAS-STING pathway in health and disease. *Nat Rev Genet* **20**, 657-674 (2019). https://doi.org:10.1038/s41576-019-0151-1

13 Chen, Y. G. & Hur, S. Cellular origins of dsRNA, their recognition and consequences. *Nat Rev Mol Cell Biol* **23**, 286-301 (2022). https://doi.org:10.1038/s41580-021-00430-1

14 Ahmad, S. *et al.* Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation. *Cell* **172**, 797-810 e713 (2018). https://doi.org:10.1016/j.cell.2017.12.016

15 Soto-Palma, C., Niedernhofer, L. J., Faulk, C. D. & Dong, X. Epigenetics, DNA damage, and aging. *J Clin Invest* **132** (2022). https://doi.org:10.1172/JCI158446

16 Yousefzadeh, M. *et al.* DNA damage-how and why we age? *Elife* **10** (2021). https://doi.org:10.7554/eLife.62852

17 Chen, H., Zheng, X., Xiao, D. & Zheng, Y. Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. *Aging Cell* **15**, 542-552 (2016). https://doi.org:10.1111/acel.12465

18 Patchsung, M. *et al.* Alu siRNA to increase Alu element methylation and prevent DNA damage. *Epigenomics* **10**, 175-185 (2018). https://doi.org:10.2217/epi-2017-0096

19 Santos-Parker, J. R. *et al.* Curcumin supplementation and motor-cognitive function in healthy middle-aged and older adults. *Nutr Healthy Aging* **4**, 323-333 (2018). https://doi.org:10.3233/NHA-170029

20 Rossman, M. J. *et al.* Chronic Supplementation With a Mitochondrial Antioxidant (MitoQ) Improves Vascular Function in Healthy Older Adults. *Hypertension* **71**, 1056-1063 (2018). https://doi.org:10.1161/HYPERTENSIONAHA.117.10787
21 Martens, C. R. *et al.* Chronic nicotinamide riboside supplementation is well-tolerated and elevates NAD(+) in healthy middle-aged and older adults. *Nat Commun* **9**, 1286 (2018). https://doi.org:10.1038/s41467-018-03421-7
22 Martens, C. R. *et al.* Short-term time-restricted feeding is safe and feasible in non-obese healthy midlife and older adults. *Geroscience* **42**, 667-686 (2020). https://doi.org:10.1007/s11357-020-00156-6
23 Kaplon, R. E. *et al.* Oral trehalose supplementation improves resistance artery endothelial function in healthy middle-aged and older adults. *Aging* **8** (2016).
24 DeVan, A. E. *et al.* Effects of sodium nitrite supplementation on vascular function and related small metabolite signatures in middle-aged and older adults. *J Appl Physiol* **4**, 416-425 (2015).
25 Martin-Willett, R. *et al.* Few Structural Brain Changes Associated With Moderate-Intensity Interval Training and Low-Intensity Continuous Training in a Randomized Trial of Fitness and Older Adults. *J Aging Phys Act* **29**, 505-515 (2021). https://doi.org:10.1123/japa.2019-0352
26 Martin-Willett, R. *et al.* The influence of a 16-week exercise program, APOE status, and age on executive function task performance: A randomized trial. *Exp Gerontol* **152**, 111431 (2021). https://doi.org:10.1016/j.exger.2021.111431
27 Weintraub, S. *et al.* Cognition assessment using the NIH toolbox. *American Academy of Neurology* **3**, 54-64 (2013).
28 Reuben, D. B. *et al.* Motor assessment using the NIH Toolbox. *American Academy of Neurology* **3**, 65-75 (2013).
29 Christou, D. D. & Seals, D. R. Decreased maximal heart rate with aging is related to reduced B-adrenergic responsiveness but is largely explained by a reduction in intrinsic heart rate. *J Appl Physiol* **1**, 24-29 (1985).
30 Wahl, D. *et al.* The reverse transcriptase inhibitor 3TC protects against age-related cognitive dysfunction. *Aging Cell*, e13798 (2023). https://doi.org:10.1111/acel.13798
31 Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593-3599 (2015). https://doi.org:10.1093/bioinformatics/btv422
32 Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15** (2014).
33 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560
34 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8
35 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019). https://doi.org:10.1093/nar/gkz369
36 Carlson, M. (R package version 3.8.2., 2019).
37 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). https://doi.org:10.1186/1471-2105-9-559
38 de Sena Brandine, G. & Smith, A. D. Fast and memory-efficient mapping of short bisulfite sequencing reads using a two-letter alphabet. *NAR Genom Bioinform* **3**, lqab115 (2021). https://doi.org:10.1093/nargab/lqab115

39 Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**, e81148 (2013). https://doi.org:10.1371/journal.pone.0081148
40 Ucar, D. *et al.* The chromatin accessibility signature of human immune aging stems from CD8(+) T cells. *J Exp Med* **214**, 3123-3144 (2017). https://doi.org:10.1084/jem.20170416
41 Genrich: detecting sites of genomic enrichment (GitHub repository, 2018).
42 Michaud, M. *et al.* Proinflammatory cytokines, aging, and age-related diseases. *J Am Med Dir Assoc* **14**, 877-882 (2013). https://doi.org:10.1016/j.jamda.2013.05.009
43 Johnson, L. C. *et al.* The plasma metabolome as a predictor of biological aging in humans. *Geroscience* **41**, 895-906 (2019). https://doi.org:10.1007/s11357-019-00123-w
44 Havecker, E. R., Gao, X. & Voytas, D. F. The diversity of LTR retrotransposons. *Genome Biol* **5** (2004).
45 Liu, X. *et al.* Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell* **186**, 287-304 e226 (2023). https://doi.org:10.1016/j.cell.2022.12.017
46 Swanson, E. C., Manning, B., Zhang, H. & Lawrence, J. B. Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence. *J Cell Biol* **203**, 929-942 (2013). https://doi.org:10.1083/jcb.201306073
47 Rostami, M. R. & Bradic, M. The derepression of transposable elements in lung cells is associated with the inflammatory response and gene activation in idiopathic pulmonary fibrosis. *Mob DNA* **12**, 14 (2021). https://doi.org:10.1186/s13100-021-00241-3
48 Cuevas, R. A. *et al.* MOV10 Provides Antiviral Activity against RNA Viruses by Enhancing RIG-I-MAVS-Independent IFN Induction. *J Immunol* **196**, 3877-3886 (2016). https://doi.org:10.4049/jimmunol.1501359
49 Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**, 1383-1393 (2006). https://doi.org:10.1016/j.jmb.2006.01.089
50 Hancks, D. C. & Kazazian, H. H., Jr. Roles for retrotransposon insertions in human disease. *Mob DNA* **7**, 9 (2016). https://doi.org:10.1186/s13100-016-0065-9
51 Dabo, S. & Meurs, E. F. dsRNA-dependent protein kinase PKR and its role in stress, signaling and HCV infection. *Viruses* **4**, 2598-2635 (2012). https://doi.org:10.3390/v4112598
52 Carter, A. C. *et al.* Challenges and recommendations for epigenomics in precision health. *Nat Biotechnol* **35**, 1128-1132 (2017). https://doi.org:10.1038/nbt.4030
53 Bouchard, C., Rankinen, T. & Timmons, J. A. Genomics and genetics in the biology of adaptation to exercise. *Compr Physiol* **1**, 1603-1648 (2011). https://doi.org:10.1002/cphy.c100059
54 Skinner, J. S., Jaskolski, A., Saskolska, A., Krasnoff, J., Gagnon, J., Leon, A.S., Rao, D.C., Wilmore, J.H., Bouchard, C. Age, sex, race, initial fitness, and response to training: the HERITAGE Family Study. *J Appl Physool* **90**, 1770-1776 (2001).
55 Lavin, K. M. *et al.* State of Knowledge on Molecular Adaptations to Exercise in Humans: Historical Perspectives and Future Directions. *Compr Physiol* **12**, 3193-3279 (2022). https://doi.org:10.1002/cphy.c200033
56 Gal-Ben-Ari, S., Barrera, I., Ehrlich, M. & Rosenblum, K. PKR: A Kinase to Remember. *Front Mol Neurosci* **11**, 480 (2018). https://doi.org:10.3389/fnmol.2018.00480
57 Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-48 (2003). https://doi.org:10.1038/ng1223
58 Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab* **29**, 871-885 e875 (2019). https://doi.org:10.1016/j.cmet.2019.02.014
59 Deininger, P. Alu elements: know the SINEs. *Genome Biol* **12** (2011).

60 Heinrich, M. J. *et al.* Endogenous double-stranded Alu RNA elements stimulate IFN-responses in relapsing remitting multiple sclerosis. *J Autoimmun* **100**, 40-51 (2019). https://doi.org:10.1016/j.jaut.2019.02.003
61 Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol* **14**, 576-590 (2018). https://doi.org:10.1038/s41574-018-0059-4
62 Barrios, M. H. *et al.* Small Extracellular Vesicle Enrichment of a Retrotransposon-Derived Double-Stranded RNA: A Means to Avoid Autoinflammation? *Biomedicines* **9** (2021). https://doi.org:10.3390/biomedicines9091136
63 Kouwaki, T., Okamoto, M., Tsukamoto, H., Fukushima, Y. & Oshiumi, H. Extracellular Vesicles Deliver Host and Virus RNA and Regulate Innate Immune Response. *Int J Mol Sci* **18** (2017). https://doi.org:10.3390/ijms18030666
64 Garatachea, N. *et al.* Exercise attenuates the major hallmarks of aging. *Rejuvenation Res* **18**, 57-89 (2015). https://doi.org:10.1089/rej.2014.1623
65 Rebelo-Marques, A. *et al.* Aging Hallmarks: The Benefits of Physical Exercise. *Front Endocrinol (Lausanne)* **9**, 258 (2018). https://doi.org:10.3389/fendo.2018.00258
66 Gleeson, M. *et al.* The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease. *Nature Reviews Immunology* **11**, 607-615 (2011). https://doi.org:10.1038/nri3041
67 Pedersen, B. K. Anti-inflammatory effects of exercise: role in diabetes and cardiovascular disease. *Eur J Clin Invest* **47**, 600-611 (2017). https://doi.org:10.1111/eci.12781
68 Lemaire, P. A., Anderson, E., Lary, J. & Cole, J. L. Mechanism of PKR Activation by dsRNA. *J Mol Biol* **381**, 351-360 (2008). https://doi.org:10.1016/j.jmb.2008.05.056
69 Fulop, T., Larbi, A., Hirokawa, K., Cohen, A. A. & Witkowski, J. M. Immunosenescence is both functional/adaptive and dysfunctional/maladaptive. *Semin Immunopathol* **42**, 521-536 (2020). https://doi.org:10.1007/s00281-020-00818-9
70 Ji, L. L., Dickman, J. R., Kang, C. & Koenig, R. Exercise-induced hormesis may help healthy aging. *Dose Response* **8**, 73-79 (2010). https://doi.org:10.2203/dose-response.09-048.Ji
71 Musci, R. V., Hamilton, K. L. & Linden, M. A. Exercise-Induced Mitohormesis for the Maintenance of Skeletal Muscle and Healthspan Extension. *Sports (Basel)* **7** (2019). https://doi.org:10.3390/sports7070170
72 Wang, K. *et al.* Epigenetic regulation of aging: implications for interventions of aging and diseases. *Signal Transduct Target Ther* **7**, 374 (2022). https://doi.org:10.1038/s41392-022-01211-8
73 Pappalardo, X. G. & Barra, V. Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics Chromatin* **14**, 25 (2021). https://doi.org:10.1186/s13072-021-00400-z
74 Johansson, A., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One* **8**, e67378 (2013). https://doi.org:10.1371/journal.pone.0067378
75 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009). https://doi.org:10.1038/nrg2640
76 Erwin, J. A., Marchetto, M. C. & Gage, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat Rev Neurosci* **15**, 497-506 (2014). https://doi.org:10.1038/nrn3730
77 Van Meter, M. *et al.* SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun* **5**, 5011 (2014). https://doi.org:10.1038/ncomms6011

CHAPTER 4: MULTI-OMICS APPROACH IDENTIFIES A CIRCULATING PROTEIN SIGNATURE THAT PREDICTS AGE-RELATED DIFFERENCES IN COGNITIVE AND MOTOR FUNCTIONS

Introduction

Age is the primary non-modifiable risk factor for disease and death[1], in part because aging itself is associated with declines in cognitive and motor functions, which can lead to neurodegeneration, sarcopenia, and other diseases[2,3]. Because the number of older adults aged >65 years is growing rapidly in the United States[4], a primary goal of current geroscience research is to prevent declines in these functional domains and extend "healthspan," the number of years spent healthy and disease-free. In this context, there is a growing interest in using "omics" to determine signatures of and predictors of healthspan[5-7], and technologies such as RNA sequencing (RNA-seq), epigenetic profiling, and proteomics have expanded the possibilities for discovering novel biomarkers and mechanisms of aging. These techniques can capture epigenetic and transcriptomic changes to known, biologically relevant genes or proteins as well as non-coding repetitive sequences, or repetitive elements (RE), which are emerging as a potential driver of aging[8-11].

Most studies to date of physiological parameters reflecting healthspan have involved only one omics strategy to identify biomarkers of aging/functional decline. For example, there are several DNA methylation-derived clocks that use epigenetic profiling to predict age and biological age[12-15]. Some of the most recently developed methylation clocks [16] incorporate physiological fitness measures such as grip strength, gait speed, and VO_2max, and are better predictors of biological age than methylation age alone. Using approaches like this, recent studies have shown that exercise is associated with slower aging and better verbal short-term memory[17]. Others have developed transcriptome clocks to accurately predict age in model organisms and humans[18,19], and recently published studies have used RE transcript accumulation as a biomarker of age that is sensitive to healthspan-altering interventions such as

exercise and high fat diet[8,9]. Finally, proteomics can also be used to establish signatures of healthspan and predict all-cause mortality in humans[20], and it is now possible to combine all of these techniques for a ”multi-omics” approach.

One way to identify omics and multi-omics signatures that are relevant to healthspan is to examine them in the context of habitual exercise. Because exercise is one of the most potent healthspan-extending strategies[21-24], it can be used as a model for determining molecular transducers/mechanisms of healthspan and biomarkers of aging (i.e., by identifying molecular signatures of aging that are reversed by exercise, and testing to see if they are related to differences in healthspan-related physiological functions). Here, we measured markers of health, including standard/cardiovascular disease risk factors, cognitive and motor functions in young, older, and older habitually exercising adults, and we performed RNA-seq, whole-genome epigenetic (methylation) profiling, and proteomics on blood samples from these individuals. Our goal was to determine if transcriptome, epigenome, and proteome signatures could be intersected to identify circulating biomarkers that predict cognitive and motor function in humans. To our knowledge, this combination of omics and functional health data in adults of varying ages and exercise habits is the first of its kind.

Methods

Human participants

A total of 40 participants were studied from the following groups: 1) young (18-35 years), healthy, sedentary (≤ 2 days/week exercise) adults (Y-SED, n=12); 2) older (60-85 years), healthy, sedentary adults (O-SED, n=12); and 3) older, healthy, habitually active (≥ 5 days/week exercise) adults (O-EX, n=16). Each participant reported to the lab for one, 2-hour visit in the morning after an overnight fast. The visit included screening via health history questionnaires, completion of surveys related to exercise habits, perceived life stress and sleep, NIH Toolbox cognitive and motor testing batteries[25,26], and a blood draw for downstream RNA/DNA

sequencing, proteomics and other analyses. This study conformed to the Declaration of Helsinki, all procedures were approved by the Institutional Review Board of Colorado State University, and written informed consent was obtained from all participants.

Surveys

Each participant reported their typical exercise per week over the last year and completed a Pittsburgh sleep quality questionnaire (nightly sleep hours), perceived stress survey (perceived stress score [PSS]) and subjective age survey (the answer to "how old do you feel in years?"), all of which are established metrics[65-67].

Cognitive and motor function testing

Following survey completion, we performed NIH Toolbox cognitive and motor function batteries (www.Healthmeasures.net/nih-toolbox) in accordance with the protocols detailed by the NIH. Instructions were provided to participants visually on an iPad screen and orally by the researcher for each test. The cognitive function battery included 8 tests: the Flanker executive function test, dimensional card change sort executive attention test, picture sequence memory test for episodic memory, list sorting working memory test, pattern comparison processing speed, Rey immediate recall test, and reading and vocabulary language tests. As described elsewhere[25], a subset of these tests reflect crystallized cognitive function (concrete knowledge), while others contribute to fluid function/score (real-time learning, short-term memory, and speed), both of which can be combined into a total composite function score. All domains are reported as raw or uncorrected scores, along with age-adjusted national percentiles, for which raw scores are normalized to age-appropriate national standards. The motor function Toolbox battery included a 2-minute walk endurance test, 4-meter gait speed test, maximal handgrip strength, and a pegboard dexterity test. All walk tests were performed on the same 50-foot

course in a straight, tiled hallway. The same pegboard and handgrip dynamometer were used for every visit, and all tests were administered by one researcher to ensure consistency.

Sample collection

Blood samples were obtained via venipuncture of the antecubital vein and collected into 4 vacutainer tubes, including a PaxGene DNA tube for downstream DNA isolation, PaxGene RNA tube for RNA isolation, and two EDTA tubes—one for blood chemistry analysis (Piccolo Xpress, Abaxis) of cholesterol, triglycerides and glucose, and one for plasma isolation for proteomics. Whole blood PaxGene DNA tubes were stored at -20°C, and PaxGene RNA tubes and plasma were stored at -80°C for future processing and analyses.

RNA extraction, sequencing, and analysis

RNA was recovered from frozen PaxGene blood collection tubes using PaxGene-specific isolation kits (Qiagen), and Poly(A) RNA-seq was performed as previously described[8]. Briefly, RNA libraries were generated using Illumina TruSeq kits and sequenced on an Illumina NovaSeq 6000 platform to produce >40M 150-bp, paired-end reads per sample. Differential gene expression was analyzed using standard techniques as previously described[8,27]. Reads were trimmed and quality filtered with *fastp*[28] then aligned to the human genome (hg38 *Homo sapiens*) using the STAR aligner[29] followed by gene and RE count generation with TEtranscripts[30]. The following STAR parameters were used during alignments to maximize unique reads and generate sufficient multi-mapping reads for RE counting: --outFilterScoreMinOverLread 0.4 --outFilterMatchNminOverLread 0.4 --winAnchorMultimapNmax 200 --outFilterMultimapNmax 100. Differentially expressed genes and RE were then identified using DESeq2[31] for R. Gene ontology (GO) analyses were performed with the gProfiler algorithm[32] to determine GO biological processes, Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways, Reactome (gene/protein interaction) signatures, and biological

processes associated with differentially expressed genes. The most specific GO biological processes were identified as terminal nodes using GO.db[33] for R. To identify gene and RE transcripts that correlated with markers of health, exercise training status, cognitive and physical performance, two weighted gene correlation network analyses (WGCNA)[34] were performed using normalized transcript counts for all samples and minimum module sizes of 500 (for genes) and 10 (for RE), followed by GO analyses on all modules correlating with functional and health markers as described above.

Whole genome bisulfite sequencing (WGBS) and analysis

To profile epigenetic (methylation) differences using WGBS, DNA was recovered from frozen PaxGene blood collection tubes using PaxGene-specific isolation kits (Qiagen). For WGBS, genomic DNA was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research), and sequencing libraries were prepared with Zymo-seq WGBS kits and sequenced (151bp paired-end fastq reads) at >20x coverage on an Illumina NovaSeq instrument. WGBS reads were mapped using abismal[35], and the methylation analysis pipeline MethPipe[36] was used to remove duplicate reads, calculate methylation levels at individual sites, estimate bisulfite conversion rates, and identify hypo- and hyper-methylated regions of the genome (i.e., differentially methylated regions). The resulting list of differentially methylated regions (bed files) were then intersected with a gtf/bed file including an annotated human genome (UCSC) and repetitive DNA sequences (obtained from repeatmasker.org) to identify the genes and RE located in these regions.

Proteomics analysis

We profiled 1,500 plasma proteins using the SOMAscan assay from SomaLogic (Boulder, Colorado, USA) as previously described[37-39] (protein detection protocols performed at SomaLogic). Briefly, plasma from each participant was diluted and incubated with fluorescently

labelled DNA aptamers. Protein-aptamer complexes were isolated, eluted, and quantified using fluorescence on a DNA microarray chip[38,39]. Technical controls were measured at the same time as all samples and used to normalize data based on signal to noise ratio as described in detail[37]. Data normalization (relative to in-plate and between-plate controls) was performed by SomaLogic, and protein measurements are reported in relative fluorescence units (RFU). Finally, we performed differential protein expression analysis among groups and analyzed correlations among protein expression and health/functional domains using SomaLogic's DataDelve Statistics platform, as well as GO analyses (using the gene name for each protein) as described above to determine the terminal biological processes related to proteins of interest.

Statistics

Differential expression of genes and RE transcripts was quantified using DESeq2[8,9,27], and significant differences were determined using the default Wald test. Simple linear regressions, normality, and *t* tests, as well as WGCNA module heat mapping were performed and constructed using GraphPad Prism software. RE comparisons by type were performed using Shapiro-Wilk normality testing followed by Mann-Whitney test versus the mean Log2 fold-change for all transcripts. JMP software was used to calculate simple Pearson correlations among the 18 predictor proteins and health/function measurements, and for regression analyses of protein predictors of age, cognitive, and motor functions. A forward stepwise regression model was used to determine the top 10 of 18 proteins that predicted function in each domain.

Results

Overall multi-omics approach

To determine mediators of healthspan-relevant physiological function, we intersected RNA-seq, whole genome methylation and proteomics data to identify genes/proteins that were

the most modulated by aging and habitual exercise and that met the following criteria: 1) related to cognitive and motor function; 2) increased ("up") transcripts with aging but decreased ("down") transcripts with exercise; 3) located in genome regions that were significantly hypomethylated with aging and hypermethylated with exercise; and 4) detectable and altered with aging/exercise in plasma via targeted proteomics.

Age and habitual exercise impact cognitive and motor function

In our overall sample group, Y-SED (27 ± 6 years), O-SED (70 ± 7 years), and O-EX (68 ± 5 years) had similar healthy body mass index (BMI) and blood pressure. O-EX self-reported exercising significantly more per week than Y-SED and O-SED (**Table 4.1**). O-SED were somewhat but significantly more physically active than Y-SED, but their average of 34 min exercise per week was still well below the recommended physical activity guidelines for older adults[40]. All groups reported similar nightly hours of sleep, and O-EX reported significantly lower perceived stress scores (PSS) compared to Y-SED. O-EX had significantly higher total cholesterol, low density lipoprotein (LDL), and higher blood glucose compared to Y-SED (but not O-SED), whereas O-SED were not different from Y-SED in these domains. Overall, these subject characteristics are generally consistent with what others have reported for differences among older vs. younger adults, and with exercise training[41-45].

Table 4.1. Participant characteristics

	Y-SED	O-SED	O-EX
Sex (M/F)	6/6	7/5	8/8
Age (years)	27 ± 6	70 ± 7 *	68 ± 5 *
BMI (kg/m^2)	23.1 ± 2.4	24.5 ± 3.0	23.7 ± 3.0
Systolic BP (mmHg)	127 ± 9	129 ± 14	123 ± 8
Diastolic BP (mmHg)	71 ± 13	80 ± 11	77 ± 6
Self-reported weekly exercise (minutes)	5 ± 15	34 ± 25 *	597 ± 319 *#
Self-reported hours of sleep (hr./night)	7.3 ± 0.9	7.0 ± 0.7	7.0 ± 1.2

Perceived stress score (0-40)	15.1 ± 4.3	11.7 ± 7.3	8.2 ± 6.9 *
Total cholesterol (mg/dL)	157.1 ± 28.3	180.2 ± 39.9	205.0 ± 38.8 *
HDL (mg/dL)	66.6 ± 20.2	73.2 ± 16.9	77.8 ± 17.4
LDL (mg/dL)	70.3 ± 15.4	86.1 ± 29.0	108.2 ± 39.4 *
VLDL (mg/dL)	20.4 ± 5.4	21.2 ± 6.7	16.8 ± 5.5
Triglycerides (mg/dL)	102.2 ± 27.2	105.4 ± 32.7	83.6 ± 28.1
Glucose (mg/dL)	93.1 ± 4.7	102.0 ± 11.0	105.9 ± 9.5 *

* = p<0.5 compared to YS
# = p<0.5 compared to OS

For the NIH Toolbox cognitive function battery, total composite score and crystallized scores were similar among all groups (**Table 4.2**). However, both O-SED and O-EX had lower uncorrected fluid scores than Y-SED. Attention/executive function was significantly lower in both O-SED (17% lower) and O-EX (11%) compared to Y-SED, but O-EX performed significantly better than O-SED. Similarly, processing speed was significantly slower in O-SED and O-EX compared to Y-SED. Executive function was also significantly lower in O-SED compared to Y-SED, but not different in O-EX vs. Y-SED or O-SED. These findings are consistent with previously published studies on aging showing that exercise interventions improve NIH Toolbox cognitive function domains[41,42]. However, to the best of our knowledge, these are the first cross-sectional data showing that older habitual exercisers have higher cognitive function[43] than their sedentary peers. Interestingly, both older groups outperformed Y-SED in list-sorting/working memory in terms of age-adjusted national percentile. However, both groups of older adults also scored higher on the picture vocabulary language test compared to Y-SED, which is consistent with published reports of these same testing batteries[25].

Table 2. Cognitive function battery results

Cognitive test	Y-SED	O-SED	O-EX
Total composite score (%)	72.7 ± 26.9	79.4 ± 19.8	84.1 ± 13.3
Crystallized score (uncorrected)	109.9 ± 4.5	117.1 ± 7.6	118.5 ± 6.0
Crystallized score (%)	81.3 ± 15.7	82.2 ± 19.7	88.1 ± 13.8
Fluid score (uncorrected)	115.5 ± 11.1	94.1 ± 7.9 *	99.7 ± 10.2 *
Fluid score (%)	64.7 ± 34.4	61.5 ± 17.6	71.4 ± 24.5
Episodic memory (picture sequence memory test) (raw score)	20.0 ± 7.2	12.6 ± 7.3	13.3 ± 5.5
Episodic memory (picture sequence memory test) (%)	65.4 ± 25.8	67.7 ± 26.0	71.5 ± 22.9
Working Memory (list sorting) (raw score)	110.6 ± 8.9	108.0 ± 8.1	108.0 ± 9.3
Working Memory (list sorting) (%)	61.3 ± 26.8	86.3 ± 15.2 *	82.3 ± 20.0 *
Attention and executive function (Flanker) (uncorrected)	108.9 ± 6.0	90.2 ± 4.2 *	96.4 ± 8.0 *#
Attention and executive function (Flanker) (%)	58.4 ± 29.0	24.2 ±10.6 *	45.4 ± 25.3
Processing speed (pattern comparison) (uncorrected)	118.0 ± 13.5	86.0 ± 5.5 *	94.1 ± 19.1 *
Processing speed (pattern comparison) (%)	63.6 ± 33.0	33.9 ± 19.5	52.2 ± 38.4
Immediate recall (Rey) (raw score)	29.8 ± 6.4	24.8 ± 7.4	25.5 ± 3.8
Executive function (card sort test) (uncorrected)	112.0 ± 6.5	99.7 ± 6.9 *	104.4 ± 8.4
Executive function (card sort test) (%)	62.3 ± 31.4	66.1 ± 24.0	76.8 ± 27.4
Language (picture vocab) (uncorrected)	107.8 ± 6.0	119.0 ± 8.8 *	119.9 ±7.6 *
Language (picture vocab) (%)	74.7 ± 22.8	81.1 ± 19.5	85.1 ± 14.2
Language (oral reading recognition) (uncorrected)	111.3 ± 5.5	93.5 ± 24.9	115.4 ± 4.4
Language (oral reading recognition) (%)	80.2 ± 22.3	80.0 ± 19.6	86.8 ± 14.1

% = age-adjusted national percentile
* = p<0.5 compared to YS
# = p<0.5 compared to OS

During the 2-minute walk test, O-EX walked significantly further than Y-SED and O-SED (**Table 4.3**) and performed significantly better in terms of age-adjusted national percentile than Y-SED, indicating a potential exercise-related benefit. O-SED also performed significantly worse on the pegboard dexterity test compared to Y-SED. Collectively, these results demonstrated important and detectable physical/functional differences between our older and young subjects, and in our exercising compared to sedentary groups, which is generally in agreement with published data on these same tests[26]. These findings also support the idea that habitual exercise may partly protect against age-related declines in motor function.

Table 4.3. Motor function battery results

Motor test	Y-SED	O-SED	O-EX
2-minute walk (ft)	690.7 ± 77.4	612.7 ± 93.9	705.1 ± 77.5 #
2-minute walk (%)	63.5 ± 24.3	74.3 ± 21.8	91.7 ± 14.6 *
4-meter walk (seconds)	3.0 ± 0.4	3.0 ± 0.5	2.8 ± 0.4
Grip strength (D, %)	42.3 ± 25.6	50.4 ± 26.9	56.0 ± 28.8
Grip strength (ND, %)	34.8 ± 26.9	47.3 ± 29.2	49.9 ± 32.9
Pegboard (D, seconds)	18.7 ± 2.0	23.7 ± 7.0 *	20.6 ± 2.9
Pegboard (ND, seconds)	19.8 ± 2.5	24.6 ± 6.3 *	22.4 ± 3.6

% = age-adjusted national percentile
* = p<0.5 compared to YS
# = p<0.5 compared to OS

Cognitive and motor functions are negatively related to transcriptome signatures of inflammation and immune signaling that are somewhat attenuated by habitual exercise

Across all of our transcriptome (RNA-seq) analyses, individual gene expression differences were modest (**Figure 4.1A-B**). However, there were clear overall transcriptomic patterns of gene expression associated with aging that were somewhat reversed with exercise. In fact, among the top increased transcripts in O-SED vs. Y-SED (based on Log2 fold change), 89% were decreased in O-EX vs. O-SED (**Figure 4.1C**). Similarly, among the top decreased transcripts in O-SED vs. Y-SED, 95% were increased in O-EX vs. O-SED, suggesting that exercise may reverse age-related changes in gene expression. This observation is consistent with other reports[46-48] and the general idea that habitual exercise exerts protective effects at the transcriptome level with aging.

To determine whether the gene expression differences we observed correlated with any of our functional measurements, we performed a weighted gene correlation network analysis (WGCNA) of transcript expression, standard health measures, and cognitive and motor function domains. This analysis resulted in 10 modules/transcript clusters related to physiological and functional parameters (**Figure 4.1D**). Notably, the transcripts in module E negatively correlated with nearly all domains of cognitive and motor function, and positively correlated with age,

subjective age, and blood glucose. Gene ontology (GO) analysis indicated that the transcripts in module E were associated with biological processes like cytokine signaling and other immune and inflammatory processes (**Figure 4.1E**). Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG)[49] pathways associated with module E included several autoimmune and inflammatory pathologies (**Figure 4.1F**), and Reactome[50] networks were also related to immune/inflammatory signaling (**Figure 4.1G**). Taken together, these results suggest that inflammatory and immune signaling may underly age-related declines in cognitive and motor function, at least at the transcriptome level.

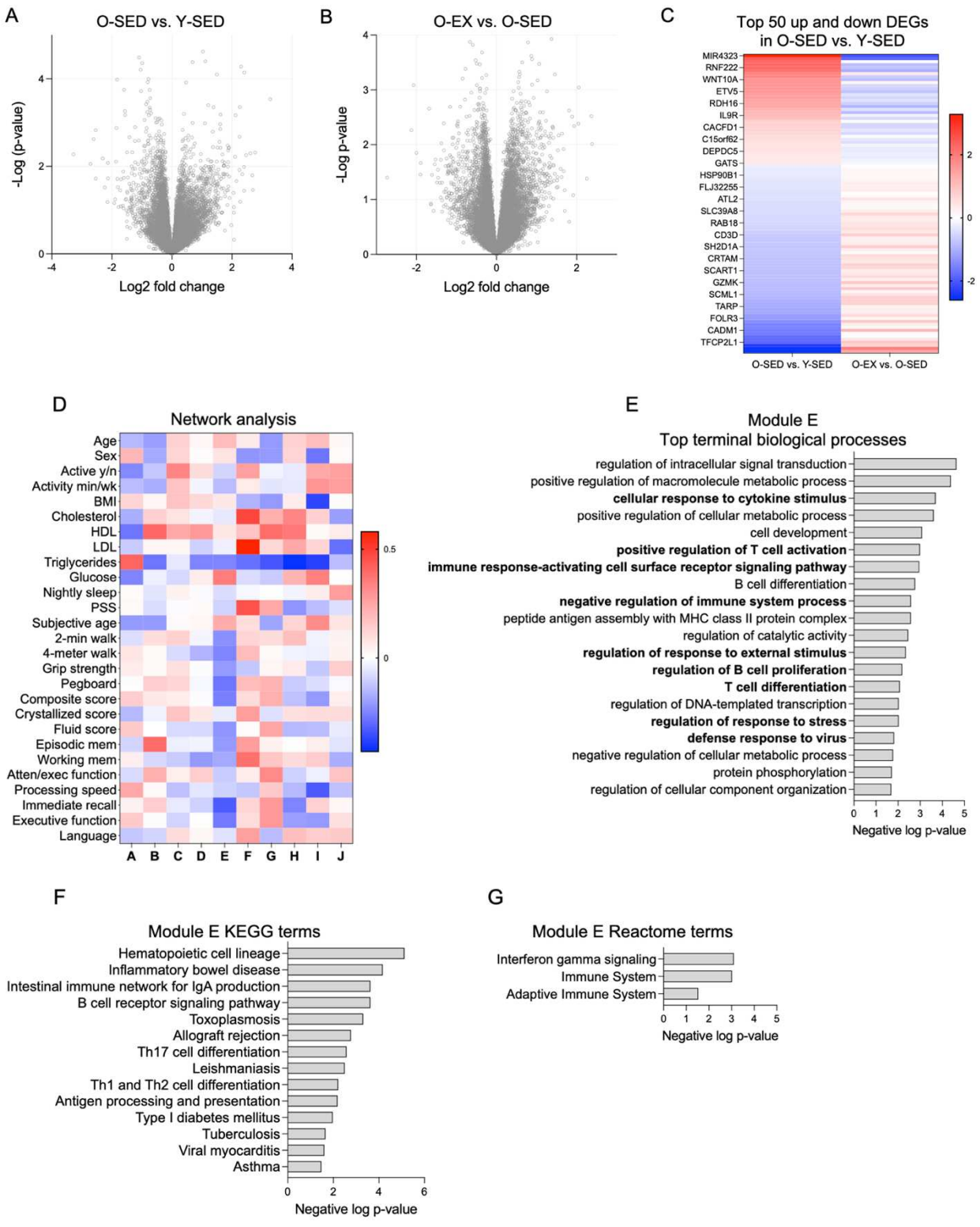


Figure 4.1. Transcriptome signatures of inflammation are negatively related to cognitive and motor functions in young and older adults. A) Volcano plot of DEGs in O-SED vs. Y-SED. **B)** Volcano plot of DEGs in O-EX vs. O-SED. **C)** Heat map of top increased transcripts (by

Log2 fold change) and top decreased transcripts in O-SED vs. Y-SED and their Log2 fold change in O-EX vs. O-SED. **D)** Weighted gene correlation network analysis (WGCNA) of all transcript counts with participant characteristics, markers of health, and cognitive and motor function domains. **E)** Terminal gene ontology (GO) biological process terms associated with the transcripts in module E. **F)** Kyoto encyclopedia of genes and genomes (KEGG)[49] terms associated with transcripts in module E. **G)** Reactome[50] terms associated with transcripts in module E.

RE transcripts accumulate with age, but are not reduced by habitual exercise in this sample

In addition to inflammatory gene expression, one possible transcriptome-related cause of inflammation with aging is an accumulation of RE transcripts, which can activate innate immune sensors. Recent findings from our lab demonstrate that RE transcript expression increases with age and is reduced by habitual exercise[8,9]. To determine the potential role of RE in the present data, we performed differential expression analysis and WGCNA on RE as above. We found that each major type of RE had significantly higher counts in O-SED vs. Y-SED (**Figure 4.2A**), consistent with our previous findings. Surprisingly though, in O-EX vs. O-SED, we also found that LINE, SINE, DNA transposon, and LTR transcripts were all significantly greater (**Figure 4.2B**). While these data were unexpected for habitual exercisers compared to their sedentary peers, in our WGCNA (**Figure 4.2C**) we found one module, D, that was negatively related to several domains of cognitive and motor function, and positively related to age, subjective age, and blood glucose. Module D consisted of 16 RE transcripts, 13 of which were long terminal repeats (LTRs), which could be involved in inflammatory signaling (**Figure 4.2D**). These data confirm previous reports of an age-related accumulation of RE transcripts and suggest that LTR transcript expression could be linked with declines in cognitive and motor function, which is consistent with ideas reported elsewhere[27,68]. However, the results for O-EX vs. O-SED indicate that additional studies should be conducted to determine whether habitual exercise is an effective strategy to mitigate age-related RE transcript accumulation.

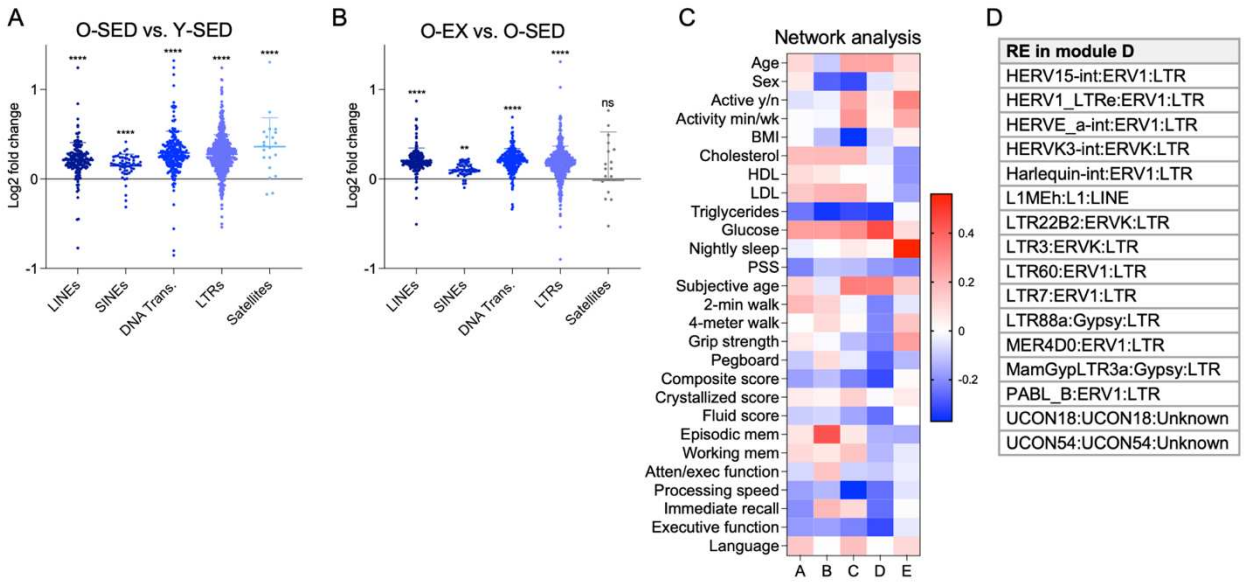


Figure 4.2. Repetitive element (RE) transcripts are increased with age and negatively correlate with function but are not reduced with habitual exercise. A) Violin plot of differential expression of the 5 major types of RE in O-SED vs. Y-SED, and **B)** O-EX vs. O-SED. **C)** WGCNA of RE transcript counts, participant characteristics, markers of health, and cognitive and motor function domains. **D)** Specific RE in module D, which negatively correlates with cognitive and motor function domains.

Genes and RE that are related to cognitive and motor function are also differentially methylated

To determine if upstream, epigenetic changes might underlie the gene expression differences we observed with aging and exercise, we performed whole genome bisulfite sequencing (WGBS) on DNA from 24 of the 40 subject samples (8 from each group) to identify differentially methylated regions (DMRs) in the genome. In O-SED vs. Y-SED, we found numerous significantly hypomethylated DMRs on each chromosome (**Figure 4.3A**), and a similar pattern was reflected in the number of significantly hypermethylated DMRs in O-EX vs. O-SED (**Figure 4.3B**). Next, we analyzed DMRs in O-SED vs. Y-SED and O-EX vs. O-SED, and then we determined whether the gene and RE transcripts that correlated with our functional data were located in these regions. We found that out of the 2,054 genes that negatively correlated with cognitive/motor function (in Figure 1F, module E), 239 met the following criteria: 1) increased (“up”) RNA transcripts with aging but decreased (“down”) transcripts with exercise; and 2) located in DMRs that were significantly hypomethylated with aging and hypermethylated

with exercise (**Figure 4.3C**). GO analysis of these 239 genes showed that the most specific biological processes they were associated with included lymphocyte differentiation, cell signaling, and nervous system development, among other processes. We also determined differential methylation of the RE that were negatively correlated with cognitive/motor function in our RNA-seq data (in Figure 2C, module D). Using the same approach as above, we identified RE transcripts that met the following criteria: 1) increased ("up") transcripts with aging but decreased ("down") transcripts with exercise; and 2) located in DMRs that were significantly hypomethylated with aging and hypermethylated with exercise. We found that only 4 RE met these criteria (L1MeEh, LTR7, MamGypLTR3a, and MER4D0) (**Figure 4.3D**). In general, L1 and LTR RE like these we identified are thought to form virus-like molecular patterns that induce inflammatory signaling. However, given this short list of RE, these results do not support further analyses using RE expression to predict functional outcomes.

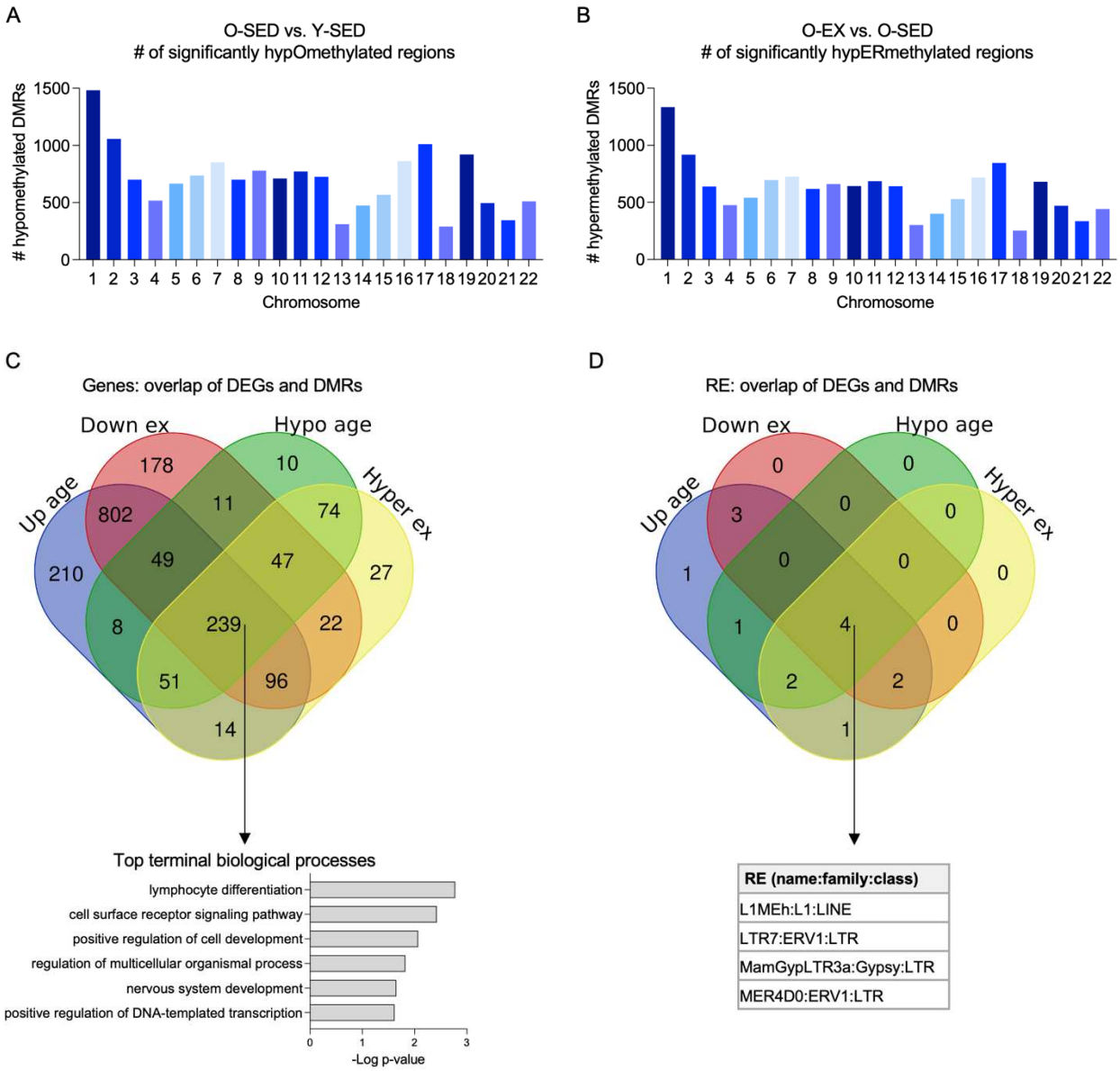
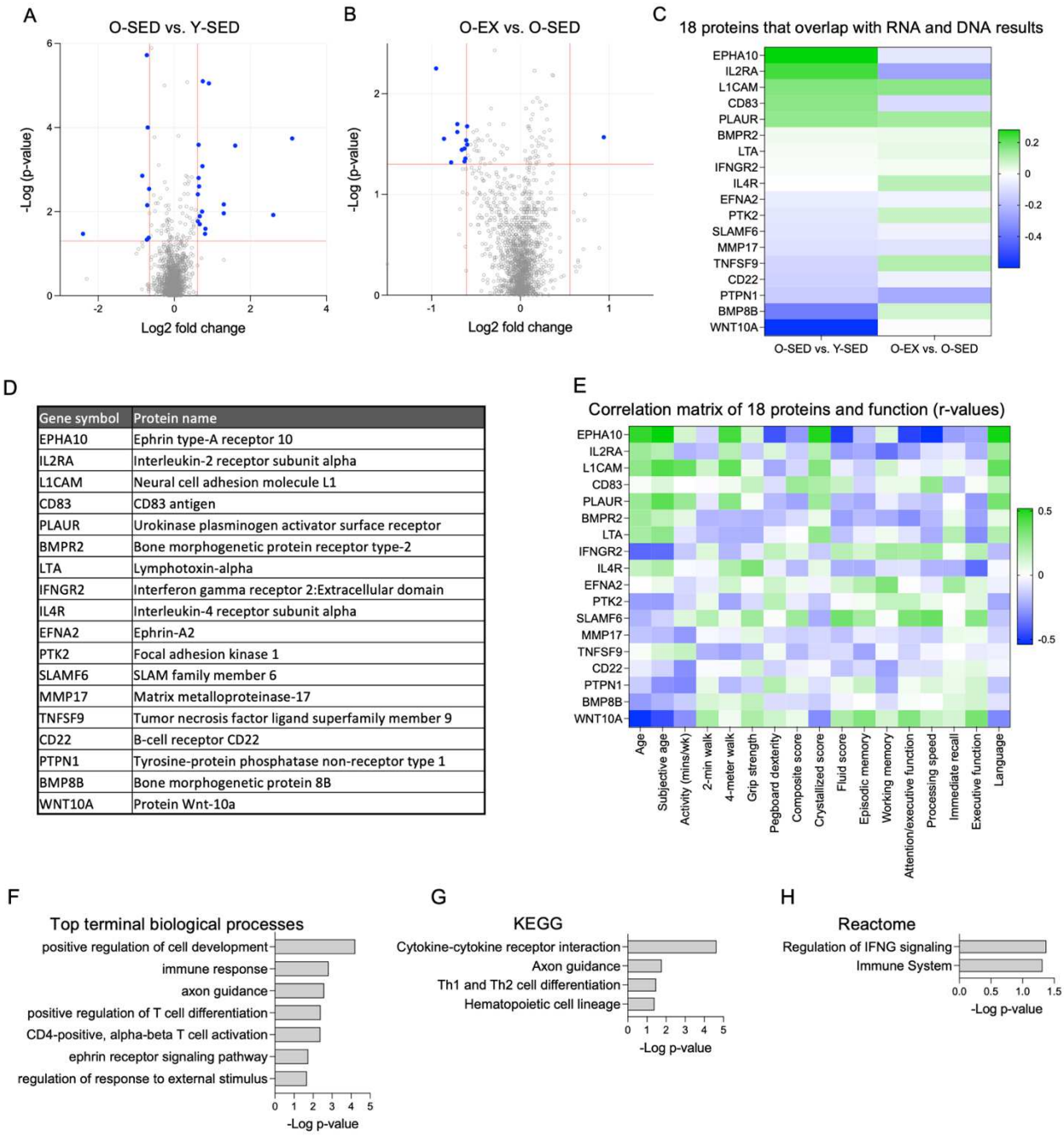


Figure 4.3. Methylation of genes that are differentially expressed with age and exercise. **A)** Hypomethylated DMRs in O-SED vs. Y-SED. **B)** Hypermethylated DMRs in O-EX vs. O-SED. **C)** Venn diagram overlap of transcripts that are increased and genes that are hypomethylated in O-SED vs. Y-SED with transcripts that are decreased and genes that are hypermethylated in O-EX vs. O-SED, along with terminal GO biological processes associated with these overlapping genes. **D)** Venn diagram overlap of RE transcripts that are modulated by age and exercise in both the RNA-seq and WGBS data.

Circulating inflammatory signaling proteins can be used to accurately predict cognitive and motor function in adults

Our transcriptome and epigenome analyses indicated that altered methylation and expression of pro-inflammatory genes might underlie and predict differences in physiological function. Therefore, we next looked for genes dysregulated at the genomic and transcriptomic level, and detectable as proteins in circulation (which could be the actual mediators of differences in physiological functions). We measured levels of 1500 circulating proteins in plasma, and differential protein expression analyses showed some significant differences in O-SED vs. Y-SED (**Figure 4.4A**) and O-EX vs. O-SED (**Figure 4.4B**). Of the 239 genes identified in our RNA-seq and differential methylation overlap analyses (Figure 3C), 18 of them coded for detectable proteins (**Figure 4.4C-D**). Of these proteins, 6 were higher in O-SED vs. Y-SED, and 9 were lower. Some proteins that increased with aging were reversed with exercise (and vice versa for some proteins that decreased with aging), but these effects were not particularly robust (**Figure 4.4C**). Nevertheless, several of these individual proteins correlated strongly with key functional measurements (**Figure 4.4E**). GO analysis of the genes that code for these 18 proteins revealed biological processes including cell development, immune response, T-cell differentiation and activation, and response to external stimuli (**Figure 4.4F**). KEGG pathways included cytokine-cytokine receptor interaction and axon guidance (**Figure 4.4G**), and Reactome processes included interferon signaling and immune system (**Figure 4.4H**). These observations suggest that protein correlates (and perhaps mediators) of age-related differences in cognitive and motor functions are largely related to inflammation and immune signaling, which is consistent with our transcriptome signature findings above.



Gene symbol	Protein name
EPHA10	Ephrin type-A receptor 10
IL2RA	Interleukin-2 receptor subunit alpha
L1CAM	Neural cell adhesion molecule L1
CD83	CD83 antigen
PLAUR	Urokinase plasminogen activator surface receptor
BMPR2	Bone morphogenetic protein receptor type-2
LTA	Lymphotoxin-alpha
IFNGR2	Interferon gamma receptor 2:Extracellular domain
IL4R	Interleukin-4 receptor subunit alpha
EFNA2	Ephrin-A2
PTK2	Focal adhesion kinase 1
SLAMF6	SLAM family member 6
MMP17	Matrix metalloproteinase-17
TNFSF9	Tumor necrosis factor ligand superfamily member 9
CD22	B-cell receptor CD22
PTPN1	Tyrosine-protein phosphatase non-receptor type 1
BMP8B	Bone morphogenetic protein 8B
WNT10A	Protein Wnt-10a

Figure 4.4. Transcriptome and epigenome differences with age and exercise are represented by 18 circulating proteins. A) Volcano plot of differentially expressed proteins in O-SED vs. Y-SED, and **B)** O-EX vs. O-SED. **C)** Heat map of Log2 fold change in the expression of the 18 proteins that are 1) increased ("up") with aging but decreased ("down") with exercise; 2) located in DMRs that were significantly hypomethylated with aging and hypermethylated with exercise; and 3) detectable via targeted proteomics. **D)** List of full protein names. **E)** Correlation matrix of the 18 proteins and age, markers of health, and cognitive and motor function domains. **F)** Top terminal GO biological processes, **G)** KEGG terms, and **H)** Reactome terms associated with the genes that code for these 18 proteins.

Importantly, the 18 proteins identified in our multi-omics approach were relevant to hallmarks of aging such as inflammation that might impact physiological function. Therefore, we performed stepwise regression analyses and found that levels of these proteins could be used to accurately predict age, fluid score, executive function, processing speed, attention, episodic memory, grip strength, and 4-meter walk times (**Figure 4.5 A-L**). Collectively, these results, generated using our multi-omics approach, indicate that a signature of inflammation and immune signaling, which involves genomic/transcriptomic dysregulation and subsequent changes in protein expression, may underlie age-related cognitive and motor function declines.

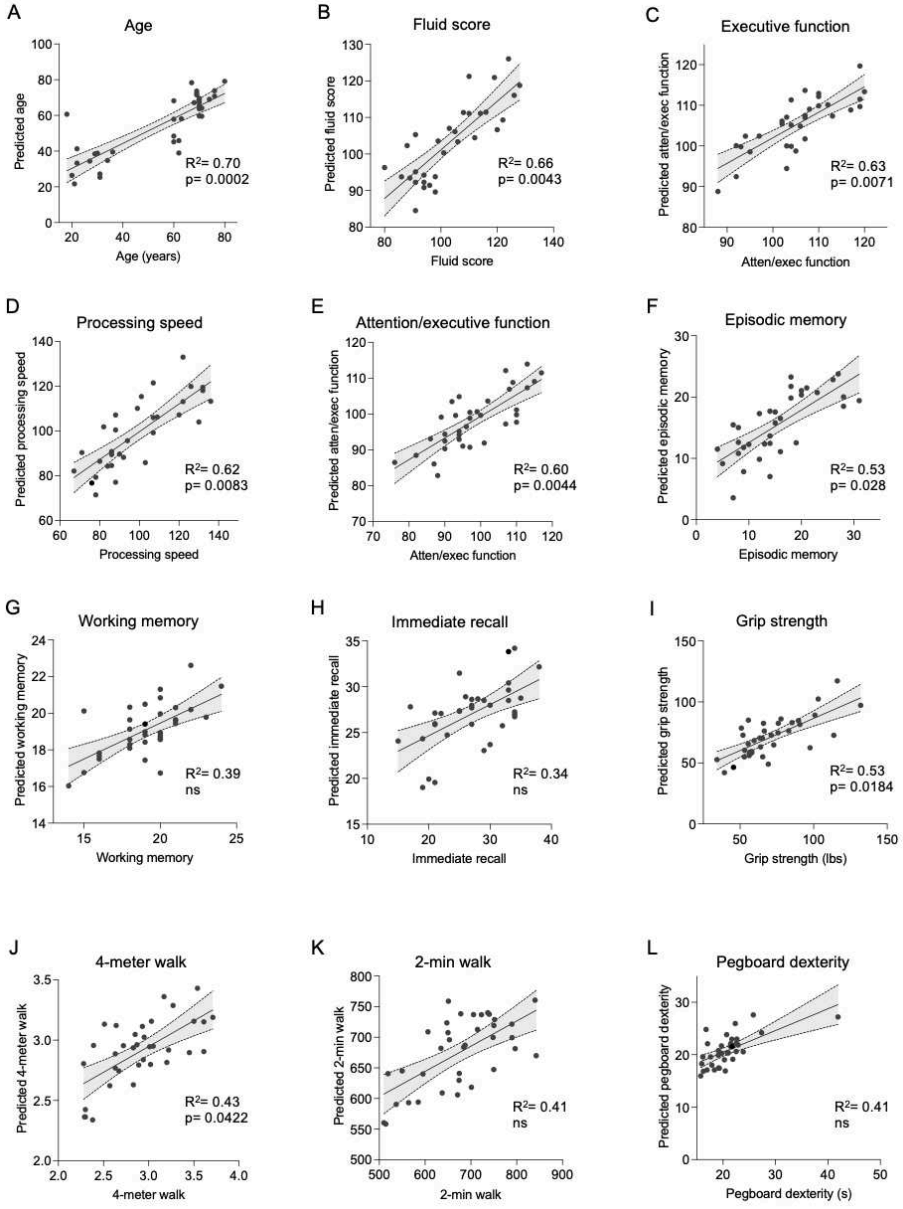


Figure 4.5. Circulating protein signature of inflammatory and immune signaling predicts age, cognitive and motor function in young and older adults. Prediction regressions for **A)** age, **B)** fluid score, **C)** executive function, **D)** processing speed, **E)** attention/executive function, **F)** episodic memory, **G)** working memory, **H)** immediate recall, **I)** grip strength, **J)** 4-meter walk time, **K)** 2-min walk distance, **L)** pegboard dexterity time.

Discussion

Determining underlying mechanisms and biomarkers of healthspan-related differences in physiological function is an important goal of research on aging. In this study, we used age, habitual exercise, and multi-omics as a framework to identify genes/proteins that might underlie differences in cognitive and motor function in healthy older adults. We found that age-related changes to the epigenome, transcriptome and proteome are largely related to inflammatory and immune signaling, key hallmarks of aging that could be mechanistically linked to changes in cognitive and motor function, and we identified 18 specific genes/proteins that could specifically drive these events. These findings provide novel insight into potential mediators of age-related functional decline that are easily-accessible in humans and may be used to detect targets for therapeutic interventions in the future.

Declines in physiological function precede and predict the risk for chronic diseases. In our study, we assessed key functions using the NIH Toolbox cognitive and motor function testing batteries. We were able to detect age-related functional differences between young and older participants, which is expected and has been demonstrated and reproduced in several studies[25,26,41,43-45]. The Toolbox scoring system provides absolute/uncorrected scores as well as age-adjusted national percentile scores, which are relative to normative values for each age group. For the Flanker attention/executive function test, which has been ranked as one of the most important domains in the Toolbox[25], we found that both older subject groups had significantly lower uncorrected scores than the young group, but that the older exercising group had a significantly higher score than the older sedentary group. In addition, the older sedentary

group performed significantly lower than young in terms of age-adjusted national percentile, while the older exercising group was not different from Y-SED or O-SED. In the card sort test, another domain of executive function, all groups were the same in terms of age-adjusted national percentile, but O-SED scored significantly lower than Y-SED, and O-EX was not different from Y-SED or O-SED. For motor function measures like the 2-minute walk test, O-EX performed significantly better than O-SED in terms of age-adjusted national percentile, indicating that habitual exercise may rescue/prevent decline in these domains as well. Some intervention studies have shown that exercise improves cognitive and motor function in previously sedentary older adults, but these studies involved resistance training as opposed to aerobic[41] exercise, and some were smaller, pilot studies[42]. One study assessed self-reported physical activity in relation to Toolbox measurements in young and older adults using the International Physical Activity Questionnaire Short Form (IPAQ-SF) but found no relationship between vigorous activity MET-minutes per week or total physical activity and cognitive test results[43]. Thus, to the best our knowledge, ours are the first cross-sectional data using NIH Toolbox measurements to show differences between age-matched sedentary and habitually active older adults. A limitation of our study is that it was limited to functional measurements in the NIH Toolbox cognitive and motor function batteries, which are designed to be short, standardized, and reproducible across research settings. Other functional measurements, such as exercise testing/maximal aerobic capacity could be more robust assessments for discerning differences among these groups in future studies. Adding tests like these to similar studies in the future could help identify additional genes/proteins linked with clinically-relevant measurements of health and function.

Microarrays and, more recently, next-generation RNA-seq have made it possible to broadly profile potential transcriptome mediators of health and functional differences like those observed here[18,19]. While there were no individual differentially expressed genes in the overall transcriptomes for O-SED vs. Y-SED and O-EX vs. O-SED in this study, we did identify general

patterns of differential gene expression that were reversed with habitual exercise. Also, using WGCNA we identified a cluster of 2,054 transcripts that was negatively related to nearly all domains of cognitive and motor function, as well as HDL cholesterol, and positively related to age, subjective age and fasting blood glucose. This cluster was largely made up of genes coding for interferons, interleukins, cytokine cell markers, growth factors, extracellular signal-regulated kinases, and serpins, and their transcripts mapped to several GO biological processes, KEGG, and Reactome terms that were mostly related to cellular responses to cytokine stimulus, stress, inflammation, and immune cell signaling. The exact processes by which these transcriptome events could cause functional declines are unclear. However, several of these genes/transcripts (such as SERPINH1, IGFBP4, TAGLN) have been linked to the senescence-associated secretory phenotype (SASP), which includes cytokines, growth factors, and serpins (serine protease inhibitors that contribute to inflammation) [51,52]. Aging is associated with accumulation and reduced clearance of senescent cells (which reduce the function of their resident tissues), and also an increase in circulating SASP factors. Recent studies have demonstrated that SASP factors are negatively associated with muscular strength, a key domain of motor function, in older adults[53], and that aerobic exercise reduces the number of senescent cells and circulating SASP factors in older adults[54,55]. Our data provide novel insight suggesting that circulating SASP factors may contribute to impaired cognitive function with aging as well, consistent with the concept of inflammation as a central hallmark of aging. Our findings could also provide a basis for future studies that might include knockdown or overexpression of the candidate proteins we identified in animal models, or exercise/other healthspan-extending intervention studies in humans in which protein expression and function are measured pre- and post-intervention.

Although not presented in the primary figures of this dissertation, we performed additional analyses of RNA transcripts specifically linked to the SASP, and found that the majority of SASP factors with higher expression in O-SED vs. Y-SED have lower expression in

O-EX vs. O-SED (**Appendix 1, Figure 6.1A**). We also profiled SASP factors in our proteomics data and found that some proteins that were differentially expressed in O-SED vs. Y-SED were changed in the reverse direction by exercise, but there was no clear pattern at the protein level (**Appendix 1, Figure 6.1B**). Circulating levels of specific proteins known to be linked with senescence such as p21, p53, and IL-6 were not different in Y-SED compared to O-SED or O-EX (**Appendix 1, Figure 6.1C**). In addition, to examine the potential role of RE transcripts (which have been reported to be linked with the SASP) in these events, we performed a cell culture experiment to determine if reverse transcriptase inhibitor Lamivudine (3TC, which inhibits RE transcript activity) could reduce inflammatory signaling in PBMCs isolated from our participants. There were no statistically significant differences in IL-6 concentration in cell culture supernatant from Y-SED, O-SED, or O-EX with SASP stressor (H_2O_2) treatment or 3TC (**Appendix 1, Figure 6.2A**). However, when we grouped Y-SED, O-SED, and O-EX PBMCs together, there was a trend ($p = 0.05$) toward reduced IL-6 with 3TC treatment in the cells treated with H_2O_2 to induce cellular stress and senescence/SASP signaling (**Appendix 1, Figure 6.2B**). These experiments bear repeating with improvements. In particular, we would establish ideal conditions for cells to be less stimulated at baseline and add another positive control such as a cytokine cocktail that would robustly increase the IL-6 signal from baseline. With these improvements, we could determine whether 3TC can be used to reduce stress-induced inflammatory signaling, which is related to aging and the SASP. Taken together with our unexpected finding that RE transcript expression was *higher* in O-EX compared to O-SED and the data described above (Appendix 1, Figure 1C), these cell culture results underscore the need for future experiments aimed at determining the effects of aging and exercise on RE transcript expression.

Recent studies from our lab and others suggest that transcriptome signatures of inflammation increase with age, along with an accumulation of transcripts from RE. These RE transcripts can form virus-like molecular patterns in the cytoplasm which may drive inflammatory

signaling. Here, we confirmed an age-related increase in RE transcripts in older vs. younger adults, but we did not find a reduction in RE transcripts in older exercising adults. In fact, LINE, SINE, DNA transposon, and LTR transcript counts were all slightly greater in older exercising vs. sedentary adults. This finding was surprising and could be a result of the different RNA-seq approach we used here (polyA rather than total RNA, which could capture a smaller/less representative portion of RE transcripts). Nevertheless, in a network analysis of RE we identified one module (D) which contained 16 RE transcripts that were negatively related top all domains of cognitive and motor function (except for crystallized score and language, which were higher in older adults compared to young). Module D was also positively related to age, subjective age, and fasting blood glucose. Interestingly, the majority of RE in module D were LTRs, which have the potential to form virus-like molecular patterns in the cytoplasm and stimulate an immune response[10 56]. Among these was LTR22B2, an endogenous retrovirus K (ERV-K) in the HML-2 group, which is considered to be the most active group of ERVs. In fact, recently published data show that ERV/HERV-Ks directly promote inflammatory signaling and tissue aging/premature senescence[57]. While our results may support other studies implicating RE in inflammatory signaling, future studies are needed to confirm their role in exercise. Such studies could include assessing the acute effects of an exercise bout on RE expression, and/or training interventions of different lengths/exercise modalities, which would provide more insight into this emerging area of research.

Gene expression differences with aging/exercise like those we observed here are often associated with upstream epigenetic differences. Therefore, we analyzed differentially methylated regions in DNA samples from our participants to characterize potential epigenetic (methylation) differences in the 2,054 genes and 16 RE that were correlated with cognitive/motor function in our RNA-seq data. To identify potential “key mediators” of functional declines, we focused on identifying genes/transcripts among these genes and RE that met the following criteria: 1) increased (“up”) transcripts with aging but decreased (“down”) transcripts

with exercise; and 2) located in genome regions that were significantly hypomethylated with aging and hypermethylated with exercise. Using this approach, we identified 239 genes/transcripts that reflected multiple biological processes (GO terms), including immune cell activity and nervous system development. Given that these genes/transcripts negatively correlated with cognitive function in our participants, “nervous system development” is an interesting finding. While these RNA-seq and WGBS data are derived from whole blood, it is possible that brain-derived transcripts are detectable in circulation, potentially via extracellular vesicles that cross the blood-brain barrier. In fact, one of the protein predictors of cognitive function in this study is L1CAM, a cell-surface marker of neuronal extracellular vesicles[58]. These findings indicate that circulating epigenetic profiles and transcriptome signatures (which are easily/clinically accessible from whole blood) also may be physiologically relevant to age-related processes.

Interestingly, when we analyzed RE transcript expression and differential methylation in our samples, we identified only four RE (L1MeEh, LTR7, MamGypLTR3a, and MER4D0) that were increased with aging but reduced with exercise, and hypomethylated with aging but hypermethylated with exercise. A recent study demonstrated that L1MEh, a LINE-1 element, is expressed by oligodendrocytes and microglia in the brain[59], and others have linked L1MEh hypomethylation with cancer[60]. LTR7 and MER4D0 expression have also been linked with various cancers[61,62], which are increasingly prevalent with age. The small number of RE resulting from this analysis is likely due to the fact that very few RE were decreased in O-EX vs. O-SED, which is a finding that differs from recently published reports on RE transcript accumulation and habitual exercise. As noted above, it is possible that PolyA RNA-seq is not an ideal method for detecting RE transcripts (as opposed to total RNA-seq), or that RE transcript accumulation is different in whole blood than in isolated peripheral blood mononuclear cells or fibroblasts (which are the sources used in other studies[8,9]). In any case, more studies profiling

age-related RE transcript expression in different cell types are necessary to determine the acute and chronic effects of exercise on RE transcript accumulation.

Finally, to identify omics changes with aging that might be the most proximal to related functional changes (and easily measured in humans), we intersected our RNA-seq and methylation data with plasma proteomics. Of the 239 genes/transcripts that were modulated by age and exercise and correlated with cognitive function, 18 of them were detected in our proteomics data, and some were differentially expressed in with aging. These 18 proteins were strongly related to inflammatory and immune signaling, and most of them proteins were significantly correlated with age and various domains of health and function. These data suggest that circulating inflammatory immune-signaling proteins are linked to cognitive and motor function in young and older adults. This approach allowed us to refine a list of candidate biological mediators of functional declines with aging, a primary goal of current research on aging. Importantly, this is not the first omics study to use exercise as a model of improved healthspan. Others have used epigenetic/methylation clocks[12,16,17], transcriptomics[8,9,63,64], and proteomics signatures[20,51] to show that habitual exercise or exercise interventions reduce biological age. Here, we combined three types of omics to narrow down predictors of biological age that are functionally relevant to cognitive and motor aging in humans. Using this multi-omics "funnel", we identified 18 functionally relevant proteins, and found that their expression accurately predicted age, cognitive fluid scores, executive function, processing speed, attention, episodic memory, grip strength, and 4-meter walk gait speed. Importantly, executive function and episodic memory are ranked as the top two most important subdomains for perceived importance for health[25]. Moving forward, longitudinal studies could be conducted to determine whether these circulating proteins can be used as biomarkers to predict future functional decline. If so, they could be employed as clinical predictors to identify areas for preventative therapeutic intervention.

While these observations add insight to potential drivers of aging and healthspan, it is important to note that our study was cross-sectional and does not provide direct, mechanistic results. Future studies including exercise and drug interventions in humans and cell/animal model studies targeting the proteins we identified are necessary to determine whether these observations are mechanistically relevant. Another limitation of our study design is that we did not include a young habitually exercising group or any middle-aged participants. In order to identify markers of healthspan, it will be important to study individuals across the entire spectrum of age, activity level, and health/disease state in the future.

To our knowledge, this is the first study to collect health, cognitive and motor function, transcriptomic, epigenomic, and proteomic data in the same individuals using age and habitual exercise as a framework for identifying mediators of healthspan. Additional studies are needed to replicate these findings in larger groups that are more diverse in terms of functional ability (e.g., people with age-related diseases that may accelerate cognitive and motor decline) to establish these proteins as biomarkers of aging/healthspan.

REFERENCES

1 Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Curr Biol* **22**, R741-752 (2012). https://doi.org:10.1016/j.cub.2012.07.024

2 Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243-278 (2023). https://doi.org:10.1016/j.cell.2022.11.001

3 Demnitz, N. *et al.* Associations between Mobility, Cognition, and Brain Structure in Healthy Older Adults. *Front Aging Neurosci* **9**, 155 (2017). https://doi.org:10.3389/fnagi.2017.00155

4 Roman, Y. The United States 2020 census data: implications for precision medicine and the research landscape. *Personalized Medicine* **19**, 5-8 (2021).

5 Ruiz, E. *et al.* An Integrative Multi-Omics Analysis of The Molecular Links between Aging and Aggressiveness in Thyroid Cancers. *Aging Dis* **14**, 992-1012 (2023). https://doi.org:10.14336/ad.2022.1021

6 Girerd, N. *et al.* Protein Biomarkers of New-Onset Heart Failure: Insights From the Heart Omics and Ageing Cohort, the Atherosclerosis Risk in Communities Study, and the Framingham Heart Study. *Circ Heart Fail* **16**, e009694 (2023). https://doi.org:10.1161/circheartfailure.122.009694

7 Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat Rev Genet* **23**, 715-727 (2022). https://doi.org:10.1038/s41576-022-00511-7

8 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R. & LaRocca, T. J. Healthy aging interventions reduce repetitive element transcripts. *J Gerontol A Biol Sci Med Sci* (2020). https://doi.org:10.1093/gerona/glaa302

9 LaRocca, T. J., Cavalier, A. N. & Wahl, D. Repetitive elements as a transcriptomic marker of aging: Evidence in multiple datasets and models. *Aging Cell* **19**, e13167 (2020). https://doi.org:10.1111/acel.13167

10 Gorbunova, V. *et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43-53 (2021). https://doi.org:10.1038/s41586-021-03542-y

11 Kreiling, J. A. *et al.* in *Human Retrotransposons in Health and Disease* (ed Gael Cristofari) 297-321 (Springer International Publishing, 2017).

12 Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)* **11**, 303-327 (2019). https://doi.org:10.18632/aging.101684

13 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* **19**, 371-384 (2018). https://doi.org:10.1038/s41576-018-0004-3

14 Zhang, M. *et al.* Genetic and epigenetic study of ALS-discordant identical twins with double mutations in SOD1 and ARHGEF28. *J Neurol Neurosurg Psychiatry* **87**, 1268-1270 (2016). https://doi.org:10.1136/jnnp-2016-313592

15 Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**, 10604-10609 (2005). https://doi.org:10.1073/pnas.0500398102

16 McGreevy, K. M. *et al.* DNAmFitAge: biological age indicator incorporating physical fitness. *Aging (Albany NY)* **15** (2023). https://doi.org:10.18632/aging.204538

17 Jokai, M. *et al.* DNA methylation clock DNAmFitAge shows regular exercise is associated with slower aging and systemic adaptation. *Geroscience* (2023). https://doi.org:10.1007/s11357-023-00826-1

18 Meyer, D. H. & Schumacher, B. BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. *Aging Cell* **20**, e13320 (2021). https://doi.org:10.1111/acel.13320
19 Shokhirev, M. N. & Johnson, A. A. Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell* **20**, e13280 (2021). https://doi.org:10.1111/acel.13280
20 Sathyan, S. *et al.* Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell* **19**, e13250 (2020). https://doi.org:10.1111/acel.13250
21 Levin, O., Netz, Y. & Ziv, G. The beneficial effects of different types of exercise interventions on motor and cognitive functions in older age: a systematic review. *Eur Rev Aging Phys Act* **14**, 20 (2017). https://doi.org:10.1186/s11556-017-0189-z
22 Garatachea, N. *et al.* Exercise attenuates the major hallmarks of aging. *Rejuvenation Res* **18**, 57-89 (2015). https://doi.org:10.1089/rej.2014.1623
23 Rebelo-Marques, A. *et al.* Aging Hallmarks: The Benefits of Physical Exercise. *Front Endocrinol (Lausanne)* **9**, 258 (2018). https://doi.org:10.3389/fendo.2018.00258
24 Valenzuela, P. L. *et al.* Effects of physical exercise on physical function in older adults in residential care: a systematic review and network meta-analysis of randomised controlled trials. *Lancet Healthy Longev* (2023). https://doi.org:10.1016/s2666-7568(23)00057-0
25 Weintraub, S. *et al.* Cognition assessment using the NIH toolbox. *American Academy of Neurology* **3**, 54-64 (2013).
26 Reuben, D. B. *et al.* Motor assessment using the NIH Toolbox. *American Academy of Neurology* **3**, 65-75 (2013).
27 Wahl, D. *et al.* The reverse transcriptase inhibitor 3TC protects against age-related cognitive dysfunction. *Aging Cell*, e13798 (2023). https://doi.org:10.1111/acel.13798
28 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560
29 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
30 Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593-3599 (2015). https://doi.org:10.1093/bioinformatics/btv422
31 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8
32 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019). https://doi.org:10.1093/nar/gkz369
33 Carlson, M. (R package version 3.8.2., 2019).
34 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). https://doi.org:10.1186/1471-2105-9-559
35 de Sena Brandine, G. & Smith, A. D. Fast and memory-efficient mapping of short bisulfite sequencing reads using a two-letter alphabet. *NAR Genom Bioinform* **3**, lqab115 (2021). https://doi.org:10.1093/nargab/lqab115
36 Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**, e81148 (2013). https://doi.org:10.1371/journal.pone.0081148
37 Timsina, J. *et al.* Comparative Analysis of Alzheimer's Disease Cerebrospinal Fluid Biomarkers Measurement by Multiplex SOMAscan Platform and Immunoassay-Based Approach. *J Alzheimers Dis* **89**, 193-207 (2022). https://doi.org:10.3233/jad-220399

38 Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010). https://doi.org:10.1371/journal.pone.0015004
39 Robbins, J. M. *et al.* Human plasma proteomic profiles indicative of cardiorespiratory fitness. *Nat Metab* **3**, 786-797 (2021). https://doi.org:10.1038/s42255-021-00400-z
40 Garber, C. E. *et al.* American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* **43**, 1334-1359 (2011). https://doi.org:10.1249/MSS.0b013e318213fefb
41 Macaulay, T. R. *et al.* 12 weeks of strength training improves fluid cognition in older adults: A nonrandomized pilot trial. *PLoS One* **16**, e0255018 (2021). https://doi.org:10.1371/journal.pone.0255018
42 Halloway, S., Wilbur, J., Braun, L. T., Schoeny, M. E. & Volgman, A. S. The Feasibility of a Combined Lifestyle Physical Activity and Cognitive Training Intervention to Prevent Cognitive Impairment in Older Women With Cardiovascular Disease. *J Phys Act Health* **18**, 70-75 (2021). https://doi.org:10.1123/jpah.2020-0206
43 Cole, K. R., Yen, C. L., Dudley-Javoroski, S. & Shields, R. K. NIH Toolbox Cognition Battery in Young and Older Adults: Reliability and Relationship to Adiposity and Physical Activity. *J Geriatr Phys Ther* **44**, 51-59 (2021). https://doi.org:10.1519/jpt.0000000000000244
44 Brunt, V. E. *et al.* The gut microbiome-derived metabolite trimethylamine N-oxide modulates neuroinflammation and cognitive function with aging. *Geroscience* (2020). https://doi.org:10.1007/s11357-020-00257-2
45 Santos-Parker, J. R. *et al.* Curcumin supplementation and motor-cognitive function in healthy middle-aged and older adults. *Nutr Healthy Aging* **4**, 323-333 (2018). https://doi.org:10.3233/NHA-170029
46 Melov, S., Tarnopolsky, M. A., Beckman, K., Felkey, K. & Hubbard, A. Resistance exercise reverses aging in human skeletal muscle. *PLoS One* **2**, e465 (2007). https://doi.org:10.1371/journal.pone.0000465
47 Kohman, R. A. *et al.* Voluntary wheel running reverses age-induced changes in hippocampal gene expression. *PLoS One* **6**, e22654 (2011). https://doi.org:10.1371/journal.pone.0022654
48 Bronikowski, A. M. *et al.* Lifelong voluntary exercise in the mouse prevents age-related alterations in gene expression in the heart. *Physiol Genomics* **12**, 129-138 (2003). https://doi.org:10.1152/physiolgenomics.00082.2002
49 Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000). https://doi.org:10.1093/nar/28.1.27
50 Good, B. M. *et al.* Reactome and the Gene Ontology: digital convergence of data resources. *Bioinformatics* **37**, 3343-3348 (2021). https://doi.org:10.1093/bioinformatics/btab325
51 Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol* **18**, e3000599 (2020). https://doi.org:10.1371/journal.pbio.3000599
52 Robbins, P. D. *et al.* Senolytic Drugs: Reducing Senescent Cell Viability to Extend Health Span. *Annu Rev Pharmacol Toxicol* **61**, 779-803 (2021). https://doi.org:10.1146/annurev-pharmtox-050120-105018
53 Fielding, R. A. *et al.* Associations between biomarkers of cellular senescence and physical function in humans: observations from the lifestyle interventions for elders (LIFE) study. *Geroscience* **44**, 2757-2770 (2022). https://doi.org:10.1007/s11357-022-00685-2

54 Zhang, X. *et al.* Characterization of cellular senescence in aging skeletal muscle. *Nat Aging* **2**, 601-615 (2022). https://doi.org:10.1038/s43587-022-00250-8
55 Englund, D. A. *et al.* Exercise reduces circulating biomarkers of cellular senescence in humans. *Aging Cell* **20**, e13415 (2021). https://doi.org:10.1111/acel.13415
56 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009). https://doi.org:10.1038/nrg2640
57 Liu, X. *et al.* Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell* **186**, 287-304 e226 (2023). https://doi.org:10.1016/j.cell.2022.12.017
58 Pulliam, L., Sun, B., Mustapic, M., Chawla, S. & Kapogiannis, D. Plasma neuronal exosomes serve as biomarkers of cognitive impairment in HIV infection and Alzheimer's disease. *J Neurovirol* **25**, 702-709 (2019). https://doi.org:10.1007/s13365-018-0695-4
59 He, J. *et al.* Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat Commun* **12**, 1456 (2021). https://doi.org:10.1038/s41467-021-21808-x
60 Abante, J., Kambhampati, S., Feinberg, A. P. & Goutsias, J. Estimating DNA methylation potential energy landscapes from nanopore sequencing data. *Sci Rep* **11**, 21619 (2021). https://doi.org:10.1038/s41598-021-00781-x
61 Mao, J., Zhang, Q. & Cong, Y. S. Human endogenous retroviruses in development and disease. *Comput Struct Biotechnol J* **19**, 5978-5986 (2021). https://doi.org:10.1016/j.csbj.2021.10.037
62 Wang, J. *et al.* Single-cell RNA sequencing highlights the functional role of human endogenous retroviruses in gallbladder cancer. *EBioMedicine* **85**, 104319 (2022). https://doi.org:10.1016/j.ebiom.2022.104319
63 Tumasian, R. A., 3rd *et al.* Skeletal muscle transcriptome in healthy aging. *Nat Commun* **12**, 2014 (2021). https://doi.org:10.1038/s41467-021-22168-2
64 Robinson, M. M. *et al.* Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans. *Cell Metab* **25**, 581-592 (2017). https://doi.org:10.1016/j.cmet.2017.02.009
65 Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989 May;28(2):193-213. doi: 10.1016/0165-1781(89)90047-4. PMID: 2748771.
66 Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983 Dec;24(4):385-96. PMID: 6668417.
67 Stephan Y, Sutin AR, Terracciano A. How old do you feel? The role of age discrimination and biological aging in subjective age. PLoS One. 2015 Mar 4;10(3):e0119293. doi: 10.1371/journal.pone.0119293. PMID: 25738579; PMCID: PMC4349738.
68 Guo C, Jeong HH, Hsieh YC, Klein HU, Bennett DA, De Jager PL, Liu Z, Shulman JM. Tau Activates Transposable Elements in Alzheimer's Disease. Cell Rep. 2018 Jun 5;23(10):2874-2880. doi: 10.1016/j.celrep.2018.05.004. PMID: 29874575; PMCID: PMC6181645.

CHAPTER 5: FUTURE DIRECTIONS: OMICS AND OTHER APPROACHES FOR IDENTIFYING MEDIATORS OF HEALTH AND HEALTHSPAN ACROSS DIVERSE AGING COMMUNITIES

Introduction

The studies in this dissertation utilized omics and multi-omics to identify the mechanisms underlying exercise adaptation, aging, and healthspan. These findings set the stage for future studies to determine new clinical biomarkers and mediators of healthspan. Important future directions for this work include:

1) Determining if these findings are reproducible in more diverse populations;
2) Determining if these findings apply to people who do not meet the definition of "healthy."

To address these future directions, I applied for a National Academy of Medicine Catalyst grant (currently under consideration), the primary text from which is included below.

Rationale

In the effort to increase human healthspan, the number of years spent healthy and disease-free, one key research strategy is to identify markers and mediators of biological age that reflect health/disease risk better than chronological age in years using "multi-omics" data. Despite significant progress and investment in this area, mediators and markers of biological aging in racial/ethnic minority, socioeconomically disadvantaged and rural populations remain majorly understudied. This is an important social and biomedical problem because these populations: 1) bear a disproportionate burden of disease; and 2) are at particular risk for adverse social, biological, behavioral, and environmental stresses that also influence health trajectories in the broader population (limiting our ability to serve these populations *and* our understanding of healthspan in general).

To address this problem, we propose to develop a mobile healthspan research unit (a.k.a., the “healthspan van”). This mobile lab will equip us to collect multi-omics (genome, transcriptome, proteome, etc.), molecular, and functional biomarkers of healthspan in any location, *and* to educate the public about healthy aging, removing barriers to entry for people in understudied groups. We have funding to purchase equipment and pay personnel to initiate this project at our home base, but a catalyst award would enable us to purchase the van itself and mobilize our research.

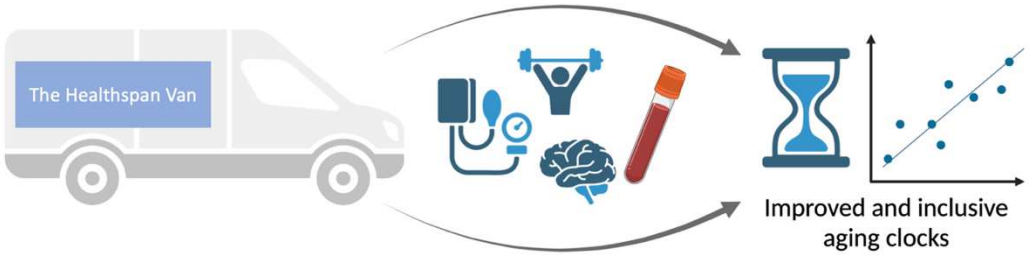


Figure 5.1. Using mobile research to develop improved and inclusive aging clocks.

Our rationale for the healthspan van/multi-omics approach is as follows: First, most biological aging research has focused on molecular markers (e.g., using omics to detect DNA/RNA or circulating metabolites). Recent studies have combined these molecular markers with measures of physiological health/function, and the newest data show that a combination of multi-omics and these functional measures accurately predicts different aging rates among individuals. These integrative “aging clocks” reflect many factors (social, biological, behavioral, environmental) that influence the pace of aging, and they may improve our ability to predict individual health trajectories and prescribe interventions. However, most of these clocks have been developed using samples from people who are racially, socioeconomically, and geographically similar (mostly living near research institutions or universities). We hypothesize that including a diverse population of samples in the construction of multi-omics aging clocks will improve our ability to predict healthspan indicators such as standard cardiovascular risk factors,

cognitive function, motor function, and perceived stress**.** The ability to collect data and perform more inclusive sampling outside of the university bubble will enable the development of these "better clocks" based on diverse, more representative populations—a necessary step toward mitigating health disparities.

A second but important part of our rationale is that despite the wealth of knowledge that *does* exist on evidence-based strategies for increasing healthspan, these same underserved populations typically have less access and exposure to this information. By coupling our research efforts with informational sessions and raising awareness about the importance of healthy aging in general (especially with a physical, local presence), we will have a broader impact beyond our research efforts.

Long-term, given the growing importance of data-driven mobile health, we believe this two-pronged strategy could be the starting point for the future of healthspan-focused medicine (e.g., a fleet of healthspan vans that collect, integrate and disseminate data, including personalized healthy aging prescriptions based on the best, population-representative multi-omics clocks we can develop). With the population of adults over the age of 65 growing rapidly, and increased public interest in longevity science, it is imperative that advances in this kind of healthspan research benefit the most people possible.

Implementation and testing

As a team, we have access to many of the resources necessary to complete this project, including an Institutional Review Board (IRB), biological sampling equipment, data analysis software, and several colleagues and mentors who are experts in biological aging and related fields. In fact, we have already collected biological aging data from 54 participants at our (non-mobile) research facility. During a ~1.5-hour protocol, we collect:

- Basic health measures such as blood pressure, height, and weight;
- Blood samples for:

    - isolation and sequencing of DNA and RNA, proteomic analysis of inflammatory and senescence biomarkers associated with aging;
    - health measures such as cholesterol and blood glucose;
- Cognitive function measures such as episodic memory, processing speed, and executive function;
- Physical/motor function measures such as grip strength, balance, and walking speed;
- Surveys to estimate perceived life stress, sleep quality, physical activity, and other lifestyle factors.

These study visits are currently completed in a small, 8x8 foot room, and can easily be adapted for a mobile unit. While we have the resources and funding available to collect these data, with the support of this award, we will purchase a van to take this research protocol on the road. We are ideally positioned to enhance our understanding of healthspan in different populations in a state that has socioeconomically and geographically diverse regions, large differences in life expectancy by county and significant populations of underrepresented minorities. To test our idea, we propose to collect data in three locations that are diverse in terms of race, ethnicity, household income (and include a mixture of rural and urban living). We have already established collaborations with local health agencies at these locations and, with them, a plan for a pilot study that would be supported by this catalyst award. Our implementation/analysis plan is as follows:

1. Purchase and equip a van with all the necessary equipment to perform our clinical protocol outlined above.
2. Conduct a first healthspan van “tour” to demonstrate feasibility and generate an initial dataset.
    a. Schedule public events based on community interests to promote evidence-based information on healthy aging and recruit research participants.

b. Visit each location as needed to host events and study 20 older adults (60 total).

3. Perform multi-omics/clock analyses (transcriptomics, epigenomics and proteomics) on samples and:
   a. Determine if established biological aging clocks predict differences in healthspan by region and in underrepresented or underserved populations.
   b. Use network analyses to relate omics data to health metrics and identify socioeconomic, behavioral, and environmental effects on them.
   c. Develop a "better aging clock" based on our findings, and determine if it predicts differences in healthspan by region and in underrepresented or underserved populations (as well as in adults studied in our existing protocols) more accurately that existing clock algorithms.

To establish that our idea is feasible/effective, we will perform the analyses outlined in 3a-c above and share our findings, including any alternative/better biological aging clock models we develop, with other researchers in the field so that they can test the models on their own datasets and help us confirm (or refute) their validity. Another measure of success will be community participation and perception. We will administer surveys to community members and participants to gauge whether they want us to come back for repeat measures/seminars. If we are successful, we will use the data collected in our pilot project to apply for funding to extend this project to cover more geography *and* to revisit the same sites for the collection of longitudinal data. Ideally, this will develop into a multi-site project so that we can collaborate with multiple universities and collect data in other states where there may be even larger health disparities.

APPENDIX 1: ADDITIONAL FIGURES

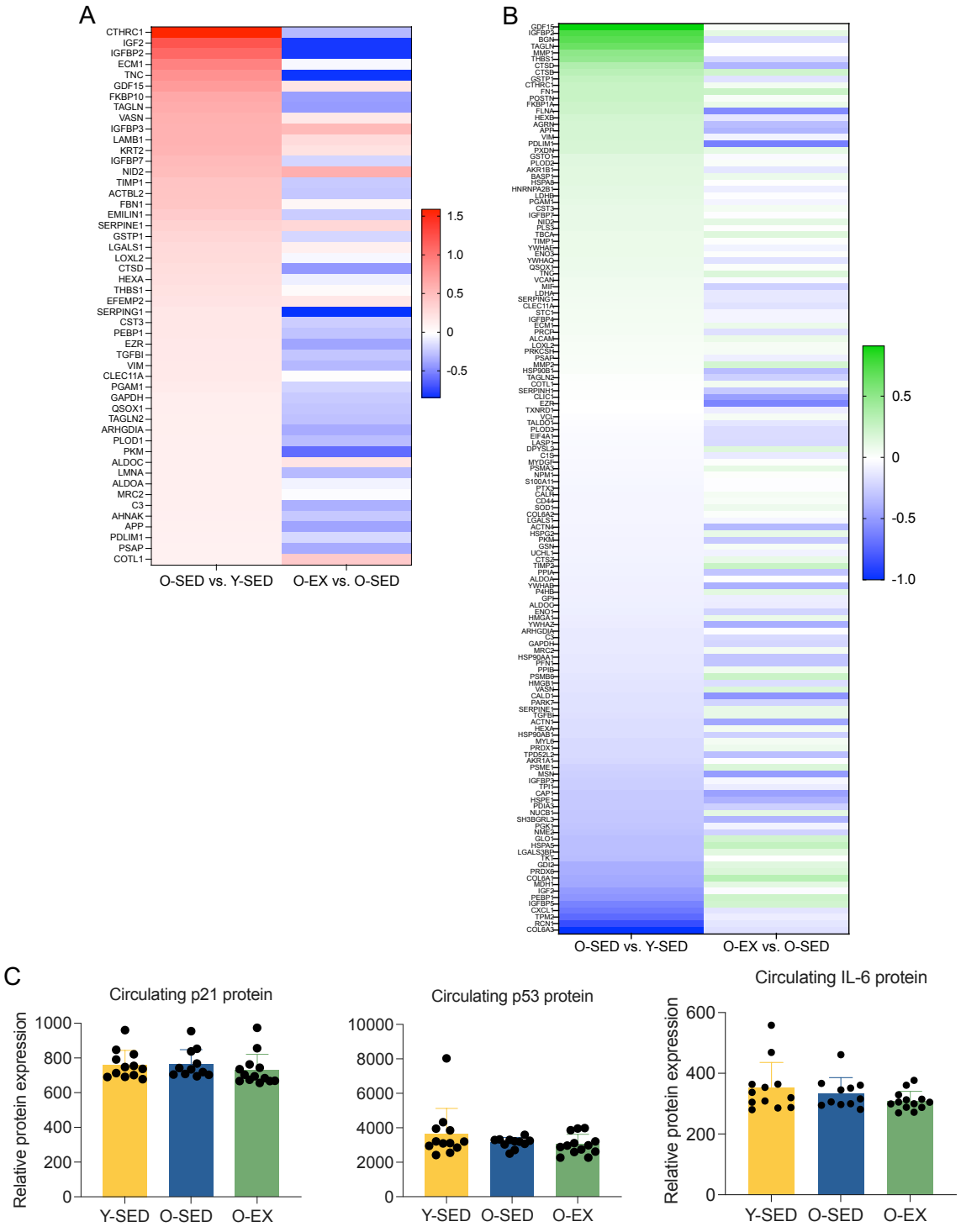


Figure 6.1. Circulating markers of senescence in RNA-seq and proteomics data. A) Top 50 senescence-associated secretory phenotype (SASP) factors that were increased with aging in RNA-seq data on O-SED vs. Y-SED and O-EX vs. O-SED subjects; **B)** 140 SASP factors from plasma proteomics and their log2 fold difference in the same O-SED vs. Y-SED and O-EX vs. O-SED subjects; **C)** Circulating levels of p21, p53, and IL-6 proteins in Y-SED, O-SED, and O-EX.

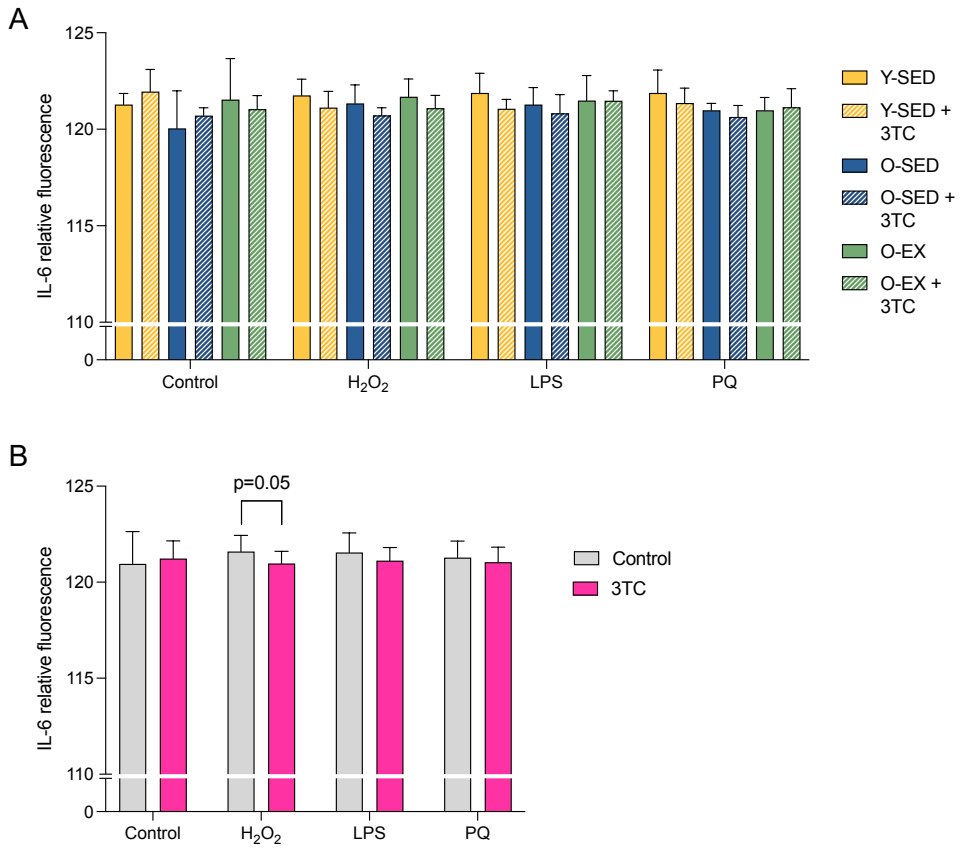


Figure 6.2. IL-6 protein production by peripheral blood mononuclear cells (PBMCs) in response to stressors -/+ 3TC. A) IL-6 in PBMC cell culture supernatant from Y-SED, O-SED, and O-EX after treatment with hydrogen peroxide (H_2O_2) -/+ 3TC, lipopolysaccharide (LPS) -/+ 3TC, and paraquat (PQ) -/+ 3TC; **B)** IL-6 in PBMC cell culture supernatant from combined age groups after treatment with stressors -/+ 3TC.