IDENTIFICATION OF TRANSCRIPTOMIC SIGNATURE IN CELLULAR SENESCENCE AND CHARACTERIZATION OF CIRCULATING SMALL NON-CODING RNA DURING HUMAN AGING

By

PING XIAO

Bachelor of Science in Animal Science
Sichuan Agricultural University
Chengdu, Sichuan
2016

Master of Science in Animal Genetics, Breeding and Reproduction
Sichuan Agricultural University
Chengdu, Sichuan
2019

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2022

IDENTIFICATION OF TRANSCRIPTOMIC SIGNATURE IN CELLULAR SENESCENCE AND CHARACTERIZATION OF CIRCULATING SMALL NON-CODING RNA DURING HUMAN AGING

Dissertation Approved:

Dr. Darren E. Hagen

Dissertation Adviser

Dr. Madhan Subramanian

Dr. Joao G. N. Moraes

Dr. Robert L. Burnap

ACKNOWLEDGEMENTS

First of all, I would like to give my heartfelt thanks to all the people who have ever helped me during my time at Oklahoma State University.

My sincere and hearty thanks and appreciations go firstly to my adviser and committee chair, Dr. Darren E. Hagen, whose suggestions and encouragement have given me much insight into these exploratory studies. It has been a great privilege and joy to study under his guidance and supervision. Furthermore, it is my honor to benefit from his personality and diligence, which I will treasure my whole life. The “hands off” way during my Ph.D. training means a lot for my education, scientific research, independent thinking and future mentorship. My gratitude to him knows no bounds.

I would also like to express my sincere gratitude to my committee members, Drs. Robert L. Burnap, Joao Gabriel N Moraes, and Madhan Subramanian for their support and guidance during my study at Oklahoma State University.

Sincere gratitude should also go to all members of Hagen Lab. I must express my appreciation to Anna Goldkamp, thank you for introducing me the different cultures and letting me get used to living in the US in an incredible speed. And our daily discussions about our research spurred my scientific advance and allowed me have more energy when I cannot think outside the box. Camila Armas, I am also appreciative of you, and thank you for giving me help in our lab and letting me know different cultures. I really enjoy and value the experience I have had in Hagen Lab.

I am also extremely grateful to all my friends, particularly Zhangyue Shi, Hao Chen, Yuxuan Li, Ziyang Zhang, Jing Liu, Zekai Wang and Jinshan Ran, who have kindly provided me assistance and companionship in both courses of study and life.

Lastly, but certainly not least, my cordial thanks also go to my family. My parents give me good environment as much as they can to ensure that I grow up happily. I am always accompanied by their encouragement, love, advice, optimistic attitude, and wisdom for living and learning, and they are well respected for their sacrifice throughout the years. I would also like to thank my grandfather and grandmother for giving all selfless love to their grandson. Thank all my family for rising me up and hope you are proud of me.

Name: PING XIAO

Date of Degree: DECEMBER, 2022

Title of Study: IDENTIFICATION OF TRANSCRIPTOMIC SIGNATURE IN CELLULAR SENESCENCE AND CHARACTERIZATION OF CIRCULATING SMALL NON-CODING RNA DURING HUMAN AGING

Major Field: ANIMAL SCIENCE

Abstract: Accumulation of cellular senescence always forebodes the initialization of aging and cancer. It is an irreversible process that leads to cell cycle arrest while senescent cells still own metabolic viability to affect tissue homeostasis. Senescent cells not only accelerate individual aging process, they are also the driver of age-related diseases such as cancer, osteoarthritis, atherosclerosis, and Alzheimer's diseases. Senescent phenotype shows heterogeneity in different cell lines under diverse triggers, and blurred traits with non-senescent cells make it difficult to identify senescence precisely. It is very necessary to identify robust shared markers, senescence-specific pathways and biological processes across different senescence models. Recently the emerging role of non-coding RNA in senescence and aging has been noticed due to its ability to control cell cycle at post-transcriptional level. Usually the highly proactive secretome from senescent cells, termed the senescence-associated secretory phenotype (SASP), can result in age-related process through intercellular communication, whereas only a number of factors have been identified in very specific scenarios and the role of secreted extracellular RNAs (exRNAs) is not well understood. Detection of exRNAs protected by EV membrane uncovered the fact that most of extracellular mRNAs are fragmentation, along with small non-coding RNAs (sncRNAs), such as miRNAs, piRNA and tRNA fragments. Therefore it is promising to uncover the role of extracellular sncRNA in aging related dysfunction during cell-cell interaction.

To better understand the nature of cellular senescence and its corresponding human aging process at transcriptome level, RNA sequencing data from different cell types and senescence inductions were collected, and significantly shared gene markers and pathways among multiple senescence models were determined through meta-analysis and machine learning-based logistic regression methods. Extensionally, the function of identified senescence associated long non-coding RNAs (lncRNAs) during cell cycle were verified through short interfering RNAs (siRNAs) knock-down treatment in lung fibroblasts (IMR-90). In parallel, the abundance of extracellular sncRNAs from healthy people aged 20 to 99 was quantified using 446 small RNA sequencing datasets. The expressional trends of each sncRNA subspecies were detected with age and a sncRNAs-based age predictors was established using high performance ensemble machine learning strategy.

TABLE OF CONTENTS

Chapter Page

I. REVIEW OF LITERATURE 1

General Introduction 1
Functional Regulation of Non-coding RNA 3
Mechanisms of Cellular Senescence 6
Human Aging 16
Machine Learning for Biomarker Discovery 18
References 21

II. META-ANALYSIS OF UNIVERSAL SIGNATURE IN CELLULAR SENESCENCE USING TRANSCRIPTOMIC STUDIES 46

Abstract 46
Introduction 48
Materials and Methods 50
Results 53
Discussion 58
Conclusion 62
References 63

III. IDENTIFICATION OF NOVEL SENESCENCE-ASSOCIATED LNCRNAS BY TESTING CELL CYCLE REGULATION 100

Abstract 100
Introduction 102
Materials and Methods 105
Results 108
Discussion 110
Conclusion 111
References 112

IV. CHARACTERISTICS OF CIRCULATING SMALL NON-CODING RNAS IN PLASMA AND SERUM DURING HUMAN AGING 125

Abstract 125
Introduction 127
Results 130
Discussion 136
Experimental Procedures 142
References 147

LIST OF TABLES

CHAPTER II: META-ANALYSIS OF UNIVERSAL SIGNATURE IN CELLULAR SENESCENCE USING TRANSCRIPTOMIC STUDIES

Table 1 85
Table 2 86
Table 3 87
Table S1 88
Table S2 92
Table S3 93
Table S4 95
Table S5 96
Table S6 97
Table S7 99

CHAPTER III: IDENTIFICATION OF NOVEL SENESCENCE-ASSOCIATED LNCRNAS BY TESTING CELL CYCLE REGULATION

Table S1 120
Table S2 123

CHAPTER IV: CHARACTERISTICS OF CIRCULATING SMALL NON-CODING RNAS IN PLASMA AND SERUM DURING HUMAN AGING

Table 1 176
Table 2 177
Table S1 178
Table S2 179
Table S3 181
Table S4 184

LIST OF FIGURES

CHAPTER II: META-ANALYSIS OF UNIVERSAL SIGNATURE IN CELLULAR SENESCENCE USING TRANSCRIPTOMIC STUDIES

Figure 1.....71
Figure 2.....72
Figure 3.....73
Figure 4.....75
Figure 5.....76
Figure 6.....77
Figure 7.....78
Figure S1.....79
Figure S2.....80
Figure S3.....81
Figure S4.....82
Figure S5.....83
Figure S6.....84

CHAPTER III: IDENTIFICATION OF NOVEL SENESCENCE-ASSOCIATED LNCRNAS BY TESTING CELL CYCLE REGULATION

Figure 1.....116
Figure 2.....117
Figure 3.....118
Figure S1.....119

CHAPTER IV: CHARACTERISTICS OF CIRCULATING SMALL NON-CODING RNAS IN PLASMA AND SERUM DURING HUMAN AGING

Figure 1.....159
Figure 2.....160
Figure 3.....162
Figure 4.....164
Figure 5.....165
Figure 6.....166
Figure S1.....167
Figure S2.....168
Figure S3.....169

Figure S4.....170
Figure S5.....171
Figure S6.....172
Figure S7.....173
Figure S8.....174
Figure S9.....175

CHAPTER I

REVIEW OF LITERATURE

GENERAL INTRODUCTION

Cell proliferation with equal chromosome segregation into two daughter cells happens at Mitosis or M phase during cell cycle (Barnum & O'Connell, 2014). Mitosis requires accurate arrangement of cellular contents, duplicated DNA, chromosome repair, and cell cycle checkpoints play important roles in surveillance process to ensure the order, integrity, and fidelity of cell division. Checkpoint genes function as cell cycle dependencies for inspecting cell size, DNA damage, DNA replication, DNA repair and mitotic spindle during growth phases (G1 and G2), synthesis (S phase) and M phase (Giono & Manfredi, 2006). DNA damage elicits cell cycle arrest for DNA repair before entering into subsequent phases of the cell cycle. For cells with irreversible defects that cannot be fixed, checkpoint genes transactivate the expression of cyclin-dependent kinase inhibitors and prolong cell cycle arrest, triggering either cell apoptosis or senescence (Carvajal & Manfredi, 2013).

Cellular senescence (CS) is a state of permanent cell cycle arrest, which can be observed after a certain number of cell divisions during cell culture. This type of senescence (called proliferative exhaustion) was first described by Hayflick and Moorhead upon observation in normal diploid fibroblasts (Hayflick & Moorhead, 1961). After that, other studies revealed that other sources of DNA damage, including oxidative stress, DNA damage, ionizing

radiation, or the expression of oncogenes, can also trigger CS (Campisi, 2013; V. Gorgoulis et al., 2019). Non-coding RNAs, including long non-coding RNA (lncRNA) and small non-coding RNA (sncRNA), have been demonstrated that they are associated with many biological processes including cell proliferation, cell fate decision, apoptosis, differentiation, stem cell maintenance and division in any transcriptional, post-transcriptional, translational or post-translational level. Some studies illustrated that non-coding RNAs play roles in CS with specific context (Abdelmohsen & Gorospe, 2015; Puvvula, 2019). With growing number of senescence models in cell and animal model level, however, senescence heterogeneity was recognized when senescent cells showed varying degrees of response to senolytics (Cohn, Gasek, Kuchel, & Xu, 2022). This heterogeneity is particularly exposed when we discuss transcriptomic profiles of diverse senescent cell populations, including different senescence inducers, cell types, and stages of the senescence process (Hernandez-Segura et al., 2017).

CS plays vital roles in maintaining tissue homeostasis and embryonic development (Yun, Davaapil, & Brockes, 2015), wound healing (Demaria et al., 2014) and tumor prevention by limiting proliferation of dysfunctional, damaged or transformed cells (Ovadya & Krizhanovsky, 2018; B. Wang, Kohli, & Demaria, 2020). Meanwhile, deleterious effects of CS have long been known to cause physiologically and pathologically age-related process due to accumulation of senescent cells, depletion of stem/progenitor cell compartments and secretion of senescence associated secretory phenotype (SASP) (Campisi, Andersen, Kapahi, & Melov, 2011; Coppe, Desprez, Krtolica, & Campisi, 2010). Age-related diseases have been documented to be associated to stagnant senescent cell elimination, and accumulated senescent cells can induce age-related inflammation, chronic diseases and tumor progression

(McHugh & Gil, 2018). Therefore, investigation of CS hallmarks for senescent cells targeting is essential for detection and removal of senescent cells in different age related pathologies, allowing prevention or delaying age-related tissue dysfunction to extend life span and improve health span (Baker et al., 2016; Hashimoto et al., 2016; Y. Zhao et al., 2018).

FUNCTIONAL REGULATION OF NON-CODING RNA

The discovery of non-coding RNAs (ncRNAs), transcribed from genomic regions without protein coding potential, has started from transfer RNA and ribosomal RNA in the 1950s (Palazzo & Lee, 2015). Over 70 years, the number of new and putative characterized ncRNAs is greatly expanded, and especially the advent of high-throughput sequencing excites the identification of genome-wide transcription (Menet, Rodriguez, Abruzzi, & Rosbash, 2012). There are different RNA types are classified based on the size and functional attributes, including small non-coding RNAs (sncRNAs) (Bissels et al., 2009; Castle et al., 2010; J. U. Guo, Agarwal, Guo, & Bartel, 2014; Waldron & Lacroute, 1975) and long non-coding RNAs (lncRNAs) (Quek et al., 2015). Multiple studies found that almost all of the mammalian genome is transcribed at some level, with the ENCODE consortium claimed that “80% of the genome has biochemical functions” by DNA elements annotation (Consortium, 2012). Under the criticisms, these “junk RNAs” have been proven that some of them have specific function based on various criteria ranging from their expression levels and splicing to conservation. The properties and corresponding functional investigation of lncRNAs and sncRNAs will be discussed below.

Long Non-coding RNA and Its Biological Function

With extensive identification of genome transcription, the products longer than 200 nucleotides that are not translated into functional proteins are defined as long non-coding RNAs (lncRNAs) (Uszczynska-Ratajczak, Lagarde, Frankish, Guigo, & Johnson, 2018). There are multiple RNA polymerases can transcribe lncRNAs, with RNA polymerase II (Pol II) plays the major role in lncRNA biogenesis. It is presumed that lncRNA shows similar pre-processing steps as mRNA, while recent studies revealed diverse transcription, processing, export and turnover of lncRNAs, which are closely linked with their cellular fates and functions (Statello, Guo, Chen, & Huarte, 2021). Compared to mRNAs, there is a higher proportion of lncRNAs retaining in the nucleus (C. J. Guo et al., 2020), since there are rapid degradation (Schlackow et al., 2017), U1 small nuclear RNA binding (Y. Yin et al., 2020), less efficient splicing (Zuckerman & Ulitsky, 2019). Through interacting with DNA, RNA and proteins, lncRNAs can regulate epigenetic modification, chromosomal structure and function, gene transcription, RNA splicing, stability and translation efficiency (Bonetti et al., 2020; Jiang et al., 2013; West et al., 2016; Q. F. Yin et al., 2012).

As participating in gene regulation, lncRNAs are associated with different aspects of biological processes, from cell proliferation, differentiation, apoptosis, stem cell homeostasis and division, to key roles in the nervous, muscular (Fatica & Bozzoni, 2014), cardiovascular (Fatica & Bozzoni, 2014), adipose (Sun & Lin, 2019), haematopoietic and immune systems (Chen, Satpathy, & Chang, 2017) and their corresponding pathologies. For example, *BACE-AS* is an antisense lncRNA of protein coding gene β-site amyloid precursor protein cleaving enzyme 1 (BACE1; also known as β-secretase 1). By binding specific microRNA, it can protect *BACE1* from degradation and increase severity of neurotoxic amyloid plates

within Alzheimer disease individuals (Faghihi et al., 2010). In a chromatin structure level, lncRNA colon cancer associated transcript 1-long (CCAT1-L) can improve proto-oncogene MYC expression by forming a long-range enhancer-MYC chromatin looping (J. F. Xiang et al., 2014). Also, there is growing number of lncRNAs that modulate cancer initiation and progression in curated databases such as Lnc2Cancer (Gao et al., 2019) or the Cancer LncRNA Census (Carlevaro-Fita et al., 2020).

Apart from RNA itself, some lncRNAs had small open reading frames (sORF, length <300 nt) that could produce steady short peptides with biological functions (Bi et al., 2017; D'Lima et al., 2017; Huang et al., 2017), which challenge the definition of lncRNAs.

Small Non-coding RNA and Its Biological Function

Small non-coding RNAs (sncRNAs) are series of ncRNAs less than 300 nucleotides in length, (Santosh, Varshney, & Yadava, 2015), and based on their biogenesis, cellular location and correspond function, they are categorized as microRNA (miRNA), piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), transfer RNA (tRNA), transfer-derived RNAs (tRFs) and small nuclear RNAs (snRNAs) (Vickers, Roteta, Hucheson-Dilks, Han, & Guo, 2015). From previous studies we are realizing that sncRNAs are involved in the regulation of numerous transcriptomic, epigenetic and proteomic events, thereby in cancer and other chronic human diseases, they are gradually employed as biomarkers for clinical utility, including subtype classification, diagnosis, and prognosis (Jacovetti, Bayazit, & Regazzi, 2021; Z. Zhang, Zhang, Diao, & Han, 2021). Due to the low reads load in high-throughput sequencing and extensive post-transcriptional modification, there are sncRNAs specific methods, including sequencing strategies (Y. Xiang, Ye, Zhang, & Han, 2018; Zheng et al.,

2015) and computational approaches (Isakova, Fehlmann, Keller, & Quake, 2020; Xing et al., 2006), are developed for detecting their abundance.

MECHANISMS OF CELLULAR SENESCENCE

Inducers of Cellular Senescence

Cells can be induced into senescence by multiple intrinsic and extrinsic stimuli, including proliferative exhaustion, radiation, oxidative and genotoxic stress, epigenetic changes, mitochondrial dysfunction and oncogene activation (Di Micco et al., 2006; Mikula-Pietrasik, Niklas, Uruski, Tykarski, & Ksiazek, 2020; Pazolli et al., 2012; Wen & Klionsky, 2016). These factors can be classified into telomere dependent replicative senescence and non-telomeric stress-induced premature senescence, for which there are more type of stress included in senescence-related process (Dierick et al., 2002). DNA damage response (DDR) shows a persistent activity once irreparable DNA damage is triggered by either intrinsic or extrinsic stimuli (Fumagalli et al., 2012). In human somatic cells lacking functional subunit of telomerase due to exhausted cell division, DRR will be triggered when shortened telomeres don have end protection (Shay & Wright, 2019). During oncogene-induced senescence (OIS), hyperactivity of oncogene will activate the mitotic signals and DDR will be initiated to prevent progression of damaged cells when chromosomal dysfunction is sensed (V. G. Gorgoulis & Halazonetis, 2010). Under chemotherapeutic drugs or oxidative stress, PARP-1/ATM/ NF-κB signaling cascade is activated to cause senescence mediated cell cycle arrest, and chemokine CCL2 from secretome will increase invasiveness of escaped melanoma cells (Ohanna et al., 2011).

Prolonged Cell Cycle

Phenotypically, multiple studies have reported that there is a reduced proliferative capacity during the senescence establishment (Y. M. Kim et al., 2013; Nassrally et al., 2019; Ponten, Stein, & Shall, 1983), with prolonged cell cycle duration after an increased population doubling level (Ogrodnik, 2021). The accompanying decline in proliferation rate may be partially explained by the averaged senescence measurements in the mixed population of cells, from which the gradual reduction of proliferative capacity correlates to proportion of non-proliferating, senescent cells in total cells (Ogrodnik, 2021). Also, based on long-term, live-cell imaging system (time-lapse cinematography), Absher et al. reported that WI-38 fibroblasts had a gradually growing doubling time (from 16.8 hours to 32.0 hours) between passage 28 and 53 (Absher, Absher, & Barnes, 1974). Therefore, there is a developing idea of gradual changes of cell cycle time and clonal capacity in human fibroblasts and it helps us identify the continuous process of cellular aging before eventual senescence (non-dividing) phenotype.

Increased Cell Soma

In addition to decline proliferative capacity before senescence, there is an enlargement of cell soma during aging process. Due to the decreased/stagnant cell division rate, aging/senescent cells show expanded cell size with an increase in organelle content and flatly spread cytoplasm (Neurohr et al., 2019; Ogrodnik, Salmonowicz, & Gladyshev, 2019). Consistent with gradually decreased replicative capacity, there is a continuously increased cell volume due to the accumulation of cellular material (Angello, Pendergrass, Norwood, & Prothero, 1987; Y. M. Kim et al., 2013). Based on the negative relation between soma size and

proliferative capacity, Angello et al. utilized real-time cell culture and demonstrated that primary cell lines lost their replicative potential when there was an increase in cell soma size by 50-100% (Angello et al., 1987). This correlation was also validated in in vivo study that overgrown mice by injection of growth hormone showed decreased cell proliferative capacity (Pendergrass, Li, Jiang, & Wolf, 1993). In primary cell lines, Neurohr et al. applied palbociclib, which is a kind of cell cycle inhibitor, to reduce replicative potential and increase cell size simultaneously (Neurohr et al., 2019). During cell culture experiment, it is very necessary to mention that appropriate nutrition and growth factor can extend the times of cell passages and excessive serum concentration may induce premature aging of treated cells (Neurohr et al., 2019).

Senescence-associated Metabolic Activities

Even though active metabolism can be observed in senescent and aging cells (Sabbatinelli et al., 2019), there are few studies about metabolic dynamics during cellular aging on a single-cell level. Yi et al. utilized nuclear magnetic resonance (NMR) analysis to identify the intracellular metabolites during proliferative exhaustion process of human primary cells, and observed metabolites changes were associated with impaired energy metabolism and inhibited protein synthesis (Yi et al., 2020). A further metabolic profiling is necessary for us to better investigate whether the dynamic of different metabolites is related to the requirements for cell cycle process. The increased glycolytic enzymes activity and more absorption of glucose were detected in senescent cells (Unterluggauer et al., 2008). In contrast to positive correlation between glycolysis and lactate production (Lunt & Vander Heiden, 2011), a decreased lactate contents were observed in CS and recent studies demonstrated that aging cells use increased

glutamate for stimulating mammalian target of rapamycin (mTOR), which is a vital pathway for cell maintenance and growth (S. G. Kim et al., 2013; Ogrodnik, 2021). Another metabolic shift during cellular aging is a decreased NAD+ concentration in cell culture medium and this is in consistency with the positive feedback relationship between NAD+ concentration and pyruvate to lactate conversion (Verdin, 2015), and supplementation of NAD+ can increase proliferative activity of primary cell lines (Lim, Potts, & Helm, 2006).

The Senescence Associated Secretory Phenotype

As one of key hallmarks of CS, senescence associated secretory phenotype (SASP) shows secretory activities of multiple senescence contexts, which can affect biological behavior of non-senescent cells (Mohamad Kamal, Safuan, Shamsuddin, & Foroozandeh, 2020). The establishment of SASP is monitored by several key factors, IGFBP3, IGFBP4 and IGFBP7, which can induce senescence through paracrine senescence pathway (Ozcan et al., 2016). IGFBP3 is a downstream target of plasminogen activator inhibitor-1 (PAI-1) system (Elzi et al., 2012), and PAI-1 can induce proliferative senescence via PI (3) K-PKB-GSK3β-cyclin D1 pathway (Kortlever, Higgins, & Bernards, 2006). The heterogeneity of SASP composition was detected across the senescence models, including the combination of senescence induction, strength of DNA damage, intercellular environment and cell type (Coppe et al., 2008; Coppe et al., 2011; Maciel-Baron et al., 2016). Among them, we still observed conserved components of SASP, including cytokines IL-6 and IL-8, serving as pro-inflammatory factors to trigger intercellular inflammation (Soto-Gamez & Demaria, 2017). Different stimuli that mentioned in inducers of CS, has different SASP contents due to the diverse durations of DDR (Coppe, Patil, et al., 2010). Interestingly, SASP can only be

detected within cells having DNA damage process, while direct $p21^{WAF1/CIP1}$ or $p16^{INK4A}$ overexpression fails to form SASP (Coppe et al., 2011). For the non-genome DNA sequence, the dysfunction of mitochondrial activity results in distinct senescence-associated secretome (Wiley et al., 2016). SASP components, such as IL-6 and IL-1A, can reinforce senescence phenotype via autocrine pathway, while other factors, which have non-cell autonomous effects, can induce DDR in nearby healthy, proliferating cells (Nelson et al., 2012).

Since DNA damage is the necessary factor to trigger SASP, some nuclear and cytoplasmic contents such as DNA fragments, degraded peptides, transposable elements, and toll like receptors (TLR) have been known to induce DDR. SASP secretion can be delivered into extracellular environment via different ways. For soluble protein factors, such as interleukins, chemokines, and growth factors, they can be directly secreted into microenvironment, while other molecules need membrane protection from extracellular degradation (Stow & Murray, 2013).

Senescence-associated Extracellular Vesicles

Extracellular vesicles (EVs) are secretions of a variety of cells with lipid membrane, in which components from host cells can be released into extracellular environment (van Niel, D'Angelo, & Raposo, 2018). Senescent cells were reported to have more EVs production than proliferating cells, to release more toxic cytoplasmic DNA fragment and maintain cellular homeostasis (Takahashi et al., 2017). It has been demonstrated that EVs from senescent cells can induce senescence phenotype via paracrine senescence in healthy cells, showing their importance in intercellular mediation (Jeon et al., 2019). Additionally, as secretome from senescent cells, EVs can enhance cancer cell proliferation in a more distal manner (Takasugi

et al., 2017). EVs have different types of extracellular RNAs (exRNAs), including mRNA fragments, miRNA, snRNA, tRFs and Y RNA, and they are found to be expressional correlated with several age-related process (Takasugi, 2018), whereas the direct effects of exRNAs from cell-cell communication should be further elucidated (Gruner & McManus, 2021).

Telomere Shortening

As a process accompanying DNA replication, telomere shortening is the dominant theory of replicative aging among primary cells (d'Adda di Fagagna, 2008). There are shortened telomeres by passaging, and when they are cut down to a certain “threshold” length, the double-strand breaks (DSBs) and cell cycle arrest pathways would be triggered (d'Adda di Fagagna, 2008). Meanwhile, whether telomere-associated DSBs can be repaired depends on their proliferative capacity (Doksani & de Lange, 2016; Hewitt et al., 2012). The dynamics of telomerase activity has been observed during cellular aging in vitro/vivo and corresponding activity is inversely correlated with senescence (Bernadotte, Mikhelson, & Spivak, 2016; Cheng et al., 2019). Previous study showed that increased cell soma, prolonged cell cycle time and also reduced telomere length were simultaneously detected during cell aging (Nassrally et al., 2019). These typical phenotypes were barely observable in telomerase overexpressed cells, indicating the potential functions of telomerase against cellular aging (Nassrally et al., 2019). During the process of telomere shortening, it has been demonstrated that telomere-binding proteins, the “shelterins”, can protect telomere from recognition by DNA damage response related proteins that initialize CS (de Lange, 2018), and oxidative stress induced senescence can reduce shelterin levels in the nucleus (Swanson, Baribault,

Israel, & Bae, 2016). As one of shelterin partners, TRF2 relocates from telomere to non-telomeric chromatin regions during telomere shortening, to regulate DNA modifications and transcription (Mukherjee et al., 2018). These results indicate that progression of cellular aging to CS is a gradual process with gradient phenotypic changes in cell cycle time, cell size, metabolic shift and telomere shortening.

Molecular Dynamic during Cellular Senescence

Molecularly, senescent cells are characterized by high expression of Cdkn2a (p16) and Cdkn1a (p21), which are involved in corresponding pathways of cell cycle inhibition (Munoz-Espin & Serrano, 2014). Other genes involved in p53/CDKN1A (p21) and CDKN2A (p16)/pRB senescence induction pathways are also commonly selected as senescence markers using multiple protein abundance assays (R. Zhang & Adams, 2007). Also, the senescence-associated secretory phenotype (SASP) is induced during cellular aging process and secretion of some senescence-associated cytokine, chemokines, growth factors and proteases triggers pro-inflammation, wound healing and growth responses in the tissue microenvironment (Cuollo, Antonangeli, Santoni, & Soriani, 2020). With growing number of senescence models in cell and animal model level, however, senescence heterogeneity was recognized when senescent cells showed varying degrees of response to senolytics (Cohn et al., 2022). These two cell cycle arrest biomarkers (p16 and p21) are insufficient conditions for senescence detections, since some p16 highly expressing cells don’t have other necessarily senescent phenotype (Hall et al., 2017), and some senescent cells do not show up-regulation of p16 due to different cell types and inductions (Casella et al., 2019; Yosef et al., 2017). Hernandez-Segura et al. used genome-wide transcriptome datasets from both human and mice

fibroblasts across multiple senescence stages, and identified specific senescence signatures under different inducers and cell types (Hernandez-Segura et al., 2017). As one of senescence-associated lysosomal enzymes, senescence-associated β-galactosidase (SA-β-galactosidase) has increased activity in aging cells, while its specificity in senescence detection was suspected as SA-β-gal is also active in proliferative nuerons and developing embryos (de Mera-Rodriguez et al., 2021; Piechota et al., 2016). Senescence heterogeneity was also observed in context of different tissues. By using single-cell RNA sequencing (scRNA-seq), previous study identified the p16 and p21 highly expressing cells in adipose tissues from obese and aged mice, and these cells were categorized into two populations based on their cell types, tissue location and physiological roles (B. Wang et al., 2021; L. Wang et al., 2022). In healthy and young mouse livers, endothelial cells have the highest p16 expression, and p16 is also abundant in epithelial cells of healthy and young mouse kidneys (Omori et al., 2020). There were different cell types (including astrocytes, pericytes, endothelial cells, and glial cells) from mouse retinas suffered from proliferative retinopathies having expression of senescence associated genes (Avelar et al., 2020). The novel senescence marker, Col1a1, was identified in senescent retina endothelial cells (Crespo-Garcia et al., 2021).

Apart from proteins participating in cell cycle arrest pathways and intercellular communication via SASP, non-coding RNAs show emerging roles in regulating CS transcriptionally, post-transcriptionally and translationally (Abdelmohsen & Gorospe, 2015; Puvvula, 2019). miRNAs are one type of non-coding RNAs that play vital roles in mediating functional activity of gene targets (Bartel, 2009). The mature miRNAs cleaved by Dicer

mainly bind to the 3'untranslated region (3'UTR) of target mRNA to induce mRNA degradation and lower translation efficiency (O'Brien, Hayder, Zayed, & Peng, 2018). Recent study also reported other anchor points of miRNAs within 5'UTR, coding sequence, and gene promoters (Broughton, Lovci, Huang, Yeo, & Pasquinelli, 2016). Several studies revealed that some miRNAs, including miR-34a, miR-22, miR-217, miR-138, miR-181a and miR-181b, induced senescence by inhibiting SIRT1 (silent mating type information regulator 2 homolog 1) expression, which can inhibit endothelial progenitor cell senescence (Jazbutyte et al., 2013; Menghini et al., 2009; Rivetti di Val Cervo et al., 2012; T. Zhao, Li, & Chen, 2010). Increased expression of miRNAs involved in let-7 family can induce senescence by reducing accumulation of proteins that essential for cell proliferation such as EZH2 and HMGA2 (Markowski et al., 2011; Tzatsos et al., 2011). Some miRNAs, including miR-25 and miR-30d, can directly target the 3'UTR of p53 mRNA, thus inhibit corresponding effects on cell cycle arrest and senescence (M. Kumar et al., 2011). On the other hand, miRNAs can target cyclin dependent kinase 2 (CDK2) and minichromosome maintenance complex component 5 (MCM5) to accumulate p53 proteins and activate downstream senescence associated genes (Afanasyeva et al., 2011). As another important axis of senescence induction, pRB/p16 pathway includes some validated miRNAs to directly or indirectly regulate the abundance of p16 (Overhoff et al., 2014; Philipot et al., 2014). There are 4 miRNAs, miR-15b, miR-24, miR-25, and miR-141, concomitantly inhibit the expression of MAPK (mitogen-activated protein kinase) kinase 4 (MKK4), and they were down-regulated in senescent human diploid fibroblasts (HDFs) (Marasa et al., 2009). When the joint reduction of these four miRNAs was performed, there was activated p38 regulated by increased MKK4,

resulting in elevated p16 and promoted senescence phenotype (Marasa et al., 2009). Meanwhile, long non-coding RNAs (lncRNAs), the single-stranded RNAs lacking the potential to encode proteins, have been reported to influence cell proliferation, differentiation, apoptosis, stem cell homeostasis and division in a spatiotemporally specific way (Puvvula, 2019). Several studies revealed that senescence induction through pRb/p16 pathway was tightly regulated by lncRNAs (Aguilo, Zhou, & Walsh, 2011; Montes et al., 2015; Sang et al., 2016). For example, lncRNA Urothelial Cancer-Associated 1 (UCA1) has been determined that it can trigger pro-senescence phenotype by avoiding the binding between heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and p16 mRNA, to promote p16 mRNA retention and translation in both autocrine and paracrine ways (P. P. Kumar et al., 2014). Also, in p53/p21 pathway, long intergenic non-coding RNA-p21 (LincRNA-p21) was activated by p53 to recruit mRNA-binding proteins hnRNP-K on the promoter of p21 and initiated CS (Dimitrova et al., 2014). Apart from that, other lncRNAs such as P21 Associated NcRNA DNA Damage Activated (PANDA), Maternally Expressed Gene 3 (MEG3), P53 Induced Noncoding Transcript (Pint), P53 Regulation Associated LncRNA (PRAL), LINC00673, Focal Amplified LncRNA On Chromosome 1 (FAL1), BRAF-Activated Noncoding RNA (BANCR) and Ovarian Adenocarcinoma Amplified lncRNA (OVAAL) showed functional association with p53/p21 pathway induced senescence by directly or indirectly regulating corresponding transcription or translation (Cho, Kim, Back, & Jang, 2005; Hu et al., 2014; Marin-Bejar et al., 2013; Puvvula et al., 2014; Roth et al., 2018; Shi et al., 2015; Su, Wang, Qi, Wang, & Zhang, 2017; Zhou, Zhang, & Klibanski, 2012).

Immunosurveillance during Cellular Senescence

The accumulation of senescent cells in our body always correlates to individual aging. To efficiently eliminate exceeding amount of senescent cells, immune surveillance through multiple immune cells (including macrophages, NK cells, T cells, CAR-T cells and dendritic cells) can specifically recognize senescent cells in different pathological conditions (Song, An, & Zou, 2020). Due to distinctive characteristic of ligands, different types of senescent cells can recruit corresponding immune cells for immune surveillance. For example, senescent hepatic stellate cells with specific cell-surface ligands MICA and ULBP2 are targeted and removed by natural killer cells (Krizhanovsky et al., 2008). Macrophages are important players in removing senescent uterine cells for maintaining postpartum uterine function in wide-type mice (Egashira et al., 2017). In murine hepatocytes, CD4+ T cells can directly remove tumors or senescent cells by recognizing MHC II surface protein, to prevent mouse liver cancer (Kang et al., 2011).

HUMAN AGING

Aging is considerably the most complex phenotype that occurs in humans. People always suffer from different aging processes in different organs with diverse strengths. Herein, we summarize the usual age-related features and corresponding diseases, and also discuss the association between senescent cell accumulation and aging process.

Age-related Phenome and Diseases

During aging process, the most common phenotype we can observe is wrinkles on the skin and gray hair. Meanwhile, our organ systems, such as skeletal system, muscular system, circulatory system, respiratory system, urinary system, digestive system, immune system,

nervous system, and endocrine system, will encounter age-related dysfunction in diverse extents. For example, normal aging process including hearing loss (Davis et al., 2016), visual acuity (Evans et al., 2002), muscle atrophy (Dodds et al., 2017), and immunosenescence (Bandaranayake & Shaw, 2016) are prevalent among elderly. Also, other chronic diseases such as cardiovascular disease, hypertension, Parkinson's disease, Alzheimer's disease and cancer show high incidence among old individuals, as chronic inflammation persistently happen in multiple senescent tissues (Sanada et al., 2018). Andreassen et al. utilized text mining and enrichment analysis to generate a comprehensive description of the human aging phenotype and with further identifying the association between features, a tissue specific aging phenome clustering was calculated (Andreassen, Ben Ezra, & Scheibye-Knudsen, 2019).

Association between Cellular Senescence and Human Aging

In old individuals, there is an increasing rate of senescent cell accumulation, and whether CS results in or is a consequence of age-related impairment of immune system is still debatable. From recent studies we know that SASP can recruit immune cells and induce immune surveillance dysfunction (Prata, Ovsyannikova, Tchkonia, & Kirkland, 2018), and extracellular matrix (ECM) remodeling of senescent cells may also hinder the access of immune cells (Fane & Weeraratna, 2020). Nevertheless, the precise mechanism of clearance of senescent cells by the immune system and corresponding impairment is still need to be further investigated. Human aging is an unavoidable process and indicates increasing susceptibility to disease and death (Lopez-Otin, Blasco, Partridge, Serrano, & Kroemer, 2013). Particularly, the long-term retention of senescent cells can lead to senescence stresses to other

proliferative cells and their interaction via chemical signaling will disrupt tissue homeostasis and lead to aging and age-related disease (Baker et al., 2011). Due to the heterogeneous human phenotype of aging that are initiated in different tissues, there are debility and dysfunction of diverse biological processes that related to accumulated CS. Combined with human aging phenome, it is therefore very necessary to determine the precursory biomarkers of CS and individual aging and to efficiently perform interventions for healthy aging.

MACHINE LEARNING FOR BIOMARKER DISCOVERY

With the rapid technology improvement and exponential growth of research and clinical data storage, there are plenty of valuable resources for us to utilize and develop approaches for senescence and aging related assessment and biomarker discovery. When machine driven methods successfully construct visible and computable data with suitable phenotypic range, appropriate analysis strategies can extract important features that contribute most to phenotype we focus on (Osborne et al., 2020). Machine learning (ML) techniques have been spread rapidly and used in inferring knowledge about molecular biology, physiology, electronic health records due to its particular ability to handle large datasets, and to make predictions (Chicco, 2017). Here the application of ML to CS and human aging for biomarker identification is discussed below.

Image Based Biomarkers

Based on cellular morphological dynamics during CS, Oja et al. utilized automated imaging system to longitudinally track mesenchymal stromal cells (MSCs) via population doubling numbers, and cell enlargement was important parameter to predict senescence level through

supervised machine learning applications (Oja, Komulainen, Penttila, Nystedt, & Korhonen, 2018). Convolutional neural networks (CNN) based image classification for distinguishing cell morphology was employed after ectopically expressing senescence inducer genes and anti-senescence drugs was identified through predicted senescence probability (Kusumoto et al., 2021). A Senescence-associated morphological profiles (SAMPs) were established for senescence marker detection across the context of senescence and heterogeneity between models of senescence was explored both through cell type and induction specific morphological feature (Wallis et al., 2022). In clinical diagnostic, magnetic resonance images from healthy individuals and patients with cognitive dysfunction were used to develop a clock predictive of cognitive ageing with multiple regression algorithms (Vemuri et al., 2018).

High-throughput Data Based Biomarkers

The advent of high-dimensional data improves our understanding of genetic regulation in global level. Meanwhile, the complexity of generated data hinders the progress of importance determination. ML can narrow down the amount of input variables and retain core factors with more predictive power for biomarker identification (Putin et al., 2016). Collection of endothelial cell (EC) RNA-seq data was employed to investigate consensus features of EC senescence in gene and pathway level via ML-based meta-analysis, which facilitates the development of therapeutic targets of numerous systemic vascular dysfunction (Park & Kim, 2021). In regression algorithms, a series of “aging-clocks” have been developed using omics like data (methylation (Horvath, 2015), transcriptomics (Shokhirev & Johnson, 2021), proteomics (Johnson, Shokhirev, Wyss-Coray, & Lehallier, 2020), metabolomics (Hertel et al., 2016) and microbiomics (Galkin et al., 2020)), to capture dynamics that occur over age or

detect age-related pathologies.

Text-based Biomarkers

When there are plenty of studies reported with certain context of senescence models, it is difficult to summarize the complexity of the aging process and text mining strategies have been employed in biomedical research. Since the growth of clinical and research data publication, we should collectively and accurately identify the most related phenotype and treatments to the certain cases (Jensen, Jensen, & Brunak, 2012). It is an efficient way to determine the power of relationship between biological process and corresponding signature that participate in. For example, Fernandes et al. established datamining endeavors and revealed relationship between certain genes and age-related processes (Fernandes et al., 2016). Additionally, a weighted interaction network between features can be developed through text-mining unstructured key-word data in large cohort studies (Westergaard, Moseley, Sorup, Baldi, & Brunak, 2019).

REFERENCES

Abdelmohsen, K., & Gorospe, M. (2015). Noncoding RNA control of cellular senescence. *Wiley Interdiscip Rev RNA, 6*(6), 615-629. doi:10.1002/wrna.1297

Absher, P. M., Absher, R. G., & Barnes, W. D. (1974). Genealogies of clones of diploid fibroblasts. Cinemicrophotographic observations of cell division patterns in relation to population age. *Exp Cell Res, 88*(1), 95-104. doi:10.1016/0014-4827(74)90622-3

Afanasyeva, E. A., Mestdagh, P., Kumps, C., Vandesompele, J., Ehemann, V., Theissen, J., . . . Westermann, F. (2011). MicroRNA miR-885-5p targets CDK2 and MCM5, activates p53 and inhibits proliferation and survival. *Cell Death Differ, 18*(6), 974-984. doi:10.1038/cdd.2010.164

Aguilo, F., Zhou, M. M., & Walsh, M. J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res, 71*(16), 5365-5369. doi:10.1158/0008-5472.CAN-10-4379

Andreassen, S. N., Ben Ezra, M., & Scheibye-Knudsen, M. (2019). A defined human aging phenome. *Aging (Albany NY), 11*(15), 5786-5806. doi:10.18632/aging.102166

Angello, J. C., Pendergrass, W. R., Norwood, T. H., & Prothero, J. (1987). Proliferative potential of human fibroblasts: an inverse dependence on cell size. *J Cell Physiol, 132*(1), 125-130. doi:10.1002/jcp.1041320117

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., . . . de Magalhaes, J. P. (2020). A multidimensional systems biology analysis of cellular senescence in

aging and disease. *Genome Biol, 21*(1), 91. doi:10.1186/s13059-020-01990-9

Baker, D. J., Childs, B. G., Durik, M., Wijers, M. E., Sieben, C. J., Zhong, J., . . . van Deursen, J. M. (2016). Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. *Nature, 530*(7589), 184-189. doi:10.1038/nature16932

Baker, D. J., Wijshake, T., Tchkonia, T., LeBrasseur, N. K., Childs, B. G., van de Sluis, B., . . . van Deursen, J. M. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature, 479*(7372), 232-236. doi:10.1038/nature10600

Bandaranayake, T., & Shaw, A. C. (2016). Host Resistance and Immune Aging. *Clin Geriatr Med, 32*(3), 415-432. doi:10.1016/j.cger.2016.02.007

Barnum, K. J., & O'Connell, M. J. (2014). Cell cycle regulation by checkpoints. *Methods Mol Biol, 1170*, 29-40. doi:10.1007/978-1-4939-0888-2_2

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell, 136*(2), 215-233. doi:10.1016/j.cell.2009.01.002

Bernadotte, A., Mikhelson, V. M., & Spivak, I. M. (2016). Markers of cellular senescence. Telomere shortening as a marker of cellular senescence. *Aging (Albany NY), 8*(1), 3-11. doi:10.18632/aging.100871

Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J. R., Shelton, J. M., . . . Olson, E. N. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. *Science, 356*(6335), 323-327. doi:10.1126/science.aam9361

Bissels, U., Wild, S., Tomiuk, S., Holste, A., Hafner, M., Tuschl, T., & Bosio, A. (2009). Absolute quantification of microRNAs by using a universal reference. *RNA, 15*(12), 2375-2384. doi:10.1261/rna.1754109

Bonetti, A., Agostini, F., Suzuki, A. M., Hashimoto, K., Pascarella, G., Gimenez, J., . . . Carninci, P. (2020). RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun, 11*(1), 1018. doi:10.1038/s41467-020-14337-6

Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W., & Pasquinelli, A. E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol Cell, 64*(2), 320-333. doi:10.1016/j.molcel.2016.09.004

Campisi, J. (2013). Aging, cellular senescence, and cancer. *Annu Rev Physiol, 75*, 685-705. doi:10.1146/annurev-physiol-030212-183653

Campisi, J., Andersen, J. K., Kapahi, P., & Melov, S. (2011). Cellular senescence: a link between cancer and age-related degenerative disease? *Semin Cancer Biol, 21*(6), 354-359. doi:10.1016/j.semcancer.2011.09.001

Carlevaro-Fita, J., Lanzos, A., Feuerbach, L., Hong, C., Mas-Ponte, D., Pedersen, J. S., . . . Consortium, P. (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol, 3*(1), 56. doi:10.1038/s42003-019-0741-7

Carvajal, L. A., & Manfredi, J. J. (2013). Another fork in the road--life or death decisions by the tumour suppressor p53. *EMBO Rep, 14*(5), 414-421. doi:10.1038/embor.2013.25

Casella, G., Munk, R., Kim, K. M., Piao, Y., De, S., Abdelmohsen, K., & Gorospe, M. (2019). Transcriptome signature of cellular senescence. *Nucleic Acids Res, 47*(21), 11476. doi:10.1093/nar/gkz879

Castle, J. C., Armour, C. D., Lower, M., Haynor, D., Biery, M., Bouzek, H., . . . Raymond, C.

K. (2010). Digital genome-wide ncRNA expression, including SnoRNAs, across 11 human tissues using polyA-neutral amplification. *PLoS One, 5*(7), e11779. doi:10.1371/journal.pone.0011779

Chen, Y. G., Satpathy, A. T., & Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol, 18*(9), 962-972. doi:10.1038/ni.3771

Cheng, L., Yuan, B., Ying, S., Niu, C., Mai, H., Guan, X., . . . Ye, Q. (2019). PES1 is a critical component of telomerase assembly and regulates cellular senescence. *Sci Adv, 5*(5), eaav1090. doi:10.1126/sciadv.aav1090

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Min, 10*, 35. doi:10.1186/s13040-017-0155-3

Cho, S., Kim, J. H., Back, S. H., & Jang, S. K. (2005). Polypyrimidine tract-binding protein enhances the internal ribosomal entry site-dependent translation of p27Kip1 mRNA and modulates transition from G1 to S phase. *Mol Cell Biol, 25*(4), 1283-1297. doi:10.1128/MCB.25.4.1283-1297.2005

Cohn, R. L., Gasek, N. S., Kuchel, G. A., & Xu, M. (2022). The heterogeneity of cellular senescence: insights at the single-cell level. *Trends Cell Biol*. doi:10.1016/j.tcb.2022.04.011

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi:10.1038/nature11247

Coppe, J. P., Desprez, P. Y., Krtolica, A., & Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol, 5*, 99-118. doi:10.1146/annurev-pathol-121808-102144

Coppe, J. P., Patil, C. K., Rodier, F., Krtolica, A., Beausejour, C. M., Parrinello, S., . . . Campisi, J. (2010). A human-like senescence-associated secretory phenotype is conserved in mouse cells dependent on physiological oxygen. *PLoS One, 5*(2), e9188. doi:10.1371/journal.pone.0009188

Coppe, J. P., Patil, C. K., Rodier, F., Sun, Y., Munoz, D. P., Goldstein, J., . . . Campisi, J. (2008). Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLoS Biol, 6*(12), 2853-2868. doi:10.1371/journal.pbio.0060301

Coppe, J. P., Rodier, F., Patil, C. K., Freund, A., Desprez, P. Y., & Campisi, J. (2011). Tumor suppressor and aging biomarker p16(INK4a) induces cellular senescence without the associated inflammatory secretory phenotype. *J Biol Chem, 286*(42), 36396-36403. doi:10.1074/jbc.M111.257071

Crespo-Garcia, S., Tsuruda, P. R., Dejda, A., Ryan, R. D., Fournier, F., Chaney, S. Y., . . . Sapieha, P. (2021). Pathological angiogenesis in retinopathy engages cellular senescence and is amenable to therapeutic elimination via BCL-xL inhibition. *Cell Metab, 33*(4), 818-832 e817. doi:10.1016/j.cmet.2021.01.011

Cuollo, L., Antonangeli, F., Santoni, A., & Soriani, A. (2020). The Senescence-Associated Secretory Phenotype (SASP) in the Challenging Future of Cancer Therapy and Age-Related Diseases. *Biology (Basel), 9*(12). doi:10.3390/biology9120485

d'Adda di Fagagna, F. (2008). Living on a break: cellular senescence as a DNA-damage response. *Nat Rev Cancer, 8*(7), 512-522. doi:10.1038/nrc2440

D'Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., . . . Slavoff, S. A.

(2017). A human microprotein that interacts with the mRNA decapping complex. *Nat Chem Biol, 13*(2), 174-180. doi:10.1038/nchembio.2249

Davis, A., McMahon, C. M., Pichora-Fuller, K. M., Russ, S., Lin, F., Olusanya, B. O., . . . Tremblay, K. L. (2016). Aging and Hearing Health: The Life-course Approach. *Gerontologist, 56 Suppl 2*, S256-267. doi:10.1093/geront/gnw033

de Lange, T. (2018). Shelterin-Mediated Telomere Protection. *Annu Rev Genet, 52*, 223-247. doi:10.1146/annurev-genet-032918-021921

de Mera-Rodriguez, J. A., Alvarez-Hernan, G., Ganan, Y., Martin-Partido, G., Rodriguez-Leon, J., & Francisco-Morcillo, J. (2021). Is Senescence-Associated beta-Galactosidase a Reliable in vivo Marker of Cellular Senescence During Embryonic Development? *Front Cell Dev Biol, 9*, 623175. doi:10.3389/fcell.2021.623175

Demaria, M., Ohtani, N., Youssef, S. A., Rodier, F., Toussaint, W., Mitchell, J. R., . . . Campisi, J. (2014). An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev Cell, 31*(6), 722-733. doi:10.1016/j.devcel.2014.11.012

Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., . . . d'Adda di Fagagna, F. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature, 444*(7119), 638-642. doi:10.1038/nature05327

Dierick, J. F., Eliaers, F., Remacle, J., Raes, M., Fey, S. J., Larsen, P. M., & Toussaint, O. (2002). Stress-induced premature senescence and replicative senescence are different phenotypes, proteomic evidence. *Biochem Pharmacol, 64*(5-6), 1011-1017.

doi:10.1016/s0006-2952(02)01171-1

Dimitrova, N., Zamudio, J. R., Jong, R. M., Soukup, D., Resnick, R., Sarma, K., . . . Jacks, T. (2014). LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell, 54*(5), 777-790. doi:10.1016/j.molcel.2014.04.025

Dodds, R. M., Granic, A., Davies, K., Kirkwood, T. B., Jagger, C., & Sayer, A. A. (2017). Prevalence and incidence of sarcopenia in the very old: findings from the Newcastle 85+ Study. *J Cachexia Sarcopenia Muscle, 8*(2), 229-237. doi:10.1002/jcsm.12157

Doksani, Y., & de Lange, T. (2016). Telomere-Internal Double-Strand Breaks Are Repaired by Homologous Recombination and PARP1/Lig3-Dependent End-Joining. *Cell Rep, 17*(6), 1646-1656. doi:10.1016/j.celrep.2016.10.008

Egashira, M., Hirota, Y., Shimizu-Hirota, R., Saito-Fujita, T., Haraguchi, H., Matsumoto, L., . . . Osuga, Y. (2017). F4/80+ Macrophages Contribute to Clearance of Senescent Cells in the Mouse Postpartum Uterus. *Endocrinology, 158*(7), 2344-2353. doi:10.1210/en.2016-1886

Elzi, D. J., Lai, Y., Song, M., Hakala, K., Weintraub, S. T., & Shiio, Y. (2012). Plasminogen activator inhibitor 1--insulin-like growth factor binding protein 3 cascade regulates stress-induced senescence. *Proc Natl Acad Sci U S A, 109*(30), 12052-12057. doi:10.1073/pnas.1120437109

Evans, J. R., Fletcher, A. E., Wormald, R. P., Ng, E. S., Stirling, S., Smeeth, L., . . . Tulloch, A. (2002). Prevalence of visual impairment in people aged 75 years and older in Britain: results from the MRC trial of assessment and management of older people in the

community. *Br J Ophthalmol, 86*(7), 795-800. doi:10.1136/bjo.86.7.795

Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., . . . Wahlestedt, C. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol, 11*(5), R56. doi:10.1186/gb-2010-11-5-r56

Fane, M., & Weeraratna, A. T. (2020). How the ageing microenvironment influences tumour progression. *Nat Rev Cancer, 20*(2), 89-106. doi:10.1038/s41568-019-0222-9

Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet, 15*(1), 7-21. doi:10.1038/nrg3606

Fernandes, M., Wan, C., Tacutu, R., Barardo, D., Rajput, A., Wang, J., . . . de Magalhaes, J. P. (2016). Systematic analysis of the gerontome reveals links between aging and age-related diseases. *Hum Mol Genet, 25*(21), 4804-4818. doi:10.1093/hmg/ddw307

Fumagalli, M., Rossiello, F., Clerici, M., Barozzi, S., Cittaro, D., Kaplunov, J. M., . . . d'Adda di Fagagna, F. (2012). Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nat Cell Biol, 14*(4), 355-365. doi:10.1038/ncb2466

Galkin, F., Mamoshina, P., Aliper, A., Putin, E., Moskalev, V., Gladyshev, V. N., & Zhavoronkov, A. (2020). Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning. *iScience, 23*(6), 101199. doi:10.1016/j.isci.2020.101199

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., . . . Li, X. (2019). Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res, 47*(D1), D1028-D1033. doi:10.1093/nar/gky1096

Giono, L. E., & Manfredi, J. J. (2006). The p53 tumor suppressor participates in multiple cell

cycle checkpoints. *J Cell Physiol, 209*(1), 13-20. doi:10.1002/jcp.20689

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., . . . Demaria, M. (2019). Cellular Senescence: Defining a Path Forward. *Cell, 179*(4), 813-827. doi:10.1016/j.cell.2019.10.005

Gorgoulis, V. G., & Halazonetis, T. D. (2010). Oncogene-induced senescence: the bright and dark side of the response. *Curr Opin Cell Biol, 22*(6), 816-827. doi:10.1016/j.ceb.2010.07.013

Gruner, H. N., & McManus, M. T. (2021). Examining the evidence for extracellular RNA function in mammals. *Nat Rev Genet, 22*(7), 448-458. doi:10.1038/s41576-021-00346-8

Guo, C. J., Ma, X. K., Xing, Y. H., Zheng, C. C., Xu, Y. F., Shan, L., . . . Chen, L. L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell, 181*(3), 621-636 e622. doi:10.1016/j.cell.2020.03.006

Guo, J. U., Agarwal, V., Guo, H., & Bartel, D. P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol, 15*(7), 409. doi:10.1186/s13059-014-0409-z

Hall, B. M., Balan, V., Gleiberman, A. S., Strom, E., Krasnov, P., Virtuoso, L. P., . . . Gudkov, A. V. (2017). p16(Ink4a) and senescence-associated beta-galactosidase can be induced in macrophages as part of a reversible response to physiological stimuli. *Aging (Albany NY), 9*(8), 1867-1884. doi:10.18632/aging.101268

Hashimoto, M., Asai, A., Kawagishi, H., Mikawa, R., Iwashita, Y., Kanayama, K., . . . Sugimoto, M. (2016). Elimination of p19(ARF)-expressing cells enhances pulmonary

function in mice. *JCI Insight, 1*(12), e87732. doi:10.1172/jci.insight.87732

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res, 25*, 585-621. doi:10.1016/0014-4827(61)90192-6

Hernandez-Segura, A., de Jong, T. V., Melov, S., Guryev, V., Campisi, J., & Demaria, M. (2017). Unmasking Transcriptional Heterogeneity in Senescent Cells. *Curr Biol, 27*(17), 2652-2660 e2654. doi:10.1016/j.cub.2017.07.033

Hertel, J., Friedrich, N., Wittfeld, K., Pietzner, M., Budde, K., Van der Auwera, S., . . . Grabe, H. J. (2016). Measuring Biological Age via Metabonomics: The Metabolic Age Score. *J Proteome Res, 15*(2), 400-410. doi:10.1021/acs.jproteome.5b00561

Hewitt, G., Jurk, D., Marques, F. D., Correia-Melo, C., Hardy, T., Gackowska, A., . . . Passos, J. F. (2012). Telomeres are favoured targets of a persistent DNA damage response in ageing and stress-induced senescence. *Nat Commun, 3*, 708. doi:10.1038/ncomms1708

Horvath, S. (2015). Erratum to: DNA methylation age of human tissues and cell types. *Genome Biol, 16*, 96. doi:10.1186/s13059-015-0649-6

Hu, X., Feng, Y., Zhang, D., Zhao, S. D., Hu, Z., Greshock, J., . . . Zhang, L. (2014). A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell, 26*(3), 344-357. doi:10.1016/j.ccr.2014.07.009

Huang, J. Z., Chen, M., Chen, Gao, X. C., Zhu, S., Huang, H., . . . Yan, G. R. (2017). A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol Cell, 68*(1), 171-184 e176. doi:10.1016/j.molcel.2017.09.015

Isakova, A., Fehlmann, T., Keller, A., & Quake, S. R. (2020). A mouse tissue atlas of small noncoding RNA. *Proc Natl Acad Sci U S A, 117*(41), 25634-25645. doi:10.1073/pnas.2002277117

Jacovetti, C., Bayazit, M. B., & Regazzi, R. (2021). Emerging Classes of Small Non-Coding RNAs With Potential Implications in Diabetes and Associated Metabolic Disorders. *Front Endocrinol (Lausanne), 12*, 670719. doi:10.3389/fendo.2021.670719

Jazbutyte, V., Fiedler, J., Kneitz, S., Galuppo, P., Just, A., Holzmann, A., . . . Thum, T. (2013). MicroRNA-22 increases senescence and activates cardiac fibroblasts in the aging heart. *Age (Dordr), 35*(3), 747-762. doi:10.1007/s11357-012-9407-9

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet, 13*(6), 395-405. doi:10.1038/nrg3208

Jeon, O. H., Wilson, D. R., Clement, C. C., Rathod, S., Cherry, C., Powell, B., . . . Elisseeff, J. H. (2019). Senescence cell-associated extracellular vesicles serve as osteoarthritis disease and therapeutic markers. *JCI Insight, 4*(7). doi:10.1172/jci.insight.125019

Jiang, J., Jing, Y., Cost, G. J., Chiang, J. C., Kolpa, H. J., Cotton, A. M., . . . Lawrence, J. B. (2013). Translating dosage compensation to trisomy 21. *Nature, 500*(7462), 296-300. doi:10.1038/nature12394

Johnson, A. A., Shokhirev, M. N., Wyss-Coray, T., & Lehallier, B. (2020). Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev, 60*, 101070. doi:10.1016/j.arr.2020.101070

Kang, T. W., Yevsa, T., Woller, N., Hoenicke, L., Wuestefeld, T., Dauch, D., . . . Zender, L. (2011). Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature, 479*(7374), 547-551. doi:10.1038/nature10599

Kim, S. G., Hoffman, G. R., Poulogiannis, G., Buel, G. R., Jang, Y. J., Lee, K. W., . . . Blenis, J. (2013). Metabolic stress controls mTORC1 lysosomal localization and dimerization by regulating the TTT-RUVBL1/2 complex. *Mol Cell, 49*(1), 172-185. doi:10.1016/j.molcel.2012.10.003

Kim, Y. M., Byun, H. O., Jee, B. A., Cho, H., Seo, Y. H., Kim, Y. S., . . . Yoon, G. (2013). Implications of time-series gene expression profiles of replicative senescence. *Aging Cell, 12*(4), 622-634. doi:10.1111/acel.12087

Kortlever, R. M., Higgins, P. J., & Bernards, R. (2006). Plasminogen activator inhibitor-1 is a critical downstream target of p53 in the induction of replicative senescence. *Nat Cell Biol, 8*(8), 877-884. doi:10.1038/ncb1448

Krizhanovsky, V., Yon, M., Dickins, R. A., Hearn, S., Simon, J., Miething, C., . . . Lowe, S. W. (2008). Senescence of activated stellate cells limits liver fibrosis. *Cell, 134*(4), 657-667. doi:10.1016/j.cell.2008.06.049

Kumar, M., Lu, Z., Takwi, A. A., Chen, W., Callander, N. S., Ramos, K. S., . . . Li, Y. (2011). Negative regulation of the tumor suppressor p53 gene by microRNAs. *Oncogene, 30*(7), 843-853. doi:10.1038/onc.2010.457

Kumar, P. P., Emechebe, U., Smith, R., Franklin, S., Moore, B., Yandell, M., . . . Moon, A. M. (2014). Coordinated control of senescence by lncRNA and a novel T-box3 co-repressor complex. *Elife, 3*. doi:10.7554/eLife.02805

Kusumoto, D., Seki, T., Sawada, H., Kunitomi, A., Katsuki, T., Kimura, M., . . . Yuasa, S. (2021). Anti-senescent drug screening by deep learning-based morphology senescence scoring. *Nat Commun, 12*(1), 257. doi:10.1038/s41467-020-20213-0

Lim, C. S., Potts, M., & Helm, R. F. (2006). Nicotinamide extends the replicative life span of primary human cells. *Mech Ageing Dev, 127*(6), 511-514. doi:10.1016/j.mad.2006.02.001

Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell, 153*(6), 1194-1217. doi:10.1016/j.cell.2013.05.039

Lunt, S. Y., & Vander Heiden, M. G. (2011). Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu Rev Cell Dev Biol, 27*, 441-464. doi:10.1146/annurev-cellbio-092910-154237

Maciel-Baron, L. A., Morales-Rosales, S. L., Aquino-Cruz, A. A., Triana-Martinez, F., Galvan-Arzate, S., Luna-Lopez, A., . . . Konigsberg, M. (2016). Senescence associated secretory phenotype profile from primary lung mice fibroblasts depends on the senescence induction stimuli. *Age (Dordr), 38*(1), 26. doi:10.1007/s11357-016-9886-1

Marasa, B. S., Srikantan, S., Masuda, K., Abdelmohsen, K., Kuwano, Y., Yang, X., . . . Gorospe, M. (2009). Increased MKK4 abundance with replicative senescence is linked to the joint reduction of multiple microRNAs. *Sci Signal, 2*(94), ra69. doi:10.1126/scisignal.2000442

Marin-Bejar, O., Marchese, F. P., Athie, A., Sanchez, Y., Gonzalez, J., Segura, V., . . . Huarte, M. (2013). Pint lincRNA connects the p53 pathway with epigenetic silencing by the

Polycomb repressive complex 2. *Genome Biol, 14*(9), R104. doi:10.1186/gb-2013-14-9-r104

Markowski, D. N., Helmke, B. M., Belge, G., Nimzyk, R., Bartnitzke, S., Deichert, U., & Bullerdiek, J. (2011). HMGA2 and p14Arf: major roles in cellular senescence of fibroids and therapeutic implications. *Anticancer Res, 31*(3), 753-761.

McHugh, D., & Gil, J. (2018). Senescence and aging: Causes, consequences, and therapeutic avenues. *J Cell Biol, 217*(1), 65-77. doi:10.1083/jcb.201708092

Menet, J. S., Rodriguez, J., Abruzzi, K. C., & Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *Elife, 1*, e00011. doi:10.7554/eLife.00011

Menghini, R., Casagrande, V., Cardellini, M., Martelli, E., Terrinoni, A., Amati, F., . . . Federici, M. (2009). MicroRNA 217 modulates endothelial cell senescence via silent information regulator 1. *Circulation, 120*(15), 1524-1532. doi:10.1161/CIRCULATIONAHA.109.864629

Mikula-Pietrasik, J., Niklas, A., Uruski, P., Tykarski, A., & Ksiazek, K. (2020). Mechanisms and significance of therapy-induced and spontaneous senescence of cancer cells. *Cell Mol Life Sci, 77*(2), 213-229. doi:10.1007/s00018-019-03261-8

Mohamad Kamal, N. S., Safuan, S., Shamsuddin, S., & Foroozandeh, P. (2020). Aging of the cells: Insight into cellular senescence and detection Methods. *Eur J Cell Biol, 99*(6), 151108. doi:10.1016/j.ejcb.2020.151108

Montes, M., Nielsen, M. M., Maglieri, G., Jacobsen, A., Hojfeldt, J., Agrawal-Singh, S., . . . Lund, A. H. (2015). The lncRNA MIR31HG regulates p16(INK4A) expression to

modulate senescence. *Nat Commun, 6*, 6967. doi:10.1038/ncomms7967

Mukherjee, A. K., Sharma, S., Sengupta, S., Saha, D., Kumar, P., Hussain, T., . . . Chowdhury, S. (2018). Telomere length-dependent transcription and epigenetic modifications in promoters remote from telomere ends. *PLoS Genet, 14*(11), e1007782. doi:10.1371/journal.pgen.1007782

Munoz-Espin, D., & Serrano, M. (2014). Cellular senescence: from physiology to pathology. *Nat Rev Mol Cell Biol, 15*(7), 482-496. doi:10.1038/nrm3823

Nassrally, M. S., Lau, A., Wise, K., John, N., Kotecha, S., Lee, K. L., & Brooks, R. F. (2019). Cell cycle arrest in replicative senescence is not an immediate consequence of telomere dysfunction. *Mech Ageing Dev, 179*, 11-22. doi:10.1016/j.mad.2019.01.009

Nelson, G., Wordsworth, J., Wang, C., Jurk, D., Lawless, C., Martin-Ruiz, C., & von Zglinicki, T. (2012). A senescent cell bystander effect: senescence-induced senescence. *Aging Cell, 11*(2), 345-349. doi:10.1111/j.1474-9726.2012.00795.x

Neurohr, G. E., Terry, R. L., Lengefeld, J., Bonney, M., Brittingham, G. P., Moretto, F., . . . Amon, A. (2019). Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. *Cell, 176*(5), 1083-1097 e1018. doi:10.1016/j.cell.2019.01.018

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne), 9*, 402. doi:10.3389/fendo.2018.00402

Ogrodnik, M. (2021). Cellular aging beyond cellular senescence: Markers of senescence prior to cell cycle arrest in vitro and in vivo. *Aging Cell, 20*(4), e13338.

doi:10.1111/acel.13338

Ogrodnik, M., Salmonowicz, H., & Gladyshev, V. N. (2019). Integrating cellular senescence with the concept of damage accumulation in aging: Relevance for clearance of senescent cells. *Aging Cell, 18*(1), e12841. doi:10.1111/acel.12841

Ohanna, M., Giuliano, S., Bonet, C., Imbert, V., Hofman, V., Zangari, J., . . . Bertolotto, C. (2011). Senescent cells develop a PARP-1 and nuclear factor-{kappa}B-associated secretome (PNAS). *Genes Dev, 25*(12), 1245-1261. doi:10.1101/gad.625811

Oja, S., Komulainen, P., Penttila, A., Nystedt, J., & Korhonen, M. (2018). Automated image analysis detects aging in clinical-grade mesenchymal stromal cell cultures. *Stem Cell Res Ther, 9*(1), 6. doi:10.1186/s13287-017-0740-x

Omori, S., Wang, T. W., Johmura, Y., Kanai, T., Nakano, Y., Kido, T., . . . Nakanishi, M. (2020). Generation of a p16 Reporter Mouse and Its Use to Characterize and Target p16(high) Cells In Vivo. *Cell Metab, 32*(5), 814-828 e816. doi:10.1016/j.cmet.2020.09.006

Osborne, B., Bakula, D., Ben Ezra, M., Dresen, C., Hartmann, E., Kristensen, S. M., . . . Scheibye-Knudsen, M. (2020). New methodologies in ageing research. *Ageing Res Rev, 62*, 101094. doi:10.1016/j.arr.2020.101094

Ovadya, Y., & Krizhanovsky, V. (2018). Strategies targeting cellular senescence. *J Clin Invest, 128*(4), 1247-1254. doi:10.1172/JCI95149

Overhoff, M. G., Garbe, J. C., Koh, J., Stampfer, M. R., Beach, D. H., & Bishop, C. L. (2014). Cellular senescence mediated by p16INK4A-coupled miRNA pathways. *Nucleic Acids Res, 42*(3), 1606-1618. doi:10.1093/nar/gkt1096

Ozcan, S., Alessio, N., Acar, M. B., Mert, E., Omerli, F., Peluso, G., & Galderisi, U. (2016). Unbiased analysis of senescence associated secretory phenotype (SASP) to identify common components following different genotoxic stresses. *Aging (Albany NY), 8*(7), 1316-1329. doi:10.18632/aging.100971

Palazzo, A. F., & Lee, E. S. (2015). Non-coding RNA: what is functional and what is junk? *Front Genet, 6*, 2. doi:10.3389/fgene.2015.00002

Park, H. S., & Kim, S. Y. (2021). Endothelial cell senescence: A machine learning-based meta-analysis of transcriptomic studies. *Ageing Res Rev, 65*, 101213. doi:10.1016/j.arr.2020.101213

Pazolli, E., Alspach, E., Milczarek, A., Prior, J., Piwnica-Worms, D., & Stewart, S. A. (2012). Chromatin remodeling underlies the senescence-associated secretory phenotype of tumor stromal fibroblasts that supports cancer progression. *Cancer Res, 72*(9), 2251-2261. doi:10.1158/0008-5472.CAN-11-3386

Pendergrass, W. R., Li, Y., Jiang, D., & Wolf, N. S. (1993). Decrease in cellular replicative potential in "giant" mice transfected with the bovine growth hormone gene correlates to shortened life span. *J Cell Physiol, 156*(1), 96-103. doi:10.1002/jcp.1041560114

Philipot, D., Guerit, D., Platano, D., Chuchana, P., Olivotto, E., Espinoza, F., . . . Brondello, J. M. (2014). p16INK4a and its regulator miR-24 link senescence and chondrocyte terminal differentiation-associated matrix remodeling in osteoarthritis. *Arthritis Res Ther, 16*(1), R58. doi:10.1186/ar4494

Piechota, M., Sunderland, P., Wysocka, A., Nalberczak, M., Sliwinska, M. A., Radwanska, K., & Sikora, E. (2016). Is senescence-associated beta-galactosidase a marker of neuronal

senescence? *Oncotarget, 7*(49), 81099-81109. doi:10.18632/oncotarget.12752

Ponten, J., Stein, W. D., & Shall, S. (1983). A quantitative analysis of the aging of human glial cells in culture. *J Cell Physiol, 117*(3), 342-352. doi:10.1002/jcp.1041170309

Prata, L., Ovsyannikova, I. G., Tchkonia, T., & Kirkland, J. L. (2018). Senescent cell clearance by the immune system: Emerging therapeutic opportunities. *Semin Immunol, 40*, 101275. doi:10.1016/j.smim.2019.04.003

Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., . . . Zhavoronkov, A. (2016). Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging (Albany NY), 8*(5), 1021-1033. doi:10.18632/aging.100968

Puvvula, P. K. (2019). LncRNAs Regulatory Networks in Cellular Senescence. *Int J Mol Sci, 20*(11). doi:10.3390/ijms20112615

Puvvula, P. K., Desetty, R. D., Pineau, P., Marchio, A., Moon, A., Dejean, A., & Bischof, O. (2014). Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat Commun, 5*, 5323. doi:10.1038/ncomms6323

Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., . . . Dinger, M. E. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res, 43*(Database issue), D168-173. doi:10.1093/nar/gku988

Rivetti di Val Cervo, P., Lena, A. M., Nicoloso, M., Rossi, S., Mancini, M., Zhou, H., . . . Melino, G. (2012). p63-microRNA feedback in keratinocyte senescence. *Proc Natl Acad Sci U S A, 109*(4), 1133-1138. doi:10.1073/pnas.1112257109

Roth, A., Boulay, K., Gross, M., Polycarpou-Schwarz, M., Mallette, F. A., Regnier, M., . . . Diederichs, S. (2018). Targeting LINC00673 expression triggers cellular senescence in lung cancer. *RNA Biol, 15*(12), 1499-1511. doi:10.1080/15476286.2018.1553481

Sabbatinelli, J., Prattichizzo, F., Olivieri, F., Procopio, A. D., Rippo, M. R., & Giuliani, A. (2019). Where Metabolism Meets Senescence: Focus on Endothelial Cells. *Front Physiol, 10*, 1523. doi:10.3389/fphys.2019.01523

Sanada, F., Taniyama, Y., Muratsu, J., Otsu, R., Shimizu, H., Rakugi, H., & Morishita, R. (2018). Source of Chronic Inflammation in Aging. *Front Cardiovasc Med, 5*, 12. doi:10.3389/fcvm.2018.00012

Sang, Y., Tang, J., Li, S., Li, L., Tang, X., Cheng, C., . . . Lv, X. B. (2016). LncRNA PANDAR regulates the G1/S transition of breast cancer cells by suppressing p16(INK4A) expression. *Sci Rep, 6*, 22366. doi:10.1038/srep22366

Santosh, B., Varshney, A., & Yadava, P. K. (2015). Non-coding RNAs: biological functions and applications. *Cell Biochem Funct, 33*(1), 14-22. doi:10.1002/cbf.3079

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell, 65*(1), 25-38. doi:10.1016/j.molcel.2016.11.029

Shay, J. W., & Wright, W. E. (2019). Telomeres and telomerase: three decades of progress. *Nat Rev Genet, 20*(5), 299-309. doi:10.1038/s41576-019-0099-1

Shi, Y., Liu, Y., Wang, J., Jie, D., Yun, T., Li, W., . . . Feng, J. (2015). Downregulated Long Noncoding RNA BANCR Promotes the Proliferation of Colorectal Cancer Cells via Downregualtion of p21 Expression. *PLoS One, 10*(4), e0122679.

doi:10.1371/journal.pone.0122679

Shokhirev, M. N., & Johnson, A. A. (2021). Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell, 20*(1), e13280. doi:10.1111/acel.13280

Song, P., An, J., & Zou, M. H. (2020). Immune Clearance of Senescent Cells to Combat Ageing and Chronic Diseases. *Cells, 9*(3). doi:10.3390/cells9030671

Soto-Gamez, A., & Demaria, M. (2017). Therapeutic interventions for aging: the case of cellular senescence. *Drug Discov Today, 22*(5), 786-795. doi:10.1016/j.drudis.2017.01.004

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol, 22*(2), 96-118. doi:10.1038/s41580-020-00315-9

Stow, J. L., & Murray, R. Z. (2013). Intracellular trafficking and secretion of inflammatory cytokines. *Cytokine Growth Factor Rev, 24*(3), 227-239. doi:10.1016/j.cytogfr.2013.04.001

Su, P., Wang, F., Qi, B., Wang, T., & Zhang, S. (2017). P53 Regulation-Association Long Non-Coding RNA (LncRNA PRAL) Inhibits Cell Proliferation by Regulation of P53 in Human Lung Cancer. *Med Sci Monit, 23*, 1751-1758. doi:10.12659/msm.900205

Sun, L., & Lin, J. D. (2019). Function and Mechanism of Long Noncoding RNAs in Adipocyte Biology. *Diabetes, 68*(5), 887-896. doi:10.2337/dbi18-0009

Swanson, M. J., Baribault, M. E., Israel, J. N., & Bae, N. S. (2016). Telomere protein RAP1 levels are affected by cellular aging and oxidative stress. *Biomed Rep, 5*(2), 181-187. doi:10.3892/br.2016.707

Takahashi, A., Okada, R., Nagao, K., Kawamata, Y., Hanyu, A., Yoshimoto, S., . . . Hara, E. (2017). Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun, 8*, 15287. doi:10.1038/ncomms15287

Takasugi, M. (2018). Emerging roles of extracellular vesicles in cellular senescence and aging. *Aging Cell, 17*(2). doi:10.1111/acel.12734

Takasugi, M., Okada, R., Takahashi, A., Virya Chen, D., Watanabe, S., & Hara, E. (2017). Small extracellular vesicles secreted from senescent cells promote cancer cell proliferation through EphA2. *Nat Commun, 8*, 15729. doi:10.1038/ncomms15728

Tzatsos, A., Paskaleva, P., Lymperi, S., Contino, G., Stoykova, S., Chen, Z., . . . Bardeesy, N. (2011). Lysine-specific demethylase 2B (KDM2B)-let-7-enhancer of zester homolog 2 (EZH2) pathway regulates cell cycle progression and senescence in primary cells. *J Biol Chem, 286*(38), 33061-33069. doi:10.1074/jbc.M111.257667

Unterluggauer, H., Mazurek, S., Lener, B., Hutter, E., Eigenbrodt, E., Zwerschke, W., & Jansen-Durr, P. (2008). Premature senescence of human endothelial cells induced by inhibition of glutaminase. *Biogerontology, 9*(4), 247-259. doi:10.1007/s10522-008-9134-x

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigo, R., & Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet, 19*(9), 535-548. doi:10.1038/s41576-018-0017-y

van Niel, G., D'Angelo, G., & Raposo, G. (2018). Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol, 19*(4), 213-228. doi:10.1038/nrm.2017.125

Vemuri, P., Lesnick, T. G., Przybelski, S. A., Graff-Radford, J., Reid, R. I., Lowe, V. J., . . . Jack, C. R., Jr. (2018). Development of a cerebrovascular magnetic resonance imaging biomarker for cognitive aging. *Ann Neurol, 84*(5), 705-716. doi:10.1002/ana.25346

Verdin, E. (2015). NAD(+) in aging, metabolism, and neurodegeneration. *Science, 350*(6265), 1208-1213. doi:10.1126/science.aac4854

Vickers, K. C., Roteta, L. A., Hucheson-Dilks, H., Han, L., & Guo, Y. (2015). Mining diverse small RNA species in the deep transcriptome. *Trends Biochem Sci, 40*(1), 4-7. doi:10.1016/j.tibs.2014.10.009

Waldron, C., & Lacroute, F. (1975). Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol, 122*(3), 855-865. doi:10.1128/jb.122.3.855-865.1975

Wallis, R., Milligan, D., Hughes, B., Mizen, H., Lopez-Dominguez, J. A., Eduputa, U., . . . Bishop, C. L. (2022). Senescence-associated morphological profiles (SAMPs): an image-based phenotypic profiling method for evaluating the inter and intra model heterogeneity of senescence. *Aging (Albany NY), 14*(10), 4220-4246. doi:10.18632/aging.204072

Wang, B., Kohli, J., & Demaria, M. (2020). Senescent Cells in Cancer Therapy: Friends or Foes? *Trends Cancer, 6*(10), 838-857. doi:10.1016/j.trecan.2020.05.004

Wang, B., Wang, L., Gasek, N. S., Zhou, Y., Kim, T., Guo, C., . . . Xu, M. (2021). An inducible p21-Cre mouse model to monitor and manipulate p21-highly-expressing senescent cells in vivo. *Nat Aging, 1*(10), 962-973. doi:10.1038/s43587-021-00107-6

Wang, L., Wang, B., Gasek, N. S., Zhou, Y., Cohn, R. L., Martin, D. E., . . . Xu, M. (2022). Targeting p21(Cip1) highly expressing cells in adipose tissue alleviates insulin resistance in obesity. *Cell Metab, 34*(1), 75-89 e78. doi:10.1016/j.cmet.2021.11.002

Wen, X., & Klionsky, D. J. (2016). Autophagy is a key factor in maintaining the regenerative capacity of muscle stem cells by promoting quiescence and preventing senescence. *Autophagy, 12*(4), 617-618. doi:10.1080/15548627.2016.1158373

West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., . . . Nakagawa, S. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J Cell Biol, 214*(7), 817-830. doi:10.1083/jcb.201601071

Westergaard, D., Moseley, P., Sorup, F. K. H., Baldi, P., & Brunak, S. (2019). Population-wide analysis of differences in disease progression patterns in men and women. *Nat Commun, 10*(1), 666. doi:10.1038/s41467-019-08475-9

Wiley, C. D., Velarde, M. C., Lecot, P., Liu, S., Sarnoski, E. A., Freund, A., . . . Campisi, J. (2016). Mitochondrial Dysfunction Induces Senescence with a Distinct Secretory Phenotype. *Cell Metab, 23*(2), 303-314. doi:10.1016/j.cmet.2015.11.011

Xiang, J. F., Yin, Q. F., Chen, T., Zhang, Y., Zhang, X. O., Wu, Z., . . . Chen, L. L. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res, 24*(5), 513-531. doi:10.1038/cr.2014.35

Xiang, Y., Ye, Y., Zhang, Z., & Han, L. (2018). Maximizing the Utility of Cancer Transcriptomic Data. *Trends Cancer, 4*(12), 823-837. doi:10.1016/j.trecan.2018.09.009

Xing, Y., Yu, T., Wu, Y. N., Roy, M., Kim, J., & Lee, C. (2006). An expectation-maximization

algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. *Nucleic Acids Res, 34*(10), 3150-3160. doi:10.1093/nar/gkl396

Yi, S., Lin, K., Jiang, T., Shao, W., Huang, C., Jiang, B., . . . Lin, D. (2020). NMR-based metabonomic analysis of HUVEC cells during replicative senescence. *Aging (Albany NY), 12*(4), 3626-3646. doi:10.18632/aging.102834

Yin, Q. F., Yang, L., Zhang, Y., Xiang, J. F., Wu, Y. W., Carmichael, G. G., & Chen, L. L. (2012). Long noncoding RNAs with snoRNA ends. *Mol Cell, 48*(2), 219-230. doi:10.1016/j.molcel.2012.07.033

Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., . . . Shen, X. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature, 580*(7801), 147-150. doi:10.1038/s41586-020-2105-3

Yosef, R., Pilpel, N., Papismadov, N., Gal, H., Ovadya, Y., Vadai, E., . . . Krizhanovsky, V. (2017). p21 maintains senescent cell viability under persistent DNA damage response by restraining JNK and caspase signaling. *EMBO J, 36*(15), 2280-2295. doi:10.15252/embj.201695553

Yun, M. H., Davaapil, H., & Brockes, J. P. (2015). Recurrent turnover of senescent cells during regeneration of a complex structure. *Elife, 4*. doi:10.7554/eLife.05505

Zhang, R., & Adams, P. D. (2007). Heterochromatin and its relationship to cell senescence and cancer therapy. *Cell Cycle, 6*(7), 784-789. doi:10.4161/cc.6.7.4079

Zhang, Z., Zhang, J., Diao, L., & Han, L. (2021). Small non-coding RNAs in human cancer: function, clinical utility, and characterization. *Oncogene, 40*(9), 1570-1577. doi:10.1038/s41388-020-01630-3

Zhao, T., Li, J., & Chen, A. F. (2010). MicroRNA-34a induces endothelial progenitor cell senescence and impedes its angiogenesis via suppressing silent information regulator 1. *Am J Physiol Endocrinol Metab, 299*(1), E110-116. doi:10.1152/ajpendo.00192.2010

Zhao, Y., Tyshkovskiy, A., Munoz-Espin, D., Tian, X., Serrano, M., de Magalhaes, J. P., . . . Gorbunova, V. (2018). Naked mole rats can undergo developmental, oncogene-induced and DNA damage-induced cellular senescence. *Proc Natl Acad Sci U S A, 115*(8), 1801-1806. doi:10.1073/pnas.1721160115

Zheng, G., Qin, Y., Clark, W. C., Dai, Q., Yi, C., He, C., . . . Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods, 12*(9), 835-837. doi:10.1038/nmeth.3478

Zhou, Y., Zhang, X., & Klibanski, A. (2012). MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol, 48*(3), R45-53. doi:10.1530/JME-12-0008

Zuckerman, B., & Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. *RNA, 25*(5), 557-572. doi:10.1261/rna.068288.118

CHAPTER II

META-ANALYSIS OF UNIVERSAL SIGNATURE IN CELLULAR SENESCENCE USING TRANSCRIPTOMIC STUDIES

ABSTRACT

Cellular senescence is an inevitable process of cell development and these senescent cells still serve as a mechanically vital part in regeneration, repair and disease progression. Cell senescence can be identified in multiple senescence-associated traits including senescence-associated protein, cell cycle arrest, DNA damage, enlarged cell area, and beta-galactosidase activity. Nevertheless, the heterogeneity of senescence features across diverse cell types and triggers of senescence arrest our further knowledge of nature of senescence. Here, we collected senescence-associated RNA sequencing datasets across different human cell types and senescence inductions to investigate the common features in a transcriptomic level. A total of 417 differentially expressed genes (DEGs) were identified in a meta-analysis way and 34 hub DEGs with strong inter-gene connectivity were found in co-expression networks. Then, machine learning-based logistic regression model was established to extract gene-level deduced features and 10 genes with non-zero coefficients were eventually confirmed in the model. Also, pathway-level features were determined through Pathifier algorithm-calculated pathway deregulation score (PDS), and processed model of 18 non-zero coefficients pathways showed

that pathways in histone modifications and tRNA modification in the nucleus and cytosol were representative in discrimination of senescent cells. Moreover, comparison analysis between senescent cells and cancer-associated fibroblasts (CAFs) projected consensus expressional changes, indicating shared transcriptome profiles may play potential roles in tumorigenesis. Taken together, we discovered the shared senescence-associated biomarkers in gene- and pathway-level analysis, and also facilitated the understanding of common properties between senescent cells and CAFs. These findings may enable possibilities to specifically target deleterious senescent cells *in vivo* and development of knowledge in cancer-associated cells.

INTRODUCTION

Cellular senescence (CS) was thought as a state of permanent cell cycle arrest after a certain number of cell division in culture. This type of senescence was first described by Hayflick and Moorhead upon observation in normal diploid fibroblasts (Hayflick & Moorhead, 1961). CS is one of the vital biological processes for tissue homeostasis, embryonic development, wound healing and tumor prevention (Ovadya & Krizhanovsky, 2018). However, senescent cells can be detrimental as well and accumulate in physiologically older tissues to trigger age-related pathologies such as tissue degradation, tissue fibrosis, arthritis, renal dysfunction, diabetes and cancer (Farr et al., 2017; Ferreira-Gonzalez et al., 2018; Martin, Soriani, & Bernard, 2020; Milanovic et al., 2018; Song, Lam, Tchkonia, Kirkland, & Sun, 2020). Notably, intercellular communication through senescence-associated secretory phenotype (SASP) secretion of the surrounded senescent cells *in vivo* can establish a microenvironment that develops the pathogenesis of age-related diseases (Faget, Ren, & Stewart, 2019; Frey, Venturelli, Zender, & Bitzer, 2018; Khosla, Farr, Tchkonia, & Kirkland, 2020).

Even though CS was mainly mirrored through cyclin-dependent kinase inhibitors and TP53/pRB tumor suppressor pathways (Campisi & d'Adda di Fagagna, 2007; Martinez-Zamudio, Robinson, Roux, & Bischof, 2017), progress to determine the common features in diverse senescent cells has been stagnated mainly due to the heterogeneity of cell type of origin and senescence triggers. Transcriptionally, significance of senescence-associated genes identified in meta-analysis can be masked by the variance of different cell types when data were integrated, even though these genes do have consistent transcriptomic alterations across senescence models (Casella et al., 2019; Dong, Wei, Zhang, & Wang, 2018).

In this study, we aimed to identify consensus feature across various senescence models in multidimensional aspects utilizing publicly available RNA sequencing (RNA-seq) datasets. Collectively, there were four types of fibroblasts, two endothelial cells and one mesenchymal

stromal cell line under four senescence inductions included in the analysis, and RNA-seq Transcripts Per Million (TPM) values were processed to eliminate the variance lays upon the datasets. Differentially expressed genes (DEGs) were determined via meta-analysis-based combined probability test, and core DEGs were identified as the most contributable variables to CS using machine learning-based logistic regression model with least absolute shrinkage and selection operator (LASSO) regularization. In parallel, we determined pathway characteristics of CS by Pathifier algorithm. Finally, transcriptomic profile of cancer-associated fibroblasts was compared to senescence-associated genes, revealing a similar expressional nature, from which helps the understanding of tumor microenvironment. With the increasing availability of data and the development of computational power, of importance, our machine learning based meta-analysis can be employed in biomarker discovery after a series of related data integration. The rapid signatures' identification through artificial intelligence methods can accelerate drug screening process and potent small molecules for senescent cells removal can be applied for treating age-related diseases caused by senescent cell accumulation.

MATERIALS AND METHODS

Data selection

Preliminarily, we searched the potential RNA-seq datasets on NCBI GEO (Clough & Barrett, 2016) database with key words “cellular senescence”. Only studies using non-immortal cell lines and the senescent phenotype identified experimentally were selected for further analysis. In total, 70 RNA-seq samples (35 senescent and 35 proliferating samples respectively) in three independent studies were included, with different cell types (including four types of fibroblasts, two endothelial cells and a mesenchymal stromal cell line) and senescence inductions (including replicative senescence, ionizing radiation, Doxorubicin and oncogene-induced senescence) (Table1 and Table S1). In addition, study GSE155343 was included for downstream comparison analysis. The complete data selection and analysis process can be viewed in Figure 1.

Data processing

The individual fastq file for each sample was obtained from sra file using fastq-dump in SRA Toolkit 2.10.9. RNA-seq data were quantified using Salmon 1.4.0 tool (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) with indexed transcript sequences (GRCh38.p13) of H. sapiens from GENCODE project (https://www.gencodegenes.org/human/). Expression profile of transcripts was merged into the gene context and was processed through transcripts per million (TPM) normalization by tximport package (Soneson, Love, & Robinson, 2015) in R.

To mitigate the inter-class effect proportion and batch effects resulted from cross-study datasets, we utilized class-specific quantile normalization of TPM values (Y. Zhao, Wong, & Goh, 2020) to better preserve useful signal for further statistical feature selection. To investigate the relevant expressional changes in individual comparison (14 comparisons in total, Table S1), following formula was employed to focus on corresponding fold changes from proliferating to senescent status:

$$\alpha = \log 2 \left[(TPM' + 1) / \ \bar{x} \right]$$

where TPM′ represented normalized TPM value and $\bar{x}$ indicated the weighted average based on the sample size of each comparison. This transformation ensured resulting values depicted differences caused by cell senescence (CS) in each comparison and eliminated the heterogeneity when using expression values across various cell types and studies.

Differential expression analysis

In a comparison group, samples with normalized TPM values for each gene were used to determine differentially expressed genes (DEGs) by Deseq2 (Love, Huber, & Anders, 2014). As a result, each gene obtained 14 raw p-values and whole p-vaue results were further processed in a meta-analysis way using fisher's combined and inverse normal methods via metaRNASeq package. Genes with Benjamini Hochberg threshold less than 0.05 in both methods were considered as statistically significant. Additionally, genes with inconsistent expression trends (upregulated or downregulated) across 14 comparisons were eventually filtered out from the DEGs list.

Weighted gene co-expression analysis

To determine the CS-related gene modules, which were composed of genes with similar expression trends among grouped samples, weighted gene co-expression analysis (WGCNA) was performed via WGCNA package (v1.69.0) in R. Briefly, log2 transformed values introduced above were as input and a soft threshold β value was set as 7 according to the criterion of approximate scale-free topology (B. Zhang & Horvath, 2005) (Figure S2).

Hub genes within CS-related module were identified with both absolute gene significance ≥ 0.2 and absolute module membership ≥ 0.8.

Model construction

Under the limitation of the sample size, especially when too many variables (genes) were presented, it is difficult to capture generalized features through conventional statistical methods

(Segata et al., 2011). Thus penalized logistic regression including ridge, least absolute shrinkage and selection operator (LASSO) and elastic net regressions impose a penalty to the logistic model for high-dimension low-sample size data, and automatically shrinking coefficients of less contributive variables toward zero. We utilized LASSO regularization with a tuning parameter $\alpha=1$ and true values of 0 and 1 were assigned to proliferating and senescent samples respectively under binomial distribution. To balance accuracy and simplicity, optimal number of genes that were the most contributive was selected with low mean absolute error (MAE) using leave-one-out cross validation (LOOCV) method (Figure S3).

Pathway-based analysis

To investigate the profiles of core pathways related to CS, a pathway-based analysis via Pathifier algorithm was performed (Drier, Sheffer, & Domany, 2013). Generally, this algorithm transforms gene expression-level data into pathway-level information based on deviations of pathway (according to known genes in pathway) among samples (assigned into multiple groups by phenotypes), and each sample in each pathway acquires a pathway deregulation score (PDS) according to the differences between case samples (senescent samples in our study) and normal samples (proliferating samples in our study) on principal curve. Specific principle about formation of principal curve and projection of each sample can be explained by Hastie and Stuetzle's algorithm (Hastie & Stuetzle, 1989).

RESULTS

Identification of differentially expressed genes across multiple cellular senescence subsets

In this study, we collected RNA-seq data of non-immortal cell lines from 3 studies that cellular senescence (CS) phenotype had been verified through senescence tests, for example, senescence-associated beta-galactosidase (SA beta-gal) activity, cell cycle arrest and/or senescence markers detection (Table 1). A total of 70 samples were included and the individual experiment comparison was performed based on their cell types and senescence inductions. As a result, there were 14 individual differential expression comparisons conducted via Deseq2 (Table S1). Respective p-values from proliferating/senescent comparisons were merged in meta-analysis methods including Fisher's and inverse normal combined methods. There were 17403 differentially expressed genes (DEGs) with an adjusted p-value ≤ 0.05 in both methods. Furthermore, we only retained 417 genes that showed consistent expression trends across 14 experiment comparisons (Figure 2A, Table S2). Expression profiling of senescence-associated genes selected from CellAge database (Avelar et al., 2020) was exhibited and there were 9 genes were DEGs, with one gene (*PEX19*) was reported to induce senescence and 8 genes (*SENP1, EZH2, CENPA, DEK, TPR, USP1, HMGB1, DHX9*) were identified to inhibit CS process (Figure 2B).

Also, we investigated the correlation between differentially expressed transcripts and their corresponding sequence conservations. Basewise and element conservation scores of human transcriptomic sequence compared to 20 vertebrates were obtained through PhastCons and Phylop computation respectively and there were 42 protein coding transcripts and 3 long non-coding RNAs were conserved beyond 99th percentile of the whole transcripts input (Figure S1, Table S3).

Hub CS-associated genes within co-expression networks

Combined with DEGs related to CS, we further investigate the gene functions based on co-expression network construction. To avoid the genetic variance from different source of senescence models, we log-transformed expression values after compared with weighted average in each paired comparisons. After choosing 8 as an appropriate soft-threshold in weighted gene co-expression network analysis (WGCNA), input genes were categorized into 17 modules with different Pearson's correlation coefficients between modules and senescence phenotype (Table S4 and Figure S2). Particularly, the turquoise module had the highest correlation with CS and corresponding gene expressions can classify proliferating and senescent samples regardless the cell type of origin (Figure 3A). Hub genes within CS co-expression networks were identified based on threshold: both absolute gene significance ≥ 0.2 and absolute turquoise membership (intra-modular connectivity) ≥ 0.8. There were 172 and 700 hub genes were up-regulated and down-regulated respectively (Figure 3B). Their gene ontology enrichment from Metascape resource (Y. Zhou et al., 2019) showed that down-regulated hub genes were enriched in pre-RNA processing, ribonucleoprotein complex biogenesis, nuclear transport, cell division, DNA replication, DNA repair and chromosome maintenance, while up-regulated hub genes were enriched in glycerophospholipid catabolism, myeloid leukocyte mediated immunity, lysosome, and cellular cation homeostasis (Figure 3C).

To further identify differentially expressed genes with potentially core CS-associated function reflected by network degree (module membership), results from meta-analysis comparison and WGCNA were overlapped. There were 58 and 91 up-/down-regulated hub DEGs were found, with 2 up-regulated hub DEGs and 32 down-regulated hub DEGs having intra-modular connectivity ≥ 0.9 respectively (Figure 4A and 4B). Previous study suggested that gene expression changes were negatively related to intron retention (IR) during CS (Yao et al., 2020), due to the unstable transcripts degradation by RNA surveillance machinery (Wong et al., 2013; Yap, Lim, Khandelia, Friedman, & Makeyev, 2012). Thus we explored the correlation

between IR and expression changes and overall IR quantification was performed by IRFinder (Middleton et al., 2017). Intriguingly, most of up-regulated DEGs were less intron retained in senescent samples and genes with increased IR were more likely down-regulated (Figure 4C), indicating the role of precursor mRNA processing during CS.

Feature selection of CS via model construction

To accurately acquire a distinguishable feature of CS, a dimensional reduction of data was performed through supervised logistic (Binomial) regression with LASSO (least absolute shrinkage and selection operator) regularization. Performance of constructed model was evaluated and optimal regularization parameter (λ) was determined via leave-one-out cross-validation (LOOCV). Based on the principle of model accuracy and simplicity, we chose the log of the λ simultaneously having small prediction error and avoiding model over-fitting (Figure S3). As a result, 10 genes were identified with non-zero coefficients (Figure 5A), and coefficient values of 1 up-regulated gene (*AL353138.1*) and 9 down-regulated genes (*DHX9*, *METTL17*, *SART3*, *SNX5*, *ELMO3*, *ITPRIPL1*, *FANCE*, *CDKAL1*, *SRSF2*) are shown in Figure 5B.

Results of model performance from LOOCV confirmed its capability of predicting senescent phenotype: with 1 area under the receiver operating curve (AUROC) in overall sample examination, and small Binomial Deviance, misclassification error (ME), mean squared error (MSE) and mean absolute error (MAE) (Table 2). Also, the fitted probability for all samples matched their actual phenotypes (Figure 5C and Table S5). We also assessed the effectiveness of remained genes in discriminating senescent cells by performing principal component analysis (PCA), and distance between senescent and non-senescent cells (variance on the PC1 axis) gradually enlarged when numbers of gene decreased (Figure S4), indicating a better performance of data dimensionality reduction.

Machine learning-based pathway-level analysis

Inspired by discriminatory pattern between proliferating and senescent cell lines in gene expression level, we further explored the possibility to determine universal pathway signals of CS across diverse cell types and senescence inductions. Logarithmic normalized data of all genes and pathway information from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000), BioCarta (Nishimura, 2001), Pathway Interaction Database (PID) (Schaefer et al., 2009), Reactome (Fabregat et al., 2017) and WikiPathways (Martens et al., 2021) were employed to integrate pathway deregulation score (PDS) through Pathifier algorithm. As a result, there were 2719 pathways scored for each sample according to the distance between proliferating and senescent samples along the projected principal curve of individual pathways. After that, we used this pathway-level matrix to perform variables minimization by logistic regression similar to gene-level analysis we described and 18 pathways with non-zero coefficients were obtained via LASSO regularization (Figure 6A, Table S6). Notably, the two most representative pathways with the largest absolute coefficients were Histone modifications and tRNA modification in the nucleus and cytosol (Figure 6B).

Transcriptomic features comparison with cancer-associated fibroblasts

Next, we sought to determine if there was a similar expression profile between senescent cells and cancer-associated fibroblasts (CAFs), both of which are embedded in tumor microenvironment that associates with tumorigenesis (Liu et al., 2019; Sahai et al., 2020). We performed transcriptome comparison analysis between CS and CAFs from GSE155343 study (Table 1), and DEGs from CAFs (compared to normal skin fibroblasts) had a consistent expression trend during CS (Figure S5A). We further investigated relative expression fold change of DEGs from CS with CAFs, and intriguingly found there was a positive correlation between CS and CAFs based on 417 DEGs (R^2 = 0.5253, Figure S5C). There were 115 genes that were differentially expressed in both CS and CAFs (Figure 7), and they were mainly enriched in cell

cycle related biological processes such as cell division (GO: 0051301), cell cycle phase transition (GO: 0044770), mitotic cell cycle phase transition (GO: 0044772), regulation of cell cycle process (GO: 0010564) and regulation of mitotic cell cycle (GO: 0007346) (Table 3).

Also, we compared the expression feature between CS and colorectal cancer cells co-cultured normal fibroblasts (CNFs), and there was no obvious correlation (Figure S5B and D). We only found 5 genes were differentially expressed in both CS and CNFs (Figure S6), indicating a heterogeneity between CAFs and CNFs.

DISCUSSION

Our study performed a quantitative meta-analysis of transcriptomic profiles in multiple senescence models (7 cell types and 4 senescence triggers) (Table 1) and identified consensus features through dimensionality reduction regression. There were 214 up-regulated and 203 down-regulated genes among all comparison experiments (Table S2). By comparing with senescence-associated genes listed in CellAge (Avelar et al., 2020), there were 9 differentially expressed genes (DEGs) previously identified to be contributable to CS (Figure 2B). Notably, we took specified normalization by weighted average after individual comparisons and data were integrated for downstream analysis, efficiently mitigating the interfering variables from cell type origin (Casella et al., 2019; Dong et al., 2018).

A total of 34 DEGs having relative high connectivity with senescence-associated phenotype were obtained from WGCNA and some of them have negative correlation between transcripts expression and intron retention (Figure 4C). For example, the down-regulated hub DEGs nurim (*NRM*) and kinesin family member 20A (*KIF20A*) showed higher intron retention rate in senescent cells while there was opportune splicing in up-regulated hub DEGs cytochrome b5 reductase 1 (*CYB5R1*) and serine incorporator 1 (*SERINC1*). As a nuclear envelope membrane protein locating at the inner nuclear membrane with a six transmembrane-structure, NRM is slightly expressed in human brain, heart and skin tissues and plays a vital role in physiological process of DNA damage and repair response (Hetzer, 2010; Hofemeister & O'Hare, 2005). Also, *NRM* acts as a mediator in embryonic heart morphogenesis in mice through its alternative splicing variants (W. Zhang et al., 2017). Noted as a significant biomarker in several cancer lines, *KIF20A* with high-expression confers the progression of malignant phenotype (Lu et al., 2018; Ma et al., 2019; Nakamura et al., 2020; X. Zhao et al., 2018). In drug therapeutic mechanisms of breast cancer cells, down-regulation of *KIF20A* was observed after paclitaxel treatment and cell senescence was achieved via mitotic catastrophe (Khongkow et al., 2016). *CYB5R1* recently has

been identified as a necessary oxidoreductase to trigger membrane damage through phospholipids oxidation during ferroptosis, a form of regulated necrotic cell death.

A machine learning-based modeling method was employed and 10 gene signals with non-zero coefficients were obtained for distinguishing senescent cells in simplicity and accuracy. The underlying function of transcriptomic features was further investigated through pathway-level analysis by Pathifier algorithm. Unlike other pathway enrichment analyses, generation of pathway deregulation score (PDS) by incorporating all included genes information helps to elaborate sample characteristics on multi-dimensional coordinate positioning. In our results, 2719 pathway information from 5 public pathway databases were reduced to 18 core pathways with non-zero coefficients via binomial model regression (Figure 6A, Table S6). The Histone modification was a pathway with the highest absolute coefficient value, and was closely related to microenvironment adaptation during genomic reorganization, which is initialized by irreparable DNA damage (Paluvai, Di Giorgio, & Brancolini, 2020). There were five histone modifications related genes (*H3C12, H4C2, EZH2, H3C3* and *SETMAR*) down-regulated in our study and three of them (*H3C12, H4C2* and *H3C3*) were members of histones that form core component of nucleosome. It is illustrious that *EZH2* as a main component of polycomb repressor complex 2 (*PRC2*), which is an essential regulator of cell growth, and its high level expression has been verified to associate with cancer aggressiveness by silencing mostly tumor-suppressing genes (Chang & Hung, 2012; Sha et al., 2016). Also, oncogene-induced senescence (OIS) is triggered by activating tumor-suppressing genes through *EZH2* repression, which leads to loosed methylation status on histone 3 lysine 27 (*H3K27*) (Paluvai et al., 2020). Notably, *CDK5* Regulatory Subunit Associated Protein 1 Like 1 (*CDKAL1*) is one of genes in our penalized model (Figure 5B) and also a member of tRNA modification in the nucleus and cytosol pathway, and it is a tRNA-modifying enzyme that specifically recognizes tRNALys3 to establish sufficient codon-anticodon binding for decoding of Lys codons (Wei et al., 2011). CDKAL1 has been

associated with the susceptibility to type II diabetes through impairing insulin biosynthesis, which can decrease insulin secretion. However, little evidence of regulatory function to the senescence model has been showed and further experimental investigation should be developed to clue the underlying mechanism of *CDKAL1* in universal CS.

We observed an interesting result that both senescence-associated genes and pathways with non-zero coefficients consistently showed reduced activity in senescent cells, reflected by negative values within logistic models (Figure 5B and Table S6). Remained genes that have strong recognizability between proliferating and stagnant cells should sufficiently and consistently exhibit expressional changes between two statuses. We found that 10 core genes in our regression model were considered as DEGs within more comparisons than other genes (Table S7). Also, 18 senescence-associated pathways were all down-regulated in senescent cells, and included genes were enriched in epigenomic modification and chromatin regulator. Previous studies mentioned that epigenetic remodeling is vital process during CS, especially resulting inactivation of cell cycle dependent genes (Glauche, Thielecke, & Roeder, 2011; Kargapolova et al., 2021; Tanaka et al., 2020). By contrast, highly expressed genes that trigger shifts of cellular aging are heterogeneous across different cell types, inducers and time courses, leaving diverse transcriptomic signatures, SASP and metabolic activity (Hernandez-Segura et al., 2017; Hernandez-Segura, Nehme, & Demaria, 2018; Pantazi et al., 2019). So we hypothesize that genes are directly related to chromosomal structure and cell cycle maintenance may show more consistent expression across multiple senescence models while those potential inducers of CS have more variations, and this should be further investigated in the future.

Our advanced aim is to investigate the homogeneity between cancer associated fibroblasts (CAFs) and senescent cells. Previous studies indicated CAFs functionally promote tumorigenesis by angiogenesis induction and extracellular matrix (ECM) remodeling (Kalluri, 2016; LeBleu & Kalluri, 2018), which also mechanically applies to aging-associated cancer development by

senescent cells SASP secretion (Bottazzi, Riboli, & Mantovani, 2018). Expressional comparison was conducted and there were 36 and 79 genes both up- and down-regulated in CS and CAFs (Figure 7) and DEGs from CS exhibited similar tendency in CAFs (Figure S5C). Among these common DEGs in two statuses of fibroblasts, it is necessary to notice that two down-regulation genes Fanconi Anemia Complementation Group E (*FANCE*) and Sorting Nexin 5 (*SNX5*) were core contributors in our constructed senescence model (Figure 5B). *FANCE* is one of members of the Fanconi anemia complementation group proteins and activated in DNA repair process through FANC-BRCA pathway (Bouffard et al., 2015; Hodson & Walden, 2012). Recent study utilized prognostic model in a pan-cancer level and uncovered *FANCE* participated in cell apoptosis through the Wnt/β-catenin pathway and its up-regulation promoted the infiltration of immune cells (Lin et al., 2021). High expression of *SNX5* is considered as a biomarker indicating poor prognosis in several tumor types (Ara et al., 2012; Cai et al., 2019; Q. Zhou et al., 2020), and also plays crucial role in insulin and glucose metabolism (F. Li et al., 2018). The advanced research is still needed to explore the function of *SNX5* in CS and its molecular mechanism that affects consistently among diverse senescent cell models.

Moreover, some identified long non-coding RNAs (lncRNAs) in our analysis provide insight into lncRNAs mode of action in senescence-related process. Particularly, a total of 5 lncRNAs (*MAP4K3-DT*, *LINC00511*, *LINC01670*, *AC091057.1* and *TMPO-AS1*) obtained in both CS and CAFs, indicating their potential regulatory function in cell cycle arrest and further microenvironment construction by intercellular signaling transduction. Recent study showed that *AC091057.1* is an immune risk signature in lung adenocarcinoma, with a positive correlation between expression and risk scores (Jin, Song, Chen, & Zhang, 2020). *TMPO-AS1* has been widely studied in several tumor development and functions in cell proliferation via competing endogenous RNA (ceRNA) network (H. Li, Zhou, Cheng, Tian, & Yang, 2020; Mitobe et al., 2019; Peng, Yan, & Cheng, 2020; Qin, Zheng, & Fang, 2019). Instead of high expression shown

in cancer cells (Agbana et al., 2020; Jiang, Xie, Bi, Ding, & Mei, 2020; Mao et al., 2019; Yu, Xu, & Yuan, 2019), in our study *LINC00511* shows a decreased signal in CS and CAFs.

CONCLUSION

In closing, we took advantage of multiple computational methods to construct senescence-associated gene modules and pathways that consistently showed expressional alterations across diverse senescence experiments, while further research should be developed to uncover whether these core features act on CS and by what mechanisms. For biomarker discovery, genes and pathways identified here can describe the dynamics of senescence or abnormal senescence associated pathologies, and constructed classification model is still suitable for predicting new samples' senescence probability. With consensus connecting CS to aging and cancer-related pathogenesis, the senescence model identified in this study provides potential molecular references for targeted tumor therapy.

REFERENCES

Agbana, Y. L., Abi, M. E., Ni, Y., Xiong, G., Chen, J., Yun, F., . . . Zhu, Y. (2020). LINC00511 as a prognostic biomarker for human cancers: a systematic review and meta-analysis. *BMC Cancer, 20*(1), 682. doi:10.1186/s12885-020-07188-3

Ara, S., Kikuchi, T., Matsumiya, H., Kojima, T., Kubo, T., Ye, R. C., . . . Ichimiya, S. (2012). Sorting nexin 5 of a new diagnostic marker of papillary thyroid carcinoma regulates Caspase-2. *Cancer Sci, 103*(7), 1356-1362. doi:10.1111/j.1349-7006.2012.02296.x

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., . . . de Magalhaes, J. P. (2020). A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol, 21*(1), 91. doi:10.1186/s13059-020-01990-9

Bottazzi, B., Riboli, E., & Mantovani, A. (2018). Aging, inflammation and cancer. *Semin Immunol, 40*, 74-82. doi:10.1016/j.smim.2018.10.011

Bouffard, F., Plourde, K., Belanger, S., Ouellette, G., Labrie, Y., & Durocher, F. (2015). Analysis of a FANCE Splice Isoform in Regard to DNA Repair. *J Mol Biol, 427*(19), 3056-3073. doi:10.1016/j.jmb.2015.08.004

Cai, J., Sun, M., Hu, B., Windle, B., Ge, X., Li, G., & Sun, Y. (2019). Sorting Nexin 5 Controls Head and Neck Squamous Cell Carcinoma Progression by Modulating FBW7. *J Cancer, 10*(13), 2942-2952. doi:10.7150/jca.31055

Campisi, J., & d'Adda di Fagagna, F. (2007). Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol, 8*(9), 729-740. doi:10.1038/nrm2233

Casella, G., Munk, R., Kim, K. M., Piao, Y., De, S., Abdelmohsen, K., & Gorospe, M. (2019). Transcriptome signature of cellular senescence. *Nucleic Acids Res, 47*(14), 7294-7305.

doi:10.1093/nar/gkz555

Chang, C. J., & Hung, M. C. (2012). The role of EZH2 in tumour progression. *Br J Cancer, 106*(2), 243-247. doi:10.1038/bjc.2011.551

Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. *Methods Mol Biol, 1418*, 93-110. doi:10.1007/978-1-4939-3578-9_5

Dong, Q., Wei, L., Zhang, M. Q., & Wang, X. (2018). Regulatory RNA binding proteins contribute to the transcriptome-wide splicing alterations in human cellular senescence. *Aging (Albany NY), 10*(6), 1489-1505. doi:10.18632/aging.101485

Drier, Y., Sheffer, M., & Domany, E. (2013). Pathway-based personalized analysis of cancer. *Proc Natl Acad Sci U S A, 110*(16), 6388-6393. doi:10.1073/pnas.1219651110

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., . . . Hermjakob, H. (2017). Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics, 18*(1), 142. doi:10.1186/s12859-017-1559-2

Faget, D. V., Ren, Q., & Stewart, S. A. (2019). Unmasking senescence: context-dependent effects of SASP in cancer. *Nat Rev Cancer, 19*(8), 439-453. doi:10.1038/s41568-019-0156-2

Farr, J. N., Xu, M., Weivoda, M. M., Monroe, D. G., Fraser, D. G., Onken, J. L., . . . Khosla, S. (2017). Targeting cellular senescence prevents age-related bone loss in mice. *Nat Med, 23*(9), 1072-1079. doi:10.1038/nm.4385

Ferreira-Gonzalez, S., Lu, W. Y., Raven, A., Dwyer, B., Man, T. Y., O'Duibhir, E., . . . Forbes, S. J. (2018). Paracrine cellular senescence exacerbates biliary injury and impairs regeneration. *Nat Commun, 9*(1), 1020. doi:10.1038/s41467-018-03299-5

Frey, N., Venturelli, S., Zender, L., & Bitzer, M. (2018). Cellular senescence in gastrointestinal diseases: from pathogenesis to therapeutics. *Nat Rev Gastroenterol Hepatol, 15*(2), 81-95. doi:10.1038/nrgastro.2017.146

Glauche, I., Thielecke, L., & Roeder, I. (2011). Cellular aging leads to functional heterogeneity of hematopoietic stem cells: a modeling perspective. *Aging Cell, 10*(3), 457-465.

doi:10.1111/j.1474-9726.2011.00692.x

Hastie, T., & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association, 84*(406), 502-516. doi:10.1080/01621459.1989.10478797

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res, 25*, 585-621. doi:10.1016/0014-4827(61)90192-6

Hernandez-Segura, A., de Jong, T. V., Melov, S., Guryev, V., Campisi, J., & Demaria, M. (2017). Unmasking Transcriptional Heterogeneity in Senescent Cells. *Curr Biol, 27*(17), 2652-2660 e2654. doi:10.1016/j.cub.2017.07.033

Hernandez-Segura, A., Nehme, J., & Demaria, M. (2018). Hallmarks of Cellular Senescence. *Trends Cell Biol, 28*(6), 436-453. doi:10.1016/j.tcb.2018.02.001

Hetzer, M. W. (2010). The nuclear envelope. *Cold Spring Harb Perspect Biol, 2*(3), a000539. doi:10.1101/cshperspect.a000539

Hodson, C., & Walden, H. (2012). Towards a molecular understanding of the fanconi anemia core complex. *Anemia, 2012*, 926787. doi:10.1155/2012/926787

Hofemeister, H., & O'Hare, P. (2005). Analysis of the localization and topology of nurim, a polytopic protein tightly associated with the inner nuclear membrane. *J Biol Chem, 280*(4), 2512-2521. doi:10.1074/jbc.M410504200

Jiang, L., Xie, X., Bi, R., Ding, F., & Mei, J. (2020). Knockdown of Linc00511 inhibits TGF-beta-induced cell migration and invasion by suppressing epithelial-mesenchymal transition and down-regulating MMPs expression. *Biomed Pharmacother, 125*, 109049. doi:10.1016/j.biopha.2019.109049

Jin, D., Song, Y., Chen, Y., & Zhang, P. (2020). Identification of a Seven-lncRNA Immune Risk Signature and Construction of a Predictive Nomogram for Lung Adenocarcinoma. *Biomed Res Int, 2020*, 7929132. doi:10.1155/2020/7929132

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer, 16*(9), 582-598. doi:10.1038/nrc.2016.73

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30. doi:10.1093/nar/28.1.27

Kargapolova, Y., Rehimi, R., Kayserili, H., Bruhl, J., Sofiadis, K., Zirkel, A., . . . Papantonis, A. (2021). Overarching control of autophagy and DNA damage response by CHD6 revealed by modeling a rare human pathology. *Nat Commun, 12*(1), 3014. doi:10.1038/s41467-021-23327-1

Khongkow, P., Gomes, A. R., Gong, C., Man, E. P., Tsang, J. W., Zhao, F., . . . Lam, E. W. (2016). Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. *Oncogene, 35*(8), 990-1002. doi:10.1038/onc.2015.152

Khosla, S., Farr, J. N., Tchkonia, T., & Kirkland, J. L. (2020). The role of cellular senescence in ageing and endocrine disease. *Nat Rev Endocrinol, 16*(5), 263-275. doi:10.1038/s41574-020-0335-y

LeBleu, V. S., & Kalluri, R. (2018). A peek into cancer-associated fibroblasts: origins, functions and translational impact. *Disease Models & Mechanisms, 11*.

Li, F., Yang, J., Villar, V. A. M., Asico, L. D., Ma, X., Armando, I., . . . Wang, X. (2018). Loss of renal SNX5 results in impaired IDE activity and insulin resistance in mice. *Diabetologia, 61*(3), 727-737. doi:10.1007/s00125-017-4482-1

Li, H., Zhou, Y., Cheng, H., Tian, J., & Yang, S. (2020). Roles of a TMPO-AS1/microRNA-200c/TMEFF2 ceRNA network in the malignant behaviors and 5-FU resistance of ovarian cancer cells. *Exp Mol Pathol, 115*, 104481. doi:10.1016/j.yexmp.2020.104481

Lin, B., Li, H., Zhang, T., Ye, X., Yang, H., & Shen, Y. (2021). Comprehensive analysis of macrophage-related multigene signature in the tumor microenvironment of head and neck squamous cancer. *Aging (Albany NY), 13*(4), 5718-5747. doi:10.18632/aging.202499

Liu, T., Han, C., Wang, S., Fang, P., Ma, Z., Xu, L., & Yin, R. (2019). Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J Hematol Oncol, 12*(1), 86.

doi:10.1186/s13045-019-0770-1

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lu, M., Huang, X., Chen, Y., Fu, Y., Xu, C., Xiang, W., . . . Yu, C. (2018). Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. *IUBMB Life, 70*(4), 328-335. doi:10.1002/iub.1726

Ma, H., Tian, T., Liu, X., Xia, M., Chen, C., Mai, L., . . . Yu, L. (2019). Upregulated circ_0005576 facilitates cervical cancer progression via the miR-153/KIF20A axis. *Biomed Pharmacother, 118*, 109311. doi:10.1016/j.biopha.2019.109311

Mao, B. D., Xu, P., Xu, P., Zhong, Y., Ding, W. W., & Meng, Q. Z. (2019). LINC00511 knockdown prevents cervical cancer cell proliferation and reduces resistance to paclitaxel. *J Biosci, 44*(2).

Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, Denise N., Hanspers, K., . . . Kutmon, M. (2021). WikiPathways: connecting communities. *Nucleic Acids Research, 49*(D1), D613-D621. doi:10.1093/nar/gkaa1024

Martin, N., Soriani, O., & Bernard, D. (2020). Cardiac Glycosides as Senolytic Compounds. *Trends Mol Med, 26*(3), 243-245. doi:10.1016/j.molmed.2020.01.001

Martinez-Zamudio, R. I., Robinson, L., Roux, P. F., & Bischof, O. (2017). SnapShot: Cellular Senescence Pathways. *Cell, 170*(4), 816-816 e811. doi:10.1016/j.cell.2017.07.049

Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J. J., . . . Ritchie, W. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol, 18*(1), 51. doi:10.1186/s13059-017-1184-4

Milanovic, M., Fan, D. N. Y., Belenki, D., Dabritz, J. H. M., Zhao, Z., Yu, Y., . . . Schmitt, C. A. (2018). Senescence-associated reprogramming promotes cancer stemness. *Nature, 553*(7686), 96-100. doi:10.1038/nature25167

Mitobe, Y., Ikeda, K., Suzuki, T., Takagi, K., Kawabata, H., Horie-Inoue, K., & Inoue, S. (2019). ESR1-Stabilizing Long Noncoding RNA TMPO-AS1 Promotes Hormone-Refractory Breast Cancer Progression. *Mol Cell Biol, 39*(23). doi:10.1128/MCB.00261-19

Nakamura, M., Takano, A., Thang, P. M., Tsevegjav, B., Zhu, M., Yokose, T., . . . Daigo, Y. (2020). Characterization of KIF20A as a prognostic biomarker and therapeutic target for different subtypes of breast cancer. *Int J Oncol, 57*(1), 277-288. doi:10.3892/ijo.2020.5060

Nishimura, D. (2001). BioCarta. *Biotech Software & Internet Report, 2*(3), 117-120. doi:10.1089/152791601750294344

Ovadya, Y., & Krizhanovsky, V. (2018). Strategies targeting cellular senescence. *J Clin Invest, 128*(4), 1247-1254. doi:10.1172/JCI95149

Paluvai, H., Di Giorgio, E., & Brancolini, C. (2020). The Histone Code of Senescence. *Cells, 9*(2). doi:10.3390/cells9020466

Pantazi, A., Quintanilla, A., Hari, P., Tarrats, N., Parasyraki, E., Dix, F. L., . . . Finch, A. J. (2019). Inhibition of the 60S ribosome biogenesis GTPase LSG1 causes endoplasmic reticular disruption and cellular senescence. *Aging Cell, 18*(4), e12981. doi:10.1111/acel.12981

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods, 14*(4), 417-419. doi:10.1038/nmeth.4197

Peng, X., Yan, J., & Cheng, F. (2020). LncRNA TMPO-AS1 up-regulates the expression of HIF-1alpha and promotes the malignant phenotypes of retinoblastoma cells via sponging miR-199a-5p. *Pathol Res Pract, 216*(4), 152853. doi:10.1016/j.prp.2020.152853

Qin, Z., Zheng, X., & Fang, Y. (2019). Long noncoding RNA TMPO-AS1 promotes progression of non-small cell lung cancer through regulating its natural antisense transcript TMPO. *Biochem Biophys Res Commun, 516*(2), 486-493. doi:10.1016/j.bbrc.2019.06.088

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., . . . Werb, Z. (2020). A framework for advancing our understanding of cancer-associated fibroblasts.

Nat Rev Cancer, 20(3), 174-186. doi:10.1038/s41568-019-0238-1

Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., & Buetow, K. H. (2009). PID: the Pathway Interaction Database. *Nucleic Acids Res, 37*(Database issue), D674-679. doi:10.1093/nar/gkn653

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol, 12*(6), R60. doi:10.1186/gb-2011-12-6-r60

Sha, M. Q., Zhao, X. L., Li, L., Li, L. H., Li, Y., Dong, T. G., . . . Wang, Z. (2016). EZH2 mediates lidamycin-induced cellular senescence through regulating p21 expression in human colon cancer cells. *Cell Death Dis, 7*(11), e2486. doi:10.1038/cddis.2016.383

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res, 4*, 1521. doi:10.12688/f1000research.7563.2

Song, S., Lam, E. W., Tchkonia, T., Kirkland, J. L., & Sun, Y. (2020). Senescent Cells: Emerging Targets for Human Aging and Age-Related Diseases. *Trends Biochem Sci, 45*(7), 578-592. doi:10.1016/j.tibs.2020.03.008

Tanaka, H., Igata, T., Etoh, K., Koga, T., Takebayashi, S. I., & Nakao, M. (2020). The NSD2/WHSC1/MMSET methyltransferase prevents cellular senescence-associated epigenomic remodeling. *Aging Cell, 19*(7), e13173. doi:10.1111/acel.13173

Wei, F. Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A., . . . Tomizawa, K. (2011). Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. *J Clin Invest, 121*(9), 3598-3608. doi:10.1172/JCI58056

Wong, J. J., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W., Huang, Y., . . . Rasko, J. E. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell, 154*(3), 583-595. doi:10.1016/j.cell.2013.06.052

Yao, J., Ding, D., Li, X., Shen, T., Fu, H., Zhong, H., . . . Ni, T. (2020). Prevalent intron retention

fine-tunes gene expression and contributes to cellular senescence. *Aging Cell, 19*(12), e13276. doi:10.1111/acel.13276

Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., & Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev, 26*(11), 1209-1223. doi:10.1101/gad.188037.112

Yu, C. L., Xu, X. L., & Yuan, F. (2019). LINC00511 is associated with the malignant status and promotes cell proliferation and motility in cervical cancer. *Biosci Rep, 39*(9). doi:10.1042/BSR20190903

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol, 4*, Article17. doi:10.2202/1544-6115.1128

Zhang, W., Bai, T., Zhang, S., Xu, S., Chen, H., & Li, C. (2017). Isoforms of the nuclear envelope protein Nurim are differentially expressed during heart development in mice. *Gene, 627*, 123-128. doi:10.1016/j.gene.2017.06.009

Zhao, X., Zhou, L. L., Li, X., Ni, J., Chen, P., Ma, R., . . . Feng, J. (2018). Overexpression of KIF20A confers malignant phenotype of lung adenocarcinoma by promoting cell proliferation and inhibiting apoptosis. *Cancer Med, 7*(9), 4678-4689. doi:10.1002/cam4.1710

Zhao, Y., Wong, L., & Goh, W. W. B. (2020). How to do quantile normalization correctly for gene expression data analyses. *Sci Rep, 10*(1), 15534. doi:10.1038/s41598-020-72664-6

Zhou, Q., Huang, T., Jiang, Z., Ge, C., Chen, X., Zhang, L., . . . Tian, H. (2020). Upregulation of SNX5 predicts poor prognosis and promotes hepatocellular carcinoma progression by modulating the EGFR-ERK1/2 signaling pathway. *Oncogene, 39*(10), 2140-2155. doi:10.1038/s41388-019-1131-9

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., . . . Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun, 10*(1), 1523. doi:10.1038/s41467-019-09234-6

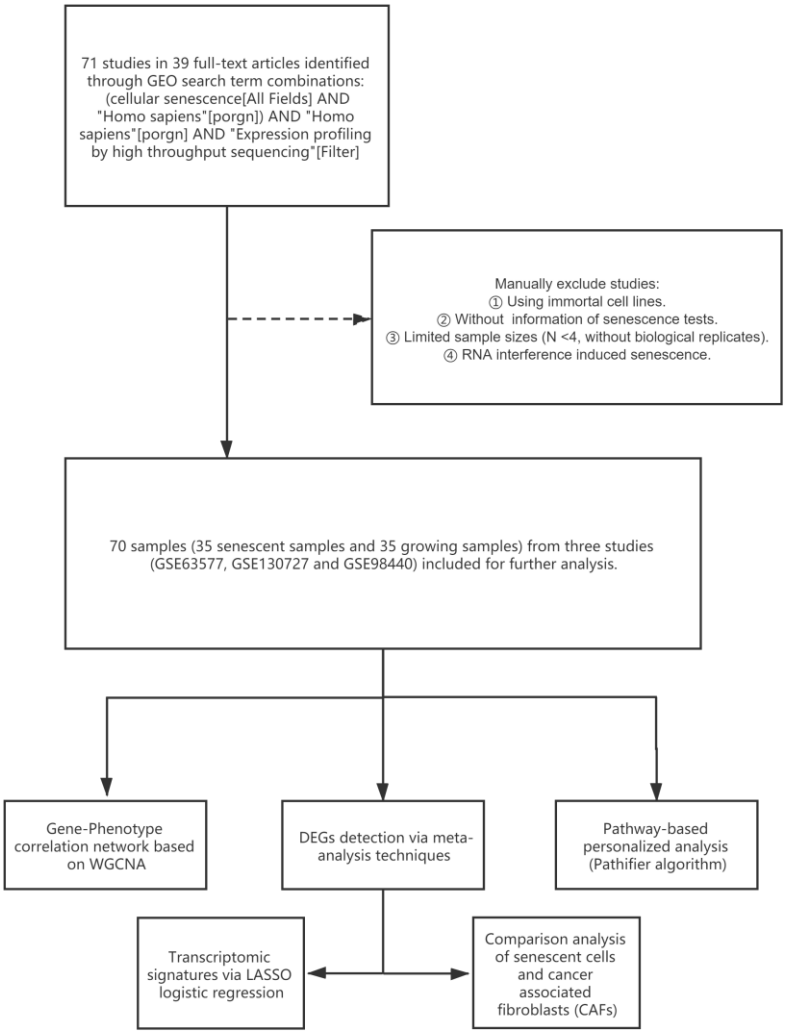


Figure 1. A diagram of the data selection and analysis contents. GEO, Gene Expression Omnibus

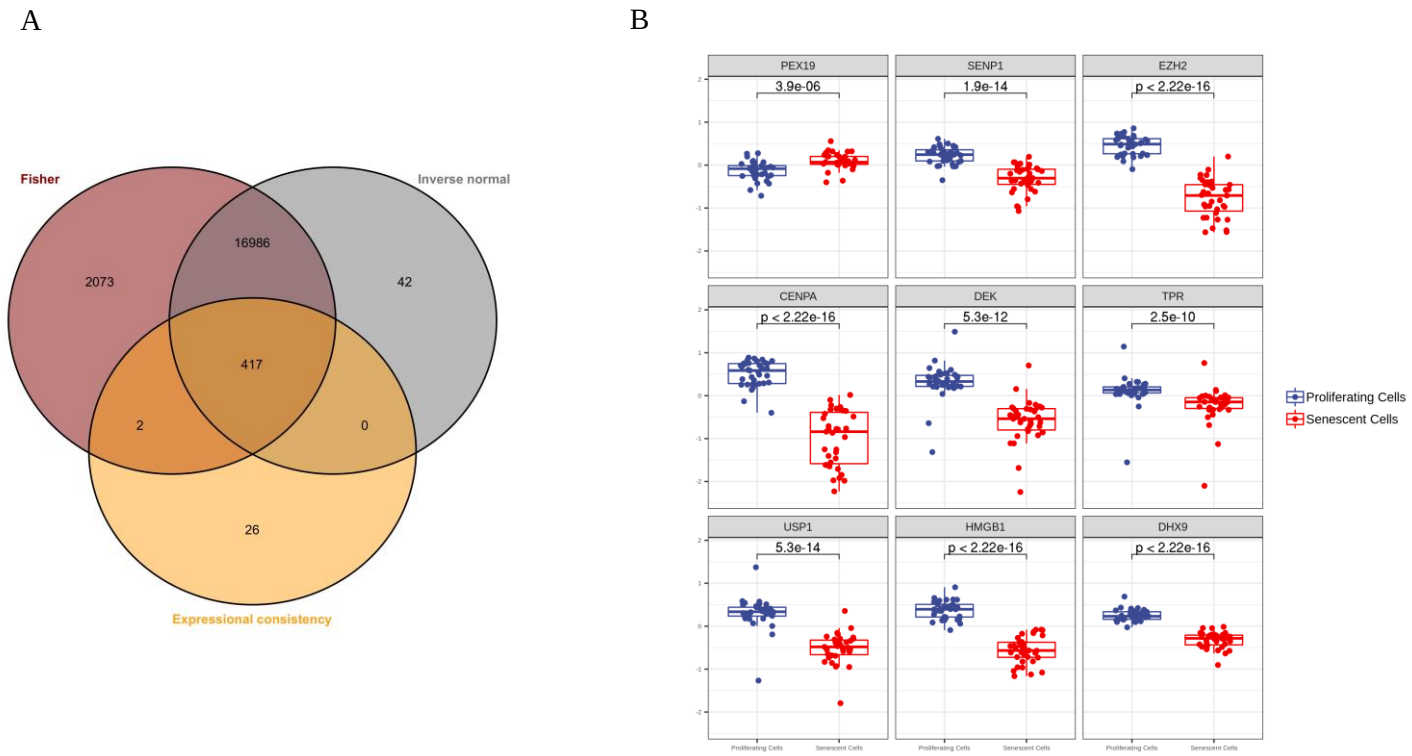
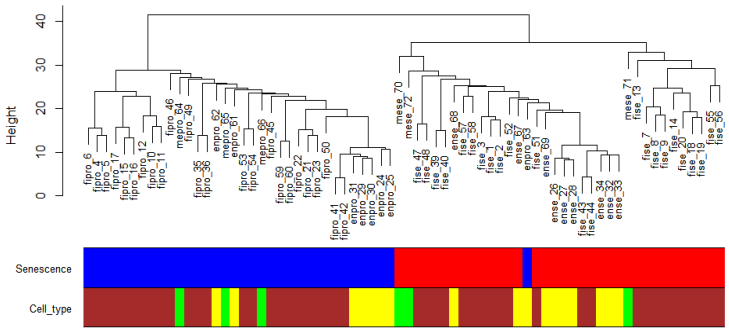
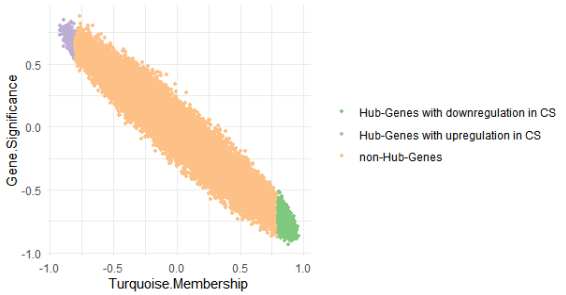


Figure 2. Identification of differentially expressed genes (DEGs) in meta-analysis. A. A venn diagram representing DEGs identification based on two meta-analysis of combined probability tests and a principle. B. Expression profile of nine example senescence-associated genes from CellAge database. Eight out of nine genes are roles in inhibiting senescence, with only PEX19 promotes senescence.

A


Height
0
10
20
30
40
Senescence
Cell_type

B


Gene.Significance
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
Turquoise.Membership
Hub-Genes with downregulation in CS
Hub-Genes with upregulation in CS
non-Hub-Genes

C

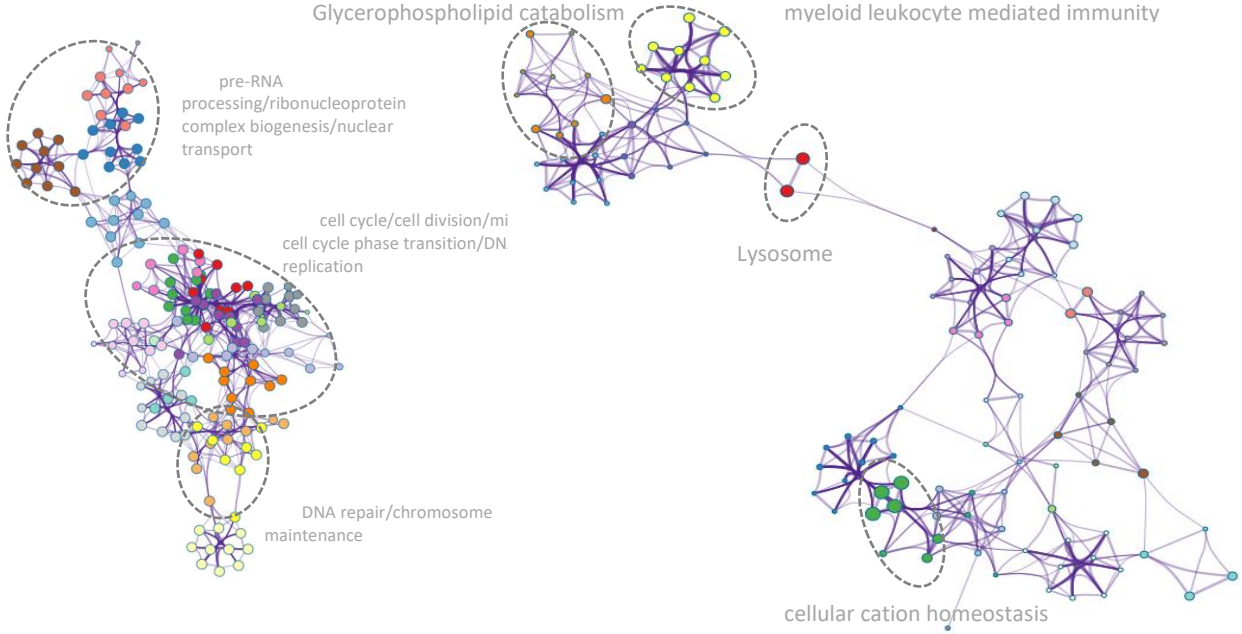

Glycerophospholipid catabolism
myeloid leukocyte mediated immunity
pre-RNA processing/ribonucleoprotein complex biogenesis/nuclear transport
cell cycle/cell division/mi
cell cycle phase transition/DN
replication
Lysosome
DNA repair/chromosome maintenance
cellular cation homeostasis

Figure 3. Senescence-associated module in weighted gene co-expression network analysis. A. Sample clustering was conducted based on genes in turquoise module. Senescence annotation row with blue and red colors representing proliferating and senescent cells respectively. For cell type, brown, yellow and green represent fibroblasts, endothelial and mesenchymal cells respectively. B. Hub genes detection in turquoise module with thresholds: absoulte membership > 0.8 and absolute gene significance > 0.2. C. Biological process in gene ontology enrichment analysis of down-regulated (left) and up-regulated (right) hub genes.

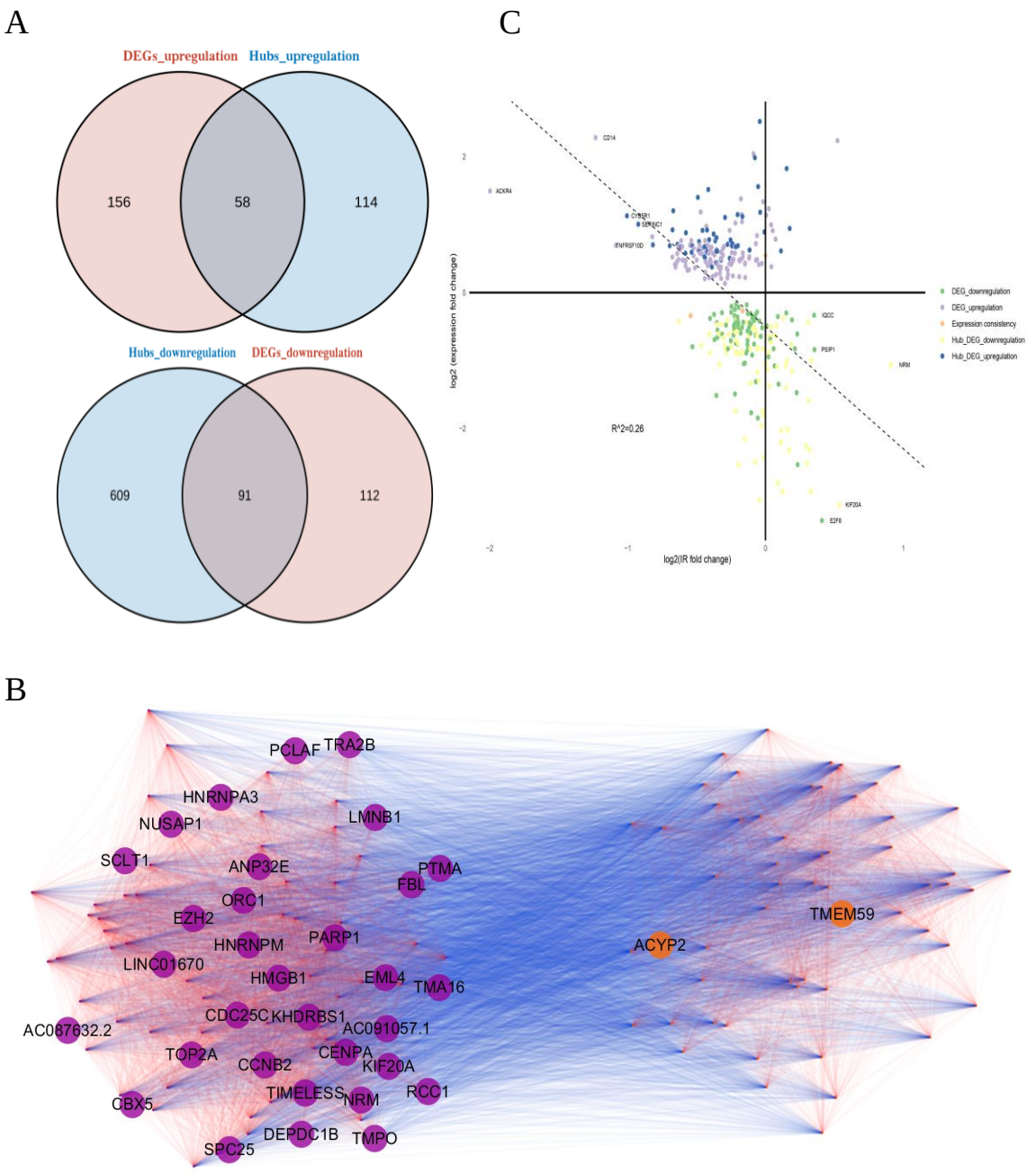


Figure 4. Identification of hub DEGs. A. Venn diagrams representing the up-regulated hub DEGs (upper) and down-regulated hub DEGs (lower). B. Representative hub DEGs co-expression networks. Genes with noted names showed absolute membership > 0.9. Purple and orange circles represent down-/up-regulated genes respectively. C. The correlation between intron retention (IR) fold change and corresponding expression fold change within DEGs.

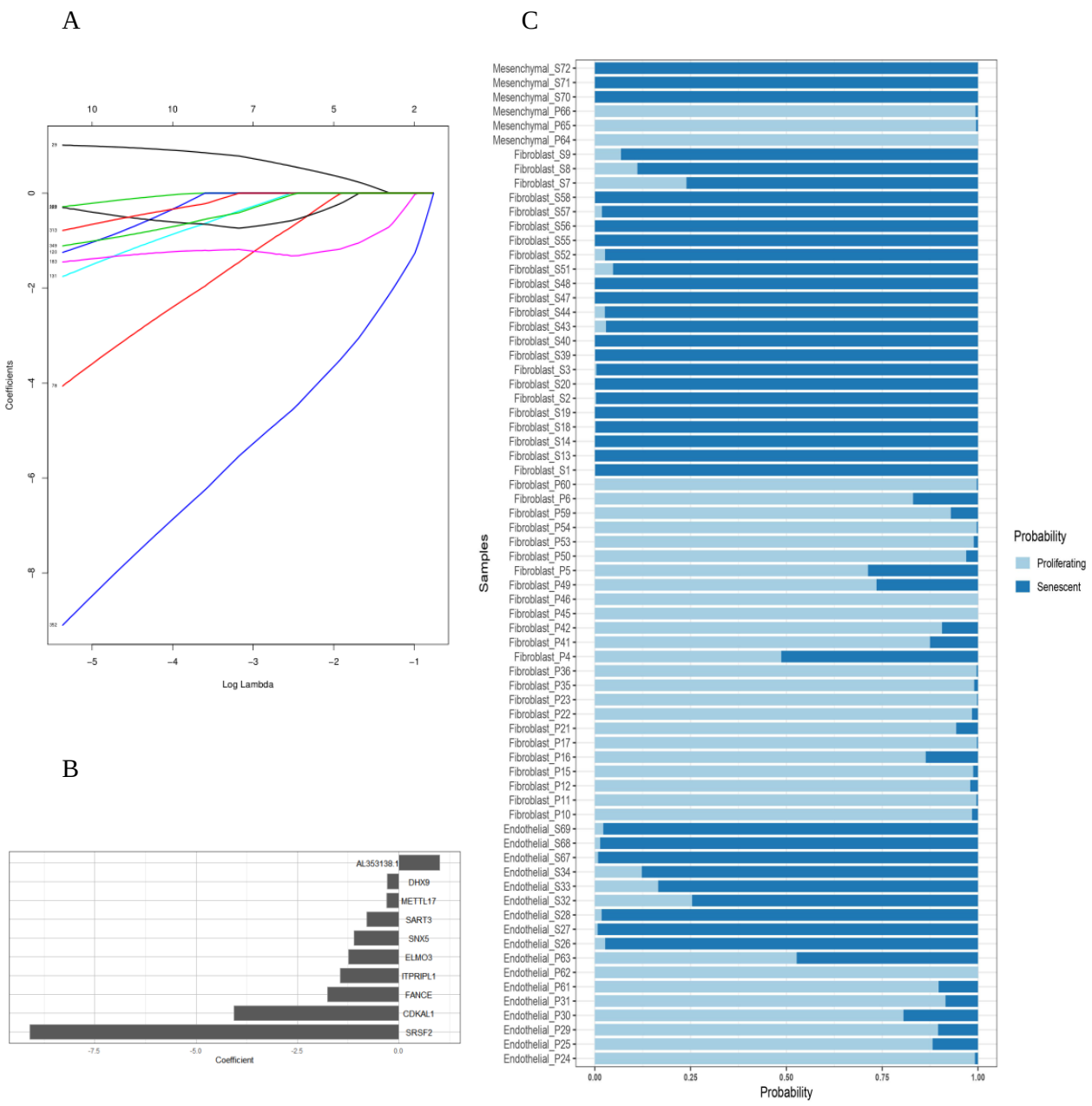
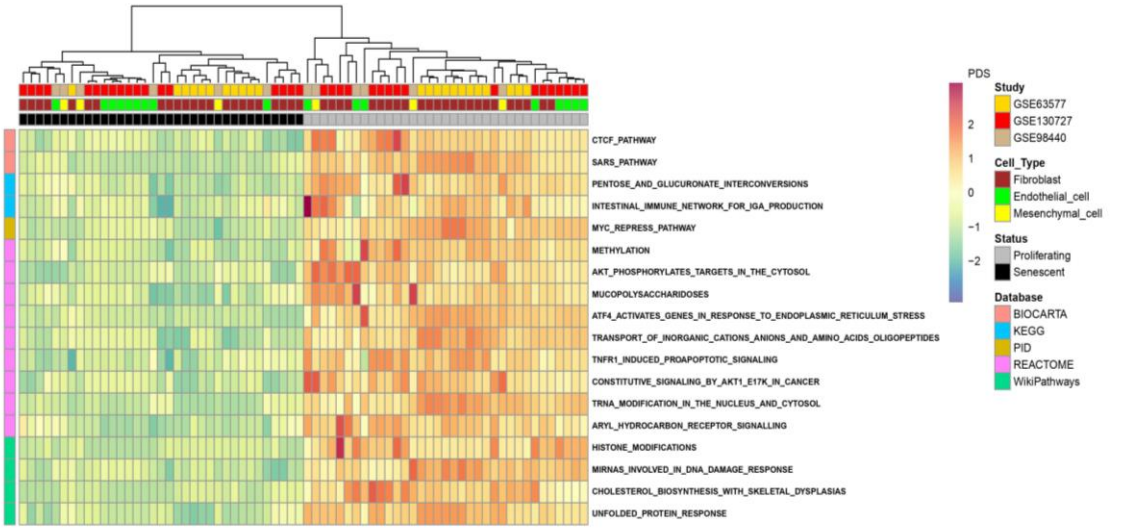


Figure 5. Dimensionality reduction through penalized logistic regression model. A. Fitted plot of variable coefficients against log of penalty strength. B. Bar graph showing non-zero coefficients of 10 senescence-associated genes. C. Probability distribution of individual samples via leave-one-out cross validation (LOOCV). Samples with P and S as a suffix indicate true proliferating and senescent phenotypes respectively in corresponding studies.

A



B

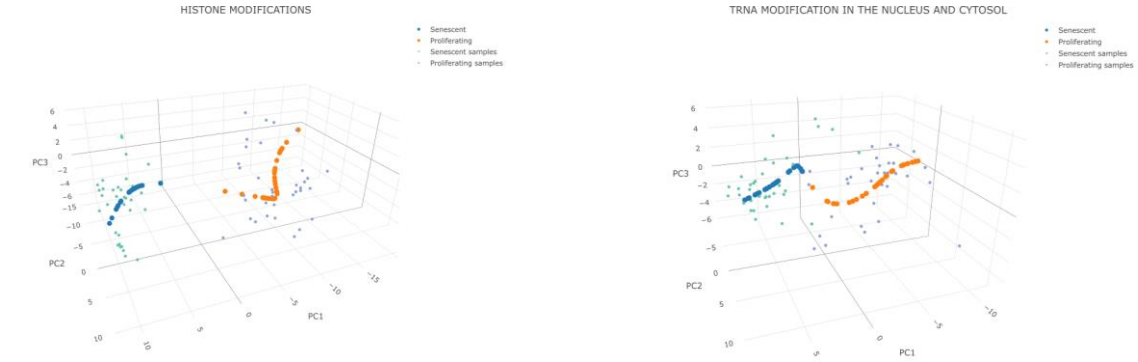


Figure 6. Senescence-associated pathway enrichment through pathway deregulation score (PDS) and logistic regression model. A. 18 pathways with non-zero coefficients identified via PDS calculated by Pathifier algorithm. B. Examples of principal curves projected by the two most contributable pathways.

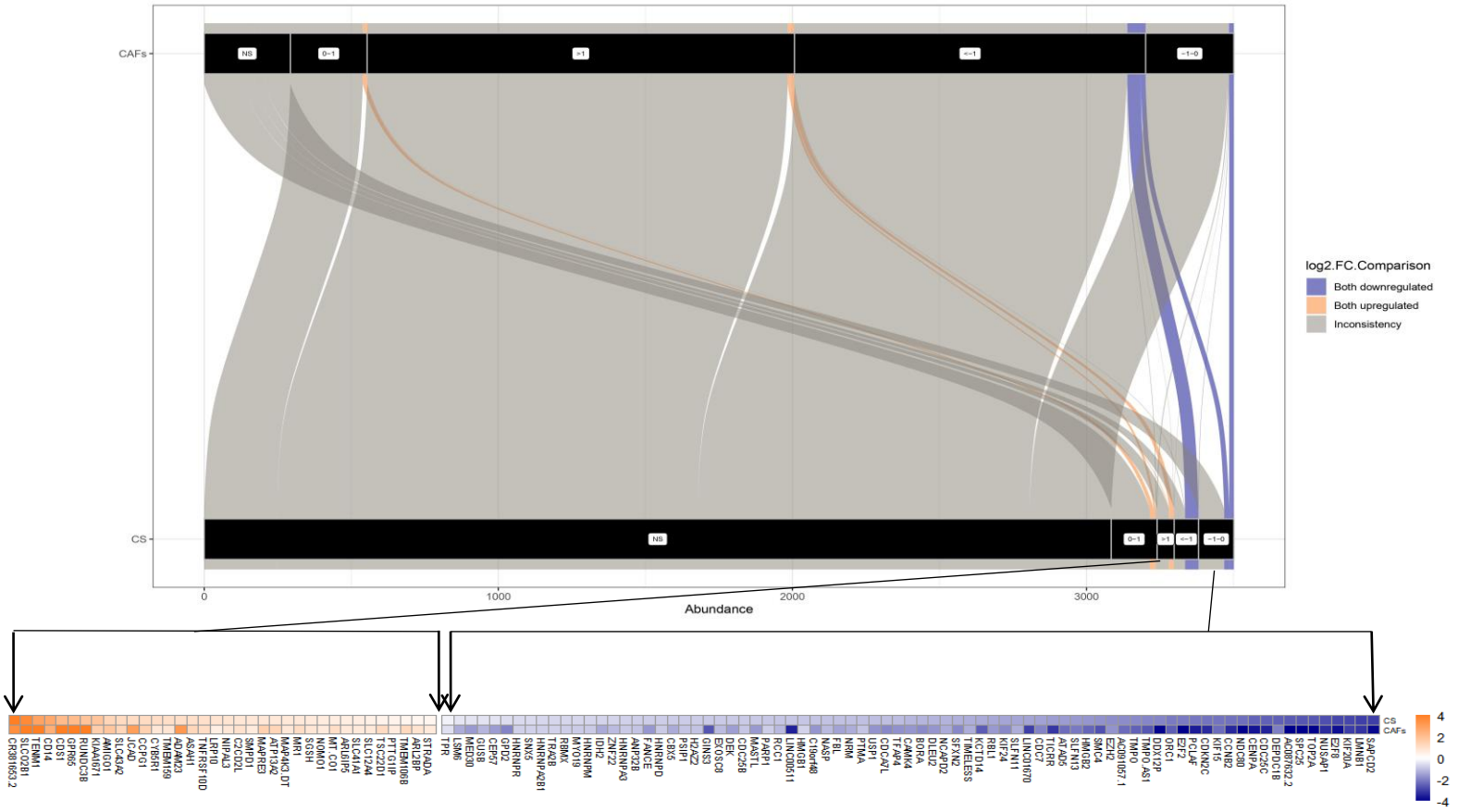


Figure 7. Comparison analysis between senescent cells (lower block) and cancer-associated fibroblasts (CAFs, upper block). Chords with orange color represent DEGs that up-regulated in both senescent cells and CAFs (with log2 (fold change) > 1 or 0~1), and chords with blue color represent DEGs that down-regulated in both senescent cells and CAFs (with log2 (fold change) < -1 or -1~0).

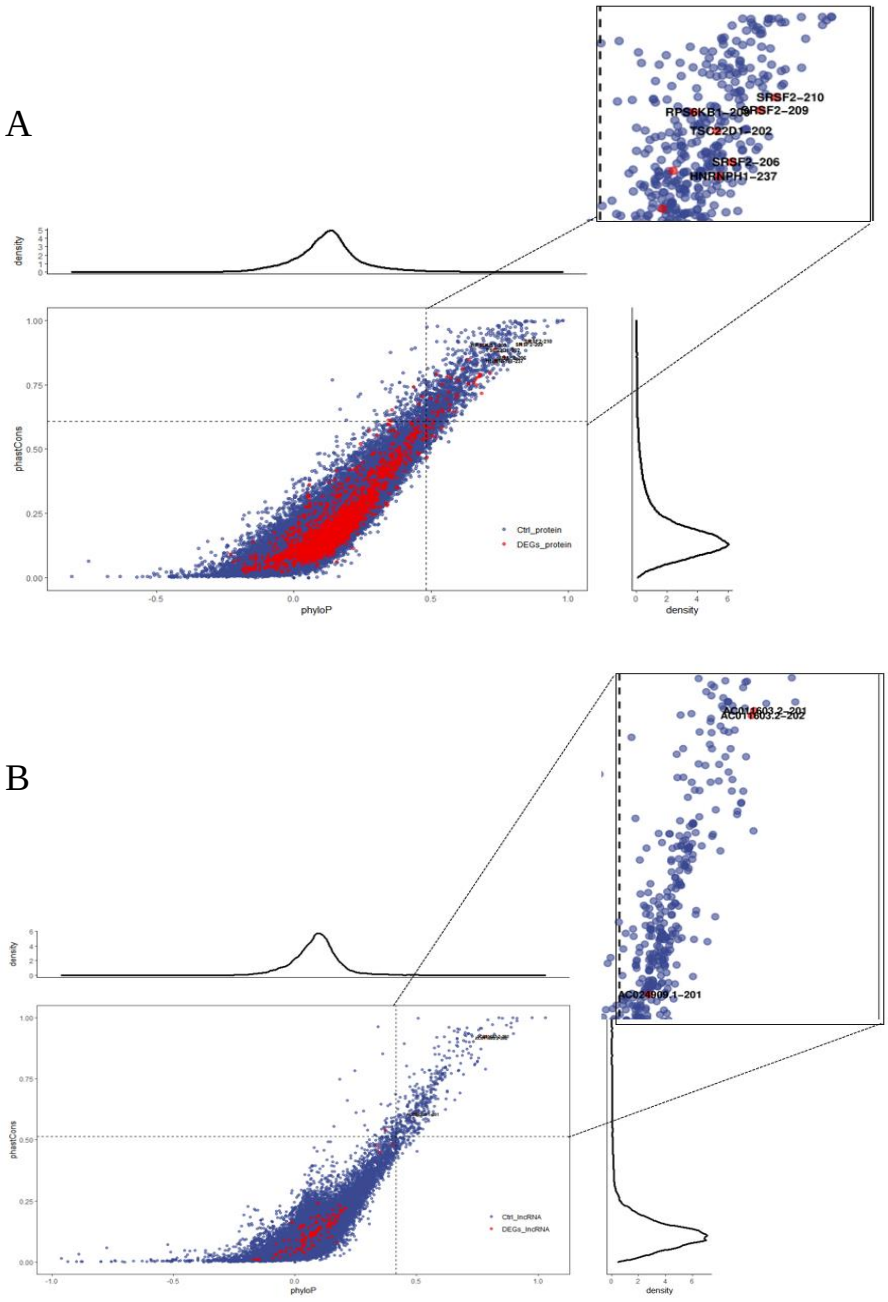


Figure S1. Transcriptomic conservation analysis of protein-coding (A) and long non-coding RNAs (B). Genes colored are differentially expressed protein coding/lncRNA genes and those conserved beyond 99th percentile of the whole inputs are labeled.

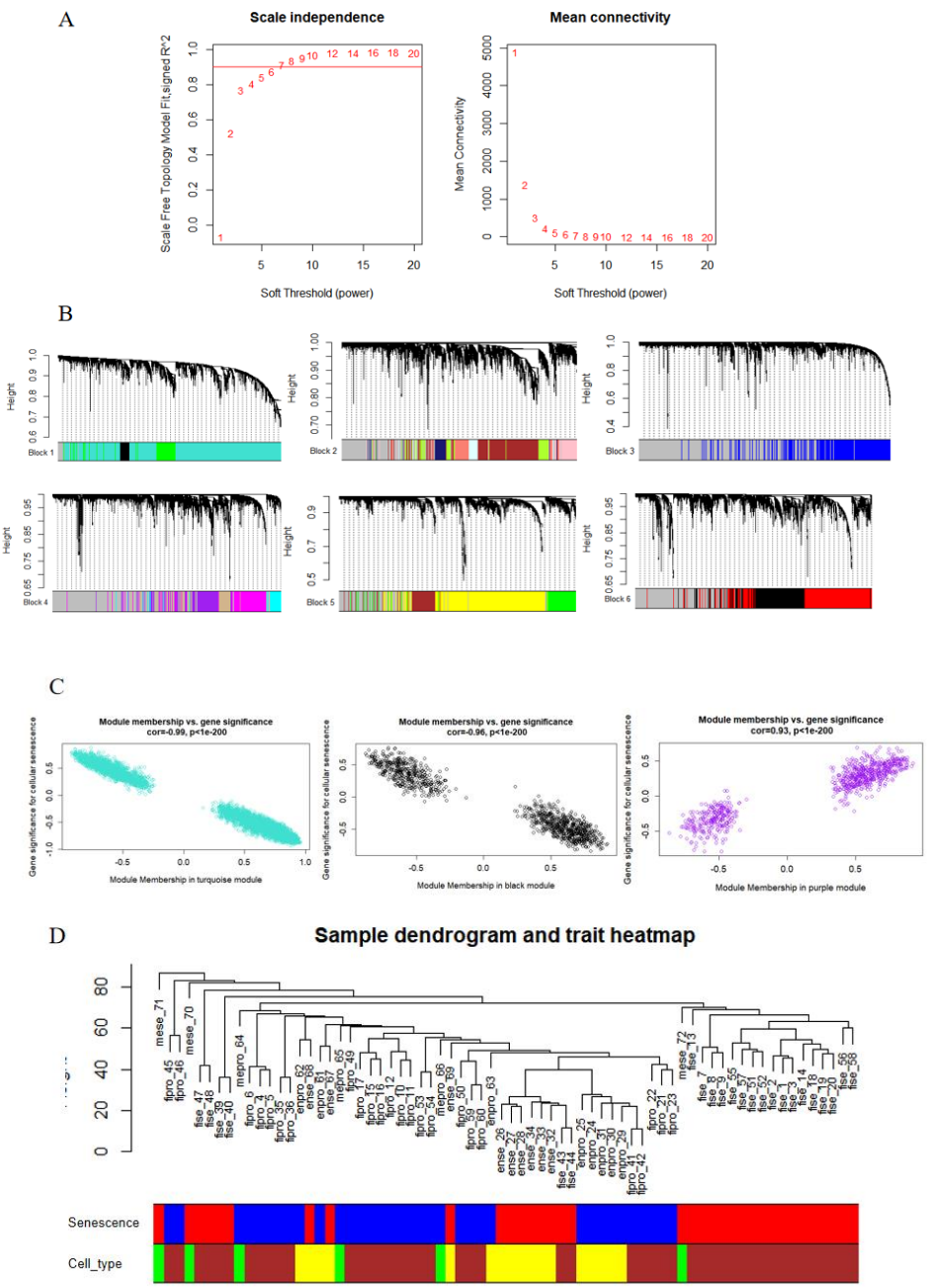


Figure S2. Weighted gene co-expression network analysis with collected samples. A. Analysis of the scale-free fit index for various soft-thresholding powers (β). B. Dendrogram of total genes clustered based on a dissimilarity measure (1-TOM). C. Examples of correlation between gene significance and module membership of genes within the three most senescence-associated modules. D. Sample clustering was conducted based on total genes input. Senescence annotation row with blue and red colors represent proliferating and senescent cells respectively. For cell type, brown, yellow and green represent fibroblasts, endothelial and mesenchymal cells respectively.

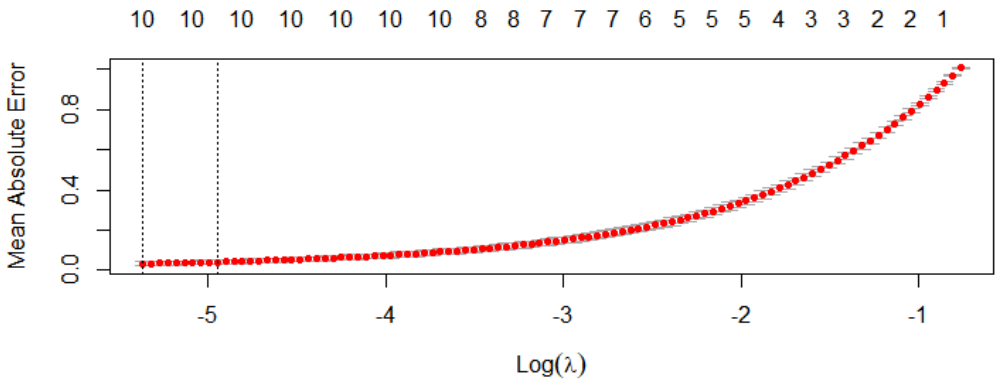


Figure S3. Mean absolute error against log of penalty strength to indentify minimum number of genes (upper x-axis) needed to discriminate proliferating and senescent cells in simplicity and accuracy.

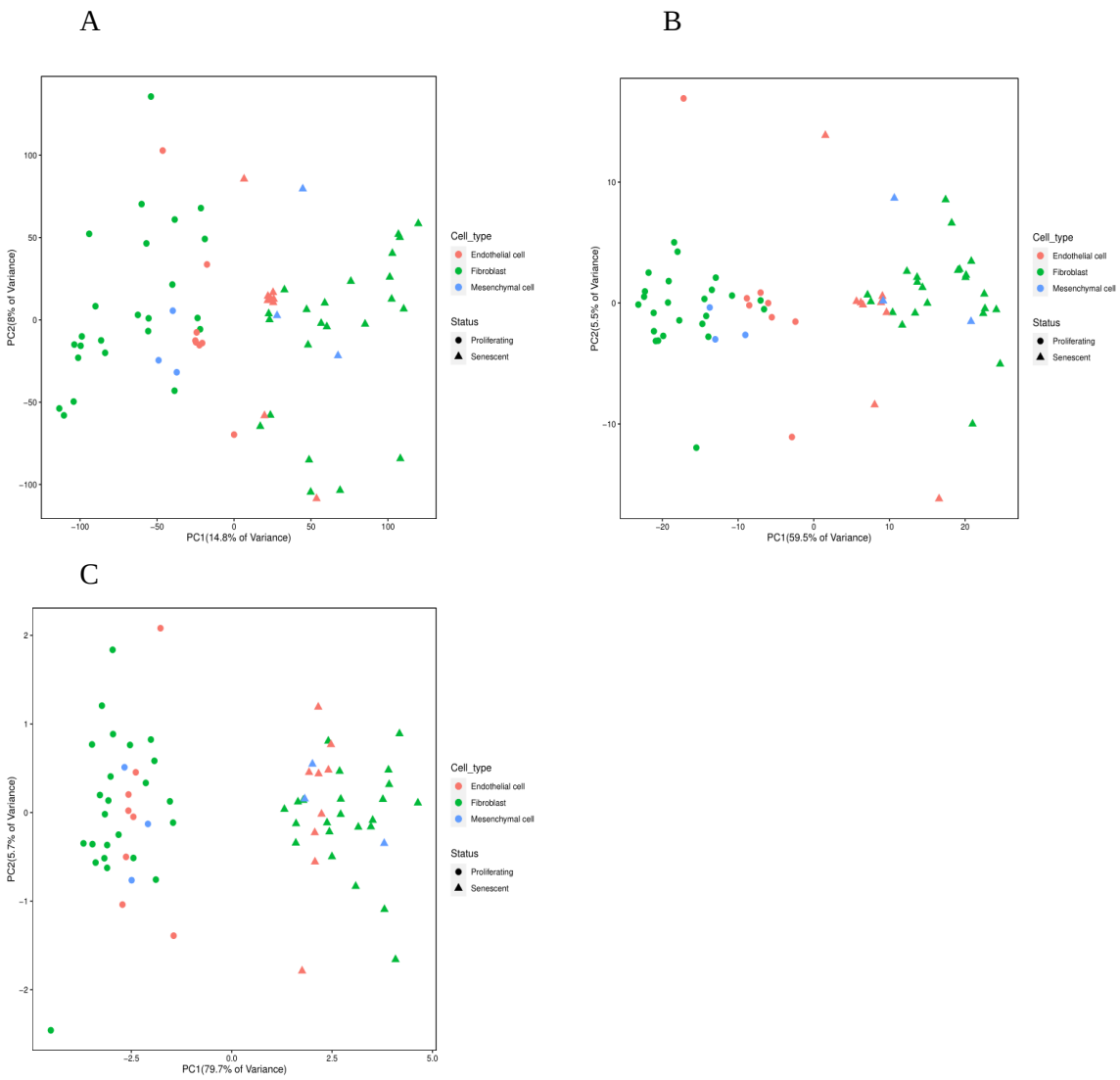


Figure S4. Principal component analysis (PCA) calculated using mormalized TPM values of all genes (A), 417 DEGs (B) and deduced 10 genes (C) from penalized regression model.

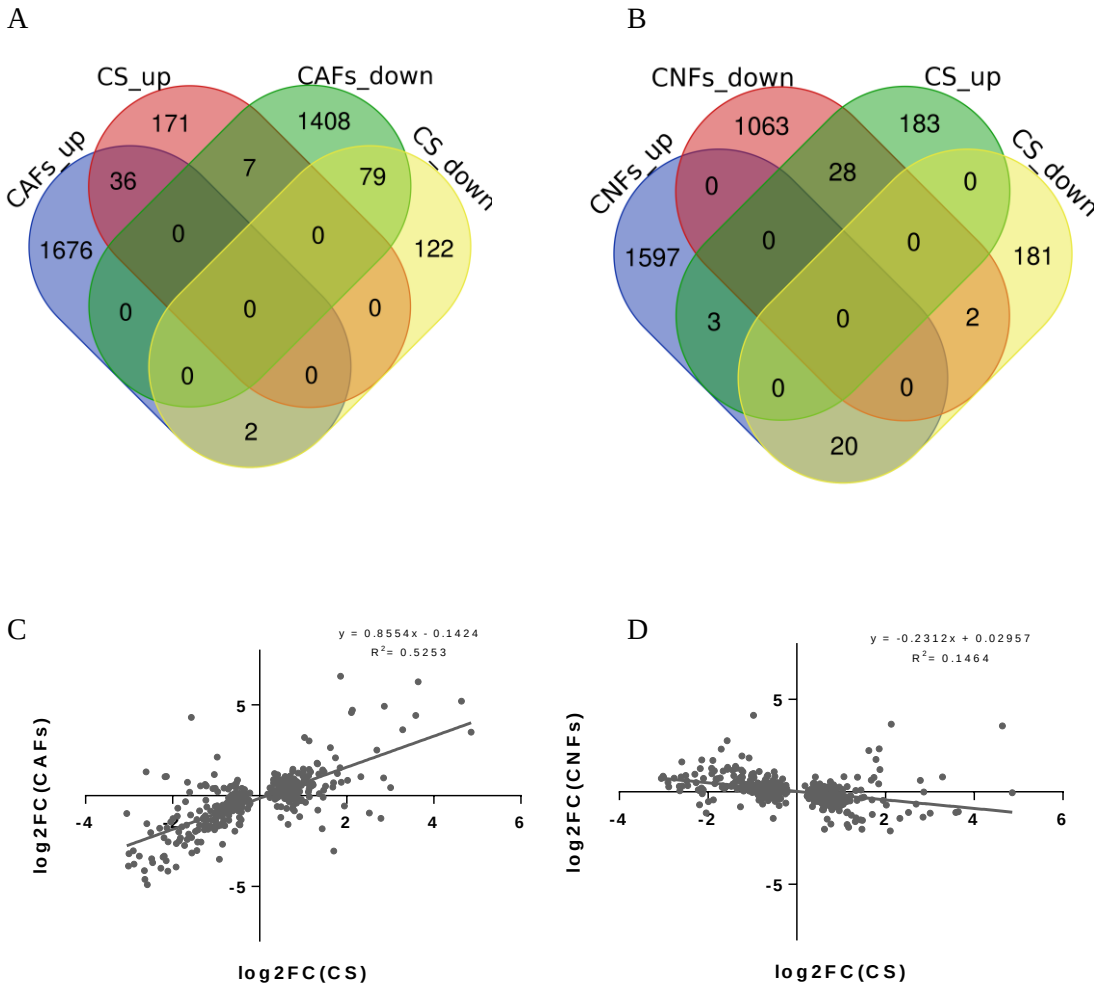


Figure S5. Consensus transcriptomic features between senescent cells and cancer-associated fibroblasts (CAFs). A. Venn diagrams representing DEGs between senescent cells and CAFs (A), and between senescent cells and colorectal cancer cells co-cultured normal fibroblasts (CNFs) (B). Expression correlation between cellular senescence (CS) and CAFs (C) and between CS and CNFs (D) by employing 417 DEGs.

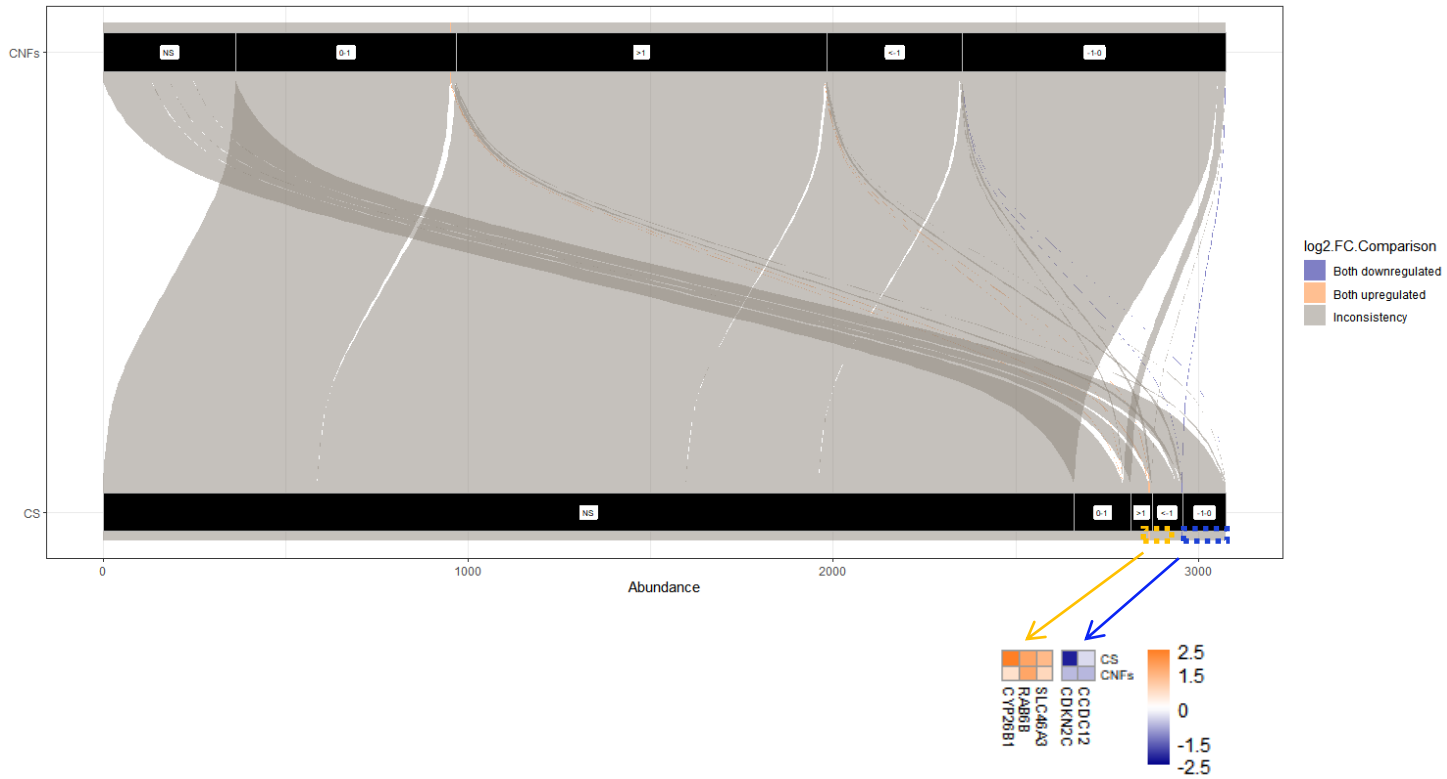


Figure S6. Comparison analysis between senescent cells and colorectal cancer cells co-cultured normal fibroblasts (CNFs). Chords with orange color represent DEGs that up-regulated in both senescent cells and CNFs (with log2 (fold change) > 1 or 0~1), and chords with blue color represent DEGs that down-regulated in both senescent cells and CNFs (with log2 (fold change) < -1 or -1~0).

Table 1. Summary of individual studies for meta-analysis.

Study	Sample Size N	Title	Platform	Cell type(s)	Senescence Factor	Senescence evaluation
GSE63577	23(45)	RNA-seq of human fibroblasts during replicative senescence	Illumina HiSeq 2000	Fibroblasts (IMR-90, WI-38, HFF, MRC-5)	RS	Cell arrest, SA-β-gal activity
GSE130727	35(37)	Transcriptome Signature of Cellular Senescence	Illumina HiSeq 4000, Illumina HiSeq 2500	Endothelial cells (HAECs and HUVECs), Fibroblasts (IMR-90 and WI-38), Mesenchymal stromal cells	IR, RS, DOX, OIS,	SA-β-gal activity, Senescence markers
GSE98440	12(26)	Topological demarcation by HMGB2 is disrupted early upon senescence entry and induces CTCF clustering across cell types	Illumina HiSeq 4000	Endothelial cells (HUVECs), Mesenchymal stromal cells	RS	Cell arrest, SA-β-gal activity
GSE155343	10(25)	Co-cultivation of colorectal cancer cells and human skin fibroblasts in 3D collagen gel and comparison of co-cultivated fibroblasts with CAFs	Illumina NovaSeq 6000	Skin fibroblasts, cancer-associated fibroblasts	–	–

RS, replicative senescence; IR, ionizing radiation-induced senescence; DOX, Doxorubicin-induced senescence; OIS, oncogene-induced senescence.

Table 2. Model performance evaluation by LOOCV.

	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10	Exp11	Exp12	Exp13	Exp14	Overall
Binomial Deviance	2.34E-02	5.20E-03	6.54E-03	5.52E-03	1.66E-02	3.72E-02	1.23E-02	3.46E-02	1.73E-02	6.83E-02	9.41E-04	3.41E-02	5.04E-03	2.11E-02	2.02E-02
ME	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AUROC	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
MSE	3.34E-04	1.57E-05	2.60E-05	1.93E-05	1.72E-04	8.68E-04	7.61E-05	6.48E-04	1.66E-04	2.46E-03	8.34E-07	6.99E-04	2.68E-05	5.57E-04	4.07E-04
MAE	2.32E-02	5.19E-03	6.53E-03	5.51E-03	1.65E-02	3.68E-02	1.22E-02	3.43E-02	1.72E-02	6.70E-02	9.41E-04	3.37E-02	5.03E-03	2.08E-02	2.00E-02

ME, misclassification error; AUROC, area under the receiver operating curve; MSE,mean squared error; MAE, mean absolute error.

Table 3. Gene ontology enrichment in biological process of genes that were differentially expressed in both CS and CAFs.

GO	Description	Log(q-value)	Hits
GO:0051301	cell division	-10.66532287	CDC25B,CDC25C,CENPA,RCC1,RBL1,TOP2A,TPR,CDC7,TIMELESS,CCNB2,NCAPD2,SMC4,KIF20A,NDC80,MAPRE3,NUSAP1,SPC25,E2F8,BORA,MYO19,MASTL,SAPCD2
GO:0044770	cell cycle phase transition	-9.177431305	CDC25B,CDC25C,CDKN2C,RCC1,EZH2,NASP,ORC1,RBL1,TFAP4,TPR,CDC7,TIMELESS,CCNB2,CEP57,NDC80,E2F8,BORA,ATAD5,MASTL,TICRR,SLFN11
GO:0044772	mitotic cell cycle phase transition	-7.857466678	CDC25B,CDC25C,CDKN2C,RCC1,EZH2,NASP,ORC1,RBL1,TFAP4,TPR,CDC7,CCNB2,CEP57,NDC80,E2F8,BORA,MASTL,TICRR,SLFN11
GO:0010564	regulation of cell cycle process	-7.636614622	CDC25B,CDC25C,CDKN2C,RCC1,EZH2,ORC1,RBL1,TFAP4,TPR,CDC7,CEP57,KIF20A,NDC80,CBX5,NUSAP1,E2F8,BORA,ATAD5,MYO19,TICRR,SLFN11
GO:0007346	regulation of mitotic cell cycle	-6.345525213	CDC25B,CDC25C,CDKN2C,RCC1,EZH2,ORC1,RBL1,TFAP4,TOP2A,TPR,CDC7,CEP57,NDC80,NUSAP1,E2F8,BORA,TICRR,SLFN11

Number	cell_type	treatment	Phenotype	Run	Study	Experimental group
1	MRC-5 fibroblasts	replicative	Senescent	SRR1660534	GSE63577	Exp 1
2	MRC-5 fibroblasts	replicative	Senescent	SRR1660535	GSE63577	
3	MRC-5 fibroblasts	replicative	Senescent	SRR1660536	GSE63577	
4	HFF fibroblasts	replicative	proliferating	SRR1660543	GSE63577	Exp 2
5	HFF fibroblasts	replicative	proliferating	SRR1660544	GSE63577	
6	HFF fibroblasts	replicative	proliferating	SRR1660545	GSE63577	
7	HFF fibroblasts	replicative	Senescent	SRR1660546	GSE63577	
8	HFF fibroblasts	replicative	Senescent	SRR1660547	GSE63577	
9	HFF fibroblasts	replicative	Senescent	SRR1660548	GSE63577	
10	IMR-90 fibroblasts	replicative	proliferating	SRR1660549	GSE63577	Exp 3
11	IMR-90 fibroblasts	replicative	proliferating	SRR1660550	GSE63577	
12	IMR-90 fibroblasts	replicative	proliferating	SRR1660551	GSE63577	
13	IMR-90 fibroblasts	replicative	Senescent	SRR1660553	GSE63577	
14	IMR-90 fibroblasts	replicative	Senescent	SRR1660554	GSE63577	
15	WI-38 fibroblasts	replicative	proliferating	SRR1660555	GSE63577	Exp 4
16	WI-38 fibroblasts	replicative	proliferating	SRR1660556	GSE63577	
17	WI-38 fibroblasts	replicative	proliferating	SRR1660557	GSE63577	
18	WI-38 fibroblasts	replicative	Senescent	SRR1660558	GSE63577	
19	WI-38 fibroblasts	replicative	Senescent	SRR1660559	GSE63577	
20	WI-38 fibroblasts	replicative	Senescent	SRR1660560	GSE63577	
21	MRC-5 fibroblasts	replicative	proliferating	SRR2751119	GSE63577	Exp 1
22	MRC-5 fibroblasts	replicative	proliferating	SRR2751120	GSE63577	
23	MRC-5 fibroblasts	replicative	proliferating	SRR2751121	GSE63577	
24	Human Aortic Endothelial Cells	ionizing radiation	Control	SRR9016146	GSE130727	Exp 7
25	Human Aortic Endothelial Cells	ionizing radiation	Control	SRR9016147	GSE130727	
26	Human Aortic Endothelial Cells	ionizing radiation	Senescent	SRR9016148	GSE130727	
27	Human Aortic Endothelial Cells	ionizing radiation	Senescent	SRR9016149	GSE130727	
28	Human Aortic Endothelial Cells	ionizing radiation	Senescent	SRR9016150	GSE130727	
29	Human Umbilical Vein Endothelial Cells	ionizing radiation	Control	SRR9016151	GSE130727	Exp 8
30	Human Umbilical Vein Endothelial	ionizing radiation	Control	SRR9016152	GSE130727	

	Cells					
31	Human Umbilical Vein Endothelial Cells	ionizing radiation	Control	SRR9016153	GSE130727	
32	Human Umbilical Vein Endothelial Cells	ionizing radiation	Senescent	SRR9016154	GSE130727	
33	Human Umbilical Vein Endothelial Cells	ionizing radiation	Senescent	SRR9016155	GSE130727	
34	Human Umbilical Vein Endothelial Cells	ionizing radiation	Senescent	SRR9016156	GSE130727	
35	IMR-90 fibroblasts	replicative	proliferating	SRR9016157	GSE130727	Exp 9
36	IMR-90 fibroblasts	replicative	proliferating	SRR9016158	GSE130727	
39	IMR-90 fibroblasts	replicative	Senescent	SRR9016161	GSE130727	
40	IMR-90 fibroblasts	replicative	Senescent	SRR9016162	GSE130727	
41	WI-38 fibroblasts	doxorubicin	Control	SRR9016163	GSE130727	Exp 10
42	WI-38 fibroblasts	doxorubicin	Control	SRR9016164	GSE130727	
43	WI-38 fibroblasts	doxorubicin	Senescent	SRR9016165	GSE130727	
44	WI-38 fibroblasts	doxorubicin	Senescent	SRR9016166	GSE130727	
45	WI-38 fibroblasts	Oncogene induced senescence (HRASG12V)	Control	SRR9016167	GSE130727	Exp 11
46	WI-38 fibroblasts	Oncogene induced senescence (HRASG12V)	Control	SRR9016168	GSE130727	
47	WI-38 fibroblasts	Oncogene induced senescence (HRASG12V)	Senescent	SRR9016169	GSE130727	
48	WI-38 fibroblasts	Oncogene induced senescence (HRASG12V)	Senescent	SRR9016170	GSE130727	
49	WI-38 fibroblasts	doxorubicin	Control	SRR9016171	GSE130727	Exp 12

50	WI-38 fibroblasts	doxorubicin	Control	SRR9016172	GSE130727	
51	WI-38 fibroblasts	doxorubicin	Senescent	SRR9016173	GSE130727	
52	WI-38 fibroblasts	doxorubicin	Senescent	SRR9016174	GSE130727	
53	WI-38 fibroblasts	ionizing radiation	Control	SRR9016175	GSE130727	Exp 13
54	WI-38 fibroblasts	ionizing radiation	Control	SRR9016176	GSE130727	
55	WI-38 fibroblasts	ionizing radiation	Senescent	SRR9016177	GSE130727	
56	WI-38 fibroblasts	ionizing radiation	Senescent	SRR9016178	GSE130727	
57	WI-38 fibroblasts	replicative	Senescent	SRR9016179	GSE130727	Exp 14
58	WI-38 fibroblasts	replicative	Senescent	SRR9016180	GSE130727	
59	WI-38 fibroblasts	replicative	Proliferating	SRR9016181	GSE130727	
60	WI-38 fibroblasts	replicative	Proliferating	SRR9016182	GSE130727	
61	human umbilical vein endothelial cells	replicative	Proliferating	SRR5494699	GSE98440	Exp 6
62	human umbilical vein endothelial cells	replicative	Proliferating	SRR5494700	GSE98440	
63	human umbilical vein endothelial cells	replicative	Proliferating	SRR5494701	GSE98440	
64	mesenchymal stromal cells	replicative	Proliferating	SRR5494706	GSE98440	Exp 5
65	mesenchymal stromal cells	replicative	Proliferating	SRR5494707	GSE98440	
66	mesenchymal stromal cells	replicative	Proliferating	SRR5494708	GSE98440	
67	human umbilical vein endothelial cells	replicative	Senescent	SRR5494710	GSE98440	Exp 6
68	human umbilical vein endothelial cells	replicative	Senescent	SRR5494711	GSE98440	
69	human umbilical vein endothelial cells	replicative	Senescent	SRR5494712	GSE98440	
70	mesenchymal stromal cells	replicative	Senescent	SRR5494717	GSE98440	Exp 5
71	mesenchymal stromal cells	replicative	Senescent	SRR5494718	GSE98440	

72	mesenchymal stromal cells	replicative	Senescent	SRR5494719	GSE98440	

Table S1. Group information of individual samples.

Table S2. Summary of differentially expressed genes.

Biotype	Upregulated in senescent cells	Downregulated in senescent cells
Protein coding	187	184
lncRNA	20	11
Pseudogene	6	7
tRNA	1	0
snoRNA	0	1

Table S3. Conservation score of mRNAs and lncRNAs from DEGs.

Transcript	phy_score	pha_score	Class
SRSF2-210	0.860632	0.921121	DEGs_protein
SRSF2-209	0.827105	0.908468	DEGs_protein
SRSF2-206	0.765233	0.857697	DEGs_protein
HNRNPH1-237	0.738281	0.8437	DEGs_protein
TSC22D1-202	0.733067	0.888036	DEGs_protein
SRSF2-211	0.71362	0.795004	DEGs_protein
HNRNPR-209	0.684226	0.718123	DEGs_protein
RPS6KB1-209	0.683355	0.9066	DEGs_protein
SRSF2-208	0.679312	0.788096	DEGs_protein
HNRNPH1-229	0.678444	0.791067	DEGs_protein
SRSF2-202	0.675244	0.783137	DEGs_protein
SRSF2-203	0.67456	0.782667	DEGs_protein
HNRNPH1-243	0.665401	0.748022	DEGs_protein
SRSF2-204	0.664155	0.773537	DEGs_protein
HNRNPH1-241	0.656736	0.760172	DEGs_protein
TTI1-205	0.63885	0.848978	DEGs_protein
SRSF2-205	0.636194	0.754587	DEGs_protein
SRSF4-204	0.616347	0.811764	DEGs_protein
AMIGO1-201	0.599703	0.767536	DEGs_protein
NDST2-204	0.594245	0.711075	DEGs_protein
HNRNPH1-239	0.593459	0.700082	DEGs_protein
HNRNPD-206	0.585873	0.644221	DEGs_protein
ITPRIPL1-203	0.574443	0.776208	DEGs_protein
HNRNPH1-227	0.568171	0.673532	DEGs_protein
HNRNPH1-236	0.566452	0.653632	DEGs_protein
HNRNPR-211	0.563183	0.722918	DEGs_protein
INSYN1-201	0.562635	0.784826	DEGs_protein
HNRNPH1-235	0.556715	0.693636	DEGs_protein
PSIP1-207	0.543557	0.645245	DEGs_protein
MCTS2P-201	0.539978	0.750568	DEGs_protein
HNRNPH1-248	0.53744	0.651717	DEGs_protein
CDKN2C-202	0.526332	0.642206	DEGs_protein
HNRNPH1-206	0.526299	0.645861	DEGs_protein
SLF2-205	0.526008	0.704398	DEGs_protein
ZNF219-208	0.524834	0.646019	DEGs_protein
ZNF219-207	0.523739	0.629575	DEGs_protein
SOX11-201	0.522041	0.685148	DEGs_protein
TSC22D1-208	0.520835	0.7916	DEGs_protein
CDKN2C-203	0.506807	0.609619	DEGs_protein
BLCAP-203	0.506098	0.646219	DEGs_protein

H2AC20-201	0.500655	0.696907	DEGs_protein
HNRNPH1-247	0.487191	0.61495	DEGs_protein
AC011603.2-201	0.785415	0.924349	DEGs_lncRNA
AC011603.2-202	0.778402	0.918934	DEGs_lncRNA
AC024909.1-201	0.495999	0.605128	DEGs_lncRNA

Table S4. Correlation between WGCNA modules and CS phenotype.

Module	Correlation	P_value	Number of genes included
MEturquoise	-0.86	4.00E-21	4824
MEblack	-0.73	7.00E-13	1042
MEgrey	-0.57	2.00E-07	8576
MEpurple	0.56	5.00E-07	792
MEtan	-0.53	2.00E-06	558
MEyellow	0.51	8.00E-06	1916
MEblue	0.48	3.00E-05	2267
MEgreenyellow	-0.46	6.00E-05	682
MEgreen	-0.43	2.00E-04	1352
MEbrown	-0.4	6.00E-04	2200
MEpink	-0.39	9.00E-04	877
MElightcyan	0.34	0.004	261
MEsalmon	0.28	0.02	480
MEmidnightblue	-0.26	0.03	347
MEmagenta	0.2	0.09	856
MEcyan	-0.19	0.1	448
MEred	-0.17	0.2	1137

Table S5. Model performance showed in confusion matrix.

		TRUE		
		Proliferating	Senescent	Total
Predicted	Proliferating	35	0	35
	Senescent	0	35	35
	Total	35	35	70
			Percent Correct:	1

Table S6. Senescence-associated pathway analysis by Pathifier algorithm. There are18 senescence-associated pathways with non-zero coefficient values.

Pathway	Coefficient	Data_source	Number of genes included
HISTONE_MODIFICATIONS	-4.835926178	WP	70
TRNA_MODIFICATION_IN_THE_NUCLEUS_AND_CYTOSOL	-3.250742649	REACTOME	43
AKT_PHOSPHORYLATES_TARGETS_IN_THE_CYTOSOL	-1.827632756	REACTOME	14
ATF4_ACTIVATES_GENES_IN_RESPONSE_TO_ENDOPLASMIC_RETICULUM_STRESS	-1.63690591	REACTOME	27
MIRNAS_INVOLVED_IN_DNA_DAMAGE_RESPONSE	-1.517198898	WP	50
CTCF_PATHWAY	-1.473152695	BIOCARTA	24
SARS_PATHWAY	-1.395818678	BIOCARTA	7
ARYL_HYDROCARBON_RECEPTOR_SIGNALLING	-1.126211459	REACTOME	7
MYC_REPRESS_PATHWAY	-1.073009449	PID	63
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	-0.533293551	KEGG	28
UNFOLDED_PROTEIN_RESPONSE	-0.507655076	WP	25
CONSTITUTIVE_SIGNALING_BY_AKT1_E17K_IN_CANCER	-0.497742696	REACTOME	26
TNFR1_INDUCED_PROAPOPTOTIC_SIGNALING	-0.424577327	REACTOME	13
CHOLESTEROL_BIOSYNTHESIS_WITH_SKELETAL_DYSPLASIAS	-0.311393491	WP	7
INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	-0.245089517	KEGG	48
MUCOPOLYSACCHARIDOSES	-0.235281627	REACTOME	11
TRANSPORT_OF_INORGANIC_CATIONS_ANIONS_AND_AMINO_ACIDS_OLIGOPEPTIDES	-0.175521187	REACTOME	106
METHYLATION	-0.099150553	REACTOME	14

WP, WikiPathways database; REACTOME, Reactome Pathway database; BIOCARTA, Biocarta Pathways database; KEGG, Kyoto Encyclopedia of Genes and Genomes Pathways database; PID, Pathway Interaction database.

Table S7. Differential expression consistency among 14 experimental comparisons.

Genes	MSC-5_replicative	HFF_replicative	IMR-90_replicative	WI-38_replicative	MSC_replicative	HUVEC_replicative	HAEC_ionizing radiation	HUVEC_ionizing radiation	IMR-90_replicative	WI-38_doxorubicin	WI-38_OIS	WI-38_doxorubicin	WI-38_ionizing radiation	WI-38_replicative	Consistency Score
AL353138.1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	11
DHX9	1	1	1	1	0	0	1	1	1	1	0	1	0	1	10
METTL17	1	1	1	1	1	0	1	1	0	1	0	0	0	0	8
SART3	0	1	0	1	0	0	1	1	1	1	0	0	0	0	6
SNX5	1	1	1	1	0	0	1	1	0	1	0	0	0	1	8
ELMO3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
ITPRIPL1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
FANCE	1	1	1	1	1	0	1	1	0	1	1	0	1	0	10
CDKAL1	1	1	0	1	0	0	1	1	0	1	1	0	1	1	9
SRSF2	1	1	1	1	1	0	1	1	0	1	0	0	0	0	8

1: Considered as DEGs in this comparison; 0: Not considered as DEGs in this comparison.

CHAPTER III

IDENTIFICATION OF NOVEL SENESCENCE-ASSOCIATED LNCRNAS BY TESTING CELL CYCLE REGULATION

ABSTRACT

Long non-coding RNAs (lncRNAs) are a class of transcripts that typically have more than 200 bases in length without protein-coding potential. It has been demonstrated that lncRNAs can participate in gene regulation through interaction with DNA, RNA and proteins. Not only transiently changed in different types of cellular processes and diseases, lncRNAs also function as primary or secondary regulators for diverse pathological responses. In previous study, we identified 8 novel lncRNAs from integrated RNA-seq data that associated with cellular senescence (CS), and the underlying role of these CS-correlated lncRNAs in cell cycle related processes is unknown. Here, by employing Dicer-Substrate Short Interfering RNAs (DsiRNAs) mediated knockdowns in human primary lung fibroblasts (IMR-90), we successfully repressed their expression. Furthermore, we primarily investigated the effect of these lncRNAs on cell cycle process by measuring genes including cyclin-dependent kinases and corresponding inhibitors, cell proliferation assay and activity of senescence-associated β-galactosidase (SA-βGal). Repression of Two lncRNAs, AL353138 (Ensembl Gene ID: ENSG00000286811) and LINC01670 (Ensembl Gene ID: ENSG00000279094) results in reduced cell proliferation activity with enhanced senescent phenotype. This study gives references for discovering novel lncRNAs

regulating CS-associated cell cycle arrest and further research related to molecular mechanism should be developed.

INTRODUCTION

Cellular senescence (CS) is generally considered an irreversible process that cells are unable to proliferate after the exposure to multiple stresses, which cause genomic instability and telomere shortening (Casella et al., 2019). In general, CS has several common features including stimulated activity of p53/p21 and p16/pRB pathways (Puvvula, 2019), increased cell soma (Nassrally et al., 2019), tendency to senescence-associated heterochromatin foci (SAHF) (Zhang & Adams, 2007) and senescence-associated secretory phenotype (SASP) (Basisty et al., 2020). It has been demonstrated that CS is beneficial for tissue homeostasis, embryonic development and tumor inhibition (Munoz-Espin & Serrano, 2014). However, long-term retention of senescent cells due to the decreased ability of immune system to recognize and eliminate these cells will trigger negatively age-related effects, including several pro-inflammatory and pro-tumorigenic processes via intercellular communication (von Kobbe, 2019). Therefore, efficiently removing senescent cells has been realized as important strategies to avoid CS-associated dysfunctions.

LncRNAs are single-stranded RNAs without protein coding potential, and usually they are located at intronic, intergenic or antisense regions relative to the coding genes (Ponting, Oliver, & Reik, 2009). Compared to protein-coding genes, lncRNAs have relative low sequence conservation between species (Oh & Lee, 2020), and there are more spatio-temporally specific expression patterns that respond to certain biological process (Necsulea et al., 2014). As for modular principles, lncRNAs can participate in RNA-protein, RNA-DNA and RNA-RNA interactions via conserved regions within lncRNA genes, to bridge together distinct complexes and regulate biological process in a particular manner (Guttman & Rinn, 2012). Specifically, lncRNAs can act as “sponges” for transcriptional activators and then enhance gene expression (Lauer et al., 2020; Niu et al., 2020). Also, lncRNAs can serve as “decoy” to compete with coding RNAs (mRNAs) for binding microRNAs, therefore ensuring mRNAs stability in post-transcription regulation (Han, Li, Xiong, & Song, 2020; Lai et al., 2022).

The initiation and maintenance of CS has been studied in the past decades and corresponding regulatory mechanism in transcription level is gradually being explored in different cell types and senescence inductions. Among genes transcribed during CS, long non-coding RNAs (lncRNAs) were found to exhibit various roles in CS-associated processes via cell type and induction specific manner (Puvvula, 2019), and an increasing amount of CS-associated lncRNAs can be identified by integration of emerging high-throughput data.

Previous studies have determined that lncRNAs were involved in senescence modulation via senescent pathways (p53/p21 and pRB/p16), as well as other senescence-associated events (LaPak & Burd, 2014; Venkatraman et al., 2013; Zhao et al., 2019). For instance, the lncRNA ANRIL is located at INK4b-ARF-INK4a locus, where encodes proteins (p15INK4b, p14ARF, and p16INK4a) for cell division and tumor growth (Popov & Gil, 2010). ANRIL represses INK4a expression by recruiting polycomb repressive complexes and increasing H3K27 methylation (Aguilo, Zhou, & Walsh, 2011). As a senescence-associated lncRNA, MIR31HG can affect senescence-triggered tumorigenesis by changing SASP contents of senescent cells, thereby shifting intercellular communication in cellular micro-environment (Montes et al., 2021).

As there were growing amount of lncRNA were identified across multiple species (Li et al., 2018; Ngoc et al., 2018; Sun et al., 2018; Yang et al., 2016), investigation of more emerging lncRNAs that may play roles in CS associated cell cycle arrest becomes possible. In our unpublished study we identified 417 common differentially expressed genes (DEGs) and 10 universal senescence-related signatures across diverse senescence contexts, and they showed strong importance score in classification model for discriminating senescent cells from proliferating ones. Among DEGs there are 31 annotated lncRNAs constantly having expressional correlation with CS. Nevertheless, the actual roles of these lncRNA and whether they can directly mediate cell cycle process and senescence phenotype is still unknown. In this study, we investigate the effects of lncRNA on senescence associated gene expression and corresponding

phenotypic changes in human lung fibroblasts (IMR-90). Among them, there are two top ranked lncRNAs, LINC01670 and AL353138, their inhibition via small interfering RNA (siRNA), decrease cyclin gene expression. Meanwhile, reduced lncRNA expressions lead to slower cell proliferation activity and more senescence-associated β-galactosidase (SA-βGal) expression. This preliminary research opens a broad perspective for CS-associated lncRNAs identification.

MATERIALS AND METHODS

Selection of senescence-correlated lncRNAs

Based on the meta-analysis described in Chapter II, we identified 31 differentially expressed lncRNAs out of 417 genes. Then senescence-associated lncRNAs were determined by integrating their normalized gene abundance (Transcripts Per Million, TPM) and adjusted p-value from fisher's combined and inverse normal methods (Chang, Lin, Sibille, & Tseng, 2013), and their sum of ranks with an ascending order were used for detecting lncRNAs with relatively high expression and significantly different expression between proliferating and senescent cells.

lncRNA knock-down and validation

To investigate the effect of selected lncRNAs on cell cycle regulation, primary human lung fibroblasts (ATCC, cat# CCL-186) were cultured in high glucose DMEM (Gibco, cat# 10566016) with 10% heat inactivated FBS (Gibco, cat# 10082147) and 1% Penicillin-Streptomycin solution (Gibco, cat# 15140122). To avoid experimental bias caused by proliferative exhaustion of primary cells, all treatments were performed before population doubling level 5 (PDL5), and cells should come from the same experimental condition before knock-down treatment. After seeded in a 6-well plate for 12h, cells in each well were treated with either 10nM negative control Dicer-Substrate Short Interfering RNAs (DsiRNAs, Integrated DNA Technology) or 10nM DsiRNAs targeting the lncRNA candidate (Table S1) by Lipofectamine™ 3000 Transfection Reagent (Invitrogen, cat# L3000015) after 10 min incubation with 250 ul reduced serum medium (Gibco, cat# 31985062), and then 1750 ul 10% FBS medium was added. Another 2 ml fresh 10% FBS medium were refreshed after 24 h transfection. The efficiency of DsiRNA transfection was validated using fluorescent TYE 563 transfection control DsiRNA (Figure S1) for 24 h, and TYE 563 labeled transfection control duplex were existed in more than 80% of total cells, indicating an efficient transfection reagent. The DsiRNA targeting housekeeping gene hypoxanthine

phosphoribosyltransferase 1 (HPRT) was used as positive control. All treatments were done with triplicate samples.

RT-qPCR analyses

Total RNA from DsiRNA treated IMR-90 cells was collected using RNAzol reagent (Molecular Research Center, cat# RN 190) according to the manufacturer's instructions. cDNA synthesis was performed using SuperScript™ III Reverse Transcriptase kit (Invitrogen, cat# 18080093) with manufacturer's manual. For qPCR process, mixture including cDNA, primers of target genes (Table S2) and iTaq Universal SYBR Green Supermix (Bio-Rad, cat# 1725121) with 20μl reactions under manufacturer's instructions. By employing HPRT as internal control, the relative expression of lncRNAs and genes related to CS and cell cycle was calculated using the Ct ($2^{-\Delta\Delta Ct}$) method. All quantitative polymerase chain reaction (qPCR) analyses were performed in technical quadruplicates.

Cell proliferation assay

To assess proliferative activity of lncRNA knockdown cells, CCK-8 assay (Abcam, cat# ab228554) was performed based on the manufacturer's instruction. In principle, water-soluble tetrazolium salt from this kit can be biological reduced by live cells, and corresponding product with orange formazan dye can be measured by absorbance meter at 460 nm. Cell viability was measured at 0h, 12h, 24h, 36h, 48 60h and 72h, with eight technical replicates at each time point.

SA-β-galactosidase (SA-β-Gal) activity

To assess the extent of senescent cells, β-galactosidase activity was measured after 5 days' lncRNA knockdown treatment at pH 6 using SA-β-galactosidase staining kit (#9680, Cell Signalling Technology). Firstly growth media was removed and 2 ml 1X PBS was used to rinse attached cells in 6-well plate. 1 ml 1X fixative solution was added to fix cells for 10 -15 min at room temperature. Then cells were rinsed with 1X PBS with two times and 1 ml of the

β-Galactosidase staining solution was added to each well. After that cells were incubated at 37 ℃ overnight (about 8 hours) and were protected from light. The development of blue color staining was observed under a microscope with 200X total magnification. SA-β-Gal activity was performed in technical triplicates.

Statistical analysis

All differences between two group comparisons (treatment vs. control) were analyzed via paired student's t-test via GraphPad Prism 9 software (GraphPad Software, Inc., San Diego, CA, USA). Differences were considered significant when there is more than 95% confidence ($P \leq 0.05$).

RESULTS

Fluctuating expression of cell cycle related gene after lncRNA inhibition

To investigate the regulatory function of top ranked lncRNAs from our meta-analysis integration in Chapter II, 5 lncRNAs were chosen as they showed detectable expression in qPCR analysis (data not shown). RNA was collected after 5 days' treatment of target knockdown, and expression of cell cycle related genes including cyclins and corresponding kinase (CDK4), and hallmark of cell cycle arrest were measured. Compared to negative control transfected with scrambled DsiRNAs, treatment groups with lncRNA knockdown showed efficient target inhibition (Figure 1A-E). Treatment groups showed different expression in cyclin dependent kinase inhibitors, with a significantly decreased expression of p16 (CDKN2A) in AL353138-KD and LINC01670-KD (Figure 1F) and reduced expression of p21 (CDKN1A) in LINC01670-KD (Figure 1G). We also determined the mRNA expression of senescence-associated secretory phenotype (SASP) marker, C-X-C Motif Chemokine 1 (CXCL1), and it showed consistently repressed expression in all lncRNA-KD groups (Figure 1H).

In parallel, we also measured mRNA level of cyclin families and their kinase (CDK4). Mitosis-involved genes G2/Mitotic-Specific Cyclin-B1 (CCNB) was not changed after inhibition (Figure 1I), while there was declined expression of G1/S-Specific Cyclin-D1 (CCND) within AC025423.4 (Ensembl Gene ID: ENSG00000257181), AL353138 and LINC01670 knockdown groups (Figure 1J). Interestingly, expression of G1/S-Specific Cyclin-E1 (CCNE) showed floating changes among different treatment groups (Figure 1K), and Cyclin Dependent Kinase 4 (CDK4) was upregulated in AC025423.4, CU634019.5 (Ensembl Gene ID: ENSG00000280018), LINC01670 and LINC00511 (Ensembl Gene ID: ENSG00000227036) inhibition groups (Figure 1L).

Cell cycle arrest was shown after the knockdown of AL353138 and LINC01670

Since there were decreased expression of CCND and CCNE after the repression of AL353138 and LINC01670, we further identified the phenotypic alterations related to cell proliferation and senescence. Cell proliferation was estimated within 0-72 h post-treatment and significantly reduced cell viability was observed after 36h incubation in both AL353138 and LINC01670 knockdown groups (Figure 2A-B). Also, based on β-galactosidase activity, senescent phenotype was more obvious within lncRNA knockdown groups (Figure 3A-B and E-F), while treatment via only transfection reagent (Figure 3C and G) can cause CS to some extent compared to group without reagent (Blank, Figure 3D and H).

DISCUSSION

In current studies, the senescence-associated lncRNAs were selected through meta-analysis based p-value combination and expression consistency. lncRNAs showing stable expression in all senescence models (inducers x cell types) were included for further analysis. To investigate these lncRNAs' functional potential during CS, expression of senescence markers (p21, p16 and CXCL1) and mitotic coordinators (CCNB, CCND, CCNE and CDK4) was measured before/after lncRNA inhibition. Furthermore, the repression of two lncRNAs, AL353138 and LINC01670, slowed down the cell proliferation activity and more β-galactosidase activity was observed. In summary, even though there is no obvious change in senescence markers in transcriptome level, two CS-correlated lncRNA candidates may participate in cell cycle and CS via post-transcription or non-canonical pathways and specific mechanism needs to be further elaborated. Also, we provide reference list of potentially senescence-associated lncRNAs and their roles in cell cycle and senescence process can be investigated through more comprehensive strategies. Since the heterogeneity of CS it is not surprising that universal biomarkers identifying the senescence phenotype have been difficult to find. The discovery of senescence associated lncRNAs expands the understanding of senescence regulation through non-coding RNAs mechanisms, allowing us to pinpoint the multitude of processes driving senescence, and perhaps allowing us to remove senescent cells. The timely senescent cell removal improves human's tissue homeostasis and reduces chronic inflammation that triggers age-related diseases.

In previous study, there were a set of lncRNAs were validated in pRb/p16 tumor-suppressor and p53/p21 DNA damage response pathways (Puvvula, 2019). We didn't observe a significant mRNA expression shift of senescence markers after lncRNAs knockdown, even though reduction of cell proliferation and elevated SA-β-gal activity were detected. We hypothesized that these lncRNAs may play functional roles in post-transcription or subsequent regulations during CS. For example, senescence-associated lncRNA 7SL has been identified that it can competitively bind to

3’UTR region of p53 to reduce its translation efficiency without affecting the presence of transcripts (Abdelmohsen et al., 2014). Similarly, Ovarian Adenocarcinoma Amplified lncRNA (OVAAL) is a down-regulated lncRNA under the status of cell cycle arrest, which is replaced by RNA-bonding protein PTBP, to promote p27 translation by binding 5’UTR regions of its transcripts (Cho, Kim, Back, & Jang, 2005). Whether these lncRNA candidates are involved in CS process as post-transcription regulators needs further verification.

There are some limitations from our present results. We only used one type of primary cell line (human lung fibroblasts, IMR-90) to validate their potential function through expressional inhibition. Future research contents should aim to systematically understand lncRNA regulation in CS by including more sources of cell type in functional validation. Also, other testing methods that can track the senescence dynamics in post-trancription level (miCLIP, ELISA, WB, etc.) should be employed to specify molecular mechanism that lncRNAs participate in. Additionally, the toxicity of transfection reagent was observed during SA-β-gal activity measurement (Figure 3A and B), which may directly trigger DNA damage response (DDR) mediated cell cycle arrest and senescence (Wang, Larcher, Ma, & Veedu, 2018). In this situation, toxicity induced senescence will confound the cellular metabolisms via lncRNA inhibition, making a complexity of explaining lncRNA-related senescence process. Advanced transfection technologies should be developed to avoid DDR as much as possible especially for senescence-related cellular research.

CONCLUSION

In summary, we propose a potential function of CS-associated lncRNAs obtained from meta-analysis of multiple senescent models, and their specific mechanism in affecting cell cycle and senescence phenotype should have detailed investigation.

REFERENCES

Abdelmohsen, K., Panda, A. C., Kang, M. J., Guo, R., Kim, J., Grammatikakis, I., . . . Gorospe, M. (2014). 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res, 42*(15), 10099-10111. doi:10.1093/nar/gku686

Aguilo, F., Zhou, M. M., & Walsh, M. J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res, 71*(16), 5365-5369. doi:10.1158/0008-5472.CAN-10-4379

Basisty, N., Kale, A., Jeon, O. H., Kuehnemann, C., Payne, T., Rao, C., . . . Schilling, B. (2020). A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol, 18*(1), e3000599. doi:10.1371/journal.pbio.3000599

Casella, G., Munk, R., Kim, K. M., Piao, Y., De, S., Abdelmohsen, K., & Gorospe, M. (2019). Transcriptome signature of cellular senescence. *Nucleic Acids Res, 47*(14), 7294-7305. doi:10.1093/nar/gkz555

Chang, L. C., Lin, H. M., Sibille, E., & Tseng, G. C. (2013). Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. *BMC Bioinformatics, 14*, 368. doi:10.1186/1471-2105-14-368

Cho, S., Kim, J. H., Back, S. H., & Jang, S. K. (2005). Polypyrimidine tract-binding protein enhances the internal ribosomal entry site-dependent translation of p27Kip1 mRNA and modulates transition from G1 to S phase. *Mol Cell Biol, 25*(4), 1283-1297. doi:10.1128/MCB.25.4.1283-1297.2005

Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature, 482*(7385), 339-346. doi:10.1038/nature10887

Han, Q., Li, J., Xiong, J., & Song, Z. (2020). Long noncoding RNA LINC00514 accelerates

pancreatic cancer progression by acting as a ceRNA of miR-28-5p to upregulate Rap1b expression. *J Exp Clin Cancer Res, 39*(1), 151. doi:10.1186/s13046-020-01660-5

Lai, L., Wang, Z., Ge, Y., Qiu, W., Wu, B., Fang, F., . . . Chen, Z. (2022). Comprehensive analysisof the long noncoding RNA-associated competitive endogenous RNA network in the osteogenic differentiation of periodontal ligament stem cells. *BMC Genomics, 23*(1), 1. doi:10.1186/s12864-021-08243-4

LaPak, K. M., & Burd, C. E. (2014). The molecular balancing act of p16(INK4a) in cancer and aging. *Mol Cancer Res, 12*(2), 167-183. doi:10.1158/1541-7786.MCR-13-0350

Lauer, V., Grampp, S., Platt, J., Lafleur, V., Lombardi, O., Choudhry, H., . . . Schodel, J. (2020). Hypoxia drives glucose transporter 3 expression through hypoxia-inducible transcription factor (HIF)-mediated induction of the long noncoding RNA NICI. *J Biol Chem, 295*(13), 4065-4078. doi:10.1074/jbc.RA119.009827

Li, X., Xing, X., Xu, S., Zhang, M., Wang, Y., Wu, H., . . . Yang, T. (2018). Genome-wide identification and functional prediction of tobacco lncRNAs responsive to root-knot nematode stress. *PLoS One, 13*(11), e0204506. doi:10.1371/journal.pone.0204506

Montes, M., Lubas, M., Arendrup, F. S., Mentz, B., Rohatgi, N., Tumas, S., . . . Lund, A. H. (2021). The long non-coding RNA MIR31HG regulates the senescence associated secretory phenotype. *Nat Commun, 12*(1), 2459. doi:10.1038/s41467-021-22746-4

Munoz-Espin, D., & Serrano, M. (2014). Cellular senescence: from physiology to pathology. *Nat Rev Mol Cell Biol, 15*(7), 482-496. doi:10.1038/nrm3823

Nassrally, M. S., Lau, A., Wise, K., John, N., Kotecha, S., Lee, K. L., & Brooks, R. F. (2019). Cell cycle arrest in replicative senescence is not an immediate consequence of telomere dysfunction. *Mech Ageing Dev, 179*, 11-22. doi:10.1016/j.mad.2019.01.009

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., . . . Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature, 505*(7485), 635-640. doi:10.1038/nature12943

Ngoc, P. C. T., Tan, S. H., Tan, T. K., Chan, M. M., Li, Z., Yeoh, A. E. J., . . . Sanda, T. (2018). Identification of novel lncRNAs regulated by the TAL1 complex in T-cell acute lymphoblastic leukemia. *Leukemia, 32*(10), 2138-2151. doi:10.1038/s41375-018-0110-4

Niu, Y., Bao, L., Chen, Y., Wang, C., Luo, M., Zhang, B., . . . Luo, W. (2020). HIF2-Induced Long Noncoding RNA RAB11B-AS1 Promotes Hypoxia-Mediated Angiogenesis and Breast Cancer Metastasis. *Cancer Res, 80*(5), 964-975. doi:10.1158/0008-5472.CAN-19-1532

Oh, H. J., & Lee, J. T. (2020). Long Noncoding RNA Functionality Beyond Sequence: The Jpx Model: Commentary on "Functional Conservation of lncRNA JPX Despite Sequence and Structural Divergence" by Karner et al. (2019). *J Mol Biol, 432*(2), 301-304. doi:10.1016/j.jmb.2019.11.011

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell, 136*(4), 629-641. doi:10.1016/j.cell.2009.02.006

Popov, N., & Gil, J. (2010). Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics, 5*(8), 685-690. doi:10.4161/epi.5.8.12996

Puvvula, P. K. (2019). LncRNAs Regulatory Networks in Cellular Senescence. *Int J Mol Sci, 20*(11). doi:10.3390/ijms20112615

Sun, S. F., Tang, P. M. K., Feng, M., Xiao, J., Huang, X. R., Li, P., . . . Lan, H. Y. (2018). Novel lncRNA Erbb4-IR Promotes Diabetic Kidney Injury in db/db Mice by Targeting miR-29b. *Diabetes, 67*(4), 731-744. doi:10.2337/db17-0816

Venkatraman, A., He, X. C., Thorvaldsen, J. L., Sugimura, R., Perry, J. M., Tao, F., . . . Li, L. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. *Nature, 500*(7462), 345-349. doi:10.1038/nature12303

Von Kobbe, C. (2019). Targeting senescent cells: approaches, opportunities, challenges. *Aging (Albany NY), 11*(24), 12844-12861. doi:10.18632/aging.102557

Wang, T., Larcher, L. M., Ma, L., & Veedu, R. N. (2018). Systematic Screening of Commonly Used Commercial Transfection Reagents towards Efficient Transfection of

Single-Stranded Oligonucleotides. Molecules, 23(10). doi:10.3390/molecules23102564

Yang, D., Lian, T., Tu, J., Gaur, U., Mao, X., Fan, X., . . . Yang, M. (2016). LncRNA mediated regulation of aging pathways in Drosophila melanogaster during dietary restriction. *Aging (Albany NY), 8*(9), 2182-2203. doi:10.18632/aging.101062

Zhang, R., & Adams, P. D. (2007). Heterochromatin and its relationship to cell senescence and cancer therapy. *Cell Cycle, 6*(7), 784-789. doi:10.4161/cc.6.7.4079

Zhao, L., Hu, K., Cao, J., Wang, P., Li, J., Zeng, K., . . . Han, L. (2019). lncRNA miat functions as a ceRNA to upregulate sirt1 by sponging miR-22-3p in HCC cellular senescence. *Aging (Albany NY), 11*(17), 7098-7122. doi:10.18632/aging.102240

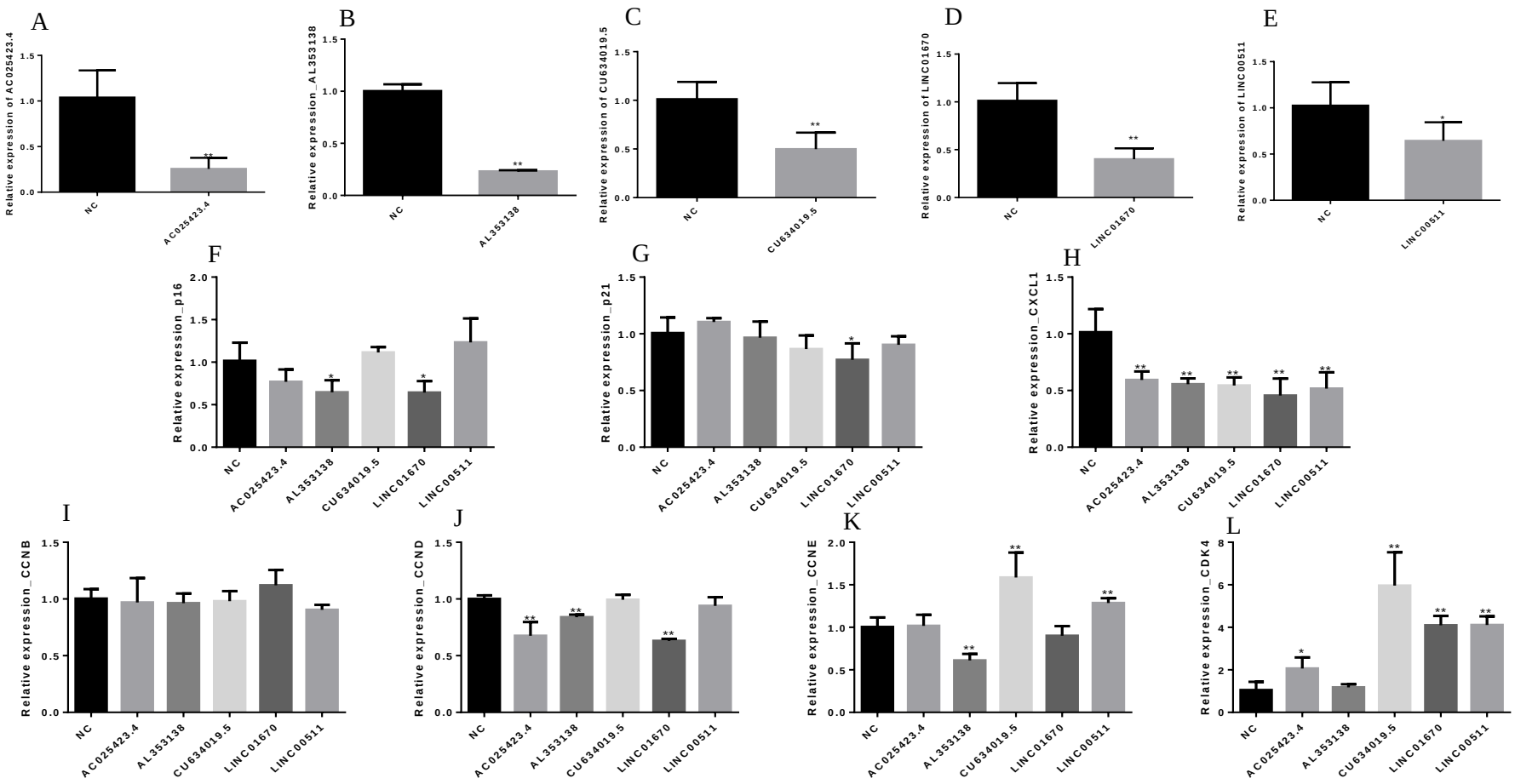


Figure 1. qPCR analysis of lncRNA targets (A-E), senescence markers (F-H), cyclin families (I-K) and their kinase (CDK4, L) after 5 days' repression of lncRNA candidates. Significant comparisons are indicated by *P $\leqslant$ 0.05; **P $\leqslant$ 0.01.

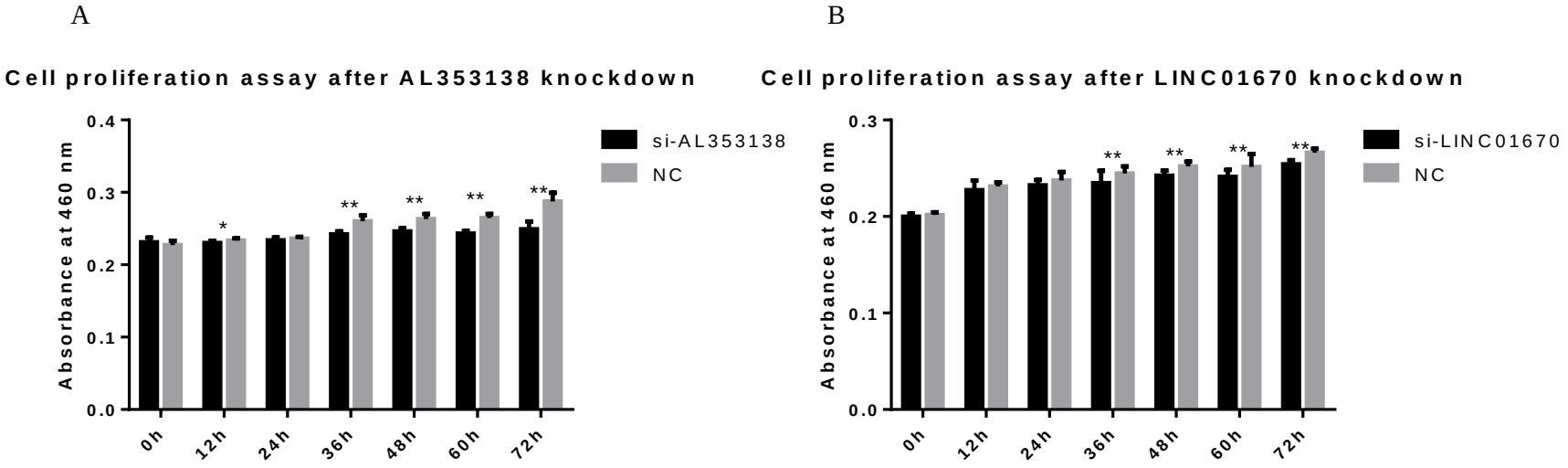


Figure 2. Identification of cell proliferation activity after AL353138 (A) and LINC01670 (B) knockdown. Significant comparisons are indicated by *P $\leqslant$ 0.05; **P $\leqslant$ 0.01.

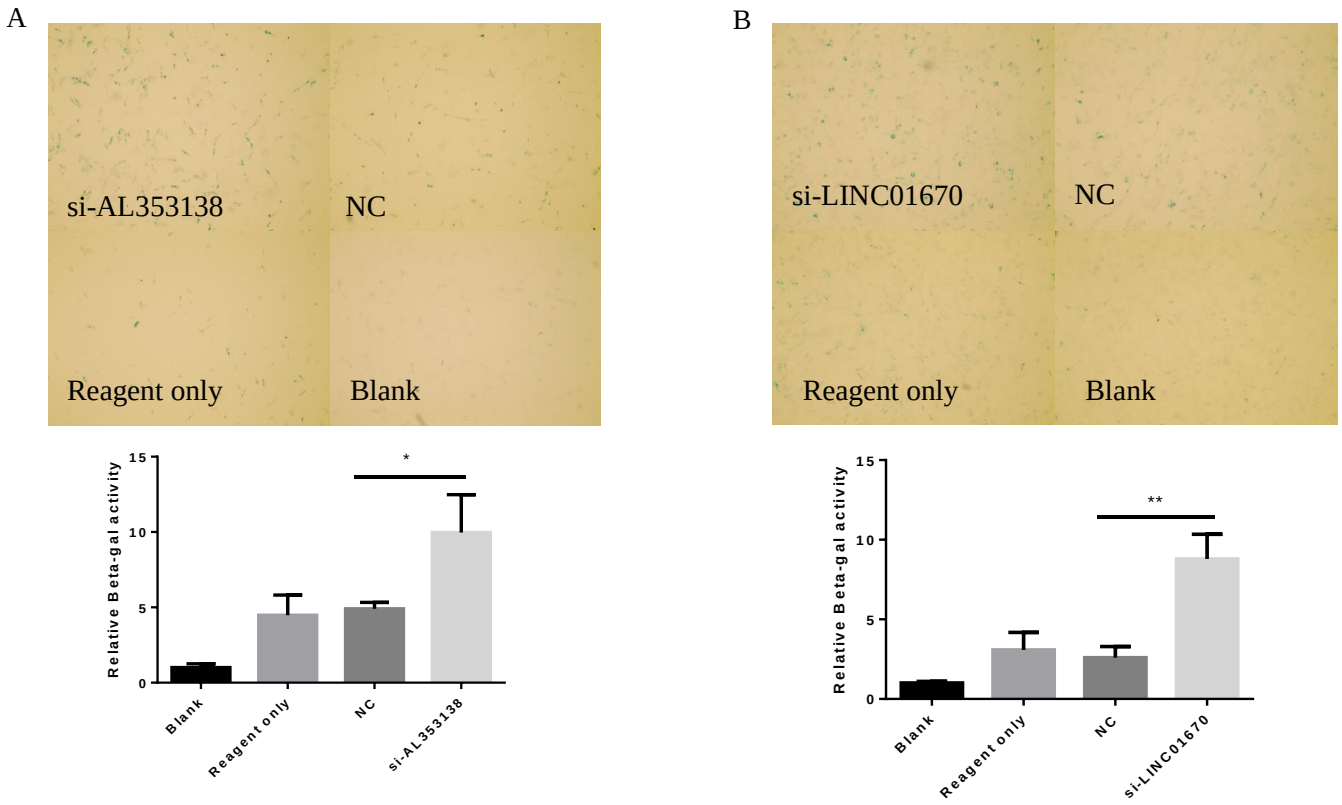


Figure 3. The senescent phenotype of AL353138 (A) and LINC01670 (B) knockdown fibroblasts assessed by monitoring SA-βGAL activity (micrographs, upper) and quantification of the% of β-galactosidase-positive cells compared to NC (lower). Significant comparisons are indicated by *P $\leqslant$ 0.05; **P $\leqslant$ 0.01.

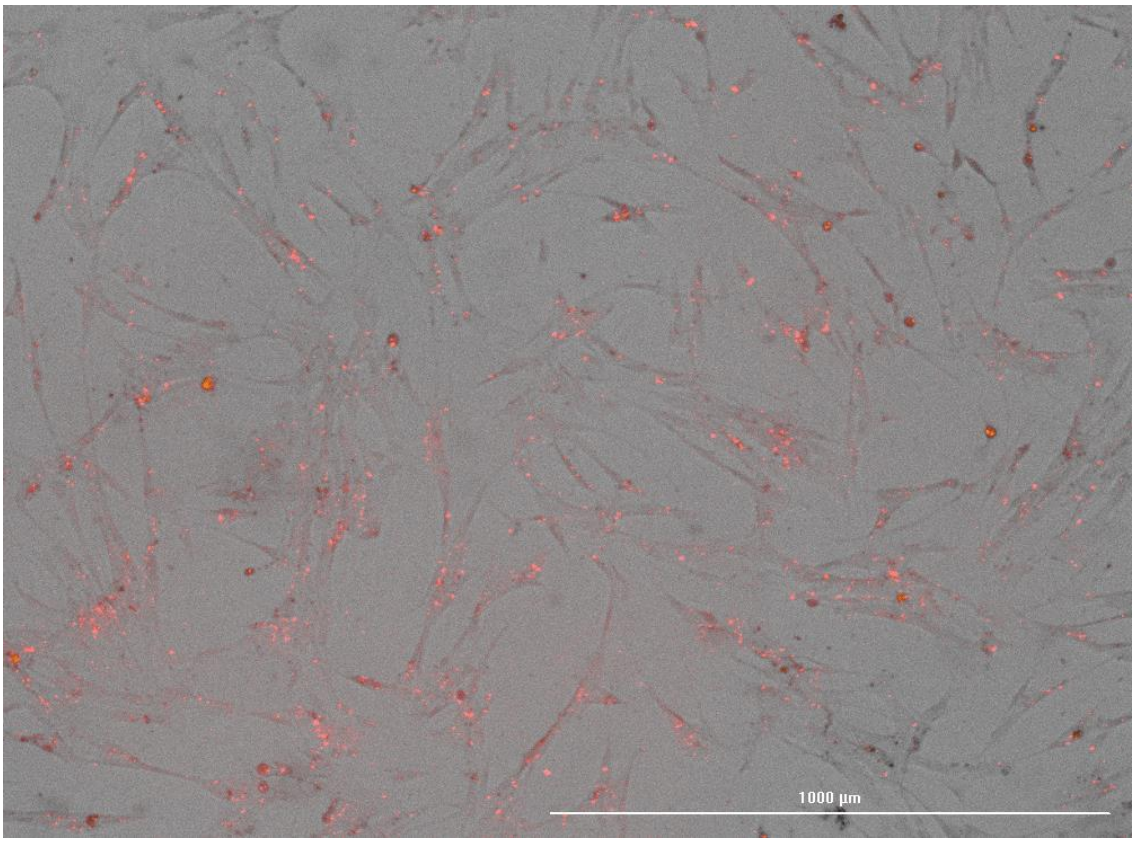


Figure S1. Evaluation of DsiRNA transfection efficiency by fluorescent TYE 563 transfection control DsiRNA. TYE 563 labeled transfection control duplex was existed in more than 80% of total cells.

Table S1. Information of Dicer-Substrate Short Interfering RNAs of 5 lncRNA candidates.

Gene	ID	Sequence
CU634019.5	ENSG00000280018-1-SEQ1	rGrUrA rCrGrU rGrCrC rCrUrG rUrUrU rGrCrU rGrArA rUrCT A
	ENSG00000280018-1-SEQ2	rUrArG rArUrU rCrArG rCrArA rArCrA rGrGrG rCrArC rGrUrA rCrArA
	ENSG00000280018-2-SEQ1	rGrGrA rUrCrU rGrArU rArArA rArCrU rGrArU rArUrU rGrGA G
	ENSG00000280018-2-SEQ2	rCrUrC rCrArA rUrArU rCrArG rUrUrU rUrArU rCrArG rArUrC rCrCrA
	ENSG00000280018-3-SEQ1	rArArU rGrCrU rArUrC rCrCrA rGrUrG rArUrU rGrUrA rCrGT G
	ENSG00000280018-3-SEQ2	rCrArC rGrUrA rCrArA rUrCrA rCrUrG rGrGrA rUrArG rCrArU rUrCrA
LINC00511	ENSG00000227036-1-SEQ1	rArCrU rCrUrC rArArG rGrUrA rGrArA rUrUrC rUrUrG rArUA A
	ENSG00000227036-1-SEQ2	rUrUrA rUrCrA rArGrA rArUrU rCrUrA rCrCrU rUrGrA rGrArG rUrUrG
	ENSG00000227036-2-SEQ1	rArUrG rGrCrA rGrArA rGrArC rGrCrU rUrArA rArArU rUrCT G
	ENSG00000227036-2-SEQ2	rCrArG rArArU rUrUrU rArArG rCrGrU rCrUrU rCrUrG rCrCrA rUrCrA
	ENSG00000227036-3-SEQ1	rGrArU rGrGrC rArGrA rArGrA rCrGrC rUrUrA rArArA

		rUrUC T
	ENSG00000227036-3-SEQ2	rArGrA rArUrU rUrUrA rArGrC rGrUrC rUrUrC rUrGrC rCrArU rCrArU
	ENSG00000279094-1-SEQ1	rArCrU rUrCrA rGrUrA rArUrG rCrUrC rArUrU rGrUrA rUrUT T
	ENSG00000279094-1-SEQ2	rArArA rArUrA rCrArA rUrGrA rGrCrA rUrUrA rCrUrG rArArG rUrArU
	ENSG00000279094-2-SEQ1	rGrGrA rArUrC rArUrU rArArG rArUrU rUrCrU rArUrU rCrUT T
LINC01670	ENSG00000279094-2-SEQ2	rArArA rGrArA rUrArG rArArA rUrCrU rUrArA rUrGrA rUrUrC rCrUrU
	ENSG00000279094-3-SEQ1	rArGrC rArArC rArUrA rUrUrG rArGrA rArCrA rGrArA rUrAA A
	ENSG00000279094-3-SEQ2	rUrUrU rArUrU rCrUrG rUrUrC rUrCrA rArUrA rUrGrU rUrGrC rUrGrC
	ENSG00000257181-1-SEQ1	rArArG rCrUrU rGrUrU rCrUrA rCrCrA rGrGrA rArUrG rArCA A
	ENSG00000257181-1-SEQ2	rUrUrG rUrCrA rUrUrC rCrUrG rGrUrA rGrArA rCrArA rGrCrU rUrUrA
AC025423.4		
	ENSG00000257181-2-SEQ1	rGrUrC rArGrA rArCrA rArUrU rArArA rArGrA rGrArU rCrAA A
	ENSG00000257181-2-SEQ2	rUrUrU rGrArU rCrUrC rUrUrU rUrArA rUrUrG rUrUrC

		rUrGrA rCrArG
	ENSG00000257181-3-SEQ1	rCrArA rGrArA rGrGrU rArCrU rUrUrA rArArG rUrGrU rCrUT A
	ENSG00000257181-3-SEQ2	rUrArA rGrArC rArCrU rUrUrA rArArG rUrArC rCrUrU rCrUrU rGrGrC
	ENSG00000286811-1-SEQ1	rGrGrA rArArU rGrGrG rArUrU rUrGrA rGrGrC rArArA rArCT A
	ENSG00000286811-1-SEQ2	rUrArG rUrUrU rUrGrC rCrUrC rArArA rUrCrC rCrArU rUrUrC rCrArG
	ENSG00000286811-2-SEQ1	rArArU rCrCrA rArArC rCrGrA rArArG rCrCrA rGrArA rGrGA A
AL353138		
	ENSG00000286811-2-SEQ2	rUrUrC rCrUrU rCrUrG rGrCrU rUrUrC rGrGrU rUrUrG rGrArU rUrGrG
	ENSG00000286811-3-SEQ1	rArArA rUrGrG rGrArU rUrUrG rArGrG rCrArA rArArC rUrAC C
	ENSG00000286811-3-SEQ2	rGrGrU rArGrU rUrUrU rGrCrC rUrCrA rArArU rCrCrC rArUrU rUrCrC

Table S2. Primer pairs used in RT-qPCR analyses.

Name	Sequence
HPRT-F	CCTGGCGTCGTGATTAGTGA
HPRT-R	CGAGCAAGACGTTCAGTCCT
ENSG00000280018-F	GATCTGTGTCTCTCGCTGGT
ENSG00000280018-R	GGCAAAACACACCTTCTTTCC
ENSG00000227036-F	TGGCCTTGGATGGAAAGTGG
ENSG00000227036-R	CTTTCCTGGTTCAAAGCACCC
ENSG00000279094-F	GGATCTGTGTCTCTCGCTGG
ENSG00000279094-R	GGGAGGTTCTGCATCCACAT
ENSG00000257181-F	TCGGCTCTCATCCCCAACTA
ENSG00000257181-R	TGGGTTACATGGTTCCCAGC
ENSG00000286811-F	TTTGCTACATCCCAGCTCCA
ENSG00000286811-R	CAAGCCATCTGGTTCAGGCTA
ENSG00000231312-F	AAACGCCCAGACCTTCTCTG
ENSG00000231312-R	TCCAGTGGCCAGGTATCTCA
CDKN1A(p21)-F	AGGCAAAAGTCCTGTGTTCCAA
CDKN1A(p21)-R	TACTCCCCACATAGCCCGTAT
CDKN2A(p16)-F	AGACACAAAGGACTCGGTGC
CDKN2A(p16)-R	CCGGACTAGGTAGGTGGAGT
CXCL1-F	CTGGCTTAGAACAAAGGGGCT
CXCL1-R	TAAAGGTAGCCCTTGTTTCCCC
CCND1-F	CAGATCATCCGCAAACACGC
CCND1-R	AAGTTGTTGGGGCTCCTCAG

CCNB1-F	CCCCTGCAGAAGAAGACCTG
CCNB1-R	AGTGACTTCCCGACCCAGTA
CCNE1-F	AGGGAGCGGGATGCGA
CCNE1-R	ATTGTCCCAAGGCTGGCTC
CDK4-F	TGTATGGGGCCGTAGGAACC
CDK4-R	GATCACGGGCCTTGTACACT

CHAPTER IV

CHARACTERISTICS OF CIRCULATING SMALL NON-CODING RNAS IN PLASMA AND SERUM DURING HUMAN AGING

ABSTRACT

Aging is a complicated process that triggers age-related disease susceptibility through intercellular communication in the microenvironment. While the classic secretome of senescence-associated secretory phenotype (SASP) including soluble factors, growth factors, and extracellular matrix remodeling enzymes are known to impact tissue homeostasis during the aging process, the effects of novel SASP components, extracellular small non-coding RNAs (sncRNAs), on human aging are not well-established. Here, by utilizing 446 small RNA-seq samples from plasma and serum of healthy donors found in the Extracellular RNA (exRNA) Atlas data repository, we successfully correlate features of human circulating sncRNAs with age. We observed the expression of a majority of transfer RNAs (tRNAs) and microRNAs (miRNAs) showed positive and negative associations with age respectively. We employed correlation analyses (including differential expression and maximum information coefficient (MIC)) and ensemble machine learning strategy to establish sncRNAs-based age predictors, resulting in a forecast performance where all R^2 values were greater than 0.94 and root-mean-square errors (RMSE) were less than 3.7 years in three ensemble machine learning methods (Adaptive Boosting, Gradient Boosting, and Random

Forest). Furthermore, age-related sncRNAs were identified based on modeling and the biological pathways of miRNAs were characterized by their predicted targets, including multiple pathways in cancer and longevity regulation. In summary, this study provides valuable insights into circulating sncRNAs dynamics in human aging and may lead to advanced understanding of age-related sncRNAs functions with further elucidation.

INTRODUCTION

Heterogeneity of human lifespan and health outcomes occurs due to differential aging process (Fleischer et al., 2018; Huan et al., 2018; Mamoshina et al., 2018). Organismal aging is often accompanied by dysregulation of numerous cellular and molecular processes that triggers age-related pathologies such as tissue degradation (Farr et al., 2017), tissue fibrosis (Gokey, 2021), arthritis (H. J. Lee et al., 2021), renal dysfunction (Chaib, Tchkonia, & Kirkland, 2021), diabetes (Palmer, Gustafson, Kirkland, & Smith, 2019), and cancer (Han, Li, Xiong, & Song, 2020). The highly proactive secretome from senescent cells, termed the senescence-associated secretory phenotype (SASP), is one of main drivers that cause age-related pathogenesis through intercellular communication (Fafian-Labora & O'Loghlen, 2020). SASP is mainly driven by persistent DNA damage response (DDR) (Rodier et al., 2009), with NF-κB and C/EBPβ signaling being activated by the transcription factor GATA4 (C. Kang et al., 2015). Also the epigenetic changes have been known to regulate the SASP and reduction of the retrotransposable element line 1 (L1) increased senescence-induced SASP (De Cecco et al., 2019). The classical SASP includes secretome of soluble factors, growth factors, and extracellular matrix remodeling enzymes (Coppe, Desprez, Krtolica, & Campisi, 2010), and it can transmit age-related information to the healthy cells via cell-to-cell contact. It has been reported that SASP can induce senescence in primary cell lines (Acosta et al., 2013) and trigger tumor progression in cancer tissues (Demaria et al., 2017; S. Lee & Schmitt, 2019) via paracrine senescence pathways. The good side is that SASP can also induce senescent cells removal in a paracrine manner, which is essential for tissue homeostasis (T. W. Kang et al., 2011; Xue et al., 2007).

As one of the emerging SASP components protected by extracellular vesicles (EVs), ribonucleoprotein (RNP) complexes, and lipoproteins (Gruner & McManus, 2021), extracellular RNAs (exRNAs) are found in many biological fluids (van Niel, D'Angelo, & Raposo, 2018) and can bridge the communication between 'donor' and 'recipient' cells through endocytosis, inducing paracrine senescence and pro-tumorigenic processes (Miyata et al., 2021; Y. Zhang et al., 2017). Deep sequencing of human plasma exRNA revealed more than 80% of sequencing reads mapped to small non-coding RNAs (sncRNAs) in human genome, including microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) (Huang et al., 2013). Extracellular miRNA expression in plasma of mice changes with age and cellular senescence can affect age-related homeostasis throughout the body by circulating miRNA (Alibhai et al., 2020). Other studies uncovered the roles of circulating miRNAs in detecting age-related dysfunction such as osteogenesis imperfecta (Davis et al., 2017), decreased myelination (Pusic & Kraig, 2014), tumorigenesis (Abels et al., 2019), and cardiovascular disease (Halkein et al., 2013). It has been demonstrated that exRNAs can retain without rapid degradation (Skog et al., 2008) and perform in vitro translation in recipient cells (Ridder et al., 2015). However, whether sncRNAs can directly regulate aging-related process in target cells needs more investigation (Gruner & McManus, 2021).Also, the molecular function of other circulating sncRNAs in aging and age-related diseases has been overlooked, and their expression profiles during human aging process must be further characterized.

To achieve this, we used 446 pre-selected small RNA-seq data from plasma and serum samples (age: 20-99 years). Combat batch effect correction or model fitting methods were used for batch effect removal, and samples showed adjusted transcriptomic feature by age without inter-datasets bias. Quasi-likelihood F-test and maximal information coefficient were employed differential expression analysis and linear or non-linear association measurements respectively to determine age-related sncRNAs as primary inputs for comprehensive machine learning modeling. Based on supervised ensemble machine learning models, aging estimators were created in high accuracy and sncRNAs candidates with top importance values in built models were considered as final age-related biomarkers. Additionally, pathway enrichment of targets of core miRNAs strengthens our viewpoint that extracellular sncRNAs change with age-related processes. This is the first study to systematically uncover the sncRNAs dynamics during healthy human aging process, providing perspective on small RNA biomarkers and small molecule therapeutics in aging and age-related diseases.

RESULTS

Overview of integrated human small RNAs dataset

To profile sncRNAs features during human healthy aging, we obtained small RNA-seq datasets from the Extracellular RNA (exRNA) Atlas data repository (https://exrna-atlas.org) (Murillo et al., 2019). This work includes the studies for which information on age, health status and gender, but only individuals having healthy aging process were retained for analysis. For datasets meeting the quality control standards established by the Extracellular RNA Communication Consortium (ERCC) (see experimental procedures), we created a bioinformatics procedure for reads mapping, processing, normalizing, categorizing and modeling (Figure 1a). As a result of these criteria, 302 plasma and 144 serum samples (Figure 1b) were used in this study, with a similar number of samples representing each gender ranging from 20–99 years old (Figure 1c). As these datasets originate from distinct studies with multiple sampling and library preparations, there are clear batch effects after Counts Per Million (CPM) normalization (Figure S1a and b). The ComBat function from the R package sva (v3.40.0) in Bioconductor (Leek, Johnson, Parker, Jaffe, & Storey, 2012) was employed to reduce or eliminate batch effect that may deviate from actual cross-study results (Figure S1c and d). These corrected data were used for correlation measurements and machine learning training described below.

Identification of expressed sncRNAs in plasma and serum

To determine sncRNAs expressed during aging, we considered sncRNAs with ≥ 1 CPM in at least 30% of individuals within an age group (young (20-30), adult (31-60) and aged (61+) groups) as expressed sncRNAs. As a result, there were 7953 and 6476 sncRNAs

observed in plasma and serum samples respectively (Figure 1a). Further, we identified highly expressed sncRNAs by increasing minimal CPM to 10, resulting in 1243 and 1139 sncRNAs retained in plasma and serum samples respectively (Figure 1a, Table S1). In terms of distribution of sncRNAs subtypes in three age groups, miRNAs account for a high proportion (26.5-63.4%) of all sncRNAs in both plasma and serum, and their abundance consistently decreased with age (Figure 2a and b). tRNAs increased and became the dominant sncRNA in aged group while expression of miRNAs were reduced in older individuals (Figure 2a and b). The corresponding mapped reads are proportional to the number of each highly expressed subtype, even though miRNA showed relatively more sequencing reads than others in both plasma and serum (Figure 2c and d).

Exploring the correlation between sncRNAs and human aging

Utilizing data from batch effect corrected expressed sncRNAs, we identified differentially expressed (DE) sncRNAs that were up or down regulated in the aged group relative to either young or adult groups (FDR < 0.05). 581 plasma and 188 serum sncRNAs were identified as differentially expressed (Figure 3a and b), with miRNAs constituting the greatest number in plasma and tRNAs accounting for the greatest number in serum (Figure S2a and b). To understand the functional role of these DE sncRNAs, an over-representation analysis of targets of the DE miRNAs was performed. miRNAs were chosen as this species of sncRNA are the best studied of sncRNAs involved in gene regulation. Gene ontology (GO) biological processes of miRNA targets were detected and for plasma samples, these targets were enriched in melanin deposition, immune response, cell proliferation and metabolic homeostasis (Figure 3e). As for DE miRNAs in serum, their targets were included

in immune and neural system development, as well as signal transduction process (Figure 3f).

In parallel, we calculated the maximum information coefficient (MIC) (D. N. Reshef et al., 2011) to investigate both linear and nonlinear associations between sncRNAs expression and corresponding individual age. By employing batch-corrected data of expressed sncRNAs, we identified 364 and 1941 age-related sncRNAs from plasma and serum respectively (Figure 3c and d). Intriguingly, piRNAs became the most abundant sncRNAs in MIC measurement, with the number of snRNAs representing the second largest (Figure S2c and d). Similarly, the over-represented biological processes of miRNA targets were identified, and cellular response and epigenetic modification were enriched in plasma (Figure 3g), while biosynthetic processes were significantly observed in serum samples (Figure 3h).

Core feature selection of age-related sncRNAs

As the expression of sncRNAs changes with age, further data-driven analysis was conducted to construct a human aging clock. DE sncRNAs or MIC-based age-correlated sncRNAs were used as inputs to train regression models in plasma and serum samples. Compared to the linear models, including the Linear Regression (without feature selection) and Elastic Net (feature selection through regularization), the tree-based ensemble machine learning methods (including Adaptive Boosting, Gradient Boosting, and Random Forest regressors) showed stronger power of prediction with better performance in accuracy (Figure 4A), due to its great capability of learning the underlying nonlinear patterns. With stably ideal performance in test subsets (Table S2), all models inputting either DE sncRNAs

(DE_plasma and DE_serum) or age-correlated sncRNAs (MIC_plasma and MIC_serum) accurately predict the ages of corresponding individuals in test sets, with average R^2 values greater than 0.94, root mean squared error (RMSE) values less than 3.7 years and mean absolute error (MAE) values less than 2 years (Figure 4).

Notably, we also observed a gender-specific model performance. When male-only samples were used as training set for predicting female-only test sets or vice versa, there were core sncRNAs unique to one gender (Figure S3a), with slightly lower performance in R^2 and RMSE values compared to the models trained in gender-mixed data (Figure S3b).

Due to the strong generalization ability in all ensemble learning methods, core sncRNAs associated with aging processes were determined by combined statistics and sum of importance ranks in the three methods was used as the criteria for core sncRNAs identification. As a result, there were 293, 169, 222 and 321 core sncRNAs overlapped in all three methods with DE_plasma, DE_serum, MIC_plasma and MIC_serum as the inputs respectively (Figure S4). Particularly, 8 piRNAs, 5 snRNAs, 4 miRNAs, 2 small cytoplasmic RNAs and one tRNA were identified as top core sncRNAs in plasma (Table 1 and Figure 5a), and 12 snRNAs, 4 miRNAs, 2 tRNAs, one snoRNA and one small cytoplasmic RNA identified as top core sncRNAs in serum samples (Table 2 and Figure 5b).

Core miRNAs are involved in aging-related processes

To gain further insight into extracellular sncRNAs potential functions in a microenvironment, we focused on miRNAs, which are well characterized in post-transcriptional gene regulation. The targets of core miRNAs in plasma and serum were predicted via the integration of 8 miRNAs databases. Their expressional profile in three age

groups is in Figure S5 and corresponding targets are included in Table S3. As expected, these miRNA targets are enriched in canonical pathways such as AMPK, GnRH, FoxO and insulin signaling pathways, as well as cellular senescence, longevity regulation and tumorigenesis pathways closely related to aging process (Figure S6).

We also investigated the association between miRNA targets and protein coding genes previously validated in the human aging process from Human Ageing Genomic Resources (Tacutu et al., 2018), and we found these targets were experimentally identified to be associated with cancer progression, senescence, aging and longevity (Figure 6), bolstering the probability that other non-miRNA sncRNAs also have functions in aging and aging related diseases.

Furthermore, to uncover the heterogeneity of age-related sncRNAs between healthy control and diseased groups, we used plasma and serum small RNA-seq data from unhealthy individuals, including plasma samples from colon carcinoma (n=100), pancreatic carcinoma (n=6) and prostate carcinoma (n=36) patients, and serum samples of patients suffered from Alzheimer's (n=44) and Parkinson's (n=47) diseases. By performing principal component analysis, there was little variance within plasma unhealthy samples, and they tightly clustered when DE sncRNAs and age-related sncRNA were employed, regardless of age difference (Figure S7a and b). To exclude the possibility that these results result from the batch effects of sequencing, expressed sncRNAs (7953 in total) were used for principal component analysis, and samples from 3 carcinoma sets showed diffuse distribution as healthy controls (Figure S7c). By contrast, for the illness that has more association with age (occurs mainly in the elderly), samples collected from patients with Alzheimer's and

Parkinson's disease showed more similar profile as healthy aging individuals (Figure S8). We hypothesize that DE/age-related sncRNAs from healthy people showed no obvious signal in patients with less age-related lesions, while these sncRNAs have mild dynamics in gradually aging process, and further research about specific molecular difference between healthy and pathological aging is required.

DISCUSSION

Our study comprehensively profiled the relationship of extracellular sncRNAs with age in blood and built an aging clock of healthy individuals using DE sncRNAs or sncRNAs linear and nonlinear correlated with age. Previously, age predictors were developed through DNA methylation sites (Lu et al., 2019), transcriptome expression (Galkin et al., 2020; Shokhirev & Johnson, 2021), repeat elements (LaRocca, Cavalier, & Wahl, 2020), microRNAs (Huan et al., 2018) and protein abundance (Johnson, Shokhirev, Wyss-Coray, & Lehallier, 2020). This study provides the first detailed analysis of relationship between circulating sncRNAs and age based on regression models and core sncRNAs whose expression changes with age, allowing reliable age prediction.

From previous human biofluids studies, differential composition of small RNA has been reported in multiple biofluids. Godoy et al. (Godoy et al., 2018) used 12 normal human biofluids including plasma and serum in their study and for mapping reads of corresponding RNA sequencing (RNA-seq), miRNA showed relative high fraction (63.8906%, median) in adult plasma compared to serum (36.0154%, median). However, the percentage of tRNA mapped reads in serum increased (42.2067%, median) and became the most abundant RNA biotype, while median value was 0.7759% in adult plasma. One study determined the diversity of small RNA in different biofluids, and tRNA showed the largest percentage of mapped reads (39.7%) in serum compared to plasma (5.8%) and whole blood (2.1%) (El-Mogy et al., 2018). Also, in Max et al. study (Max et al., 2018), they characterized extracellular RNAs (exRNAs) from both plasma and serum samples of the same healthy volunteers, and interestingly they showed substantial differences of small RNA composition,

with higher proportion of miRNA in plasma and more tRNA reads in serum. We have some serum and plasma samples from the same individuals (Study_ID: EXR-TTUSC1gCrGDH-AN) and consistent results were observed (Figure 2). They also concluded that different biofluid types, even though come from the same origin, plasma and serum show significant variable that impact exRNA profile. One of the reasons is that additional absorption and continuous degradation of exRNAs by retained blood clot will reduce exRNA abundance (Max et al., 2018). So proper exRNA isolation is essential and immediate platelet and cell debris depletion for plasma collection may avoid losses of exRNA characteristics as much as possible.

It is of interest to identify a detectable increase of highly expressed tRNAs in aged individuals, and it has been reported that spleen and brain had the highest tRNA expression (Dittmar, Goodenbour, & Pan, 2006), which may indicate unique and differential biological process happen as individuals age. A previous report similarly finds tRNAs were the second most abundant sncRNAs in healthy adults (20 – 40 years) when small cytoplasmic RNA was not mentioned (Danielson, Rubio, Abderazzaq, Das, & Wang, 2017).Unlike tRNAs driving protein synthesis, tRNA-derived small RNAs (tsRNAs) including tRNA derived fragment (tRF) and stress-induced tRNA halves (tiRNA), have been uncovered as aging process related sncRNAs (Pan, Han, & Li, 2021). Similar as human studies, the expression of tsRNAs increased during aging in *Drosophila* (Karaiskos, Naqvi, Swanson, & Grigoriev, 2015), *C. elegans* (Kim & Lee, 2019) and mouse brain cells (Dhahbi et al., 2013). Compared with healthy controls, differential expression of tsRNAs in age-related diseases has been employed in disease prediction such as Alzheimer's disease and Parkinson's

disease (S. Zhang et al., 2019), ischaemic stroke (Elkordy et al., 2019) and osteoporosis (Y. Zhang et al., 2018). tsRNAs have roles not only in potential biomarkers, but also in expressional regulation of age-related mRNAs (Pan et al., 2021). For example, 5′-tRF^{Tyr} from tyrosine pre-tRNA can silence PKM2, which is the inhibitor of p53, to cause p53-dependent neuronal death (Inoue et al., 2020). The number of highly expressed miRNA in our study displayed a decreased tendency in older group, and it has been observed in both plasma and serum. Seven of the top 8 core miRNAs identified by machine learning models were found to have reduced expression as age increased, similar to decreased expression of a majority of age-associated miRNAs in whole-blood (Huan et al., 2018), serum (H. Zhang et al., 2015) and peripheral blood mononuclear cells (Noren Hooten et al., 2010).

It has been previously demonstrated that circulating sncRNAs from serum samples show strong association with human aging (Rounge et al., 2018), while the human aging modeling based on regression relationship was not yet built. In our study, potential function of core sncRNAs was predicted via miRNA target prediction, and these genes showed enrichment in cancer, cell cycle, and longevity regulating pathways. There are overlapping genes included in both cancer and longevity regulation pathways, and this result were consistent with early study that profiled miRNAs expression between young and old individuals (Noren Hooten et al., 2010). For example, increased *PIK3R1* expression has been identified to impair anti-tumor effect through PI3K-Akt activation in breast and ovarian cancer chemotherapy (Chi et al., 2019; X. Li et al., 2019). Previous research determined that protein level of p85α, which is the subunit of *PIK3R1*, were elevated with age, and age-associated miRNAs that potentially target *PIK3R1* were downregulated (Noren

Hooten et al., 2010). Interestingly, the expression of core miRNAs (hsa-miR-203a-3p, hsa-miR-203b-3p, hsa-miR-129-2-3p and hsa-miR-372-3p) targeting *PIK3R1* were lower in aged individuals (Figure S5). Studies in human aging also show that sequence variations within *PIK3R1* gene are significantly correlated with longevity (Donlon et al., 2018), and individuals with different genotypes of *PIK3R1* were associated with longevity through reduced mortality risk in cardiovascular disease (Donlon, Chen, Masaki, Willcox, & Morris, 2021). Further, *FOXO1* is associated with resistance to oxidative stress by increasing the antioxidant capacity, thereby maintaining reactive oxygen species (ROS) homeostasis and preventing pathological processes including cancer and other age-associated diseases (Storz, 2011). The core miRNA hsa-miR-206 has been associated with chemo-resistance in breast cancer through inhibiting PI3K/Akt/mTOR signaling (H. Li et al., 2021) and its decreased expression in the aged group in our study corresponds to the tumor-prone phenotype in elderly group. Consistent with our study, increased hsa-miR-9-3p expression was observed in diabetic retinopathy patients, which further enhances abnormal angiogenesis that hampered vision therapy. The function of most of age-associated sncRNAs identified in this study is unknown and further investigation into their function may provide meaningful results.

In our study we used DE or MIC-based age related sncRNAs as inputs for machine learning modeling, and there were limited overlapped sncRNAs between two methods in either plasma or serum samples (Figure S4). Interestingly, out of 8 overlapped sncRNAs in plasma samples (Figure S4), three of them (U5-L154, hsa-miR-9-3p and piR-37253) were top 10 core sncRNAs of DE_plasma (Table 1), indicating a high proportion (3/8) of top core

sncRNAs within overlapped sncRNAs. Also, there were three (U13-L308, hsa-miR-203b-3p and U4-L151) within top core sncRNAs of DE_serum (Table 1), out of 7 overlapped sncRNAs in serum samples (Figure S4). Since MIC calculates both linear and nonlinear associations between sncRNAs expression and corresponding individual age, it is possible but not necessary to have many overlaps between MIC and DEG-based sncRNA inputs, and MIC-based core sncRNAs with nonlinear relationship, including parabolic and sinusoidal correlations (Cao, Chen, Chen, Zhang, & Yuan, 2021), may have bigger importance score, while most of core sncRNAs from DEG inputs are obviously linear.

We also observed the mild sex-dependent differences in the aging clock modeling. Similarly, a previous study indicated that sncRNAs differences between genders were minor (Max et al., 2018) and sex-specific training sets have relatively low performance score in prediction compared to the gender-mixed training sets. During this process, some gender-dependent core sncRNAs were identified, including male-specific sncRNAs piR-31143 and piR-48977 in plasma, male-specific sncRNAs piR-33527 and piR-57256 in serum, female-specific sncRNAs hsa-miR-3789 and U5-L214 in plasma and female-specific sncRNAs U6-L989 and piR-30597 in serum. Further mechanistic study is needed to uncover their prospective role in aging and aging-related disease.

A major limitation of our current study is the corresponding datasets utilized were developed by researchers for different, unique projects and with multiple RNA extraction protocols, which may bias extracellular RNA abundance (Danielson et al., 2017). Furthermore, trait information such as ethnicity, body mass and smoking habits were not considered in our study due to the lack of information, and a more sophisticated and

systematic sample processing and recording would help future research on big data-based human aging modeling.

In conclusion, we provide a novel insight into the circulating sncRNAs profile of human aging. We developed predictive models in uncovering core sncRNAs and estimated age by utilizing meta-analysis based correlation measurement and machine learning modeling. With the constructed age predictor, the determination of transcriptome-dependent healthy aging changes can be captured, more importantly these models can be applied in detecting the age-associated pathologies when abnormal outputs appeared. Compared to healthy individuals, age prediction by sncRNA profiles of patients may show large variations with their actually chronological age. The sncRNA dynamics with age provide valuable references for extracellular RNA study in aging, and rapid small RNA biomarker detection facilitates the possibility for small molecule drug discovery to intervene aging related dysfunction. Practically, in both human health and animal production areas, the developed strategy in this study can be further employed to perform preprocessing, quantification, association identification and machine learning modeling for determining core features, for example, they can be genes, environmental factors, image data, phenotypic indexes or behavior notes, that are correlated most with variables we are interested in.

EXPERIMENTAL PROCEDURES

Data acquisition and filtration

Human small RNA-Seq datasets in the extracellular RNA (exRNA) Atlas data repository (https://exrna-atlas.org) (Murillo et al., 2019) were queried with studies filtered using the following requirements: 1) data were sequenced from plasma/serum samples; 2) samples have definitive age and gender information within each study, and 3) the donor of corresponding samples should have a healthy status and was selected as a control individual for the study. As a result, two studies (Accession ID: EXR-MTEWA1ZR3Xg6-AN and EXR-TTUSC1gCrGDH-AN) were included in both plasma and serum studies, and two studies (Accession ID: EXR-TPATE1OqELFf-AN and EXR-KJENS1sPlvS2-AN) were obtained with only plasma and serum samples respectively and 366 plasma and 188 serum samples passed preliminary filtration. To avoid genes' expressional bias due to the low sequencing reads and host genome contamination, we only retained samples that met the quality control (QC) standards developed by Extracellular RNA Communication Consortium (ERCC). Specifically, individual dataset should have a minimum of 100,000 reads that aligned to annotated RNA transcript (including miRNAs, piRNAs, tRNAs, snoRNAs, circular RNAs, protein coding genes and long non-coding RNAs), and ratio of transcriptome reads over total sequencing reads should be more than 0.5. Consequently, 302 plasma and 144 serum samples were retained for further analysis.

Quantification and batch effect removal

To generate expression matrices of sncRNAs, read adaptors and low quality bases were removed using the Trim Galore (v0.6.5) wrapper (Krueger, James, Ewels, Afyounian, &

Schuster-Boeckler, 2021) with default parameters. Clean reads were aligned and quantified with bowtie2 (v2.4.4) (Langmead & Salzberg, 2012) and samtools (v1.1.4) (Danecek et al., 2021) through miRNAs and other sncRNAs annotation file from the miRBase (Release 22.1) and the DASHR (v2.0) (Kuksa et al., 2019) database, respectively. The raw sncRNAs expression results were integrated and processed in R (v4.1.1) computational environment for identifying age-related sncRNAs after preprocessing. To correct for actual expression characteristics masked by sequencing depth variability, gene read counts were transformed into CPM values after measuring normalized library sizes by edgeR (v3.14) package (Robinson, McCarthy, & Smyth, 2010). Since there were still obvious batch effects observed via principal component analysis (Figure S1), we conducted batch removal using the ComBat function in sva package (v3.40.0) (Leek et al., 2012), and processed CPM-based data showed improved sample clustering by age (Figure S1). Batch-effect corrected data were used for identifying maximum information coefficient and constructing machine learning models described below.

Differential sncRNAs expression with age

To identify sncRNAs changing with age, samples were divided into three groups, including young (20-30 years), adult (31-60 years) and aged (61+ years), and CPM-based data were used for differential expression analyses by quasi-likelihood F-test in the edgeR package after considering different datasets as batch effect in a designed model. The DE sncRNAs (up-regulated or down-regulated in aged group compared to either young or adult group, FDR < 0.05) were utilized for machine learning modeling, starting with the batch effect removed data.

Identification of association between sncRNAs and age

To select the sncRNAs representative for the age prediction model, the maximal information coefficient (MIC) (D. N. Reshef et al., 2011), which permits the identification of important, difficult-to-detect associations (Y. Zhang, Jia, Huang, Qiu, & Zhou, 2014), was used to determine and screen the linear or non-linear correlations between each sncRNA expression (*X*) and the individual's chronological age (*Y*). Reshef et al. (D. N. Reshef et al., 2011) reported that MIC $-\rho^2$ to be near zero for linear relationships and MIC $-\rho^2 > 0.2$ for nonlinear relationships, where ρ^2 is the coefficient of determination (R^2). We also employed total information coefficient (TIC) to evaluate the power of independence testing between *X* and *Y* (Y. A. Reshef, Reshef, Finucane, Sabeti, & Mitzenmacher, 2016). The sncRNAs with expression having both MIC and TIC values greater than 0.7 with actual age were retained for building models.

Comprehensive machine learning modeling

The expression data of sncRNAs selected from differential expression analysis and MIC-based correlation measurement were used for machine learning modeling. Since sncRNAs expression inputs could be seen as the explanatory variable **X,** which is a high dimensional vector, the modeling process was performed as a regression analysis problem and was formularized as:

$$y = \hat{f}(\mathbf{X}) \quad (1)$$

where **X** denotes the sncRNA inputs, *y* denotes individual's age and $\hat{f}$ denotes the fitted mapping function. Ensemble learning including Adaptive Boosting, Gradient Boosting and Random Forest were leveraged in this study, taking advantage of their strong generalization

ability achieved by multiple weak learners combination (Wang, 2006). Based on manual parameter tuning, the parameter "number of estimators", which is the number of weak learners (i.e., the regression tree in this study) to be integrated in model fitting, was determined in each specific model based on the overall performance (RMSE, R^2, and MAE, showed in Table S4). The performance of ensemble learning is compared with linear regression and elastic net. The corresponding importance of each sncRNA was calculated as impurity-based feature score (sum to 1), which can be used to determine the decisiveness of sncRNA that it makes contribution to decide individual age (Louppe, 2014). Potentially core sncRNAs were determined by sorting the corresponding sum of ranks of their importance values in each ensemble learning model.

Since the number of samples is different in each age group (young, adult and aged), simple k-fold cross-validation may cause uneven sampling and then trigger bad model performance due to over-fitting. Therefore, stratified k-fold cross-validation is a better option to avoid this issue by selecting approximately the same proportions of samples in each pre-set age group to the training set (Figure S9). In this study, we stratified 5-fold cross-validation based on the overall sample size. The regression modeling was conducted under Python 3.8.8 and scikit-learn 0.24.1 (Pedregosa et al., 2011).

Targets prediction of age-related miRNAs

To better understand the potential function of circulating sncRNAs changing with age, we primarily predicted the targets of miRNA candidates by using multiMiR R package (V3.14) (Ru et al., 2014), which integrates 8 microRNA-target databases (DIANA-microT, ElMMo, MicroCosm, miRanda, miRDB, PicTar, PITA and TargetScan). Only the top 20%

predicted targets with high sequence affinity in each database were considered as function related genes.

Functional enrichment analyses

Pathway enrichment analyses of gene targets of age-related miRNAs performed through Enrichr gene list-based enrichment analysis tool (Xie et al., 2021). We used the combined score, which is a combination of the P-value and Z-score, to offset the false positive rate caused by the different length of each term and input sets. For direct miRNAs functional enrichment, an over-representation analysis was performed via miRNA Enrichment Analysis and Annotation Tool (miEAA 2.0) (Kern et al., 2020), with expressed miRNA sets as the background set and P-values adjusted using Benjamini-Hochberg (BH) procedure (P-adj < 0.05).

REFERENCES

Abels, E. R., Maas, S. L. N., Nieland, L., Wei, Z., Cheah, P. S., Tai, E., . . . Breakefield, X. O. (2019). Glioblastoma-Associated Microglia Reprogramming Is Mediated by Functional Transfer of Extracellular miR-21. *Cell Rep, 28*(12), 3105-3119 e3107. doi:10.1016/j.celrep.2019.08.036

Acosta, J. C., Banito, A., Wuestefeld, T., Georgilis, A., Janich, P., Morton, J. P., . . . Gil, J. (2013). A complex secretory program orchestrated by the inflammasome controls paracrine senescence. *Nat Cell Biol, 15*(8), 978-990. doi:10.1038/ncb2784

Alibhai, F. J., Lim, F., Yeganeh, A., DiStefano, P. V., Binesh-Marvasti, T., Belfiore, A., . . . Li, R. K. (2020). Cellular senescence contributes to age-dependent changes in circulating extracellular vesicle cargo and function. *Aging Cell, 19*(3), e13103. doi:10.1111/acel.13103

Cao, D., Chen, Y., Chen, J., Zhang, H., & Yuan, Z. (2021). An improved algorithm for the maximal information coefficient and its application. *R Soc Open Sci, 8*(2), 201424. doi:10.1098/rsos.201424

Chaib, S., Tchkonia, T., & Kirkland, J. L. (2021). Obesity, Senescence, and Senolytics. *Handb Exp Pharmacol*. doi:10.1007/164_2021_555

Chi, Y., Xue, J., Huang, S., Xiu, B., Su, Y., Wang, W., . . . Wu, J. (2019). CapG promotes resistance to paclitaxel in breast cancer through transactivation of PIK3R1/P50. *Theranostics, 9*(23), 6840-6855. doi:10.7150/thno.36338

Coppe, J. P., Desprez, P. Y., Krtolica, A., & Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol, 5*, 99-118. doi:10.1146/annurev-pathol-121808-102144

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10*(2). doi:10.1093/gigascience/giab008

Danielson, K. M., Rubio, R., Abderazzaq, F., Das, S., & Wang, Y. E. (2017). High Throughput Sequencing of Extracellular RNA from Human Plasma. *PLoS One, 12*(1), e0164644. doi:10.1371/journal.pone.0164644

Davis, C., Dukes, A., Drewry, M., Helwa, I., Johnson, M. H., Isales, C. M., . . . Hamrick, M. W. (2017). MicroRNA-183-5p Increases with Age in Bone-Derived Extracellular Vesicles, Suppresses Bone Marrow Stromal (Stem) Cell Proliferation, and Induces Stem Cell Senescence. *Tissue Eng Part A, 23*(21-22), 1231-1240. doi:10.1089/ten.TEA.2016.0525

De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., . . . Sedivy, J. M. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature, 566*(7742), 73-78. doi:10.1038/s41586-018-0784-9

Demaria, M., O'Leary, M. N., Chang, J., Shao, L., Liu, S., Alimirah, F., . . . Campisi, J. (2017). Cellular Senescence Promotes Adverse Effects of Chemotherapy and Cancer Relapse. *Cancer Discov, 7*(2), 165-176. doi:10.1158/2159-8290.CD-16-0241

Dhahbi, J. M., Spindler, S. R., Atamna, H., Yamakawa, A., Boffelli, D., Mote, P., & Martin,

D. I. (2013). 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics, 14*, 298. doi:10.1186/1471-2164-14-298

Dittmar, K. A., Goodenbour, J. M., & Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS Genet, 2*(12), e221. doi:10.1371/journal.pgen.0020221

Donlon, T. A., Chen, R., Masaki, K. H., Willcox, B. J., & Morris, B. J. (2021). Association with Longevity of Phosphatidylinositol 3-Kinase Regulatory Subunit 1 Gene Variants Stems from Protection against Mortality Risk in Men with Cardiovascular Disease. *Gerontology*, 1-9. doi:10.1159/000515390

Donlon, T. A., Morris, B. J., Chen, R., Masaki, K. H., Allsopp, R. C., Willcox, D. C., . . . Willcox, B. J. (2018). Analysis of Polymorphisms in 59 Potential Candidate Genes for Association With Human Longevity. *J Gerontol A Biol Sci Med Sci, 73*(11), 1459-1464. doi:10.1093/gerona/glx247

El-Mogy, M., Lam, B., Haj-Ahmad, T. A., McGowan, S., Yu, D., Nosal, L., . . . Haj-Ahmad, Y. (2018). Diversity and signature of small RNA in different bodily fluids using next generation sequencing. *BMC Genomics, 19*(1), 408. doi:10.1186/s12864-018-4785-8

Elkordy, A., Rashad, S., Shehabeldeen, H., Mishima, E., Niizuma, K., Abe, T., & Tominaga, T. (2019). tiRNAs as a novel biomarker for cell damage assessment in in vitro ischemia-reperfusion model in rat neuronal PC12 cells. *Brain Res, 1714*, 8-17. doi:10.1016/j.brainres.2019.02.019

Fafian-Labora, J. A., & O'Loghlen, A. (2020). Classical and Nonclassical Intercellular Communication in Senescence and Ageing. *Trends Cell Biol, 30*(8), 628-639. doi:10.1016/j.tcb.2020.05.003

Farr, J. N., Xu, M., Weivoda, M. M., Monroe, D. G., Fraser, D. G., Onken, J. L., . . . Khosla, S. (2017). Targeting cellular senescence prevents age-related bone loss in mice. *Nat Med, 23*(9), 1072-1079. doi:10.1038/nm.4385

Fleischer, J. G., Schulte, R., Tsai, H. H., Tyagi, S., Ibarra, A., Shokhirev, M. N., . . . Navlakha, S. (2018). Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol, 19*(1), 221. doi:10.1186/s13059-018-1599-6

Galkin, F., Mamoshina, P., Aliper, A., de Magalhaes, J. P., Gladyshev, V. N., & Zhavoronkov, A. (2020). Biohorology and biomarkers of aging: Current state-of-the-art, challenges and opportunities. *Ageing Res Rev, 60*, 101050. doi:10.1016/j.arr.2020.101050

Godoy, P. M., Bhakta, N. R., Barczak, A. J., Cakmak, H., Fisher, S., MacKenzie, T. C., . . . Erle, D. J. (2018). Large Differences in Small RNA Composition Between Human Biofluids. *Cell Rep, 25*(5), 1346-1358. doi:10.1016/j.celrep.2018.10.014

Gokey, J. J. (2021). Editorial: From Development to Senescence, Bridging the Gap in Lung Fibrosis. *Front Med (Lausanne), 8*, 798164. doi:10.3389/fmed.2021.798164

Gruner, H. N., & McManus, M. T. (2021). Examining the evidence for extracellular RNA function in mammals. *Nat Rev Genet, 22*(7), 448-458. doi:10.1038/s41576-021-00346-8

Halkein, J., Tabruyn, S. P., Ricke-Hoch, M., Haghikia, A., Nguyen, N. Q., Scherr, M., . . .

Struman, I. (2013). MicroRNA-146a is a therapeutic target and biomarker for peripartum cardiomyopathy. *J Clin Invest, 123*(5), 2143-2154. doi:10.1172/JCI64365

Han, Q., Li, J., Xiong, J., & Song, Z. (2020). Long noncoding RNA LINC00514 accelerates pancreatic cancer progression by acting as a ceRNA of miR-28-5p to upregulate Rap1b expression. *J Exp Clin Cancer Res, 39*(1), 151. doi:10.1186/s13046-020-01660-5

Huan, T., Chen, G., Liu, C., Bhattacharya, A., Rong, J., Chen, B. H., . . . Levy, D. (2018). Age-associated microRNA expression in human peripheral blood is associated with all-cause mortality and age-related traits. *Aging Cell, 17*(1). doi:10.1111/acel.12687

Huang, X., Yuan, T., Tschannen, M., Sun, Z., Jacob, H., Du, M., . . . Wang, L. (2013). Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics, 14*, 319. doi:10.1186/1471-2164-14-319

Inoue, M., Hada, K., Shiraishi, H., Yatsuka, H., Fujinami, H., Morisaki, I., . . . Hanada, T. (2020). Tyrosine pre-transfer RNA fragments are linked to p53-dependent neuronal cell death via PKM2. *Biochem Biophys Res Commun, 525*(3), 726-732. doi:10.1016/j.bbrc.2020.02.157

Johnson, A. A., Shokhirev, M. N., Wyss-Coray, T., & Lehallier, B. (2020). Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev, 60*, 101070. doi:10.1016/j.arr.2020.101070

Kang, C., Xu, Q., Martin, T. D., Li, M. Z., Demaria, M., Aron, L., . . . Elledge, S. J. (2015).

The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4. *Science, 349*(6255), aaa5612. doi:10.1126/science.aaa5612

Kang, T. W., Yevsa, T., Woller, N., Hoenicke, L., Wuestefeld, T., Dauch, D., . . . Zender, L. (2011). Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature, 479*(7374), 547-551. doi:10.1038/nature10599

Karaiskos, S., Naqvi, A. S., Swanson, K. E., & Grigoriev, A. (2015). Age-driven modulation of tRNA-derived fragments in Drosophila and their potential targets. *Biol Direct, 10*, 51. doi:10.1186/s13062-015-0081-6

Kern, F., Fehlmann, T., Solomon, J., Schwed, L., Grammes, N., Backes, C., . . . Keller, A. (2020). miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res, 48*(W1), W521-W528. doi:10.1093/nar/gkaa309

Kim, S. S., & Lee, S. V. (2019). Non-Coding RNAs in Caenorhabditis elegans Aging. *Mol Cells, 42*(5), 379-385. doi:10.14348/molcells.2019.0077

Krueger, F., James, F., Ewels, P., Afyounian, E., & Schuster-Boeckler, B. (2021). FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. In.

Kuksa, P. P., Amlie-Wolf, A., Katanic, Z., Valladares, O., Wang, L. S., & Leung, Y. Y. (2019). DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. *Bioinformatics, 35*(6), 1033-1039. doi:10.1093/bioinformatics/bty709

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

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

Lee, H. J., Lee, W. J., Hwang, S. C., Choe, Y., Kim, S., Bok, E., . . . Lee, S. L. (2021). Chronic inflammation-induced senescence impairs immunomodulatory properties of synovial fluid mesenchymal stem cells in rheumatoid arthritis. *Stem Cell Res Ther, 12*(1), 502. doi:10.1186/s13287-021-02453-z

Lee, S., & Schmitt, C. A. (2019). The dynamic nature of senescence in cancer. *Nat Cell Biol, 21*(1), 94-101. doi:10.1038/s41556-018-0249-2

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics, 28*(6), 882-883. doi:10.1093/bioinformatics/bts034

Li, H., Xu, W., Xia, Z., Liu, W., Pan, G., Ding, J., . . . Jiang, D. (2021). Hsa_circ_0000199 facilitates chemo-tolerance of triple-negative breast cancer by interfering with miR-206/613-led PI3K/Akt/mTOR signaling. *Aging (Albany NY), 13*(3), 4522-4551. doi:10.18632/aging.202415

Li, X., Mak, V. C. Y., Zhou, Y., Wang, C., Wong, E. S. Y., Sharma, R., . . . Cheung, L. W. T. (2019). Deregulated Gab2 phosphorylation mediates aberrant AKT and STAT3 signaling upon PIK3R1 loss in ovarian cancer. *Nat Commun, 10*(1), 716. doi:10.1038/s41467-019-08574-7

Louppe, G. J. a. e.-p. (2014). Understanding Random Forests: From Theory to Practice. arXiv:1407.7502. https://ui.adsabs.harvard.edu/abs/2014arXiv1407.7502L

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., . . . Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY), 11*(2), 303-327. doi:10.18632/aging.101684

Mamoshina, P., Volosnikova, M., Ozerov, I. V., Putin, E., Skibina, E., Cortese, F., & Zhavoronkov, A. (2018). Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front Genet, 9*, 242. doi:10.3389/fgene.2018.00242

Max, K. E. A., Bertram, K., Akat, K. M., Bogardus, K. A., Li, J., Morozov, P., . . . Tuschl, T. (2018). Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proc Natl Acad Sci U S A, 115*(23), E5334-E5343. doi:10.1073/pnas.1714397115

Miyata, K., Imai, Y., Hori, S., Nishio, M., Loo, T. M., Okada, R., . . . Takahashi, A. (2021). Pericentromeric noncoding RNA changes DNA binding of CTCF and inflammatory gene expression in senescence and cancer. *Proc Natl Acad Sci U S A, 118*(35). doi:10.1073/pnas.2025647118

Murillo, O. D., Thistlethwaite, W., Rozowsky, J., Subramanian, S. L., Lucero, R., Shah, N., . . . Milosavljevic, A. (2019). exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell, 177*(2), 463-477 e415. doi:10.1016/j.cell.2019.02.018

Noren Hooten, N., Abdelmohsen, K., Gorospe, M., Ejiogu, N., Zonderman, A. B., & Evans, M. K. (2010). microRNA expression patterns reveal differential expression of target genes with age. *PLoS One, 5*(5), e10724. doi:10.1371/journal.pone.0010724

Palmer, A. K., Gustafson, B., Kirkland, J. L., & Smith, U. (2019). Cellular senescence: at the nexus between ageing and diabetes. *Diabetologia, 62*(10), 1835-1841. doi:10.1007/s00125-019-4934-x

Pan, Q., Han, T., & Li, G. (2021). Novel insights into the roles of tRNA-derived small RNAs. *RNA Biol, 18*(12), 2157-2167. doi:10.1080/15476286.2021.1922009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. J. t. J. o. m. L. r. (2011). Scikit-learn: Machine learning in Python. *12*, 2825-2830.

Pusic, A. D., & Kraig, R. P. (2014). Youth and environmental enrichment generate serum exosomes containing miR-219 that promote CNS myelination. *Glia, 62*(2), 284-299. doi:10.1002/glia.22606

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., . . . Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science, 334*(6062), 1518-1524. doi:10.1126/science.1205438

Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C., & Mitzenmacher, M. J. T. J. o. M. L. R. (2016). Measuring dependence powerfully and equitably. *17*(1), 7406-7468.

Ridder, K., Sevko, A., Heide, J., Dams, M., Rupp, A. K., Macas, J., . . . Momma, S. (2015). Extracellular vesicle-mediated transfer of functional RNA in the tumor microenvironment. *Oncoimmunology, 4*(6), e1008371. doi:10.1080/2162402X.2015.1008371

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package

for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rodier, F., Coppe, J. P., Patil, C. K., Hoeijmakers, W. A., Munoz, D. P., Raza, S. R., . . . Campisi, J. (2009). Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat Cell Biol, 11*(8), 973-979. doi:10.1038/ncb1909

Rounge, T. B., Umu, S. U., Keller, A., Meese, E., Ursin, G., Tretli, S., . . . Langseth, H. (2018). Circulating small non-coding RNAs associated with age, sex, smoking, body mass and physical activity. *Sci Rep, 8*(1), 17650. doi:10.1038/s41598-018-35974-4

Ru, Y., Kechris, K. J., Tabakoff, B., Hoffman, P., Radcliffe, R. A., Bowler, R., . . . Theodorescu, D. (2014). The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res, 42*(17), e133. doi:10.1093/nar/gku631

Shokhirev, M. N., & Johnson, A. A. (2021). Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell, 20*(1), e13280. doi:10.1111/acel.13280

Skog, J., Wurdinger, T., van Rijn, S., Meijer, D. H., Gainche, L., Sena-Esteves, M., . . . Breakefield, X. O. (2008). Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol, 10*(12), 1470-1476. doi:10.1038/ncb1800

Storz, P. (2011). Forkhead homeobox type O transcription factors in the responses to oxidative stress. *Antioxid Redox Signal, 14*(4), 593-605. doi:10.1089/ars.2010.3405

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., . . . de Magalhaes, J. P. (2018). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res, 46*(D1), D1083-D1090. doi:10.1093/nar/gkx1042

van Niel, G., D'Angelo, G., & Raposo, G. (2018). Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol, 19*(4), 213-228. doi:10.1038/nrm.2017.125

Wang, C. W. (2006). New ensemble machine learning method for classification and prediction on gene expression data. *Conf Proc IEEE Eng Med Biol Soc, 2006*, 3478-3481. doi:10.1109/IEMBS.2006.259893

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., . . . Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Curr Protoc, 1*(3), e90. doi:10.1002/cpz1.90

Xue, W., Zender, L., Miething, C., Dickins, R. A., Hernando, E., Krizhanovsky, V., . . . Lowe, S. W. (2007). Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature, 445*(7128), 656-660. doi:10.1038/nature05529

Zhang, H., Yang, H., Zhang, C., Jing, Y., Wang, C., Liu, C., . . . Li, D. (2015). Investigation of microRNA expression in human serum during the aging process. *J Gerontol A Biol Sci Med Sci, 70*(1), 102-109. doi:10.1093/gerona/glu145

Zhang, S., Li, H., Zheng, L., Li, H., Feng, C., & Zhang, W. (2019). Identification of functional tRNA-derived fragments in senescence-accelerated mouse prone 8 brain. *Aging (Albany NY), 11*(22), 10485-10498. doi:10.18632/aging.102471

Zhang, Y., Cai, F., Liu, J., Chang, H., Liu, L., Yang, A., & Liu, X. (2018). Transfer RNA-derived fragments as potential exosome tRNA-derived fragment biomarkers for osteoporosis. *Int J Rheum Dis, 21*(9), 1659-1669. doi:10.1111/1756-185X.13346

Zhang, Y., Jia, S., Huang, H., Qiu, J., & Zhou, C. (2014). A novel algorithm for the precise calculation of the maximal information coefficient. *Sci Rep, 4*, 6662. doi:10.1038/srep06662

Zhang, Y., Kim, M. S., Jia, B., Yan, J., Zuniga-Hertz, J. P., Han, C., & Cai, D. (2017). Hypothalamic stem cells control ageing speed partly through exosomal miRNAs. *Nature, 548*(7665), 52-57. doi:10.1038/nature23282

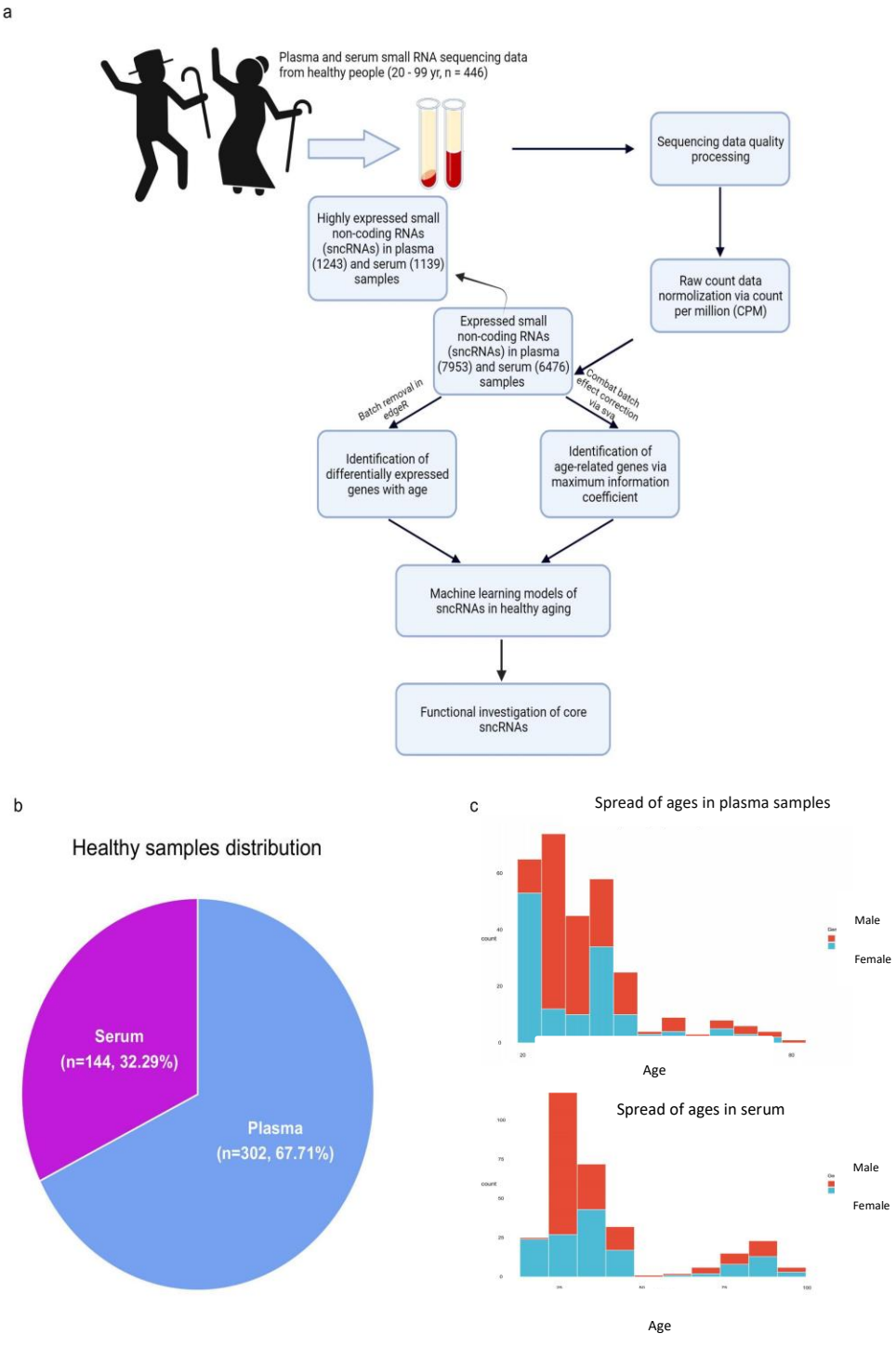


Figure 1. Identifying practical computational models of healthy aging via plasma and serum small non-coding RNAs (sncRNAs). (a) Flow chart of data preprocessing, normalizing, batch effect correcting and analyses of 446 blood samples. (b) Biofluids distribution from

healthy donors. (c) Summary form displaying sample trait information. Figure 1a was created with BioRender.com

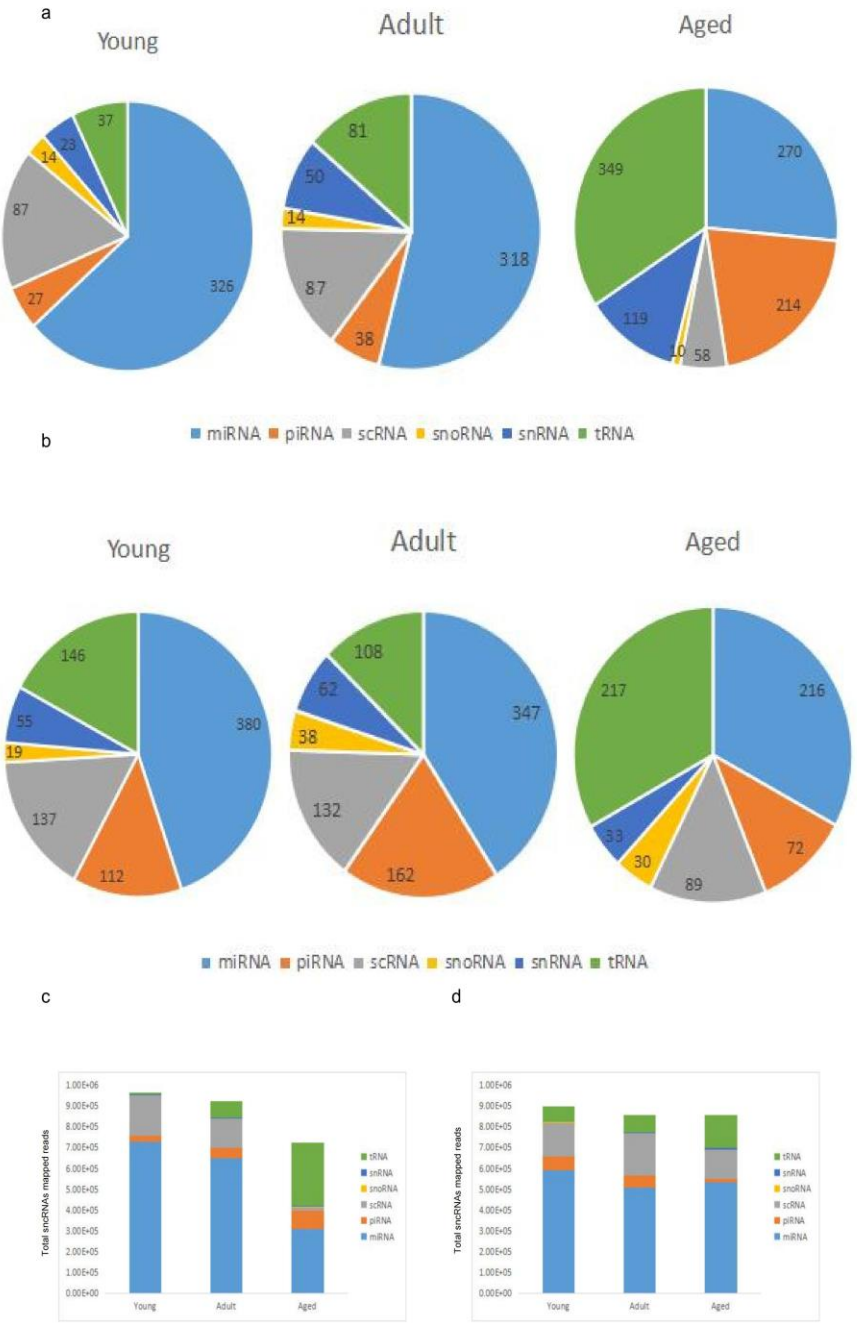
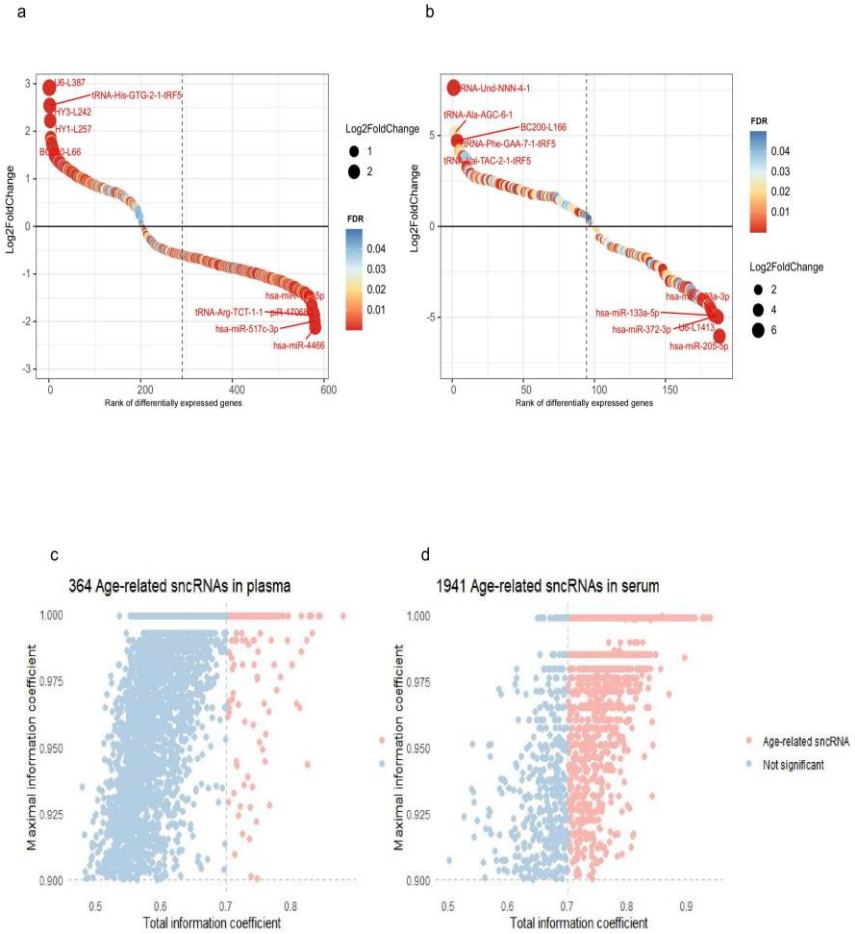


Figure 2. Highly expressed sncRNAs in plasma and serum. Subtype distribution of highly expressed sncRNAs, which meet the expression cutoff (≥ 10 CPM in ≥ 30% of samples) among young (20-30 years), adult (31-60 years) and aged individuals (≥ 61 years) in plasma

(a) and serum (b). Total sequencing reads of highly expressed sncRNAs among three age groups in plasma (c) and serum (d).


a
b
U6-L387
tRNA-His-GTG-2-1-tRF5
HY3-L242
HY1-L257
BC200-L66
hsa-miR-517c-3p
hsa-miR-4466
Log2FoldChange
FDR
Rank of differentially expressed genes
tRNA-Und-NNN-4-1
tRNA-Ala-AGC-6-1
BC200-L166
tRNA-Phe-GAA-7-1-tRF5
hsa-miR-133a-5p
hsa-miR-372-3p
U6-L1413
hsa-miR-205-5p

c
d
364 Age-related sncRNAs in plasma
1941 Age-related sncRNAs in serum
Maximal information coefficient
Total information coefficient
Age-related sncRNA
Not significant

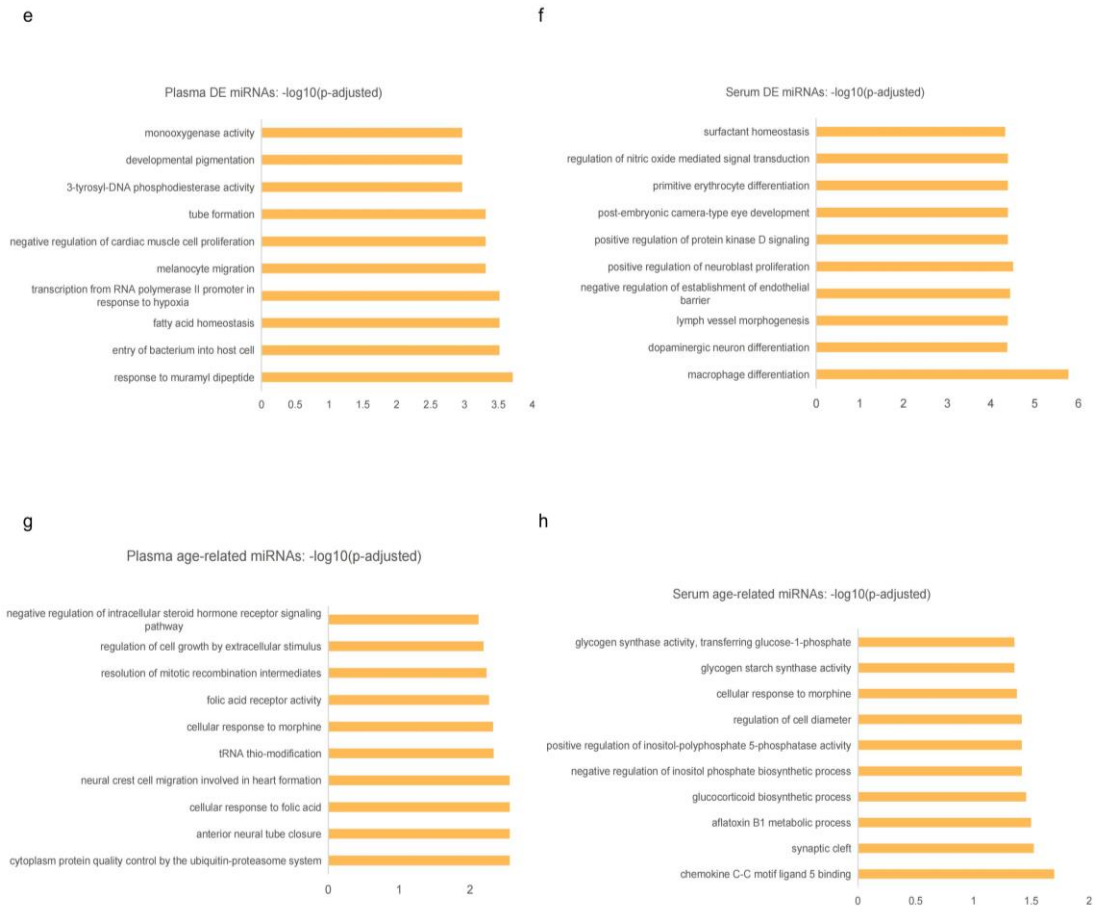


Figure 3. Identification of differentially expressed (DE) and age-related sncRNAs. DE sncRNAs (FDR < 0.05) sorted by FDR and fold change in plasma (a) and serum (b). MIC-based age-related sncRNAs in plasma (c) and serum (d), identified by both MIC and total information coefficient (TIC) values ≥ 0.7. Over-representation analysis of biological process of DE miRNAs/ MIC-based age-associated miRNAs targets in plasma (e and g) and serum (f and h) (P-adjusted value < 0.05).

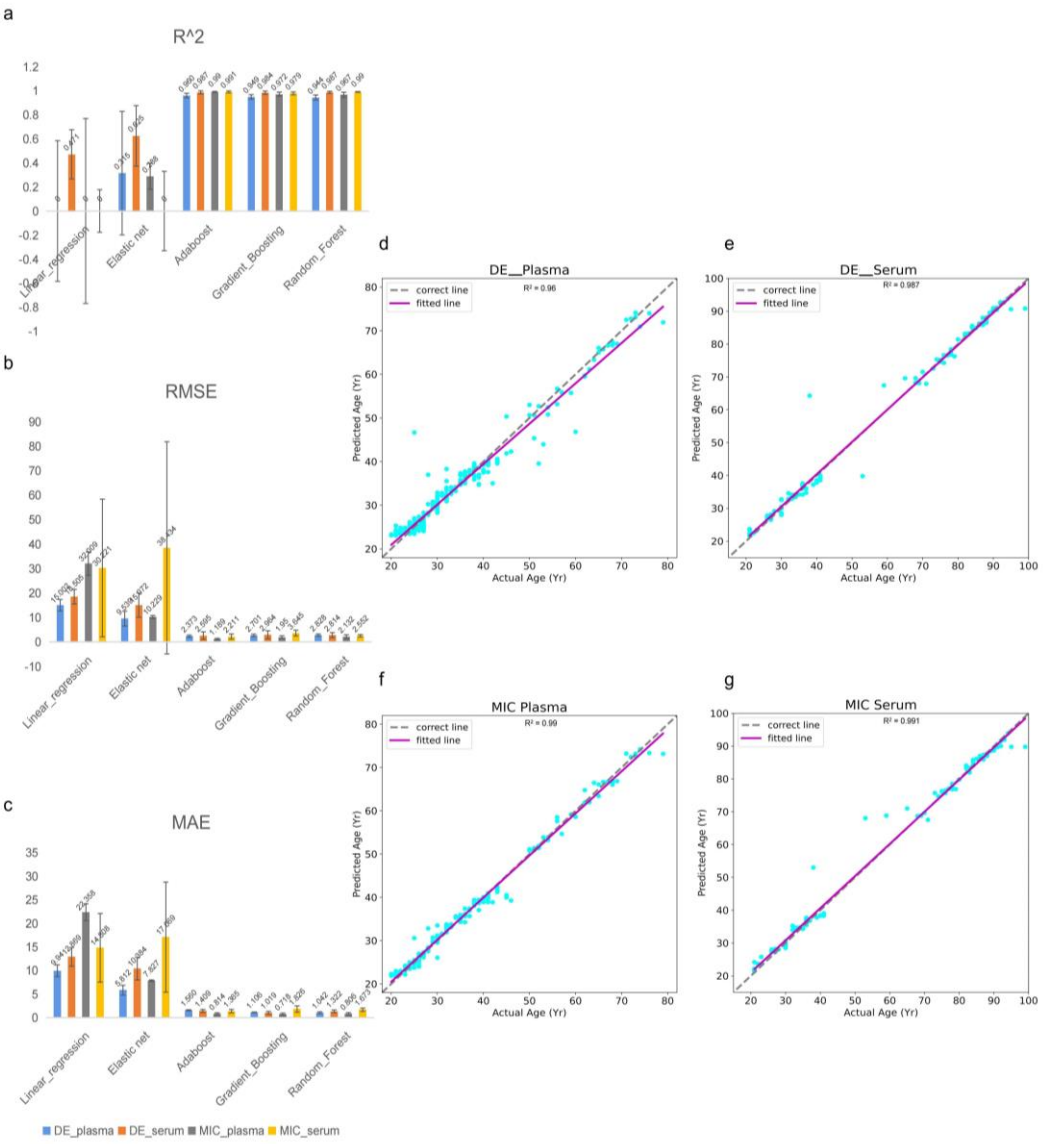


Figure 4. Performance evaluation of sncRNAs based aging clocks built by linear regression, elastic net, Adaptive Boosting, Gradient Boosting and Random Forest approaches. Summary of R^2 value (a), root mean squared error (RMSE) (b), and mean absolute error (MAE) (c). (d) Model fit based on plasma DE sncRNAs. (e) Model fit based on serum DE sncRNAs. (f) Model fit based on plasma MIC-based associated sncRNAs. (g) Model fit based on serum MIC-based associated sncRNAs. All models were constructed using Adaptive Boosting method.

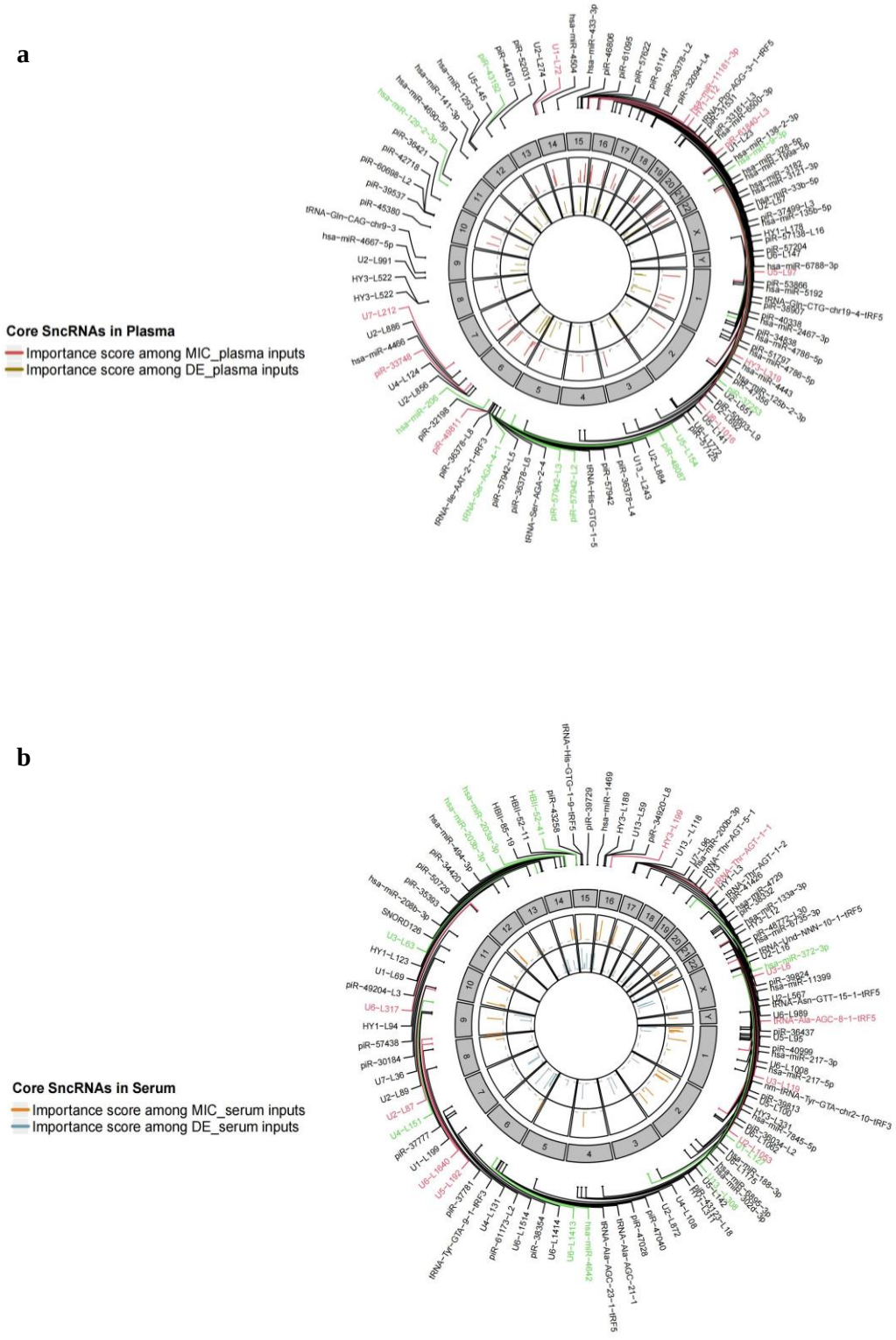


Figure 5. Circos plots representing top 100 core sncRNAs of human aging in plasma (a) and serum (b). Color bars demonstrate the integrated importance score of each sncRNA from MIC/DE based inputs. Top 10 core sncRNAs from MIC and DE based inputs are highlighted in red and green respectively.

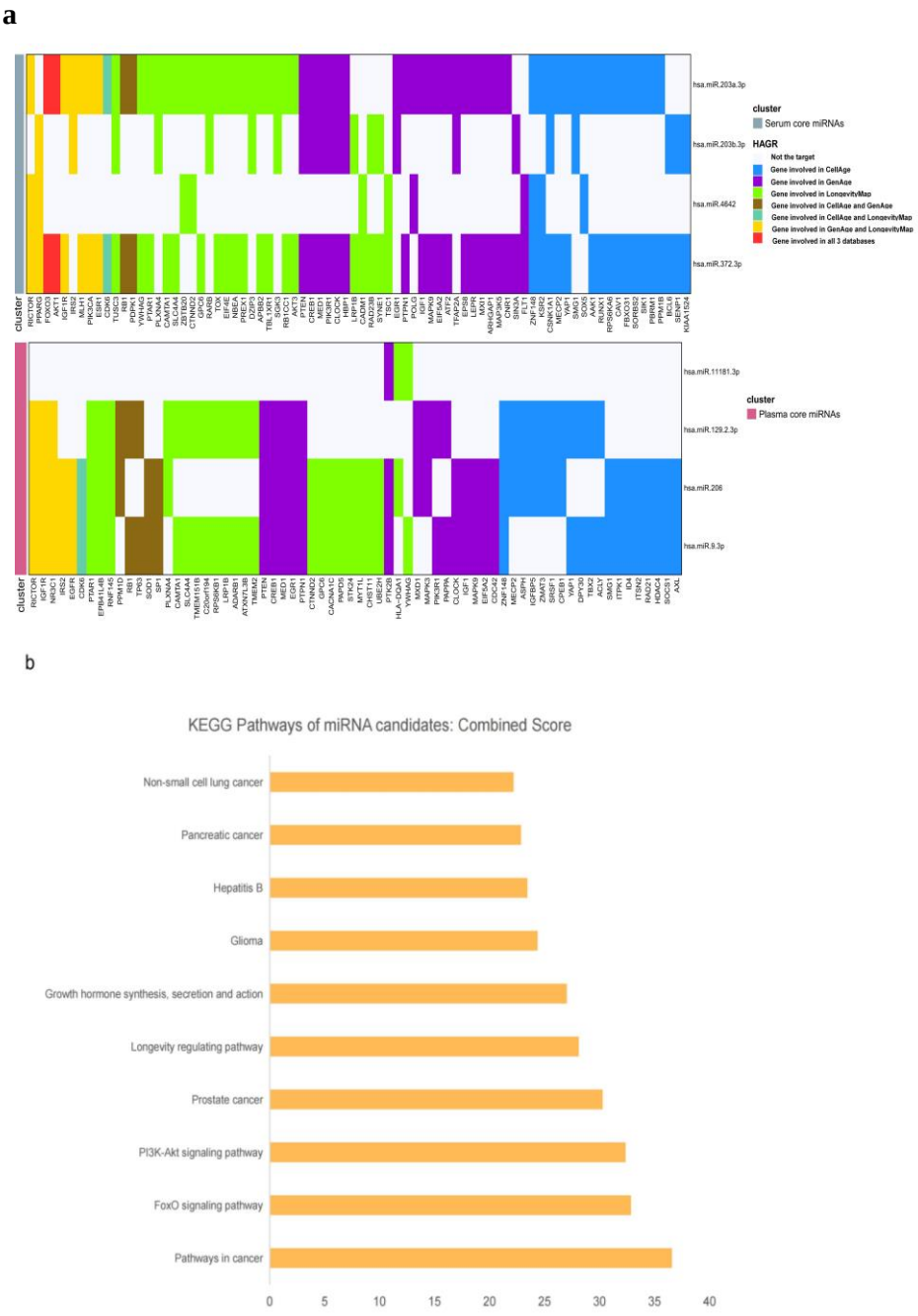


Figure 6. Top core miRNAs are associated with human aging and aging-related disease. (a) Heatmap of top core miRNAs targets involved in Human Ageing Genomic Resources. Targets that have strong sequence affinity with at least 2 core miRNAs were shown. Core miRNAs were separated into 2 clusters (core miRNAs from plasma/serum trained model), and grids represented genes were not targets (white), targets involved in CellAge (blue), targets involved in GenAge (purple), targets involved in LongevityMap (green), targets involved in both CellAge and GenAge (brown), targets involved in both CellAge and LongevityMap (cyan), targets involved in both GenAge and LongevityMap (yellow) and targets in all three databases (red) of corresponding core miRNAs. (b) Over-representation analysis of KEGG pathways of miRNA targets included in Human Ageing Genomic Resources.

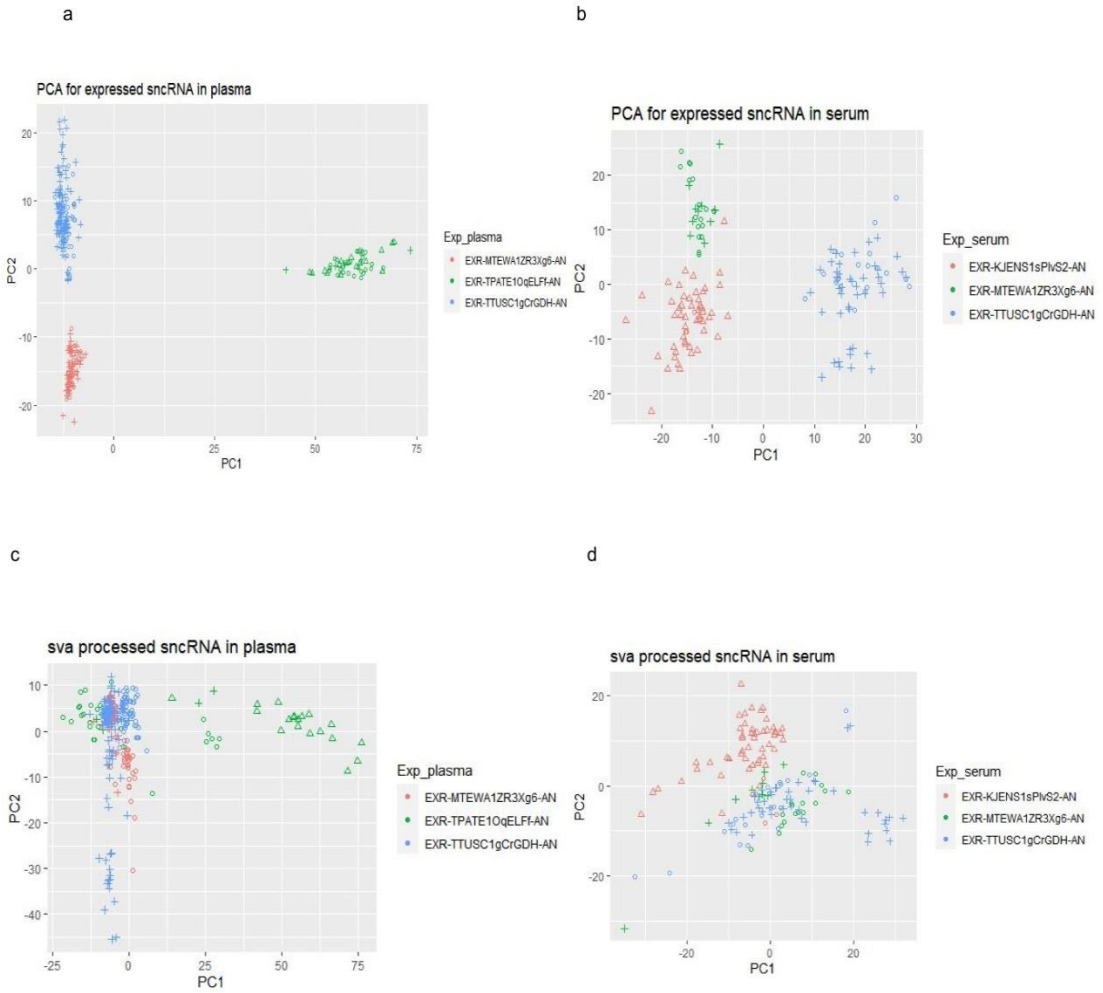


Figure S1. Principal component analysis of plasma and serum samples before (a and b) and after sva batch correction (c and d).

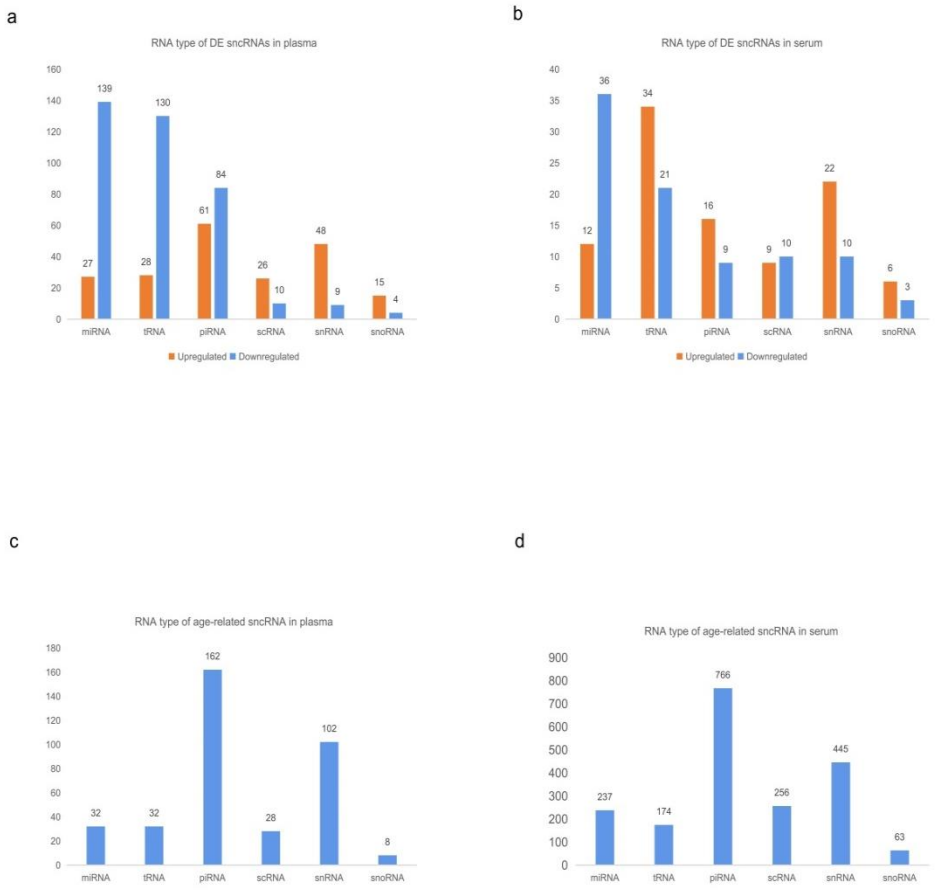


Figure S2. RNA subtype of age-related sncRNAs based on differential expression (a and b) and maximum information coefficient (c and d).

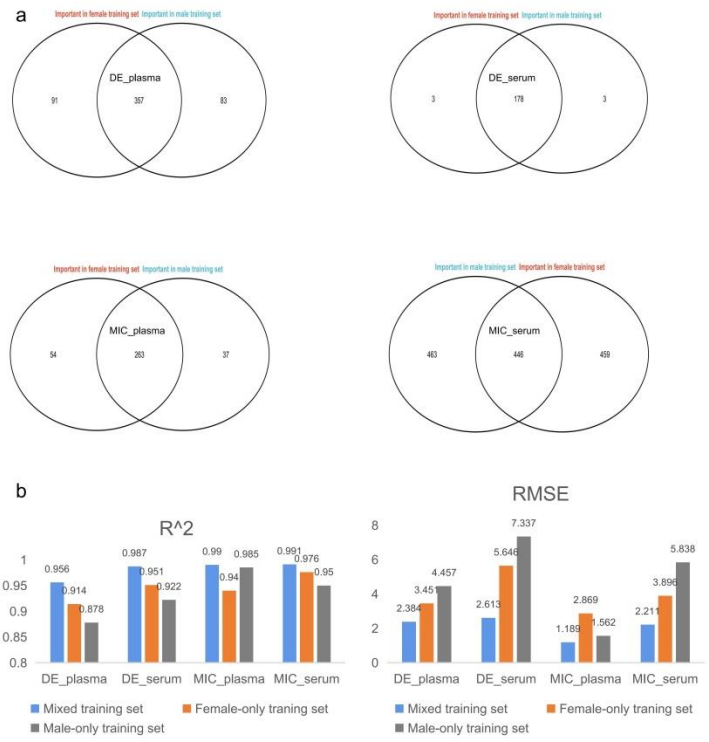


Figure S3. Difference of gender-specific training sets in model performance. Overlap of core sncRNAs between male-/female-only training sets with non-zero importance values (a) and model performance with another gender data as test sets (b)

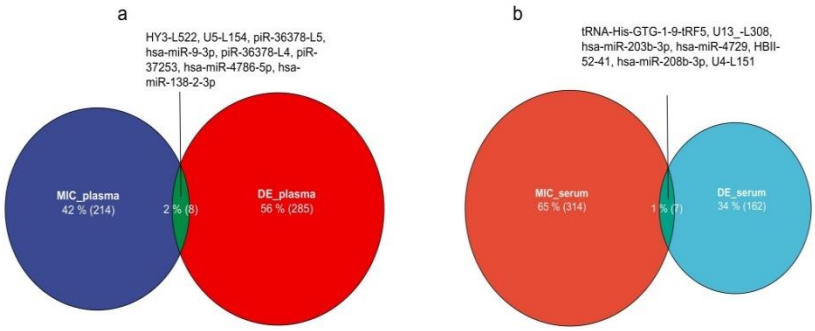
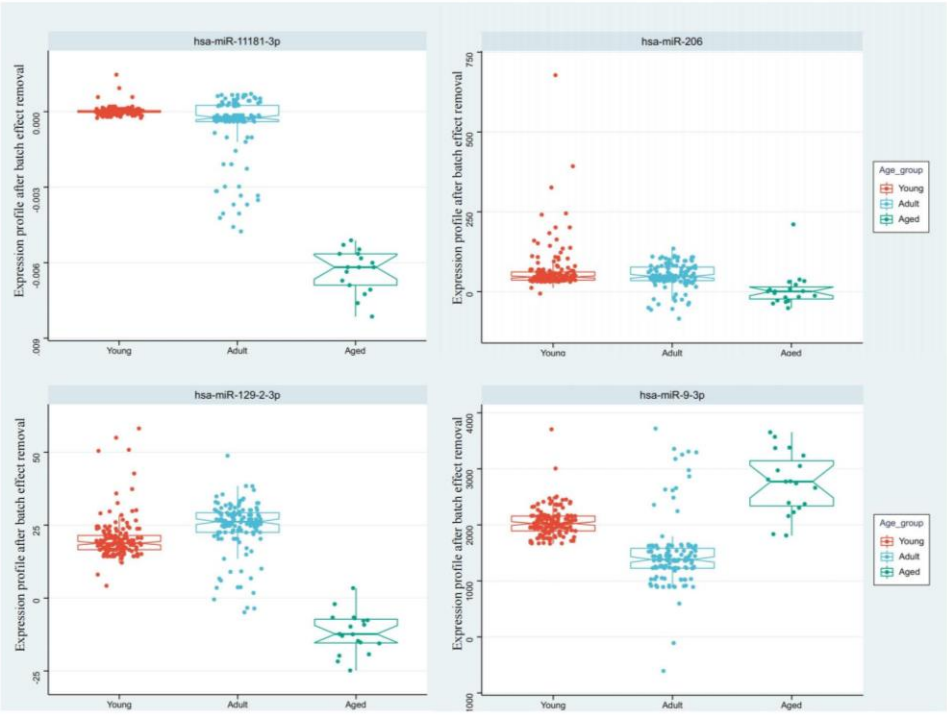


Figure S4. Overlap of core sncRNAs from DE and MIC inputs in plasma (a) and serum (b) samples.

a



b

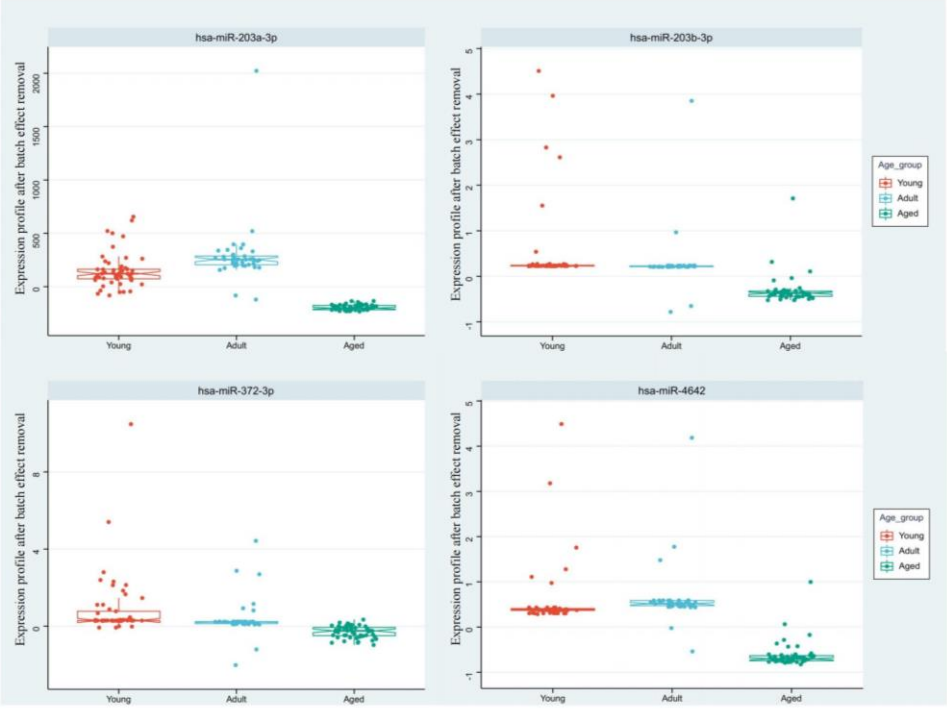


Figure S5. Expression profile of top core miRNAs in plasma (a) and serum (b).

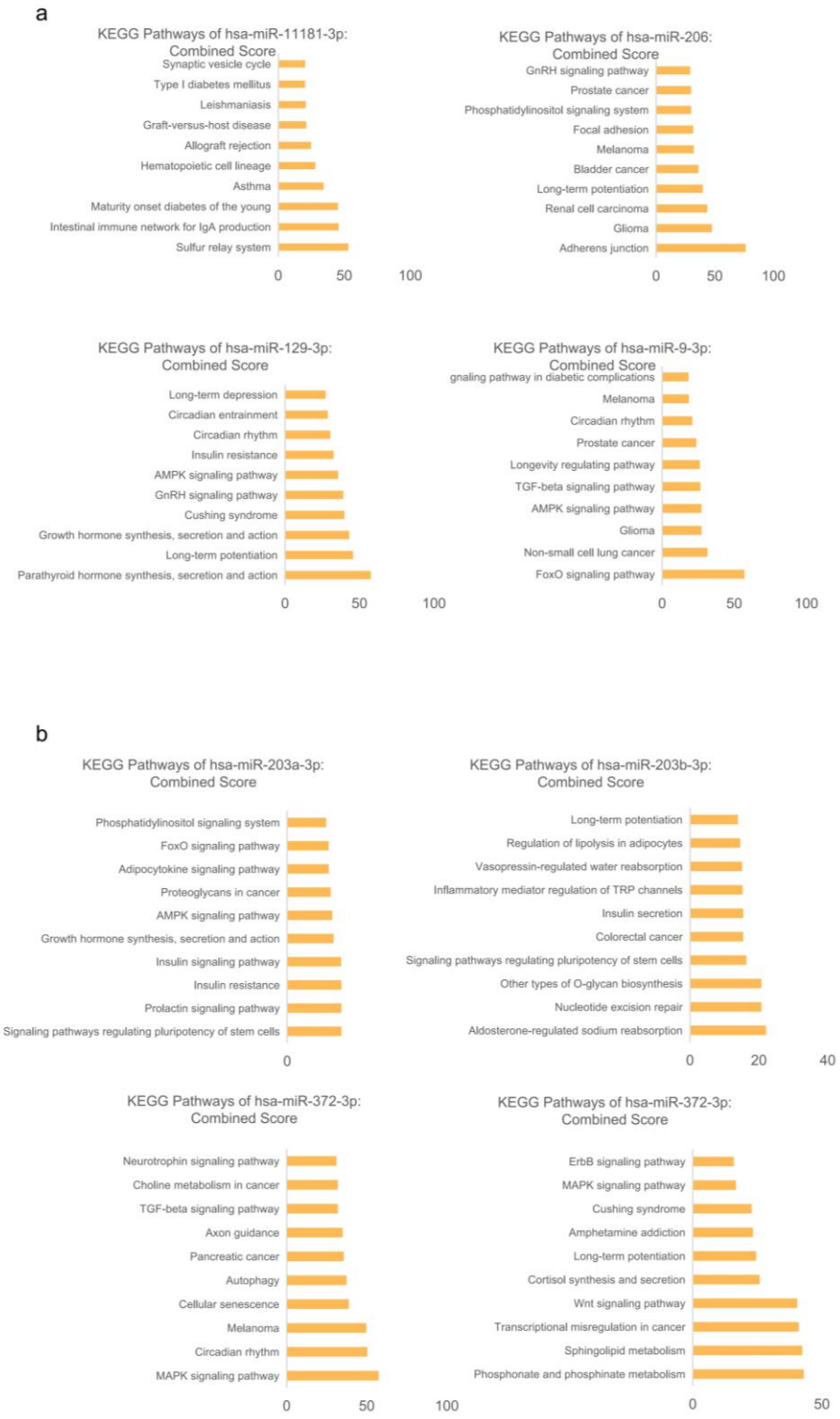


Figure S6. KEGG pathway enrichment of top core miRNA targets in plasma (a) and serum (b).

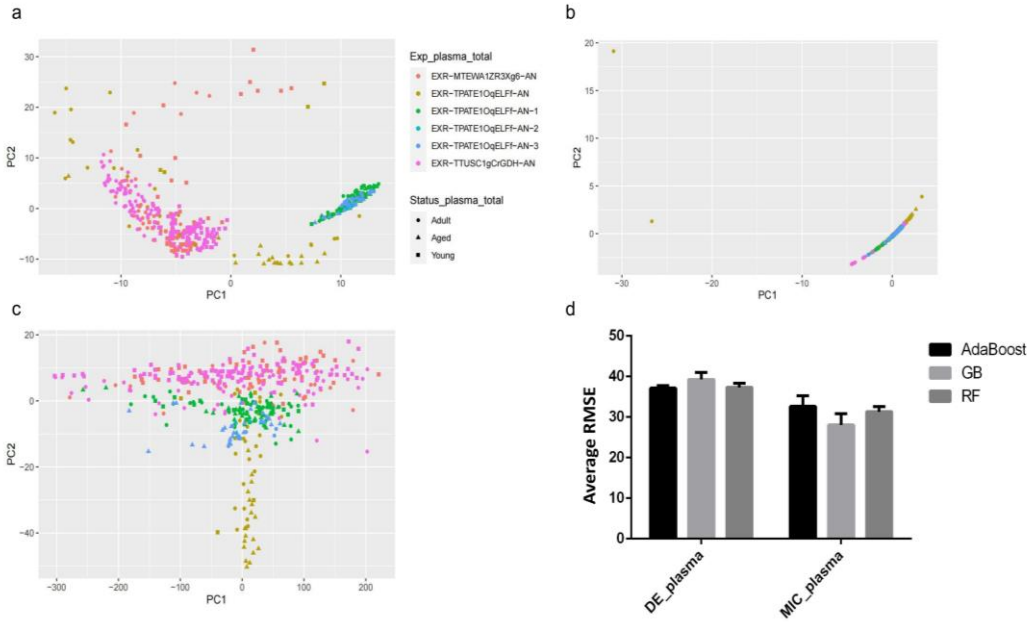


Figure S7. Profile of plasma age-related sncRNAs in unhealthy samples. Principal component analysis of all plasma samples (healthy: EXR-MTEWA1ZR3Xg6-AN, EXR-TPATE1OqELFf-AN and EXR-TTUSC1gCrGDH-AN; colon carcinoma: EXR-TPATE1OqELFf-AN-1; pancreatic carcinoma: EXR-TPATE1OqELFf-AN-2; prostate carcinoma: EXR-TPATE1OqELFf-AN-3) by DE sncRNAs (a), MIC sncRNAs (b), and expressed sncRNA (c). (d) Performance evaluation of trained model by using unhealthy samples as validation set. RMSE: root mean squared error.

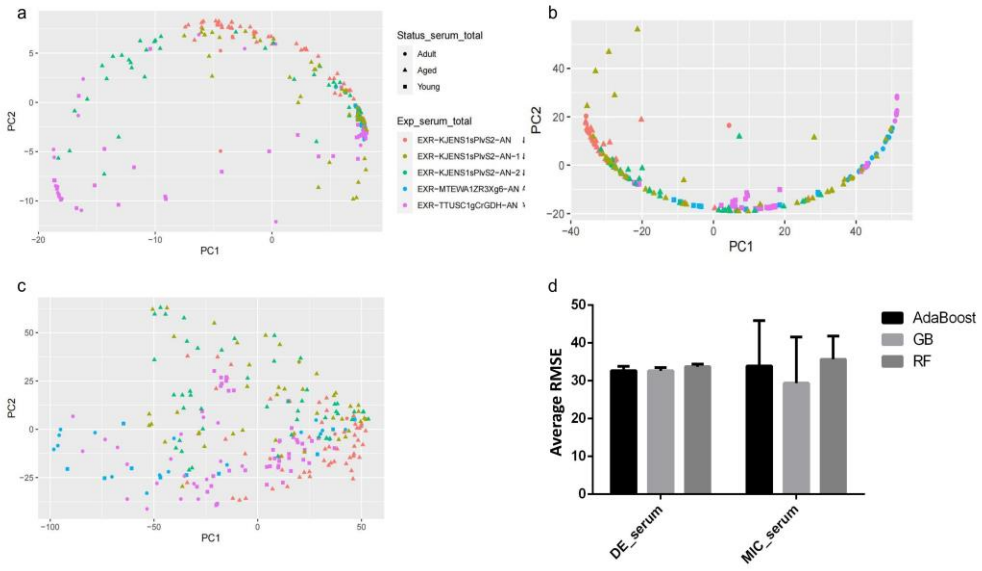


Figure S8. Profile of serum age-related sncRNAs in unhealthy samples. Principal component analysis of all serum samples (healthy: EXR-KJENS1sPlvS2-AN, EXR-MTEWA1ZR3Xg6-AN and EXR-TTUSC1gCrGDH-AN; Alzheimer's disease: EXR-KJENS1sPlvS2-AN-1; Parkinson's disease: EXR-KJENS1sPlvS2-AN-2) by DE sncRNAs (a), MIC sncRNAs (b), and expressed sncRNA (c). (d) Performance evaluation (average RMSE) of trained model by using unhealthy samples as validation set. RMSE: root mean squared error.

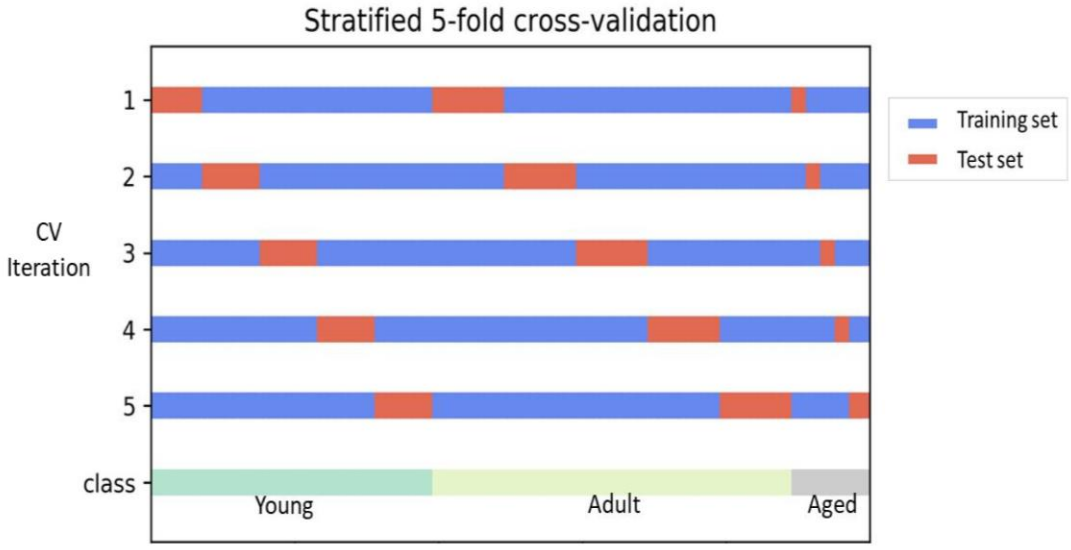


Figure S9. Stratified 5-fold cross-validation for the preprocessed data, which divides data into 3 groups according to the age range and preserves approximately relative same class frequencies in each train and validation fold.

TABLES

Table 1. Top core sncRNAs associated with age in plasma.

Model input	Gene name	RNA type	Adaboost	GB	RF	Sum of rank
MIC_plasma	piR-33748	piRNA	3	1	1	5
MIC_plasma	U5-L97	snRNA	1	2	2	5
MIC_plasma	HY3-L319	scRNA	2	6	3	11
MIC_plasma	U6-L1016	snRNA	7	7	4	18
MIC_plasma	hsa-miR-11181-3p	miRNA	6	13	14	33
MIC_plasma	HY1-L12	scRNA	22	4	10	36
MIC_plasma	piR-61840-L3	piRNA	16	3	18	37
MIC_plasma	U7-L212	snRNA	18	17	11	46
MIC_plasma	piR-49811	piRNA	20	8	23	51
MIC_plasma	U1-L72	snRNA	36	5	26	67
DE_plasma	hsa-miR-129-2-3p	miRNA	1	1	1	3
DE_plasma	hsa-miR-206	miRNA	2	2	2	6
DE_plasma	hsa-miR-9-3p	miRNA	7	3	7	17
DE_plasma	tRNA-Ser-AGA-4-1	tRNA	10	6	3	19
DE_plasma	piR-48087	piRNA	3	14	8	25
DE_plasma	piR-37253	piRNA	8	7	11	26
DE_plasma	piR-57942-L2	piRNA	14	11	9	34
DE_plasma	piR-43192	piRNA	4	18	13	35
DE_plasma	piR-57942-L3	piRNA	15	8	15	38
DE_plasma	U5-L154	snRNA	9	17	12	38

Importance ranking from three ensemble learning methods and corresponding sum of rank. Adaboost, Adaptive Boosting; GB, Gradient Boosting; RF, Random Forest.

Table 2. Top core sncRNAs associated with age in Serum.

Model input	Gene name	RNA type	Adaboost	GB	RF	Sum of rank
MIC_serum	U5-L192	snRNA	2	9	6	17
MIC_serum	U6-L317	snRNA	27	4	4	35
MIC_serum	HY3-L199	scRNA	4	5	36	45
MIC_serum	U2-L87	snRNA	34	13	28	75
MIC_serum	U3-L6	snRNA	16	56	24	96
MIC_serum	tRNA-Thr-AGT-1-1	tRNA	96	18	2	116
MIC_serum	tRNA-Ala-AGC-8-1-tRF5	tRNA	37	44	71	152
MIC_serum	U2-L1053	snRNA	135	3	15	153
MIC_serum	U3-L119	snRNA	109	41	3	153
MIC_serum	U6-L1640	snRNA	136	10	13	159
DE_serum	U6-L1413	snRNA	1	1	1	3
DE_serum	hsa-miR-203a-3p	miRNA	2	2	4	8
DE_serum	HBII-52-41	snoRNA	6	3	7	16
DE_serum	hsa-miR-203b-3p	miRNA	9	5	13	27
DE_serum	hsa-miR-4642	miRNA	4	25	3	32
DE_serum	hsa-miR-372-3p	miRNA	22	6	24	52
DE_serum	U4-L151	snRNA	12	17	29	58
DE_serum	U13-L308	snRNA	42	9	8	59
DE_serum	U3-L63	snRNA	20	16	23	59
DE_serum	U1-L127	snRNA	26	8	31	65

Importance ranking from three ensemble learning methods and corresponding sum of rank. Adaboost, Adaptive Boosting; GB, Gradient Boosting; RF, Random Forest.

Table S1. Summary of highly expressed sncRNAs in plasma and serum.

Source	RNA Type	Number
Plasma	miRNA	414
	tRNA	353
	piRNA	227
	snRNA	121
	scRNA	111
	snoRNA	17
Serum	miRNA	397
	tRNA	270
	piRNA	189
	snRNA	71
	scRNA	166
	snoRNA	46

Table S2. Model performance of each subset in 5-fold cross-validation.

			Subset1	Subset2	Subset3	Subset4	Subset5	Standard Deviation
DE_Plasma	Adaboost	R^2	0.972856161	0.958557782	0.955292624	0.979727817	0.931393977	0.018681999
		RMSE	2.052642792	2.473452512	2.454547464	1.876055217	3.008545341	0.438816996
		MAE	1.552693598	1.550845074	1.706472075	1.377019049	1.614618294	0.120473968
	GB	R^2	0.922643928	0.946930173	0.948839558	0.975180741	0.951220362	0.018636703
		RMSE	3.465173204	2.79901944	2.625723492	2.075820353	2.53684962	0.504239234
		MAE	1.077839731	1.193875168	1.016180732	1.166259813	1.075014154	0.072756059
	RF	R^2	0.952388646	0.929388723	0.943467031	0.97267156	0.919782359	0.02056475
		RMSE	2.718522984	3.228636214	2.760150358	2.178225195	3.253197145	0.441615714
		MAE	0.906557377	1.332131148	0.854333333	1.016166667	1.099833333	0.188249455

			Subset1	Subset2	Subset3	Subset4	Subset5	Standard Deviation
DE_Serum	Adaboost	R^2	0.982981969	0.9929736	0.997418573	0.96599737	0.997443135	0.013321844
		RMSE	3.139183334	2.137740287	1.366647489	5.041967245	1.289631489	1.557699223
		MAE	1.656835549	1.398204805	1.096324305	1.800079094	1.093127611	0.320945195
	GB	R^2	0.988079615	0.992856837	0.963691636	0.978831124	0.998666293	0.013678728
		RMSE	2.627284846	2.155429332	5.125425475	3.97825547	0.931412287	1.627809088
		MAE	1.129963618	0.974648823	1.26673294	1.209130942	0.513437363	0.303065455
	RF	R^2	0.985269037	0.989119605	0.994332098	0.972517888	0.994262637	0.008995496
		RMSE	2.920637035	2.660178889	2.025059386	4.532823585	1.931825561	1.047621817
		MAE	1.462413793	1.352413793	0.956896552	1.733103448	1.106428571	0.303898968

Table S2. Model performance of each subset in 5-fold cross-validation. (Continued)

			Subset1	Subset2	Subset3	Subset4	Subset5	Standard Deviation
		R^2	0.991004657	0.97320321	0.990051412	0.987287646	0.982854508	0.007254339
	Adaboost	RMSE	1.181643857	1.988948181	1.157877379	1.485622074	1.504008918	0.33581027
		MAE	0.909150037	1.196838306	0.840967126	1.09969043	0.944299043	0.146043546
		R^2	0.988150352	0.948749192	0.985735142	0.977904304	0.959976364	0.01709871
MIC_Plasma	GB	RMSE	1.356220981	2.75063167	1.386486269	1.958615569	2.297913661	0.598383334
		MAE	0.52631947	1.03521638	0.429025519	0.799240191	0.80097089	0.242052642
		R^2	0.987451067	0.948768897	0.983748381	0.939752798	0.97537761	0.02146928
	RF	RMSE	1.395664834	2.75010283	1.479892451	3.234177124	1.802355866	0.817456963
		MAE	0.576229508	1.040655738	0.5685	1.108833333	0.735	0.255447132

			Subset1	Subset2	Subset3	Subset4	Subset5	Standard Deviation
		R^2	0.978529328	0.991866732	0.99806482	0.985615433	0.996631814	0.008111834
	Adaboost	RMSE	3.526020564	2.299964129	1.183279329	3.279383175	1.480163669	1.044875906
		MAE	1.765579745	1.521807017	0.951935531	1.791460239	1.198641358	0.364865258
		R^2	0.990921045	0.984696695	0.981735514	0.95951974	0.979631193	0.011847712
MIC_Plasma	GB	RMSE	2.292872816	3.154867588	3.635216719	5.501301053	3.639942522	1.174143567
		MAE	0.986595538	1.757276313	1.870085261	2.841109461	1.676481698	0.664091253
		R^2	0.994532387	0.99238682	0.990311691	0.986181343	0.985502811	0.00390218
	RF	RMSE	1.779348001	2.225212892	2.647589014	3.214228067	3.070814806	0.594632984
		MAE	1.114827586	1.631724138	1.987931034	2.074482759	1.624642857	0.379152705

Table S3. Summary of predicted targets of 8 top core miRNAs.

Source	miRNA_ID	Database (Number of targets)	Source	miRNA_ID	Database (Number of targets)
Plasma	hsa-miR-9-3p	diana_microt (345)	Serum	hsa-miR-206	diana_microt (303)
		elmmo (1306)			elmmo (500)
		microcosm (28)			microcosm (65)
		miranda (218)			miranda (125)
		mirdb (72)			mirdb (79)
		pictar (3)			pictar (6)
		targetscan (1)			targetscan (47)
		pita (0)			pita (188)
	hsa-miR-129-2-3p	diana_microt (131)		hsa-miR-372-3p	diana_microt (334)
		elmmo (332)			elmmo (890)
		microcosm (15)			microcosm (87)
		miranda (40)			miranda (148)
		mirdb (24)			mirdb (67)

	pictar (9)
	targetscan (32)
	pita (22)
	diana_microt (138)
	elmmo (981)
	microcosm (55)
hsa-miR-203a-3p	miranda (237)
	mirdb (73)
	pictar (45)
	targetscan (0)
	pita (76)
	diana_microt (250)
hsa-miR-203b-3p	elmmo (0)
	microcosm (0)
	miranda (0)

	pictar (0)
	targetscan (33)
	pita (147)
	diana_microt (49)
	elmmo (0)
	microcosm (0)
hsa-miR-4642	miranda (0)
	mirdb (20)
	pictar (0)
	targetscan (0)
	pita (0)
	diana_microt (0)
hsa-miR-11181-3p	elmmo (0)
	microcosm (0)
	miranda (0)

	mirdb (19)	mirdb (42)
	pictar (0)	pictar (0)
	targetscan (0)	targetscan (0)
	pita (0)	pita (0)

Table S4. Parameter setting for ensemble learners.

Dataset	Parameter	Regressor	Value
DE_plasma	Num of Estimator	AdaBoost	50
		GB	200
		RF	100
DE_serum	Num of Estimator	AdaBoost	50
		GB	100
		RF	100
MIC_plasma	Num of Estimator	AdaBoost	200
		GB	100
		RF	100
MIC_serum	Num of Estimator	AdaBoost	200
		GB	200
		RF	200

VITA

Ping Xiao

Candidate for the Degree of

Doctor of Philosophy

Dissertation: IDENTIFICATION OF TRANSCRIPTOMIC SIGNATURE IN CELLULAR SENESCENCE AND CHARACTERIZATION OF CIRCULATING SMALL NON-CODING RNA DURING HUMAN AGING

Major Field: Animal Science

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Animal Science at Oklahoma State University, Stillwater, Oklahoma in December, 2022.

Completed the requirements for the Master of Science in Animal Genetics, Breeding and Reproduction at Sichuan Agricultural University, Chengdu, China in 2019.

Completed the requirements for the Bachelor of Science in Animal Science at Sichuan Agricultural University, Chengdu, China in 2017.

Professional Memberships: American Society of Animal Science, International Society for Animal Genetics.